

SURVEYS IN COVID-19 AND POST COVID-19 DAYS

TUĞBA ADALI
YASER KOYUNCU
A.SİNAN TÜRKYILMAZ



Surveys in Covid-19 and Post Covid-19 Days



This study is carried out as part of the collaboration between
Hacettepe University Institute of Population Studies and the United Nations Population Fund Turkey Country Office

Ankara, 2021

Publication No:

ISBN:

Assist. Prof. Dr. Tuğba Adalı

Yaser Koyuncu

Prof. Dr. A. Sinan Türkyılmaz

Acknowledgments

We would like to thank Attila Hanciođlu, Enver Tařtı, Frauke Kreuter, James M. Lepkowski, Ođuzhan Akyıldırım, and Raphael Nishimura (in alphabetical order) for taking part in our interviews and sharing their valuable experience and insights. We would like to acknowledge Raphael Nishimura for proofreading the Conclusion section and Akya Akarsu for language editing of the entire document. We finally would like to thank Gökhan Yıldırımkaya from UNFPA Turkey Office and our Director A. Banu Ergöçmen for their support in realizing this project.

TABLE OF CONTENTS

<i>List of Figures</i>	<i>i</i>
<i>List of Tables</i>	<i>i</i>
<i>List of Abbreviations</i>	<i>ii</i>
1. Introduction	1
2. Background on Surveys	3
2.1 Sample Surveys	3
2.2 A Brief History of Surveys	7
2.3 Types of Surveys	14
2.4 Survey Modes	18
2.4.1 Face-to-face Surveys	18
2.4.2 Telephone Surveys	20
2.4.3 Mail Surveys	21
2.4.4 Online Surveys	22
2.4.5 Mixed-Mode Surveys	22
2.5 International Survey Programs	23
2.6 Surveys in Turkey	24
2.6.1 Demographic Survey Series and Other Recent Surveys of HUIPS	25
2.6.2 Recent Surveys of the Turkish Statistical Institute (TurkStat)	27
2.6.3 Examples of Recent Surveys of the Research Companies in Turkey	30
3. Surveys at the pandemic	32
3.1 Adjustments that had to be done to ongoing surveys during COVID-19	32
3.1.1 Cross-sectional surveys	36
3.1.2 Longitudinal surveys	41
3.2 Surveys to investigate the impacts of the COVID-19 pandemic	50
3.2.1 Use of ongoing face-to-face panel studies to carry out COVID-19 related surveys	54
3.2.2 Use of ongoing online panels to carry out COVID-19 related surveys	56
3.3 Innovations in survey design and exploitation of new data sources for pandemic related research	59
3.3.1 Use of social media data	59
3.3.2 Use of Google search data and GPS mobility data	61
3.4 Surveys to estimate the prevalence of COVID-19	62
3.5 Surveys in Turkey and the COVID-19 Pandemic	66
3.6 Ethical issues	71
4. Surveys in the Post-Covid Era: What Do Experts Say?	73
4.1 Interview with James M. Lepkowski (A General Frame of the Problems)	73
4.2 Interview with Raphael Nishimura (Adaptations from Face-to-face to Phone Surveys)	87

4.3 Interview with Oğuzhan Akyıldırım (Impact on Research Companies in Turkey)	98
4.4 Interview with Enver Taştı (Impact on Official National Surveys in Turkey)	106
4.5 Interview with Frauke Kreuter (Online Surveys and the Use of Big Data)	117
4.6 Interview with Attila Hancıoğlu (Impact on International Household Survey Series)	126
5. Conclusion	141
References	161

List of Figures

Figure 3.1 The number of countries who had surveys scheduled for the pandemic period, by type of survey	33
Figure 3. 2 Responses to the question: “If fieldwork for one or more of the following surveys were planned or started, did you have to stop or postpone the fieldwork due to Covid-19 pandemic?”	34
Figure 3. 3 Main modes of labor force surveys by region	35
Figure 3. 4 Pew Research Center’s coverage of countries by two main survey modes, face-to-face, and telephone	39
Figure 3. 5 The Digital Meeting Space Experience, based on the observations of the Infotrak Company	40
Figure 3. 6 Countries responding to Eurostat survey regarding how their LFS Surveys were affected, by calendar year of 2020 (preliminary findings)	42
Figure 3. 7 Steps and processes recommended by WHO Regional Office for Europe for behavioral insight studies	51
Figure 3. 8 Screenshots from the app “Life Fits into Home”, Turkey	66

List of Tables

Table 3.. 1 List of DHS countries affected by the pandemic by their plans	38
---	----

List of Abbreviations

AAPOR	American Association for Public Opinion Research
ASR	Automatic Speech Recognition
BSA	British Social Attitudes
CAPI	Computer-assisted Personal Interviewing
CAI	Computer-assisted Interviewing
CAPAR	Computer-assisted Panel Research
CASI	Computer-assisted Self-Interviewing
CASIIP	Computer-assisted Self-Interviewing with Interviewer Present
CATI	Computer-assisted Telephone Interviewing
CAWI	Computer-assisted Web Interviewing
CNEF	Cross-National Equivalent File Project
CDS	Child Development Supplement
COVID-19	Coronavirus Disease 2019
CSAQ	Computer-assisted Self-Administered Questionnaires
DHS	Demographic and Health Survey
ECHP	European Community Household Panel
ELSA	English Longitudinal Study of Ageing
EMS	Electronic Mail Survey
EP	Institute for Employment Research Establishment Panel
ESS	European Social Survey
EU-SILC	European Union Statistics on Income and Living Conditions
EVS	European Values Study
EUROSTAT	European Statistical Office
FAO	Food and Agricultural Organization of the United Nations
FRéDA	Family Research and Demographic Analysis
FTF	Face-to-Face
GIP	German Internet Panel
GIZ	German Cooperation for International Cooperation GmbH
GGS	Generations and Gender Survey
GSS	General Social Survey
HILDA	Household Income and Labour Dynamics in Australia
HFCS	Household Finance and Consumption Survey
HOPP	High-Frequency Online Personal Panel
HRS	Health and Retirement Study
HUIPS	Hacettepe University Institute of Population Studies
IBGE	Brazilian Official Statistical Institute
ILO	International Labour Organization
ISSP	International Social Survey Programme
ISWGHS	The United Nations Intersecretariat Working Group on Household Surveys
ITA.LI	Italian Lives
JHPS	Japan Household Panel Study
JVS	Institute for Employment Research Job Vacancy Survey
KonsortSWD	Consortium for the Social, Behavioural, Educational and Economic Sciences
KLIPS	Korean Labor and Income Panel Study
LAPOP	Latin American Public Opinion Project

LFS	Labour Force Survey
LSMS	Living Standards Measurement Study
MCS	Mannheim Corona Study
MICS	Multiple Indicator Cluster Surveys
NHATS	National Health and Aging Trends Study
NLSY	National Longitudinal Survey of Youth
NORC	National Opinion Center at the University of Chicago
NIDI	Netherlands Interdisciplinary Demographic Institute
NPI	Non-pharmacological Interventions
NSO	National Statistical Offices
PAA	Population Association of America
PAPI	Paper-and-pen Interviewing
PASS	Labour Market and Social Security
PSID	Panel Study of Income Dynamics
RLMS-HSE	Russia Longitudinal Monitoring Study-Higher School of Economics
QLSP	Quality of Life and Social Participation
ReGES	Refugees in the German Educational System
RSS	Royal Statistics Society
SEAN	Societal Experts Action
SHARE	Survey of Health, Ageing and Retirement in Europe
SHP	Swiss Household Panel
SOEP	German Socio-Economic Panel
SOGI-SES	Socioeconomic Status, and Health across the Life Course
SRC	Survey Research Center of University of Michigan
SRO	Survey Research Operations at the University of Michigan
TAS	Transition into Adulthood Supplement
TIDI	Telephone In-Depth Interviews
TDE	Touchtone Data Entry
TSE	Total Survey Error
TDHS	Turkey Demographic and Health Survey
TUBİTAK	Scientific and Technological Research Council of Turkey
TURKSTAT	Turkish Statistical Institute
UAS	Understanding America Study
UKHLS	United Kingdom Household Longitudinal Study
UN	United Nations
UNDP	United Nations Development Programme
UNECE	United Nations Economic Commission for Europe
UNECLAC	United Nations Economic Commission for Latin America and the Caribbean
UNESCAP	United Nations Economic and Social Commission for Asia and the Pacific
UNESCO	United Nations Educational, Scientific and Cultural Organization
UNICEF	United Nations International Children's Emergency Fund
UNSD	United Nations Department of Economic and Social Affairs
VR	Voice Recognition
WAPOR	World Association for Public Opinion Research
WHO	World Health Organization
WVS	World Value Survey

1.Introduction

The year 2020 has been an extraordinary year, with the emergence of a global pandemic caused by the highly infectious SARS-CoV-2 virus. The World Health Organization declared the pandemic on March 11, 2020, and since then, the virus has spread everywhere, causing deaths, health system capacity overloads, job loss, the introduction of non-pharmaceutical interventions such as social distancing, lockdowns, closing of schools, distance working, and many other developments and changes that came into life rapidly.

These changes to the “normal” life we know introduced the necessity of data and research to answer all sorts of questions asked by researchers and policymakers. How much of the population was infected; what were the consequences of the pandemic on people in terms of their physical health, mental health, education, working status, financial situation, family arrangements, social interactions; how are people’s attitudes and behaviors regarding different dimensions of the pandemic, such as compliance with preventive measures, people’s concerns over the pandemic, their future evaluations, the evaluations of politicians and institutions in handling the situation, and so on.

Survey research is among the top sources for finding answers to the questions stated above and has also been affected by the pandemic. With in-person contact being the main route of transmission of COVID-19, visiting households, for instance, has become very risky for both respondents and interviewers, and such household surveys have been suspended almost everywhere. As the case in many office jobs, telephone interviewing from CATI centers have also become risky for disease spread. With service sector interruptions, postal services for mail surveys, and even online support services for any web surveys became affected. However, the need for data was real and urgent, so there have been new surveys designed, mode switches have been employed for surveys already running or planned, and usage of other data sources has been more on the forefront.

Survey researchers have been discussing these issues since early on in the pandemic. The European Survey Research Association’s journal, *Survey Research Methods*, published an extensive special issue in June, covering papers about design proposals, adaptations to ongoing surveys, and research initiatives. The World Population Association organized public webinars covering the challenges to surveys all around the world, with researchers attending from Asia, Africa, Latin America, Europe, and North America. In addition, the American Association for Public Opinion Research held workshops, and The Royal Statistics Society held two webinars on how the pandemic affected data collection from various aspects. The United Nations World Data Forum held a special session on the topic.

International organizations have also been active in guiding survey organizations. A Task Force was founded under the UN Intersecretariat Working Group on Household Surveys. Pandemic related recommendations and briefs from UN agencies, such as the World Bank, UNICEF, UN Women, ILO, UNSD, and their regional offices. The World Health Organization and EUROSTAT have also provided documents for international use.

In this project, we reviewed the developments in survey research based on similar sources listed above and survey organizations’ websites, with the aim of putting contemporary and future topics into perspective. Our desktop research covered the period from the pandemic’s beginning to mid-December 2020. We should underline that we do not cover all developments, and our review is limited to relatively more visible sources such as those above and those in English and Turkish.

To help with our understanding of issues, we decided to conduct expert interviews. We prepared separate interview guidelines for each expert, covering a wide range of topics overall while directing questions on each researcher's area of expertise. We conducted a total of six interviews between September 25 and November 20. We asked Prof. James M. Lepkowski, Emeritus Professor at the Survey Methodology Program, University of Michigan Institute for Social Research, general questions about the effects of the pandemic on various modes and their implications for survey errors. Dr. Attila Hancioğlu, the chief of the Data Collection Unit, Data & Analytic Section at the UNICEF Headquarters in New York, talked about the UNICEF Multiple Indicator Cluster Surveys program in the context of the pandemic. In our interview with Dr. Raphael Nishimura, the Director of Sampling in Survey Research Operations at the University of Michigan, he told us about his experience with household surveys from Brazil and the Institute for Social Research at the University of Michigan. We asked Prof. Frauke Kreuter, the Director of the Joint Program in Survey Methodology at the University of Maryland and Founder of the International Program for Survey and Data Science at the University of Mannheim, about more contemporary topics in survey research, such as web surveys and the use of big data. We also aimed to have a country-highlight, both in our desktop research and interviews. Enver Taştı, a senior expert and former-Deputy Director of TURKSTAT, shed light on the practices of TURKSTAT during the pandemic. Finally, our interview with Oğuzhan Akyıldırım, CEO of Tandans research company, reflected on the experiences of the private research sector in Turkey.

This publication is aimed at the general reader, anyone who has an interest in surveys. For this purpose, we start with a background section that provides a historical background for surveys. We then move on to types of surveys based on different classifications. We have separate subsections on survey modes since they lie at the heart of many discussions mentioned in this publication. We talk about some international survey programs to introduce the readers to well-known survey series, with a special focus on Turkey. Overall, we introduce the basic concepts and terminology for readers from outside the field to prepare the sections ahead.

Chapter 3 presents our review of online resources about the pandemic and surveys in the following section. Here we summarized our review separately for cross-sectional and longitudinal surveys. We also mentioned surveys designed specifically to focus on the impacts of the pandemic on various aspects of people's lives while classifying them based on the type of sampling frame they used. We briefly reviewed some innovations in data collection deemed important during the pandemic. We allocated a brief section on surveys to estimate the extent of the spread of COVID-19 is and another on survey research in Turkey. We concluded with a short section on several ethical issues discussed by the researchers in their studies during this extraordinary period.

In Chapter 4, we present the transcriptions of the interviews conducted with the experts presented above. The video recordings of these interviews are available on our Institute's YouTube channel for those interested.

We conclude our research and interviews in Chapter 5. In that section, we summarize our general conclusions and support them with observations from Chapter 3 and quotations from our expert interviews.

2. Background on Surveys

2.1 Sample Surveys

Surveys that employ sampling methods investigate a portion of the population to estimate figures about the whole population (Thompson, 2012). Sample surveys are the principal means to apply empirical research for many social science fields as they provide a substantial part of the data to monitor society's tendencies (Rossi et al., 1983). Researchers opt for sample surveys for their cost advantages as well as for their detailed data to focus on specific topics, which also enables to focus on other elements of surveys such as interviewers, respondents, and survey instruments (L. Kish, 2001). The growing demand for data and information about people's behaviors, tendencies, and preferences lays further emphasis on sampling methods. Sample surveys enable researchers to obtain richer data within a shorter time vis-à-vis censuses that mainly aim to deliver data based on the needs of governmental agencies or official enterprises (Everitt, 2003). A complete enumeration of the population takes longer and costs more. Appropriately applied, sample surveys can provide sufficient accuracy and the chance to measure the margin of uncertainty based on objective criteria at a significantly faster pace and lower costs (Mahalanobis, 1965).

Sample surveys are distinct from experimental studies. Experimental research chiefly aims to determine cause-effect relations, whose main advantage is its internal validity compared to other empirical inquiries, but it is not suitable for every type of investigation (Lavrakas, 2008). It may produce artificial results since experimental settings, which generally focus on the observed differences between the two groups, might result in overlooking some of the relevant factors (Lavrakas, 2008). Researchers perturb conditions or some part of the population to understand the results of the topic in question, whereas, in sample surveys, researchers often do not aim to disturb and manipulate the population they study (L. Neuman, 2014; Thompson, 2012). For instance, in a sample survey, researchers endeavor not to influence respondents with the wording of the questionnaire to collect the most accurate information (Thompson, 2012).

Sample surveys also differ from observational studies that aim to unearth the cause and effect relationships without employing controlled experiments and do not attempt to influence the units studied (Everitt, 2003). Sample surveys, on the other hand, seek to avoid any factors that can beget "unrepresentative" results by purposely selecting the sample (Thompson, 2012). Meaning, sample surveys apply the methods to produce population estimates.

Ideally, the sampled population would be identical to the target population. However, this rarely happens in reality since the sampled population is often smaller (Lohr, 1999). Put another way, a good sample should be representative and reflect the population's characteristics, which means: "each sampled unit will represent the characteristics of a known number of units in the population" (Lohr, 1999, p.3). It is crucial to introduce the concepts of *probability sampling* and *non-probability sampling*. For probability sampling, each member of the population has a known nonzero probability to be selected, and a well-designed probability sample should ensure that the population is known while the bias is lowered to the extent possible (Lavrakas, 2008). The probability sample, furthermore, is predicated upon three assumptions: a sample frame of all units of the population is feasible, a person or thing has a likelihood to be selected for the sample, and this probability can be calculated for each sampled unit (Cowles & Nelson, 2015).

Simple random sampling (SRS), systematic sampling, stratified sampling, and cluster sampling are the main types of probability samples. As a basic form for probability samples, *simple random sampling (SRS)* grants the same chance to every possible subset of n units in the population, while *systematic sampling* requires to select every k^{th} unit after a random selection (Biemer & Lyberg, 2003; Leslie Kish, 1965). Simply put, with *stratified sampling*, researchers structure the sampling process to lessen the sampling variation to resemble the sample to the population, whereas *cluster sampling* indicates the selection method that sampling unit involves more than one population element, and each element requires to be uniquely identified with one sampling unit (Fowler, 2014; Leslie Kish, 1965)

For non-probability samplings, nonetheless, researchers do not aim to choose a random sample, which is to say, they apply “subjective” methods to include elements into the sample (Lavrakas, 2008). Indeed, it indicates the lack of probability sampling mechanisms that include randomization, meaning that the units selected into the sample have unknown probabilities, of which some are zero (Vehovar et al., 2016). *Convenience sampling* that is based on selecting the respondents on convenience and availability; *quota sampling*, which guarantees that certain features of the population are represented in the sample; and *snowball sampling*, which additional participants are added to the sample via the information and connections provided by the respondents, are main types of nonprobability sampling (Acharya et al., 2013).

Sample selection in probabilistic sampling has three facets: the sample frame, which should correspond to the population studied, probability sampling to include each person with a known chance to be selected; and the details of the sample design involving its size and other procedures for selecting the sample which determines how far the sample is close to the characteristics of the entire population (Fowler, 2014).

The sampling frame is the complete list of all sampling units in the target population, and, to a significant degree, it determines the quality of a survey (Lavrakas, 2008). Researchers utilize geographic listing, maps, directories and membership lists for sampling frames depending on the mode of the inquiry (Everitt, 2003). It is safe to say that *enumeration areas (EAs)*, which group several households for convenient counting purposes for the census, are the ideal frames, particularly for surveys as *Demographic and Health Survey (DHS) (DHS Sampling and Household Listing Manual, 2012)*.

While establishing the sampling frame, one should assess its comprehensiveness or its coverage of the target population, and its suitability to calculate the probability of selections as well as its efficiency, which implies the extent of the target population available (Fowler, 2014). It goes without saying that obtaining a perfect sampling frame is unlikely or, most of the time, improbable, and researchers need to scrutinize the shortcomings of it or seek alternatives (Lavrakas, 2008).

Determining the sample size is another crucial step in designing sample surveys. The optimal one and the practical size is a tradeoff researchers encounter, meaning that there is the sampling error on one side, and budget, time, and effort limitations on the other (Cowles & Nelson, 2015). Put differently; the challenge is to seek a balance between the precision and purposes of analysis with the capability of implementing organization and the financial burdens (*DHS Sampling and Household Listing Manual, 2012*).

The range wherein the true value of the population is estimated to be, or *level of precision*; the *confidence level* which is based on the theories that are embodied in Central Limit Theorem¹ and the *degree of variability* that indicates the distribution of attributes in the population are the technical criteria for determining the sample size (Israel, 1992). Further, concerning the fact that researchers mostly do not have adequate information to confirm that their sample size is the most suitable, the determination of the size oftentimes might not be satisfactory (Cochran, 1953). In addition, an overestimated sample size can lead to a waste of resources or might not be big enough, which may yield dysfunctional results (Cochran, 1953). Plus, a larger sample size, theoretically, translates into a superior survey precision, but non-sampling errors can prove it to be wrong since they have a propensity to increase with the sample size (*DHS Sampling and Household Listing Manual*, 2012).

The ways to improve the data quality in a sample survey are multifarious. *Total survey error*, which assumes that every step of a survey is open to errors, is the dominant approach that focuses on various error sources to avoid them (Weisberg, 2005). In this vein, many users equate data quality solely with the amount of error in the data, meaning that high quality signifies “perfectly accurate” data, and low quality stems from the significant amount of errors (Biemer & Lyberg, 2003). Moreover, researchers endeavor to maximize data accuracy within a limited budget (Lavrakas, 2008). *Total survey error*, which is the estimate of “the difference between the estimate and the true value of the population parameter,” involves two components: *sampling error* that refers to the errors stemming from working with a sample instead of conducting a census and *non-sampling errors*, which emanates from data collection and processing procedures (Biemer & Lyberg, 2003).

Sampling errors influence the survey quality, the size, and design of a probability sample, and the distribution thereof determines the size of sampling errors (Fowler, 2014). Additionally, sampling errors, in general, are demonstrated with the standard error of an estimate, which implies the variability of sample estimates based on the true population value (Visser et al., 2014).

It is also worth noting that sampling errors can be viewed as *intentional errors* since researchers can employ some adjustments to alter the sample size to deal with them, given that there is no sampling error when the sample size is identical to the population, as Biemer and Lyberg (2003) note. Non-sampling errors, nevertheless, are unavoidable and unpredictable, especially when the survey operation expands (Biemer and Lyberg, 2003).

Specification error is one type of non-sampling error that arises from estimating the wrong parameter in a survey. That is, the survey measures a different concept than the one aimed to be evaluated (Biemer and Lyberg, 2003). More to the point, if a significant dimension of concept that should be measured is missing, that would lead to a specification error (Lavrakas, 2008).

A further dimension of non-sampling errors is *frame error* (also called *coverage error*). For instance, this error was an area of concern for telephone sampling, particularly before the 1970s, when researchers did not opt for phone surveys since telephone coverage was not as extensive as today (Weisberg, 2005). Whereas including units that are not supposed to be in the sample is *over-coverage*, a design excluding the units that must be in the sample leads to *under-coverage* (*DHS Sampling and Household Listing Manual*, 2012). By way of example, web surveys that exclude the people who do not

¹ According to the Central Limit Theorem, the sampling distribution of the mean approaches a normal distribution as the sample size increases, which means that the sample mean and standard deviation will be closer that of the population with increasing sample size.

have internet services or telephone surveys that only cover the landline phone users would lead to a coverage error (Tourangeau, 2018). Selecting the same element twice or including a vacant household in the frame are also the errors within this context (*DHS Sampling and Household Listing Manual*, 2012; Piazza, 2010).

Non-response error, as another source for non-sampling errors, is the failure to collect data from some sample members when they refuse to answer or when they are inaccessible (Steiger & Conroy, 2008). A sample unit (household, person, establishment, farm etc.) that does not participate in a survey leads to *unit-nonresponse*, whereas *item-nonresponse* corresponds to partially completed questionnaires due to skips and unfilled questions (Biemer and Lyberg, 2003). Further, the final type of non-response error is *incomplete responses*, which occur when participants deliver incomplete or inadequate answers, notably for open-ended questions (Biemer and Lyberg, 2003).

As a complicated source of non-sampling error, *measurement error* is another type that can arise from the effect of interviewers, respondents, questionnaires, and mode of data collection (Groves, 1989). Errors related to the interviewer are considered when the responses vary due to the interviewer's behavior or conduct in face-to-face interviews (Weisberg, 2005). For respondents, psychological factors and their moods can be the elements that affect the responses, while data collection mode (face-to-face, telephone, mail, self-administered) also have a critical impact on the data (Lavrakas, 2008). In a similar vein, questionnaire design is also a factor that affects the interviewees' responses (Lavrakas, 2008).

The last component of non-sampling error is the *processing error* that occurs during the data editing, entry, and processing procedures. It generally happens during the post-collection stage of survey operations that also may take place owing to transcriptions or calculation in the tabulations (*Processing Errors*, 2005). Survey planning, budget, resources, and constraints such as time, are the factors that influence this type of error (*Measuring and Reporting Sources of Error in Surveys*, 2001).

Some scholars categorize total survey error from a different viewpoint. Classification between *representation* and *measurement* is a more recent view for errors. According to that, total survey error components affecting the *representation* are coverage error, sampling error, nonresponse error, and adjustment error (samples that need to be adjusted due to the design effects), while total survey error components affecting *measurement* are specification error, measurement error, and processing error (Lavrakas, 2008).

Moreover, the total survey error paradigm finds itself in the line of criticism, based on the argument that it lacks a user perspective (Hansen et al., 2016). Some scholars argue that the suitability of the data to user's requirements is essential for data quality that the total survey error paradigm overlooks (Hansen et al., 2016). The dimension of quality hence has many more aspects, as Hansen et al. (2016) demonstrate within the *fitness for intended use* framework:

- *Comparability* (Is the data comparable with other data sources?)
- *Coherence* (Is the data forms a coherent body that can be combined or rearranged with other data sources?)
- *Relevance* (To what extent that data meet the user requirements?)
- *Accuracy* (Are the survey estimates close to the true values?)
- *Timeliness and punctuality* (time used after data collection, the time data is available, and whether the data is ready when expected)

- *Accessibility* (Can users access the data easily?)
- *Interpretability* (The consistency of hypotheses with the survey results; Are metadata², paradata³ and auxiliary data available?)
- *Cost, burden, professionalism, and design constraints* are also factors that play a role in fitness for use framework, meaning that optimizing costs, minimizing burdens, recognizing and documenting design constraints are other purposes of survey enterprises.

Although total survey error is the dominant approach and paradigm, recent approaches examine the survey quality through different lenses by including the data user's perspective on quality. It is fair to say that survey quality approaches are burgeoning, and scholars address the total survey error approach from different outlooks.

2.2 A Brief History of Surveys

Broadly accepted, the 18th century is marked as the starting point of modern censuses with the rise of statistical methodologies, internationally accepted definitions, and conceptualizations (Baffour et al., 2013). Therefore, it is necessary to differentiate between historical and modern censuses (Baffour et al., 2013). One might trace back the historical censuses to as early as 16th century BC Athens when the Cecrops, the first king of Athens, conducted a survey that only targeted men intending to evaluate the military capacity where each person cast a single stone to be counted (Missiakoulis, 2010). Intriguingly, Madansky (1986) proved that the Old Testament refers to three censuses in detail. The primary purposes of censuses were to monitor how many men pay taxes and how many of them can be used in wars (Biemer & Lyberg, 2003). Also, countries viewed censuses as a sign of political power; for example, the Swedish census, circa 1700, proved that the population of the country is considerably less than thought and the authorities hid the figures as this was believed to be an indication of weakness (Biemer & Lyberg, 2003).

The idea of sampling grew stronger with time. The combination of statistical and scientific theories laid the basis of modern sampling (Stephan, 1948). Indeed, as Stephen (1948) propounds, human beings always employed sampling and randomization in their daily activities, such as when they taste or test a thing. People arrive at conclusions in many fields of their life based on a sample; say, about food, animals, fire, and they make inferences, albeit most of the time, unconsciously (Kruskal & Mosteller, 1980). The intuition has always been there, even as a Neanderthal "sipped a bit of broth to learn whether the whole pot was cool enough to eat" (Wright & Marsden, 2010)

Theories of sampling were not present until around the 1900s; nevertheless, scientists in Europe sought ways to estimate the population's size in case of an unfeasible census (Biemer & Lyberg, 2003). Scientists as Graunt and Eden applied what is called *political arithmetic* during the 17th century based on the works of William Petty, who combined statistical records such as birthrates, family size, and the average number of people per house with "bold" calculations (Biemer & Lyberg, 2003; Porter, 2020). These attempts can be viewed as early precursors to modern sampling, while John Graunt's "Observations upon the Bills of Mortality" in 1662 sets an example for *political arithmetic* that involves

² Metadata is "a set of highly structured and/or encoded data that describes a large set of data and explains the data to be collected, processed and published and answer questions regarding every facet of the documented data" (Lavrakas, 2008, p.463)

³ Paradata is also called as process data, which involves information about the primary data collection process that helps to evaluate the quality of data (e.g survey duration, navigational errors in survey questionnaire) (Lavrakas, 2008)

compilation and manipulation of empirical observations (Wright & Marsden, 2010). Graunt's study that introduced a method to estimate London's population can also be marked as the first documented investigation based on a sample (Kuusela, 2011).

Applying statistics to population-related measurements by Quetelet in 1820 was also a significant step forward since statistical developments underpin sampling studies and censuses (Leslie Kish et al., 2003). Kish (2003) emphasizes that A.N. Kiaer's paper in 1895 can be viewed as the naissance of survey sampling. Besides, Kiaer, Director of the Norwegian Central Bureau of Statistic, was the first scholar that used the term *representative method*⁴ in an analytical sense.

Notably, Charles Booth's survey on the conditions of the workers in London around the 1890s is largely accepted as the first form of a modern survey, in which Booth demonstrated poverty maps of every street in London, unfolding the social conditions of people (Wright & Marsden, 2010). At the time, Booth and others attempted to gather data on every element of the population, and the problem was its impracticality for large populations (Groves et al., 2004). As Groves et al. (2004) explain, researchers would investigate a town and seek to resemble that to the population by distributing it to certain quotas, such as employing the survey to half men and half women.

Another essential household-by-household survey at the time was Dubois's study on the Philadelphia community, in which he utilized the forms of the 1890 census to gather basic demographic and employment data (Rossi et al., 1983). Further, despite not using the term, the need for a confidence interval (the range of values that is likely to include a population value with a certain degree of confidence) approach as outlined in 1906 by English statistician Bowley who also conducted surveys on living conditions in Reading by systematically selecting buildings from the lists (Kalton, 2019). In those years, the closer to a complete census, the better the quality was the general approach to the survey research, and researchers aimed to conduct complete enumerations for their surveys (Wright & Marsden, 2010). Another trend, particularly in sociology, was working on a few families, a single city or a tribe in detail using case studies, called monography (Kruskal & Mosteller, 1980).

Studies on agriculture, forestry, government administration, and social research can be marked as the roots of survey sampling techniques. Especially, studies on government administration and social research methodology inspired works on random sampling (Bellhouse, 1988). In particular, surveys on agriculture took sampling studies a step further with *area probability sampling*, which means to sample areas or plots of land to predict the crop yields (Groves et al., 2004). Moreover, this practice reflected upon the methods used for household sampling, which is to say, samplers started to select samples from geographic blocks of settlements, and they eventually sampled the household units listed by giving them a chance to be drawn, as Groves et al. (2004) underscore.

⁴ His paper is entitled "Observations and experiences with representative surveys". A representative sample is defined as "one that has strong external validity in relationship to the target population the sample is meant to represent. As such, the findings from the survey can be generalized with confidence to the population of interest" (Lavrakas, 2008, p.720)

In the 20th century, survey research began to crystallize when sampling designs as stratified sampling, systematic sampling, cluster sampling, multistage sampling⁵, and multiphase sampling⁶ were introduced before the 1930s (Thompson, 2012). Markedly, there was a transition from complete census and full coverage to partial investigations. Besides, in 1926, the International Statistical Institute (ISI)⁷ issued a report recognizing the representative method, which entails random selection and purposive selection, as the basis for sampling (Kalton, 2019).

Laying the groundwork for probability sampling and the idea of the bias-free survey, Neyman's article⁸ in 1934 was a breakthrough, which was even accepted as a paradigm in Kuhnian terms (Bellhouse, 1988). He also put forward the theory of confidence intervals, cluster sampling, ratio estimation, and two-phase sampling (Biemer & Lyberg, 2003). The 1940s, on the other hand, bolstered the field with new developments in universal coverage, structured questions, sampling error measurements, and statistical inference to finite populations (Groves, 2011).

In area-based sample frames, which were dominantly used around those times, researchers utilized pieces of land - through civil jurisdictions and enumeration districts, as the sampling units for stratified and multistage samples (Groves, 2011). Groves (2011) also notes that samplers used city directories along with telephone directories or subscriber lists as sample frames in the USA. However, since telephone use was rare, probability sampling was feasible, especially for political surveys during the 1940s.

Straw polls⁹ gained momentum since having data on public opinion was a growing need as, among other things, industrialization, waves of immigration, the emergence of new media of the time, such as television and radio, pushed political authorities to monitor what people think, and accordingly shape their policy-making to win the office (Carpini, 2011). For instance, there were 86 straw polls for the USA presidential campaign in 1928, and they were many times successful, notwithstanding their crude methods (Stephan, 1948). These election surveys conducted by newspapers and magazines had a "bigger the sample, better the estimation" approach, and the biases stemmed from selecting samples among telephone subscribers and automobile owners, which challenge the principle of representativeness (Wright & Marsden, 2010). By way of illustration, *The Literary Digest*, a weekly magazine in the USA, conducted a mail survey using telephone directories and automobile ownership lists. This way, they had a vast sample involving over 2 million people, but the sample was unrepresentative since the ownership of automobiles and telephones was limited in the population (Wright & Marsden, 2010). The magazine missed the mark with its prediction of Landon's victory, and Roosevelt took office.

⁵ Sampling is done with two or more hierarchical levels in multi-stage sample. For instance, the first level can be the country level, and the other levels are the census-track level, block level, household level, and within-house level (Lavrakas, 2008).

⁶ Two-phase sampling (double sampling) is the often used form of multi-phase sampling. While more phases are feasible, for this method, researchers collect basic information from a large sample of unit and collect more detailed information for a subsample of these units (*Multi-Phase Sampling*, n.d.).

⁷ The International Statistical Institute (ISI) is a professional association of statisticians. Established in 1885, it has around 4000 members.

⁸ Neyman's article is entitled "On the two different aspect of the representative method: The method of stratified sampling and the method of purposive selection"

⁹ Straw poll is "an unoffical vote that is taken to discover what people think about an idea or problem or how they intend to vote in an election" ("Straw Poll," n.d.)

Researchers began to employ quotas for political polls with a view to overcoming sampling problems; that is, quotas in sex, age, religion are thought to “mirror” the population (Rossi et al., 1983). This development was significant as it helped to advance towards the representative sample though it appears archaic from the present-day perspective (Wright & Marsden, 2010). Gallup, an American advisory company, set a successful example of a quota sampling survey as it correctly predicted the results of the US election in 1936 with a sample of 5000 people, which proved that small and representative samples could ensure better inferences as Wright & Marsden (2010) underline. Quota sampling was nevertheless not a panacea, as Gallup’s prediction for the 1948 presidential election turned out to be a failure when Truman won the election against Thomas Dewey (Langer, 2018). Other factors aside, 9600 cells were needed for cross-tabulations of ethnicity, race, education, and so forth to replicate the US population, which amounted to enormous sample size (Langer, 2018). That coalesced scholars on the power of probability-sampling as the scientifically rigorous method for representative results, Langer (2018) notes. Also, in 1948, Robert Kahn and Agnus Campbell, who were working in Survey Research Center (SRC) of the University of Michigan, ran a foreign policy survey that also included two questions about the upcoming elections of the US. Employing probabilistic sampling methods, they predicted a Truman win with a slight difference while other survey companies were bluntly estimating a clear Dewey victory (*ISR and the Truman/Dewey upset*, n.d.). That compelled the attention to the power of probabilistic surveys.

Face-to-face interviewing and mail surveys were the dominant data collection modes after the 1930s, while telephone surveys gained ground as private sector firms increasingly utilized them towards the sixties (Groves, 2011). As Groves (2011) named the period as “the glory years of non-response,” response rates were usually over 90 percent between the thirties and sixties. In addition, texts of Yates (1949), Deming (1950, Hansen *et al.* (1953), Cochran (1953), and Sukhatme (1954) that had broad repercussions on survey sampling were published in this era (Kalton, 2019).

World War 2 exerted influence on governmental survey enterprises, given the increasing need to measure public opinion, particularly in the US. For example, what public sentiment was significant for Roosevelt during the second world war, and sample surveys gained further credibility to measure public opinion for the president (Wright & Marsden, 2010). As Wright & Marsden (2010) also point out, survey expertise transferred from governmental positions to academia in the aftermath of the Second World War, which helped the development of the field of social science departments across the US.

Furthermore, US Census Bureau’s efforts on sampling methods lay the foundation of multistage area probability sampling, which is dominantly employed in national surveys since it applies sampling in large, geographically dispersed populations with no listing (Wright & Marsden, 2010). Also, scholars in US Bureau involving Morris H. Hansen founded a model for total survey error, which helps to measure the total error of an estimate as to the mean squared error of that estimate (Biemer & Lyberg, 2003).

There were also significant developments in other parts of the world, such as in India, which is to say, the Indian Statistical Institute had a significant influence in the field by training a large number of statisticians in India and a multitude of statisticians in the “least developed countries” (LDCs) (Leslie Kish et al., 2003). India was one of the top countries in using sampling techniques thanks to its founder, P.C. Mahalanobis, and his interest in survey sampling (Bellhouse, 1988). The method of

*interpenetration*¹⁰ is his brainchild. He utilized it to calculate errors that stem from interviewers, coders, and supervisors (Biemer & Lyberg, 2003) as well as the sampling errors (Rao, 2009).

In the 1960s, people began to use the telephone as the primary communication device, which brought firmer sample frames provided by telephone companies (Groves, 2011). Besides, survey operations began to incorporate computers, which were also improving, especially for data processing and further analyses as well as utilize statistical packages such as SPSS and OSIRIS (Rossi et al., 1983). Another striking development at the time, according to Rossi et al. (1983), was the mixing of sample surveys with other methods, as was the case for economists that employed the econometric models in sample surveys.

It is also imperative to note that fundings flourished for social programs in the 1960s; thereby, researchers found more opportunities to finance their sample surveys (Groves, 2011). Sample surveys gained significant momentum in social sciences with the fundings, and large projects employed face-to-face interviews, where mail surveys were preferred by smaller studies (Wright & Marsden, 2010). Surveys for the private sector were also on the rise, in conjunction with the nexus between customer survey statistics and management decisions, as Groves (2011) stresses. Further, in 1968, Cannell and Kahn introduced schemes that describe the interviewer- respondent interaction, which was not studied yet (Biemer & Lyberg, 2003). Indeed, Biemer and Lyberg (2003) argue that the respondent's role in data collection was seen as a hindrance then, which is still the case for surveys requiring mandatory participation.

The 1970s, on the other hand, marked a watershed for survey costs and response rates, as the survey operations were more costly with the increasing wage rates of interviewers, whereas a decline in response rates occurred driven by many reasons, including increasing urban crime and less time spent at home (Rossi et al., 1983). Researchers resorted to alternatives such as telephone surveys with computer assistance (Rossi et al., 1983). Interviewees, nevertheless, could terminate the interviews easier, such as by "hanging up" the phone, which resulted in more and more partially completed interviews (Groves, 2011).

In a similar vein, response rates and escalating costs were the chief concerns of the 1980s, and mail surveys also attracted further attention on the grounds that they could yield response rates similar to face-to-face interviewing with lower costs (Wright & Marsden, 2010). Still, telephone interviewing was the preferred mode because of its reduced cost -, and its data quality, reliability, and validity were close to that of face-to-face interviews (Wright & Marsden, 2010).

The appearance of mobile phones after the 1990s had a dramatic effect on surveys concerning the fact that sample frames used for telephone were not as efficient as before (Groves, 2011). Working on sample frames combining landline phones and cellphones raised significant challenges, given that many people had both (Dillman et al., 2014). Cellphones' portability entailed risks on duplication or omission of the people in the sample frames (Dillman et al., 2014). On top of that, locked apartment buildings, caller ID screening features, automatic answering machines on phones influenced the

¹⁰ "When two or more samples are taken from the same population by the same process of selection, the samples are called interpenetrating samples. The samples may or may not be drawn independently, linked interpenetrating samples being an example of the latter. There may be different levels of interpenetration corresponding to different stages in a multi-stage sampling scheme." (*Interpenetrating Samples (Sub Samples)*, 2003)

response rates for face-to-face and phone surveys negatively, let alone the extended labor to contact and interview people which increased the costs of face-to-face interviewing (Groves, 2011).

Furthermore, computers' involvement in survey research gathered pace after the 1980s (Wright & Marsden, 2010). Computer-assisted survey research methods, which facilitate the interview process, are usually outlined under the abbreviation of *CAI (Computer Assisted Interviewing)*, while conventional methods with paper and pen are titled *PAPI (Paper-and-pen interviewing)*. The main types of survey modes with computer assistance can be summarized as follows :

- CATI: Computer-assisted telephone interviewing is the first form of computer-assisted interviewing (de Leeuw & Nicholls, 1996). Interviewers work from a center and ask questions reading from a computer screen while entering the responses to the system. Supervisors control and help the interviewers for ensuring data quality in the survey center; however, decentralized CATI operations are also feasible where interviewers can work from home (de Leeuw & Nicholls, 1996)
- CAPI: Computer-assisted personal interviewing corresponds to the face-to-face version of CATI (Computer-assisted telephone interviewing). Interviewers conduct the interviews reading the questions and processing the answers via a laptop or tablet. The interviewers can send the data to the survey headquarters through an internet connection.
- CASI (Computer-assisted self-interviewing) or CSAQ (Computer-assisted self-administered questionnaires) help to employ very elaborate questionnaires without interviewers (de Leeuw et al., 1995). Participants get the computerized questionnaires via e-mail or a disc sent by mail and respond to them, eventually sending them back to the survey organization (Wright & Marsden, 2010)
- CASIIP (Computer-assisted self-interviewing with interviewer present) DBM (Disk by Mail), EMS (Electronic Mail Survey), CAPAR (Computer Assisted Panel Research), Teleinterview, TDE (Touchtone Data Entry), VR (Voice Recognition), ASR (Automatic Speech Recognition) are the other types of computer-assisted interviewing (de Leeuw & Nicholls, 1996).

On the other hand, researchers were hopeful about a transition to online modes at the end of the 1990s, believing that the internet will be the primary mean for conducting surveys considering the decline in telephone response rates (Dillman et al., 2014). The internet offered low costs, embedded consistency checks during the surveys, and rapid data presentations, which were appealing for researchers (Groves, 2011). But internet connection was not widespread, and the demographic profile of users and non-users differed crucially, not to mention the lack of sample frames that enable direct e-mail contact (Dillman et al., 2014). That is still the case due to the fact that not every region of the world has the same internet quality and penetration. No doubt that the level of income also affects internet usage around the globe. The number of internet-based surveys had nevertheless increased sharply in recent years, with substantial developments in internet usage (*Internet Surveys*, n.d.). Respondents have more time and personal space in internet surveys; it has less social desirability bias, and multimedia elements can be utilized in this research (*Internet Surveys*, n.d.)

The world witnessed a plethora of technological developments after the 2000s; technologies as voice recognition, smartphones, tablets, and many other devices that allow unfettered ways of information sharing began to be used for surveys with increasing internet usage (Dillman et al., 2014). It does not mean that traditional methods as face-to-face interviewing, mail or phone are extinct; they are still at

the heart of survey research, as Dillman et al. (2014) underscore. Mixed modes are also on the march, which will be touched upon later.

However, now that people have smartphones, tablets, and other devices that keep them connected to the world at all times, researchers have to take innovative steps to conduct surveys on various mediums. We live through the digital age where data is produced from myriad sources, including social media, video feeds, medical databases, and genome sequences (Fuller et al., 2017). These sources generate the “big data,” and social researchers encounter this buzzword increasingly, as they can not ignore its hype anymore (Mills, 2019; Salganik, 2018). Gartner report defines big data in terms of three V’s (as cited in Callegaro & Yang, 2018):

- *Volume*: Big data are those data that cannot be dealt with by traditional analytical tools
- *Velocity*: It addresses data that are coming in real-time
- *Variety*: Big data are complex datasets and include very different sources of context such as unstructured text, media content such as images and videos, logs, and other sources.

Put another way; big data indicates that “there is a lot of data, in a variety of formats, and it is being created constantly”(Salganik, 2018, p.14). It is also important to note that big data is not just for research purposes as opposed to the social research data in the analog age, which were often created for just research purposes (Salganik, 2018). Callegaro and Yang (2018) identify the primary sources of big data in the following way:

- *Internet data: Online text, videos, and sound data*: It also involves *social media data* and *website metadata, logs, cookies, transaction, and website analytics*.
- *The Internet of Things (IoT)*: It refers to the devices that communicate with each other through the internet, as devices progressively have this feature, and this data can be used for research purposes. The data originated from smartphones, wearable technology, smartwatches generate its subset: *Behavioral Data*
- *Transaction data*: Record of orders, credit card activities, billing, shipments are examples of transaction data (Ferguson, 2014).
- *Administrative data*: It is composed of data collected by public institutions such as health centers, schools, tax offices.
- *Commercially available dataset*: Data about consumers that companies collect.
- *Paradata*: Paradata concerns survey researchers as it refers to data collected by systems, or interviewers, before, during, and after the interviews. Alternatively, in a self-administrated web survey, the respondents create the paradata based on their interactions with the survey data.

Whether we reached the end of the road for survey research with big data's arrival remains the central conundrum. It might be the case in some ways. For instance, there is a decline in budget spent on market research as ESOMAR Global Market Research reports demonstrate; however, as Callegore and Yang (2018) note, the real trend in social and market surveys that need to be considered are:

- From offline data collection to web surveys
- From web surveys to mobile web surveys
- From outsourced market research to in-house market research using *Do it Yourself (DIY)* web survey platforms (e.g., SurveyMonkey)

- From outsourced market research¹¹ to in-house market research fully integrated with internal systems¹²

Big data provides various advantages: researchers obtain the data without a particular effort on primary data collection, which is generally costly and takes a serious amount of time (Japiec et al., 2015a). Conversely, skeptics believe that big data can not supersede the survey research, given many big data sources are unrepresentative (Salganik, 2018). Be that as it may, using big data and surveys combined appears as a fine choice to maximize the benefits of both. Big data can help with the behaviors and answer the “what” question, whereas surveys can examine in detail investigating attitudes and opinions; to answer the “why” question (Callegaro & Yang, 2018).

As mentioned above, the total survey error framework is the principal data quality assessment tool for surveys. For big data, Japiec et al. (2015a) present a three-step framework in the *American Association for Public Opinion for Research’s (AAPOR) Big Data Report* and state that errors can occur during these phases (Callegaro & Yang, 2018):

- *Data generation*: The problems in big data generation can spring from poor-documentation; errors can arise as missing data, self-selection, coverage, non-representativeness, and low signal-to-noise ratio, as Callegore and Yang (2018) stress.
- *Extract, transform and load (ETL)*: In this phase, researchers gather data by extracting, translating, coding, recoding, aggregating/disaggregating, and editing the data (Japiec et al., 2015a). Data cleaning, matching, coding, editing, and data cleaning might be the sources of errors (Callegaro & Yang, 2018).
- *Load Stage*: The final stage is to integrate and store the data in the data warehouse (Japiec et al., 2015a). Errors can stem from sampling, selectivity, modeling, and estimation, as well as data visualization (Callegaro & Yang, 2018).

2.3 Types of Surveys

There are many types of surveys whose taxonomy can be undertaken from various angles. Surveys could be categorized with a time dimension. While some are one-time studies that aim to investigate population characteristics, the others are continuous, aiming to measure the change in time (Biemer & Lyberg, 2003). As Biemer and Lyberg (2003) address, it is also possible that some surveys can be repeated over time, even as they are planned to be a one-time enterprise, which is not an unusual case.

Alternatively, they may also be incorporated under *cross-sectional research* and *longitudinal research* as Neuman (2014) classifies:

- *Cross-sectional surveys*: This type of survey conducts the investigation with many cases at one point in time (L. Neuman, 2014). Stated differently, they are the snapshots of the population about which researchers collect the data (Lavrakas, 2008). Whether it aims to provide descriptive estimates of the population or test hypotheses between related variables can determine if a cross-sectional sample survey is descriptive or analytic (Jupp, 2006). One

¹¹ Companies work with market research companies to carry out studies done on their behalf in outsourced market research.

¹² In-house market research is when a firm conducts the study within its business structure and with its own resources.

analogy would be seeing the process as a plank of wood with unknown length and cutting across it at one point, intending to see a single “cross-section.” Not being able to examine the whole is a disadvantage of cross-sectional surveys (Payne & Payne, 2004). Hence, cross-sectional surveys that capture the change are seldom (Neuman, 2014). Conducting research that compares cholesterol levels of walkers and people who do not walk at the same point in time would be an example for a cross-sectional study (*Cross-Sectional vs Longitudinal Studies*, 2015). As for the advantages, cross-sectional studies are the most straightforward and least expensive survey type; it also allows us to investigate many different variables concurrently (*Cross-Sectional vs Longitudinal Studies*, 2015; Neuman, 2014).

- *Longitudinal surveys*: Researchers investigate the same subjects over a period of time in the longitudinal studies (*Cross-Sectional vs Longitudinal Studies*, 2015). In contrast to cross-sectional studies that collect data just once, longitudinal research aims to analyze the social change over time (Payne & Payne, 2004). For instance, studying a single population of daily-walkers for a long period is an example of longitudinal research (*Cross-Sectional vs Longitudinal Studies*, 2015). The impetus behind conducting a longitudinal survey lies in the fact that people’s knowledge, attitude, behaviors, perceptions generally alters with time (Lavrakas, 2008). These studies are more potent since they allow us to observe the change and growth of a dependent variable as well as enable us to predict the long-term effects of the dependent variable, even though designing and conducting these surveys are more complicated and expensive (Lavrakas, 2008; L. Neuman, 2014). As Lavrakas (2008) argue, other challenges are that, some respondents might not be available during the sequences of data collection, or they might drop out at a point in addition to the problems that may stem from the death of participants. Further, Neuman (2014) addresses time-series, panel, and cohort studies as three types of longitudinal research.
 - *Time-series research*: In this type of longitudinal survey, researchers aim to measure single subject or units continually on a regular basis over time (“Time-Series Study,” 2010). By way of example, time-series data on the US birth rate demonstrate a substantial decline towards the 1940s and a spike back in the aftermath of World War 2, which is also called the “baby boom” of the post-war period (L. Neuman, 2014). As Neuman (2014) suggests, it is thus possible to see the trend and changes with the time-series that would be hard to see otherwise.
 - *Panel studies*: In some instances, panel and longitudinal studies can be used interchangeably. Nonetheless, the defining feature of panel studies lies in the emphasis on the same units. Put differently, researchers collecting data for panel surveys target precisely the same people, group, or organization at different times (L. Neuman, 2014). This type of study provides information about the given population, which helps analyze the changes over time (Lavrakas, 2008). Panel studies, therefore, are useful for the research problems on the life course, the interrelationship between life events, preferences, and behaviors, as well as the factors that influence people’s lives (Laurie, 2013). They can be preferable to cross-sectional surveys concerning their advantages to determine trends and to demonstrate the extent of the change at the individual level (Sikkel & Hoogendoorn, 2008). For instance, as

Sikkel & Hoogendoorn (2008) accentuate, researchers can investigate a particular concept such as unemployment by asking questions on employment status periodically to deliver cumulative data in time, dissimilar to cross-sectional surveys which inquire about the past. Panel studies are expensive and hard to employ as monitoring people over time is complicated owing to the deaths that can occur and the change of locations (L. Neuman, 2014). Aside from that, problems can also emanate from migration, fatigue with the research, and the response bias caused by the interviewees' experience with the study, meaning that they might cut corners to avoid detailed questions ("Panel Study," 2020). Moreover, dropouts from a panel study are called *Panel Attrition* (Lugtig, 2014). At any rate, a well-designed panel study provides beneficial results (L. Neuman, 2014).

- *Cohort Study*: The hallmark of a cohort study is that it targets people who experienced the same important life event or events within a specified period (Miller & Brewer, 2003). Cohort studies can also investigate a category of cases in a specified time (L. Neuman, 2014). Born on a specific date or time, being a university graduate in a specific year, having got divorced in a given year are the experiences and shared life events that can be subjects of a cohort study (Lavrakas, 2008). Typically, researchers will study birth cohorts and employ the questionnaire to a sample selected from that group; hence the participants forming homogenous groups, and several generations are tracked throughout their course of life for longitudinal analysis (Miller & Brewer, 2003). Further to that are the challenges that researchers encounter conducting cohort studies, which is to say, data collection from a large number of participants over long periods can be time-consuming; it requires serious financing and an intricate design (Barrett & Noble, 2019). It is also possible that, as Barrett & Noble (2019) point out, the interviewees can quit the study, and there is the risk of bias stemming from behavior change of participants' when they are cognizant of partaking in a cohort study.

One might employ other criteria to classify surveys. Stoop & Harrison (2012) have a useful framework with six more criteria other than the time notion of surveys (Who, What, By Whom, How, Where, Why). Let us recapitulate their survey classifications and examples:

- *The Target Population (Who)*: The population of inference can include households, individuals, businesses, journeys, and so forth. For instance, researchers collect data on establishments or branches in which the responses are gathered from one or many informants for *business surveys*. The UK's Workplace Employee Relations Survey (WERS) is one of the exemplars of ongoing business surveys since 1980. Moreover, in *household surveys*, the respondents are generally adults, and they commonly convey information about themselves, their households, or other household members. Providing a low-cost alternative to censuses, these studies aim to inquire about people's living conditions, their welfare, demographic characteristics, factors that impact the social and economic change in their lives (*Designing Household Survey Samples: Practical Guidelines*, 2005).

- *The Topic (What):* As Stoop & Harrison (2012) emphasize, another essential distinguishing feature of a survey is undoubtedly its topic. Any topic, such as health, political tendencies, shopping behavior, education, religion, can be the object of an investigation for researchers. The surveys that collect data on a broad range of topics are termed *omnibus surveys*, which do not have a particular topic. They are shared-cost surveys; that is to say, multiple clients pay for the survey to obtain data on their interest using a multiple-topic questionnaire.
- *Survey Agency and Sponsor (By Whom):* Governments, universities, institutions, the media, unions, non-governmental organizations, and a variety of other establishments can commission the surveys. Survey agencies can be divided into four groups: national statistical institutes, universities, market research agencies, and not-for-profit organizations. People participate more in academic or government surveys than commercial ones, which means that the response rate is affected by the survey sponsor's characteristics (Stoop, 2012). Also, survey programs can collaborate with market research firms or local agencies to conduct their projects, as is the case for the European Social Survey (ESS), which is an academic cross-sectional survey. The participating countries of ESS contract a survey agency for data collection within the country.
- *Survey mode (How):* interviewer-administered surveys carried out face-to-face or through telephone, and self-administrated surveys completed via mail and online can be marked as two types of modes. However, different classifications are also possible. Notably, multiple mode surveys that combine two or more modes for data collection gain prominence. As another example, online access panels are used mainly for commercial purposes. The survey enterprise draws the sample from the permanent pool of participants for online access panels.
- *Regional, National, Cross-National, and International Surveys (Where):* Since sampling frames barely cover multiple countries, researchers often do not have the means to conduct international surveys. Therefore, cross-national surveys, where each country has its independent sample, and the country's results are combined with that of the other countries, are more prevalent. Regional surveys investigate a particular community within a region, whereas surveys that target inhabitants of a country are called national surveys. Additionally, some surveys are classed as multi-national or multi-actor surveys. Migrant families are the target population of these types of surveys, and researchers interview them both in the sending and receiving countries.
- *Fit for Purpose (Why):* "A survey is good when it serves the purpose" is the maxim here, which means that researchers need to set their priorities amongst speed, the precision of outcomes, comparability, or other related factors that influence the survey design. Stated differently, it is often a tradeoff between accuracy, speed, and cost. Some surveys might prioritize the speed, such as exit polls, while a national survey can foreground the results' accuracy.

2.4 Survey Modes

2.4.1 Face-to-face Surveys

In terms of survey modes, face-to-face interviewing is the oldest and most popular mode, at least until recently. It is mostly associated with good data quality and is regarded as the *par excellence* of survey modes providing the fullest extent of communication between the interviewer and participant (Biemer & Lyberg, 2003). Further to that is its strength in minimizing the non-response (Lavrakas, 2008). The interviewer directly contacts the respondent and registers the answers through the questionnaire, which standardizes the interaction between them (Loosveldt, 2008). Face-to-face interviews mostly take place in the participant's natural surroundings since interviewers visit their homes or offices (Weisberg, 2005). Indeed, interviewers' task is more sophisticated than just writing down the responses; they have to accomplish the first contact procedure, endeavor to persuade the respondent, explain the respondent's role and collect the data according to the terms as Loosveldt (2008) points out. Also, the survey enterprise may send introductory letters informing the respondent about the interview, research's purpose, and procedure, which help advance the study and get better results (Lavrakas, 2008).

Face-to-face interviews can take different shapes based on the question forms and research aims. Lavrakas (2008) categorizes and describes three types of face-to-face interviewing:

- *Structured interviews*: It is the most fundamental type of face-to-face interviewing, generally employed for vast projects. It can have open-ended and close-ended questions, and interviewers ask the same questions in the same order to every participant. Quantitative studies usually apply structured interviews.
- *Semi-structured interviews*: They are usually comprised of open-ended questions. Researchers aiming to deepen the answers for particular topics or issues may want to design the questionnaires more flexibly, allowing some space for interviewers with semi-structured designs. Interviewers can also help the respondents if they lack knowledge on the topic and explore the subjects spontaneously raised by the participants in semi-structured interviews.
- *Unstructured interviews*: This type of interview almost does not have predetermined questions, which first sprouted in anthropology and sociology (Zhang & Wildemouth, 2009). Studies that utilize focus groups employ unstructured interviews wherein the interviewer or moderator lets the conversation flow (Lavrakas, 2008). That does not mean that unstructured interviews are random or uncontrolled, but the interviewer or the moderator tries to interfere minimally while encouraging the respondents to talk (Zhang & Wildemouth, 2009). The topics and issues raised during the sessions determine and direct the questions of the interviewers or moderators (Lavrakas, 2008). Semi-structured and unstructured interviews are generally used in qualitative inquiries.

It is also important to address the strengths and weaknesses of face-to-face interviews:

- The chief advantage of face-to-face interviewing is related to direct contact between the interviewer and interviewees (W. L. Neuman, 2012). As Neuman (2012) underscores, direct contact enables further cooperation and motivation. Interviewers can employ some tactics to gain more cooperation (Biemer & Lyberg, 2003). In association with that, interviewers can clarify the questions more easily and probe the answers if need be (Lavrakas, 2008). What is more, face-to-face interaction can increase the rapport and trust between the respondent and

the interviewer. This can enable the respondents to better engage with the questions and give more accurate and honest answers (W. L. Neuman, 2012)

- It has the highest response rates since it can be more difficult to refuse a person waiting at the doorway (W. L. Neuman, 2012). Also of note is that people with lower incomes, lower levels of education, and minorities are more likely to participate in face-to-face interviewing compared with other modes (W. L. Neuman, 2012).
- Interviewers can also supply the interview with visual aids such as calendar displays, lists, and pictures (Schröder, 2016). As Schroder (2016) notes, longer interviews are possible through face-to-face mode, which is advantageous for detailed investigations.
- Face-to-face interviewing is often linked with higher costs since interviewers have to travel a lot, visit the respondent at home and complete a relatively low number of interviews in a day (Biemer & Lyberg, 2003). Travelling may also require financing for the survey team's accommodation in different places. In addition to that, the training of interviewers requires more staff and time; the data collection period is also longer, and it requires more supervision and staff. That is to say, survey administration is more sophisticated than other modes, and this complexity increases the costs as well (W. L. Neuman, 2012).
- The interviewer's presence is not invariably advantageous. As Biemer & Lyberg (2003) point out, social desirability bias can be a concern for face-to-face interviewing since respondents may tend to give socially desirable answers. For instance, people underreport sensitive topics or legal issues such as drug use lest they face legal proceedings. They may also avoid being on extreme sides for political tendencies, or respondents might want to be perceived as more knowledgeable or up-to-date than they actually are (Biemer & Lyberg, 2003). People tend to give normative answers and hide embarrassing events or details (W. L. Neuman, 2012). Further to that, interviewer bias can also occur when specific characteristics of the interviewer (such as sex) influence the response behavior of particular groups (Schröder, 2016). Interviewers also have specific effects on responses, given that every interviewer has different personality traits, conduct, and style, which may also be a source of error (Biemer & Lyberg, 2003).

Conventional face-to-face interviewing with paper and pen becomes outdated with the advent of CAPI (Computer-assisted Personal Interviewing). After the 1980s, survey operations began to adapt CAPI systems; as the computer technology rocketed with time, more and more surveys began to employ it (Lavrakas, 2008). The questions appear on the screen, and the interviewer records responses through the software run on the tablet or laptop.

Whether and how far CAPI has the edge over PAPI is also a major site of discussions. Bhatia and Jaiswal (2016) argue that CAPI is more advantageous on cost-effectiveness, interviewer's productivity, days spent, and the average number of errors per interview. Notwithstanding its increased fixed costs, CAPI is preferable for its fewer variable costs (Rahija et al., 2017). By contrast, Rahija et al. (2017) also point out that CAPI interviews took longer in their study, most probably due to equipment problems. Interviewers can have more control in the rhythm of PAPI questionnaires because they can more quickly pass to the next question, reading it while recording the response of previous questions (Lavrakas, 2008). The counter-argument is that some organizations prefer CAPI over PAPI, given its shorter fieldwork and lower-costs (Biemer & Lyberg, 2003). As regards data quality, some scholars doubt CAPI's advantage. Nicholls et al. (1997) assert that CAPI is not a panacea for measurement since the positive contribution of CAPI to data quality is not as significant as expected.

2.4.2 Telephone Surveys

The telephone survey is still one of the most common modes of probability sample surveys in spite of the rapid advancements of other technologies (Vehovar et al., 2012). It was first used during the 1930s in the US when telephone penetration was 30 percent (Vehovar et al., 2012). Indeed it primarily used for follow-ups until the 1970s when the main modes were mail and face-to-face interviewing (Tucker & Lepkowski, 2008). The penetration rate increased to 90% in the US in the early 1970s, and telephone surveying became much more useful (Vehovar et al., 2012). Further, *International Conference on Telephone Survey Methodology* in 1987 was a cornerstone for telephone surveys as it stimulated new studies on the field (Tucker & Lepkowski, 2008). Nowadays, computer-assisted telephone interviewing (CATI) is widely used in the survey sector, enabling researchers to display questions on computer and record answers accordingly (Lavrakas, 2008).

To select the samples for telephone surveys, researchers use lists or random digit dialing (RDD)¹³(Steeh, 2008). As previously noted, the spread of mobile phones caused a dramatic decline in landline phone usage, and it posed a significant challenge for telephone surveys (Dillman et al., 2014). Mobile phones are portable, and people can carry them all day long, necessitating different calling procedures for survey purposes (Dillman et al., 2014). Researchers can utilize three approaches for mobile phone sampling, as Himelein et al. (2020) demonstrate:

- Using a representative baseline survey (usually face-to-face) that had a high response rate
- Obtaining a telephone-number list from a telecom company or private firm
- Employing Random Digit Dialing (RDD)

Let us consider the advantages and challenges of telephone surveying:

- Compared with face-to-face and mail modes, telephone surveys cost less and are quicker, which is of utmost importance for many survey enterprises (Lavrakas, 2008). With telephone surveys, the operation is usually centralized, which eases general administration, feedback, training, supervision, and other procedures (Biemer & Lyberg, 2003).
- Contact with an interviewer might positively affect response rates, as is the case for face-to-face interviewing (Lavrakas, 2008). It means that interviewers can elicit some degree of cooperation and rapport, albeit lower than face-to-face interviewing.
- Surveying with complex questions asking respondents to perform calculations or review records will be challenging for telephone surveys (Biemer & Lyberg, 2003). Also, as Biemer & Lyberg (2003) underline, visual aids cannot be used in telephone surveying, or they should be sent beforehand, which is not very feasible.
- Telephone surveys may also be subject to social desirability bias and interviewer variance, given the communication with an interviewer, although its relatively lower (Biemer & Lyberg, 2003)
- Landline telephone users now have various new technologies to block callers and answer machines, making it difficult for survey researchers to reach respondents (Dillman et al., 2014).
- Perhaps the greatest challenge for telephone surveying has been the dramatic decline in landline phone coverage. However, near-ubiquitous cell phone usage provides further

¹³ Computers generate digits in random digit dialing (RDD) method whereby researchers select area codes and prefixes in a representative way, and they can also target specific places of interest by intentionally drawing specific area codes (Lavrakas, 2008).

opportunities to reach people wherever they are (Kempf & Remington, 2007). Still, 30 percent of the world population do not have mobile phones, and mobile phone coverage in rural areas is lower than in urban areas (Himelein et al., 2020).

2.4.3 Mail Surveys

Using postal mail became outdated in the developed world with the advancements in technology such as e-mails, texts, apps, and social media (Dillman et al., 2014). Using postal mail for survey purposes, therefore, is also challenged with the advent of alternative ways to collect data. Still, mail surveys attract researchers, given that they are cheaper than interviewer-administrated modes. It is also simpler to operate since individuals and institutions can implement mail surveys without depending on research organizations (Dillman, 1991). Researchers generally utilize telephone directories or other lists for sampling procedures, as mail surveys require a solid sampling frame of names and addresses (de Leeuw & Hox, 2008). Respondents receive the survey packet containing the mail questionnaire, a cover letter that elucidates the study's purpose that promotes participation, a postage-paid pre-addressed return envelope, and a participation gift or a possible incentive (Lavrakas, 2008). As Lavrakas (2008) marks, mail surveys usually send some kind of an advance letter, card, or notification before communicating for the actual questionnaire.

Mail surveys have certain advantages and disadvantages:

- Mail surveys are inexpensive, given that interviewer-assistance pushes up the costs considerably (Biemer & Lyberg, 2003). Also, a smaller staff can operate, which is a huge advantage (de Leeuw & Hox, 2008). That said, it is no longer the low-cost king since the internet and web-based surveys offer better alternatives for spendings (Lavrakas, 2008).
- Mail surveys can involve visual aids and stimuli that can facilitate answering complex questions (de Leeuw & Hox, 2008). The questionnaire and packet can include graphics, charts, maps, information booklets, product samples, which adds flexibility to conduct mail surveys (Cowles & Nelson, 2015; de Leeuw & Hox, 2008).
- It is common knowledge that mail surveys generate a low response rate. (Biemer & Lyberg, 2003). Nonetheless, it can yield higher response rates than expected, with professional and diligent implementation and accurately targeted sample populations (Lavrakas, 2008).
- Participants can answer sensitive questions more comfortably since there is no interviewer; also, the interviewer's negative effect that stems from his/her characteristics is therefore eliminated (Cowles & Nelson, 2015).
- Mail surveys also offer more time, space, and privacy to respondents, and they can also consult the records if needed (de Leeuw & Hox, 2008). As de Leeuw & Hox (2008) underscore, mail survey questionnaires must be simpler and shorter than that of face-to-face or telephone interviews.
- The absence of an interviewer is also a drawback. Some elaborate questionnaires may require interviewer-assistance, instructing, and probing, which require mail surveys to be self-explanatory (de Leeuw & Hox, 2008).
- Another important disadvantage is the turnaround time, as mail surveys necessitate long field periods, it can take several weeks to receive the questionnaires back (Biemer & Lyberg, 2003; de Leeuw & Hox, 2008).

2.4.4 Online Surveys

Notwithstanding the deep digital divides across the world, online technologies have permeated all fields of life; hence online research methods have become increasingly important in the social research realm (Lee et al., 2017). With the World Wide Web (WWW) invention, researchers began to utilize the internet for data collection (Manfreda & Vehovar, 2008). At present, people are almost ceaselessly connected to the internet with mobile phones, tablets, and wearable technologies, and online-surveys adapt themselves to a variety of devices (Dillman et al., 2014).

It is also elemental to refer to the difference between web-surveys and e-mail surveys (Mesch, 2012). Respondents access web-based surveys that reside on a server through web-browsers, whereas e-mail surveys sent to the respondent generally include a text and attachment with an invitation letter (Mesch, 2012). Online surveys make use of what the internet offers: global access to people, which enables researchers to design surveys collecting data from different countries and geographically dispersed samples (Mesch, 2012). While low administration costs are among the advantages of e-mail surveys, the coverage of e-mail surveys is among the weaknesses, as Miller & Brewer (2003) argue, which continues to be the case.

Online surveys have other strengths and weaknesses:

- Internet access is common for higher socioeconomic status despite the rapid spread to the general population, which constraints online surveys (Lavrakas, 2008). Further to that is the lack of a standard or single sample frame for e-mail and web users (Lavrakas, 2008).
- Respondents' lack of internet literacy or expertise might be a problem, and it can also bring up privacy and security concerns for respondents, such as the confidentiality of their answers (Evans & Mathur, 2005).
- Like mail surveys, it gives respondents time to think on the questions, consult, and check the records (Fowler, 2014).
- Social desirability bias is less of a concern for web surveys, and participants are more likely to deliver honest answers for sensitive questions, which will positively influence some types of measurement error (Cowles & Nelson, 2015). Besides, as Cowles & Nelson (2015) point out, the absence of an interviewer will reduce the errors stemming from the interviewer's characteristics and personal traits.
- Not having interviewers, no doubt, also has adverse effects on several aspects related to cooperation, rapport, guidance, and probing (Fowler, 2014).
- Differences in web-browser usage or screen settings may alter how the participants see the questions, or potential respondents cannot partake in the surveys because of their not up-to-date system specs or inadequate software, which causes specific web-survey errors (Lavrakas, 2008).

2.4.5 Mixed-Mode Surveys

Mixed-mode was nascent at the beginning of the 1960s; thereafter, researchers increasingly choose to combine more than one method for data collection to improve the coverage, increase response rates, and decrease costs (E. D. De Leeuw & Toepoel, 2018). By way of example, the 1984 British Labor Force Survey employed telephone surveys in a quarterly panel design (E. De Leeuw, 2005). The proliferation of new technologies such as the internet, mobile phones, and other devices fueled mixed-mode surveys' rising trend (Shine & Dulisse, 2012). Participation to a survey can be through different

data collection modes in mixed-mode surveys, and different methodologies can be employed in distinct stages, such as in recruitment, screening, and questionnaire administration phases (Lavrakas, 2008).

Applying mixed-modes helps to balance the weaknesses of each mode; for instance, combining telephone survey with mail surveys can provide flexibility in question design as well as more comfort for the survey team, given that mail surveys are less flexible in questions and telephone surveying necessitates more work for the survey staff (Shine & Dulisse, 2012). Moreover, reducing measurement errors is another impetus to conduct mixed-mode surveys, such as for sensitive questions. Researchers may use appropriate modes for specific groups in a mixed-mode survey not to embarrass respondents, thereby eliminating the social desirability effect (Dillman et al., 2014). Dillman et al. (2014) also have an account of three overarching strategies for the ways modes can be combined:

- Use of multiple *contact* modes to encourage response for data to be collected by a single response mode
- Use of multiple *response* modes to collect respondent answers while using only one mode of contact
- Use of multiple *contact* and *response* modes for the same study

Furthermore, researchers have to consider some issues when deciding on a mixed-mode survey, as Lavrakas (2008) delineates:

- Researchers must analyze the target population's characteristics, such as their telephone access, literacy level, and internet penetration.
- Another point is to decide if different modes will be employed sequentially or concurrently.
- Every mode has potential drawbacks, and researchers have to be cognizant of them.
- Researchers should endeavor to decrease the measurement errors in any (E. D. De Leeuw & Toepoel, 2018) stages.

A recent trend is doing mixed-device surveys in which respondents can choose the device to complete the online survey (E. D. De Leeuw & Toepoel, 2018). In mixed-device surveys, researchers have to adapt the survey format for each device given that PCs, tablets, mobile phones have different specs such as screen sizes and processing capacity (E. D. De Leeuw & Toepoel, 2018).

2.5 International Survey Programs

International Survey Programs grew in number in the twentieth century with the need for comparable statistics across countries. The branches of international organizations as *The United Nations Statistics Division* and *Eurostat* were founded to provide high-quality statistics and develop standards and norms for statistical activities. Efforts to standardize indicators also influenced the methods and questionnaires within various programs. And now, many international survey programs collect data and publish comparable results between countries:

- *DHS (Demographic and Health Surveys)*: DHS are nationally-representative household surveys that provide data on population, health, and nutrition. *Standard DHS Surveys* are conducted every five years, while surveys employed between rounds of DHS are called *Interim DHS Surveys*, which have smaller samples and shorter questionnaires. ICF International implements, and the United States Agency for International Development (USAID) finances the project. It is active in over 90 countries.

- *Multiple Indicator Cluster Surveys (MICS)*: MICS are the international survey initiative of UNICEF that aims to collect and analyze data for children and women's situation. More than 60 countries conducted MICS within the framework of the first round. Now that UNICEF assists for more frequent intervals, which is every three years instead of five years.
- *European Social Survey (ESS)*: Began in 2002, ESS covers more than 30 countries, and the surveys investigate Europe's changing institutions, beliefs, and patterns of different populations. It is a biennial survey.
- *International Social Survey Programme (ISSP)*: Established in 1984, the ISSP is a cross-national collaboration program that conducts surveys every year on diverse social sciences topics.
- *World Values Survey (WVS)*: Started in 1981, the World Values Survey explores people's values and beliefs in around 100 countries. The seventh wave of the WVS continues.
- *Survey of Health, Ageing, and Retirement in Europe (SHARE)*: SHARE is conducted in 27 European countries plus Israel. It is a multidisciplinary and cross-national panel database of microdata on health, socio-economic status, and social and family networks.
- *European Values Study (EVS)*: Being cross-national and longitudinal survey research, the EVS focuses on basic human values. It is conducted every ten years.
- *Afrobarometer*: It is conducted in over thirty countries in Africa for data on democracy, governance, the economy, and society.
- *AsiaBarometer*: The comparative survey on the daily lives of ordinary people living in Asia, focusing on family relations, neighborhood, social and political institutions, and the markets.
- *Latinobarometro*: Public opinion survey conducted in 18 Latin American countries on the development of democracy, attitudes, behaviors, and values.
- *Household Finance and Consumption Survey (HFCS)*: It is conducted at the national level to gather data on households' finances and consumption in the euro area.
- *European Community Household Panel (ECHP)*: the ECHP is a Harmonised cross-national longitudinal survey focusing on household income and living conditions.
- *European Union Labor Force Survey (EU LFS)*: It is on labor participation of people aged 15 and over, conducted in all European Union member states.
- *European Union Statistics on Income and Living Conditions (EU-SILC)*: It was launched in 2003 to collect cross-sectional and longitudinal multidimensional microdata on income, poverty, social exclusion, and living conditions. EU-SILC now covers all EU countries with Iceland, Norway, Switzerland.
- *Generations and Gender Survey (GGS)*: Launched by United Nations Economic Commission for Europe, GGS is a series of panel surveys on families, life course trajectories, and gender relations.

2.6 Surveys in Turkey

The afore-discussed history of surveys mostly derives from the developed countries, particularly the U.S.A. One must examine the historical line of surveys in under-developed or developing countries through different lenses as they face many other challenges and might not follow the same trends of the developed countries. By the same token, the advancement of surveys in Turkey should be evaluated in a different sense.

Turkey's history of the surveys leads back to the Ottoman Era. The first census of the era was undertaken in 1831 during the reign of Mahmud II. The census was to count men in the country to

assess military capacity and tax regime (Karpat, 2010). In the imperial era, other censuses and enumerations were pursued, even though they were not comprehensive and did not conform to the scientific standards (Fişek, 1986). Importantly, the *Central Council of Statistics (Merkezi İstatistik Encümeni)* was established in 1891. It was re-established in 1926 with the name the *Central Department of Statistics (Merkezi İstatistik Dairesi)*, which was later reorganized under the *General Directorate of Statistics (İstatistik Umum Müdürlüğü)* in the republican era¹⁴. Besides, the first census of the republican period was conducted in 1927. The following censuses were employed every five years, starting with the second census in 1935 until 2000. Conventional methods for the national census are no longer in use with the advent of the *Address Based Population Registration System*.

Demographic research is a prime mover for the development of the survey sector in Turkey. The 1960s were significant in that these years were marked by the proliferation and development of demographic studies in the country (Özbay, 1986). Ministry of Public Health and Welfare was the first to employ demographic surveys. One of the most significant surveys was on the information-attitude-use about family planning in 1963 that the ministry pursued in collaboration with the Population Council (Fişek, 1986). Also, the ministry conducted *The School of Sanitation Turkey Demographic Research* in 1965, while the State Planning Organization carried out *Village Research* in 1962 and 1968 (Timur, 1971). Besides, it is not by chance that flourishing demographic studies, the foundation of *Hacettepe University Institute of Population Studies (HUIPS)* in 1967, and official policy changes that prioritize contraceptive methods were in the same period (Özbay, 1986). Ford Foundation donated over 1 million dollars to HUIPS during its establishment, and later, the institute also had financial support from UNFPA, Pathfinder, and Westinghouse (Fişek, 1986). HUIPS, in the years ahead, represented the pinnacle of the demographic studies in Turkey.

Market Study Consultancy Research (Piyasa Etüd Müşavirlik Araştırma) or PEVA is the earliest market research firm of Turkey, which was founded in 1961 (Göksu, 2018). After that, the research companies PİAR and SİAR were established subsequently in 1975 and 1981, and they also conducted inquiries on political inclinations for their clients, including political parties (Göksu, 2018). It is safe to say that marketing research and political polls gained more prominence with globalization and Turkey's further engagement in the market economy while political parties were more in need of feeling the pulse of people. Nowadays, numerous international and national research companies operate in Turkey.

It is to be noted as a sign of institutionalization of Turkey's survey sector that the *Turkish Researcher's Association (TÜAD)* was founded in 1988 by Nezih Neyzi, who was also the founder of PEVA. It is a professional association that works on national and international professional principles, laws, and quality control measures. Today, the association has many members, including research firms and individuals.

2.6.1 Demographic Survey Series and Other Recent Surveys of HUIPS

Hacettepe University Institute of Population Studies (HUIPS) is the main implementing agency of demographic studies in Turkey, being the primary provider of demographic data with nationwide and small-scale research. The hallmark of the institute is the demographic and health survey series, which

¹⁴ The name of the institute was changed to *State Institute of Statistics (Devlet İstatistik Enstitüsü)* in 1962 and finally, in 2005, it was renamed as *Turkish Statistical Institute (Türkiye İstatistik Kurumu)*.

begun in 1968 and is conducted every five years. To date, there are eleven surveys in the series with various titles:

- 2018 Turkey Demographic and Health Survey
- 2013 Turkey Demographic and Health Survey
- 2008 Turkey Demographic and Health Survey
- 2003 Turkey Demographic and Health Survey
- 1998 Turkey Demographic and Health Survey
- 1993 Turkey Demographic and Health Survey
- 1988 Turkish Population and Health Survey
- 1983 Turkish Population and Health Survey
- 1978 Turkish Fertility Survey
- 1973 Survey on Population Structure and Population Problems in Turkey
- 1968 Survey on Family Structure and Population Problems in Turkey

The HUIPS began to implement the survey within the *Demographic & Health Surveys (DHS) (The DHS Program)* framework in 1993, although the institute has undertaken the survey enterprise with national financial sources since 2003. More precisely, the *2018 Turkey Demographics and Health Survey (TDHS)*, the latest survey of the series, was funded by the *Scientific and Technological Research Council of Turkey (TÜBİTAK)* within the scope of the “1007 Support Programme for Research and Development Projects of Public Institutions”. The 2018 TDHS data and results are fully comparable with the *Demographic and Health Surveys (DHS)*, and ICF International Inc. provided support on data processing, tabulation, and the review of the final report.

What makes 2018 TDHS highly distinctive from its predecessors is the component of the Syrian Migrant Sample. The growing influx to the country reinforced the need for data on basic demographic and health indicators of the Syrian Migrant Population in Turkey. From this vantage point, HUIPS designed a separate sample for the Syrian migrants, albeit bereft of a frame that involves each Syrian household in Turkey¹⁵. To address the issue, the survey team developed a sample design hinged on each quarter's population size, whereas the camp/non-camp population variable lay the basis for the stratification. Besides, the survey team applied CAPI (Computer-assisted Personal Interviewing) for the first time in the survey series' history in Turkey.

There have been many changes and improvements in retrospect, apropos of the survey design. For instance, the survey started to use the *National Address Database System* as the sampling frame with 2008 TDHS. Before, population censuses were utilized as the sampling frames. The establishment of the *Address Based Population Registration System* in 2007 paved the way for using the current frame, which is linked to the system. Another marked development was the changes in the stratification as regards the NUTS system. Surveys henceforth are designed to deliver results accordingly for the NUTS-1 level, with the 2003 TDHS.

HUIPS has conducted numerous other studies and surveys. By way of illustration, the Institute carried out *Research on Profiles and Living Conditions of Foreigners Legally Residing in Turkey* in 2016, which aimed to collect data on the immigration and residence process, hardships and information about

¹⁵ For further information, 2018 Turkey Demographic and Health Survey Syrian Migrant Sample Main Report is available at http://www.hips.hacettepe.edu.tr/eng/tdhs2018/TDHS_2018_SR.pdf

residence permits, adaptation to the country, health services as well as plans for the future of foreigners residing legally in Turkey. Conducted in 2015, *Socio-economic Profiling of Asylum Seekers Residing in Satellite Cities in Turkey* that investigates the pre-migration status of asylum seekers living in satellite cities, the process of migration and resettlement, the conditions of living in satellite cities with the future plans, and *An Analysis on Afghan Nationals Living in Turkey, Who are International Protection Applicants, and Those Under International Protection: Their Reasons for Arrival and Stay in Turkey, Their Future Plans and Goals* that inquired about the motivations for coming to Turkey, the processes of staying in Turkey and the future plans of the holders of international protection applications and status of Afghan nationals in Turkey are the amongst studies on migration conducted by the Institute. Turkish Ministry of Interior Directorate General of Migration Management was the financier of these two research.

In 2008 and 2014, HUIPS implemented the *National Research on Domestic Violence Against Women in Turkey* with a nationwide sampling to obtain information on the prevalence and consequences of violence against women and awareness of the law on combating violence against women. An international consortium, including HUIPS, conducted the 2008 research and the European Commission was the financing source. The 2014 research, on the other hand, was conducted by HUIPS with the financial support of the Turkish Ministry of Family and Social Policies General Directorate of the Status of the Women, which was also the beneficiary institution.

Another important HUIPS enterprise is *Turkey National Maternal Mortality Study* with its distinguished design. Executed between 2004 and 2006, The Survey attempted to gauge maternal mortality rate and ratio by communicating information with an eye to improve existing registration and notification systems. An international consortium, including HUIPS, put the project into practice within the framework of the “Turkey Reproductive Health Programme.” The research utilized the RAMOS (Reproductive Age Mortality Study) data collection strategy, which is to identify and investigate the causes of death of all women of reproductive age. For that, the survey team collected data from primary informants which included cemetery officials and village headmen (mukhtars). The information concerned deceased people among their relatives. The survey also gathered information about the home address and the place of death of women aged 12-50 years. Having received the information from primary informants, the District Coordinator, alongside Project Provincial Team, reviewed the health facility records and performed verbal autopsies. The study design makes it possible to apply analysis for Turkey overall, urban and rural areas, and at the NUTS-1 level. Also, the study has a qualitative component ¹⁶.

2.6.2 Recent Surveys of the Turkish Statistical Institute (TurkStat)

TurkStat operates numerous surveys on a multitude of topics. Participation in these surveys is a legal obligation for Turkish citizens ¹⁷. Hence, TurkStat’s surveys are not on a par with the private enterprises in terms of the response rates. Let us exemplify some of the prominent studies of TurkStat in this part.

¹⁶ The main report of Turkey National Mortality Study 2005 is accessible at http://www.hips.hacettepe.edu.tr/eng/maternal_mortality/NNMS-2005_main_report.pdf

¹⁷ According to Article 54 of the Turkey Statistics Law (Law Number: 5429), citizens who do not participate without giving any excuse in TurkStats’ enumeration and surveys are subjected to fine.

Income and Living Conditions Survey is one of the prime surveys of TurkStat in which the institute aims to obtain data on income distribution, living conditions, and poverty based on income¹⁸. The first implementation of the survey was in 2006, as per the European Union Compliance Programme. Based on a panel design, the survey monitors the respondents during the course of four years with conducting the fieldwork every year; to obtain both cross-sectional and panel data annually. Following the year 2014, the survey had a larger sample size with the design that enables analysis for NUTS-2 level based on the two-stage stratified cluster sampling approach. TurkStat uses *National Address Database* and *Address Based Population Register System* as the basis for the sampling frame. It is a household survey, and individuals are monitored even if they move to obtain data from the same persons within the framework of the panel design. As for the data collection, the institute employs CAPI while also utilizing CATI for follow-ups.

Second, the *Household Labour Force Survey* stands for a significant TurkStat study that has a long-standing background. First conducted in 1988 pursuant to the International Labour Organization's norms and standards, the survey investigates the labor force structure by gathering information on economic activity, occupation, employment status, hours worked, and duration of the employment and the jobs that unemployed people seek¹⁹. It has a two-stage stratified cluster sampling method with a rotation pattern to ascertain a fifty percent overlap between two consecutive periods and during the same period of the two successive years. Similarly, the study uses the *National Address Database* as the source for the sampling frame while delivering the estimations at NUTS-1, NUTS-2, and Turkey as a whole. Crucially, the survey can be considered a semi-panel design given that the households are visited four times in 18 months; however, what makes it "semi" is that respondents are not tracked if they move from their residence. Meaning, the participants do not have to be the same persons in the consecutive rounds. It is also important to note the survey's reference periods, which are all-52 weeks of the year, and field teams complete data collection within 15 days after the reference week. Interviewers utilize the CAPI method for the data collection.

Having been carried out annually since 2002, TurkStat's *Household Budget Survey* is one of the principal sources for the data on socio-economic structures, standards of living, consumption patterns of the households²⁰. The study is instrumental in many respects; for determining the consumer price index, monitoring consumption structure, compiling data for national income and minimum wage calculations, adjusting the poverty threshold, as well as investigating the standard of living of the people. In a similar vein to the above-described TurkStat studies, the *Household Budget Survey* is also hinged on the *National Address Database* for the sampling frame, and it employs the stratified two-stage cluster sampling method taking Turkey as a whole as the estimation level. Householders record consumption expenditure for one month, amidst two visits of field teams, and interviewers inquire about the employment and income levels using laptops.

First implemented in 2006, the *Family Structure Survey* sets an example for the institute's demographic surveys. TurkStat conducted the third and most recent survey of the series in 2016, within a bilateral

¹⁸ Information on Income and Living Conditions Survey is available at http://www.tuik.gov.tr/MicroVeri/GYKA_Panel_2018/english/index.html

¹⁹ Information on the Household Labour Force Survey is available at http://www.tuik.gov.tr/MicroVeri/Hia_2019/english/meta-data/tarihcce/index.html

²⁰ Information on the Household Budget Survey can be accessed at http://www.turkstat.gov.tr/MicroVeri/HBA_2018/english/meta-data/introduction/index.html

framework, the Ministry of Family and Social Policies representing the other side. The structure of families in Turkey, lifestyles of the individuals in the family environment, and values regarding family life are matters of the survey²¹. Notably, the survey targeted households in Turkey that have individuals aged 15 and over, which amounted to a sample size of 17239 households for the 2016 implementation, and ultimately, field teams interviewed 35475 individuals. Turkey as-a-whole, NUTS-1 level, and three major provinces (İstanbul, İzmir, Ankara) are the estimation domains for the survey that holds a two-stage stratified cluster sampling method approach.

Another relevant demographic study is *Time-Use Survey*, which for the first time, was realized in 2006 in consonance with the EUROSTAT's efforts on comparable international statistics on time-use. TurkStat applied the survey for the second time between 2014-2015 with a two-stage cluster sampling method, delivering estimations on Turkey as a whole²². The survey seeks to gather data on how the individuals use their time during the day, examine time-use alterations amongst different groups of people, and use the data for the calculations of gross-domestic-product. Targeting the households with individuals aged over 18, the 2016 Time-Use Survey had a sample size of 11440 households, 9073 of which are interviewed.

Life Satisfaction Survey is another demographic research of Turkstat, whose core objectives are to investigate the happiness perception of the individuals, the societal values of the people, the satisfaction of the public services, and trends in satisfaction levels²³. The institute conducts the study regularly since 2003, and the sample was designed to carry out analyses at the NUTS-3 level in 2013. For the 2014-2018 period, the survey produced estimates for Turkey as-a-whole. With the two-stage stratified cluster sampling, the survey had a sample size of 5410 households, while field teams interviewed 9719 individuals in 2018.

TurkStat also has demographic and population studies based on enumeration. In order to gather information on the social and economic characteristics of the population and buildings and dwellings, TurkStat conducted the *Population and Housing Census* in 2011²⁴. This survey aimed to expand the data that was impossible to compile from the Address Based Population Registration System by investigating the information on the labor force, employment, migration, disability, and dwelling qualities. Survey teams enumerated institutional places besides interviewing around 2.2 million households in selected enumeration areas. For that, 22.861 enumeration areas were selected from the total number of 195.192 enumeration areas.

TurkStat's *Research and Development Activities Survey in the Business Enterprise Sector* constitutes an example for establishment surveys²⁵. It attempts to collect information on research and development, labor force, financial sources, and expenditures by using them to possess internationally comparable

²¹ Details on Family Structure Survey is available at http://www.tuik.gov.tr/MicroVeri/Aile_2016/english/meta-data/purpose/index.html

²² Information on Time-Use Survey is available at http://www.tuik.gov.tr/MicroVeri/ZKA_2014/english/meta-data/tarihcce/index.html

²³ Information on Life Satisfaction Survey is accessible at http://www.tuik.gov.tr/MicroVeri/YMA_2019/english/index.html

²⁴ Information on Population And Housing Census is available at http://www.tuik.gov.tr/MicroVeri/NKA_2011/english/index.html

²⁵ Establishment surveys investigate organizations for their behaviour, structure or output; which helps to understand trends in the economy (Lavrakas, 2008)

data²⁶. Covering enterprises that undertake R&D activities, work in Technology Development Zones, have direct and indirect support from governmental agencies, and are on TOP 500 lists in terms of turnover and value-added, the survey gathers the data via web-surveys in which users respond to the questionnaire by signing in with the user name and password assigned by TurkStat.

Furthermore, TurkStat gathers statistics in many other sectors. As an example, for fishery statistics, the institute employs complete enumeration for large-scale fishers and sampling methods for marine fishery surveys²⁷ to compile data for evaluation of the fishing sector in the country. In 2019, 3883 seasonal-small-scale, 2327 monthly-large scale, 6391 annually marine fishery surveys were applied. Another agricultural data source is poultry production statistics that TurkStat began to collect after 2010 under the frame of the European Union Accession Period²⁸. For compiling the data, enterprises fill the forms on TurkStat's data entry system on the web every month. Along similar lines, the businesses enter their data on the TurkStat's system on the web for milk and milk production statistics that first gathered in 2010²⁹. An average of 1684 enterprises take the milk and milk products survey monthly.

2.6.3 Examples of Recent Surveys of the Research Companies in Turkey

A multitude of internationally renowned research companies, national private research firms pursue surveys in Turkey. For most of the marketing surveys, the methodology and other details of the surveys are not available or not open to investigation on the internet pages of the firms. Here, it is pertinent to note some survey examples from firms and institutions that carry out political and sociological research to have insight into their methodology.

Konda, a research company that conducts surveys mainly on political and sociological matters, has the series called *Konda Barometer*, investigating people's political and sociological preferences and tendencies since 2010 March. Implemented in the first week of every month since then, field teams conduct between 1800-3600 face to face interviews, and the subscribers get the results at the end of the month³⁰. The topics include law and justice, usage of the internet and social media, immigration, happiness, security, alcohol consumption, foreign policy etc. Another example is the *Lifestyles Survey* that Konda performed in 2008, 2015, and 2018. For the last survey, 5796 individuals were interviewed between March 31 and April 1, 2018, targeting people aged 15 and over (*Konda Türkiye'de Toplumsal Cinsiyet Raporu: Hayat Tarzları 2018 Araştırması*, 2019). As the report of the survey indicates, the field teams conducted interviews in 36 provinces of Turkey. Further, the sampling is based on 12 regions, and the sample is layered with the data of Address Based Population Registration System data and General Elections. They applied quotas for age and sex for the six interviews in every quarter.

For the *Popular Culture* survey, of which the fieldwork was completed between October 2 and 3 in 2019 (*Konda Barometresi Temalar: Popüler Kültür*, 2019), Konda interviewed 2699 people in 29

²⁶ Information on Research and Development Activities Survey in the Business Enterprise Sector http://www.tuik.gov.tr/MicroVeri/ARGE_2018/english/index.html

²⁷ Information on Fishery Statistics is accessible at http://www.tuik.gov.tr/MicroVeri/SUI_2019/english/meta-data/purpose/index.html

²⁸ Information on Poultry Production Statistics is available at http://www.tuik.gov.tr/MicroVeri/KHU_2019/english/index.html

²⁹ Information on Milk and Milk Products Statistics is available at http://www.tuik.gov.tr/MicroVeri/SUT_2018/english/meta-data/purpose/index.html

³⁰ Information is available at <https://konda.com.tr/tr/konda-barometresi/>

provinces, and the sampling is designed using the Address Based Population Registration System and the 2015 General Election Results. The design also includes age and sex quotas for every 18 interviews in every quarter.

Metropoll, another research company that concentrates on political and social tendencies, regularly shares the results of *Turkey's Pulse* series with different themes. Before the pandemic, the company employed face-to-face surveys such as for *Turkey's Pulse* December 2019 research by interviewing 1783 people in 28 provinces between December 14 and 22 (*Türkiye'nin Nabzı: Medya Demokrasi*, 2019). In addition, Metropoll conducted face-to-face interviews with 1926 people in 28 provinces within one week for the January 2020 version of the series (*Türkiye'nin Aşk Raporu*, 2020). These surveys are at the NUTS-2 Level employing stratified sampling with weighting, as the reports define.

Moreover, *Istanbul Economic Research* also conducts surveys and provides services on big data. In their July 2020 *Turkey Report*, which inquires about social and political leanings, they completed 1537 interviews in 12 provinces through CATI (*Türkiye Raporu: Rapor 17*, 2020). The report underscores that the 12 provinces represent the NUTS-1 level of the country. Further, Kadir Has University has several survey projects. For the 2019 version of *Turkey's Tendencies Research* that attempts to identify the perspective of people on ongoing and future problems of Turkey, the university conducted 1000 face-to-face interviews between November 25 and December 13, 2019, in 26 provinces targeting people aged 18 and over (*Türkiye Eğilimleri - 2019*, 2020). The report of the research also informs that the sampling is representative of Turkey at the NUTS-2 level.

3. Surveys at the pandemic

The COVID-19 pandemic has affected the World in many ways, and survey research is no exception. As Sastry, McGonagle & Fomby (2020, p. 245) put it, “COVID-19 is of course, remarkable in its national and global scope and its open-ended duration”. In this section, we attempt to summarize this effect through different types of research. Doing so, we exploit a vast amount of information from the special issue of the Survey Research Methods Journal published July 2020, which was a very timely and valuable contribution to the field. The other major source of information is online Webinars, organized by various associations and institutions, namely the World Association for Public Opinion Research (WAPOR), American Association of Public Opinion Research (AAPOR), Population Association of America (PAA), Royal Statistics Society, Turkish Medical Association and some others³¹. We also looked at the websites of some surveys for pandemic-related announcements, as well as the websites of some international organizations (WHO, UN, EUROSTAT) and private research companies.

This chapter begins with a review of the effects of COVID-19 on survey research. As deLeeuw also underlined in her speech in the July 2020 Webinar by WAPOR, here we make a distinction between cross-sectional and longitudinal surveys, provided the challenges faced differ by these two types. Another topic in the agenda for survey methodology and the pandemic is the need for new data, both in terms of how people’s attitudes, behaviors and experiences in their lives are affected due to the pandemic, and in terms of disease prevalence or incidence. We thus have separate sections on surveys that focus on the effects of the pandemic, innovations in surveys and methodologies that appeared over the course of the pandemic, and surveys designed to measure prevalence and incidence. We also have a separate section dedicated to developments in Turkey at the end of the chapter.

3.1 Adjustments that had to be done to ongoing surveys during COVID-19

Although the biggest impact of the pandemic was on face-to-face surveys during the pandemic, we expected some degree of impact on most surveys. Telephone surveys with interviewers working in centralized CAPI centers may become impossible, postal surveys requiring manual paperwork could be affected operationally, and even web surveys focusing on issues other than the pandemic may become irrelevant to respondents with the big pandemic agenda. With COVID-19 prevention measures, survey institutions may put projects on hold, cancel completely, or change their mode of data collection. Changes in mode are prone to many implications regarding coverage, sampling, nonresponse, and measurement errors. Gummer et al. (2020, p. 226) underline the major measurement issue here: “In the case of mode switches, it is paramount to avoid methodological artifacts, since effects due to mode changes might be misinterpreted as substantive changes across time.” In this section, we will summarize some of the documented changes to surveys, including design details related to above errors whenever available³². As we mentioned at the beginning of the section, we make a distinction

³¹ There was also a session titled “COVID-19 crisis – is household survey at a crossroads?” in the 2020 Virtual UN World Data Forum, organized by the Intersecretariat Working Group on Household Surveys. The recording is available on YouTube: <https://youtu.be/hPdkQQgBjHA>.

Another event was conducted by the Program in Survey Methodology at the University of Michigan titled “Surveys in the Era of the Pandemic and Beyond”, as the Third Annual Likert Workshop: <https://psm.isr.umich.edu/node/315>.

³² At the webinar conducted by the International Program in Survey and Data Science at the University of Mannheim in April, Raphael Nishimura talked both about the effects of the pandemic on the surveys carried out at by the University of Michigan Survey Research Center, and on several surveys from Brazil, from a

between cross-sectional surveys and longitudinal ones, imagining the former would have higher constraints of sampling frames and fewer options of mode switches.

The United Nations Intersecretariat Working Group on Household Surveys (ISWGHS) has a Task Force on COVID-19 and household surveys, established in June 2020 (United Nations Intersecretariat Working Group on Household Surveys, 2020c). In order to assess the situation in different countries in terms of their national surveys, a survey of national statistics offices was conducted by the UN Statistics Division and World Bank, and was presented on the Task Force’s website (United Nations Intersecretariat Working Group on Household Surveys, 2020b). A total of 122 countries responded to the survey. Price surveys, labor force surveys, and enterprise surveys were the top 3 types of surveys that countries had on their agenda for the pandemic era (Figure 3.1).

Figure 3.1 The number of countries who had surveys scheduled for the pandemic period, by type of survey

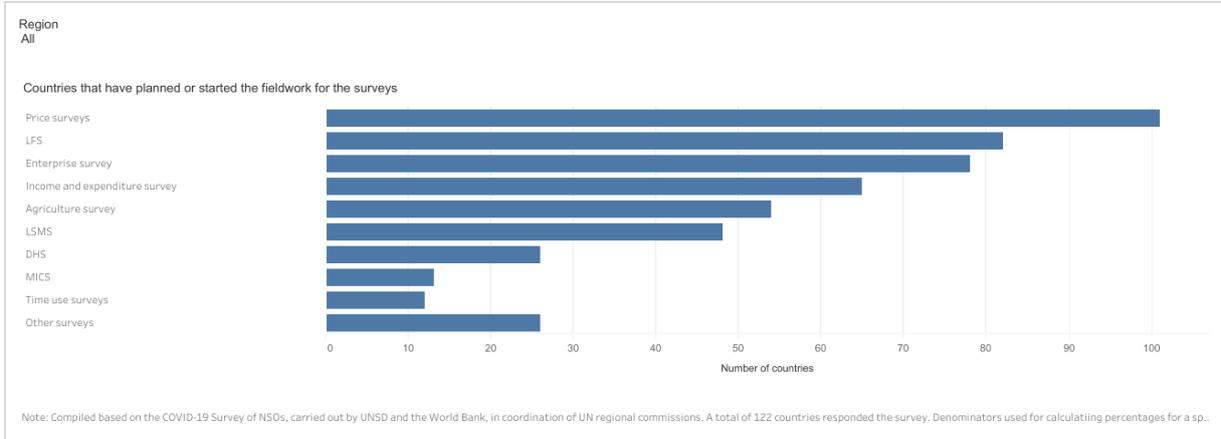


Image source: United Nations Intersecretariat Working Group on Household Surveys (2020b).

Countries were also asked about the adjustments that had to be considered due to the pandemic by types of surveys (Figure 3.2). Suspension or postponement of surveys was more common in developing regions than developed regions. By types of household surveys, the Living Standards Measurement Study (LSMS), and income and expenditure surveys were more affected in developed regions. In developing regions, MICS surveys and income and expenditure surveys were mostly stopped or postponed, whereas time use surveys were least affected.

sampling perspective: <https://survey-data-science.net/webinar-surveys-corona>. Our interview with Dr. Nishimura in Section 4 includes further details.

Figure 3. 2 Responses to the question: “If fieldwork for one or more of the following surveys were planned or started, did you have to stop or postpone the fieldwork due to Covid-19 pandemic?”

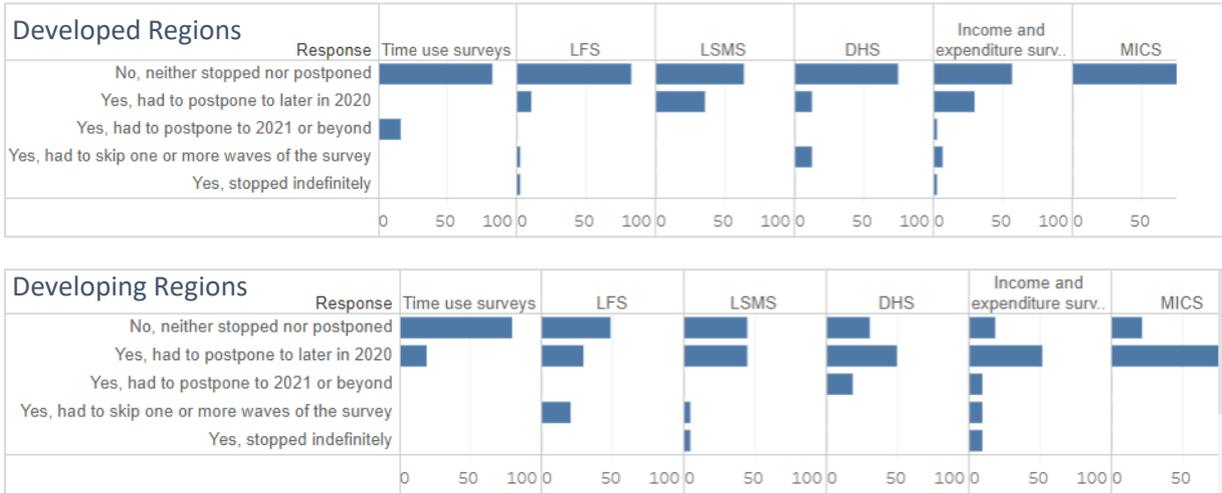


Image source: United Nations Intersecretariat Working Group on Household Surveys (2020b).

The challenges experienced by countries were also questioned. For instance, for the income and expenditure surveys, countries stated “reducing face-to-face interaction” as the main one, followed by mobility restrictions. Other challenges included staff not being available or being ill, funding limitations, etc.

When countries were asked if they had to alter their survey due to COVID-19 and what changes they implemented or considered implementing for most household surveys, the top response was changing survey mode or using alternative data sources. For MICS and DHS, however, the top response was the addition of COVID-19 related questions (also sample size reduction for MICS). Other adjustments included changing the sample design, shortening the questionnaire, and reduction of sample size.

For LSMS, DHS, income and expenditure surveys, LFS, and time-use surveys, phone surveys were the first option preferred by countries whenever a mode switch was done from FTF to another mode. For DHS, the second option was administrative data. For LSMS, income and expenditure surveys and LFS, web surveys were the second option. Other alternatives used by countries were model-based estimation, phone-call detail records, citizen-generated data/crowdsourcing, remote sensing/satellite imagery, social media, and others.

The support provided by ISWGHS agencies is documented in methodological publications (United Nations Intersecretariat Working Group on Household Surveys, 2020a). There are 32 briefs from 10 agencies as of December 10 (FAO, ILO, UN Women, UNDP, UN Economic Commission for Latin America and the Caribbean [UNECLAC], UN Economic and Social Commission for Asia and the Pacific [UNESCAP], UNESCO, UNICEF, UN Department of Economic and Social Affairs [UNSD], World Bank). Below we mention some examples of these briefs.

For instance, the FAO published a brief in mid-April (Food and Agriculture Organization of the United Nations, 2020), about National Statistical Services, based on their observations of about 30 countries. They highlighted a “disequilibrium between low-income and developed countries,” with the latter falling behind in introducing specific measures about COVID-19 and announcing them, interpreting this as a potential blackout of national statistics. They argued that labor force surveys, consumer prices

surveys, consumer trust surveys, and agricultural surveys would be immensely affected by the suspension of face-to-face surveys. They classified the effects of the pandemic on national offices into five main categories; 1) statistical operations being maintained, 2) statistical operations being maintained with alternative data collection tools, 3) postponing, suspending, or rescheduling statistical operations, 4) canceling statistical operations, and 5) no decisions being taken. Under these titles, FAO provides examples from Lithuania, Bulgaria, Ireland, Belgium, France, Peru, Georgia, Uganda, Cambodia, Philippines, Angola, Senegal, Armenia, Mexico, Liberia, Côte d'Ivoire, and others.

Another example is the brief by ILO (International Labor Organization, 2020), released in May, about data collection on labor statistics. The main topic of this brief was rapid surveys. Rapid surveys were recommended for situations when established large scale household surveys were not possible. Rapid surveys were described as surveys with shorter questionnaires, smaller sample sizes, and often limited sampling frames and sample designs compared to usual surveys. On another brief, ILO reviewed the pre-pandemic modes of labor force surveys around the World (Figure 3.3) and discussed how different surveys were affected based on which mode they have been using. One issue they mentioned in this brief was the importance of the first contact with respondents and how the first waves of surveys were at stake in the absence of telephone frames as reliable as address. ILO also underlined that nonresponse was likely to become a bigger issue in the case of a mode switch in these surveys. Another issue highlighted was that even ongoing CATI operations would be affected due to the necessity of decentralization. This, ILO argued, could be a technological challenge. Such an infrastructure issue could also arise for CAWI. The frequency of data collection was also mentioned, and how there was a potential danger of losing time series. Another important discussion in this same brief by ILO was on questionnaires, where some practices done by countries were mentioned, such as asking questions directly related to the pandemic, asking questions not as directly related but are still informative about its impacts, asking questions regarding how certain things have changed since a certain period, or adding response options to existing questions.

Figure 3.3 Main modes of labor force surveys by region

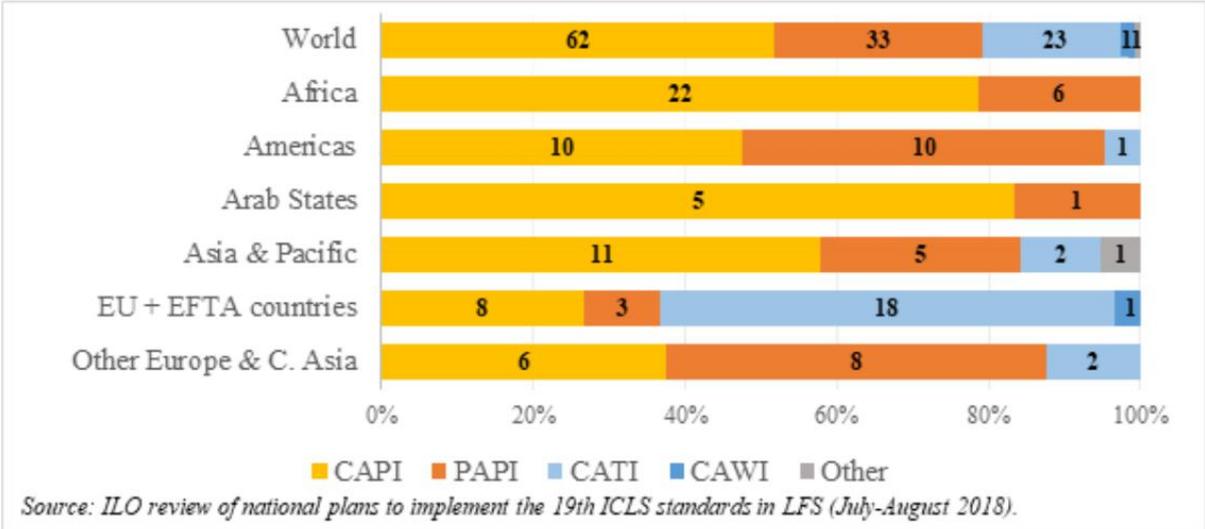


Image source: International Labor Organization Department of Statistics (2020)

Among regional agencies, UNECLAC released a brief titled “Recommendations for eliminating selection bias in household surveys during the coronavirus disease (COVID-19) pandemic” in May (United

Nations Economic Commission for Latin America and the Caribbean, 2020). The document was released because of the mode changes that had to be done by national statistical offices, from face-to-face to phone or online. The document discusses detecting bias, and certain remedies, such as propensity score adjustment, two-stage calibration method, and post-stratification based on multilevel models.

The UNICEF briefs are based on the Multiple Indicator Cluster Surveys, which we mention in Section 3.4.1. UNSD published two briefs on telephone surveys. In one, they mentioned issues to consider (United Nations Department of Economic and Social Affairs, 2020a), such as the sampling frame, response rates, interview protocol, targeted geographic area, questionnaire length, questionnaire content, interviewer training, electronic data capture/data center, and quality assessment. In another brief, UNSD published an interview about telephone surveys with Prof. James M. Lepkowski (United Nations Department of Economic and Social Affairs, 2020b).

In the rest of the section, we provide individual survey examples.

3.1.1 Cross-sectional surveys

The European Social Survey (ESS) is a multinational survey series of repeated cross-sections. It was established at the National Centre for Social Research (the current NatCen Social Research) in 2001. It has been carried out once every two years since 2002, with 38 countries participating in at least one round. Tim Hanson from the research team explained in *research methods* that round 10 was scheduled to begin in September 2020 with a FTF mode in about 30 countries (Hanson, 2020). Due to the pandemic, the ESS core scientific team has advised that countries start FTF surveys around March 2021, but still kept it flexible for countries to complete data collection any time before the end of 2021 (European Social Survey, 2020b). Hanson mentioned that respondents may not be willing to participate FTF surveys even if they become possible and that they were working on alternative modes to prepare for this. One alternative is video interviews, which the research team has already started testing. Another alternative is phone interviewing, for which showcards need to be sent to respondents. Hanson also mentioned an ongoing component of ESS, the CRONOS-2, a project to invest in online probability panels. The original plan was to initiate this project during round 10 via FTF interviews (European Social Survey, 2020a); however, Hanson mentions that recruitment may now be through the mail.

Another survey by NatCen, the British Social Attitudes (BSA) Survey Series was discussed by Gillian Prior, the Director of Survey Research, in the first webinar titled “The impact of the COVID-19 pandemic on data collection for social surveys: Challenges and opportunities” by the Royal Statistics Society (Royal Statistics Society, 2020). BSA is a cross-sectional survey series, carried out annually since 1983 (Phillips & Taylor, 2016). Its main mode is CAPI with a self-completed component, with advance letters sent beforehand. Its random sample is based on a postcode address file. Gillian Prior mentioned that there were 3 alternatives NatCen considered for their FTF surveys as a response to pandemic conditions: 1) push-to-web, 2) push-to-CATI or video, and 3) Web/CATI. While the first two alternatives were considered mostly to keep cross-sectional survey series going, the third, requiring prior frame information, was deemed possible for longitudinal ones. The first option was applied for BSA. In this new methodology, researchers reached out to 18,500 addresses, aiming for 3,000 individuals. There were reminder letters and conditional incentives. The questionnaire was revised, and would now take 25 minutes, instead of 60 minutes. The fieldwork of BSA started in late October instead of July as

originally planned after usability tests for the new questionnaire and reviews. Health Survey for England trial, Test & Trace App survey were also listed as those to switch to push-to-web.

Among NatCen surveys, in English Housing Survey, Family Resources Survey, National Travel Survey, and Scottish Health Survey, which are all cross-sectional surveys, the second option of switching from face-to-face to push-to-CATI was used. Instead of face-to-face, for these surveys, respondents sampled with address-based methods were sent invitation letters asking for telephone information through various modes. They were then called. Gillian Prior underlined that this mode proved fast responses. Response rates were around 10-15%. Questionnaire length, questionnaire wording on the telephone, managing showcards, and dealing with self-completion parts of questionnaires were among the issues researchers had to consider.

In the same webinar (Royal Statistics Society, 2020), Alex Bogdan, Associate Director at Ipsos MORI talked about the British Electoral Study. This survey, too, like the BSA, was moved from FTF to push-to-web with mail option. When the pandemic hit, fieldwork was already underway. Therefore, some respondents were interviewed FTF, some with web, and some with mail when the survey was completed. There were differential incentives planned at the time of the webinar. Comparing web respondents to mail, researchers observed younger respondents, a higher level of education, working full time, and reporting no religion. Thus, Bogdan underlined the importance of including the mail option for a more balanced sample.

Another international survey program consisting of address-based cross-sectional surveys is Multiple Indicator Cluster Surveys (MICS) by UNICEF. The survey program made an announcement on March 31 on their website, recommending suspension of certain activities, which included 1) survey staff to interact with the population, 2) international travel, or 3) face-to-face meetings, to avoid contagion of the virus (UNICEF MICS, 2020a). These activities covered a variety of survey operations, namely fieldwork, interviewer training, in-office meetings, dissemination meetings, workshops, and alike. It was further recommended that online means be used for meetings, workshops, etc. The recommendations of MICS also differed by the stage of the survey in each country. For countries that were at the stage of planning, they suggested reevaluating certain aspects such as questionnaires and sample design. For those where listing or pre-testing was underway, the program recommended countries not to speed up, to ensure data quality. In addition, for those whose fieldwork was interrupted, an assessment was recommended to see how much of the fieldwork was complete, and the extent of the geographical coverage of completed interviews, to decide whether to move on to post-fieldwork activities or pause and resume the fieldwork later. Lastly, for surveys with completed fieldwork were advised to focus on post-fieldwork activities in the pandemic conditions.

The MICS has extended its modes in the last few years, through phone surveys with CATI, with an initiative called MICS Plus. This initiative is expected to serve multiple purposes, from emergency monitoring to creating panel data sets of seasonal indicators (UNICEF, 2020). COVID-19, as a global emergency, has been a topic in MICS Plus, implemented in Belize as a pilot study, Georgia, and Mongolia, and findings from the Belize survey is presented on the survey website (UNICEF MICS, 2020b). The interview with the Global MICS Coordinator Attila Hancioğlu in Section 4 provides in-depth information on this initiative.

The Demographic and Health Surveys (DHS) Program is an international survey program, carried out in over 90 countries since 1984, focusing on issues like contraception, fertility, and mother and child health (The DHS Program, n.d.-b). DHS covers countries from Africa, Asia, Eastern Europe, and Latin

America. All DHS surveys have been conducted face-to-face to this date, using PAPI and CAPI, with nationally representative probabilistic samples. The most recent update from this program about COVID-19 is dated September 2 (The DHS Program, 2020). This update underlines that field activities were paused due to the pandemic, and there was an effort to resume them at the time, with remote support from the Program’s headquarters in the US. Countries who were affected are listed in Table 3.1, with further detail on the program’s website. Accordingly, the fieldwork of four countries was interrupted due to the pandemic, and different plans were made for each depending on how much of the fieldwork they completed before the pandemic. Those who had fieldwork scheduled for 2020 but were not interrupted are either set to start late in 2020 or early 2021. Some countries decided to postpone to 2021 altogether, especially those to conduct malaria surveys, which need to be done at a certain season. DHS also has a website that shows an up-to-date list of ongoing surveys (The DHS Program, n.d.-a).

Table 3.. 1 List of DHS countries affected by the pandemic by their plans

Status	Countries
Resuming fieldwork	Rwanda, Gabon, Madagascar, Liberia
Starting fieldwork	DHS: Afghanistan, Cambodia, Myanmar, Angola, Ethiopia Survey Provision Assessment Survey [SPA]: Ethiopia, Nepal Malaria Indicator Survey [MIS]: Kenya DHS Pilot: Uganda
Survey activities postponed until 2021	DHS: Burkina Faso, Cote d’Ivoire, Lesotho, Mozambique, Tanzania MIS: Cameroon, Guinea, Mali, Niger, Nigeria

Pew Research Center posted an entry on their website at the end of April about the effect of the pandemic on their global polling operations (The Coronavirus Pandemic’s Impact on Pew Research Center’s Global Polling, 2020). The Center has a “Global Attitudes Survey”, which are cross-sectional surveys carried out in 20-40 countries per year (Moynihan, 2020). They emphasized that they were committed to FTF methods in many parts of the world, such as Africa, Latin America, and the Middle East and North Africa (Figure 3.4); they thus had to suspend most of their ongoing operations with the pandemic in these regions. One option they considered was to switch to telephone surveys, for which the center was working at the time of the post. The other option would be to wait until the pandemic ends to resume fieldwork, the date of which is yet uncertain.

Figure 3. 4 Pew Research Center’s coverage of countries by two main survey modes, face-to-face, and telephone

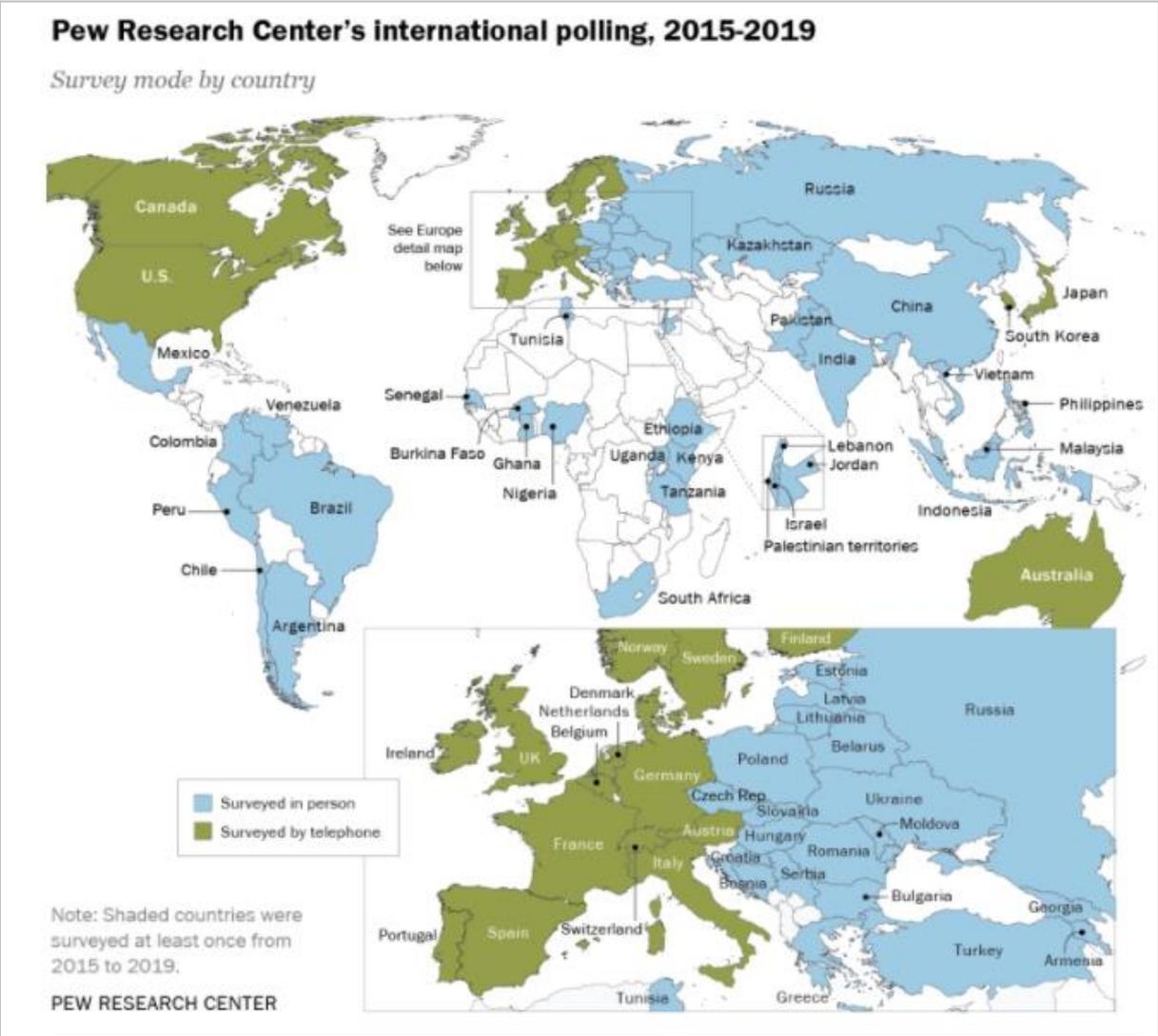


Image source: Moynihan and Letterman (The Coronavirus Pandemic’s Impact on Pew Research Center’s Global Polling, 2020)

Gallup conducts the Gallup World Poll, covering 160 countries where 99% of the world’s adult population live (Gallup, n.d.) (semiannual, annual, or biennial surveys, depending on the country). These annual surveys are based on over 100 global questions, as well as some region-specific ones. The mode differs by country. If, for any given country, telephone coverage is at 80% or higher, telephone surveys are used, like in the U.S., Canada, Japan, Australia, and Western European countries. For phone surveys, either RDD or a representative frame is used. Dual frames are also employed whenever mobile phones are commonly used. In Asia, Eastern Europe, Middle East, and Africa, face-to-face is the primary mode with random address samples. The duration of the surveys differs by mode, 1.5 hours in FTF vs. 0.5 hours on the phone. Address based samples have multistage designs. Within-household respondent selection is based on the last birthday or the Kish grid method. The sample size is around 1,000 respondents per country. Inferences are drawn from weighted data, with weights calculated based on probability of selection, nonresponse, and dual-frame adjustments when applicable. There is also poststratification by each country’s demographic structure.

In July, Gallup announced that they resumed the 2020 World Poll (Srinivasan & Clifton, 2020). They explained that this year all data would be collected by telephone, and FTF will be avoided for the safety of interviewers and respondents. For 34 countries where they used telephone methodology before, little has changed. Gallup made some assessments for the remaining countries:

- 1) They assessed telephone coverage and reduced their 80% coverage threshold to 70%, and decided to switch to this mode for countries that fit this criterion.
- 2) They also assessed mobile-only coverage for countries where only mobile phones could be sampled.
- 3) They evaluated the infrastructure of countries to switch to CATI, in terms of software, hardware, internet connection, and the possibility of decentralized CATI for interviewers.

The assessments led Gallup to switch to CATI in over 70 countries. The company's decision was to treat mobile phones as a household contact rather than an individual one to increase coverage. Questionnaires for the mode switching countries were edited according to the CATI questionnaires already used in other countries.

As a regional example, in the May Webinar of WAPOR, Angelo Ambitho, the CEO of Infotrak Research and Consulting, a company that conducts research in East and West Africa and is based in Kenya, talked about the effects of the pandemic in research and polling in Africa (World Opinion for Public Opinion Research, 2020). She mentioned that the predominant mode of survey research in Africa was FTF, which had either come to a halt due to the pandemic, or at least reduced. She mentioned that mobile devices were getting common in Africa, with some inter-country variation. As replacement modes, Ambitho mentioned CATI, CAWI, TIDI (telephone in-depth interviews), SMS Polls, and panels as alternatives. Apparently, web survey has not been popular in Kenya or East Africa, and response rates were low, as low as around 10-15%. Thus, they employed mixed modes at times, CAWI for interviews, and CATI for follow up while preparing respondents for the CAWI. Figure 3.5 below summarizes the pros and cons from this presentation regarding the shift to the online environment based on their own experience.

Figure 3. 5 The Digital Meeting Space Experience, based on the observations of the Infotrak Company

PROS	CONS
It's convenient for respondents as they are to participate from surroundings familiar to themselves	Limited internet access & poor quality connection/bandwidth affect undermine the discussions
It's cheaper to recruit respondents	Poor commitment of respondents online; many drop out at last minute
Easy to identify active respondents and thus engage those who are quite	Respondents dropping out in the course of the discussion due to poor network
No time wasted in waiting for respondents to arrive at venues	Its limited to the well to do and educated; those who have the gadgets and understand how to navigate technology smoothly

Image source: WAPOR (2020)

In the same webinar by WAPOR, the WAPOR representative for Uruguay, Ignacio Zuasnabar talked about the effect of the pandemic in survey research in Latin America Africa (World Opinion for Public Opinion Research, 2020). He underlined that face-to-face was not the predominant mode in Latin America before the pandemic and that telephone and web surveys were already common, mostly due to their cost advantage over face-to-face. He discussed that high-quality face-to-face interviews were few, and included those under the World Values Survey, Latinobarometro, or those done by Gallup World Poll or LAPOP, all of which halted field operations with the pandemic. CATI surveys were either decentralized or switched to online mode. Among phone and web modes currently used in Latin America, Zuasnabar mentioned CATI-cell phone, CATI-landline, IVR-robopolls, web-based non-probability surveys with social media recruitment, and web-based probability surveys. We mention Zuasnabar's evaluations of challenges related to mode shifts in Section 5.

3.1.2 Longitudinal surveys

As mentioned earlier, in this section we distinguish between cross-sectional studies and longitudinal surveys, because mixed-mode designs are relatively common in the latter. Even if a single mode is used for data collection, contact information is usually available from the first wave. We should underline that there is also a distinction between ongoing longitudinal surveys and those about to start their first wave at the pandemic outbreak. We can argue, for the former case, that the effect of the pandemic would perhaps be similar to that of cross-sectional surveys – depending on the availability of frame. Perhaps another distinction to make is whether or not the fieldwork had already started when non-pharmacological interventions (NPI) for COVID-19 were introduced. An interrupted fieldwork may indicate two different modes before and after NPI, as well as a complete cancellation of the remaining interviews at the time of NPI. For instance, Will et al. (2020) underline how challenging it was especially for FTF surveys interrupted by NPIs.

Here we attempt at reflecting how and if COVID affected 1) the survey mode, 2) the sample design, and 3) the questionnaire used, whenever such information is included in the sources.

3.1.2.1 Surveys done face-to-face, or involved face-to-face component prior to COVID-19

Surveys of income are among the most major surveys that are regularly conducted by national statistics offices. In Europe, these surveys are in line with the standards and definitions of EUROSTAT, and are called Survey on Income and Living Conditions (EU-SILC). In April, EUROSTAT released a methodological issue, including some recommendations to countries conducting EU-SILC (EUROSTAT, 2020b). It was highlighted that PAPI and CAPI were the dominant mode of data collection in most countries. Countries were recommended to either postpone or prolong the fieldwork or switch from PAPI or CAPI to CATI or CAWI (or both), with a bigger emphasis on mode change. Since reaching panel respondents are relatively easy with phone numbers collected in the first wave, further recommendations were made for first-wave respondents whose phone information may not exist. These recommendations were based on best practices that already took place by the time of the brief: 1) sending letters to the 1st wave sample households and ask them to contact statistics offices to provide phone numbers, 2) oversample the 1st wave, 3) prolong existing panels, 4) look into the possibility of using administrative sources more effectively, and 4) cooperate with other national institutions to match telephone information to the 1st wave sample. An example of this last recommendation can be seen in the Household Labor Force Survey in Turkey in Section 4, in the interview with Enver Taştı. EU-SILC also had some recommendations regarding questionnaire content. A question to compare income to the

previous year was recommended, as well as explicit clarifications of time periods in the questions, emphasizing the distinction between current status and pre-pandemic status.

EUROSTAT has also been releasing briefs for the EU-Labor Force Survey, with the last one dated June 30 (EUROSTAT, 2020a). The brief mentions that data collection has been affected as early as February 5 in Europe. EUROSTAT underlined the following risks to LFS surveys: fewer number of attempts at interviews, higher nonresponse, changes in interview modes – and all of these happening simultaneously with the impact of the pandemic on the topic of interest; labor force. Smaller surveys meant lower precision, higher response rates could potentially mean nonresponse bias, and changes in mode could mean coverage error. EUROSTAT conducted a survey of countries regarding their LFS surveys. Accordingly, only 5 out of 31 responding countries said their 1st quarter LFS was not affected by the pandemic. For 24 countries that provided detailed results, it was seen that mode changes and a decrease in the gross sample were common (Figure 3.6). Similar to EU-SILC, in EU-LFS too, first contacts are often done face-to-face because a phone frame is not always available for the first wave. In terms of the questionnaire, EUROSTAT underlined that no changes should be made to the questions from which ILO labor market status is derived from, to maintain comparability over time and across countries. It was emphasized that “complement rather than change existing variables is better” (EUROSTAT, 2020a, p. 5). There were variable-specific recommendations in this brief.

Figure 3. 6 Countries responding to Eurostat survey regarding how their LFS Surveys were affected, by calendar year of 2020 (preliminary findings)

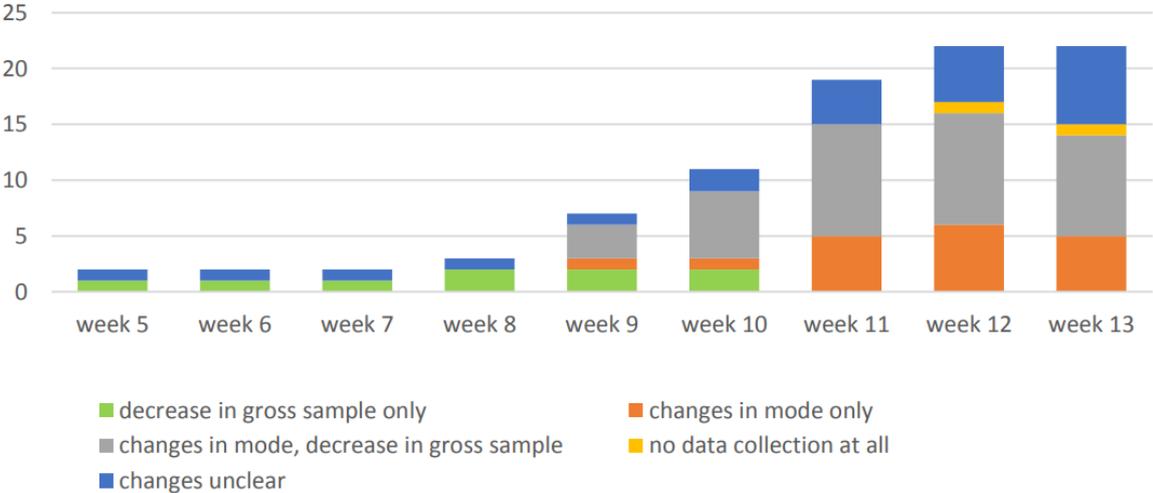


Image source: EUROSTAT (2020a)

The Generations and Gender Survey (GGS) is an international longitudinal survey series focusing on demographic and social issues. The Generations and Gender Program was launched in 2000 by the United Nations Economic Commission for Europe (UNECE), and Netherlands Interdisciplinary Demographic Institute (NIDI) has been coordinating it since 2009 (Generations & Gender Programme, n.d.). Twenty countries have conducted at least one wave of data collection within the program, and 12 have conducted at least two waves. Gummer et al. (2020) underline that the questionnaire of the new Wave of GGS 2020 is renewed extensively. The German version of GGS is part of the Family Research and Demographic Analysis (FRoDA) infrastructure, hereby will be called FRoDA-GGS. The fieldwork for FRoDA-GGS was to be conducted from November 2020 till July 2021. Respondents would be randomly recruited to CAPI or CAWI, where the former would include an invitation letter before

being contacted by interviewers, and the latter would be invited to a link through a letter. With the pandemic, the team decided to postpone recruitment to February 2021. One concern researchers have for this new date is the capacity of their institution at that time because other surveys have also been postponed or delayed. This has potential impacts on both central staff, and interviewers. Moreover, most interviewers in Germany are said to be above the age of 60, putting them in the high-risk group for COVID-19. Gummer et al. (2020) underline that in the absence of a vaccine, this staff may not be available for CAPI interviews. Another concern cited is the anticipation of lower cooperation rates because respondents may be reluctant to accept interviewers in their homes to abide by social distancing rules. One remedy the researchers are planning on is larger cash incentives. Another remedy is drawing a larger sample for CAWI and keeping some of it as a reserve in case CAPI interviews fall short. Moreover, non-contacts and soft-refusals will be invited to CAWI.

The Survey of Health, Ageing and Retirement in Europe (SHARE) is a longitudinal survey carried out since 2004 in Europe and Israel. In addition to the rich data it offers, SHARE is also special in the sense that it has data versions that are harmonized (Harmonized SHARE) with other aging surveys such as the American Health and Retirement Survey (HRS) and the English Longitudinal Study of Ageing (ELSA). We will summarize the effects of the pandemic on this survey based on the draft article by Scherpenzeel et al. (2020).

SHARE interviews are typically done face-to-face (CAPI), including some biomarker collection and physical examination. The 8th wave of this survey, which started in October 2019, was still underway when COVID-19 became an issue in Europe around February. Between March 10 and 23, all 28 countries stopped the fieldwork, as Scherpenzeel et al. (2020) report. Two issues stood out at this point: 1) a new mode was required, 2) data was needed to show the effects of the pandemic in terms of labor force participation, income, socializing, inter/intragenerational help on the major age groups of the SHARE population.

The new mode for SHARE was decided to be Computer-Assisted Telephone Interviewing (CATI). Scherpenzeel et al. (2020) argue that CAWI and PAPI³³ were not considered because they would imply a shift from an interviewer-administered mode to a self-administered one, which is a major methodological change. Another consideration regarding CAWI was the relatively low use of the internet by the elderly, especially above age 80. Although this disadvantage was not associated with PAPI, the interruptions in postal service due to the pandemic were brought up as a downside. It also may push respondents outside of their residences to send back the filled-out questionnaires.

The questionnaire, now titled SHARE Corona, was redesigned to meet the data needs mentioned before. It was now shortened and had new sections on changes the respondents may have experienced regarding health and health behavior, mental health, infections and healthcare, work and economic situation, and social networks – due to the pandemic.

By the time all CAPI interviews were suspended in March, 70% of longitudinal and 50% of refreshment interviews were completed. Once the switch to CATI was decided, and the questionnaire design was complete, the sample design was also reevaluated. For instance, no further attempts were made to contact the rest of the refreshment samples. This is similar to a cross-section being interrupted; researchers underline that telephone numbers were not available for most refreshment sample households, with a few exceptions from Scandinavian countries, which are known for well-developed

³³ Here we believe the authors used the term PAPI to imply a questionnaire sent out by post.

registration systems. Interviews were resumed for the longitudinal respondents, with telephone numbers collected at previous waves. Samples were selected both from respondents who already participated in Wave 8 before the pandemic was spread in Europe and those who have not yet participated in this Wave. Researchers underline the importance of the former enabling to observe the effects of the pandemic and the importance of the latter in terms of providing a chance to observe the effects of non-response errors, measurement errors, and survey costs.

Another related longitudinal panel study is The Health and Retirement Study (HRS) of the University of Michigan, which addresses the challenges and opportunities of aging since 1992³⁴. The survey has a representative sample, including nearly 20,000 participants in America. It is the most comprehensive data source for the work, aging, and retirement in the US, conducted in cooperation with the National Institute of Aging (NIA) (Fisher & Ryan, 2018). For the data collection, the survey team employs face-to-face interviews with half of the participants while the other half is interviewed through telephone, by alternating these modes at each wave (Fisher & Ryan, 2018). In response to the pandemic, the research team conducted the "COVID-19 module of HRS 2020", which was applied to the half of the random sample of the respondents originally assigned to the enhanced face-to-face mode (EFTF), to be now interviewed by telephone (Health and Retirement Survey, 2020). This sample was also randomly divided by half, and the two samples were interviewed at different times (starting in June or September). The data of the first subsample (EFTF1) is currently available for researchers with 3,266 respondents, and the EFTF2 data is underway as of early December. In addition to the phone interviews, respondents received self-administered questionnaires by mail. The phone interviews covered most of the regular sections of HRS, as well as questions on COVID-19 (Health and Retirement Study, n.d.). What is more, the survey initiative plans to conduct a mail survey with the full sample in 2021 (*Webinar - COVID 19 - Data Collection in Longitudinal Studies, 2020*).

Guy Goodwin (Goodwin, 2020) from the National Centre for Social Research mentioned how the pandemic impacted the English Longitudinal Study of Ageing (ELSA) in the June 2020 issue of the *research matters* magazine. ELSA is designed to represent non-institutionalized persons aged 50 (NatCen Social Research, 2018). Its sample is based on the respondents of the Health Survey for England, and the fieldwork of the first wave was conducted in 2002-2003, to be repeated at 2-year intervals. The sample is refreshed at certain waves. There have been 9 waves so far, each including a CAPI component. There have also been self-administered components, nurse visits, and online components to these surveys (English Longitudinal Study of Ageing, n.d.). Goodwin mentioned that the design was to switch to a mixed-mode web and CATI survey, where they would have a telephone line to assist with the completion of the former mode. The survey started in 2020 and was called the ELSA Covid-19 study, where the mode of contact was mail (English Longitudinal Study of Ageing, 2020), including a £10 incentive. The invitations were to the online mode, with the option of CATI. Other Surveys by NatCen, which used prior contact information of respondents to reach out to them during the pandemic were Mental Health of Children and Young People, DNAC (NDNS follow up), EHS follow-up, and NatCen Panel (Royal Statistics Society, 2020). For these surveys, there were letter or email invitations to web surveys with CATI options. Response rates were said to be around 40-75%.

Conducted by the John Hopkins University Bloomberg School of Public Health since 2011, the National Health and Aging Trends Study (NHATS) inquires about the disablement process and its consequences

³⁴ Information about Health and Retirement Study is available at <https://hrs.isr.umich.edu/about>

with a nationally representative sample of people age 65 and older in the U.S.³⁵ Its sample design is age-stratified, which is to say, individuals are selected from 5-year age groups between 65-90 and people aged 90 and over. The survey turned from in-person to the telephone mode with the COVID-19 measures. The pandemic also brought novelties to this long-term panel study; it now has an open-ended investigation for qualitative analysis of how COVID-19 affected the participants (*Webinar - COVID 19 - Data Collection in Longitudinal Studies*, 2020). Furthermore, the initiative plans to do twomail out/mail-back COVID-19 Surveys, and the questionnaire involves topics such as COVID-19 symptoms, living situation, social activities, health care, finances, well-being.

As an example from the U.S., the Panel Study of Income Dynamics (PSID), conducted by the University of Michigan, has been going on since 1986. It started out as a predominantly face-to-face survey, and then preceded as a predominantly telephone survey (Institute for Social Research, 2019), CATI, since 1993. Its sample has grown over time, with both through children in the households growing up to establish their own, and through refreshment samples. It includes a sample of over 18,000 individuals in 5,000 families in the US³⁶. The survey interrogates issues about employment, health, expenditure, income, and various related topics.

In 1997 and 2005, two supplements were established, Child Development Supplement (CDS) and Transition into Adulthood Supplement (TAS), respectively. The former is repeated at about 5-year intervals, and the latter on a biennial basis. The fieldwork for PSID-19 was completed at the end of 2019. The next wave of the PSID will start in 2021 and will likely include questions on exposure to COVID-19 as well as coping mechanisms of households with the financial problems introduced by it (Michigan Center on the Demography of Aging, n.d.).

The eligible respondents for CDS-19 and TAS-19 were identified from the PSID-19 (Sastry et al., 2020). The fieldwork for these two supplements started in fall 2019 and was expected to end in the spring of 2020; therefore, we summarize the effects on these two surveys based on the paper by (Sastry et al., 2020). Since TAS relies on FTF for follow-up, it is mentioned under the next title.

The CSD-19 was designed to include a combination of telephone and FTF interviewing. While phone surveys include screening and individual interviews, face-to-face surveys (home visits) include anthropometric measurements, biomarker collection, 24-hour diaries, child interviews, consent forms for children's record linkage, achievement tests for children, and alike. It was decided that telephone calls would be made in a decentralized manner, with interviewers calling from their homes. FTF interviews for CDS-19 were gradually suspended at the beginning of March, starting from some areas, and starting with the cancellation of long-distance travels of interviewers. By mid-March, it stopped completely, resulting in 34% of eligible home visits completed. Although the research team initially thought about shifting FTF interviews to fall 2020, it was decided that telephone and mail in/mail out would be a more feasible option, with the exception of some modules with hard to collect data. Sastry et al. (Sastry et al., 2020) underline the risk of lower response rates in this case, and potential data quality issues for some modules due to the switch from interview-administered to self-administered modes. In terms of questionnaire content, the CDS supplement was reported to include the effects of the pandemic on health, children's education and caregivers (Michigan Center on the Demography of Aging, n.d.). The pandemic had a negative effect on response rates in CDS-19. Although respondents

³⁵ Information on National Health and Aging Trends Study (NHATS) is available at <https://www.nhats.org/researcher/nhats>

³⁶ Further information on PSID is available at <https://psidonline.isr.umich.edu>

were more likely to be home, they were also more likely to have the burden of working from home, such as childcare, and alike.

At the workshop by AAPOR³⁷ titled “Changes to Field Operations” (American Association for Public Opinion Research, 2020) in November, Director of Survey Research Operations (SRO) at the University of Michigan Institute for Social Research, Survey Research Center (SRC), Stephanie Chardoul talked about the effects of the pandemic in general on the field operations done at their unit. She underlined that there were surveys at SRO that required face-to-face data collection, including biomarker collection, interviews with the elderly – a risk group for COVID-19, ranking these surveys at the bottom for returning to regular data collection. In general, Chardoul mentioned that there were 20 projects when COVID-19 restrictions were introduced in March, 11 of which were SRC projects. Rapidly, SRO prepared hardware to send FTF interviewers to switch to decentralized CATI, and prepared some remote interviewer training. An important aspect Chardoul mentioned was the need for essential staff at the center, for issues such as respondent mailing, respondent incentive payments, interviewer support for hardware, etc.

Another study from the U.S. is the General Social Survey (GSS), carried out by the National Opinion Center (NORC) at the University of Chicago. The first GSS was carried out in 1972 and has been annually done except for a few years missed, until 1994, and since then it has been carried out biennially. It started out as a cross-sectional survey series, but then it was switched to a rolling panel design (NORC, 2019). The survey is based on probability samples. At the same AAPOR workshop mentioned above (American Association for Public Opinion Research, 2020), Jodie Smylie from NORC explained the effects of the pandemic on the GSS’s field operations. She mentioned that that the 2020 plans for GSS were mainly FTF. However, NORC has been making contingency plans since January related to the pandemic, and had 12 ongoing surveys when the pandemic was declared. In the case of the GSS, NORC decided to implement 2 GSS studies for 2020, a panel one, reinterviewing respondents from previous years, and a cross-sectional one with address-based sampling. Both studies were designed as mail push to web, and researchers adapted the questionnaire design for self-administration.

At the same AAPOR workshop (American Association for Public Opinion Research, 2020), Wendy Hicks talked about the experience of Westat with face-to-face longitudinal surveys. She mentioned that they had about 1,000 interviewers in the field, and it was decided that all in-person operations would be stopped on March 16. Prior to the pandemic, the share of telephone interviews per survey was no more than 5%, but they switched all interviews to phone afterwards. Among challenges they had to overcome, she mentioned handling of incentives (face-to-face to mail), change in content and new IRB requirements because of these, the audio recording of interviews (CAPI devices vs. interviewers’ phone for decentralized CATI), and monitoring of CATI operation. Another topic Hicks underlined was interviewer training and what Westat did instead of an in-person training of interviewers.

David Hunter, project director for the National Survey on Drug Use and Health at RTI International, mentioned in the AAPOR Workshop about the safety measures they introduced for face-to-face research, and that they ran a pilot with these measures in July in places with low COVID-19 risk. At the

³⁷ AAPOR organized two other workshops within their COVID-19 Workshop series; one related to CATI, titled “Transitioning CATI to Remote Interviewing”, and another on qualitative interviewing titled “Changes to Qualitative Research”. All three webinars are available for purchase on the AAPOR website: <https://www.aapor.org/Conference-Events/COVID-19-Workshop-Series.aspx>.

time of the webinar, RTI was running fieldwork face-to-face in such places, and everywhere else, they switched to the web mode.

As an example from the U.K., Burton, Lynn & Benzeval (Burton et al., 2020) explained the effects of the pandemic on the “Understanding Society”, the UK Household Longitudinal Study³⁸. This survey has been going on annually since 2009-2010³⁹, with a mixed-mode design. While the first six waves were predominantly held face-to-face with 2% of all interviews conducted with CATI, 40% of sample members were invited to take the online mode in Wave 8. This proportion increased to 70% in Wave 10, where the follow-up to this mode would be FTF. At the beginning of the pandemic, the research team monitored the developments and kept taking necessary steps; from initially including an explanation to their website, encouraging respondents about rescheduling interviews if they were not feeling well; to finally completely suspending FTF interviews. At this point, in March, samples from Wave 10, Wave 11 and Wave 12 were ongoing. Already being a mixed-mode study, Understanding Society thus became web-first due to the pandemic, where non-respondents would be followed up by phone. Interviewers already had the phone version of the software installed on their computers and could call from home⁴⁰. Researchers also discuss cost-related issues, underlining that the saving due to the suspension of FTF was used to provide payment to interviewers who were now doing telephone interviews. The questionnaire content was also evaluated, with the decision of changing as little as possible to ensure comparability with prior waves. Nevertheless, new modules on COVID-19 were added to the Wave 11 and Wave 12 questionnaires. Researchers also decided to start a new study specific to COVID-19, as mentioned in Section 3.2.

Gummer et al. (Gummer et al., 2020) explained the effects of the pandemic on two ongoing longitudinal surveys in Germany: The German Family Panel (pairfam) and the Generations and Gender Survey (GGS). Pairfam has been going on since 2008 on an annual basis, originally starting with three birth cohorts. New cohorts and refreshment samples have been added over the course of time. Different modes are used for data collection, where CAPI is the predominant mode, and CASI is used for sensitive questions. The partners and children of respondents are interviewed, the former with paper-based self-administered questionnaires, and the latter with CAPI. The fieldwork of the Wave 12 of this survey began in November 2019 and was to be completed by the end of April 2020. It was paused in March with the pandemic. The fieldwork for the refreshment sample was scheduled for February through June 2020. Thus, the interviews for the panel and refreshment samples were completed at different rates when the survey was interrupted (78% and 28%, respectively). One concern of the researchers was the next annual wave. Unless they finished Wave 12 by August, their

³⁸ In the second webinar by the Royal Statistics Society titled “The impact of the COVID-19 pandemic on data collection for social surveys: Challenges and opportunities”, Jonathan Burton from the Institute for Social and Economic Research at the University of Essex also talked about the UK Household Longitudinal Study. In the same webinar, there were two other talks from the UK: Ian O’Sullivan, from ONS talked about how ONS redesigned social surveys during the pandemic, and Stuart Grant and Joel Williams from Katar talked about the Crime Survey for England and Wales: <https://rss.org.uk/training-events/events/events-2020/sections/2-2-the-impact-of-the-covid-19-pandemic-on-data-co/#fulleventinfo>

³⁹ Its precedent is the British Household Panel Survey from 1991. The 18th Wave of this survey constituted the frame for Understanding Society

<https://www.understandingsociety.ac.uk/sites/default/files/downloads/documentation/user-guides/mainstage/sample-design.pdf> & <https://www.understandingsociety.ac.uk/documentation/mainstage/user-guides/main-survey-user-guide/study-design>

⁴⁰ University of Essex provided ethical permission for the mode switch.

next wave would have to be postponed. Canceling the rest of the interviews would imply too much non-response and would negatively affect the participation of respondents in Wave 13, after skipping a wave. The final decision was to switch to a combination of CATI and mail survey. The reason CATI was preferred over CAWI was convenience and timeliness; the software of the former was easier to adjust. Researchers took precautions for potential mode effects: 1) all already scheduled CAPI was switched to CATI with interviewers whom respondents have already met, 2) sensitive questions that were previously collected through CASI were redesigned to be asked through a mail questionnaire following CATI interviews, and the questionnaire was edited for simplicity on paper, 3) partner interviews were not affected, already being paper-based, however, child interviews were switched from CAPI to CATI, and 4) mail incentives for CAPI were kept the same for CATI. Letters were sent to respondents, notifying them of the mode switch, and whenever phone numbers were not available, asking for them. Researchers stated that phone numbers were available for 2/3 of the respondents; thus, their overall expected contact rate was lower than their original for CAPI.

Will et al. (2020) explain the effect of the pandemic on the refugees in the German Educational System (ReGES) survey. The project started in 2016 (Steinhauer et al., 2019), the first wave of this longitudinal study was conducted in 2018 and it is conducted in 6-month intervals. It includes a cohort of young refugee children and a cohort of adolescents, where parents are interviewed for the former. The survey involves multiple modes. The main mode is CAPI, and there is a CASI component for sensitive questions and alike. Telephone and online modes have also been used, and there have also been interviews of educators and teachers. The fieldwork of the 7th wave of ReGES was scheduled between February and May 2020. A mail survey of educational staff was planned for completed interviews of children and adolescents. When COVID-19 interrupted the survey, 40% of the sample was interviewed, and the competence tests through CASI were completed in 93% of these interviews. The mail survey of educational staff was yet to be started. With resuming FTF and it no longer being an option, both due to the risk of contagion and regulations, researchers focused on 2 alternatives: to postpone the survey or switch to CATI. Because postponement would indicate a difference in terms of the academic year between the 40% of already interviewed students and 60% of yet to be interviewed students, researchers did not prefer this first option. The second option was feasible in the sense that interviewers could use existing computers and software, and they could call respondents from home. Some measures were taken to avoid a drop in response rates due to the switch from CAPI to CATI: 1) they made sure interviewers remained the same, 2) the reason for the mode switch (health) would be explained to the respondents, 3) the CAPI incentive would be kept as it is (rather than paying a CATI incentive, which is lower). Researchers also listed the disadvantages of the mode-switch, which are: 1) The competence test (CASI prior to COVID-19) could not be done for the remaining 60% of the sample, 2) consent to contact educational staff remained missing for some children, because a written consent had to be obtained during interviews and 3) the risk of mode-effects. To minimize mode-effects, researchers kept the questionnaires as similar as possible. They summarized the challenges they had with the questionnaire in two headings: content-related and technical/methodological challenges. About the former, they underlined that they had to frame questions about care, education, social interactions, living situations in a way that they would refer to the time before the pandemic. Questions similar to subjective well-being were not framed this way. Some additions were also made to the questionnaire to see the impact of the pandemic on the living situation, care, and education. Among technical/methodological challenges, researchers listed interviewers not being able to see some documents that respondents may show them (which were now read to them). Cards were shown

to respondents for some questions (e.g., life satisfaction) in CAPI; these were simplified and asked as scales on the phone. For questions in CASI, they kept response categories in CATI as similar as possible (keeping explicit refusal options in CATI, which existed in CASI). For newly added questions regarding COVID-19, researchers decided to add all of them as a block to the end of the interview to maintain comparability with prior interviews.

Sakshaug et al. (2020) explained in their paper the effects of the pandemic on ongoing surveys at the German Institute for Employment Research: IAB Job Vacancy Survey (JVS), IAB Establishment Panel (EP), Quality of Life and Social Participation (QLSP), and the Labour Market and Social Security (PASS); where JVS and EP are mentioned as establishment surveys. The PASS is an annual longitudinal household survey that has been ongoing since 2006 (Trappmann et al., 2019). The sample is based on a dual-frame of the residential population and welfare-benefit receivers. It has a mixed-mode design of face-to-face and telephone interviewing⁴¹. It is based on an “infinite degree contagion model,” where whoever moves into a PASS household becomes a member, and whoever leaves a PASS household to join a new one leads to all other members of this new household to become panel members. The fieldwork for 2020 started in February and was interrupted soon after to switch to telephone only. Researchers expected lower response rates at this point, even if calls would be made by interviewers known to respondents from prior face-to-face visits. The questionnaire was also modified to include questions on quarantine/home isolation, and the effects of the pandemic on issues such as income and work.

3.1.2.2 Surveys that did not have a face-to-face component prior to COVID-19

The TAS-19, a supplement of the PSID Survey mentioned in the previous heading, was to be carried out with a mixed-mode design, where respondents would be randomly assigned to web-first or telephone with no-web modes, with calls to be made centrally (Sastry et al., 2020). It started at the end of 2019 and was supposed to be completed by spring 2020. FTF interviews would only be conducted for those with no up-to-date phone or e-mail address information, which had to be canceled due to the pandemic. Due to its limited reliance on FTF, TAS-19 was less affected by the pandemic than the other supplement, the CDS-19. However, centralized calls were decentralized, requiring some interviewer training for the necessary technology, as well as adjustments in supervision. Researchers mentioned overall reductions in productivity among the staff due to conditions caused by the pandemic. Besides, at the time of their paper in the SRM special issue, they underlined that the effect of the pandemic on the TAS-19 response rates was uncertain. They also underlined that those who could only be reached through FTF could no longer be contacted.

The Quality of Life and Social Participation (QLSP) survey by the German Institute for Employment Research is a panel survey with telephone interviewing and aims to evaluate recent employment subsidy programs (Sakshaug et al., 2020). The first wave was supposed to start in March 2020. Although mode-wise or sample design-wise, the survey did not require any modifications, the first wave was postponed to re-design the questionnaire to include COVID-19 related questions. Researchers included questions on how working hours and income were affected in comparison to pre-COVID times as well as on working from home, concerns over becoming unemployed, changes in working conditions, and about being in quarantine/voluntary isolation. In terms of interviewers, the

⁴¹ In the first 4 waves, the first contact was through CATI whenever phone numbers existed. From wave 5 and on, it has been switched to CAPI. Mode switches between the two have been flexible, even at the request of households.

Institute took measures to avoid potential contagion by working with shifts and allowing for more physical space in between. This change in work schedule was also expected to affect overall field duration.

Another longitudinal project is the National Longitudinal Survey of Youth 1979 (NLSY79) that tracks the lives of a sample of American youth born between 1957-64⁴². Started with 12,686 respondents in 1979, the survey now has 9,964 respondents in eligible samples re-interviewed every two years. After 2002, telephone interviews became the primary mode of data collection, and in 2016, 96.3 percent of the interviews were completed via telephone. Respondents were between ages 14-22 when the first interviews were conducted, and they will be age 56-64 in the 29th wave scheduled for 2020-21 (*Webinar - COVID 19 - Data Collection in Longitudinal Studies, 2020*). The study includes questions on education, employment, household, parents, dating, sexual activity, income, health, etc. For the 2020-21 wave, telephone interviews will be the means for data collection, including content related to COVID-19.

Begun in 1994 with the first wave, Sexual Orientation/Gender Identity, Socioeconomic Status, and Health across the Life Course (SOGI-SES) is a longitudinal study that administers a web survey with the aim of gathering data on socioeconomic, developmental, and health issues of sexual and gender minorities (*Webinar - COVID 19 - Data Collection in Longitudinal Studies, 2020*). The survey enterprise aims to add COVID-19 questions in the study's sixth wave planned for 2022-24 besides the short battery of COVID-19 questions for the panel maintenance in fall 2020 (*Webinar - COVID 19 - Data Collection in Longitudinal Studies, 2020*).

In the May Webinar of the World Association for Public Opinion Research, Yashwant Deshmukh, talked about the effects of the pandemic on an ongoing RDD telephone survey in India (World Opinion for Public Opinion Research, 2020). The survey is conducted by the company Cvoter, called the Cvoter India OmniBus, and has been carried out weekly for the last ten years and daily for the last one year. The effect of the pandemic on this survey has been the decentralization of CATI. The presentation covered technical aspects of this switch, such as moving from a single outbound process to two. Deshmukh underlined some drawbacks that came with this adjustment, such as decreasing number of interviews per day. Researchers also introduced a CoronaTracker in their survey, where they asked about fear related to the virus, the government's handling of the situation, symptoms, etc. Findings are published on the survey website⁴³.

3.2 Surveys to investigate the impacts of the COVID-19 pandemic

In this section, we first mention a manual published by WHO as a set of general guidelines for conducting behavioral insight studies for COVID-19, and then we briefly mention some of the many surveys available. We classified these surveys based on frames available. Several cross-sectional surveys are mentioned in the introductory part of this section, and some of these appear in more detail in other sections. Additionally, there are surveys mentioned in Section 3.4 that include adjustments to ongoing surveys on the impacts of COVID-19.

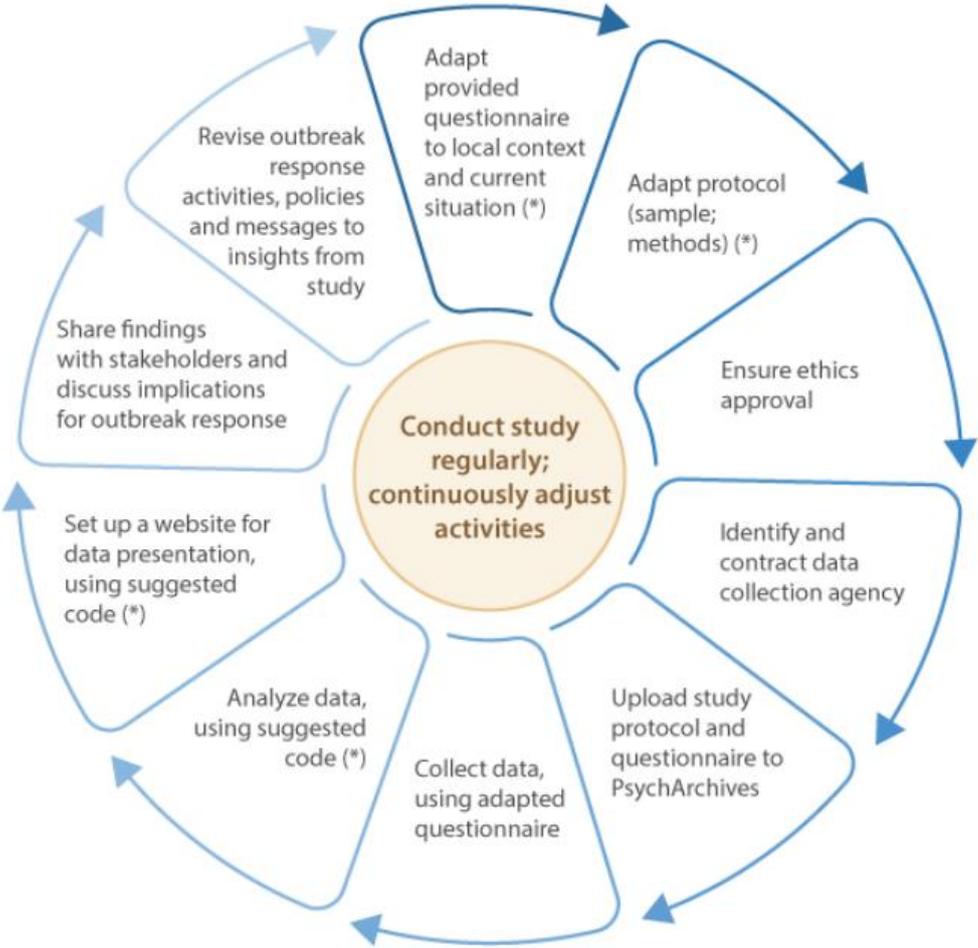
The World Health Organization Regional Office for Europe published a manual dated July 29, 2020 as a guide to conduct research on behavioral insights on COVID-19 (World Health Organization Regional

⁴² Information about National Longitudinal Survey of Youth 1979 is available at <https://www.nlsinfo.org/content/cohorts/nlsy79>

⁴³ <http://www.cvoterindia.com/>

Office for Europe, 2020). The Regional Office underlined the importance of such surveys because 1) researchers can construct scenarios about the future of the pandemic and develop necessary measures, and 2) suggest effective evidence-based measures. The manual has recommendations on how to initiate such studies, contacting the Regional Office and implementing partners, and technical details (Figure 3.7 for all steps).

Figure 3. 7 Steps and processes recommended by WHO Regional Office for Europe for behavioral insight studies



Source: World Health Organization Regional Office for Europe (2020)

The standard data collection mode recommended by the Regional Office is an online survey that takes 15-20 minutes, and CATI is suggested as an alternative. Since certain subgroups of the population are expected to be excluded with online surveys, CATI is also mentioned as a potential supplement, overall indicating a mixed-mode design. Household surveys requiring face-to-face contact is not recommended.

While there is no emphasis on probability designs in the manual, the necessity of voluntary participation is underlined. Quota samples are recommended, where the basic characteristics of the sample should match those of the general population (age, sex and residency). As for sample size, a size of about 1,000 participants is recommended at the national level. In terms of study design, repeated cross-sections are recommended. Through the repetitions, it is anticipated that developments over time, identification of emerging issues, and responses to measures taken against

the pandemic could be observed. It is recommended that no participants enroll twice so that anonymity is not compromised and that no duplicates happen.

The guide also has a recommended tool, which includes questions on socio-demographics, personal experiences with COVID-19, health literacy, COVID-19 risk perception, preparedness and perceived self-efficacy, prevention behaviors, effects on mental health, use of information sources, trust in information sources, , frequency of seeking information, trust in institutions, perceptions on policies and interventions, conspiracies, resilience, testing and tracing, fairness, lifting restrictions, unwanted behavior, well-being, and COVID-19 vaccine.

As we mentioned briefly in Section 3.1, the Task Force on COVID-19 and Household Surveys is operating under the United Nations Intersecretariat Working Group on Household Surveys. This Task Force aims to enhance the coordination of COVID-19 impact surveys among countries, support countries to maintain regular surveys and provide them guidance for COVID-19 related surveys, and help develop a collective vision for the implications of the pandemic regarding national survey programmes (United Nations Intersecretariat Working Group on Household Surveys, 2020c). Per the second aim of the Task Force, many countries are receiving support from ISWGHS for sample surveys to measure the impact of COVID-19. The ISWGH support by member institution can be seen on a country basis on the task force's website (United Nations Intersecretariat Working Group on Household Surveys, 2020a). For example, for Turkey, the institution that made contact with relevant institutions in Turkey are the Food and Agriculture Organization of the United Nations (FAO), UN Women and United Nations Development Programme (UNDP). In India, for instance, these institutions are FAO, International Labor Organization (ILO), UN Women and World Bank.

Among UN Women's releases, there is a guidance document titled "Rapid gender assessment surveys on the impacts of COVID-19" (UN Women, 2020). This publication was prepared to support those wishing to study the gender-based impacts of the pandemic. It provides a justification of why such research is necessary, provides a sample questionnaire, and possible survey modes that can be used. The modes recommended included telephone, online via SMS link, and SMS. The timing was said to depend on the needs of research objectives. For instance countries could conduct a rapid assessment survey during or after the pandemic and then do several more for deeper insight. For online surveys it was recommended that respondents be given 3 days to respond and then receive reminders. For SMS surveys, the recommended time was 8 hours. Incentives were briefly mentioned with pros and cons. Short questionnaires , taking a maximum of 15 minutes were recommended. Core questions include employment, comparison of income before the pandemic and as of survey date, knowledge on COVID-19, time spent on domestic work, and access to basic goods and services. Additional questions on feelings of safety, employment and livelihood resources, coping mechanisms, and household food production were suggested.

As an example of a guide released by UNDP, a questionnaire for the digital Socio-Economic Impact Assessment toolkit for households was released in June 2020 (UNDP Crises Bureau, 2020). The questionnaire suggested for household surveys include the following modules: ID and location; demographic information; quarantine status, knowledge, attitudes, and practices; access to food and basic services; coping strategies; and prospects and recovery needs.

As another regional agency, UNESCAP, released a brief in July, including tips for national statistical offices (NSO) for conducting COVID-19 rapid assessment surveys. The rapid assessments surveys were

recommended when “business as usual” does not work, which fits the pandemic. There 5 tips included in the brief are: 1) reaching out to relevant partners to make sure that these surveys get conducted, 2) taking leadership of the process as NSOs, 3) supervising content and questionnaire design, 4) ensuring the availability of data collected to potential users, 5) promoting partnerships for conducting these surveys (including international, or public & private sector partnerships).

The World Bank also offers some resources, such as a guideline on sampling design of a high-frequency mobile phone survey of households (The World Bank, 2020a), a guideline on CATI implementation of such a survey (The World Bank, 2020b), a questionnaire (The World Bank, 2020c), and an interviewer manual (The World Bank, 2020d). The questionnaire includes a household roster, some household characteristics, knowledge regarding the spread of COVID-19, behavior and social distancing, access to basic needs and services (health services, food, drinking water, etc.), employment, income loss, remittances, the food insecurity experience scale, concerns, opinions, and shocks/coping mechanisms.

For readers interested in surveys about the impacts of the pandemic, a website by the Consortium for the Social, Behavioral, Educational and Economic Sciences, German Data Forum (KonsortSWD/RatSWD) has a collection of studies carried out to examine the effects of the pandemic. It allows users to filter by research method (qualitative, quantitative or mixed), data collection status (whether or not it is accessible or ongoing, etc.), discipline (social, economic, education, health, etc.), and survey design (self-selected, representative, other).

Another important database of COVID-19 related surveys is kept by the Societal Experts Action Network (SEAN). The SEAN COVID-19 Survey Archive covers sources from the US, as well as international ones, and limit their collection to probability surveys. The archive aims to connect policymakers with COVID-19 related research, and share research with the general public (World Association for Public Opinion Research, 2020). The archive allows users to search for research by various features, such as topic and date, and as of June, the archive included over 225 studies from 25 countries. Users can sign up for weekly reports on the archive’s website⁴⁴ and get further information from Gary Langer’s presentation in WAPOR’s June Webinar, or his interview in Survey Practice (Aizpurua, 2020).

Here are some examples of individual surveys that were discussed in the special issue of the SRM journal or in Webinars:

Schnell and Smid (2020) proposed a design to follow up the patients diagnosed with COVID-19 to learn about disease progression. The survey design aims to estimate the proportion of mild progressions and identification of symptom clusters. They recommended using data stored at municipal health departments as the sampling frame, which includes telephone numbers. Thus, researchers recommended that both phone surveys and phone recruitment to a web survey would be possible. Consent to contact could be obtained in the initial contact, and patients could be tracked on a daily basis throughout their quarantine period. Schnell and Smid (Schnell & Smid, 2020) also recommended that the social and economic impacts of the pandemic be studied through random samples of the general population. Since FTF surveys are not an option during the pandemic, they recommend contacting respondents through the mail and recruiting them by phone, mail, or web.

⁴⁴ <https://covid-19.parc.us.com/client/index.html#/home>

See Toh Wai Yu, the Chief Operating Officer of a Malaysian polling agency called Central Force, talked about the COVID-19 Snap Poll they conducted in the June Webinar or WAPOR (World Association for Public Opinion Research, 2020). The survey covered 17,779 people from 17 countries. It was based on an RDD sample of mobile phones. Individuals above age 18 were eligible. The questionnaire included the financial effects of the pandemic on respondents, opinions on government response, expectations on the economic impact of the pandemic, opinions on international cooperation, whether or not they thought the coronavirus situation was exaggerated, fear of the coronavirus, willingness to sacrifice certain rights to prevent the spread of the virus.

In the same webinar, Constanza Cilley, the Executive Director of Voices, an Argentina-based company that provides consulting and research services, talked about the COVID-19 related research they carried out at Voices (World Association for Public Opinion Research, 2020). She mentioned that they usually use face-to-face mode to ensure representativeness in the region but had to resort to online panels with social media recruitment, e-mail invitations to web surveys, online qualitative interviews and focus groups, and telephone surveys due to the pandemic. Their research covered a variety of issues, COVID and society, COVID and economy, COVID and mental health, and COVID and the future. In addition to providing statistics, Cilley and her team provided deep insight into people's experiences through their qualitative components.

In a survey developed as a result of the collaboration between the academia and Facebook, Kreuter et al. (2020)⁴⁵ included questions on symptoms, testing, contacts outside the home, as well as household financial situation, mental health, and residence. This study is mentioned in detail in Section 3.3. The European Social Survey (ESS), discussed in Section 3.1, released a call for questions on COVID-19 to add to round 10, and two proposals were selected, one on government authority and legitimacy and another on conspiracy beliefs and government rule compliance (European Social Survey, 2020c). Two rotating modules of around 30 questions each will be added to the core questionnaire in round 10. Another example is the Multiple Indicator Cluster Surveys Plus initiative (MICS Plus), where researchers resorted to phone surveys which covered topics as information sources on COVID-19, perceived risk of contracting the disease, economic impacts of the pandemic, and opinion on government response (UNICEF MICS, 2020b). This initiative was briefly mentioned in Section 3.1, and is thoroughly mentioned in Section 4, in the interview with Attila Hancioğlu, the Global Coordinator of MICS.

3.2.1 Use of ongoing face-to-face panel studies to carry out COVID-19 related surveys

Kühne et al. (2020) described the telephone survey they carried out under the German Socio-Economic Panel (SOEP) in the Special Issue of SRM as a response to the pandemic. SOEP is an annual longitudinal household survey and has been conducted since 1984. The new survey within SOEP is called SOEP-CoV, and its sample is based on the 12,000 persons in the SOEP included since 2018. For about 25% of households, telephone numbers were not available. The sample was allocated to 9 replicate samples (tranches), where certain characteristics of the population were considered for randomization and each tranche would resemble the population of German households. A step-wise fieldwork design was implemented, four tranches were interviewed in two, and five tranches interviewed in one-week intervals. Data collection began March 30. Through this design, researchers were able to obtain time

⁴⁵ The full list of authors is: Frauke Kreuter, Neta Barkay, Alyssa Bilinski, Adrienne Bradford, Samantha Chiu, Roe Eliat, Junchuan Fan, Tal Galili, Daniel Haimovich, Brian Kim, Sarah LaRocca, Yao Li, Katherine Morris, Stanley Presser, Joshua A. Salomon, Tal Sarig, Kathleen Stewart, Elizabeth A. Stuart, Ryan Tibshirani.

trends. Repeated cross-sectional tranches can be treated as pseudo-panels, allowing the isolation of causal processes.

The data obtained in SOEP-CoV was weighted through household non-response models based on three probabilities of participation at different stages. Then weighted data was post-stratified to compensate for coverage error, using the German Microcensus as external data. Poststratification was also done at the person-level, where the household respondent was not random but was rather the person picking up the phone and being willing to answer following that.

The questionnaire in SOEP-CoV, as the authors underlined, did not present the burden of collecting core socio-demographic information because most were already available from prior waves. Questions were on health behavior and health inequality, labor market and economic situation, social life, networks and mobility, mental health and well-being and social cohesion.

Lillard (2020) mentioned the collaboration named Cross-National Equivalent file project (CNEF) at Ohio State University, between various ongoing household surveys: Household Income and Labour Dynamics in Australia (HILDA), German Socio-economic Panel (SOEP), Italian Lives (ITA.LI), Japan Household Panel Study (JHPS), Korea Labor and Income Panel Study (KLIPS), Russia Longitudinal Monitoring Study-Higher School of Economics (RLMS-HSE), Swiss Household Panel (SHP), "Understanding Society" (UK Household Longitudinal Study) (UKHLS), and the Panel Study of Income Dynamics (PSID) from the US. The CNEF aims to harmonize data between these panels and took action to add standardized COVID-19 related questions to each panel following the pandemic. The first supplemental COVID-19 survey was conducted under SOEP, the details of which are mentioned above. ITA.LI, UKHLS, SHP, KLIPS followed, and HILDA, JHPS, RLMS-HSE, and PSID were either at questionnaire development or fieldwork phases at the time the paper was published. Lillard (2020) underlined the importance of *ex-ante* harmonization (harmonizing questions and responses and wording); however, it was also discussed that this was not always possible, because surveys in different countries had different funding schemes, and different data needs were reflected on survey tools. Thus, *ex-post* harmonization is also possible. All COVID-19 related data will be available by CNEF after all countries in the collaboration complete their surveys, through *ex-ante* or *ex-post* harmonization. Lillard (2020, p. 213) explained the need for this harmonization as follows: "The goal is to use cross-country and within-country differences in the timing and implementation of various policies to explain the economic and social outcomes people experienced."

A special survey was designed to investigate the effects of the pandemic on the elderly population under the English Longitudinal Study of Ageing (ELSA). The first wave of this survey was conducted in June and July, and the second started in November to end in December (English Longitudinal Study of Ageing, 2020). The questionnaire includes sections on the experience of COVID-19 and its symptoms, other health issues and behaviors, mental health, work, and financial situation, family and living arrangements, social interactions, and connections⁴⁶.

In addition to the above studies, we can list some studies/modules developed within the German Family Panel (pairfam), Panel Study of Income Dynamics (PSID), Understanding Society, and Refugees in the German Educational System, all of which we further explain in Section 3.4. Gummer et al. (2020) designed the pairfam COVID-19 survey within the pairfam. This was designed as a CAWI survey to take place between May-June 2020. The frame was the respondents eligible for the 12th Wave of pairfam,

⁴⁶ The questionnaire can be accessed from: <https://www.elsa-project.ac.uk/covid-19>.

where respondents would be invited through mail. The survey covered changes regarding employment, relationships, domestic work, well-being, parenting, and alike. Sastry et al. (2020) mentioned that a short module would be added to the follow-up efforts of the Child Development Supplement under the PSID in the fall to investigate the effects of the pandemic on children and families. Burton et al. (2020) started a new survey called the Understanding Society COVID-19 under the Understanding Society longitudinal survey to explore the immediate effects of the pandemic. Some questions were added to existing questionnaires in the ReGES survey (Will et al., 2020).

3.2.2 Use of ongoing online panels to carry out COVID-19 related surveys

Blom et al. (2020), from Germany, started implementing the Mannheim Corona Study (MCS) within the German Internet Panel (GIP)⁴⁷. The researchers underlined the rapid need for data at the beginning of the pandemic to be able to observe the changes to this data later on the pandemic. The German Internet Panel is a probabilistic panel to which people are recruited through a variety of address-based methods. Three rounds have been done so far (2012, 2014, and 2018). Respondents are invited to a survey once every two months, with a 4 Euro incentive per completed survey.

The Mannheim Corona Study started within this panel on March 20, aiming to understand the impact of the pandemic on a variety of issues; employment and finances, childcare, attitudes towards politicians/democratic processes/measures, feelings and fears, evaluation of health risks, and people's social, health and risk behaviors. It was concluded on July 10, 16 weeks later. About 6,000 persons on the GIP were randomly assigned to one of the 8 subgroups. While each of the seven subgroup would be invited to the survey on a different day of the week, one would be left out as a control group. The questionnaire was reviewed every week for edits and additions. Respondents were given 48 hours to complete but were encouraged to reply within 24 hours, to be paid 2 Euros for participation. Results are published daily on MCS's website (*Mannheim Corona Study*, 2020), with each respondent classified according to the day of response, rather than the day of invitation. All findings are weighted, with two components to this weight: 1) response propensity weights (based on GIP characteristics), 2) raking (based on German Microcensus). Weights are trimmed to be within the 0.25-4 range. Response rates were highest on Mondays and lowest on Fridays, and people usually tended to respond on the first day of the invitation. Data over 9 weeks showed little changes in response rate across weeks.

The data was also assessed for measurement errors through Average Absolute Relative Bias (AARB). Bias was assessed using the German Microcensus for common variables such as age, gender, education, household size, marital status, region and citizenship. A daily analysis of this bias indicated that Mondays showed the highest accuracy, and Fridays the lowest, similar to response rates. The bias was said to be mostly due to a lower representation of non-German citizens, the youngest age group of 16-29 and lower education participants in the panel.

Another example where an ongoing online panel was used is from the Understanding America Study (UAS). In addition to the paper by Kapteyn et al. (2020) published in the special issue of SRM journal, this study was also addressed in the PAA Webinar titled "COVID 19 - Data Collection in Longitudinal Studies" by Arie Kapteyn (*Webinar - COVID 19 - Data Collection in Longitudinal Studies*, 2020). The UAS is one of the most significant internet panels, pursued by the University of Southern California, which

⁴⁷ Annelies Blok talked about the Mannheim Corona Study at the Webinar organized by the International Program in Survey and Data Science at the University of Mannheim earlier in 2020: <https://survey-data-science.net/webinar-surveys-corona>

investigates over 50 survey topics, including retirement planning, economic well-being, psychological constructs (Alattar et al., 2018). Representative of the entire United States, the sample comprises individuals aged over 18 in the contacted households. UAS respondents respond to the surveys via a computer, tablet, or smartphone⁴⁸. The survey enterprise has launched the Understanding Coronavirus in America to study the effects of the COVID-19 pandemic in March 2020. Through this survey, it was possible to track: diagnosis, symptoms, protective-behaviors, mental health, discrimination, social contacts who have been infected, opinion on percent chance of getting infected in 3 months and percent chance of death if infected, use/trust of information sources, opinion on likelihood that vaccination will be invented & whether respondent would pay for it, employment/working from home/ job loss, financial insecurity, health service seeking behavior, social behavior, stigmatization, social support, food insecurity, crime (Kapteyn et al., 2020). The UAS includes 9,000 individuals aged 18 and above, recruited through address-based methods. Addresses were selected in two batches. The first batch was an SRS address sample, and an adaptive design algorithm was used in the first stage selection (selection of zip-codes) of the second batch, which would bring sample characteristics closer towards population characteristics. Weights are produced to compensate for unequal probabilities of selection due to the adaptive design and post-stratification purposes. Households without means of access to the survey are provided a tablet and internet connection during the course of the survey. Respondents receive about two surveys per month, with an incentive of 20 dollars per 30 minutes spent on responding to interviews. English and Spanish languages are available.

The UAS started tracking COVID-19 related issues on March 10, and ended the collection of data for this first round on March 31. Towards the end of this round, respondents were asked to participate twice a month in the COVID-19 survey, with about 7,000 respondents agreeing. Thus, a new round began on April 1. The sample was approximately divided into 14 groups, with a different group invited to the survey every day. The assignment to the 14 groups was a nested stratified design, taking residence, age, sex and working at survey date into account. Each respondent had 14 days to respond but received an additional incentive (1 dollar) to fill out on the day of the invitation. A short-form survey was also administered half-way between the 14-day periods for Los Angeles County, funded separately. Results are presented on the study's website, updated weekly (The USC Center for Economic and Social Research, n.d.).

Sakshaug et al. (2020), in their paper published in the Survey Research Methods journal, summarized the surveys they carried out for COVID-19 related issues at the German Institute for Employment Research. These are the IAB High-Frequency Online Personal Panel (HOPP) and the IAB-BAMF-SOEP COVID-19 Survey of Refugees (CSR) (More information is available in sections 3.2.2.1 and 3.2.2.2).

In the June Webinar of WAPOR, the CEO of a company named Novus Group based in Sweden, Torbjörn Sjöström, talked about their Corona research (World Association for Public Opinion Research, 2020). Sjöström underlined that the pandemic meant "business as usual" for them because their modes of research were already telephone and web. However, he mentioned that operations were still harder to carry out and that qualitative research would be more significantly affected, with focus groups, for instance, moving to online. For their COVID-19 survey presented in June at the webinar, they used members of a probabilistic online panel. About 33,500 people were interviewed from March 18 to May 20, with a response rate of 73%. More than half (60%) of the respondents used mobile phones to respond to the survey. The questions they asked included topics such as keeping a distance from

⁴⁸ Further information is available at the survey's official page: <https://uasdata.usc.edu/index.php>

others, being worried about the virus, self-assessments of respondents regarding whether or not they have been infected, if yes, when and for how long, etc.

Ipsos conducted a survey called Ipsos Global Health Service Monitor between September 25 and October 9 in 27 countries (Ipsos, 2020b), covering ages 18-74 in some countries and ages 16-74 in others. The survey utilized the Ipsos Online Panel consisted of approximately 1,000 users per country. Ipsos underlined that while the samples were representative of their corresponding country and age group, others were biased towards more urban and more educated groups and that weights based on recent census data have been used. The survey covered a variety of health issues, including the pandemic, which was viewed as the top health problem in 26 out of 27 countries. The questionnaire also had sections on health concerns, healthcare perceptions, and healthcare challenges. In addition to this survey, Ipsos has a long list of issues, asked globally, using their online panel throughout the year (Ipsos, 2020d).

In the U.S., the Gallup Panel has been tracking COVID-19 related issues. The Panel was set up in 2004, and reaches out to respondents using various modes, telephone, web, or mail (Ipsos, n.d.). Sample selection is probabilistic, with RDD and address-based methods. The panel asked respondents about social distancing, schooling of children, media consumption habits, perceptions on self-protection from COVID-19, concern about health services and testing, pessimism about the pandemic, vaccine-related topics, wellbeing, work situation/financial effects of the pandemic, policies, etc. (Brenan, 2020).

Vezzoni et al. (2020), using the opt-in online community of a commercial research institute, devised a rolling cross-sectional survey to last 90 days. The design consists of randomly assigning respondents to days with the aim of interviewing new ones every day. So long as the number of cases is sufficient, this design would provide temporal changes over time. Vezzoni et al. (2020), did underline that while the findings from this survey cannot be generalized to Italy as a whole due to its opt-in design, as well as the general coverage issues of online surveys, it would still portray time trends in the phenomena studied regarding the pandemic. Also, it would be possible to analyze the dynamics of a mean or proportion, and to link the data with contextual data, such as data from the media on a daily basis. The survey started out with 38,000 individuals, expecting a response rate of 35%.

Respondents were invited through e-mail, starting April 6, with invitations remaining valid until the end of the data collection period. The researchers mentioned that this design were in line with the recommendations of the World Health Organization: 1) the tool is evidence-based, 2) it is fast and can be applied regularly, 3) it is simple and flexible, 4) it is cost-effective. They also mentioned that their design could be repeated to re-interview the same respondents allowing for further depth.

The questionnaire took about 20-25 minutes to fill out and had two sections. The first section was the core questionnaire, and the second one includes rotating modules. The core questionnaire included socio-demographic characteristics (age, gender, education, employment, etc.) and questions on compliance with recommendations to prevent the disease, perception of risk regarding the pandemic, well-being, attitudes towards economy and government, etc. The rotating modules were about religious attitudes, European solidarity, economy and health trade-offs, socio-economic costs of the crises with a gender perspective. In their paper published in the special issue of the SRM, researchers exemplified outcomes from their study.

3.3 Innovations in survey design and exploitation of new data sources for pandemic related research

Although web surveys by no means point out to a “new” survey mode, it is an area where innovation is the norm for reaching and recruiting respondents, designing questionnaires, and drawing inferences. With the onset of the pandemic, we observe methods of remote data collection gaining momentum. Researchers increasingly resort to data on social media and web pages as well as data from mobile phones, GPS data, and transportation usage (Sheikh et al., 2020), particularly for investigating issues concerning COVID-19.

One example is from Brazil, with innovation in its qualitative component. De et al. (2020) proposed a mixed-methods design to study the impact of the pandemic on vulnerable populations for the city of Curitiba. The authors argued that although vulnerability was often defined in economic terms, it was also related to social disadvantage, low literacy, addictions, and functional limitations in the case of the pandemic. Their mixed-method design incorporated in-depth interviews of vulnerable persons – conducted via telephone rather than traditional face-to-face mode, online surveys of the general population of the city, and online focus groups with key actors in the community life of the city. The first two phases were already completed at the time of publishing. For the in-depth interviews, researchers identified the neighborhoods where vulnerable populations lived, based on publicly available social indicators (income, income distribution, education, ethnicity, etc.). Phone numbers of potential community leaders were obtained through social connections, who referred them to regular citizens after the interviews. The online survey was non-probabilistic, yet efforts were made not to have vulnerable people underrepresented in the survey, using social media channels and telephones obtained through the first stage of the research. The last phase, a combination of online focus groups and alternative scenarios approach which would last 1 hour 30 minutes, was planned with various persons from Curitiba, when the paper was published.

3.3.1 Use of social media data

Kreuter et al. (2020), as briefly mentioned in Section 3.2, designed a global COVID-19 survey available in 56 languages, partnering with Facebook. With the need for urgent high-quality data for policymakers, they underlined the benefits of existing information systems, such as mobile phone data, mobility or bank card data. Such data, researchers mention, is useful but only to a certain degree because they cannot be linked with the health status of the persons whose information is available, making surveys still a necessity. The initiative described in Kreuter et al. (2020) consisted of researchers from the following institutions: The World Health Organization, the Delphi Group at Carnegie Mellon University, the Joint Program in Survey Methodology at the University of Maryland, Harvard, Stanford, Yale, and Johns Hopkins University. The authors identified four main issues in describing the survey: 1) privacy, 2) content, 3) sampling and nonresponse, and 4) analysis. Privacy and content are mentioned in Section 3.6. In terms of sampling and nonresponse, the authors underlined that the sampling design and the methods for correcting bias were the strengths of their survey compared to a voluntary internet survey. They drew a new random sample of Facebook users every day, stratified by region, using simple random sampling within strata. Since not everyone is a Facebook user, and not everyone sampled responds, Facebook developed weights. The weights were calculated to represent countries in terms of age, gender, and administrative region. For analysis, authors underlined that data from API systems would be about two days old at any given day because of the time it took to produce weights and prepare the data. No data was presented for locations with few observations. Munzert and Selb

(2020), in their commentary on this survey, underlined the rapidness and relevance of this survey. They argued that coverage was of particular importance in countries with low literacy levels, and discussed potential overestimation of disease symptoms due to measurement error.

Rinken et al. (2020) designed a mixed-mode study, using two different sampling strategies and a combination of sampling techniques for an online survey to observe the effects of the pandemic and non-pharmaceutical interventions in Spain, when these interventions were at their most strict level. One approach was to use a random sample from a mobile phone frame for SMS recruitment to the online survey, and the other approach was based on internet advertising. At the time, researchers had an estimate of mobile phone penetration of 79% for the population above 18. However, coverage was significantly lower for persons above age 65 (37%). Their frame included over 93 million numbers, and almost half of these were active in fall 2019. SMS recruitment invited randomly selected phone number owners to a web link. The SMSs were not sent out at once so as to protect the service against traffic. Researchers stopped the SMS sending procedure after 51,046 SMSs were sent. Their analysis of the sample revealed that some age groups were underrepresented, besides lower education groups. For the second approach, Rinken et al. (2020) used Google Ads along with Facebook/Instagram, as paid advertising services. While Google Ads covered all internet users, it differed for Facebook and Instagram, with the latter covering a younger population. This second sampling approach, the researchers underlined, was both open to coverage error and self-selection bias. The researchers did not initiate this second survey at the same time as the first but rather waited to monitor the progress in the SMS version first, similar to a responsive design approach. Once the age and education distributions were obtained from the SMS recruitment group, researchers recruited online respondents from lower education groups, ages below 30 and above 65. The overall response rate was below 1% after the advertisement was shown to 1,187,580 people, and 994 people responded. One interesting finding was that, contrary to expectations, people with lower education were again underrepresented in this sample. Researchers argued that this may indicate incorrect education recording on apps. This was not the case for age, with the advertising sample mostly compensating for people above age 65, more so for men than women. In the final sample, researchers applied weights that iteratively adjusted for the following: municipality size, NUTS-1 regions, age groups, sex and education levels.

Perrotta and et al. (2020), in their working paper, inquired about health behaviors and attitudes concerning COVID-19 in the “COVID-19 Health Behavior Survey” utilizing Facebook Ads Manager to recruit participants from eight countries (Belgium, France, Germany, Italy, the Netherlands, Spain, the United Kingdom, the United States) as a cross-sectional survey. A total of 71,612 participants responded to the questionnaire which involved sections regarding socio-demographic and health indicators, opinion, and behaviors, as well as one that examined social contacts. The authors underscored the possibility of an approximate representative survey since they were able to recruit respondents in a balanced manner across demographic subgroups, owing to the “controlled environment of targeted advertisement” and post-stratification weights based on data from Eurostat and the US census. They also highlighted that utilizing social media platforms such as Facebook and appropriate survey designs and statistical tools brought innovation and speed in terms of collecting behavior data on COVID-19.

Grow et al. (2020) also discussed how Facebook functions as a participant recruitment tool in reference to the COVID-19 Health Behaviour Survey, wherein they addressed Facebook’s contribution to the

timely design, that it enabled drawing samples from diverse-part of populations from different countries and its cost-effectiveness. The authors argued that self-selection bias was the prominent challenge when Facebook was used for sampling; and suggested stratifying ad campaigns, employing post-stratification techniques, considering possible sources of bias against these challenges (Grow et al., 2020).

Work by Mejova and Kalimeri (2020) is an example of a study where Facebook data is exploited through means other than respondent recruitment. Here the authors investigated advertisements related to COVID-19 with a view to exploring Facebook Ads usage on alternative narratives about coronavirus and misinformation. To that end, the authors analyzed 923 ads from 34 counties through keyword queries of “coronavirus” and “covid-19” on Facebook Ads. Another recent paper, furthermore, examined the urban-rural divide in Italy, using Facebook Ads as the source while also examining its reliability for estimation vis-à-vis official demographic figures (Rama et al., 2020).

3.3.2 Use of Google search data and GPS mobility data

One of the most famous examples of the use of big data was the success of Google Flu Trends in estimating flu trends in the US (Japek et al., 2015). This information, based on Google search data of certain symptoms and remedies, provided good estimations of flu incidence between 2009 and 2011, but was off for the years 2012-2013. This source has been employed for COVID-19 related research as well. Kurian et al. (2020) looked at the association between COVID-19 cases and Google search data of 10 keywords, and found that face mask, Lysol and COVID stimulus check had the largest correlations for the US. Another article from Nature looked at the predictability of COVID-19 in the US (Mavragani & Gkillas, 2020). Researchers found the correlation between Google Trends and COVID-19 data to be significant. Another article looked at Google search data for gastrointestinal symptoms and hospitalized COVID-19 cases (Ahmad et al., 2020). The observation here was that the search for such symptoms were correlated with the number of cases, with a 4-week lag. One study looked at the search data for loss of smell, to see if it increased during the epidemic in several countries; namely Italy, Spain, UK, US, Germany, France, Iran and The Netherlands (A. Walker et al., 2020). In all countries, researchers observed positive correlations between searches and confirmed positive cases/confirmed COVID-19 deaths. Another study from Iran, using linear regression and long-short term memory models to estimate positive cases found that the searches on handwashing, sanitizers and antiseptics to be effective factors of incidence (Ayyoubzadeh et al., 2020). A very recent article looked at Google Trends data and daily incidence of COVID-19 based on WHO data for the early months of the pandemic (January-April) (Peng et al., 2020), using machine learning algorithms. The researchers concluded that their model could be used to predict epidemic alert levels.

Mobility data is another focal point of research enterprises during the pandemic. For instance, Pepe et al. (2020) presented a dataset that aggregated the origin-destination movements between Italian provinces, the radius of gyration, and the average degree of a spatial proximity network to monitor the lockdown’s influence on mobility in Italy, using data of 170,000 de-identified smartphone users, which was provided by a location-intelligence and measurement platform. A recent study of Badr et al. (2020) used aggregated and anonymized cell phone data as well, for inquiring about how social distancing alter the number of infection in the 25 states of the USA where they suggest that social distancing is the foremost tool for fighting against COVID-19. Mobile data can also serve public health actions during the pandemic, as Oliver et al. (2020) argued in their recent editorial. Accordingly, technology companies and governments should collaborate for mobile data to take appropriate steps

to tackle the pandemic (Oliver et al., 2020). Further, another study employed mobility data provided by Facebook to investigate how the lockdown affected the socio-economic status of Italians (Bonaccorsi et al., 2020).

Google Mobility Reports also appears as a robust data source for studies on the impact of Covid-19. By way of illustration, Basellini et al. (2020) explored the adverse relation of excess mortality and time spent at home with a positive association of outdoor mobility, utilizing Google COVID-19 Community Mobility Reports on England and Wales. Abouk and Heydari (2020), on the other hand, studied the impact of stay-at-home policies on social interaction in the USA based on Google Mobility Reports demonstrating the effectiveness of the social distancing policies. They also marked that Google's database only enabled to monitor people who agreed on sharing their location histories in their smartphone settings which limited the study. In a similar vein, another study that used Google location history as a database, concluded that state-wide restriction on social interaction was significantly effective in the USA during the pandemic (Wellenius et al., 2020).

Moreover, Saha, Barman, & Chouhan, (2020) employed Google COVID-19 Community Mobility reports assessing the effect of lockdown in states and union territories of India as well as illustrating how mobility altered in residential, groceries, pharmacies, parks, retail and recreation, and workplaces. Comparing 131 countries based on 131 countries' Google COVID-19 Community Mobility reports, Morita et al. (2020), in their working paper, concluded that social distancing policies and socio-economic conditions of a country played critical roles for the ratio of behavior changes.

3.4 Surveys to estimate the prevalence of COVID-19

All countries register certain events such as the number of COVID-19 tests, positive cases, and deaths, which are used for obtaining certain statistics, such as R, the basic reproduction number of the infection. Schnell & Smid (2020) discussed the shortcoming of indicators obtained through such widely available statistics. They underlined that testing was often not done at random, and under which circumstances they were done depends on the protocols of different locations. As Frasier et al. (2020) pointed out, it was often people with symptoms that were tested. The proposed solution is general population surveys for estimates of prevalence (Schnell & Smid, 2020; Sperle-Heupel et al., 2020), and random testing (Frasier et al., 2020). As Frasier et al. (2020) explain, different tests are available. There are antibody tests (namely serologic testing for SARS-CoV-2 antibodies) requiring blood samples from veins. There are also diagnostic tests based on swab specimens, such as PCR (polymerase chain reaction) tests. Different survey designs may involve different types of tests. While the former produces quicker results, and are cheaper, they are less reliable than the latter, with good specificity but rather moderate sensitivity (OECD, 2020).

The World Health Organization published a document as a guide for COVID-19 studies in March, titled "Population-based age-stratified seroepidemiological investigation protocol for COVID-19 virus infection, 17 March 2020" (World Health Organization, 2020a), which was updated in May (World Health Organization, 2020b). These documents suggested two main reasons for conducting such research, 1) to estimate the extent of the infection, and 2) to determine the proportion of asymptomatic cases. Other objectives included determining risk factors and obtaining an accurate ratio of deaths to cases⁴⁹. In terms of sample design, these documents suggested cross-sections, repeated cross-sections, or longitudinal cohort studies, depending on resources available. In the first

⁴⁹ The second document also underlines an improved understanding of antibody kinetics following infection.

protocol, there was no emphasis on probability samples, and convenience samples were recommended. The second protocol discussed both convenience and random sampling, where random digit dialing and address-based household surveys such as DHS and MICS were mentioned as potential sampling frames. Readers were directed to online tools for sample size calculations. In terms of within-household sampling, the WHO did not recommend any, but rather suggested selecting all individuals regardless of age within households so that age-specific rates could be obtained. The second protocol had a larger emphasis on ages, suggesting stratification by age groups when possible. In addition to specific recommendations on medical tests, WHO suggested a questionnaire be implemented for everyone for whom specimens were taken, which would include demographic, clinical, and exposure information.

The rest of this section summarizes two design proposals from the special issue of SRM, provides two individual examples (one from Britain and one from Brazil) and mentions a few self-report or app-based studies. Another example of a prevalence survey can be found for Turkey in Section 3.5.

Schnell and Smid (2020) proposed a survey design for antibody testing in Germany. They calculated sample sizes for Germany for both a cross-sectional study and a longitudinal one as a suggestion for a national design. They considered stratification and neglected design effects due to clustering, underlining that their proposed sample sizes should be taken as a minimum. For data collection, they recommended an initial contact with mail, and suggested four different methods for venous blood collection: 1) asking the sampled citizen to visit the GP, 2) a pre-selected GP in the neighborhood, 3) municipal health department or 4) blood donation locations of the Red Cross. Incentives were recommended both for respondents and doctors taking part. Researchers underlined that non-response bias for this kind of survey was a serious threat; thus alternative modes of contact and non-response incentives should also be planned. Consent for follow up was also recommended during a PAPI interview prior to data collection⁵⁰. Schnell and Smid (2020) also recommended a large-scale, random post mortem examination to determine the cause of death during the pandemic.

Frasier et al. (2020) proposed a mail survey design to estimate the prevalence of antibodies for COVID-19 in the US. They argued that the mail mode would be safe, feasible, and also cheaper than the face-to-face mode. The addition of questionnaire data would provide information on demographic subgroups of interest, which would also include a request for consenting to a re-contact for the survey. They anticipated that the mailed envelope would include a letter, a questionnaire, a self-testing kit (with latex gloves added), and a prepaid envelope. The testing kit could either be sent to a lab or could be a type of test to reveal instant results. Authors suggested live video chat support be available to respondents, a toll-free number for assistance, a website for the survey, as well as a follow up phone call. Follow-ups to non-responders were recommended, yet only to include letters and not testing material. The details of the testing are further discussed in the paper.

Frasier et al. (2020) designed three different sample sizes assuming different magnitudes for the prevalence of antibodies: 1%, 5% and 10%, where the former would provide a 95% CI of 0.63-1.42 with a sample size of 5,000, which would increase according to the number of subgroups for which separate estimates were required. The sample size would also be adjusted for non-response and non-viable samples. For within household selection, researchers suggested three alternatives, with a preference for the first option: 1) randomly selecting a member using methods such as the last birthday method

(harder to achieve in self-administered modes), 2) letting the household decide who participates (risk of getting someone with symptoms tested), 3) testing all members (high costs). André Grow, Daniela Perrotta, Emanuele Del Fava, and Jorge Cimentada, in their commentary within the paper by Frasier et al. (2020), underlined coverage issues and cost-effectiveness about this design. They also underlined that false-positives and false-negatives be taken into consideration in the sample size.

Hallal et al. (2020) designed an antibody prevalence survey for Brazil, which they said was the largest population-based study on the prevalence of antibodies for SARS-CoV-2 in geographical scope, at the time of the article, to the best of their knowledge. They carried out two cross-sectional sample surveys, one between 14-21 May and another between 4-7 June. There are 133 intermediary regions under the 27 federation units in Brazil, and the researchers selected the most populated city from each region, calling them sentinel cities. In each sentinel city, they selected 25 urban census tracts, and within those, they randomly selected ten households. All household members were listed, and one was randomly selected⁵¹, where another member would be randomly sampled if the person initially sampled could not provide a blood sample. If the second person refused as well, the teams went on to the next address to the right. The test used was “a rapid point-of-care test, the WONDFO SARS-CoV-2 Antibody Test (Wondfo Biotech, Guangzhou, China)”, which required finger prick blood tests requiring two drops of blood. In addition to testing, there was computer-assisted personal interviewing. Interviewers used tablets to record data, randomly select a respondent, photograph the test results, and obtain GPS coordinates. Moreover, all interviews were voice recorded using the tablet, and 10% of all interviews were listened by supervisors.

In the first cross-sectional survey carried out in Brazil, interviews were conducted in all regions but one, which was excluded due to a lockdown (Hallal et al., 2020). A total of 25,025 households/individuals (54% of all contacted households) were interviewed, which was lower than what researchers had anticipated, due to interviewers having restricted mobility due to the pandemic, bureaucratic reasons, and also due to conspiracies online about interviewers being swindlers, or being sent for spreading the virus. These conditions improved in the second survey, where again one region could not be visited due to lockdowns. A total of 31,165 persons were tested (53% of contacted households). In addition to findings on various population subgroups, the authors reported that antibody prevalence increased by over 50% between the two surveys, within a period of about 2-3 weeks. The authors recommended longitudinal, or panel studies, underlined the shortcoming that the rural population was not covered and that the most developed cities were included.

Another example of a prevalence survey is from Britain. A large scale study is currently ongoing in the UK, running with the short title COVID-19 Infection Survey (A. S. Walker et al., 2020). The study aims to collect swab samples from all members of sampled households aged 2 and over for COVID-19 prevalence and aims to collect blood samples from a subsample for seroprevalence. Researchers underlined the importance of collecting samples from children in their design because of the need for understanding symptomatic and asymptomatic infection among them and provide inputs for measures to be taken about schools. The swab samples are to be collected by the respondents themselves (and by adults of the households for children under 12) to reduce the risk of infection to the study teams. The sample of the survey is based on prior participants of surveys conducted by the Office for National

⁵¹ Whenever a person was tested positive, all other members of the household were invited to be tested – not to be reflected on overall prevalence estimate. However, this information was used for reflecting family clustering.

Studies (ONS) or Northern Ireland Statistics and Research Agency, who provided consent for future surveys. Invitation letters are sent prior to visits. The survey is to include new cross-sections of households every week. If the interviewed households provide further consent, they are planned to be visited for four weeks in a row, and then every month until one year passes since the initial interview. Individuals who are not at home in the first visit are not included in the sample at later visits, and neither are individuals who later move in with the sampled household. A total of 176,602 households are expected to be recruited by the end of March 2021, corresponding to 380,292 individuals. These expected sizes are adjusted for researchers' anticipation of a 50% household response rate. Short questionnaires are designed to accompany the tests, based on WHO recommendations we mentioned earlier in this section (World Health Organization, 2020a), including knowledge of infection symptoms, health status, contacts, smoking, etc.

In addition to household surveys that provide direct estimates of COVID-19 prevalence, there have also been extensive efforts to come up with predictions of this measure using self-reports. As an example of an indirect study of prevalence, Allen et al. (2020) developed a mobile application to track COVID-19 symptoms in a longitudinal manner in the United States, called How We Feel (HWF). After filling the basic demographic information, participants are expected to answer questions every day about symptoms, testing, behavior related to virus transmission, etc. The authors underline that they drew the sample from a large user base and that it was non-probabilistic. The app had about half a million users in May, a month after it was launched. The study allowed the researchers to obtain the determinants of testing, determinants of being tested positive, as well as using models to predict COVID-19 positive cases. The researchers used data from participants who tested positive and used machine learning to develop prediction models. The app How We Feel also had questions on contact tracing, which is widely employed for watching spread (Allen et al., 2020).

An application called "Life Fits Into Home" was developed by the Ministry of Health in Turkey, which includes a daily question for users to evaluate their overall health on that particular day (Yazicioğlu, 2020), and asks about three basic symptoms in case users report themselves as unwell (Figure 3.8). The app also alerts users of the presence of COVID-19 positive cases if these cases also have the app installed, running, and both parties have Bluetooth on. As of November 2020, we are unaware of any surveys that employ the registered users of this app as a frame or unaware of any big data analysis on the data produced by the app.

Figure 3. 8 Screenshots from the app “Life Fits into Home”, Turkey

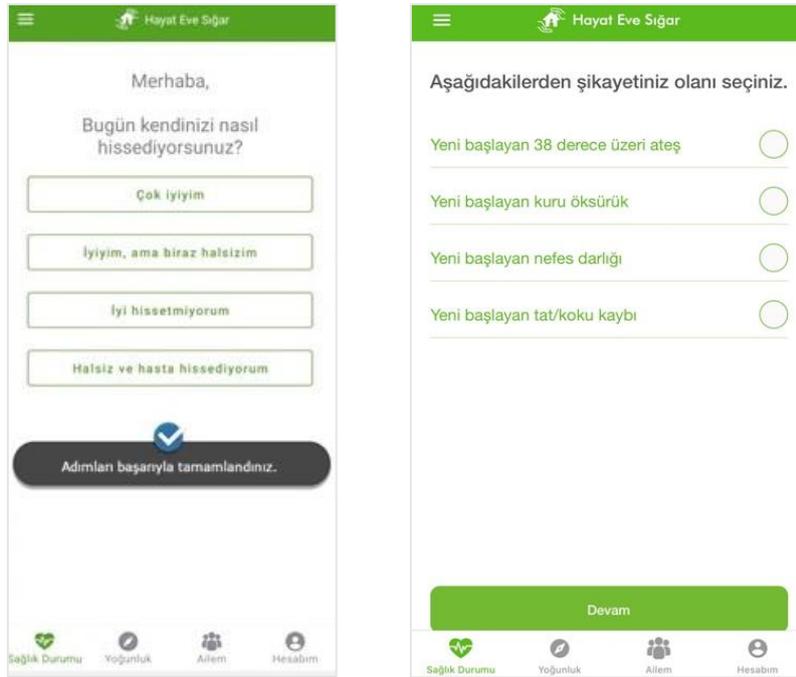


Image source: <https://www.brandingturkiye.com/hayat-eve-sigar-uygulamasi-nedir-nasil-kullanilir/>

The first image translates as: “How do you feel today? Very well/well but a little under the weather/Not well/I feel weak and sick”; the second image: “Please pick the symptoms you have. Body temperature just raised above 38 C/Recent dry cough/Recent shortness of breath/Recent loss of taste or smell”

For further examples, readers can see Azad et al. (2020), for a draft paper that reviews privacy issues in mobile applications designed for COVID-19 contact tracing. This paper includes a list of examples of apps that have been used for this purpose.

3.5 Surveys in Turkey and the COVID-19 Pandemic

Up until this section, we already talked about some survey research issues from around the world. Here we mention the COVID-19 prevalence survey, an example from TURKSTAT, some COVID-19 related surveys, and a brief overview of the activities of private research companies during the pandemic. A general observation regarding surveys in Turkey is the lack of documentation regarding the effects of the pandemic on methodology, both in public and private sectors. This gap is filled to some extent in Section 4 thanks to the experts we interviewed from the Turkish Statistical Institute (Enver Taştı), and the private sector (Oğuzhan Akyıldırım). This subsection includes a brief overview of what we could find.

The large scale COVID-19 prevalence survey was carried out in Turkey, whose sample was designed by the Turkish Statistical Institute, and whose fieldwork was implemented by the Ministry of Health. The aims were 1) to estimate the prevalence of the disease, 2) to estimate the prevalence of antibodies, and 3) to estimate the level of asymptomatic cases, with all three estimates at the national and provincial level (Çakır, 2020). The sample size was 153,577 households (*Türkiye Corona Virüsü Salgını*

Araştırmasını Nasıl Yürütüyor?, n.d.). It was reported that one PCR test and one antibody test was done per household, with a random household member being selected. Randomization was fulfilled as follows: Interview teams list household members in ascending order by age from 1 to n and then ask the household respondent to pick a random number within this range (Çakır, 2020). The interview teams were formed by the Ministry of Health, and local authorities were said to be active in notifying households for the tests. Participation in the survey was voluntary, and two levels of substitution was planned in case the selected household was not available. The tests were reported to be done at the doorstep for security reasons (so that risks of burglary be minimized with others pretending to be from the Ministry). A short questionnaire complemented the COVID-19 tests, including questions on COVID-19 symptoms, general health status (including reported height and weight), and tobacco use. Epidemiologists from the Turkish Medical Association⁵², at a webinar shortly before the start of the fieldwork, criticized the lack of methodological documentation, sample size (finding it too large, causing burden and risk for health staff), provincial allocation of the sample (in the absence of documentation of its calculation), the content of the questionnaire (lacking questions on retrospective social contacts, household members being tested positive, work status, etc.), testing one person per household (for missing data on within household transmission) and were unaware of ethics approval procedures. The methodological document and survey final report are not yet published as of the end of 2020. It was mentioned in the Ministry's website in October that the fieldwork of the second survey would be carried out between 15-30 October⁵³.

The Ministry of Health also announced in August that the tender for the "Coronavirus Barometer Survey to measure the level of perception and behavior of the public regarding the novel coronavirus" was completed and that the private company GENAR would be in charge (T.C. Sağlık Bakanlığı, 2020). The task is supposed to be complete in 270 days and involves three surveys, each with a sample of 7,680 households. One respondent per household will be selected randomly (age 18 and above) and interviewed with CAPI. Findings will be provided at the NUTS-2 level (Türkiye Cumhuriyeti Sağlık Sisteminin Güçlendirilmesi Ve Desteklenmesi Projesi (SSGDP) "Toplumun Yeni Koronavirüsle İlgili Algı Ve Davranış Düzeyini Ölçecek Koronavirüs Barometresi Araştırması" Danışmanlık Hizmeti Alımı İş Tanımı, 2020).

The Household Labour Force Survey in Turkey carried out by the Turkish Statistical Institute, was being conducted face-to-face until 2020. Just before the pandemic, it was planned that the first interviews would be conducted face-to-face, and the waves after by CATI. This shift already took place around January. A publicly available manual on the methodology of phone interviews and using phone numbers for first waves does not exist on the TURKSTAT website as of December 2020; however, the interview with Enver Taştı, presented in Chapter 4 covers the necessary details. This being said, there are no announcements regarding how the fieldwork for the 2020 survey is being conducted on TURKSTAT's website. However, for the annual Information and Communication Technology Usage Survey on Households and Individuals, it is underlined in the metadata section that the survey has been conducted via CAPI until 2020 and via CATI in 2020. No further details are provided regarding

⁵² The speakers were Banu Çakır, Gül Ergör, Onur Hamzaoğlu. The event can be accessed at: https://www.ttb.org.tr/kollar/covid19/yazdir.php?Tablo=tbl_haber&Guid=c4d2dd98-b20d-11ea-a732-c115216aae2a

⁵³ <https://www.saglik.gov.tr/TR,76505/covid-19-seroprevalans-arastirmasinin-ikincisi-basladi-21102020.html>

CATI, but it is very likely that procedures explained by Mr. Taştı in the interview, regarding matching phone numbers to addresses, apply to this survey as well.

A Kadir Has University survey, 2020 Perceptions on Turkish Foreign Policy Research, was conducted between April 3 and April 17 (*Türk Dış Politikası Kamuoyu Algıları Araştırması*, 2020), to investigate the perceptions of Turkish people on various foreign policy issues. The sample consisted of one province per NUTS-2 level, making a total of 26 provinces. The survey was designed as CAPI, but then researchers had to switch to online face-to-face interviewing due to the pandemic. The survey report did not cover how respondents were reached for online face-to-face interviewing. There was a total of 1,000 interviews conducted.

In the private sector, IPSOS has been carrying out regular surveys about the pandemic titled “Salgın ve Toplum”, “Pandemic and the Society”, where the first one was carried out between 11-17 March (Ipsos, 2020a). As of December 7, they released the findings of their 33th survey (Ipsos, 2020c) carried out between November 26 and December 1. The topics covered in these surveys included expenditure behavior, level of concern about the pandemic, things people miss, opinions on infection preventive behavior (mask use etc.), thoughts on the vaccine, schooling, opinions on curfews as non-pharmaceutical interventions, working from home, financial effects on the pandemic, and many others (Ipsos, 2020e). The Ipsos Household Expenditure Panel was used as the frame as the first report informs. The number of respondents to the first survey was 5,000 persons, and that of the last one was 800. CAWI is the mode of response. The Panel originally consists of 14,000 households from 35 provinces in Turkey, said to be representative of the country in terms of socio-economic status, household size and regions (Ipsos, 2020a). Expenditure data is collected from this sample on a weekly basis. Households are asked to keep a diary of their expenditures daily and report them weekly.

Çarkoğlu (2020) carried out a phone survey in İstanbul, using telephone numbers collected in a face-to-face, address-based survey from 2019. A total of 1,030 persons were interviewed. Comparing this sample to the original face-to-face sample, researchers found the phone respondents to have a slightly lower level of education, to be slightly older, and with slightly lower income. Among the questions asked included the contemporary problems faced by respondents (followed by unemployment, inflation, etc.). There were also questions on how the pandemic affected respondents’ economic status, their expectations for the year ahead, being worried about their health, chronic disease, COVID-19 symptoms, social isolation behavior, wearing masks, hand washing, evaluation of institutions regarding COVID-19 response, and alike. As of November 2020, we failed to locate a detailed report of this study that summarizes the sampling design of the previous address-based survey, shows response rates and provides details of the telephone interviews.

ORC Araştırma Eğitim Danışmanlık carries out qualitative and quantitative research, including electoral studies. Their study, called the “Halkın Pandemiye Bakışı ve Pandeminin Toplumsal Etkileri” (Public views on the pandemic and the effects of the pandemic on the society) was carried out between November 25-29, in a total of 42 provinces (out of 81) (ORC Araştırma Eğitim Danışmanlık, 2020). A total of 5,140 people were interviewed by CATI. There is no information provided on the frame or response rates on the company’s website. The questionnaire covered questions on what respondents think is the reason for the high number of COVID-19 cases in Turkey, how they evaluate the health system in Turkey, what measures they think could have controlled the pandemic, how they evaluate WHO’s role at the early stages when the virus was first spotted in Wuhan, how respondents evaluate the measures taken at public institutions, when they think the pandemic would end, if they accepted

any visitors during their time of quarantine (if they ever have been quarantined), if they have personally been affected by the pandemic (social bonds, eating habits), their evaluation of distance education for children, etc. ORC also carries out electoral studies. Their most recent study, *Siyasi Eğilimler Araştırması*, was carried out between November 26 – December 1, asking respondents which political party they would vote for if there were general elections held on the upcoming weekend. Their sample consisted of 3,820 participants interviewed face-to-face, but further details are not found on the website.

Areda Survey is a research company that employs quantitative and qualitative methods, where the modes they use for quantitative research include PAPI, CAPI, CATI, CAWI, and CAMI (computer assisted mobile interviewing) (AREDA Araştırma Eğitim Danışmanlık, n.d.). They have been conducting a series of studies on different aspects of the pandemic. Their first study was conducted in February, based on 1,100 respondents, before the first case was confirmed in Turkey (AREDA Araştırma Eğitim Danışmanlık, 2020c). They conducted two more surveys in March about perceptions regarding the pandemic (AREDA Araştırma Eğitim Danışmanlık, 2020a, 2020b). In the first one, carried out between March 13-16, there were 1,064 respondents recruited from the company's online panel⁵⁴ and interviewed with CAWI. The second survey followed the same mode with 3,754 respondents. These studies included questions on how people evaluate the performances of certain political figures, whether they fear contagion, what they think of curfews, how they situate Turkey among other countries in terms of the crises, etc. Other surveys conducted by the company include other COVID-19 related topics as the situation of the elderly, things people miss, experiences of youth, etc.⁵⁵.

The private research company Konda published a report titled *İstanbul Convention Research*, data for which was collected through their August 2020 Barometer (*Konda: İstanbul Sözleşmesi*, 2020). For the Barometer, despite the ongoing pandemics, field teams worked in 32 provinces where 3,569 face-to-face interviews with individuals age over 15 were completed in their households between August 8 and 9, 2020⁵⁶. The sample was based on the Address Based Population Registration System and was stratified with the results of the November 2015 General Elections, as the report describes. There were 18 interviews to be conducted in each quarter, with quotas by age and sex. The report did not mention how face-to-face interviewing was made possible during the pandemic, what measures were taken to keep interviewers and respondents safe, and if any ethics approval or Ministry approval was obtained.

The Turkey's Pulse poll series by the research company Metropoll concentrates on political and social tendencies, regularly shares the results with different themes. For the July 2020 version, the sample design was based at the NUTS-2 level, and the survey team interviewed 1,350 people using CATI between July 18-23, 2020 (Metropoll Stratejik ve Sosyal Araştırmalar, 2020). These surveys were formerly done using a face-to-face mode, and we confirmed with the company via e-mail that the shift from FTF to CATI was due to the pandemic. They explained that they had a large frame of phone numbers, yet we have no further details.

SONAR is another private research company that has been active during the pandemic, and here we will mention two of their studies conducted since March. SONAR conducted two surveys about the

⁵⁴ "Profil Bazlı Dijital Panel Araştırması" is what the company called their panel study, translating as "Profile Based Digital Panel Survey".

⁵⁵ See <https://www.areda.com/kategori/covid-19-ozel/> for the full list of studies by this company.

⁵⁶ There was no detail about the measures taken during this field operation. We sent an e-mail on November 12, yet did not hear back from the researchers as of the time of publishing this report.

assessment of the success of mayors in Turkey during the pandemic. The first one was in July and consisted of the assessment of the success of mayors of metropolitan provinces⁵⁷⁵⁸. This study was carried out in July in 20 metropolitan provinces and 39 districts in İstanbul. There were 1,000 respondents per metropolitan province and 600 respondents per district in İstanbul. The second one was carried out in August, focusing on other mayors, such as those of non-metropolitan provinces and district mayors excluding İstanbul⁵⁹. They interviewed about 500 respondents in districts of metropolitan provinces and about 800 respondents in non-metropolitan provinces. In both surveys, they used quota sampling, with criteria based on geographic location, age and sex. In the former study, both FTF and CATI modes were said to be used, and in the latter, it was purely CATI. The briefs published on the website do not include any details on the frame used for CATI and what measures were taken for FTF surveys.

Another example is the Türkiye Gündemi Araştırması (Research on Turkey's Agenda), which was carried out by MAK Danışmanlık research company in October 2020 (MAK Danışmanlık, 2020). Their sample covered 61 provinces (out of 81) and used purposive sampling (a combination of maximum variation sampling and snowball sampling) to reach respondents. PAPI was the mode of interaction, and photos were at times taken with the consent from the respondents. The report mentioned that consent was taken from respondents for ethical reasons, but no mentions of COVID-19 related measures. However, the photos provided in the report demonstrated that interviews were conducted outdoors, with the interviewers wearing masks.

The research company Area conducts social, political and market research using qualitative and quantitative techniques. One example of their 2020 studies is their Türkiye Siyasi Gündem Araştırması (Turkey Political Agenda Survey) from November (AREA Araştırma Eğitim Danışmanlık, 2020). The survey was conducted in one province per NUTS-1 region between 23-27 November. CATI was used to reach a total of 2,230 persons age 18 and above. There were quotas imposed on the sample by age and sex, and data was weighted based on figures developed according to the distribution of the population at the 2018 general elections.

İstanbul Economics Research analyzes online and offline data, using qualitative and quantitative methods (*English | İstanbul Ekonomi Araştırma*, n.d.). Among their quantitative modes are CAPI, CATI, web surveys, and digital panels. One of their projects is titled Türkiye Raporu (Report on Turkey), a bimonthly CATI survey that has been going on during the pandemic. This has been the mode of this project prior to the pandemic too, so mode-wise there has been no effect on the survey⁶⁰. The project covers 12 provinces in total, one from every NUTS-1 region. About 1,500 persons are interviewed, but there is no information if these are cross-sections or there are any panel respondents. Furthermore, no details on the sampling frame are provided on the website. COVID-19 related questions have been included in the bi-monthly surveys. For instance, on December 9, they asked people's level of concern about the pandemic, their opinions about vaccine provision to citizens, what they think of the statistics announced, etc.

⁵⁷ "Metropolitan" is an administrative distinction in Turkey, based on Law number 6360 put in force in 2012: <https://www.tbmm.gov.tr/kanunlar/k6360.html>.

⁵⁸ For details see: <https://sonararastirma.com.tr/arastirmalar/>

⁵⁹ For details see: <https://sonararastirma.com.tr/arastirmalar/>

⁶⁰ Personal communication through e-mail.

GENAR, Konsensus, Gezici Araştırma, Artibir Research, Avrasya Araştırma Şirketi, Optimar, Ada Research, and Consultancy are other research companies that carry out electoral and public opinion studies, and they have delivered findings to the media since March 2020⁶¹. However, GENAR's website is currently under construction, Konsensus' and Ada's websites are not up-to-date, Gezici Araştırma's website is down as of December 7, Avrasya's, Artibir's and Optimar's websites do not include survey reports.

Another private sector example is Tandans Data Science and Consulting. The director of the company, Oğuzhan Akyıldırım, told us about how their ongoing projects with national and international organizations were affected by the pandemic in terms of methodology in our interview presented in Section 4.

3.6 Ethical issues

Switching to new modes, changes to questionnaires and the need for collecting biomarkers for COVID-19 prevalence studies all required new ethical issues to consider and rounds of approval from institutional review boards, independent ethics committees, and alike. A thorough discussion of such issues is beyond the scope of this report, so here we briefly mention points made in the references in this chapter.

With the pandemic, any face-to-face contact brought along new safety and consent procedures. Some of these contacts were required for regular survey data collection. For instance, as mentioned above, RTI has been carrying out FTF interviews in areas with low COVID-19 risk (American Association for Public Opinion Research, 2020). As part of their safety measures, they introduced masks, hand sanitizers, plastic gloves, etc., to interviewers, where single-use items were to be replaced for each household with proper protocol. Interviewers were also told to maintain distance and offer to interview respondents outdoors. They had interviewers sign pledges that they or someone from their household has not been diagnosed with COVID-19 or have not contacted someone with COVID-19 in the last two weeks, showed at least two symptoms, and that their temperature was below 100.4 degrees (38 C) before going to interviews. They also changed their consent form for the respondent. They explicitly told respondents that an interviewer coming to their households could increase the risk of getting or spreading COVID-19. They mentioned how the virus is transmitted, what happens someone contracts it, and what RTI was doing as safety precautions.

Prevalence surveys also require face-to-face contact. For instance, in the study in Brazil by Hallal et al. (2020), all interviewers were tested before fieldwork for COVID-19. All interviewers were provided with protective equipment. They obtained ethical approval from Brazilian's National Ethics Committee and obtained written consent from respondents or parents whenever the randomly selected persons were children. Walker et al. (2020) mentioned in their study from UK that respondents would collect the swab samples themselves to reduce the risk of infection to study teams. In the study from Turkey, all contact was done at the doorstep to reduce the risk of burglary by those pretending to come for the survey (Çakır, 2020).

In cases of mode switch, researchers had to go through additional IRB review. This was explicitly mentioned in Sastry et al. (2020) for TAS-19, by Wendy Hicks for Westat surveys (American Association

⁶¹ See <https://www.haberler.com/2020-anketleri-anket-sirketlerinin-tum-secim-13484055-haberi/> for a list of electoral poll results including most of these companies. Genar's website: <https://www.genar.com.tr/>, Konsensus' website: <https://www.konsensus.com.tr/>, Gezici's website: <http://www.geziciarastirma.com/>, Avrasya's website: <http://avrasyaarastirma.org/>, Optimar's website: <http://www.optimar.com.tr/>, Artibir's website: <http://www.artibirarastirma.com/>,

for Public Opinion Research, 2020), and by Burton et al. (2020) for the UK Household Longitudinal Study.

A concern regarding the switch from CAPI to CATI was mentioned by (Scherpenzeel et al., 2020) for the SHARE survey. The target population of the SHARE survey includes nursing homes. Once the survey switched from CAPI to CATI during the pandemic, it was decided that refusal conversions will not be pressed to interviewers, considering the opposite to be ethically undesirable, putting a burden on nursing staff and people living in these institutions at the time of the pandemic.

For some survey topics, face-to-face is deemed the only way to collect information, and data on such issues may not be collected during the pandemic. For instance, UN Women recommended against collection remote data (phone or web) on gender-based violence (UN Women, 2020, p. 6), arguing “ethical and safety principles to ensure that no harm, risk or distress are imposed on women and men who take part in the data-collection process may not be possible during the lockdown phase of the crisis.” While interviewers can make sure that the respondent is isolated in a FTF survey, and that there are no bystanders, this is hard to do in CATI, despite interviewer administration. This is crucial for sensitive topics. This issue was also brought up by Oğuzhan Akyıldırım’s interview, which we present in Section 4. As part of their research, they were contacting migrants, and asking them about working conditions, which became very hard where respondents were reached by phone in workplaces.

Switching to web surveys introduce privacy concerns specific to the online medium. Rincken et al. (2020, p. 166), in their web-based study from Spain underlined that no personal data was collected: “all procedures were designed to guarantee anonymity without requiring authorization under Europe’s demanding data protection norms: no personal data were collected at any stage.” Kreuter et al. (2020), in the Facebook survey they developed, made sure that ineligible third parties could not access individual responses. Users were notified at the stage of recruitment (through an invitation placed separate from the news feed) which data would be shared. Recruitment took interested sampled persons to a separate website, Qualtrics, for the survey. For those in the European Union, further permission was obtained to comply with the regulations of the region.

Another dimension to privacy is to contact tracing apps developed for COVID-19. Azad et al. (2020, p. 1) discussed privacy issues in these apps in their draft paper, arguing that “the use of embedded sensors could disclose private information of the users thus potentially bring threat to the privacy and security of users.”. Some of the dimensions they discussed were the permissions these apps required to access other apps that could invade privacy (media, camera, microphone etc.); outsourcing data to 3rd parties; and lacking security measures for the exchange between users and data centers. Inal et al. (2020), in their brief on the “Life Fits into Home” app from Turkey, argued that no surveys had been done on this app and privacy concerns around it. It is underlined that user data is accessed by the Ministry of Health, and can be shared with the Ministry of Internal Affairs and law enforcement officers if COVID-19 positive users violate the precautions to prevent spread. Although information from this app is not shared with private companies, there is a degree of access, especially for travel purposes.

4. Surveys in the Post-Covid Era: What Do Experts Say?

4.1 Interview with James M. Lepkowski (A General Frame of the Problems)⁶²

25 September 2020



James M. Lepkowski is Research Professor at the Survey Research Center and Professor of Biostatistics at the University of Michigan. He is also Research Professor at the Joint Program in Survey Methodology at the University of Maryland. He received his Ph.D. from the University of Michigan in 1980. His research interests include sampling methods, methods for compensating for missing data, estimation strategies for complex sample survey data, and the effect of interviewer and respondent behavior on the quality of survey data.

Tuğba Adalı:

Before we start, I will give very brief information about Professor Lepkowski. He is an emeritus professor at the Survey Methodology Program, University of Michigan Institute for Social Research. He is among the authors of many scientific articles and books on survey research, in addition to teaching sampling and estimation courses. We will ask him to draw a general frame for us today to understand how we should read the developments in survey research regarding the ongoing pandemic.

To start, we would like to ask what the current developments imply for telephone interviewing. Although some surveys are already established as phone surveys, and some are established as phone interview follow-ups, phone surveys may have to become the new face to face surveys in some cases. What are the implications of this?

James M. Lepkowski:

Let us expand that a little bit because you asked about telephone interviewing. It is more complicated than that because there are things that happened when you interview by telephone and how you reach the telephone, reach subjects, and people by telephone as well as the sampling methods. In the context of what I think you are talking about, many organizations around the world, government statistical agencies, agencies like your institute do things that involved face-to-face interviewing, drawing a sample of households or housing units with people in them, and then interviewing people in those households in person. With COVID-19 that has restricted countries throughout the world, many agencies find this difficult to deal with because they do not want to endanger their subjects by possibly transmitting disease, nor do they want to endanger their employees who might be exposed to the disease. So, they are restricting, if not eliminating face to face interviewing.

The sample that they have selected is coming about methods that are specific to households. They have stopped data collection, which is a huge problem because this eliminates important data that is being used for policy purposes around the world.

⁶² Video record of the interview is available at: <https://youtu.be/v038hmbhQkM>

In individual countries and across countries, many studies are done in the same way, and this is so that studies can be compared. The data collection stoppage disrupts the understanding of what is happening within and across countries at a particularly critical time. For example, if you measure things having to do with the economy, unemployment, or social wellbeing, how people feel in a time when they may be more socially isolated. The data is now absent. The usual flow of data is not there.

Turning to the telephone makes sense. It sets up the social distance. Social distance happens automatically. The sampling method changes too, and so does the interviewing method. But you cannot use the same methods for sampling households over the phone that you do when doing them in person. Because in-person sampling involves physically looking at housing units as part of the process and typically having lists of addresses that do not have phone numbers. There is disjunction there not only in terms of the interviewing but also in terms of the sampling methods.

Telephone interviewing has been around for a long time. It is new for many of these agencies, these institutes, these organizations; they are not familiar with doing this. Moreover, it raises all sorts of issues and questions about the researcher's apparent ability to use that method for data collection. Will people answer the questions we have been asking in the past or even new kinds of questions when you call them on the phone? Will they answer less seriously or, in some cases, more seriously these questions?

Consider, for example, if you have someone calling you cannot see but can only hear. Are you likely to respond, let's say less rigorously, less formally, even casually, to questions about things like the economy, unemployment, or income and expenditures? Maybe you do not take it too seriously. You think they do not need an as accurate piece of information if they call by the phone. The phone is a personal communication device for many people.

But then there is also the issue of maybe the phone gives us better data in some cases. If we ask about sensitive behavior that people do, this sensitivity will vary from culture to culture. But things that people are uncomfortable answering questions directly, even if they had someone in front of them, make them uncomfortable: a sensitive question and a sensitive answer, both of those matter in this particular case. If you are looking directly at a person, you only hear the audio. Does that give you more social distance? Does that provide a personal distance between you and the person answering the question? Does that give them less discomfort and lead them to answer more truthfully or accurately? There may be some questions about socially desirable behaviors where maybe the telephone is better.

That is a kind of a problem having to do with what we technically call measurement error. The measurement process concerns asking the question and getting the answer. Are they answering the questions as well on the telephone as we can get them to answer face-to-face? That is important now because, for many of these surveys, many questions they want to ask are the same questions that were asked before. If they have been asking by face-to-face for mode for a long time and now all of a sudden switch from a method where they have got one level of accuracy, one level of rigor, one level of quality response to another method that has a different level, it could affect measuring change over time. If they see change over time, is the change real, or is it due to having switched methods?

Measurement error is a very complex subject. I would not characterize it as one that we understand thoroughly, but we understand it somewhat and have many theories that have been proven to describe what people do when they answer questions, whether in a face-to-face setting or by telephone.

There have been comparisons over time of face-to-face interviewing and telephone interviewing. The same questions were to equivalent groups of people by telephone and face-to-face. Do you get the

same answers on average between two modes? When we report things like unemployment, the fraction of people is employed or unemployed, do those fractions come out to be the same when data are collected by phone or face to face? Or if you ask attitudinal questions, opinions about things, do you get the same opinion level, the same opinion structure by face-to-face and telephone modes?

These comparisons have shown largely that there are not many differences between the answers you get by telephone and face-to-face. There is a little bit of indication that the telephone could be a little bit better with socially desirable questions. But it is not dramatic. There is a little bit of indication that you might get a little better factual information reporting by face to face. But it is not substantial, and certainly not at the level where it makes a difference in conclusions drawn from the data. For example, we might see unemployment reports at a rate of 8% by telephone and by face to face, it is 8.4%. That may not be systematically large enough for us to worry about because of the nature of the information we are collecting.

There is another problem here. Not everybody has a phone now. Of course, not everybody lives in a household. In populations, some people do not live in houses, but in institutions. It depends on the country as to the nature of these institutions. Most of these surveys do not attempt to interview people and institutions and mix those results with household data. Institutional surveys are more expensive. For example, for people living in a long-term health care facility, a nursing home, or a hospital for care of people with certain kinds of communicable diseases, the institutions are difficult to get into. It is difficult to get in and becomes expensive, more expensive than knocking on someone's door and finding them at home. One needs permission from gatekeepers, the people who maintain the institution, to do interviews. Consider as well people who are in prison or jail. It is more difficult to get in trying to get permission to go in and interview people whom you do not know, and they do not know who you are. The gatekeepers make it challenging to do. Jail and prison populations are usually excluded in the face-to-face household survey environment. There is a coverage problem in which we are not covering all the population.

There are also some households that we cannot locate readily in remote areas, or we cannot afford to get to them in person. For example, as in a country in the Pacific region, there may be islands with a population just too expensive to get to. The remote islands may have a very small share of the population, too.

What about on the telephone? The telephone is interesting because it reduces some of these coverage errors. If you had a population spread across islands, you might be able to reach people on islands by telephone whom you ordinarily would not visit face to face.

At the same time, people who do not have telephones used to be a larger problem when telephones were provided over a landline, telephone service hardwired into the house. There were far fewer households that had that kind of telephone coverage. But with mobile phones in the North American context, coverage occurs that is in some ways rivaling that in household face-to-face surveys. Telephone coverage has increased dramatically as cell phones, prepaid, and all sorts of other plans have come into line. You can access a very high proportion of the population by telephone in many countries. In some countries today, you can access a larger share of the population, the household population, by telephone, than you could ten years ago because cellphones or mobile phones are more readily available and accessible than a landline phone that requires wiring to install. That kept landline phones away from people who did not have the income to afford them. Today, mobile phones are much more affordable.

Measurement and coverage are two important issues in thinking about using the telephone for surveys. There is another issue: nonresponse. Once you have someone you sampled and knock on the

door, that does not mean that you will get an interview. We are talking about these survey organizations who work very hard to get interviews with every one of those sample households. Because households are typically sampled very carefully, you cannot afford to throw them away or arbitrarily substitute for them. You are looking at something that has been selected carefully as a part of a larger scheme for representing the whole population. Throwing away bits and pieces of it because you cannot get interviews leads to problems and errors when we interview only the people who are most convenient and are willing to talk to us. These readily interviewed people tend to be different from the people whom we are not getting.

We know in the sample a household has been selected. We know where it is located. We go there, but we cannot get an interview. On the telephone, this nonresponse is more challenging. The reason why the face-to-face interviews have been so successful and maintained for so long has been that the response rates in many countries in these face to face surveys are fairly high, 90% or higher. They may get 19 out of every 20 when you ask people who have got no prior contact with you. It is quite remarkable.

However, when they go to the telephone, the response rates fall. When calling on the phone, people are more reluctant to participate. This is not the same thing as measurement error. They are more reluctant to participate because they are unsure if the call is legitimate, hearing only a voice. They have no visual authentication, a name badge. Maybe they get a letter in advance from some organizations where they have got addresses attached to telephone numbers being called. A household getting a letter may realize this is an authentic, legitimate enterprise of a government or a non-government organization.

But the nonresponse rates are higher by telephone. You may see a survey done face-to-face with 95% response rates, and response rates might fall to 80, 70, 60% by telephone. That causes an organization to hesitate to use the telephone for data collection. They begin to worry if there are differences between those who respond and do not respond and that those differences could lead to an error.

But is the data worse? Remarkably, in studies that have compared respondents and non-respondents in different ways, what has been found is that for certain kinds of surveys, the difference between respondents and non-respondents is virtually zero. There can be no differences in attitudinal items. Opinions about things that the people who will answer by telephone are almost the same answers on average. No one's quite sure how this works. There are theories out there. Some very smart people have very good theories about it, and these things are being tested.

But still, organizations are being cautious when they have to make a change to telephone data collection. They are worried about coverage. They are worried about nonresponse. They are worried about measurement. Organizations that have done this have accepted the differences, concluding that their data collection is adequate for their purposes. There is a lot of telephone interviewing done with telephone sampling of households.

Are there differences between face-to-face and by telephone? You have been making a sample selection that gets you households from all the households in the country instead of a telephone sampling that may not cover some households that do not have telephones. There may be differences in measurement error. There may be differences in nonresponse. After you take all that into account, how bad is it? It depends on the circumstance. Many organizations have decided that the telephone is a suitable data collection method and gives them adequate data.

But in an emergency like today, organizations are not going to have much choice. The most important thing they could do is look at three issues: coverage, nonresponse, and measurement, and try to

understand them in the context of their particular place and survey topic. They most certainly should compare what they will get.

Tuğba Adalı:

Speaking of sampling, we have a question about our RDD's. Can you elaborate on the use of random digit dialing for telephone surveys?

James M. Lepkowski:

Random digit dialing refers to an era we do not live in anymore. No one dials a telephone anymore. It is digital now. The basic idea is that we can generate numbers at random and then call those. We imagine what the universe of all the telephone numbers is in a country. We know certain portions of telephone numbers that are part of the system.

For example, a telephone numbering system has ten digits, as in North America since 1947. Mexico, Canada, and the United States all have the same dialing system as does Caribbean parts and even parts of the Pacific. In this system, one actually knows the first six digits that are assigned by the telephone system. They are not randomly determined. I live in an area where the first three digits are called an area code, say, 734. The first three digits are also fixed, assigned to the company that provides service to my area. That is, the first six of the ten-digit number are not determined randomly but are actually known for all telephone numbers in North America.

We can buy that information very cheaply from a government agency that assigns these numbers to telephone companies. Then the random part comes in. We randomly generate the next four digits and start calling a 10-digit number. That is essentially what RDD means. The idea has been extended to other countries that have different number systems.

The random digit dialing schemes used 40-50 years ago have changed. In those earlier times, the four-digit suffix was assigned to telephones in ways that depended on electromechanical switching systems. That limited the sets numbers were actually being used. There were of the four-digit suffixes that were more likely to be used. But today, that is less true, especially with things like mobile phones that never depended on electromechanical systems. The numbers that in mobile number four digits suffixes are randomly scattered. Today, organizations buy the first six digits and generate four-digit numbers at random to append to a sample of six-digit numbers. There could be anywhere from 40 to 60% of such numbers that are assigned to telephones.

Now they may not all be households. Some could be businesses. Some could be used for security purposes, government agencies, all sorts of things. Still, the dialing system, then, is that random digit dialing. There are a lot of details that we do not have time to go into. When you get that call, there is some element in which somebody has worked with the system enough to know what the first digits of the phone number will be, and then they are randomly generating the remaining numbers. It is like we have a huge virtual array of all the possible telephone numbers that could be there, and that has got certain fixed information, the beginning of the number, but the remainder, we do not know what those numbers are. Some countries do get access to full telephone numbers, but that is unusual. We randomly generate them in that case, and that leads to some inefficiencies. You know, if half the numbers are connected to subscribers, that sounds pretty good, except it means half are not, and you got to call them. You do not know that they are connected or not, and when you call them, most of those numbers are not connected to any kind of machine that tells you this number is not in service. Most of them are not connected to anything except a ringing machine, a ringing tone. Thus, it drives up the cost of working with them.

But for RDD sampling, that kind of thing for telephone surveys is feasible. Furthermore, there are commercial companies that help in this process. Now, a government agency may find that uncomfortable because of confidentiality concerns. They want to be buying phone numbers from a company that then knows whom they are calling or what numbers they call. There are commercial companies that collect this information for 150 countries around the world. And then can generate these numbers using these kinds of systems. So, it has become commercially viable and more efficient. This time has gone on.

Sinan Türkyılmaz:

Can I add a question, a jargon or terminology used for telephone surveys? The telephone sampling, telephone surveys, and telephone interviews do not necessarily mean the same thing. They are used instead of each other in the last days. It confuses people. So, Jim, could you be more specific to determine what telephone surveys are? What is telephone sampling? And what is telephone interviews? I believe that this will give a clearer picture of the terminology and these survey types.

James M. Lepkowski:

You have identified the three major terms: telephone surveys, telephone sampling, telephone interviewing. Telephone sampling, and telephone interviewing are processes. The first is how you get to the households; how do you select the households identified through a phone number or the people identified through a phone number. How do you do that? That is telephone sampling. The part that somebody answers the phone where you have to talk to those individuals and convince them to participate in the survey: begin asking them questions, listening to their answers, recording the answers is using the telephone.

The sampling does not necessarily use the telephone. It uses information about the telephone system to generate the locations, the phone numbers. Overarching of those two terms are telephone surveys. Telephone surveys would involve operations that have both telephone sampling and telephone interviewing. However, some surveys only use telephone interviewing. They have already got addresses, or they have got telephone numbers from prior contact. This is something that many countries have been asking about. They have had prior contact with households.

They often do surveys where they go back to a household sample, what we would call a panel. So, they have interviewed them, and before they leave, they say, "We will be coming back next month to do data collection." They go back to do data collection again because there is an efficiency in such follow-up of already contacted people. But there is also some things happen statistically that get better quality data on measuring the change from the time one to time two, from 2019 to 2020, for example, and so on

Many times, they have collected the phone number in person. They want to call the phone number back. They do not need a telephone sample. They are basically doing telephone interviewing. It is kind of a face-to-face telephone survey, but some people will say it is a telephone survey since you are using the telephone for data collection. They mix the terms survey and interviewing together, not making a clear distinction.

There is a related term, questionnaire. I hear many people say, telephone survey, and what they mean is that set of questions you ask. Other people say, "No, we will call that a questionnaire." Because the word survey could be used in several different ways, it gets confusing.

Again, though, the broader term, survey, usually involves a sampling operation and a data collection operation.

Tuğba Adalı:

Our next question is about a survey from the US. We were wondering how the pandemic affected the Survey of Consumers Attitudes when the pandemic started. What about the response rates or any other effects?

James M. Lepkowski:

The survey of consumer attitudes is a good example of a telephone survey. It is one that involves selecting households across the country by telephone numbers, random digit dialing. You could use that general term "RDD sampling method" and then telephone interviewing. And that has been going on that way for a long time.

That is a survey that starts 70 years ago, face to face interview. And then they moved to telephone in the 1970s. So, they went through this transition that we are talking about now in the 1970s, and they continue interviewing by phone. And today, they are mainly interviewing people, by mobile phones what we call cell phones. It is called a cell phone survey. We only sample from telephone numbers that we know to be cell-telephone or mobile numbers. The interviewing is conducted from a centralized computer-assisted telephone interviewing facility.

It is a survey done every month, and it is used to monitor a part of the US. For the economy that we usually do not think about consumers. A lot of the U. S. economy is driven by people buying things, not companies, but individual buying. Why not understand their views about the economy and how they think it is going because they make purchase decisions? Consumers may hold back if they think the economy is not doing well, or they may be buying something they have been thinking about buying for a long time if the economy is going well.

The survey asks respondents attitudinal questions about the economy. How is the economy going? You could look this one up and see the questions; they are out there for people to see. They take these questions that they are asking and develop a numeric score, a score for the U. S. Population summarizing attitudes about the economy.

Now the score level does not matter. It was originally scaled with a score of 100 as the starting point. What matters is not the level but the change. They measure the change. Do persons' attitudes change about the economy? It turns out that as attitudes change, the economy follows. People's opinion about the economy is driving the economy. Attitudes are an indicator, a leading indicator. If you find that people's attitudes are becoming pessimistic and scores are going down, economic activity measures begin to decline within months. Companies pay for that kind of prediction. It is built into economic models.

This is what we refer to, a particular survey being done monthly and being done with cell phones. They are making phone calls at this very moment, trying to reach households. They do this every month, day by day.

Now, the response rates are low, very low, because they have had restricted how much work that can put into calling numbers. But that has not changed dramatically. There is no fall off in response rates though, related to the coronavirus pandemic.

There is a concern though: whether people are answering questions differently in the pandemic. They are monitoring this very carefully. But it is very difficult to tell because the economy is very volatile. As a retired person in the US, I have investments in the stock market or in bond markets, or other things.

Those things will go up or down, depending on the economy, and there is a lot of volatility in it right now. One day the stock index is up. Next day it takes a plunge. Next day it is back up.

What they see in the attitudes is many variabilities too. There's a lot of volatility because people are uncertain about the economy. They do not know if volatility is a measurement issue or just a real attitude about the economy.

Then add to the economic measures an election that is itself very volatile, issues with respect to race and policing in the US, and a pandemic. They are watching the data very carefully because they are trying to measure something that is itself volatile. COVID-19 probably has not changed the quality of the responses as much as it changes the responses' nature. What they are worried about is that this variability might hide economic changes, making predictions less reliable. They do not have definitive data on what is going on, because while they have been collecting for 70 years, they have only been doing it a couple of months during a pandemic and an election.

Tuğba Adalı:

Has the Survey of Consumers Attitudes switched to decentralized CATI with the pandemic?

James M. Lepkowski:

No, they have not. They have discussed the issue because the organization that runs it has a survey center at a university, and that survey center has a centralized telephone interviewing facility. They use that for the calling nationally, across the country. But they do not distribute that to another set of interviewers we have interviewing for other studies from their homes. The attitudes survey is only one of maybe 100-125 studies we do a year. The interviewers involved in some of those other studies are around the country, not in a centralized facility. We could ask them to make these calls, but the study staff has not done it.

Tuğba Adalı:

We were wondering if there were any major issues regarding decentralized CATI because we have seen some webinars where people had to switch to this because they did not want to risk their interviewers working in the same environment. So that is what they did. We were wondering if there were any drawbacks to this.

James M. Lepkowski:

Well, with respect to the survey of consumer attitudes just to continue on with that one. They made the switch to centralize telephone interviewing decades ago in the 1970s. When they did that, they had interviewing staff in the field. Centralizing data collection took away part of the work from field interviewers. Those interviewers had other work for other studies. It turned out not to be a major problem at the time. But when you have an organization that is largely doing face-to-face interviewing and stops using the interviewers, they will find other work.

Maybe the stopping is because of restrictions when you cannot go out and do interviewing due to government policy, national or regional or state. But if you also make, make it so that they stop because you moved their interviewing to the capital city, in a room with lots of telephones in it, you lose them. They may be part of the infrastructure for other studies, but if not, and you go back later to restart field interviewing, you have to start all over again because you have lost your staff.

Part of the issue for centralized CATI systems is a consideration of interviewers, field representatives, and supervisors. All of that kind of thing is in infrastructure, and so decentralized CATI in a transition

from centralized, decentralized to centralized or staying decentralized for keeping the phone of the interviewers busy by having them call by telephone from their own homes is the issue.

A centralized system has a few potential advantages because you can be doing an interview, and the supervisor can listen to what you are doing. Watch the work and make sure that you are using the right kinds of techniques. Staff can advise interviewers on dealing with particular problems in real time. Whereas in decentralized systems, that is very difficult to do. It can be done. But it is very difficult to do.

There is some indication that the centralized quality control makes the data quality a little bit better; it depends on many circumstances. There are potentially some cost savings. You would think one could just hire interviewers locally and pay them to interview only, and not pay them to travel. There are phone charges, but those are fairly nominal kinds in the phone systems we have today. There could be some cost savings in a centralized setting compared to a decentralized one.

There is also a benefit in terms of the quality of the transmission of data. In a centralized facility, the data are collected on a desktop as we are talking about right now. And when it is in real-time, that information is being uploaded to servers and backup servers and kept secure. In a decentralized system, the data is being kept on sometimes paper, sometimes laptops. And then it has to be transmitted. There is an added cost when it is on paper. Someone has to convert it to an electronic form for processing and analysis. But there is also the danger that information on the laptop could be lost because of the transmission error or failure to transmit or failure of the device. Generally, though, those kinds of problems are very rare.

Decentralized CATI has a real advantage in an era like this when you have an existing field force—moving to the telephone to give trained, knowledgeable, experienced staff the telephone assignments to continue to work. They just have to change the mode, putting on a headset.

Tuğba Adalı:

We also want to ask you about online surveys, which, as you indicated before, is a very wide subject. What we want to ask you is coverage issues regarding online surveys. We know there are different types of online surveys out there, some that recruit people through postal services or emails. There are also ways of recruiting people through social media advertising, et cetera. There are many options. What should researchers mostly consider in terms of coverage issues when doing online surveys?

James M. Lepkowski:

There is a whole lot of stuff out there in terms of what online means. Typically though, it is self-administered on a computer screen so that when you enter the data, it is right there in the system; it does not have to be converted from paper. It is transmitted right away to the group collecting the data. They do not typically do this in some kind of anonymous way.

Online data collection has a lot of attractions. But let me go back to coverage, measurement, and nonresponse because they all come into play for online surveys.

For coverage, how you draw the sample to get to people to contact them and have them complete responses often leads to coverage error. Can't we draw a sample of email addresses, just get their email addresses, and draw a sample? Unfortunately, we do not have good sources for doing that. We can figure out what numbers are being used and pretty successfully screen from a set of randomly generated numbers for telephone numbers.

But we cannot randomly generate email addresses. You can buy email addresses from commercial companies that collect them off the internet. But then you do not know whom they are connected to. You do not know where those people are. You do not know if they are even in the country that you are doing data collection for. People whose home is in one country, but they are living and working in another. You cannot really use those email addresses very well because they have no specific information about things important to the survey.

Plus, we have no idea about the coverage of email addresses available from these commercial firms. In situations like this, researchers shrink back from it, not wanting completely unknown coverage problems.

And there are other issues too. I have a friend who has six email addresses. He uses one of them for his personal things and only gives it to personal friends. Another one he gives to companies that are contacting him. Another one is his business email address. And so on. Some other people have only one, and other people have none. It poses a real difficult problem for sampling. So, what are people doing so that they can do online data collections?

They use mail systems, take addresses to buy for households, send a letter by mail, and ask households to participate in a survey. They ask if they would like to do this online. If willing, they provide a link.

The response to this process is not always as successful as even the telephone. There are again legitimacy issues around something coming by mail. It is much easier to throw away or recycle pieces of paper. Some organizations include an incentive, cash inside the envelope, telling the household that this survey is really important to us, and we are willing to send you money, even though you have not answered the questions, to get you to participate.

Because getting people to go from a mail contact online is difficult, they actually give people multiple options. Fairly typically, one would mail a questionnaire and ask them to fill it in or link to an online questionnaire in the same mailing. They can also do the online survey by smartphone.

There is another approach that commercial firms use. They contact people in advance, often by mail, sometimes by telephone, and asking them if they would be willing to participate in future online surveys. They build a panel of people willing to do this. They then collect background information about people willing to join the panel. When that commercial firm gets a contract from a government agency, from a private company, from a University, they sample from the panel the people they have already pre-screened. They can send them the questionnaire online.

The problem with panels is that there is a nonresponse issue. They have millions of people on these panels, but they are not necessarily a carefully selected sample. It is a sample built on the basis of people deciding whether they want to participate. And there you have that nonresponse problem.

These online surveys are appealing though, because they are faster and cheaper than by telephone or face-to-face. More timely information is very attractive. But the quality of the data still matters, and they are serious questions about the quality because of the coverage and the nonresponse issues and measurement. There are things that we ask in person and by telephone that are very difficult to ask online. For example, that survey of consumer attitudes has some questions in which they asked people whether they think the economy will get better or get worse in the next year. A respondent may say it is going to get better, or it is going to get worse. Or they may say, it might get better, and it might worsen, a combination of both. That was not an option in the original question. Interviewers are instructed to probe the answer. And the outcome of the probing might be that the respondent says the economy will be about the same in a year as it is now.

If you ask that question online, where they see the question and responses, and you give them “about the same” as an answer, more people choose it than they would on the telephone. That is a measurement problem. The online version is not comparable to the telephone version.

Tuğba Adalı:

Speaking of the Internet, we also want to ask you about other sources of data available through the internet. We have seen the number of studies increasing use that data, for instance, from Google search data or Twitter hashtags. Especially with the pandemic, we have seen a lot of studies with Google Mobility data.

We are seeing different ways this type of information is used to shed light on human behavior. How do you think studies as such will affect data collection in the future? Will they start serving as a replacement to surveys in some cases, or will they complement one another? Are there any pitfalls we should be aware of with the use of such data?

James M. Lepkowski:

In some ways, I think your question has the answers in it, given what we know right now. Using these alternative data sources, or "Big Data." Data is now available with many more cases and much larger files than generated from survey data collection. There is a focus on the computer aspects of managing and manipulating that data, which is called data science.

But will it substitute for what we have been doing by surveys? In some cases, yes. It is going to happen that we will find some uses of that data that will be an adequate substitute. But it will be an adequate substitute because we have been able to compare it to the survey, to the survey results as the standard.

The survey is not necessarily perfect quality, but we understand a lot about its quality. And when we move to another data source, especially these data sources, they come from commercial sources with properties that are not well understood. We have to worry about issues like coverage, nonresponse, and measurement, translated in different ways. But they are still there.

How do we understand the quality of data from such sources? We compare. We learned about how well telephone interviewing compares to face to face interviewing, comparing the two. As we move ahead, we are going to find ourselves in a situation where sometimes that Big Data, because it is fast, it is cheap, and we move in that direction.

When you look at this so far of this kind of thing has not been particularly successful. Some examples suggested that indicated that Big Data could replace survey data. For years, many countries have had survey systems in place to monitor the level of influenza infection. It goes way up in the winter months in most countries. Changes in infection levels have implications for hospitalizations, hospital care, the quality of care, illness, and death; the same issues we are dealing with COVID-19.

Infection rates go up pretty dramatically every time we go through a seasonal outbreak of influenza, for which infection rates were monitored by surveys where you would ask people about symptoms. The survey results proved to be a fairly reliable measurement of influenza infection level in a population. Surveys might ask about symptoms like fever, cough, and body aches.

Unfortunately, the survey data collections were tied into face-to-face systems, and they were not particularly timely. It took time to process the data. The data would follow weeks, maybe a month later. It was suggested to try something like reviewing online searches to monitor the level of search

topics containing the term flu "flu" or influenza. The assumption was that if people start doing more searches about these topics, it is an indication that there were more infections.

The weekly survey reports lagged by a month were compared to trends in searches mentioning "flu" or influenza. For the first year, and almost two years, the two tracked together. There was hope that the "Google Flu Trends" could replace the survey reports.

But it turned out that people were using the searches in different ways than anticipated. And the online measurement from the searches after two years did not track the same way as the survey measure anymore. There is interest in continuing to use the search mention monitoring system, but it has not been as successful an application of Big Data as anticipated.

There are lots of other examples where Big Data has similar potential and may very well work well. But we have to be careful because you do not want to go to a system that will provide you with misleading information.

So, the second part of the question: Will Big Data replace survey data? In some cases, Yes. Will everything be replaced? No. It is more likely we are going to have complements. There are things you can do in surveys that you cannot do with these Big Data systems. And the Big Data systems need to be calibrated to be measured against a standard that we know works. We are going to need to constantly measure some Big Data against survey data to see that it is working.

I think you are going to see complementarity among these systems. Surveys may decrease in importance in some cases, but they are going to be there used to evaluate the bias and reliability of the Big Data systems.

Now, I gave you a general answer. It is by its very nature wrong. There are going to be exceptions that disprove the rule. I wish that I were 40 years younger and could be active in the field for the next 40 years. For me, it is not just an intellectual thing. I care about this stuff, so I would love to see where it will go. I would love to be part of that.

Tuğba Adalı:

Let us assume the vaccine against COVID-19 is discovered and is already available to everyone. Do you think surveys can go back to the pre-COVID era instantly? What challenges do you think we should be expecting after the pandemic is eradicated?

James M. Lepkowski:

It is unlikely they would go back to the way things were because we may be anticipating the next pandemic. Who knows when the next animal transmission to human transmission might occur? Nevertheless, we will keep this information in our minds in our systems so that if something happens, we have some background, some way of thinking about what we might want to do. Would we change? Would we want to move entirely back to face to face? Is it possible that we would, depending on the country's circumstance, in terms of the data's cost and timeliness, people may find the telephone a little faster, more timely, and attractive?

Maybe we will want to blend these methods. This was something that happened when telephone surveys first became more common in Europe and the United States. You saw a blending because the telephone had attractive features, as did the face-to-face approach. Many organizations that struggled to get the telephone survey system working may welcome the opportunity to go back to what they were familiar with doing.

But I think the more important issue is that this pandemic problem is not going away. A vaccine is not going to solve all our problems.

I worked in public health for a long time, starting as a field epidemiologist with a very infectious disease decades ago. We managed to eradicate that disease, but it's the only one that's ever been eradicated. And we have lots of diseases with vaccines that have not been eradicated. The one that was eradicated had a much more effective vaccine than almost all other diseases.

It will be a good thing for us to have vaccines for COVID-19. They will reduce transmission and make this a much more manageable situation. But we are still going to need to take a layered approach. There are just things that we will need to do to protect ourselves from infection, especially older people, because we know that it strikes much more dramatically there. The death rates are higher. The infection rates are higher.

That may mean a world in which we will have restrictions or things that we do differently because of these kinds of infections. We found ways to live with it through social distance and masks. We should anticipate a future not too different from what we have now and begin thinking about a future where the restrictions remain but not as severe as now. We are going to need to manage this and protect our respondents and our staff. We are going to take advantage of things that are more timely, faster, and cheaper. We also may not necessarily have to rely on them completely because they may be compromised in quality.

We may be thinking about running this survey and have a substantial portion of it done by telephone, but keep doing face to face interviewing because that is our benchmark. I can learn on the phone and calibrate to face-to-face. I could collect more interviews for the same amount of money because the phone is less expensive. But I need to make sure that if it has a bias in it to nonresponse or coverage or measurement error problems, I can adjust from what I know, face-to-face.

I think we need to think about those things, how that works. We have had experience with this. We have what is called mixed-mode surveys. This is related to an earlier question you can do telephone interviewing and face to face interviewing, two different modes of data collection and blend them. We have also had dual-frame or multiple frame samples where you are sampling households by traditional household methods from address lists, and you're sampling households through telephone lists. You could blend those as well.

I am not pessimistic, but I think too often a lot of things have been done as though this will solve the problem. We need to do the social distance, wear masks, and have a vaccine. And this will go away. That has not been the case. There are things that we have to be smart about, more thoughtful about.

As we do that to reduce transmission rates, we need to do things in the survey realm as well. How do we be smart about this and provide quality data? That is what we continue to need. If the data quality falls, surveys will go away. And we will have to rely on these kinds of online systems, Big Data, search engine, and social media systems. And that could be a dangerous thing to do.

Tuğba Adalı:

As a final question, what kind of research topics or survey topics do you think are most affected by the pandemic? What kind of data are we most struggling to collect, and we will keep struggling for a while?

James M. Lepkowski:

Surveys collect a range of data through interviewing. That is what the questions were. But surveys collect data in many of the other modes: personal examination, physical examination, neurological examinations, tests, and things that cannot be done by telephone or online, or at least done reliably.

We have to think about ways to get these things done and combine them with the other things that we're doing. There are research topics about surveys themselves as we change the methods and understand what little we know. We do not know all the answers, so we will have to continue doing that kind of research in a redirection rather than a reinvention.

4.2 Interview with Raphael Nishimura (Adaptations from Face-to-face to Phone Surveys) ⁶³

28 September 2020



Raphael Nishimura is the Director of Sampling Operations in Survey Research Operations at the University of Michigan. He has been working with sampling and survey statistics for over ten years. He holds a PhD in survey methodology from the University of Michigan and a bachelor's degree in statistics from the University of São Paulo. His main research interest includes sampling methods, survey nonresponse and adaptive/responsive designs. He is also the director of the Sampling Program for Survey Statisticians of the SRC Summer Institute for Survey Research Techniques.

Tuğba Adalı:

We have Dr. Raphael Nishimura with us. He is the Director of Sampling in Survey Research Operations at the University of Michigan. He has a background in statistics, and he holds a PhD in survey methodology from the University of Michigan.

We had a chance to see different parts of the world coping with different challenges in terms of surveys and different types of surveys coping with different challenges. So, Dr. Nishimura, as you mentioned in the webinar by the University of Mannheim in April, surveys relying on face-to-face mode were interrupted due to the pandemic had different implications for panel studies versus cross-sectional studies. Could you elaborate on the pandemic's effect on surveys regarding this distinction of panel versus cross-sectional studies?

Raphael Nishimura:

Here is the point where my experience in Brazil might not be as relevant. I will touch base a little bit with my experience here at this research center at the University of Michigan. That is mostly because I do not have any ongoing panel study happening in Brazil. One of the things that we had in terms of the difference between panel studies and cross-sectional studies affected by the pandemic was how we contacted the respondents. Our panel studies already have an established relationship with the panel respondents; we have several ways to contact them to conduct the surveys. We typically would have face to face interviews. But, because of the pandemic, we had to put on hold those efforts, and we were able to transfer that mode of that data collection to another mode. For two of our big panel studies, we went from face to face to either web component or telephone interview. It was also relatively straightforward for the panels because, with those panel respondents, we have already conducted interviews on those kinds of modes. Those respondents were somehow already used to that sort of approach. Even those we have not done previously, we had other ways to contact them, either by telephone or e-mail, which helped us obtain responses for those cases.

On the other hand, we (usually) do not have other means to contact other than the main data collection mode for cross-sectional studies. Here in the US, I recently experienced a monthly survey that we conduct about household expenditures and consumers' behavior, a cross-sectional study in telephone mode. That was not so much affected by the pandemic, at least in obtaining the sample and

⁶³ Video record of the interview is available at: <https://youtu.be/aeWIAvZGYiU>

contacting the sample. Moreover, we also have a project in the field that started a few months ago, just right amid the pandemic. It was a push-web cross-sectional survey. For that study, we had delays at the start of the survey since the staff could not go to the office to prepare the materials. We had some delays also because of how things are working with mailing these days. However, we were able to successfully launch that survey, but we had been contacting these respondents for the first time, and contact rates are lower than our panel studies. An interesting point is slightly higher contact rates compared to previous years that we have been observing since the beginning of the pandemic, especially for the cross-sectional telephone survey that I was referring to. Other survey research centers have been observing the same pattern here in the US, maybe more broadly in the world. There is a hypothesis for that. People were stuck at home because of quarantines, lockdowns. So, they were more likely to answer the phone even though they did not recognize the number.

Interestingly, we also saw an increase in refusal rates. So, in the end, the overall response rate was pretty much the same and stable compared to the previous one. It tells us that even though people answer their phones more, those who refuse the surveys are still refusing. Those would be people who would not answer the phone anyway, and they are refusing. Nevertheless, now we are getting some idea of the level of eligibility in terms of phone numbers.

On the other hand, we did not see an increasing contact rate with our panel studies. For them, we saw the same level of contact rate that we had on previous waves of the panels after the pandemic or even a slight reduction of contact rate in a few cases. Interestingly, we are still not quite sure why that is the case. But we observed that.

Sinan Türkyılmaz:

Before we start with specific questions regarding Brazil surveys, can you give us the main idea of what the predominant survey modes are in Brazil? Has Brazil established data collection through telephones or online panels? Is face to face still popular as pre-COVID times, or is it limited to specific types of surveys like health surveys or any other surveys?

Raphael Nishimura:

Even though I am from Brazil, I have not worked in Brazil for about ten years, like in a day-to-day job. But I still have the contact having colleagues working there. As far as I know, face-to-face is still the most popular data collection mode. Most public opinion surveys, election polls, and health surveys are conducted face-to-face in Brazil. There has also been a slight increase in telephone surveys over 4-6 years, mostly in political and election polling. Also, there has been a significant exploration in marketing research. That is the case, not only in Brazil but also in many other countries. Marketing researchers have been exploring, such as in Web surveys, non-probability convenience sample types of surveys. However, in public opinion, health surveys, and official statistics, face-to-face is still definitely the primary mode of data collection as it was before the pandemic, of course. As far as I know, there is at least a couple of online panels. One of which is from someone that I know, and he has been trying to replicate an online panel, which has been proven quite successfully, I think, in the US and the UK. If you have heard of the YouGov- web panels alike, I know someone over there replicating something very similar. I have heard they have over 100,000-panel respondents that they have been collecting data over the past five years.

Since the pandemic, there has been a spike in the interest of telephone surveys in Brazil. I have been doing some work on that. Another interesting point is that the statistics agency in Brazil, IBGE, conduct most of their surveys and the census face-to-face. The census in Brazil, just like in the US, happens every year ending in zero. So, it was supposed to happen this year. Whereas here in the US, they are

still collecting and conducting the census via mail and Web. They were planning to do that face-to-face in Brazil this year, which is postponed to 2021 due to the pandemic.

Furthermore, IBGE, the official statistics institute in Brazil, believe it or not, do not have almost any telephone survey experience. When they had to transition some of their important monthly surveys to telephone, they had to get some help because they did not have any in house expertise with telephone sampling and telephone surveys. Nevertheless, they had to make the transition quickly in order to be able to continue their data collection effort. I think that they were quite successful; they have, as far as I know, two or maybe three studies ongoing in the field conducted by phone that were previously face-to-face. At least two of those were face-to-face, and they seem to be having some success in that transition.

Yaser Koyuncu:

I would like to ask a question on a survey you mentioned at the University of Mannheim as well. The Demographic Consequences of the Zika Epidemic, which was initially designed to be a household survey in Brazil. How did the pandemic affect the survey? Can you share the experience concerning the mode switch in the survey? How has the sampling design been affected? Did you miss any data in alternative modes that you would be able to collect face-to-face?

Raphael Nishimura:

That was one of two projects that I was involved in, in Brazil recently. I am going to try to tell a quick background and story of this project. I started working with them last year. Way before, we knew this was going to happen. It is a study that the target population is not the entire country; the target is a particular area in Brazil, the northeast of Brazil, the state of Pernambuco particularly. In fact, it is not even the entire state; the target is one of the largest metropolitan regions of the state and another couple of regions like Mata Grosso. As the name suggests, they were interested in studying the consequences of Zika. Their target population was women between the ages of 18 to 34 years old in that area in Brazil. Since the beginning, the plan was to conduct the study face-to-face since that is a household survey, and face-to-face gives the best coverage in Brazil. The project has the grant to be able to do that work. It was supposed to be area probability sampling.

In the first stage of this sample design, we selected a sample of municipalities in Brazil. Then, within those municipalities, we selected a sample of what is called Brazilian enumeration areas. We call it census sectors. The idea was to once we selected those census tracts, we would send enumerators to the field. They would list all the households in those census tracts. From those lists, we would select a probability sample of households. We selected a sample of municipalities and selected a sample of census tracts. We sent the field enumerators to do the household listing. They did the household listing in all the census tracts. That was back in February when they finished the household enumeration.

Then the pandemic came, while they were about to start. There were one or two weeks to start the main interview data collection - we had not selected the households' sample. However, we had all the household listings done when they had to put a hold on that effort because of the pandemic. They did not know exactly where to go with the project because we did not know how long this pandemic would take. After a couple of weeks, when they realized that they would not be allowed to go back into the field in the foreseeable future, they decided to transition on that survey switching it to a telephone survey after looking at the telephone coverage in Brazil and specifically in that area. I should say, telephone sampling or telephone survey was not that common or popular up until now. Because until around ten years ago, telephone penetration in Brazil was not that high. As we have experienced in pretty much the world, cellphone and telephone penetration have increased over the past ten years.

I think we have 90-95% telephone penetration in Brazil. The vast majority is cell phones. A tiny percentage of people have access to a landline. Most people only have a cell phone; some have both cellphone and a landline.

The project team analyzed to determine telephone penetration in that area in Brazil, especially studying the rate among women between 18 and 34. They found out that 95% of those women had a cell phone number. This means that we still would have under the coverage of about 5% of the population. But it seemed good enough for the principal investigators to transition to telephone mode. So, that is what we did. Then we had to select a sample all over again for that study, this time with Random Digit Dialing methodology, RDD. They had to change a little bit the target population because we have area codes in Brazil, just like many countries. It was possible to select telephone numbers but would not be perfect for the area they were targeting, considering that we would not cover the same municipalities. They have a list of municipalities that they were interested in. So, they slightly changed the target population to cover more than they were initially covering to select telephone numbers based on area code in that area in the country. We also had to make some changes throughout the data collection. In the beginning, they had to reduce the questionnaire drastically. For this project, an interview would take one hour when it was face-to-face, which they reduced to about 20 to 25 minutes with the modified questionnaire. The project team also added questions on the COVID-19 pandemic. It is interesting because Brazil had just been through Zika a couple of years ago, and now we are having another pandemic. They are interested in understanding the relationships of Zika-related behavior, and now the project also investigates behavioral aspects of the COVID-19 pandemic. Even though they have different vulnerable populations, they are more interested in ZIKA's impact on pregnant women. For the COVID-19 pandemic, they are interested in the consequences on the behavior of women, young women, women of reproductive age.

As I mentioned, it was an RDD telephone survey. For which, we selected a sample of telephone numbers. Although there has been some increased interest in telephone surveys, RDD methodology is not as common as it is in the US, which means implementing this methodology working with a local survey company was already another challenge. We cannot just tell the local service providers: "Well, we have a list of telephone numbers." It could be just as some sort of a supplement, given its serious coverage issues. We used a modification of RDD to improve efficiency, which can be referred to as a list-assisted RDD. Even with that, the hit rate of working residential telephone numbers was very low. For each successful contact, we had to call around a hundred numbers. They had an auto dialer, but it had a certain capacity of how many calls they can make. So, we had to also work through different ways to improve the sample's efficiency to reach the sample size they were targeting in the survey period. I should say something that I could not refer, in my presentation in Mannheim. They finished the data collection as of last week with about 4,000 interviews. Now, we are going to get into the weighting of the sample.

They also collected information about the respondent's other contact information, address, and e-mail because they want to follow those respondents for at least a year, making it a somehow longitudinal study. Moreover, they are doing some interesting experiments on how to contact and keep them engaged, not to have a high attrition rate throughout the study period. So, they use things like WhatsApp, text messaging, and e-mail messages. It is still an ongoing project, even though they are done with the baseline survey.

Did they miss any data in this mode? Yes, as I said, they had to reduce the questionnaire quite a lot, meaning that they could not collect as much information as they would through a face-to-face survey. However, as far as I remember, they did not have any biospecimen data collection.

Tuğba Adalı:

I have a minor follow up question. You said that the survey was originally designed for the northeastern part of Brazil. What if you call the mobile phone number and the person you reached now lives in the very south? Was that person excluded because of geographic definition, or how was that handled?

Raphael Nishimura:

That is a good question. Even though we could target the sample for area code, there is mobility. Some sort of screening occurs at the beginning of the survey since the project is only looking for women between 18 to 34 in that area. Geography thus was also a criterion. If they called a woman who is not in that specific area, she would be screened out, even if she is between 18 and 34 years old. Women in the northeast of Brazil but living in another city, not in the target, are eliminated.

I should also mention another point. We tweaked the sample design because it was getting hard for the survey company to get respondents. We are treating cell phone devices as a personal device. So, whoever answered the phone was the person that, we assumed, was the owner of the cell phone and the selected respondent. If we call a male number, but that male was married or lived together, had a stable relationship with a woman from 18 to 34 years old in that area, we would recruit that woman through that male. We also thought about other tweaks based on the question: "What if we call a woman in that area but is outside the age range?" She is the daughter or the mother of another woman in that area within that age range. But for that case, we decided not to pursue just because of the sample weighting concerns. We thought it is easier to do that with married males or males in a stable relationship than with other relationships. From what I have heard, there was a significant factor to push through the second half of the data collection; the project should get as many respondents as they need it.

Yaser Koyuncu:

We are also wondering about another household survey you mentioned, the Continuous National Household Sample Survey in Brazil. With the onset of the pandemic, you were involved in designing a new telephone survey whose frame was based on a former wave of the household survey. So, the questions are: Was there any coverage issue regarding phone ownership? What about the response rates? Were there any modifications to the survey instruments, such as shortening the questionnaire?

Raphael Nishimura:

I think it was back in April or May that I was contacted by some folks at the official statistics office in Brazil, IBGE, saying that they had to put on hold all their face-to-face projects because of the pandemic. They were looking at alternatives for their data collection, especially telephone surveys. But they do not have any in-house experience with telephone sampling and telephone survey methodology. So, they were looking for some people to help them design and think about doing that. They have this continuous-extensive national household survey. They are always in the field, releasing their results every quarter every year. That collects data for important indicators about the labor force, which means that they must continue gathering that information.

Given that the economy was concerned throughout this pandemic, the consequences for the labor force is significant. So, they divided this endeavor into two. On the one hand, they are continuing to collect data through this sample survey transitioning that into telephone mode, in which I was not very involved. They have an address frame working with partners to append telephone numbers linked to

those addresses. Obviously, they are not able to append for all those households. Nevertheless, it seems that they have some success.

On the other front, which I helped them, they wanted to have a similar survey to this one but collecting more information about COVID-19 and a little data about the labor force to compare findings with their main survey results. That is why I was brought on to help them design that survey. Initially, they had two options: Either selecting from scratch, which amounts to a fresh sample using some sort of an RDD sample or selecting a sample from the database of all telephone numbers in Brazil obtained from the telecommunication authority in Brazil, Anatel. Another option was to use the sample frame of the past waves from this household national household sample, which is very high quality since it is a face-to-face survey based on area probability sampling. They would append the telephone number to that sample, similar to what they were doing in the other survey. I contact those households to do some sort of a panel. As far as I understand, this one wave corresponds to a quarter of the year. They upended the telephone numbers and built a panel dividing the sample into four sub-samples to have a quarter of the sample each week in a month, at least during the pandemic. As far as I know, the data collection based on that sample is going on. No doubt that they could not use the same kind of instrument they are using on the other survey. They also wanted to ask for information about COVID-19 and, they did. Besides, they worked on a fairly reduced questionnaire, which takes about 15 to 20 minutes to implement. The questionnaire also has labor force related questions.

I think the surveys' strongest point is that they already have data from the previous wave's baseline sample, which helps them analyze and execute statistical adjustments on the sample they are currently collecting. It enables the analysis of the sample percentage without telephone numbers and the differences between the sample and the cases that they could append a telephone number. In that end, they were quite successful in getting a telephone number match to the sample as it was 93% of the cases, if I remember correctly. The sample size was over 200,000 households. Not every phone number is the actual number, of course. 75% of which were correct numbers. However, the response rate was surprisingly very high compared to the USA, for which I used to the single digits in telephone surveys. The response rate, I guess, was above sixty percent. They had a new entire advertisement campaign on the Internet and TV to let people know that the official statistics institute collects data by phone, and it is essential, which increased the response rates and aimed to ease the collaboration with the interviewers of the survey.

Most of their face-to-face interviewers were switched to telephones. They have a huge staff of interviewers calling telephone numbers every day. It was considered a big success and showcase within the institution. As I said, one of the main advantages is that they have all the sample data for non-response and coverage analysis by employing statistical adjustments such as for weighting if there is any sub-population underrepresented in the telephone sample. They can address that since they have information for the cases that they could not append a telephone number and have incorrect telephone as well as the non-responses. Non-response adjustments are possible with this information.

On top of that, they have a very high-quality sample design with an area probability sample, which enables them to use the stratification, clustering, selection probabilities from that sample design and carry them over to this sample. They are just using the selected sample to continue data collection through this longitudinal or panel study. As I said, they have much more information than we would typically have in a telephone sample survey, or even like a face-to-face survey to do those kinds of bias analysis and statistical adjustments.

Tuğba Adalı:

All those properties of area probability sampling could also apply to classical mail surveys. Are those not common in any way in Brazil? I know they are not common in Turkey, and I wondered if that was the case in Brazil. If it is the case, I was wondering why that is. Additionally, for these two surveys you mentioned, are telephone interviews done through a centralized CATI system or a decentralized CATI? Are interviewers calling from home?

Raphael Nishimura:

Unlike the US, mail surveys are not common in Brazil. There are a couple of reasons for that. Some are still concerned that Brazil's literacy rates are not high enough to justify a mail survey even though they are pretty high. In addition, postal service in Brazil is not that great and varies a lot throughout the country. Whether the mail arrives at the correct place and having the mail sent back to the survey institution is a concern. Historically, people are not used to answering surveys by mail, as they are in the US. Some institutions can also see mail surveys a big waste of time because people are just not into these types of surveys.

You are right that it does not have to be in the mail; the survey can be in the push-Web format, which is very popular these days. I probably would not use the same kind of design for area probability sampling because there is not much concern on travel costs. Besides, I probably would not apply as much clustering. Still, I would employ some stratification using geography in that sense.

The CATI is decentralized, as far as I know, for both projects. I am sure that for the project with the official statistics institute, it is decentralized CATI, the interviewers are working from home given the pandemic. For the Zika project, I believe it applies also. I have never talked to them about it. Because, from what I know, CATI work has been done decentralized. You cannot lock people in the office with the ongoing state of things.

It raises another interesting concern that we have been having here in the survey research center on monitoring data quality. Because in a centralized CATI, you have team leaders, have project managers, right there on the floor, monitoring everything. If interviews have any questions, they can rely on their supervisor, the team leader, to help them. We have been quite successful monitoring data quality for decentralized CATI in various ways, but it is less (effective). There is not a very well-controlled environment, as it would be in a centralized CATI.

Sinan Türkyılmaz:

Are you aware of any studies in Brazil, conducted to estimate the prevalence of infection? If you were to design, what would be your recommendations for reaching respondents and what would be the mode?

Raphael Nishimura:

I am aware of at least one study in Brazil. That was conducted to estimate the prevalence of COVID-19. I was not involved in that study, so I do not know a lot of the details. I know that it was a national sample and a national effort with large sample size, even though I do not remember what the sample size amounts to precisely. But I think it was over 10,000 respondents. As I remember, it is quite an ambitious study because it surprisingly was in face-to-face mode, let alone its large size. So, they sent interviewers to the field to gather data. I do not know how they protect both interviewers and respondents for that. I want to believe that they had masks and so on.

I am not sure how I would conduct this kind of study in Brazil. On the one hand, I do not know how successful they were. Nevertheless, one might expect a lower rate of success regarding response rates

and cooperation in this environment than the pre-COVID era. On the other hand, you need to do a biospecimen collection to gather this kind of data. An alternative I can think of would be sending mail to respondents involving a kit with swabs etc. However, the postal service in Brazil and the population unused to mail surveys are problems for this type of data collection. There also could be some concerns about privacy and confidentiality.

I do not think I have any strong recommendations for this type of data collection. Because I think that there are advantages and disadvantages of either approach. In terms of sample design, I would definitely consider a probability-based design vis-à-vis non-probability sampling given the amount of effort you have to go through for data collection. The design depends on the mode of data collection. If it is face to face, I would probably use area probability-based design with clustering to reduce costs and employ stratification to improve estimates. We would not rely that much on clustering for a mail survey, and we would not have the same cost concerns. Still, we would apply geographical stratification.

As I mentioned, the phone survey that Official Statistics Institute conducts in Brazil has COVID-19 related questions, including COVID-19 symptoms. However, they do not estimate a prevalence based on that. It is to have an overall idea of the symptoms that the population has been experiencing. We know that it is tough to have an estimation of prevalence just based on looking at symptoms. They would be overestimating the prevalence if they were to use that. Thus, I would probably not recommend trying to do this kind of study only based on self-reports. I do not think that relying purely on self-report is reliable at all.

Sinan Türkyılmaz:

I also spent too much time considering what if I designed this kind of study. I believe that the survey's timing is also essential because the very period will affect the prevalence of the infection across the country or sub-regions. Such surveys should be repeated. For instance, Turkey had a survey earlier with a very low prevalence. I am not talking about the other issues with the design of the survey and the implementation. Things will change if we do the survey now or later in the year. For this type of surveys, I believe, repetition and comparison are significant.

Raphael Nishimura:

Yes, that is a great point. Another thing that kind of shocked me when I saw the survey there in Brazil, which they are trying to estimate the prevalence. The intention is very good, but I was having a hard time understanding how they would operationalize that. Remember I said that it was a pretty large sample, a national sample in Brazil, a continental country. They were expecting that to be done in a very short time frame. I guess it was something between one or two weeks of data collection. So, I cannot think of a survey based on a national probability-based sample in Brazil that they could successfully finish the data collection without spending a very large amount of resources and money to gather the data within that time frame.

As you mentioned, the pandemic has its temporal dimension. Depending on when measurement takes place, you will capture different things. We cannot do area probability sampling like in the surveys that we collect the data throughout three or four months because things can change so radically as they did over the past 3 to 4 months. You can end up measuring completely different things. So, I agree with you on taking the time dimension into account.

Sinan Türkyılmaz:

I believe that this also raises the question of who we will be, the interviewers or the data collectors. In the Turkish case, the health personal collected the data, who were also in a filiation and patient tracking process. They already had clothes, the tools for protection, and the experience to conduct the PCR tests. I also believe that not only the sample design but also the implementation process is important. The characteristics, quality, and capacity of the interviewers, data collectors, or sample collectors are also very critical given that we have to do it in a very short time based on a standardized way of data collection.

Raphael Nishimura:

Exactly. Because if you want to collect information in such a short amount of time, you will also need a much larger staff of interviewers than typically needed. There are also concerns on how to train the staff for this type of data collection.

Yaser Koyuncu:

Globally speaking, what type of survey errors are going to be mostly referred to in the survey realm from today on with the abandonment or replacement of the face-to-face surveys?

Raphael Nishimura:

Most error types stemming from mode switches of face-to-face to other modes has already happened over the past twenty-thirty years. Face-to-face to telephone transition happened during the sixties and seventies in the US, and there was a lot of research and discussion about that. More recently, we switched to dual frames incorporating cell phones with the landline. Then we started to talk about cellphone only designs. In the last ten years, we also have been discussing a lot about transitioning to Web and mixed-mode designs where you use the combinations of different survey types. Discussions that are triggered by the current situation can be traced back to the discussions we previously had.

The most apparent discussion on the basis of the switch from face-to-face to telephone mode is the coverage. There is much concern about coverage, which also depends on the country. If the telephone coverage or penetration is not high enough, there will be serious concerns about potential coverage bias. In addition, measurement is another topic that will be triggered a lot since we know that the mode effect varies when the interviewer is right there in front of the respondent or not. Telephone surveys are still interviewer-administered but without visual interaction. We know from previous research that it has different effects, and there might be revisions of these error types. Furthermore, non-response will never go away. That is a problem for most data collection methods, although it is more pronounced for telephone surveys with lower response rates than face-to-face.

I do not think that the discussions will be that much different than what we had before. So, I am not sure if there is anything necessarily and particularly new regarding survey error discussions. We have been having conferences, webinars, talks about the consequences of the pandemic in survey research. However, I am yet to see anything new on any of those discussions other than observing more considerable interest of countries that did not have telephone mode as the predominant survey mode. In a way, the survey research has already been through that. Web and mixed-mode surveys have also seen more interest in developing and other countries that have not used these modes.

But again, those discussions, at some point, has already happened. I think it would help to have these discussions going forward in places that typically rely on face-to-face data collection. Now they have to consider other means of data collection and go forward. I think that is an interesting point. Maybe in Brazil, for example, telephone surveys might be considered more often given the capacity built in the past several months. That might be a part of the change regarding data collection in many countries

that previously may not have been considered. Because face-to-face was always the gold standard and default mode that countries would choose.

Tuğba Adalı:

Our final question to Dr Nishimura is a hypothetical one. We want to ask you about the long-term effects of the pandemic. Let us assume the pandemic is eradicated overnight. Do you think surveys will go back to the pre-COVID era instantly, or what challenges should we expect after the disease is eradicated?

Raphael Nishimura:

In a way, I feel that we are experiencing that. I have seen that since their infection rates are low enough, some countries already went back to face-to-face data collection, or they are starting the process to get back. It seems that people are eager to go back to their bread and butter. I think many institutions probably would go back; face-to-face if it were face-to-face before. But it would not necessarily be a good idea. It is important to take some lessons from this whole thing. One of which is that while face-to-face data collection and area probably sampling is and will probably be the gold standard, it does not mean that other data collection modes are not good enough for most surveys. I think that many places will try to resume their previous data collection mode, face-to-face or not, as soon as possible. Also, there will probably be some other institutions, I hope actually, that will think twice and consider more carefully for other alternative data collection modes. It might be telephone, Web, or even mail. I think it is a good kind of learning because it makes people think about the survey error types.

Why do we use typically face to face or area probability sampling? Well, it is mostly because of its coverage properties. As I said before, it is a gold standard for coverage as long as you have a one-to-one correspondence of geography with your population elements. In theory, you have 100 percent coverage. Does this mean that telephone survey is much worse than face to face and area probability sample? Not necessarily. It will depend on the coverage rate and telephone penetration in your target population as well as the difference between the covered and uncovered population. It might cause some small biases in estimates because of those, but I think people are thinking more about that.

As a survey methodologist, I think of a lot about survey errors, but it is not only about that. It is also about survey costs. Does the amount of bias stem from coverage properties, which sometimes are in negligible numbers, justify the more money and effort you invest for face-to-face data collection, which also can be done through telephone mode? The same goes for the Web, although I think the Web is more pronounced for the internet penetration rates and differences between the covered and uncovered populations. The literacy rate is also an issue there.

Some people are eager to go back to the pre-COVID, as I said. But I hope, and it seems that some people will probably not go instantly. They will probably need to finish their ongoing data collection with phone or Web, and hopefully, they will consider alternatives in the future. Another point to consider is the survey series, which I have been involved in a few. They were in face-to-face mode, and they had a transition to telephone or Web. If they have not been careful enough on that transition, they may see some differences due to mode effects because of measurement or response rates. So, they will also have to decide. Will they ignore this period that they had to switch modes to the previous data collection mode and have a gap in their series? Alternatively, they may start a new series and continue with the new data collection mode after the switch, or they might apply statistical adjustments on the data to make the two pieces of series more comparable. So, that is another thing to think about, especially for the longitudinal type of studies. It is one of the challenges that I would expect when the pandemic is eradicated. Many projects had to change the mode on the fly. They did

not have enough time to prepare the survey for that sort of switch. Therefore, comparability and what to do for that will be of concern thinking about these kinds of studies.

Sinan Türkyılmaz:

COVID-19 will not disappear in one night. So, I believe that we have time for revision and improvement of the surveys. COVID-19 gave us a chance to evaluate what we were doing and what we are doing and what we will do to conduct the household surveys. How about alternatives like big data? They are not the full alternative. However, this will also be a chance to see how we can combine or compare these alternatives.

Raphael Nishimura:

These days, that is a very hot topic in the survey field: how to combine big data with surveys. Interestingly, maybe because it is not part of my bubble, how alternative data sources as big data can be useful is not something that I have heard a lot about during the pandemic. But it is an ongoing talk in the survey field in terms of how to incorporate and use that.

My view on this is that even though big data at least provide an alternative data source, I still think there is a role for surveys. I mean, surveys are designed data, which means that you have more control over the data quality. It is more expensive to collect, but it requires less effort to analyze, whereas using big data may be cheaper, but it might require much more effort to analyze, get a proper meaning, and draw proper inferences. So, it depends on the allocation of the resources. You might allocate your resources in the front end through a designed data and a survey. Then, you do not have to spend as much effort analyzing the data since it is designed accordingly. Or you might spend the resources on the back end in which the data is already available or generated. But then you need to spend a lot of resources and effort to model that data to make adjustments, ultimately making it meaningful and helpful to draw proper inferences for the population you want to analyze.

4.3 Interview with Oğuzhan Akyıldırım (Impact on Research Companies in Turkey) ⁶⁴

22 October 2020



Graduated from Department of Statistics of the Middle East Technical University, Oğuzhan Akyıldırım worked for 14 years in Turkish Statistical Institute. Mr. Akyıldırım also worked as a consultant in many international projects in various fields such as research design, field application, statistics and quality control, and he is currently the CEO of Tandans Data Science Consultancy (Tandans Veri Bilim Danışmanlığı).

Tuğba Adalı:

Today, we will talk with Oğuzhan Akyıldırım from Tandans Data Science Consultancy. Mr. Akyıldırım is a statistics graduate of Middle East Technical University and a researcher who has worked for 14 years in TurkStat. Besides his private sector experience, he has also provided consultancy in various fields such as research design, field application, statistics, and quality control in many international projects.

Today we will talk with him about the pandemic's effects on the research world and private research companies and ask him to deliver his predictions. Mr. Akyıldırım, we would like to first ask about the pre-COVID-19 period. Before the pandemic, what kind of field studies Tandans Data Science Consultancy had? Were there face-to-face studies, telephone studies, online studies, or any qualitative research? What types of activities did you have in terms of research? We would be pleased if you could briefly explain them to us.

Oğuzhan Akyıldırım:

After our company was established in 2015, we employed all the modes you mentioned, and we continue to do. For instance, during the months the company was founded, we conducted research on behalf of the Council of Europe with the Turkish Ministry of Justice by applying Web-mode. In a similar vein, we have been carrying out studies on the internet since then. In addition to quantitative inquiries, we also conduct qualitative research.

Primarily focusing on the socioeconomic field, we also apply the mixed-methods, including quantitative and qualitative dimensions. So, we have a significant experience within this context. Our company generally pursues research on outsourced funds from various organizations, including the United Nations, UNICEF, and UNDP. We work with all of them. Now, we have two ongoing projects with UNICEF. Our company has joint research in mixed-methods with these organizations.

Let me talk a little bit about Tandans Data Science Consultancy. We started with two people, two statisticians. Me and a friend who is a Hacettepe University Department of Statistics graduate. Now, there are almost twenty employees in the company, including Hacettepe University Institute of Population Studies graduates and students. They are all very successful, contributing to us in every

⁶⁴ Video record of the interview is available at: <https://youtu.be/xC2KMIQhSLU>

aspect. Forty percent of our employees are statisticians, while the other forty percent are sociologists besides two psychologists. These are our permanent employees that include three support staff.

When the COVID-19 pandemic started, there were three studies. One of which was ongoing, and two were about to start. Of course, we stopped the ongoing work when the effects of COVID-19 escalated abroad. Realizing that the situation is uncertain, we asked GIZ with whom we were working at the time to call a halt to research; and they agreed. But they also asked us to take action immediately. Accordingly, we switched from face-to-face to the telephone mode for data collection that took about two weeks of time. The study was suitable for telephone data collection mode. In other words, both the questionnaires and our sample fit that. We only edited the questionnaires a little because asking questions and getting approval on the phone and face-to-face are significantly different. The control and understanding of the questions are also different. We switched to a system that would fit this, shortening our questionnaire slightly since the duration of focus in face-to-face and telephone modes varies. So, we made minor modifications without a significant change in the general framework. Besides, there was not a substantial change in the coverage.

Let me also talk about the other two studies. Both studies were qualitative fieldwork of UNICEF studies. Having no quantitative dimension, these had face-to-face interviews or focus group sessions. We also thought about whether we should start the fieldwork or not for this research and, we asked them to review this one more time, given the COVID-19 risks. They, too, were convinced that this study could not be done face-to-face in the field after considering it. We thought that it would be tough to implement focus group meetings over the Web or teleconference method. So they offered the switch to in-depth interviews. Working on it, we discussed the math of doing this switch, also talking with Professor Türkyılmaz. One of the questions was: How many in-depth interviews correspond to a focus group meeting? We considered that. Indeed, there is no such calculation; the nature of the two is different. It is impossible to compare two things that are different in nature, but we can compare the workload, including the interview duration, preparation and finalization of the interviews, and then reviewing the notes. We calculated the workload and related processes, also doing the same for the in-depth interviews. After considering the number of focus groups corresponding to the number of in-depth interviews in terms of the workload, we did the change. However, the approval of the Ethics Committee was also required. Because the previous Ethics Committee approval was obtained for face-to-face interviews. So, the Ethics Committee approval process started again.

We started to investigate the effects of conducting face-to-face interviews over the phone or online. What kinds of ethical problems might arise, what kinds of issues might there be, and how can we take measures against them? We discussed these questions, and it took a long time. Maybe you know the UNICEF Safeguarding Policy. We formed a Safeguarding Policy together. I hope next week we will get Ethics Committee approval and start. We have been working on this since March, and we have spent a long time. These two studies were affected in this way. Of course, they will also have some handicaps. Maybe we can discuss these in the next questions.

Tuğba Adalı:

I want to ask something about the first study. You said that the study allowed the switch from face-to-face to the telephone. Did you already have the phone numbers of the people you were planning to interview?

Oğuzhan Akyıldırım:

Actually, we did not. We used to go to addresses before, and we did not draw the sample based on addresses. It was a different dimension. This study tried to measure the “employability” of both Syrians and Turks. And we applied random sampling. Regions were small, which were in Adana, Gaziantep, and Kilis. When we changed the face-to-face to phone interviews, we asked for their phone directly. We signed a confidentiality agreement with the institutions—both with GIZ and with GIZ’s partners. So, we conducted the interviews based on that confidentiality agreement. Both addresses and phones were accessible. There were phones that we could not access due to some circumstances, for sure.

We saw that there are some difficulties, even if we have phone numbers. There are both advantages and disadvantages. First, let me talk about the advantages. You can catch the respondent anywhere, and at an appropriate time and position, you can conduct the interview. The disadvantages are as follows; the responder can be constantly busy and, therefore, not available. He or she may only be available outside of certain hours. Here, too, an ethical issue arises that we never thought about while doing the research. For instance, we thought these people could answer us alone. But when we called, for example, he/she was at work. We had to talk at work without the boss’s supervision. However, the owner of the business wanted to monitor this. Because questions were about his work, so this raised ethical problems. We talked to GIZ again to solve this problem, and we started to invite these people to environments where no one would be with them. The burdens brought to the respondent started to be in question.

On the other hand, the burdens on us were as follows; our staff started working in the evening rather than during the day, and interviews had to occur within a short time frame. These were the personal times of the interviewees, and we were intervening in those. Also, our team had to conduct interviews in the evening or on the weekend, which was not very convenient.

There were issues to be worked on ethics, and unfortunately, we could not do this at that point. However, later, we worked for a month about that in a follow-up study, uncovering ethical issues in telephone surveys and developed measures for that. We have already presented them to UNICEF. Something good will come out with their contributions.

Sinan Türkyılmaz:

Actually, you have answered some of the questions I was going to ask, but let me try to systematize it. You know, mod switches cause survey errors of various sizes. In this example, you switched from face-to-face mode to classic telephone mode, and you can open up the other two surveys a bit more. You know, we classify total errors in surveys as sampling errors and non-sampling errors. We also categorize non-sampling errors with the main dimensions of coverage error, measurement error, nonresponse error.

You have referred to many of them directly or indirectly, but can you talk about them more systematically? Of course, it is not possible to tell how much error has changed due to this mode switch. But if you would summarize based on your observations and thoughts; How do these forced mode changes affect survey errors in terms of coverage, measurement, and nonresponse?

Oğuzhan Akyıldırım:

I think measurement error comes to the forefront. I can say that especially based on the research conducted by GIZ. As I said, it did not affect the coverage error much. Nonresponse in face-to-face and telephone mode is almost close to each other. In fact, the nonresponse in face-to-face mode is higher than the telephone. There was no problem with that. However, as I mentioned, the problems are the questions that are not understood correctly or the cases that the interviewers cannot ask the questions

properly. When you are on the phone, you do not see the person. When their setting is not suitable for the interview, if the interviewees cannot isolate themselves from the environment, they are open to external influence. For example, the respondents may still contact their family while answering questions in their homes. It is challenging for us to prevent this. We ask them to move to a private place to conduct the interview. Unfortunately, this is not likely to happen, particularly for the low-income groups who live in a single unit or room, making it difficult to pull people out of there. Also, while talking to the interviewers, respondents may be watching the TV, or someone else's phone can ring, for instance. We experience interruptions caused by these factors often. Further, if the questionnaire is lengthy, measurement error becomes an issue, especially for the questions at the end since respondents do not immerse themselves or may not be interested anymore with a desire to end the interview. Therefore, these bring up significant losses in terms of data. Overall, I think measurement error is foremost, considering the survey errors stem from this mode change.

Apart from that, as I mentioned, we did not have a coverage error. Nor nonresponse error was much of a concern for our study. Of course, if we did not have address or telephone information, we would face huge errors on the response side.

Tuğba Adalı:

I also want to ask about the cost side of the issue. How were your research costs affected by this compulsory mode change in the study you refer to? You can also think about the other two qualitative studies. What have you experienced in terms of cost?

Oğuzhan Akyıldırım:

Frankly speaking, we were very happy at first. Our work was not canceled. We switched our modes very easily, and all our customers were convinced, already demanding about the research. We thought that it would cost cheaper. But it turned out to be more expensive. There are two main reasons for this. First, design times are getting too long. Indeed, both the design and analysis phases extended. We employ more competent, talented, and skilled people. Nevertheless, they do not have to be very competent in the data collection stage. Of course, there must be people who can collect data well. We care about training and provide it.

I thought our cost would decrease there, but as I have stated, the cost started to increase by about 50-60 percent. The design is essential. We need to design it well and need to think about it thoroughly. Once the investigation starts, we cannot go back. It is not possible to say, "the questionnaire does not work like this; let's change it." We do not have the luxury to say that.

After the data is collected, the data analysis starts to be very difficult due to the measurement errors I mentioned because the inconsistencies increase. Long questionnaires with questions over a certain number as well as increasing consistencies lay a further burden on the people who process and analyze the data. As you know, the data processors, analyzers, and staff who prepare the reports must be of high quality in terms of skills needed for managing these.

Also, while assuming a reduction in data collection cost, we never thought about the qualifications needed to talk and get answers on the phone. There is a considerable difference between ways to get a quality answer in face-to-face mode and over the phone. Asking questions on the phone requires respondents to visualize what is asked. For instance, the interviewers may ask on a scale, "For worst or lowest being 0, the highest being 10" The interviewers may need to read, say, twenty propositions and expect interviewees to give answers saying numbers between zero and ten. The respondent needs

to focus on each statement, and after the second, third, and fourth proposition, he/she might begin to get tired.

Moreover, trust in face-to-face is much higher than on the phone. Thus, the questions start to be challenging with trust problems between interviewers and respondents on the phone. Even when the participants are convinced about the research aims, for whom the study is done, how results will be used, anonymity and confidentiality; there are trust problems on the phone, making it difficult to manage.

In a nutshell, the cost started to increase, even during the data collection phase. Unforeseen costs began to emerge besides increasing telephone expenses. The interviews are interrupted, and then the staff tries to catch up again, to get an appointment. It is not a cheap mode to work with, unlike what people still think. It is not a cheaper way to work compared to face-to-face.

Tuğba Adalı:

Do you think the cost increase is just related to the mode-change? If you keep conducting telephone surveys, will cost fall below that of the face-to-face surveys in the long run?

Oğuzhan Akyıldırım:

It will not. CATI systems do not work like call centers. It is not easy to persuade researchers to talk on the phone constantly. The data collector is continuously changing, and it is not easy to find a person to collect data on the phone. The people we work with are university graduates, sociologists, psychologists, and statisticians. These people do not want to be data collectors on the telephone. You cannot train ordinary data collectors. In other words, it is tough to hire people working in those call centers as data collectors. You have to train them every time. You will train them on the research topic. It is not easy to train people at the beginning of each research, and they start phone calls. I mean, keeping these people in that system does not seem easy to me, especially in these circumstances.

Yaser Koyuncu:

Considering the pandemic and what you said, can you tell us your predictions regarding the interview environment and techniques for the future?

Oğuzhan Akyıldırım:

Actually, all of these modes can be utilized in the future. One of the studies found that the data collection techniques in the world and Turkey differ entirely. It was like that before the pandemic. For example, ninety-four percent of research in Turkey was done face to face. Only six percent of data was collected by phone or other modes, among which there could be a system called back mail-out. But abroad, this was completely the opposite. Mail-out and telephone modes are used more in other countries, and face-to-face interviews are less common. As far as I know, you can vote even by mail in American presidential elections.

Compared to other countries, trust to interviewers is lower, or respondents' views of the surveys are different. I have often seen the warning that no sellers and interviewers are allowed at the doors in my professional life. In some factories, there were warnings at the first gates that no interviewers are allowed to enter. It also relies on the data collection mentality, so it is difficult for us to collect data in any mode without changing this mentality.

I always claimed that giving data is also a national service like paying taxes. So, we have to provide information, as simple as that. But we must also trust the people we give information to. I do not think

that changing the method and modes currently used or will be used after COVID-19 will have a huge impact without changing this environment.

But let me answer your question. Other modes will now be used with the face-to-face mode. For instance, while some data collection will be done face-to-face, the other parts will be conducted by phone, mail-out, and Web modes. I think mixed-mode research will be increasingly utilized, and we will organize that according to the groups we target. For example, we will go to hard-to-reach groups face to face while surveying easy-to-reach groups via the Web or phone. And, we will employ regular mail mode to the groups we are confident about. Furthermore, we will need to develop sampling methods for mixed-mode.

Sinan Türkyılmaz:

No doubt that methods are significant, but for quantitative research, especially in large household surveys, we were having problems with sample frames in Turkey, and we still do. Almost the only option is the National Address Database connected to the Address-Based Population Registration System. However, with the increasing use of the phone, the need to obtain alternative sampling frames increases. Of course, this again depends on the research, the subject, and the universe.

Do you have any initiatives in this sense? Did you have to do any initiatives? What kind of difficulties do you face in terms of alternative sampling frames, and do you think you will encounter such hardships in the future?

Oğuzhan Akyıldırım:

You know there is a law on the protection of personal data. As long as this law exists, trying to have such databases will put the owner in complicated situations. Therefore, we, as a company, did not embark on that much. However, there are three databases we work with containing data about 1.5-2 million people. We know that it has about sixty to seventy percent of representativeness. We tried and proved it by applying face-to-face and using these databases concurrently, and the results of our research turned out to be consistent with each other. So, we rely on all three databases. However, as I said, this has 60 percent of representation, meaning that it represents sixty percent of Turkey. Our social research focuses on the remaining forty percent, which is almost half of the population. We consider the methods we can use other than face-to-face to reach these people.

In addition to the address database, the only thing that can be done is to create a communication database which can be handled by TurkStat. In fact, I wrote an article six months ago on the lack of such a database. When the COVID-19 pandemic began, it came to the fore. Meanwhile, I saw that these discussions were continuing in Europe, where some people support it and others not, especially for privacy concerns.

On the other hand, we first ask the respondents if they would like to participate in this research after explaining its purpose and who is behind it. Overall, as a researcher, I do not think there is a privacy or ethical issue here. Therefore, I believe such a database should exist. After the database is established, errors will be calculated easily. Is Address-Based Population System free of errors? No, I guess there is around 60-70 percent of error in ABPRS since, for instance, people might move from one place to another without reporting it etc.

Overall, a communication database should be created, and TurkStat, Hacettepe University Institute of Populations Studies, or another prestigious institution should lead that rather than a private institution or ministry. This database that includes data about the households and communication information should only be used for statistical purposes.

Tuğba Adalı:

Considering survey research in different sectors in Turkey, such as market, social or political surveys, which topics have the pandemic made data collection most challenging, and what are the implications of this from now on?

Oğuzhan Akyıldırım:

Due to the reasons I mentioned, the pandemic made it challenging to collect data on every subject, but mostly the data collection from people of low socioeconomic status and disadvantaged groups. It was easier before, and these groups are most prone to giving data because they make their voices heard about their complaints, using the surveys. Nonresponse rates were the lowest amongst these groups, especially if the surveys were on behalf of reputable institutions. It was hard to obtain responses from middle and high-income groups, and nonresponse levels are high for these society segments.

Unfortunately, after the pandemic, it also became difficult to get responses from low-socioeconomic groups and disadvantaged people. Besides, we are now trying a new method with the Turkish Red Crescent. We will hold focus group sessions. I hope the first meeting will be in mid-November. Tablets are distributed to households, and we are planning to conduct focus groups with these tablets. We will continue to conduct in-depth interviews, but we wanted to try that besides. Turkish Red Crescent has promised to support us. They are willing to try and implement it. It will be the first time for us, and we will start collecting data with these methods from people that we started to have low response rates. We will see the quality of data with this method. Typically, there are around 6-8 participants in focus groups, but we will limit it to 5-6 and observe if we can create a discussion environment more easily.

I never thought researchers were vulnerable to such risks before. We had many focus groups but did not think about meeting rules. We always thought that the focus group is a technique that can be done with 6-8 people gathered around a table, with a moderator and a note-taker. But we never thought about the room's size, the distance between the interviewers, and the communication between the moderators, note-takers, and others in these circumstances.

Tuğba Adalı:

You think it does not impact the research topics much. Nevertheless, it has influences on the segments and groups that the data is collected from. Right?

Oğuzhan Akyıldırım:

You got it right. Let me expand on the subject by adding one more thing. I do not know how it is abroad but getting responses in Turkey becomes very difficult now without motivating the respondents. People of higher socioeconomic status especially started to think that they spare their time for you and that they should get something in return for that. Therefore, it is necessary to motivate them in some ways. In market research, companies try to motivate respondents by giving their products or shopping vouchers of some markets. However, this is getting a little more difficult in other studies with voluntary participation. It seems easier in market research since they give people incentives by sending their products. Then they call them, and respondents participate. Also, the databases I mentioned upload credits to the accounts of people. In those cases, if you consider the consistency, you can remove unreliable surveys; hence you disqualify them in the research as well. It is debatable how realistic this is. As I said, it is easier in market research but very difficult, particularly for social and

economic research. For instance, it was already challenging to ask a person about his/her earnings and income face-to-face, but now it has become much tougher by telephone.

Sinan Türkyılmaz:

Before moving on to my question, let me ask a small additional one to see if you know whether the virtual focus group formation you mentioned for the previous question has been tried before. Have you seen any examples of that?

Oğuzhan Akyıldırım:

I have not seen any examples. I heard that the Turkish Red Crescent is conducting one or two pilot studies.

Sinan Türkyılmaz:

It sounds very interesting and creative to me. Are you the first to think about that? Or is it tried especially in other countries? It also can automate the recording. It preserves the social distance. However, how not being able to sense and not feeling each other in real will impact the interaction in the focus group is an issue of concern. But, I think it is very interesting also from a methodological perspective. I am also very curious. I would like to talk and discuss when you conduct it.

As the final question, let me ask a hypothetical one. Actually, this is not something that will happen, but I wish it would. If COVID-19 disappears overnight, how will the research sector be affected? Would you suddenly return to the methods you used before the pandemic, or did the pandemic have permanent effects?

Oğuzhan Akyıldırım:

I can speak for myself, but I cannot say much about this for the sector. This process taught us a lot. It taught us where and how to use the modes, what kind of difficulties they have, how to solve these difficulties, their costs, the type of researcher required for different modes. For example, a face-to-face interviewer may not be suitable for the telephone or online surveys using Web or software like Zoom.

We plan to use all modes that we can use in quantitative research. We are planning to use mixed-mode. We plan to employ different modes that will cover all groups in the same study. We have not done it yet. There will probably be major research and field studies after the pandemic ends. We have been waiting for the results of two studies for around six months, and we probably will obtain that when the pandemic is over. We will suggest them changing the method, that is, to use different modes besides face to face.

There is a very different environment right now. Everyone thinks they can solve all their problems with Web and online surveys. No statistician is needed. For instance, with Survey Monkey, everyone collects data and publishes their results. I tell them to send me their method for every link I receive, and I promise to answer the questionnaire. I even say that I will convince my friends and they will answer, too. But I never get information about their methods.



Enver Taştı, who graduated from the Statistics Department of the Faculty of Science and Letters of 19 Mayıs University, completed his master's degree in the Department of Statistics at Hacettepe University in 1990. Having worked in several positions in TurkStat, Mr. Taştı served as the Vice President of TurkStat between 2012-2018. He is still working as a Statistics Consultant in TurkStat.

Tuğba Adalı:

Today, we will talk with Enver Taştı from the Turkish Statistical Institute. Mr. Taştı is a graduate of Hacettepe University Department of Statistics and has served as the vice president of TurkStat. He is a senior specialist who has been working at this institution for many years. Among his areas of expertise, we can count censuses, demographic statistics, household consumption expenditure statistics, labor force statistics, living conditions statistics, and many more. Today, we will ask him about the pandemic's effect on the design and implementation of TurkStat surveys with his predictions.

Mr. Taştı, first of all, when the pandemic came to the fore around March in our country, was there any TurkStat survey going on? If so, how surveys were affected?

Enver Taştı:

After the Statistics Law of Turkey was introduced in 2005, TurkStat began to prepare annual working programs, including fieldworks. So, the program for 2020 was ready. As the pandemic started in March, monthly surveys were continuing. The fieldwork of the Household Labor Force Survey and trend surveys are among the examples of that. Also, the fieldwork of the Consumer Price Index and similar surveys were ongoing.

There was a switch to CATI in 2020 for the Household Labor Force Survey and some studies. The second, third, and fourth interviews were conducted by telephone, given the survey's design. Most of the trend surveys were already done by phone. People had serious concerns about distance, masks, hygiene, and cleanliness due to the pandemic for the surveys done face-to-face, and our interviewers also had similar worries. Households could refuse to participate despite the measures taken. So, there was a need to take new steps. The pandemic took effect around mid-March, and half of the fieldwork of the surveys were almost over. Thereupon, the switch to CATI was decided after immediately discussing what can be done considering the surveys need to be completed.

Tuğba Adalı:

So, the first measure taken was switching face-to-face to the telephone mode as much as possible.

⁶⁵ Video record of the interview is available at: <https://youtu.be/Bnrpyf1W98>

Enver Taştı:

Yes, and there was a short period of confusion. That happened in other countries as well, and we could not switch immediately. However, by making decisions very quickly, phone lines were provided, if I am not mistaken, for nearly five hundred interviewers so that they can directly conduct the surveys through this system.

Sinan Türkyılmaz:

I also want to ask a related question. We know that TurkStat has the names and addresses of people in Turkey who registered to Address-Based Population Registration System and National Address Database. What phone information is available from official or other sources that can be integrated into these databases? With what types of phone information can you integrate these? Is it possible to access these databases, which we call frame, for other public or private sector surveys? In other words, is it possible to use them as the frame for telephone surveys? How do you use it? And how about other users?

Enver Taştı:

You know, there is demographic and basic personal information in the Address-Based Population Registration System, but there is no information about telephone numbers. This information is available if the citizen has entered or declared his phone information to the General Directorate of Civil Registration and Citizenship Affairs during an address change. Unfortunately, there is no phone number of other citizens registered to ABPRS, which constitutes the majority. Under normal conditions, we draw our samples after receiving the information from the Address-Based Population Registration System. For the samples selected, we obtain phone numbers from the public institutions with phone information, namely the General Directorate of Civil Registration and Citizenship Affairs and the Information and Communication Technologies Authority. We also started to obtain the phone information based on the E-Devlet (Electronic Government) system. We match the phone numbers with the individuals in the sample and other household members to conduct the survey.

In a nutshell, we do not have an address frame that includes telephone numbers of all individuals in one place. We conduct the surveys by combining up-to-date telephone numbers obtained from other administrative records and the samples selected. Of course, there may be households that do not have a telephone. Telephone numbers of all household members are obtained as much as possible, and it might be sufficient to have a person's phone number from the household. For rural areas, if the phone numbers of headmen (mukhtars) are available, they can also be contacted to invite households. Frames are formed this way.

We cannot share personal information with the private sector or any other institution due to personal data confidentiality within the Personal Data Protection Law framework and due to a special article included in TurkStat's law. Also, we cannot share this information with public institutions as it is a legal obligation. However, if there is a collaboration on the sampling, we conduct the surveys within our structure and share their results with the public institutions, private sector, and universities. TurkStat can carry out such studies on behalf of these institutions based on a mutual protocol.

Sinan Türkyılmaz:

So, you are not forming a frame that includes phone numbers. You make a selection from the address frame and match that with the phone information. Therefore, in your first selection, you do not exclude those without telephone information, but the individuals of whom you cannot access the phone numbers in the selected samples are out of the coverage. Did I get it right?

Enver Taştı:

Yes, those can be out of the coverage in that way. A non-response form was prepared for the households without phone information or those we cannot obtain from their family or other sources. But this is mostly for the first interviews. Besides that, there is nothing to do for others. The phone number information percentage is around 80-90 in the sample. Nevertheless, its bias and other aspects are issues to be discussed.

When the pandemic was in a critical phase in March and April, there was nothing to do for households without telephone information because we could not go to the addresses. Now that the situation has returned to relatively normal, household visits are made to obtain information from those who do not have a phone and cannot be reached in other ways, within the framework of the Ministry of Health's measures.

The curfew and related factors had an impact on data collection in the first months. We did not have any other ways for data collection, so it was only conducted with those who had phones. Currently, households who cannot be contacted or whose phones cannot be reached are visited one-on-one to get information.

Yaser Koyuncu:

When we hosted you for a lecture of our department, you stated that the Household Labor Force Survey started to be carried out via CATI, which was before the pandemic. This transition has been or will be made due to the pandemic for many surveys. Can you tell us your experience with the switch from face-to-face to telephone mode?

Enver Taştı:

As I said before, we already had a preliminary preparation for data collection with CATI. We have utilized CATI for data quality control purposes since 2005. We were conducting interviews with households who had phones for data quality checks after the surveys. TurkStat had an infrastructure for that. The Household Labor Force Survey is carried out in an 18 months period with the method called "2-2-2", which refers to two periods in a row, two periods vacant; and the last "2" is two terms of interviews. For this survey, we had already completed our preparations in 2019 to use CATI for the second, third, and fourth fieldwork. I have said this before in my briefing for your lecture.

Pilot studies were also conducted. That is why we had a background. However, the pandemic's sudden outbreak caused the system to be adopted without making the necessary preparations for other studies. But at least, the interviewers/operators in all regions who will work for data collection were quickly trained to communicate and conduct the interviews since the institution had the background. The transition to the telephone mode was made based on this work. I can say that there is not much of a trouble in collecting data, apart from the households we cannot reach, based on the information provided by our interviewers, colleagues in related departments, especially from the department in charge of the data collection.

Tuğba Adalı:

It sounds like perfect timing for the institution. You have partially answered my question. I will just add a dimension to it. We know that the Household Labor Force Survey has a semi-panel design. I have a question related to the way the new households are selected and added with the telephone. Is there a difference between households with no phones and those that have phones? In other words, when we draw the sample, is there a possibility of coverage error? Even though you eliminate a certain amount

of error by including those who do not have a phone in the sampling with household visits, are there any excluded groups?

Enver Taştı:

Let me wrap it up. The Household Labor Force Survey is a semi-panel study for which we go to the same addresses in four terms. For household CATI surveys that have multiple interviews, it is suggested that the first interview should be face-to-face, and the subsequent interviews can be completed by phone within a slightly narrower scope. Many European countries and developed countries were already doing this. We have also started to prepare for that. It seems as though around twenty-five percent of the total interviews are completed for the first time in each period.

First, sample addresses are drawn, and then people are matched with the phone numbers in the administrative records obtained from other institutions. Later, interviews are conducted by calling those phone numbers. Since the second, third, and fourth interviews are conducted with CATI in the Household Labor Force Survey, the household members' telephone and contact information are obtained in the first interview. As of our fieldwork in 2019, people were informed at the first interview that the following ones would be done with CATI.

From this point, one can say that there is no significant problem in seventy-five percent of the work. The part that might be a problem is making the first call directly on the phone. There are two issues with that. The first is those who have no phone, and the other is the consistency of the information received by phone in the first call. These were the obligations arising from the March-April curfews and the fact that we could not go to the field and access the households due to the implementations of the Coronavirus Scientific Advisory Board's advice. As I said before, all of the first interviews would be conducted face-to-face in a normal environment if everything went well.

If we exclude the people we did not have the phone numbers and those we could not reach in the first months of the pandemic, there seems no problem in terms of data collection for other interviews as it is now possible to go to the field.

A pilot study was conducted in 2019 to see the effect of these two methods on the responses. This study was conducted separately, and that was the important thing. CATI and CAPI were compared. As a result of the pilot study, there was no statistically significant difference between the two methods. When the results are examined, we saw that CATI had a higher response rate since it is possible to reach people even when they are at their workplaces, who may not be available for face-to-face surveys during the day time when you go to their doorsteps.

There is no comprehensive analysis of data quality. So it is a little difficult to talk about that. Maybe we will see that in the future with further analysis. However, we can say that there is no statistically significant difference in basic indicators based on the preliminary results.

Yaser Koyuncu:

I will ask a question about bias. Have you observed any differences that may cause bias in estimates between households that can be reached on the phone and those that are not? Actually, you mentioned it, but you can open up the subject a little more.

Enver Taştı:

Let me wrap this up. Under normal circumstances, a base effect is expected in these method changes. The literature I have read on the topic and studies in other countries show that it is the result of

switching from CAPI to CATI. I speak for normal conditions. Because as I said, we had a preliminary preparation for this.

Some reports indicate that there may be a base effect for countries. When you switch from CAPI to CATI, it is argued that you can sometimes get inaccurate answers to certain questions because of the lack of face-to-face communication. However, as I said, we did not observe a big difference in the preliminary analyzes done in 2019. Bias is expected for interviews done with CATI, which I also was of the same opinion since the concern that CATI can create bias is being discussed in the studies i have read. It is recommended to revise and simplify the questionnaire used in data collection to deal with that. Also, it is suggested to ask basic questions, eliminate detailed ones, design, and ask questions in a way that prevents bias and errors stem from the respondent.

We did not see such bias in the results of the pilot study. But as I mentioned, most of the country statistics offices switched to CATI due to the pandemic. Some have switched compulsorily, and some were already using CATI. And, CATI was also started to be used in other surveys that were face-to-face before because of the pandemic. Eurostat, ILO, OECD, and the United Nations delivered their recommendations on data collection for basic statistics, which also included what countries can do about data collection techniques as well as the experiences of the countries about the issue. Each country presented its studies at these meetings. Some stated that the new mode did not create a bias on the indicators, while others talked about the problems they had for some variables. We know that there are some biases, particularly for subjects with low prevalence and limited observations. Countries discuss the question of how to improve it. I also thought that this would create bias, but we did not see such a bias in the analysis of our data. Our colleagues also prepared reports and presentations for the issue. This year, The Survey on Use of Information Technologies in Households was conducted for the first time with CATI. I have read my colleagues' reports about this survey's fieldwork, which do not indicate such bias.

Tuğba Adalı:

I want to ask a question about the questionnaire. I wonder if your questionnaire was changed while adapting it to CATI? It is recommended to conduct shorter surveys and to keep the questions shorter for CATI. Has anything been done about it?

Enver Taştı:

That is right; the CATI and CAPI questionnaires should be different, where the main questions remain the same. However, no changes have been made to the questionnaires so far in our application, which was an imperative switch due to the circumstances. As I said, this was also referred to in the Household Information Technology Usage Survey's CATI report. CATI surveys require to reduce the number of questions and design the questions accordingly, which is important for data quality. There has not been any change in our current questionnaires. But we work on that. If we are to continue with CATI, it should be studied anyway.

Yaser Koyuncu:

You mentioned that the interviewers went through special training for the switch from face-to-face to telephone mode. In your opinion, what are the points to be taken into consideration for the training?

Enver Taştı:

To sum up: communication is of great importance. Diction and way of addressing are very important for relations with people in face-to-face and CATI interviews. It is important to know how to speak and

behave in these interviews. Let me put it this way; you have to instill confidence in the person you are talking to. In face-to-face interviews, interviewers show their credentials and other official documents, which develop trust, and then they move on to questions. That is why interviewers need to receive solid communication training to build up trust in phone calls. This training should address how to communicate, behave, ask questions, and act if a problem occurs. Having all these in mind, all operators and colleagues work on this topic, particularly those in the center, have undergone a long training. Operators started to work in all regions due to the pandemic, and communication experts trained them.

In addition to this training, interviewers must have information about the research. Because when respondents ask questions about the purpose of the survey, the data, and its usage, interviewers should be able to answer them. During the interviews, they should inform the participants whenever necessary.

In short, the first important point is communication, speaking skills, and training interviewers to have skills for persuading people. The second is to train interviewers on the basic information about the research.

Tuğba Adalı:

Let us ask you some more CATI questions. We learned that the CATI center of TurkStat gained a new function this year. So, how is the current CATI operation capability? How many interviewers can work?

Enver Taştı:

Before the pandemic, we had about 25 operators in the data collection center. They were specifically conducting trend surveys and other interviews of the Household Labor Force Survey. It was thought to double the capacity of the unit. Currently, the CATI center in Ankara is purpose-made. This unit was specially designed for our new building. It has a capacity of one hundred people, and these hundred people can conduct interviews at the same time. In special cases, it is possible to increase the capacity up to three times by shifts. But we can manage the operations with the current capacity.

Besides, the establishment of CATI centers in some regions was previously on the agenda. However, at this point, and infrastructure has been developed for all regions to connect to the CATI center in Ankara and make calls. Currently, about three hundred of our operators can interview at the same time. Our colleagues in other regions connect to the center via a private line, and three-hundred calls can be made simultaneously on this line.

Tuğba Adalı:

What kind of software is used? Is there a software that organizes these interviews and a separate one for CATI data input? Or do you have software that runs both?

Enver Taştı:

The last one. It is a software called Harzemli, developed by our IT experts. It is not only used in household surveys but also in establishment surveys as a data-entry program. We are currently collecting data using laptops or tablets in our CAPI surveys. We use the Harzemli software for that. It has an interface program for CATI. Calling the respondent, making appointments, and similar operations are carried out with an interface program, and the Harzemli CAPI application is automatically activated when the interview starts. The data is directly transferred to the system. It enables us to make the basic controls between questions. In other words, the compatibility of the

responses can also be checked. The data-entry program put a minimum burden on the respondent with consistency checks. We call it smart data entry programs, which complete the surveys by making basic controls in a minimum time.

Sinan Türkyılmaz:

Although there is such a facility, interviewers may conduct the interviews from their own homes due to the pandemic. How does TurkStat manage that? What may differ between conducting the interviews from their own homes and with a supervisor at the center? What did you observe? I would like to hear your views on this as well.

Enver Taştı:

Sound insulation was made in the data collection center, and the operators' places were separated. The advantage of conducting surveys at this center is that you can monitor the interviews. Working together in one place posed a serious risk due to the increased prevalence of COVID-19. Therefore, instead of collecting data from the center, mobile connection cards were provided to the operators. They connected to the system from home via a special line and conducted the interviews.

Currently, there is no difference between interviewers making calls from their home and doing that at the center. The equipment: In particular, headsets, microphones, and other safety measures related to the network were all provided in their home environment. They can conveniently make the calls; we have not had any serious problems with the application and data collection so far. As such, there is no need for being together in one place. Information about who will make the call, how many calls interviewers need to complete, the time they are made, and how long the calls took are available in the system. We have a central monitoring platform for this that we can track the interviews.

I can even say that working from home is more efficient since it enables interviewers to use time more efficiently. It is possible to eliminate the time spent commuting to work. Interviewers cannot close the day before completing the interviews assigned to them. Thus, I can say that we are very satisfied with the application so far, and it has been efficient.

Yaser Koyuncu:

I have a question regarding the Household Labor Force Survey. How did the switch from face-to-face to phone effect costs?

Enver Taştı:

It is not possible to talk about the exact impact now. It has an economic dimension and other dimensions as well. However, the biggest benefit of switching to CATI is that transportation costs are eliminated if we exclude the cost of phone calls. The Household Labor Force Survey can be conducted anywhere. Sometimes you may have to go to a village and interview three or five households for a couple of hours. If we cannot reach the person or the household, there are the revisits. It brings the transportation costs, and now we do not have them. Unfortunately, I do not have data on the exact impact on costs. Nevertheless, I can say that eliminating transportation costs allowed us to save money in a significant amount.

Second, CATI interviews take a shorter time. While the average interview time for a face-to-face survey ranges from fifteen to twenty minutes, it is seven to ten minutes for our CATI interviews. In other words, the interview time is reduced by almost 50 percent. It also enables to reduce costs shortening the fieldwork of the survey.

In short, the effect of the mode change on costs can only be demonstrated if the cost of the new implementation is calculated. So, I cannot give an exact number, but I can say that we made serious savings given the reduced transportation costs and fieldwork time.

Tuğba Adalı:

I would also ask you to open up the subject you mentioned. The Household Labor Force Survey has a mixed-mode design that has face-to-face and telephone phone mode components. How does TurkStat view mixed-mode surveys? Has the use of mixed-mode increased with the pandemic? Are there any other examples you can tell us?

Enver Taştı:

In fact, this is not just about the application and in Turkey and TurkStat. Mixed-mode studies were conducted in the world before the pandemic. We have been classifying the data sources as censuses, administrative records, and surveys. Now, new possibilities and opportunities have arisen due to technological developments. The standardization and use of administrative records for statistical purposes have been discussed for years, and developed countries use administrative records very effectively. Many countries in Europe produce most of the significant statistics based on administrative registers.

With the increase in internet and technology usage and the advent of the 'internet of things', the data flow from the smart systems at home, workplaces, factories, and transportation has been thought to use for statistical purposes. Even while driving, you leave traces such as in recordings of city surveillance cameras and traffic lights. These all constitute what we call "big data," where the internet, social media, and smart city records are all together. There have been serious studies in the last five-six years on new data sources such as big data, which is considered to lead the data revolution. Now studies focus on which data to obtain, as well as how to get that more effectively, and the main concern is the quality and standardization of data. There are also issues regarding the classifications. Many countries discuss how to use this.

In my opinion, the pandemic accelerated this process since it caused institutions to use technology more. The pandemic has accelerated the statistical institutions' turn to new data sources. We were already using CATI in Turkey for trend surveys, rent surveys, and small-sized, short-questionnaire surveys. There were thoughts about switching to CATI for other surveys. But now it became an obligation and forced for the switch automatically. Currently, the pandemic has forced a transition to mixed-mode. There is a full switch to CATI now, and it may continue if there is no problem. As I said, there was no problem with using CATI in the Household Information Technology Use Survey. Would we have switched to CATI completely this year if there was no pandemic? Not sure.

To sum up, yes, the switch to mixed-mode would happen in the future anyway. The pandemic advanced this process. Frankly, I think it is a good development.

Yaser Koyuncu:

You have said that the transition to mixed-mode is accelerated. And, considering the pandemic, what are your predictions in terms of the interview environment and techniques for the future?

Enver Taştı:

We can summarize it as follows. In fact, I think mixed-mode will prevail, and there is no turning back. I am not just saying this for our country (Turkey). Considering the measures that other countries are

currently discussing, I do not think there will be a return in many countries. Every country will continue on its way. The only thing is the switch's implications, particularly on data quality, which are being discussed with changes that need to be made for questionnaires and methods.

Unless there is a compulsory reason to change, I think the Household Labor Force Survey's first interview should be continued to be done with CAPI, and CATI can be used for other interviews. I think that CATI can be utilized particularly for small-scale studies. But the thing is that I think administrative records are indispensable in the production of statistics processes. Administrative records have become much more important now. If we can form our administrative records in accordance with more accurate and proper classifications and standards, we will ensure that the data "speak" among themselves. We referred to the Address Based Population Registration System and the National Address Database, which we can use to produce different statistics by integrating them into other administrative registers that comply with common fields and standards; thereby, we can advance our statistical diversity.

When we improve the standardization, quality, and content of administrative records, we can produce data by integrating them with our current studies' variables. You know, some basic data such as age, gender, educational status, and marital status are obtained in household surveys. For instance, if it is possible to get the person's up-to-date educational status from the administrative records, it will reduce the respondents' burden by enabling us to integrate the survey into the administrative records. With a much more integrated and simpler questionnaire, I can use CATI for the data not available in administrative records or if it is inadequate. Or, it is possible to get the required data from administrative registers as well. I think such studies can be applied with the mixed-mode.

In recent years, there have been serious developments in administrative records usage for statistical purposes in Turkey. The Revenue Administration brought their administrative records into use for statistical purposes, and now almost all of our establishment surveys are conducted based on these data. In a similar vein, with the establishment of ABPRS, we produced many statistics in the demographic and social fields we had not produced before. Professor Türkyılmaz will agree with that. We made serious progress, particularly in updating the address frame. Before that, we were conducting listing fieldwork, and we were facing many problems about that. New settlements were not included there. There is no such problem at the moment; we have a continuously updated frame. Based on this frame, there are also statistics we produce from administrative records.

To summarize, many countries put the usage of mixed-mode on their agenda. I think mixed-mode is important to increase data quality and produce data in lower breakdowns. With mixed-mode, I think it may be more accurate and reliable to make predictions even for lower units using statistical techniques and models, let alone its contribution to the fieldwork.

Tuğba Adalı:

Household surveys can have many different topics, such as social, economic, health, market, and political surveys. TurkStat and other public institutions conduct surveys on various topics. Do you think the pandemic makes it more difficult to conduct surveys on a particular subject? What subjects, if any, are more difficult to get information about due to the pandemic?

Enver Taştı:

Some surveys are difficult even with face-to-face mode. For instance, in the Health Survey, there are many detailed questions. Conducting this with the telephone mode can cause problems. On the other hand, in income and living conditions surveys, there is not much problem with yes-no questions as well

as questions about income, social issues, or questions that the respondents can answer comfortably. But, some questions cause problems if the respondents are hesitant about its content. We encounter problems with these types of questions, even in our face-to-face surveys.

I think the implementation of CATI raised these problems further. Sometimes it may be easier to get answers to some questions on the phone. Respondents can easily answer a question on the phone that he/she cannot talk about face-to-face. We need to consider its pros and cons. That can be evaluated specifically for the survey. CATI may work much better than you might expect on some, but it can also cause significant bias in the surveys I mentioned. For instance, when their educational status is asked to respondents on the phone, the participant can hesitate and say that he/she is a university graduate when he/she actually is even illiterate. They may not tell their real profession. We have no tools to check it. There are special health-related situations; Respondents may think, “Why should I give this information on the phone to a person I do not know? “Even on the phone, they may have difficulty talking about it.

In some special studies, you interview the individual going to the household without any other household members in the environment during the interview as it is the case for Life Satisfaction Surveys and other studies related to satisfaction. The Health Survey has questions about smoking and alcohol use. You first isolate the individual and then conduct the face-to-face interview so that their answers are not affected by their environment. Moreover, there are concerns about the environments interviewees can answer the questions and the accuracy of responses depending on the respondent’s environment for CATI.

To summarize, I think it would be more appropriate to evaluate the surveys individually. There are pros and cons. In my opinion, it would be useful to consider the characteristics of the elements, such as the results you will get from the survey, the purpose, and the critical variables you want to obtain data about. You cannot talk precisely. They all have pros and cons.

Tuğba Adalı:

Do you send a letter to households before making the first call?

Enver Taştı:

Yes, we do. We give brief information about the survey, providing the phone numbers to contact us whenever they wish, stating where this data will be used, their legal rights, the confidentiality, and that it is not possible to share the data with third parties. They can call both the center and the regional offices whenever they want. In addition, with the transition to the CATI, we send an information SMS that includes the phone number the person will be called from. It is important because someone else can call and say that that he/she is calling from TurkStat. Also, a line called “TurkStat Call Center-124 “ has been established. Information about this line is provided via SMS and letters beforehand.

Sinan Türkyılmaz:

You know, we conduct interviews with different people at an international level. We ask a common question to all of them. I have received the answer to a large extent in previous answers, but I want to ask not to break the rule. If COVID-19 disappears overnight, how will research activities be affected? Will we return to the methods used before the pandemic? Did it have permanent impacts?

Enver Taştı:

As you said, I answered the question partially. I think that nothing will ever be the same, that is for sure. When we consider the technological developments and means institutions have, I think it is necessary to use new data collection techniques and data sources. With the pandemic, this process accelerated. Especially public institutions had to take faster steps for this. As I said, after the outbreak of the pandemic and its intense spread in February and March, international organizations made the following appeal to all countries and statistical institutions: Please somehow continue to collect data on basic, urgent and continuous indicators like labor, price, national income, and similar ones. Each country postponed the surveys that could be delayed except for urgent ones and continued to conduct them.

As a solution, most of them started collecting data with CATI. They started to use administrative records effectively. The importance of using big data sources has increased even more. For example, big data are used for predictions about tourism surveys. They obtain successful results. Further, by doing preliminary studies about unemployment, decent estimates are made based on those numbers.

With the pandemic, new data collection techniques have been started to be used, and I think these techniques will be developed and continued to be employed. However, I do not think that classical methods (FTF) will be abandoned. If the informality is high or there are no third data sources that we can check, the first interviews will and should be conducted with the classical method. I can cite the Netherlands as the best example for that. Administrative records in the Netherlands are extremely rich, and they use them a lot. However, the country still continues to conduct household labor force survey in the classical method. The compulsory switch to CATI may continue for 3-5 months. However, in the medium and long term, I think that the first interviews should be conducted with the classical face to face interview mode, particularly for significant household surveys. Maybe some other development can happen, such as conducting the fieldwork with a more compact integrated questionnaire. My guess is that studies that do not have a quality problem applying CATI, compiling data from administrative records or obtaining data from other data sources as big data will continue in the same way.

4.5 Interview with Frauke Kreuter (Online Surveys and the Use of Big Data) ⁶⁶

13 November 2020



Professor Frauke Kreuter is Co-Director of the Social Data Science Center and faculty member in the Joint Program in Survey Methodology at the University of Maryland, USA; faculty at the Ludwig-Maximilians Universität München, founder of the IPSDS at the University of Mannheim; and head of the Statistical Methods Research Department at the Institute for Employment Research (IAB) in Nürnberg, Germany.

Tuğba Adalı:

Professor Kreuter is the director of the Joint Program in Survey Methodology, based in the University of Maryland; she is the founder of the International Program for Survey and Data Science at the University of Mannheim and is also affiliated to the Institute for Employment Research in Nürnberg. In addition to teaching courses on data analytics, paradata, responsive design, and big data, she is among the authors of many scientific articles and books. We will ask her today about online surveys and big data in the context of the pandemic.

In the last two to three decades, the survey world has witnessed a proliferation of modes, with the three basic modes of surveys dominating the field. Would you say the Covid-19 pandemic has an impact on this increase on this proliferation? Do we see new modes or new combinations of modes emerging?

Frauke Kreuter:

I cannot speak for all the countries, but I can see closer what happens in Europe, Germany, and the U.S. It has sparked a lot of innovation. Web or cellphone-based surveys are used even more. In the developing world, we have seen that face-to-face interviews are still pretty common.

The World Bank did surveys, for example. There is now a big push to do more telephone surveys in countries where telephone surveys have been used for a while in combination with face to face. There was a push to bring people to the Web, because it is harder to reach them on the phone. There is a shift, but it depends on where they started.

Additionally, I can mention new emerging modes, because I think what we see increasingly are attempts to augment interviews with a video interviewer, even though I have not seen this fully in production yet. We are operating right now through invitations using Zoom or other products to do recordings and interviewing. That has been used before in the user experience research context, but not necessarily for large-scale surveys. They use it now as well.

⁶⁶ Video record of the interview is available at: <https://youtu.be/DW2qI34LIkY>

We have seen the other shift that data collection agencies have moved, to create panel surveys with frequent intervals. It is not a new mode, but a new way of doing things facing the need to get more information quicker. Pandemic-related, if there is the need to check people every week, or every two weeks, you need a different way of contacting them. Thus, as we call them, these access panels have also seen a greater uptake, I would say.

Sinan Türkyılmaz:

Web surveys have become another option for cases where face to face is not possible. So, from a total survey error perspective, which sources of error are the most underlined for web surveys, would you say?

Frauke Kreuter:

I am not sure if it is possible to answer for all countries and settings. Like a poorly designed interview-survey, a poorly designed web survey can have many measurement error problems. People get confused about using the instruments, and you have a huge break off as it may be too long. People may not have the patience to sit through these surveys. So, the instrument design is just as important as it always was. The most significant source of error stems from the web-access and internet penetration that depends on the country.

When we say Web surveys, I often tell our students and folks that they must think of it as a mobile device survey. People may see these invitations and cannot go to their computer. They fill them out on their smartphones. So, having a good interface that works on smartphones is key. Not every survey constructor pays attention to that.

Tuğba Adalı:

It sounds like you are saying that coverage depends on which country we are talking about, but measurement is pretty much key in any case.

Frauke Kreuter:

Yes, I would say that. It has the potential to suffer from all the problems. It depends on how much attention is paid in the design phase to different pieces. I do not think that it is, per se, inherent in a certain mode and that some modes are more prone to that. Having an eye on them and the mindset is important. Conducting a Web-survey is cheaper because you do not have to spend for interviews, which means that it is possible to pull all that money into incentives, for example. That would help with non-response. Overall, it depends on how you implement it, and I do not think we can speak of a certain and overarching mode that suffers from all TSE (Total Survey Error) elements. It has combined elements with regards to the design.

Tuğba Adalı:

Now we have a list of question about the first Facebook survey you and your colleagues designed this year, in collaboration with various institutions, including the Joint Program in Survey Methodology at the University of Maryland, the World Health Organization, the Delphi Group at Carnegie Mellon University, and Harvard, Stanford, Yale, and Johns Hopkins University.

This is a very large-scale international survey that uses Facebook users as a sampling frame. It asks respondents about COVID-19 symptoms, testing, social contact, feeling of depression, and alike. Can you tell us the innovations this survey brings? Can we say social media is more than a big data source but has features to offer for traditional surveys?

Frauke Kreuter:

It was an incredible investment from Facebook to work with the WHO's various research teams to make this happen. A lot of people are behind the scenes, at Facebook, in particular. Their data science team, as well as the demography and social science group at Facebook, are working to make sure the operation flow works. Sampling and weightings are created every day. Doing this on a global scale is a gigantic effort. You need a big partner like this.

What is unique, I would say, is that it is such a far-reaching platform, and this is the advantageous side. Also, it certainly is an innovation given that we suddenly had a global sampling frame; that, of course, has flaws as any frame does. There may be more flaws than other frames that we are used to, since obviously, we can only sample people on Facebook, by nature. Our main research question or element of interest may be the change over time and observing what happens over time in the country, so that policymakers can adjust their decisions or reshuffle resources. In this short time frame, we can be pretty sure that the sampling frame, the population, does not shift that quickly over time. So, the point estimated level at the data we get might be different. Maybe we over-report certain things or under-report certain things. However, if the population composition does not change quickly over time, then the development can be nicely monitored with it. It reminded me of the billion price project, that Roberto Regan bonded at MIT, where he tried to get alternative price indices from online postings of crisis for Latin America initially, which now is done across the world.

Yaser Koyuncu:

Referring to Mick Couper's paper titled "Is the Sky Falling? New Technology, Changing Media, and the Future of Surveys" which was published seven years ago, he goes, "For example, Facebook is not likely to make their database of members available to researchers for sampling or analysis, even at a fee." What do you think has changed in this last decade that has led us to where we are now?

Frauke Kreuter:

Yes, the world is changing. There are a couple of things I would like to refer to. One is that, even prior to the pandemic, Facebook always looked for ways to use some of the data for the public's social good. In my perception of the company, they truly believe that their mission is to bring people together and connect people. So, it falls right into the mission. Before the pandemic, they had a group called "data for good", or something like that. They worked with the World Bank and allowed them to field surveys on the platform to investigate, for example, females' participation in the labor force or as entrepreneurs. I think female entrepreneurs were the focus.

Besides, I think there was an increasing awareness at Facebook that it is possible to contribute, as it is a good way to mitigate some of those concerns related to negative media they faced. I would not take that off the table as a motivation that has changed between then and now. Google did a similar thing trying to make the mobility data available in a similar way, while being concerned not to jeopardize the people's privacy. I think, to the extent that they can, and it does not hurt their business, these companies are open to that idea. It is more likely that they do this in the fashion they do now, than they would do it for a fee, that you can regularly pay as a service. Because then it becomes a product that you have to be able to deliver. After that, if it is not aligned with your core company interest, maybe that will not work anymore at some point. I think this "social good" aspect will maintain, which I see in many companies, and I see an interest too. When someone has a good idea that does not harm the business or people's privacy, the companies will be open to that. I do not think that is a solely pandemic context behavior.

Sinan Türkyılmaz:

You already answer some part of the question, but maybe you can give some more details. Can you comment on the coverage of the survey and assess its representativeness? In other words, what would you say about the omissions, multiple memberships, non-person memberships, and fake memberships? I would like to hear more about the coverage.

Frauke Kreuter:

Fake members and trolls, I think, would have a hard time answering a survey. It is way too much action that is not pre-programmed. So, I am not worried about that in the survey. Concerning the coverage, I hope that we increase the uptake of the microdata use globally, or even just the aggregated data, which is from people of various countries with a lot of country-specific knowledge. From what we have all seen, both coverage and the data's predictive power are fairly good in the U.S., because there is still a large portion of the U.S. population on Facebook. Also, the sampling happens at an even more fine-grained level. I mean, in theory, we could adjust down to the zip code level.

For other countries, it is really hard for me to judge because I do not know. I can look into this in more detail for Germany. Just glancing at it is easier for me than doing it for other countries. One would need to look much more closely. Frankly, our team is too small for that. We have not had a chance to look at this for all the countries. However, I hope to do so that maybe your team or other folks use it, because it is a phenomenal data set, which has been going on for six months with the same questions. The translations were superb; we were impressed with the translation teams at Facebook, given that we ask the same questions for all these countries week by week and day by day.

I hope we can increase the user base and learn from those who apply and look to their own countries for validation sources, and we also can learn where to fine-tune. We make adjustments to the questionnaire around seven days coming up, and as the world learns about COVID-19, other things are needed to be asked, including symptoms and covariates.

Tuğba Adalı:

I would like to ask about the response and non-response rates in this survey. We know that the data you collect cannot be linked to Facebook data or Facebook users. Facebook provides daily weights to correct the age, gender, and region. Is there any other non-response assessment on the Facebook-side for any potential bias?

Frauke Kreuter:

First, they do the sampling internally on the platform. So, they know who were sampled and eventually participated, or at least those click and transported to the survey. Then, they get an ID from us that says that person finished or answered the survey. There may be people who dropped out in between; we weight them back. Therefore, the identification is exchanged. Afterward, they can look internally using the features variable. I call them features, because they are not necessarily designed variables that data scientists would use from the platform data to adjust for any non-response bias. Indeed, it might be better not to use the word non-response, since there are clear definitions of that term. Better still, it is to adjust for the non-participation error.

As I said that paper alluded to that in more detail. I do not know all the features and variables there and how they are coded. If we were to know this, someone smarter than I could eventually reconstruct

the original information from the weight. I have a hard time seeing how that would work, given the data that we have, but I would not put it past the realm of possibility. So, we just do not know. I mean, if we would know information about gender and etc., we might reconstruct. But we do not know that, and that is a good thing. I know that the team works very hard to use what they have and what they normally use on surveys that run on the Facebook platform.

On top of the non-response adjustment, post-stratification weighting is also employed from country data to mitigate the coverage error. But, this is the usual demographics, and there are not many variables for all the countries. Also, who knows that those variables drive being on the platform or not. Overall, from a statistician perspective, I would say that the coverage adjustment is minimal probably.

Tuğba Adalı:

And how about the levels of response rates?

Frauke Kreuter:

It varies a lot by country. Also, it reminds me that we should publish it somewhere. We looked at it internally and are monitoring it. It is given that response rates are low for these web surveys. What concerns us most, is that people start the survey and might give up in between for all kinds of reasons. That is why the COVID-19 symptom questions, which are the most important ones as early warnings, are in the beginning. We are discussing whether it is possible to shorten the questionnaire to reduce dropouts during the survey. But that comes down a bit, compared to the beginning of the survey, which is not surprising. In some countries, they are in the re-sampling phase that people see the invitations. There is a cool-down period for everybody, but then you might see the invitation again. And maybe next time around, it is not that exciting. Plus, I guess there have been many surveys and attempts to get people's data. I would not be surprised if we see the usual fatigue setting in.

Yaser Koyuncu:

We know that record linkage is one of your research areas. Could it have been possible to enrich the Facebook survey data through users' Facebook content with consent at the end of the survey?

Frauke Kreuter:

From a technological point of view, yes. From a privacy point of view, we want to avoid that absolutely. After all this Cambridge Analytica scandal, etc., Facebook had the policy of not allowing anymore. We would not have been allowed to do that in the survey, although we initially thought about asking respondents to install one of those COVID-19 apps as Germany has. This kind of "chaining" is no longer allowed, like inviting them to install that, getting consent to keep asking other survey questions, or enrolling them into a panel. It is against the policy.

I think the work I have done on record linkage might be relevant. Even more relevant is the work that I have done on the linkage consent. I would have been way too concerned that people might consent to something they do not think through and understand its implication. There was never a discussion about doing this. I think it is a significant privacy risk. On the Facebook side, the policies would probably not have allowed that.

Sinan Türkyılmaz:

Is the research team considering a longitudinal dimension to the survey?

Frauke Kreuter:

This question came up a few times. Not on an individual level, for the same reason. That would have required that IDs are kept unmergeable. To have the data usable widespread across the world, we must make sure that there is no risk of the answers leaking. It would have been nice. Nevertheless, with this kind of design, everything else would have jeopardized the goal. I would not rule it out. One could think of a design where that is possible. But not with this vehicle.

Tuğba Adalı:

This marks the end of our questions on the Facebook survey, and I have another question here regarding Facebook. What do you think of designs that use Facebook Ads or Google Ads for recruiting survey respondents?

Frauke Kreuter:

You probably do not know that, but we have started doing that. I am working with two of my graduate students, Leana and Samantha, and another professor at JPSM, Stanley Presser, on exactly that. You mentioned Groves earlier. There is a famous paper on the non-response where different topics are used to recruit people and see how the non-response bias answers and its distribution changes, depending on the recruitment. We have implemented an experiment in a similar fashion. Moreover, we used the questions in a survey done somewhere else in the world, and we can compare distributions to these other surveys. I cannot speak about the results which are not available yet. I think it is just as promising an alternative as many of the non-probability panels are.

The difference between the Facebook COVID-19 symptoms surveys we talked about and surveys you recruit respondents with advertisements is that we have no control over the algorithm that advertises to certain people. You can say that you want to target a certain quota, like people in a certain age group. However, you have no control over the aspects of who gets it, how often, and in which combination. There is a lot of algorithms going on for advertisement display and how friends of these people react to do that. For the COVID symptoms survey, people go to the news feed, and we know that if they see or not.

It is just a very different setup, but I think it has its place. If you are a researcher and you frequently use platforms like Mechanical Turk or SurveyMonkey, recruiting your respondents over the Facebook platform via advertisement might provide a better population. You can reach a much more diverse group than those that sign up to be one of those click workers. You have a higher likelihood of reaching someone on occasion who would do a survey but never sign up for any of these panels. But this is just speculation, a gut feeling from a survey methodologist. I have not done any research on that part yet, but I am watching it with great interest.

We have a colleague in Germany who studies the LGBT community. It was great for him because he could recruit smaller populations harder to find. We have seen this in the COVID-19 context too. In case you are interested in a niche subgroup, small in comparison to the whole population, and you recruit people this way setting up a longitudinal data collection, while the focus is seeing what happens to those folks over time, it can be a very terrific source. I am curious to see more research in that area, both on the content side and survey methodologists' works evaluating the data quality.

Sinan Türkyılmaz:

Let us move to big data. We would like to ask you about error sources in big data. It is a hot topic in survey research for about a decade. I want to ask your thoughts about how big data error sources differ from the Total Survey Error perspective developed by Paul Biemer, discussed in the AAPOR Task Force Report and in your co-edited book titled "Big Data and Social Science: A Practical Guide to Methods

and Tools". What are the common sources of error in surveys and in big data analysis, and what are the additional sources of error researchers be wary of when analyzing big data?

Frauke Kreuter:

The TSE framework is an important asset for survey methodologists and social science researchers to deal with big data since they are more sensitized to think about different error sources. The way you call it does not matter that much. It is more about the mindset, knowing that there are all these different error sources along the data generating process, and it is essential to keep that in mind.

We need to learn more about all the errors that could happen on the big data pipeline's technical side. TSE framework always suffered a little bit from not having enough attention to how variables were created. In the big data stream, many features need to be generated out of the data. Some of this happens algorithmically; some of this happens deterministically. Looking at that and all the decisions that I made in that segment are something different between these. However, the spirit that there are these multiple sources, some have a more social mechanism connotation and effect measurement, and others have a more technical connotation and affect the presence or absence of the data is a significant asset and helps us use these data more wisely than maybe people from other disciplines.

Yaser Koyuncu:

Do we see any new exploitation of organic data to understand COVID-19 related matters? When we think of Google Mobility Data, Google search trends for symptoms, wearables, COVID-19 related apps, mobile phone vendor data, et cetera.

Frauke Kreuter:

I have seen attempts for all of those, and I like some better than others. With any data and any measurement instrument, the question is what you want to ask. For a certain aspect of the problem, that might be a good source, but it may not for others. Let us take the example of the wearable. It is cool that you can learn health aspects depending on what the device can measure. But it often relies on the voluntary provision of the data, and people are aware that you can donate the data, not being prompted or randomly sampled. You read it on the news, and you say, "I donate my data," which means that it leaves out a huge part of the population. Even figuring out how you would do this is a problem when you consider the risk groups in the COVID-19 situation, for instance. If I think of my parents, they would be utterly overwhelmed by installing the app, uploading data, or using it. Therefore, I think it is crucial to have the TSE framework in mind and evaluate the data's certain use and the quality of these different data streams.

Tuğba Adalı:

The 2015 AAPOR task force report, which you were a part of, mentions the big data era as a "paradigm shift." Considering the face-to-face surveys which are on the verge of disappearing due to the pandemic, do you think this paradigm shift mentioned is picking up the pace?

Frauke Kreuter:

I am inclined to say yes, just because everything feels like it is picking up in pace. However, if you ask people ten-twenty years older than me, they will say that it has always been changing. I think it is often easier to look post-hoc and draw a demarcation line considering other things "happened afterward." There is a changing pace in the sense that more people talk about it and feel that this is an important thing to consider. But I think you cannot just rely on these other data sources. What is needed, more

likely, is a combination of different data sources. Surveys still have their role, and we might be well advised to think more about the products that have a variety of different data sources rather than banking on big data sources x,y, or z.

Sinan Türkyılmaz:

I have a question parallel to that, and I already get some responses. Do you think big data will eventually bring an end to classical survey research? If not, what would be the contribution of classical survey research to data science?

Frauke Kreuter:

To me, these are two very distinct questions. Will it be the end? No. Its role is shifting. It is a good idea for anyone to think of an alternative data source before collecting everything with survey data. Also, I think we had a time of too many surveys out there and unnecessarily so. Anyone could learn a lot about human behavior without necessarily having to ask.

Despite the disconnection between expressing attitudes and action, there can still be tremendous value in asking attitudes and digging a bit deeper. Qualitative interviews also belong to the classic canon of surveys, particularly when instruments are designed. I can see a stronger connection between user experience research or the qualitative in-depth interviews and how people use certain devices to better understand the data they produce. We did many surveys on smartphones lately, having the behavioral data and which apps and how these data are used. It is clear that people use multiple devices. We cannot assume that one person does everything we pick up from a mobile phone. In order to know that, you need to ask folks and observe how they use the phone. We are doing it now.

Furthermore, I spent some time in Silicon Valley in 2018 in full disclosure, working with Facebook on some other projects, and talking to many other big or small tech companies. I thought the reason why they want to talk to me is that I build data science programs. No, they wanted to talk to me because I am a survey methodologist. Despite all the data they have, they frequently need to collect some survey because they do not necessarily know the meaning of their data. For example, in Not So Standard Deviation podcast of Roger Peng and Hilary Parker, Hilary often talked about how the firm Stitch Fix, based on machine learning algorithms, suggests things like what you could be wearing or putting on your neck in terms of clothing. It was a true game-changer to these sorts of surveys to the data production pipeline.

I think that the disciplines will be moving closer together. That is the reason I built this program of survey and data science. For many research or industry questions, you just need both sources of information.

Yaser Koyuncu:

What kind of research topics do you think are most affected by the pandemic? What kind of data will the world be struggling to collect the most due to the pandemic?

Frauke Kreuter:

I think that census data collections can be affected. In ways that it might be hard to verify people, which is done in person. If it continues to be impossible in the same way, that will be a big change. It depends on the time we need to get a handle on the pandemic. The problem with the pandemic was that we had to shift modes without being able to experiment if the old and new mode can run in parallel. You can estimate the measurement difference between the two and then continue onward

with the next one. We will see jumps in the data that we will not quite know if it is the pandemic effect or if it is about the mode shift. I think that is a bit unfortunate. Some data collections have to be stopped or paused. I hope that they will be able to pick it back up. There is always a flip side too, and I have seen an increased speed of making the data available. The speed for the work we have been doing in the U.S. around the Coleridge Initiative, where we helped state and local agencies to unlock their administrative data, have them in a research environment where researchers would not have happened without the pandemic. Is this the bottom line, the net positive or negative for data availability and social, economic research? I do not know. We will see. I am curious. We should talk again in five years.

Sinan Türkyılmaz:

Let us assume that the COVID-19 pandemic is eradicated overnight. Do you think surveys can go back to the pre-COVID-19 instantly? What challenges do you think we should be expecting after the pandemic is over?

Frauke Kreuter:

One big challenge is the financial meaning of this. How are the resources going to be distributed in surveys? Good surveys cost money. That will continue to be a challenge. I think many things will not go back to pre-COVID just like that. There have been innovations in data collection that probably will not go away. I mean, people probably fast-forwarded the decentralization of telephone interviewing. Instead of having everyone in a telephone studio, interviewers can make calls from their houses and use a joint platform to distribute cases. Why would you need to go back to the old model? There is software that can be used to monitor the interviewers. I cannot see the reason why those kinds of innovations would be turned back.

I am hoping that most innovations will prevail. I have a harder time thinking of additional challenges that will happen post-COVID, given that everyone on these teams being utterly exhausted since this year was just draining. It is probably true for you guys, too; everyone is trying their best to make as much data available as possible. I could see us all falling into a big dip. I am more worried about the people who invested here than structural things that we cannot figure out.

Sinan Türkyılmaz:

I believe that even though the pandemic itself is not a positive thing; it positively affects the survey research methodology. It has a significant impact on different ways of data collections in survey methodology.

Frauke Kreuter:

And it connects people. I mean, the interview we do now, all the conferences that are now more easily accessible for people who would not travel to the U.S, for instance. I appreciate it a lot. In the seminar series in JPSM, which is open to everybody, we now see 80-90 people, which was fewer before. I think that knowledge distribution is increasing too. Of course, we all miss in person, in action communication, the possibility to exchange an idea more casually in the hallway. But again, these are all tradeoffs. It will be good to preserve the interconnectedness that we all now have based on this shared, yet negative, but still experience.

4.6 Interview with Attila Hancioğlu (Impact on International Household Survey Series)⁶⁷

20 November 2020



Dr. Attila Hancioğlu is the chief of the Data Collection Unit, Data & Analytic Section at UNICEF headquarters in New York. He is the coordinator of the Global Multiple Indicator Cluster Surveys (MICS) program. He earned his PhD from Hacettepe University Institute of Population Studies.

Tuğba Adalı:

We are today with Dr. Attila Hancioğlu. Doctor Hancioğlu is the chief of the Data Collection Unit Data and Analytic Section at UNICEF headquarters in New York. He is the coordinator of the Global Multiple Indicator Cluster Surveys program, often abbreviated as MICS surveys. We are proud to say he is a demographer and has a Ph.D. from the Hacettepe University Institute of Population Studies.

Today, we will ask him questions about his MICS experience regarding the COVID-19 pandemic. MICS surveys have been around for about 25 years, with data collected in more than a hundred countries. These surveys produced key statistics primarily on children, women, and households. The COVID-19 pandemic has become a major obstacle to face to face interviewing around the world. We had a chance to see different parts of the world, coping with different challenges in terms of surveys as well as different types of surveys coping with different challenges.

To start, can you tell us the state of MICS surveys around March when the pandemic was declared? Were there any countries in the field? Were there any countries getting ready for the field?

Attila Hancioğlu:

First of all, thank you very much for this opportunity. I am really pleased to be speaking to you today. Let me come from the context, put this all into a context. This round of MICS has been running since 2017 and will end next year, in 2021. We have had an unprecedented number of surveys joining the program.

In March, we had 71 surveys in the program, having 58-59 countries joined, I think. Some countries happened to do more than one survey in an around. For instance, Pakistan has provincial-level surveys. There is a separate story for Punjabi; there is a separate story for Sindh, et cetera. That is why the number of surveys and countries is different. We had 71 surveys at the time, and out of those 71 surveys, we had seven of them in the field, as COVID-19 came in, and the restrictions were imposed worldwide. I think; the type of restrictions and the intensity of restrictions are quite different from one country to another. However, seven surveys were in the field. A large number of surveys have already been completed. They were either published or about to be published or analyzed.

We had several countries that were in the design and planning stages. Furthermore, we had about five or six countries who are about to go to the fields. They were either training the field workers or doing

⁶⁷ Video record of the interview is available at: https://youtu.be/ie_vMpC5y4s

the logistics. Of those seven surveys in the field, we quickly assessed where they stand and immediately suspended. We recommended that the countries suspend the fieldwork in seven countries, and they did. Then we started an assessment to see how much fieldwork had been completed.

Remember, at the time, this is March and April we are talking about. There was still uncertainty as to the trajectory of the pandemic in the very early days. I remember that it was mid-March or early March when this came in, at least to the US. We thought of it as another emergency, which will happen and go away. We would go back to normal. That was the mind frame of the whole world, I think, at the point. We did not understand the extent and depth of this pandemic. Now it is a whole different story, of course.

We quickly assessed all those seven surveys in April. We assessed where they were, how much of fieldwork was completed, whether we could call them finished even. We found out that out of those seven, five had gone all the way up to 90-95% completion of fieldwork. Just a few clusters were remaining. We considered those surveys completed. We did some work to recalculate sample weights to see the distribution of the completed and incomplete clusters. Those five surveys were considered complete at that point, and we moved them to data processing, data analysis, editing, and reporting. I think, by now, most of them have been published.

There were two surveys, which were not near completion at all. One was Malawi, and the other one was Turks and Caicos. They had completed 50-60-70% of the fieldwork. It was not possible to consider it complete. That had to do with how much they had completed up to that point, which was not a good geographic distribution. There was no way that we could come up with sample weights that would compensate for that incompleteness. By the way, those two surveys went back to the field in June-July. They resumed and completed the fieldwork. The other five surveys; four of them were above 90% completed.

In Argentina, it was around 60%. However, it was good when we looked at the distribution: the way that they had planned the fieldwork, the logistics of the fieldwork, how they would move from one area of the country to the other. It allowed us to say we can do some analysis, look at the sample weights, recalculate sample weights, and see if we can complete the survey without going to the field again. And we did. I think the Argentina survey is about to publish now. That is more or less what happened, with the fieldwork going on at the time of the COVID-19.

Yaser Koyuncu:

I wonder what the response to countries for which preparations were underway was? You talked about it, but have there been any mode switches or cancellation?

Attila Hancioğlu:

I do not think any survey has been canceled because of COVID-19. A couple of surveys have been canceled since then, which has nothing to do with COVID. Those were two surveys that were uncertain primarily because of internal problems in the countries. Except for those, I do not think we have any cancellations, but there is much postponement, of course. What we did was; we first temporarily suspended all the face to face activities in all the countries. Then we thought about what we could put in place to continue or even accelerate some of the other activities, which normally depend on face-to-face interaction but can be completed in alternate ways. For instance, workshops or data interpretation workshops and meetings, steering committee meetings, technical committee meetings, and the actual technical support we provide to the countries that happen on-site and off-site. How

much of it can we do off-site? Can we continue to support countries or data processing from afar without going to the country?

One of the things that we thought was to use this as an opportunity to accelerate progress, not progress on fieldwork, of course, since they were suspended. But especially for those countries who had completed fieldwork, we thought about accelerating the data's editing, analyzing the data, and writing the reports, which we have partial success in, I would say. Not 100% success because we did not also foresee in all this clarity at the time; how much of these restrictions would put this burden on the National Statistical Offices, on our counterparts. In some countries, for instance, there were complete lockdowns for a long period. The staff members were not equipped, were not prepared to work from home to continue working on data, for instance. So, that delayed quite a number of things. Otherwise, our approach has been: "Let's suspend face to face activities but anything which is not face-to-face and try to accelerate them as much as possible virtually."

We have been doing things that we never did before. For instance, we have data interpretation reports and compilation workshops, which happens in every country after collecting and analyzing the data. That is in preparation for the writing of the reports, the final report. That has been done in several countries, virtually. It is not the real thing, as you would imagine. But I think it has been really good. We have been able to make progress on those.

Tuğba Adalı:

We have heard about MICS plus or the MICS Plus real-time monitoring survey. We read on the MICS website that this project was launched in 2018, and it involves phone interviews. Can you tell us about this initiative, please?

Attila Hancioğlu:

That is something that I am personally and quite intensely involved with. Let me go back a couple of years. We started thinking about the project and putting in place, and doing some activities about that in 2018. The idea goes way before 2018. In 2013 or 2014, as far back as that, we started to think about how we can use phones in one way or another to collect data. I will come to that in a minute.

We did not go that far, just when we were putting in place a more systematic idea of doing this. Then we realized that the World Bank had been doing similar surveys in many countries. They were mostly piloted in Latin American countries. Then they moved to a few African countries. I think they were doing two surveys in Uzbekistan and Tajikistan with a methodology similar to what we have in MICS Plus now, which is basically as follows.

This is the World Bank version of it. Commission of a survey, a household survey, collect background information, phone numbers, and ask for consent. Then immediately after the household surveys are completed, start calling a sub-sample or all households interviewed during the household survey for an extended period. For Uzbekistan and Tajikistan, this has been going on at least for two years. So, we realized that it was happening and developed a dialogue with the World Bank. We tried to learn as much from them and put MICS plus in place.

The first step was back in 2019. I think it was early 2019. We started a collaboration with the Belizean Country Office of UNICEF and the National Statistical Office. We started a pilot MICS Plus in Belize. We first did a household survey in two districts of Belize. There was a pilot that was not national, two pilots in two districts. Then, we included all of those households who gave us consents and phone numbers into the MICS Plus. We worked with a private vendor for the call center's setup, NSO trained interviewers, we started the calls.

Interestingly, there is just something slightly different from the World Bank's main objectives from the beginning, I would say. One of the objectives of MICS Plus was emergency monitoring. At the same time, we were looking for countries prone to emergencies like hurricanes and floods and that kind of thing. We thought that MICS Plus could be a great vehicle if we had it placed running, and when an emergency comes in, you have a representative sample of phone numbers that you can call. You can get an immediate response from the population to the emergency. That was one of our objectives from the beginning.

Then COVID came in March. I think it was in April, if not May, that the government of Belize declared a state of emergency, and several restrictions were put in place. We immediately reacted with MICS plus and started to collect data on COVID-19, including attitudes, behavior, etcetera. It was an unexpected gift, in a way, to MICS Plus. Then, we finished collecting COVID-19 related data several months ago with 11 waves of calls. Eleven monthly sets of calls. Because COVID was in, restrictions were in; we started thinking of doing it in other countries. Right now, we have it running in Mongolia. We have finished the first round and about to start the second round of calls. In Georgia, we have just finished piloting the main questionnaire for the first wave. We have the sample selected and everything else, and we are about to start the first wave of calls.

Let me go back to Mongolia and Georgia because we have differences there. The template of the methodology is as follows. You have to find a representative household frame and select, which could be a MICS survey. That was our original intention. Alternatively, that could be any other representative frame of households. For instance, in Mongolia's case, the frame we are working with is the population registry.

Interestingly, in Mongolia, the population registry, which was updated in February this year as part of the census, includes phone numbers. However, the MICS in Mongolia was from 2018. When we assessed, we concluded that it would be good to go with the population registry because its frame was updated. We started doing the MICS Plus based on the population registry selecting a sub-sample, and we started calling the households. In Georgia's case, the sample frame is the MICS from 2018, during which they had collected phone numbers. We have selected a sub-sample of households from the 2018 MICS sample frame, and then that is what we will be doing.

The idea is that, as I was saying, the template starts with a representative sample of households, continues with calling them over an extended period, giving yourself the chance to change the questionnaires over time, and come up with longitudinal data based on the same households. You do repeated calls from the same households every month.

It opens up lots of opportunities. One is that your reaction time to what has been happening is fast, like in Belize that I gave as an example. COVID-19 came in, and we were able to start collecting data immediately, already having the household sample frame. Second, you come up with longitudinal data with this approach. Third, you can look at the seasonality of different indicators that cross-sectional household surveys usually suffer from. Therefore, there are many different advantages to this approach.

Call centers do the whole thing. It is not Web-based and any other methodologies. Call centers just call to the households by interviewers. That is basically what the methodology is. Of course, there are lots of complex details to it. But, it may not be appropriate to go into at this point.

Sinan Türkyılmaz:

In countries where phone numbers have been collected or available from registers, what is this data's nature? Are they usually landline or mobile? If mobile phone information is provided by respondents, to which household members do the phone numbers belong? Is it usually the number for the head of household, the household respondent, or someone else? We would like to learn some details about this.

Attila Hancioğlu:

Nowadays, it is predominantly mobile phones. I think landline phones are a thing of the past in almost all countries. For instance, in many African countries, the landline concept is not there at all, but mobile phones are very common. Many countries have above 90% ownership of mobile phones, or at least one household member. There are still countries who are at the 50-60% mark, and we know this from statistics, of course, including several MICS surveys in which we do not only know the extent of ownership or mobile phone use. We also know which groups are using and not using. Methodologically, that is very useful.

So, it is predominantly mobile phones. Everybody has mobile phones now. The extent and use of mobile phones are just extraordinary. Moreover, I think it has grown exponentially in the last few years. That is something that we have been taking advantage of; not only us but also everybody else working in the data collection field.

Let me give you an example from there. In the MICS, we collect information on phone ownership and use in three different ways. One in the household questionnaire that we have, we asked if there is at least one household member who has a mobile phone. Then, when we move to the individual questionnaires -the women's and men's questionnaires- that are for 15 to 49 years old males and females; we asked about the individual use and ownership. One of the things that we have been detecting is that ownership and use are two different things, especially for women. You ask whether they own a phone, you would get a percentage, which would be significantly lower than they use. You ask a question whether they used them regularly, and they say yes. That is because most of the phones are owned by men or other people in the household. I do not think we have specific information on who owns or uses phones, at least from the MICS point of view. But we do have it from the individual questionnaires in terms of the usage, which is not for all the household members.

When it comes to collecting phone numbers, the standard practice is to collect as many phone numbers from households as possible, and I think it should be that way because there are methodological reasons for it. Also, many changes happened over time with mobile phone numbers, the area codes that vendors provide, the phone numbering system, and this kind of things to cope with those challenges. I think it is a good idea to collect more than one phone number per household.

Concerning your question, one thing is that in Belize, we had a split sample and tested two approaches. One was to see over time what our attrition rates would be if we were to call or target any adult household member as a respondent? Also, we targeted specific individuals in households. In the case of MICS in Belize, it was women who are mothers of under-fives. We have shown through this pilot that targeting a specific individual would significantly increase attrition rates. Dropouts are much more common in finding people over successive months. As you would imagine, finding the same person is much more challenging than trying to go for any knowledgeable adult household member. Therefore, it is a good idea to connect as many phones as possible for making those calls to the households to get that individual, an adult household member who can respond to the questions.

We have abandoned the idea in Georgia and Mongolia to do individual questionnaires as we do in MICS, like women's questionnaires that you guys have been doing in the DHS, for instance. We do not do that in Mongolia and Georgia. We have multiple phone numbers, calling and trying to get an adult

household member who can give us proxy information on everybody else in the household, which impacts what you can ask, what you can hope to ask, and get good answers for better data quality.

That brings another dimension of thinking into the content of what you do, the questions that you would ask. As you know, several questions, several issues, and topics you would only get from the individual. Let me give you an example from your DHS surveys. Asking contraceptive questions of contraception to men about their wives or a household head to a woman, about a woman in the household would be unthinkable. Once you target any adult household member as your respondent, those kinds of topics are completely out of the question.

Yaser Koyuncu:

Speaking of the data collection with the phone, how are the levels of providing phone number information by households? Would you expect any coverage bias due to this second round of frame preparation? We assume the first frame is the address frame, which is open to some degree of coverage error.

Attila Hancioğlu:

I can probably talk more about the experience that we have had so far, I think, in both cases, especially in Mongolia, for which we have complete data now because we have finished the first round of calls and have gone well into the process, households do not seem to reluctant at all to give phone numbers or accept our calls. Remember, in Mongolia's case, and the phone numbers are coming from the registry to the National Statistical Office, who are the ones making the calls.

I think the refusal rates are extremely low. I know several examples from other survey programs that the households in developing countries and the most middle-income countries do not hesitate to provide phone numbers to the interviewers during household surveys. No doubt, there are consent issues there. You have to approach the household or respondent in the right way. Ideally, you have to tell them exactly why you collect that information. If you do those rights in most countries, I do not think it is a big problem.

Tuğba Adalı:

As far as I understand, what you mentioned was the low level of not providing phone information. So, everyone pretty much provides this information. How about the households that do not have any mobile phones? You mentioned you could see how different these households were, because you already have MICS data from them. How about the differences between those that do not have phone information, that do not have phones actually, versus the other ones?

Attila Hancioğlu:

It goes back to your question, which I did not respond to the second part about coverage bias. Let us think about this. If households with mobile phones are quite okay to provide the phone numbers and participate in these phone surveys, your coverage bias will have to come mostly from those households you cannot access because they do not have mobile phones or landlines. It can be a big problem in some countries where mobile phone use is at low levels. For instance, in Madagascar, when we did the MICS two years ago, the mobile phone use ownership was about 49%. That is half of the country. In Sierra Leone, it was around 60%. Several countries had 70-80 coverage rates, but most of the countries have about 90%.

You have to be careful about what you are collecting and how you analyze and present the results. Several methods have been developed over time, which try to allow for correction or compensation for possible coverage biases. Some organizations calculate propensity scores to come up with correction factors or sample weights accordingly. For instance, we have been trying to compensate by doing a Euclidean distance analysis before data collection and trying to pick those households for substitution in Mongolia. Those are identical to those households that we have not been able to access, not only because they do not have mobile phone numbers but also for other reasons. That might be repeated calls to the household, which have not been answered. After that, we have tried to come up with a substitute, on a conditional basis; conditionality being the household, which is the most identical one.

In the literature, I see a surge in phone surveys conducted in the last 6-7 months. I did not realize this until we started collecting this data from UNICEF's country offices. UNICEF alone has done more than 500 worldwide phone surveys since April. It is many more. For instance, the World Bank is active in about 100 countries doing phone surveys. There are also private vendors, research organizations, and others who have been doing these phone surveys.

I see much variability in how the results have been presented, how low response plus possible bias is dealt with, how sample weights are calculated, and the language the results are presented. For example, I have seen phone surveys, including Random Digit Dialing surveys based on simple telephone number lists from a private vendor or household survey frames. Unfortunately, quite a few examples have been published based on the 50% of the population who provided phone number or what they could access with absolutely no warning or language about what the data represents. With the exception of saying, we have found that the sex of the household head is very similar to what we had in Australia three years ago, or the educational distribution, the regional distribution is quite similar. You know, trying to validate as if the survey results are really good enough and representative by doing that kind of simple comparisons.

I have seen quite a few examples in which calibration methods are being used by coming up with expansion or correction factors for different socioeconomic groups, doing it univariately, not by calculating propensity scores or anything sophisticated and robust but with a very simple analysis of what the sample would represent if it were to be expanded to the population's structure. Further, some studies are done quite professionally and methodologically correctly.

If I may, I would like to open a parenthesis here and talk about something else because what I just explained reminded me of it. With the surge of using the telephone for surveys, we have a little bit of a terminological confusion right now. We are using the term phone surveys to refer to quite different things as for MICS plus surveys, the COVID impact surveys World Bank conducts, which is similar to our methodology, as well as for surveys in Tajikistan and Uzbekistan. What is happening there in these surveys is that the phones are used as the medium of administration of questionnaires. Thus, we are talking about the phone as a tool that replaces paper or tablets to collect information. We generically call these surveys phone surveys. These are all based on representative household frames, just like a household survey would be. We also have other surveys like the Random Digit Dialing surveys, where you randomly generate phone numbers and call households. Your sample frame is a random list of phone numbers based on the country's system or phone numbers. There are other types of surveys based on phone numbers acquired from private companies, vendors, where you select a sample from and call those households. Now all of these examples are generically referred to as phone surveys, which is creating confusion. That leads to another confusion: the term survey is used sometimes to refer to household surveys only and sometimes to all kinds of population-based surveys, which might be the type of phone surveys that I gave examples of, like the Random Digit Dialing surveys. I think we

should be very clear in using the word phone, referring to a data collection, whether it is being used as a tool or constituting the starting point or the sample frame of data collection. Those are two very different methodologies. Nevertheless, we seem to be encapsulating them into one term, which I find a little problematic.

Tuğba Adalı:

Thank you for the clarification on terminology. My next question has been partially answered by your discussion earlier; I was going to ask about the data collection tool and how it was compromised using telephones when you started MICS Plus. You already mentioned some questions could not be asked through proxy respondents, such as contraceptive use. In MICS Plus, could we say that household interviews are happening rather than individual interviews? Can you please elaborate on this a little bit?

Attila Hancioğlu:

There are others who do individual interviews. I can only talk about MICS Plus for the two exercises that we have now in Mongolia and Georgia. In both cases, the questionnaire is administered to any knowledgeable household member. We are not targeting individuals, but we target individuals to collect information about. In Mongolia, for instance, one of the target groups was children 2 to 17 since we were mostly interested in education, in distance learning, which has been put in place in Mongolia. We randomly select one child per household, who is 2 to 17, and then ask the adult household member who is respondent on that child. Individual-level information is collected this way.

We have to make the distinction there. One is about talking to an individual and trying to get information from that individual about themselves. The other is talking to a proxy respondent who may or may not, depending on your questionnaire, give you information on an individual or individuals in the household. It always happens in household surveys, as you know. If your starting point is the household questionnaire, to say the least, you will collect information on demographics when you are constructing the household list from a single respondent, as is the case for regular MICS surveys or DHS surveys. You first construct the household list from which you would select the eligible people individuals for individual questionnaires.

For Mongolia and Georgia, yes, we are targeting the adult member. However, we have been asking questions about individuals in the household. I gave the example of contraception, but many others would not be suitable for phone interviews. There is now abundant literature and recommendations and guidelines out there. For instance, when it comes to topics like domestic violence and sexual abuse, it is quite difficult, if not unethical, to ask over phones.

When you have these structured questionnaires, you have plenty of time to ask questions and get to an indicator that you would create from a set of questions. You would ask several questions and probe questions, too. You would also use different parts of the questionnaire to contribute to the construction of an indicator or tabulation.

Let us remember that household survey questionnaires tend to be much longer than we can hope to have in phone interviews. I would not say standard because there is much variability in this truth. But I think most experts would agree that you would stretch it a little bit too far if you were to go beyond 20 minutes for phone interviews. I checked the first wave of calls in Mongolia; we completed the interviews 15.6 minutes on average, ranging from 12 points in the capital city to 18 minutes in one of the other regions. Therefore, it is more or less the ballpark for what you should be operating in. When it comes to the household survey, the individual questions can be 40 minutes, up to an hour in some

cases, depending on the household survey program. Phone interviewing will limit what you can ask and how many probe questions you can put in within 15 minutes. You would always be very eager to collect useful and complete information on a topic or an indicator when you have a phone survey. Then you would be economizing on the number of questions that you would ask. That would also bring in some limitations to types of topics for phone interviews.

The longitudinal MICS Plus type of surveys gives us another opportunity: we do not need to collect all the information that we want in a single call, in a single wave. In Mongolia, the first wave of calls was mostly focused, I would say, on distance learning and education. The second wave of calls will mostly focus on the nutrition and diets of both women and children. We will alternate. Therefore, you create more space for a topic like distance learning in one wave of calls by not asking questions on another topic but leaving the topic to the next wave of calls. If, of course, your plan permits it to do it that way.

Sinan Türkyılmaz:

Do you expect or did you observe or test any measurement differences due to the mode difference or interviewer involvement level? It might be because of the lack of non-verbal feedback or some rapport issues between respondents and interviewers.

Attila Hancıoğlu:

Good question. In some ways, I think it is a million-dollar question, too. You can find evidence for both arguments by looking at the literature on this. Some say that there is good evidence that a phone interview would probably give you comparable results for a specific set of indicators to what you would get in a household survey if you do things right. But there are others in the literature who say just the opposite, that you should not expect to get the same kind of comparable data from a phone interview.

My background is mostly in household surveys. Looking it from household surveys point of view and perspective, we know that these are standard for household surveys especially when we train people: talking about the methodology, face-to-face interaction, rapport that the interviewer builds with respondents usually being in a separate place without anybody else's involvement or distraction, structuring a questionnaire from a longish questionnaire, going from the general stuff to more detailed, building rapport along the way, and being able to go into more sensitive topics using the opportunity of being on 1 to 1 and face-to-face basis. We believe that these create good quality data. Now that is completely lost on phone interviews. But that does not necessarily mean that you cannot hope to get good, accurate information on a phone interview.

In my view, saying that there are two ways of thought in this, either they are comparable or not comparable, is a little bit far simplistic. I think it depends on the type of information that you are after. For some questions, it will not work. You need that face-to-face interaction for some questions; we know from household surveys that you just need it. But there are other things that maybe we are not doing as well in household surveys as you would do in a phone survey. And that might be not because of the type of questions, but because of other factors. There might be better chances on the respondent giving a better quality answer on the phone if you are creating responding fatigue in a household survey, when you are interviewing people for 50 minutes or an hour, as opposed to a simple phone call where you ask a straightforward factual question to the respondent, and it is known that the whole interview is going to be 10-15 minutes. I would be reluctant to say "No, absolutely, they are not comparable" or "Absolutely, they are comparable." I think it depends on many factors, including the context, content, and design of the questionnaires.

Yaser Koyuncu:

Related to that, are dropouts common for telephone interviews in developing countries concerning your MICS+ experience?

Attila Hancioğlu:

In Belize, we had relatively high dropout rates. The biggest dropout in these types of surveys seems to be happening when you go from the household survey itself, the phone numbers, or the frame, to the first call. We have shown it ourselves inadvertently in Belize. Also, the World Bank surveys have shown this evidence that the longer the time between you collect the phone numbers and the time you start the phone interviews, the higher the dropout rates. So, the biggest dropout happens the first time before you have the first call. You may not be able to find them; they might refuse or might have forgotten that they gave you consent before. They might not have understood why you were collecting the phone numbers from them. They might change their mind, etc.

The second biggest chunk of dropouts seems to be happening between the first call and the second call. From there on, I think the dropout rates are usually lower. At least, that is what we are learning from literature. In Belize, we did something inadvertently, a mistake, which was that, we had a very long time, four months or so between the time that we did the baseline survey, where we collected the phone numbers and the first phone calls. It was because of logistic reasons, contracts, or the vendor, et cetera. That cost us; let us put it that way.

We have several households that did not participate or dropped out very quickly. We will see what happens in the case of Mongolia and Georgia. I am very hopeful. I mean, we have got a response rate in excess of 83-84% without substitution in Mongolia, which is an incredible number for phone surveys. You see, I am also making the mistake of calling "phone survey" without qualifying.

Moreover, when we substituted, we went up to 95-96%, which is huge as well. Based on these, I am hoping that the dropout rates are going to be quite low. But you have to do things right, just like household surveys. One of my mottos in household surveys was when I used to train interviewers in the old days; I would always say that households' refusals in developing countries are based on two factors: one, the refusal of the households itself. But two: because of the interviewer. That involves how you approach the household, knocked on the door, spoke to the person, introduced yourself. Those are significant factors that may increase refusal rates. I think the same applies for phone calls to interview. It depends on how you are approaching people. If you are challenging people with putting in place 30 or 45 minutes phone call to the first round, people will think, "Uh, not really. I spent half hour with this person. This is going to go on for a year. That is too long. I do not want to spend a half-hour or 45 minutes with these challenging questions answering these people all the time" for the second round. You might cause an increase in the dropout rates. So, it depends on you as well. But we hope that in Mongolia and Georgia, we would have a stable, participating, good set of households without high dropout rates. That is what we are hoping. We will see.

Tuğba Adalı:

We are now moving towards questions regarding the future. I wonder what MICS plans for countries that had MICS surveys before but did not collect any phone information. For instance, you mentioned Mongolia having registration data for phone numbers. Are there any such plans to find alternative frames for phone numbers or anything else?

Attila Hancioğlu:

One of the decisions that we have taken that COVID has taught us is collecting phone numbers in all the MICS surveys from now on. Although we as the program had not been recommending the

collection of phone numbers in our standard questionnaires or guidelines. Our counterparts were the national statistical offices; in many cases, they collected this information themselves but not sharing it with us since it is confidential information. Mongolia MICS 2018 also had phone numbers. But we had nothing to do with those phone numbers ourselves as the MICS program.

I think there will be much more reliance on mixed methods collecting data through phones, Web-based interviews, face-to-face, etcetera. It is not only because of COVID-19. I think we were going that way as the world, but COVID-19 seems to have accelerated it. For instance, UN agencies were not systematically doing or commissioning phone surveys or partnering with statistical offices for phone surveys. COVID-19 has just caused us and most UN agencies to be very much thinking along the lines of phone surveys and big data, Web-based interviewing; those kinds of new types of data have gained prominence in the data collection field.

Going forward, I think we are all going to be considering more about mixed methods. Is this new thinking? No, it is not. It precedes the COVID. With the so-called data revolution coming in 2014-15, we started talking more about data ecosystems that no single data source will come up with all the answers needed. Household surveys will always have the weakness of not producing high-resolution data. Qualitative data will never be representative. Big data has its weaknesses and strengths. Administrative data, when it is complete, will be a great source of information. However, it will have lots of weaknesses regarding disaggregation, showing diversity, or the individuals' background characteristics in that system. Censuses will come every ten years. The census frequency is never going to be good enough when the census is face to face, of course.

We had the recognition for many years that we have to rely on more than one data source. We cannot solve the world's problems and understand the world by using only one type of data source or a few data sources. I think it is something that COVID is interestingly helping us to approach much faster than we did to using mixed methods, either within a single data collection effort or multiple data collection exercises or approaches to build the story altogether and create an ecosystem so that there is a more comprehensive understanding on what is happening. For instance, UNICEF's affinity to phone surveys have exponential increases, as I was saying earlier. It will be much less difficult now to get UNICEF into partnerships with others, including the national statistical offices to collaborate not only for household surveys or phone surveys but also other types of data collection efforts.

But we have to make this distinction. We are talking about triangulation here to some extent; we are talking about opening our horizon and trying to use different types of data with different types of robustness, representativeness, and quality. We are talking about using different data sources along those dimensions and putting them all together to build the story that can happen as a mixed-method data collection effort. Within the same survey, you might be depending on both face-to-face and phone interviews. You might be recruiting your respondents by face-to-face interviews but continue to collect data from them on the Web. There is going to be all kinds of these data collection efforts.

We have a separate initiative now in MICS we are about to embark on, called MICS Link. From the outset, it is to construct a MICS survey triangulated with another data source. We have been having negotiations with two countries to conduct a MICS survey that will collect ID information to link children going to school in these two countries to the educational system. That way, finishing the survey, National Statistical Office can access the admin system; get information from them, and merge it onto the MICS data in terms of the type of information not available from the MICS.

Another example is the use of geospatial data. Your geocode, your household survey data and suddenly horizon opens up. You can come up based on the geocodes of a cluster of a household. You can come up with lots of covariates of a climate that have to do with pollution. It is a type of triangulation or

merging data, which we intend to do in MICS Link. I think merging different types of data is really close to my heart. I also think that is something that we are going to be putting more and more emphasis on the MICS program.

Sinan Türkyılmaz:

We know that MICS collects anthropometric data from children which provide key indicators of childhood malnutrition. What would you recommend for the collection of such data under COVID conditions? Do you think the flow of this data will stop until the end of the pandemic, or will there be alternatives?

Attila Hancıoğlu:

It is going on, actually. We had anthropometry in the two surveys that we did in both Malawi and Turks and Caicos, as we have almost in all MICS surveys. Maybe there is a couple of exceptions here and there. We did continue to collect anthropometric data in those surveys, and there is no strong evidence that we had resistance from the households. However, this is June-July. Now the population see COVID-19 as a different thing. The second wave is probably coming and indeed, has come to several countries. I think populations are much more aware of COVID-19 and the risks involved.

I think you are absolutely right in pointing out to the specific topic of anthropometry. That has been a challenge for us. We are about to develop guidelines on safety recommendations and precautions that should be put in place. No doubt, the ultimate decision is always with the government who do these surveys. But we are coming up with a series of recommendations on this, on how to deal with face-to-face interaction, what to do in the venue where you are providing training, how to use PPE, how and when to use masks, what happens if somebody is Covid-positive, including all those kinds of recommendations. There is an ongoing dialogue among different survey programs on this as well.

Specifically to your question, from the outset when we started thinking about what is going to happen in surveys when we go back to face to face, we started thinking about anthropology in MICS because that is one of the two times that we have to be very close, if not interacting physically, with the respondent or the household member. One is anthropometric, and the other is when we apply the foundational learning skills module where the interview has to sit next to the child, exchange material where we are testing the child's numeracy and literacy skills. Those are two key points for us.

For anthropometry, I would say that the actual measurement during data collection is less of a problem because certain precautions can be put in place. As you know, we do validation during training; we use children measured by different measures when the child goes from one station to the other or vice versa. That is proving to be more of a challenge that we have been putting our thinking into. We have not really resolved it ourselves. There has been some thinking along those lines with regard to the extent of this validation and use dolls instead of real children during training. But one piece of news I want to give you is a survey program called Smart Surveys. They worked very closely with the CDC, the Center for Disease Control. They have come up with a set of recommendations of the guideline on how to do nutrition surveys. And there is plenty of information on how to deal with these things in that publication, which has just come out, I think. We are going to be incorporating some of those recommendations into ours. However, we have to come up with specific recommendations as they apply to the content of MICS.

Yaser Koyuncu:

From a global point of view, what type of survey errors will be most discussed in survey literature from today onwards with the abandonment of face-to-face surveys? I mean, with the at least temporarily abandonment of face-to-face surveys.

Attila Hancioğlu:

I think I have a more optimistic view of what is going to happen than most. There was a survey back in August by the World Bank and the Statistics Division of the UN. It was a survey of all the national statistics offices around the world. One of the questions to the statistical offices was, "When do you expect to go back to face to face data collection?". In that survey, close to half of the statistical offices said within the next six months. The remaining 50% were saying that they were uncertain. That was August. Now we are in November. The second wave has come in. Certain parts of the world have not been as affected as others. And things are going to develop. There is the vaccine on the horizon. So, there is still much uncertainty.

In the case of MICS, for instance, we have a few surveys which are about to go to the field. We have quite a few surveys that are planned for January, February. I am of the view that we will do just like in our normal daily social lives; we have adopted changes on certain behavior but still are trying to cope with the challenges and continue life as much as possible as usual. Once this is over, I think we will go back to what will be the new normal. What will be the new normal? How will the households react to face to face interviewing? Is this going to influence response and completion rates in countries? That remains to be seen. I do not think anyone has a very clear idea of what will happen there, but I think there is much more reliance on types of data that are not going to be necessarily representative in the way that the statistical establishment has understood it always. Quite rightly, the statistical establishment has been very much fixated on representativeness, probabilities of selection, full completion, full coverage, avoiding the bias, at least intentional bias, et cetera. But the new data sources will be different. A phone survey based on RDD will say that "Well, 30% of this country does not have phone numbers. So, we are doing this on the basis of the 70%". There will be much more participatory data collection like UNICEF's many projects in different countries where there is intentional participation on the part of the respondent. It is not representative, but it is very useful and valuable data.

I think that is what we will be increasingly looking into, including mixed methods, triangulated data, increasing use of technology in all kinds of data collection, and digitalization of almost everything. That is going to accelerate probably, from now on. So, those are the things which are coming up, I think, among many others.

Tuğba Adalı:

We have already heard your views about the data ecosystem and digitalization. But I am going to go ahead and ask, just in case you have anything else to add. The use of non-survey data, big data, or organic data has been increasing rapidly over the years, such as Google search data, Twitter hashtags, Google mobility data, and alike. With the pandemic, we see different ways such data is exploited to shed light on human behavior. How do you think studies of this kind will affect survey data collection in the future? Will it serve as a replacement to surveys, will it compliment them, or are there any specific pitfalls you would like to underline?

Attila Hancioğlu:

At least for the time being and the foreseeable future, I do not think we are. I am approaching this from a methodological perspective. I am not talking about practice. The practice is something else.

From a methodological perspective, I do not think those initiatives themselves are even talking about replacing traditional data collection. Nobody is saying that right now. Nobody is claiming that. Everybody is saying that these fit into the different places in the data ecosystem. They complement each other rather than replacing each other. You probably know that much of the analysis, which is produced based on Facebook data or Google data, remote sensing, satellite imagery, et cetera; use population census, household survey or admin data to a large extent, to cover up with models for making their data and their results more robust. There is a dependency on each other. That is one of the many reasons why these are going to remain to be complimentary.

But that is the methodological part of it. The other part of it is practice. The other approach to this could be looking at it from the practice point of view, whether the cheaper alternative ways of collecting data, which are not necessarily as robust because of resource limitations, are going to be causing fewer resources to go more expensive but more robust data collection efforts such as household surveys or censuses. Whether that is going to happen remains to be seen. I do not think it would be a good idea if that happens. I hope it does not. However, there is an attractiveness to the new types of data. They are faster and cheaper. They are not as robust, which is fine.

If resources are mostly channeled to those types of data only, and as a result, other types of traditional robust data collection instruments suffer from it, there is going to be something close to replacement, maybe. I hope it does not go that way because it is not methodologically justified at all. However, about 30 years from now, will we be doing face to face interviews in the household surveys in the same way that we are doing? Well, almost certainly not. If we were having this discussion ten years ago, 15 years ago, we would be talking about 20%, 25% of populations having phones. Doing anything on the phone, any kind of interaction with respondents on the phone, would be unthinkable. We are now discussing whether the RDD surveys or how RDD survey can be methodologically made to be representative. So, thirty years from now, I think we will be looking at a very different world. We moved from paper to CAPI, as you know. In the case of MICS, we do not have paper-based questionnaires anymore. All our surveys are tablet-based, and we moved into tablets within a couple of years. We took the decision, and within a couple of years, all our surveys were based on tablets. Maybe 30 years from now, we are not going to be thinking about tablets at all. Maybe phones are going to be the standard. Maybe Web is going to be the standard. I think it is essential to concentrate on today and innovate as much as possible and see what happens rather than fantasizing about 30 years from now. Because in terms of communication and IT technologies, the world is changing so fast. It is quite difficult to predict what is going to happen 30 years from now.

Sinan Türkyılmaz:

We came to our last question. Indeed, you already answered my question with regards to Yaser's and Tuğba's questions. You may have some last word to say on it. This is a common question that we are asking all experts that we have interviewed. Let us assume that the COVID-19 is eradicated overnight. Do you think surveys can go back to pre-COVID era instantly? What are the challenges we should be expecting after the pandemic is eradicated?

Attila Hancıoğlu:

Your question started with a hypothetical or science fiction assumption, which is that COVID-19 will be eradicated overnight. It is not going to happen. But let us think of a scenario when COVID-19 will be no more there, and the vaccines are widespread. It is going to be rolled out everywhere. Let us think of that situation. Let us think of a situation where we have a new normal. I think we are all accepting worldwide that once COVID-19 is over, it will not be back to normal; it is going to be new normal. The

new normal is going to be different from the previous normal in every aspect of life. Now, coming to household surveys. We are already making changes. As I mentioned earlier, because of COVID-19, we have decided that we should collect phone number information in all the surveys. We should be prepared for that kind of thing. We should be collecting phone numbers in preparation for MICS Plus surveys in all the countries, too.

As a household survey program, we have been thinking, about collecting data more frequently. One way that we have committed ourselves to do so is MICS Plus. It is not going to be implemented everywhere. But in countries interested in this, we will be providing an opportunity to countries to collect this data in post-COVID. One of the pessimistic views, of course, as you know, is that this is not the first and the last pandemic. There will be similar pandemics or epidemics later, so we have to be prepared for those kinds of eventualities. We have to think of mixed methods. We have to be more cognizant that we have to rely on different types of data in different types of circumstances, and each of them will have their weaknesses and strengths. But each of them will complement each other. So, we have to have a wider look at what is going to happen and how we will be collecting data.

Are household surveys going to change? I think so. We do not know what the households' behavior will be once the COVID-19 is over. Are we going to be suffering from much lower completion rates? Much lower response rates, is that what is going to happen? If households are more reluctant to open their doors to strangers remains to be seen. Most probably, there will be elements of that. I do think that in our daily lives, there are going to be behavioral changes as a result of this. I think we have been fast educated during this pandemic on several things that have to do with face-to-face and human interaction, which leaves its traces on us. So, household surveys will change as well. We should try to keep the robustness of household surveys in the future because they are the golden standard when it comes to many types of data that can be collected, including the socioeconomic disaggregation dimension. It enables us to come up with factual, attitudinal data, and behavioral data, all at the same time. These are the great strengths of a household survey. But there are weaknesses, of course. We have to keep those strengths in place and move forward.

Also, I have no doubt that we are going to be more and more technologically minded. One of the things that COVID-19 has taught all of the survey programs, us included, is that we have to rely on and use more technological gadgets and possibilities and try not to be as a perfectionist and rigid our thinking as possible. That is really important too.

5. Conclusion

In this report, we started with a background section on surveys. This chapter has shown us that over time, there has been a proliferation of modes and data sources. The developments in computer technologies have been affecting surveys for a long time, starting with data entry and data storage technologies and continuing with data collection through CAPI. The evolutions of laptops, tablets, and smartphones, combined with the spread of the internet, have added many dimensions to survey data collection in the last two decades. These devices and connection technologies are also producing some organic data, enabling researchers to study certain human behaviors without collecting survey data.

In our next chapter, we took the COVID-19 pandemic as our core issue and tried to bring together some guidelines produced on household data collection during the pandemic and some empirical examples from around the world. We mentioned how some surveys were impacted by the pandemic, making distinctions between cross-sectional and longitudinal surveys and considering former dependence on face-to-face modes. We also mentioned some surveys that relate COVID-19 with other issues and some prevalence surveys.

In our fourth chapter, we presented six interviews that we conducted with experts from the field. The recordings of these interviews are available on our Institute's YouTube channel. Below we sum up our conclusions with references to the interviews⁶⁸ and our desktop review. We should underline that these conclusions are not independent and relate to one another in many cases.

- *The effect of the pandemic on surveys depends on the country*

For some countries, the best frame available is an address frame, and face-to-face is the predominant mode, especially for national surveys. For these countries, cancellation or postponement is a bigger risk. The United Nations Intersecretariat Working Group on Household Surveys (ISWGHS) Task Force survey on national statistics offices showed that such actions were taken more often in developing regions than developed ones. The FAO also underlined that low-income countries were prone to the risk of a blackout of national statistics earlier in the pandemic. James Lepkowski indicated in the interview that cancellation or postponement was a major problem for data used for policy purposes. Moreover, in the long run, we can infer that there will be less information available from these countries regarding the effects of the pandemic on society.

For countries that already have established CATI and CAWI surveys, fewer surveys are prone to such risks. International Labor Organization's brief from May showed that the proportion of countries relying on CATI for labor force surveys was much higher in Europe and EFTA countries than the World average, where CAPI and PAPI were the main modes. For instance, Raphael Nishimura, in our interview, highlighted that the official statistics institute in Brazil did not have prior experience with telephone modes in household surveys; whereas Torbjörn Sjöström, working for a private company in Sweden, described their pandemic state as "business as usual" at the WAPOR June webinar. The Turkish Statistical Institute was just on time, since they initiated CATI at the beginning of 2020, as Enver Taştı explained. As Oğuzhan Akyıldırım underlined, data collection in Turkey has traditionally been the reverse of contemporary trends, with high reliance on face-to-face, and no history of mail surveys. This holds for other countries as well. As Raphael Nishimura underlined, mail surveys were never popular

⁶⁸ We used "..." within quotations whenever we removed some sentences for conciseness. Full transcriptions are presented in Chapter 4.

in Brazil: “Some are still concerned that Brazil's literacy rates are not high enough to justify a mail survey even though they are pretty high. In addition, the postal service in Brazil is not that great and varies a lot throughout the country. ... Historically, people are not used to answering surveys by mail, as they are in the US”.

Basically, we can say that the effects of the pandemic on surveys differ by country because the countries have different infrastructures and baselines.

- *There have been certain remedies taken to continue data collection with surveys*

In some cases, surveys have been *anceled*, such as skipping a wave. However, this did not seem to be very common according to the ISWGHS for national surveys. In other cases, we saw that surveys were *postponed*. There were many cases where survey operations were postponed to later dates in 2020, and postponing to 2021 was less common, according to the ISWGHS. Some countries postponed DHS-type surveys to 2021, which were initially planned for 2020.

In many other cases, we saw *mode switches*. These, as we will discuss in more detail below, were from FTF to another mode. The most common switch was made to *CATI* for surveys ran by national statistics offices, according to ISWGHS. Among examples, in Chapter 3, we saw SHARE, Gallup World Poll for some countries, and most of the EU-Labor Force Surveys based on EUROSTAT. In our interviews, Raphael Nishimura mentioned the switch from FTF to CATI for the survey on Demographic Consequences of the Zika Epidemic, and Oğuzhan Akyıldırım mentioned their experience on an ongoing refugee study.

There were also cases where *self-administered web modes* were introduced instead. For the British Social Attitudes Survey, the main mode of CAPI was switched to push-to-web. The General Social Survey, designed as mainly FTF for 2020, was also switched to a push-to-web mode. We did not come across examples of switching entirely to a postal survey, but rather we saw postal survey implementations as part of mixed-mode designs. For instance, although the main switch of the Health and Retirement Survey was from enhanced FTF to a combination of telephone and mail, a mail survey has been planned for 2021, as David Weir mentioned in the PAA webinar (Webinar - COVID 19 - Data Collection in Longitudinal Studies, 2020).

We also saw mixed-mode surveys designed to replace former data collection modes. Constanza Cilley mentioned in the WAPOR June webinar that mastering the mixed-mode was one of Argentina's challenges in this new period. The surveys that already had mixed-mode designs with a FTF component were often altered to other mixed-mode designs. The German Family Panel, for instance, switched from a combination of CAPI and CASI into CATI and mail. The Understanding Society, web-first with FTF follow up, would now have the follow-up by phone.

Mode switches happened in qualitative research too. In an example from Brazil, traditional in-depth interviews were carried out over the phone, and focus groups were to be conducted online (De et al., 2020). Oğuzhan Akyıldırım also mentioned how their qualitative research was affected in our interview. He underlined that they perceived focus groups to be hard to conduct online or through teleconferencing and switched to in-depth interviews instead, in one of their studies.

Rapid surveys were another proposed solution for obtaining statistics during the pandemic, with smaller samples, shorter tools, and relatively limited coverage. For instance, UN Women recommended such surveys to assess the gender-based implications of the pandemic. ILO recommended them for labor-related topics when large scale HH surveys were not possible.

Another option was the use of *alternative data sources*, such as administrative data. ISWGHS mentioned administrative data as the third alternative after phone and web surveys when a new mode or alternative data source was introduced. Enver Taştı underlined the importance of such data in our interview, saying, “The standardization and use of administrative records for statistical purposes have been discussed for years, and developed countries use administrative records very effectively.”.

Researchers highlighted certain challenges of these modifications. Losing data points in time series was one issue. Even after the mode switches were made for ensuring the timely continuation of the surveys, time series could still be affected. Gummer et al. (2020, p. 226) put this as: “In the case of mode switches, it is paramount to avoid methodological artifacts since effects due to mode changes might be misinterpreted as substantive changes across time”. This issue was also mentioned by Gillian Prior in the RSS webinar. Constanza Ciley in the WAPOR webinar mentioned losing comparability as a risk. In our interview, Frauke Kreuter explained this as: “We will see jumps in the data that we will not quite know if it is the pandemic effect or if it is about the mode shift”. Raphael Nishimura, in the interview, also said that mode effects were likely, and discussed remedies: “So, they will also have to decide. Will they ignore this period that they had to switch modes to the previous data collection mode and have a gap in their series? Alternatively, they may start a new series and continue with the new data collection mode after the switch, or they might apply statistical adjustments on the data to make the two pieces of series more comparable”. It is clear that researchers will be faced with challenges in the future when interpreting and identifying the source of the changes in estimates, whether they stem from the time or the mode switch. However, these mode changes were not optional, as James Lepkowski underlined: “But in an emergency like today, organizations are not going to have much choice.”

Interviewer training was one of the operational challenges of changing FTF to CATI. In the WAPOR May webinar, Ignacio Zuasnabar underlined that interviewers used to FTF should be trained for telephone interviewing talking about Latin America. Oğuzhan Akyıldırım also mentioned that getting a high-quality response through FTF and phone required different efforts, and said that interviewers had to be more qualified for phone interviews to succeed.

On the respondent side, Oğuzhan Akyıldırım underlined that, as an advantage, respondents could be reached at any location with the switch to CATI. However, this can also be disadvantageous for certain aspects. Reaching respondents any time any place meant could increase the possibility of contacting the respondents when they are busy or not available, and this, in turn, can affect the motivation of the respondents to participate in the survey.

- *The effect of the pandemic depends on the mode of the survey.*

All modes were and are prone to be impacted by the pandemic but in different degrees.

Face-to-face surveys were affected the most, with fieldwork no longer being possible. As James Lepkowski put it, “With COVID-19 that has restricted countries throughout the world, many agencies find this difficult to deal with because they do not want to endanger their subjects by possibly transmitting disease, nor do they want to endanger their employees who might be exposed to the disease”. Moreover, non-pharmaceutical interventions introduced by governments may not allow FTF surveys, and even if allowed, respondents may be reluctant to allow interviewers into their homes. The examples we gave earlier in this section about surveys getting canceled or postponed or subjected to mode-switches were all examples of FTF surveys or mixed-mode surveys with FTF components.

CATI was often affected when centralized CATI was used. In many cases, we saw the decentralization of CATI operations. ILO emphasized that switching to decentralized systems may propose technological challenges. Ignacio Zuasnababar, in a WAPOR webinar, mentioned that CATI surveys in Latin America were either decentralized or switched to online mode. Stephanie Chardoul mentioned in an AAPOR Webinar that they decentralized CATI in the Survey Research Operations at the University of Michigan. Yashwant Deshmukh talked about their decentralization experience in India in a WAPOR Webinar. Enver Taştı reported that TURKSTAT transitioned to decentralized CATI as well.

Enver Taştı mentioned that being able to monitor the telephone interviews simultaneously was an advantage of centralized over decentralized CATI. James Lepkowski also listed this aspect as an advantage and said: “Staff can advise interviewers on dealing with particular problems in real-time”. He also said that centralized CATI could mean higher data quality, and Raphael Nishimura had the same approach: “We have been quite successful monitoring data quality for decentralized CATI in various ways, but it is less (effective). There is not a very well-controlled environment as it would be in a centralized CATI”. James Lepkowski also underlined some potential cost savings in centralized CATI, and better and more secure transmission of data, saying, “In a decentralized system, the data is being kept on sometimes paper, sometimes laptops. And then it has to be transmitted”. Yashwant Deshmukh stated that the technical changes they had to implement translated into fewer interviews per day with decentralization. On the other hand, Enver Taştı discussed advantage of the switch to decentralization from an interviewer point of view: “I can even say that working from home is more efficient since it enables interviewers to use time more efficiently. It is possible to eliminate the time spent commuting to work.”.

Mail surveys were not as affected as the aforementioned interviewer-assisted modes. However, they were still affected due to staff working from home, for instance. Furthermore, obstacles caused by the pandemic to postal services could also affect mail surveys if respondents were reluctant to go out to send back the questionnaires. Scherpenzeel (2020, p. 218) mentioned this issue for the SHARE survey context: “Further, a switch to PAPI⁶⁹ was also seen as infeasible because in some of the SHARE countries with severe lockdowns postal delivery was (and at the time of writing still is) not possible. Moreover, collecting data in this mode would require respondents to leave their home to post the completed mail questionnaire or the interviewer to collect it from them at home. During the COVID-19 crisis, this could pose a serious risk to elderly respondents and/or interviewers”. Raphael Nishimura, talking about a push-to-web survey at the Institute for Social Research, said the following: “For that study, we had delays at the start of the survey since the staff could not go to the office to prepare the materials. We had some delays also because of how things are working with mailing these days”.

Web surveys seem to be the safest among all, although post-survey adjustments, data analysis, and report writing could be slowed down with remote working conditions of survey organizations or researchers.

Survey organizations’ operations might have been affected by the pandemic, regardless of the mode of the survey. Some of the non-fieldwork implications Attila Hancioğlu mentioned for MICS are likely to be valid for other settings as well: “For instance, workshops or data interpretation workshops and meetings, steering committee meetings, technical committee meetings, and the actual technical support we provide to the countries that happen on-site and off-site. How much of it can we do off-

⁶⁹ We believe PAPI is used to refer to a SAQ here.

site? Can we continue to support countries or data processing from afar without going to the country? ... In some countries, for instance, there were complete lockdowns for a long period. The staff members were not equipped, were not prepared to work from home to continue working on data, for instance.”

- *The effects of the pandemic depend at which stage a survey was when the pandemic hit*

The impact on a survey depends on at which stage this particular survey was when the pandemic hit. Examples from MICS and DHS showed that for countries that already completed most of their fieldwork when the pandemic hit, it could be possible to wrap up the survey. An interruption at an earlier stage though meant postponing or canceling. This could lead to time series being lost and increasing cost with potentially renewed interviewer recruitment and training.

According to the DHS update from September 2, 2020, half of the Gabon DHS was completed when lockdowns began, and it will resume when restrictions are lifted. However, it had just started in Madagascar when the pandemic reached over there, so interviewer training will be repeated, and the survey will start over in 2021. Raphael Nishimura said that the Demographic Consequences of the Zika Epidemic Survey was at a stage where the household listing was complete when they had to suspend fieldwork, and their solution was to change the mode to an RDD CATI survey.

Attila Hancioğlu explained how they proceeded with MICS surveys: “...we started an assessment to see how much fieldwork had been completed. We found out that out of those seven, five had gone all the way up to 90-95% completion of fieldwork. Just a few clusters were remaining. ... We did some work to recalculate sample weights to see the distribution of the completed and incomplete clusters. Those five surveys were considered complete at that point, and we moved them to data processing, data analysis, editing, and reporting. ... There were two surveys, which were not near completion at all. They had completed 50-60-70% of the fieldwork. It was not possible to consider it complete. That had to do with how much they had completed up to that point, which was not a good geographic distribution. There was no way that we could come up with sample weights that would compensate for that incompleteness”.

- *The effect of the pandemic depends on whether a survey is a cross-sectional or a longitudinal one.*

The impact depends on whether the survey is a cross-sectional or a longitudinal one. Household cross-sectional surveys and first waves of longitudinal surveys are more dependent on address-based frames than subsequent waves of longitudinal surveys. It is quite often the case that first wave respondents are contacted FTF, and then followed up by CATI or CAWI in later waves. Thus longitudinal surveys are less affected except for their first waves. As examples, we saw some DHS and MICS surveys being suspended, as opposed to SHARE or HRS making use of their current samples and reaching them through different modes.

Edith de Leeuw underlined this distinction at the WAPOR July webinar, saying adjustments were easier for panel designs or mixed-mode designs. Raphael Nishimura discussed this issue as follows based on his experience from the SRO at ISR: “One of the things that we had in terms of the difference between panel studies and cross-sectional studies affected by the pandemic was how we contacted the respondents. Our panel studies already have an established relationship with the panel respondents; we have several ways to contact them to conduct the surveys. We typically would have face to face interviews. But, because of the pandemic, we had to put on hold those efforts, and we were able to

transfer that mode of that data collection to another mode. ... On the other hand, we (usually) do not have other means to contact other than the main data collection mode for cross-sectional studies”.

Enver Taştı explained how they tackled the issue of contacting first wave respondents in the Household Labor Force Survey at TURKSTAT, pointing out to a new use of existing registration data in Turkey: “In a nutshell, we do not have an address frame that includes telephone numbers of all individuals in one place. We conduct the surveys by combining up-to-date telephone numbers obtained from other administrative records and the samples selected.”.

- *A variety of survey errors caused by the effects of the pandemic on surveys are to be studied*

Talking about adaptations of surveys to current conditions, many researchers discussed the types of errors to be aware of.

Coverage errors were often discussed for mode switches, in particular for cross-sectional surveys where a new mode implied the use of a new frame. Simply put, when telephone surveys replace FTF household surveys, those without telephones will not be covered. The same applies to online modes, which could only cover those with internet access. The characteristics of those not covered are very likely to be different from those who are, in terms of most social indicators, which could increase coverage error if the proportion of the uncovered population is high. For instance, Rincken et al. (2020) mentioned that mobile phone penetration was significantly lower for persons aged 65 and over in Spain, which would affect COVID-19 related information for a mobile phone survey.

Raphael Nishimura mentioned coverage error as an important source of concern, talking about errors in switching from FTF to phone, whilst underlining country-context: “The most apparent discussion on the basis of the switch from face-to-face to telephone mode is the coverage. There is much concern about coverage, which also depends on the country”. James Lepkowski underlined that telephone coverage was now higher in many countries thanks to mobile phones: “Telephone coverage has increased dramatically as cell phones, prepaid, and all sorts of other plans have come into line. ... In some countries today, you can access a larger share of the population, the household population, by telephone, than you could ten years ago because cellphones or mobile phones are more readily available and accessible than a landline phone that requires wiring to install.” Attila Hancıoğlu mentioned this distinction, and what they do in MICS Plus when collecting phone information for enhancing coverage: “For instance, in many African countries, the landline concept is not there at all, but mobile phones are very common. Many countries have above 90% ownership of mobile phones, or at least one household member.... When it comes to collecting phone numbers, the standard practice is to collect as many phone numbers from households as possible, and I think it should be that way because there are methodological reasons for it”.

Enver Taştı talked about coverage issues in the TURKSTAT context: “Yes, those can be out of the coverage in that way. ...When the pandemic was in a critical phase in March and April, there was nothing to do for households without telephone information because we could not go to the addresses. Now that the situation has returned to relatively normal, household visits are made to obtain information from those who do not have a phone and cannot be reached in other ways, within the framework of the Ministry of Health's measures.”

While coverage of telephone is often thought to be limited in comparison to FTF, James Lepkowski reminded us that there could be instances where telephone actually could offer coverage to addresses where we often cannot reach with FTF surveys, such as institutional population: “Institutional surveys

are more expensive. For example, for people living in a long-term health care facility, a nursing home, or a hospital for care of people with certain kinds of communicable diseases, the institutions are difficult to get into. ... Consider as well people who are in prison or jail. ... Jail and prison populations are usually excluded in the face-to-face household survey environment. There is a coverage problem in which we are not covering all the population". Such populations actually may be of special importance in the COVID-19 context because of the difficulty of implementing social distancing measures. He also underlines remote locations: "For example, as in a country in the Pacific region, there may be islands with a population just too expensive to get to".

Attila Hancioğlu talked about some remedies for coverage issues in our interview: "Several methods have been developed over time, which try to allow for correction or compensation for possible coverage biases. Some organizations calculate propensity scores to come up with correction factors or sample weights accordingly. For instance, we have been trying to compensate by doing a Euclidean distance analysis before data collection and trying to pick those households for substitution in Mongolia. Those are identical to those households that we have not been able to access, not only because they do not have mobile phone numbers but also for other reasons. That might be repeated calls to the household, which have not been answered. After that, we have tried to come up with a substitute, on a conditional basis; conditionality being the household, which is the most identical one".

Coverage properties of online surveys have also been discussed. Frauke Kreuter underlined how this changes by country: "The most significant source of error stems from the web-access and internet penetration that depends on the country." Munzert and Selb underlined the importance of coverage issues, especially for countries with low literacy rates, in their commentary on the Facebook survey by Kreuter et al. (Kreuter et al., 2020).

Additionally, digital literacy is an issue. Access to the internet may not translate into being able to fill out an online questionnaire. Angelo Ambitho, in her presentation at the May WAPOR Webinar, based on her experience in Africa, underlined that such modes were "limited to the educated, who have the gadgets, and understand how to navigate technology smoothly."

Nonresponse errors are likely to increase in mode switches, with the general rule that FTF almost always gives the highest response rate. This was underlined in the RSS Webinar by Gillian Prior from NatCen, that whenever a switch is made from FTF to another mode, the response rates drop, increasing the risk of bias. James Lepkowski also stated this phenomenon: "However, when they go to the telephone, the response rates fall. When calling on the phone, people are more reluctant to participate". Raphael Nishimura noted this difference between FTF and telephone while also prompting that nonresponse is an issue for virtually every mode out there: "Furthermore, non-response will never go away. That is a problem for most data collection methods, although it is more pronounced for telephone surveys with lower response rates than face-to-face". Based on the experiences with their refugee survey, Oğuzhan Akyıldırım found that levels of nonresponse were close for address-based and telephone-based fieldwork, in fact, slightly higher nonresponse at the addresses than the telephone. So whether or not response rates will fall with a switch from FTF to telephone depends on frames used, as well as the target population. Attila Hancioğlu's phone interview observations from MICS Plus showed high response rates: "...we have got a response rate in excess of 83-84% without substitution in Mongolia, which is an incredible number for phone surveys". For the Refugees in the German Educational System survey, researchers kept incentive amounts as high as FTF when moving to CATI, ensured the same interviewers contacted the respondents as before, and

explained the reasons for the switch to respondents to avoid response rate reductions in the mode switch.

There have been observations of response rates going up for surveys that were already using telephone interviews prior to the pandemic. The New York Times reported, “People are now happy to pick up the phone” in a column published on April 17⁷⁰. Raphael Nishimura confirmed this observation for some surveys at the SRO, while also stressing that the reason was higher contact rates rather than lower refusal rates: “An interesting point is slightly higher contact rates compared to previous years that we have been observing since the beginning of the pandemic, especially for the cross-sectional telephone survey that I was referring to. Other survey research centers have been observing the same pattern here in the US, maybe more broadly in the world. There is a hypothesis for that. People were stuck at home because of quarantines, lockdowns. So, they were more likely to answer the phone even though they did not recognize the number. Interestingly, we also saw an increase in refusal rates. ... It tells us that even though people answer their phones more, those who refuse the surveys are still refusing”.

For online surveys, response rates are usually lower than interviewer-administered modes. Angelo Ambitho, for example, stated that the web mode in Kenya or East Africa provided response rates around 10-15%, whereas we see very high FTF response rates from DHS surveys in these areas. On the other hand, Torbjörn Sjöström reported a response rate above 70% for their COVID-19 study with a probabilistic online panel in Sweden. Rinken et al. (2020) obtained a response rate below 1% for their design based on Google Ads. Marcus Berzofsky from RTI International underlined this issue in the commentary he made for the paper by Kapteyn et al. (2020), saying both probabilistic and non-probabilistic panels were subject to low recruitment and participation rates, and at the same time, how COVID-19 related surveys would get very high response rates because they were relevant to everyone’s lives. James Lepkowski also discussed online panels through a nonresponse perspective: “The problem with panels is that there is a nonresponse issue. They have millions of people on these panels, but they are not necessarily a carefully selected sample. It is a sample built on the basis of people deciding whether they want to participate...”, while also saying they are faster and cheaper compared to FTF and phone.

Measurement error implications have to be considered whenever mode switches happen. As Gillian Prior from NatCen underlined in the RSS webinar, mode switches from FTF to others took place without any experiments; thus, little will be known about measurement errors. Experimental designs for these studies might be considered after the pandemic to explore these issues. This issue was brought up in our interviews as well. Frauke Kreuter stated that: “The problem with the pandemic was that we had to shift modes without being able to experiment if the old and new mode can run in parallel. You can estimate the measurement difference between the two and then continue onward with the next one”, and Raphael Nishimura said that: “Many projects had to change the mode on the fly. They did not have enough time to prepare the survey for that sort of switch”. Enver Taştı confirmed this for TURKSTAT: “The pandemic's sudden outbreak caused the system to be adopted without making the necessary preparations for other studies (other than Household Labor Survey)”

As Raphael Nishimura stated, “...measurement is another topic that will be triggered a lot since we know that the mode effect varies when the interviewer is right there in front of the respondent or not.

⁷⁰ <https://www.nytimes.com/2020/04/17/us/politics/polling-coronavirus.html>

Telephone surveys are still interviewer-administered but without visual interaction". We can assume that this degree of contact will affect the rapport between the interviewer and the respondent. In cases where an interviewer-administered mode is switched to a self-administered one, we can expect respondent motivation to decrease.

James Lepkowski exemplified how respondent behavior could change over the phone relative to FTF: "Will people answer the questions we have been asking in the past or even new kinds of questions when you call them on the phone? Will, they answer less seriously or, in some cases, more seriously these questions?". Enver Taştı talked about the issue for national surveys: "In face-to-face interviews, interviewers show their credentials and other official documents, which develop trust, and then they move on to questions. That is why interviewers need to receive solid communication training to build up trust in phone calls.". Oğuzhan Akyıldırım shared the same observations in the private sector context in Turkey: "Moreover, trust in face-to-face is much higher than on the phone. Thus, the questions start to be challenging with trust problems between interviewers and respondents on the phone. Even when the participants are convinced about the research aims, for whom the study is done, how results will be used, anonymity and confidentiality; there are trust problems on the phone, making it difficult to manage."

Attila Hancıoğlu talked about rapport as a function of mode, isolating the respondent, interview length and questionnaire flow: "... talking about the methodology, face-to-face interaction, rapport that the interviewer builds with respondents usually being in a separate place without anybody else's involvement or distraction, structuring a questionnaire from a longish questionnaire, going from the general stuff to more detailed, building rapport along the way, and being able to go into more sensitive topics using the opportunity of being on 1 to 1 and face-to-face basis. ... There might be better chances of the respondent giving a better quality answer on the phone if you are creating responding fatigue in a household survey when you are interviewing people for 50 minutes or an hour, as opposed to a simple phone call where you ask a straightforward factual question to the respondent, and it is known that the whole interview is going to be 10-15 minutes". Oğuzhan Akyıldırım also mentioned the presence of third parties in telephone interviews: "For example, the respondents may still contact their family while answering questions in their homes. It is challenging for us to prevent this. We ask them to move to a private place to conduct the interview. Unfortunately, this is not likely to happen, particularly for the low-income groups who live in a single unit or room... Also, while talking to the interviewers, respondents may be watching the TV, or someone else's phone can ring, for instance.. ... Further, if the questionnaire is lengthy, measurement error becomes an issue, especially for the questions at the end since respondents do not immerse themselves or may not be interested anymore with a desire to end the interview", whilst concluding that measurement error is the most important type of survey errors in their survey on refugees.

These statements bring us to the issues of questionnaires, which were almost always altered during the pandemic. Regardless of modes being kept the same or not, COVID-19 related questions were added to most surveys. In most instances, when an FTF was switched to another mode, questionnaires had to be shortened. At the WAPOR May webinar Ignacio Zuasnabari underlined that question wording would be adapted to different modes and that the length of the questionnaire would also need to change, given FTF is the mode where the longest interviews are possible. Attila Hancıoğlu also underlined that one would have to be "economizing on the number of questions" for a phone survey.

Gillian Prior mentioned that the British Social Attitudes Survey questionnaire was designed to take an hour FTF and was reduced to take 25 minutes online. Raphael Nishimura mentioned the same phenomenon for Zika Survey questionnaire, originally planned to take 60 minutes FTF, had to be modified to last 20-25 minutes over the phone. The SHARE Corona, as a CATI survey, was a shortened version of the original SHARE questionnaire, with questions added on COVID-19 topics. Oğuzhan Akyıldırım also mentioned that they had to shorten their questionnaires to some extent. In the pairfam survey, prior self-administered CASI component filled during FTF visits was edited for simplicity as a postal SAQ. For MICS Plus, since the aims are not the same with that of a MICS survey, a questionnaire comparison was not relevant; however, it was important to demonstrate that some modules were impossible to implement in telephone surveys when a proxy respondent provides all information for the household, as Attila Hancioğlu exemplified with topics such as contraception.

Sampling errors also need to be considered. A mode change may imply a change in sample size and design. A smaller than expected sample size will be obtained for an interrupted survey that was later considered complete, yielding lower precision. If we are talking about a cross-sectional survey where we do not have phone numbers in our address frame, a new design will be needed. Also, for household surveys, the sampling unit is the household, whereas it is often persons in other modes. James Lepkowski explained this as: "...you cannot use the same methods for sampling households over the phone that you do when doing them in person. Because in-person sampling involves physically looking at housing units as part of the process and typically having lists of addresses that do not have phone numbers. There is disjunction there not only in terms of the interviewing but also in terms of the sampling methods".

The selection of individuals within households is also affected. While probability selection of individuals is straightforward in FTF upon getting a household roster, it is harder in other modes. Pseudo-probability methods, such as birthday methods, may have to be used. Gillian Prior underlined this issue in the RSS webinar. Kühne et al. (2020, p. 196) also mentioned this aspect in their paper for online surveys: "For instance, conducting multiple personal interviews in households is difficult to administer in online surveys. When such information (e.g., data about other household members) is collected, it is usually only from the person interviewed. Two problems arise here: First, it is only possible to collect proxy information about third parties inflating the risk of misreporting. Secondly, problems of missing data emerge due to the natural inability of one person to report fully on another person."

For switches to telephone surveys, telephone sampling may need to be done. James Lepkowski and Raphael Nishimura mentioned random digit dialing (RDD) among possible solutions. This method, for instance, is not used by TURKSTAT, and Raphael Nishimura underlined that it was not common in Brazil either. However, it has been used in the U.S. for decades, with some challenges regarding its use on mobile surveys. Thus when it is planned to be used for the first time in a country, certain setups are needed. As Raphael Nishimura explained: "We cannot just tell the local survey providers: 'Well, we have a list of telephone numbers.' It could be just as some sort of a supplement, given its serious coverage issues. We used a modification of RDD to improve efficiency, which can be referred to as a list-assisted RDD. Even with that, the hit rate of working residential telephone numbers was very low. For each successful contact, we had to call around a hundred numbers. They had an autodialer, but it had a certain capacity of how many calls they can make. So, we had to also work through different ways to improve the sample's efficiency to reach the sample size they were targeting in the survey period".

For national surveys, another alternative was merging address information with the address frame from other registration systems. This was one of the recommendations of EUROSTAT for EU-SILC's first wave respondents, and it was what we learned from Enver Taştı as the method of reaching first wave respondents for the Household Labor Force Survey of TURKSTAT. In this case, the sampling methodology would be the same as address-based sampling, with little sampling error implications (except for the lower number of cases achieved if lower response rates over the phone – implying less precision).

Commercial frames are another source for telephone sampling. Oğuzhan Akyıldırım mentioned commercial databases they sometimes work within Turkey. He underlined that these databases cover around 60-70% of the population and that they are thinking of ways to represent the remaining of the population for future surveys.

For voluntary or online panels, where samples are selected with non-probabilistic methods, self-selection is a problem that comes up, as Kühne, Kroh, & Liebig (2020, p. 196) put it: "...these studies face limits when it comes to providing results that can be generalized to a country's entire population. This is because people who select into a given study themselves, rather than being selected through a random statistical procedure, generally do not offer a representative picture of the larger population. ... Furthermore, by definition, online studies underrepresent population groups without Internet access or people who rarely or never use the Internet". Vezzoni et al. (2020) mentioned these types of studies, saying, "As for any non-probabilistic sampling, representativeness is a major concern," while also underlining coverage properties of internet surveys. At the same time, they mentioned increasing internet penetration and some research suggesting opt-in panels can provide similar accuracy to probability surveys.

Adjustment errors have also been mentioned. Ignacio Zuasnabari mentioned that weighting is an issue in mode switches because surveys in some Latin American countries may provide biased samples in the absence of FTF mode. Gillian Prior suggested that there would be less information available for weighting in the absence of doorstep information. The UN Economic Commission for Latin America and the Caribbean published a brief about detecting potential bias in household surveys during the pandemic and suggested methods such as propensity score adjustment, two-stage calibration method, and post-stratification based on multilevel models.

Cost implications can also be considered, although we did not focus on this issue in the former chapters. Questionnaires may be printed or programmed for fieldwork scheduled around the time when the pandemic was declared and may be outdated by the next fieldwork date in the case of a postponement. Interviewers might be trained when fieldwork was suspended and may have to be re-recruited and/or retrained. In the case of mode switches, new samples and tools may have to be designed.

Enver Taştı, for the Household Labor Force survey, mentioned that costs went down, with the elimination of travel costs. Oğuzhan Akyıldırım, on the other hand, stated otherwise, saying costs went up with the pandemic: "There are two main reasons for this. First, design times are getting too long. Indeed, both the design and analysis phases extended. We employ more competent, talented, and skilled people. ... The design is essential. We need to design it well and need to think about it thoroughly. Once the investigation starts, we cannot go back. It is not possible to say, 'the questionnaire does not work like this; let's change it.'.... After the data is collected, the data analysis starts to be very difficult due to the measurement errors I mentioned because the inconsistencies increase.", and he did not anticipate CATI costs to go down in the long run either.

- *Some survey topics are more affected than others*

Some surveys, regardless of country, are only possible through FTF data collection, such as health surveys, surveys of elderly, DHS and MICS surveys with biomarker collections, which are either impossible or very hard and open to error if collected otherwise (by self-report, for instance). Surveys with sensitive topics are also hard to conduct if they require both privacy and rapport at the same time.

Among examples, James Lepkowski listed the following: "...surveys collect data in many of the other modes: personal examination, physical examination, neurological examinations, tests, and things that cannot be done by telephone or online, or at least done reliably". Attila Hancioğlu talked about such components in MICS surveys: "...is one of the two times that we have to be very close, if not interacting physically, with the respondent or the household member. One is anthropometric, and the other is when we apply the foundational learning skills module where the interview has to sit next to the child, exchange material where we are testing the child's numeracy and literacy skills. Those are two key points for us", which would not be possible without FTF visits.

Enver Taştı mentioned the Health Survey of TURKSTAT as an example of collecting sensitive data, which would be hard to do in a mode other than FTF: "The Health Survey has questions about smoking and alcohol use. You first isolate the individual and then conduct the face-to-face interview so that their answers are not affected by their environment."

Enver Taştı also talked about surveys that collect detailed information with high respondent burden, which could only be tolerated in FTF: "Some surveys are difficult even with face-to-face mode. For instance, in the Health Survey, there are many detailed questions. Conducting this with the telephone mode can cause problems. On the other hand, in income and living conditions surveys, there is not much problem with yes-no questions as well as questions about income, social issues, or questions that the respondents can answer comfortably. But, some questions cause problems if the respondents are hesitant about its content. We encounter problems with these types of questions, even in our face-to-face surveys. I think the implementation of CATI raised these problems further. Sometimes it may be easier to get answers to some questions on the phone. Respondents can easily answer a question on the phone that he/she cannot talk about face-to-face. We need to consider its pros and cons."

As another example of what kind of data may be hard to get, Frauke Kreuter mentioned those requiring verification: "I think that census data collections can be affected. In ways that it might be hard to verify people, which is done in person".

Talking about this issue, Oğuzhan Akyıldırım stated that some groups of the population were now hard to collect data from, rather than specific survey topics: "Due to the reasons I mentioned, the pandemic made it challenging to collect data on every subject, but mostly the data collection from people of low socioeconomic status and disadvantaged groups. It was easier before, and these groups are most prone to giving data because they make their voices heard about their complaints using the surveys.. ... As I said, it is easier in market research but very difficult, particularly for social and economic research. For instance, it was already challenging to ask a person about his/her earnings and income face-to-face, but now it has become much tougher by telephone."

- *Ways of data collection that are unsusceptible to crisis situations are necessary for every country*

The effect of the pandemic on especially FTF surveys, shows how important it is to have frames that allow for rapid data collection. These could be longitudinal surveys with phone information,

longitudinal phone surveys, or online panels. Sastry, McGonagle, and Fomby (2020), in their commentary on the paper by Will et al. (2020, p. 251) underlined this necessity “...survey researchers need to begin thinking about how the pandemic will alter social interactions and fieldwork in the years ahead. ... The first is the value and importance of being able to collect data remotely—that is by telephone, the internet, or by video—from respondents who cannot be visited in person due to reasons of cost, logistics, extenuating circumstances, or preferences”.

An exemplary attempt in this manner is MICS Plus. The initiative was designed for the very purpose of “emergency monitoring”, and MICS Plus surveys could be rapidly implemented over the course of the pandemic to provide monthly panel data. For developing countries with high reliance on FTF surveys, it is worth considering to get consent for later contact from respondents and obtain telephone information from already designed probability samples. Attila Hancioğlu underlined that this was their plan for all future MICS surveys. He described the advantages of the MICS Plus approach as follows: “COVID-19 came in, and we were able to start collecting data immediately, already having the household sample frame. Second, you come up with longitudinal data with this approach. Third, you can look at the seasonality of different indicators that cross-sectional household surveys usually suffer from. Therefore, there are many different advantages to this approach”.

Vezzoni et al. (2020, p. 192), discussing their study from Italy based on an online panel, stated that: “Finally, this article is also a plea to call for the development of social research infrastructures meant to monitor social change, constantly and frequently.” Online panels are very appealing for rapid, low-cost data collection. We think it is important to make the probability and non-probability distinction at this point. Although the latter, too, allows researchers to observe change over time; an unknown coverage does not allow for population inference. Blom et al. (2020) argue that fast data collection is often possible through non-probability samples, but these do not accurately reflect the general population. Cost is another distinction between the two approaches, with online probability panels being harder to set up and maintain.

- *Innovations in respondent recruitment*

Only seven years ago, Mick Couper, in his keynote address at the 2013 European Survey Research Association Conference, said: “For example, Facebook is not likely to make their database of members available to researchers for sampling or analysis, even at a fee”. In 2020, we saw one of the largest and fastest global surveys in history, using Facebook. Despite being limited to Facebook users, the survey by Kreuter et al. (2020) allowed researchers to draw probability samples and collect a vast amount of data from around the world on a daily basis, which provided data about a once-in-a-century event. Frauke Kreuter explained the uniqueness of the survey as: “What is unique, I would say, is that it is such a far-reaching platform, and it is the advantageous side. Also, it certainly is an innovation given that we suddenly had a global sampling frame; that, of course, has flaws as any frame does”. Her discussion on social media platforms shows us that future is likely to bring more cooperation with researchers: “I think, to the extent that they can, and it does not hurt their business, these companies are open to that idea. It is more likely that they do this in the fashion they do now than they would do it for a fee that you can regularly pay this as a service. Because then it becomes a product that you have to be able to deliver. After that, if it is not aligned with your core company interest, maybe that will not work anymore at some point. I think this “social good” aspect will maintain, which I see in many companies, and I see an interest too. When someone has a good idea that does not harm the

business or people's privacy, the companies will be open to that. I do not think that is a solely pandemic context behavior".

We also saw examples of Facebook Ads or Google Ads being used to recruit respondents for COVID-19 related surveys, such as those by Rincken et al. (2020), Perrotta et al. (2020), and Mejova and Kalimeri (2020). Currently advising PhD students on this topic, Frauke Kreuter commented on this method of respondent recruitment as follows: "I think it is just as promising an alternative as many of the non-probability panels are... If you are a researcher and you frequently use platforms like Mechanical Turk or SurveyMonkey, recruiting your respondents over the Facebook platform via advertisement might provide a better population. You can reach a much more diverse group than those that sign up to be one of those click workers".

- *Organic data was unaffected by the pandemic, unlike survey data...*

Bob Groves, in his 2011 paper titled "Three Eras of Survey Research," mentioned "the growth of continuously produced process data from digital systems in all sectors, but especially those emanating from the Internet" as one of the highlights of the third era of survey research, which he defined as the 1990 and onwards (2011, p. 861). "Big Data" has really been big since, receiving attention from scientists from a variety of fields. Such data today is being used by social scientists to understand human behavior, along with other things. The flow of online data, unlike survey data, has not been affected by the pandemic and has once again proved its worth by providing insights to researchers during a time of crisis. We saw examples of Google data use, potentially shedding light on the course of the pandemic, as well as people's compliances with non-pharmaceutical interventions, such as recommendations to stay home.

The AAPOR Task Force Report on Big Data from 2015 mentioned a paradigm shift happening in terms of data. We asked Frauke Kreuter whether she thought this shift was affected by the pandemic, to which she replied, "everything feels like it is picking up in pace."

Attila Hancioğlu mentioned how such data was in their agenda at MICS: "We are talking about triangulation here to some extent; we are talking about opening our horizon and trying to use different types of data with different types of robustness, representativeness, and quality. We are talking about using different data sources along those dimensions and putting them all together to build the story that can happen as a mixed-method data collection effort. ... We have a separate initiative now in MICS we are about to embark on, called MICS Link. From the outset, it is to construct a MICS survey triangulated with another data source".

Administrative data also falls within the definition of big data (Japac et al., 2015). Enver Taştı, after touching upon the importance of administrative records, mentioned the use of big data to produce statistics: "Many countries in Europe produce most of the significant statistics based on administrative registers. With the increase in internet and technology usage and the advent of the 'internet of things', the data flow from the smart systems at home, workplaces, factories, and transportation has been thought to use for statistical purposes."

- *...not to mean that the era of surveys is over.*

Despite the advantages of big data – that it is "found," and timely, and is a cost-free way of obtaining data, a lot of effort is required to analyze it and understand its unique sources of error. Our interviews with experts indicated a tendency to describe such data sources to be rather complementary to survey

data, as Attila Hancioğlu put it: “From a methodological perspective, I do not think those initiatives themselves are even talking about replacing traditional data collection.... They complement each other rather than replacing each other”.

Frauke Kreuter stated that there actually are instances where researchers should think whether a survey is necessary for their interests or if such information was already available elsewhere: “But I think you cannot just rely on these other data sources. What is needed, more likely, is a combination of different data sources. Surveys still have their role, and we might be well advised to think more about the products that have a variety of different data sources rather than banking on big data sources x, y, or z”, and in response to our further questions added “Will it be the end? No. Its role is shifting. It is a good idea for anyone to think of an alternative data source before collecting everything with survey data. Also, I think we had a time of too many surveys out there and unnecessarily so. Anyone could learn a lot about human behavior without necessarily having to ask.”

James Lepkowski underlined that survey data was required as a benchmark and how valuable it was to have thorough knowledge on the error properties of survey data: “But will it substitute for what we have been doing by surveys? In some cases, yes. It is going to happen that we will find some uses of that data that will be an adequate substitute. But it will be an adequate substitute because we have been able to compare it to the survey, to the survey results as the standard... The survey is not necessarily perfect quality, but we understand a lot about its quality. And when we move to another data source, especially these data sources, they come from commercial sources with properties that are not well understood”.

Raphael Nishimura also mentioned higher control of researchers over survey data and discussed the trade-off in terms of design/analysis efforts: “My view on this is that even though big data at least provide an alternative data source, I still think there is a role for surveys. I mean, surveys are designed data, which means that you have more control over the data quality. It is more expensive to collect, but it requires less effort to analyze, whereas using big data may be cheaper, but it might require much more effort to analyze, get a proper meaning, and draw proper inferences. So, it depends on the allocation of the resources.”

Enver Taştı mentioned some quality issues to consider: “There have been serious studies in the last five-six years on new data sources such as big data, which is considered to lead the data revolution. Now the studies focus on which data to obtain, as well as how to get that more effectively, and the main concern is the quality and standardization of data. There are also issues regarding the classifications. Many countries discuss how to use this.”

Frauke Kreuter underlined the importance of looking at data with an error framework: “TSE framework always suffered a little bit from not having enough attention to how variables were created,” and said, “Therefore, I think it is crucial to have the TSE framework in mind and evaluate the data’s certain use and the quality of these different data streams.”

- *There was a need for new data about the multiple dimensions of the effects of the pandemic*

We saw that many surveys were done because of the need to understand how people’s lives were affected by the pandemic, in areas such as physical and mental health, health service seeking behavior, testing, compliance with disease prevention behaviors such as mask-wearing and social distancing, social life, family dynamics, childcare arrangements, work-life balance, labor force, etc. Moreover, there was a big interest in public opinion regarding people’s perception of the risk of a pandemic, attitudes on government responses, recommended safety measures, vaccines, etc. While some

surveys were based on frames of prior surveys (Mannheim Corona Study, SOEP-Cov, ELSA COVID-19 Study, etc.), some were designed anew. In some cases, questions were added to ongoing surveys.

International organizations prepared some guidelines for some core topics. For instance, the World Health Organization published a manual for research on behavioral insights on COVID-19, including a questionnaire. The World Bank suggested high-frequency mobile phone surveys be conducted, again with a core questionnaire design. UN women recommended rapid gender assessment surveys with recommended questions. We also saw international efforts of COVID-19 related data collection from private companies, such as Facebook, Gallup, and Ipsos.

- *There was a need to know the extent of the spread of the virus*

While the number of tests and infections are available for most countries, they do not tell us the prevalence of COVID-19 infections at a given time or what percentage of the population has antibodies. This is because testing is not random, and there are usually for those with symptoms, depending on country protocols, and it is known that there are asymptomatic cases.

Numerous designs have been proposed for such purposes, such as mailing out self-test kits to addresses (Frasier et al., 2020) or asking randomly selected respondents via mail to visit their GP for testing (Schnell & Smid, 2020). The World Health Organization released a protocol for such research in March to determine the extent of the infection and determine the proportion of asymptomatic cases. The protocol did not place emphasis on probability samples, which we think may be due to time or cost concerns. We looked at several country examples in this study, from the UK, Brazil, and Turkey, using different tests and probability samples. Countries differed in terms of whom they test within a household, what age limit they use, what safety measures they employ, and what kind of questionnaire accompanies the household visit for COVID-19 testing.

Raphael Nishimura underlined the importance of probability surveys for this issue: “In terms of sample design, I would definitely consider a probability-based design vis-à-vis non-probability sampling given the amount of effort you have to go through for data collection”.

In addition to surveys, we saw some efforts using self-reports or apps. Similar to the Google Flu Trends series, we saw researchers looking into correlations between COVID-19 symptoms searches and disease statistics and mentioned some of these studies in Chapter 3. James Lepkowski reminded us of the potential drawbacks of such data, going back to the Google Flu Trends example from the last decade: “But it turned out that people were using the searches in different ways than anticipated. And the online measurement from the searches after two years did not track the same way as the survey measure anymore”.

- *Survey researchers and organizations have paid close attention to developments in 2020*

The survey research community has been active throughout 2020. International organizations, such as the World Health Organization, UN agencies, EUROSTAT, research institutions, academic organizations, professional associations have been sharing experiences and information from the very beginning. We have seen briefs to guide surveys, webinars, and online workshops to discuss survey issues, calls for questionnaire modules, journal articles to explain the effects on surveys, propose new designs, and websites devoted to enlisting COVID-19 related surveys for researchers. We tried listing as many of these as possible in Chapter 3.

Frauke Kreuter described this unexpected positive impact on the survey research community as follows: “And it connects people. I mean, the interview we do now, all the conferences that are now more easily accessible for people who would not travel to the U.S, for instance ... I think that knowledge distribution is increasing too ... It will be good to preserve the interconnectedness that we all now have based on this shared, yet negative, but still experience”.

- *Increased reliance on mixed-mode seems to be the future*

Mick Couper mentioned the “rise of mixed-mode approaches” as one of the three main trends in survey research back in 2011 (Couper, 2011). This trend, as we observe, is getting stronger, or appearing recently in some countries with the COVID-19 pandemic, with respondents being recruited, interviewed or followed up with different modes. Raphael Nishimura mentioned that such surveys were receiving attention in developing countries: “Web and mixed-mode surveys have also seen more interest in developing and other countries that have not used these modes.” Enver Taştı stated that “I think mixed-mode will prevail, and there is no turning back. I am not just saying this for our country (Turkey). Considering the measures that other countries are currently discussing, I do not think there will be a return in many countries ... I can use CATI for the data not available in administrative records or if it is inadequate. Or, it is possible to get the required data from administrative registers as well. I think such studies can be applied with the mixed-mode.” Attila Hancioğlu also expected increased use of mixed-methods: “I think there will be much more reliance on mixed methods collecting data through phones, Web-based interviews, face-to-face, etcetera. I think it is something that COVID is interestingly helping us to approach much faster than we did to using mixed methods, either within a single data collection effort or multiple data collection exercises or approaches to build the story altogether and create an ecosystem so that there is a more comprehensive understanding on what is happening”. Oğuzhan Akyıldırım also anticipated a similar trend: “Other modes will now be used with the face-to-face mode. For instance, while some data collection will be done face-to-face, the other parts will be conducted by phone, mail-out, and Web modes. I think mixed-mode research will be increasingly utilized, and we will organize that according to the groups we target. For example, we will go to hard-to-reach groups face to face while surveying easy-to-reach groups via the Web or phone. And, we will employ regular mail mode to the groups we are confident about. Furthermore, we will need to develop sampling methods for mixed-mode”. From a methodological point of view, Sastry, McGonagle, and Fomby in their commentary on the paper by Will, Becker, and Weigand (2020, p. 251) underlined the necessity of implementing mixed-mode data collection “to develop an understanding of the nature, causes, and magnitudes of mode effects.”

- *Nothing is “new”, it is mostly change picking up pace*

Most of what we talked about in this section are developments and ideas that survey researchers have been familiar with for a long time; be it switching from FTF to other modes, be it the increasing use of mixed-mode, or new uses of the internet for data collection and respondent recruitment. However, they are still new to some countries and contexts.

Raphael Nishimura stated how established these topics are in the literature: “Most error types stemming from mode switches of face-to-face to other modes has already happened over the past twenty-thirty years. Face-to-face to telephone transition happened during the sixties and seventies in the US, and there was a lot of research and discussion about that. ... We have been having conferences, webinars, talks about the consequences of the pandemic in survey research. However, I am yet to see anything new on any of those discussions other than observing more considerable interest of countries that did not have telephone mode as the predominant survey mode. In a way, the survey research has

already been through that. Web and mixed-mode surveys have also seen more interest in developing and other countries that have not used these modes” while underlining differences by countries. Attila Hancioğlu put it as: “I think we are all going to be considering more about mixed methods. Is this new thinking? No, it is not. It precedes the COVID”.

Turkey is an example of countries where such topics are new; Enver Taştı explained the changes for TURKSTAT as: “Currently, the pandemic has forced a transition to mixed-mode. There is a full switch to CATI now, and it may continue if there is no problem. Would we have switched to CATI completely this year if there was no pandemic? Not sure. To sum up, yes, the switch to mixed-mode would happen in the future anyway. The pandemic advanced this process”.

Sastry, McGonagle and Fomby (2020), on their commentary in Will et al. (2020, p. 252) also underlined how the changes in methods are occurring faster with the pandemic: “In conclusion, survey research methodologists and practitioners should use the disruption and upheaval caused by COVID-19 to identify and capitalize on important new trends in data collection that will occur faster in the coming months and years than could have previously been imagined”.

- *The future holds challenges*

We think that the big challenge COVID-19 created on in-person interactions will affect FTF surveys for at least another year. As of December 2020, we are in the second global wave of the pandemic; there is the news of vaccines being developed and news of new virus mutations being formed. Lockdowns and restrictions are in effect in most places. James Lepkowski, with his prior expertise in the field of epidemiology, stated that “going back to normal” would not be so easy: “But I think the more important issue is that this pandemic problem is not going away. A vaccine is not going to solve all our problems.”

Although the situation is hard to read, researchers have been thinking about ways to make FTF interviews possible for countries, for surveys, or survey topics for which this mode is assumed a must. Sastry, McGonagle and Fomby (2020), in their commentary on the study by Will et al. (2020, p. 251) stated that “ how the pandemic will alter social interactions and fieldwork in the years ahead” was an issue to be considered. Gillian Prior from NatCen also mentioned that many commissioners wished to return to FTF mode as soon as it was declared safe at the RSS webinar. James Lepkowski underlined that this demand was dependent on several aspects: “Would we want to move entirely back to face to face? Is it possible that we would, depending on the country's circumstance, in terms of the data's cost and timeliness, people may find the telephone a little faster, more timely, and attractive?”.

Attila Hancioğlu said he was optimistic about the return to FTF, based on the approaches of national statistics offices: “I think I have a more optimistic view of what is going to happen than most. There was a survey back in August by the World Bank and the Statistics Division of the UN. It was a survey of all the national statistics offices around the world. One of the questions to the statistical offices was, “When do you expect to go back to face to face data collection?”. In that survey, close to half of the statistical offices said within the next six months.... . Once this is over, I think we will go back to what will be the new normal. What will be the new normal? How will the households react to face-to-face interviewing? Is this going to influence response and completion rates in countries? That remains to be seen”.

Enver Taştı pointed out the need to continue FTF despite the emergence of new techniques: “With the pandemic, new data collection techniques have been started to be used, and I think these techniques will be developed and continued to be employed. However, I do not think that classical methods will

be abandoned (FTF) . If the informality is high or there are no third data sources that we can check, the first interviews will and should be conducted with the classical method.”

Raphael Nishimura suggested that most organizations would be keen to return to FTF as soon as possible. However, he underlined that permanent change is also possible, and actually helpful for cost reasons: “...while face-to-face data collection and area probably sampling is and will probably be the gold standard, it does not mean that other data collection modes are not good enough for most surveys. I think that many places will try to resume their previous data collection mode, face-to-face or not, as soon as possible. Also, there will probably be some other institutions, I hope actually, that will think twice and consider more carefully for other alternative data collection modes. Does the amount of bias stem from coverage properties, which sometimes are in negligible numbers, justify the more money and effort you invest for face-to-face data collection, which also can be done through telephone mode?”. Frauke Kreuter also underlined the cost aspect of surveys for post-COVID times: “One big challenge is the financial meaning of this. How are the resources going to be distributed in surveys? Good surveys cost money. That will continue to be a challenge. I think many things will not go back to pre-COVID just like that. There have been innovations in data collection that probably will not go away. ... I cannot see the reason why those kinds of innovations would be turned back”. James Lepkowski also underlined cost while suggesting how to deal with potential bias: “I can learn on the phone and calibrate to face-to-face. I could collect more interviews for the same amount of money because the phone is less expensive. But I need to make sure that if it has a bias in it to nonresponse or coverage or measurement error problems, I can adjust from what I know, face-to-face”.

While these cost discussions underline that FTF is not necessarily the only and best way to collect data, some researchers underlined its advantages, such as Kühne et al. (2020, p. 197) saying, “...(future) face-to-face surveying with an interviewer present offers possibilities to integrate medical or psychological testing into an ongoing survey”, comparing interviewer vs. self-administered modes. Attila Hancioğlu emphasized how FTF is optimal in some cases: “We should try to keep the robustness of household surveys in the future because they are the golden standard when it comes to many types of data that can be collected, including the socioeconomic disaggregation dimension. It enables us to come up with factual, attitudinal data and behavioral data, all at the same time”. Gillian Prior underlined that FTF surveys facilitated lengthy, detailed questionnaires, ensured higher response rates, elimination of mode effects (if former time series is FTF based), and provided a collection of objective measurement.

Evaluating the measures that can be taken to make FTF possible, we saw that a variety of safety measures were employed for this kind of interviewing. For instance, in the UK, the prevalence survey included self-testing of respondents. Gillian Prior, at the RSS webinar, underlined how important it was to prioritize the safety of interviewers and respondents and to introduce protocols to ensure these. Attila Hancioğlu explained their own preparations at MICS: “But we are coming up with a series of recommendations on this, on how to deal with face to face interaction, what to do in the venue where you are providing training, how to use PPE, how and when to use masks, what happens if somebody is Covid-positive, including all those kinds of recommendations.” David Hunter from RTI International explained safety protocols they used for FTF interviewing (masks, hand sanitizers, single-use plastic gloves, interviewing outside when possible, etc.) at an AAPOR workshop for places where they were conducting FTF interviews.

There were hybrid solutions mentioned by researchers too. Gillian Prior mentioned in-home interviews could be made optional at doorstep contacts, and web, phone, or video options could be made available to respondents. Sastry, McGonagle, & Fomby (2020), in the context of PSID and its supplements, CDS, and TAS, mentioned the use of video-based calls for “recruiting respondents, for

providing explanations to respondents, and assisting complicated modules such as anthropometric measurements or collection of biomarkers.”

In any case, researchers underlined the risk of decreasing response or cooperation rates in FTF surveys due to COVID-19, such as Attila Hancioğlu: “We do not know what the households' behavior will be once the COVID-19 is over. Are we going to be suffering from much lower completion rates? Much lower response rates, is that what is going to happen? If households are more reluctant to open their doors to strangers remains to be seen”.

- *We need to work on survey documentation: Example of Turkey*

One problem we observed was the availability of survey documentation. Attila Hancioğlu touched upon this issue as such: “I see much variability in how the results have been presented, how low response plus possible bias is dealt with, how sample weights are calculated, and the language the results are presented.... Unfortunately, quite a few examples have been published based on the 50% of the population who provided phone number or what they could access with absolutely no warning or language about what the data represents”.

It is safe to argue that countries have different survey research traditions, reflecting how methodologies are presented. Our brief review of the status in Turkey revealed a general transparency problem. As we observed in our interview with Mr. Taştı, the Turkish Statistical Institute is transparent about the effects of the pandemic on their practice; however, no documentations are available on the TURKSTAT web page just yet, making this information on demand rather than readily supplied. There is no manual on the transition from CAPI to CATI for the Labor Force Survey of Turkey, a shift that took place almost a year ago. This will probably be shared as part of metadata eventually, but it will not be timely.

As far as other private institutions, we seldom see methodological details of the research in Turkey. The mode and sample size are often stated, but we almost never see the frame/a coverage discussion or response rates. Companies do not have a tendency to state where they obtain phone numbers for CATI and how they recruit their respondents in the case of CAWI. In terms of sampling, we see some non-probabilistic approaches, such as quotas; or we see examples of providing one province per NUTS-1 region and calling these provinces “representative” of their corresponding region. These, too, could be about the lack of demand in Turkey for methodology by users. We conclude that we do need some basic awareness about survey methodology issues in Turkey, which we think is likely to apply to many other countries out there.

References

- About, R., & Heydari, B. (2020). The Immediate Effect of COVID-19 Policies on Social Distancing Behavior in the United States. *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.3571421>
- Acharya, A. S., Prakash, A., Saxena, P., & Nigam, A. (2013). *Sampling: Why and How of it?* 4, 330–333.
- Ahmad, I., Flanagan, R., & Staller, K. (2020). Increased Internet Search Interest for GI Symptoms May Predict COVID-19 Cases in US Hotspots. In *Clinical Gastroenterology and Hepatology* (Vol. 18, Issue 12, pp. 2833-2834.e3). W.B. Saunders. <https://doi.org/10.1016/j.cgh.2020.06.058>
- Aizpurua, E. (2020). Interview the Expert: The Societal Experts Action Network (SEAN) COVID-19 Survey Archive, with Gary Langer. *Survey Practice*, 13(1), 1–4. <https://doi.org/10.29115/sp-2020-0006>
- Alattar, L., Messel, M., & Rogofsky, D. (2018). An Introduction to the Understanding America Study Internet Panel. *Social Security Bulletin*, 78(2).
- Allen, W. E., Altae-Tran, H., Briggs, J., Jin, X., McGee, G., Shi, A., Raghavan, R., Kamariza, M., Nova, N., Pereta, A., Danford, C., Kamel, A., Gothe, P., Milam, E., Aurambault, J., Primke, T., Li, W., Inkenbrandt, J., Huynh, T., ... Lin, X. (2020). Population-scale longitudinal mapping of COVID-19 symptoms, behaviour and testing. *Nature Human Behaviour*, 4(9), 972–982. <https://doi.org/10.1038/s41562-020-00944-2>
- American Association for Public Opinion Research. (2020). *COVID-19 Workshop Series*. <https://www.aapor.org/Conference-Events/COVID-19-Workshop-Series.aspx>
- AREA Araştırma Eğitim Danışmanlık. (2020, November). *Türkiye Siyasi Gündem Araştırması Kasım 2020*. <http://www.areaarastirma.com/files/turkiye-siyasi-gundem-arastirmasi-2020-kasim.pdf>
- AREDA Araştırma Eğitim Danışmanlık. (n.d.). *Kantitatif*. Retrieved December 21, 2020, from <https://www.areda.com/arastirma-yontemlerimiz/kantitatif/>

AREDA Araştırma Eğitim Danışmanlık. (2020a). *16 Mart Koronavirüs Algı Araştırması*.

<https://www.areda.com/koronavirus-almi-arastirmasi/>

AREDA Araştırma Eğitim Danışmanlık. (2020b). *21 Mart Koronavirüs Algı Araştırması*.

<https://www.areda.com/sokaga-cikma-yasagi-gelmeli/>

AREDA Araştırma Eğitim Danışmanlık. (2020c, February 11). *Koronavirüsü Algı Araştırması*.

<https://www.areda.com/koronavirusu-almi-arastirmasi/>

Ayyoubzadeh, S. M., Ayyoubzadeh, S. M., Zahedi, H., Ahmadi, M., & R Niakan Kalhori, S. (2020).

Predicting COVID-19 Incidence Through Analysis of Google Trends Data in Iran: Data Mining and Deep Learning Pilot Study. *JMIR Public Health and Surveillance*, 6(2), e18828.

<https://doi.org/10.2196/18828>

Azad, M. A., Arshad, J., Muhammad, S., Akmal, A., Riaz, F., Abdullah, S., Imran, M., & Ahmad, F.

(2020). *A First Look at Privacy Analysis of COVID-19 Contact Tracing Mobile Applications*.

Badr, H. S., Du, H., Marshall, M., Dong, E., Squire, M. M., & Gardner, L. M. (2020). Association

between mobility patterns and COVID-19 transmission in the USA: a mathematical modelling study. *The Lancet Infectious Diseases*, S1473309920305533. [https://doi.org/10.1016/S1473-3099\(20\)30553-3](https://doi.org/10.1016/S1473-3099(20)30553-3)

Baffour, B., King, T., & Valente, P. (2013). The Modern Census: Evolution, Examples and Evaluation.

International Statistical Review, 81(3), 407–425.

Barrett, D., & Noble, H. (2019). What are cohort studies? *Evidence Based Nursing*, 22(4), 95–96.

<https://doi.org/10.1136/ebnurs-2019-103183>

Basellini, U., Alburez-Gutierrez, D., Del Fava, E., Perrotta, D., Bonetti, M., Camarda, C. G., & Zagheni,

E. (2020). *Linking excess mortality to Google mobility data during the COVID-19 pandemic in England and Wales*. SocArXiv.

Bellhouse, D. R. (1988). 1 A brief history of random sampling methods. In *Handbook of Statistics* (Vol.

6, pp. 1–14). Elsevier. [https://doi.org/10.1016/S0169-7161\(88\)06003-1](https://doi.org/10.1016/S0169-7161(88)06003-1)

- Bhatia, M. S., & Jaiswal, A. (2016). Empirical analysis of data acquisition techniques: PAPI vs. CAPI. *2016 6th International Conference - Cloud System and Big Data Engineering (Confluence)*, 326–330. <https://doi.org/10.1109/CONFLUENCE.2016.7508137>
- Biemer, P. P., & Lyberg, L. E. (2003). *Introduction to Survey Quality*. John Wiley & Sons, Inc.
- Blom, A. G., Cornesse, C., Friedel, S., Krieger, U., Fikel, M., Rettig, T., Wenz, A., Juhl, S., Lehrer, R., Möhring, K., Naumann, E., & Reifenscheid, M. (2020). High-Frequency and High-Quality Survey Data Collection: The Mannheim Corona Study. *Survey Research Methods*, 14(2), 171–178. <https://doi.org/10.18148/srm/2020.v14i2.7735>
- Bonaccorsi, G., Pierri, F., Cinelli, M., Flori, A., Galeazzi, A., Porcelli, F., Schmidt, A. L., Valensise, C. M., Scala, A., Quattrocioni, W., & Pammolli, F. (2020). Economic and social consequences of human mobility restrictions under COVID-19. *Proceedings of the National Academy of Sciences*, 117(27), 15530–15535. <https://doi.org/10.1073/pnas.2007658117>
- Brenan, M. (2020, December 15). *Roundup of Gallup COVID-19 Coverage*. <https://news.gallup.com/opinion/gallup/308126/roundup-gallup-covid-coverage.aspx>
- Burton, J., Lynn, P., & Benzeval, M. (2020). How Understanding Society: The UK Household Longitudinal Study adapted to the COVID-19 pandemic. *Survey Research Methods*, 14(2), 235–239. <https://doi.org/10.18148/srm/2020.v14i2.7746>
- Çakır, B. (2020). Türkiye COVID-19 Seroprevalans Çalışması: Epidemiyoloji Perspektifinden Değerlendirme ve Öneriler. In Türk Tabipleri Birliği COVID-19 İzleme Kurulu (Ed.), *COVID-19 PANDEMİSİ 4. AY DEĞERLENDİRME RAPORU* (pp. 235–251).
- Callegaro, M., & Yang, Y. (2018). The Role of Surveys in the Era of “Big Data.” In *The Palgrave Handbook of Survey Research* (pp. 175–198).
- Çarkoğlu, A. (2020). *Salgın günlerinde İstanbul’da yaşam- Bir araştırma—Sarkaç*. https://sarkac.org/2020/05/salgin-gunlerinde-istanbulda-yasam-bir-arastirma/#_ftn1
- Carpini, M. X. D. (2011). *Constructing Public Opinion*. Oxford University Press. <https://doi.org/10.1093/oxfordhb/9780199545636.003.0018>

- Cochran, W. G. (1953). *Sampling Techniques*. Asia Publishing House.
- Couper, M. P. (2011). THE FUTURE OF MODES OF DATA COLLECTION. *Public Opinion Quarterly*, 75(5), 889–908. <https://doi.org/10.1093/poq/nfr046>
- Cowles, E. L., & Nelson, E. (2015). *An Introduction to Survey Research* (First Edition).
- Cross-sectional vs. Longitudinal studies*. (2015). The Institute for Work & Health.
<https://www.iwh.on.ca/what-researchers-mean-by/cross-sectional-vs-longitudinal-studies>
- De, D. R., Silva, M., & Mont'alverne, C. (2020). Identifying impacts of Covid-19 pandemic on vulnerable populations: A mixed-methods approach. *Survey Research Methods*, 14(2), 141–145. <https://doi.org/10.18148/srm/2020.v14i2.7742>
- De Leeuw, E. (2005). To Mix or Not to Mix Data Collection Modes in Surveys. *Journal of Official Statistics*, 21(5), 233–255.
- de Leeuw, E. D., & Hox, J. J. (2008). Self-administered Questionnaires: Mail Surveys and Other Applications. In *International Handbook of Survey Methodology* (pp. 239–264). Taylor & Francis Group.
- de Leeuw, E. D., Hox, J. J., & Snijkers, G. (1995). The Effect of Computer-assisted Interviewing on Data Quality. A Review. *Market Research Society. Journal.*, 37(4), 1–19.
<https://doi.org/10.1177/147078539503700401>
- De Leeuw, E. D., & Toepoel, V. (2018). Mixed-Mode and Mixed-Device Survey. In *The Palgrave Handbook of Survey Research* (pp. 51–63). Palgrave Macmillan.
- de Leeuw, E., & Nicholls, W. (1996). Technological Innovations in Data Collection: Acceptance, Data Quality and Costs. *Sociological Research Online*, 1(4), 23–37. <https://doi.org/10.5153/sro.50>
- Designing Household Survey Samples: Practical Guidelines* (Studies in Methods). (2005). United Nations Department of Economic and Social Affairs Statistics Division.
<https://unstats.un.org/unsd/demographic/sources/surveys/Handbook23June05.pdf>
- DHS Sampling and Household Listing Manual*. (2012). [Manual]. Demographic and Health Survey.
https://dhsprogram.com/pubs/pdf/DHSM4/DHS6_Sampling_Manual_Sept2012_DHSM4.pdf

- Dillman, D. A. (1991). The Design and Administration of Mail Surveys. *Annual Review of Sociology*, 17(1), 225–249. <https://doi.org/10.1146/annurev.so.17.080191.001301>
- Dillman, D. A., Smyth, J. D., & Christian, L. M. (2014). *Internet, Phono, Mail, and Mixed-Mode Surveys: The Tailored Design Method*. Wiley.
- English | *Istanbul Ekonomi Araştırma*. (n.d.). Retrieved December 21, 2020, from <https://www.researchistanbul.com/copy-of-türki-ye-raporu>
- English Longitudinal Study of Ageing. (n.d.). *The data we collect*. Retrieved December 21, 2020, from <https://www.elsa-project.ac.uk/the-data-we-collect>
- English Longitudinal Study of Ageing. (2020). *ELSA COVID-19 Substudy*. <https://www.elsa-project.ac.uk/covid-19>
- European Social Survey. (2020a, January 6). *ESS awarded major Horizon 2020 grant*. <https://www.europeansocialsurvey.org/about/singlenew.html?a=/about/news/essnews0079.html>
- European Social Survey. (2020b, April 8). *COVID-19 implications on the ESS*. <http://www.europeansocialsurvey.org/about/singlenew.html?a=/about/news/essnews0083.html>
- European Social Survey. (2020c, July 7). *COVID-19 questions selected for Round 10*. <https://www.europeansocialsurvey.org/about/singlenew.html?a=/about/news/essnews0089.html>
- EUROSTAT. (2020a). *DATA COLLECTION FOR THE EU-LABOUR FORCE SURVEY IN THE CONTEXT OF THE COVID-19 CRISIS*.
- EUROSTAT. (2020b). *EU-SILC GUIDANCE NOTE ON THE 2020 DATA COLLECTION IN THE CONTEXT OF THE COVID-19 CRISIS*.
- Evans, J. R., & Mathur, A. (2005). The value of online surveys. *Internet Research*, 15(2), 195–219. <https://doi.org/10.1108/10662240510590360>

- Everitt, B. S. (2003). *The Cambridge Dictionary of Statistics* (Second Edition). Cambridge University Press.
- Ferguson, M. (2014). *Big Data—Why Transaction Data is Mission Critical to Success*. Intelligent Business Strategies.
- Fişek, N. H. (1986). Türkiye’de Demografinin Evrimi. In *Türkiye’de Sosyal Bilim Araştırmalarının Gelişimi*. Türk Sosyal Bilimler Derneği.
- Fisher, G. G., & Ryan, L. H. (2018). Overview of the Health and Retirement Study and Introduction to the Special Issue. *Work, Aging and Retirement*, 4(1), 1–9.
<https://doi.org/10.1093/workar/wax032>
- Food and Agriculture Organization of the United Nations. (2020). *The coronavirus and the potential blackout of national statistics: Reorganizing national statistical systems in the context of COVID-19*.
- Fowler, F. J. (2014). *Survey Research Methods*. SAGE Publications.
- Frasier, A. M., Guyer, H., Digrande, L., Domanico, R., Cooney, D., & Eckman, S. (2020). Design for a Mail Survey to Determine Prevalence of SARS-CoV-2 Antibodies in the United States. *Survey Research Methods*, 14(2), 131–139. <https://doi.org/10.18148/srm/2020.v14i2.7757>
- Fuller, D., Buote, R., & Stanley, K. (2017). A glossary for big data in population and public health: Discussion and commentary on terminology and research methods. *Journal of Epidemiology and Community Health*, jech-2017-209608. <https://doi.org/10.1136/jech-2017-209608>
- Gallup. (n.d.). *How Does the Gallup World Poll Work?* Retrieved December 21, 2020, from <https://www.gallup.com/178667/gallup-world-poll-work.aspx>
- Generations & Gender Programme. (n.d.). *About: The Infrastructure is run by leading Social Scientists and National Statistical Offices*. Retrieved December 21, 2020, from <https://www.ggp-i.org/about/>

- Göksu, O. (2018). Kamuoyu Araştırmalarında Tutarlılık Sorunsalı: 2010-2018 Dönemi Kamuoyu Araştırmaları Üzerine Bir İnceleme. *Türkiye İletişim Araştırmaları Dergisi*.
<https://doi.org/10.17829/turcom.469863>
- Goodwin, G. (2020, June). Testing boundaries during the pandemic. *Research Matters*, 1–2.
- Groves, R. M. (1989). *Survey Errors and Survey Costs*. Wiley-Interscience.
- Groves, R. M. (2011). Three Eras of Survey Research. *American Association for Public Opinion Research*, 75(5).
- Groves, R. M., Fowler, F. J., Couper, M. P., Lepkowski, J. M., Singer, E., & Tourangeau, R. (2004). *Survey Methodology*. John Wiley & Sons, Inc.
- Grow, A., Perrotta, D., Del Fava, E., Cimentada, J., Rampazzo, F., Gil-Clavel, S., & Zagheni, E. (2020). *Addressing Public Health Emergencies via Facebook Surveys: Advantages, Challenges, and Practical Considerations*. SocArXiv.
- Gummer, T., Schmiedeberg, C., Bujard, M., Christmann, P., Hank, K., Kunz, T., Lück, D., & Neyer, F. J. (2020). The impact of COVID-19 on fieldwork efforts and planning in pairfam and FReDA-GGS. *Survey Research Methods*, 14(2), 223–227.
<https://doi.org/10.18148/srm/2020.v14i2.7740>
- Hallal, P. C., Hartwig, F. P., Horta, B. L., Silveira, M. F., Struchiner, C. J., Vidaletti, L. P., Neumann, N. A., Pellanda, L. C., Dellagostin, O. A., Burattini, M. N., Victora, G. D., Menezes, A. M. B., Barros, F. C., Barros, A. J. D., & Victora, C. G. (2020). SARS-CoV-2 antibody prevalence in Brazil: Results from two successive nationwide serological household surveys. *The Lancet Global Health*, 0(0). [https://doi.org/10.1016/S2214-109X\(20\)30387-9](https://doi.org/10.1016/S2214-109X(20)30387-9)
- Hansen, S. E., Benson, G., Bowers, A., Pennell, B.-E., Lin, Y. (Jay), Duffey, B., Hu, M., & Hibben, K. C. (2016). *Survey Quality* [Guideline].
https://ccsg.isr.umich.edu/images/PDFs/CCSG_Survey_Quality.pdf
- Hanson, T. (2020, June). Covid-19 implications for the European Social Survey. *Research Matters*, 0, 7.

- Health and Retirement Study. (n.d.). *Questionnaires*. Retrieved December 21, 2020, from <https://hrs.isr.umich.edu/documentation/questionnaires>
- Health and Retirement Survey. (2020, November). *2020 HRS COVID-19 PROJECT Early, Version 1.0 Data Description and Usage*. https://hrsdata.isr.umich.edu/sites/default/files/documentation/data-descriptions/2020COVID_DD_0.pdf
- Himelein, K., Eckman, S., Lau, C., & McKenzie, D. (2020). *Mobile Phone Surveys for Understanding COVID-19 Impacts: Part I Sampling and Mode*. World Bank Blogs. <https://blogs.worldbank.org/impac evaluations/mobile-phone-surveys-understanding-covid-19-impacts-part-i-sampling-and-mode>
- Inal, E., Cwalina, C., Evans, M., Flockhart, F., & Gamvros, A. (2020). *Contact tracing apps in Turkey A new world for data privacy*.
- Institute for Social Research. (2019). *PSID Main Interview User Manual: Release 2019*.
- International Labor Organization. (2020). *COVID-19: Guidance for labour statistics data collection Capturing impacts on employment and unpaid work using Rapid Surveys*.
- International Labor Organization Department of Statistics. (2020). *COVID-19 impact on labour market statistics*. <https://ilostat.ilo.org/topics/covid-19/covid-19-impact-on-labour-market-statistics/>
- Internet Surveys*. (n.d.). Pew Research Center. Retrieved August 31, 2020, from <https://www.pewresearch.org/politics/methodology/collecting-survey-data/internet-surveys/>
- Interpenetrating Samples (Sub samples)*. (2003). OECD Glossary of Statistical Terms. <https://stats.oecd.org/glossary/detail.asp?ID=3682#:~:text=When%20two%20or%20more%20samples,an%20example%20of%20the%20latter.>
- Ipsos. (n.d.). *How Does the Gallup Panel Work?* Retrieved December 21, 2020, from <https://www.gallup.com/174158/gallup-panel-methodology.aspx>

- Ipsos. (2020a, March 21). *Koronavirüs, Hane İçi Hızlı Tüketim Ürünleri Harcamalarına Nasıl Etki Ediyor?* <https://www.ipsos.com/tr-tr/koronavirus-hane-ici-hizli-tuketim-urunleri-harcamalarına-nasil-etki-ediyor>
- Ipsos. (2020b). *IPSOS GLOBAL HEALTH SERVICE MONITOR 2020 A Global Advisor survey.*
- Ipsos. (2020c, December 7). *Son Koronavirüs Kısıtlamalarından İşinin/ İşyerinin Etkilendiğini Söyleyenlerin Oranı %70.* <https://www.ipsos.com/tr-tr/son-koronavirus-kisitlamalarindan-isinin-isyerinin-etkilendigini-soyleyenlerin-orani-70>
- Ipsos. (2020d, December 17). *Public Opinion on the COVID-19 pandemic [17th December].* <https://www.ipsos.com/en/public-opinion-covid-19-outbreak>
- Ipsos. (2020e, December 21). *Koronavirüs Salgınıyla İlgili Kamuoyu ve Tüketicinin Nabzı.* <https://www.ipsos.com/tr-tr/koronavirus-salginiyla-iligili-kamuoyu-ve-tuketicin-nabzi>
- ISR and the Truman/Dewey upset.* (n.d.). University of Michigan Institute for Social Research. <https://isr.umich.edu/news-events/insights-newsletter/article/isr-and-the-truman-dewey-upset/>
- Israel, G. (1992). *Determining Sample Size* [Fact Sheet PEOD-6]. University of Florida Florida Cooperative Extension Service.
- Japac, L., Kreuter, F., Berg, M., Biemer, P., Decker, P., Lampe, C., Lane, J., O'Neil, C., & Usher, A. (2015a). *AAPOR Report: Big Data.* American Association for Public Opinion Research. <https://www.aapor.org/Education-Resources/Reports/Big-Data.aspx#8.%20Conclusions%20and%20Research%20Needs>
- Japac, L., Kreuter, F., Berg, M., Biemer, P., Decker, P., Lampe, C., Lane, J., O'Neil, C., & Usher, A. (2015b). *AAPOR Report: Big Data.* American Association for Public Opinion Research.
- Jupp, V. (2006). *The SAGE Dictionary of Social Research Methods.* SAGE Publications.
- Kalton, G. (2019). Developments in Survey Research over the Past 60 Years: A Personal Perspective. *International Statistical Review*, 87(S1), S10–S30. <https://doi.org/10.1111/insr.12287>

- Kapteyn, A., Angrisani, M., Bennett, D., Bruin, W. B. de, Darling, J., Gutsche, T., Liu, Y., Meijer, E., Perez-Arce, F., Schaner, S., Thomas, K., & Weerman, B. (2020). Tracking the Effect of the COVID-19 Pandemic on American Households. *Survey Research Methods*, 14(2), 179–186. <https://doi.org/10.18148/srm/2020.v14i2.7737>
- Karpat, K. (2010). *Osmanlı Nüfusu 1830-1914* (Second Edition). Timaş.
- Kempf, A. M., & Remington, P. L. (2007). New Challenges for Telephone Survey Research in the Twenty-First Century. *Annual Review of Public Health*, 28(1), 113–126. <https://doi.org/10.1146/annurev.publhealth.28.021406.144059>
- Kish, L. (2001). Statistical Systems: Censuses of Population. In *International Encyclopedia of the Social & Behavioral Sciences* (pp. 15049–15053). Elsevier. <https://doi.org/10.1016/B0-08-043076-7/00004-8>
- Kish, Leslie. (1965). *Survey Sampling*. John Wiley & Sons, Inc.
- Kish, Leslie, Kalton, G., & Heeringa, S. (2003). *Leslie Kish: Selected papers*. John Wiley.
- Konda Barometresi Temalar: Popüler Kültür. (2019). Konda.
- Konda: İstanbul Sözleşmesi. (2020). Konda.
- Konda Türkiye’de Toplumsal Cinsiyet Raporu: Hayat Tarzları 2018 Araştırması. (2019). Konda. https://konda.com.tr/wp-content/uploads/2019/11/KONDA_ToplumsalCinsiyetRaporu.pdf
- Kreuter, F., Barkay, N., Bilinski, A., Adrienne Bradford, S. C., Eliat, R., Fan, J., Galili, T., Haimovich, D., Kim, B., LaRocca, S., Li, Y., Morris, K., Presser, S., Salomon, J. A., Sarig, T., Stewart, K., Stuart, E. A., & Tibshirani, R. (2020). Partnering with Facebook on a university-based rapid turn-around global survey. *Survey Research Methods*, 14(2), 159–163. <https://doi.org/10.18148/srm/2020.v14i2.7761>
- Kruskal, W., & Mosteller, F. (1980). Representative Sampling, IV: The History of the Concept in Statistics, 1895-1939. *International Statistical Review*, 48(2), 169–195.

- Kühne, S., Kroh, M., & Liebig, S. (2020). The Need for Household Panel Surveys in Times of Crisis: The Case of SOEP-CoV. *Survey Research Methods*, 14(2), 195–203.
<https://doi.org/10.18148/srm/2020.v14i2.7748>
- Kurian, S. J., Bhatti, A. ur R., Alvi, M. A., Ting, H. H., Storlie, C., Wilson, P. M., Shah, N. D., Liu, H., & Bydon, M. (2020). Correlations Between COVID-19 Cases and Google Trends Data in the United States: A State-by-State Analysis. *Mayo Clinic Proceedings*, 95(11), 2370–2381.
<https://doi.org/10.1016/j.mayocp.2020.08.022>
- Kuusela, V. (2011). *Paradigms in Statistical Inference for Finite Populations Up to the 1950s* (No. 257; Research Reports). Statistics Finalnd. <https://core.ac.uk/download/pdf/14921463.pdf>
- Langer, G. (2018). The Importance of Probability-Based Sampling Methods for Drawing Valid Inferences. In *The Palgrave Handbook of Survey Research* (pp. 7–13).
- Laurie, H. (2013). *Panel Studies*. Oxford Bibliographies.
<https://www.oxfordbibliographies.com/view/document/obo-9780199756384/obo-9780199756384-0108.xml>
- Lavrakas, P. (2008). *Encyclopedia of Survey Research Methods*. SAGE Publications.
- Lee, R. M., Fielding, N. G., & Blank, G. (2017). Online Research Methods in the Social Science: An Editorial Introduction. In *The SAGE Handbook of Online Research Methods* (Second Edition, pp. 3–17). SAGE Publications.
- Lillard, D. R. (2020). A Cross-National Design to Estimate Effects of COVID-Induced Non-Pharmacological Interventions. *Survey Research Methods*, 14(2), 211–216.
<https://doi.org/10.18148/srm/2020.v14i2.7766>
- Lohr, S. L. (1999). *Sampling: Design and Analysis*. Duxbury Press.
- Loosveldt, G. (2008). Face-To-Face Interviews. In *International Handbook of Survey Methodology*. Taylor & Francis Group.

- Lugtig, P. (2014). Panel Attrition: Separating Stayers, Fast Attriters, Gradual Attriters, and Lurkers. *Sociological Methods & Research*, 43(4), 699–723.
<https://doi.org/10.1177/0049124113520305>
- Madansky, A. (1986). On Biblical Censuses. *Journal of Official Statistics*, 2(4), 561–569.
- Mahalanobis, P. C. (1965). Statistics as a Key Technology. *The American Statistician*, 19(2), 43.
<https://doi.org/10.2307/2682378>
- MAK Danışmanlık. (2020). *Ekim 2020 Türkiye Gündemi Araştırması*.
- Manfreda, K. L., & Vehovar, V. (2008). Internet Surveys. In *International Handbook of Survey Methodology* (pp. 264–285). Taylor & Francis Group.
- Mannheim Corona Study*. (2020). <https://www.uni-mannheim.de/en/gip/corona-study/>
- Mavragani, A., & Gkillas, K. (2020). COVID-19 predictability in the United States using Google Trends time series. *Scientific Reports*, 10(1), 1–12.
- Measuring and Reporting Sources of Error in Surveys* (Statistical Policy Working Paper 31). (2001). [Working Paper]. Statistical Policy Office. <https://nces.ed.gov/FCSM/pdf/spwp31.pdf>
- Mejova, Y., & Kalimeri, K. (2020). COVID-19 on Facebook Ads: Competing Agendas around a Public Health Crisis. *COMPASS '20: ACM SIGCAS Conference on Computing and Sustainable Societies*, 22–31. <https://doi.org/10.1145/3378393.3402241>
- Mesch, G. (2012). E-Mail Surveys. In *Handbook of Survey Methodology for the Social Sciences* (pp. 313–327). Springer.
- Metropoll Stratejik ve Sosyal Araştırmalar. (2020). *Türkiye'nin Nabzı—Temmuz 2020*.
<http://www.metropoll.com.tr/arastirmalar/turkiyenin-nabzi-17/1858>
- Michigan Center on the Demography of Aging. (n.d.). *Demographic Data for COVID-19 & Aging Research*. Retrieved December 21, 2020, from <https://micda.isr.umich.edu/research/covid-19/>
- Miller, R. L., & Brewer, J. D. (2003). *The A-Z of Social Research*. SAGE Publications.
- Mills, K. A. (2019). *Big Data for Qualitative Researchers*. Routledge.

- Missiakoulis, S. (2010). Cecrops, King of Athens: The First (?) Recorded Population Census in History. *International Statistical Review*, 78(3), 413–418.
- Morita, H., Kato, H., & Hayashi, Y. (2020). International Comparison of Behavior Changes with Social Distancing Policies in Response to COVID-19. In *SSRN*.
- Moynihan, P. (2020, June 25). *The coronavirus pandemic's impact on our polling*.
<https://www.pewresearch.org/methods/2020/06/25/the-coronavirus-pandemics-impact-on-our-polling/>
- The coronavirus pandemic's impact on Pew Research Center's global polling, Pew Research Center (2020).
- Multi-phase sampling*. (n.d.). Unesco Glossary. Retrieved August 28, 2020, from
<http://uis.unesco.org/en/glossary-term/multi-phase-sampling>
- NatCen Social Research. (2018). *ELSA 50+ Health&Life User Guide to the Main Interview Datasets—Waves 1 to 8*.
- Neuman, L. (2014). *Social Research Methods: Qualitative and Quantitative Approaches* (Seventh Edition). Pearson Education Limited.
- Neuman, W. L. (2012). Designing the Face-to-Face Survey. In *Handbook of Survey Methodology for the Social Sciences*. Gideon.
- Nicholls II, W. L. (1997). The Effect of New Data Collection Technologies on Survey Quality. In *Survey Measurement and Process Quality*. John Wiley & Sons, Inc.
- NORC. (2019). *GENERAL SOCIAL SURVEYS, 1972-2004: Cumulative Codebook*.
- OECD. (2020, December 4). *Testing for COVID-19: How to best use the various tests?*
<http://www.oecd.org/coronavirus/policy-responses/testing-for-covid-19-how-to-best-use-the-various-tests-c76df201/>
- Oliver, N., Lepri, B., Sterly, H., Lambiotte, R., Deletaille, S., De Nadai, M., Letouzé, E., Salah, A. A., Benjamins, R., Cattuto, C., Colizza, V., de Cordes, N., Fraiberger, S. P., Koebe, T., Lehmann, S., Murillo, J., Pentland, A., Pham, P. N., Pivetta, F., ... Vinck, P. (2020). Mobile phone data for

- informing public health actions across the COVID-19 pandemic life cycle. *Science Advances*, 6(23), eabc0764. <https://doi.org/10.1126/sciadv.abc0764>
- ORC Arařtırma Eđitim Danıřmanlık. (2020). *HALKIN PANDEMİYE BAKIŐI VE PANDEMİNİN TOPLUMSAL ETKİLERİ KASIM 2020*.
- Özbay, F. (1986). Nüfus Bilim Arařtırmaları Oturumu Üzerine Yorum. In *Türkiye’de Sosyal Bilim Arařtırmalarının Geliřimi* (pp. 411–417). Türk Sosyal Bilimler Derneđi.
- Panel Study. (2020). In *Encyclopedia.com*. <https://www.encyclopedia.com/social-sciences/dictionaries-thesauruses-pictures-and-press-releases/panel-study>
- Payne, G., & Payne, J. (2004). *Key Concepts in Social Research*. SAGE Publications, Ltd. <https://doi.org/10.4135/9781849209397>
- Peng, Y., Li, C., Rong, Y., Chen, X., & Chen, H. (2020). Retrospective analysis of the accuracy of predicting the alert level of COVID-19 in 202 countries using Google Trends and machine learning. *Journal of Global Health*, 10(2), 020511. <https://doi.org/10.7189/jogh.10.020511>
- Pepe, E., Bajardi, P., Gauvin, L., Privitera, F., Lake, B., Cattuto, C., & Tizzoni, M. (2020). COVID-19 outbreak response, a dataset to assess mobility changes in Italy following national lockdown. *Scientific Data*, 7(1), 230. <https://doi.org/10.1038/s41597-020-00575-2>
- Perrotta, D., Grow, A., Rampazzo, F., Cimentada, J., Del Fava, E., Gil-Clavel, S., & Zagheni, E. (2020). *Behaviors and attitudes in response to the COVID-19 pandemic: Insights from a cross-national Facebook survey*. Public and Global Health.
- Phillips, M., & Taylor, E. (2016). *Introduction to the British Social Attitudes survey*.
- Piazza, T. (2010). Fundamentals of Applied Sampling. In *Handbook of Survey Research* (Second Edition).
- Porter, T. M. (2020). Probability and Statistics. In *Encyclopædia Britannica*. <https://www.britannica.com/science/probability/The-probability-of-causes#ref849466>
- Processing Errors*. (2005). OECD Glossary of Statistical Terms. [https://stats.oecd.org/glossary/detail.asp?ID=2138#:~:text=A%20processing%20error%20is%](https://stats.oecd.org/glossary/detail.asp?ID=2138#:~:text=A%20processing%20error%20is%20)

20the,Context%3A&text=Most%20processing%20errors%20occur%20in,implementation%20of%20systems%20and%20estimates.

Rahija, M., Mwisomba, T., Kamwe, M. A., Muwonge, J., & Pica-Ciamarra, U. (2017). Are CAPI based surveys a cost-effective and viable alternative to PAPI surveys? Evidence from agricultural surveys in Tanzania and Uganda. *ICAS VII 2016 : Seventh International Conference on Agriculture Statistics Proceedings*. <https://doi.org/10.1481/icasVII.2016.f39e>

RAJESH SRINIVASAN, & JON CLIFTON. (2020, June 21). *Gallup Keeps Listening to the World Amid the Pandemic*. <https://news.gallup.com/opinion/gallup/316016/gallup-keeps-listening-world-amid-pandemic.aspx>

Rama, D., Mejova, Y., Tizzoni, M., Kalimeri, K., & Weber, I. (2020). Facebook Ads as a Demographic Tool to Measure the Urban-Rural Divide. *WWW '20: The Web Conference 2020*, 327–338. <https://doi.org/10.1145/3366423.3380118>

Rao, J. (2009). *Jackknife and Bootstrap Methods for Variance Estimation from Sample Survey Data*. <https://www.semanticscholar.org/paper/Jackknife-and-Bootstrap-Methods-for-Variance-from-Rao/387b3cd95087a3eb69a78ded379a2000233107fa?p2df>

Rinken, S., Domínguez, J.-A., Trujillo, M., Lafuente, R., Sotomayor, R., & Serrano-Del-Rosal, R. (2020). Combined mobile-phone and social-media sampling for web survey on social effects of COVID-19 in Spain. *Survey Research Methods*, 14(2), 165–170. <https://doi.org/10.18148/srm/2020.v14i2.7733>

Rossi, P. H., Wright, J. D., & Anderson, A. B. (1983). Sample Surveys: History, Current Practice, and Future Prospects. In *Handbook of Survey Research* (pp. 1–20). Academic Press.

Royal Statistics Society. (2020, November 17). *1/2 The impact of the COVID-19 pandemic on data collection for social surveys: Challenges and opportunities*. <https://rss.org.uk/training-events/events/events-2020/sections/the-impact-of-the-covid-19-pandemic-on-data-collec/#speakers>

- Saha, J., Barman, B., & Chouhan, P. (2020). Lockdown for COVID-19 and its impact on community mobility in India: An analysis of the COVID-19 Community Mobility Reports, 2020. *Children and Youth Services Review, 116*, 105160. <https://doi.org/10.1016/j.chidyouth.2020.105160>
- Sakshaug, J. W., Beste, J., Coban, M., Fendel, T., Haas, G.-C., Hülle, S., Kosyakova, Y., König, C., Kreuter, F., Küfner, B., Müller, B., Osiander, C., Schwanhäuser, S., Stephan, G., Vallizadeh, E., Volkert, M., Wenzig, C., Westermeier, C., Zabel, C., & Zins, S. (2020). Impacts of the COVID-19 Pandemic on Labor Market Surveys at the German Institute for Employment Research. *Survey Research Methods, 14*(2), 229–233. <https://doi.org/10.18148/srm/2020.v14i2.7743>
- Salganik, M. (2018). *Bit by bit: Social Research in the Digital Age*. Princeton University Press.
- Sastry, N., Mcgonagle, K., & Fomby, P. (2020). Effects of the COVID-19 crisis on survey fieldwork: Experience and lessons from two major supplements to the U.S. Panel Study of Income Dynamics. *Survey Research Methods, 14*(2), 241–245. <https://doi.org/10.18148/srm/2020.v14i2.7752>
- Scherpenzeel, A., Axt, K., Bergmann, M., Douhou, S., Oepen, A., Sand, G., Schuller, K., Stuck, S., Wagner, M., & Börsch-Supan, A. (2020). Collecting survey data among the 50+ population during the COVID-19 pandemic: The Survey of Health, Ageing and Retirement in Europe (SHARE). *Survey Research Methods, 14*(2), 217–221.
- Schnell, R., & Smid, M. (2020). Methodological Problems and Solutions for Sampling in Epidemiological SARS-CoV-2 Research. *Survey Research Methods, 14*(2), 123–129. <https://doi.org/10.18148/srm/2020.v14i2.7749>
- Schröder, J. (2016). *Face-to-Face Surveys* (GESIS Survey Guidelines). GESIS. https://www.gesis.org/fileadmin/upload/SDMwiki/Schröder_Face-to-Face_Surveys.pdf
- Sheikh, A., Sheikh, Z., & Sheikh, A. (2020). Novel approaches to estimate compliance with lockdown measures in the COVID-19 pandemic. *Journal of Global Health, 10*(1), 10348. <https://doi.org/10.7189/jogh.10.010348>

- Shine, B., & Dulisse, B. (2012). Does Paying More Mean Getting a Better Product: Comparison of Modes of Survey Administration. In *Handbook of Survey Methodology for the Social Sciences* (pp. 361–372). Springer.
- Sikkel, D., & Hoogendoorn, A. (2008). Panel Surveys. In *International Handbook of Survey Methodology* (pp. 479–500).
- Sperle-Heupel, I., Gassowski, M., Nielsen, S., Bremer, V., Gößwald, A., Marcus, U., Schlaud, M., & Zimmermann, R. (2020). *Technical protocol for hepatitis C prevalence surveys in the general population*. Robert Koch-Institut.
- Steeh, C. (2008). Telephone Surveys. In *International Handbook of Survey Methodology* (pp. 221–239). Taylor & Francis Group.
- Steiger, D. M., & Conroy, B. (2008). IVR: Interactive Voice Response. In *International Handbook of Survey Methodology*. Lawrence Erlbaum Associates.
- Steinhauer, H. W., Zinn, S., & Will, G. (2019). Sampling Refugees for an Educational Longitudinal Survey. *Survey Methods: Insights from the Field*. <https://doi.org/10.13094/SMIF-2019-00007>
- Stephan, F. F. (1948). History of the Uses of Modern Sampling Procedures. *Journal of the American Statistical Association*, 43(241), 12–39.
- Stoop, I. (2012). Unit Non-Response Due to Refusal. In *Handbook of Survey Methodology for the Social Sciences* (pp. 121–147). Springer.
- Stoop, I., & Harrison, E. (2012). Classification of Surveys. In *Handbook of Survey Methodology for the Social Sciences* (pp. 7–18).
- Straw Poll. (n.d.). In *Cambridge Dictionary*. Retrieved September 21, 2020, from <https://dictionary.cambridge.org/dictionary/english/straw-poll>
- T.C. Sağlık Bakanlığı. (2020a). *“Toplumun Yeni Koronavirüsle ilgili Algı ve Davranış Düzeyini Ölçecek Koronavirüs Barometresi Araştırması Danışmanlık Hizmeti Alımı*. <https://sggm.saglik.gov.tr/TR,68312/toplumun-yeni-koronavirusle-ilgili-algı-ve-davranis-duzeyini-olcecek-koronavirus-barometresi-arastirmasi-danismanlik-hizmeti-alimi.html>

- T.C. Sağlık Bakanlığı. (2020b, November 7). *07.10.2020 Tarihinde Gerçekleşen Koronavirüs Bilim Kurulu Toplantısına İlişkin Basın Açıklaması (07.10.2020)*.
<https://www.saglik.gov.tr/TR,73776/07102020-tarihinde-gerceklesen-koronavirus-bilim-kurulu-toplantisina-iliskin-basin-aciklamasi-07102020.html>
- The DHS Program. (n.d.-a). *Survey Characteristics Search, By Status: All Ongoing Surveys*. Retrieved December 21, 2020, from https://dhsprogram.com/methodology/survey-search.cfm?sendsearch=1&sur_status=Ongoing&crt=1&listgrp=2
- The DHS Program. (n.d.-b). *Team and Partners: Who We Are*. Retrieved December 21, 2020, from <https://dhsprogram.com/Who-We-Are/About-Us.cfm>
- The DHS Program. (2020, September 2). *COVID-19 Update: Some DHS surveys return to the field; others postponed until 2021*. <https://dhsprogram.com/Who-We-Are/News-Room/COVID-19-Update-Some-DHS-surveys-return-to-the-field-others-postponed-until-2021.cfm>
- The USC Center for Economic and Social Research. (n.d.). *Understanding America Study: Understanding Coronavirus in America*. Retrieved December 21, 2020, from <https://covid19pulse.usc.edu/>
- The World Bank. (2020a). *High Frequency Mobile Phone Surveys of Households to Assess the Impacts of COVID-19 Guidelines on Sampling Design*.
- The World Bank. (2020b). *High Frequency Mobile Phone Surveys of Households to Assess the Impacts of COVID-19 (Vol. 3): Guidelines on CATI Implementation*.
<https://documents.worldbank.org/en/publication/documents-reports/documentdetail/189691588696451053/guidelines-on-cati-implementation>
- The World Bank. (2020c). *Questionnaire Template*.
- The World Bank. (2020d). *COVID-19 CORE-PLUS Questionnaire INTERVIEWER MANUAL*.
- Thompson, S. K. (2012). *Sampling* (Third Edition). John Wiley & Sons, Inc.
- Time-Series Study. (2010). In N. Salkind, *Encyclopedia of Research Design*. SAGE Publications, Inc.
<https://doi.org/10.4135/9781412961288.n465>

- Timur, S. (1971). Geniş Kapsamlı Alan Araştırmalarında Veri Toplama Teknikleri: 1968 Aile Yapısı ve Nüfus Sorunları Araştırması Örneği. In *Türkiye’de Sosyal Araştırmaların Gelişmesi* (pp. 153–170).
- Tourangeau, R. (2018). Choosing a Mode of Survey Data Collection. In *The Palgrave Handbook of Survey Research*.
- Trappmann, M., Bähr, S., Beste, J., Eberl, A., Frodermann, C., Gundert, S., Schwarz, S., Teichler, N., Unger, S., & Wenzig, C. (2019). Data Resource Profile: Panel Study Labour Market and Social Security (PASS). *International Journal of Epidemiology*, 48(5), 1411–1411g.
<https://doi.org/10.1093/ije/dyz041>
- Tucker, C., & Lepkowski, J. M. (2008). Telephone Survey Methods: Adaption to Change. In *Advances in Telephone Survey Methodology* (pp. 3–28).
- Türk Dış Politikası Kamuoyu Algıları Araştırması*. (2020). Kadir Has Üniversitesi.
- Türkiye Corona Virüsü Salgını Araştırmasını Nasıl Yürütüyor?* (n.d.). Retrieved October 24, 2020, from <https://www.amerikaninsesi.com/a/turkiye-corona-virusu-salgini-arastirmasini-nasil-yapiyor/5463899.html>
- TÜRKİYE CUMHURİYETİ SAĞLIK SİSTEMİNİN GÜÇLENDİRİLMESİ VE DESTEKLENMESİ PROJESİ (SSGDP)
“TOPLUMUN YENİ KORONAVİRÜSLE İLGİLİ ALGI VE DAVRANIŞ DÜZEYİNİ ÖLÇECEK
KORONAVİRÜS BAROMETRESİ ARAŞTIRMASI” DANIŞMANLIK HİZMETİ ALIMI İŞ TANIMI,
(2020).
- Türkiye Eğilimleri—2019*. (2020). Kadir Has University.
https://www.khas.edu.tr/sites/khas.edu.tr/files/inline-files/TE2019_TUR_BASIN_15.01.20%20WEB%20versiyon%20powerpoint_0.pdf
- Türkiye Raporu: Rapor 17* (No. 17). (2020). İstanbul Ekonomik Araştırma.
https://docs.google.com/presentation/d/1DZu3V7oyAMlP1kdojzkob_cOZHM1Fa5jGsg1ouy_ygk/present?slide=id.p1
- Türkiye’nin Aşk Raporu*. (2020). Metropoll.

Türkiye'nin Nabzı: Medya Demokrasi. (2019). Metropoll.

UN Women. (2020). *Rapid gender assessment surveys on the impacts of COVID-19.*

UNDP Crises Bureau. (2020). *Digital Socio-Economic Impact Assessment Households Questionnaire Guidance notes.*

UNICEF. (2020). *MICS Plus: A Step-by-Step Guide to Implementation.*

UNICEF MICS. (2020a, March 31). *COVID-19 PANDEMIC AND IMPLEMENTATION OF MULTIPLE INDICATOR CLUSTER SURVEYS (MICS).* [https://mics.unicef.org/news_entries/157/COVID-19-PANDEMIC-AND-IMPLEMENTATION-OF-MULTIPLE-INDICATOR-CLUSTER-SURVEYS-\(MICS\)](https://mics.unicef.org/news_entries/157/COVID-19-PANDEMIC-AND-IMPLEMENTATION-OF-MULTIPLE-INDICATOR-CLUSTER-SURVEYS-(MICS))

UNICEF MICS. (2020b, June 2). *CHANGING PERCEPTIONS OF COVID-19 RISK IN BELIZE - UNICEF MICS.* http://mics.unicef.org/news_entries/165/CHANGING-PERCEPTIONS-OF-COVID-19-RISK-IN-BELIZE

United Nations Department of Economic and Social Affairs. (2020a, April 14). *Carrying out a telephone survey under the impact of COVID-19—What to consider.* <https://covid-19-response.unstatshub.org/statistical-programmes/telephone-surveys-what-to-consider/>

United Nations Department of Economic and Social Affairs. (2020b, May 15). *Using telephone interview for household surveys: A conversation with Prof. Jim Lepkowski.* <https://covid-19-response.unstatshub.org/statistical-programmes/using-telephone-interview-for-household-surveys/>

United Nations Economic Commission for Latin America and the Caribbean. (2020). *COVID-19 Reports. Recommendations for eliminating selection bias in household surveys during the coronavirus disease (COVID-19) pandemic.*

United Nations Intersecretariat Working Group on Household Surveys. (2020a). *Task Force on COVID-19 and household surveys: Methodology on COVID-19 impact surveys.* <https://unstats.un.org/iswghs/task-forces/covid-19-and-household-surveys/methodology-on-COVID-19-impact-surveys>

United Nations Intersecretariat Working Group on Household Surveys. (2020b). *Task Force on COVID-19 and household surveys: National responses to COVID-19*.

<https://unstats.un.org/iswghs/task-forces/covid-19-and-household-surveys/national-responses-to-COVID-19/>

United Nations Intersecretariat Working Group on Household Surveys. (2020c). *Taskforce on COVID-19 and household surveys*.

Vehovar, V., Solavec, A., & Berzelak, N. (2012). Costs and Errors in Fixed and Mobile Phone Surveys. In *Handbook of Survey Methodology for the Social Sciences* (pp. 277–297). Springer.

Vehovar, V., Toepoel, V., & Steinmetz, S. (2016). Non-probability Sampling. In *The SAGE Handbook of Survey Methodology*. SAGE.

Vezzoni, C., Ladini, R., Molteni, F., Sani, G. M. D., Biolcati, F., Chiesi, A. M., Guglielmi, S., Maraffi, M., Pedrazzani, A., & Segatti, P. (2020). Investigating the social, economic and political consequences of the COVID-19 pandemic: A rolling cross-section approach. *Survey Research Methods*, 14(2), 187–194. <https://doi.org/10.18148/srm/2020.v14i2.7745>

Visser, P. S., Krosnick, J. A., & Lavrakas, P. J. (2014). Survey Research. In *Handbook of Research Methods in Social and Personality Psychology* (Second Edition).

Walker, A., Hopkins, C., & Surda, P. (2020). The use of google trends to investigate the loss of smell related searches during COVID-19 outbreak. *International Forum of Allergy & Rhinology*.

Walker, A. S., Bell, I., Diamond, I., & Benton, P. (2020). *Incidence of COVID-19 (SARS-CoV-2) infection and prevalence of immunity to COVID-19 (SARS-CoV-2) in the UK general population as assessed through repeated cross-sectional household surveys with additional serial sampling and longitudinal follow-up-an Office for National Statistics: Clinical Research Protocol Template version 15.0*.

Webinar—COVID 19—Data Collection in Longitudinal Studies. (2020).

Weisberg, H. F. (2005). *The Total Survey Error Approach: A Guide to The New Science of Survey Research*. The University of Chicago Press.

- Wellenius, G. A., Vispute, S., Espinosa, V., Fabrikant, A., Tsai, T. C., Hennessy, J., Williams, B., Gadepalli, K., Boulanger, A., Pearce, A., Kamath, C., Schlosberg, A., Bendebury, C., Stanton, C., Bavadekar, S., Pluntke, C., Desfontaines, D., Jacobson, B., Armstrong, Z., ... Gabrilovich, E. (2020). Impacts of State-Level Policies on Social Distancing in the United States Using Aggregated Mobility Data during the COVID-19 Pandemic. *ArXiv:2004.10172 [q-Bio]*.
- Will, G., Becker, R., & Weigand, D. (2020). COVID-19 lockdown during field work—Challenges and strategies in continuing the ReGES study. *Survey Research Methods, 14*(2), 247–252. <https://doi.org/10.18148/srm/2020.v14i2.7753>
- World Association for Public Opinion Research. (2020, June 4). *WAPOR June 2020 Webinar—COVID-19: How We Ask Questions, Collect the Data and What Do We Learn?* <https://www.youtube.com/watch?v=m3PmY9TYCqc&t=6s>
- World Health Organization. (2020a). *Population-based age-stratified seroepidemiological investigation protocol for COVID-19 virus infection*.
- World Health Organization. (2020b). *Population-based age-stratified seroepidemiological investigation protocol for coronavirus 2019 (COVID-19) infection Version 2.0*.
- World Health Organization Regional Office for Europe. (2020). *Survey Tool and Guidance: Rapid, simple, flexible behavioural insights on COVID-19*.
- World Opinion for Public Opinion Research. (2020, May 21). *WAPOR Webinar May 2020—Changes in Survey Research in Times of COVID-19*. <https://www.youtube.com/watch?v=ZnpfWPWreac&t=19s>
- Wright, J. D., & Marsden, P. V. (2010). Survey Research and Social Science: History, Current Practice, and Future Prospects. In *Handbook of Survey Research* (Second Edition).
- Yazıcıoğlu, Y. (2020). “Hayat Eve Siğar” Uygulamasına “Detaylı Bilgi” Tepkisi. Amerika’nın Sesi (Voice of America). <https://www.amerikaninsesi.com/a/hayat-eve-sigar-uygulamasina-detayli-bilgi-tepkisi/5378063.html>

Zhang, Y., & Wildemouth, B. M. (2009). Unstructured Interviews. *Applications of Social Research Methods to Questions in Information and Library Science*, 222–231.