The Impact of COVID-19 on Large-scale Phone Survey (Sample) Productivity and Response Rates

Matt Jans[1], James Dayton[2], Randy ZuWallack[2], Don Allen[3], Josh Duell[4], Andy Dyer[2], Thomas Brassell[4], Sam Collins[2], Traci Creller[4], Zoe Padgett[1]
[1]ICF, 530 Gaither Rd., Rockville, MD 20850
[2] ICF, 126 College St Suite 2, Burlington, VT 05401
[3] ICF, 980 Beaver Creek Dr, Martinsville, VA 24112
[4] ICF, Remote

Abstract
COVID-19 has had a complex impact on survey productivity and response rates. Anecdotal reports from a range of organizations and surveys suggest that both productivity and final response rates are up. At least two things may be responsible for this. First, more people are at home. Due to stay-at-home orders and lay-offs, more people are at home full-time than ever in recent history, making households easier to contact and sampled respondents (potentially) more available to be interviewed. Second, more interviewers are making calls from home. Individual interviewer productivity may improve when calling from home instead of from a centralized call center. Certainly, shift staffing, call-back scheduling, and other logistics are much simpler with home-based interviewing. This paper compares sample productivity (i.e., sample outcome rates including contact, refusal, cooperation and response rates) from several ICF phone surveys conducted before and during COVID-19. Surveys were selected to minimize differences in target populations, topics, and calling protocols. The paper addresses the following three questions about the effect of COVID-19 on phone surveys: 1) how is COVID-19 affecting common sample outcome rates, specifically contact, cooperation, response rates, and refusal conversions, 2) can we identify different types of "COVID effects", including the initial *shock* of the pandemic and any ongoing effects, and 3) do surveys using similar protocols show the same effects, or is there variability across surveys? Findings are discussed in the context of COVID-19's impact on data collection, interviewer staffing and productivity, and the general difficulty disentangling interviewer effects from effects of the broader data collection context.

Key Words: response rates, COVID-19, RDD survey, survey productivity

1. Introduction

The success of population-based surveys relies, in part, on certain population habits and preferences, such as population members' at-home patterns, interest in specific survey topics, and tolerance for response burden. The population habits and preferences that we count on to conduct surveys efficiently apply to our interviewers as well as the populations we survey. The COVID-19 pandemic has interrupted everyone's (both sample households' and interviewers') habits and may even be changing preferences and burden tolerance (e.g., work style and expectations, daily habits, time demands, and survey topic interest). Beginning in March 2020, most people in the general population were required to work from home, while others lost jobs because they were neither "essential workers," nor could work from home. Essential workers faced unique challenges at work and home due to COVID exposure risk and social distancing guidelines. Healthcare workers in particular saw a dramatic increase in work hours and degradation of their working conditions. School and day care closures disrupted parenting habits, and teachers had to figure out how to reach their students remotely. Rarely in social or statistical science do we make broad generalizations, but it's safe to say that COVID-19 has affected everyone in one way or another. For most of us, this includes major life disruption. For everyone, health has been top-of-mind. Our field wondered whether sample would be worked more or less efficiently when interviewers dialed from home, as well as whether the general household population would be easier to contact.

From the survey research perspective, these disruptions presented major operational challenges, but also a potential silver lining. Interviewers could not work from centralized call centers, upending work habits and plans that assume physically-centralized work. To follow social distancing requirements, many if not most call centers had to close. Interviewers who were used to working in centralized facilities with close, physical supervision and strictly-defined shifts had to adjust to a different type of day-to-day management and monitoring. Their supervisors had to adjust, too. On one hand, working from home could lead to declines in productivity and work quality. Would interviewers just "phone it in"? Would they be more likely to falsify work? On the other hand, allowing interviewers to work from home afforded them the personal benefits previously only afforded to other types of staff. Further, working from home could have performance benefits, such as flexibility in scheduling call-backs and the ability to make calls at unusual hours. Some staffing logistics would certainly be eased as well. Would known benefits of teleworking (e.g., Madsen, 2003) apply to interviewers?

In addition to changes in interviewers' lives and work, the general population from which we often sample was home more often. Despite new challenges in their lives, such as working and schooling from home, social distancing, and dealing with unemployment and COVID-19 infections within their home or social networks, would they be more available to participate in an interview? Would health surveys be uniquely poised to benefit from the COVID environment?

With all the inherent complexities of separating interviewer effects from sample unit effects (e.g., Schnell and Kreuter 2003), and challenges establishing causal relationships in an ongoing society-wide event like COVID-19, this paper does not claim to answer, or even address all these hypotheses. However, it focuses on several general questions that could be addressed with sample productivity metrics data:[1]

> **Question #1)** How is COVID-19 affecting contact, cooperation, response rates, and refusal conversions?
>
> **Question #2)** Can we identify different types of "COVID effects", including the initial *shock* of the pandemic and any *ongoing* effects?
>
> **Question #3)** Do health surveys using similar protocols show the same effects, or is there variability across surveys?

2. Methods

2.1 General Design and Outcomes

To identify possible COVID-19 effects on survey production, several common phone survey sample productivity metrics were used (see Table 1). We refer to these as "sample productivity" to distinguish them from "interviewer productivity". The former are aggregate metrics of how well a sample is performing, and the latter are person-level metrics of how individual interviewers perform. In summary, a sample's contact rate can be interpreted as a measure of success at the first step of data collection (i.e., reaching the sampled household). Cooperation rate is a measure of success obtaining participation among those who have been contacted and are known to be eligible (i.e., screened). Response rate, which is often difficult to interpret as an operations metric by itself, shows how the whole sample performs with respect to the population. That is, the numerator is the same as cooperation rate (completes and partials), but the denominator includes an estimate of the number of eligible households. Response rates are sometimes weighted, but they are not in this paper. This paper uses an older AAPOR Response Rate 4 (RR4) definition, which only includes one estimate of unknown eligibility in the denominator. Finally, refusal conversions are a metric of how easy it was to gain participation from resistant households.[2]

---

[1] Operational details about how ICF adapted to COVID are discussed elsewhere (Dayton, Penn, Hairston, and Allen, 2021). For perspectives from other data collectors see other presentations at the 2021 AAPOR mini-conference on surveying during COVID-19 (e.g., Danforth, 2021; Jannett, Billington, and Keating, 2021; Johnson, Lynn, and Lancaster, 2021; Keirns and Rees, 2021; Rodriquez, B., 2021; Wright, Trucano, Walter, Allen, Kallaur, Shrestha, and McCoy, 2021; Young, Wivagg, Yan, Carusi, Delnevo, and Gundersen, 2021).

[2] Current AAPOR dispositioning and response rate best practices are always available online at: https://www.aapor.org/Standards-Ethics/Standard-Definitions-(1).aspx

Table 1: Sample Productivity Rates Assessed

Contact Rate: Proportion of eligible and expected eligible sampled numbers that have been contacted

$$\frac{Contacted\ Eligible}{Known\ Eligible + e(Unknown\ Eligible)}$$

Cooperation Rate: Proportion of known eligible phone numbers that produce an interview

$$\frac{Completed\ and\ Partial\ Interviews}{Known\ Eligible}$$

Response Rate (AAPOR 4): Proportion of phone numbers with known and estimated eligibility that produce an interview

$$\frac{Completed\ and\ Partial\ Interviews}{Completed\ and\ Partial\ Interviews + e(Unknown\ Eligible)}$$

Refusal Conversions: Proportion of cases with at least one refusal that become completed or partial interviews

$$\frac{Completed\ Refusals}{Refusals}$$

The four metrics summarized in Table 1 were calculated monthly for 12 ongoing dual-frame (landline and cell) random digit dial (RDD) health surveys conducted across the United States. The 12 surveys employed very similar sample designs, contact strategies, and interview instruments. While the surveys were conducted around the United States, no single survey reported here is national in scope, and the twelve surveys together should not be interpreted as “national results.” All months of 2019 and 2020 were included in the analysis to facilitate comparisons between each month of 2019 and 2020, as well as overall yearly averages and month-to-month change. Only surveys with data in both years were included.

2.2 Analysis approach

The “analysis” for this paper was conducted through a basic “plot and review” process in which sample productivity rates were plotted by month, with a blue reference line placed between December 2019 and January 2020 to denote the change between years, and a red reference line at March 2020 (the month during which most stay-at-home orders were enacted) to help identify shock effects of COVID. This is the month during which the vast majority of ICF interviewers were sent home. Annual average rates were also plotted to assist visual interpretation (visualized with a horizontal dashed red line). No statistical testing was conducted, but plot review focused on identifying the following patterns:

1) "Shock effect" of the COVID-19 pandemic and stay-at-home orders, which was operationalized as a spike or dip in March 2020 (i.e., difference between February and March 2020)
2) Possible "ongoing effect" of COVID, which was operationalized as an increase or decrease through some or all months of 2020 relative to 2019

Shock and ongoing effects were first assessed for all surveys combined (i.e., the average rate calculated based on all surveys in a given month and the annual average across all surveys and months). After that, survey-level plots were reviewed to see if each survey followed the same pattern as the monthly average across all surveys. Trend lines in the survey-level plots were bolded for individual surveys that exhibited both the shock and ongoing effects that were observed in the overall plots. Because no statistical testing was conducted, the analysis should be considered exploratory and informative for future, more rigorous research. Noteworthy for interpreting differences is a sample design change in 2020, specifically doubling the listed-to-unlisted number ratio for all 12 surveys in 2020 (from 2:1 in 2019 to 4:1 in 2020). Listed sample is generally easier to contact than unlisted sample.

3. Results

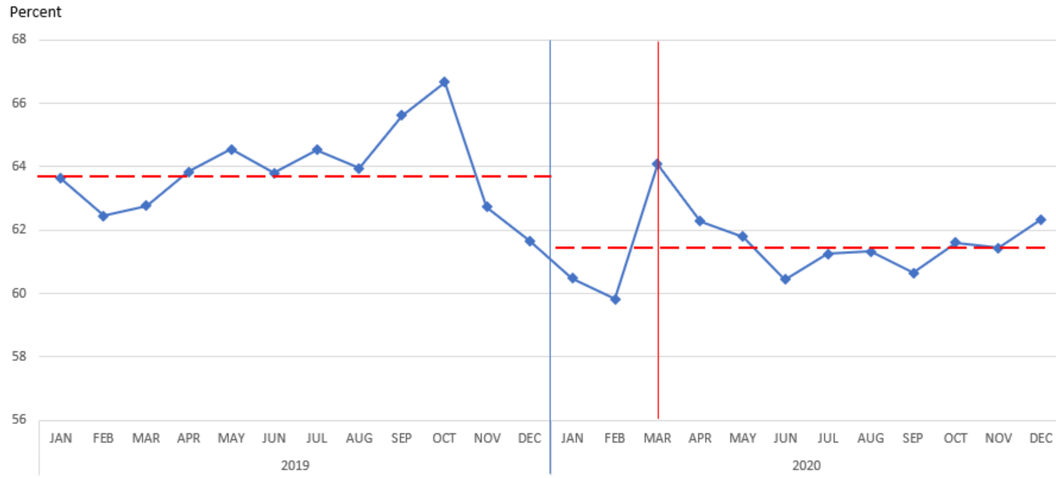
Figures 1a through 4b show the overall and survey-level production rates respectively. Contact rate (Figure 1a) shows a possible COVID-19 shock effect (the spike of about four percentage points in March).[3] In April through December of 2020, however, the contact rate declined and nearly flattened to the annual rate, which was lower than 2019. It would be tenuous to assert this lower rate as a COVID effect given the increased listed-to-unlisted number ratio and its known effect on contact.

Figure 1b exhibited a large amount of survey-to-survey variability in contact rate.[4] Notice that some surveys (e.g., survey 1), saw a dip in April 2021 that quickly recovered. Other surveys (e.g., survey 8) climbed over the first year of COVID.

---

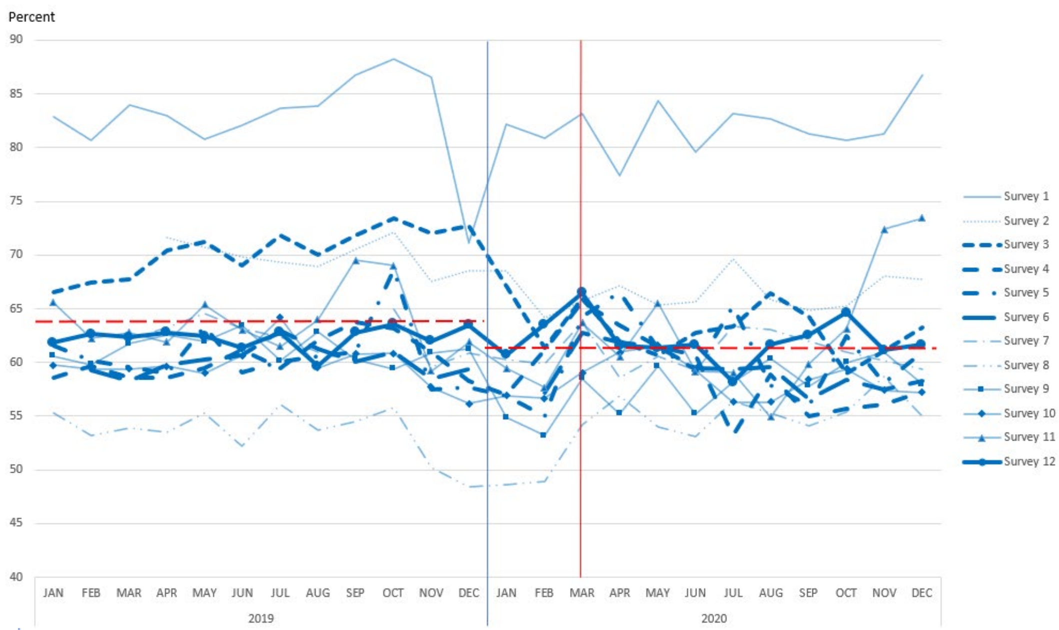
[3] For ease of interpretation the y-axis of each plot us optimized for the data presented in the plot, and thus, varies across plots. Readers should keep this in mind and be sure to consult the numeric categories on each plot's axis.

[4] In Figure 1b and the other survey-level plots, each survey has its own trend line defined by the type of line (solid, dash, etc.) and the type of symbol (or lack of symbol) at each month's data point. After heuristic review, surveys trends that follow the overall trend closely were bolded. Surveys with unique trends have been made more transparent.



2019 v. 2020
Start of COVID work-from-home
Annual Average

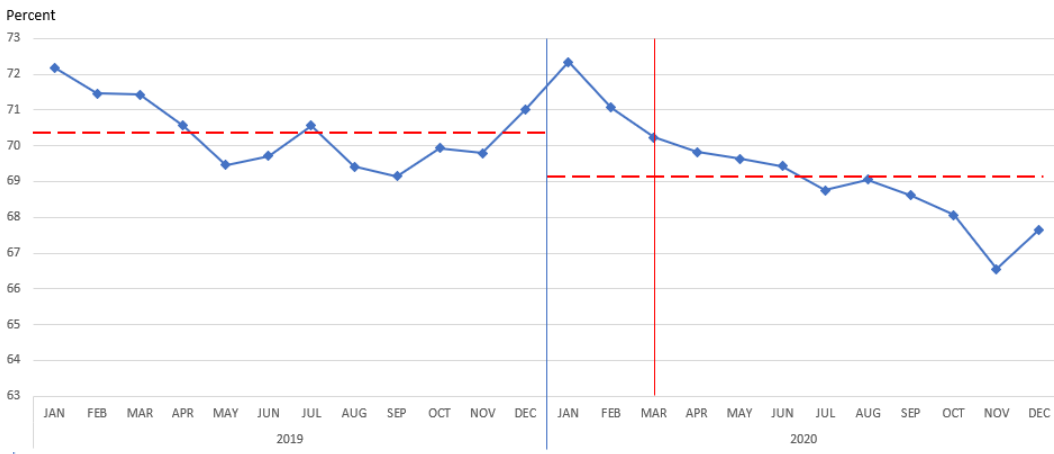
Figure 1a: Contact Rate (monthly before and during COVID)



2019 v. 2020
Start of COVID work-from-home
Annual Average

Figure 1b: Contact Rate by Survey (monthly before and during COVID)

Cooperation rate (Figure 2a) showed no shock effect of COVID, but a clear downward trend during 2020. The initial decline before the pandemic made it difficult to assert an ongoing COVID effect, and whether the higher listed-to-unlisted number ratio affected cooperation. Survey-by-survey variability in contact rate (Figure 2b) showed that some surveys (e.g., survey 6 and survey 12) climbed after COVID began, but most declined. Thus, if there was any ongoing COVID effect, it decreased cooperation, but the effect was not universal.

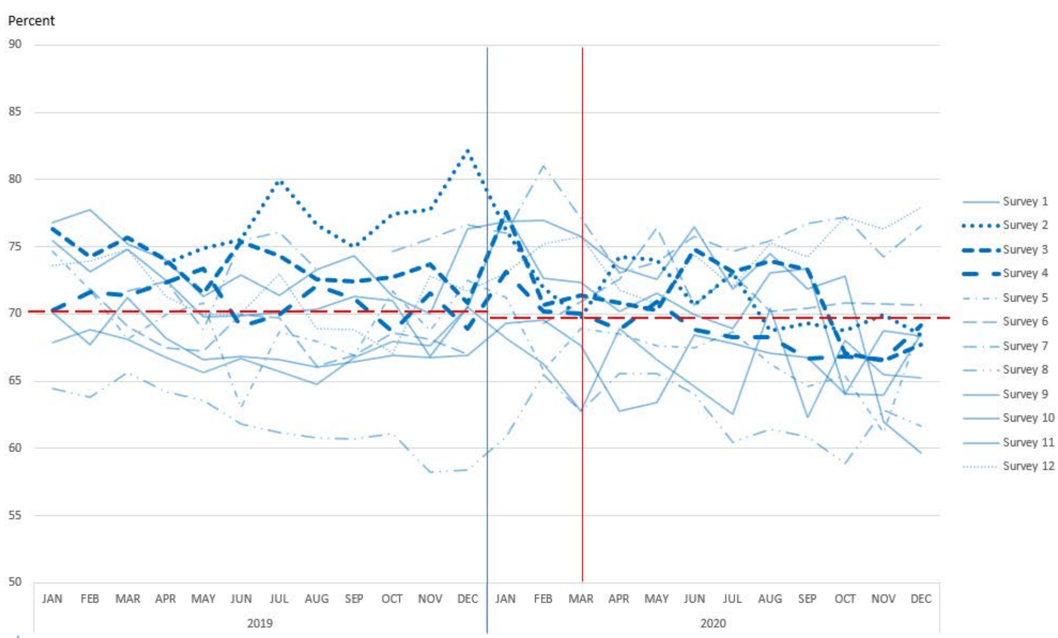


2019 v. 2020

Start of COVID work-from-home

Annual Average

Figure 2a: Cooperation Rate (monthly before and during COVID)



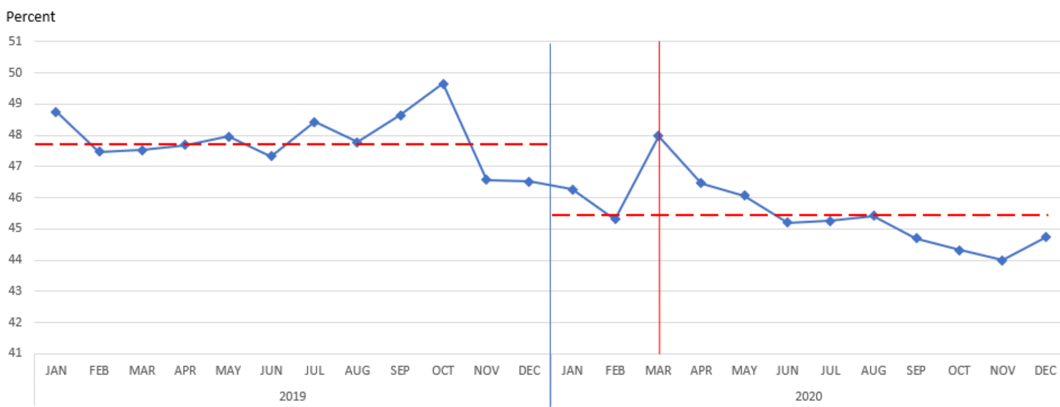
2019 v. 2020

Start of COVID work-from-home

Annual Average

Figure 2b: Cooperation Rate by Survey (monthly before and during COVID)

Response rate (Figure 3a) showed a potential shock effect of COVID and a possible ongoing effect (September-November 2020). The 2020 annual average was below the 2019 average by about 2.5 percentage points), with January and February have a response rate near December 2019. This may be a seasonal effect. Between-survey variability (Figure 3b), emphasized that some surveys (e.g., survey 1) experienced no discernable shock effect or continued effect, while others did (bolded). Further, part of the overall shock effect may be due to survey 1. While its March 2020 response rate was not high compared to the survey's average, its value relative to the other surveys could pull the March average up when combined with small increases in other surveys during March.

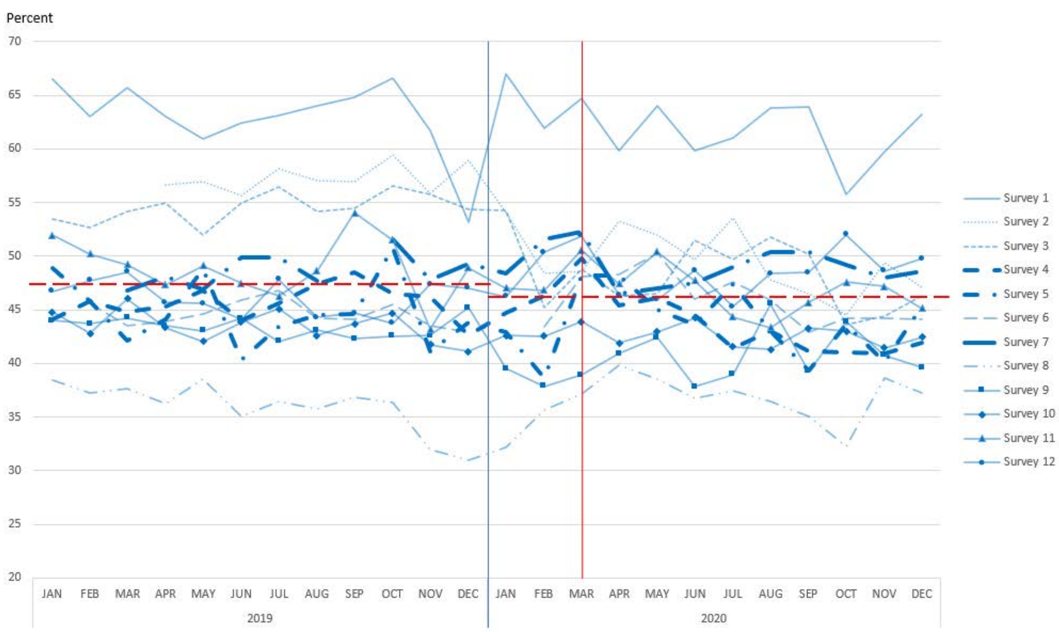


2019 v. 2020

Start of COVID work-from-home

Annual Average

Figure 3a: Response Rate (monthly before and during COVID)



2019 v. 2020

Start of COVID work-from-home

Annual Average

Figure 3b: Response Rate by Survey (monthly before and during COVID)

Refusal conversion (Figure 4a) experienced a clear shock effect of about 6 percentage points in the first month of the COVID pandemic. The first two months of 2020 were almost identical to 2019, so there did not appear to be any effect of the listed-to-unlisted number ratio. Notice that refusal conversion was relatively low on average, so the shock effect doubled the rate. Refusal conversion declined over 2020, leveling out from September through December, and returning to a rate only slightly higher than the 2019 refusal conversion rate (about one percentage point).

The between-survey variability in refusal conversion (Figure 4b) provided further, clear evidence of a COVID shock effect (and possible declining ongoing effect). In addition to the uniform spike across all surveys in March 2020, there was much less between-survey variability than in other metrics. That variability appeared to be lowest in March through May 2020, and increased over the year. Such entropy (i.e., a system returning to a disorganized state) is one way to characterize such behavior in time series analysis (e.g., Ponce-Flores, Frausto-Solis, and Santmaria-Bonfil, 2020). Despite declining toward the 2019 average in April through December of 2020, the 2020 average remained about two percentage points higher than 2019. Notice that the variability across surveys appears to increase (i.e., wider spread in refusal conversions in October through December 2020 than in March 2020). This is expected if the effect of a specific event (e.g., COVID) is immediate and strong. After the initial shock, other influences take over and variability is introduced across surveys.

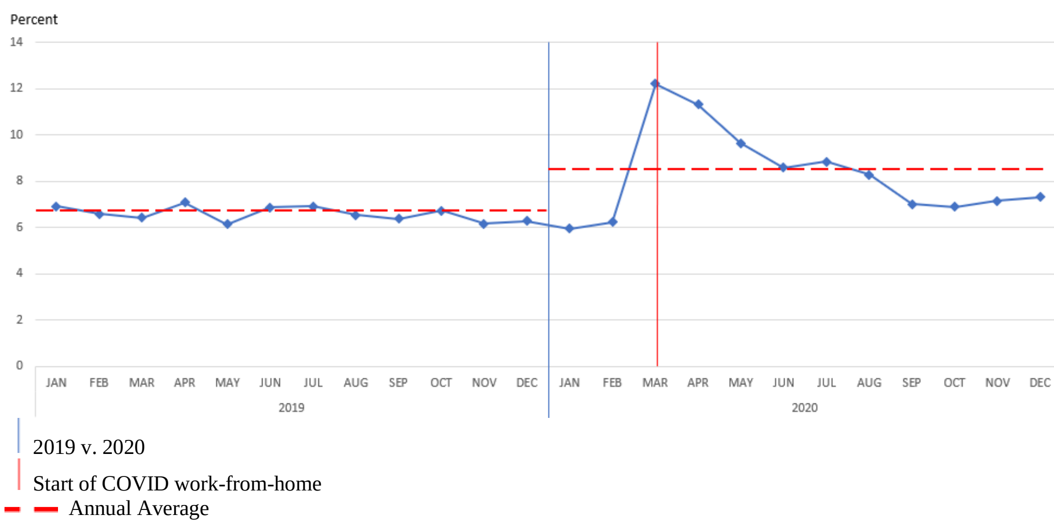
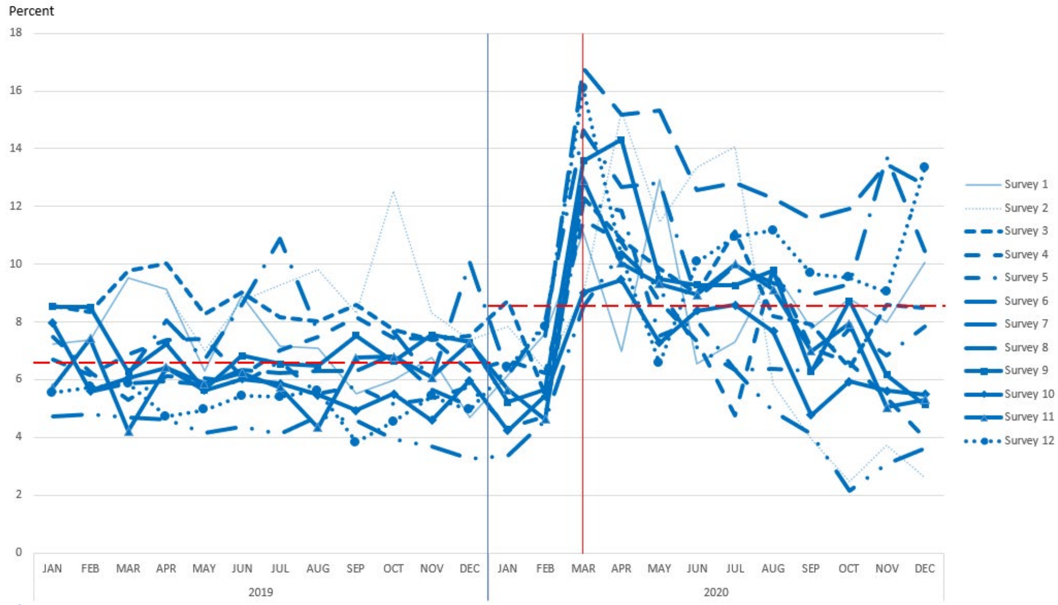


Figure 4a: Converted Refusal Rate (monthly before and during COVID)



2019 v. 2020

Start of COVID work-from-home

Annual Average

Figure 4b: Converted Refusal Rate by Survey (monthly before and during COVID)

4. Discussion

In summary, we saw a potential shock effect of COVID on contact rate, response rate, and refusal conversion. Continued effects were much more difficult to discern due to the nonstatistical "plot and review" method and the confounding listed-to-unlisted number ratio increase in 2020. There was no discernable effect of COVID on cooperation rate, although 2020 was slightly lower than 2019 on average, and ended much lower, a trend which began prior to the COVID pandemic. Across all sample performance metrics, there was high between-survey variability. Refusal conversion showed the lowest amount of between-survey variation, at least by visual inspection, in March 2020 and the months shortly after. However, that initial shock effect diminished over the year. Survey-level variability is another reminder that the COVID pandemic (and the listed-to-unlisted number ratio) are not the only causes of sample productivity. Survey productivity in any month or year is due to multiple cause and can be considered a dynamic system (e.g., Singh, Mourelatos, and Nikolaidis, 2011; Huffaker, 2010).

4.1 Limitations and Reflections for the Future

4.1.1 Importance of the data included

Ignoring interviewer-level performance outcomes certainly hindered our ability to understand the full processes that led to the rates reviewed. We plan to evaluate sample and interviewer productivity once COVID is over, whenever that may be.

4.1.2 Importance of the chosen metrics

The various trends observed in this paper underscore the importance of reviewing multiple sample performance metrics in analyses like these. Some (e.g., response rates) may be more sensitive to methods changes (sampling or otherwise), and unique trends will almost always be found when additional metrics are reviewed. However, reviewing too many metrics can leave the researcher in the position of the proverbial person with 10 watches; mesmerized by the complexity of measuring time, but not sure exactly what time it is. This study started with 12 metrics, which were quickly whittled down to the four presented here due to the uniqueness of the observed trends and the ability to tell a story about various aspects of data collection (i.e., interpretability).

Similarly, other production metrics not presented here (and not in the original 12), such as sampled phone numbers per complete, showed that interviewers were much more productive at dialing in the early months of COVID. While it may seem counter-intuitive to observe increases in interviewer productivity that do not appear in other sample outcome metrics, this is not altogether odd. For example, if productivity is defined as interviewers' ability to dial numbers (i.e., "work sample"), higher productivity would lead to more contacts, but possibly less cooperation as a percentage of contacts. Response rate is one step further removed from interviewer productivity, taking into account eligibility (detailed limitations of AAPOR RR4 discussed below).

4.1.3 Limitations of the "plot and review" method

Reviewing sample outcome and productivity rates is a reminder that summary rates alone do not fully reflect data collection processes. Further, this paper highlights difficulties assessing causation in nonexperimental situations. While no statistical testing was employed, it is also not clear whether sophisticated modeling would have been more informative than the visual review conducted, at least for the scope of this study. Statistical testing and modeling are planned for future analyses. This may include relatively simple statistical tests of the differences discussed in this paper, or more complex modeling, such as multilevel models predicting specific call dispositions with appropriate terms for interviewer, month (including whether pre-, during-, or post-COVID), and survey. Of course, the choice of model depends on future studies' stated goals and specific research questions. For example, predicting call dispositions would be most useful if interviewer-level productivity is a goal. Alternatively, if monthly sample productivity is the goal, each month's sample productivity rates could be conceptualized as predictable values, with the appropriate fixed or random effects for survey, month, etc. While almost anything can be modelled, the most sophisticated analyses are not always the most optimal method for every context. For example, insight would be gained from simply plotting interviewer productivity by month and survey in a similar way to what was done for sample productivity. This could be done at the aggregate level, or by individual interviewers (i.e., profile plots). Further, simple comparisons between interviewer work locations, with basic controls for nonrandomness in that status would help establish causal relationships.

4.1.4 Sample design changes and the importance of consistency in general data collection method (and comment on AAPOR Response Rate 4)

Complicating this analysis was a sample design change implemented in 2020, in which the ratio of listed to unlisted numbers was doubled. In general, listed numbers should be easier to contact than unlisted numbers, so a higher listed number rate should increase contact. Increasing contact, all other things being equal, should decrease cooperation after contact (i.e., a smaller proportion of the contacted phone numbers will participate). The potential for this change to impact refusal rate is not intuitive. However, the impact on response rate may be more straightforward. AAPOR's Response Rate (RR) 4 penalizes a more productive sample (i.e., one in which it is easy to contact households but the rate of not completing contacts is higher). In other words, having more unknown eligibility and nonworking numbers helps increase the traditional RR4 (with one *e* term). Thus, if contact and eligibility screening are higher due to having a greater concentration of listed sample, but the rate of completing interviews with those cases is lower, RR4 will suffer. This explanation matches the pattern of results, in which response rate was lower on average in 2020 than 2019, despite a spike in March. That said, the higher listed-to-unlisted number ratio in 2020 cannot completely explain average differences in sample performance metrics. For example, the cooperation rate was highest in January 2020, contradicting the notion that the increased listed-to-unlisted number ratio alone would decrease cooperation. In fact, the first two months of 2020 (prior to COVID) are very similar to the first two months of 2019. Future research could disentangle this issue.

4.2 Next Steps

In addition to addressing the limitations above, we plan to include 2021 data to see whether monthly trends continue or change as the COVID pandemic evolves and, hopefully, ends. This will allow us to start assessing the overall impact of COVID. Once interviewers begin returning to call centers, we will have several interviewer groups to compare and can conduct within-interviewer comparisons before, during and after COVID. For example, we will be able to compare call center productivity versus work-from-home productivity before, during, and after COVID (including, potentially, interviewers who work from home versus call centers post-COVID). While our "return to normal" plans are not finalized as of the writing of this paper, having home-based and center-based staff working at the same time would allow comparisons we were not able to make in the past. More broadly, we encourage other researchers to continue investigating the impact of COVID on interviewer and survey production on their surveys, particularly surveys of other topics, in other modes, or with other populations. Our field certainly has much more to learn, and results will help prepare our statistical infrastructure for future pandemics and disasters.

References

Danforth, L., Creating and utilizing a remote call center during a pandemic. Presentation at the 2021 AAPOR Annual Conference (virtual). https://aapor.secure-platform.com/a/solicitations/12/sessiongallery/587/application/5137

Dayton, J., Penn, M., Hairston, A., and Allen, D., (2021). Creation of remote call center during COVID-19: Moving from centralized calling to remote calling. Presentation at the 2021 AAPOR Annual Conference (virtual). https://aapor.secure-platform.com/a/solicitations/12/sessiongallery/587/application/5081

Huffaker, R. G. (Ed.). (2010). Phase Space Reconstruction from Economic Time Series Data: Improving Models of Complex Real-World Dynamic Systems. International Journal on Food System Dynamics. https://doi.org/10.22004/ag.econ.97021

Jannett, A., Billington, C., and Keating, C. (2021) Now you see it, now you don't: Adapting visual aids designed for in-person use to interviewing by phone. Presentation at the 2021 AAPOR Annual Conference (virtual). https://aapor.secure-platform.com/a/solicitations/12/sessiongallery/587/application/5045

Johnson, K., Lynn, K., and Lancaster, M (2021). Altered state: Performance impacts of a rapid pivot from on-site to virtual call center. Presentation at the 2021 AAPOR

Annual Conference (virtual). https://aapor.secure-platform.com/a/solicitations/12/sessiongallery/587/application/5128

Keirns, T. and Rees, J. (2021). Differential nonresponse in a high responding COVID-19 surveillance survey. Presentation at the 2021 AAPOR Annual Conference (virtual). https://aapor.secure-platform.com/a/solicitations/12/sessiongallery/579/application/5017

Madsen, S. R. (2003). The Benefits, Challenges, and Implications of Teleworking: A Literature Review. Undefined. https://www.semanticscholar.org/paper/The-Benefits%2C-Challenges%2C-and-Implications-of-A-Madsen/bc31d69850fa83d9ac5d953677c320d7cd14b91d

Ponce-Flores, M., Frausto-Solís, J., Santamaría-Bonfil, G., Pérez-Ortega, J., & González-Barbosa, J. J. (2020). Time Series Complexities and Their Relationship to Forecasting Performance. Entropy, 22(1), 89. https://doi.org/10.3390/e22010089

Rodriguez, B. (2021). Transitioning from in-person training to virtual training in response to COVID-19. Presentation at the 2021 AAPOR Annual Conference (virtual). https://aapor.secure-platform.com/a/solicitations/12/sessiongallery/587/application/4812

Schnell, R., & Kreuter, F. (2003). Separating interviewer and sampling-point effects. http://escholarship.org/uc/item/7d48q754.pdf

Singh, A., Mourelatos, Z., & Nikolaidis, E. (2011). Time-Dependent Reliability of Random Dynamic Systems Using Time-Series Modeling and Importance Sampling. SAE International Journal of Materials and Manufacturing, 4(1), 929–946.

Wright, D., Trucano, T., Walter, K., Allen, K., Kallaur, P., Shrestha, A., and McCoy, M., (2021). Conducting a recurring national health survey during the COVID-19 pandemic: Response patterns and lessons learned. Presentation at the 2021 AAPOR Annual Conference (virtual). https://aapor.secure-platform.com/a/solicitations/12/sessiongallery/579/application/4552

Young, W., Wivagg, J., Yan, T., Carusi, C., Delnevo, C. D., and Gundersen, D. A. (2021). Impact of COVID-19 on survey response rates in a longitudinal study of young adults. Presentation at the 2021 AAPOR Annual Conference (virtual). https://aapor.secure-platform.com/a/solicitations/12/sessiongallery/579/application/4876