REPORT FROM THE AAPOR TASK FORCE ON

"THE FUTURE OF U.S. GENERAL POPULATION TELEPHONE SURVEY RESEARCH"

Paul J. Lavrakas (Chair), Independent Consultant

Grant Benson, University of Michigan

Stephen Blumberg, National Center for Health Statistics

Trent Buskirk, University of Massachusetts Boston

Ismael Flores Cervantes, Westat

Leah Christian, Nielsen

David Dutwin, SSRS

Mansour Fahimi, GfK

Howard Fienberg, Marketing Research Association

Tom Guterbock, University of Virginia

Scott Keeter, Pew Research Center

Jenny Kelly, NORC at the University of Chicago

Courtney Kennedy, Pew Research Center

Andy Peytchev, University of Michigan

Linda Piekarski, Survey Sampling Inc.

Chuck Shuttles, Symphony Advanced Media

EXECUTIVE SUMMARY

The primary goal of this Task Force has been to apprise AAPOR members and others in the public opinion and survey research professions about what is likely to happen in the next decade (and beyond) with telephone surveying of the general public in the United States. Herein there is a main report and six appendices (history, coverage/sampling, weighting, nonresponse, costs, and legal/operational) that provide more details to supplement the main report. Our work was organized to address sampling, recruitment, and data collection issues for (a) surveys that are only conducted via telephone and (b) surveys in which the telephone is one mode, but not the only mode, that is used to sample, recruit, and/or gather data. As part of the Task Force's work, two original surveys of survey organizations in the U.S. were conducted.

The traditional approach for general public telephone surveying was to use only the telephone mode for the entire survey, whereas during the past decade and anticipated into the future, the telephone is being, and will continue to be, used, along with other sampling, recruitment, and/or data collection modes in many surveys.

The Task Force's main findings and implications follow.

Coverage and Sampling via Telephone. Telephone numbers continue to be used to cover and sample the general population of the United States. However, the use of landline RDD numbers seems to be quickly becoming unattractive for most general population surveys, due to the rising costs associated with gaining completions via that frame and the shrinking portion of the population with only landline service. For most surveys in the future that use a telephone frame for the general public, the cell phone RDD frame will be sufficient because it will not lead to a meaningful amount of unit-level coverage error. And, as RDD cell phone completions become more attractive in terms of costs, researchers will be drawn to conducting single frame RDD surveys; an added attraction being that weighting will be less complex. As such, surveys with a given sample size will have lower sampling errors, since design effects due to variability in the weights will be lower than for dual-frame RDD surveys and effective sample sizes will be larger.

The field will benefit from quality research on within-unit coverage for the cell phone frame. Currently, there is little reliable information about the amount of sharing of individual cell phones. More importantly, we are unaware of reliable evidence as to whether the sharing of cell phones leads to nonignorable coverage errors of the general population when the person who answers a cell phone automatically becomes the designated respondent.

Recruitment via Telephone. Telephone surveying is likely to be used for many years to come, especially in mixed-mode surveys that can afford the added expense of using interviewers to persuade those who remain nonresponders after being invited to participate via mail or email. The persuasion that can be carried out by the interviewer is qualitatively different from the persuasion possible via mail, email, or an IVR system. It is likely that human contact differentially motivates different types of respondents, so that having interviewers contact certain types of respondents may bring about nonresponse bias reduction. However, interviewers are known to vary widely in their success in gaining respondent cooperation and in the types of respondents they succeed in recruiting. Thus, surveys using telephone interviewers to recruit

respondents in the future should pay more attention to this interviewer-level variation and the effects it may have on the nonresponse bias in a survey.

Data Collection via Telephone. There should be continued attention to the extent that telephone interviewers create measurement error, especially with survey items that are sensitive or prone to social desirability bias. But there are good reasons to choose live interviewers to gather data; e.g., there is generally less missing data when a live interviewer gathers data.

Another measurement consideration is how being in different locations and/or multitasking while providing data via telephone affects respondent-related measurement error(s), such as straight-lining and other forms of satisficing. Still another is how the audio fidelity, when a respondent is using a mobile device, may contribute to measurement error.

It is also likely that interviewers will continue to vary substantially in the quality of the data that they gather via the telephone. Thus, future researchers who use telephone interviewers to gather some or all of the data for a survey should pay more attention to the amount of this interviewer-variation and the effects it on data quality.

Conclusion. There are many researchers who believe that the telephone as a viable and attractive mode for surveying the general public is dying or is already dead, particularly advocates of nonprobability online surveys. The attraction of nonprobability approaches for conducting survey research is their low cost and very quick turnaround. And there may well be many times when the data from nonprobabilty online surveys are fit for purpose. Unfortunately, too often it is uncertain whether such methods are or are not suitable for a given purpose.

In contrast, there are many clients and researchers who still rely on the telephone for all or at least a part of the surveys that they sponsor and conduct. And, the Task Force anticipates that the telephone will remain an important mode for surveying the general public of the United States for many years to come. We say this for several reasons:

- The RDD cell frame will continue to provide extensive coverage of the U.S. population, yielding little unit-level coverage error for most topics of interest.

- As time passes, more auxiliary data that are accurate enough to be useful will be able to be appended to cell phone frames, to aid sampling, weighting, nonresponse bias investigations, and other analytical needs.

- Samples that are based on probability selection methods can be drawn from the cell RDD frames, thereby providing a measurable amount of sampling error (imprecision).

- Interviewers can be used to help recruit respondents from the initial sample.

- Interviewers can help accurately screen for eligibility, which often is a complex process.

- Nonresponse often will be very high, but there will be continual improvements in the means for studying of nonresponse bias in telephone surveys.

- Interviewers can continue to be used to gather data from cooperating respondents, helping respondents better understand survey questions, motivating respondents to provide accurate responses, overcoming language barriers, and using unbiased probing to elicit detailed responses to open-ended questions. Also, advances in conversational interviewing methods eventually are expected to provide benefits that have not as yet been realized.

- There will be well-established methods for weighting to try to adjust biases in telephone surveying that are associated with noncoverage, sampling, and nonresponse.

However, we expect that over time, fewer and fewer surveys will be conducted using only the telephone for sampling, recruiting, and data collection and that there will be more surveys that use the telephone for some, but not all, of their recruiting and data collection needs; whereas proportionally fewer surveys will use a telephone frame for coverage and sampling purposes.

THE FUTURE OF U.S. GENERAL POPULATION TELEPHONE SURVEY RESEARCH

1. BACKGROUND

1.1 Task Force's Mandate

The primary goal of the Future of Telephone Surveying Task Force, from the time of its establishment, has been to apprise American Association for Public Opinion Research (AAPOR) members and others in the public opinion and survey research professions about what is likely to happen in the next decade (and beyond) with telephone surveying of the general public in the United States.

This AAPOR Task Force on the Future of U.S. General Population Telephone Surveys was established in the summer of 2014. The Task Force soon learned that there would be important new regulatory changes and related turmoil affecting telephone surveying in the United States, which would affect the timing of the Task Force's work. These issues included new federal interpretations of the 1994 Telephone Consumer Protection Act (TCPA), affecting the way that telephone survey organizations in the United States would have to dial cell phone numbers (see AAPOR White Paper on the TCPA at https://www.aapor.org/getattachment/Education-Resources/TCPA/TCPA_FINAL.pdf.aspx, but only available to AAPOR members).

In addition, the Task Force decided to carry out two new surveys to gather important data related to the mission of the Task Force.

These issues, and others, kept the Task Force from completing its work until this year, when the regulatory environment stabilized enough for the Task Force members to be more confident about the future of telephone surveying in the United States, and after both of the surveys were completed and had been analyzed.

1.2 A Brief History of Telephone Surveying in the United States

The emergence of the telephone as a primary mode to conduct general population surveys in the United States had its roots in the 1970s – the time by which household telephone penetration in the United States reached a level that provide adequate coverage of the general population – and then matured in the 1980s. With the publication of Groves and Kahn's (1979) book, *Surveys by Telephone: A National Comparison with Personal Interviews*, enough evidence was provided to give the survey research community (and their clients) confidence that the samples (both initial and final samples) and data quality from a random-digit-dialing (RDD) landline telephone surveys of the general public was of sufficient reliability and validity for researchers to take advantage of the considerable savings in costs and time of data collection compared to in-person surveying. By the end of the 1980s, two books had been published explaining the operational details for conducting RDD landline telephone surveys (Frey 1989; Lavrakas 1987), as well as a book that presented invited papers from a 1987 international conference on various aspects of telephone survey methodology (Groves, Biemer, Lyberg, Massey, Nicholls, and Waksberg 1988).

Thus, for reasons of quality, timing, and costs, landline RDD telephone surveys of the general population of the United States became the preferred mode to survey residents of the United States through the 1990s. These were surveys that (1) sampled all their respondents via the landline RDD

frame (or in some cases via listed telephone numbers), (2) used telephone interviewers to recruit all their respondents, and (3) used telephone interviewers to gather data from all the respondents.

However, starting in the early 2000s, three societal changes that began in the 1990s rapidly changed the dynamics of survey research of the general population of the United States. The first of these was the pervasive use of the Internet by the public. Then came the rapid adoption of cellular telephone service among the general population. And, as these were occurring, there was a continued worsening of the United States public's willingness to participate in social, behavioral, and marketing research studies, including telephone surveys.

Even before cell phones supplanted landline telephones, the use of the Internet had pervaded the lives of most Americans. This rapid change stimulated researchers and their clients in the 1990s to begin to embrace low cost and quick-turnaround survey data collection via the Internet. With this rapidly growing interest in spending research funding on Internet data collection from those sampled mostly via nonprobability online methods – including opt-in Internet access panels (cf. Baker, Blumberg, Brick, Couper, Courtright et al. 2010) – the client base for traditional U.S. telephone surveying began to diminish.

And as these developments were occurring, the public's growing unwillingness to participate in surveys continued to make it more difficult to use the telephone to gain a representative sample of the general public, even when using dual-frame RDD (DFRDD) surveys. It certainly made these surveys more expensive, as response rates for many DFRDD frame telephone surveys in the United States approached single digit values, thus forcing survey organizations to dial more and more numbers to get the number of completions needed for their telephone surveys. (See Appendix A for more details about these changes and the role of AAPOR and its some of its members in identifying new approaches to maintain the viability and value of RDD telephone surveying of the U.S. general public.)

Thus, because of these and other factors, an escalating proportion of survey researchers during the past 15 years have been investigating means other than the telephone to sample, recruit, and collect data from the general population.

But, as discussed in this report, the telephone remains an important surveying mode for many researchers and likely will remain so for many years to come.

2. THIS TASK FORCE REPORT

2.1 Work of the Task Force

Many of the members[1] of this Task Force also served on the 2007-08 and 2009-10 AAPOR Task Forces on telephone surveying. Following the success of those prior task forces, the current Task Force established seven working subcommittees: Coverage and Sampling, Nonresponse, Weighting, Measurement, Operations, Legal and Ethics, and Costs. Most members of the Task Force served on more than one of these subcommittees.

As part of the work of the subcommittees, two surveys were conducted to gather information directly related to the work of the Task Force. One of the surveys was led by Tom Guterbock and Grant Benson and focused on gathering data from a national sample of survey organizations about the costs of landline and cell phone RDD surveying (see Appendix E for details). The other survey

1. Grant Benson, Stephen Blumberg, Trent Buskirk, Ismael Flores Cervantes, Leah Christian, David Dutwin, Mansour Fahimi, Howard Fienberg, Tom Guterbock, Scott Keeter, Jenny Kelly, Courtney Kennedy, Paul J. Lavrakas, Andy Peytchev, Linda Piekarski, and Chuck Shuttles.

was led by David Dutwin and focused on gathering data from survey organizations about their response dispositions and response rates during the past decade for landline RDD and cell phone RDD samples in the United States; see Dutwin and Lavrakas (2016) in Appendix D for details.

2.2 Contents of Report

Main Report. This main report of the Task Force is aimed at providing a relatively brief and easily read document that gives a concise and broad understanding of what to expect about the use of the telephone as a survey mode in the coming decade and beyond. Although the Task Force has focused on general population surveys in the United States, much of what we expect to happen should apply to other regions of the world, especially Europe and Australia, and to the use of the telephone for surveying target populations other than the general public.

As explained in more detail in the next section, the report is organized by addressing the use of the telephone by survey researchers for three related, but separate, purposes:

- First, telephone numbers can serve as a frame that researchers choose to use to sample households and persons.
- Second, the telephone can be used by interviewers (or interactive voice response [IVR] technology) to try to recruit sampled persons and gain their cooperation in providing data for the survey.[2]
- Third, the telephone can be used by interviewers (or IVR technology) to administer the survey's questionnaire to gather data from respondents.

In many future surveys of the general public, the telephone will be used for all three purposes and there will be no other survey mode used for sampling, recruitment, or data collection.

But as already is happening, many other future surveys of the general public will not only use the telephone for one or more of these purposes but also will use other survey modes (mail, in-person, and/or Internet) to help achieve the goals of a particular survey. These choices will be made by the researchers in terms of what mode or modes are most "fit for purpose" for the given study.

Appendices to the Report. The appendices provide more detail for those readers who are interested about certain aspects of telephone surveying than are presented in this main report. In creating these appendices, the subcommittees decided to focus on what is new in the field of telephone surveying since the 2010 release of the last AAPOR Task Force on telephone surveying (Lavrakas, Blumberg, Battaglia, Boyle, Brick et al. 2010).

The appendices for the report include:

- Appendix A on additional history of advancements in telephone surveying in the U.S.
- Appendix B on coverage and sampling;
- Appendix C on weighting;

2. Currently, in the U.S., and with the exception of Federal surveys, no autodialer can be used to recruit those reached by a cell phone number – whether by an interviewer or IVR technology – unless that person has given the survey organization her/his prior consent to be called.

- Appendix D on nonresponse trends;
- Appendix E on costs; and
- Appendix F on legal and operational issues.

3. FRAMEWORK FOR THINKING ABOUT THE FUTURE OF GENERAL POPULATION TELEPHONE SURVEYING

Surveys have three main stages that are used sequentially to yield the data that will be analyzed: (1) selecting a frame from which to draw a sample and drawing the initial sample; (2) devising ways to recruit the sampled persons from the initial sample, then trying to make contact with eligible respondents and trying to gain their cooperation to provide data; and (3) gathering the data from the cooperating portion of the initial sample that is eligible for the survey. As recently as 15 years ago, traditional telephone surveying meant that researchers were exclusively using telephone methods to sample, recruit, and gather data from cooperating respondents. Thus, "traditional" telephone surveys were ones that used the telephone for all three of these purposes and that no other survey mode was used for any of the three purposes.

Nowadays, many traditional telephone surveys are being conducted and will be into the future. But more and more surveys are using mixed modes for one or all of these three stages, with the telephone being only one of the modes being used for at least one, but not necessarily all, of these stages.

It is this organizational framework that structures most of the remainder of this main report and this approach works closely in parallel with the total survey error (TSE) perspective. The Task Force believes that this approach to addressing the future of general population telephone surveying in the United States is a valuable one to help survey researchers think carefully, on a survey-by-survey basis, about the choices they make for using the telephone for all, part, or none of a given survey.

3.1 Coverage and Coverage Error, Sampling and Sampling Error, Weighting and Adjustment Error

In selecting a frame, a researcher will decide, either implicitly or explicitly, whether or not a telephone frame will be used. If the decision is made to use telephone numbers for sampling purposes, then the researcher will next need to decide whether *all* of the sample will be telephone numbers and which telephone frames should be used for the telephone portion of the sample. In the United States, there are now at least three telephone frames available for sampling the general public via the telephone: (1) the cell phone RDD frame, (2) the landline RDD frame, and (3) the listed "Electronic White Pages" (EWP) frame (which is essentially comprised of listed landline numbers).

In 2017, most surveys of the general public of the United States that use telephone numbers for sampling use both cell phone and landline RDD frames. However, there are other surveys that use only the cell phone RDD frame; and still others that use both the cell phone RDD frame and the EWP frame. Essentially no credible U.S. general population survey in 2017 uses only the landline RDD frame or only the EWP frame.

The decision about which frame to use will affect how well the survey covers the general population and the extent to which the survey will have nonignorable coverage error. The frame selection decisions also will affect sampling decisions; for example, in a DFRDD sampling design, researchers will need to decide on sample allocation (i.e., what portion of the final sample will come from the cell phone RDD frame and what portion from the landline RDD frame, and thus how large should be the initial samples drawn from each frame). Furthermore, the frame selection will affect sampling error calculations (given that the surveys being discussed here are probability samples), because the decision will determine the stratification that can be done, the length of the field period and the sample size that can be achieved with a survey's budget. It also will affect the weighting adjustments that will be applied to correct for issues related to coverage and the sampling design.

3.2 Recruitment, Unit Nonresponse, and Unit Nonresponse Bias

When using telephone numbers for all or some of the initial sample, researchers should consider how their interviewers (or IVR scripts) will go about gaining cooperation from those who are contacted for recruitment via telephone, thereby avoiding unit nonresponse whereby no data are gathered from the sampled household or person. As part of the recruitment efforts, researchers also will need to think carefully about the dialing technologies and calling rules used to process the telephone sample so as to maximize the rate of contact with individuals at the sampled numbers. The survey's telephone calling rules may need to vary by the telephone number frame being used to achieve the most cost-beneficial outcomes for making contact and gaining cooperation. Because coverage differs across the telephone frames, each frame will contact demographically and behaviorally different members of the general public. So, for example, people most likely to be contacted via a landline number will be most available at different times of the day and days of the week than will people most likely to be contacted via a cell phone number. Not giving adequate attention to the differences between those most likely to be interviewed via a landline number compared to those most likely to be interviewed via a cell phone number may lead to nonignorable nonresponse bias in the final dataset of a DFRDD survey.

Traditionally, the scripts used by telephone interviewers to recruit respondents (i.e., gain their cooperation) were highly structured and often were expected, or even required, to be read verbatim. But in the past 20 years, it has been learned that more cooperation is gained when recruitment is tailored to the individual to whom the interviewer is speaking (cf. Bauer 2008, Groves and Couper, 1998). Interviewers often are most successful at gaining cooperation from reluctant respondents compared to solely relying on mail, internet or IRV recruitment. Thus, in mixed-mode surveys using a frame of addresses for all of the sample, having interviewers call nonrespondents to previous unsuccessful mail contact attempts may be a cost-effective means for recruitment purposes (regardless of whether data will also be gathered via telephone).

3.3 Data Collection and Measurement Error

When the telephone is used to gather all or part of the data for a survey, researchers should monitor the data quality that is achieved via cell phone respondents versus landline respondents, assuming

both types of telephone service technologies are being used.[3] Currently, there are no recognized best practices regarding how to carry out this monitoring. The quality of the data may also vary by whether respondents are in a fixed location while being interviewed (e.g., sitting at their kitchen table) versus those who are mobile while providing data (e.g., walking from the subway to their office). Similarly, there may be data quality differences between those who are solely engaged in being interviewed and those who are providing data while multitasking, both of which cell phones and cordless landline phones readily allow (Lavrakas, Tompson, Benford, and Fleury 2010; Lavrakas et al. 2010).

The telephone as a mode for gathering data may have potential effects on data quality compared to a self-administered mode. For example, survey questions that offer a list of choices for a respondent in an interviewer-administered survey often suffer from *recency effects* whereby those choices heard last are more likely to be chosen by respondents (cf. Holbrook, Krosnick, Moore, and Tourangeau 2007).[4] These mode effects on data quality are especially important in mixed-mode surveys that use more than one mode to gather data. In such cases, it is important for researchers to try to measure or at least estimate how the data gathered via the telephone differ because they were gathered via the phone instead of via a self-administered mode (cf. Kolenikov and Kennedy 2014).[5] (This, of course, also is true when all data are gathered via the telephone; i.e., researchers should consider how their data may have differed in nonignorable ways had all the data been gathered via a self-administered mode.)

4. CURRENT STATUS AND FUTURE EXPECTATIONS FOR TELEPHONE SURVEYING IN THE UNITED STATES

This section addresses what is now known and what is expected to happen in the coming years regarding the use of telephone surveying methods in the United States. Where appropriate, each section is broken into two parts: (1) Surveys that use only the telephone for sampling, recruitment, and data collection; and (2) surveys that use the telephone as only one of their sampling, recruitment, and/or data collection modes.

4.1 Coverage and Sampling Considerations for Future Telephone Surveying

4.1.1 Telephone Only Frame(s)

The most recent data from the National Health Interview Survey (NHIS) for the first half of 2016 indicate that nearly half (49.3 percent) of U.S. households had only cell phone service (Blumberg and Luke 2016). Furthermore, the cell phone only (CPO) percentage varied considerably by several demographic characteristics: e.g., more than 70 percent of adults aged 25-34 years have only a cell phone; nearly 70 percent of renters (69.7 percent); more than three-fifths (63.1 percent) of adults living in poverty; and more than three-fifths (63.7 percent) of Hispanic adults. At the

3. Demographic and psychographic differences between cell phone respondents and landline respondents need to be controlled for in these data quality comparisons.
4. It also is known that survey questions that offer a list of choices for a respondent in a self-administered survey will suffer from *primacy effects* whereby those choices read first are more likely to be chosen by respondent (cf. Scanlan, 2008).
[5] AAPOR members are reminded that a detailed discussion of how to analyze data gathered for the same survey via different survey modes took place via AAPORNet in early January 2017. There was no consensus on how this should be carried out.

same time, approximately one in 14 U.S. households (7.2 percent) had only landline telephone service. In the first half of 2016, 3.1 percent of U.S. households had no telephone service.

From these statistics, it is clear that the cell phone frame in the United States covers approximately 9 of 10 households in the country and covers nearly 19 out of 20 households with telephone service. However, these statistics for cell phone service coverage somewhat overstate the "practical" coverage of adults in the United States, because not every adult can be reached via a cell phone associated with her/his household (e.g., instances in which only one person in the household has a cell phone and she/he does not allow others in the household to receive calls or otherwise use it). But even with this caveat, coverage of the United States adult population by the cell phone RDD frame is very extensive and can be expected to continue to grow.

Peytchev and Neely (2013) have provided strong rationale for suggesting that traditional RDD telephone surveys in the United States, in which telephone numbers make up the entire sample, will suffer only negligible coverage bias if they are limited only to cell phone RDD numbers; that is, telephone surveys in which no landline numbers are included in the sample. However, this conclusion when applied to the last part of the second decade of the 21st century is likely to be limited to studies on certain topics. For example, a cell-only sample may not be wise for a telephone survey of house-bound elderly aged 65 years and older. So, for the next decade or longer, there may be some surveys conducted solely via telephone sampling that will still need to use a landline frame (and a cell phone frame) to select an unbiased initial sample of their target population. But for most future telephone surveys it appears that adequate coverage of the general population (and most subgroups of the general population) can be obtained with only the cell phone RDD frame.

Furthermore, in instances where future researchers are planning a traditional telephone survey of the U.S. general population and they believe that they need to have a small portion (e.g., 5 percent-10 percent) of their initial sample composed of landline numbers, it may be adequate to use the much more efficient EWP frame than the landline RDD frame.

In addition to the RDD frames and the EWP frame for telephone surveys of the general population (or substantial subsets of the general population) of the United States, there are some other telephone frames that might be fit for purpose for a given survey. One of these frames comes from voter registration lists that are available in certain states that include telephone numbers for citizens of those states who are registered to vote. These frames would support surveys and polls that have registered voters as their target population. However, with such frames, there may be nonignorable coverage error associated with missing and incorrect telephone numbers of the registered voters in the geo-political area(s) being surveyed. Furthermore, such frames will undoubtedly contain cell phone numbers and U.S. researchers will need to be able to determine which are cell phone numbers and thus must be dialed without the use of an auto-dialer.

There is a new frame that may prove to be valuable for sampling members of the general population in the United States via their telephone number. In 2017, this is called the "RICS frame" and represents telephone numbers of people who have misdialed numbers and reached nonworking numbers, or have dialed other numbers that are not in service or otherwise do not connect with anyone. RICS stands for "redirected inbound call sampling" Of note, there are an estimated 6 billion RICS calls each month in the United States (Levine, Krotki, and Bobashev 2016). The organization that pioneered this sampling concept (Reconnect Research) has a business agreement with many telephone service providers (e.g., AT&T and Verizon), and these providers can redirect their RICS calls to the organization, which in turn can serve up an IVR invitation to the caller to participate in a survey right then and there. This approach to gaining completed

telephone survey questionnaires is low cost and data are provided very quickly, and it appears to in no way violate any TCPA-related restrictions. The use of RICS in the United States creates a nonprobability sample of telephone numbers, and it remains unknown whether the approach has nonignorable coverage bias of the general population. But early research with this sampling approach shows sufficient promise to merit additional research to further investigate under what circumstances the RICS frame would be fit for purpose for a given survey's needs (Levine et al. 2016). Time will tell whether the promising early findings with RICS telephone sampling are due to the novelty of the approach and whether when that novelty fades the quality of the survey datasets produced by the RICS approach erodes.

Apart from how well the various telephone frames cover the U.S. general population there is a legitimate concern about another coverage issue for which very little appears to be known. The issue is that of within-unit coverage with the cell phone RDD frame.

Heretofore, researchers have assumed *and acted* as though a cell phone number in the United States is exclusively a one-person number and that the person who answers it is the one person who should be recruited into the survey being conducted. However, studies conducted during the past few years by the AP-NORC Center for Public Affairs Research (using large national DFRDD surveys) consistently have found that approximately 40 percent of cell phones are reported to be shared with at least one other adult on a weekly basis.[6,7] This suggests that much more research is needed on this issue – and on the issues of how to ask reliable questions about cell phone sharing. That information is needed because this within-unit selection issue could be creating a nonignorable coverage bias of adults when using the current approach for telephoning cell phone RDD numbers in the United States. Furthermore, if it were decided that within-unit selection is needed for cell phone surveys, it will increase nonresponse and raise costs. In addition, the impact on coverage will depend on whether the sampling design defines the sampled units as individual adults or as households.[8]

4.1.2 Combining Telephone Frame(s) and Nontelephone Frames

It is not expected that there will be many instances in which a general population survey in the United States will use a telephone frame for drawing only a portion of the survey's initial sample. However, if such a survey were to be planned, the researchers should take into account the issues covered above regarding the coverage and sampling of the general population via only the telephone. Thus, in most instances, the cell phone RDD frame would prove adequate, but if landline numbers were deemed necessary (as they might be for a locally targeted sample in a small community area), then the EWP frame would likely be fit for purpose.

6. The wording of the question that has gathered those data is: *How many adults, in addition to you, carry and use this cell phone at least once a week or more?*

7. These findings are for cell phones used by people who have completed these AP-NORC surveys. Thus, it is possible, though unlikely, that the findings do not generalize to the larger population of persons in the U.S. who are the sole or primary user of a cell phone.

[8] Other issues related to the sampling of cell phones and their users that merit additional research include how often cell phones are turned off and how business cell phones affect the selection probabilities of individuals.

4.2 Recruitment and Nonresponse for Future Telephone Surveying

4.2.1 Telephone-Only Recruitment Surveying

Evidence from the Task Force's study of general population DFRDD survey response rates documents that U.S. survey organizations have experienced a significant drop in response rates, by as much as a factor of four, in the past 15 years (see Dutwin and Lavrakas, 2016; included as Appendix D).

But the general trend in declining RDD rates goes back even further. Curtin, Singer, and Presser (2000) and Curtin, Presser, and Singer (2005) measured response over time in the University of Michigan's monthly Survey of Consumers[9], finding that landline RDD response rates were at 72 percent in 1979 and then had a linear decline to 48 percent by 2003. Ten years later (2013), the response rate had dropped to 16 percent. Many others researchers have noted the same trends in other RDD studies (e.g., Tourangeau and Plewes 2013). For example, Pew (2012) reported on trends in its own RDD studies, with a response rate decline from 35 percent to 9 percent from 1997 to 2012.

The survey that the Task Force conducted of recent cell phone RDD and landline RDD response rate trends for survey organizations in the United States suggests that DFRDD surveys are continuing to see response rate declines. Landline rates declined from an average of 15.7 percent in 2008 to an average of 9.3 percent in 2015 (a relative decline of 41 percent), and cell phone response rates declined at the same rate, from an average of 11.7 percent to an average of 7.0 percent (a relative decline of 40 percent).

Regarding refusal rates, surprisingly, there is evidence of only a small increase in refusals for landline RDD surveys, from about 50 percent in 2008 to 55 percent in 2015; and refusal rates stayed basically unchanged during that eight-year time period for cell phone RDD surveying (with percentages in the low 40s).

In contrast, making contact has grown more difficult, with a slight upward trend in answering machine/voice mail rates and in no answer rates. No answer/answering machine/voicemail rates have increased 10 percentage points in the past eight years for landline RDD surveys and a staggering 24 percentage points for cell phone RDD surveys. This large growth in cell phone voicemail outcomes may be associated with voicemail becoming a more frequently used (and "hidden") way of refusing unwanted incoming calls on cell phones. And this, in turn, may explain why cell phone refusals rates have not changed in recent years. The nonworking rate for landline RDD surveys has increased from 28 percent to 40 percent in the study's timespan (a relative *increase* of 43 percent), whereas for cell phone RDD surveys, the rate has dropped from 39 percent to 24 percent (a relative *decrease* of 38 percent).

Overall, the reduced contactability of landline households using the landline RDD frame is likely to continue to grow because, as households continue to give up their landlines, the number of nonworking numbers in valid 1+ landline telephone banks is growing. So, looking to the future, fewer surveys are likely to use the landline RDD frame, if for no other reason than it will cost more to make contact with eligible persons/households via a landline number. This, in turn, should make the EWP frame more attractive when landline phone numbers are needed for sampling and recruitment purposes (Guterbock, Diop, Ellis, Holmes, and Le, 2009, 2011).

The Task Force's study of survey organizations also gathered data to address the number of sampled RDD phone numbers needed for a completed interview. This amount has increased from

[9] Formerly known as the Survey of Consumer Attitudes (SCA).

an average of about 14 sampled landline numbers needed to attain a single interview in 2008 to an average of about 46 sampled landline numbers needed in 2015 (i.e., a relative increase in the amount of landline sample needed of more than 300 percent). For cell phone RDD surveys the trend is similar but not so extreme: In the 2008–2011 period, there was an average of about 17 sampled cell phone numbers needed for one completion, compared to 36 needed in the time period spanning 2012–2015 (i.e., a relative increase in the amount of cell phone sample needed of approximately 118 percent).

Also related to the issue of contactability rates for landline RDD and cell phone RDD surveying are the *calling rules* that survey organizations use to process the phone numbers sampled for a given survey.[10] There appears to be little evidence available about the need for differential calling rules depending on the type of telephone number frame and which numbers are being processed. But logic suggests that people differ quite a bit in how they use cell phones and landlines, and because of this the calling rules should be tailored to the type of number being called. However, anecdotal experience suggests that not many telephone calling centers in the United States have taken that tailored approach to determining their calling rules. In the future, research needs to be conducted to identify best practices in calling rules for general population telephone surveying in the United States.

Another factor that affects response rates for surveys that use the telephone for all recruitment and data collection is the level of success that interviewers have in gaining cooperation when they make contact with a person at a sampled phone number. Research has attempted to identify ways to improve success in recruiting respondents sampled via a telephone number. This includes research into gaining cooperation at households by tailoring the interviewer's approach, including with households contacted via phone that were sent advance contact communications where an address is accurately matched to the sample phone number (e.g., Conrad, Broome, Benki, Kreuter, Groves et al. 2013; Lavrakas, Ward, Geng, Welch, Skalland et al. 2015; Maynard and Schaeffer, 1997). In addition, research has examined how interviewers can more effectively tailor their recruitment approaches to the different sampled households/persons (e.g., Lavrakas, Kelly, and McClain 2016; Maynard and Schaeffer 2002). This research also includes the use of response propensity modeling to devise tailored recruitment strategies for sampled telephone numbers based on information that is linked to the particular telephone number itself. Heretofore, these approaches have mostly been explored for landline numbers in the United States. But societal changes related to the continued adoption of a CPO-lifestyle and technology changes in the ability of vendors to accurately match auxiliary data to cell phone numbers is likely to make the use of these approaches more fruitful when dialing cell phone numbers in the coming years.

Furthermore, a recent announcement by the U.S. Federal Government has lifted the restriction on dialing cell phone numbers (for which the owner has not given prior permission to be contacted) via an autodialer if the telephone survey being conducted is for a Federal Government agency. This will make calling the cell phone portion of these surveys much less costly. (Whether this easing of restrictions on the way that cell phone numbers can be dialed in the United States will eventually apply to more or all types of legitimate research surveys is uncertain at this time.)

Apart from using live telephone interviewers to recruit sampled respondents, it is possible for telephone surveys to use IVR to recruit sampled respondents but the Federal Government

10. Calling rules represent the logic programmed into a computer-assisted telephone interviewing (CATI) system to control when a telephone number is dialed. This logic includes what time of day and day of week the number is dialed on its first call attempt and on subsequent call attempts. Calling rules also control how long a time lag is assigned to a given number before it is called again, which may be based in part on the call history outcomes from previous dialings to that number (cf. Stec, 2008).

restrictions mentioned above also apply to any survey that tries to recruit via IVR. That is, cell phone numbers cannot be dialed via an autodialer if the survey is sponsored by anyone other than the Federal Government, unless the cell phone owner has given the caller prior permission to do so. Furthermore, little is known about under what circumstances recruitment via IVR is cost-effective in balancing survey costs with survey quality (cf. Corkrey and Parkinson 2002).

4.2.2 Recruitment via Telephone and other Mode(s)

In surveys that use the telephone as one of the modes for gaining cooperation from sampled households and persons, the frame for the survey will often not be a telephone frame. Many times, it will be an address-based frame. Furthermore, in many of these mixed-mode surveys, the telephone will not be the first mode or even the primary mode that is used for recruitment. Instead, some form of mail contact will be used first and likely will be the primary mode for recruitment.

Research has shown that telephone interviewers who have become successful in recruiting respondents using one approach will have difficulty learning to be successful with a new approach (cf. Burks, Camayd-Freixas, Lavrakas, and Bennett 2007; Lavrakas et al. 2016; Shuttles, Hoover, Welch, and Lavrakas 2002). Recent experience has shown that many interviewers experienced in trying to recruit persons whose phone number was sampled from a telephone frame will have problems converting their recruitment approaches to a study in which a telephone number is being called to try to reach someone at an address that was sampled from a non-telephone frame. These are cases in which the sampling unit (an address) has had a telephone number matched to it. Reliable evidence is not available on the accuracy of these phone-to-address matching processes that survey vendors perform. But two things are known: First, for many reasons, a sizable minority of matches of phone numbers to addresses are incorrect, and second, the accuracy for such matching in the United States is much higher for landline numbers than for cell phone numbers.

Because of these issues, when a telephone interviewer reaches a person at a telephone number that has been matched to a sampled address in the United States, the person/household will often *not* be eligible for the survey because she/he does not live at the sampled address. Thus, the first priority of the initial contact that an interviewer makes in such a mixed-mode survey will be to determine whether the correct address has been reached. Thus, the interviewer's first priority will not be to try to gain the cooperation of the person being spoken to for data collection purposes. This is especially challenging for many interviewers because it is well known that a large portion of telephone refusals occur within the first 10-15 seconds after someone answers the phone (cf. Dutwin, Loft, Darling, Holbrook, Johnson et al. 2014). Asking about eligibility at the very start of contact, especially if no effort is made to build some rapport with the person to whom the interviewer is speaking, often leads respondents to quickly refuse.

Screening for address-eligibility is not what many experienced telephone interviewers are used to doing at the start of contact. Instead, they are oriented to doing what is needed to quickly try to gain cooperation. Experience and past research suggests that (re)training is needed to help these interviewers "unlearn" their old expectations and habits (cf. Burks et al. 2007; Lavrakas et al. 2016). Special introductory scripts also are required to screen these households to determine whether they are in fact the household at the address that was sampled. Only after it has been confirmed that the correct address has been reached can the telephone interviewer begin the process of gaining cooperation to gather data. Telephone survey organizations also may need to revise the "reward structure" for their interviewers to credit an interviewer for a positive outcome when she/he successfully screens out people reached at a wrong address.

Furthermore, the telephone interviewer in these mixed-mode surveys may not be gathering data from everyone (or possibly anyone) found at the correct address and who is eligible for the survey. Instead, the interviewer may instead be trying to persuade the eligible respondent to go to a website to complete the questionnaire online. So, these types of surveys also may require special training for the interviewers and special scripts devised to achieve these data collection goals.

Another form of mixed-mode surveying that uses the telephone for only a portion of recruitment will very likely increase in frequency in the future. These are surveys that select a sample from a telephone frame and then follow-up nonresponders to telephone recruitment by mailing recruitment (and possibly data collection) materials to addresses that can be matched to the previously nonresponding phone numbers. Fowler, Roman, Mahmood, and Cosenza (2016) have recently reported a study like this in which they increased the response rate by 34 percentage points for a general population survey that used RDD phone numbers for their initial sample. In these surveys, researchers must gather information from the household to make certain that the address that was reached matches the telephone number that was initially sampled.

In sum, it is very likely that telephone recruitment in the United States will continue to be used well into the future in general population mixed-mode surveys.

4.2.3 Differential Nonresponse Associated with Telephone Interviewers

This is yet another area of telephone survey research for which very little appears to be known and therefore more research is needed on it. It encompasses the considerable variation that often occurs between telephone interviewers in terms of (1) the response rate they achieve in a given survey and (2) the nature of the nonresponse that is associated with them. The latter refers to the likely possibility that the make-up of the nonrespondents associated with one interviewer differs from the make-up of nonrespondents associated with another interviewer. These interviewer variations may affect the nonresponse bias in a given survey and more attention should be given to their study (cf. Tarnai and Moore 2008; West and Groves 2013).

4.2.4 Calculating Telephone Surveying Response Rates

When the telephone is the only mode that is used for sampling, recruitment, and data collection, calculating response rates is relatively straightforward. It is more complex for dual frame telephone surveys because rates must be calculated for each frame and then also for the dual frames combined (cf. AAPOR 2016).

When the telephone is used as part of recruitment along with other modes, response rate calculations become more complex. This is especially true for sampled cases that have been recruited via multiple modes (e.g., mail, telephone, and in-person). In those instances researchers are advised to familiarize themselves with AAPOR's guidelines on mixed-mode survey response rates (AAPOR 2016).

4.3 Weighting in Future Telephone Surveying

4.3.1 Telephone-Only Surveying

Appendix C provides considerable detail about issues related to the weighting of traditional telephone surveys for which phone numbers make up the entire sample, and all recruitment and data collection takes place via telephone.

During this and the previous decade, the inclusion of cell phone RDD numbers has offered an effective remedy for improving coverage compared to the traditional landline-only RDD samples of the past. However, the current practice of DFRDD is subject to potential methodological problems. Most survey researchers rely on ad-hoc assumptions to determine the allocation of landline and cellular numbers in their DFRDD samples. This practice, which is mostly due to unavailability of current counts of CPO households, has implications for both sample selection as well as subsequent methods used to weight the resulting DFRDD survey data (cf. Fahimi 2014). Furthermore, as noted in Section 4.1.1, the issue of within-household coverage has not been given adequate attention as it relates to the weighting of cell phone RDD samples.

Many survey researchers contend that because the traditional sampling methods, including DFRDD, are subject to such high levels of nonresponse, the resulting data can no longer be treated fully as probability-based samples. Accordingly, investigations are being carried out to examine the viability of sampling options that do not conform to the orthodox Neyman paradigm of statistical inference (cf. DiSogra 2011). Moreover, different methods of weighting of survey data are gaining popularity – ones that go beyond the traditional geodemographic adjustments. In particular, relying on calibration adjustments techniques popularized by Deville and Särndal (1992), these new techniques support the need to include an expanded set of auxiliary variables in computing survey weights.

Arguably, such auxiliary variables can include attitudinal and behavioral measures for which reliable benchmarks are available. In fact, this view suggests that variables included in the weighting process should include those that are highly correlated with key outcome measures – again, so long as their corresponding benchmarks are secured from reliable sources. While inclusion of more variables during weight adjustments can increase the resulting variability of the weights, thereby increasing the survey's design effects and reducing its effective sample size, this trade-off may be justified. After all, when a survey experiences over 90 percent nonresponse, the corresponding pool of respondents may be too skewed to be balanced adequately with only simple geodemographic weighting. Thus, more aggressive weighting/calibration adjustments may be all but inevitable for future telephone surveys to compensate for the growing rates of differential nonresponse.

It will be important in the future to conduct further research on three issues. First, it is crucial to develop sampling methods to improve coverage. In particular, telephone surveys should capitalize on advances in the telecommunication technologies and become progressively open to methods of sampling other than DFRDD sampling. Second, new methods are needed to supplement the existing statistical machinery that surveys have relied on for decades. Undoubtedly, this will include more robust methods for weighting and calibration for surveys that are subject to high rates of undercoverage and nonresponse. And finally, the preceding two initiatives will become even more important as survey researchers are expected to do more with fewer resources. Future investigations should also focus on identifying alternative methods of sampling and data collection that are more cost-effective. It also should be noted that currently there is no general consensus on the best way to weight for nonresponse in telephone surveys. This

was already recognized by the 2010 AAPOR Cell Phone Task Force, but despite many recommendations, there are still many variants; see Appendix C for more discussion of these issues).

4.3.2 Telephone in Mixed-Mode Surveying

Using a telephone frame with an address-based or other non-telephone frame is a very rare design choice in survey research and is likely to remain very rare. However, if a survey were to use a telephone frame as only one of its sampling frames, all the issues discussed above come into play. Furthermore, researchers will need to devise ways to accommodate the complexities of mixed-frame surveys (apart from DFRDD ones) when they approach the weighting that is needed (cf. Brick and Lepkowski 2008). However, we do not envision that these types of survey designs – for example, a survey that would use RDD phone numbers and household addresses for their sampling frames – will be used much at all in the future.

5. DATA COLLECTION AND DATA QUALITY

Regardless of whether data for a survey are gathered only by telephone, or if the telephone is simply one of multiple data collection modes for a particular survey, the data gathered by telephone are interviewer-administered, unless IVR is used.

Interviewer-Related Measurement Error. Interviewer-administered data collection has its advantages and its disadvantages vis-a-vis the reliability and validity of the data. This report is not the place to review those strengths and weaknesses, but Schaeffer, Dykema, and Maynard (2010) and Dillman, Smyth, and Christian (2014) provide good overviews of the issues involved in the decision to use telephone interviewers as a data collection mode for a particular survey.

It is clear that there can be (and most often is) great variation across telephone interviewers in how they affect measurement bias and variance. Matters such as the hiring, training, and monitoring of telephone interviewers will remain very important considerations in the future whenever the telephone mode is used for data collection because they always will affect the quality of the data that a group of interviewers produces.

Telephone Audio Fidelity. The quality of data gathered via the telephone also will depend upon the quality of the telecommunications technologies of the phone and the telephonic service connecting an interviewer or an IVR system with a respondent. In the United States, in 2017, the quality of cell phone service, on average, remains inferior to the quality of landline phone service. These service quality differences may differentially affect the quality of the telephone data that are gathered for a DFRDD survey. This issue should continue to be of concern to conscientious telephone survey researchers for several years going forward. Only time will tell whether – and when – cell phone service will achieve the same (or better) quality of audio fidelity that is now commonplace in the United States for landline service. This issue merits continued research into the future; for example, data could be gathered in a survey about the quality of the telephonic connection from the respondent's standpoint and from the interviewer's standpoint.

Respondent-Related Measurement Error, Respondent Location, and Respondent Multitasking. The location in which a respondent is completing a telephone interview, whether administered by an interviewer or via IVR, and the activities in which a respondent is engaged while completing the interview via telephone, very likely affect the quality of the data that are

gathered by telephone (Ward, Reimer, Elam-Evans, Yankey, and Khare 2014; Ward, Reimer, Khare, and Black 2015). Findings have been reported regarding the varied activities (including many that are cognitively distracting) in which a substantial minority of respondents engage while completing a telephone interview, especially respondents contacted via a cell phone or a cordless landline phone (e.g., Lavrakas et al. 2010). Respondents completing an interview on a cell phone are able to engage in a wider range of these distracting activities while they are being interviewed than are landline respondents. But given the wide prevalence of cordless landline phones, many persons complete a telephone interview via landline while they also are engaged in other distracting activities. This issue merits continued future research to assess how these respondent behaviors impact the quality of data that they provide via telephone.

Mode-Related Measurement Error. Data gathered via telephone interviewers can differ substantially from data gathered via a self-administered questionnaire, whether a mail-back questionnaire or a self-administered ACASI, CASI, CAWI, or PAPI questionnaire (cf. Kreuter, Presser, and Tourangeau 2008; Krosnick, Presser, Fealing, and Ruggles 2012) Sensitive questions, questions prone to social desirability bias, and choose-all-that apply formats are examples of where data gathered via a telephone interviewer may be materially different from data gathered from the same person for the same measure via a self-administered mode. To this end, researchers who have some or all of a survey's data gathered by telephone interviewing or via an IVR system, now and in the future, should explicitly consider whether mode-related measurement error associated with telephone data collection has affected (biased) the data enough to merit modifying the conclusions that are drawn from such data (cf. Kolenikov and Kennedy 2014).

IVR Data Collection. Some telephone surveys use IVR technology to gather data.[11] The quality of these data now and in the future will, in part, be a function of the quality of the phone line and telephone system via which respondents are providing answers and the quality of the IVR technology that is being used. The bigger issue, however, is whether or not IVR data collection is fit for purpose with a particular survey. IVR limits the nature of the question wording that is practical to use in a questionnaire, it limits the nature of the response alternatives for a question that can be offered to respondents and it limits the length of the questionnaire (cf. Currivan, 2008a). When the question wording and response alternatives used in a survey questionnaire are appropriate for IVR data collection, this telephone mode of data collection should not add substantial measurement error.

6. COSTS OF TELEPHONE SURVEYING

Regardless of whether a survey uses the telephone to recruit and/or gather data from all or some respondents, the type of telephone frame used to sample the respondent will affect survey costs. And, as reported below, there are many factors related to telephone survey costs that are continuing to evolve – some that lessen the cost differential between cell phone and landline surveying, and others that increase that differential. But overall, the differential is closing and is expected to continue to close in future years.

The Task Force conducted a survey of telephone survey organizations in order to gather detailed data about the costs of their DFRDD surveys. The details about the methods used for this

11. Again, we note that autodialers combined with IVR cannot be used in the United States to recruit anyone reached via a cell phone number unless the owner of the cell phone number has given prior permission to the survey organization to call her/his number or unless the survey is sponsored by a Federal agency.

study and the findings are presented in Appendix E. What follows are the high points of the findings and their implications for the costs of conducting future telephone surveying of the general population in the United States.

The working number rate in DFRDD surveys is a function of the working number rate in each frame – that is, in the number blocks from which the initially designated landline and cell phone samples are drawn. The difference across frames in the working number rate may vary in a given sampling area. With the continued migration of the U.S. population to CPO service, the working number rate in the landline frame is falling, whereas the existing cell phone number banks are continuing to fill up with subscribers, thereby increasing the cell phone working number rate. As reported by Dutwin and Lavrakas (2016; see Appendix D), there was a substantial *increase* in the percentage of non-working numbers in the landline frame and a substantial *decrease* in non-working percentages in the cell phone frame. These changes are likely to continue and thus the differential costs between processing landline numbers and cell phone numbers will lessen compared to previous years when interviews with cell phone numbers were generally twice as expensive as interviews with landline numbers.

Cell phone-only users are more likely to answer calls to their cell phones than are dual service users (Guterbock 2009; Brick, Flores Cervantes, Lee, and Norman 2011). Therefore, as the proportion of CPO users increases within the cell phone user population, contact rates for cell phone sample might be expected to increase, if other countering forces where not to occur. Furthermore, it is becoming more common for dual-service phone users to ignore calls to their landline phones. In addition, dual-service users with cable/VoIP/DSL packages (TV, Internet, and phone) will have a telephone number that may or may not actually be connected to a telephone, resulting in persistent No Answer or Busy dispositions. These trends will further reduce the landline contact rate.

However, call screening/caller ID technology is essentially universal on cell phones, as is voicemail; both are thought to promote greater screening of incoming calls by respondents. The result may be a lower live contact rate for sample cell phone numbers. It may be that, in the past, cell phone users expected any call to their cell phone to be from someone they knew and they may have readily answered all calls to their cell phones based on this expectation. If cell phone users are now experiencing more unsolicited calls from unknown parties, including survey organizations, they may be becoming more selective in which calls they choose to answer. It is difficult to predict how the costs of contacting respondents on RDD cell phone numbers due to answering or not answering incoming calls from unknown parties on a cell phone will change in the coming years. But the costs of contacting respondents on RDD landline numbers can be expected to increase as contact rates continue to decrease.

The use of samples scrubbed of numbers that were unlikely to be working – so called "enhanced samples" – on the cell phone side will become standard practice. Since it is now possible to determine in advance the activity history or working-number status of sampled cell phone numbers, researchers who use samples that append this information, or drop non-working numbers, gain a significant advantage in efficiency and cost that may outweigh the extra per-number cost of purchasing such enhanced samples (see Dutwin and Malarek 2014).

Predictive dialing of landlines yields substantial costs advantages, but recent rulings by the Federal Communications Commission (FCC) on the interpretation of TCPA, and the rise of litigation based on alleged TCPA violations, have caused some survey organizations to consider

abandoning predictive dialing altogether.[12] It is also clear that more and more organizations are seeing the usefulness of various kinds of enhanced cell phone samples. In the future, it may be that the typical DFRDD survey will use enhanced sample for cell phones that will be dialed manually. These forces would further reduce the differential costs associated with calling cell phone versus calling landline numbers in the United States.

Sampling companies are offering sampling products (sometimes referred to as "consumer cell samples") that match cell phone numbers to other individual and household information available in various public and proprietary sources. Although a sample drawn from such a frame is certainly not a full probability sample of all cellular households or individuals, it already is an attractive alternative for some research purposes because it allows targeting to specific populations and offers higher calling efficiencies than RDD cell phone samples. It is likely that some future telephone studies will draw from these commercial samples in much the same way that some researchers have been combining directory-listed and RDD samples to cover the landline population. If these samples continue to improve in their coverage or if more ways are found to append household or individual data to RDD cell phone samples, the cost of cell phone interviewing may be further reduced in the future, with a possible reduction of the risk of coverage bias.

The survey of telephone survey organizations that the Task Force conducted found that cell phone RDD interviews still cost more, on average, than landline RDD interviews. Using representative cost figures for sample numbers and fixing the overall cost of an hour of interviewing time at $30, we arrive at estimated cost per interview (CPI) rates of about $47 per RDD cell phone completion compared to about $36 for a RDD landline completion. This is an approximate cell/landline ratio of 1.4:1.0. Cell phone interviewing incurs higher costs for purchased numbers, but most of the difference in cost comes from the difference in hours-per-completion rates between the average cell phone RDD sample and the average landline RDD sample. Although the enhanced cell phone samples are priced higher per number than conventional cell phone samples, the enhanced samples have much higher yields, requiring fewer sampled numbers per completion, so that the total sample cost comes out a bit lower for the enhanced cell phone samples. The enhanced sample also delivers a somewhat lower hours-per-completion rate, so that cell phone completions from enhanced samples cost about $45 versus $48 for those from conventional cell phone RDD samples.[13]

As telephony technology changes in the United States and additional vendor services are created to make the calling of cell phone RDD numbers more efficient, the differential cost of surveying landline numbers vs. cell phone numbers is likely to be reduced in the coming years.

There also are new legal developments that likely will affect the future costs of telephone surveying, especially those that affect whether numbers must be manually dialed. For example, new interpretations of the Telephone Consumer Protection Act of 1994 may eventually lead to permission to place some calls via equipment that connects to an autodialer (Lange and Zielinski 2016). In addition, a 2016 ruling by the Federal Communications Commission (FCC 16-72, CG Docket No. 02-278) stated that telephone research calling for a Federal survey is exempt from the restrictions of the TCPA that prohibit using an autodialer to call cell phone numbers. Rulings such as these can be expected to lower the cost of cell phone surveying.

[12] Predictive dialing is a calling technology that in a CATI survey research setting is used to lower the costs of reaching sampled respondents. It does this by trying to maximize the likelihood that the available interviewing staff at any given time are as productive as possible in speaking to human beings at sampled phone numbers (Kelly and Kulp 2008).

[13] Readers are reminded that these rates and ratios may change in the future and that using these figures to formulate future telephone survey budgets may lead to nonnegligible errors in budgeting.

This cost section has addressed what is known and what is expected for surveys of the general population in the United States that use *only* the telephone for sampling, recruitment, and data collection. However, the patterns of differential costs between processing landline numbers and cell phone numbers are also likely to hold for surveys that use the telephone for only part of their sampling, recruitment, and/or data collection.

7. LOOKING FORWARD

As we look to the future, it is again useful to consider the distinct roles of the telephone for purposes of sampling, recruitment, and data collection.

Coverage and Sampling via Telephone. Telephone numbers are still being used to cover and sample the general population of the United States. However, the use of landline RDD numbers seems to be quickly becoming unattractive for most general population surveys, due to the rising costs associated with gaining completions via that frame and the shrinking portion of the population with only landline service. If a survey that samples only telephone numbers needs to include the dwindling segment of the population that only can be reached via landline, then some form of listed landline frame, such as the EWP, is likely to be the most cost-effective solution.

For most surveys in the future that use a telephone frame to reach the general population of the United States, the cell phone RDD frame will be sufficient because it will not lead to a meaningful amount of unit-level coverage error. Furthermore, as RDD cell phone completions become more attractive in terms of costs, researchers will be drawn to conducting single frame RDD surveys as was the case in the 1980s and 1990s but this time using only cell phone numbers rather than only landline numbers. Another attraction to using single frame RDD surveys is that weighting is less complex. As a result, surveys with a given sample size will have lower sampling errors, since design effects due to variability in the weights will be lower than for DFRDD surveys and effective sample sizes will be larger.

However, the field would benefit from quality research on within-unit coverage with a cell phone frame. Currently, there is very little reliable information available about the prevalence and amount of sharing of individual cell phones in the United States. More importantly, the Task Force is unaware of any reliable evidence that addresses whether and how much of the sharing of cell phones leads to nonignorable coverage errors of the general population when the person who answers a cell phone automatically becomes the designated respondent.

Finally, it will be interesting to watch the evolution of the value of the RICS frame because of the very low costs associated with such surveys, especially when data are gathered by the IVR, and the very quick turnaround time that the use of the RICS frame provides. Research on the circumstances under which RICS sampling is fit for purpose will be very valuable.[14]

Recruitment via Telephone. Telephone surveying is likely to be used for many years to come, especially in mixed-mode surveys that can afford the added expense of using interviewers to persuade those who remain nonresponders after being invited to participate via mail or email. The "real-time" persuasion that can be carried out by the interviewer is qualitatively different from the persuasion possible via mail, email, or a recorded voice system. It also is likely that human contact will differentially motivate different types of respondents, so that having telephone interviewers contact certain types of respondents may bring about nonresponse bias reduction. Furthermore,

[14] The RICS approach to sampling is/was the subject of a paper panel at the 2017 AAPOR conference but those papers were not available at the time this report was written.

interviewers are likely to vary widely in their success in gaining cooperation from respondents and in the types of respondents they have more or less success in recruiting. Thus, studies using telephone interviewers to recruit respondents in the future should pay special attention to this interviewer-related variation and the effects it has on the level of nonresponse bias in a survey.

It also will be of interest in the coming years to watch the development of methods, such as those used by Fowler et al. (2016) in a mixed-mode survey, to conduct so-called NRFU (nonresponse follow-up) studies for surveys that use a telephone frame for their initial sample but then follow up all or a subsample of nonresponders whose phone number can be accurately matched to an address using another mode of recruitment.

Data Collection via Telephone. There should be continued attention in the future to the extent that use of telephone interviewers creates measurement error, especially with survey items that are sensitive or prone to social desirability bias. But there are good reasons to choose live interviewers to gather data. For example, there is generally less missing data when a live interviewer engages in data collection. Another consideration is how being in different locations and/or multitasking while providing survey data via a telephone will contribute to respondent-related measurement error(s), including straight-lining and other forms of satisficing. Still another is how the audio fidelity, especially when a respondent is using a mobile device while being interviewed, may contribute to measurement error.

It is also likely that interviewers will continue to vary substantially in the quality of the data that they gather via the telephone. Thus, future researchers who use telephone interviewers to gather some or all of the data for a survey should pay special attention to the amount of this interviewer-related variation and the effects it has on the nature of the measurement error that results in a survey.

Conclusion. There are many survey researchers who believe that the telephone as a viable and attractive mode for surveying the general public is dying or is already dead, particularly advocates of nonprobability online surveys. The attraction of these nonprobability approaches for conducting survey research is their low cost and very quick turnaround. And there may well be many times when the data that nonprobabilty online surveys generate are fit for purpose. Unfortunately, too often it is uncertain whether such methods are or are not suitable for a given purpose.

In contrast, there are many clients and researchers who still rely on the telephone for all or at least a part of the surveys that they sponsor and conduct. And, as the Task Force has projected, the telephone will remain an important mode for surveying the general public of the United States for many years to come. We say this for several reasons:

- The RDD cell frame will continue to provide extensive coverage of the U.S. population, yielding little unit-level coverage error for most topics of interest to survey sponsors.
- As time passes, more auxiliary data that are accurate enough to be useful will be able to be appended to telephone frames, including cell phone frames, to aid sampling, weighting, nonresponse bias investigations, and other analytical needs.
- Samples that are based on probability selection methods can be drawn from these telephone frames, thereby providing a measurable amount of sampling error (imprecision).
- Interviewers can be used to help recruit respondents from the initial sample.
- Interviewers can help screen for eligibility in a given survey, which often can be a complex process that puts off many respondents, especially if it occurs at the very start of contact.

- Nonresponse often will be very high, but there will be continual improvements in the means for the study of nonresponse bias in telephone surveys.
- Interviewers can continue to be used to gather data from cooperating respondents, helping respondents better understand the survey questions, motivating respondents to provide accurate responses, overcoming language barriers, and using unbiased probing to elicit detailed responses to open-ended questions. Furthermore, it seems possible that advances in conversational interviewing methods eventually will provide benefits that have not as yet been realized (cf. Currivan 2008b).
- There will be well-established methods for weighting to try to adjust biases in telephone surveying that are associated with noncoverage, sampling, and nonresponse.

However, we expect that over time, fewer and fewer surveys will be conducted using only the telephone for sampling, recruiting, and data collection and that there will be more surveys that use the telephone for some, but not all, of their recruiting and data collection needs; whereas fewer surveys will use a telephone frame for coverage and sampling purposes.

References

American Association for Public Opinion Research (2016), *Standard Definitions: Final Dispositions of Case Codes and Outcome Rates for Surveys, 9th edition*, Deerfield IL: AAPOR, Available at: http://www.aapor.org/AAPOR_Main/media/publications/Standard-Definitions20169theditionfinal.pdf

Baker, R., S. Blumberg, J.M. Brick, M.P. Couper, M. Courtright, J.M. Dennis, D. Dillman, M. Frankel, P. Garland, R.M. Groves, C. Kennedy, J. Krosnick, P.J. Lavrakas, S. Lee, M. Link, L. Piekarski, K. Rao, R. Thomas, and D. Zahs (2010), *AAPOR Report on Online Panels*, Deerfield IL: American Association for Public Opinion Research, Available at https://www.aapor.org/AAPOR_Main/media/MainSiteFiles/AAPOROnlinePanelsTFReportFinalRevised1.pdf

Bauer, J. (2008), Tailoring. In P.J. Lavrakas (ed.), *Encyclopedia of Survey Research Methods*. Thousand Oaks, CA: Sage, 873-875.

Blumberg, S., and J. Luke (2016), *Wireless Substitution: Early Release from the National Health Interview Survey, January-June 2016, Washington DC: Centers for Disease Control and Prevention, National Center for Health Statistics, Available at: https://www.cdc.gov/nchs/data/nhis/earlyrelease/wireless201612.pdf*

Brick, J.M., I. Flores Cervantes, S. Lee, and G. Norman (2011), "Nonsampling Errors in Dual Frame Telephone Surveys," *Survey Methodology*, 37, 1-12.

Brick, J.M. and J. M. Lepkowski (2008). "Multiple Mode and Frame Telephone Surveys." In Lepkowski, J.M., C. Tucker, J.M. Brick, E. de Leeuw, L. Japec, P.J. Lavrakas, et al. (eds.), *Advances in Telephone Survey Methodology*, Hoboken, NJ: Wiley; 149-169.

Burks, A.T., E. Camayd-Freixas, P. J. Lavrakas, and M. A. Bennett. (2007) "The Use of Progressive Involvement Principles in a Telephone Survey Introduction to Reduce Immediate Refusals." Paper presented at the 62nd Annual Conference of the American Association for Public Opinion Research, Anaheim, CA.

Conrad, F., J. Broome, J. Benki, F. Kreuter, R., Groves, D. Vannette, and C. McClain (2013). "Interviewers Speech and the Success of Survey Introductions." *Journal of the Royal Statistical Society*, Series A, 176(1), 191-210.

Corkrey, R., and L. Parkinson (2002). "A Comparison of Four Computer-based Telephone Interviewing Methods: Getting Answers to Sensitive Questions." *Behavior Research Methods, Instruments, & Computers, 34*(3), 354-363.

Currivan, D.B. (2008a). "Interactive Voice Response." In P. J. Lavrakas (ed.) *Encyclopedia of Survey Research Methods.* Thousand Oaks, CA: Sage, 342-344.

Currivan, D.B. (2008b). "Conversational Interviewing." In P. J. Lavrakas (ed.) *Encyclopedia of Survey Research Methods.* Thousand Oaks, CA: Sage, 151-152.

Curtin, R., S. Presser, and E. Singer (2000), "The Effects of Response Rate Changes on the Index of Consumer Sentiment," *Public Opinion Quarterly*, 64(4), 413–428.

Curtin, R., S. Presser, and E. Singer (2005), "Changes in Telephone Survey Nonresponse Over the Past Quarter Century," *Public Opinion Quarterly*, 69(1), 87–98.

Dillman, D., J. Smyth, and L. Christian (2014), *Internet, Phone, Mail, and Mixed-Mode Surveys.* Hoboken, NJ: Wiley & Sons.

Dutwin, D., J. Loft, J. Darling, A. Holbrook, T. Johnson, R. Langley, P.J. Lavrakas, K. Olson, E. Peytcheva, J. Stec, T. Triplett, and A. Zuckerberg (2014), *Current Knowledge and Considerations Regarding Survey Refusals,* Deerfield IL, AAPOR. Available at: https://www.aapor.org/AAPOR_Main/media/MainSiteFiles/RefusalTF_FINAL090814.pdf

Dutwin, D., and P.J. Lavrakas (2016), "Trends in Telephone Outcomes, 2008-2015," *Survey Practice*, 9(3), Available at: http://www.surveypractice.org/index.php/SurveyPractice/article/view/346

Dutwin, D., and D. Malarek (2014), "The Use of Recent Activity Flags to Improve Cellular Telephone Efficiency," *Survey Practice,* 7(1), 1-10.

Fahimi, M. (2014), "Practical Guidelines for Dual-Frame RDD – Now That the Dust Is Settling," *Survey Practice*, May Issue.

Frey, J. (1989), *Survey Research by Telephone, 2nd edition,* Newbury Park, CA: Sage.

Groves, R.M., P. Biemer, L. Lyberg, J. Massey, W. Nicholls, and J. Waksberg (1988), *Telephone Survey Methodology,* New York: Wiley.

Groves, R.M. and M. Couper. (1998) *Nonresponse in Household Interview Survey.* New York: Wiley.

Groves, R.M., and R. Kahn (1979), *Surveys by Telephone: A National Comparison with Personal Interviews*, New York: Academic Press.

Guterbock, T. (2009), "Estimating Local Phone Service and Usage Percentages: How to Weight the Data from a Local, Dual Frame Sample Survey of Cell Phone and Landline Telephone Users in the United States," Paper presented at the Annual Meetings of the American Association for Public Opinion Research, Hollywood, FL.

Guterbock, T.M., Diop, A., Ellis, J., Holmes, J. and Le. T.K. (2009). "Who Needs RDD? (Part II)." Poster presented at the Annual Meetings of the American Association for Public Opinion Research, Hollywood, FL.

Guterbock, T.M., Diop, A., Ellis, J., Holmes, J. and Le. T.K. (2011).. "Who Needs RDD? Combining Directory Listings with Cell Phone Exchanges for an Alternative Telephone Sampling Frame." *Social Science Research* 40(3), 860-872.Holbrook, A. (2008) Recency Effects. In P.J. Lavrakas (ed.) *The Encyclopedia of Survey Research Methods*. Thousand Oaks CA: Sage Pub., 695-696.

Holbrook, A. L., J. A. Krosnick, D. Moore, and R, ourangeau, R. (2007). "Response Order Effects In Dichotomous Categorical Questions Presented Orally: The Impact of Question and Respondent Attributes." *Public Opinion Quarterly*, 71(3), 325-348.

Kelly, J. and Kulp, D. (2008). Predictive Dialing. In P.J. Lavrakas (ed.), *The Encyclopedia of Survey Research Methods*. Thousand Oaks CA: Sage Pub., 602-605.

Kolenikov, S. and C. Kennedy. (2014) Evaluating Three Approaches to Statistically Adjusting for Mode Effects. *Journal of Survey Statistics and Methods,* 2, 126-158.

Kreuter, F., Presser, S. and Tourangeau, R. (2008). Social Desirability Bias in CATI, IVR, and Web Surveys: The Effects of Mode and Question Sensitivity. *Public Opinion Quarterly*, 72(5), 847-865.

Krosnick, J., Presser, S. Fealing, K., and Ruggles, S. (2012). *The Future of Survey Research: Challenges and Opportunities*. https://www.nsf.gov/sbe/AC_Materials/The_Future_of_Survey_Research.pdf

Lange, L. and Zielinski, S. (2016) "TCPA Argument Foreshadows Unforeseen D.C. Circuit Ruling on Manually Dialed Calls." http://www.jdsupra.com/legalnews/tcpa-argument-foreshadows-unforeseen-d-58902/

Lavrakas, P.J. (1987), *Telephone Survey Methods: Sampling, Selection, and Supervision*, Newbury Park, CA: Sage.

Lavrakas, P.J. (Ed.) (2007), "Cell Phone Numbers and Telephone Surveys in the U.S.," *Public Opinion Quarterly*, 71(5), Special Issue.

Lavrakas, P.J., Kelly, J. and C. McClain. (2016). "Using a Progressive Engagement Introduction to Gain Cooperation in an Interviewer-Administered Telephone Survey." Paper presented at the 71st Annual Conference of the American Association for Public Opinion Research, Austin, TX.

Lavrakas, P.J., S. Blumberg, M. Battaglia, J. Boyle, J.M. Brick, T. Buskirk, C. DiSogra, D. Dutwin, M. Fahimi, H. Fienberg, A. Fleeman, T. Guterbock, J. Hall, S. Keeter, C. Kennedy, M. Link, L. Piekarski, C. Shuttles, C. Steeh, T. Tompson, and R. ZuWallack (2010), *New Considerations for Survey Researchers When Planning and Conducting RDD Telephone Surveys in the U.S. with Respondents Reached via Their Cell Phone Numbers,*

Deerfield, IL: American Association for Public Opinion Research, Available at: https://www.aapor.org/AAPOR_Main/media/MainSiteFiles/2010AAPORCellPhoneTFReport.pdf

Lavrakas, P.J., C. Steeh, S. Blumberg, J. Boyle, J.M. Brick, M. Callegaro, H. Fienberg, A. Fleeman, D. Gillin, J. Hall, S. Keeter, C. Kennedy, M. Link, L. Piekarski, L.C. Shuttles, and T. Tompson (2008), *Guidelines and Considerations for Survey Researchers When Planning and Conducting RDD and Other Telephone Surveys in the U.S. with Respondents Reached via Their Cell Phone Numbers*, Deerfield, IL: American Association for Public Opinion Research, Available at: http://www.aapor.org/Education-Resources/Reports/Cell-Phone-Task-Force-Report-2008.aspx

Lavrakas, P.J., Ward, C., Geng, C. Welch, V., Skalland, B, Jeyarajah, J. and C. Knighton. (2015) "Testing Envelope Features and Interviewer Training in a Large-Scale Advance letter Experiment." Paper presented at the 70st Annual Conference of the American Association for Public Opinion Research, Ft. Lauderdale, FL.

Lavrakas, P.J., Tompson, T.N., Benford, R. and C. Fleury. (2010) "Investigating Data Quality in Cell Phone Surveying." Paper presented at the 65th Annual Conference of the American Association for Public Opinion Research Conference, Chicago IL.

Lepkowski, J.M., C. Tucker, J.M. Brick, E. de Leeuw, L. Japec, P.J. Lavrakas, M. Link, and R.L. Sangster, (2008), *Advances in Telephone Survey Methodology*, Hoboken, NJ: Wiley.

Levine, B., K. Krotki, and G. Bobashev (2016), "Introducing Inbound Calling Survey: A New Sampling Methodology." Paper presented at the 71st Annual Conference of the American Association for Public Opinion Research, Austin, TX.

Maynard, D. and Schaeffer, N. C. (1997). "Keeping the Gate: Declinations of the Request to Participate in a telephone Survey Interview." *Sociological Methods and Research*, 26(1), 34-79.

Maynard, D. and Schaeffer, N. C. (2002). "Refusal Conversions and Tailoring." In D. W. Maynard et al (eds.), *Standardization and Tacit Knowledge: Interaction and Practice in the Survey Interview.* New York: Wiley, 219-239.

Peytchev, A., and B. Neely (2013), "RDD Telephone Surveys: Toward a Single Frame Cell-Phone Design," *Public Opinion Quarterly*, 77(1), 283-304.

Pew Research Center (2012), *Assessing the Representativeness of Public Opinion Surveys,* Available at: http://www.people-press.org/2012/05/15/assessing-the-representativeness-of-public-opinion-surveys/

Pew Research Center (2015), *Advances in Telephone Survey Sampling,* Available at: http://www.pewresearch.org/2015/11/18/advances-in-telephone-survey-sampling/

Reconnect Research (2016), *Inbound Calling Survey*, http://www.reconnectresearch.com/inbound-calling-survey/

Scanlan, C. R. (2008). "Primacy Effects." In P.J. Lavrakas (ed.), *The Encyclopedia of Survey Research Methods*. Thousand Oaks CA: Sage Pub., 609-610.

Schaeffer, N. C., Dykema, J. and Maynard, D.W. (2010). Interviewers and Interviewing. In J. D. Wright and P. V. Marsden (eds.), *Handbook of Survey Research, Second Edition*. United Kingdom: Emerald Group Publishing Limited, 437-470.

Shuttles, C. D., Welch, J., Hoover, J. B., and Lavrakas, P. J. (2001) The Development and Experimental Testing of an Innovative Approach to Training Interviewers. Paper presented at the meeting of the American Association of Public Opinion Research, St. Petersburg Beach, FL.

Stec, J. (2008). Calling Rules. In P.J. Lavrakas (ed.) *The Encyclopedia of Survey Research Methods*. Thousand Oaks CA: Sage Pub., 75-77.

Tarnai, J. and Moore, D. (2008). Measuring and Improving Telephone Interviewer Performance and Productivity. In Lepkowski, J.M., C. Tucker, J.M. Brick, E. de Leeuw, L. Japec, et al. (eds.), *Advances in Telephone Survey Methodology*, Hoboken, NJ: Wiley. Tourangeau, R., and T.J. Plewes (Eds.) (2013) *Nonresponse in Social Science Surveys: A Research Agenda.* Washington DC: National Academies Press.

Ward, C.D., Reimer, B., Elam-Evans, L., Yankey, D., and Khare, M. (2014) "Tradeoffs in Quality: Examining the Relationship between Cell Telephone Respondent Location and Data Quality. In *Joint Statistical Meetings Proceedings*, Survey Research Methods Section. Alexandria, VA: American Statistical Association.

Ward, C.D., Reimer, B., Khare, M., and Black, C. (2015) "Home Is Where the Cooperation Is: Association between Interview Location and Cooperation among Cellphone Users." *Survey Practice*, 8, 1-8.

West, B. and Groves, R.M. (2013). "A Propensity-adjusted Interviewer Performance Indicator." *Public Opinion Quarterly*, 77(1), 352-374.

APPENDIX A

ADDITIONAL HISTORY OF ADVANCEMENTS IN TELEPHONE SURVEYING IN THE UNITED STATES

Paul J. Lavrakas, Independent Consultant

ABSTRACT: AAPOR and many of its members have played major roles in identifying the ways in which telephone survey methods have remained viable and valuable for researchers and their clients that are well served by using the telephone mode for all or part of their survey needs. This appendix provides more detailed of what occurred along these lines during the past two decades.

The rapid adoption of cell phone technology early in the first decade of the 21st century brought about the demise of the type of telephone surveying that sampled only landline telephone numbers (cf. Pew 2015). This occurred because as people began to abandon their landline phones for cell phones, landline RDD surveys could not adequately cover the general population of the United States. AAPOR Members were at the forefront of the efforts to devise reliable and valid research methods to include those with cell phones in general population telephone surveys.

In February 2002, a business dinner conversation in Crystal City MD among Ed Cohen (Arbitron), Paul J. Lavrakas (Nielsen), Linda Piekarski (SSI), and Clyde Tucker (BLS) led to the Nielsen Company sponsoring two "Cell Phone Summits" (2003 and 2005) to bring together recognized public and private sector telephone research experts to address "what should be done about cell phones?" At the 2004 AAPOR annual conference in Phoenix, Linda Piekarski organized and chaired a session on "Cellular Phones and Telephone Sampling." Through the rest of the decade, AAPOR went on to become the primary forum through which telephone researchers in the United States discussed and vetted how dual frame random-digit dialing (DFRDD) survey sampling should be carried out, including (a) having three-day "mini-conference" of consecutive sessions about various aspects of cell phone surveying at the 2007 annual AAPOR conference in Anaheim CA, (b) publishing the 2007 special issue of *Public Opinion Quarterly* devoted to "Cell Phone Numbers and Telephone Surveys in the U.S." (Lavrakas 2007), (c) organizing the conference, *Telephone Survey Methodology II*, in Miami and (d) publishing the resulting book on *Advances in Telephone Survey Methodology* (Lepkowski, Tucker, Brick, de Leeuw, Japec et al. 2008). Following that, reports from two AAPOR task forces (Lavrakas, Steeh, Blumberg, Boyle, Brick, et al. 2008; Lavrakas, Battaglia, Blumberg, Boyle, Brick et al. 2010) addressed the state of knowledge at the time about DFRDD surveying in the United States.

APPENDIX B

TELEPHONE PROBABILTY-BASED SAMPLING ADVANCEMENTS

Scott Keeter, Kyley McGeeney, and Courtney Kennedy
Pew Research Center

ABSTRACT: In recent years a number of advancements have been made in sampling for telephone surveys. Much of the activity has focused on cellphone samples. Innovations include activity flags, which give an indication as to the working status of a number. There are also sample flags indicating whether the phone is prepaid rather than on a contract with a carrier; prepaid phones are disproportionately used by a number of hard-to-reach demographic groups. There have been improvements in geographic targeting in cellphone samples as well. This is important as 10% of U.S. adults have a cellphone number associated with a state different from where they live. Finally, landline sampling now offers a new assignment-based sample frame with slightly higher coverage of U.S. adults.

1. INTRODUCTION

Telephone surveys face numerous challenges, but there are some positive developments as well, principally with respect to sampling. Companies that provide telephone samples have developed several products that have the potential to improve the efficiency of surveys in both the landline and cell phone frames and allow for more accurate geographic and demographic targeting of calling in the cell phone frame. A number of tests have been conducted with these products and the relative costs and benefits of each are becoming clearer. This section describes some of these innovations and presents the results of experiments conducted with each of them. These experiments were designed to assess the benefits of each.

2. CELL PHONE SAMPLING

The most important change in telephone surveys in the past decade has been the adoption of dual frame survey designs that include cell phone numbers. Many survey organizations now conduct at least as many interviews by cell phone as by landline in typical national surveys. With nearly half of all U.S. adults now reachable only by cell phone, cell phones samples reach a more diverse sample of the public than do landlines. But this benefit is partially offset by the fact that less is known in advance about cell phone samples – in particular, where people actually live. Moreover, cell phone surveys are especially costly because of the government requirement that numbers be dialed manually. Therefore, any improvement in the accuracy and efficiency with which these samples reach the targeted respondents could yield significant benefits in terms of the cost of such surveys. Fortunately, new products are helping researchers make more effective use of cell phone samples.

3. ACTIVITY FLAGS

One factor contributing to the cost of cell phone interviewing is the substantial share of the numbers sampled and dialed by interviewers that turn out to be nonworking, resulting in wasted

interviewer time. As of 2015, the incidence of nonworking numbers on the U.S. cell phone RDD frame was approximately 38 percent. To reduce this inefficiency, random digit dialing (RDD) sample vendors have developed services to identify and discard cell phone numbers that are likely to be nonworking before they are dialed.

These services leverage "activity flags," which indicate whether a sampled cell phone number is flagged as "active" (and is likely working) or flagged as inactive (likely nonworking). For a small fraction of cell numbers, typically less than 5 percent, the flag is not available. In 2012, vendors started making these flags available so that survey designers could exclude or subsample flagged-inactive cases from their samples and, in turn, save money by reducing the amount of time that interviewers spend dialing unproductive numbers.

Unfortunately, the flags are not perfect, containing both false positives (numbers flagged-active but actually nonworking) and false negatives (numbers flagged-inactive but actually working). False positives are basically harmless, aside from the loss in efficiency. False positives, however, have the potential to reduce the survey coverage rate and increase the risk of noncoverage error. Dutwin and Malarek (2014) estimated that using the Targus or Cell-WINS recent activity flags to exclude suspected inactive numbers reduces the coverage of all "cell phone households" by 5 to 6 percentage points. A study by Pew Research Center (2016) using a sample five times larger estimated the cell frame coverage reduction to be 7.4 percentage points.

For a national cell phone sample, the estimated *net* population coverage rate drops from about 90.6 percent to 83.2 percent as a result of exclusion. These net rates are based on estimates from the National Health Interview Survey (NHIS) that about 3.1 percent of adults have no telephone and another 6.3 percent have a landline but no cell phone (Blumberg and Luke 2015). For national dual frame RDD surveys of adults, the inclusion of a landline sample substantially dampens the estimated reduction in net coverage from excluding flagged-inactive numbers (from 96.9 percent coverage to 93.1 percent, rather than 83.2 percent).

If the survey designer uses the recent activity flags to subsample flagged-inactive numbers rather than to wholly exclude them, there is no coverage rate penalty. Two downsides of subsampling with these flags, however, concern precision and cost. The weighting adjustment to correct for the subsampling carries a design effect and may reduce the precision of survey estimates. Also, subsampling entails retaining some fraction of the flagged-inactive numbers in the sample, which reduces interviewer productivity relative to the exclusion approach. Considering that the recent activity flag currently costs 7 cents per number to append, the net cost savings may be marginal at best under the subsampling approach. Pew Research Center (2016) found no clear benefit to subsampling over exclusion with recent activity flags when examining effects on weighted point estimates and precision.

While excluding flagged-inactive cell phones yields an undesirable reduction in the survey coverage rate, it does not appear to have a meaningful effect on bias, at least for public opinion surveys. Adults with cell phones erroneously flagged as inactive do differ systematically from adults with flagged-active cell phones on several dimensions (e.g., age, race, education, income, voter registration status), but they constitute too small a fraction of the population to meaningfully move full sample or even some common subsample estimates. In the Pew Research Center study, 2.0 percent of all n=5,003 completed cell phone interviews were with numbers erroneously flagged as inactive using the flag, which compares to 3.5 percent of the n=3,645 cell phone interviews in the Dutwin and Malarek study. Dutwin and Malarek found that the average change in estimates associated with excluding flagged-inactive numbers is less 0.5 percentage points and is even less for RDD surveys that also feature a landline sample.

4. PREPAID PHONE FLAG

Another new service that telephone survey researchers have been testing identifies cell phone numbers that belong to prepaid or "pay as you go" cell phones. One reason that prepaid phones are of potential interest to survey researchers is because they are disproportionately used by demographic groups that are often underrepresented in telephone surveys. For example, Dutwin (2014) and McGeeney (2015) found that prepaid phone users were more likely than other cell phone users to be non-white, lower income, and less educated.

The prepaid phone flag is currently not available as a standalone product; one must purchase the activity flag in order to have it appended. McGeeney reported that 12.4 percent of the dialed sample (n=23,120) and 15.4 percent of the completed cell phone interviews (n=978) were flagged as prepaid in a 2015 national cell phone RDD sample conducted for the Pew Research Center. While the flag succeeded in identifying traditionally harder to reach groups, the fact that it is appended after the sample is drawn substantially limits its cost effectiveness for demographic targeting purposes. Using this flag to oversample traditionally hard-to-reach groups is generally considered too costly and inefficient relative to other approaches.

5. GEOGRAPHIC TARGETING

Another notable development is that accurate geographic targeting is increasingly possible with cell phone samples. A critical limitation of cell phone samples for research below the national level is that they suffer from both overcoverage and undercoverage. Samples of cell phone numbers drawn for a particular area inevitably include people who do not live there (overcoverage). And depending on the area, significant numbers of people may live in the area but have cell phone numbers corresponding to a different geographic area (undercoverage). Pew Research Center estimates that 10 percent of U.S. adults have a cell phone number from another state.

A new product to help address undercoverage is a database containing a convenience sample of cell phone numbers with additional information appended to each record. This information includes demographics and, more importantly, address. The address can be used to help alleviate undercoverage in dual frame telephone sample designs of subnational geographies. Researchers can use this database as a third stratum to sample cell phone numbers from rate centers excluded from the dual frame design but with appended addresses that fall within the study's geography.

Cell phone samples are also now available with billing ZIP code appended, which is based on where the current bill is being sent, as opposed to rate center location, which is based on where the phone was initially issued. This is only available for all cell phone numbers that are sent a bill; prepaid phones are excluded, but unlike the database of cell phone numbers, the billing ZIP information is not limited to a convenience sample of cell phone numbers. However, billing ZIP codes are appended after the sample is drawn, which means that researchers may have to pay for sample that end up not using.

Luckily, the ability to draw the sample based on billing ZIP, and even full billing address, is now a reality. A new product allows researchers to not only *sample* on billing zip code, as opposed to *appending* it after the sample was drawn, but would include full addresses as well. This information is available for the entire cell phone sample frame apart from prepaid phones. It allows for geographic targeting but also provides a unique opportunity to send advance materials to addresses from a cell phone sample, including a prepaid incentive. Pew Research Center tested doing just that and was able to increase the response rate by 3 percentage points among cell phone numbers with an address available.

Moreover, this enhanced cell phone sample frame contains demographic information from large consumer databases maintained by companies such as Experian. This feature helps telephone survey researchers target specific demographic groups more efficiently than they can using existing approaches, for instance by assigning bilingual interviewers to make first calls to flagged Hispanic phone numbers. Finally, this enhanced frame also includes the recent activity and prepaid cell phone flag, allowing for samples to be drawn based on these characteristics rather than having them appended after the fact.

6. LANDLINE SAMPLING

Innovations are also occurring with landline samples. One common option to improve landline efficiency is restricting the sampling frame to blocks of 100 consecutive numbers ("100-banks") where at least one (or two or three) numbers are listed in a directory (Casady and Lepkowski 1993). However as more and more landline numbers are unlisted, in part due to the proliferation in voice over internet protocol (VoIP) phones, studies indicate that the list-assisted landline sample frame may have increasing coverage error due to the exclusion of numbers on 100-banks with no directory-listed numbers ("zero banks") (Boyle et al. 2009; Fahimi et al. 2009). To overcome the effects of this change, telephone sample vendors are now offering "assignment-based" landline sample frames, where a 100-block is included in the frame if one of its numbers is assigned for use by a telecom company, regardless of whether it is directory listed.

The only independent assessment of the assignment-based frame that we are aware of was conducted by the Pew Research Center. They used the more inclusive assignment-based frame for drawing the landline sample for a large national dual frame RDD survey of adults (the 2014 Religious Landscape Survey). A flag was included in the landline sample to indicate whether the number would have been included in the 1+ list-assisted landline sample frame. They found that respondents in the assignment-based landline sample who would have been excluded in a list-assisted landline sample were distinct demographically. They were younger, less educated, lower income, more likely to be non-Hispanic black, live in an urban area and to not be married; they were less likely to be registered to vote. However, they constituted only 1.4 percent of the landline numbers dialed and 0.9 percent of completed landline interviews. Thus, their incidence was too low to meaningfully affect survey estimates.

References

Blumberg, S.J., J.V. Luke (2014), *Wireless Substitution: Early Release of Estimates from the National Health Interview Survey, July–December 2014.* National Center for Health Statistics. June 2015. Available from: http://www.cdc.gov/nchs/nhis.htm.

Boyle, J., M. Bucuvalas, L. Piekarski, and A. Weiss (2009), "Zero Banks: Coverage Error and Bias in RDD Samples Based on Hundred Banks with Listed Numbers," *Public Opinion Quarterly*, 73(6), 729-750.

Casady, R.J., and J.M. Lepkowski (1993), "Stratified Telephone Survey Designs." *Survey Methodology*, 19, 103–13

Dutwin, D., and D. Malarek. (2014). "The Use of Recent Activity Flags to Improve Cellular Telephone Efficiency," *Survey Practice*, 7(1).

Dutwin, D. (2014), "Cellular Telephone Methodology: Present and Future" [Webinar]. In: AAPOR Webinar Series. Available at: http://www.aapor.org/AAPORKentico/Education-Resources/Online-Education/Webinar-Details.aspx?webinar=WEB0114.

Fahimi, M., D. Kulp, and M.J. Brick (2009), "Reassessment of List-Assisted RDD Methodology," *Public Opinion Quarterly*, 73(4), 751-760.

McGeeney, K., and C. Kennedy (2016), "A Next Step in the Evolution of Landline Sampling: Evaluation the Assignment-Based Frame," *Journal of Survey Statistics and Methodology*, 4(4).

McGeeney, K. (2015). "Appending a Prepaid Phone Flag to the Cell Phone Sample," *Survey Practice*, 8(3).
Pew Research Center, (October, 2016), "Cell Phone Activity Flags." Available at:
1250 http://www.pewresearch.org/2016/10/24/cellphone-activity-flags/
Pew Research Center, (August, 2016), "Moving Without Changing Your Cell Phone Number: A Predicament for Pollsters." Available at: http://www.pewresearch.org/2016/08/01/moving-without-changing-your-cellphone-number-a-predicament-for-pollste

APPENDIX C

WEIGHTING IN CURRENT AND FUTURE TELEPHONE SURVEYS

Major developments in the design of surveys in which telephone numbers comprise the sampling frame have had important implications on the way weights are constructed for random digit dial (RDD) telephone surveys in the United States. These developments affect how samples are designed, which in turn affect how the weights are computed. Other developments have made available new sources of information that have the potential to improve adjustments for nonresponse when survey recruitment is carried out via telephone.

Here we focus on how these developments are shaping current weighting practices for telephone surveys while providing a summary of current weighting practices. Much of the weighting content from prior American Association for Public Opinion Research (AAPOR) Task Force reports is still very relevant, and it is not repeated here. This appendix is organized into sections that describe the need of weights and types of weighting procedures in telephone surveys, the role of weighting in future telephone surveys, and the main steps that are often done to compute weights.

1. THE NEED FOR WEIGHTS IN TELEPHONE SURVEYS

In most situations, weights are needed in the analysis of telephone surveys. The main reasons for producing weights are to compensate for differential probabilities of selection, reduce bias occurring when respondents have different answers from nonrespondents, and compensate for sampling frame coverage problems. Although there is no unique way to produce weights, we can identify three common steps during weighting: (1) creation of the base weights, (2) adjustment of base weights to compensate for differential nonresponse, and (3) benchmarking to some known population totals so the sample is representative of the population of interest. With the use of a dual frame random digit dialing methodology (DFRDD), where a separate cell phone sample is drawn to address the undercoverage in the landline frame, an additional weighting step—the combination of the samples drawn from the different frames--is needed.

Although the weighting steps are straightforward, their creation varies greatly among survey organizations and, in most cases, by survey (Fahimi 2014). Part of this variation is the result of the different ways these organizations responded to the fast changes of telephone usage in the population, increasing survey costs, and changing attitudes toward surveys in a competitive business environment. Other sources are the result of the methodological inconsistencies of the current DFRDD approach in combination with researchers' ad hoc assumptions that impact the landline and cell phone samples at the design stage. In summary, there is no general consensus on the best way to weight for nonresponse in telephone surveys. This was already recognized by the 2010 AAPOR Cell Phone Task Force. Despite many recommendations, there are still many variants.

Despite the different methods, we refer to the weighting approaches based on three additional criteria: (1) the sample design strategy used to collect the data, such as the use of a single or dual frames, (2) the type of nonresponse/undercoverage adjustments implemented, and (3) the sequence of implementation of weighting adjustments. The last two criteria are related to the use of available auxiliary information in weighting to try to reduce nonresponse bias and to align the sum of weights to known population totals through benchmarking. This classification is presented in Table

1 and is not by any means complete or exhaustive. However, it highlights the different opinions toward weighting.

TABLE 1. WEIGHTING APPROACHES IN TELEPHONE SURVEYS IN THE UNITED STATES

CLASSIFICATION OF WEIGHTING PROCESS BASE ON	EXAMPLES
Sample design strategy	• Overlapping dual-frame designs (dual frame: cell phone and landline design) • Non-overlapping designs (screening designs, single-frame cell phone design) • Landline frame • Cell phone frame
Type of nonresponse/undercoverage adjustment	• Specific adjustments for nonresponse, such as weighting classes, propensity score stratification, etc. • Benchmarking to known population totals: poststratification or raking
Sequence of weighting adjustments	• Sequential weighting adjustments, such as separate household and person nonresponse adjustments • One-step weighting adjustment

The DFRDD design with samples drawn from landline and cell phone RDD sampling frames is currently the most common approach for general population telephone surveys in the United States. As a result, weighting for such designs is the main focus of this appendix. However, it is important to highlight that as the use of cell phones continues to increase, single cell phone frame weighing methods may become more prominent in the future as researchers abandon the dual-frame approach (Peytchev and Neel 2012). Furthermore, due to the flexibility of telephones in surveys, there are other alternative designs such as telephone surveys of respondents selected using address-based sampling (ABS) designs that have different weighting issues not discussed in this appendix.

2. IMPORTANCE OF WEIGHTING IN CURRENT AND FUTURE OF TELEPHONES SURVEYS

When response rates are low and sampling frames are incomplete, the main focus of weighting is to reduce the impact of nonresponse and undercoverage on the population estimates. As response

rates continue to decline in the coming years, the role of weighting will be more prominent and subject to close scrutiny. As mentioned in the main report, there are concerns in the field that the levels of noncoverage and the nonresponse rates achieved in current telephone surveys make it difficult to claim these samples and estimates are representative of the population of interest. Currents trends for addressing this problem include implementing more sophisticated weighting methods such as calibration adjustments with auxiliary variables that go beyond the simple demographic adjustments used in the past surveys. However, these techniques heavily rely on auxiliary information from sources that may not be readily available to all researchers and organizations. Furthermore, the auxiliary variables need to be correlated to the key outcomes in order to be effective in removing nonresponse biases. Although this situation is likely to improve in the future as more and better quality sources are available for weighting, it is still not clear if more sophisticated weighting methodologies are able to address the nonresponse bias in surveys with very low response rates.

We repeat calls from previous AAPOR tasks forces in the disclosure of the weighting procedures used in the analyses including the decision of not to weight if this is the case. This information will enable the survey community to evaluate current and new weighting approaches, identify those methods that work, and determine limitations of weighting.

3. WEIGHTING STEPS IN CURRENT TELEPHONE SURVEYS

The current weighting methodology with few variations is based on the DFRDD design discussed in the AAPOR Cell Phone Task Force report of 2010 (Lavrakas et al., 2010). A general discussion of the steps that are typically implemented to weight the data is provided; however, the number and type of weighting steps vary and depend on how the sample design used to select the data, project resources, and the researcher's experience.

3.1 Design Weights

Weighting is a process needed for producing estimates from survey respondents that are representative of the total population from which they were sampled. There are different types of weights associated with samples from telephone surveys; however, here we focus on design weights. The design weights reflect that the data were drawn using a sample design based on the randomization of telephone numbers in the sampling frame.

The initial or base weights are computed as the inverse of the probability of selection of the sampled telephone number. However, in contemporary telephone research, particularly in DFRDD samples, it is unusual for all respondents to have the exact same chance of selection. It is common to implement special sampling procedures such as subsampling or oversampling to target certain subpopulations or geographic areas, increase the sampling efficiency, or reduce costs. All these special procedures affect how the base weights are adjusted. In addition to these subsampling procedures, other adjustments are used to account for the increased probability of selection in the households that have multiple telephone numbers and the links among respondents to the cell phone and landline telephone numbers. For example, if according to the telephone survey sample design, a member of a sampled household is selected in the second stage (i.e., the within-unit selection of a designated respondent), the probability of selection of the sampled person among all persons eligible in the household is also used to adjust the base weight.

Recent developments in record linkage have enabled samplers to match auxiliary information to sampled cell phone numbers to make further distinctions among telephone numbers that can be used in sampling. Information can be used to make the survey more cost effective by targeting specific types of telephone numbers during sampling. Examples of this information include:

- Telephone activity (i.e., activity flag, which indicates if the telephone is currently active or not, or has been active in the past six months),
- Geographic location where the phone was activated (i.e., rate center), which can be used as linked to other sources with demographic and economic characteristics like the census, and/or
- Geographic information of the location of the user (i.e., ZIP code of the billing address associated with the number)

This information can be used to reduce the number of calls to unproductive numbers (i.e., inactive numbers) and exclude ineligible numbers (i.e., numbers outside geographic areas of interest). Although this information is used mainly to increase the efficiency of the sample, it is also needed to adjust base weights. Below are details of these forms of adjustments to the base weights when this information is used at the design stage.

3.2 Base Weights, Linkage, and Subsampling Adjustments

The base weight in DFRDD depends on the type of sample design used to collect the data. In a non-overlapping DFRDD design (also known as screening designs), the sample from one frame is screened to identify and interview respondents not present in the second frame. This approach was used in the early days of cell phone sampling when cell phone samples were used to identify cell-only users who are not found in the landline frame. However, this approach has been mostly abandoned in favor of the overlapping design where the samples are not screened to identify specific types of telephone users.

In a non-overlapping design, the design weights need only account for the probability of selection from the frame in which each respondent was actually reached. In contrast, in overlapping designs, the base weights need to additionally account for the fact that dual users – users of both cell phone and landline telephone services – have a chance of selection in both frames. Composite weight adjustment factors and multiplicity sampling estimation are two approaches for combining samples and adjusting weights for frame overlap. Details of these approaches, as well as their advantages and disadvantages, are considered next.

Composite estimation is perhaps the most intuitive way to correct for frame overlap of dual telephone users (Hartley 1962; 1974). Before the adjustment, the dual (or common) telephone user population (i.e., person who use both a landline and a cell phone) is essentially twice represented in the combined sample (once in the cell phone sample and again in the landline sample). Under compositing, the two dual-user samples are averaged together – not necessarily in equal proportion – so that the dual-user population is counted exactly once in the weighted survey estimate. Specifically, the weights for dual users are multiplied by a factors θ and θ-1 with values of θ between 0 and 1 determined by the researcher. The main issue with this approach is that there is no unique optimal value for the composite factor for all key outcomes of interest. In most situations, the compositing factors are set 0.5 independently of the outcome being estimated. Other more formal approaches involve using the effective sample size of the outcome (e.g., Frankel et al. 2007) or incorporating the response rates of dual users in each sample in the calculation of the factor (Brick et al. 2011). A sensitivity analysis should be conducted because the value selected

for θ affects the overall variability of the weights. Additional details of the use of composite weights can be found in the 2010 AAPOR Cell Phone Task Force report.

The multiplicity sampling estimation approach addresses the frame overlap issue from a rather different perspective (Bankier 1986; Kalton and Anderson 1986). Instead of viewing the design as two partially overlapping sampling frames, it can be viewed as one frame with three strata. One stratum contains the landline-only cases; another stratum contains the cell phone-only cases, while the third stratum contains the dual-user cases. In other words, a multi-frame sample can be viewed as a special case of selecting two or more samples independently from a single frame. In in most cases, computing the multiplicities or different probabilities of selection of sampled numbers in the frames requires access to the frames. The researcher computes a single selection probability for all dual users, regardless of the frame in which they were reached. For example, if the probability of selection of dual user i was f_i^C in the cell frame is and f_i^L in the landline frame, then the dual-user design weight is computed as $1/(f_i^C + f_i^L)$. Mathematically, this is equivalent to selecting θ so that the weights in both frames are the same, and equal to the inverse of the unconditional total probability of selection. In other words, single-frame estimation is a special case of composite estimation. Composite weight adjustment factors and multiplicity sampling estimation are closely connected, but they have different properties and can look rather different when described in a methodology report.

There are advantages and disadvantages of the methods for overlap adjustment. In DFRDD designs – where all telephone numbers have the same probability of selection within each frame – both overlap adjustments tend to perform reasonably well. One advantage of single-frame estimation is that it generalizes well to surveys that have more than two sampling frames. In addition, some practitioners find that at the design stage, it is easier to estimate the design effect from the overlap adjustment using single-frame estimation. Compositing, by contrast, has the advantage of being fairly intuitive and, in its simplest form (i.e., setting θ equal to 0.5), it is easy to implement. It also can have more desirable statistical properties. The framework is more flexible, allowing researchers to combine the dual-user samples in a way that yields some desirable outcome, such as reduced variance or reduced nonresponse bias.

One major disadvantage of single-frame estimation is that the procedure requires knowing the probability of selection of each selected dual user from both the landline frame and from the cell phone frame. This situation does not pose a problem in DFRDD surveys with no oversampling because the sampling fraction in each frame is known (with some assistance from the sample vendor). However, in surveys with oversampling, single-frame estimation may become extremely complicated to the point of being unworkable.

To illustrate this point, consider a national DFRDD survey where less populous states are oversampled so that the final sample supports both national- and state-level estimates for all 50 states. Under compositing, the researcher only needs to know the sampling fraction within each frame for each state. That is a sizable number of parameters, but they are readily available based just on the design. In contrast, under single-frame estimation, the researcher needs to collect unusual information from each dual user. For example, consider the case of a Texas resident who is interviewed on her landline and who also has a cell phone number from Idaho. Typically, a DFRDD survey records that this person is a dual user but not the fact that the cell phone is from a different state. A naïve approach is to assume the respondent's cell phone number to be also from Texas and weight it accordingly. This assumption is incorrect and results in a larger design weight than appropriate. In order to correctly compute design weights for a study with single-frame estimation like this, the study needs to collect the state their cell phone number originated from (or

alternatively, the area code) from each dual user. In other words, when respondents can be reached on more than one cell phone, this information needs to be collected for each additional cell number. This type of measurement is likely to have a sizable item nonresponse rate (and cost implications), forcing the researcher to impute the missing information. The self-reported data are subject to response error, leading to additional variability in the weights. Even if all the necessary information is collected, the researcher has the daunting task of computing weights for hundreds of potential combinations of state of residence and state of cell phone origination. For such surveys, it seems unlikely that any potential benefit from single-frame estimation would outweigh these substantial complications. In contrast to dual-frame approaches, when the survey design features only a cell RDD sample a single frame (either cell phone, landline, or from a list), computing design weights is relatively straightforward. The researcher needs only account for any factors that resulted in different probabilities of selection among the sample members. In the case of cell phone samples, this includes the number of cell phones on which the person can be reached, the numbers of persons who share the cell phone number, and the number of eligible persons in the household if a within-household person selection is used.

The use of activity flags to design cell RDD samples is one of the main major developments since the previous AAPOR 2010 cell phone task force report. The activity status flag or Cell-WINS flag indicates if a cell phone number is active or not. In recent surveys, the observed distribution of the activity flag on samples from the national cell RDD frame has been approximately 30 percent flagged-inactive and 70 percent flagged-active or unknown. This information can be used for sampling; and as a result, it has an impact on the creation of the weights. However, the activity flag is subject to measurement error where a proportion of numbers are misclassified as active while inactive or otherwise. While this misclassification has an impact of the efficiency of the estimates, in weighting, we need to distinguish whether the activity flags are used to exclude or to over- or under-sample certain numbers. For example, if a researcher wants to reduce cost by excluding all flagged-inactive numbers, the base weight is then computed reflecting the reduced number of sampled telephone numbers. On the other hand, if there is both a concern in the misclassification errors in the values of the activity flag and there is also the need to lower costs, then those numbers flagged as inactive can be under-sampled. For example, the composition of a sample for such a study might look something like 10 percent flagged-inactive and 90 percent flagged-active or unknown. In this case, the achieved cost savings are the result of not dialing all cases that are likely to be nonworking. For weighting, the base weight of the cases flagged as inactive is then adjusted by the inverse of the sampling rate. If the design weights are not adjusted to account for this, the survey estimates will under-represent adults with flagged-inactive cell phones (who are known to exist) and will be biased. In this example, the weights of the flagged-inactive cases are adjusted by a factor of 30%/10%=3.00 while the weights for the flagged-active or unknown cases are adjusted by a factor of 70%/90%=0.78.

The cell phone weights are also adjusted in sample designs where one eligible respondent is selected for the interview within households and that respondent shares the same cell phone number. Although many researchers appear to assume that this situation is not very common because cell numbers often are viewed as mostly personal devices, some surveys attempt to collect the linkage data for this type of adjustment. As in the case of respondents with multiple cell numbers, the full information needed to adjust for shared cell phone requires identifying the number of cell phones in the households and the linkages among the adults who share the cell phone numbers through the questionnaire. This information is difficult (and costly) to collect, and requires multiple questions that are subject to measurement error and/or nonresponse error. For

example, the linkages can be different and not well defined in surveys that collect that information about children because parents or guardians who grant permission or provide information about the child. One alternative suggested by the AAPOR 2010 cell phone task force report was to adjust the weights in cases when the cell phone is shared by two eligible persons by assuming that cell phone can only be shared by at most two persons. Since the percentage of adults who share a cell phone number is assumed to be small and is expected to become smaller as cell phones ownership become more prevalent, most current surveys ignore these linkages by assuming that no cell phone is shared and there is a one-to-one linkage between a one cell phone and one adult.

Another challenge that DFRDD telephone surveys based on geographic areas (i.e., state and local area surveys) is the portability of cell phone numbers. Quite often in non-national telephone surveys, a high proportion of respondents report living outside the study area, rendering them ineligible for the survey. As discussed before, it is increasingly common to use sample designs that stratify the frame, and samples are drawn from strata constructed using geographic location of the rate center and/or appended billing address ZIP code. The researchers can increase the study's overall geographic eligibility rate by using a higher sampling rate in some strata and a lower rate in others. When rate center, billing address ZIP code, or similar information is used to design a cell RDD sample, the implications for weighting are analogous to those discussed above for activity flags. If the choice is made to exclude a certain class or stratum of numbers, the coverage rate declines and the weights can still be computed based on the reduced frame. On the other hand, if the researcher chooses to vary the sampling rate based on the geographic information to target specific geographic areas, then the weights need to be adjusted accordingly to account for the over- or under-sampled geographic areas. As in the case of the activity flags, there are misclassification errors associated to the appended geographic information that affects the coverage of the frame when these are used to exclude ineligible geographic areas from sampling and the efficiency of the estimates.

3.3 Nonresponse und Undercoverage Adjustments

In the ideal situation without nonresponse and no undercoverage, the creation of the weights requires only information on how the sample was selected. When the survey response rates are low (and there is differential nonresponse) and the sampling frames are incomplete (and there is differential undercoverage), the focus of weighting shifts to reducing the impact of nonresponse and undercoverage on the population estimates. Since the weighting adjustments rely mainly on auxiliary variables and statistical models, the form of the adjustment depends on the available auxiliary information and the implicit or explicit assumptions made about the nonresponse and undercoverage mechanisms.

Although response rates in DFRDD telephone surveys tend to be low, a low response rate does not imply that the survey estimates necessarily suffer from nonresponse bias after the weights have been adjusted for nonresponse. Furthermore, the level of nonresponse bias is different for each survey estimate. Unless there is a strong relationship between the likelihood of participating and the outcome of interest not adjusted during weighting, the bias of the estimates is expected to be negligible. Unfortunately, the bias is only observable in estimates that are available from other sources or surveys. Therefore, it is difficult to assess when nonignorable nonresponse bias exists in telephone surveys.

The main goal in the nonresponse adjustments is to adjust the weight for the differential nonresponse among subgroups in the population. A missing at random (MAR) assumption is

employed for elements of these groups (also known as weightings classes). It is assumed that sampled persons within the weighting class respond to the survey at the same rate and this rate varies across groups. Any nonresponse weighting scheme involves analyses for identifying these subgroups with differential response rates. Once these classes have been determined, the weights are ratio adjusted so the weights of respondents within weighting classes represent both respondents and nonrespondents. Unfortunately, in most telephone surveys, there are a limited number of auxiliary variables that can be used for nonresponse adjustments. Furthermore, since these variables are either related to the geographic location of the telephone number or the demographic characteristics of respondents, they are powerful predictors of nonresponse. For example, geographic characteristics that can be linked to the telephone number may be too large (i.e. too imprecise because they apply to areas that are too large in size) to effectively differentiate respondents within the geographic area. We expect that more detailed information about the telephone number (i.e., information that is measured at a smaller and thus "more local" geographic area) will be available in the future as additional and better quality data sources are able to be matched to the individual telephone numbers.

There are also differences in the way the nonresponse adjustments are implemented. In the direct weighting or one-step weighting, the adjustments are done to the person base weights that are computed using the probability of selection of the telephone number adjusted by the probability of selection of the person and any other subsampling adjustments. This person-level weight is then adjusted for nonresponse and undercoverage in one single step by benchmarking demographic characteristic such as age group, gender, education, or race/ethnicity to population totals from external sources such as the American Community Survey (ACS). In the future, we expect new sources for external population totals to be available to researchers.

In contrast to the one-step weighting, the nonresponse adjustments can be implemented sequentially reflecting how the data are collected and how nonresponse affects the sample at every stage. This sequential approach computes weights following the same path in how the sample was selected and how the interview is conducted. For example, interviews can be divided into a screening interview and extended interview. During screening the eligibility of both households and persons are determined, and a person is selected for the subsequent part of the interview. In the second part of the questionnaire, the extended interview focuses on collecting data from the sampled adult. In the sequential approach, weights are created reflecting these steps. For example, a nonresponse adjusted household weight is computed for the screener interview. Then, the nonresponse adjusted household weight is used as a component to compute the initial person weight (i.e., unconditional person weight), which is, in turn, adjusted for nonresponse at the person level at the extended interview. One advantage of this approach is that different auxiliary variables can be used to adjust for nonresponse at different stages. For example, information collected during the screener interview, which is available for both extended interview respondents and nonrespondents, can be used to adjust for nonresponse in the extended interview.

The weighting procedures also differ on the form of nonresponse adjustments that result from the method used to adjust the weights. One difference among methods is the implicit or explicit mathematical models assumed to describe the nonresponse mechanism. Most of these methods and the corresponding adjustments are well described in the literature and have been applied to telephone surveys (cf. Kalton & Flores Cervantes, 2003). We can distinguish those methods based on modeling response propensities from those based on response homogeneity groups (Särndal, Swensson, & Wretman, 1992). Among the methods used to compute the propensities are logistic regression, segmentation/classification procedures such as CHAID or CART, and more recently,

methods based on statistical learning algorithms such as random forests (Buskirk and Kolenikov 2015). However, most methods produce similar results when the nonresponse model holds.

After adjusting for nonresponse, the weights can be adjusted for undercoverage of the sampling frame by benchmarking the sum of weight to known population totals from external sources such as the ACS. Some methods are the simple ratio estimator, poststratification, and raking (Deville and Särndal, 1992). As in previous methods, these can be used stand-alone or in combination with other forms of weighting adjustments.

Newer methods of weighting of survey data are gaining popularity and go beyond the traditional geodemographic adjustments. In particular, by advancing calibration adjustments techniques popularized by Deville and Särndal (1992), these new techniques require the inclusion of an expanded set of auxiliary variables when computing survey weights. Such auxiliary variables can include attitudinal and behavioral measures for which reliable sources for benchmarks are available. In fact, expansion of this position suggests that the variables included in the weighting process should include those that are highly correlated with key outcome measures – again, so long as their corresponding benchmarks are secured from reliable sources. While inclusion of more variables during weight adjustments can increase the resulting unequal weighting effects, and, hence, reduce the effective sample size, this compromise may have a simple justification. After all, when a survey has over 90 percent nonresponse, it would be an easy argument to make that its corresponding pool of respondents are too skewed to be balanced adequately with simple geodemographic weighting. As such, more aggressive weighting/calibration adjustments may be all but inevitable to compensate for the growing rates of nonresponse.

References

American Association for Public Opinion Research (2010), *New Considerations for Survey Researchers When Planning and Conducting Rdd Telephone Surveys in the U.S. With Respondents Reached Via Cell Phone Numbers*, AAPOR Cell Phone Task Force report. Available at http://aapor.org/Cell_Phone_Task_Force.htm

Bankier, M.D. (1986), "Estimators Based on Several Stratified Samples With Applications to Multiple Frame Surveys," *Journal of the American Statistical Association*, 81, 1074-1079.

Brick, M., I. Flores Cervantes, S. Lee, and G. Norman (2011), "Nonsampling Errors in Dual Frame Telephone Surveys," *Survey Methodology*, *37*(1), 1-12.

Buskirk, T.D., and S. Kolenikov (2015), "Finding Respondents in the Forest: A Comparison of Logistic Regression and Random Forest Models for Response Propensity Weighting and Stratification," *Survey Insights: Methods from the Field, Weighting: Practical Issues and 'How to' Approach*. Available at: http://surveyinsights.org/?p=5108

Deville, J.C., and C.E. Särndal (1992), "Calibration Estimation in Survey Sampling," *Journal of American Statistical Association*, 87, 376-382.

Fahimi, M. (2014), "Practical Guidelines for Dual-Frame RDD (Now That the Dust Is Settling)," *Survey Practice*, *7*(2).

Frankel, M.R., M.P. Battaglia, M. Link, and A.H. Mokdad (2007), *Integrating Cell Phone Numbers into Random Digit-Dialed (RDD) Landline Surveys*, Proceedings of the Survey Research Methods Section, 3793-3800, American Statistical Association, Alexandria, VA.

Hartley, H.O. (1962), *Multiple Frame Surveys*, Proceedings of the Social Statistics Section, 203-206, American Statistical Association, Alexandria, VA.

Hartley, H.O. (1974), "Multiple Frame Methodology and Selected Applications," *Sankhya*, 36, 99-118.

Kalton, G., and D.W. Anderson (1986), "Sampling Rare Populations," *Journal of the Royal Statistical Society*, Series A, 149, 65-82.

Kalton, G., and I. Flores Cervantes (2003), "Weighting Methods," *Journal of Official Statistics, 19*(2), 81-97.

Lavrakas, P.J., S. Blumberg, M. Battaglia, J. Boyle, J.M. Brick, T. Buskirk, C. DiSogra, D. Dutwin, M. Fahimi, H. Fienberg, A. Fleeman, T. Guterbock, J. Hall, S. Keeter, C. Kennedy, M. Link, L. Piekarski, C. Shuttles, C. Steeh,

T. Tompson, and R. ZuWallack (2010), *New Considerations for Survey Researchers When Planning and Conducting RDD Telephone Surveys in the U.S. with Respondents Reached via Their Cell Phone Numbers*, Deerfield, IL: American Association for Public Opinion Research, Available at: https://www.aapor.org/AAPOR_Main/media/MainSiteFiles/2010AAPORCellPhoneTFReport.pdf

Peytchev, A., and N. Neel (2012), "Feasibility of a Single-Frame Cell Phone RDD Sampling Design," Proceedings of the American Statistical Association, Section of Survey Research, San Diego, CA.

Särndal, C., B. Swensson, and J. Wretman (1992), *Model Assisted Survey Sampling,* New York: Springer.

APPENDIX D[15]

TRENDS IN TELEPHONE OUTCOMES, 2008–2015

David Dutwin, SSRS/SSRS

Paul J. Lavrakas, Independent Consultant

ABSTRACT: The appendix documents trends in telephone survey dispositions for many major survey firms since the start of the "cell phone period" of survey research, roughly 2008 to the present. Findings for this period show a stark decline in productivity on landlines with some encouraging news for the future of cell phone research.

There is a considerable degree of consternation about the future of telephone surveys. Evidence from the field has documented a significant drop in response rates, by as much as a factor of four in the past 15 years. But the trend has run even longer than that. Curtin and colleagues measured response over time in the Survey of Consumer Attitudes (SCA) (2000, 2005), finding random digit dial (RDD) response rates of 72 percent in 1979 and a linear decline to 48 percent in 2003. By 2013, the SCA response rate had dropped to 16 percent. Many others have noted the same trends in other RDD studies (Bradburn 1992; Steeh 1981; Tourangeau and Plewes 2013). Of course, perhaps the most cited article on nonresponse in these kinds of survey is the Pew report (2012) on trends in their own RDD studies, which documents a response rate decline from 35 percent to 9 percent from 1997 to 2012.

These reports from the field make clear that we are seeing an unprecedented drop in response rates, particularly in the past decade. But many questions remain unanswered: What specific changes in telephone dispositions are causing the drop in response? Are trends in RDD response rate moving in lockstep for landlines and cell phones? And what is the net result of the drop in response rate on effort and cost of RDD telephone surveys?

As part of the American Association for Public Opinion Research (AAPOR) Future of Telephone Surveying Task Force, we volunteered to lead an effort to gather data to document the response rates of a number of polling firms, and in addition to gather data on specific response rate outcomes in order to investigate the questions posed above.

Data were requested from a number of well-known firms. In order to attain relatively equivalent data, the request was constrained to RDD dual-frame surveys of the general population that, if not full trend studies, are nevertheless repeated cross-section in which the method is consistent from year to year. An example would be Pew political surveys or ABC polls, and although the topic varies by study, place great importance on maintaining a consistent methodology from study to study. Also important was that each study utilized a dual-frame RDD telephone methodology from at least 2009, so that trend data could be analyzed separately for each frame.

[15] An earlier version of this Appendix appeared in *Survey Practice*: Dutwin, D. and P. J. Lavrakas. 2016. Trends in Telephone Outcomes, 2008–2015. *Survey Practice,* 9 (2). ISSN: 2168-0094

Overall, 15 large and prestigious research firms[16] were asked if they could provide detailed dispositions by year and by frame, preferably from 2007 to 2015. Eight firms reported that they did not have data that fit our criteria, and the remaining provided what information they had available. The net result is data from seven firms, whose data are summarized below in Table 1.

TABLE 1. PARTICIPATING STUDY SURVEYS

DATA	STUDY	SCRUB[A] LL	SCRUB CELL	FIRST YEAR	LAST YEAR	LL SAMPLE	CELL SAMPLE
ABC	ABC polls	Biz Purge 2010–2015	None	2008	2015	259,677	188,177
Gallup	Gallup daily tracking surveys	None	None	2009	2015	18,490,017	14,465,292
GfK	AP polls	Aug 2012 to present	None	2009	2014	434,405	100,586
NBC	NBC polls			2012	2015	125,382	140,384
PSRAI	Pew omnibus	Biz Purge	None	2010	2015	285,708	165,711
Pew	Pew internet & American Life polls	Biz Purge	None	2007	2015	369,301	185,385
RTI	Survey of consumer attitudes	None	None	2010	2013	197,878	432,149
SRBI	Confidential	Biz Purge	Inactive 2014+	2007	2014	280,880	85,329
SSRS	SSRS omnibus	MSG ID+	None	2009	2015	696,688	622,684

[a] "Scrub" refers to processes use in the sample generation process to eliminate numbers known or thought to be known as business or nonworking numbers.

Because not every firm provided data for every year and because of the small sample of companies, the analyses must be interpreted with some caution, and indeed, we provide graphs that show in the background results from each individual firm, and make note of trends that go against the norm of other firms, as well as document the impact on the overall trend if outlier firms are excluded. As well, some firms utilize scrubbing procedures for their samples while others do not. Again, we make note of when such data affect the overall trend. Finally, the reported trend begins in 2008 since only two firms provided data for 2007 (mainly because most firms did not begin large-scale cell phone dialing until 2008).

Of course, the first step is to confirm that response rates are in fact declining for the firms in the study. Response rates are provided in Table 2, and indeed, they are declining during the study time period. Landline rates decline from an average of 15.7 percent in 2008 to 9.3 percent in 2015 (a relative decline of 41 percent), while cell phone response rates decline at the same rate, from 11.7 percent to 7.0 percent (a relative decline of 40 percent).[17]

16. ABC, CBS, Gallup, GfK, ICF, Ipsos Nielsen, NORC, ORC, Pew, RAND, RTI, SRBI, TNS, Westat.
17. Firms are not all consistent in how they categorize outcome dispositions into response rate calculators, nor are they consistent in how they deal with screening on cell phones. For example, some firms delineate between answering machines in which a household is confirmed vs. those in which it is unclear whether

TABLE 2. RESPONSE RATES

	LANDLINE	CELL
2008	15.7%	11.7%
2009	13.7%	10.3%
2010	13.0%	11.2%
2011	13.6%	10.4%
2012	10.9%	7.2%
2013	9.8%	6.9%
2014	8.2%	6.5%
2015	9.3%	7.0%

Having confirmed the expected drop in response rate in the data, we turn to trends in dispositions. We report four principal statistics. The first is a traditional refusal rate, AAPOR REF3 [R/((I+P)+(R+NC+O)+e(UH+UO))] (AAPOR Standard Definitions 2016). Second is a combined no answer/answering machine rate which is defined as [(NA+AM)/total sample]. The nonworking rate is simply nonworking sample/total sample, and the yield is total sample/ completed interviews.

First are REF3 refusal rate trends; see Figures 1 and 2:

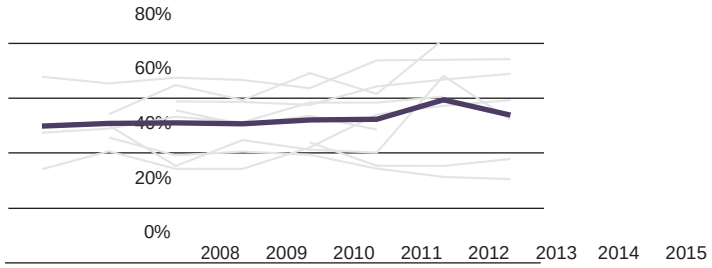

FIGURE 1. REFUSAL RATE, LANDLINES

the answering machine is for a household (vs. a business phone), while others clump answering machines into one disposition. The response rates here have been calculated fresh, using the same categorization rules for every firm.

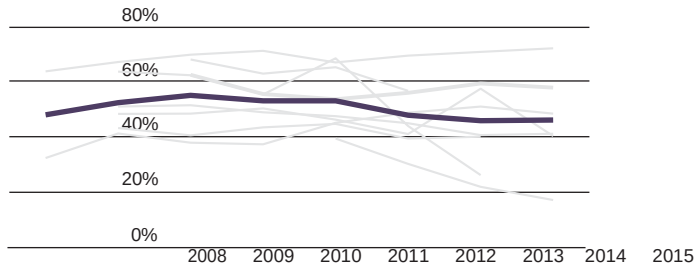

FIGURE 2. REFUSAL RATE, CELL PHONES

Perhaps surprisingly, there is evidence of only a small increase in refusals for landlines, and, in fact, a decline on cell phones. Data from each firm is largely linear and consistent with other firms.

On the other hand, as shown in Figures 3 and 4, there is some upward trending of no answer and answering machine rates. Specifically, no answer/ answering machine rates have increased 10 percentage points in the past 8 years for landlines and 24 percentage points for cell phones. There is a notable drop in this metric in one firm for 2014, but elimination of this data point does not impact the 2015 trend and does not significantly affect the overall pattern. Generally, firms were again largely consistent with one another in their trends, though the actual reported percent of no answer and answering machine dispositions differed significantly.

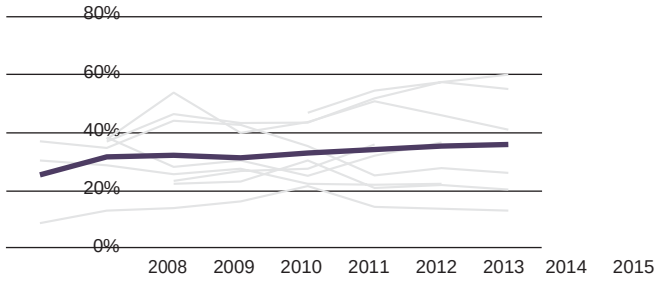

FIGURE 3. NA/AM RATE, LANDLINES

FIGURE 4. NA/AM RATE, CELL PHONES

With regard to nonworking rates, as shown in Figures 5 and 6, the trends on each frame are moving in opposite directions. The nonworking rate on landlines has increased from 28 percent to 40 percent in the study timespan (a relative increase of 43 percent), while on cell phones, the rate has dropped from 39 percent to 24 percent (a relative decrease of 38 percent). There was considerable variance in the individual reports on nonworking numbers, unsurprising given that some firms utilize at least business purging on landlines and others do not. Because the nonworking rates of the NBC and RTI data were inconsistent with other data and that these data run only partial years, the figure trendline excludes their data.

Finally, there is the summative measure of yield (Figures 7 and 8). Yield is the number of sample pieces required to secure a single completed interview (completes/total sample). Yields on landlines have been decreasing significantly through the study timeline, from a high of about 0.08 (14 sample records to attain a single interview) in 2008 to 0.03 in 2015 (46 sample records, a relative increase in sample needed of 329 percent). However, yields on cell phone have declined at a more modest pace. In 2008–2011, there was an average yield of 0.05 (17 sample records per interview). Yields then dropped to 0.04 (36 sample records) in the time period spanning 2012–2015, a relative decrease in yield of 47 percent). There is no clear explanation for the pattern of response here, the potential for there to be two levels of response for cell phones, and the general flat distribution within each time period. Taken as a whole, yields have undoubtedly declined for cell phones, though there is some small ray of hope in that yields have been somewhat flat since 2012.

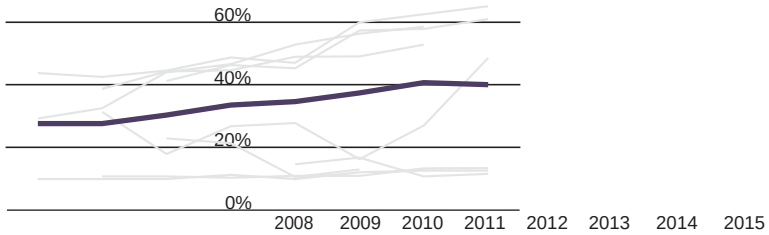

FIGURE 5. NW RATE, LANDLINES

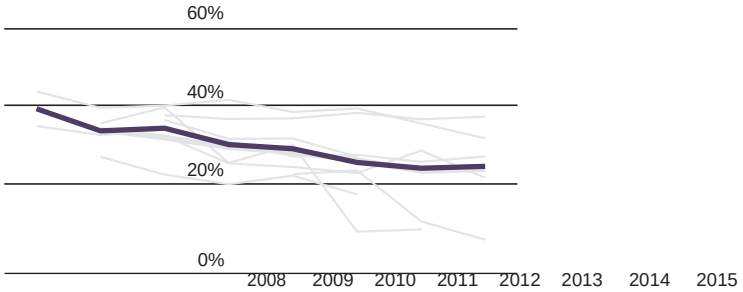

FIGURE 6. NW RATE, CELL PHONES

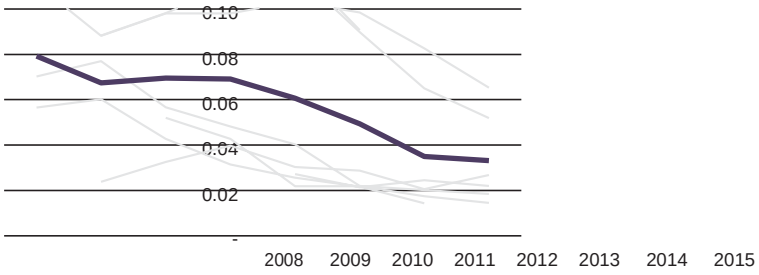

FIGURE 7. YIELD, LANDLINES

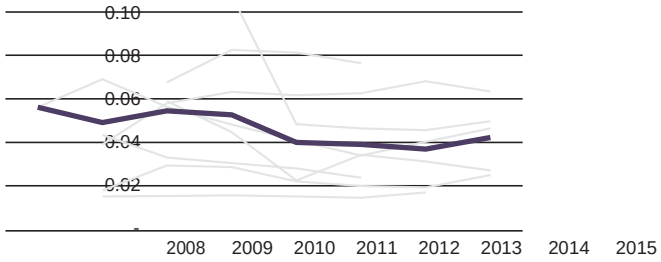

FIGURE 8. YIELD, CELL PHONES

Overall trends paint a fairly different picture for landlines and cells. It is fairly well understood and clear that the landline frame is living on borrowed time in terms of coverage, as the number of households with landlines is now in less than 50 percent with no evidence that the saturation will not continue to decline over time. Indeed, projecting the quite linear trend in Blumberg and Luke's data (2015) suggests that landline ownership would be under 10 percent in about 10 years. The data reported here suggest that the impact of the loss of coverage is an increase in nonworking numbers. Simply put, as households have eschewed their landlines, the number of nonworking numbers in valid 1+ telephone banks is growing.

The data, however, show that there is more to the story with regard to landlines. Yield has been cut in half in 8 years, the result of not just declines in working rates but increases in refusals, modestly, and increases in screening, more substantially, as evidenced by the growth of the no answer/answering machine rate. A breakdown of these two types of dispositions (not shown) reveals that both are decreasing at about the same rate. As it is likely that most no answers are "hidden" nonworking numbers (see Smith 2009), part of the increase in the no answer rate suggests an even stronger impact of nonworking numbers. Nevertheless, answering machine dispositions have nearly doubled on landlines from 2009 to 2015 (21 percent to 39 percent). Clearly, either those households remaining with landlines were more likely owners of answering machines than those who have dropped landlines, or they have become more likely to screen calls from unknown numbers, or both.

With regard to cell phones, thankfully for the future of telephonic research, the news is much better, though there is some reason for concern. Cell phone yields have dropped relatively modestly and much less than for landlines. That said, the devil is in the details. Nonworking rates have fallen, likely because the number of available exchanges has grown little in the past 8 years while the number of numbers going to working phones has increased in line with increased cell phone penetration, according to industry experts.

At the same time, however, the no answer/answering machine rate has increased significantly. Unlike for landlines, it is less plausible to think that "no answer" dispositions are evidence of an increased nonworking rate since that runs counter to the overall nonworking rate. No, it is in the authors' experience that no answer dispositions are commonly hidden voice mail dispositions on cell phones, as many cell phones will not transfer to voice mail until after the fourth ring (and even fifth) and many call centers do not wait that long before abandoning a call attempt. In other words,

it seems clear that the American public, while not necessarily refusing to participate in cell phone surveys more than in the past decade (as evidenced by the slightly declining refusal rate on cell phones), is increasingly avoiding calls from unknown telephone numbers by just letting them go to voice mail.

The million dollar question, then, with regard to cell phones is whether the increase in the no answer/answering machine rate will continue to increase over time, for it is the principal, indeed, the only clear source for lower response rates on cell phones. It is notable that for about half the firms, this rate has been flat for a number of years, while for others it has continued to rise. Given this variation across firms, it is difficult to predict the future of cell phone research on this issue. If the firms that have seen increased no answer/answering machine rates fall in line with those who have enjoyed flat rates, then the future of cell phones would look much like it does today. If the converse is true, however, cell phones will continue to see lower response rates and decreased productivity.

Overall, however, it is encouraging that nonworking rates on cell phones are declining and refusal rates are remaining steady, if not slightly declining. Only time will tell how long cell phones will continue to remain viable as a mode to sample and recruit survey respondents, but the evidence reported here suggest that is more likely to hold true for the near future than not.

References

American Association for Public Opinion Research (2016), *Standard Definitions: Final Dispositions of Case Codes and Outcome Rates for Surveys (9th Ed.),* American Association for Public Opinion Research, Ann Arbor, MI.

Blumberg, S.J., and J.V. Luke (2015), *Wireless Substitution: Early Release of Estimates from the National Health Interview Survey, July–December 2012,* National Center for Health Statistics. June 2013. Available at http://www.cdc.gov/nchs/nhis.htm.

Bradburn, N. (1992), "A Response to the Nonresponse Problem. 1992 AAPOR Presidential Address," *Public Opinion Quarterly*, *56*(3), 391–397.

Curtin, R., S. Presser, and E. Singer (2000), "The effects of Response Rate Changes on the Index of Consumer Sentiment," *Public Opinion Quarterly*, *64*(4), 413–428.

Curtin, R., S. Presser, and E. Singer (2005), "Changes in Telephone Survey Nonresponse Over the Past Quarter Century," *Public Opinion Quarterly*, *69*(1), 87–98.

Pew Research Center (2012), *Assessing the Representativeness of Public Opinion Surveys,* Available at http://www.people-press.org/2012/05/15/assessing-the-representativeness-of-public-opinion-surveys/.

Smith, T. (2009). *A Revised Review of Methods to Estimate the Status of Cases With Unknown Eligibility,* American Association for Public Opinion Research, Available at https://www.aapor.org/AAPOR_Main/media/MainSiteFiles/ FindingE.pdf.

Steeh, C.G. (1981), "Trends in Nonresponse Rates, 1952–1979," *Public Opinion Quarterly*, *45*(1), 40–57.

Tourangeau, R., and T.J. Plewes, eds. (2013), *Nonresponse in Social Science Surveys: A Research Agenda*, A Report by the National Research Council of the National Academies, Washington, DC: National Academies Press.

APPENDIX E

THE CHANGING COSTS OF RDD CELL PHONE INTERVIEWING

Tom Guterbock, University of Virginia
Grant Benson, University of Michigan
Paul J. Lavrakas, Independent Consultant

ABSTRACT: Using data on dual-frame telephone surveys collected in three waves from a large number of US survey organizations, this appendix discusses the current cost ratio of cell phone RDD interviewing versus landline RDD surveying, and shows the recent trends in costs and cost ratios, and examines some of the key factors that cause variation in the cost of cell phone RDD surveying versus landline RDD surveying. Although the overall cost of RDD telephone interviewing has increased substantially over the last several decades, it is now not uncommon for a dual frame survey to report cell phone RDD surveying to be equal to or even lower in cost than landline RDD surveying.

From the first experiments with cell phone interviewing through the present, it has been generally understood that the cost of completing an interview with a random digit dial (RDD) cell phone sample is substantially higher than the cost of a completion with an RDD landline sample. However, in the 10 years that have passed since those first experiments, there have been substantial changes in the system of telephony and in available sampling methods in the United States. These changes potentially affect the cost of cell phone interviewing and/or the cost of landline interviewing. These shifts include continuing changes in the ways that members of the public respond to telephone calls from survey organizations, changes in the operational and technological choices made by survey organizations, and changes in the commercially available methods for drawing samples from the cell phone frame.

Using newly collected data on recent dual frame surveys from a large number of survey organizations, this report discusses the current cost ratio of cell phone RDD interviewing versus landline interviewing, shows the recent trends in costs and cost ratios, and examines some of the key factors that cause variation in the cost of cell phone interviewing versus landline interviewing. While the overall cost of telephone interviewing has increased substantially over the last several decades, it is now not uncommon for a dual frame survey to report cell phone RDD interviewing to be equal to or even lower in cost than landline RDD interviewing.

1. MEASURING THE COST TREND

Survey organizations operate in a competitive business environment and are not disposed to freely disclose their costs of operations to outsiders. But interviewing costs are closely tied to response rates, sample yields, completions per hour, and other related production measures that are often disclosed routinely or that may be shared by organizations interested in assessing production trends. Several research designs can be applied to assess changes over time in survey costs or production rates. One strategy is to focus on a single, recurrent survey project that uses a fairly constant method to survey the same population year after year, as reported, for example, by Martonik et al. (2016). This design holds constant many of the variables and factors that affect interviewing costs, allowing the trends in production rates and associated costs to be seen as a function of trends external to the (fixed) survey design and data collection operation (e.g., changes

in working number rates or trends in cooperation rates). The limitation of this design is that one cannot be sure what the trends might look like in a survey of a different design, a survey with a different instrument, sponsor, or topical focus, or one applied to a different study population. The recent work of David Dutwin and Paul Lavrakas (2016) offers an important elaboration on this design. They examine multiple longitudinal dual frame studies, each of which has a relatively fixed design over time but in aggregate represent various sponsors and somewhat varying designs. (The results of that study are presented in Appendix C.) All of the studies included in their report are recurrent national surveys conducted by large organizations that agreed to share their data with the researchers.

This section relies on data collected via a different design: a trend study including three waves of data collection. In each wave, an independent sample of dual frame surveys was reported on by a diverse range of U.S. survey organizations.

The first wave was conducted during late 2009 by the Cost subcommittee of the second AAPOR Cell Phone Task Force to gather data from a select group of telephone survey organizations in the United States (Guterbock, Lavrakas, Tompson, and ZuWallack 2010). This survey used a purposive nonprobability sample of eight nationally known survey organizations (four commercial and four academic).[18] Members of the Cost subcommittee conducted telephone interviews with a senior researcher at each organization who was knowledgeable about the cost information that was to be gathered. Each of those senior researchers then submitted a spreadsheet containing available information about each of the dual telephone frame surveys that the organization had conducted. These spreadsheets were shared in confidence with the Cost subcommittee members, providing production information for 38 separate dual frame RDD surveys. These surveys represented a mix of national, state, and local surveys. The type of information that was gathered about the RDD cell and RDD landline samples in these surveys included: (1) number of completions, (2) average length of a completion in minutes, (3) geography covered, (4) screening criteria, (5) completes per hour (CPH), (6) incentive amounts, and (7) cost per interview (CPI), where available. These data were used to generate various ratios for the analyses reported in the 2010 Task Force Report.

The second wave of data collection was conducted by Guterbock, Peytchev, and Rexrode (2013), using an online questionnaire that included the same items along with other, more specific production details. They sent requests for participation to people in 98 survey organizations, using two sources to develop their sample list. They compiled a list of researchers who had reported on cell phone surveys in the three preceding annual meetings of American Association for Public Opinion Research (AAPOR), and combined that with the list of member organizations of the Association of Academic Survey Research Organizations (AASRO). Participating organizations could each report on up to three different dual frame surveys. The survey instrument invited detailed breakdowns of production details for each sample (total dialings, total numbers attempted, number of interviewing hours, number of completes). Recognizing the sensitivity of production details (a concern especially salient for commercial organizations), the questionnaire also allowed organizations to simply provide, for each study, the overall ratio of completions per hour for cell phones over the completions for hour for landlines (the CPH ratio) and withhold the additional detail. This effort succeeded in gathering usable data on 37 dual frame surveys from 27 different organizations.

A third wave of data collection was undertaken for the present report, using a virtually identical online instrument to that used in 2013 (Guterbock et al. 2016). As in 2013, the primary sampling

18. The survey organizations were promised their names would not be disclosed.

frame was based on a list of authors who had presented cell phone studies at the previous two AAPOR meetings, supplemented by the membership list of AASRO. In an effort to expand the study's coverage, the study team obtained cooperation from two commercial sampling companies, Marketing Systems Group and ASDE Survey Sampler,[19] each of whom circulated a survey invitation and a reminder letter to all of their clients who had recently purchased RDD cell phone samples from their firm. These survey organizations were invited to respond using an anonymous link to the survey instrument, and they were allowed to respond without identifying their firm name if they chose. The 2016 survey gathered data on 53 dual frame studies conducted by 25 survey organizations, 11 of which were commercial firms. Altogether the surveys conducted in 2010, 2013, and 2016 provide CPH ratio data on 118 dual frame RDD studies, but since the detailed production statistics are lacking for the 2010 cohort of studies, as well as for firms that elected to provide only their CPH ratios, some of the results reported here rely on the 54 studies reported in 2013 and 2016 with full production details.

2. BREAKING DOWN INTERVIEWING COSTS

The 2010 Task Force report laid out many of the reasons why cell phone interviewing has tended to cost substantially more per interview than traditional landline interviewing (Lavrakas et al 2010, pp. 95-102). As discussed there, the per-interview cost of an RDD interview can be broken down into the cost of interviewer time required to obtain a completion, the cost of the sample numbers used to achieve that completion, the cost of any remuneration or incentives given to respondents, and the cost of any possible mailings sent to respondents. By far the largest of these cost components is interviewer time, which is the fundamental cost unit of telephone survey budgets. An hour of an interviewer's time in the computer-assisted telephone interviewing (CATI) center entails not only that employee's wages and benefits but also a portion of the time of the supervisors, an hour of phone dialing and accrual of telephone charges, and other infrastructure overhead. The main reason that cell phone interviews cost more than landline interviews is because the former have required substantially more interviewer time to complete.

As discussed in the 2010 Task Force Report, there are several factors that can affect the interviewing production rate, measured as hours per completion (HPC = 1/CPH). These include the method of dialing used, the length of the interview, the number of dialing attempts devoted to each sampled phone number, and four independent properties of the telephone number sample and frame from which it is drawn: the working number rate, the contact rate, the eligibility rate, and the cooperation rate. (The product of these four rates is the sample yield—the ratio of completed interviews to the total number of sampled numbers attempted in the study.) Changes over time in any of the above factors can affect the production rate for either cell phone or landline calling, and if the change affects one type of sample more than the other, then the cost ratio of cell phone calling vs. landline calling will be affected.

3. WHY COST RATIOS MAY BE CHANGING

Dialing Rate. One reason cell phone interviewing takes longer than landline interviewing is that cell phone numbers must be manually dialed in the United States, while landline numbers can legally be dialed with auto-dialing equipment. With the Federal Communications Commission's

19. Special thanks to Trent Buskirk at M-S-G and Michel Durocher at ADSE Survey Sampler for their gracious assistance with our survey recruitment effort.

(FCC's) recent clarifications of the Telephone Consumer Protection Act (TCPA) and the increased risk of costly litigation for alleged TCPA violations, survey organizations may have changed their dialing practices in recent years. On the cell phone side, where telephone labs may formerly have used "power-dialing" or "one-finger" dialing methods that allowed each phone number to be dialed with one press of the key at the initiation of the interviewer, they now are more likely to dial all 10 digits of the phone number manually. On the landline side, there is a risk of unintentionally reaching a cell phone number by dialing a landline number that was recently ported to a cell phone number. To avoid this risk, some survey labs may have abandoned predictive dialing for landline samples on some or all of their studies. As will be seen below, the use of a predictive dialer on the landline side has a strong effect on the HPC ratio because it makes landline calling more efficient (more completions per hour) than cell phone calling. Thus, changes in dialing practice on either the cell phone side or the landline side could change the HPC ratios and hence the cost ratio.

Interview Length. When survey organizations first began to call cell phones, there was broad concern that respondents would not be willing to do interviews of standard length on their cell phones. This was based in part on the fact that many cell phone users were on plans that required them to pay for cell phone usage by the minute. As more cell phone users have moved to unlimited plans (or plans without limits for evening and weekend calling, when most interviewing takes place), survey organizations have found that cell phone interviews can be of equal length to landline interviews. When cell phone interviews are no longer shortened (as compared to the landline version), the HPC ratio and cost ratios will be somewhat increased as the cell phone calling becomes a bit less efficient.

Working Number Rate. A fundamental property of an RDD sampling frame is that it may contain nonworking numbers. In the early years of cell phone calling, the proportion of nonworking numbers in the cell phone frame was far higher than in the landline frame. In most cases, list-assisted landline RDD samples select only "working" number blocks into the sampling frame, resulting in greater efficiency. In contrast, no such selection was initially possible for RDD cell phone samples because there were no publicly available directories or other sources listing cell phone subscribers. In addition, landline RDD samples could be prescreened to eliminate nonworking and business numbers, yielding a significant gain in calling efficiency, but this prescreening was not possible for RDD cell phone numbers in the United States. The overall result was a much lower working number rate for cell phone RDD samples.

These circumstances have changed considerably, as sampling companies now offer products that identify cell phone numbers that are currently (or recently) active, or services that "ping" candidate RDD cell numbers to see if they are in service. These higher-priced "enhanced" cell phone samples provide much higher working number rates in the delivered cell phone sample, generating cost savings from higher CPH rates that more than offset the additional sample cost in most applications.

In addition, the working number rate is a function of working number density in the frame—that is, in the number blocks from which sample is drawn. The differential in the working number rate is dependent on the relative density of working numbers within the cell phone and landline exchanges in use in a given sampling area. With the continued migration of the U.S. population to wireless-only service, the working number rate in the landline frame is falling, while the existing cell phone number banks are continuing to fill up with subscribers, increasing the working number rate. As reported by Dutwin and Lavrakas in their analysis of longitudinal survey production data from seven organizations (2016; Figures 5 and 6), in the period from 2008 to 2015 there was a

substantial increase in the percentage of nonworking numbers in the landline frame and a substantial decrease in nonworking percentages in the cell phone frame.

Contact Rate. The contact rate is affected by cultural and technical differences in how people in the United States use cell phones as contrasted with how they use household landline phones. Initially, many people used cell phones as a supplementary communication device. Many cell phones that interviewers called were turned off when called. However, as more cell phone users migrate into the cell phone-only and cell phone-mostly categories, this usage pattern declines in frequency. It is known that cell phone-only users are more likely to answer their cell phone calls than those who have dual service (Guterbock 2009, Brick et al 2011). As the proportion of cell phone-only subscribers increases within the cell phone user population, contact rates ought to increase. In fact, it has become ever more common for dual-service phone users to ignore calls to their landline phones, expecting any really important call to come to their cell phones; this trend will reduce the landline contact rate. Call screening/caller ID technology is essentially universal on cell phones, as is voicemail; both are thought to promote greater screening of incoming calls by respondents. The result may have been a lower live contact rate for the cell phone numbers that are sampled. It is also quite possible that cell phone users in the early years expected any call on the cell phone to be from someone known to them, and they may have tended to answer unknown callers on the expectation that the caller was an acquaintance using a different phone number. If cell phone users are now experiencing more unsolicited calls from unknown parties, including survey organizations, they may be becoming more selective in which calls they choose to answer.

The contact rate is in part a result of the calling effort, and some survey designs specify a lower maximum number of attempts for numbers in the cell phone sample. If this practice is decreasing and calling effort is becoming more equal across the frames, then the contact rate on the cell phone side will be increased relative to the landline contact rate.

Dutwin and Lavrakas report that the percentage of no-answer and answering machine calls is going up in both the landline calls and in cell phone calls, with the latter increasing somewhat more rapidly (2016: Figures 3 and 4).

Eligibility Rate. In most U.S. dual frame designs, the cell phone sampling will have a lower overall eligibility rate, resulting in more time spent screening and recruiting for the cell phone side, a higher HPC rate for cell phones, and a larger cost ratio. One reason is age eligibility since so many minors (persons under 18) have their own cell phones. However, as more older people migrate to cell phone-only or cell-mostly status, the percentage of ineligible minors among sampled cell phone numbers may be declining, somewhat lessening the impact of this source of ineligibility. In surveys within a single state or locality, another source of ineligibility is geography; many people reached by cell phone using the initial RDD cell phone methods would turn out to reside outside the state or the survey area. However, sampling companies are now offering cell phone sample products that identify—for many of the sampled numbers—the ZIP code of the cell phone subscriber's billing address. Local dual frame studies that use cell phone samples with appended geographic information will experience less of a differential in rates of ineligibility due to geography than those using samples lacking this information.

Another reason that eligibility rates were initially lower for cell phones was that some prominent survey organizations used a screening study design in their dual frame studies, in which each cell phone household was screened for dual telephone usage, with all or most of the dual-usage households being dropped as ineligible. Over the years, the screening design has become less and less common and is now not commonly used. This change in practice again means that

eligibility rates for cell phone samples will be less different from the landline rate than they once were.

It must be noted that the difference in eligibility rates between cell phone and landline samples will depend on study-specific screening requirements. For example, a survey seeking young Hispanic males likely will reach more eligible cases with cell phones than on landlines; the opposite would be true for a survey of married female retirees. There is no reason to think that the mix of study-specific eligibility requirements in dual frame telephone studies is changing markedly over time, but the composition of the cell phone user and landline user populations is continuing to change. As the cell phone population begins to look more and more like the population in general, it is probable that eligibility rates in the cell phone frame will go up. Moreover, as younger people continue to eschew the landline, it is even possible that eligibility rates in the landline frame will go down for many studies.

Cooperation Rate. The 2010 Task Force Report noted that there was less difference in the cooperation rates than had been experienced earlier in the previous decade, when cell phone interviewing was beginning to be deployed in the United States (Lavrakas, et al 2010: p. 38). Dutwin and Lavrakas show that refusal rates have been roughly equal for cell phone and landline calling in the period from 2008 to 2015, and that the increases in refusal rates have been quite small over that period. Refusal and cooperation rates are measured in relation to those calls that result in live contact with the sample member. However, if one takes the larger view that, in the era of caller ID and call screening, the respondent's decision to not answer a survey organization's call could be taken as a type of refusal, then the lower contact rates we have been experiencing can also be interpreted as a decrease in cooperation. As noted above, contact rates seem to be decreasing for both cell phones and landlines, so whether one thinks of cooperation in the narrower or broader definition, the two frames remain fairly similar and there is little indication that changes in cooperation will affect the cost ratio in dual frame surveys.

4. SUMMARY OF COST FACTORS

The cost per completion for either part of a dual frame RDD survey (cell phone or landline) can be thought of as a sum of:

- The cost per completion of remuneration/incentives and/or advance mailings, if any;
- The cost per completion of the purchased sample phone numbers; and
- The interviewing costs per completion.

This third term is by far the largest factor in most telephone surveys and can be calculated as a product of the billing rate (or full cost) for an interview hour and the HPC. HPC can, in turn, be thought of as a sum of the interview length and the hours spent (per completion) on screening and recruiting (i.e., all interviewer time that is not devoted directly to completing the interview).

It is this last cost component – screening and recruiting hours per completion (SRHPC) – that was found in the 2010 Task Force Report to be markedly higher for cell phone interviewing. Any differentials in the productivity factors listed above have a direct, multiplicative effect on the ratio of SRHPC in cell phone interviewing to SRHPC in landline interviewing. The SRHPC ratio (cell phone SRHPC divided by landline SRHPC) drives the HPC ratio and the overall cost ratio as well.

5. CHANGING PRODUCTIVITY RATIOS IN U.S. DUAL FRAME RDD SURVEYS

Since some survey organizations chose not to report detailed production statistics, we have reports of HPC rates for cell phone and landline samples for 94 dual frame surveys conducted between 2010 and 2016. **Table 1** shows how these rates have changed across the three survey waves. Since the sample sizes are not large and there is much variation across studies, this table reports median values for each wave of data collection. Most notable is that the median SRHPC for cell phone samples dropped from 2.0 in the first wave to 1.0 in the most recent wave. Over the same period, the median SRHPC for landline samples went up from three-quarters of an hour to a full hour. In general, the HPC rates are about one-third of an hour higher than the corresponding SRHPC rates (since the average survey was about 20 minutes in length), except

TABLE 1. MEDIAN PRODUCTION RATES FOR CELL PHONE AND LANDLINE SAMPLES

DATA COLLECTION WAVE	SCREENING AND RECRUITMENT HOURS PER COMPLETION (SRHPC)		HOURS PER COMPLETION (HPC)	
	CELL PHONE	LANDLINE	CELL PHONE	LANDLINE
2010 (N = 27)	2.0	.77	2.3	1.2
2013 (N = 32)	1.1	.68	1.4	.91
2015 (N = 35)	1.0	1.0	1.3	1.3

that in the 2010 data collection wave some studies were using longer surveys on the landline side than on the cell phone side.

As detailed above, the three waves of data collection on productivity rates yielded usable reports on HPC ratios for 118 dual frame surveys conducted between 2010 and 2016. (These include reports from survey organizations that reported their production ratios but declined to share more detailed production statistics.) **Table 2** shows how the production ratios changed across the three survey waves.

TABLE 2. PRODUCTIVITY STATISTICS FOR DUAL FRAME RDD SURVEYS

	SCREENING AND RECRUITING HOURS PER COMPLETION (SRHPC)			HOURS PER COMPLETION (HPC)		
RATIO (CELL/LANDLINE)	SRHPC RATIO			HPC RATIO		
	2010	2013	2015	2010	2013	2015
Mean	2.5	1.7*	1.5	2.0	1.5*	1.4
Minimum	1.2	.77	.43	1.2	.85	.45
Maximum	5.4	3.7	3.8	3.5	2.9	3.3
N	27	37	50	27	38	53

Std. deviation	1.0	.71	.80	.63	.51	.64

* 2013 and 2015 are significantly different from 2010. 2015 is not significantly different from 2013.

The 27 surveys that were included in the first wave of data collection in 2010 had a mean HPC ratio of 2.0 and a mean SRHPC ratio of 2.5. That is, in the average dual frame survey at that time, the interviewer expended two-and-a-half times the hours trying to connect with a willing, eligible respondent in the cell phone sample as were expended in the landline sample. When the actual interview time is added into the numerator and denominator of the ratio, to yield the overall HPC ratio, the 2010 data showed that the HPC for cell phone calling was, on average, twice the HPC for landline calling. Just three years later, in the second wave of data collection, completed in late 2012 and reported in 2013, the average HPC ratio for the 38 surveys reporting had decreased significantly to a mean of 1.5, and the SRHPC ratio had also decreased significantly, from an average of 2.5 to an average of 1.7. In the third and most recent wave of data collection, involving 53 surveys, the HPC ratio went down slightly to a mean of 1.4, and the average SRHPC ratio also went down a little, to a new mean of 1.5. The decreases between the 2013 report and the 2015 report were not statistically significant.

Figure 1 displays the distribution of HPC ratios in the three waves of data collection. The HPC ratios in the earliest wave were widely dispersed, with some surveys reporting HPC values for cell phone calling that were more than three times the landline HPC value. The HPC values in 2013 were somewhat less dispersed, and for the first time a few surveys (four surveys in 2013) reported HPC ratios less than one, meaning that the cell phone calling was actually more efficient than the landline calling. In the most recent round of data collection, the main group of HPC values moves further down, and 20 of the 53 reporting surveys show HPC ratios below the parity level of 1.00. However, there are again several surveys reporting HPC ratios of around 3; these turn out to be surveys that used predictive dialing for landline calling, a technique that has a strong effect on the HPC ratio because it substantially lowers the HPC for the landline calling.

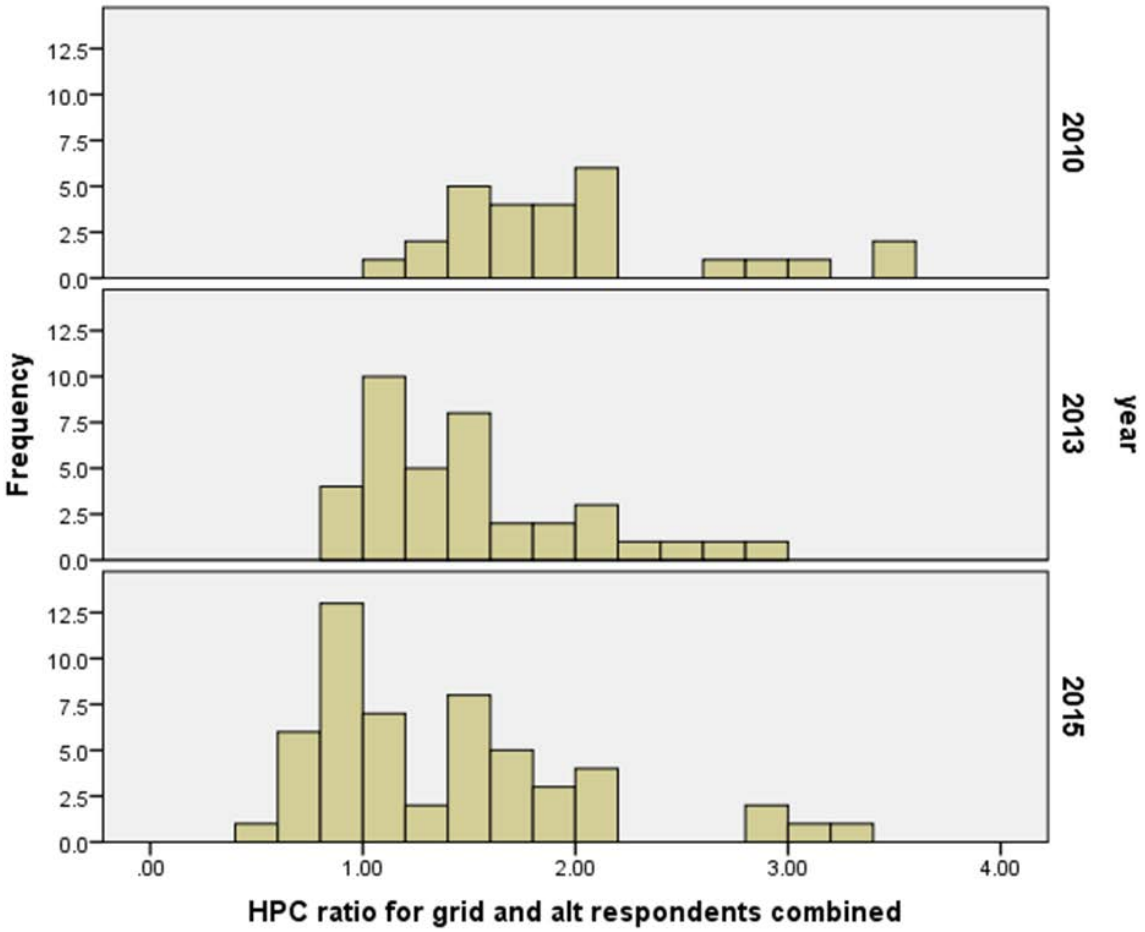

FIGURE 1. DISTRIBUTION OF HPC RATIOS IN THREE WAVES.

Figure 2 displays results for the SRHPC ratio in similar format. In general, the SRHPC ratio is higher than HPC ratio, but the distributions are similar and contrasts between waves of the data collection are similar. Again, there is a notable increase in the number of surveys in which screening and recruitment took less time on the cell phone than on the landline, resulting in an SRHPC ratio that is less than 1.00.

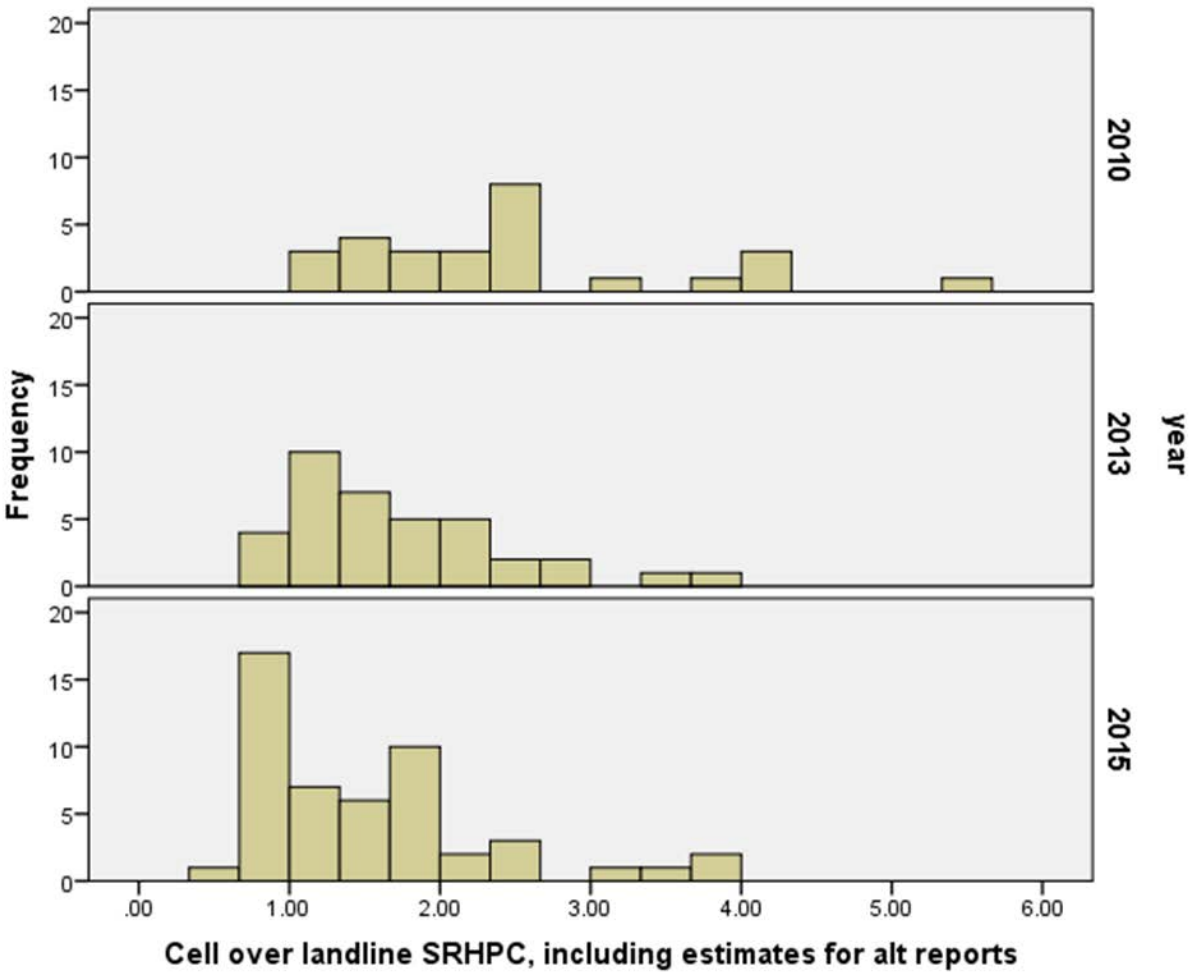

FIGURE 2. DISTRIBUTION OF SRHPC RATIOS IN THREE WAVES.

6. FACTORS THAT AFFECT PRODUCTIVITY RATIOS

To examine various factors that affect the SRHPC ratio and the HPC ratio, we analyzed the data from survey organizations that reported detailed production data in either the second or third wave of data collection—a total of 91 dual frame surveys conducted between 2011 and 2015. We looked at five factors that might affect production rates: 1) whether or not predictive dialing was used for the landline sample; 2) the geography of the survey (national, statewide, or regional/local); 3) whether the study was a part of the Centers For Disease Control and Prevention (CDC's) Behavioral Risk Factor Surveillance System (BRFSS), which requires extended dialing protocols for landline calling and uses a modified "screened" sampling procedure for selecting cell phone respondents; 4) whether cash incentives were offered to cell phone respondents, and 5) whether the cell phone sample was of the newer "enhanced" sample type that includes appended information on recent phone activity, information from "pinging" to detect whether the phone number is working, or ZIP code information for the cell phone billing address. Out of the 91 surveys reported, 25 used predictive dialing to reach landlines, 26 were national surveys, 10 were

state-level BRFSS surveys, 17 used cash incentives for their cell phone respondents, and 31 reported using some kind of enhanced cell phone sample. As seen in **Figure 3**, our data reflect the fact that the use of cash incentives for cell phone respondents has declined markedly during the study period, in part because experiments comparing surveys with and without such incentives showed little gain in the cell phone production rate when incentives were offered (Guterbock et al 2012; Oldendick and Lambries 2013). Six years ago, about 4 out of 5 cell phone surveys offered cash to cell phone participants, contrasting with only about 1 in 10 today.

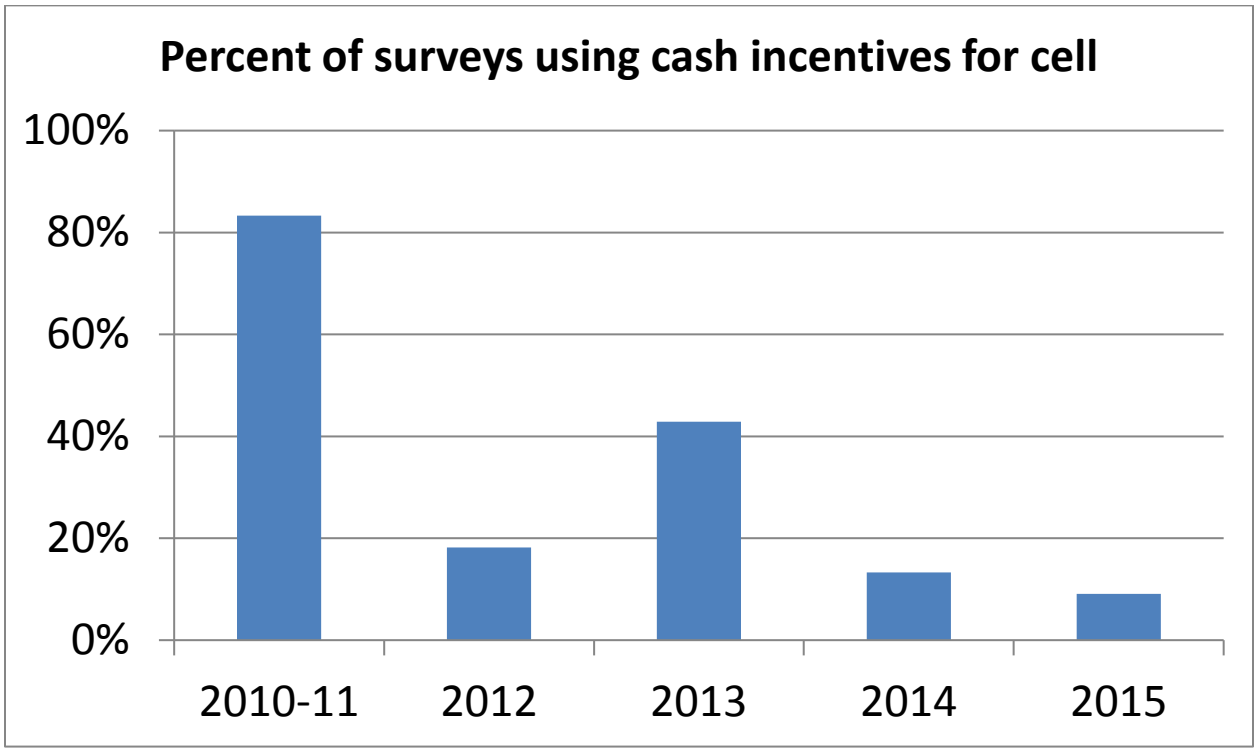

FIGURE 3. DECLINING USE OF CELL PHONE INCENTIVES.

Our analysis used ordinary least squares (OLS) regression to examine the effect of each factor on the HPC ratio and the SRHPC ratio (cell phone over landline). The results, seen in **Table 3**, are similar for both of the productivity ratios. By far the largest effect is from the use of a predictive dialer to call landline phones. Surveys in which landlines are dialed with predictive technology have lower HPC and SRHPC rates (higher calling efficiency) for landlines, thus raising the production ratios. The use of enhanced cell phone samples also has a strong effect, significantly lowering the production ratios because when these samples are used, the HPC and SRHPC rates for cell phones are lower (cell phone efficiency is higher). Sample geography has a smaller but

TABLE 3. EFFECTS OF KEY FACTORS ON PRODUCTION RATIOS (CP/LL)

FACTOR	HPC RATIO	SRHPC RATIO
	STANDARDIZED REGRESSION COEFFICIENTS	
LL Predictive dialer	.632**	.628**
National sample	-.262*	-.212+
BRFSS	-.120	-.105
CP Cash incentive	.011	.028
Enhanced CP sample	-.277**	-.290
R^2	.466	.507

N = 73. **p <.01. *p < .05 +p < .10

still significant on the production ratios, with lower production ratios in national samples as compared to state, local, or regional samples. In the latter types of geography, many cell phone numbers must be disqualified as being outside the study area, so in national surveys the cell phone efficiency is higher, resulting in somewhat lower HPC and SRHPC ratios in national samples, other factors being equal. BRFSS studies report slightly lower production ratios than do other studies (all other factors held equal), but the effect is not statistically significant. And cell phone incentives show no net effect on production rates in this analysis, reinforcing the negative results of studies that have tested cell phone incentives versus no incentives in split-half field experiments (Guterbock et al 2012; Oldendick and Lambries 2013).

Figure 4 shows a simple graphic comparison of the distribution of SRHPC ratios for surveys that used predictive dialing as compared to those using manual or interviewer-controlled methods to reach landline phones. (The survey did not ask organizations to report on how they dialed the cell phone sample, due to the legal sensitivity of that information in the current TCPA litigation environment.) Among the 79 surveys in which SRHPC and dialing mode could be determined, the mean SRHPC ratio for the studies using predictive dialers was 2.1, while the mean was 1.2 for those using manual or interviewer-controlled (one-finger) dialing. While many of the studies using manual dialing recorded SRHPC ratios less than one (meaning cell phone calling was more efficient than landline), this was not the case for any of the studies using predictive dialing. An important reason for the difference in ratios is the fact that a predictive dialing system moves some of the calling time "off the clock" of interviewer time, as the time of dialing and much of the time waiting for an answer elapses behind the scenes before the call is routed to an interviewer. However, a more detailed analysis that breaks down the SRHPC rates into separate dialing rates shows not only that predictively dialed landline samples have significantly lower average measured time given to each calling attempt but also a significantly lower number of dialing attempts per number. These differences reflect, in part, choices that calling labs and software vendors make as they set the various parameters that control their dialing systems. In addition, our data show that predictively dialed landline samples have a far lower yield (completes over count of numbers attempted), so that a larger number of landline phone numbers must be purchased for use in these studies. To put it another way: On average, the organizations that dial their landlines without predictive dialers (typically smaller labs, many

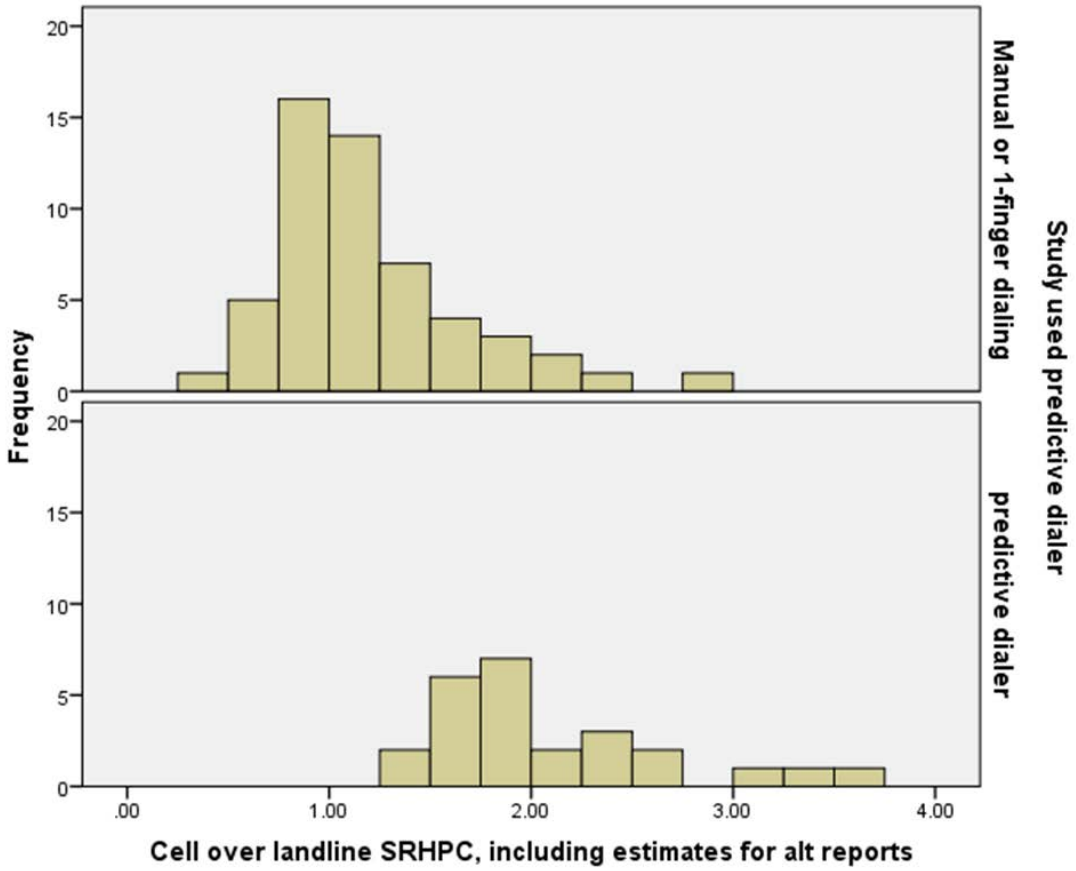

FIGURE 4. SRHPC-RATIO HISTOGRAMS FOR MANUAL DIALING VS. PREDICTIVE DIALING

of which are in the academic sector) give more time to each call attempt, give more attempts to each number, and achieve higher landline yields at the cost of a higher HPC rate for the landline numbers.

Figure 5 graphically compares the distribution of SRHPC ratios for surveys in which an enhanced cell phone sample was used compared to those in which a conventional sample was purchased. Enhanced cell phone samples include samples with an activity flag appended, those with ZIP code information appended, those screened by "pinging" to determine if a number is working, and those drawing on proprietary databases of cell phone users with known demographic characteristics. (We did not ask survey organizations to report which vendor they purchased their samples from.) Among the 81 surveys for which SRHPC data were available and cell phone sample type was reported, the mean SRHPC ratio was 1.9 for surveys using a conventional cell phone sample compared to just 1.1 for those using an enhanced sample. Out of 30 studies that used enhanced samples, 14 reported SRHPC ratios less than 1.0. On average, surveys that used enhanced cell phone samples experienced calling efficiency on the cell phone side that was nearly equal to that of the landline side.

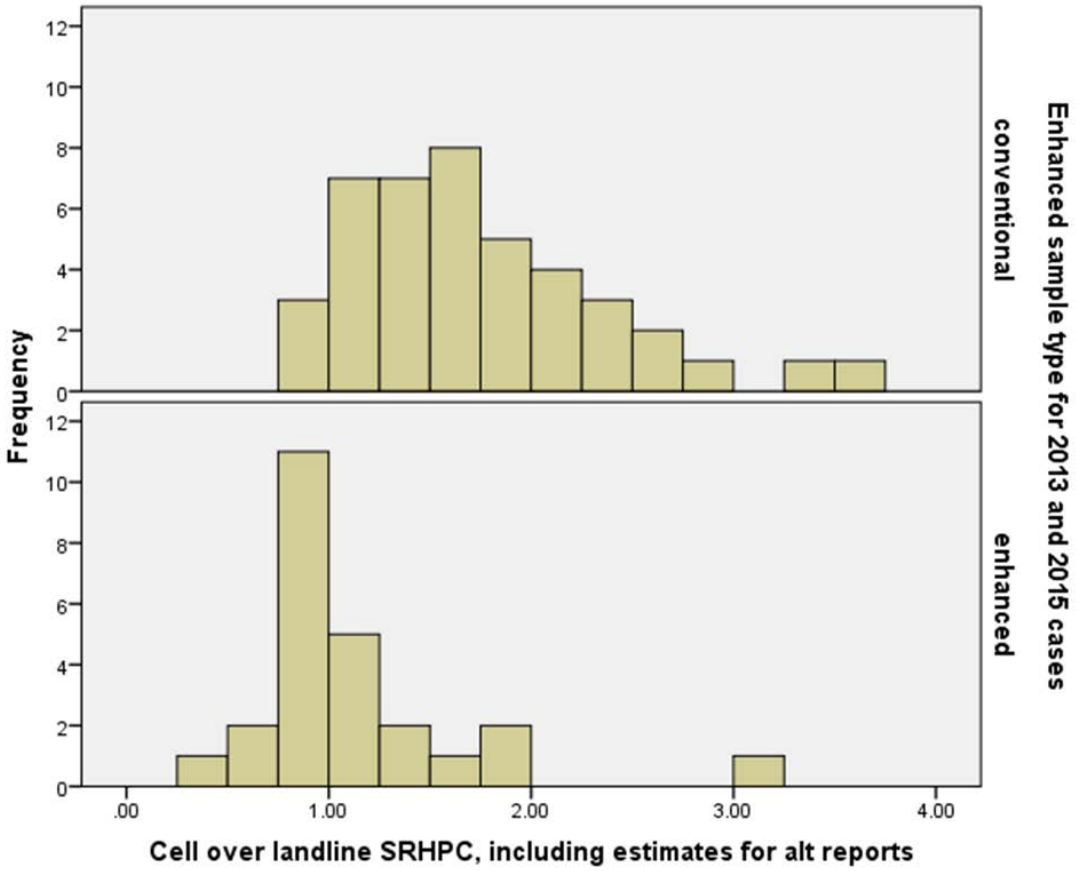

FIGURE 5. SRHPC-RATIO HISTOGRAMS FOR CONVENTIONAL AND ENHANCED CELL PHONE SAMPLES

Interestingly, in a bivariate comparison by sample geography the production rates are higher for national samples than for the state, local, and regional samples; this is mainly because many of the national studies in our data set were conducted by large national firms that use predictive dialers for landlines. As seen above in the multivariate analysis in **Table 3**, when other factors are taken into account, national studies have somewhat *lower* production ratios on average than do local, state, and regional studies because they have no need to screen out cell phone numbers based on geography.

7. OVERALL COSTS

As noted above, the cost of a telephone interview can be thought of as the sum of the cost of the hours of interviewing required to get a completion, the cost of the purchased sample numbers, and the cost of any mailings or cash incentives used. Using the data from organizations that provided detailed production information, we can calculate averages for each of these values from our collected data. To simplify the calculations, we ignore any costs of cell phone incentives, noting that these are now rarely used; advance mailings to landline households were also quite rare and so are not factored into these calculations. The per-number cost of purchased phone numbers is

our estimate based on discussions with sample vendors of representative prices for typical sample buys of various types. (Our respondents did not report the cost of their samples, but they did report the count of numbers they attempted for cell phones and landline in each survey). The averages reported in **Table 4** must be interpreted with the knowledge that all the cost factors vary widely across specific studies and situations. The cost of an hour of survey lab time varies across organizations and regions; the price of sampled numbers varies across vendors, purchase volumes and specific sample products; and most importantly, as has clearly been seen above, production rates vary widely depending on the nature of the survey and the characteristics of the study population. That said, it is still useful to consider the averages shown in **Table 4**.

TABLE 4. ESTIMATED OVERALL COST PER INTERVIEW (CPI)

	N OF SURVEYS	NUMBERS ATTEMPTED	COST PER NUMBER	SAMPLE COST PER COMPLETE	HPC	HPC COST ($30/HR)	TOTAL COST PER INTERVIEW
All CP	56	18	0.09	$1.62	1.51	$45.30	$46.92
All LL	56	14	0.07	$0.98	1.16	$34.80	$35.78
Conventional CP sample	38	24	0.07	$1.68	1.54	$46.20	$47.88
Enhanced CP sample	18	13	0.11	$1.54	1.44	$43.20	$44.74
Predictive dial LL	23	31	0.07	$2.17	0.94	$28.20	$30.37
Regular dial LL	45	13	0.07	$0.91	1.46	$43.80	$44.71

NOTE.—Cost of any cash incentives or advance mailing not included. Cost per number estimated and is meant to reflect net cost per usable number after screening is complete.

As can be seen in the top two rows of the table, cell phone interviews still cost more, on average, than landline surveys. Using representative cost figures for sample numbers (9 cents for cell sample, 7 cents for landline) and fixing the overall cost of an hour of interviewing time at $30, we arrive at estimated CPI rates of about $47 per cell phone completion compared to about $36 for a landline completion. Cell phone interviewing incurs higher costs for purchased numbers, but most of the difference in cost comes from the difference in HPC rates between the average cell phone sample and the average landline sample.

The third and fourth lines of **Table 4** examine cell phone samples only, comparing the cost of a cell phone interview using a conventional sample to one using an enhanced sample. The cost difference, although not large, favors the enhanced sample. Although the enhanced samples are priced higher per number than conventional samples, the enhanced samples have much higher yields, requiring fewer sampled numbers per completion, so that the sample cost comes out lower for the enhanced samples. On top of that, the enhanced sample produces a somewhat lower HPC rate, so that cell phone completions from enhanced samples cost about $45 versus $48 for those from conventional samples. A more detailed analysis shows that organizations in our data set using enhanced samples (many of them in the academic sector) tend to devote more time per attempt

and make more attempts per sampled number. If these factors were equal, then the HPC ratios would differ more strongly due to the higher yield of the enhanced samples.

The last two lines of **Table 4** focus only on landline calling and compare the estimated average overall cost of landline completions using a predictive dialer versus manual or interviewer-controlled dialing. While predictive dialing surveys have a far lower yield and, therefore, cost more in terms of purchased sample per complete, the far lower HPC rate for predictive dialing more than makes up for this cost disadvantage, so that the average completion using predictive dialing costs about $30 compared to $45 for manual dialing. It is no wonder that large survey labs have adopted predictive dialing technology, still, the cost savings are not solely the result of reducing labor costs for time for dialing and waiting for the phone to be answered. Organizations using predictive dialers also tend to make fewer attempts for each sampled number. The low yields they experience undoubtedly result in lower response rates for their studies, as compared to the response rates achieved by organizations that put more interviewer time into each call attempt and make more attempts per sampled number.

As discussed elsewhere in this report, the recent rulings by the FCC on the interpretation of TCPA, and the rise of litigation based on alleged TCPA violations, have caused some survey organizations to consider abandoning predictive dialing altogether. It is also clear that more and more organizations are seeing the usefulness of various kinds of enhanced cell phone samples. In the future, it may well be that the typical dual frame telephone survey will use enhanced sample for cell phones, and fully manual dialing for landlines. In such a design, **Table 4** suggests that, although the cost of telephone interviewing will rise further, cell phone interviews and landline interviews will no longer differ appreciably in cost. We estimate the average overall CPI for manually dialed landline interviews to be $44.71, while the average overall CPI for enhanced cell phone interviews is $44.74. We thus may be arriving at a point where—literally—there will not be a dime's worth of difference in average per-interview cost between landline and cell phone interviewing.

8. LOOKING AHEAD

While telephone interviewing trends are hard to predict, there are two trends that we can project with confidence into the near future. First, it seems clear that the trend toward allocating increasing shares of the total sample to the cell phone side (seen in **Figure 3**) will continue. This trend will continue to be driven by several considerations. The change in the cost ratio of cell phone calling versus landline calling, documented here, means that there is less of a cost savings to be realized from allocating calls to the landline sample. As more people abandon their landlines and adopt a cell phone-only or cell phone-mostly lifestyle, the cell phone frame becomes more closely representative of the general population, while the landline frame loses its ability to represent the population. And survey researchers are becoming aware that under-allocation to the cell phone side requires the application of larger design weights, reducing the effective sample size, so that spending more to call more cell phones may be cost effective in terms of effective sample size achieved. In fact, some researchers are already making the case for using samples that are drawn 100 percent from the cell phone frame (Peytchev and Neely 2013).

A second trend that seems certain is that the use of enhanced samples for cell phones will become standard practice. Since it is now possible to determine in advance the activity history or working-number status of sampled cell phone numbers, researchers who use samples that append this information, or scrub out nonworking numbers, gain a significant efficiency and cost advantage in the calling lab that clearly outweighs the extra per-number cost of purchasing such

samples (see Dutwin and Malarek 2014). Research into coverage and bias issues in using enhanced samples has generally been reassuring in that coverage bias from exclusion of phone numbers seems to be minimal for activity flags, and potentially manageable for appended billing ZIP code information, which can increase efficiency for locally targeted sampling designs (Dutwin and Malarek 2014; Dutwin 2014).

A third trend is highly probably but far from certain: Increasing amounts of information about sample cell phone numbers will become available to survey researchers. Already, sampling companies are offering sampling products from commercial sources that match cell phone numbers to other individual and household information available in various public and proprietary sources. The sources for these matched data tend to be shrouded in mystery. While these samples are certainly not full probability samples of all cellular households or individuals, they are already an attractive alternative for some research purposes because they allow targeting to specific populations and offer higher calling efficiencies than RDD cell phone samples. It is likely that some future telephone studies will draw from these commercial samples in much the same way that some researchers have been combining directory-listed and RDD samples for landline samples. While there may be costs in terms of coverage bias, if these samples continue to improve in coverage, or if (alternatively) more ways are found to append household or individual data to RDD cell phone samples (without violating privacy agreements with cell phone subscribers), the cost of cell phone interviewing may be further reduced in the future.

The outlook for predictive dialing in the United States is currently doubtful, yet quite difficult to predict. The TCPA outlaws automatic dialing of cell phone numbers, including any repeated dialing of numbers recently ported from landline to cell phone. The latter restriction renders predictive dialing of landlines highly risky. Yet there is some hope that the recent regulatory interpretations of the law from the FCC will be modified in the future, either through legislative action, reconsideration by the FCC, or by rulings from the bench in current or future TCPA lawsuits. This report has documented the extent to which predictive dialing can lower the cost of landline calling, and suggested that, if survey organizations switch to manual dialing of all calls, and use enhanced samples on the cell phone side, they will experience very similar calling efficiencies in cell phone and landline interviewing. However, if predictive dialing remains viable on the landline side, then cell phone interviewing will continue to be at a cost disadvantage compared to landline interviewing, especially in the larger calling labs that can benefit most from dialing predictively. The resulting cost differential may lessen the enthusiasm for switching to designs with very high percentages of cell phones in the sample. Smaller, mostly academic calling labs have not generally used predictive dialers, and so their costs will not be greatly affected by future changes in the TCPA regulatory climate.

Two cautions are in order with respect to the summary cost estimates provided in this report. First, the production rates for each frame depend in large part on respondent eligibility, so that some studies may experience very different cost ratios from the averages reported here. For example, a survey that targets elderly, low-income respondents will find few respondents to be eligible in the cell phone sample, while many would be eligible in the landline sample. Such a study could have a very high CPH ratio and cost ratio (cell phone over landline). In fact, one study reported in our third wave of data collection had an astonishing CPH ratio of over 13; it was a study by a small academic lab targeting elderly people lacking full Medicaid benefits, and was excluded from our analyses as an outlier. In contrast, a survey targeting young Hispanic males would have the opposite experience, finding high eligibility rates in the cell phone sample, few

eligible in the landline sample. A study with that target population would most probably experience a CPH ratio less than 1.00.

Second, survey costs must always be considered in balance with survey error. In particular, dual frame designs require weighting to take into account differing inclusion probabilities for different segments of the telephone universe, and most studies also use post-stratification weights to further adjust estimates. But the decrease in bias from weighting comes at a cost in increased variance. Larger design effects from weighting mean smaller effective sample sizes. While this report has estimated overall costs and cost ratios for dual frame surveys, these estimates have been based on the actual number of completes, without taking design effects into account. The true cost of landline and cellphone interviewing would more properly be calculated in relation to the effective sample size, yielding an estimate of costs per effective completion (see Benford et al 2009; Peytchev and Neely 2013).

Overall, there is some reason to believe (or hope) that the trend of increased cost for phone surveys is levelling off, albeit at a level of cost per interview that makes the telephone mode less competitive with other modes than it once was. When considering the telephone mode and various possibilities for single frame, dual frame, or multiframe sampling designs, study designers must consider the balance of cost vs. quality in the context of each study's features and requirements. Even though the costs of telephone interviewing have risen sharply over the last decade, the relative cost of cell phone interviewing (compared to landline interviewing costs) has lessened considerably, and telephone sampling and interviewing will likely persist as the preferable mode choice for many important studies in the future.

References

Benford, R., T. Tompson, C. Fleury, G. Feinberg, B. Feinberg, N. Speulda, and A. Weber (2009), **"**Cell Phone and Landline – Considerations for Sample Design, Estimates, Weighting, and Costs," Paper presented at the 64th annual conference of the American Association for Public Opinion Research; Hollywood, FL.

Brick, J.M., I. Flores Cervantes, S. Lee, and G. Norman (2011), "Nonsampling Errors in Dual Frame Telephone Surveys," *Survey Methodology,* 37, 1-12.

Dutwin, D. (2014), "Billing ZIP Codes in Cellular Telephone Sampling," *Survey Practice*, 7(4).

Dutwin, D., and D. Malarek (2014), "The Use of Recent Activity Flags to Improve Cellular Telephone Efficiency," *Survey Practice*, 7(1), 1-10.

Dutwin, D., and P.J. Lavrakas (2016), "Trends in Telephone Outcomes, 2008-2015," *Survey Practice*, 9(2).

Guterbock, T., P.J. Lavrakas, T. Tompson, and R. ZuWallack (2010), "The Variable Costs of Cell Phone Interviewing: Understanding Cost and Productivity Ratios in Dual-Frame Telephone Surveys," Paper presented at the 65th annual conference of the American Association for Public Opinion Research; Chicago, IL.

Guterbock, T. (2009), "Estimating Local Phone Service and Usage Percentages: How to Weight the Data from a Local, Dual Frame Sample Survey of Cell Phone and Landline Telephone Users in the United States," Paper presented at the Annual Meetings of the American Association for Public Opinion Research, Hollywood.

Guterbock T, R.A. Bebel, J.L.P. Holmes, and P.A. Furia (2012), "Why We No Longer Need Cell Phone Incentives," Presented at the Annual Meetings of the American Association for Public Opinion Research, Orlando.

Guterbock T, G. Benson, P.J. Lavrakas, D. Dutwin, and J. Kelly (2016), "How Cell Phone Interviewing Costs Are Changing—and Why," Presented at the Annual Meeting of the American Association for Public Opinion Research, Austin.

Guterbock, T., A. Peytchev, and D.L. Rexrode (2013), "Cell Phone Costs Revisited: Understanding Cost and Productivity Ratios in Dual-Frame Telephone Surveys," Presented at the Annual Meeting of the American Association for Public Opinion Research, Boston.

Martonik, R., T. Merry, N. Lee, S. Immerwahr, and M. Sanderson (2016), "Managing Efficiency in Telephone Surveys: Insights from Survey Paradata Trends 2010-2015," Paper presented at the Annual Meeting of the American Association for Public Opinion Research, Austin.

Oldendick, R.W., and D.N. Lambries (2011), "Incentives for Cell Phone Only Users: What Difference Do They Make?" *Survey Practice*, 4(1). Available at: http://www.surveypractice.org/index.php/SurveyPractice/article/view/94

Peytchev, A., and B. Neely (2013), "RDD Telephone Surveys: Toward a Single Frame Cell Phone Design," *Public Opinion Quarterly*, 77 (Spring), 283-304.

APPENDIX F – LEGAL & OPERATIONAL CONSIDERATIONS

Chuck Shuttles, Symphony Advanced Media
Jenny Kelly, NORC at the University of Chicago
Howard Fienberg, Marketing Research Association

ABSTRACT: Researcher, companies, and clients considering telephone surveying should be aware of the legal and ethical considerations that can impact how the research is performed (i.e., operationalized). This appendix lists the broad range of issues stemming from legal protections afforded to the general public; e.g., those addressed by the Telephone Consumer Protection Act (TCPA) and its evolving interpretations by enforcement agencies. We draw expert advice from sources such AAPOR, CASRO, MRA, and FCC Declaratory Rulings, and list concerns and questions to explore when designing/evaluating surveying operations. It is important that researchers understand the legal and operational considerations before performing telephone research and always consider the need to consult with legal counsel.

In this appendix we elected to combine legal and operational considerations into one section due to the fact that any update from the 2010 guidance given in these areas has been profoundly affected by changes to the requirements and enforcement of the Telephone Consumer Protection Act (TCPA). By reviewing changes enacted in July 2015 by the Federal Communications Commission (FCC) and subsequent operational changes, researchers will have a better understanding of the current list of considerations for telephone surveying. (Note: there is no substitute to consulting with your own legal counsel to review the current state of law and your risk of legal actions.)

In broad stokes, the 2010 AAPOR Cell Phone Task Force Report covered the following legal and ethical topics:

- U.S. Legal Restrictions on Calling Cell Phones – the TCPA [Updated guidance below]
    - The Need for Manual Dialing [Updated guidance below]
    - The Permitted Use of Neustar – to scrub residential lines that have ported to cell
- Legal Considerations Regarding Text Messaging and Spam – information from the CAN-SPAM Act
- Legal and Ethical Considerations Regarding Possible Harassment Due to the Number of Callbacks Used[20]
- Ethical Considerations for Time-of-Day Calling Restrictions
- Ethical Considerations for Taking Safety and Respondent Privacy into Account
- Ethical Considerations for Remunerating Cell Phone Respondents
- The Ethics of Transmitting Accurate Caller Identification Information
- The Ethics of Maintaining an Internal *Do Not Call* List

While the above issues still apply, the 2015 changes to the TCPA have impacted research / data collection operations. AAPOR, the Council of American Survey Research Organizations

[20] See AAPOR's 2014 Survey Refusals Task Force Report on additional guidance on callbacks. http://www.aapor.org/AAPOR_Main/media/MainSiteFiles/RefusalTF_FINAL090814.pdf.

(CASRO), and the Market Research Association (MRA) offer a variety of resources to researchers to better understand TCPA, including the following:

1. AAPOR's TCPA Task Force White Paper[21]
2. Addendum to the 2010 AAPOR Cell Phone Task Force Report[22]
3. A joint AAPOR/CASRO/MRA TCPA webinar[23] with advice from:
    a. Duane Berlin, General Counsel, CASRO/AAPOR
    b. Abby Devine, Director of Government and Public Affairs, CASRO
    c. Peter Milla, Consultant, CASRO
    d. David Almy, CEO, MRA
4. Additional materials and summaries for specific areas of concerns and impact to telephone survey operations[24]

While major aspects of the TCPA have been reinterpreted by federal courts and the FCC, the risk to researchers, call centers, and even the client sponsoring the research has not abated. *Each contact attempt, regardless of completion* could lead to $500 and up to $1,500 for each violation. Even small sample sizes could lead to multi-million dollar judgements or settlements.[25] This has given rise to a cottage industry of law firms seeking cases to pursue. Accordingly, researchers should be aware of the following operational considerations (as stated in the resources above):

- There is a tremendous amount of risk for all parties (the researcher, the data collection vendor, and the client/research sponsor).
- The definition of an "autodialer" has been broadened. Simply stating that calling attempts to mobile phone numbers are manual/hand dialed 10 digits does not insure compliance. Any connection to equipment or a system that has "capacity" of an autodialer (including systems that can record, track, or route calls or store/produce and dial random/sequential numbers…even if that capacity is not turned on) leaves you in jeopardy.
- The consumer may freely revoke consent at any time in any reasonable manner. The researcher / caller may not be limited in how consent is revoked.
- For numbers that are reassigned to a mobile number, the caller has a 1-call safe harbor for autodialed calls. As long as the caller believes that they have consent to make a call and did not know about the reassignment before making the call, you have one (but only one) attempt (even if it does not connect) without liability. You are liable for all subsequent call attempts.[26]

[21] See AAPOR TCPA Task Force White Paper. https://www.aapor.org/getattachment/Education-Resources/TCPA/TCPA_FINAL.pdf.
[22] See Addendum to 2010 AAPOR Cell Phone Task Force Report. https://www.aapor.org/AAPOR_Main/media/MainSiteFiles/CellPhoneTaskForceAddendum.pdf.
[23] See joint AAPOR/CASRO/MRA webinar presentation slides. https://www.aapor.org/getattachment/Education-Resources/TCPA/TCPA-Webinar_Presentation.pdf.
[24] See Additional Materials. https://www.aapor.org/Education-Resources/TCPA.aspx.
[25] For example, see "Gallup Settles Costly TCPA Lawsuit for $12 million." http://www.marketingresearch.org/article/gallup-settles-costly-tcpa-lawsuit-12-million
[26] As FCC Commissioner Michael O'Reilly observed in his statement on the 2015 TCPA Order, "A person could take a call, never let on that it's the wrong person, and receive subsequent calls solely to trip the liability trap." And as Commissioner Ajit Pai explained in his statement, this has already happened. Pai cited the example of Rubio's restaurant, which thought it was

- Telephone carriers and VoIP providers are encouraged to offer call-blocking technology, which may prevent your calls from reaching the consumer in the first place.[27]
- If you work with a vendor, be wary of their assurances. For example:
    - Do not rely on their claim to be "completely manual" or that their calling requires "human intervention". As stated above, there is still risk if that manual calling is performed on a system that has the capacity to autodial.
    - Do not rely on claims that the vendor will indemnify you. While most vendors would not agree to indemnification in the first place, those that do would need to have the financial ability to withstand a judgement or settlement (given the $500-$1,500 risk per call attempt). The vendor's indemnification may not prevent your liability regardless.
    - Clearly document responsibility and liability for full adherence to TCPA compliance with any vendor (e.g., vendor documents that the dedicated calling stations/booths have direct connection to telephonic dialing via a push button handset with no connection to autodialer-capable systems).
- If the number is from a supplied list (e.g., a telephone survey of members belonging to a professional association), (1) ensure that you know how the number was collected (i.e., did the association's terms state that the number could be used by 3rd party to perform research?) and (2) require the client to indemnify you against TCPA claims (as noted above, that indemnity may not fully cover the large potential judgement / settlement costs).
- Be prepared for legal actions!
    - Perform extensive due diligence with your data collection process and vendor(s). Regularly reevaluate your vendor due diligence and perform risk assessments.
    - Build additional fees into your cost structure to fund additional labor and precautions due to TCPA. Consider informing your client as to why you are doing this and that you are attempting to protect yourself and your client.
    - Consider setting aside funding for legal defense of any claims.
    - Have a well-documented and vociferous response prepared for any initial queries from attorneys/legal firms. If you can show that you are fully TCPA compliant and are prepared to defend yourself, you may be able to dissuade further action from those fishing for a potential claim.
- While most of the ruling has resulted in additional risk and further precautions for research operations, there is one beneficial outcome to the 2015 ruling. If a consumer supplies you with a mobile number, without placing restrictions on calling them on that number and the call is not telemarketing based, then the consumer is deemed to have given consent.

texting employees about food safety issues and warnings. Unknown and unreported to Rubio's, one employee had lost his phone and his number was reassigned to a new subscriber. "The new subscriber never asked Rubio's to stop texting him – at least not until he sued Rubio's in court for nearly half a million dollars."

[27] "What the Rise of Call-Blocking Technology Could Mean for Telephone Research" http://www.marketingresearch.org/article/what-rise-call-blocking-technology-could-mean-telephone-research

- Additional good news is that "the TCPA does not apply to calls made by or on behalf of the federal government in the conduct of official government business, except when a call made by a contractor does not comply with the government's instructions," as stated in a FCC declaratory ruling released on July 5, 2016[28].

Operationally, the TCPA has placed a lot of burden on research and call centers to take additional steps to ensure compliance. However, given the penetration of mobile phones in the U.S., it is likely critical to include mobile phones in your research. Thus, take time to understand the legal and operational considerations before performing telephone research and always consult with your own legal counsel.

[28] See FCC Declaratory Ruling issued July 5, 2016. http://transition.fcc.gov/Daily_Releases/Daily_Business/2016/db0706/FCC-16-72A1.pdf