Evaluating the Impact of Using Pre-Recorded Voicemail Messages in the National Immunization Surveys

Benjamin Skalland[1], Becky Reimer[1], Qiao Ma[1], Vince Welch[1], Sarah Kornylo[1], Kate Hobson[1], Holly A. Hill[2], Benjamin Fredua[2,3]
[1]NORC at the University of Chicago
[2]National Center for Immunization and Respiratory Diseases, Centers for Disease Control and Prevention
[3]Leidos Health, Inc.

Abstract

The National Immunization Surveys (NIS) are large random-digit-dial (RDD) telephone surveys to assess vaccination coverage in the United States among children age 19-35 months (NIS-Child) and adolescents age 13-17 years (NIS-Teen), and to assess influenza vaccination coverage among children 6 months-17 years (NIS-Flu). Among working telephone numbers in the RDD samples, the most common NIS call outcome is reaching a potential respondent's answering machine or voicemail. When messages are left for the respondent, under the standard NIS procedure, the message is left live by the interviewer making the call. An alternative is the use of a pre-recorded message. Relative to a live message, a pre-recorded message should result in a reduction in interviewer time, potentially improving the cost-effectiveness of survey administration, as the interviewers will no longer be spending time leaving messages. A pre-recorded message could also affect response rates, either positively or negatively. The use of a pre-recorded message allows for more control over the clarity, pacing, and vocal characteristics of the message left and therefore could result in higher response rates on subsequent calls. However, if the respondent can tell the message is a recording, response rates could suffer.

Beginning in Quarter 4 of 2017, an evaluation was conducted in the NIS-Child, NIS-Teen, and NIS-Flu to examine the impact of pre-recorded messages on interviewer time and on response rates. Households were randomly assigned to receive either a live message (control group) or one of four versions of a pre-recorded message (experimental group) when a respondent's answering machine or voicemail was encountered. Pre-recorded messages differed by the age and gender of the person recording the message. We examined the interviewer time spent on calls on which messages were left and compared post-message contact rates among the control group and treatment groups. We found that the use of pre-recorded messages significantly reduced interviewer time spent on these calls compared with live messages, and we found similar post-message contact rates between pre-recorded and live messages in both the landline and cell-phone samples. Post-message contact rates did not differ by the age of the person recording the message. Post-message contact rates were significantly higher for female recorded voices than male recorded voices in the landline sample but not in the cell-phone sample. We speculate that this may be due to the availability of message transcription on many cell-phone devices. In this paper, we present the study design, study results, conclusions, limitations, and recommendations for future research.

Key Words: National Immunization Surveys, Answering Machine Message, Voicemail Message, Pre-Recorded Message

1. Introduction

In surveys with telephone data collection, answering machine and voicemail messages are a way to inform potential respondents about the survey (Holbrook et al. 2007). The literature on the effectiveness of such messages is mixed, with some studies finding no effect of messages on survey response (e.g., Tuckel and Schulman, 2000; Link and Mokdad 2005) and others finding a positive effect (e.g., Harlow et al., 1993; Xu et al., 1993; Koepsell et al., 1996; Roth et al., 2001). Even if messages have a positive impact on response, messages left live by interviewers add time to telephone calls, and this additional interviewer time can add to the cost of survey administration, particularly for large-scale telephone surveys that involve millions of dials.

One way to reduce interviewer time when answering machine and voicemail messages are left is the use of pre-recorded messages. Under this approach, the interviewer listens to the potential respondent's outgoing message, waits for the beep, and then initiates a pre-recorded message instead of reading the message script live. After initiating the pre-recorded message, the interviewer is freed up to place the next call while the message is being left. This approach should result in reduced interviewer time for calls on which messages are left, as the interviewer is no longer spending the time reading the message; however, this approach could also impact response rates. A pre-recorded message allows control over the clarity, pacing, and vocal characteristics of the message, and so could improve response relative to a message left live by interviewers; alternatively, if the potential respondent can tell that the message has been pre-recorded, response could be negatively affected.

To our knowledge, there is no literature on the use of pre-recorded answering machine or voicemail messages in surveys. In this paper, we present results of a study to gauge the impact of pre-recorded messages in the National Immunization Surveys (NIS), which are large, national RDD telephone surveys. Our research questions are:

1. Do pre-recorded messages reduce interviewer time on calls on which messages are left?
2. Can a pre-recorded message be as effective as a live interviewer message in inducing contact with the respondent after the message is left?
3. Do the voice characteristics (in particular, the age and gender) of the person making a pre-recorded message affect the rate of contact after the message is left?

Further details of our evaluation are provided below. In Section 2, we provide a description of the NIS and the design of our pre-recorded message experiment; in Section 3, we present the results of the experiment; in Section 4, we discuss the results and their implications; in Section 5, we state the limitations of our study; and in Section 6, we present recommendations for future research.

2. Description of the National Immunization Surveys and the Design of the Experiment

The NIS family of surveys are annual RDD telephone surveys to provide U.S. national, state, and selected local area estimates of vaccination coverage among children age 19-35 months (NIS-Child) and adolescents age 13-17 years (NIS-Teen), and influenza vaccination coverage estimates among children age 6 months-17 years (NIS-Flu). They are

sponsored by the Centers for Disease Control and Prevention (CDC), and, since 2005, have been conducted by NORC at the University of Chicago. NIS data are collected via an RDD telephone survey of parents and guardians of children in the target age ranges. From 2011 through 2017, the NIS was a dual-frame RDD survey, sampling both landline and cell-phone numbers; beginning in 2018, the NIS is a single-frame RDD survey, sampling only cell-phone numbers. For NIS-Child and NIS-Teen, the RDD survey is followed by a mail survey sent to the vaccination providers of the children and adolescents identified in the RDD phase to obtain their vaccination histories with consent from a parent or guardian.

Because it targets a rare population (children in specific age ranges) and has sample size requirements at the state- and local-area levels, the NIS places a very large number of dials each year – over 60 million in 2017. While a large portion of those dials are to numbers that turn out to be not in service, among working numbers the most common call outcome is reaching a potential respondent's answering machine or voicemail; over 70 percent of NIS calls to working numbers reached an answering machine or voicemail in 2017. Leaving live messages on all of those calls would be quite expensive, and therefore the current NIS protocol is to leave a message on the fourth answering machine or voicemail event, and on every third answering machine or voicemail event thereafter.[1]

Because leaving live messages before the fourth event is cost-prohibitive, beginning in Quarter 4 of 2017 and continuing into 2018, the NIS conducted an experiment on the use of pre-recorded messages, as shown in Figure 1. Cases were randomly divided to receive either a live interviewer message (control) or a pre-recorded message on the fourth event (treatment). Within the pre-recorded message treatment, cases were further randomized to receive one of four versions of the pre-recorded message in a 2 x 2 design. The four versions of the pre-recorded message differed by the characteristics of the voice used on the message: a mature female voice, a mature male voice, a young female voice, and a young male voice. The text of the message was the same for the control and all four treatment groups:

> "Hello. I am calling on behalf of the Centers for Disease Control and Prevention. We are conducting a survey about childhood immunization. Would you please call us at 1-877-220-4805 to let us know whether or not there are any children between 12 months and 4 years old living or staying in this household? The number again is 1-877-220-4805. Thank you."

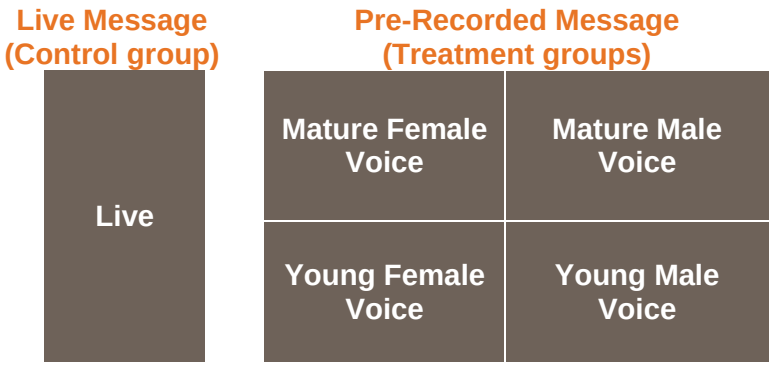
Two key outcome measures were examined:

- Average interviewer time for calls on which a message was left.
- Among cases receiving a message, the rate of contact with a respondent after the message was left.

The experiment was conducted in both the landline and cell-phone samples in Quarter 4 of 2017 and in the cell-phone sample in 2018, as there was no landline sample in the 2018 NIS. The experiment began November 17, 2017; results presented in the paper include data collected through March 31, 2018.

[1] There are a few rare exceptions to this protocol. For example, if the respondent had scheduled an appointment but an answering machine or voicemail was reached when the respondent was called for that appointment, a message was always left on that call.

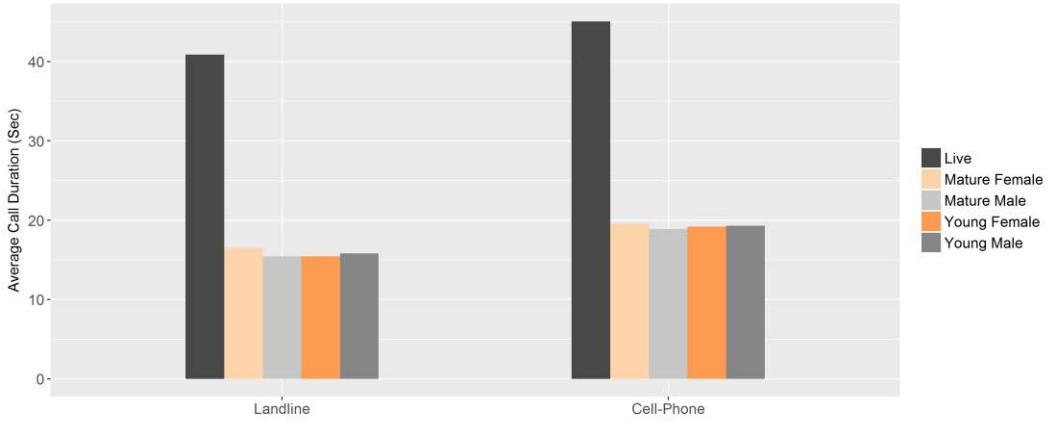
Figure 1: Design of the Pre-Recorded Message Experiment



3. Results

We first examine the interviewer time spent on calls on which messages were left. Figure 2 presents the average interviewer time (in seconds) for message-left calls for the control group and four treatment groups, separately for the landline and cell-phone samples. As expected, the use of pre-recorded messages results in less time spent by interviewers on these calls. Live message calls lasted about 40 seconds on average in the landline sample and 45 seconds on average in the cell-phone sample; pre-recorded message calls lasted about 15 seconds and 20 seconds in the landline and cell-phone samples, respectively, or about 25 seconds less than live-message calls in each sample type.

Figure 2: Average Interviewer Time for Calls on which a Message Was Left



Next, we examine the rate of contact after the message was left for live interviewer messages versus pre-recorded messages. Figure 3 presents the post-message contact rates and 95% confidence intervals for the control group and for all four pre-recorded message groups combined, overall and separately for the landline and cell-phone samples. The point estimates of the post-message contact rate are slightly lower for the pre-recorded message group compared to the live-message group, but these differences are not statistically significant (p=0.271 for the overall comparison; p=0.448 for the comparison within the landline sample; p=0.437 for the comparison within the cell-phone sample).

Finally, we examine the post-message contact rates for pre-recorded messages by treatment. Figure 4 compares the post-message contact rates between the two female voice treatment groups and the two male voice treatment groups, overall and by sample type. Overall, we observe no difference in the post-message contact rate between the female and male voices, but the rate was lower for the male voices in the landline sample (p=0.018) and higher for the male voices in the cell-phone sample (p=0.065).

Figure 3: Post-Message Contact Rates and 95% Confidence Intervals: Live vs. Pre-Recorded Message

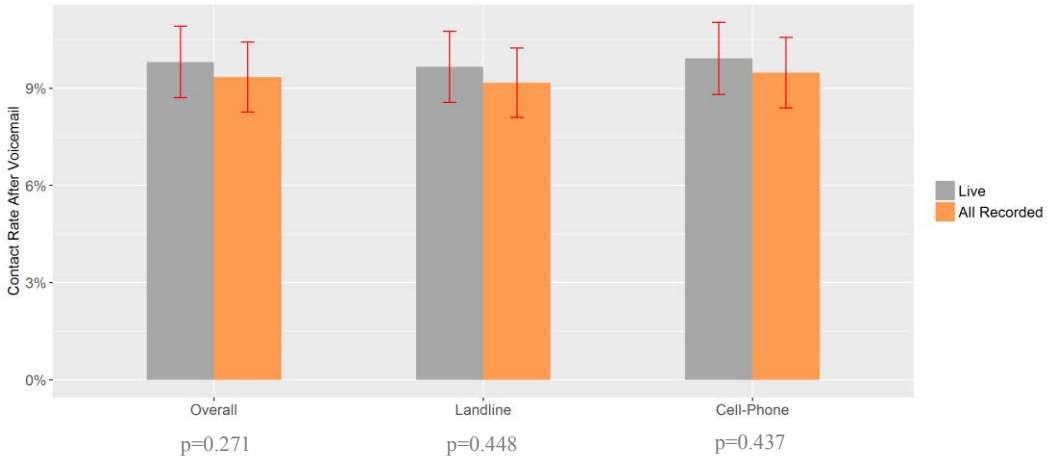


Figure 4: Post-Message Contact Rates and 95% Confidence Intervals: Female vs. Male Pre-Recorded Voices

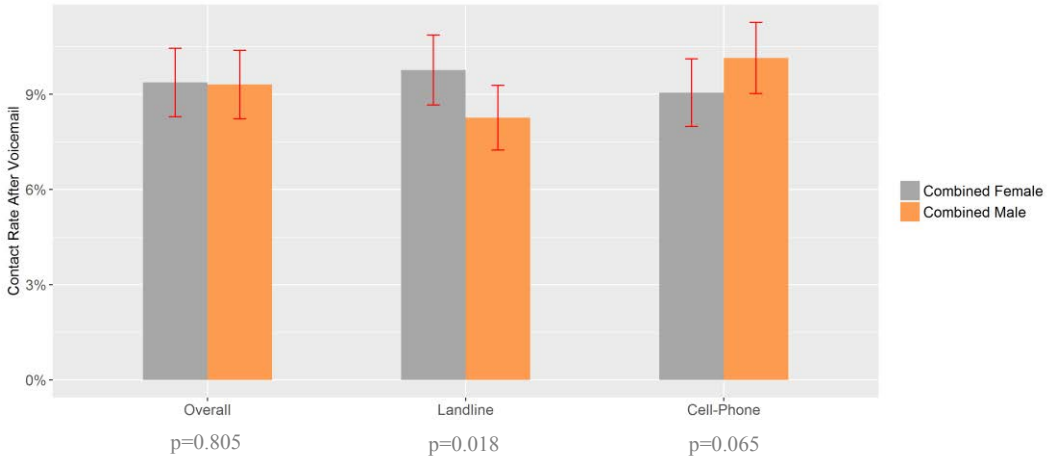


Figure 5 compares the post-message contact rates between the two mature voice treatment groups and the two young voice treatment groups. Here we observe no significant differences.

Figure 6 presents the post-message contact rates separately for each of the four treatment groups. We do not observe any interaction effects between the gender of the voice (female, male) and the age of the voice (mature, young) overall or within sample type. Within gender, the post-message contact rates are similar for the mature and young voices, and within age, the difference in post-message contact rates between genders is similar.

Figure 5: Post-Message Contact Rates and 95% Confidence Intervals: Mature vs. Young Pre-Recorded Voices

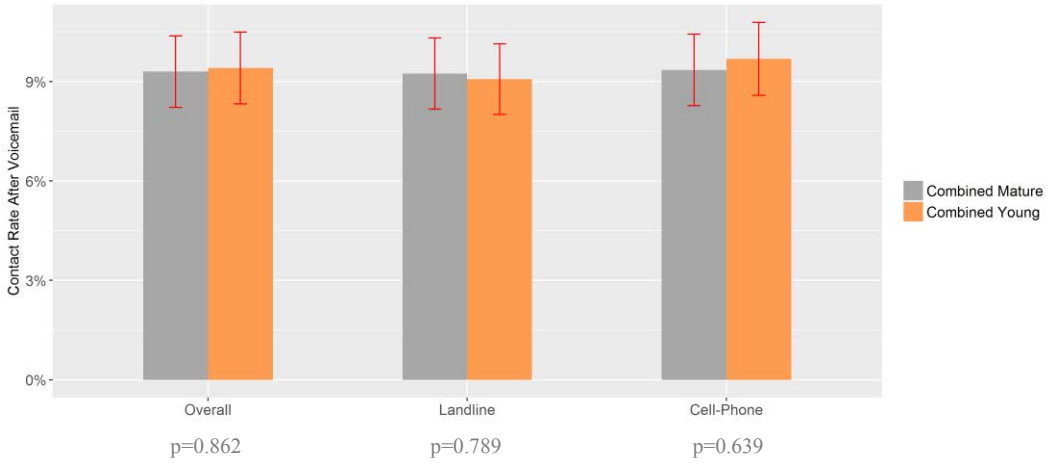
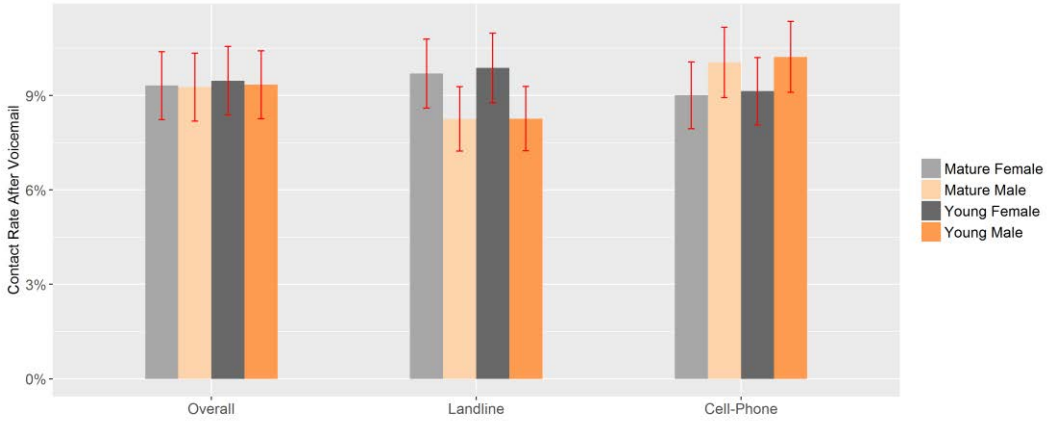


Figure 6: Post-Message Contact Rates and 95% Confidence Intervals: Mature Female vs. Mature Male vs. Young Female vs. Young Male Pre-Recorded Voice



4. Conclusions and Discussion

The use of pre-recorded answering machine and voicemail messages can save interviewer time. We found that pre-recorded messages saved about 25 seconds, on average, for calls on which messages were left.[2] While 25 seconds may not seem like a lot of time saved on a call, when multiplied by the millions of calls on which a message might be left in a large-scale telephone survey, this can translate into a savings of thousands of hours of interviewer time.

We did not observe statistically-significant differences in post-message contact rates between the live interviewer message (control) and the pre-recorded message (treatments), suggesting that the use of pre-recorded messages can save interviewer time while having little or no impact on response rates.

We also did not observe statistically-significant differences in post-message contact rates between the mature-voice versions of the pre-recorded message and the young-voice versions of the pre-recorded message. However, we did observe differences by the gender of the recorded voice in the landline sample: the male voice recordings were associated with lower contact rates than the female voice recordings in the landline sample ($p=0.018$) but not in the cell-phone sample.

It is not clear why the female pre-recorded voices performed better than the male pre-recorded voices in the landline sample but not in the cell-phone sample. One possibility is that this is due to differences the characteristics of respondents with landline phones and respondents with cell-phones. Adults with landline phones are more likely to be older, non-Hispanic, female, college-educated, non-poor, home-owners, and living in the Northeast Census region (Blumberg and Luke, 2018). Such adults could respond differently to male versus female pre-recorded messages compared to adults with cell-phones.

Another possibility is that the variation arises due to differences between landline and cell-phone devices. Many smartphones transcribe voicemail messages. If a respondent reads the transcription rather than listening to the message, then the voice on the message would not matter and a difference in post-message contact rate by voice characteristic would not be expected; instead, the quality of the transcription may be important.

In our study, transcriptions did vary across the four pre-recorded messages, even for the same device. Figure 7 shows transcriptions by treatment made by an AT&T iPhone 7. None is perfectly transcribed, as each contains multiple blanks, representing words that the algorithm was unable to decipher, as well as incorrect words; moreover, the transcriptions differ for the four recordings. This suggests that when using pre-recorded messages, researchers should test the messages on various smartphones and re-record if necessary until a clean transcription is obtained. However, we also found that the transcription can vary even for the *same* recording on the *same* device. Figure 8 shows transcriptions of the "Young Male" version of the recording, all made on an AT&T iPhone 7 just minutes apart. While the last transcription is nearly flawless, the first two contain multiple blanks. It is

[2] In our study, interviewers waited for the beep before initiating the pre-recorded messages. Savings could be greater if the data collection system itself could detect when to initiate the pre-recorded message; under this scenario, instead of waiting for the beep, interviewers would simply code the call as an answering machine or voicemail and move onto the next call while the data collection system detects when to initiate the message.

not clear why the transcription would vary for the same recording on the same device; this is an area in need of further investigation if pre-recorded messages are to be widely used.

5. Limitations

Although we found that pre-recorded answering machine and voicemail messages can save interviewer time with little or no impact on post-message contact rates, there are several limitations to our study. First, NIS interviewers do not leave messages before the fourth voicemail event, and respondents with four voicemail events are, by definition, hard-to-reach; our conclusions may not be applicable if messages are left earlier. Second, the NIS uses a particular message script; our conclusions may not be applicable to other message scripts. Third, we tested four particular pre-recorded message voices; our conclusions may not be applicable to other voices or vocal characteristics (e.g., accent, pacing, pitch, etc.). Finally, nearly all of the messages left in our study (both live and pre-recorded) were in English; our conclusions may not apply to other languages.

Figure 7: Pre-Recorded Message Transcriptions on an AT&T iPhone 7, by Treatment Group

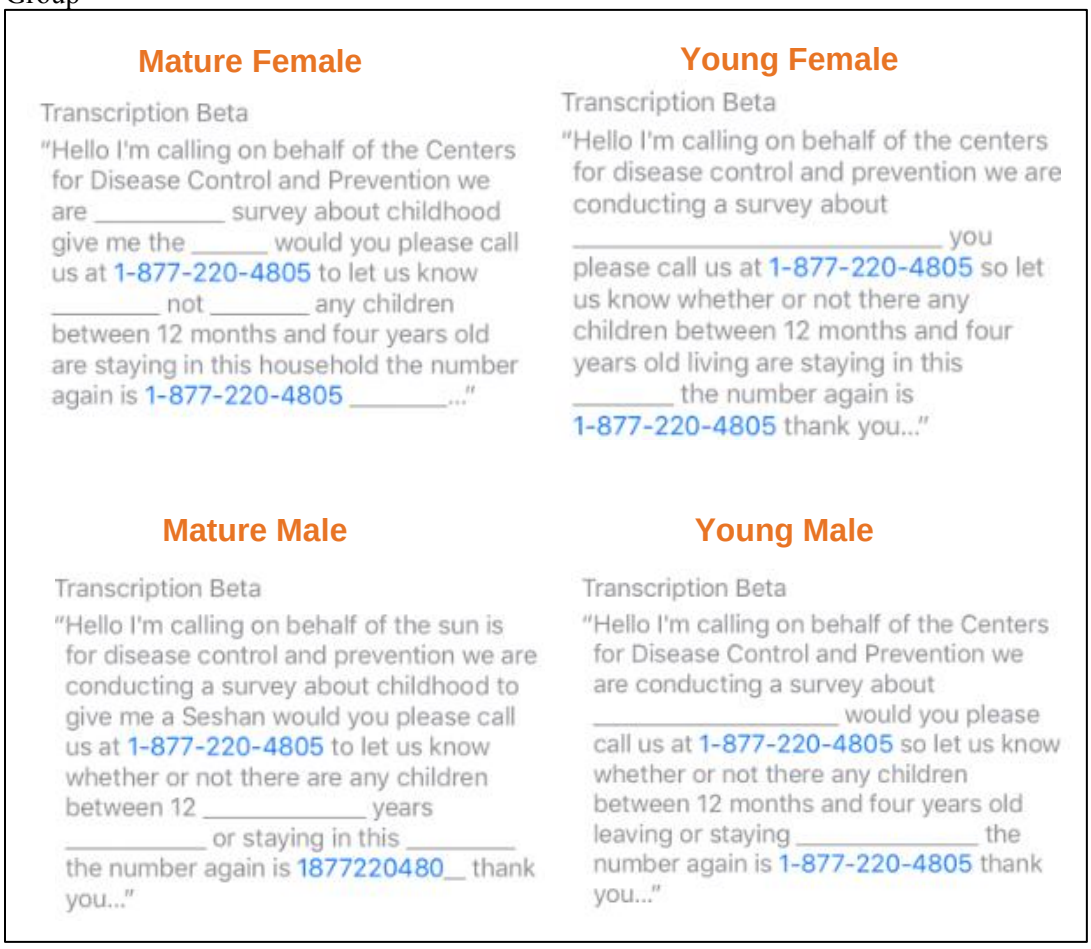
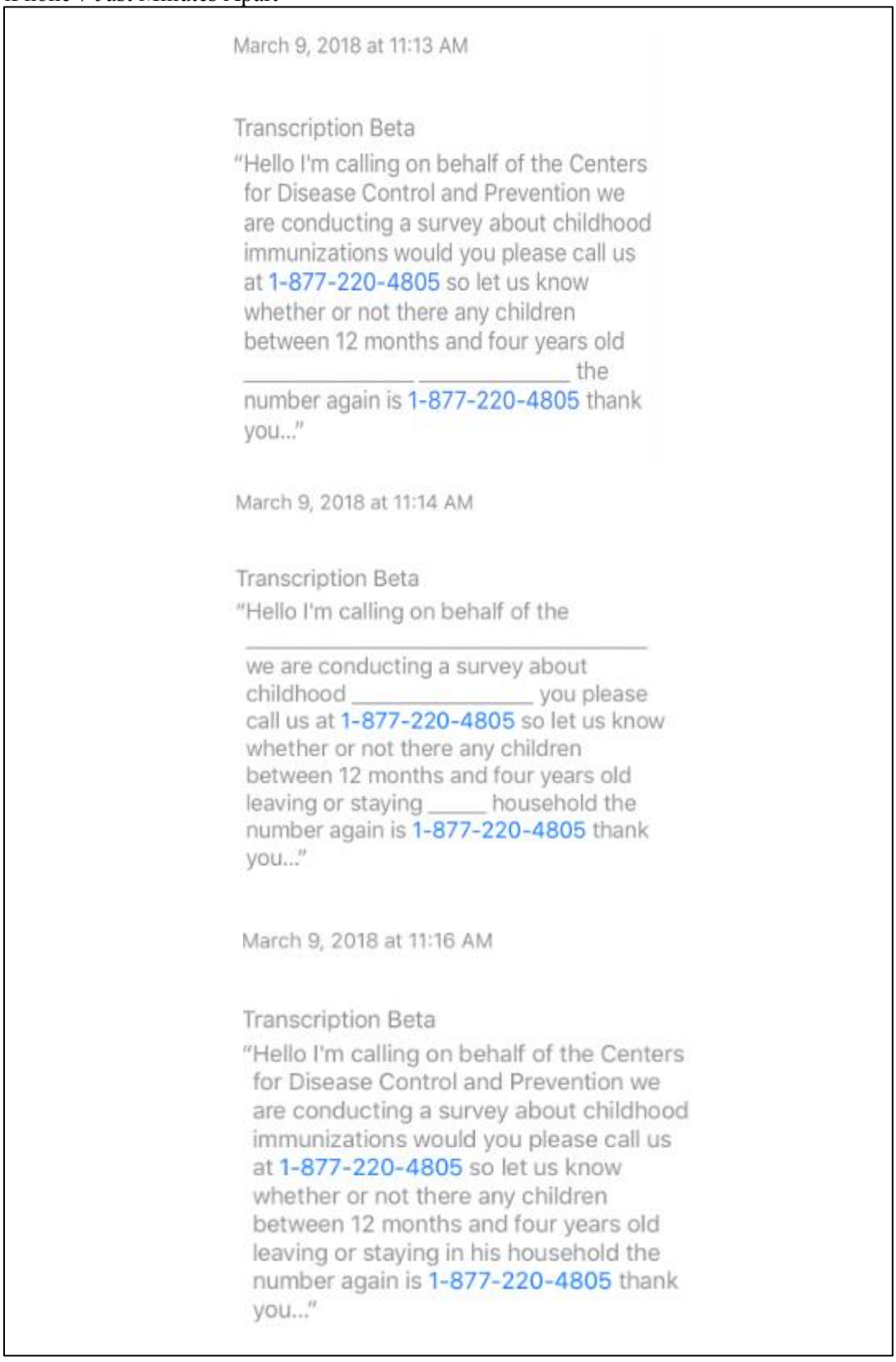


Figure 8: Transcriptions of the "Young Male" Pre-Recorded Message Made by an AT&T iPhone 7 Just Minutes Apart



6. Future Research

Based on the results of our study, beginning in Quarter 2 of 2018, the NIS stopped leaving live interviewer messages on the fourth answering machine or voicemail event and instead leaves pre-recorded messages on these events. Because the NIS no longer fields a landline sample (beginning in 2018), and because none of the four pre-recorded message voices performed conclusively better than the other three on cell phones, we have continued to use all four pre-recorded voices.

As mentioned previously, in the current study, interviewers waited for the beep before initiating the pre-recorded message. In the future, the NIS will explore the possibility of automated message initiation, i.e., having the data collection system itself detect when to begin leaving the pre-recorded message rather than having the interviewer remain on the phone until the beep. This should save even more interviewer time on message-left calls.

With the implementation of automated message initiation, there should be little or no difference in NIS interviewer time for calls on which a message is left and calls on which a voicemail is reached but a message is not left. If true, this would allow the NIS to leave messages earlier in a case's call history instead of waiting until the fourth voicemail event. We plan to experiment with when messages are first left (i.e., first event, second event, third event, etc.) to determine the optimal protocol.

Beyond the savings in interviewer time, a benefit of pre-recorded messages is that different versions can be implemented and tested very easily. For example, the current study experimented with four voices that differed by age and gender, but other voice characteristics could be explored, such as pitch, pacing, accent, etc. Furthermore, in the current study, the same message script was used for all treatments, but new recordings using different message scripts could be easily implemented and tested without having to alert or re-train the interviewers.

Finally, additional research is needed on voicemail transcription. As transcription becomes more widely available, accuracy of the transcription may become more and more important, and we need to gain a better understanding about how to ensure accurate transcription across calls, devices, and carriers.

References

Blumberg SJ, Luke JV. Wireless substitution: Early release of estimates from the National Health Interview Survey, July-December 2017. National Center for Health Statistics. June 2018. Available from: http://www.cdc.gov/nchs/nhis.htm.

Harlow BL, Crea EC, East MA, Oleson B, Fraer CJ, and Cramer DW (1993). Telephone answering machines: the influence of leaving messages on telephone interviewing response rates. Epidemiology. 4(4):380-383, July 1993.

Holbrook A, Krosnick J, and Pfent A (2007). The causes and consequences of response rates in surveys by the news media and government contractor survey research firms. In Advances in telephone survey methodology, ed. James M. Lepkowski, N. Clyde Tucker, J. Michael Brick, Edith D. De Leeuw, Lilli Japec, Paul J. Lavrakas, Michael W. Link, and Roberta L. Sangster. New York: Wiley.

Koepsell TD, McGuire V, Longstreth WT, Nelson LM, and van Belle G (1996). Randomized trial of leaving messages on telephone answering machines for control recruitment in an epidemiologic study. Am J Epidemiol 1996; 144:704–6

Link MW and Mokdad A (2005). Leaving answering machine messages: do they increase response rates for the behavioral risk factor surveillance system? International Journal of Public Opinion Research, 17, 239–250.

Roth SB, Montaquila J, and Brick JM (2001). Effects of telephone technologies and call screening devices on sampling, weighting and cooperation in a random digit dialing (RDD) survey. Paper presented at the Annual Conference of the American Association for Public Opinion Research, Montreal, Quebec, Canada

Tuckel P and Schulman M (2000). The impact of leaving different answering machine messages on nonresponse rates in a nationwide RDD survey. Proceeding of the Section on Survey Research Methods, American Statistical Association, 901–906.

Xu M, Bates BJ, and Schweitzer JC. The impact of messages on survey participation in answering machine households (1993). Public Opinion Quarterly 57(2): 232-237.