

Survey of Massachusetts Congressional District #4

Methodology Report

Prepared by Robyn Rapoport and David Dutwin
Social Science Research Solutions
53 West Baltimore Pike
Media, PA, 19063



Contents

Overview 3

Sample Design..... 3

Field Preparations 4

Data Collection Procedures 4

Weighting Procedures..... 6

 1. Phone-Status Correction (W_{PS}): 7

 2. Within household selection correction (W_{HC}): 7

 3. Post stratification weighting: 7

Response Rate..... 9

Overview

The University of Massachusetts Lowell contracted with Social Science Research Solutions/SSRS to conduct the Massachusetts Congressional District #4 (MA CD-4) Study from February 2 through February 4 and February 6 through February 8, 2012. The purpose of the MA CD-4 Study was to conduct the first valid and reliable poll on the possibility that Joseph Kennedy III will run for Congress in the newly redrawn MA Congressional District #4 for the seat being vacated by Barney Frank. This report provides information about the methods used to collect the data and report the survey results.

The study collected data from a representative sample of 408 registered voters living in the newly redistricted area of MA CD-4. The study consisted of a landline component (n = 304) and a cell phone component (n = 104).

Sample Design

To address concerns about coverage, the study employed a dual-frame landline/cell phone random digit dial (RDD) telephone design. Both samples were generated by SSRS's sister company, Marketing Systems Group (MSG).

RDD landline sample was drawn from telephone exchanges within the new MA CD-4. Using Marketing Systems Group's Genesys database of telephone exchanges, we were able to select telephone exchanges that would result in a 92 percent incidence of reaching households in MA CD-4. These telephone exchanges cover 95 percent of all households in the District. Following generation, landline sample was prepared using MSG's proprietary procedures that not only limit sample to non-zero banks, but also identify and eliminate approximately 90% of all non-working and business numbers and ported cell phones.

For the RDD cell phone sample, numbers were initially drawn from the four switch-points (central routing mechanisms that send cell phone calls to different parts of the country) located in MA CD-4. After the initial sample was drawn, additional analyses were conducted through the Telcordia database in order to better align the cell phone sample with the borders of the new MA CD-4 and improve coverage. Based on these analyses, the cell sample was refined as follows: First, the analysis identified a number of 1,000 blocks of telephone numbers connected with switch-points outside of the District that are most often routed to households within the District. These 1,000 blocks were therefore included in the sample file. Second, the analysis tagged several 1,000 blocks in the four switch points within the District as being owned by telephone resellers that typically do not provide numbers used by households. SSRS dialed a portion of these exchanges and confirmed that indeed, these exchanges are non-residential; therefore, these exchanges were excluded from further dialing. Third, Telcordia flagged 1,000 blocks within the four in-District switch-points that target households outside of MA CD-4.

SSRS also dialed several of these exchanges. After confirmation that none of the households were part of MA CD-4, telephone numbers associated with these 1,000 blocks were removed from the active sample.

Survey incidence before the sample refinements outlined above was less than four percent; following the refinements, the sample attained a 15 percent incidence of registered voters living with MA CD-4, closer to the original estimate of a 20 percent incidence.

Field Preparations

The questionnaire was developed by UMass Lowell in consultation with the SSRS project team. Prior to the field period, SSRS programmed the study into CfMC Computer Assisted Telephone Interviewing (CATI) software. Extensive checking of the program was conducted to assure that skip patterns followed the design of the questionnaire.

The field period for this study was February 2 through February 4 and February 6 through February 8, 2012. All interviews were done through the CATI system. The CATI system ensured that questions followed logical skip patterns and that complete dispositions of all call attempts were recorded.

CATI interviewers received both written materials on the survey and formal training. The written materials were provided prior to the beginning of the field period and included an annotated questionnaire that contained information about the goals of the study as well as detailed explanations of why questions were being asked, the meaning and pronunciation of key terms, potential obstacles to be overcome in getting good answers to questions, and respondent problems that could be anticipated ahead of time as well as strategies for addressing the potential problems.

Interviewer training was conducted immediately before the survey was officially launched. Call center supervisors and interviewers were walked through each question from the questionnaire. Interviewers were given instructions to help them maximize response rates and ensure accurate data collection.

Data Collection Procedures

Interviews were conducted from February 2 through February 4 and February 6 through February 8, 2012; interviews were not conducted on Sunday, February 7, "Super Bowl Sunday," because of the likelihood that cooperation and response rates would be low on that day.

For the landline sample, interviewers asked to speak with the youngest adult male or female currently at home. In order to produce a sample that would more closely resemble the general population in the area by gender and age when combined with the cell completes, the program

asked for youngest males first preferentially, 70% of the time. Callbacks were set up if no adult was available to complete the interview at the time of the call.

For the cell phone sample, interviewers first determined whether the person who answered the phone was an adult and then confirmed that the respondent was not driving or doing anything that required their full attention. If possible, callbacks were set up if the respondent was not able to complete the interview at the time of the call.

Respondents were asked their zip code in order to determine geographic eligibility. Interviews with out-of-area respondents were terminated. Interviews were continued with respondents who provided in-area or 'borderline' zip codes. 'Borderline' zip codes are zip codes associated with residential areas that are both inside and outside the borders of MA-CD4. Screening questions were asked to determine if the respondent was registered to vote at their current address. Respondents who said that they were not registered to vote or were not certain of their registration status, either in general or at their current address, were asked demographic questions necessary for weighting the sample. Registered voters continued with the main interview.

Notably, the survey instrument used the respondent-reported zip code to ascertain whether a respondent resided within MA CD-4. For the majority of the households in MA CD-4, geographic eligibility is knowable based on the zip code alone; for the remaining respondents – those living in households with borderline zip codes – it was necessary to determine geographic eligibility using 'geo-coding' information (i.e., 100 block and cross-street), collected at the end of the survey. Since geographic eligibility for these cases could not be determined programmatically, SSRS needed to conduct additional interviews in order to ensure that the final sample of completed interviews would contain a minimum of 300 landline and 100 cell completes with registered voters known to live in MA CD-4.

Overall, SSRS completed 41 full interviews (27 landline and 14 cell) with registered voters living in a 'borderline' zip code and asked demographic questions of 22 respondents living in a 'borderline' zip code who did not qualify for the full survey as registered voters. SSRS 'mapped' geocoding information for each borderline case and compared the location with boundaries of MA CD-4. Of the 'borderline' completes, SSRS determined that 11 (four landline and seven cell phone) were out of the area of MA CD-4; of the 'borderline' demographic-only interviews, SSRS determined that five (one landline and four cell) were out of the area. Thus, while SSRS completed 419 complete interviews and 119 demographic-only interviews, the final sample used for weighting included 408 interviews with registered voters and 114 demographic-only interviews.

In order to maximize survey response, SSRS enacted the following procedures during the field period:

- An average of 3 follow-up attempts were made to contact non-responsive numbers (no answer, busy, answering machine).

- Each non-responsive number was contacted multiple times, varying the times of day, and the days of the week that call-backs were placed using a programmed differential call rule.
- Respondents were offered the option to set a schedule for a call-back.
- Phone numbers received a daytime call attempt, if necessary.

Weighting Procedures

The final data were weighted to correct for variance in the likelihood of selection for a given case and to balance the sample to known population parameters in order to correct for systematic under- or over-representation of meaningful social categories.

Typically, data are weighted to Census parameters via the American Community Survey or the Current Population Survey. However, these data are only reliable down to the PUMA (Public Area Microdata) level. Because the Congressional District does not perfectly overlap with PUMA, SSRS utilized counts from Claritas, a Nielsen company, to weight the data for this survey. Claritas takes data from the decennial Census and models it from a variety of sources to update the 2010 Census counts quarterly, until the next Census in 2020. These data are therefore quite accurate, given our proximity to the 2010 Census.

We selected Claritas data for the block groups in MA CD-4. We then compared demographic frequencies for age, race, education, and gender to the “best fit” overlap of PUMA from the 2010 American Community Survey. The estimates were quite close. This is an important ‘check’ used to ascertain the reliability of the Claritas data in providing meaningful weighting targets for our sample because Claritas data provides race and ethnicity separately but these data are weighted in our sample in a single step. In addition, Claritas provides education for the 25+ population; thus, educational attainment for 18-24 year olds needs to be imputed using the ACS estimates to produce counts for the full 18+ adult population. SSRS has enacted this procedure for dozens of local-area studies and are quite confident in the accuracy of the results.

Phone use (cell phone only, dual users, and landline only) was modeled utilizing the same procedure used by the National Health Interview Survey to estimate phone use at the state level. Namely, a logistic regression was run within NHIS data, predicting these three phone use types separately. Then, Claritas and ACS estimates of the District were utilized to solve the regression equation for CD-4 specifically. This procedure found that 30.2% of CD-4 households are cell phone only, compared to only 12 percent that are landline only.

Demographic data were collected and weighting procedures were executed for all geographically-eligible respondents. These steps were necessary because universe counts for registered voters living in MA CD-4 are not available. After weighting the data of all respondents who are geographically eligible to the universe counts, the final step is to remove

cases that were not eligible as voters registered to vote at an address located within MA CD-4. This results in a final “self-weighted” sample of registered voters in CD-4.

The weighting procedure involved the following steps:

1. Phone-Status Correction (W_{PS}): Respondents whose household members answer both landlines and cell phones have a higher likelihood of inclusion in the sample. To correct for this, cases from dual-frame households were assigned a weight equal to half the weight assigned to single-mode households.

2. Within household selection correction (W_{HC}): To correct for the fact that only one qualifying adult was selected in any given household, landline cases from households with a single qualifying adult received a weight of 1, those with two received a weight of 2, and those with 3 or more qualifying adults received a weight of 3. Respondents with missing data were assigned the mean weight. Cell phone respondents received a weight of 1, as there was no within-household selection on the cell phones.

The product of these two stages was the baseweight for the sample.

$$BW = W_{PS} \times W_{HC}$$

3. Post stratification weighting: The baseweight was used as a balancing weight in the iterative proportionate fitting (IPF) process, or ‘raking.’ Universe counts were attained—through the procedure described earlier—for age, educational attainment, gender, phone use, and race.

Table 1: Comparison of Benchmark Data, Unweighted Sample, and Weighted Sample

Parameter	Value Label	Benchmark*	Unweighted*	Weighted*
Education	Less than High School	10.1%	5.4%	10.1%
	High School Graduate	23.4%	18.8%	23.4%
	Some College	25.9%	23.8%	25.9%
	College+	39.9%	51.3%	39.9%
Gender	Male	47.4%	46.6%	47.4%
	Female	52.6%	53.4%	52.6%
Age	18-24	12.8%	9.6%	12.8%
	25-34	14.4%	8.2%	14.4%
	35-44	17.4%	18.0%	17.4%
	45-54	21.2%	22.4%	21.2%
	55-64	16.2%	18.4%	16.2%
	65+	17.0%	22.4%	17.0%
Race	White	88.4%	88.7%	88.4%
	Black (non-Hispanic)	2.0%	1.5%	2.0%

	Hispanic	3.0%	3.8%	3.0%
	Other (non-Hispanic)	4.9%	4.2%	4.9%
Phone Use	Cell phone only	30.2%	10.9%	30.2%
	Dual Frame	57.8%	83.5%	57.8%
	Landline only	12.0%	5.6%	12.0%

*-Percentages may not add to 100% to account for cases where respondents refused to provide this demographic information.

Weighting procedures increase the variance in the data, with larger weights causing greater variance. Complex survey designs and post-data collection statistical adjustments affect variance estimates and, as a result, tests of significance and confidence intervals. The final design effect for the survey was 1.7, and the margin of sampling error was 4.85 (6.39 with design effect).

Response Rate

The landline response rate was 27.3% and the cell phone response rate was 14.0%, for an overall response rate of 20.9%, using AAPOR's RR3 formula. Below is a full disposition of the sample selected for the survey.

Table 2: Sample Dispositions

	LL	Cell	Total
Eligible, Interview (Category 1)			
Complete	304	104	408
Eligible, non-interview (Category 2)			
Refusal (Eligible)	17	15	32
Break-off	5	7	12
Answering Machine (Eligible)	0	3	3
Physically or mentally unable	0	0	0
Language problem	0	0	0
Unknown eligibility, non-interview (Category 3)			
Always busy	90	344	434
No answer	2375	7746	10121
Answering machine, don't know if household	117	4574	4691
Call blocking	36	396	432
Technical phone problems	3	19	22
Housing unit, unknown if eligible respondent	891	3687	4578
No screener completed	801	2917	3718
Not eligible (Category 4)			
Fax/data line	848	125	973
Non-working number	9420	4314	13734
Business, government office, other organizations	1076	344	1420
No eligible respondent	94	1509	1603
Total phone numbers used	16,077	26,104	42,181