METHODOLOGY REPORT:
EVERY VOICE, EVERY VOTE SURVEY

Prepared for The Lenfest Institute

Prepared by:

Robyn Rapoport

Rob Manley

Hope Wilson

FEBRUARY 2023

AAPOR Transparency Initiative
AMERICAN ASSOCIATION FOR PUBLIC OPINION RESEARCH

TABLE OF CONTENTS

OVERVIEW

The Lenfest Institute engaged SSRS to conduct the Every Voice, Every Vote survey ahead of the 2023 Philadelphia mayoral and city council election to help "elevate and amplify the voices" of Philadelphia residents to ensure that their concerns are accounted for. The goal of the survey is to understand how community groups and media organizations can engage Philadelphia's diverse communities and neighborhoods in the electoral process.

The Every Voice, Every Vote survey included both a qualitative and quantitative phase. Prior to fielding the survey, focus groups were carried out by the Institute for Survey Research (ISR) at Temple University in collaboration with SSRS. The questionnaire for the Every Voice, Every Vote survey was informed by the findings of the focus groups. This report details the methodology of the quantitative phase.[1]

The Every Voice, Every Vote survey obtained surveys via mixed-mode online, inbound telephone, and mail survey design. N=1,247 Philadelphia residents completed the survey, with n=939 completing via web, n=40 via inbound telephone, and n=268 via mail. Data collection was conducted in English and Spanish from Dec. 5, 2022, to Jan. 9, 2023.

Details on the sampling, questionnaire design, data collection, data processing, and weighting are discussed below.

SSRS PROFILE

SSRS is a full-service survey and market research firm managed by a core of dedicated professionals with advanced degrees in the social sciences. SSRS designs and implements research solutions for complex strategic, tactical, public opinion, and policy issues in the U.S. and in more than 40 countries worldwide. The SSRS team specializes in creative problem-solving and informed analysis to meet its clients' research goals. SSRS provides the complete set of analytical, administrative, and management capabilities needed for successful project execution. We partner with clients interested in conducting high-quality research. In the industry, SSRS is renowned for its sophisticated sample designs and its experience with all facets of data collection, including those involving multimodal formats. SSRS also has extensive statistical and analytical capabilities for extracting important insights from the survey data and suggesting strategies based on those insights.

SAMPLE DESIGN

The target population for this survey was adults ages 18 or older living in Philadelphia, Pennsylvania. SSRS drew a representative sample of the target population, using a full probability design.

[1] For more information about the qualitative phase, see final qualitative report provided by ISR.

Sampling Frame

The sampling frame uses an address-based sample (ABS) drawn from the United State Postal Service (USPS) Computerized Delivery Sequence File (CDSF). The CDSF is a computerized file that contains information on all delivery addresses serviced by the USPS. The sample was first stratified by region. Then within region the sample was stratified by ZIP code, and ZIP codes with higher incidences of Hispanic and African American residents were oversampled. Additionally, Hispanic and African American flags were appended to the sample frame and flagged records were sampled at higher rates to help increase sample representativeness among these populations that are less likely to respond. In total, n=14,424 addresses were selected. Table 1 shows the seven regions used for stratification.

Table 1: Sample Region Definitions

Region	ZIP Code
South and Center Philadelphia	19102, 19103, 19106, 19107, 19109, 19112, 19145, 19146, 19147, 19148
Southwest Philadelphia	19142, 19143, 19153
West Philadelphia	19104, 19131, 19139, 19151
Lower Northeast Philadelphia	19133, 19122, 19125, 19134, 19137, 19124, 19135, 19149
Upper Northeast Philadelphia	19111, 19152, 19136, 19114, 19115, 19116, 19154
North Philadelphia	19120, 19121, 19123, 19126, 19130, 19132, 19138, 19140, 19141, 19150
Northwest Philadelphia	19128, 19127, 19144, 19119, 19118, 19129

QUESTIONNAIRE DESIGN

Questionnaire Design and Pretest

Based on findings from the qualitative phase and discussions with The Lenfest Institute, SSRS drafted a survey outline with questions to include in the survey. Once SSRS and The Lenfest Institute iterated on the outline and decided on the content of the survey, SSRS created the full questionnaire and formatted it for online, phone, and paper administration. SSRS, in consultation with The Lenfest Institute, reviewed the pretest version for question wording, order, clarity, logic/programming, and other issues related to questionnaire quality and design across modes.

From Thursday, Nov. 17 through Monday, Nov. 21, SSRS completed six cognitive pretest interviews to evaluate the usability of the online and paper survey instruments and to identify questions that might be associated with measurement error because of possible confusion or misunderstanding. The online survey was pretested with four participants and the paper survey was pretested with two participants. For each cognitive pretest interview, an SSRS team member was on Zoom with the participant as they went through the survey, asking the participant to "talk out loud" as they went through the survey and asking various

probing questions. SSRS provided a detailed pretest memo of the findings and made minor changes based on discussion with The Lenfest Institute.

Programming

Prior to the field period, SSRS formatted the questionnaire and translated the survey instrument into Spanish. SSRS programmed the survey into its Forsta Plus (formerly Confirmit) platform for web administration in both English and Spanish. The program was optimized for administration via smartphone or other mobile handheld devices. Extensive checking of the program was conducted to ensure that skip patterns followed the design of the questionnaire. The web program was checked on multiple devices, including desktop computers and handheld mobile devices, and different web browsers to ensure consistent and optimized visualization across devices and web browsers.

SSRS generated unique survey passwords that were assigned and provided via mail to potential respondents. The web survey was accessed directly by respondents, using their unique passwords. This feature also gave respondents the ability to return to the survey later if they chose to suspend their interview.

DATA COLLECTION

The Every Voice, Every Vote survey was fielded from Dec. 5, 2022, to Jan. 9, 2023. The mailing protocol consisted of an invitation letter and follow-up reminders (a postcard and a letter with a paper survey) to all households included in the sample (n=14,424). SSRS crafted the invitation letter, reminder postcard and final reminder letter in consultation with The Lenfest Institute to make each material as appealing as possible.

The initial invitation letter was sent to each household in a No. 11 envelope. The invitation included a one-page letter inviting a member of the household to participate in an important research study. Samples identified as being more likely to be a Spanish-speaking household received a double-sided invitation with Spanish translations on the back of the letter. The invitation letter included a link (URL), an individual passcode to log on to the study, a QR code for easy scannable entry into the survey, and a toll-free number for respondents to call in to complete the survey with a trained interviewer. To increase the cooperation rate, a $1.25 cash pre-survey incentive was visible through the invitation letter envelope window. Additionally, the invitation letter offered $10 to respondents upon completion of the survey. The $10 payment via an electronic code was disbursed immediately after respondents completed the web survey and sent via a check in the mail to respondents completing and returning the paper survey questionnaire.

The reminder postcard included the same information provided on the invitation letter (i.e., the survey link, passcode, QR code, and toll-free number for respondents to call in to complete the survey) and was folded and sealed. Similar to the invitation letter, the postcard asked respondents to participate in this important research and included Spanish translations of the key points.

Finally, a reminder letter was sent to each household that had not responded to the initial invitation letter or reminder postcard. These final reminders were sent in a 9 x 12 envelope and contained the following materials:

- Personalized cover letter explaining the nature of the survey.
- One 8-page questionnaire booklet in English or two 8-page questionnaire booklets (one in English and one in Spanish) for ABS records identified as being more likely to be a Spanish-speaking household.
- Postage-paid business reply envelope (BRE).

For the Every Voice, Every Vote survey, the survey administration schedule was as follows:

Table 2: Survey Administration Schedule

		Date
1	Invitation letter	12/5/2022
2	Postcard	12/10/2022
3	Reminder letter and paper survey	12/20/2022

DATA PROCESSING AND INTEGRATION

SSRS implemented several quality assurance procedures in data file preparation and processing. Prior to launching data collection, extensive testing of the web survey was completed to ensure that it was working as anticipated. Data was also checked by the SSRS team following the pretests and throughout the field period to confirm that skip patterns were correctly followed.

The returned paper questionnaires were scanned, cleaned, and edited per the programming and skip instructions. They were then combined with the web and phone data. All data (collected online and through the paper surveys) was thoroughly cleaned with a computer validation program written by one of our data processing programmers. This program establishes editing parameters to locate any errors including data that does not follow skip patterns, out of range values, and errors in data field locations. The program confirmed that data was consistent with the definitions of codes and ranges and matched the appropriate bases of all questions.

As a standard practice, quality checks were incorporated into the survey. Quality control checks for this study included a review of "speeders" and reviewing the internal response rate (number of questions answered divided by the number of questions asked). Final data was reviewed for internal consistency, and any questions that have been asked in previous surveys[2] were reviewed for logic and consistency over time.

All hardcopy surveys that had conflicting data were reviewed by the project staff and adjudicated to ensure the best quality data. A total of 16 paper surveys were excluded from the final data set due to respondents skipping the screening questions at the beginning of the survey.

[2] A subset of questions in the survey was drawn from Pew's State of the City survey.

WEIGHTING

Overview

For this research, weighting is used to compensate for sample designs and patterns of nonresponse that might bias results. The weighting ensures that the demographic profile of the sample is representative of adult residents of Philadelphia County. The first stage of the weighting applies a sampling weight to adjust for any disproportional sample stratification included. Second, an adjustment is made to account for sampling of one adult within each household. The final stage rakes sample demographics to match population parameters.

For this study, two weights were developed for varying analytical purposes:

WEIGHT

- This weight should be used for analysis across the entire sample of completed interviews.

NEIGHBORHOOD WEIGHT

- This weight should be used when producing estimates within the following key subpopulations or when comparing the subgroups to each other:
    - South and Center Philadelphia
    - Southwest Philadelphia
    - West Philadelphia
    - Lower Northeast Philadelphia
    - Upper Northeast Philadelphia
    - North Philadelphia
    - Northwest Philadelphia

The next few sections detail the procedures for each of these weights, as well as final design effects and margins of error.

Base Weight

The first step in the weighting process is to assign a base weight. The base weight is simply the inverse of the sampling fraction and can be expressed as $d0_i = N_i/n_i$ where N_i is the sample frame count in stratum i and n_i is the number of addresses mailed stratum i.

Unknown eligibility adjustment

The first adjustment to the base weight is to adjust for cases with unknown eligibility. This is done by distributing the weights of cases with unknown eligibility among the cases for which eligibility is known. The unknown eligibility adjustment, a_1, is expressed as:

$$a_1 = \begin{cases} \sum_{i \in s} d_{0i} \Big/ \sum_{i \in s, KN} d_{0i}, & i \in s, KN \\ 0, & i \in s, UNK \end{cases}$$

where d_{0i} is the base weight for case i, s is the entire sample, s, KN is the subset of sample for which eligibility status has been determined, either eligible or ineligible, and s, UNK is the subset of sample for which eligibility status could not be determined. Note that these two groups account for the entire sample and do not overlap, i.e., $s, KN \cup s, UNK = s \; and \; s, KN \cap s, UNK = \emptyset$.

This adjustment was made within the seven regions. The unknown eligibility adjusted base weight, d_1, for unit i is the product of the base weight and the unknown eligibility adjustment, or $d_{1i} = d_{0i}a_{1i}$.

Nonresponse Adjustment

The next base weight adjustment is a nonresponse adjustment that distributes the weights of eligible nonresponders among eligible responders. The nonresponse adjustment, a_2, can be expressed as:

$$a_2 = \begin{cases} \sum_{i \in s,E} d_{1i} \Big/ \sum_{i \in s,ER} d_{1i}, & i \in s, ER \\ 1, & i \in s, IN \\ 0, & i \in s, ENR \end{cases}$$

Where d_{1i} is the unknown eligibility adjusted base weight for case i, s, E is the set of all eligible cases, s, ER is the set of all eligible respondents, s, IN are cases that are ineligible and s, ENR are eligible nonrespondents (i.e., cases that were determined to be eligible but never completed the survey).

This adjustment was also made within the seven regions. The nonresponse adjusted base weight, d_2, for unit i is the product of the base weight, the unknown eligibility adjustment, and the nonresponse adjustment, or $d_{2i} = d_{0i}a_{1i}a_{2i}$.

Number of Adults Adjustment

The final adjustment to the base weight is to account for the random selection of one adult in each sampled household. This adjustment, a_3, is the inverse of the selection probability within household and is simply the number of adults in the household.[3]

The final base weight is the product of the initial base weight and the subsequent base weight adjustments, or $d_{3i} = d_{0i}a_{1i}a_{2i}a_{3i}$.

Calibration

The final step in weighting is to calibrate sample demographic distributions to match known target population benchmarks. This is done using the ANESRAKE package in R.[4] Dimensions for raking include gender, age, education, race/ethnicity, and region. Benchmarks were derived from ACS 2021 1-Year and 5-Year estimates. Weights were trimmed at the 2nd and 98th percentile to prevent individual respondents having too much influence on survey-derived estimates.

[3] The number adults adjustment is capped at three to help contain the variance of the weights.
[4] https://rdrr.io/cran/anesrake/man/anesrake.html

Similarly, the sample across each of the seven regions was weighted to match population benchmarks per Claritas 2023 CY estimates on age, gender, education, and race/ethnicity. The following tables compare unweighted and weighted sample demographic distributions to target population benchmarks.

Table 3: Overall Sample Demographics

	Benchmark	Unweighted	Weighted
Sex by age			
Male 18-24	4.8%	2.2%	4.5%
Male 25-34	11.9%	8.0%	11.7%
Male 35-44	8.6%	5.8%	8.4%
Male 45-54	6.7%	3.6%	6.5%
Male 55-64	6.8%	6.2%	7.0%
Male 65+	7.4%	11.5%	7.5%
Female 18-24	5.1%	4.6%	5.2%
Female 25-34	12.8%	11.8%	12.9%
Female 35-44	9.2%	9.5%	9.3%
Female 45-54	7.4%	8.5%	7.5%
Female 55-64	8.1%	10.9%	8.2%
Female 65+	11.0%	17.4%	11.3%
Sex by education			
Male HS or less	20.3%	11.8%	19.8%
Male some college/associate	10.6%	6.8%	10.1%
Male college+	15.4%	18.8%	15.7%
Female HS or less	21.9%	20.8%	22.2%
Female some college/associate	13.2%	14.8%	13.3%
Female college+	18.6%	27.1%	18.9%
Age by education			
18-34, HS or less	10.7%	7.0%	10.5%
18-34, Some college/associate	9.4%	4.3%	8.9%
18-34, College+	14.6%	15.3%	14.9%
35-64, HS or less	21.5%	13.7%	21.2%
35-64, Some college/associate	10.5%	11.4%	10.6%
35-64, College+	14.9%	19.3%	15.2%
65+, HS or less	10.1%	11.9%	10.3%
65+, Some college/associate	3.9%	5.9%	4.0%
65+, College+	4.4%	11.2%	4.5%
Race/ethnicity			
White, not Hispanic	36.2%	39.4%	36.5%
Black, not Hispanic	37.9%	38.5%	37.5%
Hispanic	13.7%	12.4%	13.9%
Asian, not Hispanic	7.9%	6.4%	7.8%
Other, not Hispanic	4.3%	3.3%	4.3%

Internet use			
Yes	93.4%	93.3%	93.3%
No	6.6%	6.7%	6.7%

Table 4: Overall Sample Demographics (continued)

	Benchmark	Unweighted	Weighted
Region			
South and Center	14.7%	13.0%	14.8%
Southwest	6.8%	9.1%	6.9%
West	11.3%	13.6%	11.1%
Lower Northeast	19.6%	17.7%	19.9%
Upper Northeast	17.5%	17.4%	17.5%
North	21.5%	19.5%	21.0%
Northwest	8.7%	9.8%	8.9%

Table 5: Regional Sample Gender

		Benchmark	Unweighted	Weighted
Sex				
South and Center Philly	Male	47.8%	42.6%	46.7%
	Female	52.2%	57.4%	53.3%
	Total	100.0%	100.0%	100.0%
Southwest Philadelphia	Male	44.4%	32.7%	43.5%
	Female	55.6%	67.3%	56.5%
	Total	100.0%	100.0%	100.0%
West Philadelphia	Male	46.1%	36.7%	45.5%
	Female	53.9%	63.3%	54.5%
	Total	100.0%	100.0%	100.0%
Lower Northeast Philly	Male	47.0%	31.2%	46.7%
	Female	53.0%	68.8%	53.3%
	Total	100.0%	100.0%	100.0%
Upper Northeast Philly	Male	47.5%	42.9%	47.0%
	Female	52.5%	57.1%	53.0%
	Total	100.0%	100.0%	100.0%
North Philadelphia	Male	45.2%	35.0%	45.1%
	Female	54.8%	65.0%	54.9%
	Total	100.0%	100.0%	100.0%
Northwest Philadelphia	Male	44.9%	41.8%	45.1%
	Female	55.1%	58.2%	54.9%
	Total	100.0%	100.0%	100.0%

Table 6: Regional Sample Age

		Benchmark	Unweighted	Weighted
Age				
South and Center Philly	18-24	7.4%	7.4%	7.5%
	25-34	27.8%	30.2%	27.4%
	35-44	18.7%	13.0%	18.0%
	45-54	14.3%	9.9%	14.6%
	55-64	12.7%	11.1%	12.9%
	65+	19.2%	28.4%	19.6%
	Total	100.0%	100.0%	100.0%
Southwest Philadelphia	18-24	10.4%	5.3%	10.6%
	25-34	24.4%	15.9%	23.1%
	35-44	17.3%	15.0%	17.6%
	45-54	13.9%	12.4%	14.1%
	55-64	15.0%	14.2%	15.2%
	65+	19.0%	37.2%	19.3%
	Total	100.0%	100.0%	100.0%
West Philadelphia	18-24	20.0%	10.1%	19.5%
	25-34	21.7%	23.7%	22.0%
	35-44	16.2%	15.4%	16.1%
	45-54	11.8%	9.5%	11.9%
	55-64	12.5%	15.4%	12.6%
	65+	17.7%	26.0%	17.9%
	Total	100.0%	100.0%	100.0%
Lower Northeast Philly	18-24	12.6%	5.9%	12.7%
	25-34	24.2%	19.0%	23.8%
	35-44	18.6%	21.7%	18.7%
	45-54	14.9%	17.2%	15.0%
	55-64	14.3%	17.6%	14.4%
	65+	15.4%	18.6%	15.5%
	Total	100.0%	100.0%	100.0%
Upper Northeast Philly	18-24	7.6%	5.1%	7.7%
	25-34	19.7%	11.5%	18.5%
	35-44	17.2%	15.2%	17.5%
	45-54	14.3%	12.9%	14.5%
	55-64	15.8%	19.4%	16.1%
	65+	25.4%	35.9%	25.8%
	Total	100.0%	100.0%	100.0%

Table 7: Regional Sample Age (continued)

		Benchmark	Unweighted	Weighted
Age				
North Philadelphia	18-24	12.2%	9.1%	12.3%
	25-34	24.0%	21.0%	24.1%
	35-44	16.9%	14.0%	16.8%
	45-54	13.4%	11.5%	13.2%
	55-64	14.6%	18.9%	14.7%
	65+	18.8%	25.5%	18.9%
	Total	100.00	100.0%	100.0%
Northwest Philadelphia	18-24	10.0%	3.3%	10.0%
	25-34	22.9%	18.0%	23.0%
	35-44	17.0%	9.0%	17.0%
	45-54	13.8%	9.0%	13.5%
	55-64	15.0%	21.3%	15.0%
	65+	21.4%	39.3%	21.5%
	Total	100.0%	100.0%	100.0%

Table 8: Regional Sample Education

		Benchmark	Unweighted	Weighted
Education				
South and Center Philly	18-24 years old	7.4%	7.4%	7.5%
	HS or less	29.1%	14.8%	27.7%
	Some college/associate	10.4%	9.9%	10.6%
	College+	53.1%	67.9%	54.1%
	Total	100.0%	100.0%	100.0%
Southwest Philadelphia	18-24 years old	10.4%	5.3%	10.6%
	HS or less	46.2%	35.4%	45.3%
	Some college/associate	18.4%	29.2%	18.7%
	College+	25.0%	30.1%	25.4%
	Total	100.0%	100.0%	100.0%
West Philadelphia	18-24 years old	20.0%	10.1%	19.5%
	HS or less	34.3%	30.2%	34.3%
	Some college/associate	17.5%	20.1%	17.7%
	College+	28.2%	39.6%	28.5%
	Total	100.0%	100.0%	100.0%
Lower Northeast Philly	18-24 years old	12.6%	5.9%	12.7%
	HS or less	54.2%	35.7%	54.0%
	Some college/associate	13.0%	15.8%	13.1%
	College+	20.1%	42.5%	20.3%
	Total	100.0%	100.0%	100.0%
	18-24 years old	7.6%	5.1%	7.7%

Upper Northeast Philly	HS or less	48.7%	37.8%	47.9%
	Some college/associate	14.6%	24.0%	14.8%
	College+	29.1%	33.2%	29.5%
	Total	100.0%	100.0%	100.0%
North Philadelphia	18-24 years old	12.2%	9.1%	12.3%
	HS or less	45.6%	32.9%	45.4%
	Some college/associate	16.0%	24.7%	16.1%
	College+	26.1%	33.3%	26.3%
	Total	100.0%	100.0%	100.0%
Northwest Philadelphia	18-24 years old	10.0%	3.3%	10.0%
	HS or less	28.1%	16.4%	27.8%
	Some college/associate	14.4%	13.9%	14.4%
	College+	47.6%	66.4%	47.7%
	Total	100.0%	100.0%	100.0%

Table 9: Regional Sample Race/Ethnicity

		Benchmark	Unweighted	Weighted
Race / Ethnicity				
South and Center Philly	White, not Hispanic	61.7%	72.8%	61.8%
	Black, not Hispanic	12.0%	8.6%	12.2%
	Hispanic	8.1%	7.4%	8.3%
	Asian, not Hispanic	15.0%	8.6%	14.4%
	Other, not Hispanic	3.2%	2.5%	3.3%
	Total	100.0%	100.0%	100.0%
Southwest Philadelphia	White, not Hispanic	12.2%	15.0%	12.5%
	Black, not Hispanic	75.0%	73.5%	74.6%
	Hispanic	4.1%	4.4%	4.2%
	Asian, not Hispanic	5.3%	2.7%	5.4%
	Other, not Hispanic	3.3%	4.4%	3.4%
	Total	100.0%	100.0%	100.0%
West Philadelphia	White, not Hispanic	19.6%	18.9%	19.1%
	Black, not Hispanic	63.1%	65.1%	63.4%
	Hispanic	4.4%	3.0%	4.4%
	Asian, not Hispanic	9.5%	8.3%	9.6%
	Other, not Hispanic	3.4%	4.7%	3.4%
	Total	100.0%	100.0%	100.0%
Lower Northeast Philly	White, not Hispanic	32.0%	41.6%	31.7%
	Black, not Hispanic	23.7%	23.5%	23.8%
	Hispanic	32.6%	27.6%	32.7%
	Asian, not Hispanic	8.2%	4.5%	8.2%
	Other, not Hispanic	3.5%	2.7%	3.5%
	Total	100.0%	100.0%	100.0%
Upper Northeast Philly	White, not Hispanic	57.0%	60.4%	56.4%
	Black, not Hispanic	14.7%	14.7%	14.9%
	Hispanic	11.4%	12.9%	11.6%

	Asian, not Hispanic	13.1%	9.7%	13.3%
	Other, not Hispanic	3.8%	2.3%	3.9%
	Total	100.0%	100.0%	100.0%
North Philadelphia	White, not Hispanic	16.5%	17.3%	16.6%
	Black, not Hispanic	60.3%	58.8%	60.0%
	Hispanic	14.7%	15.6%	14.8%
	Asian, not Hispanic	5.6%	4.5%	5.6%
	Other, not Hispanic	3.0%	3.7%	3.0%
	Total	100.0%	100.0%	100.0%
Northwest Philadelphia	White, not Hispanic	48.5%	48.4%	48.7%
	Black, not Hispanic	39.3%	37.7%	39.1%
	Hispanic	4.8%	4.9%	4.8%
	Asian, not Hispanic	3.3%	5.7%	3.3%
	Other, not Hispanic	4.1%	3.3%	4.1%
	Total	100.0%	100.0%	100.0%

Design Effect and Margin of Sampling Error

Post-data collection statistical adjustments require analysis procedures that reflect departures from simple random sampling. SSRS calculates the effects of these design features so that an appropriate adjustment can be incorporated into tests of statistical significance when using this data. The so-called "design effect" or *deff* represents the loss in statistical efficiency that results from a disproportionate sample design and systematic nonresponse. The total sample design effect for this survey is 1.55.

The final design effects and margins of error are outlined below.

Table 10: Design Effect and Margin of Error

	N-size	Design effect	Margin of error
Region			
South and Center Philly	162	1.66	9.9
Southwest Philadelphia	113	1.58	11.6
West Philadelphia	169	1.48	9.2
Lower Northeast Philly	221	1.83	8.9
Upper Northeast Philly	217	1.51	8.2
North Philadelphia	243	1.55	7.8
Northwest Philadelphia	122	2.09	12.8
Total sample	1,247	1.55	3.5

SSRS calculates the composite design effect for a sample of size n, with each case having a weight, w, as:[5]

$$deff = \frac{n\sum w^2}{(\sum w)^2}$$

[5] Kish, L. (1992). Weighting for Unequal Pi. *Journal of Official Statistics*, Vol. 8, No.2, 1992, pp. 183-200.

The survey's margin of error is the largest 95% confidence interval for any estimated proportion based on the total sample—the one around 50%. For example, the margin of error for the entire sample is ± 3.5 percentage points. This means that in 95 of every 100 samples drawn using the same methodology, estimated proportions based on the entire sample will be no more than 3.5 percentage points away from their true values in the population. Margins of error for subgroups will be larger.

It is important to remember that the sampling fluctuations captured in the margin of error are only one possible source of error in a survey estimate. Other sources, such as respondent selection bias, questionnaire wording, and reporting inaccuracy, may contribute additional error of greater or lesser magnitude.

How to Analyze Data with Oversamples

It is a common practice to oversample certain groups of interest to provide larger sample sizes for analysis. When groups are oversampled, weighting will correct for the oversampling by "weighting down" the groups to their proper proportion of the sample.

It is important for researchers to understand the weighting implications of these oversamples. SSRS typically computes "balancing weights," which means that the weights across the entire sample sum to the total number of interviews. If we have oversampled a group, the sum of that group's balancing weight will then be less than the number of interviews we completed with the group because that group has been weighted down in the aggregate. If such data was analyzed with a basic statistics package like SPSS, the margin of error for the oversample population would reflect the weighted n-size and not the number of interviews that would lead to an overestimate of the sample variance.

The following table shows an example of population and interview n-sizes when an oversample is used. For this example, a main cross-section sample of 1,000 was combined with an oversample of 800 among some subpopulation of interest. Although the researcher did 920 interviews with the oversample population, the statistical software will run statistical tests as though only 216 interviews were completed.

Table 11: Example of Oversample N-Sizes

	Natural population distribution (%)	Example study sample completes:			
		Main sample	Over-sample	Total	Weighted n-size
Non-oversample population	88%	880 (88%)	0	880 (49%)	1,584 (88%)
Oversample population	12%	120 (12%)	800	920 (51%)	216 (12%)
Total	100%	1,000	800	1,800	1,800

There are two solutions to this problem. The first is to utilize a statistics package that can apply a Taylor Series Linearization to the data. Under this procedure, the researcher would enter a strata variable[6] into the statistics package that indicates the sample selections upon which under/oversampling occurred. In effect, this will allow the statistics package to calculate proper margins of error for estimates based on the true sample sizes of groups. Taylor Series Linearization will also account for the impact of any complex sample design features, such as stratification, on sample variances. The researcher will also attain a margin of error appropriate to the number of interviews rather than the weighted n-size, which can be a problem in some statistical software packages such as SPSS. Statistics packages with the capability to compute linearized variances estimates include SAS with the survey procedures module, R with the *survey* package, Stata, and SPSS with the Complex Samples module.

If one does not have access to such a package, SSRS will provide a secondary weight to be used to conduct analyses within oversampled groups or between oversampled groups and other respondents, as the main weight supplied with the data will be appropriate for analysis of the overall population only.

Researchers should be aware that these two methods will obtain equivalent point estimates; however, they may not obtain equivalent sample variances, meaning that results of statistical tests could differ depending on the method used. In general, when the two methods differ, Taylor Series Linearization will obtain the most accurate sample variances and statistical tests, both overall and within subgroups. Therefore, if the researcher has access to software that can conduct Taylor Series Linearization, this is the preferred method.

Regardless, SSRS will identify the strata and PSU variables whenever they are applicable, so that researchers can properly analyze their data with the correct margins of error.

RESPONSE RATE

The response rate for this study was calculated using AAPOR's RR3.

Table 12: Response rate

Total records	14,424
Ineligibles[7]	51
Returned mail	332
Valid sample	12,794
Completes	1,247
Response rate	11.1%

[6] Or a Primary Sampling Unit (PSU) for a multistage sample design.
[7] Includes n=16 paper surveys removed due to skipping the screening questions.

DELIVERABLES

SSRS provided The Lenfest Institute with the following deliverables from the quantitative phase of the study:

- SPSS dataset with an overall city weight (weight), and a region-based weight (neighborhood_weight)
- Three (3) weighted banners of cross tabulations
- Topline questionnaire
- Final methodology report

CONTACT

Contact Robyn Rapoport for additional information.

rrapoport@ssrs.com | 484.840.4354 | @RobynRapoport

1 Braxton Way
Suite 125
Glen Mills, PA 19342