Methodology

The survey involved a representative sample of 1,000 adults who moved out of Philadelphia in the 2014-18 period. Data collection was conducted in English and Spanish from Nov. 13 to Dec. 12, 2018. Respondents were reached via mail and were asked to complete the survey online or through a paper version of the survey. Statistical results were weighted to correct known demographic discrepancies. The margin of sampling error for the complete set of weighted data is plus or minus 4.6 percentage points. The questionnaire was developed by The Pew Charitable Trusts, and the survey was conducted by SSRS, a nationally recognized public opinion research firm based in Glen Mills, Pennsylvania.

The target population for this study was adults age 18 or older who were former residents of Philadelphia. Working with Pew, SSRS used an address-based sample (ABS) design. Under this design, Pew and SSRS started with Philadelphia County addresses from the U.S. Postal Service's Computerized Delivery Sequence File (CDSF). Since the CDSF includes only addresses, Marketing Systems Group (MSG) was contracted to append names to the addresses using licensed commercial databases. Addresses where a name was successfully appended were then matched against the USPS' National Change of Address (NCOA) database to identify the people who had moved out of Philadelphia within the previous 48 months. NCOA-flagged movers constituted the sampling frame for this study.

All selected movers received a one-page, single-sided study invitation letter from Pew, explaining the purpose of the survey. For records flagged with a Hispanic surname, the letter was printed double-sided, with one side in English and the other in Spanish. The letter included a short URL for the survey and a personalized PIN for respondents to use to access it online. The mailing also included a $2 bill as a pre-incentive and an offer of a $10 Visa debit card upon completion of the survey.

Two days after the mailing of the study invitation letter, a reminder postcard was sent out. Approximately two weeks after the study invitation letters were mailed, questionnaire packets were sent to nonresponders. This mailing contained a personalized cover letter explaining the nature of the survey; one eight-page questionnaire booklet in English, or two eight-page questionnaire booklets (one in English and one in Spanish) for ABS records flagged with Hispanic surnames, and a postage-paid business reply envelope.

To detect any issues related to the questionnaire, sampling, or response rate, SSRS conducted the mailing in two phases. Phase I involved sending out 1,000 invitation letters, reminder postcards, and mail questionnaire packets. Within approximately two weeks of the point of sending the Phase I study invitation letters, a Phase II mailing for another 4,667 cases was conducted. Phase II addresses received only the invitation letter and reminder postcard—mail questionnaire packets were not sent to this group due to the high response rate achieved. In total, 94 percent of the surveys were completed via web, while 6 percent were completed via mail.

Data from both web and paper modes were thoroughly cleaned with a computer validation program written by one of SSRS' data processing programmers. This program established editing parameters to locate any errors, including data that did not follow skip patterns, out-of-range values, and errors in data field locations.

WEIGHTING AND ANALYSIS

Weighting is generally used in survey analysis to compensate for sample designs and patterns of nonresponse that might bias results. The weighting ensures that the demographic profile of the sample matches the profile of the target population.

The sample was balanced to match former Philadelphia County resident parameters for sex, age, education, race/ethnicity, and census region. The weighting parameters were derived from the American Community Survey (ACS) 2012-2016 IPUMS file downloaded from the usa.ipums.org website.[1] Former Philadelphia residents were defined as adults who currently live in households outside Philadelphia County but lived in the county one year ago. Table 1 lists the dimensions used in the raking.

Table 1: Raking Dimensions

Dimension	Value label
Sex	Male
	Female
Age breaks	18-24
	25-29
	30-35
	36-49
	50-60
	61-64
	65+
Education	Less than high school
	High school graduate
	Some college
	College graduate or more
Race/ethnicity	White/not Hispanic
	Black/not Hispanic
	Hispanic
	Other/not Hispanic
Census region	Northeast
	Midwest
	South
	West

Weighting was accomplished using SPSSINC RAKE, an SPSS extension module that simultaneously balances the distributions of all variables using the GENLOG procedure. Weights were trimmed to prevent individual interviews from having too much influence on the final results. The use of these weights in statistical analysis ensures that the demographic characteristics of the sample closely approximate the demographic characteristics of the target population. Table 2compares weighted and unweighted total sample distributions to population parameters.

---

[1] Steven Ruggles, Sarah Flood, Ronald Goeken, Josiah Grover, Erin Meyer, Jose Pacas, and Matthew Sobek. IPUMS USA: Version 8.0 [dataset]. Minneapolis: University of Minnesota, 2018. https://doi.org/10.18128/D010.V8.0.

Table 2: Population Parameters and Weighted and Unweighted Total Sample Distribution

Characteristic	Value label	Parameter	Unweighted	Weighted
Sex	Male	48.9%	47.3%	51.1%
	Female	51.1%	52.7%	48.9%
Age breaks	18-24	19.4%	2.0%	11.4%
	25-29	23.5%	13.5%	25.9%
	30-35	19.8%	23.8%	21.4%
	36-49	21.1%	28.8%	23.3%
	50-60	8.1%	11.6%	9.0%
	61-64	2.8%	5.7%	3.1%
	65+	5.3%	14.6%	5.9%
Education	Less than high school	5.7%	1.5%	5.6%
	High school graduate	16.6%	11.4%	14.0%
	Some college	26.1%	18.2%	25.2%
	College graduate or more	51.6%	68.9%	55.2%
Race/ethnicity	White/not Hispanic	54.6%	65.8%	56.8%
	Black/not Hispanic	21.9%	14.7%	20.0%
	Hispanic	10.7%	6.9%	10.3%
	Other/not Hispanic	12.8%	12.6%	12.9%
Census region	Northeast	62.6%	69.1%	62.3%
	Midwest	4.4%	5.3%	4.4%
	South	22.5%	17.6%	22.9%
	West	10.5%	8.0%	10.4%

Effects of Sample Design on Statistical Analysis

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 these 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 2.20.

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

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

In a wide range of situations, the adjusted standard error of a statistic should be calculated by multiplying the usual formula by the square root of the design effect ($\sqrt{deff}$). Thus, the formula for computing the 95 percent confidence interval around a percentage is:

$$\hat{p} \pm \left(\sqrt{deff} \times 1.96 \sqrt{\frac{\hat{p}(1 - \hat{p})}{n}} \right)$$

where $\hat{p}$ is the sample estimate and n is the unweighted number of sample cases in the group being considered.

The survey's margin of error is the largest 95 percent confidence interval for any estimated proportion based on the total sample—the one around 50 percent. For example, the margin of error for the entire sample is plus or minus 4.6 percentage points. This means that in 95 out of every 100 samples drawn using the same methodology, estimated proportions based on the entire sample will be no more than 4.6 percentage points away from their true values in the population. Margins of error for subgroups will be larger. It is important to remember that sampling fluctuations 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.

RESPONSE RATES

Table 3 reports the response rate, which is computed according to American Association for Public Opinion Research standards.[2] The response rate for this study was 18.7 percent.

---

[2] The American Association for Public Opinion Research. 2016. Standard Definitions: Final Dispositions of Case Codes and Outcome Rates for Surveys. 9th edition. AAPOR.

Table3: Sample Disposition

Disposition	N
1. Complete (I)	1,000
2. Eligible, noninterview (R)	26
Refusal and break-off	26
3. Unknown eligibility, noninterview (UH)	4,581
Nothing ever returned	4,437
Refused, unknown if eligible	2
Undeliverable	142
4. Not eligible, returned (IN)	60
Currently lives in Philadelphia	37
Never lived in Philadelphia	3
Deceased	20
Total records contacted	5,667
e=(I+R)/(I+R+IN)	94.5%
RR3=I/[I+R+(e*UH)]	18.7%