Technical Report

Panel Study of Income Dynamics
Construction and Evaluation of the 2019 Longitudinal Individual and Family Weights

April 2021

Wen Chang, Raphael Nishimura, Steven G. Heeringa, David Johnson,
and Narayan Sastry

Survey Research Center, Institute for Social Research
University of Michigan,
Ann Arbor, MI

This technical report documents the sample design of the 2017 Immigrant sample and the methodological approach to creating the longitudinal weights constructed for the family units and individuals from the 2019 Panel Study of Income Dynamics (PSID-2019).

This technical report is organized into five sections. Section I is the introduction. Section II provides an overview of the PSID 2017 Immigrant sample and its joint inclusion probabilities. Section III lists the steps to construct the longitudinal weights for the PSID 2019 sample. Section IV provides the details of the methods used at each step of the weight construction. The report concludes in Section V with a descriptive analysis of the weights, including comparisons of distributions of U.S. socioeconomic characteristics using weighted estimates from the Current Population Survey (CPS), American Community Survey (ACS) and PSID.

I. Introduction

The 2019 PSID panel is based on the dynamic, longitudinal follow-up of individuals and their families originally identified in a combination of four probability samples of U.S. households. The PSID started in 1968 and was comprised of two separate samples: an equal probability national sample of households selected from the Survey Research Center 1960 National Sample (SRC sample) and a subsample of families interviewed in 1967 by the Bureau of the Census for the Office of Economic Opportunity (SEO sample) (McGonagle and Schoeni, 2006). Sample persons and their linear descendants identified in the baseline SRC and SEO samples have been interviewed since 1968[1]. In 1997 and 1999, a baseline sample of post-1968 immigrants (1997 Immigrant sample) was added and these immigrant sample persons have been followed continuously since the late 90s (Heeringa and Connor, 1998). In 2017 and 2019, a baseline sample of the post-1997 immigrants (2017 Immigrant sample) was added to the PSID panel.

Under the "dynamic" sample follow-up design, PSID interviewed 9,569 families in 2019. Among them, 9,119 families are members of the 1968 SRC/SEO or 1997 Immigrant samples (which we will refer to as Core families in this technical report) and 450 families are members of the 2017 Immigrant sample. Included in the Core and 2017 Immigrant families are 26,084

1 PSID has developed a measure of cumulative response rates for the original sample of persons interviewed in the initial 1968 wave. It estimates the proportion of the surviving members of the original 1968 census who were interviewed at each wave of data collection. The cumulative response rate report is available from the PSID website (see Heeringa et al., 2018).

individuals: 19,055 PSID "sample persons" (see Table 1) and 7,029 "non-sample" spouses and family members.

PSID traditionally categorizes individuals into one of two groups: sample persons and nonsample persons. Sample persons are individuals who were living in the original family unit at the time of the very first interview and their lineal descendants[2] born after 1968. For subsequent samples, such as the immigrant samples, the year of the first interview serves as the base for determining who is an original sample member, and all individuals present in the family at that time qualify. For the 1997 Immigrant sample, the sample status determination for individuals is 1997 or 1999, depending on when the family was first interviewed. For the 2017 Immigrant sample, the sample status determination for individuals is 2017 or 2019, depending on when the family was first interviewed.

PSID sample persons, including all those leaving to establish separate family units (split-offs), are tracked and followed. In 2019, the following rule was the same as in the prior 2017 wave[3]. Specifically, sample persons who participated in the previous wave survey were followed. Additionally, the PSID attempted to obtain an interview with sample individuals who did not respond in the prior wave (2017 survey year), but responded in the 2015 survey year. The 2017 Immigrant sample was followed in 2019 regardless of their response status in 2017.

Each sample person successfully interviewed for 2019 received a positive value for their 2019 longitudinal individual weight. The PSID rules for following household members were designed to maintain a nationally representative sample of families and individual family members at any point in time as well as across time, excluding the immigrants entering the U.S after 1968 (Gouskova et al 2008). The immigrant refreshment samples were added in 1997 and 2017 with

[2] The definition of sample person, especially linear descendant, has changed slightly over the years. From 1968 to 1993, a sample person was defined as someone who was either an original sample person; i.e., resident of a PSID sample family in 1968, or offspring born to or adopted by a sample individual who was actively participating in the study at the time. A newborn child had to appear in the study at the wave immediately following their birth to be considered a sample person. In 1994, the definition of a sample person was expanded to include children born to or adopted by a sample person when the sample person was not participating in the study; i.e., the child need not be residing with a responding panel family unit at birth or adoption. The same current PSID definition of sample persons (implemented in 1994) applies to the 1997/1999 Immigrant sample and 2017/2019 Immigrant sample.

[3] For more detail on the following rules in 1993-2005 survey years see Table 1 in Gouskova et al. (2008) (http://psidonline.isr.umich.edu/data/weights/Long-weights-doc.pdf)

the goal of allowing the PSID sample to remain nationally representative of all U.S. families.

All other members of PSID families are considered nonsample persons. They are typically spouses and partners or other family unit members. See McGonagle and Schoeni (2006) for a detailed background on the PSID. Nonsample persons received a 2019 PSID individual longitudinal weight equal to zero (0), and thus are excluded from any properly weighted longitudinal analysis of the PSID individual data.

Because 2019 is the first "follow-up" year that we will create longitudinal weights for the 2017 Immigrant sample, the weight construction will go through different processes depending on the type of sample. The following section provides an overview of the 2017 Immigrant sample design that makes it easier to understand the methods used to construct the PSID 2019 longitudinal weight.

II. Overview of 2017 Immigrant Sample

In 2017, a baseline sample of post-1997 immigrant families and individuals was added to the PSID. The 2017 Immigrant sample was recruited by either the New Immigrant Supplement (NIS-2017) study or the New Immigrant Multiplicity Supplement (NIMS) study. The NIMS is a special multiplicity sample that was needed because the 2017 Immigrant sample was recruited concurrently with the 2016/2017 new cohort recruitment for the Health and Retirement Study (HRS, see below). For a family to be eligible, the 2017 Immigrant sample screening process required that at least one family member who was a reference person or spouse/partner immigrated to the U.S. after 1997. Therefore, a subgroup of the 2017 Core families could also be eligible for the 2017 Immigrant sample if one member (i.e., reference person or spouse/partner) had immigrated to the U.S. and joined the family after 1997. This differs from the eligibility rule used in the 1997 Immigrant sample where both reference person and spouse/partner for families headed by a couple were required to be post-1968 immigrants. A consequence of the 2017 Immigrant sample eligibility rules is that the weight construction for PSID 2019 for different types of sample will go through different processes because of the existence of a joint inclusion probability from the different sample sources. To make it easier to describe the methods that we use for calculating the weights, we have categorized the PSID 2019 sample families into six

types and listed them in Table 1.

Concurrent Recruitment with HRS

Individuals who were born outside of the U.S. and entered the U.S after 1997 comprise 7.5% of the total U.S. population in 2016 based on ACS 2016 one-year Public Use Microdata Sample (PUMS) data. In order to reduce the cost of probability sampling and screening for eligible post 1997 immigrants, the PSID NIS-2017 sample was recruited in 2016 concurrently with the screening of "Late Baby Boomer (LBB)" and "Early Generation X (EGX)" cohorts for the Health and Retirement Study (HRS). The HRS screener identified households with any individuals who were born between 1960 and 1971. For households that did not have anyone born in these target years, the HRS screener asked two additional questions for the PSID to identify households in which anyone was born outside the U.S. and moved to the U.S. since 1997. The PSID conducted an additional screening phase to identify the family units in which the reference person and/or spouse/partner were recent immigrants. Due to overlap with the HRS screening for its new cohorts, recent immigrants born between 1960 and 1971 (as well as post-1997 immigrants who co-reside with individuals born in these years) were not part of the PSID NIS-2017 sample. In the rest of this report, 1960 to 1971 will be referenced as 'donut hole years'; recent immigrants born between 1960 and 1971 will be referenced as the 'donut hole' group; and recent immigrants born before 1960 or after 1971 will be referenced as the 'donut' group. The families in which the reference person and/or the spouse/partner are in the donut were recruited through the NIS study. The families, in which the reference person and/or the spouse/partner are in the donut hole, were recruited through a separate process, the PSID-2017 New Immigrant Multiplicity Supplement (NIMS or NIS-Donut). The NIMS sample was screened in 2017 and was added to the PSID panel in 2019.

Joint Inclusion Probability

The 2017 Immigrant sample consists of the families in which the reference person and/or the spouse/partner are recent immigrants who were either: 1) born abroad[4] and entered the U.S. after

[4] Another difference in the screening criteria for 1997 and 2017 Immigrant samples are the definition of abroad. To be eligible for the 2017 Immigrant sample, the reference person and/or the spouse/partner must have been recent immigrants who were born abroad and moved to the U.S. after 1997. This includes the reference person and/or the spouse/partner who were born in U.S.

1997; or 2) born in 1997 or later to foreign parents who were not in the U.S. in 1997. As noted above, the screening criterion for the recruitment of the 2017 Immigrant sample is slightly different from that used to recruit the 1997 Immigrant sample.

For families headed by a couple to be eligible for the 1997 Immigrant sample, both members of the couple, reference person and spouse/partner, must have immigrated to the U.S. after 1968. In contrast for families headed by a couple in 2017, if either the reference person or spouse/partner (or both) of a screened family unit were recent immigrants, the family unit was eligible for the PSID NIS-2017 or NIMS recruitments. Under the 2017 eligibility rule, families who were eligible for the NIS-2017 or NIMS included: 1) single reference person – recent immigrant; 2) reference person – recent immigrant; spouse/partner – recent immigrant; 3) reference person – recent immigrant; spouse/partner – not recent immigrant; and 4) reference person – not recent immigrant; spouse/partner – recent immigrant. Families in categories (1) and (2) could only enter the PSID through the NIS-2017 or NIMS. However, the families in categories (3) and (4) could have been added to the PSID panel prior to 2017 if either the reference person or the spouse/partner were herself/himself Core-eligible and a sample person. For this reason, such families have a dual chance of being selected in PSID. We identified 140 families from the NIS-2017 sample, 15 families from NIMS sample and 98 families from the PSID Core sample as being in categories (3) and (4) in 2019, hereafter referenced as 'Joint' families. An integration of weights to account for the dual chance of selection is needed for the Joint families. In addition, the individuals living in the Joint Core families should be defined as sample persons as if 2017, or 2019 for recontact[5] in 2019, is their baseline year. The nonsample persons in the identified Joint Core families are changed to sample persons in 2019 and will be followed and tracked as sample persons. The details of the longitudinal weight construction for the Joint families are described in section IV. For ease of communication, families are classified into different subgroups described in Table 1 based on the existence of a joint inclusion probability and the sample sources involved with the joint inclusion. Table 1 also lists the number of families and sample persons in each of these subgroups.

territories and entered the U.S. after 1997. For the 1997 Immigrant sample, if the reference person or the spouse/partner was born in the U.S. territories, the family unit was not eligible.

[5] If a family did not respond in 2017 and are followed in 2019, they are a 'recontact' family in 2019.

New Immigrant Multiplicity sample (NIMS or NIS-Donut)

From the HRS/PSID screening activities in 2016, PSID obtained a sample of post-1997 immigrants to the U.S. Due to exclusions of persons who were age-eligible for the HRS LBB and EGX cohorts, the standard PSID post 1997-immigrant screen did not cover families that included one or more new immigrants individuals who were born between 1960 and 1971 (i.e., a "donut hole"). A multiplicity sampling technique (Birnbaum and Sirken, 1965; Sirken, 1970) was used to obtain a representative sample of individuals in the donut hole via the individuals who joined the PSID sample as new post-1997 immigrants. This approach was followed only for Reference persons and Spouse/Partners who themselves met the new immigrant screening criteria. If both the Reference person and the Spouse/Partner were classified as immigrants, then both were eligible to nominate related individuals for the multiplicity sample (persons in their "network").

Each reference person or spouse/partner who met the criteria for being a "primary new immigrant" was asked to identify all of their relatives from three categories: (a) parents; (b) adult siblings; and (c) adult children, who meet the following criteria:

B1. Were born outside the U.S.; did not live in the U.S. in 1997; moved to the U.S. after 1997; are currently living in the U.S.; and
B2. Were born between 1960 and 1971.
B3. Are not living with the respondent.

After the pool of eligible individuals was identified for the multiplicity sample, the PSID field staff contacted the family, performed the 2017 Immigrant sample screen, determined their "multiplicity" (number of relatives that in the three aforementioned categories who meet the multiplicity sampling criteria) and interviewed them. In 2019, 25 NIMS families (75 sample persons) were added to the PSID panel.

III. Overview of methods to construct the longitudinal weights for the PSID-2019 sample

For the families in the Core sample that do not have joint inclusion probability with the 2017 Immigrant sample, we used the conventional method for computing 2019 PSID longitudinal

weights (Section IV.1). The construction of the longitudinal individual weight for the 2017 Immigrant sample (Section IV.3) starts by accounting for the unequal probability of selection for these families and goes through additional adjustments to address potential nonresponse bias and non-coverage errors. The existence of the Joint Core sample (Core-J-D; Core-J-DH; Core-J-O) requires additional steps to re-create their longitudinal weights. For the 2017 Immigrant sample, including the NIMS sample, the longitudinal weight is derived from a preliminary weight that we calibrate to 2019 ACS 1-year PUMS population totals (Section IV.4). For Joint Core families, we adjust the longitudinal weights to account for the joint probability of selection. Therefore, the longitudinal weight also needs to be derived from a calibrated weight as for the 2017 Immigrant sample. It is a "preliminary" longitudinal weight because it includes the families responding in either 2017 or 2019. We then do an additional nonresponse adjustment to get the longitudinal weight for the 2017 Immigrant sample or Joint Core sample respondents (Section IV.5).

The basic steps to produce the longitudinal individual weights in 2019 are as follows:

1. Conduct attrition adjustment for Core Families
2. Integrate the adjustment for the 1997 CDS family reintroduction to the 2019 longitudinal individual weight
3. Prepare the starting weight for calibration for Joint Core and 2017 Immigrant families by sample type
4. Calibration for Joint Core and 2017 Immigrant families
5. Create longitudinal individual weight for persons in Joint Core and 2017 Immigrant families
6. Create the longitudinal family weight

IV. Methods to construct the longitudinal weights

1. Conduct attrition adjustment for Core Families

The methodology for the calculation of PSID longitudinal weights follows a four-year (two-wave) cycle. At the beginning of each cycle, the calculation of weights incorporates an explicit adjustment for panel attrition due to nonresponse that has occurred over the past four years. At the second wave of each four year weight development cycle, a simpler procedure is used to

carry forward the individuals' weights from the previous wave and to update the weights for new births, and for sample panel members who "reappear" and are interviewed again after one or more waves of nonresponse. The detailed description of this approach is provided in Gouskova et al. (2008).

The 2019 weight is an attrition-adjusted weight. The last attrition adjustment of the PSID longitudinal weights was done in 2015. Thus, the construction of the 2019 longitudinal weights starts with the 2015 longitudinal weight as the basis. For those who responded in 2015, the 2019 weights are obtained by multiplying the 2015 longitudinal weight by the attrition adjustment factor. For new sample members (sample newborns and sample persons who moved in), the 2019 weight is calculated as an average of reference person's and spouse/partner's weight in 2015. To account for attrition between the 2015 and 2019 waves among the sample respondents who participated in the 2015 study, an adjustment factor was calculated.

Table 2 reports the results of the multinomial logistic regression estimating probabilities of three possible states in 2019 for those who responded in 2015: 1. response, 2. nonresponse and 3. death. Using the estimates, the probability of response conditional on surviving has been calculated as described in section 4.3 in Gouskova et al. (2008). To decrease the weight variance and reduce reliance on correct model specification a propensity score stratification (Little, 1986) was used, in which ten weighting classes were created based on the decile of the estimated mortality-adjusted response propensities. Then adjustment factors for each class were calculated as the inverse of the average response propensities across both respondents and nonrespondents in each class and the 2015 individual weight was multiplied by these nonresponse adjustment factors to obtain the 2019 individual weight.

2. Integrate the adjustment for the 1997 CDS family reintroduction to the 2019 longitudinal individual weight

A modest upward distortion in the weighted estimates of Black individuals with children has been identified in PSID, beginning in the late 1990s, for selected cohorts. (Freedman and Schoeni, 2016) This distortion is linked to the nearly 600 Black PSID families with children under the age of 13 who were identified to be dropped as part of a larger sample size reduction in

1997, but were retained in the PSID sample (or "reinstated") so that they could be part of the original Child Development Supplement (CDS). Freedman and Schoeni (2016) have provided the adjustment factors needed to post-stratify the assigned PSID weights for Black individuals to the CPS totals by presence of a child under the age of 13 for the years from 1997 to 2015. We have integrated this adjustment in the 2019 longitudinal weights so that the analysts do not need to do additional adjustment in longitudinal analysis that employs these weights. To calculate the adjustment factors for 2019, we use the same method described in Freedman and Schoeni (2016) to reconstruct the adjustment factor by individuals' age, sex, race of reference person and whether living in a family unit (FU) with children <13 years in the CPS and PSID in 1997. For the individuals living in a FU with children <13 years in 1997 and who responded in 1997 and 2019, we reconstructed the adjustment factor so that it would match the CPS estimates in 1997 after the adjustment. The reconstructed adjustment factors for each cell are listed in Table 3a. A comparison of the CPS and PSID weighted percentage of individuals with a child < 13 in the FU is shown in Table 3b. For the born-in and move-in sample persons, we identify the reference person and spouse/partner in the family unit when they were first born-in or move-in in to the PSID panel. If the reference person or spouse/partner of their original family unit is a sample person responding in both 1997 and 2019, we apply the same adjustment factor as for their reference person or spouse/partner to these born-in or move-in sample persons.

The construction of the longitudinal individual weight for Core-S families is complete after applying the adjustment to account for the 1997 CDS family reintroduction. For individuals in the joint families, additional adjustments are needed and described in the sections below.

3. Prepare the starting weight for calibration for Joint Core and 2017 Immigrant families by sample type

Population totals used in this step are based on ACS 2019 1-year PUMS data. In addition, both ACS PUMS data and Integrated Public Use Microdata Series (IPUMS) are needed to define PSID equivalent family units and classify different family types listed in Table 1. See Chang et al. (2019) for additional detail on how we defined PSID equivalent family units to get estimated ACS population totals for each family type.

Joint Core Families

The longitudinal individual weight calculated from two steps described in Section IV.1 and IV.2 and resulting family weights for the Joint Core families (Core-J-D; Core-J-DH; Core-J-O) are their initial longitudinal weights. This initial longitudinal family weight is first integrated with the weight from the joint NIS families based on their age group (both donut and donut hole) to account for the joint inclusion probability. The resulting integrated Core/NIS family weight is their starting weight for the calibration to get their preliminary longitudinal weight. We describe the calculation of the integrated Core/NIS family weight in the next section.

When one member of the couple is native-born or an earlier immigrant and another is a recent immigrant in the donut, and if they co-reside with other family members born in the donut hole years, this family unit would not get a chance to be selected from the NIS-2017 study or NIMS study. However, families in this special subgroup are present in the Core and it is possible to identify them and develop a weight that represents the population of such families. These families are classified as Core-J-O families in Table 1. Their starting weight for the calibration for the preliminary longitudinal weight is directly scaled from the initial longitudinal weight to match the population totals of this type of family. Table 4a shows the integration scaling factor for the Core-J-O families interviewed in 2019.

2017 New-Immigrant Supplement (NIS-2017)

Probability of selection

The 2017 New-Immigrant Supplement was recruited through a sequence of three steps:

1. HRS 2016 Screening,
2. PSID 2016 New Immigrant Screening and
3. PSID 2017 Main Interview.

Therefore, the probability of selection in HRS 2016 Screening sample is the starting point for the weighting for the NIS Donut sample. The HRS probability samples for its new cohort sample updates oversamples addresses in Census Blocks with higher density of Black and Hispanic families so a sampling weight factor is needed to adjust for these unequal probabilities of

selection to avoid potential bias in the resulting estimates. The base probability of selection of each housing unit screened by HRS can be computed as the product of the Primary Sampling Unit (PSU) probability of selection, the probability of the selection for the segment and the probability of selection for the housing unit.

Nonresponse adjustment

Nonresponse can occur at each of the three steps in the NIS recruitment process. In 2019, three nonresponse weighting adjustments—one for each step-- were performed using a response propensity procedure. In this approach, the nonresponse adjustment factor at each step is inversely proportional to estimates of the probability of response at the screening or to the main interview. These estimated probabilities of responding to the survey, also referred to as response propensities, are computed using a logistic regression model of the survey response indicator over a set of covariates available for both respondents and nonrespondents. In order to reduce nonresponse bias while not increasing sampling variance of the survey estimates, the covariates used in this adjustment should be correlated with both the survey response and the study outcomes (Little and Vartivarian, 2003). For this reason, the following PSID key outcomes were selected to assist in this adjustment:

- mean age of reference person
- percentage of reference persons with health insurance
- percentage of spouse/partner with health insurance
- mean family unit income
- mean reference person labor income
- mean family unit wealth
- percentage of black
- percentage of white
- percentage of foreign born - reference person
- percentage of foreign born – spouse/partner
- percentage of families owning a home
- percentage of families with food stamps/SNAP
- mean total food spending

Ideally, we would like to use these survey outcomes as covariates in the response propensity modelling. However, they are only observed for the survey respondents. Instead, each of these survey outcomes was imputed (predicted) for both nonrespondents and respondents using regression models based on covariates available for every sampled housing unit, including

Census variables, paradata and observational data collected by PSID field staff at the time of contact for the screening or main interview.

The probability of responding to HRS screening, PSID screening and PSID main interview were estimated separately using three logistic regression models using as covariates the predicted values of the selected survey outcomes, the stratification variable and a few additional paradata variables. To reduce variation in response propensity weights and lower the reliance on correct model specification of the logistic regression models, ten nonresponse adjustment classes were created based on deciles of the estimated response probabilities (propensity score stratification; Little, 1986). The inverse of the mean response probabilities for the cases in each decile was assigned as the nonresponse adjustment factor for that weighting class. The same process is done separately for the three stages and the final nonresponse adjustment factor is the product of nonresponse adjustment factors for HRS screening, PSID screening and for the PSID main interview. The nonresponse-adjusted base weight for the NIS-2017 was calculated as the product of the base probability of selection weight and the final combined nonresponse adjustment factor. The estimated parameters and standard errors for these logistic models are reported in Tables 5a, 5b, and 5c.

The PSID will follow the families that responded in either 2017 or 2019 in 2021. Therefore, in 2019, we include not only the families that responded in 2019 but also the families that responded in 2017, but not in 2019, for the nonresponse adjustment process. This ensures that there will be a reference longitudinal weight for these families if they are respondents in 2021 regardless of their 2017/2019 response status. In other words, if a family responded in either 2017 or 2019, they are treated as "respondents" for the PSID main interview nonresponse adjustment.

Integration to account for Joint Inclusion Probabilities

For the New Immigrant sample cases that do not have joint probabilities (D-S) under the Core samples, the nonresponse-adjusted base weight is rescaled and serves as the starting weight for the calibration for preliminary cross-sectional individual weight. However, for the NIS Donut sample that could possibly be recruited from the Core sample (D-J), we use the proportion of Joint Donut families observed in each sample source as a proxy for the conditional probability of

being observed from each sample. We include the Joint families that responded in either 2017 or 2019 in this process for the same reason described in the nonresponse adjustment subsection above.

The estimated conditional probability to include the Joint families for each sample source (s) is calculated as:

$$p_{(s)} = \frac{n_{(s)}}{n_{j1D\ \ from\ \ Core} + n_{j1D\ \ from\ \ NIS}}$$

where $n_{(s)}$ is the number of families in the respective sample source, $n_{j1D\ \ from\ \ Core}$ is the number of Core-J-D (see Table 1), and $n_{j1D\ \ from\ \ NIS}$ is the number of D-J cases from the NIS.

The starting point for the integrated Core/NIS family unit weight is the initial longitudinal family weight for Core-J-D, and is the nonresponse-adjusted base weight for D-J. The joint families that responded in 2017 but not 2019 are not part of the attrition adjustment in step 1 so we use their 2017 longitudinal family weight as the start point for integration. An integration scaling factor must be applied to the starting weight in order to have their share of the weights proportional to the estimated inclusion probability of being observed from each sample source. The integration scaling factor was calculated as:

$$Intf_{(s)} = \frac{p_{(s)} \times \sum_m W_m}{\sum_{i(s)} W_{i(s)}}$$

where *m* indexes the families headed by couples in ACS in which only one of the reference person or spouse/partner is a recent immigrant and W_m is the family unit weight for each ACS family. The collection of *i*(*s*) includes the PSID Joint families from the sample source (*s*). $W_{i(s)}$ is the starting weight for the respective family source. Table 4b shows the integration scaling factor for the Joint Donut families that completed interviews in either 2017 or 2019. Please note, there are more families (83 for Core and 163 for NIS) in this table than the counts listed on table 1 (78 for Core and 140 for NIS) because the families that respondent in 2017 but not in 2019 are also included in this process.

The integration scaling factor is applied to the to the initial longitudinal family weight for the 83

families from Core sample, and to the nonresponse adjusted base weight for the 163 families from NIS sample to produce the integrated Core/NIS weight that serves as the starting weight for the calibration for the preliminary longitudinal weight.

2017 New-Immigrant Mulitplicity Supplement (NIMS or NIS-Donut)

The reference person and/or spouse/partner in the 2019 New Immigrant Multiplicity sample was nominated by a reference person or spouse/partner who met the criteria for being a "primary new immigrant" and responded in 2017. Therefore, the probability of inclusion into the PSID panel for the multiplicity sample would depend on the probability of selection into HRS screening project for the nominator of the multiplicity individuals. Because of the relatively small sample size for the multiplicity sample, we made no adjustment for their 'multiplicity' and rescaled the nonresponse-adjusted base weight of the nominator to align to the estimated ACS population totals of the donut hole families. This calibrated multiplicity weight is used as the starting weight for the calibration for the preliminary longitudinal weight if they do not have joint inclusion probabilities (DH-S families).

If a multiplicity family has a head or spouse who is not a post-97 immigrant, it is a joint family (DH-J) and we need to create an integrated Core/NIS weight before performing the calibration for the preliminary longitudinal weight. We use the same method to create the integrated Core/NIS weight as we do for the donut group. Table 4c shows the integration scaling factor for Joint Donut Hole families that completed interviews in either 2017 (Core only) or 2019.

Please note, there are more families (10 for Core) in this table than the counts listed on table 1 (9 for Core) because the families that respondent in 2017 but not in 2019 are also included in this process.

4. Calibration for Joint Core and 2017 Immigrant families

From the above three steps, we have prepared the starting weight at the family level for each of the sample types defined in Table 1. The next step is the calibration of these starting weights for

the Joint Core and 2017 Immigrant families. This calibration of the preliminary longitudinal weight for these families is summarized in this section. Core families that do not have a joint inclusion probability under the 2017 Immigrant sample recruitment are not included in this calibration.

Recall from the previous section that each subgroup of the combined Core/2017 Immigrant sample has a preliminary family weight. By subgroup, these can be summarized as:

Core-J-D (including 2019 NR): Integrated Core/NIS family weight that is derived from attrition adjusted and CDS reinstatement adjusted longitudinal family weight and is integrated with NIS donut sample
Core-J-DH (including 2019 NR): Integrated Core/NIS family weight that is derived from attrition adjusted and CDS reinstatement adjusted longitudinal family weight and is integrated with NIS donut hole sample
Core-J-O: Attrition-adjusted and CDS reinstatement adjusted longitudinal family weight that is rescaled to its share of population total
D-S (including 2019 NR): NIS nonresponse-adjusted base weight that is rescaled to its share of population total
D-J (including 2019 NR): Integrated Core/NIS family weight that is derived from NIS nonresponse adjusted base weight and is integrated with Core-J-D sample
DH-S: The nonresponse-adjusted base weight of the nominator that is rescaled to its share of population totals
DH-J: Integrated Core/NIS family weight that is derived by rescaling the nonresponse-adjusted base weight of the nominator and then integrating with Core-J-DH sample

Similar to the NIS-2017 nonresponse adjustment, if the variables used as calibration dimensions are predictive of the survey outcomes, the calibration adjustment can reduce non-sampling biases (such as noncoverage and nonresponse) and improve the precision of the survey estimates. Therefore, we select the calibration control variables based on their association with the same key outcome variables that we used for the nonresponse adjustment. We regress the 13 selected outcome variables on the following demographic and socio-economic characteristics and their two-way interactions:

- age of individual (0-9/10-19/20-29/30-39/40-49/50-59/60-69/70+)
- sex of individual (Male/Female),
- race of reference person (Black alone or in combination with one or more other races /Non-Black)
- race of reference person (Asian alone or in combination with one or more other races /Non-Asian)
- ethnicity of reference person (Hispanic/Non-Hispanic)
- region (Northeast/Midwest/South/West)
- education of individual (15 years old or younger/11 years or less/12 years/13-15 years/16 years/17 years or more)
- family unit type and employment status (LF=Labor Force)
    - family unit headed by a couple: reference person and spouse/partner in LF
    - family unit headed by a couple: reference person or spouse/partner in LF
    - family unit headed by a couple: Neither reference person or spouse/partner in LF
    - male reference person, no spouse/partner present, in LF
    - female reference person, no spouse/partner present, in LF
    - Non-Couple, reference person not in LF
- family unit size (1/2/3/4 or more)
- presence of children (Yes/No)

Using the results of these 13 regression models, the final set of controls for the weight calibration was chosen to include all the main effects (regardless of their level of explanation on the survey outcomes) and any two-way interactions of these predictors that were significant, at a 10% level, in the regression models for at least seven of the thirteen key survey outcomes. The selected interactions employed in the calibration included:

- ethnicity of reference person x presence of children
- Black x region
- education of individual x region

In order to avoid undue increase in the variability of the weights, the following calibration cells with small sample sizes were collapsed for the calibration procedure:

- education of individual x region: individual education of 16 years and 17 years or more were collapsed

The calibration adjustment was performed using a raking ratio (or iterative proportional fitting) method (Deming and Stephan, 1940) through a SAS macro developed by Battaglia et al (2009). The appropriate starting weight created from step 3 for each family type was used as the input weight for raking procedure. An advantage of this SAS macro is that apart from running the

raking procedure to adjust the weights to enforce the weighted sample distribution to match the population margins in the selected calibration dimensions, it also simultaneously trims the weights according to trimming parameters, in order to mitigate some potential increase of the variance of estimates due to weight variability.

The preliminary longitudinal weight for Joint Core and 2017 Immigrant individuals is derived from the output weights of this calibration adjustment with trimming.

5. Create longitudinal individual weight for Joint Core and 2017 Immigrant families

As we mentioned earlier, the sample persons who responded in 2017 but not in 2019 will be followed in 2021. Therefore, we include them in the earlier steps and the preliminary longitudinal weight created at step 4 for these individuals will serve as their reference longitudinal weight in 2021 if they respond that year.

To calculate the final longitudinal weights for the 2019 respondent data, only the individuals from families responding in 2019 are kept in this step and a nonresponse adjustment is performed to account for the nonresponse from 2017 to 2019. The nonresponse propensity is estimated by a logistic regression where outcome is 0 for nonresponse in 2019 and 1 for response in 2019. The covariates used to predict the response propensity for the usual attrition adjustment are used in this nonresponse adjustment except: 1. being in SEO sample and the interactions with SEO indicators are not included because there is only a handful of SEO cases in the Joint Core and 2017 Immigrant sample; 2. An indicator of 2017 Immigrant sample is added to account for the potential differential response propensity of the 2017 Immigrant sample; and 3. instead of 1^{st} and 99^{th} income percentile, 10^{th} and 90^{th} income percentiles are used in this model due to smaller sample size. Table 6 reports the results of the multinomial logistic regression estimating the probabilities of responding in 2019. To minimize weight variability and reduce the reliance of correct model specification, we again create ten weighting classes and use the inverse of the mean response propensities for the cases in each decile as the nonresponse adjustment factor for that weighting class.

Historically, (as far back as 1968) a linear scale factor was applied to PSID longitudinal weights

so the sum of the longitudinal weights is proportional to but not equal to the same scale as the U.S. population total. Therefore, after applying the nonresponse adjustment factor to the preliminary longitudinal weight, we apply the same scaling factor[6] to create the longitudinal individual weight for the joint Core and 2017 Immigrant sample.

6. Create the longitudinal family weight

Once final individual longitudinal weight is constructed for each sample person interviewed in 2019, the 2019 longitudinal family weight was computed as the average of the positive individual weights for sample persons and the zero-value weights for the nonsample persons in the family. For example, consider a 2019 PSID family that consisted of a young married couple in which the female spouse was a PSID sample person and had an individual longitudinal weight of 60. Her new spouse was PSID "nonsample" and therefore is assigned a "0" value for his longitudinal individual weight. The 2019 family weight for this two-person family is (60+0)/2=30. Note that the family units do change from year to year. See the PSID family level data set documentation and codebooks for more information. See also Duncan and Hill (1985) for a discussion of the issues involved in longitudinal analysis of family units.

IV. Descriptive Statistics for the 2019 PSID Longitudinal Weights

Tables 7 through 11 provide descriptive information on the 2019 PSID longitudinal weights. To enable comparison of the longitudinal weights across years, the same set of descriptors is reported for the longitudinal weights since 2001.

Tables 7 and 8 summarize the total number of cases with positive, zero, and missing values for individual and family weights and the total numbers of sample and non-sample individuals (families with and without sample members). For individual weights, the number of weights with a positive value is equal to the number of sample persons, and the number of the zero-valued

[6] The scaling factor is calculated by dividing the current PSID weight totals by estimated ACS totals for Core-S families. The estimated population totals for the individuals in the Core-S families based on ACS data is 278,268,180. The sum of the PSID 2019 longitudinal individual weights for the individuals in the Core-S families is 478,921. Scaling factor is calculated. As 478,921 /278,268,180 and is about 1/581.

individual weights is the same as the number of non-sample persons (Table 7). As with the 2017 survey, in 2019 all families had at least one sample member (Table 3). As a result, all PSID families in 2019 carry a non-zero, positive longitudinal family weight.

Tables 9 and 10 report summary statistics for the longitudinal individual and family weights. Based on the summary statistics, the distributions of the 2019 longitudinal weights are similar to those in the recent survey waves. Across years, the measures of dispersion indicate that there is an increasing trend in variability of the distribution in the individual and family weights. This steady increase in the variability of the PSID longitudinal weights can be attributed to the periodic nonresponse adjustment (every four years) and to the reweighting that is required to reflect changes in family composition (e.g. new family formations). There is a slightly larger increase in weight variability in 2019 resulting from the addition of 2017 Immigrant sample, especially the "donut hole" group. Table 15A and 15B and Figures 1A and 1B shows the distribution of the longitudinal weights by sample type. The average weights for the 2017 immigrant sample are higher than the one for Core sample.

Table 11 provides a key to the PSID variables names for longitudinal individual and family weight variables.

V. Evaluation of the PSID Longitudinal Weights: Comparisons with the CPS or the ACS

Tables 12 through 14 compare PSID, CPS and ACS weighted estimates for selected demographic statistics based on characteristics including age, gender and race of family reference person. Each table reports the unweighted PSID estimates, PSID estimates weighted (as applicable) by the PSID family or individual longitudinal weight, the CPS weighted estimates and the ACS weighted estimates. For age (Table 12) and race (Table 14), the first panel of the table compares weighted estimates for the family reference person and the second panel of the table provides estimates of mean or percent values for individuals. The statistics in the right most columns of each table are simple ratios of the weighted PSID and CPS estimates and the ratios of the weighted PSID and ACS estimates. These tables are useful for examining three features of

the PSID data: consistency of unweighted and weighted estimates across years, the effect of the longitudinal weights on the distributions of estimates of family and individual population characteristics, and, finally, the consistency of the PSID weighted estimates with those obtained from the CPS data[7] and from the ACS data[8]. The comparison with ACS estimates are added to these tables since 2015 when we changed to use ACS data as the benchmark population totals for the cross-sectional weight.

Comparison of the unweighted and weighted PSID distributions with the CPS or with ACS distributions reveals that in a majority of cases, the weighted estimates are closer to CPS or ACS estimates than are the estimates obtained without weights. This is to be expected since due to the 1968 SEO oversample for which the baseline inclusion probabilities for African American and lower income PSID families and individuals were substantially greater than for other domains of the U.S. household population.

While there are some noticeable difference in the weighted distribution by race, the weighted PSID, CPS and ACS estimates align fairly closely for age and gender. However, caution is advised in placing too much emphasis on minor differences between the PSID and CPS or between the PSID and ACS weighted distribution. Analysts should keep in mind that for any given wave before 2017, the simple comparison of weighted demographic distributions does not explicitly take into account PSID non-coverage of immigrant populations after 1997. Immigrants arriving after 1997 when the immigrant sample was added to the PSID are not fully represented in the PSID. The comparison with ACS estimates, reported since 2017, provides a more comparable comparison by excluding the foreign-born persons entering U.S. after 1997

[7] Some characteristics are not strictly comparable between the two surveys. For example, in the PSID, race is not asked of all individuals while in the CPS data all individuals are asked to provide detailed race information. To calculate proportions of black and non-black individuals in the PSID data, individual race was approximated using the race of the family reference person. Age is top-coded at 85 years old in CPS data while it is not top-coded in PSID data. CPS estimates are calculated based on CPS March supplement data collected in the same year of the PSID data collection.

[8] Some characteristics are not strictly comparable between the two surveys. For example, in the PSID, race is not asked of all individuals while in the ACS data all individuals are asked to provide detailed race information. To calculate proportions of black and non-black individuals in the PSID data, individual race was approximated using the race of the family reference person. Age is top-coded at 99 years old in ACS data while it is not top-coded in PSID data. ACS estimates are calculated based on ACS one-year PUMS data collected in the same year of the PSID data collection.

from the calculation of ACS estimates. Another limitation of this comparison is that the CPS does not cover the institutionalized population while PSID, due to the dynamic nature of the sample, may include institutionalized persons. There are differences in the definitions that PSID, CPS and ACS use to code household composition and disaggregate households into family and non-family units. Finally, the PSID longitudinal weights for families and individual do not include any recent adjustment to external population controls (e.g. 2000 Census, annual CPS or annual ACS population totals).

References

Battaglia, M. P, Hoaglin D. C., and Frankel M. R. (2009). "Practical Considerations in Raking Survey Data." Survey Practice 2 (5). https://doi.org/10.29115/SP-2009-0019.

Birnbaum, Z. W., & Sirken, M. G. (1965). "Design of sample surveys to estimate the prevalence of rare diseases: Three unbiased estimates" (No. 11). US Department of Health, Education, and Welfare, Public Health Service.

Chang, W., Nishimura, R., Heeringa, S., Johnson, D., and Sastry N. (2019). "PSID Cross-Sectional Individual Weights, 1997-2017", Panel Study of Income Dynamics Technical Report. Survey Research Center, University of Michigan, Ann Arbor. Available at: http://psidonline.isr.umich.edu.

Deming, W. E., & Stephan, F. F. (1940). "On a least squares adjustment of a sampled frequency table when the expected marginal totals are known." The Annals of Mathematical Statistics, 11(4), 427-444.

Duncan, G.J. and Hill, M.S. (1985). "Conceptions of longitudinal households: Fertile or futile?", Journal of Social and Economic Measurement, 13, 361-375.

Freedman, V. and Schoeni, R. (2016) Generating Nationally Representative Estimates of Multi-Generational Families and Related Statistics for Blacks in the PSID. https://psidonline.isr.umich.edu/publications/Papers/tsp/2016-01_Est_Multi_Gen_Black_Fam.pdf

Gouskova E., Heeringa S., McGonagle K., Schoeni R., and Stafford F. (2008). "Panel Study of Income Dynamics Revised Longitudinal Weights 2007", PSID Technical Report #08-05, ISR, University of Michigan. http://psidonline.isr.umich.edu/Publications/Papers/tsp/2008-05_PSID_Revised_Longitudinal_Weights_1993-2005%20.pdf

Heeringa, S.G. and Connor J.H. (1998). "Technical documentation for the 1997 PSID Immigrant Supplement". Panel Study of Income Dynamics Technical Report. Survey Research Center, University of Michigan, Ann Arbor.

Heeringa, S.G., Chang, W. and Johnson, D. (2018). "Panel Study of Income Dynamics (PSID) 1968-2015 Cumulative Response Rates for 1968 Sample Persons". Panel Study of Income Dynamics Technical Report. Survey Research Center, University of Michigan, Ann Arbor. Available at: http://psidonline.isr.umich.edu.

Little, R. J. (1986). "Survey nonresponse adjustments for estimates of means". International Statistical Review/Revue Internationale de Statistique, 139-157.

Little, R.J.A., and Vartivarian, S. (2003). "On weighting the rates in nonresponse weights". Statistics in Medicine, 22, 1589-1599.

McGonagle, K. and Schoeni, R. (2006). “The Panel Study of Income Dynamics: Overview and Summary of Scientific Contributions After Nearly 40 Years.” Panel Study of Income Dynamics Technical Paper Series. Available at: http://psidonline.isr.umich.edu/Publications/Papers/tsp/2006-01_PSID_Overview_and_summary_40_years.pdf.

Sirken, M. G. (1970). “Household surveys with multiplicity.” Journal of the American Statistical Association, 65(329), 257-266.

Appendix

Table 1. Number of family units by Joint categories

Category	Type	Abbreviation		Definition	Families		Sample Persons	
		From Core	From 2017 Immigrant		SRC/SEO/1997IMM	2017IMM	SRC/SEO/1997IMM	2017IMM
Core non-immigrant family	Sole Source	Core-S		Neither reference person nor spouse/partner is a recent immigrant	9021	--	17170	--
2017 Immigrant (donut)	Sole Source		D-S	Single reference person or both members of the couple are recent immigrants born before 1960 or after 1971 (donut) and no family members in the donut hole	--	286	--	948
2017 Immigrant (donut hole)	Sole Source		DH-S	Single reference person or both members of the couple are recent immigrants born between 1960 and 1971 (donut hole)	--	9	--	21
Core mixed native-born family and immigrant (donut)	Joint	Core-J-D	D-J	Couple headed family units; one of the couple is native-born or an earlier immigrant; another one of the couple is a recent immigrant in the donut; No family members in the donut hole	78	140	272	535
Core mixed native-born family and immigrant (donut hole)	Joint	Core-J-DH	DH-J	Couple headed family units; one of the couple is native-born or an earlier immigrant; another one of the couple is a recent immigrant in the donut hole	9	15	24	54
Core immigrant family with other family member born in the donut hole years	Joint	Core-J-O		Couple headed family units; one of the couple is native-born or an earlier immigrant; another one of the couple is a recent immigrant in the donut; There are other family members born in the donut hole years	11	Not covered	31	Not covered
Total					9119	450	17497	1558

Table 2. Multinomial Logistic Regression Using 2015 Covariates to Predict Response, Non-Response or Died in 2019

Died before 2019 Interview[1]						Non-Response in 2019 Interview[1]				
2015 Covariates	**Estimate**	**SE**	**Wald ChiSq**	**P Value**		**2015 Covariates**	**Estimate**	**SE**	**Wald ChiSq**	**P Value**
Intercept	**-2.0544**	**0.8509**	**5.8299**	**0.0158**		**Intercept**	**-3.9634**	**0.4427**	**80.1361**	**<.0001**
1st Percentile Income	**-2.9972**	**0.8933**	**11.2564**	**0.0008**		1st Percentile Income	0.2927	0.677	0.187	0.6654
Log Income	**-0.4279**	**0.0697**	**37.6327**	**<.0001**		**Log Income**	**0.1455**	**0.0375**	**15.0162**	**0.0001**
99th Percentile Income	-0.4059	1.0372	0.1532	0.6955		99th Percentile Income	-0.044	0.2303	0.0365	0.8485
Age	0.0237	0.015	2.5023	0.1137		Age	0.00743	0.00507	2.148	0.1428
Age*Age	**0.000515**	**0.000131**	**15.3999**	**<.0001**		**Age*Age**	**-0.00018**	**0.00007**	**6.5594**	**0.0104**
Midwest	0.104	0.2124	0.24	0.6242		Midwest	0.0103	0.1062	0.0093	0.9231
South	-0.0371	0.2102	0.0311	0.8599		South	0.1026	0.1018	1.0165	0.3134
West	0.0481	0.2264	0.0452	0.8317		**West**	**0.2991**	**0.1038**	**8.2999**	**0.004**
Male	**0.3431**	**0.1401**	**5.9993**	**0.0143**		**Male**	**0.1866**	**0.0619**	**9.0922**	**0.0026**
SMSA	-0.1014	0.144	0.4964	0.4811		SMSA	-0.013	0.0661	0.0387	0.844
Might Move	0.2305	0.1716	1.8039	0.1792		Might Move	-0.0253	0.0681	0.1379	0.7104
SEO	**-3.6509**	**1.4146**	**6.6614**	**0.0099**		SEO	-0.107	0.7673	0.0194	0.8891
SEO*1st percentile	2.0212	1.2158	2.7636	0.0964		SEO*1st percentile	0.884	0.8893	0.9881	0.3202
SEO*Log Income	**0.2527**	**0.1129**	**5.0074**	**0.0252**		SEO*Log Income	0.0277	0.0661	0.1763	0.6746
SEO*Age	**0.0532**	**0.0251**	**4.4835**	**0.0342**		SEO*Age	-0.00215	0.0105	0.0418	0.8381
SEO*Age*Age	**-0.00058**	**0.00023**	**6.3129**	**0.012**		SEO*Age*Age	-0.00006	0.000162	0.1262	0.7224
SEO*Midwest	-0.1695	0.5192	0.1066	0.7441		**SEO*Midwest**	**-0.6716**	**0.2559**	**6.8863**	**0.0087**
SEO*South	0.3208	0.4768	0.4526	0.5011		**SEO*South**	**-0.4611**	**0.2253**	**4.19**	**0.0407**
SEO*West	0.2072	0.5706	0.1318	0.7165		SEO*West	-0.1789	0.2711	0.4351	0.5095
SEO*Male	-0.00409	0.2408	0.0003	0.9865		SEO*Male	0.1659	0.1189	1.9473	0.1629
SEO*SMSA	0.4696	0.2618	3.2187	0.0728		SEO*SMSA	-0.1956	0.1297	2.2749	0.1315
SEO*Might Move	-0.2026	0.2723	0.5536	0.4568		**SEO*Might Move**	**0.2623**	**0.1243**	**4.4519**	**0.0349**
Immigrant Sample	-0.4226	0.2499	2.8594	0.0908		**Immigrant Sample**	**0.2696**	**0.086**	**9.8183**	**0.0017**

1. Omitted outcome category is responded in 2019. Bold indicates significant at the alpha=0.05 level

Table 3a. PSID Individual Longitudinal Weight Adjustment Factors for Individuals in Households with Black Reference Persons, 2019

		Male		Female	
Age in 1997	Age in 2019	Adjustment for PSID individual with child < 13 in HH in 1997	Adjustment for PSID individual without child < 13 in HH in 1997	Adjustment for PSID individual with child < 13 in HH in 1997	Adjustment for PSID individual without child < 13 in HH in 1997
--	0-21	NA	NA	NA	NA
0-12	22-34	1.00	1.00	1.00	1.00
13-19	35-41	0.86	1.21	1.03	0.96
20-29	42-51	0.89	1.07	0.96	1.08
30-39	52-61	1.05	0.96	0.97	1.08
40-49	62-71	0.88	1.08	0.94	1.04
50-59	72-81	0.72	1.12	1.38	0.93
60+	82+	1.43	0.97	1.08	0.99

Table 3b. Percentage of Individuals with a Child < 13 Years Old in the Household (CPS) or in the FU (PSID)

		PSID 2019 Individuals also responded in 1997 (97R and 19R) (% with child<13 in 1997)			
		Black Head (total sample persons=3655)			
		Male (n=1530)		Female (n=2125)	
Age in 1997	Age in 2019	**CPS % child <13 in HH**	PSID % of 97R and 19R who are in a FU with child <13 in 1997 (2019 longitudinal weight)	**CPS % child <13 in HH**	PSID % of 97R and 19R who are in a FU with child <13 in 1997 (2019 longitudinal weight)
0-12	22-34	100	100	100	100
13-19	35-41	51	51	60	60
20-29	42-51	35	36	66	66
30-39	52-61	48	48	68	68
40-49	62-71	35	35	37	38
50-59	72-81	22	22	21	21
60+	82+	9	9	13	13
	Total	53	60	59	64

*It is worth noting that the overall 2019 PSID percentage of individuals with a child < 13 in FU in 1997 would not be the same as CPS estimate in 1997 since it is determined by the percentages by age group and these percentages change with time. However, it is reasonable to assume that it would also be the case in the population as these individuals get older. Therefore, it is expected that the 2019 overall percentage of individuals with children <13 in 1997 is different from the 1997 estimate

Table 4a. PSID 2017 Integration scaling factor by sample source – other group

Sample Source	Number of families	Sum of Preliminary Longitudinal Family Unit Weights	ACS total	Integrating scaling factor	Sum of Integrated Core/NIS weights	Percentage of total Integrated Core/NIS weights
Core	11	533	714,617	1341	714,617	100%

Table 4b. PSID 2017 Integration scaling factor by sample source – donut group

Sample Source	Number of families	Sum of Preliminary Longitudinal Family Unit Weights*	Sum of Nonresponse Adjusted Base Weights	ACS total	Integrating scaling factor	Sum of Integrated Core/NIS weights	Percentage of total Integrated Core/NIS weights
Core	83	**2694**		2,870,647	360	968,552	34%
NIS	163		1,367,850		1.39	1,902,095	66%

*The PSID longitudinal family weight incorporates a linear scaling factor from the population total due to historical reasons. Therefore, it is much smaller than the nonresponse adjusted base weights for the NIS-2017 that are computed on a current population scale.

Table 4c. PSID 2017 Integration scaling factor by sample source – donut hole group

Sample Source	Number of families	Sum of Preliminary Longitudinal Family Unit Weights*	Sum of Nonresponse Adjusted Base Weights of the Nominators	ACS total	Integrating scaling factor	Sum of Integrated Core/NIS weights	Percentage of total Integrated Core/NIS weights
Core	10	**316**		1,393,010	1763	557,204	40%
NIMS	15		123,410		6.77	835,806	60%

*The PSID longitudinal family weight incorporates a linear scaling factor from the population total due to historical reasons. Therefore, it is much smaller than the nonresponse adjusted base weights for the NIMS that are computed on a current population scale.

Table 5a. Estimated parameters and standard errors for the HRS Screening Propensity Model

	Variable	Coefficient	Std. Error	Pr(>|z|)	
	(Intercept)	-0.0582	0.0437	0.1826	
Predictive Values for Survey Outcomes (standardized)	Age of reference person	0.5694	0.0142	< 0.001	***
	Reference persons with health insurance	-0.1562	0.0196	< 0.001	***
	Spouse/partner with health insurance	0.0317	0.0157	0.0433	*
	Total family income	-0.1660	0.0286	< 0.001	***
	Reference person labor income	0.0685	0.0261	0.0086	**
	Family wealth	-0.1221	0.0190	< 0.001	***
	Reference person race: White	0.0549	0.0169	< 0.001	**
	Reference person race: Black	0.0356	0.0153	0.0196	*
	Foreign born-reference person	-0.2175	0.0178	< 0.001	***
	Foreign born-spouse/partner	-0.0407	0.0124	0.0010	**
	Home owner	0.0305	0.0202	0.1311	
	With food stamp	0.1211	0.0156	< 0.001	***
	Food Spending	-0.0260	0.0170	0.1274	
	Sampling Stratum 3	-0.2729	0.0762	0.0003	***
	Sampling Stratum 4	-0.1799	0.0552	0.0011	**
	Sampling Stratum 5	0.1814	0.0474	< 0.001	***
	Sampling Stratum 6	0.2982	0.0594	< 0.001	***
	Sampling Stratum 7	0.0478	0.0493	0.3323	
	Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1				
Summary Statistics:					
	Number of observations			34347	
Response profile:					
	Response			17006	
	Nonresponse			17341	
Measures of fit:					
	Pseudo R-squared (Nagelkerke)			0.107	
	p-value for Hosmer and Lemeshow Goodness-of-Fit Test			<0.001	

Table 5b. Estimated parameters and standard errors for the PSID Screening Propensity Model

	Variable	Coefficient	Std. Error	Pr(>|z|)	
	(Intercept)	0.1934	0.2321	0.4047	
Predictive Values for Survey Outcomes (standardized)	Age of reference person	0.2443	0.0795	0.0021	**
	Reference persons with health insurance	0.0433	0.0989	0.6615	
	Spouse/partner with health insurance	-0.1058	0.0830	0.2028	
	Total family income	-0.2613	0.1453	0.0720	.
	Reference person labor income	0.2464	0.1387	0.0757	.
	Family wealth	0.1211	0.1256	0.3349	
	Reference person race: White	0.0498	0.0802	0.5343	
	Reference person race: Black	-0.0677	0.0642	0.2918	
	Foreign born-reference person	-0.0926	0.0817	0.2571	
	Foreign born-spouse/partner	-0.0196	0.0634	0.7577	
	Home owner	0.1113	0.1024	0.2771	
	With food stamp	0.0046	0.0632	0.9423	
	Food Spending	-0.1015	0.0681	0.1364	
	HRS screening in Spanish	0.4538	0.1596	0.0045	**
	Sampling Stratum 3	0.6872	0.3797	0.0703	.
	Sampling Stratum 4	-0.0848	0.3021	0.7788	
	Sampling Stratum 5	0.5383	0.2475	0.0297	*
	Sampling Stratum 6	0.0320	0.3073	0.9169	
	Sampling Stratum 7	0.3277	0.2505	0.1908	
	Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1				
Summary Statistics:					
	Number of observations			1453	
Response profile:					
	Response			920	
	Nonresponse			533	
Measures of fit:					
	Pseudo R-squared (Nagelkerke)			0.062	
	p-value for Hosmer and Lemeshow Goodness-of-Fit Test			0.862	

Table 5c. Estimated parameters and standard errors for the PSID 2017/2019 Survey Interview Propensity Model

	Variable	Coefficient	Std. Error	Pr(>|z|)	
	(Intercept)	1.4162	0.5450	0.0094	**
Predictive Values for Survey Outcomes (standardized)	Age of reference person	-0.1059	0.1510	0.4833	
	Reference persons with health insurance	0.1491	0.1784	0.4034	
	Spouse/partner with health insurance	0.0330	0.1508	0.8267	
	Total family income	-0.3105	0.2844	0.2750	
	Reference person labor income	-0.0015	0.2439	0.9951	
	Family wealth	0.1735	0.2318	0.4540	
	Reference person race: White	0.1198	0.1485	0.4200	
	Reference person race: Black	0.0084	0.1201	0.9442	
	Foreign born-reference person	0.0553	0.1395	0.6919	
	Foreign born-spouse/partner	0.0737	0.1064	0.4884	
	Home owner	-0.0618	0.1817	0.7337	
	With food stamp	0.1138	0.1042	0.2751	
	Food Spending	0.1031	0.1498	0.4914	
	PSID screening in Spanish	-0.4794	0.2707	0.0765	.
	HRS screening in Spanish	0.4251	0.2978	0.1535	
	Reference person is a new immigrant	-0.1522	0.2740	0.5785	
	Couple family	0.2753	0.2293	0.2300	
	Sampling Stratum 3	1.3264	0.8680	0.1265	
	Sampling Stratum 4	-0.2974	0.5745	0.6047	
	Sampling Stratum 5	0.0620	0.4748	0.8961	
	Sampling Stratum 6	-0.2544	0.5912	0.6669	
	Sampling Stratum 7	-0.3267	0.4732	0.4900	
	Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1				
Summary Statistics:					
	Number of observations			641	
Response profile:					
	Response			501	
	Nonresponse			140	
Measures of fit:					
	Pseudo R-squared (Nagelkerke)			0.045	
	p-value for Hosmer and Lemeshow Goodness-of-Fit Test			0.896	

Table 6. Logistic Regression Using 2017 Covariates to Predict Response in 2019

Response in 2019 Interview				
2017 Covariates*	**Estimate**	**SE**	**Wald ChiSq**	**P Value**
Intercept	**2.8126**	**0.974**	**8.3391**	**0.0039**
10th Percentile Income	-0.5462	0.3473	2.4741	0.1157
Log Income	0.0478	0.0782	0.3739	0.5409
90th Percentile Income	-0.3057	0.2843	1.1562	0.2822
Age	**-0.025**	**0.0112**	**4.972**	**0.0258**
Age*Age	**0.000384**	**0.000178**	**4.6764**	**0.0306**
Midwest	0.1832	0.2308	0.63	0.4273
South	0.0232	0.1837	0.016	0.8993
West	-0.0705	0.1925	0.1339	0.7144
Male	0.0841	0.1275	0.4357	0.5092
SMSA	-0.2571	0.1579	2.6507	0.1035
Might Move	0.1783	0.1364	1.708	0.1912
1997 Immigrant Sample	-0.4464	0.4389	1.0344	0.3091
2017 Immigrant Sample	**-1.1843**	**0.3571**	**10.9961**	**0.0009**

*For the cases that did not respond in 2017, 2019 covariates are used as a proxy of 2017 covariates. Age in 2017 was calculated by the age in 2019.

Table 7. PSID Longitudinal Individual Weights, 2001-2019

Year	Total number of individuals	Total number of "sample persons"	Total number of "non-sample persons"	Number of cases with positive individual weight*	Number of cases with zero individual weight*	Number of cases with missing individual weight
2001	21400	15646	5754	15646	5754	0
2003	22290	16012	6278	16012	6278	0
2005	22918	16620	6298	16620	6298	0
2007	23508	16906	6602	16906	6602	0
2009	24385	17471	6814	17471	6814	0
2011	24661	17643	7018	17643	7018	0
2013	24952	17785	7167	17785	7167	0
2015	24637	17505	7132	17505	7132	0
2017*	26445	19258	7187	17643	8802	0
2019	26084	19055	7029	19055	7029	0

* The sample persons from 2017 New Immigrant Supplement did not have longitudinal weights in 2017

Table 8. PSID Longitudinal Family Weights, 2001-2019

Year	Total number of families	Number of families with no "sample person"	Number of families with positive weight	Number of families with zero weight	Number of families with missing weight
2001	7406	211	7195	211	0
2003	7822	257	7565	257	0
2005	8002	0	8002	0	0
2007	8289	0	8289	0	0
2009	8690	0	8690	0	0
2011	8907	0	8907	0	0
2013	9063	0	9063	0	0
2015	9048	0	9048	0	0
2017*	9607	0	9155	0	0
2019	9569	0	9569	0	0

* The sample persons from 2017 New Immigrant Supplement did not have longitudinal weights in 2017. Therefore, 2017 New immigrant families have zero family weight in 2017

Table 9. Summary Statistics for the PSID Longitudinal Individual Weights, 2001-2019 (Sample Persons Only)

Year	N	Mean	Standard Deviation	Min	Max	Coefficient of Variation
2001	15646	25.07	18.97	0.25	167.68	0.76
2003	16012	25.62	19.54	0.25	173.56	0.76
2005	16620	24.81	19.33	0.23	173.56	0.78
2007	16906	25.38	20.09	0.20	181.45	0.79
2009	17471	24.57	19.9	0.23	181.45	0.81
2011	17643	25.65	21.47	0.25	196.44	0.84
2013	17785	24.75	21.11	0.25	196.44	0.85
2015	17505	26.96	23.91	0.28	225.82	0.89
2017*	17643	26.02	23.50	0.20	167.07	0.90
2019	19055	28.88	26.98	0.21	255.16	0.93

*2017 New Immigrant sample persons did not have longitudinal weights in 2017 so they are excluded from this table

Table 10. Summary Statistics for the PSID Longitudinal Family Weights, 2001-2019

Year	N	Mean	Standard Deviation	Min	Max	Coefficient of Variation
2001*	7195	22.03	16.74	0.06	167.68	0.76
2003*	7565	22.06	17.06	0.12	132.64	0.77
2005	8002	21.04	16.82	0.12	136.03	0.8
2007	8289	21.32	17.4	0.1	139.34	0.82
2009	8690	20.66	17.28	0.1	139.34	0.84
2011	9807	21.71	18.75	0.12	150.89	0.87
2013	9063	20.85	18.44	0.08	150.89	0.89
2015	9048	22.80	21.06	0.10	156.12	0.92
2017**	9155	22.11	20.66	0.08	142.78	0.93
2019	9569	24.73	24.06	0.06	154.57	0.97

*2001 and 2003 Based on families with positive weights only

**2017 New Immigrant families did not have longitudinal weights in 2017 so they are excluded from this table

Table 11. Names of the PSID Longitudinal Weight Variables, 1993-2019

Year	Core Longitudinal Weight	
	Individual	**Family**
1993	ER30864	V23361
1994	ER33119	ER4160
1995	ER33275	ER7000
1996	ER33318	ER9251
	Core/Immigrant Longitudinal Weight	
	Individual	**Family**
1997	ER33430	ER12084
1999	ER33546	ER16518
2001	ER33637	ER20394
2003	ER33740	ER24179
2005	ER33848	ER28078
2007	ER33950	ER41069
2009	ER34045	ER47014
2011	ER34154	ER52436
2013	ER34268	ER58257
2015	ER34413	ER65492
2017	ER34650	ER71570
	Core/1997 Immigrant/2017 Immigrant Longitudinal Weight	
	Individual	**Family**
2019	ER34863	ER77631

Table 12. Comparison of PSID,CPS, and ACS Weighted Estimates of Mean and Median Age, 2001-2019

A. Family Level Data (age of reference person)

	PSID** unweighted		PSID** weighted***		CPS**** weighted		ACS***** weighted		PSID/CPS Ratio		PSID/ACS Ratio	
	Mean	Median	Mean	Median	Mean	Median	Mean	Median	Mean	Median	Mean	Median
Year	Years	Years	Years	Years	Years	Years	Years	Years	Years	Years	Years	Years
	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]	[3]/[5]	[4]/[6]	[3]/[7]	[4]/[8]
2001	44.91	43	49.39	47	48.72	46	Not Used		1.01	1.02	Not Used	
2003	44.98	43	49.6	48	48.69	47			1.02	1.02		
2005	45.08	44	49.96	48	49.04	47			1.02	1.02		
2007	45.04	44	50.13	49	49.30	48			1.02	1.02		
2009	45.79	44	49.82	49	47.60	47			1.05	1.04		
2011	45.21	43	50.60	50	48.11	47			1.05	1.06		
2013	45.68	43	51.21	51	48.56	48			1.05	1.06		
2015	45.65	43	52.02	52	Not Used		51.86	52	Not Used		1.00	1.00
2017*	46.20	43	53.14	54			52.96	53			1.00	1.02
2019	46.44	43	52.36	53			51.93	52			1.01	1.02

*2017 New Immigrant families did not have longitudinal weights in 2017 so they are excluded from the 2017 estimates

** Missing value of age of reference person in PSID data was imputed.

*** PSID weighted estimates were weighted by PSID longitudinal family weight

**** Age in CPS data is top-coded at 85 years old.

***** Age in ACS data is top-coded at 99 years old. The families with reference person who was foreign-born and entered the U.S. after 1997 were excluded from the data used for calculating 2017 ACS estimates

‡Prior to 2015, we used CPS estimates as the population totals for post-stratification so CPS data was used as the benchmark for this table. We started to use ACS estimates as the population totals for calibration since 2015 and thus changed the benchmark for the comparison.

B. Individual Level Data

	PSID** unweighted***		PSID** weighted****		CPS***** weighted		ACS****** weighted		PSID/CPS Ratio		PSID/ACS Ratio	
	Mean	Median	Mean	Median	Mean	Median	Mean	Median	Mean	Median	Mean	Median
Year	Years	Years	Years	Years	Years	Years	Years	Years	Years	Years	Years	Years
	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]	[3]/[5]	[4]/[6]	[3]/[7]	[4]/[8]
2001	30.86	29	36.30	36	35.65	35	Not Used		1.02	1.03	Not used	
2003	31.25	29	36.53	36	35.82	35			1.02	1.03		
2005	31.41	29	36.93	36	36.17	36			1.02	1.00		
2007	31.61	29	37.35	37	36.44	36			1.02	1.03		
2009	32.30	29	37.90	37	36.80	36			1.03	1.03		
2011	31.95	29	38.75	38	37.00	36			1.05	1.06		
2013	32.91	30	39.27	38	37.64	37			1.04	1.03		
2015	32.55	30	40.18	39	Not Used		38.31	37	Not Used		1.05	1.05
2017*	32.93	31	40.75	39			38.88	38			1.05	1.03
2019	33.04	31	39.69	38			39.10	38			1.02	1.00

*2017 New Immigrant families did not have longitudinal weights in 2017 so they are excluded from the 2017 estimates

** Missing value of age in PSID data was imputed

*** Unweighted individual level PSID estimates were calculated based on sample and non-sample individuals

**** PSID weighted estimates were weighted by PSID longitudinal individual weight

***** Age in CPS data is top-coded at 85 years old

****** Age in ACS data is top-coded at 99 years old. The individuals who were foreign-born and entered the U.S. after 1997 were excluded from the data used for calculating 2017 ACS estimates

‡Prior to 2015, we used CPS estimates as the population totals for post-stratification so CPS data was used as the benchmark for this table. We started to use ACS estimates as the population totals for calibration since 2015 and thus changed the benchmark for the comparison.

Table 13. Comparison of PSID,CPS and ACS Weighted Estimates of % Population by Gender, 2001-2019

	PSID unweighted**		PSID weighted***		CPS weighted		ACS**** weighted		PSID/CPS Ratio		PSID/ACS Ratio	
Year	**Male**	**Female**	**Male**	**Female**	**Male**	**Female**	**Male**	**Female**	**Male**	**Female**	**Male**	**Female**
	[1]	**[2]**	**[3]**	**[4]**	**[5]**	**[6]**	**[7]**	**[8]**	**[3]/[5]**	**[4]/[6]**	**[3]/[7]**	**[4]/[8]**
2001	47.93	52.07	48.08	51.92	48.86	51.14			0.98	1.02		
2003	47.98	52.02	48.17	51.83	48.92	51.08			0.98	1.01		
2005	47.88	52.12	48.23	51.77	49.03	50.97			0.98	1.02		
2007	47.88	52.12	48.58	51.42	49.08	50.92			0.99	1.01		
2009	47.48	52.52	48.40	51.60	49.10	50.90			0.99	1.01		
2011	47.87	52.13	48.74	51.26	49.21	50.79			0.99	1.01		
2013	47.69	52.31	48.83	51.17	48.96	51.04			1.00	1.00		
2015	47.53	52.47	48.70	51.30			49.20	50.80			0.99	1.01
2017*	47.69	52.31	48.62	51.38			49.28	50.72			0.99	1.01
2019	47.72	52.28	49.19	50.81			49.23	50.77			1.00	1.00

*2017 New Immigrant families did not have longitudinal weights in 2017 so they are excluded from the 2017 estimates
** Unweighted individual level PSID estimates were calculated based on sample and non-sample individuals
*** PSID weighted estimates were weighted by PSID longitudinal individual weight
**** The individuals who were foreign-born and entered the U.S. after 1997 were excluded from the data used for calculating 2017 ACS estimates
ǂPrior to 2015, we used CPS estimates as the population totals for post-stratification so CPS data was used as the benchmark for this table. We started to use ACS estimates as the population totals for calibration since 2015 and thus changed the benchmark for the comparison.

Table 14. Comparison of PSID, CPS, and ACS Weighted Estimates of % Population by Race, 2001-2019

A. Family Level Data (age of reference person)

	PSID** unweighted		PSID** weighted***		CPS**** weighted		ACS***** weighted		PSID/CPS Ratio		PSID/ACS Ratio	
Year	**Non-black**	**Black**	**Non-black**	**Black**	**Non-black**	**Black**	**Non-black**	**Black**	**Non-black**	**Black**	**Non-black**	**Black**
	[1]	**[2]**	**[3]**	**[4]**	**[5]**	**[6]**	**[7]**	**[8]**	**[3]/[5]**	**[4]/[6]**	**[3]/[7]**	**[4]/[8]**
2001	69.60	30.40	87.40	12.60	87.80	12.20	Not Used		1	1.03	Not Used	
2003	68.40	31.60	87.20	12.80	87.90	12.10			0.99	1.06		
2005	66.70	33.30	86.10	13.90	87.80	12.20			0.98	1.14		
2007	65.70	34.30	85.90	14.10	87.60	12.40			0.98	1.14		
2009	64.60	35.40	84.40	15.60	87.50	12.50			0.96	1.25		
2011	62.93	37.07	85.18	14.82	87.35	12.65			0.98	1.17		
2013	61.84	38.16	83.54	16.46	86.97	13.03			0.96	1.26		
2015	61.20	38.80	83.73	16.27	Not Used		87.09	12.91	Not Used		0.96	1.26
2017*	60.62	39.38	83.64	16.36			86.85	13.15			0.96	1.24
2019	60.84	39.16	83.45	16.55			86.53	13.47			0.96	1.23

*2017 New Immigrant families did not have longitudinal weights in 2017 so they are excluded from the 2017 estimates

** Black was defined based on the race first mention of reference person for PSID estimates. Missing value of race first mention of reference person in PSID data was imputed.

*** PSID weighted estimates were weighted by PSID longitudinal family weight

**** Black was defined by black alone or in combination with one or more other races for CPS

***** Black was defined by black alone or in combination with one or more other races for ACS estimates. The families with reference person who was foreign-born and entered the U.S. after 1997 were excluded from the data used for calculating 2017 ACS estimates

‡Prior to 2015, we used CPS estimates as the population totals for post-stratification so CPS data was used as the benchmark for this table. We started to use ACS estimates as the population totals for calibration since 2015 and thus changed the benchmark for the comparison.

B. Individual Level Data

	PSID** unweighted***		PSID** weighted****		CPS***** weighted		ACS****** weighted		PSID/CPS Ratio		PSID/ACS Ratio	
Year	**Non-black**	**Black**	**Non-black**	**Black**	**Non-black**	**Black**	**Non-black**	**Black**	**Non-black**	**Black**	**Non-black**	**Black**
	[1]	**[2]**	**[3]**	**[4]**	**[5]**	**[6]**	**[7]**	**[8]**	**[3]/[5]**	**[4]/[6]**	**[3]/[7]**	**[4]/[8]**
2001	67.00	33.00	86.90	13.10	87.30	12.70	Not Used		1.00	1.03	Not Used	
2003	66.10	33.90	86.60	13.40	87.50	12.50			0.99	1.07		
2005	64.60	35.40	86.00	14.00	87.40	12.60			0.98	1.11		
2007	64.20	35.80	85.90	14.10	87.40	12.60			0.98	1.12		
2009	63.70	36.30	85.20	14.80	86.70	13.30			0.98	1.11		
2011	63.35	36.65	84.19	15.81	86.43	13.57			0.97	1.17		
2013	61.88	38.12	84.79	15.21	85.95	14.05			0.99	1.08		
2015	61.51	38.49	84.85	15.15	Not Used		86.10	13.90	Not Used		0.99	1.09
2017*	61.19	38.81	84.42	15.58			85.70	14.30			0.99	1.09
2019	62.03	37.97	84.40	15.60			85.75	14.25			0.98	1.09

*2017 New Immigrant families did not have longitudinal weights in 2017 so they are excluded from the 2017 estimates

** Individual race in PSID data was approximated using the race of the family reference person. Black was defined based on the race first mention of reference person for PSID estimates. Missing value of race first mention in PSID data was imputed.

*** Unweighted individual level PSID estimates were calculated based on sample and non-sample individuals.

****PSID weighted estimates were weighted by PSID longitudinal individual weight

*****Black was defined by black alone or in combination with one or more other races for CPS

****** Black was defined by black alone or in combination with one or more other races for ACS estimates. The individuals who were foreign-born and entered the U.S. after 1997 were excluded from the data used for calculating 2017 ACS estimates

‡Prior to 2015, we used CPS estimates as the population totals for post-stratification so CPS data was used as the benchmark for this table. We started to use ACS estimates as the population totals for calibration since 2015 and thus changed the benchmark for the comparison.

Table 15A. Distribution of Longitudinal Individual Weights by Sample Type

Sample Type	N	Mean	Std Dev	Minimum	Maximum
Core	17497	28.15	25.99	0.21	255.16
2017 Immigrant - Donut	1483	34.25	32.43	2.80	159.92
2017 Immigrant - Donut Hole	75	92.35	42.60	10.73	164.47

Table 15B. Distribution of Longitudinal Family Weights by Sample Type

Sample Type	N	Mean	Std Dev	Minimum	Maximum
Core	9119	23.90	22.81	0.06	154.57
2017 Immigrant - Donut	426	38.32	35.69	1.61	154.57
2017 Immigrant - Donut Hole	24	97.81	42.31	22.67	154.57

Figure 1A. Distribution of Longitudinal Individual Weights by Sample Type

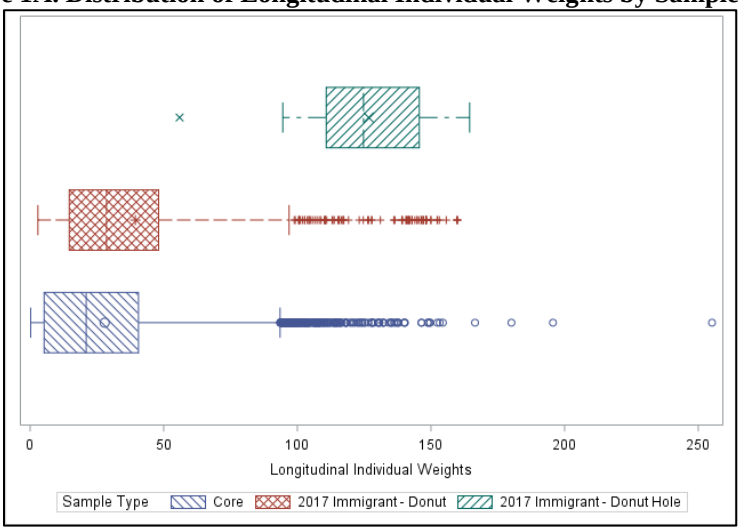

Figure 1B. Distribution of Longitudinal Family Weights by Sample Type