Panel Study of Income Dynamics (PSID) 1968-2015 Cumulative Response Rates for 1968 Sample Persons

Steven Heeringa, Wen Chang, David Johnson
SRC Statistical Design Group and
Panel Study of Income Dynamics
Survey Research Center
University of Michigan

August 14, 2018

Summary

The purpose of this technical report is to describe the statistical methodology and results for an investigation of cumulative response rates (CRRs) for the original sample of persons interviewed in the initial 1968 wave of the Panel Study of Income Dynamics (PSID).

Introduction

The PSID is based on an innovative design for a dynamic longitudinal sampling of U.S. families and individuals (Hill, 1992). The Panel Study of Income Dynamics (PSID) began data collection on U.S. households and individuals in 1968. The original 1968 PSID sample was the combination of two national probability samples. The first of the two PSID sample components was an equal probability sample of approximately 3000 U.S. households. This nationally-representative "SRC sample" was a multi-stage sample selected from the Survey Research Center's 1960 master area probability frame. The second component of the original 1968 PSID sample was a supplemental sample of approximately 2000 low income families subselected from the full sample for the Survey of Economic Opportunity (SEO) that was conducted by the U.S. Census Bureau on behalf of the Office of Economic Opportunity (OEO). Following the initial 1968 data collection on the combined SRC and SEO samples of families and individuals, SRC survey statisticians developed analysis weights (both family and individual) that reflected: 1) the joint probability of selection in the SRC national sample and SEO subsample; 2) differential nonresponse for sample subgroups; and 3) potential noncoverage due to administrative procedures involved in fielding the SEO sample.

Beginning the following year, regardless of current residence, each sample person who was a member of an original 1968 PSID family was recontacted and asked to provide data describing their individual work-related and financial characteristics. Data was also collected for their current family unit. The "dynamic sampling" of families and individuals that has characterized the PSID's longitudinal design since its inception in 1968 was certainly unique at the time. Over time, children born to sample persons or their spouses/partners were added to the PSID panel, dynamically extending PSID's representation of post 1968 birth cohorts in the U.S. population. Over the same period, the PSID panel also suffered attrition of 1968 sample persons due to mortality and follow-up nonresponse. Following each wave of PSID data collection, analysis weights were computed for children born to a 1968 sample person. In 1969 and every five years until 1993 (1974, 1979, 1984, 1989, 1993), the core longitudinal individuals weights for 1968 sample persons and their children were updated to compensate for panel attrition due to nonresponse. At each successive wave, family weights were also updated to reflect changes in composition (primarily due to marriage) that affected the family's joint inclusion probability for the panel.

This basic post-survey weighting approach to maintaining the PSID's "representativeness" of the U.S. population underwent very little change for almost 25 years. However, beginning in the early 1990s there were several important changes in the PSID sample. Concern over cumulative attrition in the PSID panel led to a major nonrespondent recontact effort in 1993 and 1994. That two-year follow-up effort was successful in reintroducing several thousand former nonrespondents to the PSID panel—many of whom had not been interviewed for over 5 years. For the 1994-1996 period immediately following this PSID nonresponse follow-up initiative, PSID longitudinal weight computations did include adjustments for sample persons who were "restored" to the panel. However, for this three year period, the longitudinal weight calculations did not include an explicit adjustment for attrition due to nonresponse (Gouskova et al. 2007). For the current calculations, the decision to not adjust the weights for nonresponse in 1994,1995,1996 may introduce a small positive bias to later waves' estimates of cumulative response rates.

Three years later, in1997, the PSID sample underwent even more significant changes. A national probability sample of individuals and families representing post-1968 immigrants to the U.S. was added to the panel. In addition, the overall size of the permanent longitudinal panel was reduced through a probability subsampling of original 1968 "family trees"—reducing the size of the original panel to roughly 2/3rds the number of 1968 families interviewed prior to the 1997 panel reduction. Following the 1997 panel "reduction", the core longitudinal individual weights for 1968 sample persons were revised to reflect the subsampling and beginning in 2003 periodic adjustment of the individual weights to account for nonresponse was again performed every four years (2003, 2007, 2011 and 2015; Gouskova et al. 2007; Berglund et al. 2017).

3. Statistical Methods to Estimate PSID Cumulative Response Rates (CRRs)

3.A Estimators

The objective here is to develop and report a measure of the cumulative response rate for the cohort of PSID sample persons interviewed in the first wave (1968) of panel data collection. It should be noted that the derived response rate statistics reported here do not take into account nonresponse or noncoverage for the original SRC and SEO samples from which the original panel of 1968 families and sample persons were derived. Inclusion of 1968 unit nonresponse by sample households in the total sample cumulative response rate estimate would be possible for the SRC National Sample component. However, due to the confidential process by which the U.S. Census Bureau selected, consented and transferred the SEO sample of families component to SRC for the 1968 PSID contact, required information needed to compute unit nonresponse rates for that sample component is not available.

As a statistical measure of panel retention (or the complement, panel attrition), estimates of CRRs can take several forms. The simplest of these is the ratio in which the numerator is the

unweighted count of 1968 sample persons responding at wave t and the denominator is the unweighted count of 1968 sample persons alive at time t:

$$UCRR(t) = \frac{r_t}{(n_{1968} - mx_t)} \qquad (1)$$

where:

r_t = the unweighted count of 1968 sample persons responding at time t;

n_{1968} = the unweighted count of 1968 sample persons; and

mx_t = the time t cumulative total of deaths of 1968 sample persons.

Despite its simplicity, there are several reasons why this statistic is not a completely accurate or satisfactory metric for measuring the potential impact that cumulative panel attrition may have on the population representation of the retained sample of panel members who are interviewed at time t. The first of these is that in terms of population representation, not all 1968 sample persons were "created equal". Under the 1968 dual-frame sample recruitment for PSID, low income, predominantly African-American families were included with higher probabilities (and therefore lower sampling weights) than other U.S. families. As shown in Figure 1, the disproportionate sampling that was used to generate the 1968 baseline sample resulted in a high degree of variability in the 1968 individual weights.

Figure 1. Frequency of 1968 Individual Weights
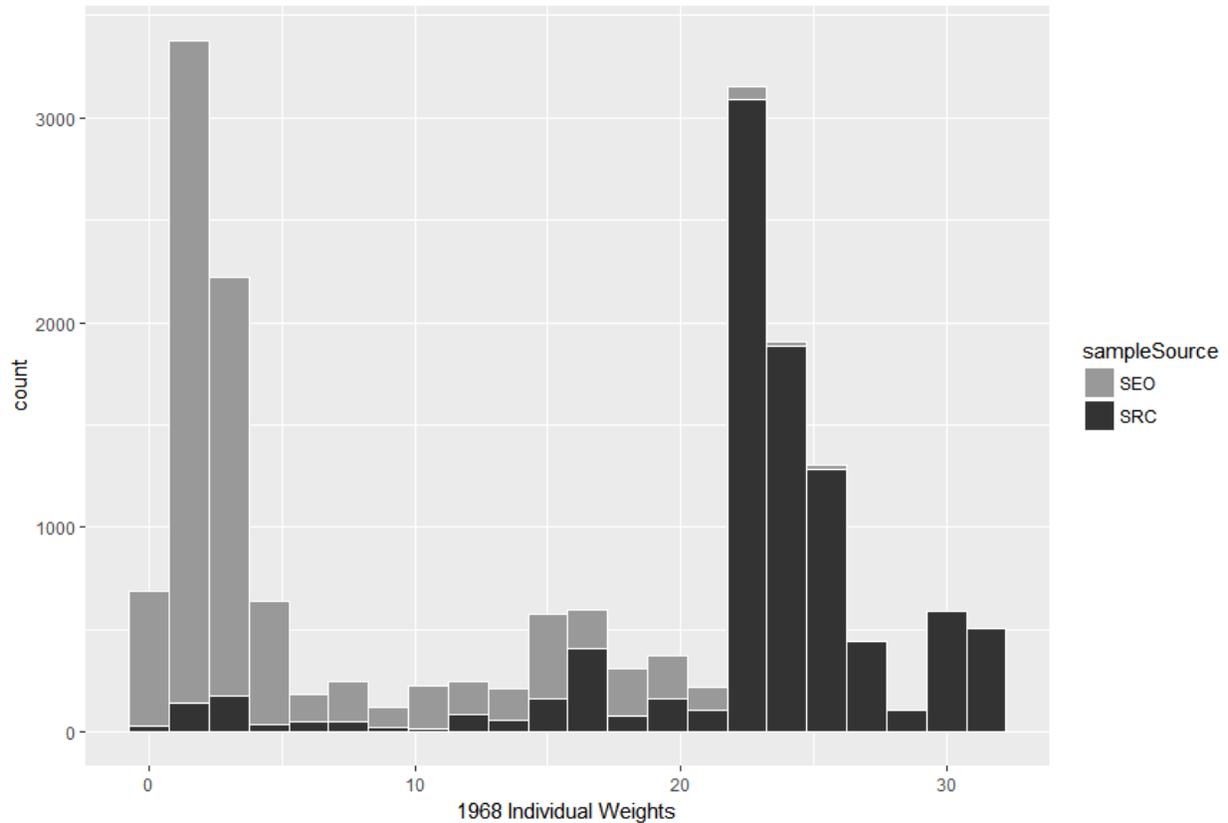
Second, in practice, the simple unweighted ratio, UCRR(t), is likely to be biased downward due to incomplete recording of deaths for original 1968 sample persons. That is, in practice, mx_t is measured by cumulative deaths known to PSID.   The PSID rigorously tracks and records deaths to panel members, and in 2006 PSID conducted an extensive effort to determine vital status of attritors (McGonagle, Smith, and Schoeni, 2008). Regardless, vital status is not 100% complete and becomes more difficult with the passage of time since attrition.

An alternative measure of panel retention/attrition is a statistic that we will label the weighted estimate of the panel cumulative response rate. The ratio form of this statistic is:

$$WCRR(t) = \frac{\hat{N}_{t,r}}{\hat{S}_{1968}} = \frac{\sum_i W_{i,1968} \cdot I_{i \in r(t)}}{\sum_i W_{i,t} \cdot I_{i \in r(t)}}$$ (2)

where :

$\hat{N}_{t,r}$ = the weighted estimate of the count of the 1968 study population
   "represented" by 1968 sample persons responding at time t;

$\hat{S}_{t,r}$ = the weighted estimate of the count of the 1968 study population
   members alive at time t;

i = indexes the individual 1968 sample persons (n=18,233)

$I_{i \in r(t)}$ = indicator that 1968 sample person i is a PSID respondent at time t,
   = 1 if respondent at time t, 0 otherwise;

$W_{i,1968}$ = 1968 base sampling weight for 1968 sample person i;

$W_{i,t}$ = time t PSID longitudinal individual weight for 1968 sample person i,
   includes mortality and nonresponse adustments at and prior to time t.

This weighted form of the CRR estimator is sometime referred to as the "census response rate". If instead of starting with a probability sample of persons the 1968 PSID Wave 1 interview had been administered to a complete census of the population (i.e. $W_{i,1968}=1.0$ for all i=1,…,N) then this statistic would estimate the proportion of the *surviving* members of the original 1968 census who were interviewed at time t. Since the 1968 PSID was in fact based on a sample (and not a census) the value of $W_{i,1968}$ can be interpreted as the number of U.S. population members like themselves that a 1968 sample person represented in 1968. Taking this simple interpretation of weighting, summing over the respondent values of $W_{i,1968}$, the numerator of the ratio is therefore an estimate of the count of persons in the study population represented by the living 1968 sample persons responding at time t. The numerator implicitly accounts for mortality (the respondent is alive at time t) but does not include any adjustment for the surviving nonrespondents. In contrast, the weights in the denominator, $W_{i,t}$, are the time t analysis weights for the 1968 sample persons and these weights do adjust for both panel losses due to both mortality and nonresponse. The summation in the denominator therefore estimates the count of persons in the 1968 study population who are still alive at time t.

The Results section of this report will also present weighted estimates of the CRR for major subpopulations of the 1968 PSID sample. For characteristics that are not time dependent, CRR

estimates for subpopulations are easily generated by including a simple indicator of subpopulation membership in the ratio estimator:

$$WCRR(t, sub) = \frac{\hat{N}_{r(t),sub}}{\hat{S}_{1968,sub}} = \frac{\sum_i W_{i,1968} \cdot I_{i \in r(t),sub}}{\sum_i W_{i,t} \cdot I_{i \in r(t),sub}} \qquad (3)$$

$where:$

$I_{i \in r(t),sub}$ = indicator that 1968 sample person i is a PSID respondent at time t and

belongs to the subpopulation of interest,

= 1 if respondent at time t and a subpopulation member, 0 otherwise.

The 1968 PSID sample of families and individuals is based on a complex, probability sample design. The weighted estimates, WCRR, of the cumulative response rate for the total sample and its subpopulations take the form of a combined ratio estimator. Correct standard errors and confidence intervals for the estimates can be computed using one of several methods for complex samples including a Taylor Series Linearization (TSL) Method, Jackknife Repeated Replication (JRR), Balanced Repeated Replication (BRR) and the Rao-Wu Rescaling Bootstrap (Heeringa et al., 2017). Estimates published in this report were computed in SAS V 9.4 using the TSL method.

3.B Special Procedures for Post-1996 Weighted Estimation of WCRRs

As described in Section 2, 1997 was a year of major changes to the PSID panel and data collection. The 1997 change that is most relevant to the topic of cumulative response rates is the removal of a probability subsample of 1968 family trees from the Core panel. The Core panel transition from 1996 to 1997 was further complicated by a subsequent decision by the PSID Study team to not remove some of the 1997 families that were initially selected to be dropped, specifically, families that included children who were eligible for the PSID's 1997 Child Development Supplement (CDS)—a set African-American families whose PSID family tree had "roots" in the 1968 SEO sample. To account for the associated changes in the base sample inclusion probabilities for all Core sample family trees and individuals, expression (2) for computing the WCRR in 1997 and later waves must incorporate an adjustment to the values of the $W_{i,1968}$ in the numerator and the $W_{i,t}$ in the denominator of the estimator. (See expression 4).

$$WCRR(t > 1996) = \frac{\sum_i W^*_{i,1968} \cdot I_{i \in r(t)}}{\sum_i W^*_{i,t} \cdot I_{i \in r(t)}} \qquad (4)$$

where :

$W^*_{i,1968} = W_{i,1968} \cdot k1_{h,i} = 1968$ base sampling weight for 1968 sample person i

adjusted for 1997 Core Panel reduction and CDS

family "restoration" to the Panel.

$W^*_{i,t} = W_{i,t} \cdot k2_{h,i} = $ time t PSID longitudinal individual weight for 1968 sample person i

adjusted for 1997 Core Panel reduction and CDS

family "restoration" to the Panel.

$k1_{h,i}$ and $k2_{h,i}$ = weight adjustment factors that account for 1997 stratified

sampling of 1968 PSID family trees and CDS family restoration.

The following approach was used to derive the adjustment needed to bridge the PSID Core sample transitions that occurred between 1996 and 1997. Each 1968 PSID sample person who was a respondent in 1996 or 1997 was assigned to one of the three sample strata that determined the probability that their 1968 family tree would be retained for a Core panel interview in 1997.

Table 1: 1997 PSID Core Sample Reduction Strata

Strata for 1997 PSID Core Reduction	Description
1	SRC National Sample, 1968 address not in SEO Low Income Domain
2	SRC National Sample or SEO Sample, 1968 address in SEO Low Income Domain, non-Black Family Head
3	SRC National Sample or SEO Sample, 1968 address in SEO Low Income Domain, Black Family Head.

Nested within 1997 PSID Core Panel Reduction strata 1 and 2, 1996 and 1997 individual respondents were further assigned to one of 16 demographic groups defined by the cross-classification of age (28-39,40-49,50-59,60+), gender (male, female), and race of family head (Black, non-Black) resulting in post-strata for a 1996 to 1997 "bridging calibration" of the PSID individual weights. The post-strata for stratum 3 were defined by cross-classification of age, gender, race of family head, and CDS subgroup indicator (with age 12 or younger children in the family units in 1997, without age 12 or younger children in the family units in 1997). The bridging adjustments, k1 and k2 in expression 4 were then computed as:

$$k1_h = \frac{\sum\limits_{i \in (1996,h)} W_{i,1968} \cdot I_{1i}}{\sum\limits_{i \in (1997,h)} W_{i,1968}}; \quad k2_h = \frac{\sum\limits_{i \in (1996,h)} W_{i,1996} \cdot I_{2i}}{\sum\limits_{i \in (1997,h)} W_{i,1997}} \qquad (5)$$

where :

   h denotes the poststratum defined by 1968 Core sample reduction strata
      and demographic categories of 1968 sample persons.

   $I_{1i} = 1$ if individual "i" responded in 1996 and either responded in 1997
      or was dropped from the study because of 1997 reduction, 0 otherwise.

   $I_{2i} = 1$ if the individual responded in 1996 and was not deceased in 1997, 0 otherwise.

4. Results

Using the appropriate choice of expressions (2)-(4), Table 1 presents the weighted (WCRR) and unweighted (UCRR) estimates of the cumulative response rate for 1968 PSID sample persons. At the conclusion of the 2015 wave of data collection, the estimate of the WCRR for surviving members of the full 1968 baseline sample is 40.4%. This pooled estimate of the WCRR blends the population-weighted representation of the PSID 1968 SRC National Sample of individuals (WCRR=41.0%) with the low-income SEO component which experienced a higher attrition rate (WCRR=33.4%). Year-by-year until 1997 when the unweighted rates are no longer applicable, the computed values of the UCRR are consistently lower than the weighted WCRR estimates of cumulative response rates.

There are two explanations for the difference in the two rates and the increased disparity over time. The first is that, with time, attrition losses due to nonresponse and noncontact were greater for the lower income SEO sample individuals who had significantly higher 1968 PSID Core inclusion probabilities and therefore smaller population weights. The weighted estimates, WCRR, therefore down-weight the experience in the oversampled segment of SEO Core sample component. A second reason that UCRR values may trend below the WCRR estimates lies in the treatment of mortality. The UCRR values incorporate information only for known deaths for the 1968 sample persons—with the passing of time, deaths occurring to panel members lost to nonresponse in the early years of the study may not be easily detected and recorded as a final PSID sample disposition. Although standardized mortality corrections have not been consistently applied since in 1968, they were used to develop PSID Core longitudinal weights from 1990 to 1996. It was also possible to poststratify the PSID core longitudinal weights to U.S. population totals in 1997 when the PSID Immigrant supplement was introduced to the panel. This one time poststratification control for the PSID Core longitudinal weights also served to recalibrate the PSID weights for the population mortality that occurred between 1968 and 1997.

Table 2a and Table 2b provides more detail on the cumulative response rates for PSID, looking at the estimates of WCRR for major demographic subpopulations of the 1968 sample persons. The times series of estimates of WCRR suggest that cumulative response rates for male and female members of the 1968 Core panel did not differ significantly over the 1968-2015 follow-up period. For individuals who were less than 65 years old in 1968, the 2015 estimates of the WCRR for surviving members of the 1968 age 18-39 cohort are slightly lower than those for the 1968 sample persons who were children or teens (age 0-17) or middle age (40-64) in 1968. Likewise, WCRR estimates for the cohort of 1968 sample persons who were age 65+ in 1968 follow a lower trend line than those for the younger age groups. The observed pattern may be due in part to the previously mentioned challenge of confirming deaths among nonrespondents in this higher mortality age group. As noted, failure to account for deaths in the 1968 baseline cohort of individuals will lead to positive bias in the denominator of the WCRR estimator and therefore negative bias in the WCRR ratio estimates.

WCRR estimates for the 1968 sample persons in African-American families ($WCRR_{2015}$=32.9%) also trend lower than those of other 1968 families ($WCRR_{2015}$=41.5%). The underlying cause of the race/ethnicity differential in WCRR is unclear but the observed difference is strongly associated with the pattern for the SEO sample in which lower income African-American families are disproportionately represented. Looking at the trend over time, the WCRRs for these two race/ethnicity subpopulations began to diverge in the late 1970s and the majority of the ultimate Black/Non-black differential (~10%-13%) in estimated WCRR is already present by the 1984 and 1989 data collection waves.

5. Summary

This brief PSID technical report has summarized the statistical methodology and results (through 2015) for estimates of the weighted cumulative response rates (WCRR) for PSID sample persons who were recruited and participated in the 1968 PSID Core panel.   Estimates of the WCRR statistic have been presented for the full panel, the two components (SRC and SEO) of the original dual-frame sample of U.S, households and for major demographic subpopulations of the 1968 sample persons. The weighted ratio estimator for subpopulations (3) may be used to estimate the WCCC for any subpopulation of the 1968 sample persons. Furthermore, studies of panel retention/attrition for different time periods (e.g. 1997 to present) require only the alteration of the base time period, t_0, in the numerator of the ratio estimator.

$$WCRR(t \mid \mathrm{t}_0) = \frac{\hat{N}_{t_0,r}}{\hat{S}_t} = \frac{\sum_i W_{i,t_0} \cdot I_{i \in r(t)}}{\sum_i W_{i,t} \cdot I_{i \in r(t)}} \qquad (6)$$

where :

$\hat{N}_{t_0,r}$ = the weighted estimate of the count of the time t_0 study population

"represented" by sample persons responding at time t;

$\hat{S}_{t,r}$ = the weighted estimate of the count of the time t_0 study population

members alive at time t;

Similar adaptations of the numerator and denominator of the estimator may be used to investigate longitudinal attrition for other PSID core sample subgroups of interest. For example, to estimate the cumulative response rate at time t for any sample person born during the interval [1968, t) one only need take the value of $W_{i,t0}$ in expression 6 to be the weight of the newborn at the data collection time t_0 immediately following their birth—the first time they appear in the panel and receive a non-zero longitudinal weight.

Table 1: Estimates and Standard Errors of PSID Cumulative Response Rate (UCRR, WCRR) for the Original 1968 Sample of Individuals. Estimates for Selected Years: 1968-2015

Panel Year	1968 Sample Person Respondents*	UCRR Unweighted Cumulative Response Rate**	WCRR- Weighted CRR (Std. Error)		
			Total 1968 Sample	Sample Frame	
				SRC	SEO
1968	18233	100.0%	100.0%	100.0%	100.0%
1969	16050	88.0%	91.6% (0.12%)	91.3% (0.14%)	93.3% (0.15%)
1974	13917	76.3%	80.6% (0.16%)	80.3% (0.18%)	82.0% (0.23%)
1979	12064	66.2%	73.2% (0.17%)	73.5% (0.15%)	71.9% (0.48%)
1984	10524	57.7%	67.1% (0.23%)	67.9% (0.20%)	63.7% (0.64%)
1989	8938	49.0%	59.2% (0.28%)	60.2% (0.26%)	54.7% (0.72%)
1993	8236	45.2%	55.6% (0.28%)	57.2% (0.26%)	49.0% (0.60%)
1997***	5718	***	55.9% (0.34%)	56.7% (0.32%)	43.7% (0.53%)
2003	5124	***	53.9% (0.35%)	54.8% (0.32%)	42.3% (0.58%)
2007	4724	***	51.5% (0.34%)	52.3% (0.32%)	40.1% (0.54%)
2011	4298	***	46.9% (0.32%)	47.6% (0.31%)	37.1% (0.61%)
2015	3702	***	40.4% (0.28%)	41.0% (0.27%)	33.4% (0.61%)

*Includes attrition due to nonresponse and mortality.

**See text for discussion of potential downward bias of the UCRR values.

*** PSID Core Sample panel reduction implemented in 1997. UCRR values cannot be calculated due to differential subsampling of 1968 family trees.

Table 2a: Weighted Estimates and Standard Errors of PSID Cumulative Response Rate (WCRR) for Subpopulations of the Original 1968 Sample of Individuals, 1968-2015.

Panel Year	Demographic Subpopulation of 1968 PSID Sample Persons			
	Sex		Race of HH Head	
	Male	Female	Black	Non-Black
1968	100.0%	100.0%	100.0%	100.0%
1969	91.9% (0.13%)	91.4% (0.14%)	91.9% (0.33%)	91.6% (0.13%)
1974	81.0% (0.18%)	80.3% (0.19%)	80.0% (0.49%)	80.7% (0.16%)
1979	73.7% (0.19%)	72.8% (0.19%)	68.3% (0.51%)	73.9% (0.15%)
1984	67.5% (0.25%)	66.8% (0.26%)	58.1% (0.61%)	68.4% (0.18%)
1989	59.8% (0.29%)	58.6% (0.32%)	47.3% (0.48%)	60.9% (0.19%)
1993	56.0% (0.29%)	55.3% (0.34%)	43.8% (0.44%)	57.3% (0.20%)
1997*	56.2% (0.41%)	55.6% (0.37%)	43.8% (0.52%)	57.6% (0.26%)
2003	54.1% (0.37%)	53.8% (0.42%)	42.5% (0.59%)	55.6% (0.25%)
2007	51.5% (0.38%)	51.4% (0.40%)	40.5% (0.62%)	53.0% (0.27%)
2011	46.6% (0.35%)	47.1% (0.37%)	37.1% (0.61%)	48.3% (0.25%)
2015	39.8% (0.32%)	40.9% (0.32%)	32.9% (0.58%)	41.5% (0.22%)

* PSID Core Sample panel reduction implemented in 1997.

Table 2b: Weighted Estimates and Standard Errors of PSID Cumulative Response Rate (WCRR) for Subpopulations of the Original 1968 Sample of Individuals, 1968-2015.

Panel Year	Demographic Subpopulation of 1968 PSID Sample Persons			
	Age in 1968			
	0-17	18-39	40-64	65+
1968	100.0%	100.0%	100.0%	100.0%
1969	98.1% (0.09%)	87.8% (0.20%)	88.1% (0.15%)	87.6% (0.33%)
1974	85.7% (0.18%)	76.5% (0.19%)	79.2% (0.21%)	75.4% (0.96%)
1979	77.2% (0.28%)	69.5% (0.20%)	72.6% (0.25%)	67.8% (0.78%)
1984	68.8% (0.40%)	64.5% (0.26%)	68.2% (0.26%)	63.1% (1.05%)
1989	60.2% (0.45%)	57.4% (0.28%)	60.2% (0.29%)	52.5% (1.26%)
1993	55.6% (0.44%)	54.5% (0.28%)	57.8% (0.33%)	49.8% (1.27%)
1997*	55.8% (0.52%)	54.7% (0.38%)	58.7% (0.39%)	52.2% (3.32%)
2003	53.6% (0.50%)	53.0% (0.38%)	58.1% (0.48%)	50.4%**
2007	51.2% (0.47%)	50.6% (0.39%)	56.7% (0.51%)	
2011	47.6% (0.46%)	45.5% (0.37%)	48.8% (0.49%)	
2015	42.3% (0.38%)	38.0% (0.35%)	37.9% (0.46%)	

* PSID Core Sample panel reduction implemented in 1997.

** The number of responded sample person was 1.

References

Berglund, P.A., Chang, W., Heeringa, S.G., McGonagle, K., Brown, C., and Johnson, D. (2017). "Panel Study of Income Dynamics Construction and Evaluation of the Longitudinal Sample Weights 2015", PSID Technical Report, ISR, University of Michigan - https://psidonline.isr.umich.edu/data/weights/long_weight_15.pdf

Gouskova E., Heeringa, S.G., McGonagle, K., Schoeni, R., and Stafford, F. (2008). "Panel Study of Income Dynamics Revised Longitudinal Weights 1993-2005", 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, West, B.T., Berglund, P.A. (2017). *Applied Survey Data Analysis, 2nd Edition.* Chapman and Hall, London.

Hill, M. (1992) The *Panel Study of Income Dynamics – A User's Guide*. Sage Publications.

McGonagle, K., Smith, J.P., and Schoeni, R. (2008). "The PSID Sample Leaver Tracking Project", PSID Technical Report, ISR, University of Michigan - https://psidonline.isr.umich.edu/Publications/Papers/tsp/2008-03_PSID_Sample_Leaver_Tracking.pdf