

## Information on the PSID Immigrant Sample Addition of 1997/1999

PSID Staff

This project was supported by funding from the National Science Foundation (SES 0518943).

## Information on the PSID Immigrant Sample Addition of 1997/1999

### Overview of Sample Design

A design goal of the PSID since its inception is that it represents the national population of U.S. families with sampling weights. In addition to following the children of the adult sample as they leave home, the sample is periodically refreshed over time to reflect changes in the composition of the national population. In 1997, in order to keep the study representative of the U.S. population, two major changes were made to the PSID sample: 1) a reduction of the core sample and 2) the introduction of a refresher sample of post 1968 immigrant families and their adult children. The original core sample was reduced from nearly 8,500 families in 1996 to approximately 6,168 in 1997 and a sample of 511 immigrant families was added in 1997/1999. The sample so refreshed was 6,434 for 1999 and has grown to approximately 8,400 in 2007.

It is important to note that the size of the immigrant refresher sample was driven by budget constraints and is not designed for immigrant-focused subgroup analysis. Its central advantage is in maintaining the national representation of the overall sample.

The PSID immigrant sample was designed to update the PSID sample of U. S. households by adding a representative sample of recent immigrants to the U.S. Immigrants eligible for the supplemental sample were those who immigrated to the U.S. after 1968 and who were not married to persons who were living in the U.S. at the time of the original PSID sample selection in 1968. These recent immigrants were not represented in the PSID sample as of 1997.

Data from the 1990 Census was merged with the 1990 National Sample Database at the county level in order to obtain the numbers of post-1968 immigrants living in each PSU. This allowed us to group PSUs together to form 28 collapsed strata each of which had a minimum number of post-1968 immigrants.

The Immigrant sample was selected from 28 collapsed strata of the Survey Research Center's 1990 National Sample. Each stratum represented approximately 1.5 - 2 percent of the total post-1968 immigrant population.

The Kish-Scott<sup>1</sup> procedure was used to select one primary stage unit (PSU) from each stratum and retained 24 PSUs from the current "A" half-sample PSUs of the 1990 SRC National Sample. The other 4 PSUs used for this sample are new PSUs specifically selected for this PSID Immigrant Supplement.

A total of 194 second stage units (SSUs) were allocated to the strata in proportion to their share of the post-1968 immigrant population. The Kish-Scott procedure was also used in the second stage of selection in order to retain as many as possible of previously listed SSUs. If more SSUs were required to meet the number allocated to a PSU, the SSUs were selected from the 1990 SRC sampling frame of SSUs. Of the 194 SSUs, 103 were retained listed SSUs and 91 were newly selected SSUs.

### Weights for the Immigrant Sample

Each of the interviewed families in the 1997/1999 Immigrant Supplement was assigned an initial base weight value that reflected the probability of selection and screening for the area segment in which they resided. These base weights for immigrant supplement households were scaled to U.S. population values. The 1997 PSID preliminary family weight = 16.198\* Immigrant household weight. 1997 and 1999 PSID weights for families and individuals are scaled to represent the corresponding United States population estimates (in 1000s). The first step in this scaling process is to apportion the total population weight between the domain of families and

---

<sup>1</sup> Leslie Kish and Alastair Scott were the first to describe the probability sampling procedures which are designed to optimize the reselection or "retention" of sample units during a transition from an old to a new sample design. A description of this can be found in: Kish, Leslie, and Scott, Alastair "Retaining units after changing strata and probabilities." *Journal of the American Statistical Association*, Vol. 667, Number 335, Applications Section, Sept. 1971.

individuals who were eligible for the 1968 PSID Core panel and those who were only eligible for inclusion under the 1997 Immigrant Supplement. Based on data from the 1997 CPS, an estimated 7.5 percent of U.S. households have immigrated to the United States after 1968. Therefore the Immigrant Supplement sample represents about 7,500,000 of the 100,000,000 U.S. households and the 1968-based PSID sample represents the remaining 92,500,000 households. The 1997 Immigrant Supplement sample includes 441 respondent families who participated in 1997 and an additional 70 families who became participants in 1999 for a total of 511 families.

### Identifying Immigrant Families in the Data

Immigrant families and individuals were assigned a unique range of values so that they can easily be distinguished from the other PSID samples. The 1997 Interview (ID) Number is in the range 10001-10441, and the 1968 Interview (ID) Number is in the range 3001-3441. For the 1999 addition, this series is continued, with values of 3442 through 3511. The table below displays this information about the 1968 Interview Number. Person Numbers for immigrants were assigned in the usual manner, with all those present in the first (1997) wave receiving values from 1-20. Those in institutions received values of 20-29, and movers-out were given nonsample Person Numbers (170 or greater).

All new heads and wives/"wives", both immigrant and core sample, were asked in what year they came to the U.S. to stay. These questions served as screening questions for the immigrant sample and were useful in development of 1997 and 1999 sample weights for both the core and the immigrants. These variables are located on the 1997 and 1999 portion of the cross-year individual file.

Name	Label	Number	Type	Width	Decimals	Location
V1	1968 INTERVIEW NUMBER OF INDIVIDUAL	1	Numeric	4	0	1-4

**Question** 1968 Interview Number (1968 ID Number) of the Individual

**Explanation** This variable is the identification number for the 1968 Family Unit with which the individual is associated, whether or not he or she actually resided in that Family Unit. The combination of this variable and the individual's Person Number (V2) provides unique identification for him or her. These two variables are the same identifiers as those on the cross-year Individual File. Ranges of values are also meaningful, as indicated below.

Frequency	Code	Description
18974	1-2930	Individual is associated with a family from the SRC cross-section sample.
1673	3001-3511	Individual is associated with a family from the New Immigrant sample.
16342	5001-6872	Individual is associated with a family from the Census sample.
8211	7001-9308	Individual is associated with a family from the Latino sample.

### **Race of Immigrant Family Heads**

As the table below shows, most family heads in the immigrant sample identify their race as Latino (52.4%), followed by Asian (21.1%), white (11.7%), black (7.8%), and other (6.8%).

### **Race of 1997 & 1999 Immigrant Family Heads (N=511)**

<b>Self-Identified Race</b>		
White	60	11.7%
Black	40	7.8%
Asian	108	21.1%
Latino	268	52.4%
Other	35	6.8%
Total	511	100%