

United States 2002

1 Survey Description

Survey: The Panel Study of Income Dynamics (PSID) is the longest longitudinal household survey in the world. The Study started in 1968 and is currently being developed by the University of Michigan. It includes data related to employment, income, wealth, expenditures, health, marriage, childbearing, child development, philanthropy, education, and other topics. This document refers specifically to the data collected in 2002.

Link to the document: <https://simba.isr.umich.edu/data/data.aspx>

Sample: The sample of the PSID was originally composed of approximately 18,000 people in 5,000 households, making it a representative sample of the entire country. This sample was complemented with a non-representative oversample of low-income families to facilitate the investigation of issues related to poverty and vulnerability. Over the years, the sample size has changed as the number of members in the families considered originally has grown. There are 7,494 individuals in the total sample and 6,787 individuals in the analysis sample. Section 3 of this document describes the prevalence and pattern of missing data.

Weights: Population weights are computed as the ratio between specific weights for the Panel Study of Income Dynamics (PSID) and the Current Population Survey (CPS). These weights, based on sex, age, and ethnicity, are tailored for different demographic groups. Age groups are formed, and missing values are imputed using historical data. The process ensures a more representative combined dataset for analysis.

Outcome: The outcome variables are annual equivalized household disposable total (*eq_inc*) income in dollars PPP 2017.¹

Circumstances:

- Race 'ethnicity'(white, black, american indian, aleut, eskimo, asian, pacific islander, hispanic and others described in Table 1)
- Sex (female, male)
- Region of upbringing (6 categories described in Table 1)
- Fathers's edu. (years of education, described in Table 2)
- Mother's edu. (years of education, described in Table 2)
- Father's occ. (3 categories, High: includes ISCO 1, 2 and 3, Medium: includes ISCO 4, 5 and 6, and Low: ISCO 7, 8, 9 and 0, described in Table 3)
- Mother's occ. (3 categories, High: includes ISCO 1, 2 and 3, Medium: includes ISCO 4, 5 and 6, and Low: ISCO 7, 8, 9 and 0, described in Table 3)

¹Income variable was equivalized using the square root scale.

2 Descriptive Statistics

Table 1: Respondant's socio-demographics - 2002

	Analysis sample	Total sample
	(N=6,787)	(N=7,494)
Gender		
0 Female	3,723 (54.9%)	4,064 (54.2%)
1 Male	3,064 (45.1%)	3,430 (45.8%)
ethnicity		
1 White	4,776 (70.4%)	5,031 (67.1%)
2 Black	1,453 (21.4%)	1,765 (23.6%)
3 Am. Indian, Aleut, Eskimo	117 (1.7%)	137 (1.8%)
4 Asian, Pac. Islander	114 (1.7%)	151 (2.0%)
5 Hispanic	27 (0.4%)	43 (0.6%)
7 Other	300 (4.4%)	367 (4.9%)
Region of upbringing		
1 Northeast	1,162 (17.1%)	1,212 (16.2%)
2 North central	1,895 (27.9%)	2,012 (26.8%)
3 South	2,220 (32.7%)	2,474 (33.0%)
4 West	887 (13.1%)	947 (12.6%)
5 Alaska, Hawaii	4 (0.1%)	4 (0.1%)
6 Foreign country	619 (9.1%)	743 (9.9%)
Missing	0 (0%)	102 (1.4%)

Table 2: Parental education - 2002

	N	Mean	SD	Median	Min	Max	Missing
Analysis sample - Mother's edu	6,787	3.822	2.316	4	0	8	0
Analysis sample - Father's edu	6,787	3.845	2.124	4	0	8	0
Total sample - Mother's edu	7,494	3.747	2.352	4	0	8	219
Total sample - Father's edu	7,494	3.778	2.165	4	0	8	97

Table 3: Parental occupation - 2002

	Analysis sample	Total sample
	(N=6,787)	(N=7,494)
Father's occupation (ISCO)		
1 Low	2,651 (39.1%)	2,919 (39.0%)
2 Middle	2,226 (32.8%)	2,329 (31.1%)
3 High	1,910 (28.1%)	1,978 (26.4%)
Missing	0 (0%)	268 (3.6%)
Mother's occupation (ISCO)		
1 Low	4,430 (65.3%)	4,707 (62.8%)
2 Middle	517 (7.6%)	558 (7.4%)
3 High	1,840 (27.1%)	1,966 (26.2%)
Missing	0 (0%)	263 (3.5%)

Table 4: Respondant's income - 2002

	N	Mean	SD	Median	Min	Max	Missing
Analysis sample	6,787	45,157	59,189	35,580	30.31	1,747,830	0
Total sample	7,494	43,973	58,053	34,525	30.31	1,747,830	0

3 Missing data analysis

3.1 Missing patterns

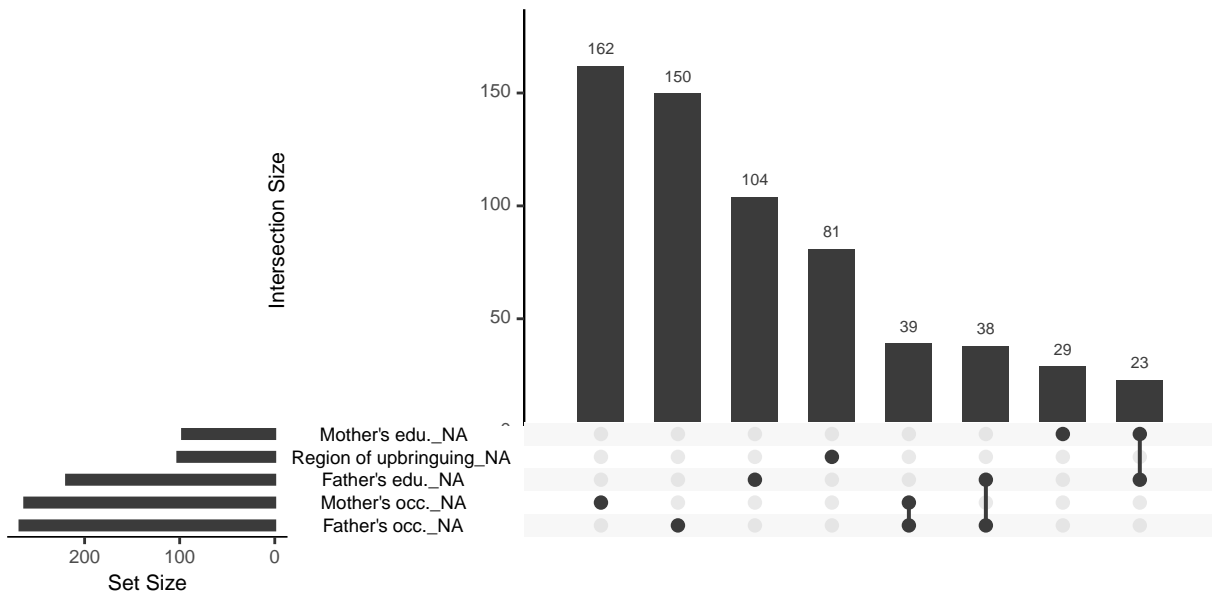


Figure 1: Missing patterns: *Left*: Marginal distribution of missing observations per variable. *Right*: Combination of missingness across cases

3.2 Differences in expected total equivalized household income between samples

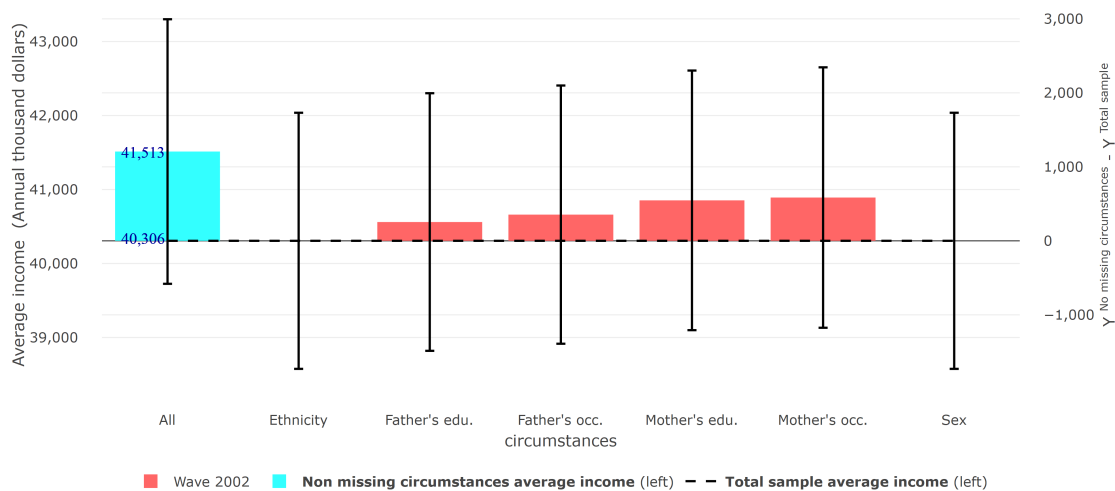


Figure 2: Differences in expected total equivalized household income between the sample with non-missing circumstances and the total sample

3.3 Gini coefficient

Table 5: Gini coefficient in analysis sample and total sample

Wave	Sample	Gini	Lower bound	Upper bound	Average income
Wave 2002	Analysis sample	0.395	0.380	0.417	41,513
Wave 2002	Total sample	0.400	0.386	0.421	40,306

3.4 Differences in Gini coefficient between samples

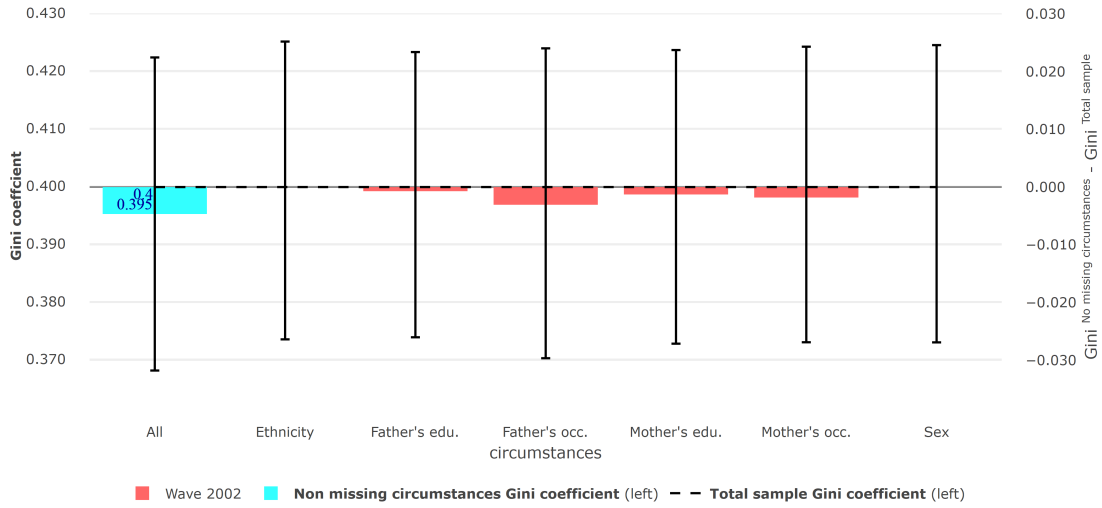


Figure 3: Differences in Gini coefficient between the sample with non-missing circumstances and the total sample