

Attrition and the National Longitudinal Surveys' Mature Women Cohort

by

Jay Zagorsky

and

Pat Rhoton

Center for Human Resource Research

Ohio State University

July 1998

Zagorsky is a Research Scientist at Ohio State University. Rhoton is a Senior Research Associate at Ohio State University and the data archivist for the Mature Women cohort. The authors thank Amanda McClain for her editing assistance.

Authors' email addresses:

Zagorsky	zagorsky.1@osu.edu
Rhoton	rhoton.1@osu.edu

Executive Summary

This article examines attrition in the National Longitudinal Survey of Mature Women. The focus of the article is threefold: it documents how much attrition has occurred, examines who has exited the survey and compares the NLS data to other national data for the same age group. The major findings of this research are:

- As of 1995, 2,372 or approximately 47% of the original 5,083 respondents have exited the survey. After adjusting for respondents who have died, attrition falls to 38.7%.
- Attrition happens at a steady pace each survey round. On average, each time the survey is fielded about 0.6% of the cohort dies and 2.3% do not respond for other reasons. There have been no years when very sharp losses occurred.
- Whites leave the survey at a slower rate than blacks and other races. Whites also have a higher probability of completing every interview.
- Over time, individual refusals are the primary reason the vast majority of individuals do not complete an interview. In 1995, 55% of the noninterviews were classified as refusals.
- The typical individual who leaves the Mature Women survey had a lower socio-economic status in 1967 than those remaining in the survey; those remaining had higher income and educational attainment at the start of the survey, not controlling for race.
- While a relatively large number of respondents have attrited, the survey still closely matches characteristics of similar women found in the Current Population Survey (CPS), another national labor survey.

I. Introduction

In 1967 the National Longitudinal Survey (NLS) of Mature Women began with 5,083 respondents. As of the 1995 survey 2,711, or 53.3%, of the original respondents were interviewed. This research focuses on quantifying who left, or attrited, from the survey and the ramifications of losing these individuals. In brief, the two major findings of this research are: (1) The NLS of Mature Women has primarily lost those at the lower end of the socio-economic scale in 1967. (2) Nonetheless, while a large number of individuals have left, overall the sample still matches its portion of the nation.

This paper continues the focus of previous NLS data quality research that investigates the effects of attrition and sample representativeness. Attrition presents two distinct problems for researchers. First, if large numbers of respondents drop out or attrite from the survey, the remaining respondents may no longer constitute a valid sample. Second, even if overall response rates are high, attrition can be a problem if a select group of respondents leave. For example, if less wealthy respondents tend to leave the survey, data on income and assets may overrepresent the wealth of the cohort.

Selective and massive attrition are not the only problems which should concern researchers. Even with zero attrition, longitudinal surveys can become unrepresentative if the national population changes. For example, since the NLS of Mature Women was sampled in the 1960s, it does not represent the large number of Hispanics who have immigrated since that time.

The remainder of this paper examines the effects of attrition on the NLS of Mature Women. The next section reviews the previous literature on this subject. Section three discusses the statistical problems caused by attrition. Section four describes the Mature Women's cohort. Section five provides detailed tables on completion and attrition rates for this cohort. Section six

investigates the characteristics of individuals who have attrited over time. Section seven investigates how well NLS data match national data, represented by the CPS (Current Population Survey), over time. Finally, a conclusion summarizes the entire paper.

II. Literature Review

While the NLS cohorts have generated thousands of research papers and projects, the subject of attrition has received relatively little attention. This section overviews the research on sample attrition and the NLS cohorts.

One of the early sample representativeness reports was written by June O'Neill (1982). An appendix to this report contains an investigation of sample bias in the four original NLS cohorts. This appendix, written by members of the U.S. Census Bureau, suggests that after 10 years of surveying little bias had occurred due to attrition.

Other attrition findings are reported by Parsons (1987). This research investigates only the Young Men's cohort and breaks down response rates by IQ, father's education and ethnicity. This study shows that attrition among whites is negatively related to IQ but that there is no relationship between IQ and attrition for blacks.

Rhoton (1984) examined reasons why the original NLS cohorts have a lower response rate than the NLSY79. She suggests that higher NLSY79 response rates are due to a number of procedural factors. For example, the NLSY79 interviewers tell respondents that they will be interviewed next year, while original cohort interviewers are instructed not to be specific about when the next contact will be made. NLSY79 questionnaires include more extensive locating information, such as names and contact information for parents and other relatives, which helps interviewers to find respondents who move between rounds. Unlike the original cohorts, NLSY79 respondents are given a nominal sum of money for participating in the survey. Finally, NLSY79

interviewers can fall back on a central locating shop in Chicago for help; at that time original cohort Census interviewers had no centralized locating organization to assist in finding difficult-to-locate respondents.

Juster and Kuester's 1991 article investigated a variety of wealth measures in the NLS of Older Men. Their research suggests that wealthier respondents dropped out of the survey at a significantly higher rate than other respondents did.

Rhoton and Nagi (1989) have also examined the relationship between attrition and wealth in all four of the original cohorts. Defining wealthy individuals as respondents who were in the top 5% of either income or assets in the first year of the survey, they determined whether wealthy respondents attrited at a different rate than other respondents. Their findings about attrition are at odds with Juster and Kuester: "The non-wealthy respondents showed a consistent tendency toward greater attrition." Zagorsky (1997) discovered a similar problem with the NLSY79; in this cohort, non-wealthy respondents have also dropped out at a higher rate.

MaCurdy, Mroz and Gritz (1998) examined attrition in the NLSY79. They compared the NLSY79 and CPS for three key variables: wages, earnings and returns to education. Their research concludes that "our analysis offers little basis for suspecting that the NLSY[79] presents an inaccurate picture of youths' labor market experiences."

Overall, previous NLS attrition research has not found any serious problems. Nevertheless, researchers should understand the impact of attrition. The next section describes the statistical problems caused when individuals drop out of the survey.

III. Statistical Problems

A large amount of NLS research is concerned with measuring the impact a set of input variables has on an outcome. In equation form this is written:

$$(1) \quad Y = \alpha + \beta X + \varepsilon$$

where Y is the outcome measure, α is a constant value, β is the set of estimated coefficients, X is the vector of input variables, and ε is the error term. The original sampling framework of the NLS survey was designed so that weighted estimates of β from each cohort match the “true” values of β derived from the matching portion of the U.S. population.

One simple example elaborates on the problem attrition causes researchers. Suppose researchers want to use an equation similar to (1) to investigate how various factors affect an individual’s net worth. In such research, net worth is Y and the vector X contains variables such as educational attainment, marital status, occupation, industry, and work experience. Researchers estimate β using statistical packages such as SAS or SPSS. These coefficient estimates show how much changing one input, such as increasing work experience, affects the output, or net worth.

For example, assume that in the first survey there are 100 individuals who hold different amounts of wealth. Researchers who estimate equation (1) for the entire group will end up with a societal estimate of β that combines wealthy and non-wealthy individuals. If the top 10 wealth holders leave the survey, then the coefficient estimate in equation (1) will fall. Hence, attrition of the high wage earners causes the coefficient estimate in this case to be downwardly biased and underrepresent the “true” value of the factor on wealth. If just 10 of the lowest wealth holders leave the survey then coefficient estimate rises, and researchers are faced with a coefficient which overrepresents the true value. Attrition by either group results in biased coefficient estimates.

However, if respondents attrite randomly because of a factor not related to wealth, coefficient estimates will not be biased. As this simple example shows, not all survey attrition leads to biased results. Random attrition of respondents means the survey continues to be useful for researching the U.S. economy. However, nonrandom attrition may cause some specific results

from the survey to be suspect. The goal of the rest of the paper is to determine whether the NLS of Mature Women suffers from random or nonrandom attrition. The next section first describes the cohort so that readers understand its composition.

IV. Cohort Description

The Mature Women's cohort was started in 1967 by the U.S. Department of Labor to understand the important choices made by middle-aged women as they reentered the work force after raising their children. The enormous success of the Older and Young Men's cohorts, started in 1966, convinced the Department of Labor to fund similar surveys of American women.

To reduce costs, all four original NLS cohorts were taken from the same sampling frame. In 1965, Census Bureau employees created a list, or sampling frame, by surveying a random set of households. Men in these households ages 45 to 59 in 1966 were selected for the Older Men's cohort. Women in these households ages 30 to 44 in 1967 were selected for the Mature Women's cohort. Young adults ages 14 to 24 in 1966 and 1968 respectively created the basis for the Young Men and Young Women cohorts. Due to this sample design, a number of respondents have relatives in one or more cohorts. In particular, many Mature Women have husbands in the Older Men survey and children in the Young Men's and Young Women's surveys.

The Mature Women's cohort actually comprises two samples. The first group is a nationally representative sample of women living in the United States in 1965. At the same time, a second group of black Americans was oversampled. This second group is needed to ensure that estimates based on race have enough black respondents to provide statistically accurate results; the total sample contains about three times as many blacks as the expected population rate.

The U.S. Census Bureau interviewed the Mature Women's cohort seventeen times between 1967 and 1995.¹ Ten of these interviews were conducted in person, six were administered by telephone and one by mail. From the cohort's beginning until the middle of the 1980s, the Census Bureau followed a simple interview policy. If a respondent refused to be interviewed, she was dropped from the interview rolls. Additionally, if a respondent was a noninterview for any other reason two years in a row, she was dropped from consideration. As will be shown in the data tables, these two policies significantly decreased the number of respondents. In 1986, these policies were discontinued.

Each survey includes questions on a number of key topics. Respondents are asked detailed questions about the number and characteristics of people living in their household. Additionally, respondents are asked about changes in their educational achievement, employment, marital status, and income. Periodically, the survey has collected information about topics such as pensions, transfers of money and time, leisure time activities, cigarette and alcohol use, and household chores. Overall, the information provides researchers with a detailed image of this select segment of the U.S. population.

V. Attrition

This section provides a detailed examination of completion and attrition rates in the Mature Women's cohort. The data are broken down to provide a variety of different perspectives on this subject. The most important information is contained in Table 1, which shows the number of interviews completed over time by cohort members. In 1967, 5,083 women completed the first survey. By 1995 only 2,711, or 53.3%, of these original participants were interviewed.

¹ In 1997 2,605 respondents were interviewed. Details about these individuals are not yet available.

In table 1 the third and fifth columns shows what percentage of the cohort participated in each year. The third column shows the percentage who were interviewed based on the original sample of 5,083 respondents, while the fifth column is based only on the number of respondents who are known to be living. Both columns show a steady drop in participation over time.

Column three shows on average the cohort loses 2.9% or about 148 of its members, in each survey round. This number, however, is upwardly biased since death eliminates many respondents who normally would participate. Revising the data to remove respondents who have died shows in column five that on average the cohort really loses 2.3%, or about 117 of its members, in each survey round. Death has clearly caused a significant portion of attrition. For example, comparing columns three and five in 1995 reveals that removing deceased respondents increases response rates by 8%.

Readers are cautioned when examining this table and those that follow that many surveys were done at two- or even three-year intervals. Therefore, the number of respondents lost from round to round is not interpretable in yearly terms.

Table 1. Number of Interviews Completed

<i>Year</i>	<i>Number interviewed</i>	<i>Percent interviewed</i>	<i>Percent lost from previous round</i>	<i>Number known dead</i>	<i>Percent of living respondents interviewed</i>
1967	5,083	100.0%			100.0%
1968*	4,910	96.6%	-3.4%	22	97.0%
1969	4,712	92.7%	-3.9%	35	93.3%
1971	4,575	90.0%	-2.7%	60	91.1%
1972	4,471	88.0%	-2.0%	72	89.2%
1974*	4,322	85.0%	-2.9%	101	86.8%
1976*	4,172	82.1%	-3.0%	131	84.2%
1977	3,964	78.0%	-4.1%	140	80.2%
1979*	3,812	75.0%	-3.0%	176	77.7%
1981*	3,677	72.3%	-2.7%	216	75.5%
1982	3,542	69.7%	-2.7%	238	73.1%
1984*	3,422	67.3%	-2.4%	285	71.3%
1986*	3,335	65.6%	-1.7%	341	70.3%
1987	3,241	63.8%	-1.8%	364	68.7%
1989	3,094	60.9%	-2.9%	431	66.5%
1992	2,953	58.1%	-2.8%	546	65.1%
1995	2,711	53.3%	-4.8%	664	61.3%
1997	2,605	51.2%	-2.1%	unknown	unknown

Note: A * means the interview was a mail or telephone survey.

Readers should note the attrition suggested in table 1 greatly overrepresents the amount of lost information. Starting in the 1992 survey a detailed retrospective work history section for both the respondent and husband was added. Other sections like education, training, marital status and fertility are designed to capture information lost due to missing interviews. Hence, a perfect response record is not needed for researchers to understand how the respondent's life changes over time, unless the respondent leaves the survey forever.

Table 2 shows the number of interviews each respondent completed out of all 17 surveys fielded to date. Overall, the Mature Women participate in the surveys quite regularly. Over three quarters (75.4%) have completed more than half the surveys. The modal respondent (44.1%) completed all surveys. Respondents who either completed all surveys or missed just one make up more than half (55.2%) of the entire group. Relatively few respondents have not wanted to

participate; only 2.6% completed just a single survey and slightly more than ten percent completed four or fewer surveys. This table shows that the Mature Women have maintained relatively high participation rates over the almost three decades of interviewing.

Table 2. Number of Interviews Completed out of 17 Surveys

<i>Number of surveys</i>	<i>Number who completed</i>	<i>Percent completing surveys</i>	<i>Cumulative percent</i>
17	2241	44.1	44.1
16	563	11.1	55.2
15	308	6.1	61.3
14	193	3.8	65.1
13	117	2.3	67.4
12	121	2.4	69.8
11	102	2.0	71.8
10	52	1.0	72.8
9	131	2.6	75.4
8	149	2.9	78.3
7	205	4.0	82.3
6	154	3.0	85.3
5	155	3.0	88.3
4	127	2.5	90.8
3	144	2.8	93.6
2	188	3.7	97.3
1	133	2.6	99.9

Table 3 expands on the information presented in the previous table by breaking down the number of interviews completed by race. Readers should note that Census procedures in the 1960s divided almost all individuals into white or black; most Hispanics were classified as white. The “other” classification includes primarily American Indians and Asians, not Hispanics.

As the table shows, non-blacks, who are primarily whites and Asians, have a slightly stronger attachment to the survey than blacks and are more likely to continue participating. First, active non-black respondents, those completing either all surveys or missing just one, comprise 56.2% of the non-black group. Active black participants, however, comprise 52.3% of their racial

group, a difference of 3.9 percentage points. Additionally, while 75.6% of non-blacks completed at least half of the surveys, 74.5% of blacks finished at least half (nine) of the surveys.

Table 3. Number of Interviews Completed out of 17 Surveys by Race

<i>Number of surveys</i>	<i>Non-Black</i>		<i>Black</i>	
	<i>Number who completed</i>	<i>Cumulative percent</i>	<i>Number who completed</i>	<i>Cumulative percent</i>
17	1688	45.7%	553	39.8%
16	389	56.2%	174	52.3%
15	214	62.0%	94	59.1%
14	133	65.6%	60	63.4%
13	88	68.0%	29	65.5%
12	86	70.3%	35	68.0%
11	63	72.1%	39	70.8%
10	34	73.0%	18	72.1%
9	97	75.6%	34	74.5%
8	108	78.5%	41	77.5%
7	154	82.7%	51	81.2%
6	111	85.7%	43	84.2%
5	110	88.7%	45	87.5%
4	85	91.0%	42	90.5%
3	92	93.5%	52	94.2%
2	144	97.4%	44	97.4%
1	97	100.0%	36	100.0%

Table 4 examines the percentage of Mature Women who have answered the survey in every year through a given year. For example, 2,241 respondents (44.1%) answered every survey up to and including 1995. This table provides the reader with a picture of high levels of participation over time, which is important for researchers trying to create unbroken pictures of the respondent's life. Continuous participation is important since retrospective questions are not asked about all activities. In addition, retrospective questions can capture less reliable information than contemporaneous questions. Finally, answering every survey is a good indicator that respondents take the interviews seriously and are committed to future participation.

The table shows that the percentage of respondents who have answered the survey in every year has dropped steadily over time, from 96.6% in 1968 down to a low of 44.1% in 1995. Each time the survey has been fielded, the percentage who have responded to every survey through that year has dropped by an average of 3.5%, or 178 respondents. The biggest drop (4.9%) occurred late in the survey, between the years 1992 and 1995, while the smallest drop (2.6%) occurred between 1986 and 1987.

Table 4. Percent of Mature Women Answering in Every Survey Year

<i>Year</i>	<i>Number</i>	<i>Percent</i>	<i>Year</i>	<i>Number</i>	<i>Percent</i>
1967	5083	100.0	1981	3487	68.6
1968	4910	96.6	1982	3348	65.9
1969	4673	91.9	1984	3176	62.5
1971	4487	88.3	1986	3043	59.9
1972	4342	85.4	1987	2912	57.3
1974	4177	82.2	1989	2715	53.4
1976	4015	79.0	1992	2493	49.0
1977	3801	74.8	1995	2241	44.1
1979	3637	71.6			

While the above data examine completion rates from a variety of perspectives, Table 5 presents the specific reasons why a respondent was not interviewed. Before examining this table, readers should understand that the Census Bureau procedures automatically exclude a number of respondents from the interview process. In all years, respondents who are currently institutionalized (i.e. in jail or hospitalized) are not interviewed. If a respondent has joined the Armed Forces, she is removed from the valid interview rolls. Additionally, respondents who are living outside the U.S. are dropped from contact consideration.

The Census Bureau also had a number of special procedures in force from 1968 to 1984. During these years, if a respondent refused to grant an interview, she was dropped automatically from all further interviews. Additionally, if a respondent was a noninterview for two consecutive

rounds for one of the reasons listed above, she was permanently dropped from consideration. However, this particular policy was changed in 1986 and respondents were no longer dropped after two interview years. Finally, a small number of respondents have written or called their congressional representative and asked to be removed from the list of eligible respondents; these respondents, called congressional refusals, are permanently dropped.

Table 5. Reason for Noninterview

<i>Year</i>	<i>In survey</i>	<i>Unable to contact</i>	<i>Moved</i>	<i>Repeated attempts to interview failed</i>	<i>Temp. absent</i>	<i>In an institution</i>	<i>Refused</i>	<i>Dead</i>	<i>Not in survey for 2 years, dropped</i>	<i>Other</i>
1968	4,910	3	56	3	0	1	76	22	0	12
1969	4,712	4	62	21	16	7	210	35	0	16
1971	4,575	1	61	22	12	6	292	60	29	25
1972	4,471	4	47	15	8	2	389	72	61	14
1974	4,322	8	34	9	6	5	479	101	104	15
1976	4,172	4	33	11	5	7	580	131	119	21
1977	3,964	2	22	17	10	6	761	140	141	20
1979	3,812	1	21	13	4	4	867	176	169	16
1981	3,677	1	17	10	5	6	963	216	183	5
1982	3,542	1	17	10	8	4	1,061	238	193	9
1984	3,422	7	25	19	1	6	1,110	285	194	14
1986	3,335	0	46	17	8	5	1,127	341	194	10
1987	3,241	4	31	15	20	8	1,190	364	194	16
1989	3,094	3	28	21	20	14	1,258	431	194	20
1992	2,953	17	48	5	5	15	1,271	546	194	29
1995	2,711	13	59	33	11	26	1,299	664	194	73

Table 5 shows that refusals are the major reason why eligible respondents were not interviewed. A total of 1,299 respondents either refused to participate in the 1995 interview or had been dropped for previously refusing. To provide some perspective, as of 1995 more than one quarter (25.6%) of all non-respondents are coded as refusing to participate; as a percentage of all noninterviews, refusals comprise 54.8%. However, over time the number of new refusals has been decreasing. For example, in the 1995 survey only 28 more women refused to participate than in

1992. This is a dramatic reduction in the trend seen earlier in the survey when, on average, 82 more women refused to participate each time the survey was fielded. Detailed information on why people refuse to participate is not available.

While refusals are cumulatively the most important reason for not participating in an interview, the table shows a variety of other reasons that have some relevance in explaining why individuals leave the survey over time. Ever since the cohort's inception, when 22 respondents died between the first and second surveys, death has been a steadily increasing reason for attrition. As of 1995, 664 Mature Women respondents, or 13.1% of the cohort, are known to have died.

U.S. life tables show that from 1967 to 1995 the typical white respondent has an 11.2% chance of dying, while black respondents have a 19.1% chance.² Multiplying these death probabilities by the number of respondents in 1967 suggests that 413 white respondents and 265 black respondents should have died by 1995. This estimate for blacks matches almost exactly the 266 black deaths found in the NLS data. The estimate for whites, however, is slightly higher than the 398 actually found. While some individuals whom the Census Bureau has dropped or been unable to locate have actually died, these numbers are very close to the expected values.

Census rules that prohibit interviewers from contacting individuals who are institutionalized eliminate only a small number of interviews. These rules are designed to have the NLS parallel the CPS rules for sample selection. The maximum number of interviews missed for this reason was 26 cases in 1995; information on the type of institution is not available.

The last column, labeled "other," contains a mixture of reasons for non-participation. In 1995, for example, seven respondents lived outside the U.S. borders and were not contacted. Twenty two respondents either had a lawyer or congressperson contact the Census Bureau and

² We thank Frank Mott for calculating these probabilities.

request that the interviewing be stopped. These and other varied reasons resulted in 73 individuals not being interviewed in the 1995 survey.

VI. Characteristics of Attriters

While the above section provides information on completion rates and reasons why respondents are not interviewed, it does not provide a picture of what type of individual has attrited over time. This section shows readers that the typical individual who left the survey by 1995 had a lower socio-economic status in 1967 than those who remain.

To examine the characteristics of survey leavers, all members of the Mature Women cohort were divided into two groups. Attriters were classified as individuals who did not complete the 1995 survey, while participants were those who completed the 1995 survey. Information on these two groups was then gathered from key variables drawn from the 1967 data set: educational attainment, labor force status, income (1966), rural/urban status, home ownership status and father's education. These variables are drawn from 1967 and not later years because only the first survey contains information collected for all respondents at the same point in time.³

Table 6 provides readers with an overview of the information contained in this section. This table shows that respondents missing the 1995 survey were more likely to be black; 49.7% of blacks missed the survey compared to 45.5% of nonblacks. Additionally, those who missed the survey were on average half a year older than those still participating. Since death is highly correlated with age it is reasonable to expect that participating living members of the Mature Women's cohort are younger than nonparticipating dead members.

³ Attrition is also related to age. Slightly higher death rates among the older women of the cohort are shifting the average age slightly downward.

A number of differences are clearly seen between the two groups' status during the first survey round held in 1967. Those participating in 1995 were more likely to be married during the first survey round, less likely to be employed, less likely to live in a central city, and more likely to be home owners. Additionally, participants had about a quarter of a year more education, lived in a family that on average was half a person larger, had roughly 10% more income and had a father who went to school roughly half a year more than those who did not participate in 1995.

Table 6. Overall Characteristics of Attriters and Nonattriters as of 1995 Survey.

	<i>Not In 95</i>	<i>In 95</i>
Number of Respondents by Race		
Nonblack	1681	2012
Black	691	699
Age in 1995	65.8	65.3
Status in 1967		
Married	78.8%	82.7%
Employed	50.9%	48.8%
Living In Central City	41.5%	38.8%
Living In A Home Owned By Family	63.2%	67.7%
Mean Values in 1967		
R's Highest Grade	10.73	11.15
Family Size in Persons	4.5	5.0
Family Income*	\$7,531	\$8,225
Father's Highest Grade	7.88	8.26

* Family income is for calendar year 1966.

The next six sets of graphs and discussions expand on the information presented in table 6. Each investigates a different area to provide a more in-depth look at the differences in characteristics between those interviewed in 1995 and those not participating.

The first graph examines educational attainment and shows that respondents who continued to participate as of 1995 had a slightly higher educational attainment in 1967 than those attriting; over 17.8% of participants had attended college, compared to only 13.9% of attriters.

Overall, the typical participant spent about a quarter of a year more in school than the typical respondent who left the survey.⁴

Figure 1. Educational Attainment in 1967

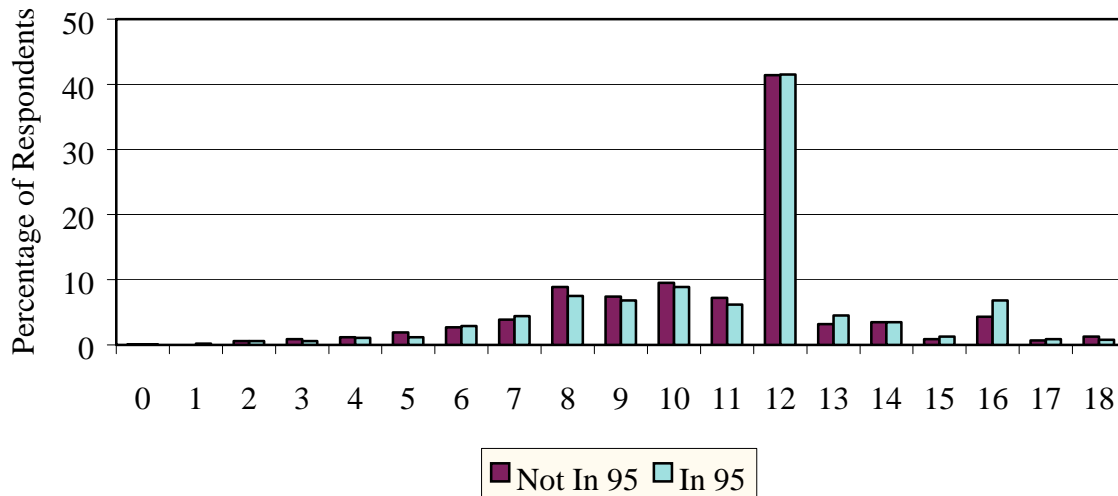


Figure 2 examines the labor force status of both groups in 1967. Individuals who left the survey had slightly higher levels of labor force participation in 1967 than those who remained. Attriters had a slightly higher rate for both employment (50.9%) and unemployment (3.4%) than participants (employment 48.8% and 2.5% unemployment). Therefore, attriters have a slightly lower incidence of not being in the labor market at the time of the original survey (45.7% vs. 48.7%).⁵

⁴ Educational differences between the two groups are statistically significant at both the 5% and 1% level.

⁵ Labor market differences between the two groups are statistically significant at only the 10% level.

Figure 2. Labor Force Status in 1967

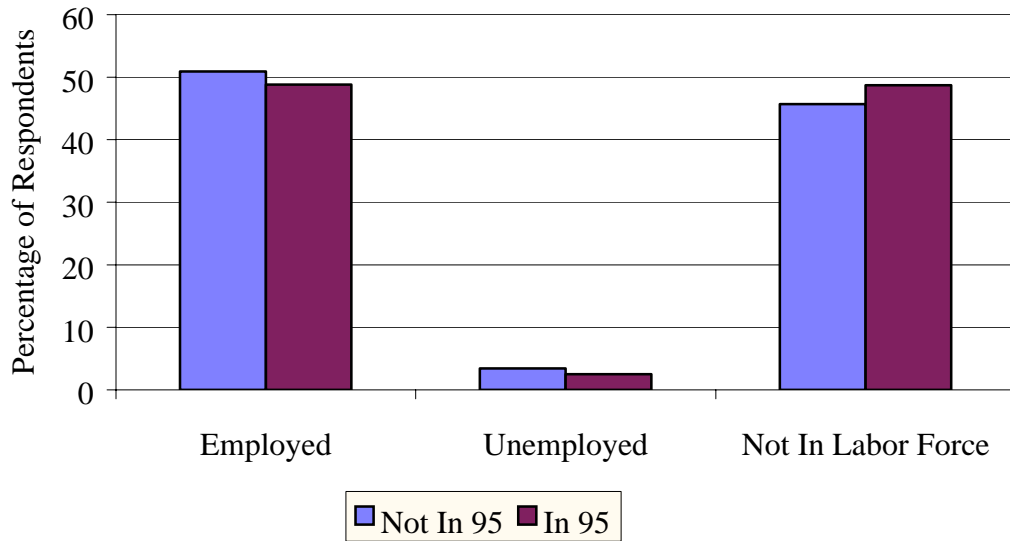


Figure 3 examines the total family income of respondents during 1966. This graph shows that participants in 1995 and their families earned slightly more money in 1966 than current nonparticipants. Overall, the family of the typical attriter earned \$8,091 while the family of the typical participant earned \$8,456.⁶ These differences are also seen in the various income categories of the graph. A larger percentage of attriters earned between \$0 and \$5,000; above \$5,000, however, the percentage of participants consistently outpaces attriters. Readers should note that income data in this table were rounded to the nearest figure used in the chart. For example, \$2,500 actually represents the range from \$1,251 to \$3,750.

⁶ Income differences between the two groups are statistically significant at the 5% level but not at the 1% level.

Figure 3. Total Income in 1966

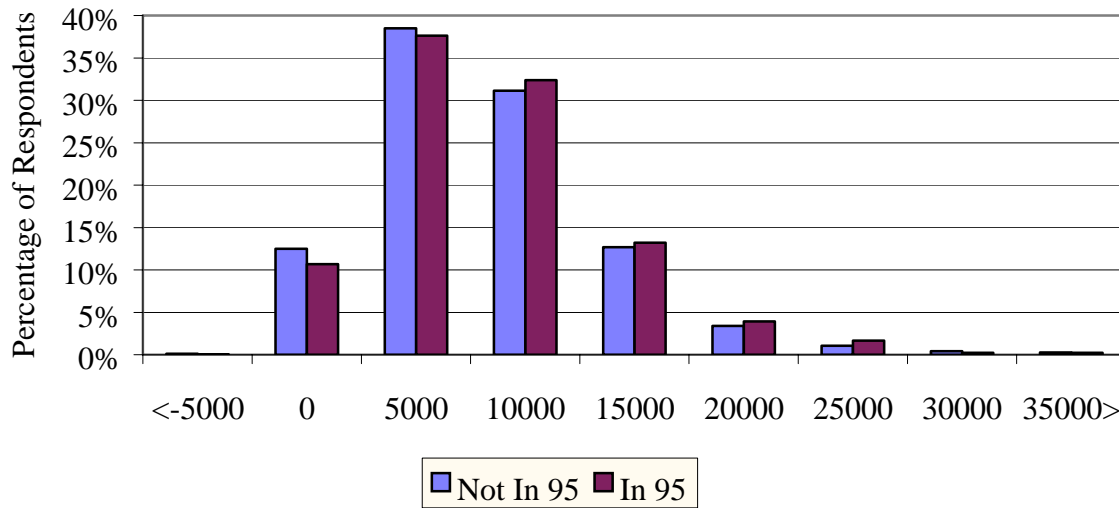


Figure 4 examines participation based on the type of area in which respondents lived during 1967. This graph shows that central city dwellers in 1967 are much less likely to participate in 1995 than those who lived in rural or non-urban areas. Central city dwellers comprise 41.5% of the nonparticipants but only 33.8% of participants. Rural dwellers in 1967, those who did not live in an SMSA, comprise 24.8% of the nonparticipants but 30.4% of those continuing to be interviewed.

Figure 4. Live in SMSA in 1967

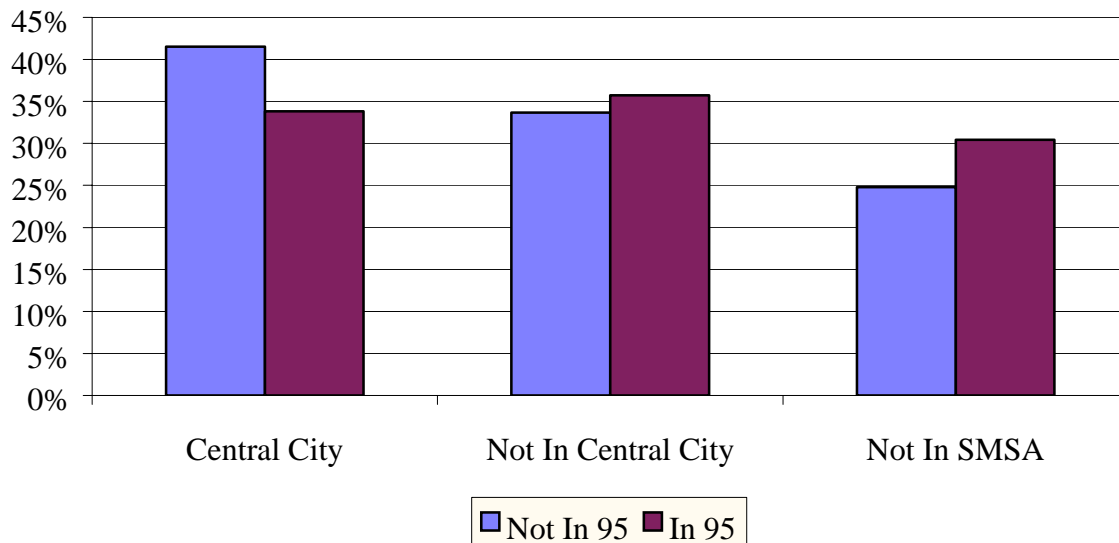
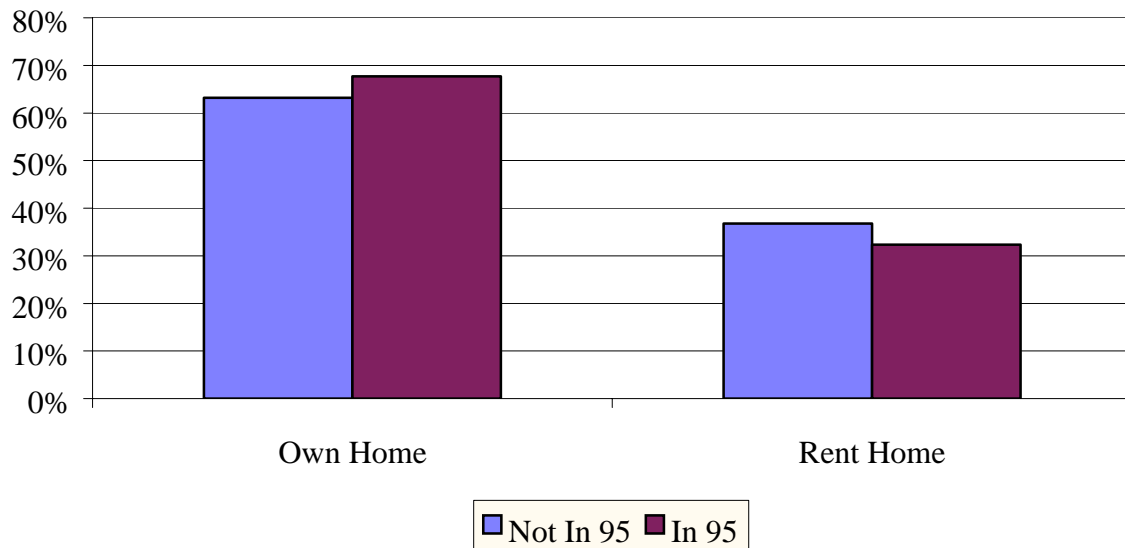


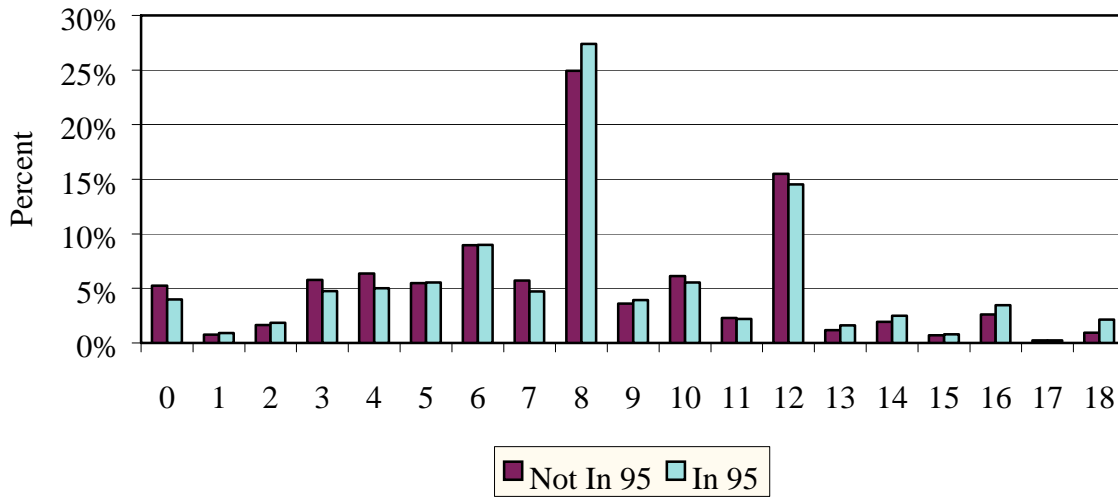
Figure 5 examines whether participants come from families that owned their home in 1967 or whether they were renters at that time. Overall, the graph suggests that renters in 1967 have slightly lower participation rates in 1995, while individuals who lived in an owned home have slightly higher rates. This graph provides additional evidence that attriters tend to be individuals with a lower socio-economic standing.

Figure 5. Home Ownership or Rental Status in 1967



The last figure, number 6, examines the highest grade each respondent's father attended. The data show that the typical father completed only the eighth grade. This graph, like the one showing the respondent's educational attainment, reveals that higher educational attainment is associated with lower attrition.

Figure 6. Father's Highest Grade Attended as of 1967



Overall, this section shows that Mature Women participating in the 1995 survey lived in slightly more affluent families in the 1960s, had a slightly higher educational level in the 1960s and worked less in the labor market than those who did not participate. Moreover, these participants more likely lived in rural areas during the 1960s and owned their homes. Additionally, current participants had fathers who were more educated than nonparticipants. While uncontrolled for possible racial differences, this limited evidence suggests that individuals with higher socio-economic standing at the start of the survey have continued with the survey at a greater rate than those with lower socio-economic standing.

VII. Time Series Comparison of NLS and CPS Data

The previous sections have shown that 46.7% of the Mature Women respondents had left the survey by 1995. This section investigates the effect of this attrition by examining how well the sample matches other national survey data from 1967, when the NLS began, and from 1995, the most recent year of available data. Overall this section finds that, while a large number of individuals have left, the sample still closely matches its portion of the nation.

Prior research has not documented a precise numerical measure of survey changes due to attrition. Since a standard measure has not been defined, this section outlines the theory and general form for measuring changes caused by attrition and then uses this measure to calculate the change over time for the Mature Women.

Our measure of change compares the NLS survey against matching national information. Measuring the difference between the survey and another national survey data set makes possible computation of the amount of change. For example, comparing the number of children NLS Mature Women have given birth to against national birth records would indicate the quality of the sample's match to national information.⁷

While this study uses the CPS as a comparison source of information, like the NLS of Mature Women, the CPS is also a survey and subject to bias. Nevertheless, except for the decennial census, which was not conducted in either 1967 or 1995, there is no comparable source of national information. The CPS surveys an extremely large number of households each month and continuously refreshes the sample to ensure it presents an accurate picture of the current U.S. population, and CPS staff perform a broad range of procedures designed to ensure consistency and data quality.

In one important respect, immigration, the CPS does not directly match the NLS. The NLS is a panel study, while the CPS is primarily a cross-sectional study. This means that if large groups of immigrants enter the U.S. the closed NLS sample will not pick up these new individuals and their economic and demographic characteristics. The CPS, however, continuously refreshes its sample each month. To more closely match the NLS, all CPS respondents who stated they

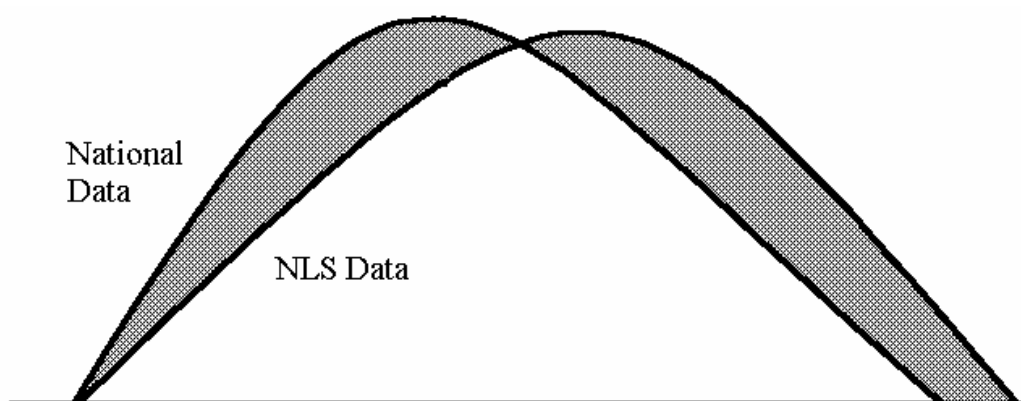
⁷ The implicit assumption is that other national survey data accurately measure the variable in question.

arrived in the U.S. after the NLS sample was drawn in the mid-1960s are dropped from this analysis. This ensures that the effects of immigration are removed from both samples.

CPS information used below was taken from the March 1967 and March 1995 surveys. Only females whose ages matched the NLS of Mature Women were selected from the data. Each respondent's answers are weighted using the weights provided with the respective data set. Beyond combining other non-black race/ethnic groups with whites and removing immigrants, this research did not subject CPS data to any special cleaning or processing.

The concept of comparing the two data sets is shown graphically in Figure 7. In this figure, the hypothetical difference between the national data set and the NLS of Mature Women is shown by the shaded area. If such a figure was used to graph number of children, for example, the highest point on the national data curve could represent the average (modal) number of children for all women in the United States, while the corresponding point on the NLS curve would represent the average (modal) number of children of Mature Women. When the two curves are a better match the difference tends to zero. Conversely, when the two curves do not represent exactly the same underlying information, there is a divergence and the difference is measurable.

Figure 7. Measure of Difference Diagram



A formula for calculating the difference in the shaded areas is

$$(2) \quad D_f = \left[\sum_{c=1}^m (N_{fc} - S_{fc})^2 \right]^{1/2}$$

In this equation, D stands for the difference term for a given factor, f. This term is calculated by subtracting the percentage of individuals in category, c, for survey data, S, from the corresponding percentage of individuals in that category in the national data, N. To ensure that negative and positive differences do not offset each other, the differences are first squared to remove the sign effects; at the end, the square root is taken to restore the answer to its correct magnitude.

To make these equations useful, this research uses the CPS as its national data set. While the CPS is not perfect, it is both easily available and widely used. Below is an example of the equations used to calculate a difference term for weeks worked.

Table 7A below shows the percentage employed in 1967, while Table 7B shows the percentage in 1995. The percentage employed is the factor, f, under consideration. National averages, N, are represented by data for the Current Population Survey, while NLS data is the survey, S. The first column, labeled “weeks worked,” contains 7 categories, c.

Two difference terms were calculated, one for white women and one for black women.⁸ In all tables white women are separated from black. This is necessary since the NLS of Mature Women contains an oversample of black women. The effects of this oversample are eliminated both by using the sampling weights provided with the data set and by separating the races.

The difference term for each of the racial groups was calculated by comparing the NLS values and CPS values for each category. These differences were squared and summed. Then the

⁸ Other races were combined with whites, as is typical of previous NLS original cohort research (e.g., Parnes, Fleisher, Miljus, and Spitz 1968).

square root of the summed series was computed. For 1967, this calculation results in a 2.89% difference for white women and a 5.63% difference for black women. Repeating this procedure with 1995 data yields a difference of 10.42% for white women and 9.39% for black women.

Table 7A. Percentage Employed by Weeks Worked in the Last Year, 1967

<i>Weeks worked</i>	<i>NLS white women</i>	<i>CPS white women</i>	<i>NLS black women</i>	<i>CPS black women</i>
0	46.04%	48.13%	27.55%	32.38%
1 to 13	6.48%	7.34%	6.53%	7.17%
14 to 26	6.27%	5.95%	9.53%	8.35%
27 to 39	6.41%	4.94%	7.35%	6.97%
40 to 47	4.66%	3.90%	6.29%	5.91%
48 to 49	2.14%	1.54%	3.17%	1.54%
50 to 52	27.99%	28.21%	39.58%	37.69%

Table 7B. Percentage Employed by Weeks Worked, 1995

<i>Weeks worked</i>	<i>NLS white women</i>	<i>CPS white women</i>	<i>NLS black women</i>	<i>CPS black women</i>
0	57.66%	66.48%	59.42%	67.13%
1 to 13	4.04%	3.27%	5.79%	1.61%
14 to 26	3.90%	3.23%	5.04%	1.84%
27 to 39	4.15%	2.51%	2.84%	3.03%
40 to 47	3.09%	2.14%	1.53%	1.40%
48 to 49	0.53%	0.85%	1.27%	1.74%
50 to 52	26.62%	21.52%	24.11%	23.26%

Note: Weeks worked for the NLS is calculated since the last survey, which for most respondents is over a two-year interval. CPS data are based on weeks worked in the last year.

The second step in calculating the overall attrition difference is to subtract the base year difference from the final year difference. Subtracting the base year eliminates the original baseline differences in the sample and shows how much the sample has changed over time for this particular factor. This change is represented by

$$(3) \quad D_{f \text{ diff}} = D_{f \text{ Base}} - D_{f \text{ final}}$$

Again using the data from Tables 7A and 7B as an example, we can calculate the change for weeks worked over time. The 1967 $D_{f \text{ Base}}$ for NLS white women was 2.89%. This number, calculated

above, represents the percentage that NLS data deviate from the corresponding national information. In 1995, $D_{f \text{ final}}$ was different from the national information by 10.42%, and $D_{f \text{ diff}}$ shows a change of 7.53%. It is important to note that the two surveys' employment measures are not perfectly comparable. The CPS directly asks respondents how many weeks they worked in the previous year; the NLS calculates the amount respondents worked since the last survey (two years or more) based on answers to a work history questionnaire.

Since the survey of Mature Women covers a broad range of topics, the difference measure calculated in equation (3) needs broadening. Equation (4) expands the attrition difference calculation by including multiple factors.

$$(4) \quad C_{\text{total}} = \sum_{f=1}^N W_i * D_{f \text{ diff}}$$

In this equation the total change, C_{total} , is the weighted sum of the component differences, which are computed individually using equation (3). This research computes C_{total} for the NLS of Mature Women using five key factors: employment status, marital status, education, household size, and income. The actual data and calculations are discussed below.⁹

While a variety of weighting schemes are possible, no information gathered during this research suggests that one of these five factors is more important or more prone to change than any of the others. Hence, instead of using different weights for particular factors, this research gives each of the five factors an identical weight of 0.20.

Computing equation (4) for white women shows that the total difference increased by 2.03 percentage points from 1967 to 1995. Computing the same measure for black women shows the difference increased by a slightly larger 2.97 percentage points. This level of change clearly

⁹ These factors represent all the major areas of overlap between the two surveys.

shows that losing respondents has had a small but visible effect on the ability of the NLS of Mature Women to match its part of the nation's demographic structure.

Details on Computing the Difference Terms. The rest of this section contains the detailed tables used to calculate equations (2), (3) and (4) for the various factors in the study. Table 8 examines marital status, Table 9 household size, Table 10 education and Table 11 wage and salary income. The total difference for each of the five factors is computed in Table 12.

Tables 8A and 8B examine the current marital status of NLS and CPS women. Although the NLS and CPS both allow finer breakdowns of marital status categories, for reasons of sample size we have reduced the number of groups to three. While overall from 1967 to 1995 white NLS women have moved closer to national averages (change -0.03%), black NLS women have tended to move away (change +3.46%).

Table 8A. Current Marital Status, 1967

<i>Marital status</i>	<i>NLS white women</i>	<i>CPS white women</i>	<i>NLS black women</i>	<i>CPS black women</i>
Married	87.15%	87.88%	65.44%	68.91%
Sep/widow/div	8.03%	7.84%	26.92%	26.14%
Never married	4.82%	4.29%	7.64%	4.95%

Table 8B. Current Marital Status, 1995

<i>Marital status</i>	<i>NLS white women</i>	<i>CPS white women</i>	<i>NLS black women</i>	<i>CPS black women</i>
Married	62.49%	62.99%	34.26%	38.69%
Sep/widow/div	34.28%	33.58%	59.66%	53.37%
Never married	3.23%	3.43%	6.08%	7.94%

Tables 9A and 9B investigate the number of residents in the households of NLS and CPS respondents. Over time, white NLS females have moved closer to national averages (change -0.45%), while black females have moved only slightly farther away (change +0.34%).

In addition to measuring household size, this research would ideally include total fertility, as the number of children born to respondents is an important social indicator. Unfortunately, the March CPS data do not provide a reliable fertility measure because the survey only records children currently residing in the household.¹⁰ Since no data are gathered regarding children who reside elsewhere or who have died, CPS data in this supplemental survey significantly underrepresent actual fertility and do not match NLS information.

Table 9A. Size of Household, 1967

<i>Number of residents</i>	<i>NLS white women</i>	<i>CPS white women</i>	<i>NLS black women</i>	<i>CPS black women</i>
1	2.97%	1.55%	5.83%	4.17%
2	9.74%	9.38%	10.89%	15.63%
3	14.71%	15.54%	13.96%	12.20%
4	26.03%	25.37%	16.33%	15.16%
5	21.26%	20.92%	13.55%	15.41%
6	13.63%	13.18%	11.68%	11.65%
7	6.10%	7.42%	8.28%	8.60%
8	2.63%	2.94%	7.33%	5.71%
9	1.50%	1.45%	4.67%	3.67%
10	0.72%	0.80%	2.91%	3.14%
10+	0.70%	1.45%	4.59%	4.66%

¹⁰ The authors did not have access to the CPS May fertility supplemental files for this time period.

Table 9B. Size of Household, 1995

<i>Number of residents</i>	<i>NLS white women</i>	<i>CPS white women</i>	<i>NLS black women</i>	<i>CPS black women</i>
1	26.51%	25.73%	34.95%	29.55%
2	56.81%	58.39%	32.46%	34.50%
3	11.28%	10.38%	16.18%	18.34%
4	3.02%	2.89%	8.26%	7.81%
5	1.48%	1.47%	3.66%	5.28%
6	0.48%	0.76%	2.53%	2.08%
7	0.26%	0.30%	0.86%	0.92%
8	0.16%	0.05%	0.56%	0.58%
9	0.00%	0.03%	0.35%	0.79%
10	0.00%	0.00%	0.19%	0.11%
10+	0.00%	0.01%	0.00%	0.03%

Tables 10A and 10B track education. During the 1960s the CPS asked respondents to report their highest grade completed. Beginning in the early 1990s, the CPS switched from the educational completion question to one focusing on attainment. The new question asked respondents about the highest degree or educational level they had ever attained. While this enables researchers to distinguish, for example, whether a respondent with four years of college received a degree, it makes time series comparisons difficult. Since education is a key area of interest, a special computer program was written to compute the highest degree or grade attained by NLS respondents as of 1995, enabling NLS data to be matched to the CPS. Over time, both white (change +1.49%) and black (change +1.26%) NLS females have diverged somewhat from national averages.

Table 10A. Highest Grade Completed, 1967

<i>Grade</i>	<i>NLS white women</i>	<i>CPS white women</i>	<i>NLS black women</i>	<i>CPS black women</i>
0	0.05%	0.46%	0.10%	0.18%
1	0.06%	0.10%	0.37%	0.13%
2	0.28%	0.23%	0.85%	0.68%
3	0.45%	0.47%	1.35%	1.49%
4	0.73%	0.78%	1.48%	1.04%
5	0.98%	0.85%	2.41%	2.12%
6	1.44%	1.75%	5.34%	5.26%
7	2.21%	2.20%	7.50%	6.24%
8	6.86%	7.09%	10.34%	9.30%
9	5.81%	6.14%	9.94%	8.64%
10	8.10%	7.70%	11.09%	10.67%
11	5.14%	5.67%	10.72%	9.06%
12	48.58%	47.79%	27.66%	30.66%
13	4.80%	4.11%	2.72%	2.97%
14	4.20%	4.57%	1.97%	3.41%
15	1.47%	1.70%	0.24%	2.22%
16	7.06%	6.54%	3.34%	3.73%
17	0.86%	0.88%	0.72%	1.34%
18	0.92%	0.97%	1.85%	0.84%

Table 10B. Highest Grade Completed or Degree Received, 1995

<i>Grade or degree</i>	<i>NLS white women</i>	<i>CPS white women</i>	<i>NLS black women</i>	<i>CPS black women</i>
<1	0.17%	0.45%	0.00%	1.02%
1-4	1.19%	1.23%	2.88%	3.54%
5-6	1.72%	1.72%	4.35%	3.67%
7-8	6.17%	6.55%	13.30%	12.20%
9	3.91%	3.40%	7.45%	6.59%
10	5.08%	5.72%	7.73%	9.04%
11	3.34%	3.70%	7.68%	9.13%
12	2.54%	1.01%	6.45%	1.54%
H.S. diploma	45.05%	43.65%	29.31%	31.49%
Some college	15.18%	15.30%	10.75%	11.15%
Associate	2.25%	3.19%	1.19%	0.99%
Bachelor's	1.17%	2.17%	0.68%	0.88%
Master's	8.98%	7.78%	4.02%	5.18%
Prof. degree	2.92%	3.71%	3.48%	3.17%
Doctorate	0.32%	0.42%	0.72%	0.42%

Tables 11A and 11B examine total wage and salary income earned by NLS and CPS women over the past calendar year. This income does not include money earned by spouses or partners at their jobs, only money earned by the respondent. Over time, NLS income data diverged slightly from the national averages for white women (change +1.59%) and significantly for black women (change +6.02%). One probable cause for this difference is that the CPS allows proxy respondents while the NLS does not. Since table 9B shows that significantly more CPS black women live in multi-person households than NLS black women, the likelihood of someone else estimating the wages and salary earned by a CPS black women is much higher. Differences between estimates by CPS proxy respondents and NLS self-reports potentially explain the gap in wage and salary income among black women.

Table 11A. Wage and Salary Income, 1967

<i>Yearly income</i>	<i>NLS white women</i>	<i>CPS white women</i>	<i>NLS black women</i>	<i>CPS black women</i>
\$0 to \$500	13.96%	14.85%	17.15%	16.24%
\$501 to \$1500	17.48%	17.31%	24.80%	22.68%
\$1501 to \$2500	14.19%	14.94%	18.92%	19.82%
\$2501 to \$3500	14.73%	16.32%	15.49%	17.57%
\$3501 to \$4500	12.78%	13.72%	7.25%	7.92%
\$4501 to \$5500	11.69%	11.00%	8.16%	7.37%
\$5501 to \$6500	7.11%	6.04%	3.16%	3.13%
\$6501 to \$7500	3.65%	2.98%	1.63%	3.22%
\$7501 to \$8500	1.93%	1.35%	1.64%	0.61%
\$8501 to \$9500	1.25%	0.51%	0.50%	0.48%
> \$9501	1.24%	0.98%	1.30%	0.96%

Table 11B. Wage and Salary Income, 1995

<i>Yearly income</i>	<i>NLS white women</i>	<i>CPS white women</i>	<i>NLS black women</i>	<i>CPS black women</i>
\$0 to \$2500	15.92%	14.77%	13.95%	11.46%
\$2501 to \$7500	21.21%	19.54%	21.83%	25.11%
\$7501 to \$12500	17.02%	18.10%	25.22%	18.86%
\$12501 to \$17500	12.95%	11.49%	8.69%	12.60%
\$17501 to \$22500	8.39%	10.64%	10.61%	9.64%
\$22501 to \$27500	6.75%	6.83%	8.09%	5.02%
\$27501 to \$32500	4.59%	6.93%	3.59%	7.26%
\$32501 to \$37500	3.90%	3.12%	1.78%	2.44%
\$37501 to \$42500	2.21%	2.70%	2.29%	2.83%
\$42501 to \$47500	1.75%	1.66%	0.00%	0.32%
\$47501 to \$52500	2.15%	1.93%	0.00%	0.85%
\$52501 to \$57500	0.64%	0.40%	0.49%	0.57%
> \$57501	2.51%	1.88%	3.46%	3.04%

Table 12 summarizes the differences between the NLS and national data found in the preceding tables. Each line of the table calculates equation (4) for a particular factor. Equation (5), which measures the total change, is computed by doing a weighted sum of the columns marked “Change.”

Table 12. Percentage Difference in Key NLS Measures in 1967 and 1995

<i>Variable</i>	<i>1967 white</i>	<i>1995 white</i>	<i>Change</i>	<i>1967 black</i>	<i>1995 black</i>	<i>Change</i>
Employment	2.89%	10.42%	7.53%	5.63%	9.39%	3.76%
Marital status	0.92%	0.89%	-0.03%	4.46%	7.92%	3.46%
Education	1.56%	3.05%	1.49%	4.93%	6.19%	1.26%
Household size	2.45%	2.00%	-0.45%	6.08%	6.42%	0.34%
Income	2.80%	4.39%	1.59%	3.90%	9.92%	6.02%

Overall, from 1967 to 1995 the total weighted change for NLS white women compared to CPS white women of the same age is an increase of 2.03 percentage points. Computing the same measure for black women shows the difference increased by a slightly larger 2.97 percentage points. Even with a relatively large amount of attrition, the NLS of Mature Women still closely matches similar CPS individuals.

VIII. Conclusions

This research has investigated attrition in the NLS of Mature Women. Attrition is a concern because almost 47% of the original respondents have left the survey. The section on reasons for noninterviews suggests a few steps that may slightly increase participation. First, one area that has the potential for increasing participation rates by up to 8% is to move all individuals dropped from the survey because they missed more than two years back onto the interview lists.

Second, very little information is available about reasons why respondents refuse to participate. Future surveys should include a mini-questionnaire for interviewers to fill out that pinpoints the specific reasons why a respondent is refusing. This information may help us to improve response rates in the future.

This research showed that even though 2,372 respondents no longer participate, continuing survey members still provide information that is similar to individuals in the Current Population Survey (CPS). Comparing the NLS with national CPS information shows that an attrition difference measure increased 2.03 percentage points for white women from 1967 to 1995. Computing the same measure for black women shows their difference increased by a slightly larger 2.97 percentage points. This suggests that, while attrition has somewhat reduced the ability of the NLS of Mature Women to match other national data, over time this survey still is a good representation of the cohort of women ages 30–44 in 1967.

We have measured the differences between the NLS of Mature Women data and the CPS for the same age cohort and compared those differences over time. While the differences may well indicate the effects of attrition, they may reflect other factors as well. For example, the NLS uses fewer proxy reports than the CPS. The CPS uses a single household informant to collect data on everyone in the household. While this methodological difference has always existed, in the

1990s the relative frequency with which the household informant could have been a person age-eligible for the NLS was likely much higher than in the 1960s. This is because, with the passage of time, household size has gotten much smaller (see tables 9A and 9B). In turn, this should have caused the relative differences to shrink.

On the other hand, the NLS respondents are much better “trained” by decades of participation and have established a rapport with interviewers. Over time, CPS and NLS instruments may have become more different, such as in question wording, use of telephone versus face-to-face interviewing, or in top-coding of income. Each of these factors could contribute to a divergence, inasmuch as the first round of the NLS was likely more similar to an incoming rotation group of the CPS than is the case after thirty years of interviewing NLS respondents.

Thus, we need to be cautious in assigning a clear cause to the changes in the difference measures developed here. Doubtless attrition, especially differential attrition by race, is responsible for many of the distributional discrepancies, but other factors may also be at work to either increase or decrease the differentials. As is always the case, we can assess differences between surveys, but without external validation and exact specification of the concept to be measured we cannot judge which survey estimates are closer to the population. Nevertheless, since attrition is not causing large biases, continued fielding of the survey will provide policy makers and researchers with important information on this aging portion of the U.S. population.

References

- Juster, F. Thomas and Kathleen A. Kuester (1991) "Difference in the Measurement of Wealth, Wealth Inequality and Wealth Composition Obtained from Alternative U.S. Wealth Surveys." *Review of Income and Wealth*, Series 37, Number 1, March, pp. 33-62.
- MaCurdy, Thomas, Thomas Mroz, and R. Mark Gritz (1998) "An Evaluation of the National Longitudinal Survey of Youth." *Journal of Human Resources*, Vol. 33, No. 2., pp. 346-436.
- O'Neill, June E. (1982) *Review of the National Longitudinal Surveys*, Office of Research and Evaluation, Employment and Training Administration, U.S. Department of Labor.
- Parnes, Herbert S., Belton Fleisher, Robert Miljus, and Ruth Spitz (1968) *The Pre-Retirement Years*, Vol. I, Center for Human Resource Research, The Ohio State University.
- Parsons, Donald O. (1987) *Summary Tables: The National Longitudinal Surveys of Mature Males 1966-1981*. Center for Human Resource Research, The Ohio State University.
- Rhoton, Patricia (1984) *Attrition and the National Longitudinal Surveys of Labor Market Experience: Avoidance, Control and Correction*, Center for Human Resource Research, The Ohio State University.
- Rhoton, Patricia and Karima Nagi (1989) *Attrition and Wealth in the Original Cohorts*, Center for Human Resource Research, The Ohio State University.
- Zagorsky, Jay L. (1997) *The NLSY79 Wealth Data Evaluation*, Center for Human Resource Research, The Ohio State University