Patterns of Item Nonresponse for Income in the National Survey on Drug Use and Health

Peter Frechtel,[1] Glynis Ewing,[1] Kristen Brown,[1] Jonaki Bose,[2] Rebecca Ahrnsbrak[2]

[1] RTI International (a registered trademark and a trade name of Research Triangle Institute), Research Triangle Park, NC, USA
[2] Substance Abuse and Mental Health Services Administration, Rockville, MD, USA

1. Introduction

Conducted annually, the National Survey on Drug Use and Health (NSDUH) is the primary source of information on alcohol use, tobacco use, illicit drug use, substance use disorders, and mental health issues for the U.S. civilian, noninstitutionalized population aged 12 years or older. In the 2014 NSDUH, this population included residents of noninstitutional group quarters (e.g., shelters, rooming houses, dormitories, and group homes) and civilians residing on military bases. The target population excluded people with no fixed household address (e.g., homeless or transient people not in shelters), residents of institutional group quarters (e.g., jails and hospitals), children younger than 12, and active military personnel. As it has since 1999, the 2014 NSDUH utilized a 50-state, multistage cluster design that enables the Substance Abuse and Mental Health Services Administration (SAMHSA) to provide representative estimates for each state and the District of Columbia. Both direct and model-based state and substate estimates are produced on a variety of measures based on a combination of multiple years of data.

Low response rates to income questions and resulting nonresponse bias are well documented in the survey research literature (Bollinger, Hirsch, Hokayem, & Ziliak, 2014; Tourangeau & Yan, 2007; Pleis & Dahlhamer, 2004; Moore, Stinson, & Welniak, Jr., 2000; Juster & Smith, 1997). Like many other household surveys, the family income variables measured in NSDUH have much lower response rates than the vast majority of other questionnaire items. Typically, approximately 90 percent of variables that underwent statistical imputation required less than 5 percent of their records to be logically assigned or statistically imputed (Center for Behavioral Health Statistics and Quality, 2015, Appendix A). In 2014, 6,589 cases were missing for finer categories of total family income resulting in a weighted nonresponse rate of 10.46 percent. This relatively high item nonresponse rate is of interest because the NSDUH family income variables and their recodes are used in many analyses, and the distribution of public health variables by income levels has implications for policy decisions. Any steps that could be taken to reduce the impact of item nonresponse, to improve the imputation method, or to improve the questionnaire would be helpful with reducing nonresponse bias and improving overall data quality. By improving the understanding of the mechanisms of item nonresponse for income, this appendix describes solutions and recommendations for further addressing and reducing item nonresponse and/or nonresponse bias in variables measuring total family income and poverty through better imputation methods or by making changes to the questionnaire to reduce item nonresponse.

2. Measurement of Total Income in NSDUH

NSDUH estimates respondents' total income for adults and youths aged 12 to 17 by asking about total personal income and total family income, based on two questions:

1. Of these income groups, which category best represents (your/SAMPLE MEMBER's) total personal income during [the previous calendar year]?
2. Of these income groups, which category best represents (your/SAMPLE MEMBER's) total combined family income during [the previous calendar year]?

Respondents receive these questions after being routed through an unfolding bracket of related income questions in order to minimize income item nonresponse as much as possible.

Family is defined as any related member in the household roster, including all foster relationships and unmarried partners (including same-sex partners). Roommates, boarders, and other nonrelatives are excluded from the definition of family for total family income. Responses from proxies are accepted for items of health insurance and

income from a family member living in the same household who is identified as being better able to give the correct information. The NSDUH questionnaire allows respondents to decline to answer any question (except age) by entering "Don't know" (DK) or "Refused" (REF) as a response.

Total family income is an important measure because it is used to establish a respondent's poverty level using the NSDUH data. Poverty level is determined by comparing a respondent's total family income with the U.S. Census Bureau's poverty thresholds (both measured in dollar amounts), with a respondent's family size and composition (i.e., number of children) taken into consideration. The resulting variables indicating levels of poverty are often used in NSDUH analyses. When total family income is missing, the poverty level is unknown.

2.1 Questionnaire Skip Logic

Total income is measured in NSDUH through an unfolding bracket of questionnaire items that can be understood as steps. These steps are illustrated in the flowchart in Figure 1. There are 29 finer categories of personal and family income that are captured by the variables PINC2 and FINC2, respectively. Table 1 shows the binary and finer categories of income.

Figure 1. 2014 NSDUH Questionnaire Measurement of Total Income

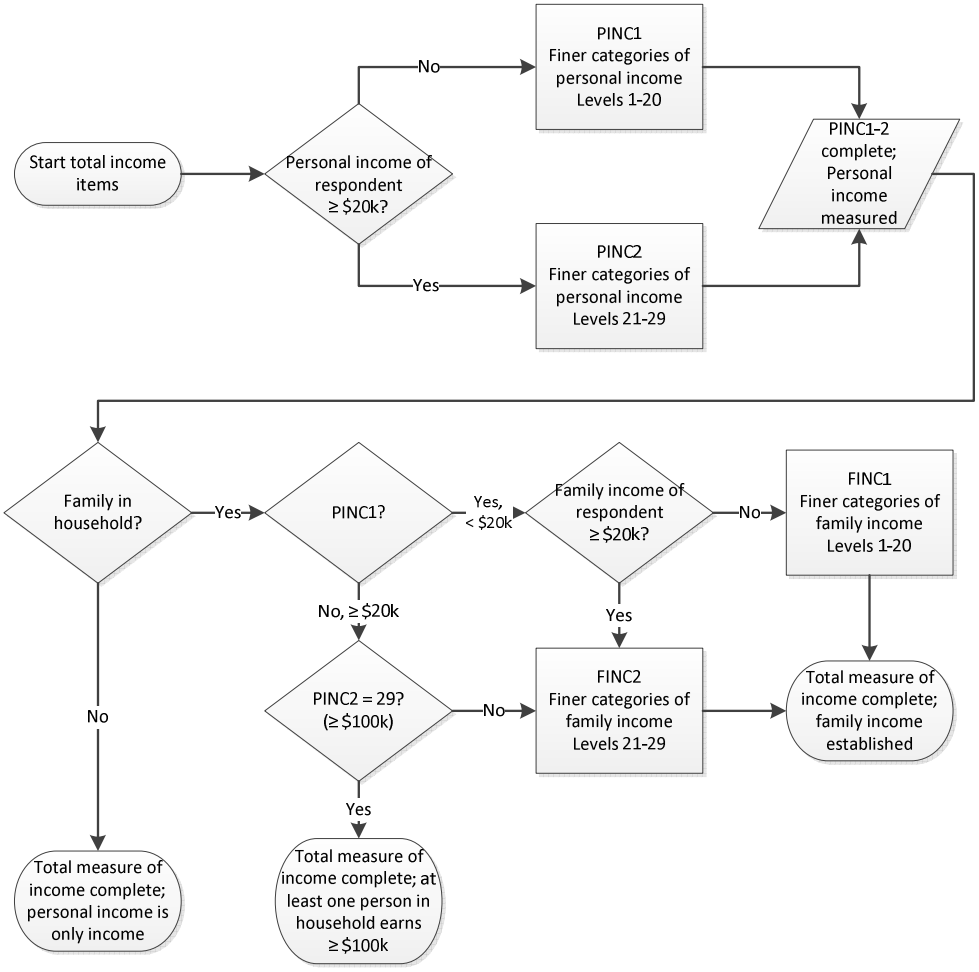


Table 1. Binary Categories of Personal and Family Total Income: 2014 NSDUH

Binary Category (PINC1, FINC1)		Finer Category (PINC2, FINC2)	
Level	**Description**	**Level**	**Description**
2	< $20,000	1	< $1,000
		2	$1,000-$1,999
		⋮	
		20	$19,000-$19,999
1	≥ $20,000	21	$20,000-$24,999
		⋮	
		26	$45,000-$49,999
		27	$50,000-$75,999
		28	$75,000-$99,999
		29	≥ $100,000[1]

[1] This final measurement category of income was measured in the 2014 and previous iterations of NSDUH. Beginning in 2015, this category became $100,000-$149,999 and a 30th measurement category of ≥ $150,000 was added.

The first step of the income questionnaire module establishes whether the personal income of the respondent is greater than or equal to $20,000 with a binary questionnaire item. This response is captured by PINC1. The next step in measuring total personal income is to determine the finer category of personal income for each respondent.

Respondents reporting personal incomes that are less than $20,000 are directed to the finer personal income categories with levels ranging from less than $1,000 (PINC2 = 1) to $19,000-$19,999 (PINC2 = 20). Respondents reporting personal incomes that are greater than or equal to $20,000 are directed to the finer personal income categories, ranging from $20,000-$24,999 (PINC2 = 21) to $100,000 or more (PINC2 = 29). This response is captured by PINC2.

After measuring finer categories of total personal income, respondents are routed to questionnaire items measuring total family income depending on the presence of family members in the household. This information is reported in the household roster. If the respondent did not report any family members in his or her household, then the questionnaire item on total family income is not asked and the personal income response (PINC2) is used for total family income for that respondent. If the respondent reported family members in the household and the personal income level that is reported in PINC1 is greater than or equal to $20,000, then the respondent is directly routed to the finer categories of family income question, FINC2, skipping the binary family income questionnaire item FINC1.

If the respondent reports the highest level of response available for personal income, $100,000 or more (PINC2 = 29), total family income is automatically completed as $100,000 or more (FINC2 = 29). If reported personal income is less than $20,000, the questionnaire then asks respondents whether the total family income is greater than or equal to $20,000 (FINC1). Based on this response, respondents are directed to one of the two questionnaire items measuring finer categories of family income (captured by FINC2). FAMINC2 is the final resulting variable measuring the finer categories of family income. It is equal to PINC2 for those with no other family members in the household roster, and it is equal to FINC2 for those with other family members in the household roster.

2.2 Paths to Income Item Nonresponse

Because of the skip logic and questionnaire routing, there are six ways, or "paths," in which the respondent could have a valid value for total family income, and there are nine opportunities for respondents to become defined as missing in the total family income measure (see Table 2 and Figure 2 for paths of nonresponse). The flowchart in

Figure 2 illustrates how and where these valid or missing values occur based on the structure of the income questionnaire items. The numbered missing nodes correspond with the paths of nonresponse shown in Table 2.

Table 2. Paths of Finer Categories of Family Income Item Nonresponse: 2014 NSDUH

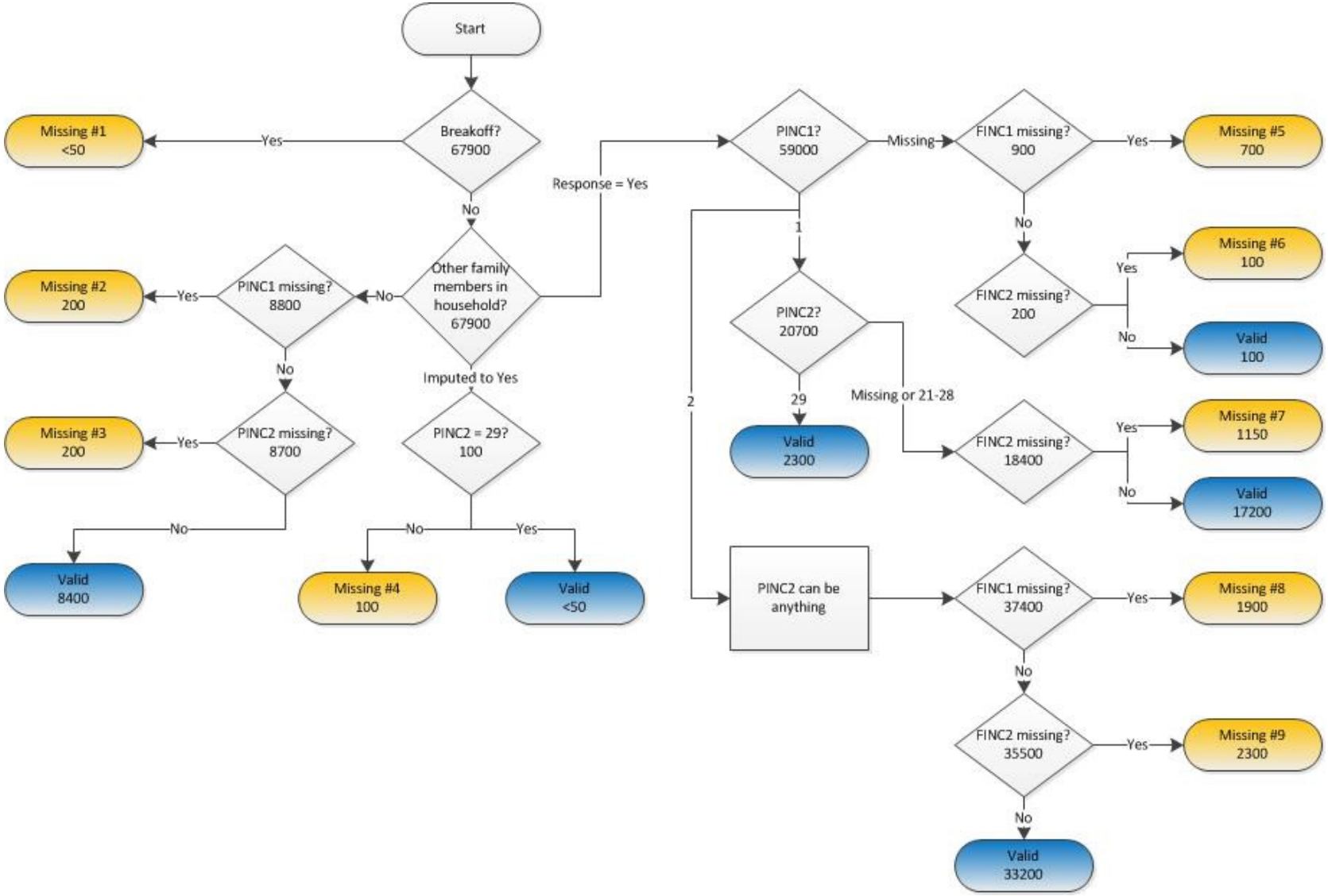
	Path of Income Item Nonresponse	Frequency	Percent	Rank by Frequency Missing
1	Breakoff	27	0.41	9
2	No Family in HH, PINC1 Missing	182	2.76	6
3	No Family in HH, PINC2 Missing	241	3.66	5
4	Imputed to Family in HH, PINC2 ≠ 29	56	0.85	7 (tie)
5	Family in HH, PINC1 Missing, FINC1 Missing	686	10.41	4
6	Family in HH, PINC1 Missing, FINC1 Valid, FINC2 Missing	56	0.85	7 (tie)
7	Family in HH, PINC1 = 1, PINC2 ≠ 29, FINC2 Missing	1,150	17.45	3
8	Family in HH, PINC1 = 2, FINC1 Missing	1,929	29.28	2
9	Family in HH, PINC1 = 2, FINC1 Valid, FINC2 Missing	2,262	34.33	1
Total Income Item Nonresponse		6,589	100.00	N/A

HH = household.

Source: SAMHSA, Center for Behavioral Health Statistics and Quality, National Survey on Drug Use and Health, 2014.

The two most common paths of income item nonresponse occur for sample members with personal incomes that are less than $20,000 living in households with family members who either (1) responded to the binary total family income questionnaire item but did not respond to the finer categories of total family income question (nonrespondents = 2,262) or (2) did not respond to the binary total family income questionnaire item and thus did not receive the finer categories of total family income question (nonrespondents = 1,929). The third most common path to item nonresponse for total family income occurs for sample members with personal incomes that are greater than $20,000 but less than $100,000 living in households with family members who do not answer the finer categories of family income question (nonrespondents = 1,150). This is followed by the fourth most common path of family income item nonresponse for respondents living with family members in the household who did not answer the binary personal or family income items at all (nonrespondents = 686). The next most common paths of family income item nonresponse occur for respondents without any family members living in the same household and either do not respond to the binary personal income questionnaire item (nonrespondents = 182) or respond to the binary personal income item and do not respond to the finer categories of personal income (nonrespondents = 241). The three remaining paths of family income item nonresponse are less common with fewer than 60 nonrespondents each.

Figure 2. Income Item Response and Nonresponse Paths with Frequencies: 2014 NSDUH



Source: SAMHSA, Center for Behavioral Health Statistics and Quality, National Survey on Drug Use and Health, 2014.

3. Potential Solutions and Recommendations

In NSDUH, income item nonresponse has historically been handled using imputation. Income is a unique measure because the observed nonresponse depends, in part, on the unknown values of the variable that are not actually observed, so income item nonresponse can be described as not missing at random (NMAR) (Bollinger et al., 2014; Pleis & Dahlhamer, 2004). Because income item nonresponse is NMAR, imputation involving only auxiliary variables does not completely correct for the present nonresponse bias and resulting measurement error (Little & Rubin 1987; Frechtel & Copello, 2007). However, NSDUH has certain survey features that can be used to more accurately decrease income item nonresponse. This section presents potential solutions and recommendations for the future based on reclaiming[1] missing income responses from family pair members and explores how further use of proxy responses can improve overall data quality.

3.1 Reclaiming with Other Family Pair Member

In each household selected for NSDUH, zero, one, or two household members are selected for interviewing. When two members of the same household are selected and both complete an interview, a "pair" is formed. Pairs are quite common. For example, in the 2014 NSDUH, 53.7 percent of the unit respondents were members of a pair. The pair relationship can be parent-child, sibling-sibling, spouse-spouse, or some other relationship.

The report on the evaluation of imputation methods for NSDUH (Center for Behavioral Health Statistics and Quality, in press) discusses the use of the other pair member as a donor in imputation when exactly one pair member is missing. In that report, family income is considered a good candidate for this sort of logical imputation whenever the pair members are members of the same family. Because most of the questions in the income module ask about the family income of those in the household, given the case where the pair members are members of the same family, the only source of disagreement should be measurement error (Frechtel, Scott, Couzens, Moore, & Bose, 2012). This section describes reclaiming values of family income that sometimes involves the data reported by the respondent and sometimes involves the data reported by the other family pair member (OFPM) about the family. The following two key recoded variables involving family income are considered:

1. INCOME5, a straightforward five-level recode of finer income; and
2. POVERTY2, a three-level recode of not only family income but also of roster information and type of household.

Although recodes on NSDUH typically do not have imputation indicators associated with them, it would be beneficial to add them because the imputation indicators associated with INCOME5 and POVERTY2 would differ from the imputation indicator for the family finer categories income variable IRFAMIN2, called IIFAMIN2.

3.1.1 Reclaiming Missing Values for INCOME5

Table 3 shows how the 29 levels of IRFAMIN2 map to the 5 levels of INCOME5. The highest three levels are the same for both variables, but the lower levels are heavily aggregated.

Table 3. Mapping of IRFAMIN2 Levels to INCOME5 Levels

IRFAMIN2		INCOME5	
Level	**Description**	**Level**	**Description**
1	< $1,000	1	< $20,000
2	[$1,000, $2,000)		
⋮	⋮		
20	[$19,000, $20,000)		
21	[$20,000, $25,000)	2	[$20,000, $50,000)
⋮	⋮		

[1] To "reclaim" in this context is to assign a value for a recode using logic when some or all of the parent variables for the recode have missing values. It might also be called "logical imputation."

IRFAMIN2		INCOME5	
Level	**Description**	**Level**	**Description**
26	[$45,000, $50,000)		
27	[$50,000, $75,000)	3	[$50,000, $75,000)
28	[$75,000, $100,000)	4	[$75,000, $100,000)
29	≥ $100,000	5	≥ $100,000

There are exactly 661 respondents in the 2014 data whose missing values for IRFAMIN2 can be reclaimed using only the self's data. For these types of respondents, it is known that INCOME5 = 1 even though IRFAMIN2 is missing. This can occur in two ways:

1. The respondent has no other family members in the household, reported having a personal income of less than $20,000 (PINC1 = 2), and had a missing value for the personal finer categories income variable (PINC2).
2. The respondent has other family members in the household, reported having a family income of less than $20,000 (FINC1 = 2), and had a missing value for the personal finer categories income variable (FINC2).

Using the self, 661 (10.03 percent) of the 6,589 missing values for IRFAMIN2 can be reclaimed. Of the remaining 5,928 cases with missing values, 1,267 (21.37 percent) cases had an OFPM with a nonmissing value for FAMINC2, and 60 (1.01 percent) other cases had an OFPM with family binary income of less than $20,000. In total, the reclaiming process can reduce the item nonresponse rate for INCOME5 from 9.70 percent to 6.78 percent (Table 4).

Table 4. Reclaiming of Missing Values of INCOME5

	Count	**Nonresponse for INCOME5 (%)**
Cases with Missing FAMINC2	6,589	9.70
Minus Cases with FAMINC1 = 2	661	8.73
Minus Cases Where OFPM Has Nonmissing FAMINC2	1,267	6.86
Minus Cases Where OFPM Has FAMINC1 = 2	60	6.78
Final Nonresponse Rate for INCOME5	**4,601**	**6.78**

3.1.2 Reclaiming Missing Values for POVERTY2

The variable POVERTY2 is a complex recode involving the imputation-revised family finer categories income variable IRFAMIN2, plus the respondent's age (AGE) and the roster variables IRFMLYSZ (imputation-revised household size including fosters) and IRKDFMLY (imputation-revised number of children in household including fosters). The variable is created in two steps. In the first step, the poverty threshold is calculated based on a formula from the U.S. Census Bureau involving values that are captured by the variables AGE, IRFMLYSZ (imputation-revised family size), and IRKDFMLY (imputation-revised number of children younger than 18 in the household).[2] In the second step, the family income variable IRFAMIN2 (actually the midpoint of the interval associated with the value of IRFAMIN2) is compared with the poverty threshold. POVERTY2 has four levels:

1. The respondent is 18 to 22 years old and lives in a college dorm (POVERTY2 = missing).
2. The family income is less than the poverty threshold (POVERTY2 = 1).
3. The family income is greater than or equal to the poverty threshold but less than twice the poverty threshold (POVERTY2 = 2).
4. The family income is greater than twice the poverty threshold (POVERTY2 = 3).

Note that the pre-imputation versions of IRFMLYSZ and IRKDFMLY have missing values as well, but the item nonresponse rates for these (less than 1 percent) are much smaller in comparison with the item nonresponse rate associated with the pre-imputation version of IRFAMIN2 (9.70 percent weighted, 10.46 percent unweighted).

[2] See https://www.census.gov/hhes/www/poverty/data/threshld/.

Because of the relatively low item nonresponse rates associated with these household size variables, it is assumed that there are no missing data for the two household composition variables, and reclaiming efforts for POVERTY2 are focused on obtaining valid responses from income variables for this investigation.

In total, 1,964 cases with missing values for FAMINC2 can be reclaimed if an imputation indicator were created for POVERTY2. The first step in reclaiming is to calculate bounds for the family income based on both the person's responses and on the OFPM's responses. The possible lower bounds for FAMINC2 are the following:

- (L1) 21, if PINC1 = 1;
- (L2) PINC2, if $2 \leq$ PINC2 ≤ 28; and
- (L3) 21, if FINC1 = 1.

The possible upper bounds for FAMINC2 are the following:

- (U1) 20, if PINC1 = 2, PINC2 is missing, and the person lives with no other family members; and
- (U2) 20, if FINC1 = 2.

All of these bounds were calculated for both the self and the OFPM, except U1, which was only calculated for the self. If the self has no other family members in the household, then it is assumed that the OFPM's data would not be available.

Table 5 shows the reclaiming of missing values for POVERTY2 relative to FAMINC2. Probably the most interesting category is the 612 cases that can be reclaimed using the self data. Of the 612 cases, 277 (45.26 percent) can be reclaimed because the family income is definitely greater than twice the poverty threshold, 320 (52.29 percent) can be reclaimed because the family income is definitely less than the poverty threshold, and 15 (2.45 percent) can be reclaimed because the family income is definitely greater than the poverty threshold and less than twice the poverty threshold. Overall, the reclaiming process can reduce the item nonresponse rate for POVERTY2 from 9.70 percent to 6.81 percent.

Table 5. Reclaiming of Missing Values of POVERTY2

	Count	Nonresponse for INCOME5 (%)
Cases with Missing FAMINC2	6,589	9.70
Minus College Students in Dorms	8	9.69
Minus Cases Where POVERTY2 Value Can Be Determined Based on the Self Data	612	8.79
Minus Cases Where OFPM Has Nonmissing FAMINC2	1,288	6.89
Minus Cases Where POVERTY2 Value Can Be Determined Using Bounds from the OFPM Data	56	6.81
Final Nonresponse Rate for POVERTY2	**4,625**	**6.81**

3.1.3 Next Steps for Reclaiming Income with Other Family Pair Member

One problem with editing income using the OFPM data is that a direct assignment can create an inconsistent record. There are 130 cases in the 2014 data where a direct assignment of the OFPM's value for FAMINC2 would create an inconsistent record. These are mostly cases where the self has FINC1 = 1 and the OFPM has FINC1 = 2, or the self has FINC1 = 2 and the OFPM has FINC1 = 1. There are also 24 cases in the 2014 data where the OFPM has a missing value for FAMINC2, but the bounds based on the OFPM's responses are inconsistent with bounds based on the self's responses. For example, the self might have FINC1 = 2, the OFPM might have FINC1 = 1, and both pair members might have FINC2 missing.

The direct-assignment approach rests on an implicit assumption that these pair members agree when they both respond. This is not always the case. When both family pair members responded in the 2014 survey, they agreed 68.64 percent of the time, suggesting that there is non-negligible measurement error present (Table 6). Sometimes the responses of the family pair members are quite different. For example, they disagreed by greater than four or more levels of FAMINC2 8.59 percent of the time.

Table 6. Disagreement in Family Income among Family Pair Members: 2014 NSDUH

Number of Income Levels between Paired Responses	Number of Family Pairs	Percentage of Family Pairs
0	9,441	68.64
1	2,013	14.63
2	754	5.48
3	366	2.66
4+	1,181	8.59

Source: SAMHSA, Center for Behavioral Health Statistics and Quality, National Survey on Drug Use and Health, 2014.

The response agreement between the responding family pair members across pair types shows that the agreement rates for grandparent-grandchild pairs (56.99 percent) and spouse-spouse pairs (57.36 percent) are lowest, and the agreement rates among sibling-sibling pairs where one of the pair members is aged 12 to 14 are highest (84.51 percent) (Table 7). The low agreement rate for spouse-spouse pairs may suggest measurement error associated with a lack of knowledge. The relatively high agreement rates among parent-child pairs where one of the pair members is aged 12 to 14 (77.66 percent) or 15 to 17 (77.65 percent) may be a result of the higher number of proxy responses for the family income item in this age group, and it is possible that the parent pair member gave responses as a proxy for the child pair member.

Table 7. Agreement by Pair Type among Responding Pairs for Family Finer Categories Income: 2014 NSDUH

Pair Type	Responding Pairs		Percent Agreement among Pairs
	Number	Percent	
Parent-Child, Child Aged 12-14	2,596	41.63	77.66
Parent- Child, Child Aged 15-17	2,139	34.40	77.65
Parent- Child, Child Aged 18-20	711	11.40	59.92
Parent- Child, Child Aged 21+	790	12.67	56.33
Parent-Child Total	**6,236**	**100.00**	**72.93**
Sibling-Sibling, Youngest 12-14/Oldest 15-17	1,291	31.00	84.51
Sibling-Sibling, Youngest 12-17/Oldest 18-25	1,330	31.94	63.53
Sibling-Sibling, Other Age Pairings	1,543	37.06	66.88
Sibling-Sibling Total	**4,164**	**100.00**	**71.28**
Spouse-Spouse with Children	1,631	51.55	56.59
Spouse-Spouse without Children	1,511	47.76	58.37
Spouse-Spouse, Children Not Clear	22	0.70	45.45
Spouse-Spouse Total	**3,164**	**100.00**	**57.36**
Grandparent-Grandchild	**193**	**100.00**	**56.99**
Overall Total	**13,757**	**100.00**	**68.63**

Source: SAMHSA, Center for Behavioral Health Statistics and Quality, National Survey on Drug Use and Health, 2014.

A caveat to using OFPM data is that sometimes the pair members disagree on whether they are part of a family. Other-pair-member roster editing has revealed that, especially for blended families and families involving unmarried partners, the pair members occasionally disagree on the nature of their relationship. The imputation-revised family-skip variable IRFAMSKP does not have to be consistent among the pair members (though the imputation-revised

pair relationship variable IRPRREL does), and those with IRFAMSKP = 1 are automatically assigned skip codes for edited and imputation-revised income variables related to the family.

For these reasons, if these methods were to be put into practice, it is recommended that the OFPM's values be used only if all three of the following conditions are met:

1. According to IRPRREL, the pair members are members of the same family.
2. Both pair members have IRFAMSKP = 0.
3. The values of the other pair member are consistent with the nonmissing responses given by the respondent.

3.2 Proxy Respondents

Respondents' selection of a proxy tends to improve the response rate for income variables. The next step is to examine whether proxy responses tend to improve agreement between the pair members. Table 8 shows the agreement among pairs with at least one proxy respondent compared with no proxy respondents, by pair type. For every pair type, the agreement is higher, usually *much* higher, for pairs with at least one proxy respondent. Pair types with high levels of agreement in the previous table (Table 7) are the pair types that tend to use proxies frequently: namely, the pairs with children aged 12 to 17.

Table 8. Influence of Proxy Respondents on Agreement by Pair Type among Responding Pairs for Family Finer Categories Income: 2014 NSDUH

	Responding Pairs		Percent Agreement among Pairs	
Pair Type	**Number**	**Percent with at Least One Proxy**	**At Least One Proxy**	**No Proxies**
Parent-Child, Child Aged 12-14	2,596	97.34	77.92	68.12
Parent-Child, Child Aged 15-17	2,139	92.01	79.98	50.88
Parent-Child, Child Aged 18-20	711	49.93	75.49	44.38
Parent-Child, Child Aged 21+	790	30.25	78.66	46.64
Parent-Child Total	**6,236**	**81.61**	**78.58**	**47.86**
Sibling-Sibling, 12-14/15-17	1,291	98.22	85.02	56.52
Sibling-Sibling, 12-17/18-25	1,330	91.05	66.31	35.29
Sibling-Sibling, Other	1,543	69.09	80.21	37.11
Sibling-Sibling Total	**4,164**	**85.13**	**77.18**	**37.48**
Spouse-Spouse with Children	1,631	20.48	74.55	51.97
Spouse-Spouse without Children	1,511	16.68	75.79	54.88
Spouse-Spouse, Children Not Clear	22	4.55	100.00	42.86
Spouse-Spouse Total	**3,164**	**18.55**	**75.13**	**53.32**
Grandparent-Grandchild	193	75.65	60.96	44.68
Overall Total	**13,757**	**68.09**	**77.56**	**49.57**

Source: SAMHSA, Center for Behavioral Health Statistics and Quality, National Survey on Drug Use and Health, 2014.

These results suggest that consistency would be increased by having only one member of the household respond to the family income questions, but this may not increase the accuracy of the family income measures. The next logical next step would be to use the pair data to estimate the measurement error. The family income questions are among the few NSDUH questions that are asked of two individuals who should be giving the same response. For these variables, it is probably true that the measurement error is larger than the sampling error.

3.3 Promising Potential of Introducing Nonresponse-Specific Probe Items

The final recommended next step for further reducing income item nonresponse is to explore the addition of probe items to the questionnaire.

There is evidence from drug measures in NSDUH that the inclusion of probe items asking for coarser estimates has some impact on nonresponse. Past month users of certain drugs are asked to report the number of days in the past 30 days in which they have used the drug, and the resulting measure is referred to as 30-day frequency of use. If they respond "Don't know" or refuse this question, they receive a probe asking for an estimate of their drug use within the past 30 days from the following categories:

- 1 or 2 days;
- 3 to 5 days;
- 6 to 9 days;
- 10 to 19 days;
- 20 to 29 days; and
- all 30 days.

The 30-day frequency-of-use probe questions have surprisingly high response rates, often greater than 80 percent (Table 9). If this approach is applied to income, perhaps the probe would include response levels matching the categories of the INCOME5 variable (Table 3), or even the levels of the POVERTY2 variable, based on the age and the responses to the household roster questions (Section 3.1.2).

Table 9. Item Response Rates of 30-Day Frequency Probe Questions: 2014 NSDUH

Drug	Number of Nonrespondents to Original 30-Day Frequency Question	Number of Respondents to Probe	Response Rate to Probe
Cigarettes	103	94	91.26
Snuff	20	16	80.00
Chewing Tobacco	14	11	78.57
Cigars	28	26	92.86
Alcohol	234	196	83.76
Inhalants	14	11	78.57
Marijuana	81	69	85.19
Hallucinogens	4	3	75.00
Cocaine	1	1	100.00
Crack	1	1	100.00
Heroin	1	1	100.00

Source: SAMHSA, Center for Behavioral Health Statistics and Quality, National Survey on Drug Use and Health, 2014.

Because it has been established that there are fundamental differences between nonresponse types, the current unfolding bracket structure of the NSDUH income questions could be improved by adding probe items based on types of nonresponse (i.e., “Don’t know” and refusals). Based on the understanding of cognitive factors surrounding income item nonresponse, it is recommended that the probe item for nonrespondents answering “Don’t know” should emphasize the reduction of response burden and help respondents come to a reasonable estimate. Furthermore, the probe item for nonrespondents refusing to answer the income items should focus on reassuring the respondent of anonymity and the importance in the accuracy of the survey estimates. The next step for examining the effectiveness of such probes in reducing income item nonresponse would require further refinement and field testing.

References

Bollinger, C. R., Hirsch, B. T., Hokayem, C. M., & Ziliak, J. P. (2014). *Trouble in the tails: Earnings non-response and response bias across the distribution.* Unpublished manuscript.

Center for Behavioral Health Statistics and Quality. (2015). *2014 National Survey on Drug Use and Health: Methodological summary and definitions*. Retrieved from http://www.samhsa.gov/data/

Center for Behavioral Health Statistics and Quality. (in press). *National Survey on Drug Use and Health: Evaluation of imputation methods for the National Survey on Drug Use and Health*. Rockville, MD: Substance Abuse and Mental Health Services Administration.

Juster, T., & Smith, J. P. (1997). Improving the quality of economic data: Lessons from the HRS and AHEAD. *Journal of the American Statistical Association, 92*, 1268-1278.

Frechtel, P., & Copello, E. (2007). Patterns of nonresponse for key questions in NSDUH and implications for imputation. In *Proceedings of the American Statistical Association, Survey Research Methods Section*.

Frechtel, P., Scott, V., Couzens, A., Moore, A., & Bose, J. (2012). Imputation using the other pair member. In *Proceedings of the American Statistical Association, Survey Research Methods Section* (pp. 4248-4258).

Little, R. J. A., & Rubin, D. B. (1987). *Statistical analysis with missing data*. New York: John Wiley & Sons.

Moore, J. C., Stinson, L. L., & Welniak, Jr., E. J. (2000). Income measurement error in surveys: A review. *Journal of Official Statistics, 16*(4), 331-361.

Pleis, J. R., & Dahlhamer, J. M. (2004). Family income response patterns for varying levels of income detail: An analysis of the National Health Interview Survey (NHIS). In *Proceedings of the Joint Statistical Meetings, American Statistical Association, Survey Research Methods Section* (pp. 4200-4207). Alexandria, VA: American Statistical Association.

Tourangeau, R., & Yan, T. (2007). Sensitive questions in surveys. *Psychological Bulletin, 133*(5), 859.