Donor Selection for MCBS Income and Asset Imputation

Kanru Xia[1], Meimeizi Zhu[1], Edward Mulrow[1]

[1]NORC at the University of Chicago, 55 E. Monroe St. FL. 30, Chicago, IL 60603

Abstract

Income and asset (IA) survey questions are known to have high non-response rates. To mitigate this, surveys often allow respondents to report a value-range in lieu of an exact value. Still, most population based surveys implement item non-response imputation based on an appropriate respondent donor pool to enhance IA data quality. Prior research of elderly populations has suggested that those who provide value-range responses to asset questions have higher asset values than exact-value responders. Moreover, asset value-range respondents are more representative of item non-responders. We will examine whether IA exact-value respondents differ from value-range respondents in IA-related key demographic variables in the Medicare Current Beneficiary Survey (MCBS), a continuous, multipurpose survey of a nationally representative sample of the Medicare population that implemented a new IA questionnaire in 2015. In addition, we will compare IA item non-response imputation using a donor pool of all respondents versus only value-range respondents and see if restricting the donor pool would increase the imputed IA means. Outcomes will provide guidance for future MCBS IA imputation.

Key Words: Imputation, MCBS

1. Introduction

Income and asset (IA) survey questions are known to have high item non-response rates (Riphahn & Serfling 2005). To mitigate this, surveys often allow respondents to report a value-range in lieu of an exact value. Still, most population based surveys implement item non-response imputation based on an appropriate respondent donor pool to enhance IA data quality. Prior research (Juster & Smith 1997) of elderly populations has suggested that those who provide value-range responses to asset questions have higher asset values than exact-value responders. Moreover, asset value-range respondents are more representative of and therefore would be better donors for item non-respondents.

The aim of this study is to explore whether using only value-range respondents as donors during the imputation of item non-respondents provides more accurate IA estimates than using both value-range and exact-value respondents as donors. And if this is true, then we would like to examine the impact of selecting only value-range donors on the post-imputation IA estimates. In what follows, we describe the data source and methodology used for the study and present the results of the analyses. The results of this study are mixed. For "income earned from an asset," we conclude that only value-range respondents should be used as donors to impute for item non-response. For the rest of the IA items, we should use all item respondents as donors.

2. Data

For the analyses undertaken in this study, we used the Medicare Current Beneficiary Survey (MCBS)[1] IA data collected in 2016[2]. The MCBS is a continuous, multipurpose survey of a nationally representative sample of the Medicare population, conducted by the Centers for Medicare & Medicaid Services (CMS) through a contract with NORC at the University of Chicago. The MCBS collects data from Medicare beneficiaries at three points per year for four consecutive years. The survey covers many topics including health care utilization and expenditures, all sources of health insurance coverage, and health status and functioning.

An IA questionnaire is administered annually to MCBS respondents. It collects 17 IA items[3] (Table 1), such as Social Security income and present value of home. Probe questions are first asked to check whether the beneficiary and/or the spouse have an IA item. If the answer is yes, then dollar amount questions follow. For dollar amount questions, respondents are first asked for an exact amount. If the respondent refuses or says "don't know," a set of value ranges is presented for 16 IA items[4]. If the respondent still refuses or says "don't know," item non-response occurs. In this paper, we analyzed 15 IA items with value ranges from the 2016 MCBS IA data[5]. Among these 15 IA items, 73% of the respondents provided an exact amount, 12% provided a value range, and 15% were item non-respondents.

To enhance MCBS data quality, an exact dollar amount is imputed for both value-range responses and item non-responses. Hot deck imputation, which imputes the value from a donor, is used when the value cannot be imputed by the prior-year carry-over method[6]. For value-range respondents, donors are exact-value respondents. For item non-respondents, we compared imputation results using only value-range respondents to using value-range and exact-value respondents as donors.

---

[1] MCBS data are made available via two annual releases of Limited Data Set (LDS) files that contain roughly 40 linkable data sets and over 2,000 variables. A public use file is also available.
[2] These data are collected in 2016. But since most of the questions ask about prior year information, the data are included in the 2015 data files.
[3] The MCBS IA questionnaire also collects liability and expense items, such as rent and total mortgage owed. Since these are not IA items, they are not used in this study.
[4] MCBS did not ask range values for "salary". This is because the "salary" questions have complicated question routing and asks for amount information in separate variables (pay check, hourly rate, or daily rate).
[5] IA item "cars" is excluded from our analyses. This is because the "cars" questions ask about the total value of cars owned by the beneficiary and spouse. But during imputation, we first calculate the average value per car owned by the beneficiary and spouse, impute it, and then multiple by reported number of cars owned to create the imputed total value. The value range of total car value could not be used directly during this imputation.
[6] When prior year amount is available, we take that amount, adjust the amount by inflation, and set the adjusted amount as the imputed value.

Table 1: Income and Asset Questions Asked in MCBS

Type	Description
Income	Social Security / Railroad Retirement (SSRR)
Income	Supplemental Security Income (SSI)
Income	Veteran Administration (VA)
Income	Lump Sum
Income	Total Income Excluding Lump Sum
Income	Pension Payments
Income	Salary
Asset	Home Present Value
Asset	Retirement Plans (401K, etc.)
Asset	Mutual Fund / Bond
Asset	Bank Accounts / CD
Asset	Land, Business, Farm, etc.
Asset	Cars
Income from Asset	Retirement Plans (401K, etc.) Payments Last Month
Income from Asset	Retirement Plans (401K, etc.) Payments Last Year
Income from Asset	Interest (Mutual fund / Bond / Bank accounts / CD)
Income from Asset	Income from Land, Business, Farm, etc.

3. Method

To explore whether using only value-range respondents as donors during the imputation of item non-respondents provides more accurate IA estimates than using both value-range and exact-value respondents as donors, ideally we would like to compare the IA values among the three groups. However, even though the MCBS collects IA information over three years, item non-respondents in the current year tend to be item non-respondents in past years as well. Thus, we could not gather enough past year IA values for current year item non-respondents and could not compare IA values directly among the three groups. Instead, we compared IA-related key variables (Table 2), such as poverty and age, among exact-value respondents, value-range respondents, and item non-respondents for each of the 15 IA items we analyzed. We hoped that similarity in these demographics would provide evidence that the IA values would also be similar. We considered item non-respondents more similar to value-range respondents than to exact-value range respondents when (1) item non-respondents and value-range respondents did not differ significantly at the 0.05 alpha level in demographics and (2) item non-respondents and exact-value respondents had significantly different demographics.

To see how IA estimates would change by selecting different donors, we first checked whether exact-value and value-range respondents reported significantly different IA values. Next, we compared the after-imputation IA estimates using all item respondents versus only value-range respondents as donors for item non-respondents.

Table 2: IA-Related Key Variables

IA-Related Key Variables
Age
Gender
Race/Ethnicity
Education
Marital Status
Household Size
Poverty
Metro Area
Census Division

4. Results

Overall, results were mixed. As shown in Table 3, of the 15 IA items analyzed, only 7 showed that item non-respondents are more similar to value-range respondents in key demographics. In a different set of 7 IA items, value-range respondents had significantly higher IA values than exact-value respondents (Table 4). And in 3 IA items, value-range respondents had significantly lower IA values. When we grouped the IA items into three categories: income, asset, and income earned from an asset, we saw that three of the four "income earned from asset" IA items showed similarity between item non-respondents and value-range respondents. Moreover, value-range respondents provided significantly higher IA values in all four IA items. Results remain mixed for the "income" and "asset" categories.

Table 3: Key Demographics Comparison Among Item Non-Respondents, Range-Value Respondents, and Exact-Value Respondents

	Number of IA Questions Analyzed	
IA Question Type	*Total*	*Non-Respondent More Similar to Range-Value Respondent*
Income	6	3
Asset	5	1
Income from Asset	4	3
Total	15	7

Table 4: IA Value Comparison Between Range-Value and Exact-Value Respondents

	Number of IA Questions Analyzed		
IA Question Type	*Total*	*Reported Range Values Significantly Different From Exact Values*	*Reported Range Values Significantly Higher Than Exact Values*
Income	6	3	2
Asset	5	3	1
Income from Asset	4	4	4
Total	15	10	7

In conclusion, for "income earned from an asset," we recommend using only value-range respondents as donors to impute for item non-response. The after imputation IA mean is higher. For the rest of the IA items, we should use all item respondents as donors.

Acknowledgements

References

Juster, F. Thomas, and James P. Smith. "Improving the Quality of Economic Data: Lessons from the HRS and AHEAD." Journal of the American Statistical Association 92, no. 440 (1997): 1268-1278.

Riphahn, Regina T., and Oliver Serfling. "Item non-response on income and wealth questions." Empirical Economics 30, no. 2 (2005): 521-538.