Using Administrative Data to Validate HRS Survey Responses on Application for DI and SSI Disability Benefits

John Bound, Charles Brown, and Chichun Fang [*]

Social scientists usually rely on two types of information to assess participation in government-funded programs such as Social Security Disability Insurance (DI) or Supplement Income Security (SSI). One involves survey data, where researchers ask respondents about program participation as well as other information about the respondent, such as education, race, and household composition. Another source of information is the administrative records that government agencies keep as they administer the programs.

While administrative records typically include large numbers of individuals and are considered as the "gold standard" in terms of accuracy, they are not always readily available for research use. Administrative data also usually only include information needed for the purpose of program administration, and lack other important information that may interact with plan participation. On the other hand, surveys are easier to access and contain a much broader set of information about program participants. However,

to the extent that respondents may not fully understand the survey questions or fail to recall the correct answers, survey answers might not always be accurate, resulting in "reporting errors." Evaluating the magnitude of reporting errors is hence an important question that informs social science research, and the results would inform researchers about what source(s) of data to use when balancing data availability and data quality.

In this research, we combine the two types of data — the Health and Retirement Study (HRS) survey and Social Security Administrative (SSA) administrative records of DI/SSI applications and awards, to assess the magnitude of reporting errors and its implications. We use each source of data to (independently) construct a history of DI/SSI applications and awards for a group of survey respondents that are representative of the U.S. population between ages 51 and 61 (the decade before they are eligible for Social Security early retirement). We then evaluate the consistency between

* **John Bound** is the George E. Johnson Collegiate Professor Emeritus of Economics and Professor Emeritus of Economics at the University of Michigan (U-M), and Research Professor Emeritus at the Survey Research Center (SRC), Institute for Social Research at the University of Michigan. **Charles Brown** is Professor Emeritus of Economics at U-M and Research Professor Emeritus at SRC. **Chichun Fang** is an assistant research scientist at SRC. This research brief is based based on working paper [MRDRC WP 2023-462](), UM20-01.

the two sets of application/award history that we construct. We also estimate a simple empirical model of the determinants of applications using both data sets and contrast the results. We ask two research questions: (a) How accurate is the HRS survey on DI/SSI applications and awards, and (b) how much difference does it make if a researcher uses HRS versus SSA's Form 831 to assess DI/SSI applications?

We find that, even for salient experiences such as participation in DI/SSI programs, there are some discrepancies between the HRS survey and Form 831 records. Among those whose administrative records indicate that they have applied for DI/SSI between 1992 and 2016, about 40% either did not report doing so in the survey or misreported application outcomes (for example, misreported a denial as an approval). There is some evidence that respondents who are less well educated, have cognitive limitations, or experience health-related limitations on their capacity to work are more likely to misreport.

Part of the discrepancy between the two data sources can be attributed to the nature of these data. That is, even if HRS respondents report their DI/SSI applications accurately in the survey, HRS and Form 831 may still not match. Only respondents who reported health-related work limitations are asked about DI/SSI applications, so there are a number of respondents who, despite the administrative

data showing they had applied for disability benefits, never got the opportunity to report their DI/SSI applications in the survey. Additionally, the administrative data we use do not include technical denials, denials made by the Social Security Administration regarding whether the individual has a sufficient work history to qualify for DI or a sufficiently low level of assets and income to qualify for SSI. A survey respondent may correctly report such a technical denial as a denied application in the survey, but this denial will not appear in the administrative data we use. While we can, to some extent, tailor analysis to minimize the importance of these features of the two data sets, we nonetheless find discrepancies that seem very likely to represent misreporting by respondents.

We then explore the impact of such discrepancy in reporting/recording of DI/SSI applications between the HRS' and SSA's data. We employ a simple research setting and study the factors that affect the likelihood survey respondents apply for DI/SSI benefits. We estimate the same model twice — one identifying DI/SSI applications using the HRS data and another using SSA's administrative data. We find that, while the information on DI/SSI applications between the two sources of data is not always consistent, we obtain qualitatively similar results regardless whether HRS or administrative data are used to identify DI/SSI applications. ❖