

REPRESENTATIVENESS OF THE LOW-INCOME POPULATION IN THE HEALTH AND RETIREMENT STUDY

ERIK MEIJER AND LYNN A. KAROLY *

July 2013

Introduction

The Health and Retirement Study (HRS) is a key data source used to analyze the health and economic status of the middle-aged and older population in the United States. Many analyses are specifically interested in the low-income population that is eligible for various means-tested programs such as Supplemental Security Income (SSI) and Medicaid. However, the validity of studies that rely on the HRS depends on whether it accurately represents the size and composition of the low-income population. Prior research has provided some evidence that the HRS may not be representative of the low-income population in particular. The goal of this paper then is to examine whether the HRS is representative of all income groups, but with a particular emphasis on low-income groups.

Aims and Methods

This paper undertakes a rigorous assessment of the representativeness of the low-income population in the HRS by using matched Social Security Administration (SSA) administrative data on earnings and beneficiary payments and comparing the resulting distributions to those for the same population taken directly from SSA records. The SSA records cover the entire population, so they provide a benchmark distribution, for any given income component, with which we can compare the distribution of the same administrative data measure in the SSA data sets matched to the HRS survey sample. Because the variables originate in the same SSA records, discrepancies between the distributions found in the direct SSA records and the distributions found in the administrative data sets matched to the HRS must be due to differences in sample composition.

Findings

Based on our preliminary analysis, we find that overall, for the age cohorts and years that can be most reliably compared, the distributions are very similar and conclude that the HRS is representative for the population it covers. For example, for the age cohort and year that can be most reliably compared—the original HRS cohort of individuals born between 1931 and 1941 and measures as of 2003—the combined income from earnings, Social Security benefits, and Supplemental Security Income (SSI) benefits is very similar between the HRS

* *Erik Meijer* is an economist at the University of Southern California and RAND. *Lynn A. Karoly* is a senior economist at RAND. This Brief is based on MRRC Working Paper 2013-273.

sample and the SSA administrative data sample. However, for some subgroups in the low-income population, there are some differences. For example, the HRS underestimates the share of the 1931–1941 cohort that is receiving SSI benefits as of 2003 by 1.5 percentage points or about 28%. For this reason, we caution against estimating population totals for such small subpopulations such as SSI recipients or Medicaid beneficiaries. We expect these findings to continue to hold when the analysis is performed with revised analytic files to address several data issues.

For some analyses, the matched administrative data on income sources and program participation are preferred because they are less likely to be measured with error compared with the same variables based on survey responses. Our analysis demonstrates that the HRS samples for which restricted matched administrative data are available are often not representative of the population of interest, because not all HRS respondents were asked permission to match their SSA records to their survey response in any given year. Researchers who intend to use the restricted data for their analyses should take note of which respondents were asked permission to match in which wave. Although the restricted HRS data files are generally not suitable for estimating population distributions, they are still very useful for modeling purposes.

Conclusion

In the course of this study, we identified a number of limitations with the SSA administrative data that are matched with the HRS survey data, issues which complicated our ability to implement our research design. Some of these limitations are unavoidable, but we believe that the usefulness of the HRS data can be improved in a number of ways that should be feasible to implement. Thus our paper lists a series of potential improvements to the data files that will enhance the usefulness of the restricted HRS data files that are matched to SSA records in ways that will support a wider array of analyses.

University of Michigan Retirement Research Center
Institute for Social Research 426 Thompson Street Room 3026
Ann Arbor, MI 48104-2321 Phone: (734) 615-0422 Fax: (734) 615-2180
mrrcumich@umich.edu www.mrrc.isr.umich.edu

Sponsor Information. The research reported herein was performed pursuant to a grant from the U.S. Social Security Administration (SSA) through the Michigan Retirement Research Center (MRRRC). The findings and conclusions expressed are solely those of the author(s) and do not represent the views of SSA, any agency of the federal government, or the MRRRC.

Regents of the University of Michigan: Mark J. Bernstein, Ann Arbor; Julia Donovan Darlow, Ann Arbor; Laurence B. Deitch, Bloomfield Hills; Shauna Ryder Diggs, Grosse Pointe; Denise Ilitch, Bingham Farms; Andrea Fischer Newman, Ann Arbor; Andrew C. Richner, Grosse Pointe Park; Katherine E. White, Ann Arbor; Mary Sue Coleman, Ex Officio