**CID Technical Notes Series 2024-02:**

# Measuring Income from OASDI Benefits with Administrative Universe Data [1]

**Angela Wyse**
University of Chicago

**Connacher Murphy**
University of Chicago

**Derek Wu**
University of Virginia

**Bruce D. Meyer**
University of Chicago,
NBER, and AEI

May 6, 2024

## I. Overview

This document summarizes our methodology for measuring income from Old Age, Survivors, and Disability Insurance (OASDI) benefits with universe data from the Social Security Administration (SSA). OASDI is a near-universal program designed to replace lost income due to retirement, disability, or death. In 2020, the program paid $1.1 trillion in benefits to 65 million beneficiaries.[2]

We combine the 2015 Master Beneficiary Record (MBR) with the 2019 Payment History Update System (PHUS) files to measure monthly OASDI benefits from 2008 through 2018. This memo describes our procedure for combining the data sources, assigning or imputing benefit types and amounts, validating imputations of benefit types and amounts, and determining whether OASDI benefit amounts brought in from the PHUS are net of Medicare premium amounts. The cleaned dataset benchmarks well against public aggregates for 2014-2018, covering at least 97.5% of benefits and 98.7% of recipients in these years. The dataset's coverage is less complete going back from 2013 to 2008, falling to about 77% in 2008. Even in these years, however, our combined dataset improves coverage by about 12 percentage points relative to aggregates based on the PHUS alone, which misses about this share of people due to individuals dying or aging out of benefit eligibility.

[1] This memo is released to inform interested parties of research and to encourage discussion. Any views expressed are those of the authors and not those of the U.S. Census Bureau. The Census Bureau has reviewed this data product for unauthorized disclosure of confidential information and has approved the disclosure avoidance practices applied to this release, authorization number: CBDRB-FY2023-CES005-011. Authors can be contacted at the following email addresses: Angela Wyse (awyse@uchicago.edu); Connacher (Connor) Murphy (connormurphy@uchicago.edu); Derek Wu (derek.wu@virginia.edu); Bruce D. Meyer (meyer1@uchicago.edu).

[2] Social Security Administration's Annual Statistical Supplement, 2021

## II. Administrative Data Sources

The 2015 MBR, which SSA uses to generate OASDI benefit checks, provides a snapshot of beneficiaries and their payment status in its extraction month, December 2014. When we restrict to individuals with "current" payment status we closely match the public SSA OASDI beneficiary total for December 2014.

The 2019 PHUS provides monthly payment history from January 1984 to December 2018 for nearly all people who received benefits in its extraction month, October 2019. Payments reported in the PHUS are net of Medicare premium deductions.[3] Unlike the MBR, it reports disbursement of, not entitlement to, benefits. We estimate that the PHUS includes payments for about 99.5% of people who were entitled to benefits in the extraction month, or 97-98% of all dollars paid.[4] We hypothesize that this gap in coverage arises because public SSA aggregates indicate payments to people with current payment status, while the PHUS indicates the actual disbursement of payments, including retroactive benefit payments and lump-sum death benefits. Accordingly, we do not expect our imputation procedure to recover 100% of aggregate benefit amounts and we use these estimates as the targeted share of public aggregates recovered by our cleaned dataset in the benchmarking section below.

Besides the extraction month and coverage dates, a key distinction between the MBR and PHUS lies in the information about beneficiaries included in each dataset. The MBR includes linkage keys (Protected Identification Keys, or PIKs) for both claimants (the person whose work history generated the entitlement) and beneficiaries, as well as a code indicating the type of beneficiary (e.g. widower, child, or claimant). The PHUS includes linkage keys for claimants but not for beneficiaries. For beneficiaries, the PHUS only indicates beneficiary type.

The following table summarizes key attributes of the MBR and PHUS datasets:

| **Key Attributes of 2015 MBR and 2019 PHUS** | | | | | |
|---|---|---|---|---|---|
| **Dataset** | **Extraction Month** | **Coverage Date(s)** | **Level** | **Contents Summary** | **Linkage key availability** |
| **2015 MBR** | Dec 2014 | Dec 2014 | Claimant-beneficiary | Payment status, beneficiary type | PIKs for both claimant and beneficiary |
| **2019 PHUS** | Oct 2019 | Jan 1984-Dec 2018 | Claimant-beneficiary | Payments (disbursed and net of Medicare premium deductions), beneficiary type | PIK for claimant only; beneficiary type only |

[3] If an individual is eligible and signed up for both Social Security and for Medicare Part B (the portion of Medicare that provides standard health insurance), SSA automatically deducts the Medicare B premium amount from monthly benefits. We verify that payments reported in the PHUS are net of Medicare premium deductions (e.g., that the payments have already subtracted what the individual owes in Medicare Part B premiums) by comparing the PHUS amounts to benefit amounts recorded in the CPS-PHUS extracts, which contain detailed OASDI payment history for a subset of the individuals found in the 2019 PHUS, including benefit amounts net of Medicare premiums. Each year between 2014 and 2016, about 98.6% of the payments matched between the PHUS and the CPS-PHUS extracts, confirming that the PHUS benefit amounts are net of Medicare premiums.

[4] We obtain this estimate by comparing the 2020 PHUS files to SSA aggregates; we find that, as a share of public aggregates, the totals from the PHUS peak at 97.6% of benefits paid and 99.6% of beneficiaries. We cannot compare the 2019 PHUS to SSA aggregates because its payment amounts data ends in December 2018, well before the dataset's October 2019 extraction date.

## III. Challenges in Creating a 2008-2018 OASDI Dataset

We address two main challenges in combining the MBR and PHUS to form our target dataset:

- **Attrition between MBR and PHUS extraction months.** We lack payment amounts for those who received payments in December 2014 (MBR extraction month) but not October 2019 (PHUS extraction month). Approximately 15.6% of beneficiaries who had current payment status in the 2015 MBR do not link to the PHUS, suggesting that they stopped receiving benefits (e.g. aged out, died) during this window.[5] For these individuals, we impute a mortality-adjusted benefit amount according to a procedure described in the next section.

- **Beneficiary identification.** We cannot merge the MBR and PHUS using beneficiary PIKs because these are not available in the PHUS. Instead, we merge based on claimant PIK and beneficiary type. About 96% of beneficiaries in the MBR are uniquely identified by a given claimant/beneficiary-type combination. [6] The remaining 4% of beneficiaries cannot are not uniquely identified by claimant PIK and beneficiary type, because there are multiple beneficiaries of the same type, e.g. several child beneficiaries of a single claimant. To address this, we apportion benefit amounts equally within each claimant and beneficiary-type group, preserving the total amount paid to people in that group.

We face a third challenge with no ready solution, which will remain a limitation of our cleaned dataset:

- **Beneficiaries missing from both the MBR and PHUS.** Our 2008-2018 dataset contains neither indicators nor amounts for people who a) received benefits at some point in 2008-2014 but stopped before December 2014, or b) only received benefits for some window between the two extraction dates. Results from our benchmarking analysis suggest that the latter group is small while the former group is substantial, representing about 25% of 2008 OASDI beneficiaries.[7]

## IV. Combining the MBR and PHUS

As described above, we merge the MBR and PHUS at the monthly level using claimant PIK and beneficiary type code, apportioning PHUS benefit amounts equally among beneficiaries within these groups when necessary. We then proceed to assign benefit types and amounts where possible and impute them if not. Section V of the memo describes our assignment and imputation methodology in detail. Here, we summarize our methods by considering the set of cases where a) a beneficiary appears in both sources, b) a beneficiary appears in the PHUS only, and c) a beneficiary appears in the MBR only.

*a) Beneficiary appears in both MBR and PHUS.* In general, we use the MBR to identify beneficiaries and merge on monthly benefit amounts from the PHUS using claimant PIK and beneficiary type code. This is the procedure by which benefits are assigned to most records that appear in our final cleaned dataset.[8]

*b) Beneficiary appears in PHUS only.* In some cases, especially after 2014, no MBR record exists for a given claimant PIK/beneficiary type combination in the PHUS. This can occur if someone began

[5] There were approximately 59 million individuals with current payment status in the 2015 MBR and about 9.2 million of these individuals did not link to the PHUS.
[6] About 57 million of the 59 million people with current payment status in the 2015 MBR were in a uniquely identified claimant/beneficiary type group.
[7] See Figures 1 and 2.
[8] The exact share varies by month but ranges from about 65-85%.

receiving benefits after the MBR extraction date. In these cases, we assign all PHUS payments to the claimant, since this is the only individual for whom we have a PIK. This rule will overstate benefits paid to the individual but help us identify benefits received by the family. This assignment procedure is rarely used before 2015 but becomes increasingly important afterwards. In 2018, this is the procedure by which benefits are assigned to about 18% of beneficiaries. Moreover, in cases where the beneficiary appears in the PHUS only, we are left with no indication of benefit type—just monthly payment history and amount. For those missing benefit types but with record of positive receipt, we impute as receiving OASI for people below the age of 18 or over the age of 62 and as receiving DI for everybody else.

c) *Beneficiary appears in MBR only.* About 15.6% of those with current payment status in the MBR do not link to the PHUS, leaving us with no indication of monthly payment history or amount. In this case, we impute monthly benefit amounts from 2008-2018 via hotdeck. We define cells by variables available in the MBR (age, beneficiary type, benefit type, and sex) and draw donors from the Current Population Survey (CPS) extract files, which contain OASDI benefit payment and eligibility history for respondents to various years' CPS Annual Social and Economic Supplement (ASEC) surveys. We restrict the donor sample to people who received benefits in December 2014 but not October 2019. We then impute a sequence of monthly benefit amounts over 2008-2018 from a single donor individual, thereby preserving the intertemporal dependency in benefit amounts. We adjust for mortality using the SSA Numident file by recoding benefit amounts for all months following an individual's death to zero.

## V. Validating Imputed OASDI Type and Receipt in the Combined Administrative MBR-PHUS

We validate how accurate our imputation methodology is by comparing how well our imputations agree with true benefit types and amounts recorded in an additional data source, the CPS-PHUS extract files, which contain detailed information on OASDI payment history (including both benefit types and amounts) from 1995 to 2018 for individuals interviewed in the Current Population Survey (CPS) across a range of years (1973, 1991, 1994, 1996-2018). As expected, benefit types are least missing in 2014 (0.75%) and steadily increase to most missing in 2018 (19.44%), while benefit amounts are most missing in 2014 (14.70%) and steadily decline to least missing in 2018 (5.68%).[9]

Our age-based imputations accurately attribute benefit type (e.g., OASI or DI) among at least 97.5% of recipients in 2015 and 2016 and at least 94.3% of recipients in 2017 and 2018. Overall, including both imputed and non-imputed observations, benefit type matched those recorded in the PHUS extracts at least 98.3% of the time between 2015 and 2018. We also correctly attribute positive or no receipt of OASDI benefits to at least 94.7% of individuals who link from the PHUS to extracts to our cleaned MBR-PHUS dataset. The mean absolute difference between amounts reported in the PHUS extracts and the universe data for recipients with imputed amounts is around $6,300, relative to mean true receipt as reported in the extracts of around $12,800.

## VI. Estimating Medicare premium deductions

As in the PHUS, our combined dataset indicates benefit payments net of Medicare premium deductions. Users of this dataset may wish to know the value of such deductions to calculate income more accurately. We draw on the CPS-PHUS extract files, which indicate Medicare premium amounts in addition to net

[9] This is consistent with the fact that the extraction month for the MBR, which indicates benefit type but not amount, is December 2014, and the extraction month for the PHUS, which indicates benefit amount but not but not type, is October 2019.

OASDI payments, to identify the modal Medicare premium deduction amount by age in 2014. At the beginning of 2014, Medicare deductions were zero for about 68.39% of OASDI recipients under 65, $104 for about 25.43% of OASDI recipients under 65, and $104 for about 65.15% of those 65 and older. This latter amount of $104 is the basic premium amount people paid in 2014 if they were signed up for Medicare B and their yearly income was $85,000 or below (as a single filer) or $170,000 or below (filing jointly). Each year, this basic premium increases to adjust for inflation and rising healthcare costs[10]. These modal values may offer data users reasonable estimates for Medicare premium deduction amounts by age in 2014.

## VII. Benchmarking the Cleaned Dataset

In Figures 1 and 2 (next page), we compare aggregates from our combined administrative data to publicly available aggregates for benefits and beneficiaries, respectively.[11] We also show aggregates from the PHUS alone. Each graph displays a target, which is based on the ratio of the 2020 PHUS total to public aggregates in its extraction month, March 2020, as indicated in Section II, and reflect the fact that public OASDI aggregates reflect total disbursements whereas the internal PHUS file only reflect disbursements to people with current payment status. We rely on the 2020 PHUS to obtain this target because the 2019 PHUS stops reporting benefit amounts in December 2018, ten months before its extraction month.

Between 2014 and 2018, we cover at least 97.5% of benefits and 98.7% of recipients. Our coverage deteriorates going back from 2014 to about 77% in 2008, although we note that the combined administrative data represent a substantial improvement – about 12 percentage points – over the PHUS alone. In part of 2014 and between 2015 to 2018, our cleaned dataset covers a larger share of beneficiaries and dollars than the target, suggesting that we may be slightly over-covering both benefits and beneficiaries in these years.

[10] Tables indicating Medicare Part B annual premium amounts by income category and year can be found at Q1Medicare.com, a non-government resource that archives Medicare-related material for research purposes.

[11] Publicly available aggregates for OASDI benefits and beneficiaries between 2008 and 2018 can be found using the Social Security Administrations databases for Social Security Benefit Payment Data and Benefits Paid by Type of Beneficiary, respectively. Aggregates for OASDI benefit payments include total benefit payments minus Medicare premium deduction amounts at the end of each month and exclude reductions for unnegotiated checks and payments (if any) for vocational rehabilitation services to disabled beneficiaries. Aggregates for OASDI beneficiaries include the total number of OASI and/or DI beneficiaries at the end of each month, including retired workers and their dependents, surviving family members of deceased workers, and disabled workers and their dependents.

## VIII. Figures

**Figure 1. PHUS and Combined Administrative Data Coverage of Public Benefits Aggregates, 2008-2018**

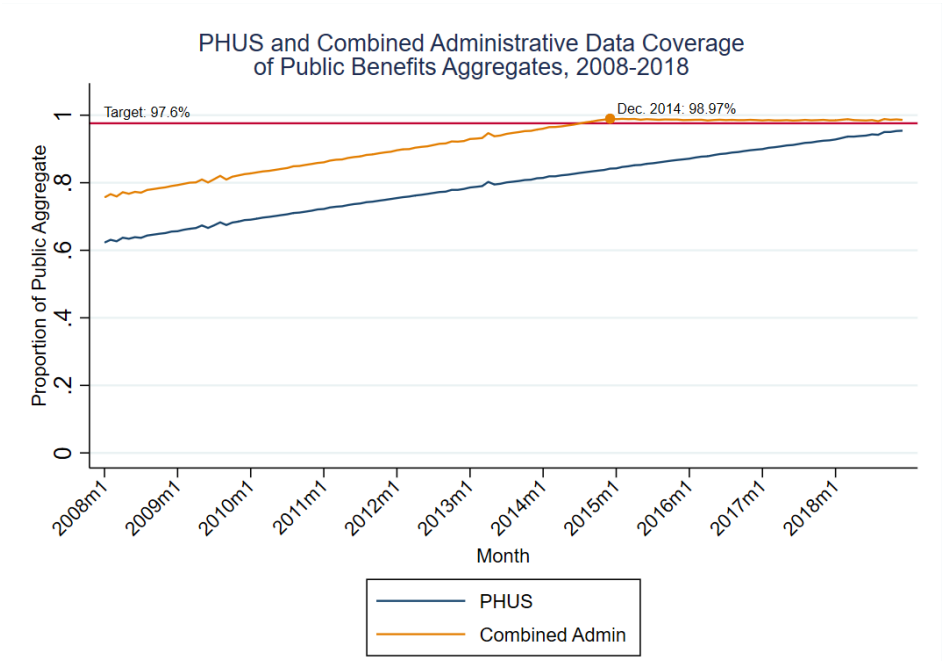


**Figure 2. PHUS and Combined Administrative Data Coverage of Public Beneficiaries Aggregates, 2008-2018**