Overview and Documentation of the Social Security Administration's Disability Analysis File (DAF) Public Use File for 2018 (PUF18)

October 2020

Submitted to:

Social Security Administration
Office of Retirement and Disability Policy
Office of Program Development and Research
Washington, DC 20024-2796
Project Officers: Paul O'Leary and Debra Tidwell-Peters
Contract Number: SS00-16-60003

Submitted by:

Mathematica
1100 1st Street, NE
12th Floor
Washington, DC 20002-4221
Telephone: (202) 484-9220
Facsimile: (202) 863-1763
Project Director: Jody Schimmel Hyde
Reference Number: 50214.Y3.T05.530.360

This page has been left blank for double-sided copying.

CONTENTS

SUGGESTED CITATIONS V

I. INTRODUCTION 1

A. What is the DAF? 2

B. Which Beneficiaries Have Records in the DAF? 3

C. Why Work with the DAF? 4

D. How Does the DAF PUF Differ from the DAF? 5

E. What's New in the DAF PUF? 6

II. WORKING WITH PUF FOR THE FIRST TIME 9

A. Basic Structure of the PUF 9

Demographic Component 9

Annuals Component 10

B. Selecting Which File to Use to Start a Research Project 12

C. Critical Considerations for Working with Variables in the PUF 13

D. Issues Specific to Variables in the PUF 16

Special missing values. 16

Determining whether a variable applies to SSDI or SSI. 16

Missing values indicating data not available. 18

Missing values indicating all values masked. 18

The PUF unique identifier 19

Data masking. 19

III. PUF VARIABLE LISTS 23

IV. PUF VARIABLE DETAIL PAGES 27

DAF PUF Variable Detail Pages: Demographic File 29

PUF Variable Detail Pages: Annuals Component 78

TABLES

Table II.1. Record Counts in DAF18 PUF Files 11

Table II.2. Components in the DAF PUF 13

Table II.3. Mapping Broad Areas to SSA Regions for Purposes of Masking Move Patterns 21

Table III.1. Variables in the PUF Demographic Component 23

Table III.2. Variables in the PUF Annuals Component 25

Table IV.1. List of commonly used terms in variable names and DAF documentation 28

SUGGESTED CITATIONS

Users of the DAF PUF should cite the data files and/or documentation using the following citations, updated to include the date the files were accessed, and the years of data used:

1. For the DAF PUF Demographic file (*<YYYY>* is the year accessed):

U.S. Social Security Administration, Office of Retirement and Disability Policy, Office of Research, Demonstration, and Employment Support. (*<YYYY>*). *Disability Analysis File (DAF) Public Use File (PUF), DAF PUF Demographic File* [Data set]. Retrieved from https://www.ssa.gov/disabilityresearch/daf_puf.html#files. Accessed on *<DATE>*.

2. For the DAF PUF Annual files (*<YYYY>* is the year accessed):

U.S. Social Security Administration, Office of Retirement and Disability Policy, Office of Research, Demonstration, and Employment Support. (*<YYYY>*). *Disability Analysis File (DAF) Public Use File (PUF), DAF PUF Annual File(s) (<DATA YEARS USED>)* [Data set]. Retrieved from https://www.ssa.gov/disabilityresearch/daf_puf.html#files. Accessed on *<DATE>*.

3. For DAF documentation (*<YYYY>* is documentation publication date):

U.S. Social Security Administration, Office of Retirement and Disability Policy, Office of Research, Demonstration, and Employment Support. (*<YYYY>*). *Overview and Documentation of the Social Security Administration's Disability Analysis File Public Use File* [Data file and code book]. Retrieved from https://www.ssa.gov/disabilityresearch/daf_puf.html#files.

This page has been left blank for double-sided copying.

I. INTRODUCTION

This document describes the public use file (PUF) version of the Social Security Administration's (SSA) Disability Analysis File (DAF). Despite the generic term "PUF," we use it throughout this document to refer to the DAF PUF. All potential users of the PUF, even those experienced with working with the DAF, should read this document. This document describes the DAF, explains the PUF and its contents in more detail, and highlights differences between the full DAF and the PUF.

The DAF is a set of files containing SSA administrative data on federal disability beneficiaries, culled from a variety of SSA sources. In particular, it contains data related to program participation and benefits for Social Security Disability Insurance (SSDI) and Supplemental Security Income (SSI) beneficiaries between birth and Full Retirement Age (FRA) who have received disability benefits in any month since 1996. There are millions of disability beneficiaries in each year of the DAF.

Access to the DAF itself is limited to SSA staff and selected researchers. SSA developed this PUF to make the DAF widely available to researchers interested in those receiving federal disability payments through the SSI and SSDI programs. This PUF contains a random 10 percent (10%) sample of beneficiaries included in the DAF. The 10% file is large enough to avoid disclosure risk and small enough to keep the file size manageable. The file is available on https://www.data.gov/ in a .CSV format that can be converted for use in common statistical packages. The version of the file stored at SSA will be available in .CSV format as well as SAS and Stata datasets. To release the PUF, we made modifications to the underlying data in keyways to avoid disclosing personal information about beneficiaries. The changes should not affect researchers' analysis plans and are described in this document.

This document begins by describing the DAF before proceeding to explaining the PUF and its contents in more detail. If after reading this document, you think that your research project cannot be completed using the PUF, but could be using the full DAF, please contact: ORDES.DAF@ssa.gov.

A. What is the DAF?

There are millions of disability beneficiaries in each year of the DAF. For each beneficiary, the DAF contains information about the disability determination process, program participation, and benefit payments. The DAF contains both longitudinal and one-time data about beneficiaries, with data from SSI and SSDI combined into a single record per beneficiary.

The DAF is currently constructed for SSA on an annual basis by Mathematica. Each year, we build a new version of the DAF to add records for beneficiaries who began participating in the SSDI or SSI programs during the most recent year and to update records for beneficiaries who enrolled in an earlier year and are in previous versions of the DAF. The DAF was originally designed for the purpose of evaluating the TTW program and was initially called the Ticket Research File (TRF), but we changed the name in 2011 to reflect its use in studying a variety of topics related to SSDI or SSI.[1]

Each version of the DAF is named based on the last year of data it contains. For example, the DAF18 contains data through December 2018. One-time variables in the DAF include data such as Social Security Number (SSN) and date of birth, while longitudinal variables track beneficiaries over time and include data such as state of residence, impairment codes, and benefit payment amounts. The DAF contains data as early as January 1994, and many longitudinal

[1] In this document, we use the term "DAF" to refer to its current incarnation as well as generically to the database in all its iterations since 2004. We use the term "TRF" when discussing specific previous versions of the database. The name change reflects the broadening use of the database beyond its original purpose in evaluating the TTW program, and users should understand that it does not mark a sharp break in how the database is structured.

variables are available for the full time period (through the end of the most recent DAF year). Due to limitations in the source data used to build the DAF, some variables are only available for a portion of the time range, and information in DAF documentation provides users with the range available for each variable.

The DAF is in SAS format with one-time information contained in a Demographic file, and with monthly (longitudinal) data contained in Annual files. There is one Annual file for each calendar year, beginning with 1994 and continuing through 2018 in the current version of the DAF. We provide extensive documentation for the DAF, with twelve comprehensive volumes available at: https://www.ssa.gov/disabilityresearch/daf.html. In this document, we focus on aspects of the DAF that are unique to the PUF, but researchers considering projects using the PUF are encouraged to consult the full set of documentation, particularly Volumes 1 through 3, to better understand the selection criteria for the DAF and the information it contains.

B. Which Beneficiaries Have Records in the DAF?

DAF18 contains over 30 million records. This includes all beneficiaries between birth and FRA who have received SSI or SSDI since 1996. When building the DAF, the criteria for selection are slightly different for SSI and SSDI. [2]

We select SSI beneficiaries if they meet the following conditions in any month during the years that the DAF has been collected:

- are currently in active status. temporary suspended status, or non-pay status for the SSI program (if in non-pay status, benefits must not have been denied); and
- are blind or disabled; and
- are between birth and FRA;[3] and

[2] The selection criteria and changes over time are described in more detail in Volume 1 of the full DAF documentation, available at https://www.ssa.gov/disabilityresearch/daf.html.

[3] SSI beneficiaries receiving benefits between 65 and FRA are doing so as old age beneficiaries.

- have non-blank SSN

We select SSDI beneficiaries if they meet the following conditions in any month during the years that the DAF has been collected:

- are currently in active or temporary deferred or suspended status for the SSDI program; and
- are disabled and either a primary claimant, dependent adult child, or a widow(er); and
- are between 18 years old and FRA; and
- have a non-blank SSN

C. Why Work with the DAF?

The DAF repackages existing SSA administrative data with research and analysis in mind. It is available to help researchers quickly answer questions about beneficiaries in SSDI, SSI, or both. Some of the questions that can be answered using the DAF (and the DAF PUF) include:

- How many beneficiaries first received SSDI in 1999 and what was their age distribution, by gender and impairment category?
- How does the distribution of primary disabling condition of beneficiaries change according to age when benefits first began?
- How many SSDI beneficiaries had their cash benefits suspended or terminated because they were working above Substantial Gainful Activity (SGA), and what was the dollar value of those forgone benefits?

Relative to working with other SSA administrative data, the DAF eliminates tedious data development steps, letting researchers quickly focus on data analysis. Benefits of working with the DAF include:

- It combines information across many SSA administrative data files, significantly cutting time spent merging across sources.
- It offers in-depth documentation including helpful hints for analysis and variable descriptions that are not readily available elsewhere.
- It is structured as one-observation-per-beneficiary, spanning SSDI and SSI, so that it is easy to account for concurrent beneficiaries and movement between programs.

- It is already available in SAS format and can be converted to Stata or SPSS for researchers more familiar with those analysis packages.
- Key measures, such as monthly flags for program participation, indicators of suspension or termination for work, and the dollar value of benefits forgone for work are already constructed based on information across sources, significantly cutting time spent constructing variables from scratch.
- It makes it easy to limit analysis to specific years and to conduct cross-sectional or longitudinal analysis.

D. How Does the DAF PUF Differ from the DAF?

The PUF contains a random 10 percent sample of records available in the DAF. Like the DAF, the PUF is divided among sub-files so you only use what you need. Unlike the SSNs contained in the DAF, PUF files have a unique random identifier for each observation (beneficiary), allowing researchers to link across the various sub-files of the PUF. Similar to the DAF, the PUF is available in SAS format, is also available in Stata and CSV formats, and can be converted easily for use with other statistical packages. Relative to the DAF, the PUF contains a limited set of variables, selected based on those most useful to the general research audience. The PUF file also excludes the Ticket to Work and employment services sub-files that are part of the full DAF because the populations are too small to adequately avoid disclosure. Over time, additional measures may be added to the PUF; if there are measures that you think would be particularly beneficial, please contact ORDES.DAF@ssa.gov.

In addition to these variable restrictions on the PUF, we also masked the data in a variety of ways. For certain categorical variables, we combine categories to limit the possibility of identifying beneficiaries based on rare characteristics. For example, we collapsed the PUF variable for BIC, BIC_PUF**,** to three possible codes based on about a dozen codes for BIC in the restricted use DAF: A if BIC=A, C if BIC=C, and O if BIC in any other category. In a similar way, for primary diagnosis code DXPRIBEST_PUF, we collapsed to eight groups based on

hundreds of four-digit diagnosis codes in the DAF. For date variables, we recoded all days to the 15th of the month to avoid identifying individuals based on their exact date of birth or the date they experienced certain outcomes. Additionally, we placed a lower bound on dates in some cases to minimize the possibility of identifying individuals based on outlier values. For dollar values, we used Census guidelines to round and top-code values to also minimize the possibility of outlier values that could potentially be used to identify a beneficiary.

Chapter IV of this document contains variable detail pages for each variable included in the PUF. Each page describes any recoding we applied to the full DAF data so users know when information is based on rounded, masked, or regrouped values. Researchers should review the variable detail pages thoroughly before conducting analysis.

E. What's New in the DAF PUF?

This section describes changes reflected in the DAF18 PUF relative to the two previous versions of the PUF (DAF15 and DAF17). In each year of the DAF, we make improvements designed to improve the user experience and make updates to constructed measures. Consistent with changes in the DAF18 files, this year's DAF PUF files have changes in the constructed variables related to suspense or termination for work (STW) and benefits forgone due to work (BFW). In particular:

- In the case of SSDI, we made a small change to the STWDI algorithm to account for better information on completion of the Trial Work Period (TWP). The previous algorithm used only one source for the TWP completion date, the TWP Completion Date variable from the DCF-Claim table. The new algorithm uses a second, monthly variable, the Work Determination Code from the DCF-Title II Work Determination table. Together these provide a more complete list of beneficiaries who have completed a TWP.
- In the case of SSI, the STW and BFW measures have been refined to include more complete information on income accrued by SSI beneficiaries; these changes relate to SSI's system of retrospective monthly accounting and calculating deemed income. Volume 3 of the DAF18 documentation available at https://www.ssa.gov/disabilityresearch/daf.html contains a detailed explanation of the alternate measures of STWSSI and BFWSSI.

- As a result of changes to the individual program measures, the combined version of those measures, STWCM and BFWCM have also been revised to reflect those changes.

Additionally, we have made small revisions to the documentation below and to the handling of missing data for some variables to make the PUF easier to work with. None of these revisions substantively change the variables from earlier years; our aim was to provide better information to users to promote understanding of the individual measures.

This page has been left blank for double-sided copying.

II. WORKING WITH PUF FOR THE FIRST TIME

This section highlights some key considerations for working with the PUF. It describes the Demographic and Annuals components contained in the PUF and how those files are structured; discusses issues specific to individual variables, including the importance of consulting the variable detail pages and understanding the administrative source file before working with particular variables; and highlights the variable cleaning and masking procedures for each type of variable.

A. Basic Structure of the PUF

The PUF is sectioned into two core components, the Demographic component and the Annuals components, linkable using the unique identifier variable in each file. Table II.1 in the next section provides more information about the file that might be the best to start with for particular research projects, but we begin by describing each component in more detail.

Demographic Component

The Demographic component is a beneficiary-level file that contains demographic and other one-time information, such as date of birth, date of death, and information collected at the time of disability application. The Demographic file contains one record for each beneficiary who has ever met the DAF selection criteria since 1996 and is not limited to those still receiving benefits. After selecting a random 10 percent of beneficiaries, the Demographic component in the PUF18 contains just under 3.5 million beneficiary records (see Table II.1).

The variables in the PUF Demographic component are either one-time variables or so-called *n-variables* with multiple occurrences. For example, the date variable BDOE_START_PUFn indicates when a beneficiary's eligibility for SSDI began, where *n* denotes the occurrence. In other words, a beneficiary with multiple periods of eligibility for SSDI would have multiple

occurrences of this variable, each with a date that SSDI eligibility for that occurrence started. Some beneficiaries may have not have data in one or more occurrences because that occurrence did not apply to them. In the PUF, the number of occurrences is constrained to avoid identifying individuals with unusual patterns of occurrences.

Many, but not all, *n*-variables are date variables. In cases where the *n-variable* does not contain date information but rather contains information about an occurrence, there is often a corresponding date variable and vice versa. For example, in the case of TOC1_PUF, the corresponding date variable is TOC_START1_PUF, "SSDI Type of Claim Start Date (n=1-20)," and for TOC2_PUF, the corresponding date variable is TOC_START2_PUF, etc. If used in ascending order, the values in the *n-variable* are usually also in ascending order by date, but this is not necessarily the case. To determine the correct chronological order of the values in variables in which *n* denotes an occurrence (e.g., TOCn), researchers should first examine the dates in the corresponding date variable (e.g., TOC_STARTn). Researchers can also use the MIN or MAX SAS function to determine the earliest or latest date. The DAF page also offers suggestions for how to recode n-variables to chronological order and can be found here: https://www.ssa.gov/disabilityresearch/daf.html.

Annuals Component

The Annuals component of the DAF is a set of beneficiary-level files, one per calendar year from 1994 to 2018, which contains monthly and yearly longitudinal data related to program participation and benefits.[4] Such data include benefit amounts due, SSA region where benefits were received, participation in the TTW program, and information related to beneficiary income.

[4] Beneficiaries are selected for the PUF if they have one month of benefits from 1996 onward. The Annuals files contain historical data on those beneficiaries back through 1994.

Table II.1. Record Counts in DAF18 PUF Files

File	Total records (approximate)	Share eligible for SSDI-related variables (approximate %)[1]	Share eligible for SSI-related variables (approximate %)[2]
Demographic	3.48 million	65	54
Annual-1994	0.95 million	49	52
Annual-1995	1.07 million	48	52
Annual-1996	1.19 million	48	53
Annual-1997	1.25 million	49	53
Annual-1998	1.29 million	50	53
Annual-1999	1.33 million	51	52
Annual-2000	1.38 million	50	51
Annual-2001	1.42 million	50	51
Annual-2002	1.49 million	51	51
Annual-2003	1.54 million	51	53
Annual-2004	1.60 million	53	43
Annual-2005	1.65 million	53	43
Annual-2006	1.69 million	53	43
Annual-2007	1.74 million	54	43
Annual-2008	1.79 million	54	42
Annual-2009	1.85 million	54	42
Annual-2010	1.90 million	55	43
Annual-2011	1.94 million	56	43
Annual-2012	1.97 million	57	42
Annual-2013	1.98 million	58	42
Annual-2014	1.98 million	58	42
Annual-2015	1.97 million	59	42
Annual-2016	1.94 million	59	42
Annual-2017	1.90 million	60	42
Annual-2018	1.83 million	60	42

[1] In the Demographic file, we set to missing SSDI-related variables for beneficiaries who do not have a populated initial date of SSDI eligibility (DOEI). In the Annuals files, we set to missing SSDI-related variables for beneficiaries who do not have a populated Ledger Account File status code for SSDI in the month (LAFyymm). Note that the number of non-missing values does not correspond to SSA published statistics of beneficiaries in the year because we include all non-missing LAF codes, including suspense and termination.

[2] In the Demographic file, we set to missing SSI-related variables for beneficiaries who do not have a populated initial date of SSI eligibility (SSIELIG_FIRST). In the Annuals files, we set to missing SSI-related variables for beneficiaries who do not have a populated payment status code for SSI in the month (PSTAyymm). We observed a change in the share of beneficiaries with a populated PSTA code starting in September 2003; virtually all of the cases in T31 and T51 (system-generated termination codes) in August 2003 had a missing PSTA code starting in September 2003, thus explaining the drop in the share with a non-missing PSTAyymm in 2004. Like SSDI, the number of non-missing PSTA values includes those in current pay as well as those with a suspense or termination code.

Each Annual file contains beneficiaries who were under FRA at any time during that year and had data in any of twelve monthly variables indicating program participation at some point during the year. [5] Since program participation changes over time, the sample size of each Annual file differs, but is smaller than the Demographic file, reflecting the fact that not all beneficiaries received benefits in every year. For example, the 2012 Annual file has about 1.96 million records (Table II.1). To construct a complete profile of the beneficiary's data over time, users need to merge the Annuals files across the multiple years of interest. In the Annuals components, the variables take on two forms: *yymm* variables and *yy* variables. The *yymm* variables indicate the year and month that the data reflects. For example, yymm=1205 contains data from May 2012. For purposes of building the PUF, in a few instances there are *yy* variables, where monthly occurrences were converted to annual occurrences. This approach was used for variables that rarely changed at the monthly level (e.g., education), as well as for variables where monthly patterns could have identified individuals (e.g., region of residence).

B. Selecting Which File to Use to Start a Research Project

In total, the PUF contains a single Demographic file and more than 20 Annual files. While some research projects may require combining data across all of these files, others may be completed using a subset of files. Identifying the best file to start with may help streamline your research plan.

The file you should use to start your research depends on the question you are trying to answer and the way in which you'd like to select beneficiaries (Table II.2). In general, questions that rely on identifying groups of beneficiaries based on their characteristics are best started in

[5] Beneficiary records are included in the Annual file if the person had data in any of twelve monthly variables during the year including LAFyymm, PSTAyymm, PAYDyymm, PAYSyymm, STWDIyymm, STWSSIyymm, STWCMyymm, BFWDIyymm, BFWSSIyymm, or BFWCMyymm.

the Demographic component. Questions that rely on identifying beneficiaries at a point in time are best started in the Annuals files. When merging multiple years of the Annuals files together, you may want to consider working with a limited number of variables, as the monthly occurrence of variables combined with many years of data quickly makes the file size unwieldy.

Table II.2. Components in the DAF PUF

	Demographic Component	Annuals Component
Number of files	One	More than 20 (one for each calendar year)
Number of records in each file	Just under 3.5 million	Fewer than the demographic component; ranging from under a million in 1994 to just under 2 million in 2013. There were about 1.94 million records in 2015, the year used for the frequencies in this document.
Start a research project with this file if…	Your selection criteria are based on beneficiary characteristics	Your selection criteria are based on a particular point in time
Examples of questions that this file can answer	• How many SSI beneficiaries were first eligible for benefits in 2002? • What is the distribution of disabling conditions among youth receiving SSI? • How many beneficiaries had a ticket assigned through the Ticket to Work program in each year from 2010 through 2015? • How many SSI and SSDI beneficiaries have a representative payee, and among those, what is the relationship between the beneficiary and their representative payee?	• How many SSDI beneficiaries had their benefits suspended for work in June 2010? • What was the average SSI benefit forgone for work in December of each year from 2000 through 2010? • How many SSDI beneficiaries in 2014 were expected to medically improve?

C. Critical Considerations for Working with Variables in the PUF

In addition to the issues highlighted in this section, users are strongly encouraged to consult Volume 3 of the DAF documentation available at https://www.ssa.gov/disabilityresearch/daf.html. Volume 3 highlights tips for working with the DAF data and is worth reading prior to beginning analysis. Users are also encouraged to consult the volume after starting analysis to ensure each data element used in the analysis is consistent with its intended use. Specifically, Volume 3 covers topics including:

- Determining monthly participation based on payment status variables
- Determining initial benefit entitlement date, first payment date, and disability onset date

- Working with benefit paid and benefit due amounts, for SSDI and SSI
- Working with beneficiary income data, for SSDI and SSI

In addition to the information provided in Volume 3 of the full documentation, this section highlights a few additional caveats and considerations for PUF users who may be less familiar with SSA administrative data, including the importance of considering the original administrative source of each variable and how to identify whether a variable applies to SSDI-only, SSI-only, or concurrent beneficiaries.

The administrative source of each variable matters! The data contained in the DAF are extracted or derived from other SSA administrative records which were designed for purposes of program administration rather than research.[6] As such, each source file has nuances that may affect the interpretation of particular variables. Some of these nuances include:

- **Variables from MBR.** The Master Beneficiary Record (MBR) contains information on beneficiaries enrolled in the SSDI or Old Age Survivors Insurance (OASI) programs. Many of the variables stored in the MBR are overwritten as the beneficiary's status changes, reflecting the most current information about the beneficiary. Beginning in May 2009, SSA began to record historical occurrences of some of the key information in the MBR. For example, the field TOC indicates the Type of Claim, or reason that the beneficiary is receiving benefits. Prior to May 2009, an SSDI beneficiary who reached FRA and transferred to OASI would have had the TOC field overwritten to reflect retirement instead of disability. Since May 2009, the TOC field in the MBR is historical, retaining both the current and previous TOC values, stored in the *n-variable* fields. In this example, it would be possible to identify both SSDI and OASI benefit receipt using the TOC if that transition occurred after May 2009.
- **Variables from the 831 & 832/833 files.** The 831 & 832/833 variables contain information collected at the time a determination for SSDI or SSI was made. The 831 Disability File contains records of initial medical decisions made by the DDS, as reported on the SSA-831 disability determination form. The 832/833 Disability files contain records of medical

[6] These include the SSI Longitudinal File (LF), the CER100% files, the Master Beneficiary Record (MBR), the Payment History Update System (PHUS), the Disabled Beneficiaries and Dependents (DBAD) file, the Disability Control File (DCF), Employment Network (EN) Payment Cumulative Payment Report, the Vocational Rehabilitation Reimbursement Management System (VRRMS), 831 & 832/833, and NUMIDENT.

decisions made in the course of Continuing Disability Reviews, with each new determination resulting in a new record. These files were used beginning in 1988, meaning information will be missing for beneficiaries who applied for benefits prior to this time as will any data collected from pre-1988 Continuing Disability Reviews completed for such individuals. Additionally, even among applicants since 1988, information is not available for all beneficiaries. Several variables will have missing information unless the information contained in the variable was necessary to complete the determination or review. For example, education, industry, and occupational information are only relevant to a determination for those beneficiaries whose application reached Step 5 of the sequential determination process.[7] Those allowed at Step 3 (if they meet or equal the impairment listings), are not asked to provide education, industry, and occupational information because it is not relevant for the decision. Consequently, data on such variables will be missing for the large segment of beneficiaries whose benefits were allowed at step 3.

- **Variables from the CER100% and DBAD files.** The variables from these files are monthly snapshots of SSI and SSDI program activity, extracted from the SSR and MBR, respectively. SSA creates these files on a monthly basis approximately one month after the data month (for example, SSA extracted data for June 2015 in July 2015) and SSA does not retroactively update these files. The snapshot aspect of the data means that the information contained the variables in these files for a given month may not reflect the most current picture of a beneficiary's status. These files are useful for showing beneficiary information and status as it looked in "real time" as SSA acted on the information available at the time benefits were paid. For example, a beneficiary whose January 2013 CER snapshot indicates no earned income will show the benefit status and payments that actually occurred for January 2013. If SSA discovers earnings six months later and inputs them to the SSR, from which the CER are extracted, the SSR will now show this new information for January 2013. If the earnings are large enough to affect the payment status and the payment amount due for that month, SSA will modify these values in the SSR to reflect SSA's latest information for this earlier month. In this case, the updated SSR values for the payment status and the amount due for the month will no longer match the CER values for these variables.

[7] The sequential determination process is: Step 1 decisions are denials only--working at substantial gainful activity (SGA); Step 2 decisions are denials only—impairments do not meet the severity requirements for eligibility; Step 3 decisions are allowances only--the applicant meets the criteria at steps 1 and 2 and has an impairment that meets or equals SSA's Listing of Impairments. Step 4 decisions are denials only—applicants who meet the criteria from the first three steps, but whose residual functional capacity means they are found able to perform past work. Step 5 decisions include both allowances and denials based on whether the applicant's residual functional capacity in conjunction with age, education and work experience mean that the applicant is found able to perform any work in the United States economy. While education and job information will be relevant at Step 4, Step 4 can only result in a denial so that information will not be in the DAF, because the DAF only includes those who were allowances and became beneficiaries.

The CER100% and DBAD files will also tend to undercount the number of beneficiaries in the SSI and SSDI programs. This is because SSA usually awards benefits retroactively. For example, a person awarded benefits in June of 2012 that are retroactive to January will not show up in the CER files for January through May of 2012. They will first show up in the July 2012 CER file. The SSR in July will now show benefits due but not paid in January through May of 2012 and will likely show a lump sum payment in July to adjust for these earlier missed payments. While retrospective updating occurs for both the SSI and SSDI programs, it is less of an issue for SSDI because SSDI program dynamics change less relative to SSI. Still researchers need to use caution in using and interpreting the data from these snapshots files.

More information about the source data used in the DAF can be found in Volume 9 of the DAF documentation, available at: https://www.ssa.gov/disabilityresearch/daf.html.

D. Issues Specific to Variables in the PUF

Special missing values. In the DAF PUF, we use several special missing values to provide information about why the data is likely missing. These values are:

- .N (or N for numeric variables)—not applicable, based on criteria identifying participation in the SSDI or SSI programs (see Table II.1)
- .M (or M for numeric variables)—applicable based on program participation, but no data available
- .P (or P for numeric variables)—data not yet populated
- .X (or X for numeric variables) –too few non-missing values so set to missing

We explain each of these cases below.

Determining whether a variable applies to SSDI or SSI. Because the DAF contains information on SSDI and SSI variables, not all of the information contained in it applies to all beneficiaries in all months. But, because beneficiaries enter and exit both programs, they may have data for SSDI in some months, SSI in some months, and some months with data in both programs. The DAF file includes variables for both the SSI program and the SSDI program regardless of the program in which the person participates each month. Those in SSI-only status

in a given month will have missing values for all SSDI variables in that month and those in SSDI only status will have all missing values for their SSI variables in that month. If data fields are populated that we would expect to be missing, we do not change those values. In other words, we did not overwrite any of the information housed in the administrative record, even if it seems illogical.

We distinguish between data that is missing and applicable and data that is missing and not applicable using two of the types of special missing characters described above. In particular, .N (or N for the case of numeric variables) denotes "not applicable." If a value for a variable is .N, it indicates that the data element was not relevant given the beneficiary's program participation. Specifically, if a monthly variable applies to SSI beneficiaries only, but the person did not receive SSI in the month, the value would be .N. We determined participation in SSI in a given month using the PSTAyymm variable; if PSTAyymm was missing, we determined the person was not an SSI beneficiary (and therefore SSI-related variables were not applicable). Similarly, a value would be .N for a data element that applies to SSDI beneficiaries only, in a month that the person was not an SSDI beneficiary. We determined participation in SSDI in a given month using the LAF*yymm* variable; if LAFyymm was missing, we determined the person was not an SSDI beneficiary (and therefore SSDI-related variables were not applicable). In addition, the PUF contains a special missing value of .M (or M for the case of numeric variables), which indicates "missing data" for other reasons. We used an .M value to indicate that the variable *could* have applied to the beneficiary based on program participation, but the person did not have data available. Because we determine the type of missing based on PSTA*yymm* for SSI and LAF*yymm* for SSDI, missing values for these variables can only be .N—if no data are present, the data do not apply by definition.

The criteria used to define these special missing characters depends on the particular data element and are detailed on the variable detail pages in Chapter IV.[8] We caution that the information we used to separate .M from .N may not be exactly correct, and is used instead to give users some sense of the reason that information might be missing. For most purposes, users can treat .M and .N as analogous for purposes of analyzing populated data. Additionally, in some cases, some variables had values that fell outside of the designated range and did not get captured into one of the special missing groups. These instances are rare and will simply be shown as missing ("."). Users should treat such cases as missing without trying to infer anything about the meaning.

Missing values indicating data not available. Some Annual variables in the PUF are not available in the earliest years of the Annuals files. For example, PSTyymm, the variable we used from the DAF to develop the region in which a beneficiary resided during the year, was not populated until 1996 and as such, REGION_PUFyy does not have values in 1994 and 1995. We have assigned a value of .P to variables in the years in the Annuals files that preceded the year in which data became available. The variable detail page for each variable indicates which years and months the data is populated.

Missing values indicating all values masked. For some monthly variables in the annual data where topcoding applies (described further below), we had fewer than 30 positive

[8] In general, we attempted to identify whether data was missing on a particular element because it was not applicable (for example, an SSI-specific measure for a beneficiary only receiving SSDI) or potentially applicable but missing (an SSI variable for a beneficiary who appeared to be receiving SSI). To do this for variables that are time-based (in the *yymm* format in the Annuals files), we used LAFyymm and PSTAyymm to determine program participation in a given month. For variables that were not time-based (in the DMG file), we used Date of Initial Eligibility (DOEI) for SSDI and the first date of eligibility (SSIELIG_FIRST) for SSI. We did not overwrite information that was populated for a given element based on these measures. In other words, if a variable applied only to SSI but our review indicated the person received SSDI, we did not overwrite what was populated. Rather, we only used these criteria to assign missing values to groups.

values. In such cases we had too few values to effectively top code the data and so we set all values to .X. This special missing indicates that, for this particular month, the data in the PUF is not useful and should be ignored in all analyses.

The PUF unique identifier. Beneficiaries are identified in the PUF using a unique, random identifier called the PUFPIN. This variable is constructed with a DAFyy prefix, where yy is the DAF year, to indicate that it should not be merged across versions of the PUF, but it can be merged across files from the same PUF version. Use the PUFPIN to merge DMG and Annuals files together to get records for the same beneficiary.

Data masking. To protect beneficiaries' identities, PUF data is masked in a variety of ways, as described below.

Dollar values. Exact values for income or benefits may make individuals identifiable. To address this privacy risk, we therefore mask each dollar value in the DAF PUF based on Census guidelines. In particular, we recoded dollar values between $1 and $7 to $4, rounded values between $8 and $999 to the nearest $10, rounded values between $1,000 and $49,000 to the nearest $100, and rounded values of $50,000 or higher to the nearest $1,000.

To make sure that extreme dollar values do not pose a disclosure risk, we applied top codes to the top ½ of 1 percent of positive values of all dollar variables. We then replaced the individual values for these top cases with the average value across those in the top coded group. For example, if there are 7,000 observations with a positive value for a given variable, there are 35 observations in the top ½ of 1 percent that we top coded. For each of those observations, we replaced the individual values for each of these observations with average value for these 35 cases. For some months and variables, relatively few of the populated values were positive. In particular, if there were fewer than 6,000 positive values for a variable, the top ½ of 1 percent

would yield fewer than 30 observations. Because we regarded 30 observations as the lower limit of cases to top code, we top coded the top 30 observations whenever there were fewer than 6,000 positive values for a variable. Consequently, the top-coded group will have more than ½ of 1 percent for these variables. In cases where there were fewer than 30 positive values for a variable in a month, the data cannot be effectively masked with top coding so we replaced all observations with a special missing character, .X. Each variable detail page in Chapter IV describes the modifications made as compared to the full DAF.

Date Variables. Dates can also pose a privacy risk. For all date variables, days were set to the 15th of the month to avoid identifying individuals based on their date of birth or the date they experienced particular outcomes. Additionally, in some cases, a lower bound was placed on the dates to minimize the possibility of identifying cases based on particularly early dates. Each variable detail page in Chapter IV describes the modifications made as compared to the full DAF.

Categorical values. There are other characteristics that may make a beneficiary highly visible that cannot be bottom or top coded because they are nonordinal. In these cases, the level of detail for the PUF was collapsed into larger categories. For example, primary and secondary diagnosis groups were aggregated into larger bins so that no group was particularly small or identifiable (see DXPRIBEST_PUF and DXSECBEST_PUF). Each variable detail page in Chapter IV describes the modifications made as compared to the full DAF. The values from the full DAF are shown in parentheses for each value in the PUF. For example, for BIC_PUF we modify the values for BIC from the full DAF as follows:

= A (BIC=A) means a BIC_PUF of A corresponds to a full DAF BIC of A

= C (BIC=C) means a BIC_PUF of C corresponds to a full DAF BIC of C

= O (BIC in other categories) means a BIC_PUF of O corresponds to all other values of BIC in the full DAF.

Location values. The only location information in the DAF PUF is in the REGION_PUFyy variables. Because these data are provided over time, we have masked the region data to avoid detection by way of movement patterns. For beneficiaries with more than two moves across regions in the DAF (three or more moves), representing less than one percent of beneficiaries, we masked each region of residence. Specifically, we assigned the location of residence for these individuals to be in a broadly similar area of the US to the actual location (Northeast, South, Midwest, and West; see Table II.3), but the masked value is randomly assigned to a region within that area. Thus, a person that moved from the northeast to the south to the Midwest would still be placed within regions in those geographic areas, but the specific regions assigned to that person would randomly differ from the true set of values. Because we retain this broad regional element even after masking, and because relatively few beneficiaries have more than two moves, the data retain their ability to support the examination of regional movements over time.

Table II.3. Mapping Broad Areas to SSA Regions for Purposes of Masking Move Patterns

Geographic Area	SSA Region	REGION_PUFyy values
Northeast	Boston, New York, Philadelphia	1, 2, 3
South	Atlanta, Dallas	4, 6
Midwest	Chicago, Kansas City	5, 7
West	Denver, San Francisco, Seattle	8, 9, 10

This page has been left blank for double-sided copying.

III. PUF VARIABLE LISTS

This chapter includes a list of variables available on the PUF Demographic and Annuals components. Additional information about these PUF variables can be found in Chapter IV. In addition to the tables in this chapter, users are strongly encouraged to consult Volume 4 of the full DAF documentation available at https://www.ssa.gov/disabilityresearch/daf.html.

Table III.1. Variables in the PUF Demographic Component

Variable Name	Variable Label	Variable Description
PUFPIN	Unique identifier formatted as DAFyy_PUFPIN	Variable that can be used to merge across files in the same version of the DAF PUF
BDOE_START_PUFn (n=1, 2)	SSDI DATE OF ENTITLEMENT START (n=1,2)	Entitlement start date for SSDI beneficiary.
BDOE_TERM_PUFn (n=1, 2)	SSDI DATE OF ENTITLEMENT TERMINIATION (n=1,2)	Entitlement termination date for SSDI beneficiary.
BDOF_PUFn (n=1, 2)	SSDI DATE OF FILING ENTRY (n=1,2)	Date SSDI beneficiary application was filed.
BIC_PUF	SSDI BENEFICIARY IDENTIFICATION CODE (BIC)	SSDI beneficiary identification code (identifies type of claimant).
CLM_FIL_PUF	SSI CLAIM FILED DATE	Date that an application for SSI benefits was received at SSA; different from the application date.
DD_PUFn (n=1-5)	MEDICAL CONTINUING DISABILITY REVIEW DATE (n=1-5)	Date of medical re-examination; data for five periods are recorded.
DOBBEST_PUF	BEST DATE OF BIRTH	Best available date of birth after reviewing multiple sources.
DODBEST_PUF	BEST DATE OF DEATH	Best date of death after reviewing multiple sources.
DODEC_PUFn (n=1 to 5)	DATE OF MEDICAL DECISION (SSA OR STATE DDS) (n=1-5)	Date of DDS or SSA Decision.
DOEC_PUF	SSDI MOST RECENT ENTITLEMENT DATE	The date of entitlement to SSDI that corresponds to the most recent period of disability.
DOEI_PUF	SSDI INITIAL ENTITLEMENT DATE	Date of initial entitlement gives the earliest month of entitlement to SSDI benefits and is considered to be the enrollment date.
DR_PUFn (n=1-5)	MEDICAL CONTINUING DISABILITY REVIEW REASON (n=1-5)	Reason code for medical re-examination; data for five re-examinations are recorded.
DT_PUFn (n=1-5)	MEDICAL CONTINUING DISABILITY REVIEW REASON TYPE (n=1-5)	Type code for medical re-examination; data for five re-examinations are recorded.
DXPRIBEST_PUF	PRIMARY DIAGNOSIS GROUP, MOST RECENT, BEST FROM ALL SOURCES	Primary diagnosis group, based on 4-digit diagnosis code.
DXSECBEST_PUF	SECONDARY DIAGNOSIS GROUP, MOST RECENT, BEST FROM ALL SOURCES	Secondary diagnosis group, based on 4-digit diagnosis code.
FIRSTDD_PUF	MEDICAL CONTINUING DISABILITY REVIEW DATE, EARLIEST	The date of the earliest medical re-examination.

Table III.1. Variables in the PUF Demographic Component

Variable Name	Variable Label	Variable Description
FIRSTDR_PUF	MEDICAL CONTINUING DISABILITY REVIEW REASON, EARLIEST	Reason code for earliest medical re-examination.
FIRSTDT_PUF	MEDICAL CONTINUING DISABILITY REVIEW REASON TYPE, EARLIEST	Type code for earliest medical re-examination.
FRA_PUF	BENEFICIARY'S FULL RETIREMENT AGE	Beneficiary's full retirement age
FRADATE_PUF	DATE WHEN FULL RETIREMENT AGE IS REACHED	The date the beneficiary reaches full retirement age.
MFT_PUF	SSI MASTER FILE TYPE	Identifies the person in terms of his/her program eligibility and is used to identify the type of SSI beneficiary.
PIA_MAX_PUF	SSDI PRIMARY INS. AMOUNT, MAXIMUM	The maximum primary insurance amount (PIA) after selecting 50 available earnings values in the DAF, with indexing applied to each earnings value.
PIA_MIN_PUF	SSDI PRIMARY INS. AMOUNT, MINIMUM	The minimum primary insurance amount (PIA) after selecting 50 available earnings values in the DAF, with indexing applied to each earnings value.
PIED_MAX_PUF	SSDI MAX PIA EFFECTIVE DATE	Earliest date associated with maximum primary insurance amount.
PIED_MIN_PUF	SSDI MIN PIA EFFECTIVE DATE	Most recent date associated with minimum primary insurance amount.
REPPAYTP_PUF	SSI TYPE OF PAYEE CODE	Relationship between the beneficiary and the representative payee
REPPYSSD_PUF	SSDI REPRESENTATIVE PAYEE INDICATOR	Indicates whether SSDI beneficiary has a representative payee.
REPPYSSI_PUF	SSI REPRESENTATIVE PAYEE INDICATOR	Indicates whether SSI beneficiary has a representative payee.
SEX_PUF	SEX	Sex of beneficiary
SSDI_ONSET_PUF	SSDI DISABILITY ONSET DATE EARLIEST PERIOD OF ELIGIBILITY	The earliest date of SSDI disability onset among all periods of SSDI disability.
SSIELIG_FIRST_PUF	SSI EARLIEST ELIGIBILITY DATE	The earliest date an individual is eligible for SSI.
TKT0205_PUF	TICKET ACTIVE STATUS IN YEARS (2002-2005)	Indicates whether a beneficiary had an active Ticket (assigned or in use) in the 2002-2005 period.
TKT_PUFyy (yy=06-18)	TICKET ACTIVE STATUS IN YEAR	Indicates whether a beneficiary had an active Ticket (assigned or in use) in the year, for years ranging from 2006 (06) to 2018 (18).
TOC_NUM_PUF	SSDI NUMBER OF TYPE OF CLAIM (TOC) OCCURRENCES	Number of occurrences of the Type of Claim (TOC) fields for SSDI beneficiary.
TOC_START_PUFn (n=1, 2)	SSDI TYPE OF CLAIM (TOC) START DATE (n=1,2)	Date the Type of Claim (TOC) is set
TOC_PUFn (n=1, 2)	SSDI TYPE OF CLAIM (N=1, 2)	Helps to explain BIC, and in conjunction with BIC, can be used to determine the beneficiary type.
TOP_PUF	SSDI TYPE OF PAYEE	Relationship between the beneficiary and the representative payee.
TWPCMPLMNTH1_PUF	FIRST SSDI TRIAL WORK PERIOD (TWP) COMPLETION MONTH	Trial Work Period completion date, first occurrence.

Table III.2. Variables in the PUF Annuals Component

Variable Name	Variable Label	Variable Description
PUFPIN	Unique identifier formatted as DAFyy_PUFPIN	Variable that can be used to merge across files in the same version of the DAF PUF
BFWCM_PUFyymm	COMBINED (SSDI/SSI) ESTIMATED BENEFITS FORGONE FOR WORK (DRAFT) (yymm = year/month)	Indicates the amount of combined SSI and SSDI benefits forgone when beneficiary is in non-payment status due to suspense or termination for work in the SSDI program and/or not receiving the full SSI Federal Benefit Rate (FBR) due to earnings from work.
BFWDI_PUFyymm	SSDI ESTIMATED BENEFITS FORGONE FOR WORK (yymm = year/month)	Indicates the amount of SSDI benefits forgone when beneficiary is in non-payment status due to suspense or termination for work in the month indicated.
BFWSSI_PUFyymm	SSI ESTIMATED BENEFITS FORGONE FOR WORK (DRAFT) (yymm = year/month)	Indicates the amount of SSI benefits forgone when beneficiary is in non-payment status due to suspense or termination for work in the SSI program or is not receiving the full SSI Federal Benefit Rate (FBR) due to earnings from work in the month indicated.
CDR_PUFyymm	SSDI CESSATION OF DISABILITY REASON (yymm = year/month)	Indicates the cessation reason that applies in the month indicated.
DIERNLVL_PUFyymm	SSDI EARNING LEVEL INDICATOR	A categorical variable that indicates the level of earnings, in the month indicated, for SSDI beneficiaries based on earnings, self-employment earnings, and income exclusions. The variable is then related to the Trial Work Period (TWP) and Substantial Gainful Activity (SGA) levels.
DIRPAY_PUFyymm	SSDI ACTUAL BENEFIT PAID AMOUNT (yymm = year/month)	Actual amount of SSDI benefits disbursed in the month indicated.
DUED_PUFyymm	SSDI BENEFIT DUE (yymm = year/month)	Indicates the benefit amount due to beneficiary, after rounding, but before obligations such as Supplemental Medical Insurance (SMI) premium, in the month indicated.
DUES_PUFyymm	SSI BENEFIT DUE (yymm = year/month)	Indicates the total (federal and state) SSI benefit (expected amount of check to be paid) in the month indicated.
EDX_PUFyy	EDUCATION LEVEL in the last month observed during the calendar year	This variable gives the beneficiary's years of education as of the last month recorded during the calendar year.
EICM_PUFyymm	SSI COUNTABLE EARNED INCOME (yymm = year/month)	The amount of earned income used to for the SSI payment due calculation in the month indicated. It is generally half of actual earned income, minus deductions and exclusions.
LAF_PUFyymm	SSDI LEDGER ACCOUNT FILE (LAF)- PARTICIPATION STATUS (yymm = year/month)	SSDI Ledger Account File (LAF) reflects the MBR payment status in the month indicated.
LIVF_PUFyymm	SSI LIVING ARRANGEMENT CODE (yymm = year/month)	Identifies the living arrangement for a beneficiary in the month indicated and is used to adjust payments.
MEDEXyymm	MEDICAL IMPROVEMENT CODE (yymm = year/month)	Medical improvement indicator, in the month indicated.
MEDE_PUFyymm	SSI MEDICAID ELIGIBILITY (yymm = year/month)	Indicates whether an SSI beneficiary is eligible for Medicaid in the month indicated (as of the month indicated).

Table III.2. Variables in the PUF Annuals Component

Variable Name	Variable Label	Variable Description
MEDPREM_PUFyymm	SSDI PHUS MEDICARE PREMIUM (yymm = year/month)	Medicare Premium amount paid on behalf of the beneficiary in the month indicated.
MEDR_PUFyymm	SSDI MEDICARE ELIGIBILITY (yymm = year/month)	Medicare eligibility for SSDI in the month indicated.
PAYD_PUFyymm	SSDI PHUS BENEFIT PAID (yymm = year/month)	The sum of the direct pay (DIRPAYyymm) and the Medicare payment (MEDPREMyymm) fields for the SSDI beneficiary in the month indicated.
PAYO_PUFyymm	SSDI PHUS DEPENDENT BENEFIT PAID (yymm = year/month)	The payment made for the SSDI dependents to a primary beneficiary in the month indicated.
PAYS_PUFyymm	SSI BENEFIT PAID (yymm = year/month)	Total SSI benefit paid in the month indicated.
PROA_PUFyy	SSI USE OF 1619A PROVISION INDICATOR (yymm = year/month)	Indicates the use of provision 1619a in at least one month of the calendar year; provision 1619a indicates the continuance of SSI benefit receipt though working at SGA level.
PROB_PUFyy	SSI USE OF 1619B PROVISION INDICATOR (yymm = year/month)	Indicates whether the 1619b continuing Medicaid eligibility provision was used in the month indicated. Indicates the use of provision 1619b in at least one month of the calendar year; provision 1619b indicates the continuance of Medicaid receipt though working at SGA level.
PSTA_PUFyymm	SSI PAYMENT STATUS (PSTAT) (yymm = year/month)	Indicates payment, suspension, termination, etc. in the month indicated. It can be compared to the date of enrollment to determine prior suspensions and terminations.
REGION_PUFyy	SSA REGION NUMBER (YY=YEAR)	Constructed measure derived from using the last monthly occurrence of state of residence reported during the year.
SSICOUPL_PUFyymm	SSI COUPLE INDICATOR (yymm = year/month)	Indicates whether SSI beneficiaries are part of a couple in the month indicated.
STWCM_PUFyymm	COMBINED (SSDI/SSI) SUSPENSE OR TERMINATION DUE TO WORK (yymm = year/month)	Beneficiary's least restrictive non-payment status due to suspense or termination for work in the month indicated.
STWDI_PUFyymm	SSDI SUSPENSE OR TERMINATION DUE TO WORK (yymm = year/month)	Beneficiary's non-payment status due to suspense or termination for work in the SSDI program in the month indicated.
STWSSI_PUFyymm	SSI SUSPENSE OR TERMINATION DUE TO WORK (yymm = year/month)	Beneficiary's non-payment status due to suspense or termination for work in the SSI program in the month indicated.
T16GRSAMT_PUFyymm	SSI EARNINGS GROSS AMOUNT (DCF) (yymm = year/month)	SSI gross earnings in the month indicated.
T16NETAMT_PUFyymm	SSI SELF-EMP NET INCOME AMOUNT (DCF) (yymm = year/month)	Net amount of self-employment earnings in the month indicated.
T2GRSAMT_PUFyymm	SSDI EARNINGS GROSS AMOUNT (DCF) (yymm = year/month)	SSDI gross earnings in the month indicated.
T2NETAMT_PUFyymm	SSDI SELF-EMP NET INCOME AMOUNT (DCF) (yymm = year/month)	The net amount of monthly self-employment earnings in the month indicated.
TWPDATA_PUFyymm	SSDI TRIAL WORK PERIOD CODE (yymm = year/month)	Identifies whether earnings are considered as a trial work month or as SGA in the month indicated.
UINC_PUFyymm	SSI COUNTABLE UNEARNED INCOME (yymm = year/month)	Unearned income minus deductions and exclusions in the month indicated.

IV. PUF VARIABLE DETAIL PAGES

This chapter provides detail variable information for each variable in the Demographic and Annuals file. In addition to the variable detail pages in this chapter, users are strongly encouraged to consult Volume 5 of the DAF documentation available at https://www.ssa.gov/disabilityresearch/daf.html.

The following list explains the variable fields in this volume.

Variable Name. The SAS variable name used in the PUF.

Variable Label. The SAS label assigned to the variable.

Variable Description. The text description of the variable.

Data Type. Variables are standardized to the following possible data types:

- Binary (0/1) – numeric 0/1 flag
- Categorical – variables that can be used to group (categorize) observations but that are not binary (e.g., values such as E or S for language or 1-9 for Type of Claim)
- Continuous – numeric variables, other than dollar amounts and dates, that can be any number and with which arithmetic computations can be performed (e.g., number of dependents)
- Dollars – numeric variables that represent dollar amounts
- Date – numeric field that represents the number of days before or after January 1, 1960 (formerly called SAS Date)

Data Available For. This field indicates to which type of beneficiary the variable applies. Some variables apply only to SSI beneficiaries, some only to SSDI beneficiaries, and some to both.

Variable Location. This field indicates whether the data element is contained in the Demographic or Annuals component of the DAF PUF.

DAF Source Variable. This field indicates the DAF variable from which the PUF variable was derived. This field also indicates the SSA administrative source file of the DAF variable.

When constructed variables have multiple data sources, all are listed. Data in the DAF are extracted from various SSA data files, including the SSI-LF, the CER, the MBR, the PHUS, the DBAD, the DCF, Ticket payments, 831 & 832/833, NUMIDENT, and VRRMS.

Related PUF Variables. This field indicates other PUF variables to which the detailed variable is related. This is not an exhaustive list, but rather a list of the most related PUF variables.

Value categories in the PUF. This field provides detail about the variable values. Information about approximate distribution of the variable values is also provided, where applicable. The frequencies shown for each variable are only meant to be illustrative of the distribution of values users might expect to see. These were drawn from DAF18 data; as future iterations of the DAF PUF are developed and SSA data are updated, we expect slight changes in these distributions.

Notes. This field provides additional information or context, such as limits on variable use or further refinement of variable values.

In addition to the standard data fields listed above, the terms in Table IV.1 are commonly used in variable names and elsewhere in this volume.

Table IV.1. List of commonly used terms in variable names and DAF documentation

Variable Name	Reference
BEST	The best available approximation of data for which there may be conflicting values in SSA administrative files
DX	Diagnosis
ELG or ELIG	Eligibility or Eligible
FIRST	Indicates the earliest available occurrence of a particular event described in DAF data
MAX	Indicates the most recent occurrence of a particular event described in DAF data
MFT	Master File Type
MIN	Indicates the earliest available occurrence of a particular event described in DAF data
TOA	Type of Action
yymm	Year and month—the format for many DAF date variables (some variables have a yy-only format if they are annual)

DAF PUF Variable Detail Pages: Demographic File

Note: The frequencies shown for each variable are only meant to be illustrative of the distribution of values users might expect to see and are rounded, so may not sum to exactly 100 percent. These were drawn from DAF18 data; as future iterations of the DAF PUF are developed and SSA data are updated, we expect slight changes in these distributions.

Variable Name	BDOE_START_PUFn (n=1, 2)
Variable Label	SSDI DATE OF ENTITLEMENT START (n=1,2)
Variable Description	Entitlement start date for SSDI beneficiary.
Data Type	Date
Data Available For	SSDI
Variable Location	Demographic Component
DAF Source Variable	BDOE_STARTn; DAF data extracted from the MBR
Related PUF Variables	BDOF_PUFn; BDOE_TERM_PUFn
Value categories in the PUF	**Approximate distribution**
	BDOE_START_PUF1
= Masked date	64.7
= .M **(**missing, may be applicable)	0.7
= .N (missing, not applicable)	34.6
	BDOE_START_PUF2
= Masked date	24.2
= .M **(**missing, may be applicable)	41.2
= .N (missing, not applicable)	34.6
Notes	Masked date range limited to the last 45 years for privacy reasons. Special missing condition (.M): No data available, but applicable to the beneficiary having a populated date of initial SSDI eligibility (DOEI). Special missing condition (.N): No data available, and not applicable to the beneficiary because the initial SSDI eligibility date (DOEI) is not populated.

Variable Name	BDOE_TERM_PUFn (n=1, 2)
Variable Label	SSDI DATE OF ENTITLEMENT TERMINIATION (n=1,2)
Variable Description	Entitlement termination date for SSDI beneficiary.
Data Type	Date
Data Available For	SSDI
Variable Location	Demographic Component
DAF Source Variable	BDOE_TERMn; DAF data extracted from the MBR
Related PUF Variables	BDOF_PUFn; BDOE_START_PUFn
Value Categories in the PUF	**Approximate Distribution**
	BDOE_TERM_PUF1
= Masked date	39.6
= .M **(**missing, may be applicable)	25.8
= .N (missing, not applicable)	34.6
	BDOE_TERM_PUF2
= Masked date	9.9
= .M **(**missing, may be applicable)	55.4
= .N (missing, not applicable)	34.6
Notes	Masked date range limited to the last 35 years for privacy reasons. Special missing condition (.M): No data available, but applicable to the beneficiary having a populated start date of SSDI entitlement (BDOE_STARTn) or a populated date of initial SSDI eligibility (DOEI). Special missing condition (.N): No data available, and not applicable to the beneficiary because the start date of SSDI entitlement (BDOE_STARTn) and the date of initial SSDI eligibility (DOEI) are both not populated.

Variable Name	BDOF_PUFn (n=1, 2)
Variable Label	SSDI DATE OF FILING ENTRY (n=1,2)
Variable Description	Date SSDI beneficiary application was filed.
Data Type	Date
Data Available For	SSDI
Variable Location	Demographic Component
DAF Source Variable	BDOFn; DAF data extracted from the MBR
Related PUF Variables	BDOE_START_PUFn; BDOE_TERM_PUFn
Value Categories in the PUF	**Approximate Distribution**
	BDOF_PUF1
= Masked date	64.8
= .M **(**missing, may be applicable)	0.6
= .N (missing, not applicable)	34.6
	BDOF_PUF2
= Masked date	24.2
= .M **(**missing, may be applicable)	41.2
= .N (missing, not applicable)	34.6
Notes	Masked date range limited to the last 50 years for privacy reasons. Special missing condition (.M): No data available, but applicable to the beneficiary having a populated date of initial SSDI eligibility (DOEI). Special missing condition (.N): No data available, and not applicable to the beneficiary because the initial SSDI eligibility date (DOEI) is not populated.

Variable Name	BIC_PUF
Variable Label	SSDI BENEFICIARY IDENTIFICATION CODE (BIC)
Variable Description	SSDI beneficiary identification code (identifies type of claimant).
Data Type	Categorical
Data Available For	SSDI
Variable Location	Demographic Component
DAF Source Variable	BIC; DAF data extracted from the MBR.
Related PUF Variables	MFT_PUF; LAF_PUFyymm; TOC_PUFn
Value Categories in the PUF	**Approximate Distribution**
= A (BIC=A)	60.1
= C (BIC=C)	4.2
= O (BIC in other categories)	1.2
= .M missing (may be applicable)	0.0
= .N missing (not applicable)	34.6
Notes	Special missing condition (.M): No data available, but applicable to the beneficiary having a populated date of initial SSDI eligibility (DOEI). Special missing condition (.N): No data available, and not applicable to the beneficiary because the initial SSDI eligibility date (DOEI) is not populated. First Character = Type of Benefit. Common first characters in the DAF: A = Primary claimant/ number holder B = Spouse of the number holder C = Child of the number holder D = Widow/widower of the number holder (at least 60 years old) E = Young widow caring for a child of the number holder/worker. Children must be under 16 or became disabled before 18 years. F = Parents of the number holder J, K = Prouty beneficiaries M, T = Hospital Insurance only involved W = Disabled widow/widower of number holder (at least 50 years old) Second Character = Benefit Description (if first character is not sufficient to identify the type of beneficiary)

Variable Name	CLM_FIL_PUF
Variable Label	SSI CLAIM FILED DATE
Variable Description	Date that an application for SSI benefits was received at SSA; different from the application date.
Data Type	Date
Data Available For	SSI
Variable Location	Demographic Component
DAF Source Variable	CLM_FIL; DAF data extracted from the SSI-LF
Related PUF Variables	None
Value categories in the PUF	**Approximate distribution**
= Masked date	44.9
= .M **(**missing, may be applicable)	12.1
= .N (missing, not applicable)	43.0
Notes	Masked date range limited to the last 30 years for privacy reasons. Special missing condition (.M): No data available, but applicable to the beneficiary having a populated date of initial SSI eligibility (SSIELIG_FIRST). Special missing condition (.N): No data available, and not applicable to the beneficiary because the initial SSI eligibility date (SSIELIG_FIRST) is not populated.

Variable Name	DD_PUFn (n=1-5)
Variable Label	MEDICAL CONTINUING DISABILITY REVIEW DATE (n=1-5)
Variable Description	Date of medical re-examination; data for up to five periods are recorded.
Data Type	Date
Data Available For	SSI/SSDI
Variable Location	Demographic Component
DAF Source Variable	DDn; DAF data extracted from the 831 & 832/833
Related PUF Variables	DR_PUFn; DT_PUFn; FIRSTDD_PUF; FIRSTDR_PUF; FIRSTDT_PUF
Value Categories in the PUF	**Approximate Distribution**
	DD_PUF1
= Masked date	53.1
= .M **(**missing, may be applicable)	44.7
= .N (missing, not applicable)	2.2
	DD_PUF2
= Masked date	31.6
= .M **(**missing, may be applicable)	66.1
= .N (missing, not applicable)	2.4
	DD_PUF3
= Masked date	20.7
= .M **(**missing, may be applicable)	76.9
= .N (missing, not applicable)	2.4
	DD_PUF4
= Masked date	11.7
= .M **(**missing, may be applicable)	85.9
= .N (missing, not applicable)	2.4
	DD_PUF5
= Masked date	8.4
= .M **(**missing, may be applicable)	89.2
= .N (missing, not applicable)	2.4
Notes	Masked date range limited to dates after 1/15/1988, reflecting the start of the 831 & 832/833 files. Special missing condition (.M): No data available, but applicable to the beneficiary having a populated date of initial SSDI or SSI eligibility (DOEI or SSIELIG_FIRST).

Variable Name	DD_PUFn (n=1-5)
	Special missing condition (.N): No data available, and not applicable to the beneficiary because the initial SSDI and SSI eligibility date (DOEI and SSIELIG_FIRST) are not populated.

Variable Name	DOBBEST_PUF
Variable Label	BEST DATE OF BIRTH
Variable Description	Best available date of birth after reviewing multiple sources.
Data Type	Date
Data Available For	SSI/SSDI
Variable Location	Demographic Component
DAF Source Variable	DOBBEST, as constructed in the DAF using information from multiple sources
Related PUF Variables	FRADATE_PUF; DODBEST_PUF
Value Categories in the PUF	**Approximate Distribution**
= Masked date	100.0
Notes	Masked date range limited to dates after 1/15/1930, the earliest date of birth a person could have to be included in the DAF. Dates outside of this range (due to data errors) set to missing.

Variable Name	DODBEST_PUF
Variable Label	BEST DATE OF DEATH
Variable Description	Best date of death after reviewing multiple sources.
Data Type	Date
Data Available For	SSI and SSDI beneficiaries who have a date of death recorded
Variable Location	Demographic Component
DAF Source Variable	DODBEST, as constructed in the DAF using information from multiple sources
Related PUF Variables	DOBBEST_PUF
Value Categories in the PUF	**Approximate Distribution**
= Masked date	30.8
= .M **(**missing, may be applicable)	69.2
Notes	Masked date range limited dates prior to 1/15/1996, the earliest date a person included in the DAF could have died. Special missing condition (.M): No data available.

Variable Name	DODEC_PUFn (n=1 to 5)
Variable Label	DATE OF MEDICAL DECISION (SSA OR STATE DDS) (n=1-5)
Variable Description	Date of DDS or SSA Decision.
Data Type	Date
Data Available For	SSI/SSDI
Variable Location	Demographic Component
DAF Source Variable	DODECn, as constructed in the DAF; source data extracted from the 831 & 832/833
Related PUF Variables	None
Value Categories in the PUF	**Approximate Distribution**
	DODEC_PUF1
= Masked date	95.2
= .M (missing, may be applicable)	4.0
= .N (missing, not applicable)	0.8
	DODEC_PUF2
= Masked date	65.6
= .M (missing, may be applicable)	33.0
= .N (missing, not applicable)	1.4
	DODEC_PUF3
= Masked date	40.2
= .M (missing, may be applicable)	58.0
= .N (missing, not applicable)	1.8
	DODEC_PUF4
= Masked date	27.8
= .M (missing, may be applicable)	70.2
= .N (missing, not applicable)	2.1
	DODEC_PUF5
= Masked date	16.7
= .M (missing, may be applicable)	81.2
= .N (missing, not applicable)	2.1
Notes	Masked date range limited to dates after 1/15/1988, reflecting the start of the 831 & 832/833 files. Special missing condition (.M): No data available, but applicable to the beneficiary having a populated date of initial SSDI or SSI eligibility (DOEI or SSIELIG_FIRST).

Variable Name	DODEC_PUFn (n=1 to 5)
	Special missing condition (.N): No data available, and not applicable to the beneficiary because the initial SSDI and SSI eligibility date (DOEI and SSIELIG_FIRST) are not populated.

Variable Name	DOEC_PUF
Variable Label	SSDI MOST RECENT ENTITLEMENT DATE
Variable Description	The date of entitlement to SSDI that corresponds to the most recent period of disability.
Data Type	Date
Data Available For	SSDI
Variable Location	Demographic Component
DAF Source Variable	DOEC; DAF source data extracted from the MBR
Related PUF Variables	DOEI_PUF
Value Categories in the PUF	**Approximate Distribution**
=Masked date	65.4
=.M (missing, may be applicable)	0.0
=.N (missing, not applicable)	34.6
Notes	Masked date range limited to the last 50 years for privacy reasons. Special missing condition (.M): No data available, but applicable to the beneficiary having a populated date of initial SSDI eligibility (DOEI). Special missing condition (.N): No data available, and not applicable to the beneficiary because the initial SSDI eligibility date (DOEI) is not populated.

Variable Name	DOEI_PUF
Variable Label	SSDI INITIAL ENTITLEMENT DATE
Variable Description	Date of initial entitlement gives the earliest month of entitlement to SSDI benefits and is considered to be the enrollment date.
Data Type	Date
Data Available For	SSDI
Variable Location	Demographic Component
DAF Source Variable	DOEI; DAF data extracted from the MBR
Related PUF Variables	DOEC_PUF
Value Categories in the PUF	**Approximate Distribution**
=Masked date	65.4
=.N (missing, not applicable)	34.6
Notes	Masked date range limited to the last 50 years for privacy reasons. Special missing condition (.N): No data available.

Variable Name	DR_PUFn (n=1-5)
Variable Label	MEDICAL CONTINUING DISABILITY REVIEW REASON (n=1-5)
Variable Description	Reason code for medical re-examination; data for five re-examinations are recorded.
Data Type	Categorical
Data Available For	SSI/SSDI
Variable Location	Demographic Component
DAF Source Variable	DRn; DAF data extracted from the 831 & 832/833
Related PUF Variables	DD_PUFn; DT_PUFn; FIRSTDD_PUF; FIRSTDR_PUF; FIRSTDT_PUF
Value Categories in the PUF	**Approximate Distribution**
	DR_PUF1
=MIE (DRn=A-U, Y, Z)	5.6
=MINE (DRn=5, 7)	24.4
=MIP (DRn=3)	23.1
=.M (missing, may be applicable)	44.8
=.N (missing, not applicable)	2.2
	DR_PUF2
=MIE (DRn=A-U, Y, Z)	2.2
=MINE (DRn=5, 7)	11.1
=MIP (DRn=3)	15.3
=.M (missing, may be applicable)	69.0
=.N (missing, not applicable)	2.4
	DR_PUF3
=MIE (DRn=A-U, Y, Z)	1.1
=MINE (DRn=5, 7)	6.2
=MIP (DRn=3)	10.9
=.M (missing, may be applicable)	79.4
=.N (missing, not applicable)	2.4
	DR_PUF4
=MIE (DRn=A-U, Y, Z)	0.6
=MINE (DRn=5, 7)	3.5
=MIP (DRn=3)	6.2
=.M (missing, may be applicable)	87.2
=.N (missing, not applicable)	2.4

Variable Name	DR_PUFn (n=1-5)
	DR_PUF5
=MIE (DRn=A-U, Y, Z)	0.4
=MINE (DRn=5, 7)	2.4
=MIP (DRn=3)	4.8
=.M (missing, may be applicable)	90.0
=.N (missing, not applicable)	2.4
Notes	Special missing condition (.M): No data available, but applicable to the beneficiary having a populated date of initial SSDI or SSI eligibility (DOEI or SSIELIG_FIRST). Special missing condition (.N): No data available, and not applicable to the beneficiary because the initial SSDI or SSI eligibility dates (DOEI or SSIELIG_FIRST) are not populated. DRn values in the DAF: A-U, Y and Z (MIE) 3 = Periodic Review (3 Years) (MIP) [V, W or X and (PDI NOT = P)] (MIP) 5 = Periodic Review (5 Years) (MINE) 7 = Periodic Review (7 Years) (MINE) [V, W or X and (PDI = P)] (MINE) 9 = Unknown (MIP) Blank = Not Available (MIP)

Variable Name	DT_PUFn (n=1-5)
Variable Label	MEDICAL CONTINUING DISABILITY REVIEW REASON TYPE (n=1-5)
Variable Description	Type code for medical re-examination; data for five re-examinations are recorded.
Data Type	Categorical
Data Available For	SSI/SSDI
Variable Location	Demographic Component
DAF Source Variable	DTn; DAF data extracted from the 831 & 832/833
Related PUF Variables	DD_PUFn; DR_PUFn; FIRSTDD_PUF; FIRSTDR_PUF; FIRSTDT_PUF
Value Categories in the PUF	**Approximate Distribution**
	DT_PUF1
=MIEP (DTn=3)	28.3
=MINE (DTn=4)	24.4
=.M (missing, may be applicable)	45.1
=.N (missing, not applicable)	2.2
	DT_PUF2
=MIEP (DTn=3)	17.3
=MINE (DTn=4)	11.2
=.M (missing, may be applicable)	69.2
=.N (missing, not applicable)	2.4
	DT_PUF3
=MIEP (DTn=3)	11.2
=MINE (DTn=4)	6.7
=.M (missing, may be applicable)	79.7
=.N (missing, not applicable)	2.4
	DT_PUF4
=MIEP (DTn=3)	6.7
=MINE (DTn=4)	3.6
=.M (missing, may be applicable)	87.3
=.N (missing, not applicable)	2.4
	DT_PUF5
=MIEP (DTn=3)	4.6
=MINE (DTn=4)	2.9
=.M (missing, may be applicable)	90.1
=.N (missing, not applicable)	2.4

Variable Name	DT_PUFn (n=1-5)
Notes	Special missing condition (.M): No data available, but applicable to the beneficiary having a populated date of initial SSDI or SSI eligibility (DOEI or SSIELIG_FIRST). Special missing condition (.N): No data available, and not applicable to the beneficiary because the initial SSDI or SSI eligibility dates (DOEI or SSIELIG_FIRST) are not populated. DTn value range in the DAF: 3 = Medical Reexam, Nonpermanent (MIE) and (MIP) 4 = Medical Reexam, Permanent (MINE) 5 = Medical Reexam (Permanency Of Impairment Not Known) 6 = No Diary Required 9 = Title XVI VR Child Referral Z = Invalid

Variable Name	DXPRIBEST_PUF
Variable Label	PRIMARY DIAGNOSIS GROUP, MOST RECENT, BEST FROM ALL SOURCES.
Variable Description	Primary diagnosis group, based on 4-digit diagnosis code.
Data Type	Categorical
Data Available For	SSI/SSDI
Variable Location	Demographic Component
DAF Source Variable	DXPRIBEST, as constructed in the DAF; source data extracted from the MBR and 831 & 832/833
Related PUF Variables	DXSECBEST
Value categories in the PUF	**Approximate distribution**
= 1	5.0
= 2	7.2
= 3	20.7
= 4	26.8
= 5	21.5
= 6	10.3
= 7	2.6
= 8	5.7
= .M (missing, may be applicable)	0.1
=.N (missing, not applicable)	0.0
Notes	Special missing condition (.M): No data available, but applicable to the beneficiary having a populated date of initial SSDI or SSI eligibility (DOEI or SSIELIG_FIRST). Special missing condition (.N): No data available, and not applicable to the beneficiary because the initial SSDI or SSI eligibility dates (DOEI and SSIELIG_FIRST) are not populated. DXPRIBEST value category descriptions: 1 = autistic disorders; developmental disorders; or childhood and adolescent disorders not elsewhere classified 2 = intellectual disability 3 = mood disorders; organic mental disorders; schizophrenic and other psychotic disorders; or other mental disorders 4 = endocrine, nutritional, and metabolic

Variable Name	DXPRIBEST_PUF
	diseases; circulatory system; digestive system; genitourinary system; nervous system and sense organs; or respiratory system 5 = musculoskeletal system and connective tissue 6 = infectious and parasitic diseases or injuries 7 = congenital anomalies; blood and blood-forming organs; skin and subcutaneous tissue; or other 8 = populated with an unknown value. Group 8 consists of those with populated diagnosis codes, but in the vast majority of those cases, the code listed indicates that the diagnosis is unknown, that there is no/insufficient medical evidence in the file. In most cases, users should consider group 8 with the =.M group.

Variable Name	DXSECBEST_PUF
Variable Label	SECONDARY DIAGNOSIS GROUP, MOST RECENT, BEST FROM ALL SOURCES
Variable Description	Secondary diagnosis group, based on 4-digit diagnosis code.
Data Type	Categorical
Data Available For	SSI/SSDI
Variable Location	Demographic Component
DAF Source Variable	DXSECBEST, as constructed in the DAF; source data extracted from the MBR and 831 & 832/833
Related PUF Variables	DXPRIBEST
Value Categories in the PUF	**Approximate Distribution**
= 1	3.6
= 2	0.9
= 3	17.2
= 4	20.8
= 5	10.9
= 6	3.4
= 7	1.1
= 8	42.0
= .M (missing, may be applicable)	0.1
=.N (missing, not applicable)	0.0
Notes	Special missing condition (.M): No data available, but applicable to the beneficiary having a populated date of initial SSDI or SSI eligibility (DOEI or SSIELIG_FIRST). Special missing condition (.N): No data available, and not applicable to the beneficiary because the initial SSDI or SSI eligibility dates (DOEI and SSIELIG_FIRST) are not populated. DXSECBEST value category descriptions: 1 = autistic disorders; developmental disorders; or childhood and adolescent disorders not elsewhere classified 2 = intellectual disability 3 = mood disorders; organic mental disorders; schizophrenic and other psychotic disorders; or other mental disorders 4 = endocrine, nutritional, and metabolic diseases; circulatory system; digestive system; genitourinary

Variable Name	DXSECBEST_PUF
	system; nervous system and sense organs; or respiratory system 5 = musculoskeletal system and connective tissue 6 = infectious and parasitic diseases or injuries 7 = congenital anomalies; blood and blood-forming organs; skin and subcutaneous tissue; or other 8 = populated with an unknown value. Group 8 consists of those with populated diagnosis codes, but in the vast majority of those cases, the code listed indicates that the diagnosis is unknown, that there is no/insufficient medical evidence in the file. In most cases, users should consider group 8 with the =.M group.

Variable Name	FIRSTDD_PUF
Variable Label	MEDICAL CONTINUING DISABILITY REVIEW DATE, EARLIEST
Variable Description	The date of the earliest medical re-examination.
Data Type	Date
Data Available For	SSI/SSDI
Variable Location	Demographic Component
DAF Source Variable	FIRSTDD; DAF data extracted from the 831 & 832/833
Related PUF Variables	DD_PUFn; DR_PUFn; DT_PUFn; FIRSTDR_PUF; FIRSTDT_PUF
Value Categories in the PUF	**Approximate distribution**
= Masked date	53.1
= .M (missing, may be applicable)	44.7
= .N (missing, not applicable)	2.2
Notes	Masked date range limited to dates after 1/15/1988, reflecting the start of the 831 & 832/833 files. Special missing condition (.M): No data available, but applicable to the beneficiary having a populated date of initial SSDI or SSI eligibility (DOEI or SSIELIG_FIRST). Special missing condition (.N): No data available, and not applicable to the beneficiary because the initial SSDI and SSI eligibility date (DOEI and SSIELIG_FIRST) are not populated.

Variable Name	FIRSTDR_PUF
Variable Label	MEDICAL CONTINUING DISABILITY REVIEW REASON, EARLIEST
Variable Description	Reason code for earliest medical re-examination.
Data Type	Categorical
Data Available For	SSI/SSDI
Variable Location	Dem
DAF Source Variable	FIRSTDR; DAF data extracted from the 831 & 832/833
Related PUF Variables	DD_PUFn; DR_PUFn; DT_PUFn; FIRSTDD_PUF; FIRSTDT_PUF
Value Categories in the PUF	**Approximate Distribution**
= MIE (FIRSTDR=A-U, Y and Z)	5.6
= MINE (FIRSTDR=5 or 7)	24.4
= MIP (FIRSTDR=3)	23.1
= .M (missing, may be applicable)	44.8
= .N (missing, not applicable)	2.2
Notes	Special missing condition (.M): No data available, but applicable to the beneficiary having a populated date of initial SSDI or SSI eligibility (DOEI or SSIELIG_FIRST). Special missing condition (.N): No data available, and not applicable to the beneficiary because the initial SSDI or SSI eligibility dates (DOEI or SSIELIG_FIRST) are not populated. FIRSTDR values in the DAF: A-U, Y and Z (MIE) 3 = Periodic Review (3 Years) (MIP) [V, W or X and (PDI NOT = P)] (MIP) 5 = Periodic Review (5 Years) (MINE) 7 = Periodic Review (7 Years) (MINE) [V, W or X and (PDI = P)] (MINE) 9 = Unknown (MIP) Blank = Not Available (MIP)

Variable Name	FIRSTDT_PUF
Variable Label	MEDICAL CONTINUING DISABILITY REVIEW REASON TYPE, EARLIEST
Variable Description	Type code for earliest medical re-examination.
Data Type	Categorical
Data Available For	SSI/SSDI
Variable Location	Demographic Component
DAF Source Variable	FIRSTDT; DAF data extracted from the 831 & 832/833
Related PUF Variables	DD_PUFn; DR_PUFn; DT_PUFn; FIRSTDD_PUF; FIRSTDR_PUF
Value Categories in the PUF	**Approximate Distribution**
= MIEP (FIRSTDT=3)	28.3
= MINE (FIRSTDT=4)	24.4
= .M (missing, may be applicable)	45.1
= .N (missing, not applicable)	2.2
Notes	Special missing condition (.M): No data available, but applicable to the beneficiary having a populated date of initial SSDI or SSI eligibility (DOEI or SSIELIG_FIRST). Special missing condition (.N): No data available, and not applicable to the beneficiary because the initial SSDI or SSI eligibility dates (DOEI or SSIELIG_FIRST) are not populated. FIRSTDT values in DAF: 3 = Medical Reexam, Nonpermanent (MIE) And (MIP) 4 = Medical Reexam, Permanent (MINE) 5 = Medical Reexam (Permanency Of Impairment Not Known) 6 = No Diary Required 9 = Title XVI VR Child Referral Z = Invalid

Variable Name	FRA_PUF
Variable Label	BENEFICIARY'S FULL RETIREMENT AGE
Variable Description	Beneficiary's full retirement age
Data Type	Continuous
Data Available For	SSI/SSDI
Variable Location	Demographic Component
DAF Source Variable	FRA, as constructed in the DAF; source data extracted from the MBR
Related PUF Variables	FRADATE_PUF, DOBBEST_PUF
Value categories in the PUF	**Approximate distribution**
= 65	6.4
= 65.167	1.3
= 65.333	1.3
= 65.5	1.4
= 65.667	1.6
= 65.833	1.7
= 66	28.7
= 66.167	2.9
= 66.333	2.9
= 66.5	2.9
= 66.667	2.7
= 66.667	2.6
= 67	43.6
Notes	Not applicable.

Variable Name	FRADATE_PUF
Variable Label	DATE WHEN FULL RETIREMENT AGE IS REACHED
Variable Description	The date the beneficiary reaches full retirement age.
Data Type	Date
Data Available For	SSI/SSDI
Variable Location	Demographic Component
DAF Source Variable	FRADATE, as constructed in the DAF using DOBBEST and FRA
Related PUF Variables	DOBBEST_PUF, FRA_PUF
Value categories in the PUF	**Approximate distribution**
= Masked date	100.0
Notes	Masked date range limited to dates from 2/15/1996 through 12/15/2085, to reflect the age range captured by the DAF and corresponding FRA for each year. Dates outside of this range (due to data errors) set to missing).

Variable Name	MFT_PUF
Variable Label	SSI MASTER FILE TYPE
Variable Description	Identifies the person in terms of his/her program eligibility and is used to identify the type of SSI beneficiary.
Data Type	Categorical
Data Available For	SSI
Variable Location	Demographic Component
DAF Source Variable	MFT; DAF data extracted from SSI-LF
Related PUF Variables	BIC_PUF; LAF_PUFyymm; TOC_PUFn
Value categories in the PUF	**Approximate distribution**
= 1 (MFT=AI, BI, DI)	44.3
= 2 (MFT=AS, BS, DS)	1.5
= 3 (MFT=DC, BC)	11.9
= .M (values of XF, XM, XS and missing, may be applicable)	0.1
= .N (missing, not applicable)	42.1
Notes	Special missing condition (.M): No data available, but applicable to the beneficiary having a populated date of initial SSI eligibility (SSIELIG_FIRST). Special missing condition (.N): No data available, and not applicable to the beneficiary because the initial SSI eligibility date (SSIELIG_FIRST) is not populated. MFT values in the DAF: AI, BI, DI = Aged, blind or disabled individual AS, BS, DS = Aged, blind or disabled eligible spouse (first position will be blank if TOA = ES) BC, DC = Blind or disabled eligible child XF, XM, XS = Ineligible father, mother or spouse XP = Ineligible essential person who is not an ineligible parent or spouse

Variable Name	PIA_MAX_PUF
Variable Label	SSDI PRIMARY INS. AMOUNT, MAXIMUM
Variable Description	The maximum primary insurance amount (PIA) after selecting 50 available earnings values in the DAF, with indexing applied to each earnings value.
Data Type	Dollars
Data Available For	SSDI
Variable Location	Demographic Component
DAF Source Variable	PIAn; DAF data extracted from the MBR
Related PUF Variables	PIA_MIN_PUF; PIED_MAX_PUF; PIED_MIN_PUF
Value Categories in the PUF	**Approximate Distribution**
= Max(PIAn) if at least one PIAn is non-missing, then modified based on Census guidelines	65.3
= .M (missing, may be applicable)	0.0
= .N (missing, not applicable)	34.6
Notes	Use with PIED_MAX_PUF to determine date of amount. Special missing condition (.M): No data available, but applicable to the beneficiary having a populated date of initial SSDI eligibility (DOEI). Special missing condition (.N): No data available, and not applicable to the beneficiary because the initial SSDI eligibility date (DOEI) is not populated.

Variable Name	PIA_MIN_PUF
Variable Label	SSDI PRIMARY INS. AMOUNT, MINIMUM
Variable Description	The minimum primary insurance amount (PIA) after selecting 50 available earnings values in the DAF, with indexing applied to each earnings value.
Data Type	Dollars
Data Available For	SSDI
Variable Location	Demographic Component
DAF Source Variable	PIAn; DAF data extracted from the MBR
Related PUF Variables	PIA_MAX_PUF; PIED_MAX_PUF; PIED_MIN_PUF
Value Categories in the PUF	**Approximate Distribution**
= Min(PIAn) if at least one PIAn is non-missing, then modified based on Census guidelines	65.3
= .M (missing, may be applicable)	0.0
= .N (missing, not applicable)	34.6
Notes	Use with PIED_MIN_PUF to determine date of amount. Special missing condition (.M): No data available, but applicable to the beneficiary having a populated date of initial SSDI eligibility (DOEI). Special missing condition (.N): No data available, and not applicable to the beneficiary because the initial SSDI eligibility date (DOEI) is not populated.

Variable Name	PIED_MAX_PUF
Variable Label	SSDI MAX PIA EFFECTIVE DATE
Variable Description	Earliest date associated with maximum primary insurance amount.
Data Type	Date
Data Available For	SSDI
Variable Location	Demographic Component
DAF Source Variable	PIEDn; DAF data extracted from the MBR
Related PUF Variables	PIA_MAX_PUF; PIA_MIN_PUF; PIED_MIN_PUF
Value categories in the PUF	**Approximate distribution**
= Masked date	65.3
= .M(missing, may be applicable)	0.0
= .N (missing, not applicable)	34.7
Notes	Use with PIA_MAX_PUF to determine associated amount. Masked date range limited to the last 55 years for privacy reasons. Special missing condition (.M): No data available, but applicable to the beneficiary having an associated amount (PIA_MAX_PUF). Special missing condition (.N): No data available, and not applicable to the beneficiary because an associated amount (PIA_MAX_PUF) is not populated.

Variable Name	PIED_MIN_PUF
Variable Label	SSDI MIN PIA EFFECTIVE DATE
Variable Description	Most recently available date associated with minimum primary insurance amount.
Data Type	Date
Data Available For	SSDI
Variable Location	Demographic Component
DAF Source Variable	PIEDn; DAF data extracted from the MBR
Related PUF Variables	PIA_MAX_PUF; PIA_MIN_PUF; PIED_MAX_PUF
Value categories in the PUF	**Approximate distribution**
= Masked date	0.6
=.M (missing, may be applicable)	64.8
=.N (missing, not applicable)	34.7
Notes	Use with PIA_MIN_PUF to determine associated amount. We used the last date populated that corresponded to the minimum value occurrence. In most cases, the minimum value was populated past this occurrence, but without a date value. Masked date range limited to the last 55 years for privacy reasons. Special missing condition (.M): No data available, but applicable to the beneficiary having an associated amount (PIA_MIN_PUF). Special missing condition (.N): No data available, and not applicable to the beneficiary because an associated amount (PIA_MIN_PUF) is not populated.

Variable Name	REPPYSSD_PUF
Variable Label	SSDI REPRESENTATIVE PAYEE INDICATOR
Variable Description	Indicates whether SSDI beneficiary has a representative payee
Data Type	Binary
Data Available For	SSDI
Variable Location	Demographic Component
DAF Source Variable	REPPYSSD, as constructed in the DAF; source data extracted from the MBR
Related PUF Variables	TOP_PUF; REPPYSSI_PUF; REPPAYTP_PUF
Value Categories in the PUF	**Approximate Distribution**
= 0 (REPPYSSD=0)	56.6
= 1 (REPPYSSD=1)	8.8
= .N (missing, not applicable)	34.6
Notes	REPPYSSD is populated in DAF for all beneficiaries. For REPPYSSD_PUF, we overwrote cases where REPPYSSD=0 to missing if the date of initial SSDI eligibility (DOEI) was not populated. Special missing condition (.N): REPPYSSD=0, but not applicable to the beneficiary because the initial SSDI eligibility date (DOEI) is not populated.

Variable Name	REPPYSSI_PUF
Variable Label	SSI REPRESENTATIVE PAYEE INDICATOR
Variable Description	Indicates whether SSI beneficiary has a representative payee
Data Type	Binary
Data Available For	SSI
Variable Location	Demographic Component
DAF Source Variable	REPPYSSI, as constructed in the DAF; source data extracted from the SSI-LF
Related PUF Variables	REPPYSSD_PUF; REPPAYTP_PUF; TOP_PUF
Value Categories in the PUF	**Approximate Distribution**
= 0 (REPPYSSI=0)	31.9
= 1 (REPPYSSI=1)	22.8
= .N (missing, not applicable)	45.2
Notes	REPPYSSI is populated for all beneficiaries in the DAF. For REPPYSSI_PUF, we overwrote cases where REPPYSSI=0 to missing if the date of initial SSI eligibility (SSIELIG_FIRST) was not populated. Special missing condition (.N): REPPYSSI=0, but not applicable to the beneficiary because the initial SSI eligibility date (SSIELIG_FIRST) is not populated.

Variable Name	REPPAYTP_PUF
Variable Label	SSI TYPE OF PAYEE CODE
Variable Description	Relationship between the beneficiary and the representative payee
Data Type	Categorical
Data Available For	SSI
Variable Location	Demographic Component
DAF Source Variable	REPPAYTP; DAF data extracted from SSI-LF
Related PUF Variables	REPPYSSD_PUF; REPPYSSI_PUF
Value Categories in the PUF	**Approximate Distribution**
= 1 (REPPAYTP =AGY, FIN)	1.0
= 2 (REPPAYTP = SPO, FTH, MTH, SFT, SMT, GPR, CHD, REL)	18.8
= 3 (REPPAYTP=NPM, PRM, SLM, FDM, NPO, PRO, SLO, FDO)	1.6
= 4 (REPPAYTP=OTH, OFF, PYE, SEL, RPD, INP, ESP, BLANK)	1.4
= .M (missing, may be applicable)	31.9
= .N (missing, not applicable)	45.2
Notes	Special missing condition (.M): No data available, but applicable to the beneficiary having a populated date of initial SSI eligibility (SSIELIG_FIRST). Special missing condition (.N): No data available, and not applicable to the beneficiary because the initial SSI eligibility date (SSIELIG_FIRST) is not populated. REPPAYTP values in the DAF: Blank = Not applicable/unknown SEL = Beneficiary is his own payee SPO – Spouse FTH = Natural or adoptive father MTH = Natural or adoptive mother SFT = Stepfather SMT = Stepmother GPR = Grandparent CHD = Natural or adoptive child or stepchild ESP = Essential person REL = Other Relative

Variable Name	REPPAYTP_PUF
	FDO = Federal Non=mental institution FDM = Federal mental institution SLO = State/local non=mental institution SLM = State/local mental institution PRO = Proprietary non=mental institution PRM = Proprietary mental institution NPO = Non=profit non=mental institution NPM = Non=profit mental institution FIN = Financial organization AGY = Social agency OFF = Public official OTH = Other RPD = Payee being developed PYE = Payee has custody INP = Legally incompetent w/ no representative payee

Variable Name	SEX_PUF
Variable Label	SEX
Variable Description	Sex of beneficiary
Data Type	Categorical
Data Available For	SSI/SSDI
Variable Location	Demographic Component
DAF Source Variable	SEX, as constructed in the DAF using multiple administrative sources
Related PUF Variables	None
Value Categories in the PUF	**Approximate Distribution**
= F	46.0
= M	54.0
= .M (missing, may be applicable)	0.0
Notes	Special missing condition (.M): SEX is unknown (=U).

Variable Name	SSDI_ONSET_PUF
Variable Label	SSDI DISABILITY ONSET DATE EARLIEST PERIOD OF ELIGIBILITY
Variable Description	The earliest date of SSDI disability onset among all periods of SSDI disability.
Data Type	Date
Data Available For	SSDI
Variable Location	Demographic Component
DAF Source Variable	SSDI_ONSET, as constructed in the DAF; source data extracted from the MBR
Related PUF Variables	None
Value Categories in the PUF	**Approximate Distribution**
= Masked date	64.8
= .M (missing, may be applicable)	0.5
= .N (missing, not applicable)	34.6
Notes	Masked date range limited to the last 50 years for privacy reasons. Special missing condition (.M): No data available, but applicable to the beneficiary having a populated date of initial SSDI eligibility (DOEI). Special missing condition (.N): No data available, and not applicable to the beneficiary because the initial SSDI eligibility date (DOEI) is not populated.

Variable Name	SSIELIG_FIRST_PUF
Variable Label	SSI EARLIEST ELIGIBILITY DATE
Variable Description	The earliest date an individual is eligible for SSI.
Data Type	Date
Data Available For	SSI
Variable Location	Demographic Component
DAF Source Variable	SSIELIG_FIRST, as constructed in the DAF; source data extracted from the SSI-LF
Related PUF Variables	None
Value Categories in the PUF	**Approximate Distribution**
= Masked date	54.3
=.N (missing, not applicable)	45.7
Notes	Due to limitations of the administrative data, not all periods of SSI eligibility are captured in the DAF. This variable describes the first available SSI eligibility date, but there may have been earlier periods of eligibility that are no longer available. Masked date range limited to dates after 1/1/1974 reflecting the start of the SSR. Special missing condition (.N): No data available, and not applicable to the beneficiary because the initial SSI eligibility date (SSIELIG_FIRST) is not populated.

Variable Name	TKT_PUF0205
Variable Label	TICKET ACTIVE STATUS IN YEARS (2002-2005)
Variable Description	Indicates whether a beneficiary had an active Ticket (assigned or in use) in the 2002-2005 period.
Data Type	Binary
Data Available For	SSI/SSDI
Variable Location	Demographic Component
DAF Source Variable	TKTyy, as constructed in the DAF; source data extracted from the DCF
Related PUF Variables	TKT_PUFyy (yy=06-18)
Value Categories in the PUF	**Approximate Distribution**
= 0 (TKT02=0, TKT03=0, TKT04=0, and TKT05=0)	73.7
= 1 (TKT02=1 or TKT03=1 or TKT04=1 or TKT05-1)	0.6
= .M (missing, may be applicable)	23.4
= .N (missing, not applicable)	2.4
Notes	TKT0205_PUF is a single indicator for whether the ticket was assigned ever during that period. Because relatively few tickets were assigned in the early years of the program, we needed to combine across these years to minimize disclosure risk. Special missing condition (.M): No data available, but applicable to the beneficiary having a populated date of initial SSDI or SSI eligibility (DOEI or SSIELIG_FIRST). Special missing condition (.N): No data available, and not applicable to the beneficiary because the initial SSDI or SSI eligibility dates (DOEI or SSIELIG_FIRST) are not populated.

Variable Name	TKT_PUFyy (yy=06-18)
Variable Label	TICKET ACTIVE STATUS IN YEAR
Variable Description	Indicates whether a beneficiary had an active Ticket (assigned or in use) in the year, for years ranging from 2006 (06) to 2018 (18).
Data Type	Binary
Data Available For	SSI/SSDI
Variable Location	Demographic Component
DAF Source Variable	TKTyy, as constructed in the DAF; source data extracted from the DCF
Related PUF Variables	TKT0205_PUF
Value Categories in the PUF	**Approximate Distribution**
	TKT_PUF06
= 0 (TKTyy=0)	73.7
= 1 (TKTyy=1)	0.6
= .M (missing, may be applicable)	23.4
= .N (missing, not applicable)	2.4
	TKT_PUF10
= 0 (TKTyy=0)	73.3
= 1 (TKTyy=1)	0.9
= .M (missing, may be applicable)	23.4
= .N (missing, not applicable)	2.4
	TKT_PUF_15
= 0 (TKTyy=0)	73.0
= 1 (TKTyy=1)	1.3
= .M (missing, may be applicable)	23.4
= .N (missing, not applicable)	2.4
Notes	Values from 2006, 2010, and 2015 are shown for illustrative purposes; values from other years were within the range of these values. Special missing condition (.M): No data available, but applicable to the beneficiary having a populated date of initial SSDI or SSI eligibility (DOEI or SSIELIG_FIRST). Special missing condition (.N): No data available, and not applicable to the beneficiary because the initial SSDI or SSI eligibility dates (DOEI or SSIELIG_FIRST) are not populated.

Variable Name	TOC_NUM_PUF
Variable Label	SSDI NUMBER OF TYPE OF CLAIM (TOC) OCCURRENCES
Variable Description	Number of occurrences of the Type of Claim (TOC) fields for SSDI beneficiary.
Data Type	Categorical
Data Available For	SSDI
Variable Location	Demographic Component
DAF Source Variable	TOC_NUM; DAF data extracted from the MBR
Related PUF Variables	TOC_PUFn; TOC_START_PUFn
Value Categories in the PUF	**Approximate Distribution**
= 1 (TOC_NUM=1)	50.5
= 2 (TOC_NUM=2)	12.7
= 3 (TOC_NUM>=3)	2.0
= .M (missing, may be applicable)	0.2
= .N (missing, not applicable)	34.6
Notes	Special missing condition (.M): No data available, but applicable to the beneficiary having a populated date of initial SSDI eligibility (DOEI). Special missing condition (.N): No data available, and not applicable to the beneficiary because the initial SSDI eligibility date (DOEI) is not populated.

Variable Name	TOC_START_PUFn (n=1, 2)
Variable Label	SSDI TYPE OF CLAIM (TOC) START DATE (n=1,2)
Variable Description	Date the Type of Claim (TOC) is set
Data Type	Date
Data Available For	SSDI
Variable Location	Demographic Component
DAF Source Variable	TOC_STARTn; DAF data extracted from the MBR
Related PUF Variables	TOC_PUFn; TOC_NUM_PUF
Value Categories in the PUF	**Approximate Distribution**
	TOC_START_PUF1
= Masked date	18.7
= .M (missing, may be applicable)	46.5
= .N (missing, not applicable)	34.8
	TOC_START_PUF2
= Masked date	14.7
= .M (missing, may be applicable)	0.0
= .N (missing, not applicable)	85.3
Notes	TOC_STARTn (n=1,2) should be used with TOCn (n=1,2) to determine the dates that the TOC was active. TOC_STARTn was added to SSA administrative records following the rewrite of the Master Beneficiary Record in May 2009. Our assessment is that this variable remains incomplete and should be used with caution. Dates before 12/15/2007 were masked to 12/15/2007 to avoid disclosure risk; this change affected 0.1% of records in DAF18. Special missing condition (.M): No data available, but applicable to the beneficiary having a populated date of initial SSDI eligibility (DOEI). Special missing condition (.N): No data available, and not applicable to the beneficiary because the initial SSDI eligibility date (DOEI) is not populated.

Variable Name	TOC_PUFn (n=1, 2)
Variable Label	**SSDI TYPE OF CLAIM (n=1, 2)**
Variable Description	Helps to explain BIC, and in conjunction with BIC, can be used to determine the beneficiary type.
Data Type	Categorical
Data Available For	SSDI
Variable Location	Demographic Component
DAF Source Variable	TOCn; DAF data extracted from the MBR
Related PUF Variables	BIC_PUF; LAF_PUFyymm; MFT_PUF; TOC_NUM_PUF; TOC_START_PUFn
Value Categories in the PUF	**Approximate Distribution**
	TOC_PUF1
= 1 if (TOC1=1)	7.6
= 2 if (TOC1=0, 2, 4, 9)	5.2
= 3 if (TOC1=3, 8)	2.1
= 5 if (TOC1=5, 6)	49.7
= 7 if (TOC1=7, R)	0.6
= .M (missing, may be applicable)	0.2
= .N (missing, not applicable)	34.6
	TOC_PUF2
= 1 if (TOC2=1)	11.0
= 2 if (TOC2=0, 2, 3, 4, 8, 9)	1.8
= 3 if (TOC2 =5, 6, 7, R)	1.9
= .M (missing, may be applicable)	50.5
= .N (missing, not applicable)	34.8
Notes	Prior to May 2009, SSA recorded only the current TOC and did not retain historical information, meaning that a beneficiary's status would be overwritten when moving from SSDI to retirement benefits at the full retirement age. After the Master Beneficiary Record was rewritten in May 2009, TOCn now retains historical data as beneficiaries' status changes Special missing condition (.M) for TOC_PUF1: No data available, but applicable to the beneficiary having a populated date of initial SSDI eligibility (DOEI). Special missing condition (.N) for TOC_PUF1: No data available, and not applicable to the beneficiary

Variable Name	TOC_PUFn (n=1, 2)
	because the initial SSDI eligibility date (DOEI) is not populated. Special missing condition (.M) for TOC_PUF2: No data available, but applicable to the beneficiary having a populated date of initial SSDI eligibility (DOEI) and populated TOC1. Special missing condition (.N) for TOC_PUF2: No data available, and not applicable to the beneficiary because the initial SSDI eligibility date (DOEI) is not populated or TOC1 is not populated.
TOC values in the DAF: R = Disability case where beneficiary is student child 0 = Survivor case where the beneficiary is: a) Young child b) Widow(er) whose benefitit is limited by primary's amount c) Widow(er), remarried widow, or surviving divorced wife whose benefit is not reduced for age d) Mother/father entitled because of young child-in-care e) Parent of deceased worker f) Widow(er), remarried widow, or surviving divorced wife whose benefit is not actuarially reduced (or actuarial reduction is involved but the benefit amount is limited by the primary's amount) 1 = Life (Retired) case where beneficiary is: a) Retired primary not reduced for age or not increased by Delayed Retirement Credit (DRC) (MBR and LBI) b) Aged or divorced wife/husband not reduced for age c) Young wife entitled for young child-in-care d) Young child e) Prouty beneficiary (BIC J) or spouse of Prouty (BIC K) 2 = Life (Retired) or survivor case where beneficiary is: a) Retired primary reduced for age with no DRC b) Retired primary and spouse beneficiary is reduced for own age c) Widow(er) reduced for age but not limited by primary's amount d) Disabled widow(er) 3 = Life or survivor case where beneficiary is: a) Retired primary reduced for age with DRC b) Surviving Disabled child of deceased primary c) Mother/father entitled solely for Disabled Adult Child (DAC) in care	

Variable Name	TOC_PUFn (n=1, 2)
4 = Life (Retired) case where beneficiary is: a) Retired primary not reduced for age with DRC b) Young wife entitled solely for DAC in care c) DAC 5 = Disability case where beneficiary is: a) Disabled primary not reduced for age b) Aged or divorced wife/husband not reduced for age c) Young wife entitled for young child-in-cared) Young child 6 = Disability case where beneficiary is: a) Disabled primary reduced for age b) Aged or divorced wife/husband reduced for age 7 = Disability case where beneficiary is: a) Young wife entitled solely for DAC in care b) Disabled Adult Child 8 = Survivor case where beneficiary is student child 9 = Life case (retired primary) where beneficiary is student child	

Variable Name	TOP_PUF
Variable Label	SSDI TYPE OF PAYEE
Variable Description	Relationship between the beneficiary and the representative payee
Data Type	Categorical
Data Available For	SSDI
Variable Location	Demographic Component
DAF Source Variable	TOP; DAF data extracted from MBR
Related PUF Variables	REPPYSSD_PUF; REPPYSSI_PUF; REPPAYTP_PUF
Value Categories in the PUF	**Approximate Distribution**
= 1 (TOP=R, S)	0.6
= 2 (TOP=B, C, D, E, F, G, H, I)	5.6
= 3 (TOP=Q, O, M, K, P, N, L, J)	1.8
= 4 (TOP=U, T, A, V, W, Y)	3.7
= .M (missing, may be applicable)	53.7
= .N (missing, not applicable)	34.6
Notes	Special missing condition (.M): No data available, but applicable to the beneficiary having a populated date of initial SSDI eligibility (DOEI). Special missing condition (.N): No data available, and not applicable to the beneficiary because the initial SSDI eligibility date (DOEI) is not populated. TOP values in the DAF: A - Beneficiary Direct (SEL) B - Spouse (SPO) C - Natural or Adoptive Father (FTH) D - Natural or Adoptive Mother (MTH) E - Stepfather (SFT) F - Stepmother (SMT) G - Grandparent (GPR) H - Natural or Adoptive Child or Stepchild (CHD) I - Other Relative (REL) J - Federal Nonmental Institution (FDO) K - Federal Mental Institution (FDM) L - State or Local Nonmental Institution (SLO) M - State or Local Mental Institution (SLM) N - Privately-Owned Nonmental Institution (PRO) O - Privately-Owned Mental Institution (PRM)

Variable Name	TOP_PUF
	P - Nonprofit Nonmental Institution (NPO) Q - Nonprofit Mental Institution (NPM) R - Financial Organization (FIN) S - Social Agency (AGY) T - Public Official (OFF) U - Other (OTH) V - Payee Being Developed (RPD) W - Unknown for Minor Child Y - Non Classified

Variable Name	TWPCMPLMNTH_PUF1
Variable Label	FIRST SSDI TRIAL WORK PERIOD (TWP) COMPLETION MONTH
Variable Description	Trial Work Period completion date, first occurrence
Data Type	Date
Data Available For	SSDI
Variable Location	Demographic Component
DAF Source Variable	TWPCMPLMNTHn; DAF data extracted from the DCF
Related PUF Variables	None
Value Categories in the PUF	**Approximate Distribution**
= Masked date	6.2
= .M (missing, may be applicable)	59.2
= .N (missing, not applicable)	34.6
Notes	This value indicates the first date of TWP completion; subsequent dates were not populated frequently enough to include. Masked date range limited to the last 40 years for privacy reasons. Special missing condition (.M): No data available, but applicable to the beneficiary having a populated date of initial SSDI eligibility (DOEI). Special missing condition (.N): No data available, and not applicable to the beneficiary because the initial SSDI eligibility date (DOEI) is not populated.

PUF Variable Detail Pages: Annuals Component

Note: The Annuals Component is a set of files, one per calendar year, with the same file structure in each year. The frequencies shown for each variable are only meant to be illustrative of the distribution of values users might expect to see, and are rounded, so may not sum to exactly 100 percent. These were drawn from DAF18 data for January 2015 (updated from January 2014 in the DAF17 version of the PUF documentation); the distribution in other months may be slightly different. As future iterations of the DAF PUF are developed and SSA data are updated, we expect to see minor changes in these distributions.

Variable Name	BFWCM_PUFyymm
Variable Label	COMBINED (SSDI/SSI) ESTIMATED BENEFITS FORGONE FOR WORK (DRAFT) (yymm = year/month)
Variable Description	Indicates the amount of combined SSI and SSDI benefits forgone when beneficiary is in non-payment status due to suspense or termination for work in the SSDI program and/or not receiving the full SSI Federal Benefit Rate (FBR) due to earnings from work.
Data Type	Dollars
Data Available For	SSI/SSDI beneficiaries with non-missing STWCMyymm. Variable populated from March 1994 onward.
Variable Location	Annuals Component
DAF Source Variable	BFWCM_yymm, as constructed in the DAF; source data extracted from the MBR
Related PUF Variables	BFWDI_PUFyymm; BFWSSI_PUFyymm; STWCM_PUFyymm; STWDI_PUFyymm; STWSSI_PUFyymm
Value Categories in the PUF	**Approximate Distribution**
= modified BFWCMyymm (>0) based on Census guidelines	3.2
=0	84.1
=.N (missing, not applicable)	12.7
Notes	BFWCM_PUF is non-missing for all beneficiaries with a non-missing STWCM_PUF value in the month. If BFWCMyymm is not populated and STWCMyymm is non-missing, then BFWCM_PUFyymm is set to zero. Special missing condition (.N): Not applicable to the beneficiary because combined (SSI/SSDI) suspense or termination due to work (STWCMyymm) is not populated. This measure is a constructed measure using other administrative data; given its construction, we condition the missing codes on STWCMyymm instead of PSTAyymm and LAFyymm that we use for other variables.

Variable Name	BFWCM_PUFyymm
	Due to the complexities in constructing this variable, it cannot be relied upon as 100% accurate at the beneficiary level. Visit https://www.ssa.gov/disabilityresearch/daf.html and consult Volume 3 of the documentation for more information about its construction.

Variable Name	BFWDI_PUFyymm
Variable Label	SSDI ESTIMATED BENEFITS FORGONE FOR WORK (yymm = year/month)
Variable Description	Indicates the amount of SSDI benefits forgone when beneficiary is in non-payment status due to suspense or termination for work in the month indicated.
Data Type	Dollars
Data Available For	SSDI beneficiaries with non-missing STWDIyymm.
Variable Location	Annuals Component
DAF Source Variable	BFWDIyymm, as constructed in the DAF; source data extracted from the MBR
Related PUF Variables	BFWCM_PUFyymm; BFWSSI_PUFyymm; STWCM_PUFyymm; STWDI_PUFyymm; STWSSI_PUFyymm
Value Categories in the PUF	**Approximate Distribution**
= modified BFWDIyymm (>0) based on Census guidelines	1.5
=0	56.3
=.N (missing, not applicable)	42.2
Notes	BFWDI_PUF is non-missing for all beneficiaries with a non-missing STWDI_PUF value in the month. If BFWDIyymm is not populated and STWDIyymm is non-missing, then BFWDI_PUFyymm is set to zero. Special missing condition (.N): Not applicable to the beneficiary because SSDI suspense or termination due to work (STWDIyymm) is not populated. This measure is a constructed measure using other administrative data; given its construction, we condition the missing codes on STWDIyymm instead of LAFyymm that we use for other variables. Due to the complexities in constructing this variable, it cannot be relied upon as 100% accurate at the beneficiary level. Visit https://www.ssa.gov/disabilityresearch/daf.html, and consult Volume 3 of the documentation for more information about its construction.

Variable Name	BFWSSI_PUFyymm
Variable Label	SSI ESTIMATED BENEFITS FORGONE FOR WORK (DRAFT) (yymm = year/month)
Variable Description	Indicates the amount of SSI benefits forgone when beneficiary is in non-payment status due to suspense or termination for work in the SSI program or is not receiving the full SSI Federal Benefit Rate (FBR) due to earnings from work in the month indicated.
Data Type	Dollars
Data Available For	SSI beneficiaries with non-missing STWSSI. Variable populated from March 1994 onward.
Variable Location	Annuals Component
DAF Source Variable	BFWSSI_yymm, as constructed in the DAF; source data extracted from the MBR
Related PUF Variables	BFWCM_PUFyymm; BFWDI_PUFyymm; STWCM_PUFyymm; STWDI_PUFyymm; STWSSI_PUFyymm
Value Categories in the PUF	**Approximate Distribution**
= modified BFWSSI_DRAFTyymm (>0) based on Census guidelines	1.8
=0	52.7
=.N (missing, not applicable)	45.5
Notes	BFWSSI_PUF is non-missing for all beneficiaries with a non-missing STWSSI_PUF value in the month. If BFWSSIyymm is not populated and STWSSIyymm is non-missing, then BFWSSI_PUFyymm is set to zero. Special missing condition (.N): Not applicable to the beneficiary because SSI suspense or termination due to work (STWSSIyymm) is not populated. This measure is a constructed measure using other administrative data; given its construction, we condition the missing codes on STWSSIyymm instead of PSTAyymm that we use for other variables. Due to the complexities in constructing this variable, it cannot be relied upon as 100% accurate at the beneficiary level. Visit https://www.ssa.gov/disabilityresearch/daf.html, and consult Volume 3 of the documentation for more information about its construction.

Variable Name	CDR_PUFyymm
Variable Label	SSDI CESSATION OF DISABILITY REASON (yymm = year/month)
Variable Description	Indicates the cessation reason that applies in the month indicated.
Data Type	Categorical
Data Available For	SSDI
Variable Location	Annuals Component
DAF Source Variable	CDRyymm; DAF data extracted from the MBR
Related PUF Variables	None
Value Categories in the PUF	**Approximate Distribution**
= 1 (CDRyymm = E, S)	2.4
= 2 (CDRyymm = M)	1.2
= 3 (CDRyymm = F, N, V, W)	0.1
= .M (missing, may be applicable)	54.4
= .N (missing, not applicable)	41.9
Notes	Special missing condition (.M): No data available, but applicable to the beneficiary having a populated value for SSDI ledger account file participation status (LAFyymm). Special missing condition (.N): No data available, and not applicable to the beneficiary because SSDI ledger account file participation status (LAFyymm) is not populated. The range of values for CDRyymm is: E = TWP SGA Cessation (EPE) F = Failure to cooperate M = Medical Cessation N = DA&A Termination (12 consecutive months of noncompliance and termination effective 1/97 for P.L. 104-121) S = Substantial gainful activity (SGA) cessation - No Trial Work Period (TWP) V = Vocational Rehabilitation W = Whereabouts Unknown

Variable Name	DIERNLVL_PUFyymm
Variable Label	SSDI EARNING LEVEL INDICATOR (yymm = year/month)
Variable Description	A categorical variable that indicates the level of earnings, in the month indicated, for SSDI beneficiaries based on earnings, self-employment earnings, and income exclusions. The variable is then related to the Trial Work Period (TWP) and Substantial Gainful Activity (SGA) levels.
Data Type	Categorical
Data Available For	SSDI
Variable Location	Annuals Component
DAF Source Variable	DIERNLVLyymm, as constructed in the DAF and extracted from the DCF
Related PUF Variables	T2GRSAMT_PUFyymm; T2NETAMT_PUFyymm
Value Categories in the PUF	**Approximate Distribution**
= 0 if gross earnings are zero	0.1
= L if gross earnings are greater than zero, but less than the TWP limit	1.6
= T if gross earnings are at or above the TWP limit, but less than SGA	0.7
= G if gross and countable earnings are above SGA	0.8
= .M (missing, may be applicable)	55.1
= .N (missing, not applicable)	41.7
Notes	Special missing condition (.M): No data available, but applicable to the beneficiary having a populated value for SSDI ledger account file participation status (LAFyymm). Special missing condition (.N): No data available, and not applicable to the beneficiary because SSDI ledger account file participation status (LAFyymm) is not populated.

Variable Name	DIRPAY_PUFyymm
Variable Label	SSDI ACTUAL BENEFIT PAID AMOUNT (yymm = year/month)
Variable Description	Actual amount of SSDI benefits disbursed in the month indicated.
Data Type	Dollars
Data Available For	SSDI beneficiaries in current pay status (LAFyymm=C)
Variable Location	Annuals Component
DAF Source Variable	DIRPAYyymm; DAF data extracted from PHUS
Related PUF Variables	MEDPREM_PUFyymm; PAYD_PUFyymm
Value Categories in the PUF	**Approximate Distribution**
= modified DIRPAYyymm based on Census guidelines	50.8
= 0 if DIRPAYyymm is missing and LAFyymm is in C	12.1
= .M (missing, may be applicable)	0.0
= .N (missing, not applicable)	37.0
Notes	Special missing condition (.M): No data available, but applicable to the beneficiary having a populated value for SSDI ledger account file participation status (LAFyymm) that is not in C. Special missing condition (.N): No data available, and not applicable to the beneficiary because SSDI ledger account file participation status (LAFyymm) is not populated.

Variable Name	DUED_PUFyymm
Variable Label	SSDI BENEFIT DUE (yymm = year/month)
Variable Description	Indicates the benefit amount due to beneficiary, after rounding, but before obligations such as Supplemental Medical Insurance (SMI) premium, in the month indicated.
Data Type	Dollars
Data Available For	SSDI beneficiaries in current pay status (LAFyymm=C)
Variable Location	Annuals Component
DAF Source Variable	DUEDyymm; DAF data extracted from the MBR
Related PUF Variables	DUES_PUFyymm
Value Categories in the PUF	**Approximate Distribution**
= modified DUEDyymm based on Census guidelines	58.0
= 0 if DUEDyymm is missing and LAFyymm is C	5.0
= .M (missing, may be applicable)	0.0
= .N (missing, not applicable)	37.0
Notes	Special missing condition (.M): No data available, but applicable to the beneficiary having a populated value for SSDI ledger account file participation status (LAFyymm) that is not in C. Special missing condition (.N): No data available, and not applicable to the beneficiary because SSDI ledger account file participation status (LAFyymm) is not populated. This variable should be used only in conjunction with appropriate LAF codes; visit https://www.ssa.gov/disabilityresearch/daf.html, and consult Volume 5 of the documentation for more details about usage of this variable.

Variable Name	DUES_PUFyymm
Variable Label	SSI BENEFIT DUE (yymm = year/month)
Variable Description	Indicates the total (federal and state) SSI benefit (expected amount of check to be paid) in the month indicated.
Data Type	Dollars
Data Available For	SSI beneficiaries in current pay status (PSTAyymm=C01)
Variable Location	Annuals Component
DAF Source Variable	DUESyymm, DAF data extracted from the SSI-LF
Related PUF Variables	DUED_PUFyymm; PAYS_PUFyymm
Value Categories in the PUF	**Approximate Distribution**
= modified DUESyymm based on Census guidelines if PSTAyymm=C01	33.6
= 0 if PSTAyymm=C01, but DUESyymm is missing	31.8
= .M (missing, may be applicable)	0.0
= .N (missing, not applicable)	34.6
Notes	Special missing condition (.M): No data available, but applicable to the beneficiary having a populated value for SSI payment status (PSTAyymm) that is not C01. Special missing condition (.N): No data available, and not applicable to the beneficiary because SSI payment status (PSTAyymm) is not populated. Visit https://www.ssa.gov/disabilityresearch/daf.html, and consult Volume 5 of the documentation for more details about usage of this variable.

Variable Name	EDX_PUFyy
Variable Label	EDUCATION LEVEL in the last month observed during the calendar year (yy=year)
Variable Description	This variable gives the beneficiary's years of education as of the last month recorded during the calendar year.
Data Type	Categorical
Data Available For	SSI/SSDI
Variable Location	Annuals Component
DAF Source Variable	EDXyymm; DAF data extracted from 831 & 832/833
Related PUF Variables	None
Value Categories in the PUF	**Approximate Distribution**
= 1 (if EDXyymm is 0 to 11)	19.3
= 2 (if EDXyymm is12)	26.4
= 3 (if EDXyymm is 13-15)	9.4
= 4 (if EDXyymm is 16-26)	4.3
= .M (missing, may be applicable)	40.6
Notes	The last populated EDXyymm variable for a beneficiary during the year is used to classify education. EDX_PUFyy value range: 1 = less than high school 2 = high school 3 = some college 4 = college or more Special missing condition (.M): No data available (or having an invalid value such as ZZ or 99), but applicable to the beneficiary. This field is only populated for beneficiaries whose application received an initial determination at step 5 of SSA's sequential evaluation process, at which point education is used to make a decision.

Variable Name	EICM_PUFyymm
Variable Label	SSI COUNTABLE EARNED INCOME (yymm = year/month)
Variable Description	The amount of earned income used to for the SSI payment due calculation in the month indicated. It is generally half of actual earned income, minus deductions and exclusions.
Data Type	Dollars
Data Available For	SSI
Variable Location	Annuals Component
DAF Source Variable	EICMyymm; DAF data extracted from the SSI-LF
Related PUF Variables	UINC_PUFyymm
Value Categories in the PUF	**Approximate Distribution**
= Modified EICMyymm (>0) based on Census guidelines	1.8
= 0	63.7
=.M (missing, may be applicable)	0.0
=.N (missing, not applicable)	34.6
Notes	Special missing condition (.M): No data available, but applicable to the beneficiary having a populated value for SSI payment status (PSTAyymm). Special missing condition (.N): No data available, and not applicable to the beneficiary because SSI payment status (PSTAyymm) is not populated.

Variable Name	LAF_PUFyymm
Variable Label	SSDI LEDGER ACCOUNT FILE (LAF)-PARTICIPATION STATUS (yymm = year/month)
Variable Description	SSDI Ledger Account File (LAF) reflects the MBR payment status in the month indicated.
Data Type	Categorical
Data Available For	SSDI
Variable Location	Annuals Component
DAF Source Variable	LAFyymm; DAF data extracted from the MBR
Related PUF Variables	BIC_PUF; DUED_PUFyymm; MFT_PUF; TOC_PUFn
Value Categories in the PUF	**Approximate Distribution**
= C (LAFyymm =C*)	53.6
= S (LAFyymm = S*)	1.5
= T (LAFyymm = T*)	2.8
= O (LAFyymm in other non-missing value)	0.1
= .N (missing, see below)	41.9

Notes

Special missing condition (.N): No data available. The beneficiary did not have a populated LAFyymm code during the month.

The first character in LAFyymm is a status code, and the second character is a "reason" code that expands on the status. Often, only the first character is used for selection purposes. The range of values for LAFyymm is:

A = withdrawal for adjustment
AD = Adjusted for dual entitlement
AS = Adjusted for simultaneous entitlement
A9 = All other adjust actions
B = abatement status
C = Current payment status (except railroad payment)
D = Deferred payment status
DP = Deferred because of public assistance
DW = Deferred because of Workers Compensation (WC) offset
D1 = Deferred for foreign work test
D2 = Deferred for annual retirement test
D3 = Deferred as an auxiliary of a beneficiary whose status is deferred for annual retirement test (LAF = D2)
D4 = Deferred for no child-in-care
D5 = Deferred as an auxiliary of a beneficiary whose status is deferred for foreign work test (LAF = D1)
D6 = Deferred to cover overpayments not covered by any of the above reasons

D9 = Miscellaneous deferment not provided with a specific code
E = Current payment through Railroad Board (RRB)
F = Advanced filling for current payment through RRB
J = Advanced filing for current payment (except railroad payment)
K = Advanced filing for deferred payment (low order position same as LAF D)
L = Advanced filing for conditional payment (low order position same as LAF S)
N = Disallowed claim (technical disallowance) e.g., not enough quarters of coverage (QC)
ND = Denied claim (medical denial) e.g., not proven disabled
P = Pending claim (adjudication pending)
PB = Delayed claim (another beneficiary in this claim has a LAF other than B, N, P, T, U, or X). This may apply to a child who does not live with the wage earner and for whom information is not yet available.
PT = Claim has been terminated from delayed status (LAF PB)
S = Conditional payment status
SD = Technical entitlement
SF = Prouty beneficiary fails to meet residency requirement
SH = Government pension completely offsets benefits due to a spouse, widow, mother, father or special age 72SJ - Alien suspension
SK = Deportation
SL = Barred payment country
SM = Refused old-age insurance benefits to get Medicare-only coverage (prior to 1/81)
SP = Prouty beneficiary receiving public assistance
SS = Non-payment to post-secondary students during summer months
SW = WC Offset
S0 = ending determination of continuing disability
S1 = Worked outside the United States (U.S.)
S2 = Worked inside the U.S.
S3 = Insured person worked in the U.S.
S4 = Failure to have child-in-care
S5 = Insured person worked outside
S6 = For better address
S7 = Prisoner suspension; suspension because of extended trial work period (EPE SGA); or suspension for refusing Vocational Rehabilitation (VR) Services. (See Reason for Suspension or Termination (HRFST) for why suspension occurred)
S8 = Payee not determined
S9 = All other suspension reason not specifically defined (miscellaneous suspension) (see RFST for why suspension occurred)
T = Terminated (not used)
TA = Advanced filing claim terminated before maturity
TB = Mother, father terminated-entitled to disabled widow(er)'s benefits
TC = Disabled widow or widower attained age FRA
TJ = Advance file claim terminated after maturity
TL = Termination of student (post-secondary) benefits because of legislative changes in student requirements
TP = Terminated for change of payment identification code (PIC) on post-entitlement actions
T& = Claimed withdrawn

TX = DIB attained FRA (also used for auxiliary beneficiaries)
T = DIB attained 65, converted to RIB (same as TX)
T0 = Benefits payable by some other agency
T1 = Death of beneficiary
2 = Dependent status terminated because of death of primary beneficiary. This is a very ephemeral code. Most people who lose benefits because of the death of a primary beneficiary will soon convert to a survivors beneficiary.
T3 = Divorce, marriage, remarriage. (See RFST for why suspension occurred.)
T4 = Attainment of age 18 or 22 and not disabled; mother/father terminated based on last child's attainment of age 16. Beneficiary Identification Code (BIC) = B2 mother < 62 with child <16
T5 = Entitled to other benefits equal or larger
T6 = Termination of mother/father benefits because of death or marriage of last child; termination of benefits because child no longer disabled or attending school. Examination of BIC will determine the meaning of this value.
T7 = Adoption of child; mother terminated, last entitled child adopted. Valid only if Date of Suspension or Termination (DOST) is earlier than 10/72. Probably not happening any more.
T8 = DIB no longer disabled; mother/father terminated (BIC = B2, E), child no longer disabled
T9 = All other termination reasons
U = Active title XVIII status only
W = Withdrawal before entitlement
X = Title XVIII adjustment or termination status
XD = Withdrawal for adjustment
X8 = Payee not determined
XK = Deportation
XR = Withdrawn from Supplemental Medical Insurance Benefits (SMIB)
X0 = Claim transferred to RRB
X1 = Death of beneficiary
X5 = Entitled to other benefits
X7 = Hospital Insurance Benefits (HIB)/SMIB terminated
X9 = All other reasons

Variable Name	LIVF_PUFyymm
Variable Label	SSI LIVING ARRANGEMENT CODE (yymm = year/month)
Variable Description	Identifies the living arrangement for a beneficiary in the month indicated and is used to adjust payments.
Data Type	Categorical
Data Available For	SSI
Variable Location	Annuals Component
DAF Source Variable	LIFVyymm; DAF data extracted from the SSI-LF
Related PUF Variables	None
Value Categories in the PUF	**Approximate Distribution**
= A if (LIVFyymm=A)	30.4
= B if (LIVFyymm=B)	2.4
= C if (LIVFyymm=C)	6.5
= D if (LIVFyymm=D)	0.7
= .M (missing, may be applicable)	1.5
= .N (missing, not applicable)	58.4
Notes	Special missing condition (.M): No data available, but applicable to the beneficiary having a populated value for SSI payment status (PSTAyymm). Special missing condition (.N): No data available, and not applicable to the beneficiary because SSI payment status (PSTAyymm) is not populated. The range of values for LIVFyymm is: A = Living in own household or alone or "intervening" B = Living in the household of another receiving support and maintenance (not C) C = Living in parent's own household and child less than 18 D = Living in medical facility and Medicaid pays more than 50 percent of expenses. * = Error in transmission of data.

Variable Name	MEDE_PUFyymm
Variable Label	SSI MEDICAID ELIGIBILITY (yymm = year/month)
Variable Description	Indicates whether an SSI beneficiary is eligible for Medicaid in the month indicated (as of the month indicated).
Data Type	Binary
Data Available For	SSI. Variable populated from March 1996 onward.
Variable Location	Annuals Component
DAF Source Variable	MEDEyymm, as constructed in the DAF; DAF data extracted from the CER100%
Related PUF Variables	None
Value Categories in the PUF	**Approximate Distribution**
= 0 if (MEDEyymm =0)	11.5
= 1 if (MEDEyymm =1)	25.8
= .M (missing, may be applicable)	6.1
= .N (missing, not applicable)	56.7
Notes	Special missing condition (.M): No data available, but applicable to the beneficiary having a populated value for SSI payment status (PSTAyymm). Special missing condition (.N): No data available, and not applicable to the beneficiary because SSI payment status (PSTAyymm) is not populated. Variables from the CER100% Field Files are based on monthly snapshots and not retroactively updated. Data in these variables may therefore not reflect the most recent update in the administrative data.

Variable Name	MEDEX_PUFyymm
Variable Label	MEDICAL IMPROVEMENT CODE (yymm = year/month)
Variable Description	Medical improvement indicator, in the month indicated.
Data Type	Categorical
Data Available For	SSI/SSDI
Variable Location	Annuals Component
DAF Source Variable	MEDEXyymm, as constructed in the DAF; DAF data extracted from the 831 & 832/833
Related PUF Variables	None
Value Categories in the PUF	**Approximate Distribution**
= E	3.8
= N	29.9
= P	44.7
= 0	20.8
= .M (missing, may be applicable)	0.3
= .N (missing, not applicable)	0.5
Notes	Special missing condition (.M): No data available, but applicable to the beneficiary having a populated value for SSDI ledger account file participation status (LAFyymm) or SSI payment status (PSTAyymm). Special missing condition (.N): No data available, and not applicable to the beneficiary because SSDI ledger account file participation status (LAFyymm) and SSI payment status (PSTAyymm) are not populated. The range of values for MEDEXyymm is: E = medical improvement expected N = medical improvement not expected P = medical improvement possible 0 = medical improvement unknown

Variable Name	MEDPREM_PUFyymm
Variable Label	SSDI PHUS MEDICARE PREMIUM (yymm = year/month)
Variable Description	Medicare Premium amount paid on behalf of the beneficiary in the month indicated.
Data Type	Dollars
Data Available For	SSDI
Variable Location	Annuals Component
DAF Source Variable	MEDPREMyymm; DAF data extracted from the PHUS
Related PUF Variables	DIRPAY_PUFyymm; PAYD_PUFyymm
Value Categories in the PUF	**Approximate Distribution**
= modified MEDPREMyymm (>0) based on Census guidelines	23.1
= 0 if MEDPREMyymm is missing and LAFyymm=C	39.7
= .U if MEDPREMyymm is missing, but LAFyymm=C	0.0
= .M (missing, may be applicable)	0.2
= .N (missing, not applicable)	37.0
Notes	Special missing condition (.M): No data available, but applicable to the beneficiary having a populated value for SSDI ledger account file participation status (LAFyymm) than is not C. Special missing condition (.N): No data available, and not applicable to the beneficiary because SSDI ledger account file participation status (LAFyymm) is not populated.

Variable Name	MEDR_PUFyymm
Variable Label	SSDI MEDICARE ELIGIBILITY (yymm = year/month)
Variable Description	Medicare eligibility for SSDI in the month indicated.
Data Type	Binary
Data Available For	SSDI
Variable Location	Annuals Component
DAF Source Variable	MEDRyymm, as constructed in the DAF; DAF data extracted from the MBR
Related PUF Variables	None
Value Categories in the PUF	**Approximate Distribution**
= 0 if (MEDRyymm =0)	11.7
= 1 if (MEDRyymm =1)	48.4
= .M (missing, may be applicable)	2.1
= .N (missing, not applicable)	37.8
Notes	Special missing condition (.M): No data available, but applicable to the beneficiary having a populated value for SSDI ledger account file participation status (LAFyymm). Special missing condition (.N): No data available, and not applicable to the beneficiary because SSDI ledger account file participation status (LAFyymm) is not populated.

Variable Name	PAYD_PUFyymm
Variable Label	SSDI PHUS BENEFIT PAID (yymm = year/month)
Variable Description	The sum of the direct pay (DIRPAYyymm) and the Medicare payment (MEDPREMyymm) fields for the SSDI beneficiary in the month indicated.
Data Type	Dollars
Data Available For	SSDI
Variable Location	Annuals Component
DAF Source Variable	PAYDyymm; DAF data extracted from the PHUS
Related PUF Variables	PAYO_PUFyymm; PAYS_PUFyymm
Value Categories in the PUF	**Approximate Distribution**
= modified PAYDyymm (>0) based on Census guidelines if PAYDyymm is populated	50.8
= 0 if PAYDyymm is missing or negative	12.1
= .M (missing, may be applicable)	0.0
= .N (missing, not applicable)	37.0
Notes	Special missing condition (.M): No data available, but applicable to the beneficiary having a populated value for SSDI ledger account file participation status (LAFyymm) that is not C. Special missing condition (.N): No data available, and not applicable to the beneficiary because SSDI ledger account file participation status (LAFyymm) is not populated. Visit https://www.ssa.gov/disabilityresearch/daf.html, and consult Volume 5 of the documentation for more details about usage of this variable.

Variable Name	PAYO_PUFyymm
Variable Label	SSDI PHUS DEPENDENT BENEFIT PAID (yymm = year/month)
Variable Description	The payment made for the SSDI dependents to a primary beneficiary in the month indicated.
Data Type	Dollars
Data Available For	SSDI
Variable Location	Annuals Component
DAF Source Variable	PAYOyymm, DAF data extracted from the PHUS
Related PUF Variables	PAYD_PUFyymm; PAYS_PUFyymm
Value Categories in the PUF	**Approximate Distribution**
= modified PAYOyymm (>0) amount based on Census guidelines	6.5
= 0 if LAFyymm is C, but PAYOyymm is missing	54.8
= .M (missing, may be applicable)	1.1
= .N (missing, not applicable)	37.7
Notes	Special missing condition (.M): No data available, but applicable to the beneficiary having a populated value for SSDI ledger account file participation status (LAFyymm) that is not C. Special missing condition (.N): No data available, and not applicable to the beneficiary because SSDI ledger account file participation status (LAFyymm) is not populated.

Variable Name	PAYS_PUFyymm
Variable Label	SSI BENEFIT PAID (yymm = year/month)
Variable Description	Total SSI benefit paid in the month indicated.
Data Type	Dollars
Data Available For	SSI
Variable Location	Annuals Component
DAF Source Variable	PAYSyymm, DAF data extracted from the SSI-LF
Related PUF Variables	DUES_PUFyymm
Value categories in the PUF	**Approximate distribution**
= modified PAYSyymm based on Census guidelines if PSTAyymm=C01	30.9
= 0 if PSTAyymm is C01, but PAYSyymm is missing or negative	2.6
= .M (missing, may be applicable)	8.1
= .N (missing, not applicable)	58.4
Notes	Special missing condition (.M): Applicable to the beneficiary having a populated value for SSI payment status (PSTAyymm) that is not C01. Special missing condition (.N): Not applicable to the beneficiary because SSI payment status (PSTAyymm) is not populated. Visit https://www.ssa.gov/disabilityresearch/daf.html, and consult Volume 5 of the documentation for more details about usage of this variable.

Variable Name	PROA_PUFyy
Variable Label	SSI USE OF 1619A PROVISION INDICATOR (yymm = year/month)
Variable Description	Indicates the use of provision 1619a in at least one month of the calendar year; provision 1619a indicates the continuance of SSI benefit receipt though working at SGA level
Data Type	Binary
Data Available For	SSI
Variable Location	Annuals Component
DAF Source Variable	PROAyymm, constructed in the DAF; source data extracted from the SSI-LF
Related PUF Variables	PROB_PUFyy
Value Categories in the PUF	**Approximate Distribution**
= 1 if any month during the calendar year of PROAyymm=1	0.4
= 0 if PROA_PUFyy is not 1 in any month and in at least one month: PROAyymm is 0 or PROAyymm is missing, but STWSSIyymm is in (1,2,3,4)	44.9
=.M (missing, may be applicable)	19.7
=.N (missing, not applicable)	35.0
Notes	Special missing condition (.M): No data available, but applicable to the beneficiary having a populated value for SSI payment status (PSTAyymm) or SSI suspense or termination due to work (STWSSIyymm) value is 8 or 9 in at least one month during the year. Special missing condition (.N): No data available, and not applicable to the beneficiary because SSI payment status (PSTAyymm) is not populated in any month during the year.

Variable Name	PROB_PUFyy
Variable Label	SSI USE OF 1619B PROVISION INDICATOR (yymm = year/month)
Variable Description	Indicates whether the 1619b continuing Medicaid eligibility provision was used in the month indicated. Indicates the use of provision 1619b in at least one month of the calendar year; provision 1619b indicates the continuance of Medicaid receipt though working at SGA level
Data Type	Binary
Data Available For	SSI
Variable Location	Annuals Component
DAF Source Variable	PROByymm, constructed in the DAF; source data extracted from the SSI-LF
Related PUF Variables	PROA_PUFyy
Value Categories in the PUF	**Approximate Distribution**
= 1 if any month during the calendar year of PROByymm=1	1.1
= 0 if PROB_PUFyy is not 1 in any month and in at least one month: PROByymm is 0 or PROByymm is missing, but STWSSIyymm is in (1,2,3,4)	44.3
= .M (missing, may be applicable)	19.7
= .N (missing, not applicable)	35.0
Notes	Special missing condition (.M): No data available, but applicable to the beneficiary having a populated value for SSI payment status (PSTAyymm) or SSI suspense or termination due to work (STWSSIyymm) value is 8 or 9 in at least one month during the year. Special missing condition (.N): No data available, and not applicable to the beneficiary because SSI payment status (PSTAyymm) is not populated in any months during the year.

Variable Name	PSTA_PUFyymm
Variable Label	SSI PAYMENT STATUS (PSTAT) (yymm = year/month)
Variable Description	Indicates payment, suspension, termination, etc. in the month indicated. It can be compared to the date of enrollment to determine prior suspensions and terminations.
Data Type	Categorical
Data Available For	SSI
Variable Location	Annuals Component
DAF Source Variable	PSTAyymm; DAF data extracted from the SSI-LF
Related PUF Variables	None
Value Categories in the PUF	**Approximate Distribution**
= 1 (PSTAyymm=C01, M01, M02)	33.6
= 2 (PSTAyymm in N31-N33, N42, N43)	1.3
= 3 (PSTAyymm = E* or N* not specified in group 2)	5.8
= 4 (PSTAyymm = T*)	0.6
= 5 (PSTAyymm = other non-missing values)	0.2
= .N (missing, see below)	58.4

Notes

Special missing condition (.N): No data available. The beneficiary did not have a populated PSTAyymm code during the month.

The range of values for PSTAyymm is:

C01 = Current pay (eligible for a payment).
E01 = Eligible for federal and/or state benefits based on eligibility computation, but no payment is due based on the payment computation.
E02 = Non-pay = First month of eligibility; eligible but not payable (effective 08/22/96). Under welfare reform, the second month of eligibility is the earliest month payment may begin.
H10 = Hold = living arrangements change in process
H20 = Hold = marital status change in process
H30 = Hold = resource change in process
H40 = Hold = student status change in process
H50 = Hold = head of household change in process
H60 = Hold = pending receipt of date of death
H70 = Hold = pending posting of payment made outside the system
H80 = Hold = early input case– waiting on a disability determination
H90 = Hold = systems limitation regarding computation, field office must manually compute

and input payment amounts.
M01 = Force payment case = Recipient may be in current pay or non-pay status depending on payment history. If in current pay, it is a manual payment.
M02 = Force due case = Field office controls case through MSSICS. Recipient may be in current pay or non-pay status, depending on the payment history. This is a temporary situation. The benefit amount is manually computed, but offsets from computation history may be computed and sent to the field office for possible adjustments. (effective 08/1999)
N01 = Non-pay = Claimant's chargeable income exceeds both the applicable SSI payment and his state's payment standard.
N02 = Non-pay – Claimant is an ineligible resident of a public institution
N03 = Non-pay – Claimant is absent from the U.S.
N04 = Non-pay = Claimant's non-excludable resources exceed the SSI limitations.
N05 = Non-pay – Unable to determine if eligibility for some period of nonpayment or failure to provide information for children overseas. (Obsolete)
N06 = Non-pay – Claimant refuses to file for other benefits.
N07 = Non-pay – Cessation of recipient's disability.
N08 = Non-pay – Cessation of recipient's blindness.
N09 = Non-pay – Recipient refused vocational rehabilitation without good cause. (Obsolete)
N10 = Non-pay – Failure to comply with drug or alcohol treatment plan (Prior to 03/95 – Refused treatment for drug addiction). (Obsolete)
N11 = Non-pay – Benefit sanction month because of failure to comply with approved treatment plan (Prior to 03/95 – Refused treatment for alcoholism). (Obsolete)
N12 = Non-pay – Claimant voluntarily withdrew application.
N13 = Non-pay – Not a U.S. citizen, U.S. national, or eligible alien.
N14 = Non-pay = Aged claimant under 65 and not blind or disabled.
N15 = Non-pay = Blind claim denied; applicant not blind. No visual impairment.
N16 = Non-pay = Disability claim denied; applicant not disabled.
N17 = Non-pay = Failure by applicant to pursue claim.
N18 = Non-pay = Failure by applicant to cooperate on development of claim, or unable to locate.
N19 = Non-pay = Recipient voluntarily terminates participation in SSI program. (Obsolete)
N20 = Non-pay = Recipient fails to provide information.
N22 = Non-pay = Claimant is an inmate of a penal institution (effective 12/97). This denial code is a breakout of the current code N02.
N23 = Non-pay = Claimant is not a resident of the United States. Effective 12/97 N13 will no longer be used for non-residents.
N24 = Claimant provided false or misleading statements affecting benefit eligibility or amount and administrative sanction is imposed. N24 stop payment cases are eligible for Medicaid (MEDIC = N). The N24 definition was changed by the Foster Care Independence Act of 1999. A Systems cleanup run removed N24 from any months prior to 12/1999.
N25 = Non-pay – Claimant is fleeing to avoid prosecution for, or custody or confinement for a crime which is a felony (or in New Jersey, a high misdemeanor) under the laws of the place from which he/she flees, or is violating a condition of probation or parole imposed under Federal or State law (effective 12/97). N05 will no longer be used for this welfare reform provision.
N27 = Non-pay = Termination due to Substantial Gainful Activity (SGA). (Obsolete)

N30 = Non-pay = Slight impairment, medical consideration alone; no visual impairment.
N31 = Non-pay = Capacity for SGA – customary past work; no visual impairment.
N32 = Non-pay = Capacity for SGA = other than relevant past work; no visual impairment.
N33 = Non-pay = Engaging in SGA despite impairment; no visual impairment.
N34 = Non-pay = Impairment prevented SGA for period of 12 months; no visual impairment.
N35 = Non-pay = Impairment is severe at time of adjudication but not expected to last 12 months; no visual impairment.
N36 = Non-pay = Insufficient or no medical data furnished.
N37 = Non-pay = Failure or refusal to submit to consultative examination.
N38 = Non-pay = Applicant does not want to continue development of claim.
N39 = Non-pay = Applicant willfully fails to follow prescribed treatment.
N40 = Non-pay – Individualized functional assessment shows impairment(s) not of comparable severity; no visual impairment.
N41 = Non-pay – Impairment not severe – medical consideration alone; visual impairment.
N42 = Non-pay = Capacity for SGA = customary past work; visual impairment.
N43 = Non-pay = Capacity for SGA = other work; visual impairment.
N44 = Non-pay – Impairment not severe.
N45 = Non-pay = Impairment prevented SGA for a period of less than 12 months; visual impairment.
N46 = Non-pay = Impairment prevented SGA at time of adjudication but is not expected to prevent SGA for a period of 12 months; visual impairment.
N47 = Non-pay = Insufficient or no medical evidence furnished; visual impairment. (Obsolete)
N48 = Non-pay = Failure or refusal to submit to consultative examination; visual impairment. (Obsolete)
N49 = Non-pay = Applicant does not want to continue development of claim; visual impairment. (Obsolete)
N50 = Non-pay = Applicant willfully fails to follow prescribed treatment; visual impairment. (Obsolete).
N51 = Non-pay = Individualized functional assessment shows impairment(s) not of comparable severity; visual impairment.
N52 = Non-pay = Deleted from state rolls prior to 12/73 payment.
N53 = Non-pay = Deleted from state rolls after 12/73 payment.
N54 = Non-pay = District Office unable to locate applicant. (Obsolete)
P01 = Possible reinstatement pending development of SGA. (Obsolete)
S01 = Suspended = Report of death by Treasury. Potential automated death case. (Obsolete)
S04 = Suspended = System is awaiting disability determination input (system generated).
S05 = Suspended = System unable to determine prerequisite month for 1619(a) eligibility.
S06 = Suspended = Recipient's address unknown.
S07 = Suspended = Returned checks for other than death, identification, address, death of payee, or duplicate check (system generated).
S08 = Suspended = Representative payee development pending.
S09 = Suspended – Temporary institutionalization suspense (system generated).
S10 = Suspended – Recipient has a bank account and refuses to receive payments via direct deposit.

S20 = Suspended – Recipient is presumptively disabled or blind and has received 6 months of payments (FO input).
S21 = Suspended = Recipient is presumptively disabled or blind and has received 6 months of payments (system generated).
S90 = Suspended – PR1 change in process.
S91 = Suspended – PR1 change in process.
T01 = Terminated = Death of recipient.
T20 = Terminated = Received a duplicate payment based on two different numbers (FO or CO input).
T22 = Terminated = Received a duplicate payment based on the same number on different SSRs or on two different numbers (system generated).
T30 = Terminated = Received payments, but record must be reestablished to correct SSR (such as when the individual gets married). Manual termination by FO.
T31 = Terminated – Payment previously made or refund on record. System generated termination.
T32 = Terminated – Automated system termination of a paid record that has exceeded certain size limitations.
T33 = Terminated = Received payments, but record must be reestablished to correct SSR. Manual termination by FO through MSSICS.
T50 = Terminated = No payment received, but record must be reestablished to correct SSR (false start). Manual termination by FO.
T51 = Terminated – No payment made on record. System generated termination.

Variable Name	REGION_PUFyy
Variable Label	SSA REGION NUMBER (yy=YEAR)
Variable Description	Constructed measure derived from using the last monthly occurrence of state of residence reported during the year
Data Type	Categorical
Data Available For	SSI/SSDI. Variable populated from 1996 onward.
Variable Location	Annuals Component
DAF Source Variable	PSTyymm, as constructed in the DAF; source data extracted from the CER100% and DBAD/ZIP
Related PUF Variables	None
Value Categories in the PUF	**Approximate Distribution**
= 1	4.0
= 2	8.2
= 3	8.4
= 4	19.9
= 5	13.8
= 6	10.8
= 7	3.7
= 8	2.0
= 9	9.9
= 10	3.2
= .M	16.2

Notes

PSTyymm is derived from the beneficiary's zip code in SSI data sources, if available, and from SSDI data sources otherwise.

Special missing condition (.M): No data available.

REGION_PUFyy values:

1 = Connecticut, Maine, Massachusetts, New Hampshire, Rhode Island, Vermont
2 = New Jersey, New York, Puerto Rico, Virgin Islands
3 = Delaware, District of Columbia, Maryland, Pennsylvania, Virginia, West Virginia
4= Alabama, Florida, Georgia, Kentucky, Mississippi, North Carolina, South Carolina, Tennessee
5 = Illinois, Indiana, Michigan, Minnesota, Ohio and Wisconsin
6 = Arkansas, Louisiana, New Mexico, Oklahoma, Texas
7 = Iowa, Kansas, Missouri, Nebraska
8 = Colorado, Montana, North Dakota, South Dakota, Utah, Wyoming
9 = American Samoa, Arizona, California, Guam, Hawaii, Nevada, Saipan
10 = Alaska, Idaho, Oregon, Washington

Variable Name	SSICOUPL_PUFyymm
Variable Label	SSI COUPLE INDICATOR (yymm = year/month)
Variable Description	Indicates whether SSI beneficiaries are part of a couple in the month indicated.
Data Type	Binary
Data Available For	SSI beneficiaries with non-missing PSTAyymm
Variable Location	Annuals Component
DAF Source Variable	CURCOMPyymm, as extracted from the SSI-LF
Related PUF Variables	BFWSSI_PUFyymm; STWSSI_PUFyymm
Value Categories in the PUF	**Approximate Distribution**
= 1 (CURCOMPyymm = 6)	1.1
= 0 (CURCOMPyymm in other non-missing value is 0, 1-4, 7, 8)	64.3
= .M (missing, may be applicable)	0.0
= .N (missing, not applicable)	34.6

Notes

Available for SSI beneficiaries with a non-missing PSTAyymm value.

Special missing condition (.M): No data available, but applicable to the beneficiary having a populated value for SSI payment status (PSTAyymm).

Special missing condition (.N): No data available, and not applicable to the beneficiary because SSI payment status (PSTAyymm) is not populated.

CURCOMP only identifies cases where the SSI couple came into existence as a result of successful application for SSI filed by the ineligible spouse of a beneficiary who is already receiving SSI. Because there are two other ways of that an SSI couple might come into existence (see below), CURCOMP does not successfully identify all SSI couples.

The other two ways of establishing an SSI couple are by the joining, either through marriage, residency, or both, of two beneficiaries already eligible for SSI, and by the simultaneous application of two individuals neither of whom were previously eligible for SSI. CURCOMP cannot be used to identify SSI couple status in these two cases.

The range of values for CURCOMPyymm is:

1 = Presumptive eligibility established for an eligible individual or child (TOA = AI, DI, BI, DC or BC)
2 = Presumptive eligibility established for a disabled child living with his/her mother (TOA = DM or BM)
3 = Presumptive eligibility established for a disabled child living with both parents (TOA = DB or BB)
4 = Presumptive eligibility established for an individual with an eligible spouse (TOA = AX, DX, or BX) or a disabled child living with his/her father (TOA = DF or BF)
6 = Posted to both halves of an eligible couple record to indicate it was formerly computed as

if it were a record for an eligible individual with an ineligible spouse.
7 = Posted to the portion of an eligible couple record that is being computed as an eligible individual with an ineligible spouse. (The portion of the record being treated as an ineligible spouse will contain a blank entry.)
8 = Posted to the survivor portion of a former eligible couple's record beginning with the effective month of an automated case composition.

Variable Name	STWCM_PUFyymm
Variable Label	COMBINED (SSDI/SSI) SUSPENSE OR TERMINATION DUE TO WORK (yymm = year/month)
Variable Description	Beneficiary's least restrictive non-payment status due to suspense or termination for work in the month indicated.
Data Type	Categorical
Data Available For	SSI/SSDI
Variable Location	Annuals Component
DAF Source Variable	STWCMyymm, as constructed in the DAF; DAF data extracted from the MBR
Related PUF Variables	STWCM_PUFyymm; STWDI_PUFyymm; STWSSI_PUFyymm
Value Categories in the PUF	**Approximate Distribution**
= 0 (STWCMyymm = 0)	72.5
= 1 (STWCMyymm = 1)	0.5
= 2 (STWCMyymm = 2, 3)	1.3
= 8 (STWCMyymm = 8, 9)	13.0
= .M (missing, may be applicable)	8.3
= .N (missing, not applicable)	4.4

Notes

Special missing condition (.M): No data available, but applicable to the beneficiary having a populated value for SSI payment status (PSTAyymm) or SSDI ledger account file participation status (LAFyymm).

Special missing condition (.N): No data available, and not applicable to the beneficiary because SSI payment status (PSTAyymm) and SSDI ledger account file participation status (LAFyymm) are not populated.

For concurrent beneficiaries, this variable takes on the lower number in STWDIyymm or STWSSIyymm, erring on the value closer to current pay status.

The range of values for STWCMyymm is:

0 = in current pay status in this month for either SSDI or SSI
1 = suspended due to work in this month for either SSDI or SSI and not in current pay in this month for the other program
2 = terminated due to work in this month for either SSDI or SSI and not in current pay or suspended for work in this month for the other program
3 = presumed terminated due to work in this month for either SSDI or SSI and not in current pay or suspended for work in this month for the other program: not in current pay, suspended due to work, or terminated due to work status in either program, but benefits terminated due to work in an earlier month in at least one program and never reinstated, and beneficiary is

alive and has not attained the FRA.
8 = in suspense status for reasons not determined to be work for either SSDI or SSI and not in current pay or suspended for work in this month for the other program.
9 = In terminated status for reasons not determined to be work in either SSDI or SSI and not in current pay or suspended for work in this month for the other program.
. (Missing Value) = Beneficiary is not yet entitled to benefits, has died, or has reached FRA

This measure is a constructed measure using other administrative data. Visit https://www.ssa.gov/disabilityresearch/daf.html, and consult Volume 3 of the documentation for more information about its construction.

Variable Name	STWDI_PUFyymm
Variable Label	SSDI SUSPENSE OR TERMINATION DUE TO WORK (yymm = year/month)
Variable Description	Beneficiary's non-payment status due to suspense or termination for work in the SSDI program in the month indicated.
Data Type	Categorical
Data Available For	SSDI
Variable Location	Annuals Component
DAF Source Variable	STWDIyymm, as constructed in the DAF; DAF data extracted from the MBR.
Related PUF Variables	BFWDI_PUFyymm; STWCM_PUFyymm; STWSSI_PUFyymm
Value Categories in the PUF	**Approximate Distribution**
= 0 (STWDIyymm = 0)	53.5
= 1 (STWDIyymm = 1)	0.4
= 2 (STWDIyymm in 2, 3)	1.1
= 8 (STWDIyymm = 8, 9)	2.8
= .M (missing, may be applicable)	0.3
= .N (missing, not applicable)	41.9

Notes

Special missing condition (.M): No data available, but applicable to the beneficiary having a populated value for SSDI ledger account file participation status (LAFyymm).

Special missing condition (.N): No data available, and not applicable to the beneficiary because SSDI ledger account file participation status (LAFyymm) is not populated.

The range of values for STWDIyymm is:

0 = in current pay status in this month
1 = suspended due to work in this month
2 = terminated due to work in this month
3 = presumed terminated due to work in this month: terminated due to work in an earlier month and never reinstated, and beneficiary is alive and has not attained FRA
8 = in suspense status in this month for a reason not determined to be work
9 = in terminated status in this month for a reason not determined to be work.

This measure is a constructed measure using other administrative data. Visit https://www.ssa.gov/disabilityresearch/daf.html, and consult Volume 3 of the documentation for more information about its construction.

Variable Name	STWSSI_PUFyymm
Variable Label	SSI SUSPENSE OR TERMINATION DUE TO WORK (yymm = year/month)
Variable Description	Beneficiary's non-payment status due to suspense or termination for work in the SSI program in the month indicated.
Data Type	Categorical
Data Available For	SSI
Variable Location	Annuals Component
DAF Source Variable	STWSSIyymm, as constructed in the DAF; DAF data extracted from the SSI-LF
Related PUF Variables	BFWDI_PUFyymm; STWCM_PUFyymm; STWDI_PUFyymm
Value Categories in the PUF	**Approximate Distribution**
= 0 (STWSSIyymm =0)	26.9
= 1 (STWSSIyymm = 1)	0.5
= 2 (STWSSIyymm = 2, 3)	0.3
= 8 (STWSSIyymm = 4, 8, 9)	26.9
= .M (missing, may be applicable)	8.2
= .N (missing, not applicable)	37.3

Notes

Special missing condition (.M): No data available, but applicable to the beneficiary having a populated value for SSI payment status (PSTAyymm).

Special missing condition (.N): No data available, and not applicable to the beneficiary because SSI payment status (PSTAyymm) is not populated.

The range of values for STWSSIyymm is:

0 = in current pay status in this month
1 = suspended due to work in this month
2 = terminated due to work in this month
3 = presumed terminated due to work in this month: terminated due to work in an earlier month and never reinstated, and beneficiary is alive and has not attained FRA
4 = suspended due to excess income in this month, part of that income is earnings from work, but countable unearned income is sufficient to cause suspension without consideration of earned income
8 = in suspense status in this month for a reason not determined to be work
9 = in terminated status in this month for a reason not determined to be work.

This measure is a constructed measure using other administrative data. Visit https://www.ssa.gov/disabilityresearch/daf.html, and consult Volume 3 of the documentation for more information about its construction.

Variable Name	T16GRSAMT_PUFyymm
Variable Label	SSI EARNINGS GROSS AMOUNT (SSR) (yymm = year/month)
Variable Description	SSI gross earnings in the month indicated.
Data Type	Dollars
Data Available For	SSI
Variable Location	Annuals Component
DAF Source Variable	T16GRSAMTyymm; DAF data extracted from the SSR
Related PUF Variables	T2GRSAMT_PUFyymm
Value Categories in the PUF	**Approximate Distribution**
= modified T16GRSAMTyymm (>0) based on Census guidelines	2.3
= 0	0.0
= .M (missing, may be applicable)	39.2
= .N (missing, not applicable)	58.4
Notes	Special missing condition (.M): No data available, but applicable to the beneficiary having a populated value for SSI payment status (PSTAyymm). Special missing condition (.N): No data available, and not applicable to the beneficiary because SSI payment status (PSTAyymm) is not populated. Visit https://www.ssa.gov/disabilityresearch/daf.html, and consult Volume 5 of the documentation for more details about usage of this variable.

Variable Name	T16NETAMT_PUFyymm
Variable Label	SSI SELF-EMP NET INCOME AMOUNT (SSR) (yymm = year/month)
Variable Description	Net amount of self-employment earnings in the month indicated.
Data Type	Dollars
Data Available For	SSI
Variable Location	Annuals Component
DAF Source Variable	T16NETAMTyymm; DAF data extracted from the SSR
Related PUF Variables	None
Value Categories in the PUF	**Approximate Distribution**
= modified T16NETAMTyymm (>0) based on Census guidelines	0.0
=0	0.0
= .M (missing, may be applicable)	41.6
= .N (missing, not applicable)	58.4
Notes	Special missing condition (.M): No data available, but applicable to the beneficiary having a populated value for SSI payment status (PSTAyymm). Special missing condition (.N): No data available, and not applicable to the beneficiary because SSI payment status (PSTAyymm) is not populated.

Variable Name	T2GRSAMT_PUFyymm
Variable Label	SSDI EARNINGS GROSS AMOUNT (DCF) (yymm = year/month)
Variable Description	SSDI gross earnings in the month indicated.
Data Type	Dollars
Data Available For	SSDI beneficiaries with positive earnings, based on DIERNLVL (DIERNLVL>0).
Variable Location	Annuals Component
DAF Source Variable	T2GRSAMTyymm; DAF data extracted from the DCF
Related PUF Variables	T16GRSAMT_PUFyymm
Value Categories in the PUF	**Approximate Distribution**
= modified T2GRSAMTyymm (>0) based on Census guidelines	2.8
= 0	0.4
= .M (missing, may be applicable)	55.1
= .N (missing, not applicable)	41.7
Notes	Special missing condition (.M): No data available, but applicable to the beneficiary having a populated value for SSDI ledger account file participation status (LAFyymm). Special missing condition (.N): No data available, and not applicable to the beneficiary because SSDI ledger account file participation status (LAFyymm) is not populated.

Variable Name	T2NETAMT_PUFyymm
Variable Label	SSDI SELF-EMP NET INCOME AMOUNT (DCF) (yymm = year/month)
Variable Description	The net amount of monthly self-employment earnings in the month indicated.
Data Type	Dollars
Data Available For	SSDI with positive earnings, based on DIERNLVL (DIERNLVL>0).
Variable Location	Annuals Component
DAF Source Variable	T2NETAMTyymm; DAF data extracted from the DCF
Related PUF Variables	None
Value Categories in the PUF	**Approximate Distribution**
= modified T2NETAMTyymm based on Census guidelines	0.4
= 0	57.9
= .M (missing, may be applicable)	0.0
= .N (missing, not applicable)	41.7
Notes	Special missing condition (.M): No data available, but applicable to the beneficiary having a populated value for SSDI ledger account file participation status (LAFyymm). Special missing condition (.N): No data available, and not applicable to the beneficiary because SSDI ledger account file participation status (LAFyymm) is not populated.

Variable Name	TWPDATA_PUFyymm
Variable Label	SSDI TRIAL WORK PERIOD CODE (yymm = year/month)
Variable Description	Identifies whether earnings are considered as a trial work month or as SGA in the month indicated.
Data Type	Categorical
Data Available For	SSDI
Variable Location	Annuals Component
DAF Source Variable	TWPDATAyymm, as constructed in the DAF and extracted from the DCF
Related PUF Variables	TWPCMPLMNTH1_PUF
Value Categories in the PUF	**Approximate Distribution**
= N (TWPDATAyymm =N and STWDIyymm in 0-3)	1.0
= O (TWPDATAyymm populated with another value and STWDIyymm in 0-3)	1.1
= S (TWPDATAyymm in E, S, or U and STWDIyymm in 0-3)	0.4
= T (TWPDATAyymm in 1-9 and STWDIyymm in 0-3)	0.5
= .A (STWDIyymm in 0-3 and TWPDATAyymm is missing)	52.1
= .M (missing, may be applicable)	2.9
= .N (missing, not applicable)	42.2
Notes	This variable contains data back to 1994, but is of lower quality prior to 2000; users should take caution when using in the earlier years. Special missing condition (.M): No data available, but possible applicable to the beneficiary based on having a populated value of 8 or 9 for SSDI suspense or termination due to work (STWDIyymm), meaning that at some point, the beneficiary could have completed a TWP. Special missing condition (.N): No data available, and not applicable to the beneficiary because SSDI suspense or termination due to work (STWDIyymm) is not populated. Available for SSDI beneficiaries with an applicable STWDIyymm code (0-3).

Variable Name	TWPDATA_PUFyymm
	The range of values for TWPDATAyymm is: 1 - 9 = Trial work month N = Not a trial work month and not SGA S = SGA month after end of TWP U = Not a TWP month but SGA determined to be an unsuccessful work attempt after TWP O = Other E = EPE

Variable Name	UINC_PUFyymm
Variable Label	SSI COUNTABLE UNEARNED INCOME (yymm = year/month)
Variable Description	Unearned income minus deductions and exclusions in the month indicated
Data Type	Dollars
Data Available For	SSI
DAF Source Variable	UINCyymm; DAF data extracted from the SSI-LF
Related PUF Variables	EICM_PUFyymm
Value Categories in the PUF	**Approximate Distribution**
= modified UINCyymm (>0) based on Census guidelines	16.5
= 0	48.9
= .M (missing, may be applicable)	0.0
= .N (missing, not applicable)	34.6
Notes	Special missing condition (.M): No data available, but applicable to the beneficiary having a populated value for SSI payment status (PSTAyymm). Special missing condition (.N): No data available, and not applicable to the beneficiary because SSI payment status (PSTAyymm) is not populated.

This page has been left blank for double-sided copying.

Mathematica

Princeton, NJ • Ann Arbor, MI • Cambridge, MA
Chicago, IL • Oakland, CA • Seattle, WA
Tucson, AZ • Woodlawn, MD • Washington, DC

EDI Global, a Mathematica Company

Bukoba, Tanzania • High Wycombe, United Kingdom

mathematica.org