Volume 3: Tips for Conducting Analysis with the DAF22

April 2024

Suggested Citation: U.S. Social Security Administration, Office of Retirement and Disability Policy, Office of Research, Demonstration, and Employment Support. (2022). "Disability Analysis File 2022 (DAF22) Documentation: Data from January 1994 through December 2022." Washington, DC: Mathematica, April 2024. [Volume 3: Tips for Conducting Analysis with the DAF22]. Retrieved from https://www.ssa.gov/disabilityresearch/daf.html.

See page vi for suggested citations for DAF datasets.

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

Contents

Tables

Suggested Citations

Users of the DAF 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 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), DAF Demographic File* [Data set]. Accessed on *<DATE>*.

2. For the DAF Awardee Data Mart (ADM) 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), DAF Awardee Data Mart (ADM) File [Data set]. Accessed on <DATE>.

3. For the DAF 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), DAF Annual File(s) (<DATA YEARS USED>) [Data set]. Accessed on <DATE>.

4. For the DAF Ticket 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), DAF Ticket File(s)* (*<DATA YEARS USED>*) [Data set]. Accessed on *<DATE>*.

5. For the DAF Payment 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), DAF Payment File(s)* (*<SUBFILE USED>*) [Data set]. Accessed on *<DATE>*.

6. For the DAF Standalone CDR 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), DAF Standalone CDR File* [Data set]. Accessed on *<DATE>*.

7. For the DAF Surveys and Demonstrations Extract 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), DAF Surveys and Demonstrations Extract File* [Data set]. Accessed on *<DATE>*.

8. For the DAF Rehabilitation Services Administration (RSA) 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), DAF Rehabilitation Services Administration (RSA) File* [Data set]. Accessed on *<DATE>*.

9. For the DAF Master Earnings (MEF) 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), DAF Master Earnings (MEF) File* [Data set]. Accessed on *<DATE>*.

Glossary

AB	Accelerated Benefits Demonstration
ADM	Awardee Data Mart
AIME	Average Indexed Monthly Earnings
BEST	Benefits Entitlement Services Team
BFW	Benefits forgone due to work
BIC	Beneficiary Identification Code
BMF	Budget Month Factor
BOAN	Beneficiary's Own Account Number
BOND	Benefit Offset National Demonstration
BOPD	Benefit Offset Pilot Demonstration
CAN	Claim Account Number
CDR	Continuing Disability Review
CER	Characteristics Extract Record 100% Field File
COLA	Cost-of-Living Adjustment
COSSN	Claimants Own Social Security Number
DAC	Disabled Adult Child
DAF	Disability Analysis File (previously known as TRF)
DBAD	Disabled Beneficiary and Dependents Extract
DCF	Disability Control File
DDS	Disability Determination Services
DER	Detailed Earnings Record
DI	Disability Insurance, also referred to as SSDI
DMG	Demographic component of the DAF
DRF	Disability Research File
DWB	Disabled Widow Beneficiaries
EDW	Enterprise Data Warehouse
EN	Employment Network (also called a TTW provider)
EPE	Extended Period of Eligibility
EVS	Enumeration Verification System
EXR	Expedited Reinstatement
FBR	Federal Benefit Rate
FCI	Federal Countable Income
FIPS	Federal Information Processing Standards (in reference to U.S. Census standardized codes for uniform identification of geographic entities)

FRA	Full Retirement Age
HI	Hospital Insurance (Medicare Part A)
HOPE	Homeless Outreach Projects and Evaluation Demonstration
HUN	Housed Under Number
ICD-9	International Classification of Diseases Coding Scheme
IPE	Individualized Plan for Employment, developed by State VR Agency
IRS	Internal Revenue Service
IRWE	Impairment-Related Work Expense
LAF	Ledger Account File
LAUS	Local Area Unemployment Statistics
MBR	Master Beneficiary Record
MEF	Master Earnings File
MHTS	Mental Health Treatment Study
MIE	Medical Improvement Expected
MO	Milestone + Outcomes payment system
MPR-EVS	Mathematica's EVS
NBS	National Beneficiary Survey
NSCF	National Survey of SSI Children and Families
NUMIDENT	Numerical Identification File
OIM	Office of Information Management
OO	Outcomes-Only payment system
PAN	Person's Account Number
PASS	Program to Achieve Self-Support
PHUS	Payment History Update System
PIA	Primary Insurance Amount
PIN	Personal Identification Number
POD	Promoting Opportunity Demonstration
POMS	SSA's Program Operations Manual System
PROMISE	Promoting Readiness of Minors in SSI
Provider	Service provider under TTW (also called an EN)
PUF	Public Use File
RECS	Race and Ethnicity Classification System
REMICS	Revised Management Information Counts System
RIB	Retirement Insurance Benefits
RMA	Retrospective Monthly Accounting

RSA	Rehabilitation Services Administration
RSA-911	RSA Case Service Report
SAIPE	Small Area Income and Poverty Estimates
SAS	Statistical Analysis Software, used to produce the DAF
SCWF	Standalone Companion Work File
SED	Supported Employment Demonstration
SER	Summary Earnings Record
SGA	Substantial Gainful Activity
SMI	Supplemental Medical Insurance (Medicare Part B)
SNAP	Supplemental Nutrition Assistance Program
SSN	Social Security Number
SSA	Social Security Administration
SSDI	Social Security Disability Insurance (also referred to as DI)
SSI	Supplemental Security Income
SSI-LF	SSI - Longitudinal File
SSR	Supplemental Security Record
STW	Suspension or termination of cash benefits for work
T2	Title II, the SSDI Program
T16	Title XVI, the SSI Program
TANF	Temporary Assistance for Needy Families
TCNEI	Total countable non-earned income
TKT	DAF component containing data related to TTW participation
TRF	Ticket Research File, now called the DAF
TTW	Ticket to Work
TWP	Trial Work Period
VR	Federal/State Vocational Rehabilitation program/agency
VRRMS	Vocational Rehabilitation Reimbursement Management System; data from this system is contained in the Payments component
YTD	Youth Transition Demonstration

Overview of DAF Documentation

The documentation for the DAF consists of the eleven volumes described below. Questions about these documents should be directed to ORDES.DAF@ssa.gov. All of these documents are available at https://www.ssa.gov/disabilityresearch/daf.html.

- **Volume 1:Getting Started with the DAF22.** Provides an overview of the structure and contents of the DAF and related linkable files.

- **Volume 2: Working with the DAF22.** Contains practical suggestions such as how to extract data and interpret blank or missing variables as well as more detailed information on DAF data marts and linkable files.

- **Volume 3: Tips for Conducting Analysis with the DAF22.** Contains suggestions for working with common research concepts in the DAF such as program participation, benefits paid versus benefits due, and constructed measures related to beneficiary work activity resulting in the loss of cash benefits.

- **Volume 4: Lists of DAF22 Variables.** Contains lists of new, changed, and deleted variables, as well as lists of variables by DAF component and analytic category.

- **Volume 5: DAF Variable Detail Pages.** Contains specifications for each DAF variable, including name, definition, data format, identification of the DAF component to which it belongs, data source, availability, and (where applicable) SAS code used to construct the variable.

- **Volume 6: Validating the DAF22 Against Other Sources.** Provides an explanation of validation methods and summary of validation results.

- **Volume 7: DAF22 Development History and Construction Methods.** Describes key changes in DAF construction methodology over time as well as a description of each step in the current year DAF construction process.

- **Volume 8: DAF22 Construction Workflow Charts and Task Tables.** Provides detailed information in both chart and table format on each step in the current year DAF construction process.

- **Volume 9: DAF22 Source File Descriptions.** Describes the administrative source files used to construct the DAF.

- **Volume 10: DAF22 Administrative Source File Documentation.** Contains documentation from SSA or other agencies on the administrative source files described in Volume 9.

- **Volume 11: DAF22 Construction Code**. Contains all SAS code used to construct the DAF.

- **Volume 12: DAF22 RSA Administrative Source File Documentation.** Contains a description of the processing of Rehabilitation Services Administration (RSA) data for linkage to the DAF, along with documentation from RSA on the RSA-911 files.

The following table provides specific locations for common research-related questions and issues.

In order to …	Refer to …
Get started with a research task	Volume 2, "Working with the DAF22," for information about selecting beneficiaries using finder files versus selection criteria
Identify what's changed in the latest version of the DAF	Volume 1, "Getting Started with the DAF22"
View lists of DAF variables	Volume 4, "Lists of DAF22 Variables"
Understand individual variable definitions, specifications, and value ranges	Volume 5, "DAF Variable Detail Pages"
Understand the structure of the DAF data files at a high level	Volume 1, "Getting Started with the DAF22"
Identify variables for a specific research task	Volume 4, "Lists of DAF22 Variables," for a list of variables contained within each DAF file and by analytic category
Understand the beneficiaries for which the DAF does and does not contain data	Volume 1, "Getting Started with the DAF22"
Identify administrative data sources for the DAF	Volume 9, "DAF22 Source File Descriptions"
Understand the linkage of the DAF to RSA-911 data and contents of the RSA files	Volume 12, "DAF22 RSA Administrative Source File Documentation"
Generate ideas for using the DAF more efficiently	Volume 1, "Getting Started with the DAF22" and Volume 2, "Working with the DAF22"
Find suggested ways to identify common research concepts in the DAF, such as calculating age of retirement, or disability title	Volume 3, "Tips for Conducting Analysis with the DAF22"
Understand what variables have changed in the most recent DAF	Volume 4, "Lists of DAF22 Variables"
Read about how information in the DAF is validated against other sources	Volume 6, "Validating the DAF22 Against Other Sources"

I. Calculating Age-Based Measures

A. Calculating age at a point in time

While birthdate information is available from the Disability Analysis File (DAF), there is no age variable in the database, meaning that users must define it using date of birth relative to another point in time. As discussed in Volume 2, DOBBEST is the most accurate DAF variable for date of birth. Age can simply be calculated as reference date minus DOBBEST. One common method for performing this calculation is using a Statistical Analysis Software (SAS) INTCK statement such as *age=INTCK('Months',DOBBEST,ref date,'C')/12*. This statement will count the number of months between DOBBEST and the reference date and then divide by 12 to get an age in years and whole month fractions. This method is particularly useful because it can be used to compare age to Full Retirement Age (FRA), which is calculated in two-month intervals based on birth year.

Selection of a reference date or dates depends on the research design. In many cases, users just want to calculate age on a particular calendar date. In other cases, users want to calculate age at the time of a particular program activity, such as at program entry. In this case, users would define the reference date from the variable they choose to use as the point of benefit start (discussed elsewhere in this volume). In other cases, users want to calculate age over a period such as a calendar year. This may be the case if one wants to create an extract of beneficiaries between the ages of 18 and 65 during the calendar year 2010. In this case, the user should calculate two age variables using two reference dates—January 1, 2010 and December 31, 2010—then keep all records that meet the age selection criteria at either time.

Users should keep in mind that the methods described for coding age at a point in time will code valid age values for deceased beneficiaries. To code age as missing for deceased beneficiaries, use the variable DODBEST. If it is blank or missing, no death date has yet been recorded for the beneficiary. If it is populated and has a value that is earlier than the reference date, code age as missing.

B. Determining FRA

A beneficiary's FRA for purposes of Social Security benefits was previously 65 years, but amendments to the Social Security program in 1983 gradually increased FRA from 65 years, for those born in 1937 or earlier, to 67 years, for those born in or after 1960 (Table I.1). There are two variables on the DAF related to FRA. The first is FRA, the beneficiary's full retirement age as defined by the table below. The second is FRADATE, which is the date the beneficiary reaches FRA. Note that for beneficiaries born on January 1 of any year, the Social Security Administration (SSA) considers the previous year as their year of birth. The two FRA variables are populated for all beneficiaries in the DAF, regardless of whether they are still alive at the date they would attain FRA.

Table I.1. Retirement age, by year of birth

Year of Birth	Full Retirement Age
1937 or earlier	65
1938	65 + 2 months
1939	65 + 4 months
1940	65 + 6 months
1941	65 + 8 months
1942	65 + 10 months
1943-1954	66
1955	66 + 2 months
1956	66 + 4 months
1957	66 + 6 months
1958	66 + 8 months
1959	66 + 10 months
1960 and later	67

Note: If beneficiary was born on January 1 of any year SSA considers the previous year as their year of birth.

II. Program Participation and Disability Onset

To determine Social Security Disability Income (SSDI) and Supplemental Security Income (SSI) program participation at a given point in time, researchers often consult monthly payment status variables in the DAF along with one-time measures such as date of first payment receipt, first entitlement, and disability onset. Additionally, it is important that researchers use the appropriate variables for SSDI and SSI; variables are different for SSDI and SSI and are indicated throughout the chapter.

Note that in this chapter we are discussing program participation in terms of eligibility (i.e., entitlement to payment), not actual payment. There are many reasons why a beneficiary may have been eligible for but did not receive a payment in a given month, including recovery of an overpayment previously incurred and months prior to a favorable disability decision being made for which eligibility is retroactively determined. Similarly, it is possible for beneficiaries to have received payments in a month when no benefit payment was due in that month, such as when an underpayment for prior months has been retroactively determined but there is no current-month eligibility.

By description, there are numerous DAF variables that appear to delineate a beneficiary's program participation, but researchers should exercise caution and ensure that they fully understand these variables before basing analyses on them. For example, the Demographic component of the DAF (DMG) variable DOEC (SSDI Most Recent Entitlement Date) is the most recent entitlement date to SSDI, but it remains populated even after benefits have terminated and so cannot be used to determine ongoing eligibility. However, it can generally be used with DOST (Date of Suspension or Termination) to identify the most recent period of eligibility. DOEI (Date of Initial Entitlement) gives the first date of entitlement to SSDI benefits. If there are multiple periods of eligibility these variables cannot be used in combination to identify the dates of all periods of eligibility because all three—DOEC, DOST, and DOEI—are one-time only variables. In addition, DOEI may identify a non-disability-based period of entitlement, such as a person who initially receives Social Security early retirement benefits prior to receiving an SSDI award, or one who received auxiliary benefits, such as a dependent child, before receiving SSDI benefits.

Similar issues exist for comparable SSI variables such as ELG_RD (most recent date of SSI eligibility), SSIELIG_FIRST (earliest available SSI eligibility date) and CURSTAT (most recent SSI payment status). These issues are compounded by the fact that SSI data are maintained primarily in a multiple-record format of the Supplemental Security Record (SSR) rather than the single-record format of the Master Beneficiary Record (MBR) for SSDI (see the SSR and MBR descriptions in Volume 9 for more information about these data sources). Data are sometimes overwritten when new SSR records are created, which is why, for example, SSIELIG_FIRST is characterized as the earliest *available* date rather than the earliest actual date.

Absent compelling reasons to the contrary, researchers are therefore advised to follow the guidance outlined in this chapter to determine program participation. The variables SSDIyy and

SSIyy have been constructed for researchers who only need to identify program participation on an annual basis. These binary flag variables indicate whether or not a beneficiary was in current pay status as defined below for any month in a given year. The SSDIyy and SSIyy flags also require that SSI recipients are under the age of 65 and SSDI beneficiaries have not reached FRA. It is important to note that this may not meet all research needs, especially in the case of SSDI, where the current pay status flag may show continued program participation for individuals who claim retirement benefits prior to FRA. The variable detail pages for these variables offer suggestions for users who may be interested in narrowing the population identified by SSDIyy to align more closely to published statistics on the number of SSDI beneficiaries.

A. Determining monthly participation based on payment entitlement status variables

Receipt of SSDI or SSI can vary from month to month as participants gain and lose their eligibility to benefits according to changing circumstances such as incarceration or a return to employment. In other words, once eligible for benefits, a participant will not remain eligible indefinitely. The DAF contains monthly variables that indicate whether a beneficiary was entitled to payments for a given month, regardless of whether or not that payment was actually received by the beneficiary in that month. These variables are LAFyymm for SSDI and PSTAyymm for SSI. After reading this section, users interested in identifying beneficiaries who were entitled to SSDI or SSI during a certain period of time may want to consult the DAF code library, available at https://www.ssa.gov/disabilityresearch/daf.html#library, which provides sample code for that task.

1. SSDI beneficiaries

All SSDI beneficiaries in the DAF received disability benefits at one point between March 1996 and December of the DAF year, and were selected based on their participation in SSDI. That status was in part determined by the Type of Claim (TOC).[1] The TOC variable identifies whether a beneficiary received retirement or disability benefits and can be aligned to a particular month to be used in conjunction with LAFyymm. More information about working with TOC is available in the "Identifying SSDI beneficiary type" later in this chapter.

To determine program participation status based on eligibility for an SSDI beneficiary, use the SSDI payment status code variable (also known as the "ledger account file status" variable), LAFyymm, for a specific month or range of months. For example, to determine participation status for an SSDI beneficiary in 2003, use LAF0301 through LAF0312. Current pay status is indicated by a value beginning with C or E, with C being much more common (E only applies to beneficiaries of the Railroad Retirement Board). If a beneficiary was in current pay status for January through October 2003 then terminated from the SSDI program in November 2003 due to death (for example), the LAFyymm variables for the year 2003 would contain values as shown in

---

[1] Beneficiaries are selected for inclusion in the DAF based on the TOC variable appearing on the DBAD snapshot files but the TOC variable on the DAF is drawn from the MBR which is subject to retroactive updates.

Table II.1, with a Ledger Account File (LAF) value of T1 referring to termination due to death. See the LAFyymm entry in Volume 5 for a full description of possible variable values.

Table II.1. Illustration of monthly participation variables for SSDI beneficiaries

Variable	LAF 0301	LAF 0302	LAF 0303	LAF 0304	LAF 0305	LAF 0306	LAF 0307	LAF 0308	LAF 0309	LAF 0310	LAF 0311	LAF 0312
Value	C	C	C	C	C	C	C	C	C	C	T1	T1

The DMG file includes variables that allow users to identify the most recent LAF status of a beneficiary. LAF_MR indicates the most recent LAF code for the beneficiary, while the LAF_MR_DT indicates the month in which that most recent status occurred.

The STWDIyymm variable can also be used to establish eligibility in a month or range of months. Months in which beneficiaries have a status of "0" for this variable are months in which the beneficiary was eligible for payment. See Chapter VIII in this volume for additional information on how to use this variable.

2. SSI beneficiaries

To determine the participation status for an SSI beneficiary, use the SSI payment status code variable, PSTAyymm, for a specific month or range of months. For example, to determine participation status for an SSI beneficiary in 2003, use PSTA0301 through PSTA0312. Current pay status is indicated by a value of C01.[2] If a beneficiary was in current pay status for January thru October 2003 then suspended from the SSI program in November 2003 as a result of incarceration (for example, PSTA value of N22), the PSTAyymm variables for the year 2003 could contain values as shown in Table II.2 below. See the PSTAyymm entry in Volume 5 for a list of possible variable values.

Table II.2. Illustration of monthly participation variables for SSI beneficiaries

Variable	PSTA0 301	PSTA0 302	PSTA0 303	PSTA0 304	PSTA0 305	PSTA0 306	PSTA0 307	PSTA0 308	PSTA 0309	PSTA0 310	PSTA 0311	PSTA0 312
Value	C01	C01	C01	C01	C01	C01	C01	C01	C01	C01	N22	N22

As is the case for SSDI beneficiaries, the DMG file includes variables that allow users to identify the most recent PSTA status of a beneficiary. PSTA_MR indicates the most recent PSTA code for the beneficiary, while the PSTA_MR_DT indicates the month in which that most recent status occurred.

---

[2] In addition, payment status codes of M01 and M02 indicate that eligibility has been manually determined. Although M01 and M02 are sometimes associated with ineligibility, this can be difficult to determine because of how manual computations display (or do not display) in the administrative data. Many researchers consider M01 and M02 payment status codes to indicate program participation analogously to C01 because these codes generally indicate manually computed eligibility rather than ineligibility.

The STWSSIyymm variable can also be used to establish eligibility in a month or range of months. Months in which beneficiaries have a status of "0" for this variable are months in which the beneficiary was eligible for payment. See Chapter VIII in this volume for additional information on how to use this variable.

B. Determining initial benefit entitlement date

Conceptually, the initial benefit entitlement is the month for which the beneficiary is first due payments. This is often different than the first benefit payment date because when beneficiaries are first deemed eligible, they may receive retroactive benefits for some of the months since their disability onset date. For example, someone receiving benefit payment for the first time in September 2004 may also receive retroactive payments for the previous three months at the same time as the September benefit payment is made. In this case, the initial benefit entitlement date would be June 2004, while the first payment date would be September 2004. The payment status codes described above will reflect eligibility in a situation such as this one without regard to whether a payment was actually made in that month.

Two constructed variables on the DAF ADM–DIELIGDT for SSDI beneficiaries and SSIELIGDT for SSI beneficiaries–identify the first month of eligibility for benefits. These variables, however, are only available for beneficiaries on the Awardee Data Mart (ADM), i.e., only beneficiaries who receive their *first benefit payment* as a non-retiree adult in 1996 or later. Because beneficiaries who received their first payment before 1996 are excluded from the ADM, the ADM contains millions fewer beneficiaries than the full DAF DMG (see Volume 2 for more information about the ADM).

In addition to data in the ADM, variables from the MBR in the DMG file can identify additional entitlement dates for SSDI beneficiaries but should be used with caution. There are two variables that could be of interest: 1) the DOEI ("Date of Initial Entitlement") variable and 2) the earliest date stored in the variable series ENTDATn, which records entitlement dates for up to twelve possible periods of eligibility. In the vast majority of MBR records, DOEI and the earliest value from ENTDATn contain the same date. In instances when the dates do not match, there does not seem to be a consistent logic for the difference, though they may make sense in particular analytical circumstances. For SSI beneficiaries, there is no comparable entitlement date in the administrative data. SSI eligibility can be determined by evaluating monthly SSI program status codes (PSTA) for the first instance of current pay.

C. Identifying first benefit payment date

The first payment date for SSDI or SSI benefits is often used to determine when a beneficiary began participating in the disability program. As described above, in many cases for SSDI beneficiaries, this is different than the date of initial benefit eligibility. Two variables on the ADM identify the first benefit payment date received as an adult under SSDI or SSI, if that date was in 1996 or later: DIAWARDDT and SSIAWARDDT, respectively.

The variable MINFRSTP (SSI Earliest Available Application, First SSI Payment Date) on the DMG contains the initial SSI payment date for the earliest available entitlement period, regardless of adult or child status. However, a word of caution regarding MINFRSTP—SSA periodically overwrites the data from which this variable is drawn (FIRST-PAY-DTE in the SSR) and sometimes the initial payment date for the first entitlement period is replaced by the initial payment date for a subsequent entitlement period. To illustrate, suppose a beneficiary has three entitlement periods on record in the SSR spanning a period of twenty-five years. In this example, the first entitlement period began twenty-five years ago, the second began twenty years ago, and the third began two years ago. The date of the initial payment for all three entitlement periods is stored in the SSR, and the earliest occurring initial payment date represents the initial payment date for the first entitlement record, twenty-five years ago. Later, the SSR data are updated, and perhaps because the earliest entitlement period began such a long time ago, the update routine overwrites the data for the earliest entitlement period. After the overwrite all that remains are the two most recent entitlement periods, with the initial payment date for the first entitlement period occurring twenty years ago. In the absence of any other information, it appears that this beneficiary's initial payment date occurred twenty years ago, not twenty-five years ago. In actuality, the earliest initial payment date from the SSR reflects the earliest **available** payment date, which is not necessarily the date the SSI beneficiary first received a payment.

For SSDI beneficiaries there is no administrative variable that identifies first payment. Instead, when not using the ADM the first SSDI payment for a beneficiary is determined by examining the LAFyymm and PAYDyymm variables to determine the first month that the beneficiary is in current pay status and received a payment. Because these variables only go back to 1994 in the DAF, caution should be used because payments may have started before this time. Also, the first payment date under SSDI using the combination of LAFyymm and PAYDyymm could have been for a child auxiliary payment unrelated to a later period of adult disability. In this case, the first payment variable as an adult recorded in the ADM may differ from what a user might calculate.

D. Determining disability onset date

Researchers are sometimes interested in the date a beneficiary first had an onset of disability. In most cases, this is distinct from the date of initial benefit eligibility and the first payment date. To determine the earliest date of onset of a beneficiary's date of disability:

- for SSDI beneficiaries, use the first occurrence of DDOn (SSDI Disability Onset Date), where n ranges from 1 to 12 (data for up to 12 periods of disability are recorded);[3] and

- for SSI beneficiaries, use SSIELIG_FIRST (SSI Earliest Eligibility Date (Constructed))— note that this variable contains the earliest available eligibility date; in some cases data on earlier periods of eligibility, particularly if they were far in the past, are overwritten.

---

[3] As discussed in Volume 2, n=1 may not be the earliest occurrence and it is important to inspect the data to determine that before working with this variable as described.

E. Identifying SSDI beneficiary type

To identify whether a beneficiary was in current pay status for disability or retirement for a particular point in time, there are multiple variables that users should consult. Type of Claim variables (TOC) indicate whether a beneficiary received retirement or disability benefits, and thus should be used in conjunction with LAFyymm to determine whether a beneficiary was in current pay status for disability (versus retirement) benefits in a given month. Beneficiary Identification Code (BIC) variables indicate whether the beneficiary received benefits as a primary beneficiary or as a spouse or child, which may be important for certain research uses.

The variables used to determine the type of benefits is different depending on the time period of interest. From May 2009 onward, Type of Claim (TOC), Beneficiary Identification Code (BIC), and the historical version of BIC (HBICn) provide the relevant information. Prior to May 2009—reflecting different source information, discussed below—the relevant variables are named TOC_HISTn and BIC_HISTn.

The MBR rewrite in May 2009 and implications for TOC and BIC variables

TOC and BIC are drawn from the MBR, which underwent a significant rewrite in May 2009. Since that time, TOC occurrences have been stored as "n" variables, which means that along with TOC_STARTn, a history of the type of benefits received can be constructed. Historical occurrences of BIC are contained in the variable Historical BIC (HBICn) with analogous date variables to identify the relevant period of each BIC (see Volume 5).[4]

The rewrite of the MBR, meant that only the most recent value of the TOC and BIC variables were available for the period preceding May 2009. This left a serious deficiency in the data when trying to conduct longitudinal research that relied on these variables; as data prior to 2009 had been overwritten as beneficiaries' status changed. To help users assess historical TOC and BIC status before that date, we have accessed snapshot extracts from the MBR to construct a set of historical values that cover the period from January 1995 through May 2009. To compile this information, we relied on two sources. We used the semiannual MBR100 files from January 1995 to December 2000 and the monthly Disabled Beneficiary and Dependents Extract (DBAD) from May 2001 to May 2009.[5] These files contain the TOC and BIC values as they were at the point in time the file was extracted, but the files differ in their structure requiring separate processing. In particular, the MBR100 is a semiannual file, meaning that TOC and BIC variables for the period from January 1995 to December 2000 are based on six-month intervals.[6]

---

[4] The 'n' suffix for TOC is not necessarily associated with the 'n' suffix for BIC (i.e. TOC3 is not necessarily concurrent with BIC3). We discuss this in more detail in the section "Working with TOC and BIC variables" below.

[5] The MBR100 files were last available in December 2000, and there are periods after that time in which the DBAD records are no longer stored on SSA's system; from January to April 2001 and again from June to August 2006. The historical TOC and BIC were built as if that gap didn't exist, meaning that we do not observe any changes during that period, and any changes that are present in April 2001 and August 2006 are assigned to that month.

[6] To build the historical TOC and BIC values, the MBR100 was sorted by SSN, BIC, and Date of Current Entitlement (DOEC) and then re-sorted by just SSN and DOEC keeping records with unique combinations of SSN and DOEC. This method results in retention of only the Primary (BIC=A) record when the beneficiary is dually entitled. This set of data is then reduced to one record per SSN with a historical series of n-suffixed variables that

The DBAD is a monthly file with a limited set of variables for disability beneficiaries and their dependents, meaning that the TOC and BIC variables during this period are closer to those since the rewrite, potentially changing on a monthly basis.[7]

Using MBR100 and DBAD data spanning the period from 1995 through May 2009, we constructed a set of historical TOC and BIC values, and their associated date variables, listed in Table II.3. Beginning in May 2009, the TOCn and HBICn values (in conjunction with their associated date values) can be used to span the period through the present. We therefore have both TOC_HISTn and TOCn and both BIC_HISTn and HBICn values for May 2009. Our inspection of the values for the current and historical values suggests that they overlap in almost all cases.[8]

Table II.3. Historical TOC and BIC variables in the DAF

Variable name	Variable availability	Source file	Variable label
TOC variables			
HISTORICAL (through May 2009)			
TOC_HISTn	January 1995 to May 2009	MBR100, DBAD	SSDI HISTORICAL TYPE OF CLAIM
TOC_HIST_STARTn	January 1995 to May 2009	MBR100, DBAD	SSDI HISTORICAL TYPE OF CLAIM (TOC) START DATE
TOC_HIST_NUM	January 1995 to May 2009	MBR100, DBAD	SSDI NUMBER OF HISTORICAL TYPE OF CLAIM (TOC) OCCURRENCES
CURRENT (May 2009 onward)			
TOCn	May 2009 onward	MBR	SSDI TYPE OF CLAIM
TOC_STARTn	May 2009 onward	MBR	SSDI TYPE OF CLAIM (TOC) START DATE
TOC_NUM	May 2009 onward	MBR	SSDI NUMBER OF TYPE OF CLAIM (TOC) OCCURRENCES
BIC variables			
HISTORICAL (through May 2009)			
BIC_HISTn	January 1995 to May 2009	MBR100, DBAD	SSDI HISTORICAL BENEFICIARY IDENTIFICATION CODE (BIC)
BIC_HIST_STARTn	January 1995 to May 2009	MBR100, DBAD	SSDI HISTORICAL BENEFICIARY IDENTIFICATION CODE (BIC) START DATE

represent the progression of TOCs and BICs for the beneficiary during the period captured by the MBR100. The progression is determined by using changes in DOEC.

[7] We followed a similar process to build the historical TOC and BIC values using the DBAD as we did in the MBR100 described in the footnote above, using monthly data instead of semi-annual. One difference is that in the MBR100, we used DOEC to set the start date for TOC and BIC. For the DBAD-based TOC and BIC, we set the start date based on the month in which a change in the value was identified. This means that the start date from the MBR100 can be as early as the SSDI program, but for the DBAD, cannot be earlier than May 2001.

[8] Among SSDI beneficiaries in current pay status or suspense for work status (STWDI=0,1) in the DAF17 10% file in May 2009, the latest historical TOC value was the same as TOC1 in 92.5 percent of cases. For the same comparison using BIC, the values matched in 85.6 percent of cases. Among cases that did not match, most of the mismatch reflected different populated values in the historical and current occurrence, though in a minority of cases it was due to missing data on the historical occurrence.

Table II.3. Historical TOC and BIC variables in the DAF

Variable name	Variable availability	Source file	Variable label
BIC_HIST_NUM	January 1995 to May 2009	MBR100, DBAD	SSDI HISTORICAL BENEFICIARY IDENTIFICATION CODE (BIC) OCCURRENCES
CURRENT (May 2009 onward)			
HBICn	May 2009 onward	MBR	HISTORICAL BIC n
BDOE_STARTn	May 2009 onward	MBR	SSDI DATE OF ENTITLEMENT START
BDOE_TERMn	May 2009 onward	MBR	SSDI DATE OF ENTITLEMENT TERMINATION
Variables used to process historical TOC/BIC			
DOEC_DOEI_SAME	January 1995 to May 2009	MBR100, DBAD	HISTORICAL TOC/BIC BUILT FROM RECORD WITH SAME DOEC AND DOEI
DOEC_SET_TO_DOEI	January 1995 to May 2009	MBR100, DBAD	HISTORICAL TOB/BIC BUILT FROM RECORD WHERE A MISSING DOEC WAS SET TO THE DOEI

Working with TOC and BIC variables

DAF users often are interested in identifying sub-groups of beneficiaries, such as disabled workers or disabled widows, which require using TOC and BIC. It is important to first determine the time period of the analysis, to narrow in on whether the MBR variables from 2009 and later should be used, or whether the historical data before May 2009 are applicable. Table II.3 can be used for that purpose. Even after identifying the right variables, users often have questions about working with TOC and BIC. We attempt to answer some of those questions in this section.

Why is TOC_START1 missing for many SSDI beneficiaries? TOC_START1, the start date for the first TOC occurrence after the MBR rewrite, is blank in most cases. It seems safe to presume based on our investigation, that for these cases the start date occurred prior to May 2009, and that the latest occurrence of TOC_HIST_STARTn can be used instead.

How do I know the relevant start and end *dates for historical BIC records?* Unlike the occurrence-based BIC values, there is not a termination date for BIC_HISTn that aligns to BDOE_TERMn. Our presumption is that the BIC remained active until or unless the next value of BIC is populated, and that the BIC_HIST_STARTn date can be used as the termination date for the previous BIC value.

How often do beneficiaries have multiple TOC and BIC records? Based on DAF18 data, about three-quarters (77.2 percent) of SSDI beneficiaries have a single populated TOCn value, and almost all beneficiaries 99.2 percent) have 3 or fewer populated TOC values. Among SSDI beneficiaries with DOEI prior to May 2009, 99.0 percent have 3 or fewer populated TOC_HISTn variables.

Based on data from DAF18, virtually all SSDI beneficiaries (99.9 percent) have 3 or fewer populated HBIC values; almost two-thirds have a single populated value (61.8 percent) and one-third (33.4 percent) have two populated values. However, HBIC values are often duplicates. For example, of the set of beneficiaries with two populated values of HBIC, 98.5 percent have the

same value for both HBIC1 and HBIC2. Some, but not all, of this repetition is caused by a new *n* for HBIC being generated when TOC changes. Specifically, of beneficiaries with HBIC1 equal to HBIC2, 56 percent have BDOE_START2 equal to TOC_STARTn.

Using BIC_HISTn in DAF18, 99.7 percent of SSDI beneficiaries have 3 or fewer populated BIC_HIST values, with around 85 percent having only one populated value and 10 percent have two populated values.

Are the 'n' occurrences for TOC and BIC in chronological order? The short answer is not necessarily. Thus, it is important for users to sort occurrences using the relevant date variables for each measure, rather than selecting based on the order of the n variables.

Most, but not all, of observations of TOCn are arranged in chronological order with n = 1 being the earliest; 93 percent of beneficiaries with two populated observations of TOC have TOC_START1 preceding TOC_START2, while 89 percent of beneficiaries with three populated observations of TOC have TOC_START2 preceding TOC_START3. Because the TOC_HIST measures were constructed for the DAF with research use in mind, those 'n' occurrences are in chronological order with the n = 1 being the earliest.

Unlike TOC, the BICn variables are in chronological order with n=1 being the first occurrence, based on DAF18 data. In general, current BIC is duplicated in the most recent historical BIC. For 98 percent of SSDI beneficiaries, HBIC (largest n) equals BIC. We also found that for 99 percent of SSDI beneficiaries, BDOE_START (largest n) equals the most recent SSDI entitlement date (DOEC).

Does the 'n' for TOC align with the 'n' for BIC? In general, it is not appropriate to assume TOC1 is associated with HBIC1. Based on DAF18, we found that 20 percent of beneficiaries have a different number of observations of TOC compared to HBIC, with 16.5 percent of beneficiaries having more populated observations for HBIC than for TOC. It may be that this simply reflects the fact that HBIC values seem to repeat across subsequent values of *n* (i.e. HBIC1 equals HBIC2), which may mean that HBIC gets "updated" to the same value whenever TOC changes. We have not verified whether a new *n* for HBIC is generated every time TOC changes. Users should use the associated date variables (e.g., TOC_STARTn, BDOE_STARTn, BDOE_TERMn) to establish the alignment of TOC and BIC at different points in time.

III. Benefit Amount Data

Monthly benefit amount data are available separately for SSDI and SSI benefits. Categories of benefit variables include the amount due to the beneficiary, the amount paid, the federal or state dollar portion of the SSI benefit payment, and the amount due or paid for dependents. The amount paid may differ from the amount due as deductions may be applied to account for a variety of factors including Medicare premiums or previous overpayments. The benefit amount variables require special handling for accurate use and details are provided below.

A. Formats for amount variables

A variety of currency formats are in use for benefit and earnings amounts variables. SSDI benefit amount variables generally include cents and use SAS formats such as "$$$$.c" or "$$$$.cc". SSI benefit amount variables do not include cents and use SAS formats such as "$$$$$$" or "$$$". Other variables related to SSI earnings, such as IEA_BLINDyymm (SSI earned income amount – blind work expense) and IEA_PASSyymm (SSI earned income exclusion amount – PASS), do include cents, using SAS formats such as "$$$$.cc."

B. SSDI benefit due variables

There are three variables for SSDI benefits due:

- **MBAyymm–Federal SSDI Benefit Due.** Monthly Benefit Amount (MBA) is the benefit amount due to the beneficiary *before* rounding or adjustments and can be thought of as the "base" amount due. The source variable is the SSA variable MBA from the MBR file.

- **DUEDyymm–SSDI Benefit Due.** DUED is the benefit amount due to the beneficiary *after* rounding to the nearest lower dollar but *before* adjustments are made. SSA refers to this as "amount credited" and it is the basic benefit amount field to use for analysis as other amount fields are affected by factors such as overpayments or deductions for Medicare. Mathematica uses DUED to determine a beneficiary's total benefit amount for a given month. The source variable is the SSA variable MBC (Monthly Benefit Credited) from the MBR file

- **DUEOyymm–SSDI Dependent Benefit Due.** This variable is similar to DUEDyym but represents the benefit amount due to the beneficiary's dependents who are entitled to cash benefits as a result of the beneficiary's entitlement, *after* rounding to the nearest lower dollar but *before* adjustments are made. It is based on the same SSA source variable, MBC from the MBR file, and is calculated by summing only the benefit amounts due to dependents of the beneficiary. It excludes benefits due to the beneficiary but includes all benefits payable to dependents, including Disabled Adult Children (DACs) and Disabled Widow Beneficiaries (DWBs).

To accurately determine the monthly benefit amount due to an SSDI beneficiary, it is not enough to use only the benefit due variables—you must also use the payment status code (also known as the "ledger account file status," or LAF variable) for the corresponding month, LAFyymm. Our investigations of the SSDI benefit amount due variables and ensuing discussions with SSA staff

revealed that the SSDI benefit due variables are often populated even when the beneficiary is not in current pay status. Such situations occur because SSA computer systems often store placeholder values in the benefit due variables for future months and these placeholders are not always removed from the variables when a beneficiary changes from current pay status. Therefore, to accurately determine the amount due for an SSDI beneficiary, SSA staff advise first determining whether a beneficiary is in current pay status in the month in question using the LAF variable for the corresponding month and year. LAF values of "C" or "E" in the first position indicate the beneficiary is in current pay status.

C. SSDI benefit paid variables

Benefit paid variables reflect payments made in a given month and are conceptually distinct from benefit due variables. For example, a positive value for a benefit paid variable in a given month does not mean the beneficiary is entitled to benefits for that month; the amount in PAYDyymm may reflect a delayed entitlement for a prior month. To determine entitlement for a specific month, use the benefit due variables described in the previous section (MBAyymm, DUEDyymm, DUEOyymm). Unlike benefit due variables, benefit paid variables are only populated when the beneficiary received a payment. There are three variables for SSDI benefits paid:

- **MBPyymm—Federal SSDI Benefit Paid:** MBP represents the benefit amount payable to the beneficiary. The source variable is the SSA variable MBP, which is based on the variable MBC that is used to generate DUEDyymm (see above), with adjustments such as Medicare premiums and previous under or overpayments. It may differ from the actual check amount paid to the beneficiary (PAYDyymm); the latter may include under or overpayments.[9]

- **PAYDyymm—PHUS SSDI Benefit Paid:** PAYD reports the actual amount paid to the beneficiary in a given month. It is derived from two SSA variables: Direct Pay (DP—the actual amount of Social Security benefits disbursed for a specific month) plus Medicare Premium (MD – Part A and Part B premiums) from the Payment History Update System (PHUS).

- **PAYOyymm—PHUS SSDI Dependent Benefit Paid:** PAYO reports the actual amount paid to the beneficiary's dependents. It is similar to PAYDyym and is based on the same SSA source variables, DP and MD, from the PHUS. It is calculated by summing only the check amounts paid for qualifying dependents of the beneficiary. It excludes check amounts for the beneficiary, as those check amounts are available in the PAYDyymm variable for the individual in question. It includes dependents, such as DACs and DWBs.

---

[9] Researchers should note that, over time, MBPyymm is overwritten as retroactive adjustments are made.

D. SSI benefit due variables

Unlike the SSDI variables, the SSI benefit due variables are populated only when a payment is actually made, so there is no need to determine whether a beneficiary was in current pay status before using these variables. Benefit due variables for SSI beneficiaries include the following:[10]

- **FAMTyymm—Federal SSI Benefit Due.** FAMT reports the federal benefit amount for which the SSI beneficiary is eligible. FAMT can be retroactively adjusted to reflect under or overpayments as a result of changes in income or other factors such as living arrangement that affect benefit amount. The SSA source variable is FEDAMT (also called FAMT) from the SSR.

- **SAMTyymm—State SSI Benefit Due.** SAMT reports any federally-administered state SSI benefit the beneficiary is eligible to receive. It can be retroactively adjusted to reflect under or overpayments as a result of changes in income or other factors such as living arrangement that affect benefit amount. The SSA source variable is SUPAMT (also called SAMT) from the SSI – Longitudinal File (SSI-LF).

- **DUESyymm—SSI Benefit Due.** DUES (the sum of FAMT and SAMT) reports the total federal and federally administered state benefit amount for which the SSI beneficiary is eligible. It can be retroactively adjusted to reflect under or overpayments as a result of changes in income or other factors such as living arrangement that affect benefit amount. The SSA source variables are FEDAMT and SUPAMT (also called FAMT and SAMT) from the SSI-LF.

E. SSI benefit paid variables

The SSI benefit actually paid may differ from the amount due because retroactive changes may be applied to the amount due to reflect changes in income or other factors that affect the benefit amount, such as living arrangements or payment withheld to recover an overpayment. Similarly, the amount paid in a given month may differ from the amount due in that month, due to under or overpayments in prior months. There are three SSI benefit paid variables:

- **FPMTyymm—Federal SSI Benefit Paid:** FPMT reports the federal benefit amount paid to the SSI beneficiary. The SSA source variable is FEDPMT (also called FPMT) from the SSI-LF. Unlike FAMTyymm, FPMT is not retroactively adjusted.

- **SPMTyymm—State SSI Benefit Paid:** SPMT reports the federally-administered state SSI supplement paid to the SSI beneficiary. The source is the SSA variable STATPMT (also called SPMT) from the SSI-LF. Unlike SAMTyymm, it is not retroactively adjusted.

- **PAYSyymm—SSI Benefit Paid:** PAYS reports the total federal and state benefit amount paid to the SSI beneficiary. It is the sum of the SSA source variables FEDPMT and STATPMT (also called FPMT and SPMT) from the SSI-LF. Unlike DUESyymm, this variable is not retroactively adjusted.

---

[10] In DAF17, we identified and corrected an error related to the processing of benefits due and earnings variables for SSI beneficiaries derived from the SSR. This error had been present in all earlier versions of the DAF, meaning that statistics derived from these measures in DAF17 and onward will appear different to those derived from earlier versions of the DAF. A complete discussion of the error, including the affected variables, and implications for research is contained in Volume 7.

IV. Variables Related to Beneficiary Income

There are a variety of useful income variables in the DAF, though these data are generally less reliable than the official earnings data recorded by the IRS. For example, in the early years of the DAF, monthly income amounts of $201 occur frequently for SSDI beneficiaries because the Trial Work Period (TWP) limit was $200 and SSA staff did not distinguish specific levels of earnings so long as they were above that amount. Yet, for both SSI and SSDI, DAF earnings variables offer a significant advantage over the IRS data in that they provide monthly rather than annual data.

Researchers working with DAF earnings data should pay particular attention to the limitations of the data. For example, when SSDI beneficiaries have low levels of earnings, they are often not recorded or requested by SSA. In addition, SSDI earnings may not be present in the Disability Control File (DCF) variables described below when no benefit is due. These limitations are particularly evident in earlier years covered by the DAF. On the other hand, income information has long been collected from SSI beneficiaries because the SSI benefit amount varies according to income. For this reason, we begin by discussing SSI-related income variables.

A. Income information for SSI beneficiaries

The income received by SSI beneficiaries are generally self-reported and, where possible, independently verified by SSA staff. Accurate and timely information about SSI beneficiary income is important because the monthly SSI benefit paid is calculated as a function of income received, with different forms of income being accounted for in different ways, per program rules. As such, the administrative data capturing SSI beneficiary income—and hence, the DAF— contain variables both for the total amount of income received and the amount of income actually counted in calculating the monthly SSI benefit amount.

Income received by SSI beneficiaries is recorded in the month in which the earnings are received. For example, a beneficiary who works in May and is paid on June 5th will have June earnings recorded in the SSI data. Often the initial income report is for projected income and may be subsequently revised if the amount actually earned differs from the projection.

In addition to capturing countable and total income, the DAF contains two sets of SSI earnings variables. Researchers should be aware that neither set of income variables is of the same overall quality as the official data recorded by the IRS. However, the monthly level of the data combined with the fact that these variables comprise the actual basis for SSA adjudications of eligibility and payment may make them well suited for some analyses related to SSA program eligibility.

The first set of SSI earnings information comes from the CER100% Field Files, monthly snapshot files that are not revised as new information becomes available. Because SSI income variables are subject to frequent retrospective revision as new information is discovered, investigated, and verified, researchers should be aware that the Characteristics Extract Record 100% Field File (CER) source file does not reflect these revisions. Yet, for researchers interested

in what was known by SSA at the time these benefits were paid, these source variables are particularly useful and will align with data on benefits paid.

The second set of SSI earnings variables in the DAF come from the SSR. In DAF16, we changed the source of these variables from the DCF to the SSR. We did this because we had determined that the DCF was missing some earnings months and showed fewer workers with earnings than was shown in SSA published statistics. In particular, we found cases where the countable earnings in the month (EICMyymm), which we have always sourced from the SSR, were positive, but the gross earnings and related variables, which we sourced from the DCF, were 0 or missing. In discussions with the SSA Office of Systems, we determined that the data are in the SSR but that in some cases the batch process that creates these values in the DCF are not always copying them from the SSR to the DCF.[11]

In DAF18, we substantially revised the measures from those available in DAF16 and DAF17, though the variable names and labels remain the same. The current DAF reflects the approach we developed in DAF18, in which we revised our SSR-sourced earnings measures. We did so after identifying that earnings estimates in DAF16 and DAF17 were overstated as compared to SSA published statistics and determined that the issue was due to how we filled in monthly earnings based on SSR records. Specifically, the SSR shows the start and stop dates for earnings and we use that information to "forward populate" the starting earnings value. This value is automatically carried forward into the next month and all succeeding months until there is a manual change, at which point the new value is carried forward. Problems arise when there is no ending date on an earnings record. Beginning in DAF18, we curtailed the propagation of earnings in instances where other variables indicated that the value was no longer valid, for example, when a beneficiary is terminated or has died. More information is available in in Volume 7.

Table IV.1 identifies the SSI earnings variables from the CER100%, the SSR, and the DCF. We provide the SSR variables for users more familiar with using them, either directly from the SSR or from earlier versions of the DAF.

---

[11] Further, we believe this issue may have existed since the inception of the DCF in 2002, and so existed in the DAF15 and all prior versions of the DAF. Our analysis of the DAF15 suggested that the erroneous batch process was not transferring about 8% of the earnings and related information from the SSR to the DCF. Subsequent comparisons with SSA published statistics suggest that the DCF for 2015 was actually missing 30% of SSI earnings, and that the error has been growing over time. This latter finding suggests the error was likely smaller in earlier versions of the DCF and hence the DAF. Because the data in the DCF are incomplete, we developed a new process to source these gross earnings and related variables directly from the SSR starting in DAF16.

Table IV.1. SSI earnings measures from various SSA administrative data files, as named in the DAF

Earnings/Exclusion Description	DAF variable name		
	Source: CER100%	Source: DCF	Source: SSR
Earned income	EARNyymm		
Countable earned income			EICMyymm
Countable unearned income			UINCyymm
Wages/earnings	IEA_WAGEyymm	T16GRSAMTyymm	T16GRSAMTyymm
Self-employment	IEA_NESEyymm	T16NETAMTyymm[a]	T16SEAMTyymm
Net loss from self-employment	IEA_LOSSyymm		T16NETAMTyymm
Student earned income exclusion (SEIE)	IEA_SEIEyymm	T16EXLAMTyymm	T16EXLAMTyymm
Blind work expense (Blind IRWE)	IEA_BLINDyymm	T16EXPAMTyymm[b]	T16BEXPAMTyymm
Disabled workers impairment-related work expenses (IRWE)	IEA_IRWEyymm		T16EXPAMTyymm
Program to Achieve Self Support (PASS)	IEA_PASSyymm	T16PASAMTyymm	T16PASAMTyymm
Unearned income from SSDI benefit	IUA_SSDIyymm		
Unearned income from workers' compensation	IUA_WCyymm		
Unearned income from TANF	IUA_TANFyymm		

[a] The T16NETAMTyymm variable contained in the DCF was a composite measure of self-employment income (T16SEAMTyymm) and net loss from self-employment (T16NETAMTyymm); beneficiaries in the SSR may have one, though not both, of those values populated in a given month.

[b] The T16EXPAMTyymm variable contained in the DCF was a composite measure of Blind work expenses (T16BEXPAMTyymm) and non-blind impairment-related work expenses (IRWE—T16EXPAMTyymm); beneficiaries in the SSR may have one, though not both, of those values populated in a given month.

An indicator variable provides summary information about the earnings of SSI beneficiaries using the T16 variables available in the DAF. This measure was designed to be similar to the earnings indicator that was added in the previous version of DAF for SSDI beneficiaries. The SSI measure is called SSIERNLVLyymm and takes on the following values:

- 0=gross earnings are zero

- L=gross earnings are greater than 0 but less than the Federal Benefit Rate (FBR)

- F=gross earnings are at or above FBR but lower than the Substantial Gainful Activity (SGA) amount, and earnings less exclusions are less than FBR

- B=gross earnings are at or above FBR but lower than SGA, and earnings less exclusions are greater than or equal to FBR

- A=gross earnings are at or above SGA, but net earnings are less than FBR

- G=gross earnings are at or above SGA, and earnings less exclusions are greater than or equal to FBR but less than SGA

- S=earnings less exclusions are at or above SGA

The values for SSIERNLVL are determined by taking T16GRSAMT and subtracting the T16NETAMT to create total SSI Gross Earnings. We compare that value to the FBR and SGA thresholds for the year. We then sum the values of T16PASAMT, T16EXPAMT, T16EXLAMT to create total SSI Income Exclusions, subtracting that from SSI Gross Earnings to create earnings less exclusions. We then compare that second value to the FBR and SGA to complete the measure.

B. Income information for SSDI beneficiaries

Although earned income from working affect SSDI benefits, benefit amounts for SSDI are not subject to change as frequently as they are for SSI. Unlike SSI earnings, which are applied based in the month they are paid (regardless of when the work was performed), SSDI earnings have historically been applied based on the month in which they were earned. Using the earlier example, an SSDI beneficiary who works in May and is paid on June 5th will have May earnings recorded in the SSDI data. This changed in September 2016, as SSA enacts the provisions of the Bipartisan Budget Act of 2015.[12] Starting in September 2016, SSDI earnings will reflect a mix of when they were earned and when they were paid. Based on the change over time in the definition of the SSDI earnings data, our recommendation is to use the SSI earnings data (and associated timing of earnings) for beneficiaries receiving SSI only and concurrently receiving SSI and SSDI. Users should also assume that after September 23, 2016, SSDI earnings will reflect when they were paid, similar to the method used for SSI. Still, it is important to recognize that this will not be the case in all instances, meaning that annual earnings recorded in the DAF for SSDI beneficiaries will not necessarily match IRS earnings.

SSDI contains two sets of earnings variables, one that reflects the earnings received during SSDI entitlement, and another that is based on historical earnings prior to entitlement. The latter is necessary for determining eligibility, while the former is used by SSA in determining the monthly benefit amount.

Current monthly income information (during SSDI entitlement): SSDI income amounts are derived from the DCF and include:[13]

- T2GRSAMTyymm: SSDI Earnings Gross Amount (DCF)

- T2NETAMTyymm: SSDI Self-Employment Net Income Amount (DCF)

- ALLGAMTyymm: Alleged Earnings Amount (DCF—captured both for SSDI and SSI beneficiaries)

---

[12] Section 825 of the Bipartisan Budget Act of 2015 (PL 114-74) that was effective September 23, 2016 requires that when determining if an individual is engaged in substantial gainful activity to determine SSDI eligibility, SSA considers earnings to have been earned: (1) for the purposes of initial eligibility in the month in which such services were performed and (2) in all other cases, in the month such earnings were paid. Unless SSA can establish based on readily-available evidence the actual month such earnings were earned or the person (after having been denied benefits due to excess earnings) can demonstrate to SSA that such earnings were actually earned in a different month, the month they were paid will be used.

[13] The issue noted above in which some earnings from the SSR were not recorded in the DCF does not affect the DCF data for SSDI beneficiaries. The DCF remains the best SSA administrative source of earnings data for SSDI beneficiaries.

Work-related expenses and income exclusions are reported in:

- T2CDNAMTyymm: SSDI Special Condition Amount (DCF)

- T2EXPAMTyymm: SSDI Work Expense Amount (DCF)

- T2SBDYAMTyymm: SSDI Earnings Subsidy Amount (DCF)

The earnings variable contained on the DCF resulted from the advent of Ticket to Work (TTW), which required that SSA track SSDI beneficiaries' current income more closely in order to make payments to providers. These new variables are available in the DAF as early as January 2003. Some SSA field offices began using these new systems later than others, leading to inconsistencies in the early months of data availability. When possible, the amounts are independently verified by SSA.

Earnings recorded for SSDI beneficiaries may not always be reliable, particularly at lower levels of earnings, which do not affect benefit eligibility. They are reliable at higher earnings levels, after allowing for an administrative processing lag. Moreover, while they are not always consistent with IRS annual earnings data, they are monthly and therefore can be more suitable for analyses of SSA programs for which eligibility is calculated on a monthly basis.

Beginning in DAF15, we added a measure of SSDI earnings that takes into account the variables above to create a single summary measure of earnings relative to the TWP and SGA levels in each month. This new measure is called DIERNLVLyymm (SSDI Earning Level Indicator) and is described in Volume 5. To create this variable, we summed T2GRSAMT and T2NETAMT to create SSDI Gross Earnings. We then subtract the sum of T2CNDAMT, T2EXPAMT and T2BDYAMT from the SSDI Gross Earnings to create SSDI Countable Earnings. Using those values, we generated DIERNLVLyymm by comparing Gross and Countable earnings to the TWP and SGA limits in place during each month. The values of the newly constructed variable are:

- 0=gross earnings are zero

- L=gross earnings are greater than zero, but less than the TWP limit

- T=gross earnings are at or above the TWP limit, but less than SGA

- G=gross earnings are at or above SGA, but countable earnings are less than SGA

- S=countable earnings are at or above SGA

Earnings history information (prior to SSDI entitlement): To qualify for SSDI benefits, a worker must have had a work history with earnings above a certain level for a minimum number of quarters. This earnings history is then indexed to inflation and averaged across the months covered by the period of the earnings history; a calculation captured in the Indexed Monthly Earnings (IMEn) variable. The PIEDn variable indicates the effective date represented by the IMEn variable. The highest thirty-five years of IME, in turn, are used by SSA to calculate a Primary Insurance Amount (PIAn). The monthly benefit amount due derives from and closely tracks the PIA.

According to SSA program rules, IME are based on earnings that are averaged over the highest 35 years of earnings. As individuals work additional years, they may accumulate additional occurrences of IMEn and PIAn. Including the years from age 18 until FRA, and considering the highest 35 years of earnings, each of these variables may have up to 50 occurrences. All of these variables are ascending such that the beneficiary's earliest PIA is PIA1, their next PIA is PIA2, etc.

Only the 50 most recent occurrences of the variable are retained. More than 99 percent of beneficiaries have 38 or fewer occurrences. However, less than 1 percent of beneficiaries have more than 50 occurrences of these variables. For beneficiaries with more than 50 occurrences, the first occurrence (i.e., IME1, PIA1, PIED1, etc.) corresponds to the total number of occurrences minus 49. For example, the first occurrence for a beneficiary with 73 occurrences will correspond to the 24th occurrence overall and ascend from there.

V. Impairment Codes

Impairment information for beneficiaries in SSDI and SSI is recorded by SSA only at select times—initially at disability award and updated later during periodic medical CDRs, generally scheduled to occur every three or seven years.[14] Primary and secondary impairment codes reflect the evidence used by SSA to determine whether a person is disabled at that time. However, SSA staff generally record only sufficient impairment information to justify disability benefits so if a primary impairment is sufficient to qualify a beneficiary for benefits, a secondary impairment may not be recorded.

When stored in the DAF, these periodic impairment codes are converted to monthly occurrences. The preferred variables to use are DXPRIBESTyymm and DXSECBESTyymm in the DAF Annual files; these incorporate information across multiple sources to develop the most up-to-date diagnostic information in a particular month. The impairment code is listed in each "yymm" occurrence in which the diagnosis was in effect. For example, for a beneficiary who became entitled to benefits in January 2002 with a disability due to a musculoskeletal disorder, the variable DXPRIBEST0201 would indicate the relevant code for the musculoskeletal disorder. If the diagnosis was revised during a CDR in January 2005 to a psychiatric disorder, the primary impairment code would be updated in the SSA administrative data, and the DAF variable DXPRIBEST0501 would contain the revised impairment code, while the intervening DAF variables—from DXPRIBEST0202 to DXPRIBEST0412—would contain the initial impairment code. For the months after January 2005, the subsequent variables, DXPRIBEST0502 and onwards, would reflect the revised impairment code, up to the time of the next Continuing Disability Review (CDR) or the month when the beneficiary's entitlement ceased. Converting the periodic codes to monthly occurrences permits a beneficiary's most recent impairments to be identified at any time while they are entitled to benefits.

A. Coding schemes and categorization

SSA records impairment information using four-digit SSA Impairment Codes, a system of diagnosis codes that has been in use since the mid-1980s. Prior to the advent of the SSA impairment codes currently in use, SSA used the International Classification of Diseases Coding Scheme (ICD-9) coding scheme. In some cases, the ICD-9 codes are still present in the DAF data because medical data has not been updated, particularly for beneficiaries who have been receiving benefits for a long time. There is not currently a single standard at SSA for grouping the diagnosis codes, however, SSA and Mathematica researchers have jointly developed a grouping scheme suitable for research projects that accommodates both the SSA Impairment Codes as well as the ICD-9 codes. This categorization is presented in Table V.1. An alternative grouping is shown below in Table V.2. This is the grouping that SSA uses in the SSI Annual Statistical Report. After reading this section, users interested using either of these categorization approaches should consult the DAF code library, available at https://www.ssa.gov/disabilityresearch/daf.html#library.

---

[14] The term "CDR" also has meaning in the context of a review to determine how an SSDI beneficiary's work affects their benefits. Such a review is called a work CDR as opposed to the medical CDRs under discussion here.

Table V.1. Suggested impairment code groupings and labels

Group number*	Group label	Impairment codes
1	Major affective disorders	2960-2969
		3110-3119
2	Schizophrenia and psychoses	2950-2959
		2980-2989
3	Anxiety and neurotic disorders	3000-3019
		3080-3099
4	Other mental disorders	2900-2949
		2990-2999
		3030-3079
		3100-3109
		3120-3129
		3138-3169
		3195 only
5	Intellectual disability	3170-3194
		3196-3199
6	Back disorders	7221-7249
7	Musculoskeletal system	7100-7200
		7250-7399
8	Infectious & parasitic diseases	0110-0119
		0450-0459
		0930-1359
		1380-1389
9	HIV/AIDS	0070-0079
		0201-0449
		0540-0559
		0780-0789
		1360-1369
10	Neoplasms	1400-2399
11	Endocrine/ nutritional	2400-2479
		2500-2559
		2630-2799
12	Blood/ blood-forming diseases	2800-2899
13	Severe visual impairment	3610-3699
		3780-3789
14	Severe hearing impairment	3890-3899
15	Severe speech impairment	7840-7849
16	Nervous system	3200-3419
		3430-3599
		3860-3889
17	Circulatory system	3420-3429
		3750-3759
		3900-4599

Table V.1. Suggested impairment code groupings and labels

Group number*	Group label	Impairment codes
18	Respiratory system	4600-4869
		4910-5199
		7690-7699
19	Digestive system	5200-5799
20	Genitourinary system	5800-6299
21	Skin/ subcutaneous tissue	6900-7099
22	Congenital anomalies	7400-7599
23	Injuries	8000-9599

* The assignment of group number is arbitrary and serves as a suggestion for coding primary diagnosis variables.

Table V.2. SSI annual statistical supplement categories

Group number*	Diagnostic category	Impairment codes
1	Congenital anomalies	7400-7599
2	Endocrine, nutritional, and metabolic diseases	2400-2469
		2500-2539
		2550-2559
		2600-2779
3	Infectious and parasitic diseases	0020-0189
		0200-0279
		0300-0419
		0430-0579
		0600-0669
		0700-0889
		0900-1049
		1100-1189
		1200-1359
		1370-1399
		7710-7719
4	Injuries	8000-8489
		8500-8549
		8600-8879
		8900-8979
		9000-9059
		9070-9099
		9200-9299
		9400-9599
5	Autism spectrum disorders	2990-2999
6	Developmental disorders	3150-3159
7	Childhood and adolescent disorders not elsewhere classified	3120-3139
8	Intellectual disorders	3170-3194
		3196-3199

Table V.2. SSI annual statistical supplement categories

Group number*	Diagnostic category	Impairment codes
9	Depressive, bipolar, and related disorders	2960-2969
		3110-3119
10	Neurocognitive disorders	2900-2909
		2940-2949
		3100-3109
11	Schizophrenia spectrum and other psychotic disorders	2950-2959
		2970-2989
12	Other mental disorders	2910-2939
		3000-3099
		3140-3149
		3160-3169
		3195
13	Neoplasms	0420-0429
		1400-1659
		1700-1769
		1780-2089
		2100-2399
14	Diseases of the blood and blood-forming organs	2800-2899
		7720-7739
		7760-7769
15	Diseases of the circulatory system	3750-3759
		3900-3989
		4010-4059
		4100-4179
		4200-4389
		4400-4449
		4460-4489
		4510-4599
16	Diseases of the digestive system	5200-5379
		5400-5439
		5500-5539
		5550-5589
		5600-5609
		5620-5629
		5640-5799
		7770-7779
17	Diseases of the genitourinary system	5800-6089
		6100-6119
		6140-6299
18	Diseases of the musculoskeletal system and connective tissue	7100-7399
19	Diseases of the nervous system and sense organs	3200-3269
		3290-3379
		3400-3749
		3760-3899

Table V.2. SSI annual statistical supplement categories

Group number*	Diagnostic category	Impairment codes
20	Diseases of the respiratory system	4600-4669
		4700-4789
		4800-4879
		4900-4969
		5000-5089
		5100-5199
		7680-7709
21	Diseases of the skin and subcutaneous tissue	6800-6869
		6900-6989
		7000-7099
		7780-7789
22	Other	7600-7609
		7640-7669
		7800-7809
		7830-7849
		9330
23	Unknown	0000-0019
		0190-0199
		0280-0299
		0580-0599
		0670-0699
		0890-0899
		1050-1099
		1190-1199
		1360-1369
		1660-1699
		1770-1779
		2090-2099
		2470-2499
		2540-2549
		2560-2599
		3270-3289
		3380-3399
		3990-4009
		4060-4099
		4180-4199
		4390-4399
		4450-4459
		4490-4509
		4670-4699
		4790-4799
		4880-4899
		4970-4999
		5090-5099
		5380-5399
		5440-5499

Table V.2. SSI annual statistical supplement categories

Group number*	Diagnostic category	Impairment codes
23 (*continued*)	Unknown	5540-5549
		5590-5599
		5610-5619
		5630-5639
		6090-6099
		6120-6139
		6300-6799
		6870-6899
		6990-6999
		7610-7639
		7670-7679
		7790-7799
		7810-7829
		7850-7999
		8490-8499
		8550-8599
		8880-8899
		8980-8999
		9060-9069
		9100-9199
		9300-9329
		9331-9399
		9600-9999

Note: Starting in DAF20, attention deficit/hyperactivity disorder (ADHD) diagnoses have been reallocated from the "childhood and adolescent disorders not elsewhere classified" group to the "other mental disorders" group to reflect SSA's adoption of an ADHD diagnostic code for both children and adults.

* The assignment of group number is arbitrary and serves as a suggestion for coding primary diagnosis variables.

B. How to determine statutory blindness within the limitations of the administrative data

SSA's definition of statutory blindness is central visual acuity of 20/200 or less in the better eye with best correction, or a limitation in the field of vision in the better eye so that the widest diameter of the visual field subtends an angle of 20 degrees or less. Beneficiaries who meet this definition are determined to be statutorily blind and are accorded some special provisions, including employment supports and a higher threshold for considering whether earnings are above the level of SGA.[15] The best way to identify beneficiaries who are subject to the provisions of statutory blindness is to use the BLINDDT variable. A valid date in BLINDDT indicates that a determination of statutory blindness was made at one time. If BLINDDT is blank, no onset of blindness has been recorded for the beneficiary. If it contains a valid date, compare it to the reference date, if appropriate. For example, if you need to know if a beneficiary was blind when they entered the SSDI rolls, and the initial date of program participation was June 2005, determine if the date in BLINDDT is before June 2005. In addition to BLINDDT, for

---

[15] People who have a visual impairment that does not meet the definition of statutory blindness may still be eligible for disability benefits based on other criteria.

SSI beneficiaries, the presence of data in IEA_BLINDyymm or a variable value beginning with the letter "B" in COMP_STA or TOA (Type of Action) also means that the beneficiary was, at some point, determined to be statutorily blind.

It is important to note that BLINDDT does not identify all beneficiaries who have blindness or low vision. Any diagnosis code in DXPRIBESTyymm or DXSECBESTyymm that begins with "36" indicates a visual impairment, including "3690" which indicates blindness or low vision but not necessarily *statutory* blindness. The impairment code "3694" was introduced in 2012 to indicate statutory blindness although the overlap of this diagnosis code with BLINDDT is not one-to-one (Table V.3). Approximately 87 percent of beneficiaries with a date of statutory blindness onset (BLINDDT) in 2015 through 2020 had "3694" as either a primary or secondary diagnosis. Of those with a primary or secondary diagnosis code of 3694, 83 percent have a populated blind date.

Table V.3. Percentage of statutorily blind beneficiaries who have the statutory blindness diagnosis code ("3694") as primary or secondary diagnosis, by year of statutory blindness onset

Year of statutory blindness onset (using BLINDDT)	Percentage of beneficiaries with "3694" as primary diagnosis	Percentage of beneficiaries with "3694" as secondary diagnosis	Percentage of beneficiaries with "3694" as primary or secondary diagnosis
2012	29.7	1.2	30.8
2013	55.3	3.1	58.2
2014	75.0	3.4	78.3
2015	83.5	2.8	86.2
2016	84.9	2.3	87.1
2017	86.0	2.6	88.6
2018	86.2	2.9	89.0
2019	86.6	3.0	89.5
2020	80.3	2.7	82.9

Source: DAF20 10% DMG.

It is not unexpected that some individuals with a populated value for BLINDDT do not have a corresponding impairment code for blindness in the primary or secondary impairment fields. This may reflect the fact that many beneficiaries who are blind have a number of other significant disabilities that are sufficient to qualify the beneficiary for benefits, and thus these other disabilities are recorded in the impairment code fields while blindness is indicated by entering a date in the BLINDDT field. Another factor is that a beneficiary may become blind after their initial qualification for disability benefits, for instance, from the progression of diabetes. This would likely be reflected in BLINDDT, but might not be recorded in the impairment codes if it is not the primary or secondary reason for benefit eligibility. In addition, it is possible for a finding of statutory blindness to be revised or revoked, even if a beneficiary continues to be disabled as a result of blindness or low vision.

VI. Identifying SSI Couples

When two individuals of the opposite sex who are married or holding themselves out to the community as married, or when same-sex couples are married, and both receive SSI and live together, they are considered one unit for benefit purposes.[16] As a couple, the thresholds for financial eligibility will be higher, and the income and resources of both members of the couple affect the cash benefits to which they are both entitled. This can mean that an individual who is working at a level that would make him ineligible for SSI as an individual may be eligible to receive SSI as a couple.[17]

Identifying SSI couples in the DAF is challenging, and not always possible with certainty, in part because being a member of an SSI couple is based on residence and so a beneficiary may go back and forth between individual and couple status. The SSR variable for household composition, CUR-COMP, known as CURCOMP on the DAF, can be used to identify couple status. However, our initial investigations revealed that CURCOMP only identifies SSI couples approximately 60 percent of the time. Specifically, 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, CURCOMP does not successfully identify all SSI couples.[18] In a small number of cases, CURCOMP also erroneously identifies beneficiaries as members of SSI couples.[19]

---

[16] The "holding out" concept is similar but not identical to common-law marriage. See POMS SI 00502.145 for a full discussion. As of 2015, "same-sex couples will be recognized as married for purposes of determining entitlement to Social Security benefits or eligibility for SSI payments" (https://www.ssa.gov/people/same-sexcouples/).

[17] For example, the SSI FBR (or maximum monthly federal payment) in 2021 was $914 for individuals and $1,371 for couples. An individual with between $914 and $1,371 in countable (i.e., after all exclusions and income disregards) income, from working or from some other source, is ineligible for SSI so long as he or she is not a member of a couple. However, if that individual is married to and living with another SSI beneficiary, he or she becomes eligible in that month so long as his or her spouse's countable income is not sufficient to add up to more than $1,371.

[18] 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.

[19] These erroneous identifications include the one or two months at the beginning or end of a period of SSI couple status when a new SSI record is being established and when one member of the couple is incarcerated. Because there is no valid external benchmark, it is impossible to be certain how frequent these erroneous identifications occur. Our investigations suggest that they are quite infrequent.

VII. Identifying the Expectation of Medical Improvement

The DAF contains two measures related to the expectation of medical improvement: MEDXyymm (previously named MEDEXyymm, and before that, MIEXyymm), which indicates the medical CDR diary type in effect in a given month and year, and FIRSTMIE (previously named FRST_MIE), which indicates the date in which a beneficiary was first in the MIE category.

The MEDXyymm variable, as detailed in Volume 5, contains the following categories:

- P – Medical improvement is possible (MIP)

- N – Medical improvement not expected (MINE)

- E – Medical improvement expected (MIE)

- 0 – No medical improvement information

Monthly measures of medical improvement among beneficiaries incorporate data from the DCF in May 2008 onward; prior to that time, the data is sourced from the 831 & 832/833 files.[20,21] After the completion of DAF18, the DAF team was alerted to the fact that the MEDXyymm (called MEDEXyymm in DAF18) values had a higher than expected proportion of missing values. Upon review, we determined that a coding error in combining information from the DCF and 831 & 832/833 was the cause. In the vast majority of cases, after the code correction was implemented, populated values remained unchanged; the main effect was replacing missing data with populated values. The DAF18 was subsequently corrected and all versions since the DAF18 are correct, but the error was never corrected for DAF16 and DAF17. As a result, prior analyses that used the DAF16 or DAF17 and then revisited those analyses using more recent DAF data will show fewer cases with missing MEDXyymm (called MEDEXyymm in DAF18) information. More information about the nature of the error and the effect of the revision is available in Volume 7, "Correction After the DAF18 in the Variable Indicating the Expectation of Medical Improvement."

---

[20] Prior to DAF15, the expectation of medical improvement was drawn from the 831 & 832/833 files, which have rates of missing information for medical CDR diaries of about 20%. In 2016, we learned that more complete medical CDR diary information is available from the Medical table in the DCF from 2008 forward, so we are combining data drawn from the 831 & 832/833 files and the DCF Medical tables to provide the most complete data available for medical CDR diaries.

[21] In April 2008, SSA converted the historical data from one system to another (from VSAM to DB2) to populate the DCF Medical table. During that conversion, some historical dates were overlain making it impossible to determine the timing of the determinations.

VIII. Constructed Measures Related to Suspension or Termination for Work and Dollar Value of Cash Benefits Forgone Due to Work

Because the relationship between work and benefit receipt is a subject of intense interest, we have constructed two families of analytic variables to aid researchers. These variables consist of monthly indicators for cash benefits in suspense or termination for work (STW) and a constructed monthly measure of the dollar value of cash benefits forgone for work (BFW). The core STW and BFW measures are included on the Annuals files and described in detail in Sections VIII.A through VIII.C below and the relevant variable detail pages in Volume 5.

The algorithm for developing these measures has been refined over time, so we recommend using the STW variables from the most recent DAF file available. The most significant refinement occurred in DAF14 when we updated the algorithm for STWSSI and BFWSSI to better account for deemed income and the retrospective monthly accounting (RMA) process that SSA uses to calculate SSI payments. In turn, revisions to STWSSI affect STWCM and revisions to BFWSSI affect BFWCM, as described in what follows. These measures were initially included on the DAF in a standalone SSI Companion Work File (SCWF) but as of DAF18 have been fully incorporated into the core DAF files.

After reading this section, users interested in identifying beneficiaries in STW during a certain period of time may want to consult the DAF code library, available at https://www.ssa.gov/disabilityresearch/daf.html#library, which provides sample code for that task.

A. Overview of STW and BFW measures

The DAF contains three variables that provide monthly information on the non-payment of cash benefits following suspense or termination for work, each with names that start with "STW." These measures are described in Section VIII.B. There is one variable for SSI beneficiaries, one for SSDI beneficiaries, and also a combined variable that takes concurrent beneficiaries into account.[22] In many instances, the data in a given month are not definitive about whether a beneficiary whose benefits were suspended or terminated for work are, in fact, continuing to work at a level that would make them ineligible for benefits in the current month. Instead, they only tell us that benefits were suspended or terminated in the past because of work and that benefits have not been reinstated as of the current month.[23]

The second set of variables indicates the dollar value of BFW. These are described in Section VIII.C. As with STW, there are three such variables, one each for SSI and SSDI and also a combined variable for concurrent beneficiaries. Conceptually, the value of a BFW variable for a month represents the amount of benefits that the individual has forgone in the month because he or she worked at a level that caused the suspension, reduction, or termination of benefits. For

---

[22] In the Ticket Research File (TRF) 10, now known as 'DAF', and earlier versions of the database, these began with "LDW" for "left due to work."

[23] These measures cease to be populated when a beneficiary dies, turns 65 (SSI recipients), or reaches FRA (SSDI beneficiaries).

SSDI beneficiaries, the STWDI value must indicate suspense or termination for work in order for the BFWDI value to be populated, reflecting the SSDI program rules. SSI beneficiaries, however, may forgo benefits due to work even when they are still receiving a cash benefit, as described below. Hence, BFWSSI may be populated even if the beneficiary is not suspended or terminated for work per STWSSI. In certain instances, BFW values for SSI beneficiaries are "special missing" characters, described below.

Determination of the value of BFW requires counterfactual information that can only be estimated: how much the individual's benefits would have been in the current month if benefits had not been suspended, reduced, or terminated for work. For BFW, estimates are based on program rules and information on benefit amounts prior to suspense or termination for work. For SSDI, these rules are such that the estimates are likely exact in a large majority of cases, and very close in the remainder. The SSI rules and data are more complex, making it more difficult to exactly estimate BFW. Yet, our development of the algorithm over many years and inspection of records leads us to believe the current version is a close approximation to the true value.

SSA and Mathematica researchers collaboratively developed the complex algorithms required to create the STW and BFW variables. The inputs to these algorithms are DAF variables that describe program status, benefit payment, income, and other indicators of work. These elements are not complete in all records. In cases where one or more of these variables is absent, the algorithms use second-best alternatives. As a result, these variables are sometimes inexact. STW accurately reflects whether or not the beneficiary is in current pay status (that is, due a cash benefit) according to the latest information in the administrative records, but determination of the exact reason for benefit suspense or termination requires examination of several administrative data elements and is not always definitive.

B. STW indicators

In both the SSDI and SSI programs, a suspension or termination status implies a $0 cash benefit for that month. In the SSDI program, a suspension or termination for work requires the beneficiary to have engaged in SGA after completing the 9-month TWP and three grace period months. For SSI, however, an indication of suspension or termination for work does not indicate any specific minimum level of work or earnings. SSI suspensions and terminations are based on all sources of beneficiary income, including earnings from work, SSDI benefits (for concurrent beneficiaries), deemed spousal income, income from any other source, and in-kind support and maintenance as monetized in SSI rules. Minimal earnings from work may therefore be enough to push a beneficiary into suspense or terminated status. Conversely, an SSI beneficiary may be suspended for excess income as a result of other unearned income alone, so the beneficiary is not considered suspended for work, regardless of the amount of beneficiary earnings.

We developed three longitudinal indicator variables of non-payment status due to suspense or termination for work, included in each Annual file and populated on a monthly basis. The STWDI variables indicates how work has affected the status of SSDI benefits, the STWSSI variable does the same for SSI benefits. The values for the STWDI and STWSSI variables show

the beneficiary's work status in the specified program in any given month but have somewhat different sets of values. The STWCM variable indicates how work has affected the combination of SSDI and SSI benefits. Each variable is named according to the year and month using the same *yymm* convention as elsewhere in the DAF (for example, STWDI1901, STWDI1902, and STWDI1903 are the STWDI values for the first three months of 2019).

1. *The values of the STWDI indicator*

The values for STWDI are as follows, all referring to SSDI status:

- 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[24]
- 9 = in terminated status in this month for a reason not determined to be work
- . (missing value) = beneficiary is not yet entitled to benefits, has died, or has attained FRA

2. *The values of the STWSSI indicator*

There is one more possible value for the STWSSI variable than STWDI due to the greater complexity in determining suspension or termination for work in the SSI program. Because SSA administrative data do not distinguish between suspensions and terminations as a result of excess income from work versus excess income from other sources, STWSSI can take on a value of 4. For SSI beneficiaries with a suspension due to excess income, an STWSSI value of 1 indicates that the beneficiary was working during the months of suspension and that the earnings from work were necessary for suspension (even if they were not sufficient to induce suspension in the absence of unearned income). In contrast, an STWSSI value of 4 indicates that the beneficiary was working, but that the beneficiary would have been suspended for excess income in that month even without that earned income (that is, as a result of the unearned income alone).

The values for STWSSI are as follows:

- 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 age 65

---

[24] For previous users of STWDI, note that the value of 8 was first introduced in the DAF12.

- 4 = suspended due to excess income in this month, part of that income is earnings from work, but total countable nonearned 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

- . (missing value) = beneficiary is not yet entitled to benefits, has died, or has attained age 65

When developing STWSSI, we first identify beneficiaries in suspense due to excess income (PSTA=N01). We then compare total countable nonearned income (TCNEI) to the FBR; TCNEI is the sum of UINC, deemed income, and income support and maintenance (ISM). If TCNEI is above the FBR, then we determine an SSI beneficiary is STWSSI=4 because their unearned income was sufficient to preclude a benefit payment. If TCNEI is less than the FBR, then the earned income was necessary to result in benefits being suspended, in which case STWSSI=1.

In DAF18, we have updated the STWSSI calculation in the core Annuals files relative to recent versions. With this change, we also dropped the _DRAFT suffix on the STWSSI measure that we have used since its inception, in recognition that further major changes to the measure are unlikely. The values the variable can take have not changed, but the revised version of STWSSI more accurately uses TCNEI in the construction of STWSSI, as described above. Using TCNEI improved the STWSSI calculation to more fully account for deemed income using a variable derived from the SSI Longitudinal File (SSI-LF), FCIyymm. This variable more completely accounts for all non-earned sources of income, including deemed income and in-kind support and maintenance (ISM).[25] Before we had access to FCI, we calculated STWSSI using only EICM (countable earned income) and UINC (countable unearned income). In most cases, EICM+UINC is the countable income amount used for computation of the monthly benefit. A known limitation of that sum, however, was that it excluded deemed income.

Using FCI to calculate TCNEI affects the STWSSI calculation. Up until DAF18, when PSTA=N01, we determined whether STWSSI was 1, 4, or 8 by first identifying whether there was countable earned income (EICM) in the current month. If there was not (i.e., EICM=0), then we set STWSSI=8, meaning in suspense for reasons not determined to be work. If there was countable earned income in that month (EICM>0), we next compared the amount of the countable unearned income (UINC) in the month to the FBR. If the UINC was greater than the FBR, we set STWSSI=4 because the unearned income alone was sufficient to produce a suspension for excess income. If the UINC<=FBR, STWSSI=1 because EICM>0 was necessary to put the total countable income for that month over the FBR. In the current version of STWSSI, we now make the same comparisons, but use TCNEI instead of UINC.

Deemed income is an important consideration for calculating STWSSI, though in most months the STWSSI value remains unchanged (see Appendix A for statistics from the DAF18). When we first developed the SCWF in DAF14, we considered the share of beneficiaries in PSTA=N01

---

[25] We refer to this as deemed income in the remainder of what follows for simplicity's sake, but our measure cannot distinguish between the two forms of income.

status in each month of 2012 (preceding the end of DAF14 enough to minimize the effects of data processing lags). Among beneficiaries in PSTA=N01, we found that approximately two out of three had some amount of deemed income (for example, 62 percent of those in N01 status in January 2012 had deemed income based on TCNEI> UINC). Among beneficiaries who were originally STWSSI=1 in months during 2012, more than nine in ten had deemed income, meaning that they were potential candidates for moving from STWSSI=1 to some other status (STWSSI_FCI=4 or 8) when we incorporated FCI. For example, in January 2012, 94 percent of those in STWSSI=1 had TCNEI>UINC.

In the last several versions of the DAF, the version of STWSSI now on the Annuals files was housed on the standalone SSI Companion Work File (SCWF). We opted to maintain two STWSSI measures to allow time for user feedback. Having not received any, we have updated the measures on the core files and no longer create the SCWF.

3.   *The combined STW indicator*

The combined STW indicator was developed by taking into account the STW status for both the SSI and SSDI programs. This variable was designed to support the evaluation of TTW, but it is useful in most other examinations of work and benefits as well. Under TTW, if a qualified service provider, called an Employment Network (EN), has accepted assignment of a Ticket from the beneficiary, the EN is entitled to an outcome payment if and only if the beneficiary is not receiving benefits from either SSDI or SSI. STWCM was designed to reflect EN eligibility for an outcome payment.

The STWDI and STWSSI variables alone do not provide complete information for beneficiaries in both programs. The combined indicator is useful in any context where a DAF user wants to know beneficiaries' status across both programs. If a beneficiary only ever receives benefits from SSDI or SSI, the STWCM indicator may also be used, reflecting their STW status in the relevant program. But, for beneficiaries receiving benefits from SSDI and SSI in the same month, or for beneficiaries who might have previously received benefits from one or both programs but who now only receive benefits from one program or the other, the STWCM measure can be incredibly useful.

Construction of STWCM is based on the separate SSI and SSDI STW indicators and <u>errs toward current pay status in cases where the two indicators are not the same</u>. For example, if the separate indicators show suspense or termination for work in one program (1, 2 or 3), but current pay status (0) in the other program, STWCM indicates the beneficiary is "in current pay status." More generally, <u>the STWCM indicator takes on the value of the lower of the two values for STWDI and STWSSI</u>; for example, if STWDI=0 and STWSSI=1, then STWCM=0, and if STWDI=3 and STWSSI=2, then STWCM=2. It is important to note that the STWCM value may therefore not comprehensively reflect benefits in both programs for all research cases. For example, a beneficiary might be working to a level that puts them in suspense for SSI, so that STWSSI=1, but remain in STWDI=0. In that case, STWCM would be 0. Or, a beneficiary may have been terminated from both programs for work at some point in the past, so that STWSSI

and STWDI were 3 for some period. If that person returned to benefits for STWDI, the STWCM value would be 0 and would potentially mask a former termination from SSI for work. Depending on the analysis, the separate indicators may be more beneficial for cases like these.

The only exception to this rule is when STWSSI=4, which is treated as STWSSI=8 in setting the value of the corresponding STWCM variable. For example, if STWSSI=4 and STWDI=. (missing), 8, or 9, then STWCM=8. This exception was instituted because an STWSSI value of 4 does not indicate a loss of benefits specifically due to income from work. Rather, the value of 4 indicates that the beneficiary is working but would be in suspense for income status even if not working.

Researchers should choose the STW indicator that is appropriate for their purposes taking into account how all three STW indicators are constructed and work together, and how SSDI and SSI interact.

The STWCM values are as follows:

- 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, suspended, or terminated for work in this month for the other 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 FRA.
- 8 = in suspense status for reasons not determined to be work for either SSDI or SSI and not in current pay, suspended, or terminated 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, suspended, or terminated for work in this month for the other program.
- . (missing value) = beneficiary is not yet entitled to benefits, has died, or has reached FRA

The change to STWSSI in DAF18 mentioned in the previous section also carries over into changes in STWCM, by construction.

4. *Presumption of work after benefit termination*

The STW measures take on a value of 3 in the months after SSA terminates benefits due to work, provided the beneficiary has not regained eligibility, reached FRA (SSDI) or age 65 (SSI), or is known by SSA to be deceased. In these months, it is not possible to determine from the administrative data whether the former beneficiary is actually earning at a level that would preclude eligibility for benefits. Thus, an STW value of 3 indicates that work is the reason that benefits were terminated prior to that month but does not imply that earnings during that month make the beneficiary ineligible for benefits during that month. For example, STWDI0606=3

means that benefits were terminated for work prior to June 2006 and the beneficiary had not died or reached FRA and had not returned to receiving cash benefits as of June 2006. But, it does not mean that the beneficiary is necessarily engaged in SGA or worked at any level in June 2006—he or she may be, but it is not possible to determine that from the administrative data. SSA has no reason to collect beneficiary work information after benefits have been terminated unless the beneficiary applies for reinstatement.

Researchers should not misinterpret a value of 3 as indicating that the former beneficiary is necessarily earning enough in that month to be ineligible for benefits. In many cases, we believe it is reasonable to presume continued work is the reason for not returning to benefits, though we have not verified this to be true in all cases. Even if the beneficiary is not engaged in SGA, a "3" STW value indicates that past work is the reason that SSA is not paying benefits to the individual in the current month.

5. *Auxiliary SSDI beneficiaries*

Special rules are applied to the DAF records of auxiliary Social Security beneficiaries with disabilities whose entitlement is based on their own medical eligibility and the earnings record of another person who is a primary Social Security beneficiary: DACs and DWBs.[26] For DWB, STW values that occur after the primary beneficiary is deceased apply to work of the DWB, not the primary beneficiary. For DAC, in many cases the primary beneficiary is retired or deceased, but in some cases the primary beneficiary is an SSDI worker beneficiary. In such cases, any evidence of suspension or termination on the record of a DAC might actually belong to the primary beneficiary, because a suspension or termination on the record of the primary beneficiary applies equally to auxiliary beneficiaries *even if* the auxiliary beneficiaries are not working. Hence, if a DAC record is flagged as in suspense or termination following the above methods, further checks are done to ascertain whether the suspension or termination is due to the earnings of the auxiliary beneficiary.

SSDI records contain two identifying variables: 1) the Claim Account Number (CAN) (the Social Security Number, SSN, of the primary beneficiary) and 2) the BOAN (the beneficiary's own account number). For primary beneficiaries, the CAN is the same as the BOAN. For DAC beneficiaries, the CAN is the SSN of the primary beneficiary (i.e., the beneficiary's parent), and the BOAN is the DAC's own SSN. When constructing STW indicators, we therefore take note of the CAN when a primary beneficiary is flagged as suspended or terminated due to work. Later, if a DAC record is flagged as STW, we check to see if its CAN variable is the same as the CAN of a flagged primary beneficiary. If so, we assume that the suspension/termination information belongs to the primary beneficiary and not to the DAC, and the STWDI value for the DAC is set to 8 or 9 for suspension or termination, respectively.

---

[26] Approximately 10.7% and 2.6% of SSDI beneficiaries were entitled under DAC and DWB provisions, respectively, in December 2016 according to Chart 1 in the 2016 Annual Statistical Report on the SSDI Program.

6. SSI couples

STWSSI, and thus also STWCM, algorithms attempt to account for the difference in program rules associated with members of SSI couples using the CURCOMP variable described in Chapter VI above. An SSI beneficiary is a member of an SSI couple if he or she is married to and living with another SSI beneficiary. Under SSI program rules, SSI couples are treated as a single unit for the purposes of determining eligibility for an SSI payment and payment amount. The payment to SSI couples is not double that of an individual. For example, in 2022, the SSI benefit amount for an individual was $841, for a couple it was $1,261.

Although CURCOMP is not a perfect indicator of SSI couple status, we determined in consultation with SSA that incorporation of monthly CURCOMP values into STWSSI construction was worthwhile. As a result, STWSSI values starting in DAF13 are derived differently from the values in prior versions of the database for beneficiaries now identified as members of SSI couples. The effect of incorporating couples' status through CURCOMP into this calculation was in changing the relevant FBR; instead of comparing to the individual FBR, members of an SSI couple are compared to one-half of the couples FBR. This effectively reduced the FBR by 25 percent for beneficiaries identified to be part of a couple. As a result, incorporating CURCOMP made it more likely for members of an SSI couple to have unearned income that is sufficient for suspension of benefits due to excess income, resulting in a higher number of STWSSI=4 months, and fewer STWSSI=1 month.

SSA's rules for allocating non-SSI income of any sort to the two members of an SSI couple implies that income received by either member of the couple, alone, can affect both members' values for STW and BFW. That is, all earned and unearned income is attributed equally to both members of the couple. In fact, each member of a couple will generally have the same STWSSI status and it is not possible to determine which member of the couple did the work that resulted in any suspension or termination.[27]

We include CURCOMPyymm on the DAF so that users can determine for themselves which beneficiaries are potentially affected by this issue if they are interested in doing so. For more information about how CURCOMP was incorporated into STWSSI construction, please consult the SAS program STWSSI in Volume 11.

C. BFW indicators

We developed three sets of monthly variables capturing the dollar value of benefits forgone due to work, included in each Annual file and populated on a monthly basis. As with the STW indicators, there is one BFW indicator for SSDI, one for SSI, and a combined indicator. Each is named according to the year and month through the last month covered by the DAF as described above.

---

[27] There are exceptions to SSI couples having identical statuses. For example, one member of the couple may be incarcerated. In this case, CURCOMP continues to identify each beneficiary as a member of a couple during the months of incarceration, but the STWSSI values in those months are likely to be different.

1. BFWDI

BFWDI is a monthly estimate of the SSDI benefit amount forgone by an SSDI beneficiary and any associated auxiliaries because of work. After an SSDI beneficiary completes the TWP, the beneficiary and any auxiliaries are not entitled to a benefit payment in any month in which he or she engages in SGA after three grace period months. Therefore, for months in which the STWDI indicator is 1, 2, or 3, BFWDI is equal to the SSDI benefit and auxiliary amount for the most recent month in which the STWDI value was 0, i.e., when the beneficiary was last in current pay.

The benefit amounts used in BFWDI are contained in the DUEDyymm and DUEOyymm variables for the last month in current pay, increased for the annual Cost-of-Living Adjustments (COLA) that take place each year in January. To illustrate, these increases are in DAF-year dollars such that a beneficiary whose most recent STWDI value of 0 was in 1998 would have a 2001 BFWDI value that reflects COLAs that occurred between 1998 (when benefits ceased) and 2001 applied to DUED and DUEO (or in other words, the COLAs in January 1999, 2000, and 2001). The values in DUED and DUEO are used in the BFWDI calculation for as long as the STWDI value remains 1, 2, or 3, subject to the imputed COLA increase, until the beneficiary reaches FRA, dies, or becomes re-entitled to SSDI cash benefits (comes back into current pay status).[28]

2. BFWSSI

Unlike SSDI, SSI benefit receipt is not binary. Rather, an SSI beneficiary may receive any amount between $0 and the FBR, which is the maximum monthly federal benefit. Many types of income in addition to that received from work can affect SSI eligibility and reduce the amount due to a beneficiary. In addition to calculating this benefit reduction, there are several programmatic factors that complicate the calculation of BFWSSI, the most important of which are described below.

Starting in DAF18, we updated the version of BFWSSI in the core Annuals files. With this change, we also dropped the _DRAFT suffix on the measure that we have used since its inception, in recognition that further major changes are unlikely. We opted to maintain two STWSSI measures to allow time for user feedback. Having not received any, we have updated the measures on the core files.

a. A refresher on the computation of monthly SSI benefits

In each month, SSI benefits are determined using a two-part assessment. The first step determines whether a benefit is due at all, while the second part determines the amount of the benefit. An SSI recipient is due benefits in a given month provided that the current month's income does not exceed the SSI FBR. In other words, current month income determines whether

---

[28] SSDI beneficiaries who participated in the Benefit Offset National Demonstration (BOND) were eligible to receive reduced SSDI benefits when they worked above SGA. In other words, for SSDI beneficiaries who participated in BOND, benefit receipt is not binary. Because this is outside normal program rules and because BOND participants represent a small share of total beneficiaries, we did not separately change the BFWDI calculation for these participants.

a benefit is paid. If income (either earned or unearned) exceeds FBR, then the SSI recipient does not receive a benefit. If income is under FBR, the benefit may be reduced, but the beneficiary is still in current pay status.

If a benefit is due, the next step is calculating the amount. As explained in SSA's Program Operations Manual System (POMS) SI 02005, SSI payments are computed using a method known as Retrospective Monthly Accounting (RMA). Under RMA, income received in a given month generally affects the payment computation for two months later, with the following primary exceptions:

- When the income in a given month is sufficient to make a beneficiary ineligible due to excess income, the two-month look-back does not apply and the beneficiary is ineligible in that month.

- In the first month of overall eligibility or the first month of eligibility after a period of ineligibility for any reason, the two-month look-back does not apply. In that first month, the SSI payment is based on the income in that month; in the following month, assuming that eligibility continues, the payment is based on the same income as for the first month (one month look-back); in the third month, assuming that eligibility continues, the normal two-month RMA look-back resumes and continues until another exception occurs. Thus, as long as eligibility continues, the income amounts used in the computation of benefits in each of the first three months are the same and the income in that first month is triple counted; that is, the income amounts used in the computation of benefits in each of the first three months after a period of ineligibility for any reason are identical with each other and with the income amounts in the first month in that string.

For any given month, the term budget month (BM) is used to refer to the month the income of which is used to calculate the given month's benefit amount. Depending on the circumstance, the BM is either 0 (the current month), 1 (the month prior), or 2 (two months prior). In other words, the benefit computation for any given month's SSI benefit is based on the income in either the current month, the month prior, or two months prior. While countable income below the FBR in the current month places the SSI recipient in current pay (PSTA=C01) status, the BM, which determines the payment amount, is usually two months prior.[29] So, for example, an SSI recipient in current pay status (PSTA=C01) in February 2002 has countable income less than FBR in that moth but will have a computed benefit based on income in December 2001, the relevant BM in that case. When an SSI recipient is in nonpayment status due to excess income (PSTA=N01), however, the BM is the current month. In other words, when countable income is greater than FBR, that month's income factors into the determination of both a benefit being paid and the benefit amount (which in N01 months is always $0).

---

[29] Exceptions to this rule are noted above. This list of exceptions to RMA is not exhaustive, as explained in the previous section on POMS. However, these are the most significant exceptions and the only ones that are addressed in the construction of BFWSSI.

b.   Calculating BFWSSI

An issue in the construction of the BFW variable for SSI beneficiaries is how to interpret the conceptual intent of the BFW variable—to calculate benefits forgone due to work—when the BM for the calculation of BFW is affected by earnings.[30] Specifically, the BM in reality depends on whether or not a beneficiary had earnings, meaning that under the scenario of no earnings, the BM would be different. In other words, the BFW calculation is based on a counterfactual: it is the difference between what a month's benefit *would* have been in the absence of earned income and the benefit actually paid in that month (if any). Under RMA rules, the BM under the BFW counterfactual might differ from the BM that was actually used for the same month, and the beneficiary might have had earned income in the counterfactual BM.

For example, a beneficiary with $1,000 of countable earned income in August 2013 will have an actual BM of August 2013 in that month, because $1,000 is high enough to preclude benefit receipt in that month. Yet, the counterfactual BM (the month that would have been used in the absence of earnings) could have been either June or July 2013 (depending on the pattern of earnings in prior months). In those months, the beneficiary might have still had countable earned income. The intent of the BFW calculation, however, is to determine the difference between benefits received and benefits that *would* have been received in the absence of any earned income, regardless of *when* that income was earned. Hence, the BFWSSI variable ignores earned income in the counterfactual BM, regardless of when that month occurs. This means that even if the beneficiary had earnings in June or July 2013, the calculation of BFWSSI would ignore them.

The BFWSSI estimate previously on the Annuals files was too high in months where there is deemed income, because EICM and UINC—used to measure total countable income—did not account for all income that the beneficiary might have received. To update BFWSSI, we initially thought we could perform an identical substitution based on the TCNEI formula above. This was not possible, however, because the FCI variable is based on countable income in *the budget month*, while EICM and UINC variables indicate the countable income received *in the current month*, regardless of budget month, payment status, or benefit amount. The budget month when PSTA=N01 (non-payment due to excess income) is the current month, but in C01 (current pay) months, the budget month is usually (though not always) two months prior.

The difficulty incorporating FCI into the BFWSSI calculation arose because BFWSSI is based on computing a counterfactual situation—how much *would* benefits have been without earnings? This counterfactual relies on the assumption that an SSI recipient had no earnings in any month that could possibly affect the benefit amount. Because the counterfactual relies on a situation that did not actually occur, the budget month may be different under the counterfactual relative to the

---

[30] The construction of BFWSSI also accounts for SSA policy on nonrecurring income, described in POMS SI 02005.005.A.4. Prior to April 2005, it was possible for an SSI beneficiary to receive one-time income in the first month of eligibility after a period of ineligibility that, because of RMA, would reduce the SSI benefit in that first month, the second month, and the third month, assuming that eligibility continued. To address this issue, beginning in April 2005, SSA revised its policy on income received in the first month after a period of ineligibility, so that the non-recurring income would affect only the first month and not the two following months. This avoids triple-counting of non-recurring income.

budget month used in the actual computation. The change in budget month under the counterfactual when FCI is only recorded for the actual budget month becomes problematic in cases where STW=1, 2, or 3. An example below illustrates this issue more clearly.

Incorporating the deemed income information contained in FCI is valuable enough to warrant a revised approach to calculating BFWSSI. In particular, when using FCI to estimate deemed income, we now rely on actual budget month information, contained in a variable from the SSI-LF called BMF. In previous versions of the DAF, we estimated the budget month based on data about the beneficiary's status otherwise available in the DAF, following the rules outlined above. Because our method for determining budget month was only an estimation of the true value, it may have differed from the budget month used by SSA. Of course, the true budget month is not necessarily the counterfactual of what would have happened in the absence of earnings. Yet, when we did not account for the sum of deemed income and ISM, we could say with certainty that BFWSSI was overestimated. Changing the budget month under the counterfactual will result in some error, though we do not have a strong indication that the error is systematically upward or downward.

We also impute a value for deemed income and ISM in the last two current pay months before an SSI beneficiary moves from current pay status to suspense. By accounting for all months, this imputed value further reduces the likelihood that our estimate of BFWSSI is biased upward. This estimate is necessary because FCI while in suspense is based on the actual budget month (the suspense month) so for the two months prior to the suspense month, no information on deemed income is available. In DAF14 and earlier versions, we did not attempt to populate values in the two months preceding suspense and instead relied on using the actual budget month.

We estimate deemed income in the months prior to suspense by using the average of deemed income in the two most recent current pay months where FCI is available (provided they are within the 12 months immediately preceding the suspense month) and the two earliest current pay months immediately succeeding the suspense month (provided they are within 12 months of the first suspense month).[31] It is important to note that this estimate of deemed income is based on the assumption that the deemed income in the current pay months surrounding the suspense month are a reasonable proxy for deemed income in the months immediately preceding the suspense month.

c. Accounting for SSI couples

BFWSSI construction accounts for SSI couple status using CURCOMP. The same counterfactual principle described above applies to members of SSI couples, but as described above, suspense or termination for work is less frequent for couples after incorporating the couple's indicator. In addition, the same caveat concerning the impossibility of determining which member of the couple earned the income that is equally attributed to each member of the couple in the administrative data applies to BFWSSI, i.e., a positive BFWSSI value for a member of an SSI

---

[31] We use data from the maximum number of months that has deemed income information available during this period, ranging from 1 to 4.

couple does not necessarily mean that it was the individual's own (rather than his or her spouse's) work that resulted in forgone benefits.

d. Limitations of BFWSSI

There are several known limitations of the BFWSSI variable as it is currently constructed. It is not possible to construct an entirely accurate BFW variable for SSI beneficiaries given the data currently in the DAF and the complexity of the rules for determining SSI benefit amounts. There are two program rules that we cannot fully account for in the construction of BFWSSI because of limitations of the data in the current DAF.

- **SSI couples.** Although BFWSSI accounts for SSI couple status using the CURCOMP variable described above, CURCOMP does not identify all beneficiaries who are members of SSI couples. Therefore, BFWSSI values of those beneficiaries who are members of SSI couples but not identified as such are systematically too high.

- **Proration.** As described in POMS SI 02005.008, benefits are prorated for the first month in which eligibility is obtained after a period of ineligibility when eligibility is reinstated on a date other than the first day of the month, such as when an individual returns to the U.S. after a period of absence of more than 30 days on a day other than the first of the month. It is not possible to calculate an accurate BFWSSI value in these cases because the DAF does not contain the specific day on which eligibility is re-attained. As a result, the BFWSSI amount is too high if proration applies in the month for which the value is being calculated.

Of these known limitations, proration is less frequent and significant, because it only affects one month in each instance. The limitation surrounding SSI couples is more significant and it is important for researchers to be aware of it in their research design and conclusions. We continue to look for ways to improve these measures.

In addition to these known limitations associated with specific SSI policies, certain anomalies in the DAF data also affect the BFWSSI computation. In rare instances, the regular formula for BFWSSI produces values that are incompatible with SSI policy, such as values less than zero or greater than the FBR. In these cases, we have assigned a "special missing" value of ".m" so that researchers are aware that no BFW value can be calculated for these SSI beneficiaries in those months given available data.

Researchers should also be aware that BFWSSI applies only to *federal* SSI benefits forgone due to work. The policies on state supplementation vary so much on a state-by-state basis that it is not feasible to incorporate anything other than federal SSI benefits in the calculation of the BFW counterfactual.

3. Values of the BFW indicators

The values of all three BFW indicators are expressed in dollars, with a minimum value of zero. BFWSSI variable values are constrained at the upper end by that year's individual FBR, or 50% of the couples FBR if the beneficiary is identified as a member of an SSI couple by the

CURCOMP variable value in that month. The combined BFW variable, BFWCM, is calculated simply by adding the value of BFWDI and BFWSSI together for each month. Note that this means that a beneficiary may have an STWCM value of 0 but a positive BFWCM value in any given month when benefits are suspended for work in one program but not the other.

Because the STW variables presume a termination due to work (value of 3) in months after termination when SSA is no longer collecting data on the beneficiary, the BFW variable continues to be positive in such months as well. Beneficiaries whose benefits terminate due to work activity are eligible for a period of 60 months after benefit termination for an immediate return to cash benefits once they cease engaging in SGA under SSA's rules for expedited reinstatement (see POMS DI 13050). Therefore, it is reasonable to assume that the lack of reinstatement implies that these terminated beneficiaries continue to work, at least for the five years following the termination of benefits due to work, but researchers should be aware that this is an assumption and possible to be violated in some months.

None of the BFW values are adjusted for inflation. BFWDI values include imputed COLA increases from the last monthly benefit on record, but these are not adjustments for inflation. Rather, the COLA increases are imputed in order to approximate what benefits *would* have been had they not been suspended or terminated for work.

IX. Race and Ethnicity Categories

We reintroduced race data into the DAF20, after omitting it from DAF15 through DAF19. In those years, we did not include race information given limitations in the data and SSA's decision to stop reporting race in published statistics, in 2009 for OASDI and in 2002 for SSI.[32] Considering recent interest in racial and ethnic disparities, we thought it was important that the information be made available to users. **It is critical that DAF users understand the limitations of the race information available on the DAF and recommend that users read this section in its entirety before proceeding**.

The DAF measure of race (RACE; housed on the DMG file) has the following mutually exclusive categories, that combines racial background with Hispanic ethnicity:

- A - Asian American, Pacific Islander

- B - Black (not Hispanic)

- H - Hispanic

- I - North American Indian/Alaskan native

- W - White (not Hispanic)

- O - Other

- U - Unknown

A. Data collection

The primary source for race and ethnicity information is the application for a Social Security card, via Form SS-5. The race and ethnicity information collected on Form SS-5 has changed twice. Before February 1981, applicants for a Social Security card were allowed to select one of three races: white, negro, or other (specify). SSA classified as "unknown" the race of those who did not make a selection, though the question usually seemed required, so most applicants provided a response. Starting February 1981, the number of race/ethnicity categories on Form SS-5 expanded to five: White (non-Hispanic), Black (non-Hispanic), Hispanic, Asian or Pacific Islander, and American Indian or Alaska Native. Form SS-5 was updated again in 2009 with seven race categories (of which applicants can select more than one): White, Black/African American, Asian, Native Hawaiian, Other Pacific Islander, Alaska Native, and American Indian. A separate question that allows the applicant to select his or her ethnicity—Hispanic or Latino—by answering yes or no was also added to Form SS-5 in 2009.

Race and ethnicity information obtained between 1935 and 2009 is stored in the Master File of Social Security Number Holders and Applications file, known as the Numerical Identification (NUMIDENT) file. Starting in 2010 race and ethnicity information is stored in SSA's Race and Ethnicity Collection System (RECS).

---

[32] Martin, P. P. (2016). Why Researchers Now Rely on Surveys for Race Data on OASDI and SSI Programs: A Comparison of Four Major Surveys. *Research and Statistics Note No. 2016-01*.

B. Limitations of the race data

Because the way in which SSA has collected race data has changed over time, measures of a beneficiary's race may be a function of when that information was collected.[33] Users should be aware that the quality and completeness of race information in the DAF will depend on when beneficiaries applied for a Social Security card and whether or not updated race information was obtained in subsequent agency interactions.

The second important limitation to be aware of is that, reporting race is voluntary. Because administering programs do not require knowledge of a participant's race or ethnicity, the agency no longer requires individuals to provide such information. The fact that the information is collected only on a voluntary basis means that those who report the information may be a self-selected and non-representative sample.

Finally, in 1987 SSA began to rollout a process by which parents can request a Social Security card as part of the hospital birth registration process. This process is called Enumeration at Birth. When the Social Security card is requested in this way, SSA does not receive race or ethnicity information. Over time the vast majority of Social Security card applications have been through Enumeration at Birth and not via Form SS-5 which has limited the collection of race and ethnicity information for more recent cohorts. In a November 2023 press release, SSA announced intentions to begin collecting race and ethnicity information from parents during the birth registration process.

C. DAF Processing

Prior to DAF22, race and ethnicity information for the DAF was obtained from the NUMIDENT file. Starting in DAF22 RECS is our primary source of race information and we use information from NUMIDENT only to fill in missing information from RECS. Because the majority of DAF beneficiaries race information was collected prior to 2009, we have retained the pre-2009 categorization scheme.

Unlike the information in NUMIDENT, RECS allows multiple race category selections. About one percent of DAF beneficiaries had multiple race categories selected in RECS. In those cases, we revert to the information contained in the NUMIDENT file, if it is available. If the NUMIDENT information is not populated, we randomly assign one race from the list of multiple races that were selected. We added a new indicator, RACEMULT_FLAG, to identify cases in which multiple races were selected on the RECS record.

---

[33] Scott, C. G. (2000). Identifying the Race or Ethnicity of SSI Recipients. Social Security Bulletin, 62(4), 9 20.

X. Geographic Measures in the DAF

The DAF contains two types of geographic measures—one-time measures show a person's location at a point in time like birth or disability award, while monthly measures relate to their current location. These variables are listed in the "Variables related to geographic location of beneficiaries" table in Volume 4. In this section, we provide more details about the variables that fall into this category that may be of particular research interest; this is not the full list of geographically-related measures in the DAF.

A. Point-in-time geographic measures

Beginning with DAF18, the DAF contains information related to beneficiaries' country of origin and citizenship status that may be useful for research purposes. These variables are:

- **FOREIGN_BORN**: This is a one-time indicator based on the NUMIDENT that identifies whether beneficiary was born outside of the US. This information is available for SSDI and SSI beneficiaries.[34]

- **COUNTRY_OF_BIRTH**: This is a code that indicates the location of a beneficiary's birth, as recorded in the NUMIDENT. This information is available for SSDI and SSI beneficiaries. The variable contains both state and nation codes and those codes sometimes have dual meanings; this variable must be used with the FOREIGN_BORN indicator to determine the meaning of the code.

- **CITIZEN_NUM**: Derived from the NUMIDENT, this is a measure of citizenship status. This information is available for SSDI and SSI beneficiaries.

- **CITIZEN_CER**: This is a citizenship status measure derived from the CER100% file and therefore is only available for SSI beneficiaries. In most, but not all cases, it aligns with the information in CITIZEN_NUM. Because the fields in the two citizenship measures are not identical, however, users interested in citizenship status for SSI beneficiaries may want to consider both measures.

---

[34] We recommend using this variable in conjunction with COUNTRY_OF_BIRTH and cross-referencing the state and country codes to identify any cases where the information recorded may be inaccurate. In our review of the two variables, there are instances where the foreign-born indicator suggests being born in the United States, but the country of birth does not align to a U.S. state code. Users should use their discretion in handling these cases.

B. Monthly geographic measures

The DAF also contains monthly measures of beneficiaries' location of residence, which can be used to analyze local-level trends in beneficiary outcomes. These measures are based on two snapshot files: the CER for SSI and the DBAD for SSDI beneficiaries.[35] Most of the monthly geographic identifiers in the DAF are derived from four variables:

- **ZIP_SSI_BASEyymm**: Monthly five-digit zip code, as populated in the CER file (in versions of the DAF prior to DAF21 this variable was named ZIP_SSIyymm).

- **ZIP_SSI_EXTyymm**: Monthly four-digit zip extension, as populated in the CER file (available from January 2016 through the end of the current DAF year).

- **ZIP_DI_BASEyymm**: Monthly five-digit zip code, as populated in the DBAD (in versions of the DAF prior to DAF21 this variable was named ZIP_DIyymm).

- **ZIP_DI_EXTyymm**: Monthly four-digit zip extension, as populated in the DBAD (available from January 2016 through the end of the current DAF year).

We first use those measures to construct six-digit county-level FIPS codes, FIPS_SSIyymm and FIPS_DIyymm.[36] The information in the two versions of these variables should be analogous; if ZIP_SSI is missing, so too will be FIPS_SSI; this simply repackages the information from zip codes to county codes.

We also use these measures to construct "best" versions of the monthly geographic measures, which we suggest users work with unless they have strong reason to not do so. To construct these measures, we first identify the correct source for the data. In particular, if only the SSDI or SSI measure is available, we use that. In months where both measures are available, we use the SSI measure, because SSI benefits are reviewed monthly and therefore location is more likely to be accurate. Additionally, we propagate the information forward in months where data is missing. For example, suppose a beneficiary's five-digit zip code was 92692 from January 2006 through March 2009, missing for multiple months, then 80206 from April 2010 onward. The 'best' version of the measure would fill in the missing months with 92692. Of course, users who want to inspect the missing data could revert to the ZIP_SSI_BASEyymm and ZIP_DI_BASEyymm values as needed. The "best" version of monthly geographic variables include:

- **ZIP_BEST_BASEyymm**: Monthly best five-digit zip code of residence for a beneficiary (in versions of the DAF prior to DAF21 this variable was named ZIP_BESTyymm).

- **ZIP_BEST_EXTyymm**: Monthly best four-digit zip extension for a beneficiary

---

[35] Starting in DAF21, we utilize the DBAD and CER files at two points in time. We first use recent data from these files in the initial steps of the DAF construction process to identify new SSDI and SSI beneficiaries based on the DAF selection criteria and we collect geographic data as part of this process. Later when populating the DAF data fields, we obtain available DBAD and CER data in all months for those beneficiaries. Before this change, we only collected DBAD and CER data in months in which beneficiaries met the DAF selection criteria.

[36] Starting in DAF21 we use the 9-digit ZIP to match one-to-one with county in the years from 2016 through the current DAF year. County codes in the years prior to 2016 were established using five-digit ZIP. A given five-digit ZIP code may encompass more than one county. In cases in which the CER or DBAD five-digit ZIP code spans more than one county we use the FIPS code with the plurality of the geographic area within the ZIP code boundary.

- **FIPS_BESTyymm**: Monthly best FIPS code of residence for a beneficiary

- **PSTyymm**: Monthly best state of residence for a beneficiary

The PSTyymm measure is in turn used to construct information about state of residence at benefit award in the ADM, AWARDST. Separate measures are available for each program in DIAWARDST and SSIAWARDST.

C. Linking LAUS and SAIPE Data to Geographic Identifiers in the DAF

SAS formats containing economic data from the LAUS and SAIPE are available for linking to the DAF. These formats contain county-level annual unemployment rates from the LAUS, and county-level monthly poverty rates and median income from the SAIPE. If stored on each DAF record, individual monthly variables would be highly duplicative because each data point (for example, the unemployment rate in Burlington County, New Jersey, in March 1996) applies to many beneficiaries. For this reason, we have stored these data as SAS formats in a separate SAS format library.

The format library name for the LAUS data is OPDR.TG.PRD.ETTW.FINAL.DAF22P.LAUS.FMTLIB and the internal format name is $LAUSYYF (YY=94-22). When $LAUSYYF is applied to a character variable containing the five-digit FIPS code, two-digit year, and two-digit month (for example, 240319905), the variables take on the value of the poverty rate in May 1999 for Montgomery County, Maryland, which has a FIPS code of 24031.

The format library name for the SAIPE data is OPDR.TG.PRD.ETTW.FINAL.DAF22P.SAIPE.FMTLIB and it has two internal formats, $$SAIPEINCYYF (YY=95, 97-22) and $SAIPEPOVYYF (YY=95, 97-22). When these formats are applied to character variables containing five-digit FIPS codes, two-digit years, and two-digit months as in the example above, the variables take on the value of the median income and poverty rate, respectively, for that given FIPS code, year and month.

To extract the formats, the user would include the permanent SAS format library statement in their Job Control Language (JCL) and execute a SAS statement in their extraction DATA step. For instance, the user would need to include the following in the JCL:

//LIBLAU DD DSN=OPDR.TG.PRD.ETTW.FINAL.DAF22P.LAUS.FMTLIB,DISP=SHR

OPTIONS FMTSEARCH = (LIBLAU);

The following code provides an example of how to extract the unemployment, annual county poverty rate, and median income for each SSN in the DAF.DMG component at a fixed time, March 1999.

```
DATA EXAMPLE;

  SET ANN.Y1999;
```

```
    WHERE FIPS_BEST9903 NE "";

    LENGTH FIPSYR $ 7  FIPSYRMO $ 9;

  IF FIPS_BEST9903 NE "" THEN DO;

    FIPSYR = TRIM(FIPS_BEST9903||"99");

    FIPSYRMO = TRIM(FIPS_BEST9903||"99"||"03");

    UNEMPL_RATE = INPUT(PUT(FIPSYRMO,$LAUS99F.),8.);

    MEDIAN_INC  = INPUT(PUT(FIPSYR,$SAIPEINC99F.),8.);

    POVERTY_RATE  = INPUT(PUT(FIPSYR,$SAIPEPOV99F.),8.);

   END;

   RUN;
```

The SAIPE data is not available at the county-level in 1996, and therefore also not available in the SAS formats available in the DAF; researchers attempting to access the 1996 SAIPE data in the SAS formats will receive an error message. Instead, researchers can manually code 1996 data using Table X.1, which provides the national- and state-level poverty and median household income estimates from that year.

Table X.1. State-level poverty and median household income values, 1996

State name	Estimated percent of people of all ages in poverty, 1996	Estimate of median household income, 1996
United States	13.7	35,492
Alabama	16.7	29,618
Alaska	10.6	44,797
Arizona	16.3	32,708
Arkansas	17.6	27,367
California	16.7	38,691
Colorado	10.5	38,923
Connecticut	9.0	44,981
Delaware	9.8	39,701
District of Columbia	21.9	34,436
Florida	14.2	31,008
Georgia	15.3	33,763
Hawaii	11.7	43,677
Idaho	11.9	33,279
Illinois	12.0	39,490
Indiana	9.3	35,502
Iowa	9.5	33,721
Kansas	10.8	33,610

Table X.1. State-level poverty and median household income values, 1996

State name	Estimated percent of people of all ages in poverty, 1996	Estimate of median household income, 1996
Kentucky	17.5	30,630
Louisiana	20.7	28,921
Maine	12.0	33,002
Maryland	9.7	44,196
Massachusetts	9.8	40,686
Michigan	12.2	38,266
Minnesota	8.8	39,791
Mississippi	20.8	26,901
Missouri	12.6	32,947
Montana	15.1	28,714
Nebraska	9.4	33,562
Nevada	9.6	38,213
New Hampshire	6.3	40,153
New Jersey	9.0	46,872
New Mexico	20.6	27,014
New York	16.3	35,696
North Carolina	12.8	34,487
North Dakota	11.5	30,798
Ohio	11.7	34,198
Oklahoma	17.1	27,648
Oregon	12.3	35,144
Pennsylvania	11.5	35,109
Rhode Island	11.8	36,402
South Carolina	15.2	32,728
South Dakota	13.0	29,810
Tennessee	14.8	31,097
Texas	17.5	32,773
Utah	9.4	36,360
Vermont	11.0	33,352
Virginia	11.5	38,510
Washington	11.5	37,847
West Virginia	19.9	25,760
Wisconsin	8.7	38,598
Wyoming	11.1	31,173

Source: United States Census Bureau.

XI. Ticket Event Dates

The SSA administrative data provide a history of events for the Tickets received by a beneficiary in the TTW program. Historically, Ticket events began when a beneficiary was selected for inclusion in the TTW program, after which a Ticket was mailed to the beneficiary. Physical possession of a Ticket has never been a factor in Ticket eligibility, and between December 2012 and March 2015, SSA did not automatically mail Tickets—instead, they were only mailed upon request. Mailings to new beneficiaries resumed in April 2015, along with mailings to those celebrating their first, second, or third benefit anniversaries. Catch-up mailings were also sent to all beneficiaries who had not been mailed a ticket during the hiatus.

The DAF incorporates the data from multiple SSA source files to populate the TKTMAILDTn, a variable reflecting the date a ticket was mailed to the beneficiary. This variable is drawn from multiple sources, reflecting the various ticket mailing schemes that have been in place over time.[37] Additional types of mailings that SSA has implemented since the "hiatus" from mailing tickets that occurred from July 2011 through March 2015 are anniversary mailings—one-year and three-year, ad hoc mailings in response to beneficiary requests, and catch up mailings. These new types of mailings are not reflected in the TKTMAILDTn variable. Instead, as of DAF17 and later, they are stored in a series of other new variables: TKTINTMLDTn (initial mailing), TKTADHMLDTn ('ad hoc' mailing), TKT1YRMLDTn (one-year anniversary), TKT3YRMLDTn (three-year anniversary), TKTCUPMLDTn (catch-up mailing). As with the original "initial" ticket mailing, each of these variables for the new types of ticket mailings since 2015 are repeated for each period of ticket eligibility.

Once a beneficiary is eligible for the TTW program, he can choose to assign his Ticket to a provider of employment services. The Ticket can, though does not need to, be later unassigned and re-assigned to another provider. The Ticket remains active until the participant stops receiving benefits for reasons other than work, such as medical recovery, retirement, death, or not meeting timely progress requirements stipulated by the TTW program. If one of those events occurs, the Ticket is terminated. If a beneficiary has a subsequent period of disability eligibility, he can receive another Ticket, and event dates for each subsequent Ticket are also tracked. In this situation, the new ticket mail dates would begin to be populated as described above, as would ticket assignment, unassignment, reassignment, and termination. For example, if a beneficiary receives disability benefits for cancer, recovers, and then later has a new period of disability

---

[37] Initial mailings (those through June 2011) and mailings that occurred during the "hiatus" when SSA did not automatically send tickets (from July 2011 through March 2015) are recorded in the DAF component containing data related to TTW participation (TKT) table. Mailings that have occurred since the hiatus ended are stored in the TKTSENT table. TKTMAILDTn is populated with either the mail date on the TKT table or an initial or catch-up mail date from the TKTSENT table. Beneficiaries who do not appear on the TKTSENT table, those that have not had a new initial, catch-up, or anniversary mailing, but do appear on the TKT table will have a TKTMAILDTn that reflects the mail date appearing on the TKT table. If that date occurs from July 2011 through March 2015, then the TKTMAILDTn value reflects the date the beneficiary was selected for participation in the TTW program, even though no physical ticket was mailed. When both the information in TKT and TKTSENT are populated, the mail dates from TKTSENT takes precedent over those from TKT.

benefits due to an injury, there would be two periods of disability and two Tickets, with the first Ticket showing a terminated status and the second Ticket an active status.

A. Sequence of Ticket event dates and Ticket records in DAF

This section provides more information about Ticket event information, including some potential pitfalls, explains how the data are stored in the DAF, and illustrates Ticket events more fully.

Though not always the case, Ticket events are expected to proceed in the following logical sequence: (1) beneficiary becomes eligible for the TTW; (2) Ticket is mailed; (3) the beneficiary assigns Ticket to provider; (4) beneficiary unassigns their ticket (after completing services or prior to moving to a different provider); (5) beneficiary reassigns their ticket and then (6) the Ticket is terminated. It is not necessarily the case that tickets will be unassigned or terminated, though may occur. We present examples of how recorded data may deviate from this model in what follows.

During construction of the DAF, dates for all Ticket entries are numbered and re-organized within a reference system, based on Ticket assignments. Each numbered Ticket assignment entry consists of the mail date, assignment date, unassignment date (often blank), and termination date, using the following variables, each of which can have up to 22 occurrences:

- TKTMAILDDT1 to TKTMAILDDT22
- TKTASGNDDT1 to TKTASGNDDT22
- TKTUNASGDT1 to TKTUNASGDT22
- TKTTERMDDT1 to TKTTERMDDT22

Most participants have fewer than 22 Ticket assignment entries and the number of entries for each participant is reflected in the variable NOE (Number of Entries). Therefore, a DAF record where NOE = 1 would use the following variables to store the SSA administrative Ticket data:

- TKTMAILDDT1
- TKTASGNDDT1
- TKTUNASGDT1
- TKTTERMDDT1

The remaining occurrences, 2-22, for each of the four variables would be blank. A record where NOE = 2 would use the following DAF variables:

- TKTMAILDDT1 and TKTMAILDDT2
- TKTASGNDDT1 and TKTASGNDDT2
- TKTUNASGDT1 and TKTUNASGDT2
- TKTTERMDDT1 and TKTTERMDDT2

The remaining occurrences, 3-22, for each of the four variables would be blank.

B. Examples of conceptual Ticket assignment information and corresponding DAF records

In this section, we present hypothetical examples of Ticket assignment data. Through these examples, we demonstrate that while the majority of cases will contain the expected series of data, sometimes anomalies occur. We do not attempt to clean records with seemingly contradictory information. Thus, researchers should expect that a small number of date values will be out of alignment with others and therefore sequences of events should be carefully examined before use for research purposes.

Example 1

Example 1 is a conceptual example of the SSA administrative data for the Ticket event dates for a participant. In this example, shown in Table XI.1a, a Ticket was mailed in February 2005, assigned in July 2005, and terminated—perhaps due to medical recovery—in February 2007.

Table XI.1a. Example 1 conceptual Ticket assignment information

Ticket 1			
Mail Date	Assign Date	Status Code	Status Effective Date
Feb 05	Jul 05	T	Feb 07

In the DAF, the corresponding variables for Example 1 would appear as shown below in Table XI.1b; all 22 occurrences for each of the four Ticket date variables are present in a single wide record. Note that in this example there would be just one Ticket assignment entry, with data in just the first occurrence of three variables. The first unassignment date would be blank, as would all variables for occurrences 2-22.

Table XI.1b. Example 1 corresponding DAF variables

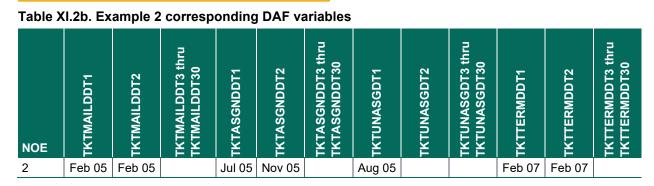| NOE	TKTMAILDDT1	TKTMAILDDT2 thru TKTMAILDDT30	TKTASGNDDT1	TKTASGNDDT2 thru TKTASGNDDT30	TKTUNASGDT1	TKTUNASGDT2 thru TKTUNASGDT30	TKTTERMDDT1	TKTTERMDDT2 thru TKTTERMDDT30
1	Feb 05		Jul 05				Feb 07	

Example 2

Sometimes a participant re-assigns a Ticket to another provider, in which case the cycle includes two intermediate steps, unassignment, and reassignment (Table XI.2a).

Table XI.2a. Example 2 conceptual Ticket assignment information

Ticket 1					
Mail Date	Assign Date	Unassign Date	Re-assign Date	Status Code	Status Effective Date
Feb 05	Jul 05	Aug 05	Nov 05	T	Feb 07

Note that this example does not reveal whether the beneficiary assigned their Ticket to different providers or to the same provider twice, but the DAF does contain this information (though not shown here) in the DUNSn variable on the Ticket Component. In the DAF, as shown below in Table XI.2b, there would be two Ticket assignment entries. Both entries have the same mail date and termination dates, but unique assignment dates. The assignment date for the first entry is the date the Ticket was originally assigned while the assignment date for the second entry is the date the Ticket was reassigned.

Table XI.2b. Example 2 corresponding DAF variables

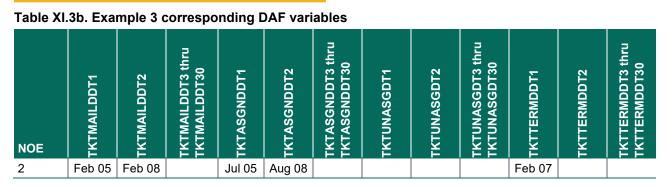| NOE	TKTMAILDDT1	TKTMAILDDT2	TKTMAILDDT3 thru TKTMAILDDT30	TKTASGNDDT1	TKTASGNDDT2	TKTASGNDDT3 thru TKTASGNDDT30	TKTUNASGDT1	TKTUNASGDT2	TKTUNASGDT3 thru TKTUNASGDT30	TKTTERMDDT1	TKTTERMDDT2	TKTTERMDDT3 thru TKTTERMDDT30
2	Feb 05	Feb 05		Jul 05	Nov 05		Aug 05			Feb 07	Feb 07	

Example 3

A beneficiary with multiple periods of eligibility may receive more than one Ticket (Table XI.3a). Note that the first Ticket is terminated before the cycle begins again for the second Ticket and that the current status of the second Ticket is assigned, as indicated by the "I" for Status Code.

Table XI.3a. Example 3 conceptual Ticket assignment information

Ticket 1				Ticket 2			
Mail Date	Assign Date	Status Code	Status Effective Date	Mail Date	Assign Date	Status Code	Status Effective Date
Feb 05	Jul 05	T	Feb 07	Feb 08	Aug 08	I	Aug 08

For this example also, there would be two populated occurrences of each variable in the DAF, with the rest blank as shown in Table XI.3b.

Table XI.3b. Example 3 corresponding DAF variables

NOE	TKTMAILDDT1	TKTMAILDDT2	TKTMAILDDT3 thru TKTMAILDDT30	TKTASGNDDT1	TKTASGNDDT2	TKTASGNDDT3 thru TKTASGNDDT30	TKTUNASGDT1	TKTUNASGDT2	TKTUNASGDT3 thru TKTUNASGDT30	TKTTERMDDT1	TKTTERMDDT2	TKTTERMDDT3 thru TKTTERMDDT30
2	Feb 05	Feb 08		Jul 05	Aug 08					Feb 07		

Example 4

Occasionally, the administrative data that documents the Ticket events may exhibit some anomalies (Table XI.4a), in which the first assignment, status code, and date fields contain no data. Furthermore, the value for the second assignment date field occurs prior to the second mail date. The likely scenario here is that the second assignment date is actually the first assignment date and possibly likewise for the second status code and status date.

Table XI.4a. Example 4 conceptual Ticket assignment information

Ticket 1				Ticket 2			
Mail Date	Assign Date	Status Code	Status Effective Date	Mail Date	Assign Date	Status Code	Status Effective Date
Feb 05				Apr 07	Jul 06	T	Jan 08

In the DAF, such a case would look like the data in Table XI.4b: note that although it is likely that the second assignment date should pair with the first mail date, we cannot be sure of this and therefore *we do not attempt to apply corrections to the data*.

Table XI.4b. Example 4 corresponding DAF variables

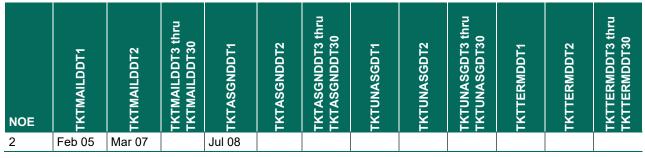| NOE	TKTMAILDDT1	TKTMAILDDT2	TKTMAILDDT3 thru TKTMAILDDT30	TKTASGNDDT1	TKTASGNDDT2	TKTASGNDDT3 thru TKTASGNDDT30	TKTUNASGDT1	TKTUNASGDT2	TKTUNASGDT3 thru TKTUNASGDT30	TKTTERMDDT1	TKTTERMDDT2	TKTTERMDDT3 thru TKTTERMDDT30
2	Feb 05	Apr 07			Jul 06						Jan 08	

Example 5

In Example 5, the first assignment date occurs after the second mail date (Table XI.5a); in this case it is likely that the first Ticket was never assigned, and that the first assignment date should actually be the second assignment date.

Table XI.5a. Example 5 conceptual Ticket assignment information

	Ticket 1				Ticket 2		
Mail Date	Assign Date	Status Code	Status Effective Date	Mail Date	Assign Date	Status Code	Status Effective Date
Feb 05	Jul 08			Mar 07			

In the DAF, this data anomaly would look like the data in Table XI.5b. Because we cannot be sure that the first assignment date actually belongs with the first Ticket, we do not attempt to apply corrections to the data.

Table XI.5b. Example 5 corresponding DAF variables

NOE	TKTMAILDDT1	TKTMAILDDT2	TKTMAILDDT3 thru TKTMAILDDT30	TKTASGNDDT1	TKTASGNDDT2	TKTASGNDDT3 thru TKTASGNDDT30	TKTUNASGDT1	TKTUNASGDT2	TKTUNASGDT3 thru TKTUNASGDT30	TKTTERMDDT1	TKTTERMDDT2	TKTTERMDDT3 thru TKTTERMDDT30
2	Feb 05	Mar 07		Jul 08								

A separate question allows the applicant to select his or her ethnicity—Hispanic or Latino—by answering yes or no.