

Volume 2: Working with the DAF21

April 2023

Submitted to:

Social Security Administration
Office of Retirement and Disability Policy
Office of Research, Demonstration, and Employment
Support
Washington, DC 20024-2796
Project Officers: Paul O’Leary and Debra Tidwell-Peters
Contract Number: 28321322D00060010

Submitted by:

Mathematica
1100 1st Street, NE
12th Floor
Washington, DC 20002-4221
Telephone: (202) 484-9220
Facsimile: (202) 863-1763
Project Director: Matthew Urato
Reference Number: 51390.BY.T02.260.000

Suggested Citation: U.S. Social Security Administration, Office of Retirement and Disability Policy, Office of Research, Demonstration, and Employment Support. (2021). “Disability Analysis File 2021 (DAF21) Documentation: Data from January 1994 through December 2021.” Washington, DC: Mathematica, April 2023. [Volume 2: Working with the DAF21]. 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

Suggested Citations	vi
Glossary.....	viii
Overview of DAF Documentation	xi
I. Extracting Data from the DAF	1
A. Extracting records using a finder file.....	2
B. Extracting records using selection criteria	2
II. DAF Variable Conventions	4
A. DAF conventions for variables with multiple chronological occurrences	4
1. The “yymm” variables.....	4
2. The “n” variables	4
B. DAF conventions for date variables.....	5
1. Formats for date variables.....	5
2. Birth and death dates	5
C. Missing values in the DAF.....	6
III. The Importance of the Administrative Source for Understanding How to Use DAF Data	7
IV. Details About the DAF Payments Component.....	11
A. EN payments files (TTW payments to ENs or State VR Agencies acting as ENs—two files)	11
B. VRRMS file (payments related to SSA cost-reimbursement to State VR Agencies—two files).....	11
1. VRRMS Data through February 2017.....	12
2. VRRMS Data from March 2017 forward	13
V. Using DAF Data Marts	15
A. DAF 10% Data Mart.....	15
B. DAF ADM.....	17
1. Method Used to Identify Award Date	20
2. Method for Identifying Eligibility Date.....	22
VI. Using DAF Extracts.....	23
A. TTW participant extract	23

B. NBS extract.....	25
C. Demonstrations and surveys extract	26
VII. Using the DAF-Linkable Files with DAF Components.....	28
A. DAF-RSA files	28
1. Summary of Changes in RSA-911.....	28
2. Guide to Using DAF-RSA files.....	29
B. MEF DAF-linkable file with DAF components	33
VIII. Using Information on Continuing Disability Reviews in the DAF	35
A. Overview of the DAF standalone CDR file.....	35
B. Structure and contents of the standalone CDR file	36
C. Working with information from the Waterfall File.....	39
D. Working with information from the DCF review and event tables.....	40
E. How does information in the standalone CDR file relate to other information in the DAF?.....	41
Appendix A DAF21 Ticket to Work (TTW) Linkable File Extract	A.1
Appendix B DAF21 National Beneficiary Survey (NBS) Linkable File Extract.....	B.1
Appendix C DAF21 Survey and Demonstration Projects Linkable File Extract	C.1
Appendix D DAF21 Master Earnings File (MEF) Documentation and JCL/SAS Code.....	D.1

Note: User interested in LAUS and SAIPE should consult Volume 3; we have moved a section formerly in Volume 2 into that volume to combine all geographic-related content into a single location.

Tables

III.1.	Example of using the 831 & 832/833 files to build time-based variables	9
IV.1.	Variables Available for Each VRRMS Claim in File Covering the Period from March 2017 Onward	14
V.1.	Number and percent of records in the DAF21 10% data mart.....	17
VI.1.	Frequency of latest Ticket assignment dates (by month, by year) ^a	24
VI.2.	Frequency of participation flags for de-duplicated combined demonstration and surveys extract finder files in DAF21	27
VII.1.	File structure of the RSA-DAF Files under the new and old layouts	30
VII.2.	Variables in the RSA-DAF linkable summary file	32
VII.3.	Variables in the MEF DAF-linkable file.....	34
VIII.1.	Variables in the standalone DAF CDR file	37
VIII.2.	Values for Adjudicative Level in the Waterfall File	40

Figures

I.1.	Flowchart of ways to select the starting file for DAF analysis	3
V.1.	Layout of the DAF21 10% data mart components	16

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
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
DWB	Disabled Widow Beneficiaries
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
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	SSI 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 DAF21. Provides an overview of the structure and contents of the DAF and related linkable files.
- **Volume 2:** Working with the DAF21. 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 DAF21. 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 DAF21 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 DAF21 Against Other Sources. Provides an explanation of validation methods and summary of validation results.
- **Volume 7:** DAF21 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:** DAF21 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:** DAF21 Source File Descriptions. Describes the administrative source files used to construct the DAF.
- **Volume 10:** DAF21 Administrative Source File Documentation. Contains documentation from SSA or other agencies on the administrative source files described in Volume 9.
- **Volume 11:** DAF21 Construction Code. Contains all SAS code used to construct the DAF.
- **Volume 12:** DAF21 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 DAF21," 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 DAF21"
View lists of DAF variables	Volume 4, "Lists of DAF21 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 DAF21"
Identify variables for a specific research task	Volume 4, "Lists of DAF21 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 DAF21"
Identify administrative data sources for the DAF	Volume 9, "DAF21 Source File Descriptions"
Understand the linkage of the DAF to RSA-911 data and contents of the RSA files	Volume 12, "DAF21 RSA Administrative Source File Documentation"
Generate ideas for using the DAF more efficiently	Volume 1, "Getting Started with the DAF21" and Volume 2, "Working with the DAF21"
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 DAF21"
Understand what variables have changed in the most recent DAF	Volume 4, "Lists of DAF21 Variables"
Read about how information in the DAF is validated against other sources	Volume 6, "Validating the DAF21 Against Other Sources"

I. Extracting Data from the DAF

The Disability Analysis File (DAF) is comprised of many files and millions of beneficiary records. In total, there are 52 files associated with the DAF21 components, including:

1. Demographic component of the DAF (DMG) component (1 file)
2. Annual component 1994-2021 (28 files; one for each calendar year)
3. Ticket component (1 base file and 19 annual files—one for each calendar year between the start of the Ticket program in 2002 and 2021)
4. Payments component (4 files; 2 files related to payments made by the Social Security Administration (SSA) to State Vocational Rehabilitation (State VR) Agencies under the cost reimbursement system, 2 other files—one horizontal and one vertical—related to payments made under the Ticket to Work (TTW) program to Employment Networks (EN) and State VR Agencies operating as ENs)

In addition to the core components, there are also separate files available for use for the standalone Continuing Disability Review (CDR) file (new in DAF19), data marts, and extracts. Collectively, the DAF components contain thousands of variables. Volume 4 of the DAF documentation provides an alphabetical list of variables, by component, along with lists of variables from the DAF data marts, extracts, and linkable files, described in more detail in this volume. In addition, Volume 4 also contains variable lists sorted by analytic category to provide researchers with common variables used in certain research contexts. For example, lists of variables are provided for: (1) record identification, (2) identification of widows and auxiliary beneficiaries, (3) geographic location, (4) beneficiary demographic information, (5) the calculation of beneficiary age and other important beneficiary dates, (6) diagnosis codes, (7) disability status and medical information, (8) application and award information, (9) continuing disability status, (10) post-entitlement status, (11) benefit and payment amounts, (12) earnings, income, and work, (13) Ticket participation, and (14) payments to EN providers.

Some of the DAF files have over 30 million records, often making it unwieldy to use the full DAF without paring down observations or variables (or both). Because the DAF is composed of Statistical Analysis Software (SAS) files, this is a relatively easy process, but as an alternative, researchers may consider using one of the existing extracts or data marts before proceeding with a plan to use the full DAF. These are described in Volume 1 and Chapters V and VI of this volume. These data extracts were developed for the express purpose of easing the accessibility of the DAF and were designed to be useful for a broad array of research purposes. Testing a research design, in particular, should be done on the 10% data mart (or the DAF public use file described in Volume 1) when feasible. If none of the existing extracts or data marts suit the research being conducted, then the full DAF should be used.

Using the full DAF generally starts with making an extract, for which there are two approaches, as shown in Figure I.1. One is to match DAF records to a finder file of Social Security Numbers (SSN) for a subpopulation of beneficiaries, and the other is to select DAF records according to

selection criteria designed for a specific analysis need. The decisions about which parts of the DAF to use and how to extract data are critical. The figure below provides a model for how the decision might be made along with examples of the types of research questions that might be most appropriate if using a given component as a starting place for DAF analysis. Additional information about how one might use these selection criteria is described in Sections A and B below.

A. Extracting records using a finder file

The finder file approach is common when records from another file are to be combined with DAF data, such as Medicare participants from a particular state, nationwide Supplemental Nutrition Assistance Program (SNAP) recipients, or recipients of services from a State VR Agency. It is generally more straightforward than extracting records by selection criteria, and because the DAF is a SAS file, it is easy to select only the variables needed for a specific analysis without reading all of the DAF's data. However, if the finder file contains SSNs obtained outside SSA, for example from a state agency or another federal agency, they must be verified using SSA's Enumeration Verification System (EVS), and only records that pass the EVS checks can be used with the DAF.¹ Users who are unfamiliar with the EVS process should contact ORDES.DAF@ssa.gov for assistance with EVS. After SSN verification, the next step is to merge the verified records from the finder file to one of the DAF components, often the DMG component, keeping only the DAF records that match the finder file. As needed, researchers can merge the extracted records to other DAF components using SSN to obtain the remaining variables for the task.

B. Extracting records using selection criteria

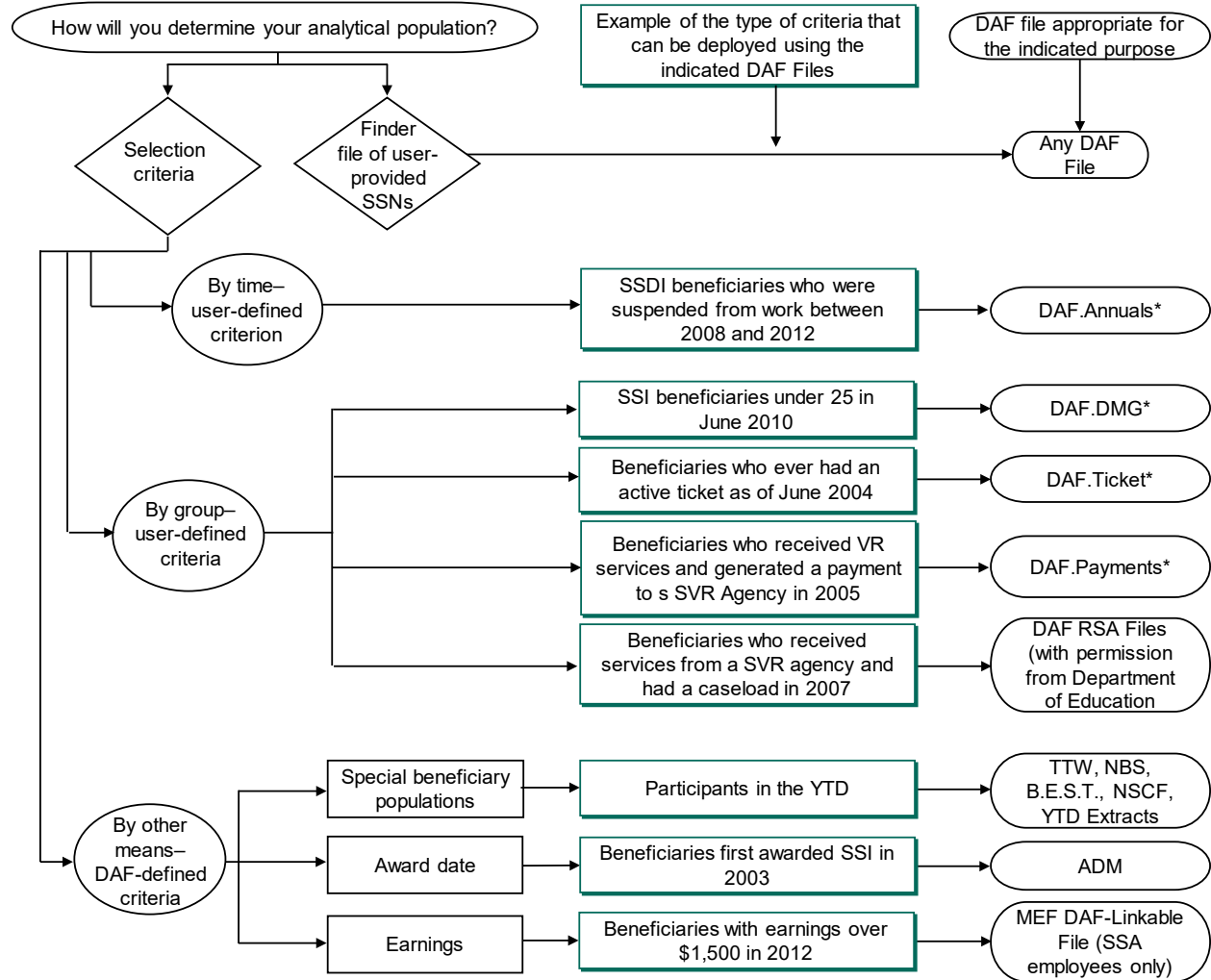
The selection criteria method is common when the targeted participants are not known ahead of time, but their characteristics can be specified. An example is selecting all beneficiaries who were between ages 21 and 25 in 2007 and received a Social Security Disability Insurance (SSDI) payment for one or more months during that year. Another example is all beneficiaries with at least one month of cash benefit suspension during the calendar year. When using this method, we caution against specifying the selection criteria too narrowly to start. If a researcher discovers the extraction criteria were too narrow, it would later be necessary to repeat the extraction process from the beginning, which can be time-consuming given the size of the source files. At the same time, care should be taken not to specify the selection criteria too broadly. If you select far more beneficiaries than you need, the resulting extract files may be large and unwieldy, leading to extra time and cost for your research task.

To ease the selection of beneficiaries, several commonly used variables are now included on all DAF components and may minimize the need to merge across components to select beneficiaries. As detailed in Volume 1, these variables are: DOBBEST, DODBEST, DOEI,

¹ Note that EVS will not produce a match on accent marks and other diacritical characters such as the “ñ” in Peña or “é” in Renée. As a result, bias could be introduced into the data if, for example, people who are Hispanic are excluded from analysis because of a non-matching “ñ.” To address this potential problem, it is best to substitute for diacritical marks using a SAS function such as “cl_name = tranwrd(cf_name,'ñ','n');”.

DXIPRIBEST, FRADATE, SEX, SSDI_ONSET, SSDIyy, SSIELIG_FIRST, SSIyy, and TKTy.

Figure I.1. Flowchart of ways to select the starting file for DAF analysis



* Indicates a file for which the DAF 10% Data Mart can be substituted, if a 10% sample will suffice or for code testing.

II. DAF Variable Conventions

A. DAF conventions for variables with multiple chronological occurrences

There are many variables in the DAF with multiple occurrences, and these fall into two major categories: 1) monthly variables (these have a “yymm” suffix) and 2) multiple-occurrence variables not related to months (these have an “n” suffix).

1. *The “yymm” variables*

The “yymm” variables in DAF are also referred to as longitudinal or monthly variables and are prevalent in the Annual, Ticket Annual, and Payments (Horizontal) files. They are named according to the month and year using the variable suffix “yymm,” which indicates the year and month in question. For example, the generic variable name DUEdyymm represents “SSDI Benefit Payment Due in yymm” while DUEd9805 is the specific variable name for “SSDI Benefit Payment Due – May 1998.”

2. *The “n” variables*

The “n” variables also have multiple occurrences but are not tied to a specific point in time in the same way as the “yymm” variables. These “n” variables are found in the DMG component, Ticket Base, and Payments (Horizontal) files. For reasons outlined in this section, we suggest carefully examining an “n” variable’s contents before using the data. A summary table of the “n” variables by topic area can be found in Volume 4.

Many, but not all, “n” variables are date variables. Unlike the “yymm” monthly variables, the chronological order of the “n” variables is not always readily apparent. Some “n” variables may be arranged in ascending order, that is, the first occurrence of the “n” variable contains the chronologically earliest data, and the second occurrence contains the chronologically second data, and so on. But this is not always the case. In cases where the “n” variable is derived directly from the administrative source file, we retain the order of data as it is stored.

In other cases, the “n” variable does not contain date information, but rather contains information about an occurrence. In these instances, the “n” variable usually has a corresponding date variable. An example of this an occurrence “n” variable is TOCn, or “Type of Claim (n=1-20).”² The corresponding date variable is TOC_STARTn, “SSDI Type of Claim Start Date (n=1-20).” In this context, the suffix “n” indicates the number of the occurrence of the variable under consideration. The number of occurrences is indicated in the variable’s label; in this case, it indicates the variable has up to 20 occurrences. The date variable contains the relevant dates for the values in the same-numbered “n” variable. For instance, the date in TOC_START1 indicates the relevant date for the value in TOC1. To determine the correct chronological order of the values in variables in which “n” denotes an occurrence, such as TOCn, first examine the dates in

² As discussed in Volume 3 in the section on “Determining monthly participation based on payment entitlement status variables,” working with TOCn is a bit challenging depending on the time period of the analysis due to a change over time in how the information was recorded.

the corresponding date variable, TOC_START_n in this example. If they are in ascending order, then the values in the “n” variable are usually also in ascending order by date.

In addition to a date that aligns to occurrence-type “n” variables, some “n” variables have a complementary variable that indicates the maximum number of populated occurrences. For example, the variable TOC_NUM indicates the highest number of populated occurrences of TOC_n. For example, if TOC_NUM contains the value “5”, then 5 occurrences of TOC_n are populated while the rest have missing data. It is possible that one or more of the variables presumed to be populated is missing.

In Volume 5, the information provided for each DAF variable includes a field that indicates related variables, to assist users in linking “n” variables to other variables that might be of strong interest to researchers.

B. DAF conventions for date variables

1. *Formats for date variables*

Most DAF date variables are unformatted SAS dates tied to January 1, 1960, a SAS convention. The values of these date variables represent the number of days the date occurs before or after January 1, 1960. To display these variables as recognizable dates, they must be formatted with a SAS format statement, such as: *FORMAT VAR yymmdd10*. This statement displays the internal SAS numeric value of 16649 as “2005-08-01.” Other SAS formats with different display options are also available and can be found in SAS documentation. Some DAF date variables already have a SAS format attached to them and these variables will display in a recognizable format while the variable will retain the underlying numeric value.

2. *Birth and death dates*

There are multiple birth and death date fields in the DAF because many of the SSA administrative files used to construct the DAF contain these dates. When the various DAF source files are combined to form a record in the DAF for each beneficiary, discrepancies among birth and death dates for some beneficiaries become apparent. For example, a concurrent beneficiary, one who participates in both SSDI and Supplemental Security Income (SSI), may have the birthdate “06/01/1957” recorded in the SSDI administrative data but “01/06/1957” recorded in the SSI administrative data. Another problem occurs if only a partial date was recorded, with perhaps the day or month missing, resulting in that the date cannot easily be used for calculations or comparisons.

During construction of the DAF, we develop a “best” variable for birthdate and death date—DOBBEST and DODBEST—to resolve some of these issues. We developed code to compare the birth/death dates from the various administrative files, identify the date that was most likely to be correct, and set any missing days or months to a default value. We only populate DOBBEST and DODBEST if Numerical Identification File (NUMIDENT) has a valid year, otherwise we set those variables to missing. For observations with fully valid dates in NUMIDENT, we populate

DODBEST and DOBBEST with those dates. If the NUMIDENT birth/death date has a valid year but is missing either day or month (or both), we populate DOBBEST and DODBEST with the day or month from the Supplemental Security Record (SSR) and the Master Beneficiary Record (MBR) in that order. If this process does not yield a valid month and day for DOBBEST or DODBEST but the NUMIDENT has a valid birth/death year then we impute these dates by setting missing months and/or days to “01” and identify the case with a special flag, DOBFLAG for birth and DODFLAG for death.

We also retain all original birthdate variables from the different source files, including DOBMBR (from the MBR), DOBSSR (from the SSR), DOBCER (from the Characteristics Extract Record 100% Field), DOBNUM (from the NUMIDENT), and DOBTKT (from the Disability Control File).

Most DAF date variables are stored in date format, but the birthdate variable from the NUMIDENT (DOBNUM) is stored in a character format instead, following the convention mmddyyyy. The reason is that some NUMIDENT records have birthdates with the month or day missing, and in these cases, SSA’s convention is to fill in the missing day or month with “xx” or “99”, e.g., “03991954” or even “xxxx1957”. The resulting value is not a true date and cannot be stored in a date-format variable or used for computations or comparisons without first correcting the missing date components.

C. Missing values in the DAF

Some monthly variables in the DAF are not available for the full range of months and years, perhaps because they were not available in earlier versions of the SSA source files or because they did not contain meaningful or useful data until more recent years. For instance, variables related to the Ticket to Work program are not available in the earlier years of the DAF because the program was not rolled out until 2002 (and then phased in across states through 2004).

Even for variables that were available in a given month, not all beneficiaries will have data populated. This could occur, for example, because the beneficiary had not yet started receiving benefits. If a beneficiary began to receive benefits in June 1999, all variables in the DAF (e.g., DUEdyymm) for that beneficiary from January 1994 to May 1999 are missing. In those cases, the value for a particular month may be missing, blank, or zero-filled. We suggest that before working with a given variable that users consider the frequency of each variable to understand how missing values might have been recorded; the variable detail pages in Volume 5 provide a set of categories that are applicable for each variable.

III. The Importance of the Administrative Source for Understanding How to Use DAF Data

The data contained in the DAF are extracted or derived from SSA administrative records, which were designed for purposes of program administration rather than research.³ As such, each source file has nuances that may affect the interpretation of particular variables. Here, we discuss some of the issues to keep in mind based on the source of each variable. Note that each variable detail page in Volume 5 identifies the underlying administrative source for the data element.

Variables from MBR. The MBR contains information on beneficiaries enrolled in the SSDI or Old Age Survivors Insurance (OASI) programs. The MBR has been used for many years, with important changes that occurred in 2004 and 2009 that affects how to use data before and after this time. We explain these changes in what follows.

In September 2004, SSA conducted a rewrite of the MBR, which reformatted some of the elements contained in the file. As a result, some variables in the DMG component that were new at the time of MBR extraction are blank or missing for SSDI beneficiaries who are on the DAF but were no longer entitled to benefits when the MBR was rewritten. These new variables are listed below. However, there is no easy way to determine whether blank or missing values for these variables are a result of this MBR rewrite process or a result of other anomalies in the administrative data.

- SSDI disability cessation date (ADCn)
- SSDI appeals date (APSn)
- SSDI basis for denial code (BDCn)
- SSDI disability award code (DACn)
- SSDI disability cessation date (DDBCn)
- Date of medical decision (SSA or state DDS) (DODECn)
- SSDI disability adjudication date (DSDn)
- SSDI primary diagnosis code from MBR (DXPRI_MBRn)
- SSDI secondary diagnosis code from MBR (DXSEC_MBRn)
- SSDI hearing decision date (HDDn)
- Level of adjudication (JUDLVLn)
- SSDI level of denial code (LODn)
- Result of disability determination (RDTn)
- Program identification (RIDn)

³ These include the Supplemental Security Income Longitudinal File (SSI-LF), the CER100% files, the MBR, the Payment History Update System (PHUS), the DBAD file, the DCF, EN Payment Cumulative Payment Report, the VRRMS, 831 & 832/833, and NUMIDENT. These sources are more fully described in Volume 9.

- SSDI SGA disability cessation (SDSn)
- SSI medial decision appeal date (T16APPLn)
- SSI benefit entitlement date (T16STARTn)
- SSI benefit cessation date (T16STOPn)

Beginning in May 2009, SSA added historical occurrences of some of the key information in the MBR. This was a change from earlier years, in which MBR records were overwritten as new information became available. For example, the TOC field indicating the Type of Claim, or reason that the beneficiary is receiving benefits, would have been overwritten as a beneficiary's status changed. Of particular importance is that prior to May 2009, SSA would have overwritten the TOC field to reflect retirement instead of disability for an SSDI beneficiary who reached Full Retirement Age (FRA) and transferred to OASI.⁴ For example, since May 2009, the TOC field in the MBR is historical, retaining both the current and previous TOC values, stored in the *n-variable* fields. In this example, it would be possible to identify both SSDI and OASI benefit receipt using the TOC if that transition occurred after May 2009. To better account for historical occurrences of OASDI receipt, the DAF includes data on TOC and the related Beneficiary Identification Code (BIC) derived from snapshot files before 2009; see section II.E in Volume 3 for more details.

Variables from the 831 & 832/833 files. Variables from the 831 & 832/833 files contain disability data from initial medical determinations, appeals, and medical Continuing Disability Reviews (CDRs).⁵ Data in the 831 file represent the initial medical determination and any reconsiderations, as well as a limited set of elements related to further appeals, for both SSI and SSDI applicants. While the full 831 file includes all medical determinations, the DAF will only include 831 records for those who have received SSA disability benefits. Once someone has received disability benefits since 1996, they will be in the DAF, and all of their 831 records will be included. As a result, the 831 data will include denials for those with both allowances and denials across multiple applications, but it will not include anyone who only had denials. Further, because the 831 is a medical determination file, it will not include technical denials for anyone since these occur before starting the medical determination process and are not in 831 data. The 832 portion of the 832/833 file contains data from the CDR process for SSI beneficiaries, and the 833 portion of the 832/833 file contains CDR data for SSDI beneficiaries. The CDRs are generally scheduled to be conducted every 3 or 7 years, although the schedule on which they actually occur can be erratic. Each decision, whether initial or CDR, is stored as a separate record in the 831 & 832/833 files.

To convert the data in these files to records that conform to the DAF format, we merged all the records for each beneficiary and constructed variables with multiple occurrences to fit in the

⁴ Because the TOC variable is of such importance to understanding SSDI benefits, we have constructed historical TOC (and Beneficiary Identification Code, or BIC) data for the period before May 2009 using data from other administrative sources. These variables and the process for constructing them are described in Volume 3.

⁵ There is a separate process that applies to certain SSDI beneficiaries who return to work that is known as a work CDR. The data in the 831 & 832/833 files is not related to this process.

DAF’s one-row-per-beneficiary structure. For example, if a beneficiary has one initial determination and three subsequent decisions, there are four records in total on the 831 & 832/833 files, and the DAF will have four occurrences of the variables DODECn (Date of Decision). The same would be true for other variables related to disability determinations.

Some 831 & 832/833 variables are used to build time-based variables, such as education level, EDXyymm, where “yymm” represents a year and month. The construct of such variables is best explained with an example. Table III.1 shows a beneficiary with three records in the 831 & 832/833 files, each associated with a different DODEC value. To generate monthly occurrences for EDXyymm, we sort the records by ascending date (DODEC) and use that date to determine which month-year field to update. We copy the value for ED (9) from the first record into the appropriate EDXyymm field for June 1997, e.g., EDX9706. We copy ‘11’ into EDX0009 and ‘15’ into EDX0308. Intervening occurrences of EDXyymm are forward-filled with the previous value, e.g. EDX9707 through EDX0008 is set to ‘9’, and EDX0010 through EDX0307 is set to ‘11’. We finish up by applying the last value of ED (15) to all subsequent occurrences of EDXyymm from September 2003 onward.

Table III.1. Example of using the 831 & 832/833 files to build time-based variables

Record number	DODEC: Date of decision	RDT: Result of determination	ED: Education years
1	June 1997	Continuance	9
2	September 2000	Continuance	11
3	August 2003	Continuance	15

These 831 & 832/833 files came into use beginning in 1988, meaning information will be missing for beneficiaries who applied for benefits prior to this time as will any data collected from pre-1988 CDRs completed for such individuals. Additionally, even among applicants since 1988, information is not available for all beneficiaries. Several variables will have missing information unless the information contained in the variable was necessary to complete the determination or review. For example, education, industry, and occupational information are only relevant to a determination for those beneficiaries whose application reached Step 5 of the sequential determination process.⁶ Those allowed at Step 3 (if they meet or equal the impairment listings), are not asked to provide education, industry, and occupational information because it is not relevant for the decision. Consequently, data on such variables will be missing for the large segment of beneficiaries whose benefits were allowed at step 3.

⁶ The sequential determination process is: Step 1 decisions are denials only--working at SGA; Step 2 decisions are denials only--impairments do not meet the severity requirements for eligibility; Step 3 decisions are allowances only--the applicant meets the criteria at steps 1 and 2 and has an impairment that meets or equals SSA’s Listing of Impairments. Step 4 decisions are denials only--applicants who meet the criteria from the first three steps, but whose residual functional capacity means they are found able to perform past work. Step 5 decisions include both allowances and denials based on whether the applicant’s residual functional capacity in conjunction with age, education and work experience mean that the applicant is found able to perform any work in the United States economy. While education and job information will be relevant at Step 4, Step 4 can only result in a denial so that information will not be in the DAF, because the DAF only includes those who were allowances and became beneficiaries.

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

The CER100% and Disabled Beneficiary and Dependents Extract (DBAD) files will also tend to undercount the number of beneficiaries in the SSI and SSDI programs. This is because SSA usually awards benefits retroactively. For example, a person awarded benefits in June of 2012 that are retroactive to January will not show up in the CER files for January through May of 2012. They will first show up in the July 2012 CER file. The SSR in July will now show benefits due but not paid in January through May of 2012 and will likely show a lump sum payment in July to adjust for these earlier missed payments. While retrospective updating occurs for both the SSI and SSDI programs, it is less of an issue for SSDI than SSI because of the rules governing each program. Still, researchers need to use caution in using and interpreting the data from these snapshot files; the detail pages in Volume 5 indicate the source for each variable for users to consult.

IV. Details About the DAF Payments Component

In this section, we provide more information about working with the DAF.Payments component. As we described in Volume 1, the component is comprised of two sets of files, the first relates to payments to ENs and State VR agencies acting as ENs under the Ticket to Work program, while the second relates to cost reimbursement payments to State VR agencies.

A. EN payments files (TTW payments to ENs or State VR Agencies acting as ENs—two files)

The first set of files in the DAF.Payments component relates to payments made by SSA to ENs and State VR Agencies acting as ENs under the TTW program. These data come from the Cumulative Payment Report by EN provided by MAXIMUS, Inc., SSA's contractor responsible for TTW payment administration, and are retained in two files that present the same data in different formats. The first file includes a record for each payment made, called the Vertical Payments File. In the Vertical Payments File, each record represents a Milestone or Outcome payment made to a provider and may include multiple records for beneficiaries. The file includes payments made between 2002 and the end of the DAF year. The second file includes one record per beneficiary who generated a payment, called the Horizontal Payments File. In the Horizontal Payments File, individual payment records made to a provider for services rendered to an SSDI or SSI beneficiary are rolled up into a single record per beneficiary. This component includes person-level observations of beneficiaries who generated at least one Milestone or Outcome payment made between 2002 and the end of the DAF year.

The Horizontal and Vertical Payments Files contain essentially the same data, reflecting the potential for different user needs. Researchers interested in beneficiary-level data would be more interested in the Horizontal file, while those interested in per-payment data would likely want to start with the Vertical file. Reflecting their different uses, the specific variables included in each of the two EN Payment Files varies slightly. In the Horizontal file, each record contains beneficiary-level variables including SSN, the total amount of payments made to the provider for that beneficiary, the type of payment, the payment triggering event date, the payment amount, the payment date, and the type of payment, i.e., Milestone or Outcome. Variables are named according to the month in which the payment triggering event occurred. For example, the generic variable name EN_MPAYAMTyymm represents "Milestone payment amount" while EN_MPAYAMT0702 is the specific variable name for "Milestone payment amount—[triggered by work activity in] 2007 February." In the Vertical file, each record contains variables with information about each payment transaction to a provider including the SSN of the beneficiary generating the payment, the amount of the payment, the triggering event date, the payment date, and type of payment.

B. VRRMS file (payments related to SSA cost-reimbursement to State VR Agencies—two files)

The second set of files in DAF.Payments contains data about traditional cost-reimbursement payments made by SSA to State VR Agencies. The DAF data cover reimbursement payments

made from January 1994 through December 2021. The records are drawn from the Vocational Rehabilitation Reimbursement Management System (VRRMS), an SSA administrative file that records data about Federal/State Vocational Rehabilitation program (VR) claims for payment from SSA for beneficiaries who received VR services, have closed cases, and for whom the State VR Agencies believes worked at Substantial Gainful Activity (SGA) levels for 9 months within a 12-month period and therefore qualify for reimbursement from SSA.⁷ Each record in the DAF VRRMS file is at the beneficiary level and identified by the beneficiary's SSN.

Beginning in March 2017, SSA redesigned and reformatted the VRRMS data. VRRMS data from January 1994 through February 2017 are recorded in the DAF using one file structure, while VRRMS data from March 2017 onward use a different structure. Both files contain information that allows for analysis of payments at the beneficiary level and by state VR Agency, but beyond that, there are conceptual differences in the data elements included in the two files and substantial differences in how the data are stored. As such, the new variables are not simply renamed versions of variables from the prior VRRMS file and so cannot be used in that way. As a consequence, and to highlight the differences in the meaning of the data, the variable naming conventions differ across the two VRRMS versions.

In what follows, we describe the layout for each of the VRRMS versions separately. Care must be taken to align the information across the files; the variable information contained in Volumes 4 and 5 and the source file layouts in Volume 10 can be highly beneficial in these cases. In Volume 5, the detail pages for VRRMS variables indicate both the source and the applicable years for a particular variable.

Though not all of the VRRMS data are new, the source has not been widely used by DAF researchers and therefore has not been as heavily scrutinized as other DAF components. For this reason, we suggest that users proceed cautiously when working with these data, and report any anomalies, particularly across the new and old file types, to ORDES.DAF@ssa.gov.

1. VRRMS Data through February 2017

Each beneficiary VR participant can have multiple service spells with a state VR agency, and each spell can have multiple claims (and payments). There are two primary ways that a beneficiary could have multiple payments in the VRRMS file.⁸ First, the beneficiary could have records for multiple spells of VR participation. Second, for the same VR spell, there could be multiple payments made from SSA to the state VR agency. In the VRRMS data through February 2017, payments were rolled up to the spell level. That is, beneficiaries with multiple payments in the same period of participation (i.e., in the same spell) have a single value indicating the sum of the payments across all records for that period of VR participation.

The variable `COUNT_SPELLS` indicates the total number of spells, or periods, of VR service for a beneficiary on which the State VR Agency made a claim for, and received reimbursement

⁷ Though the VRRMS source file contains data related to claims for payment, the VRRMS sub-file in the DAF includes only payments reimbursed.

⁸ Again, the VRRMS source data contain information on claims, meaning that there could be claims made that are not paid, and therefore would not be captured in the DAF VRRMS file.

from, SSA. For example, when COUNT_SPELLS = 3, the beneficiary had generated a State VR Agency claim on three separate spells of VR service. In the DAF, nearly 92 percent of beneficiaries have only one spell and more than 98 percent have two or fewer. The maximum number of spells for any beneficiary on the file is nine.

For beneficiaries with multiple spells of VR participation, the DAF VRRMS file contains all the variables for the most recent spell. For all earlier spells, selected key variables are retained: C_CLM_ENTEREDn, C_DISP_DTE n, C_EMP_BEGANn, C_RPB_RECVDn, C_SVRA_CLOSEDn, C_SVRA_ENTERn, and TL_ALLOWEDn. That is, *all* data are retained only for the most recent period of VR participation.

2. VRRMS Data from March 2017 forward

Like the old version of the VRRMS data, the new version also records multiple claims across multiple spells of VR participation. In the new version of the files, however, information is retained at the claim level, rather than aggregating all claims to the spell level. As is the case in the DAF VRRMS prior to March 2017, we retain data only on claims for which a reimbursement was made.⁹

Reflecting the claims-level structure of the new version of the DAF VRRMS, the ‘n’ suffix indicates the claim with the earliest payment recorded as having n=1, and then numbered sequentially from 2 onward. The variable NCLAIM indicates the total number of paid claims generated by a given beneficiary; NSPELL indicates the number of spells. To construct a spell-level payment value similar to what was available in the older version of the DAF VRRMS, the variable VR_CLNT_ENTRY_DT (start date of VR spell) can be used to identify reimbursements associated with the same spell. For example, if claims n=1 to n=4 all have VR_CLNT_ENTRY_DT=May 5, 2015, then these claims are all related to the same spell. If there is a second spell for this client, they could have, say, claims n=5 to n=7 all with a second VR_CLNT_ENTRY_DT that is after May 5, 2015, and so on.

As was the case in the previous version of the VRRMS, more information is retained for the most recent data than earlier claims. There are 16 variables that are retained for all claims, suffixed with ‘n’ (Table IV.1), and about three dozen other variables that are only retained for the most recent claim (see Volume 4 for a list of these variables).

⁹ In DAF19, after consultation with SSA, we modified the process we had used in DAF17 and DAF18 to collapse records to one per claim. Specifically, we added a step to combine records from the Claims and Payments files first by COSSN and VR_CLM_SNUM, before collapsing to the COSSN level. As a result of this method, the DAF aligns very closely with SSA’s published statistics (see Volume 6 for more details), but DAF19 has fewer claim occurrences (7) than it would have if we had continued processing records as we did in the previous two DAF cycles (11).

Table IV.1. Variables Available for Each VRRMS Claim in File Covering the Period from March 2017 Onward

Variable	Description
VR_CLM_SNUM	Systems assigned unique identifier for the VR claim
NPAYSEQ	Number of payment sequence records included in the VR total variables
VR_CLMBSS_CD	Indicates the basis for the VR claim
VR_CLNT_ENTRY_DT	Date the individual entered VR
SSA_RCPDT	Date on which SSA received the VR claim
PSTUS_TS	VR Claim Payment Status Timestamp
TOT_PAYBL_AMT	The systems calculated amount of the final VR claim payment amount, if allowed
TOT_DUE_AMT	The total amount approved for the VR claim
TOT_PAID_AMT	The amount payable on the VR claim after all overpayments were recovered
CLNT_EMPT_STDT	Date the employment began for VR spell
IPE_SGN_DT	The date the client and VR agency signed the Individual Plan for Employment (IPE)
ST_VR_ABBR_CD	The State VR agency to whom the VR payment is being made
VR_FNL_CLOSR_DT	Date the individual VR case was closed
VR_PMT_ACTN_CD	Code which reflects SSA's disposition of the VR claim
VR_PRD_ENDT	The earlier of the State VR provided date of final closure and the SSA-determined payment period end date
VR_PRD_STDT	The later of the State VR provided date of client entry into VR or the SSA determined payment period begin date.

V. Using DAF Data Marts

Due to file size and current computing limitations, the DAF is often too cumbersome for simple analyses. To make the DAF more accessible to researchers, SSA contracted with Mathematica to construct two “Data Marts,” or subsets of records from DAF that were created based on criteria often of interest to researchers:

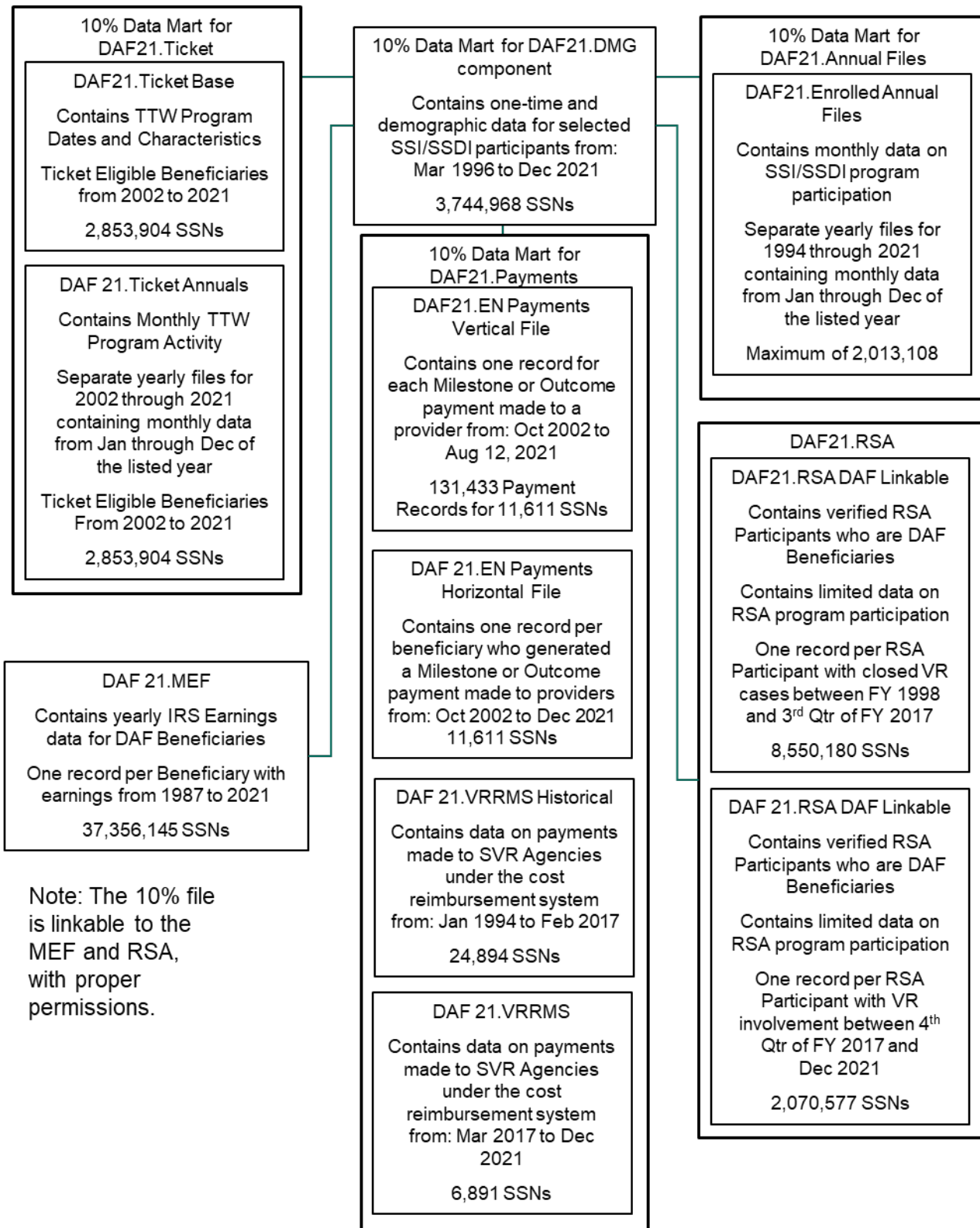
- **The DAF 10% Data Mart** is a ten percent sample of records based on the complete DAF. Its file structure and data components are identical to those described for the DAF and make it ideal for testing programs and analyses prior to running on the full DAF, if necessary. It can also be used to conduct analyses when the subpopulation sample size is large enough.
- **The DAF ADM** includes all the variables from the full DAF DMG component file as well as constructed variables related to the date of first payment and month in which eligibility was first obtained. It only contains beneficiaries whose first SSDI payment as an adult or whose first child or adult SSI payment was between 1996 and the DAF year. It therefore does not contain records for all beneficiaries on the DAF. One benefit of the Awardee Data Mart (ADM) is that researchers will not have to go through the somewhat lengthy process of determining first non-retiree adult payment before conducting their research.

A. DAF 10% Data Mart

The DAF 10% Data Mart has an identical structure to the full DAF, but with a file size approximately one-tenth the size (Figure V.1). Data elements contained in the complete DAF are detailed in Volumes 4 and 5, and the DAF 10% Data Mart contains the same data elements. For researchers interested in analyzing data in the complete DAF, the DAF 10% Data Mart will reduce computing time, and still contain a representative sample of all beneficiaries, though sample sizes may be problematic for rare subgroups. It can be used independently for analysis, or for testing programs on a smaller sample of data before being run on the complete DAF.

To create the DAF 10% Data Mart, we select beneficiaries from each DAF component based on the value of the eighth and ninth position of beneficiaries’ nine-digit SSN. Similar selection algorithms are used within SSA to create 10 percent extracts from other administrative files, but the DAF 10% Data Mart files are not expected to match every 10% extract file created by SSA. Information about SSA files that match to the DAF 10% Data Mart is available upon request.

Figure V.1. Layout of the DAF21 10% data mart components



Note: The 10% file is linkable to the MEF and RSA, with proper permissions.

Note: The 10% file is linkable to the MEF and RSA, with proper permissions.

Table V.1 displays the total number of records in each component of the DAF21 10% Data Mart and the percent of records from the complete DAF21. As expected, the application of the selection criteria resulted in very close to 10 percent of the records in the full DAF. To ensure comparability of the 10% Data Mart to the full DAF, each year we compare basic demographic and disability statistics from the two files to verify that the differences between the population in the DAF and the 10% file are not significant, a process discussed in more detail in Volume 6.

Table V.1. Number and percent of records in the DAF21 10% data mart

File	N (DAF21 10% Data Mart)	N (complete DAF21)	% Records
DMG component	3,744,968	37,486,887	10.0
Annual files* (1994 to 2021)	956,890 to 2,013,108	9,578,805 to 20,140,142	10.0
Ticket Base	2,853,904	28,570,366	10.0
Ticket Annual files	2,853,904	28,570,366	10.0
EN payments (horizontal)	11,611	116,519	10.0
EN payments (vertical)	131,433	1,317,390	10.0
VRMS (through Feb.. 2017)	24,894	248,321	10.0
VRMS** (after March2017)	6,891	68,780	10.0

*Due to the removal of beneficiaries from the Non-Enrolled Annual files, the number of observations varies across the Annuals files. Note that as described in Volume I, the Ticket Annual and Base files may include a small number of records not in the DMG file, reflecting differences in when the source files are accessed.

**Our selection algorithm for 10% files reaches a 10 percent sample with large enough sample size, but because the VRMS file is small, it is not exactly 10 percent.

While the 10% Data Mart allows for similar analyses as the complete DAF, the reduction in number of observations can be potentially problematic in certain contexts. For example, certain subpopulations, such as beneficiaries with earned income, may be prohibitively small for certain research questions. It is also important to be mindful that the DAF 10% Data Mart was not constructed using a stratified sampling technique and therefore, the Data Mart may not be representative of subpopulations.

B. DAF ADM

The DAF ADM includes all the variables from the full DAF DMG component file and is constructed based on a beneficiary's SSDI or SSI award date, which is defined in the ADM as the month of the beneficiary's first receipt of SSDI or SSI payment.¹⁰ More specifically, it contains only beneficiaries whose first SSDI payment as an adult or whose first child or adult SSI disability payment was received between 1996 and the DAF year, provided that the beneficiary's age at the time of first payment was under FRA. Beginning in DAF20, child and

¹⁰ In other words, the ADM award date is *not* defined in terms of month of eligibility. There is frequently a gap, sometimes quite large, between first month of eligibility and first month of payment because of the time it takes to make a determination that the beneficiary is disabled. We use first payment month rather than first eligibility month for ADM award dates because this is the earliest month at which eligibility affects the beneficiary financially, but the ADM also contains variables based on first month of eligibility, as described below.

adult SSI awardees are separately identified in the ADM, and a given beneficiary may be identified as both a child and adult awardee; more information on how we identify each type of award is available below.

The ADM excludes DAF beneficiaries whose first payment was prior to 1996.¹¹ It is important to note that our evaluation for inclusion in the ADM is independent for each program, SSDI and SSI. This means a beneficiary qualifies for inclusion in the ADM when he or she meets the criteria in *either* program, regardless of his or her award status in the other program. For example, a beneficiary who received his first non-retiree adult SSDI award in May 2001 will be included in the ADM, even if that beneficiary had his first SSI award in July 1991. They would not have an SSI award date in the file but would have an SSDI award date. Further, a beneficiary who first received a first payment for SSI in March of 1998 and then received a first non-retiree adult SSDI award in November of 2004 will be in the ADM and will show these ADM first-award dates for both SSI and SSDI and the ADM variable “Overall Award Month” (OVERAWARDDT) will show the earlier of these two dates (i.e., March 1998). Researchers wanting to identify the first month of pre-1996 eligibility across programs (e.g., to remove those with earlier periods of eligibility in the other program) can use DOEI to identify the first month of SSDI eligibility and SSIELIG_FIRST or SSIELIG_ADULT to identify the first month of SSI eligibility (overall or as an adult). Those dates can be compared to ADM award dates to identify earlier eligibility spells.

Because of the exclusions to the file, the ADM contains millions fewer records than the full DAF, which is not insignificant when considering the reduction in computing resources needed to conduct research on this group. Because this data mart includes all the variables from the full DAF DMG component file, it can be used in place of the DMG component file when constructing an analysis file that is limited to select beneficiaries.

Prior to the availability of the ADM, identifying first SSDI and SSI award dates required extracting data from each of the DAF Annual files as well as processing the SSI Longitudinal files for 1974 through 1993. Researchers interested in first benefit cohort analyses using award date information can avoid the necessity of processing the DAF Annual files and/or SSI Longitudinal files, thus saving time and computing resources, by using the DAF ADM.

The ADM includes all the variables from the full DAF DMG component file and additional ADM-specific variables (as noted above, award in these variables means first *payment*):

- Date of first adult SSDI award (DIAWARDDT)
- Age at first adult SSDI award (DIAWARDAGE)
- State of residence at first adult SSDI award (DIAWARDST)

¹¹ For SSI beneficiaries, the ADM excludes any beneficiary who was 18 or older, was in current pay status, and received a payment in the same month in any month from January 1973 through December 1995. For SSDI beneficiaries, the ADM excludes any beneficiary who (1) had an initial period of eligibility before 1994 and who has a subsequent period of eligibility, or (2) who was 18 or older, in current pay status, and received a payment in the same month in any month from January 1994 through December 1995. The SSDI exclusion criteria reflect the availability of SSDI data in conjunction with time period of DAF data.

- Date of first child SSI award (SSICHLDAWARDDT)
- Age at first child SSI award (SSICHLDAWARDAGE)
- State of residence at first child SSI award (SSICHLDAWARDST)
- Date of first adult SSI award (SSIADLTAWARDDT)
- Age at first adult SSI award (SSIADLTAWARDAGE)
- State of residence at first adult SSI award (SSIADLTAWARDST)
- Date of earliest adult SSI or SSDI award (ADLTAWARDDT)
- Age at earliest adult SSI or SSDI award (ADLTAWARDAGE)
- State of residence at earliest adult SSI or SSDI award date (ADLTAWARDST)
- Date of earliest SSI or SSDI award, including child SSI awards (OVERAWARDDT)
- Age at earliest SSI or SSDI award, including child SSI awards (OVERAWARDAGE)
- State of residence at earliest SSI or SSDI award date, including child SSI awards (OVERAWARDST)
- Date of first adult SSDI eligibility (DIELIGDT)
- Age at first adult SSDI eligibility (DIELIGDT) (DIELIGAGE)¹²
- Date of first child SSI eligibility (SSICHLDELIGDT)
- Age at first child SSI eligibility (SSICHLDELIGAGE)
- Date of first adult SSI eligibility (SSIADLTELIGDT)
- Age at first adult SSI eligibility (SSIADLTELIGAGE)
- Date of earliest adult SSI or SSDI eligibility (ADLTELIGDT)
- Age at earliest adult SSI or SSDI eligibility (ADLTELIGAGE)
- Date of earliest SSI or SSDI eligibility, including child SSI eligibility (OVERELIGDT)
- Age at earliest SSI or SSDI eligibility, including child SSI eligibility (OVERELIGAGE)
- Flag for Disabled Adult Child (DAC) whose Date of Initial Entitlement is before the DAC's 18th birthday (DIAWARDDAC)
- Flag for beneficiaries who received SSI payments as a child (SSICHILD)
- Flag for beneficiaries who were receiving SSI child payments at the time of their 18th birthday (SSICHILDAT18)
- Flag for beneficiaries who received their first non-retiree adult SSI payment during the immediate 5 months preceding receipt of his/her first non-retiree adult SSDI payment (POSSERIAL)

¹² To clarify, the name of this variable is "DIELIGAGE" and "(DIELIGDT)", the other variable to which it refers, is part of DIELIGAGE's SAS label.

- Flag indicating that a SSDI beneficiary has a DOEI before 1994, has not had a subsequent period of eligibility (DOEI=DOEC), was 18 or older at the time of DOEI, but whose first observed non-retiree adult payment (DIAWARDDT) was in 1996 or after (SUSPECT)

1. Method Used to Identify Award Date

The variables indicating award date in the ADM are based on identifying the first payment received, either as an adult (for SSDI), or as a child or adult (for SSI). Due to programmatic differences and data element availability between SSDI and SSI, the criteria for determining first payment as an adult differs slightly (Volume 11 contains the SAS/JCL code developed for SSDI and SSI beneficiaries).

SSDI Beneficiaries. The receipt of the first SSDI payment as an adult under FRA is determined by examining the status codes Ledger Account File (LAF) from DAF for current pay status (beginning with a “C” or “E”) combined with the payment received field (PAYD) showing a benefit amount greater than zero in or after: (1) the month the beneficiary turns 18, or (2) the month the beneficiary was deemed eligible for benefits (DOEI), or (3) January 1994 (the earliest available data); whichever came latest.¹³

Once first payment receipt as an adult is established, we exclude beneficiaries who received their first payment as an adult after attaining FRA and those that received their first payment in 1994 or 1995, as the data mart range starts in 1996.¹⁴ Also excluded are SSDI beneficiaries who turned 18 before 1994 and had dates of initial entitlement before 1994 as it is likely his/her first payment was received before 1996. Likewise, we exclude beneficiaries whose DOEI is before 1994 and who have a subsequent period of eligibility, identified as beneficiaries who have a date of current entitlement (DOEC) that is greater than DOEI. Finally, we exclude those who were FRA or older at the time of first payment as an adult.

SSI Beneficiaries. We separately identify first payment as a child, if any, and first payment as an adult. First payment receipt as an adult under FRA is determined by examining status codes (PSTA) for current pay status (equal to “C01”, “M01”, “M02”), combined with the payment receipt field (PAYS) showing a benefit amount greater than zero starting in the first month after the beneficiary turns 18 or January 1974, whichever is later. This information is from the DAF for years 1994 and later and from the SSI Longitudinal file for years 1974 through 1993.¹⁵ Once first payment as an adult under FRA is established, beneficiaries whose first payment as a non-

¹³ Approximately 1.6% of SSDI beneficiaries received SSA retirement benefits prior to their entitlement to SSDI, which we determined using payment data in conjunction with the TOC variable from the MBR. For these beneficiaries, this method incorrectly identifies their retirement award date as their SSDI award date. Researchers interested in a population that include beneficiaries who may have received a retirement benefit prior to their first disability-based payment may want to use the SSDI eligibility-based award variable (DIELIGDT) described below, which is not subject to this issue.

¹⁴ Note that FRA varies according to each beneficiary’s birth date.

¹⁵ As there is no reliable variable available on the DAF indicating the date of initial entitlement to the SSI program it is necessary to use the data from the SSI Longitudinal file for years 1974 through 1993 to ensure that the first non-retiree adult benefit observed on the DAF is actually the first ever received by the beneficiary. As the SSDI data provides a reliable date of initial entitlement using the historical pre-1994 data, no additional data for that program was necessary.

retiree adult under FRA was before 1996 are excluded. We use the same logic to identify first child payment except that the payment must have occurred prior to the first month after the beneficiary turned 18.

If a beneficiary was receiving SSI payments as a child at the time of his or her 18th birthday, he or she would be subject to an age 18 redetermination to establish adult eligibility. The ADM includes these beneficiaries with an adult SSI award date constructed in the same manner as for non-child beneficiaries regardless of whether an age 18 redetermination has occurred. In other words, in some cases, the adult award date for SSI may not identify new awards, but rather the month in which a child SSI beneficiary reaches age 18, regardless of whether or not an age-18 redetermination has begun. As a result, the adult SSI award date will be the first month after the beneficiary's 18th birthday, provided a payment was received in that month. These beneficiaries are identified with a flag, SSICHILDAT18.

SSI applicants who have certain conditions may be awarded presumptive benefits for up to six months while awaiting a disability determination decision. While most frequently favorable, resulting in benefits continuing, that decision is in some cases unfavorable. No attempt was made to identify these probationary SSI beneficiaries whose applications ended up being denied, so the SSI award date for these beneficiaries will be the first presumptive disability payment they received as a non-retiree adult even if that application for which a presumptive determination was made results in a denial. This results in seemingly illogical data in which the record would show a person receiving payments but never being allowed benefits; this information seems odd but is in fact correct.

While oftentimes SSDI and SSI are received concurrently, there are certain situations in which they are received in sequence. One more common occurrence is the receipt of SSI in the period before becoming entitled to SSDI benefits for the first time. Beneficiaries must wait five full months after their disability onset date before becoming eligible for their first SSDI payment. Some beneficiaries, called serial beneficiaries, may receive SSI during this five-month waiting period before becoming ineligible for SSI once their SSDI payments start. These beneficiaries are flagged in the ADM using the variable POSSERIAL, which indicates that the first adult SSI payment received was in the immediate five months preceding the first non-retiree adult SSDI payment. Beneficiaries who receive SSI payments during SSDI's five-month waiting period will have an SSI award date corresponding to the first SSI payment.

Beneficiaries who were entitled to serial SSI payments but who did not receive those payments until after they began to receive monthly SSDI benefit payments will not have an adult SSI award date in the ADM if they were no longer eligible for SSI after the serial period. This is because although their payment status (PSTA) in the serial months would retroactively indicate current pay status, they would not have received a payment in those months and therefore would not meet the criteria for having an adult SSI award date assigned. If the serial beneficiary later becomes re-eligible for and begins receiving SSI payments after they start receiving SSDI payments, the SSI Adult Award Date in the ADM will reflect the month he or she began receiving the later SSI payments.

Concurrent Beneficiaries. Beneficiaries who received both SSDI and SSI will usually have both an SSDI and SSI award date. The one exception is serial SSI beneficiaries (as described above) who did not receive a first SSI payment until after they were already permanently ineligible for SSI due to SSDI benefit receipt, and who were never eligible for SSI benefits again. Concurrent beneficiaries are not separately identified in the ADM because the definition of concurrent is dependent upon the research question. However, two data elements, capture the earliest of the SSDI and SSI award dates: ADLTAWARDDT, which does not consider child SSI awards, and OVERAWARDDT which does include child SSI awards.

2. Method for Identifying Eligibility Date

The ADM also includes ten variables based on first eligibility month and age at first eligibility month. These variables include DIELIGDT and DIELIGAGE for SSDI, SSIADLTELIGDT and SSIADLTELIGAGE for first adult SSI award, and SSICHLDELIGDT and SSICHLDELIGAGE for first child SSI award. Additionally, ADLTELIGDT and ADLTELIGAGE are for the earliest eligibility across both programs considered together but limited to first eligibility at age 18 or older. Finally, OVERELIGDT and OVERELIGAGE allow first eligibility to include eligibility for child SSI benefits. These variables may be preferable to researchers in some research contexts.

Like the award date variables, the calculation we made to determine eligibility date was specific to the program under which benefits were received. To calculate DIELIGDT, we looked at the disability onset date, applied the five-month SSDI waiting period, and then looked for the first month in which the LAF code started with C or E and a non-zero benefit was due (based on DUED). To calculate first SSI eligibility, we looked for the first month in or after the disability onset date in which PSTA was C01, M01 or M02. In both cases, the algorithm for identifying first eligibility month followed the relevant program rules. The overall eligibility-based ADM variables are simply the earliest of the two program-specific variables. In all cases, the first eligibility month is equal to or earlier than the award date as defined by payment, as would be expected.

VI. Using DAF Extracts

In addition to the data marts described above, the DAF also contains three simple extracts to facilitate analysis on subpopulations of the full DAF. Because these files are simply extracted from the larger DAF, no special rules exist for working with these files and the information presented in this volume also applies to the extract files.

A. TTW participant extract

The TTW participant extract is a set of DAF component files (DMG component, Annual, Ticket, Payments) limited to those who had a Ticket assignment date (TKTASGNDDTn) occurring after December 31, 2005. This collection of files contains all the variables in the full DAF. A frequency for the latest Ticket assignment date for each beneficiary is included in Table VI.1 to show the range of Ticket dates in the extract as well as the distribution by month and year. See Appendix A for a memorandum that provides detailed information about how these files are constructed and their contents.

Table VI.1. Frequency of latest Ticket assignment dates (by month, by year)^a

Month	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022
January	4,358	5,059	6,058	5,584	5,981	6,369	6,584	6,629	8,402	8,123	8,601	8,732	9,258	10,326	6,499	7463
February	4,154	5,085	5,799	4,958	5,879	6,912	10,366	6,506	8,906	8,905	8,287	8,254	8,538	9,332	6,579	7413
March	4,342	4,508	5,604	6,990	6,885	8,907	6,187	7,031	8,392	9,586	9,346	8,983	9,208	7,917	7,521	8543
April	3,831	4,524	5,183	6,322	6,205	9,200	7,385	8,543	8,472	8,518	8,377	8,801	9,534	7,640	7,684	6790
May	4,748	4,680	4,729	6,154	7,074	7,441	6,571	7,193	8,009	8,978	9,416	8,943	10,073	6,699	7,034	3365 ^b
June	4,657	4,484	5,147	7,525	5,987	8,985	5,850	6,599	8,158	9,108	9,525	9,125	9,937	7,063	8,153	
July	4,379	5,247	5,198	6,711	5,452	7,654	6,412	10,8392	8,402	8,070	8,385	8,678	9,281	6,448	7,907	
August	5,411	5,320	5,465	6,721	6,784	7,491	7,930	8,883	9,079	9,584	10,201	10,093	9,691	6,675	8,425	
September	3,510	5,951	5,330	6,073	9,025	6,164	8,034	7,611	8,002	8,320	8,168	8,163	9,197	6,754	8,287	
October	4,804	5,424	5,618	6,161	5,811	7,934	8,884	7,885	8,662	8,215	9,116	9,445	10,164	6,710	8,301	
November	4,207	4,420	4,670	5,983	5,314	6,151	6,186	6,481	7,428	7,980	8,070	7,969	8,336	5,939	7,614	
December	3,868	5,781	5,121	5,419	64,218	5,439	5,814	8,009	8,082	7,908	7,479	7,418	8,730	6,375	7,016	
Total	52,270	60,483	63,922	74,601	137,377	87,647	86,203	92,209	99,994	103,295	103,2	104,604	111,947	87,878	91,020	

Note: Total Obs =1,441,012

^a Because the data for the DAF Ticket component is extracted later in the year than data are from other SSA administrative data sources (MBR, SSR, and 831) the available Ticket counts extend beyond the end of the DAF year.

^b Due to the timing of the extract and data lags, this figure represents a partial month.

B. NBS extract

The National Beneficiary Survey (NBS) extract is a set of DAF component files (DMG, Annual, Ticket, Payments as well as RSA and MEF-linkable file) limited to those beneficiaries identified in a finder file of NBS participants provided by SSA. This finder file contains all NBS participants in any of the 7 rounds of the survey; in DAF21, this included R1 (2004), R2 (2005), R3 (2006), R4 (2010), R5 (2015), R6 (2017), and Round 7 (2019). This collection of files contains all the variables in the full DAF as well as a set of NBS LINKID variables, unique identifiers assigned to participants in each round of the NBS survey.

In DAF21, we used the NBS finder file provided by SSA in February 2020. This NBS finder contained 42,881 records and is duplicated by SSN. When we merged the finder to some components of DAF, not all finder records were found in all DAF components. Specifically, from the DMG component, we extracted 42,878 records, a shortage of three records, and from the Annuals component, we extracted 42,876 records, a shortage of five records. From the Ticket component, we extracted 42,284, a shortage of 597 records. We believe these discrepancies stem from three factors:

1. The selection criteria for inclusion in the NBS beneficiary samples and the DAF database were different in that the NBS selection criteria drew from the CER100% and DBAD for a single month while the DAF selection criteria drew from a span of months, usually a year. It is possible that a beneficiary who was in the CER100% or DBAD for only a single month may not have made it into the SSR or MBR files from which the DAF was ultimately constructed.
2. The DAF component containing data related to TTW participation (TKT) is built from SSA's DCF data which is subject to periodic update and removal of records by SSA. It is possible that records were selected from the DCF that were later deleted from the DCF and thus also from the TKT component of the DAF. These deleted records would still be in the NBS participant sample.
3. Fewer records may have been extracted from the Annuals component than from the DMG component as a result of anomalies in the application of DAF selection criteria and possible anomalies in the NBS finder process. It is possible that mismatches involving the Claimant Account Number (CAN) of a deceased primary beneficiary and the Beneficiary's Own Account Number (BOAN) of an auxiliary DAC or widow(er) are a contributing factor.

The Payments files also contain fewer records than NBS respondents, but those differences are due to differences in program participation rather than to sampling differences.

See Appendix B for a memorandum that provides detailed information about how these files are constructed and their contents. Further documentation on the NBS, including instructions for how to access and linking to the NBS Public Use Files with the NBS_LINKID, can be found at <http://www.ssa.gov/disabilityresearch/nbs.html>. Information on the MEF linkable file for the NBS is provided below in section B, MEF DAF-linkable file with DAF components. Information

on accessing the NBS Restricted Use File, which can be linked to using SSN, can be found at <http://www.ssa.gov/disabilityresearch/publicusefiles.html>.

C. Demonstrations and surveys extract

The Demonstrations and Surveys extract is a set of DAF component files (DMG, Annual, Ticket, Payments) limited to those beneficiaries who participated in one or more of the following SSA demonstrations and surveys. Use of these data are restricted to projects that meet the privacy and disclosure restrictions as disclosed to the participants in these data collections. Contact ORDES.DAF@ssa.gov for assistance in obtaining permission to use these files. The demonstrations and surveys include:

- Accelerated Benefits Demonstration (AB)
- Benefits Entitlement Services Team (BEST) demonstration project;
- Benefit Offset National Demonstration (BOND);
- Benefit Offset Pilot Demonstration (BOPD);
- Homeless Outreach Projects and Evaluation Demonstration (HOPE);
- Mental Health Treatment Study (MHTS);
- National Survey of SSI Children and Families (NSCF);
- Ohio Direct Referral demonstration project (ODRD);
- Promoting Opportunity Demonstration (POD);
- Promoting Readiness of Minors in SSI (PROMISE);
- Retaining Employment and Talent After Injury/Illness Network (RETAIN) projects;
- Supported Employment Demonstration (SED);
- Youth Transition Demonstration (YTD)

Participants were identified in a finder file provided by SSA. This collection of files contains all the variables in the full DAF as well as flags for each demonstration or survey in which that individual participated. These variables are AB_FLAG, BEST_FLAG, BOND_FLAG, BOPD_FLAG, HOPE_FLAG, MHTS_FLAG, NSCF_FLAG, ODRD_FLAG, POD_FLAG, PROMISE_FLAG, RETAIN_FLAG, SED_FLAG, and YTD_FLAG. Table VI.2 shows the frequency of beneficiaries in each sub-sample in the finder file; not all records in the finder file match to beneficiaries in the DAF. See Appendix C for a memorandum that provides detailed information about how these files are constructed and their contents.

Table VI.2. Frequency of participation flags for de-duplicated combined demonstration and surveys extract finder files in DAF21

Survey or Demonstration	Flag Variable	Number of Beneficiaries with Indicator
Accelerated Benefits Demonstration	AB_FLAG	2,005
Benefits Entitlement Services Team Demonstration	BEST_FLAG	1,183
Benefit Offset National Demonstration	BOND_FLAG	994,099
Benefit Offset Pilot Demonstration	BOPD_FLAG	1,838
Homeless Outreach Projects and Evaluation Demonstration	HOPE_FLAG	6,941
Mental Health Treatment Study	MHTS_FLAG	2,238
National Survey of SSI Children and Families	NSCF_FLAG	9,242
Ohio Direct Referral Demonstration	ODRD_FLAG	663
Promoting Opportunity Demonstration	POD_FLAG	10,070
Promoting Readiness of Minors in SSI	PROMISE_FLAG	13,444
Retaining Employment and Talent After Injury/Illness Network	RETAIN_FLAG	771
Supported Employment Demonstration	SED_FLAG	2,972
Youth Transition Demonstration	YTD_FLAG	6,177

VII. Using the DAF-Linkable Files with DAF Components

There are several additional administrative files that, while not part of the core DAF components, are often used in conjunction with those components to study SSDI and SSI beneficiaries. These data often require additional permissions beyond those required for accessing the core DAF and thus are described as DAF-linkable files.

A. DAF-RSA files

Note: Use of the DAF-RSA files is permitted only for projects that have permission from the Department of Education to use the RSA data. Contact ORDES.DAF@ssa.gov for assistance in obtaining permission to use these files.

Because many SSA disability beneficiaries seek services from State VR Agencies in setting and achieving work goals, administrative files containing information on State VR service receipt are of special interest to many researchers working with the DAF. For this reason, SSA and RSA have developed files that simplify the merging of the DAF with the Rehabilitation Services Administration Case Service Report (RSA-911) files. We build a number of files using the RSA-911 data in combination with the DAF; we refer to these files collectively as the DAF-RSA files. We describe these files in what follows, after a brief description of the RSA-911 files.

1. *Summary of Changes in RSA-911*

As described in Volume 1, the RSA-911 files changed significantly starting with data covering the period July 2017 forward. Prior to that time, the data were annual fiscal year files that contained information collected by State VR Agencies about all individuals whose cases were closed during the fiscal year, meaning they had applied for VR services at some point (and may or may not have received VR services before their case was closed). Because they were closure files, a time lag was necessarily built into the files because their data was not recorded until the case closed. For example, a beneficiary who began receiving VR services in 2004 and had her case closed in fiscal year 2006 will not be present on the 2004 or 2005 RSA source files because the case was not closed until fiscal year 2006.

Beginning in July 2017 and reflecting changes in VR service provision resulting from the Workforce Innovation and Opportunity Act (WIOA), the structure of the RSA-911 files fundamentally changed. RSA-911 files are now prepared quarterly,¹⁶ and contain records on all cases open at any point in that quarter (both cases that closed in the quarter as well as cases that opened or remained open). Additionally, the universe of records includes students who are receiving pre-employment transition services (pre-ETS) through the VR agency, even if not formally receiving VR services.¹⁷ Clients receiving pre-ETS only have a select set of information recorded. Finally, the data elements in the new version of the RSA-911 data are significantly different than those in the earlier version. While it may be possible to use some of the elements

¹⁶ DAF-RSA files are still processed on an annual (calendar year) basis.

¹⁷ Clients who receive only pre-ETS are not required to report SSN, so many of these records do not contain SSN. Without SSN, records are of little value for purposes of matching to the DAF.

across both the pre-July 2017 and July 2017-onward files to measure similar concepts, the Department of Education has not documented a crosswalk that formalizes the relationship across variables, and thus DAF users should proceed with caution.

In addition to the major change in the RSA-911 data in 2017, there are other smaller changes to the file (in 2002, 2014, and 2020). Those interested in analyzing RSA-911 data should consult Volume 12 for information on the RSA source file, including agency instructions for form completion by the State VR agencies, and descriptions of changes in the source file structure across versions. Even when using the three variations of the pre-July 2017 RSA-911 data, special attention should be paid to variables that may have changed over time, including those potentially that were dropped or added, changed their format type, or had differing valid values. In July 2020, RSA discontinued use of more than 90 variables that were added in July 2017, while adding another 14 new variables. Volume 12 details the variables that were modified and motivation for the changes made by RSA.

2. Guide to Using DAF-RSA files

There are three main types of DAF-RSA files that researchers might want to consider. In each case, there is one file structure for the period from 1998 to June 2017 and a second file structure from July 2017 onward. Because each of the DAF-RSA files meets slightly different research needs, users should consult the descriptions below to identify the file most appropriate for their situation.

It is important to note that the data from July 2017 onward include both open and closed VR cases, compared to earlier data which contained only records once they were closed. A critical implication of the difference is that in the earlier data, each record was unique; a VR spell could only close in a single year. In the more recent data, the same VR spell may exist in multiple years of data. For instance, a VR applicant in January 2018 would have data in the 2018 VR file and would have another record in the 2019 file, whether the spell was ongoing, or even it had closed (as RSA-911 reporting requirements indicate that closed cases remain on the file for eight quarters after closure). Based on our analysis of the data, we recommend that users identify unique VR spells in the post-July 2017 data using PIN-Application date-Agency Code-Individual ID. To look across years, users should first append selected multiple annual files together, identifying the most recent record for each spell using the latest REPORT_THROUGH date.

Table VII.1 highlight features of each of the files under the new and old layouts. Volume 12 contains additional information about how these files are constructed and their contents, as well as source file layouts for the RSA-911 files. In each of these files, we substitute a de-identified Personal Identification Number (PIN) variable for the SSN variable. Records are included only if the RSA SSNs are verified by the Enumeration Verification System (MPR-EVS) process.¹⁸ After this verification takes place, SSA de-identifies the files and creates and applies the PINs.

¹⁸ Mathematica worked with SSA staff to develop a verification system for instances in which a beneficiary's name is unavailable, as is the case in RSA-911. Using this EVS process, we validate the SSNs from the RSA-911 data by comparing the SSN, gender, and year of birth from the RSA data with the SSN, gender, and year of birth from the NUMIDENT.

SSA maintains an SSN-PIN crosswalk file. The PINs are specific only to the DAF-RSA linkable files and allow users to link RSA-911 files with PINs to other DAF files with PINs, in the same way they would if they had SSN available to them. Other files can be linked to the RSA-911 files if SSA first applies PINs to these files using the SSN-PIN crosswalk file.

Table VII.1. File structure of the RSA-DAF Files under the new and old layouts

File	Pre-July 2017 Layout	July 2017 onward layout
DAF-Linkable RSA Summary File	One file containing data from fiscal year 1998 through June 2017. Seven variables, each with up to 24 occurrences, and PIN for a total of 169 variables.	One file containing data from July 2017 – December 2021. ^a Eight variables, each with up to 8 occurrences, and PIN for a total of 64 variables.
RSA-911 Verified SSN Files	One file for each fiscal year from 1998 through 2017. ^b	One file for each reporting year from 2018 ^c 2021
RSA Mini DAF	54 DAF Component file extracts (1 DMG, 28 Annuals, 21 Ticket, 4 Payment)	54 DAF Component file extracts (1 DMG, 28 Annuals, 21 Ticket, 4 Payment)

^a As we describe above, the July 2017-onward file includes open VR cases and retains closed cases for 8 quarters after closure. Because of this, VR clients may have records in multiple files. To construct the linkable summary file, we identify the most recent record (using the REPORT_THROUGH variable) for the observation, identified using PIN-Application date-Agency Code-Individual ID. We assume that records that match on these values are the same case, with updated data for purposes of constructing this closure file.

^b The fiscal year 2017 file ends in June 2017.

^c The 2018 file contains data covering the period from July 2017 through December 2018, reflecting a time of transition across layouts to get more updated information from RSA.

The DAF-RSA files include:

- **DAF-Linkable RSA Summary File.** This file contains select RSA data for *all* RSA participants with an SSN validated by the MPR-EVS. In other words, this file is not limited to beneficiaries in the DAF, though as the name suggests, can be linked to any DAF component. In this file, records are stored horizontally as a single record per PIN. Because records with missing SSNs do not receive a PIN, this file does not include information on RSA cases that received pre-ETS services only and did not have an SSN recorded. It contains seven series of variables, with the number of occurrences dependent on the file layout (see Table VII.1). These variables are described and mapped to the RSA source data in Table VII.2 below. Permission from the Department of Education is required before using these files.
- **RSA-911 Verified SSN Files.** For researchers who need more detail about the nature of provided State VR services, these files might be ideal. They contain the full RSA-911 data for all RSA participants who have an SSN verified by MPR-EVS.¹⁹ These files do not contain any DAF data; they are the full RSA-911 files, only subject to SSN verification by

¹⁹ One exception is that the Illinois gender field from 2002-2007 contained an error that had to be corrected in DAF processing. Illinois gender data during this time period therefore does not match what is in the original RSA-911 data.

MPR-EVS.^{20, 21} Permission from the Department of Education is required before using these files.

- **RSA Mini-DAF.** The Mini-DAF is identical to the full DAF in terms of component files and most variables,²² but it is limited to RSA participants whose SSNs are verified by the MPR-EVS process and match to a record in the DAF. These files do not contain any of the RSA data that is in the two RSA files above. Instead, these files contain DAF data only on RSA participants. Researchers can link the RSA Mini-DAF to the DAF Linkable RSA Summary File or the RSA-911 verified SSN file using PIN variable also contained in the files above.

²⁰ All elements are converted to SAS variables, including some that are converted to multiple SAS variables (such as the element “PUBLIC SUPPORT DURING VR,” which is converted to separate binary variables for each type of public support that are 0/1 switches indicating whether each type of public support was or was not received).

²¹ The variable IN_DAF, indicates whether the RSA participant is also in the DAF (value of 1 indicates the RSA participant has a record in the DAF, otherwise value is 0).

²² The file excludes certain PII such as first and last name.

Table VII.2. Variables in the RSA-DAF linkable summary file

Variable name	Variable label	RSA source element ^a
Variables in the RSA-911 files prior to July 2017		
RSAAGYn ^b	RSA Agency	Agency Code (pre-2002 #2, 2002-onward #1)
RSACLTYPn	RSA Closure Type	Type of Closure (pre-2002 #38, 2002-2013 #36, 2014-onward # 213)
RSADOAn	RSA Date of Application	Date of Application (#5)
RSADOCn	RSA Date of Closure	Date of Closure (pre-2002 #40, 2002-216 #38, 2014-onward #215)
RSADOEn	RSA Date of Eligibility Determination	Date of Eligibility Determination (pre-2002 #25, 2002-2013 #22, 2014 - onward #46)
RSAIPEn	RSA Date of Individualized Plan for Employment	Date of Individualized Plan for Employment (pre-2002 n/a, 2002-2013 #23, 2014-onwards #49)
RSASVSn ^c	RSA Service Use Flag	Type of Services Provided (pre-2002 #31, 2002-2014 #25, 2014 - onwards ##51-190)
Variables in the RSA-911 files from July 2017 to December 2021^d		
RSAAGYn	RSA Agency	Agency Code (July 2017-onward #4)
RSAEXTTYPn	RSA Exit Type	Type of Exit (July 2017-onward #354)
RSADOAn	RSA Date of Application	Date of Application (July 2017-onward #7)
RSADOEXTn	RSA Date of Exit	Date of Exit (July 2017-onward #353)
RSADOEn	RSA Date of Eligibility Determination	Date of Eligibility Determination (July 2017-onward #38)
RSAIPEn	RSA Date of Individualized Plan for Employment	Date of Most Recent or Amended IPE (July 2017-December 2019 #48), Date of Initial IPE (January 2021 -onward #398)
RSADOPREm	RSA Date of Pre-Employment Transition Services	Date of Pre-Employment Transition Services (July 2017-onward #96)
RSASVSDTn	RSA Start Date of Initial VR Service	Start Date of Initial VR Service on or After IPE (#127)

^a Numbers in the "RSA Source Element" column, e.g. #1, indicate the element number in the original RSA source file.

^b Pre-2002 records have an Agency Code field that is 3 positions long but contains only the two-digit Agency Code from the original data.

^c Pre-2002: if any of the 13 Types of Services were received during the rehabilitation process (there is a 1 in that record position) then this variable is set to 1; otherwise, this variable is set to 0. 2002-2013: if any of the 22 Types of Services were received during the rehabilitation process (there is something other than a '00' in that record position) then this variable is set to 1; otherwise, this variable is set to 0. 2014-2016: if any of the 28 Types of Services were received during the rehabilitation process (there is something other than '0' in that record position) then this variable is set to 1; otherwise, this variable is set to 0.

^d See the footnote to Table VII.1 for an explanation on how we combine data across years to construct these measures.

To accommodate non-SAS users, the Local Area Unemployment Statistics (LAUS) and Small Area Income and Poverty Estimates (SAIPE) information in the RSA Mini-DAF is available in

variable format (rather than in SAS format libraries as on the full DAF).²³ DERSEC8.DER_MEF_PROCESSING_WIDE (from 1994 through the DAF year),²⁴ each containing one record per RSA participant (with a verified SSN) along with the annual county-level poverty rate and annual county-level median income data from the SAIPE, and monthly county-level unemployment rates from the LAUS. The statistic for each participant is based on the county of residence for the RSA participant during each month. Due to limitations of county code availability in earlier years, the LAUS unemployment rates are not available in the 1994 or 1995 annual files and become available in the 1996 annual file for March 1996, and the SAIPE estimates become available starting in the 1997 annual file. More information about these measures as incorporated into the DAF are available in Volume 3, Section X “Geographic Measures in the DAF.”

To summarize, all of the DAF-RSA files contain information on RSA participants whose SSN contained in the RSA-911 files have been verified using MPR-EVS. Two file types (DAF-Linkable RSA Summary File and RSA-911 Verified SSN Files) contain RSA-911 data for all RSA participants, regardless of whether they are contained in the DAF, but do not incorporate data from the DAF. The other file type (RSA Mini-DAF) contains DAF data, but is limited to those receiving RSA services. All files contain PINs rather than SSNs and can be linked to other files with these PINs but cannot be merged to files with SSNs. However, SSA can apply PINs to non-DAF files using the SSN-PIN crosswalk file so that those files can be merged by PIN to the RSA-911 files. Permission from the Department of Education is needed to access files containing RSA-911 data.

B. MEF DAF-linkable file with DAF components

Note: Use of this data is available only to selected staff at SSA.

The Master Earnings File (MEF) contains annual earnings information based on Internal Revenue Service (IRS) Form W-2, quarterly earnings records, and annual income tax forms. These data include regular wages and salaries, tips, self-employment income, and deferred compensation (contributions or distributions). SSA accesses these data via section 6103 of the Internal Revenue Code, under which SSA can use it only to record wages in their systems and cannot share it with any other entity including other federal agencies. As a result, these data are not available to researchers outside of SSA. Outside researchers interested in merging this information with the DAF components need to seek permission from SSA and work with an SSA staff member with permission to access this file. The variables in the MEF DAF-linkable file are listed in Table VII.2.

²³ This information is found in the OPDR.TG.PRD.ETTW.RSAVER.DAF21.Lyyyy.PIN files. Note that the same type of economic data is available in the full DAF, but is provided as SAS formats as described in Volume 1. In this case, to facilitate research by staff at the Department of Education, we built files providing this information as variables instead of SAS formats. The content is the same as for the full DAF, just in a different format and limited only to RSA participants.

²⁴ The 1996 SAIPE data is not available at the county-level. SAIPE variable values for 1996 are therefore missing, and a table of state-level SAIPE data can be found in Chapter VII of this volume.

Table VII.3. Variables in the MEF DAF-linkable file

Variable name	Variable description	SAS label
ALLEARNyyyy	The total annual earnings, including any positive self-employment earnings and wages from all employers, during the year (excludes payments from Third-Party Payers).	TOTAL NON-TPP EARNINGS (TLWAGES+SELF-EMPL) IN YEAR (1987-2021)
FEINyyyy	The Federal Employer Identification Number of the sole employer, or the employer with the highest wages in the year when there are multiple employers.	FEDERAL EMPLOYMENT IDENTIFICATION NUMBER (FEIN) OF PRIMARY (HIGHEST\$) NON-TPP EMPLOYER IN YEAR (1987-2021)
LAST_DER_YR	The last year after 1987 in which positive earnings from wages or self-employment were recorded in the MEF.	LAST YEAR OF EARNINGS IN DER DATA
N_EMPLRSyyyy	The total number of employers in the year (excludes Third Party Payers).	NUMBER OF NON-TPP EMPLOYERS IN YEAR (1987-2021)
N_TPPEINyyyy	The total number of Third-Party Payers in the year.	NUMBER OF TPP EINS IN YEAR (1987-2021)
PRWAGEyyyy	The annual earnings from the sole employer in the year or the employer with the highest wages in the year when there are multiple employers (excludes payments from Third-Party Payers).	WAGES PRIMARY (HIGHEST\$) NON-TPP EMPLOYER IN YEAR (1987-2021)
SE_EARNyyyy	Net positive earnings from self-employment for the year. Net negative earnings from self-employment (a net loss), are not recorded in the MEF.	SELF-EMPLOYMENT EARNINGS IN THE YEAR (1987-2021)
SSN	Social Security Number	SOCIAL SECURITY NUMBER
TLWAGEyyyy	The annual earnings from the sole employer in the year or the total in wages across all employers in the year when there are multiple employers (excludes payments from Third-Party Payers).	TOTAL WAGES NON-TPP EMPLOYERS IN YEAR (1987-2021)
TPPFEINyyyy	The Federal Employer Identification Number of the sole Third Party Payer, or the Third Party Payer with the highest total transfer payments in the year when there are multiple Third Party Payers.	FEDERAL EMPLOYMENT IDENTIFICATION NUMBER (FEIN) FOR HIGHEST TPP EIN IN YEAR (1987-2021)
TPPWAGEyyyy	The total annual transfer payments from the sole Third-Party Payer in the year or the total transfer payments across all Third-Party Payers in the year when there are multiple Third-Party Payers.	EARNINGS FROM ALL TPP EINS IN YEAR (1987-2021)
WORKyyyy	A binary indicator of work during the year (1=working, 0=not working). The individual is indicated as working if either wage employment or self-employment earnings are positive in the year. Transfer payments from Third Party Payers do not indicate work.	BINARY WORK INDICATOR (1=WORKING, 0=NOT WORKING) FOR THE YEAR (1987-2021)

Note: For more information about the MEF DAF-linkable file, see Appendix D of this volume.

We also create a version of the MEF that aligns with the NBS extract files. It has all of the MEF variables but is limited to those SSA beneficiaries who participated in one of SSA's NBS rounds.

VIII. Using Information on Continuing Disability Reviews in the DAF

SSA periodically reviews the records of disability beneficiaries to confirm that they remain eligible for benefits. This review process is called a “continuing disability review” (CDR) and occurs in one of two forms. First, medical CDRs are done on a periodic basis depending on SSA’s assigned expectation of medical improvement for a beneficiary.²⁵ For those with medical improvement expected, CDRs are scheduled at intervals from 6 to 18 months following the most recent determination. For those with medical improvement possible, CDRs are scheduled at least once every three years, and for those with medical improvement not expected, CDRs are scheduled once every five to seven years. Medical CDRs are also performed for SSI child recipients after they reach age 18, called age-18 redeterminations, to determine that the individual’s medical conditions meet the adult disability criteria which differ from the child disability criteria and remain significant enough to maintain benefits. The CDR process may begin with a mailer, typically for beneficiaries with low probability of medical improvement, to gather information about medical conditions, recent treatment, as well as work activity. SSA uses the information received with the mailer response to determine whether the beneficiary should proceed with a full medical review or defer the medical review to a later date.

Second, work CDRs are done when SSA receives information (via self-report, a mailer response, or IRS earnings reports) about a beneficiary’s earnings that suggest he or she is engaging in substantial gainful activity (SGA) and therefore no longer meets the program’s eligibility requirements.

A. Overview of the DAF standalone CDR file

Starting in DAF19, based on user suggestions, we incorporated additional information about CDRs. This information has been incorporated into a single standalone CDR file, available at OPDR.TG.PRD.ETTW.FINAL.DAF21.SCDR. Starting in DAF21 we added additional information to the standalone file from the DCF Medical table. Any beneficiary in the DAF with a CDR derived from one of two sources (described below) has a record in this file, even if the CDRs predated 1994. We have prepared this information as a separate file to allow user testing and analysis before incorporating into the full DAF. If you work with the information and identify questions or comments, please email ORDES.DAF@ssa.gov.

The information in the standalone CDR file comes from two sources, SSA’s Waterfall File and the DCF, which overlap to some extent, but each contains distinct information. We verified when building the file for the first time in DAF19 that the data we collected for the DAF aligns with available statistics published by SSA, as documented in Volume 6.

- Data from the **Waterfall File** includes 21 variables with information culled from several SSA sources on centrally-initiated full medical reviews for SSDI and SSI beneficiaries who received a full medical review from 1993 to present.²⁶ The Waterfall File does not include

²⁵ Program Operations Manual System (POMS) Part 04 / Chapter 280 / Subchapter 01, DI 28001.020.

²⁶ The Waterfall File is commonly used by SSA to support review and research related to CDRs. It is used because its record layout makes it easier to work with than pulling information from multiple SSA files.

information on mailer CDRs that did not progress to a full medical review. Information from the Waterfall File in the DAF includes all levels of adjudication relevant to the particular CDR from the initial review up through the federal district court, if cases were appealed to that level. The Waterfall file contains information on SSI child recipients as well as SSDI and SSI adult beneficiaries.

- Data from the **DCF Review and Event Tables** includes seven variables with information on initial decisions for all CDRs from 1981 to present, with centrally-initiated claims distinguished from others using the REVIEW table variable ORIG_CD. It contains data on both work (SSDI) and medical (SSDI and SSI) CDRs, including mailer-only CDRs that did not progress to a full medical review. Variables in the file allow users to determine whether the CDR was work or medical. The DCF data also includes information on SSI child recipients as well as SSDI and SSI adult beneficiaries.
- Data from the **DCF Medical Table** includes 11 variables that characterize the most recent review in the Review table including the date that the diary was established and the date that the scheduled diary will mature. In addition, the new variables provide information on the type of diary, the diary reason, residual functional capacity, regulation basis code, and whether the diary is permanent.

To construct the DAF CDR file, we include only CDRs that have concluded (those with a non-missing value for “CDR end date” (ENDT in the Review table). The DCF tables include information on events that occurred throughout the initial determination (EVNT_CD in the Event table), though we did not retain all of the intermediate information in that file for the DAF. Instead, per guidance from SSA, we identify the final event associated with the CDR using the latest event sequence number (EVNT_SEQ_NUM) in the Event table.

B. Structure and contents of the standalone CDR file

The DAF CDR file has the typical one-row-per-beneficiary structure used in the rest of the DAF. To convert records into that structure, we linked records based on Beneficiary’s Own Account Number (BOAN) in the Waterfall file and Claimant’s Own Social Security Number (COSSN) in the DCF.

Because of the number of CDRs that each beneficiary has, and the information retained on each CDR, the file is relatively wide. In the Waterfall File through 2021, the maximum number of CDRs is 7; we retained information from all seven occurrences. For a beneficiary with all seven occurrences populated, 175 variables might be populated (21 variables for each of seven occurrences). The total number of occurrences in the Waterfall file is recorded in the variable CDR_WFAL_NUM. The DCF includes a broader range of CDRs and spans a longer period than the Waterfall File; as such, it has more CDRs recorded. The 99th percentile of beneficiaries had 12 or fewer CDRs, so we retained the twelve most recent occurrences.²⁷ The total number of CDRs that a beneficiary had in the DCF is reported in the variable CDR_DCF_NUM. For a

²⁷ This means that the earliest of CDR records for the top 1 percent of beneficiary CDR counts will be missing, but data on the 12 most recent occurrences will be retained.

beneficiary with all 12 CDR occurrences populated in the DCF, 84 variables might be populated (7 variables for each of twelve occurrences). Over time, the number of CDRs that each beneficiary has will increase, and we expect that as future DAF versions are built, we will retain additional occurrences of these variables. Variables from the DCF Medical table are associated with the most recent review and are populated for beneficiaries whose own SSN matches the controlling SSN established during the application process for a Social Security administered program.

The variables in the DAF standalone CDR file that are derived from each source are listed and described in Table VIII.1. We retained information in the DAF for CDRs even if they preceded the start of monthly data in the DAF in 1994; this is consistent with other one-time information in the DAF recorded in ‘n’ variables. In addition to information in this volume, each of the variables in the DAF CDR file are listed with their own detail pages in Volume 5; we recommend that in addition to this section, users consult those pages before they begin to work with information in the DAF standalone CDR file.

Table VIII.1. Variables in the standalone DAF CDR file

DAF variable name	Description (from the source file)
SSN	Beneficiary's Own Account Number
Variables from the Waterfall File	
CDR_WFAL_NUM	Number of CDRs in the Waterfall file
CDR_WFAL_ROMn	Date the mailer was received
CDR_WFAL_TYPEn	Type of CDR (direct release, mailer, etc.)
CDR_WFAL_WRMn	Why the review was made (from the 833 file)
CDR_WFAL_INIT_STRDTn	Date that the state DDS received the record (from the 833 file)
CDR_WFAL_INIT_DODECn	Date of initial decision (from the 833 file)
CDR_WFAL_INIT_RDTn	Result of initial determination (from the 833 file)
CDR_WFAL_ND_STRDTn	Date that the file was received by the DDS, no decision
CDR_WFAL_RCN_STRDTn	Date that the file was received at the reconsideration level
CDR_WFAL_RCN_DODECn	Date of reconsideration decision
CDR_WFAL_RCN_RDTn	Result of reconsideration determination
CDR_WFAL_ALJ_RCPTDTn	Date that the file was received at the Administrative Law Judge (ALJ) level
CDR_WFAL_ALJ_DODECn	Date of ALJ decision
CDR_WFAL_ALJ_RDTn	Result of ALJ determination
CDR_WFAL_AC_DODECn	Date of the Appeals Council (AC) decision
CDR_WFAL_AC_RDTn	Result of the AC decision
CDR_WFAL_DC_DODECn	Date of the District Court (DC) decision
CDR_WFAL_DC_RDTn	Result of DC determination
CDR_WFAL_ALn	Highest adjudicative level
CDR_WFAL_FIN_ACTn	Final Action
CDR_WFAL_FIN_ACT_DTn	Date of Last (Final) Action

DAF variable name	Description (from the source file)
Variables from the DCF Review and Event Tables	
CDR_DCF_NUM	Total number of CDRs in the DCF (only records for the 12 most recent retained)
CDR_DCF_REV_TYPn	Review type, identifies the type of CDR, for example: T2 medical, T16 medical, T2 work, etc.) (Review table)
CDR_DCF_REV_DETn	Review type detail, indicates further breakdown of review type, some categories are unique to a specific review type (Review table)
CDR_DCF_ORIG_CDn	Origination code, used to distinguish centrally-initiated claims from others (Review table)
CDR_DCF_STDTn	Start date of CDR (Review table)
CDR_DCF_ENDTn	End date of CDR (Review table)
CDR_DCF_FIN_EVTn	Event code, the final event in a CDR is most often either continuance or cessation (Event table)
Variables from the DCF Medical Table	
CDR_DCF_MED_STDT	Medical start date
CDR_DCF_MED_TYP	Medical type
CDR_DCF_MED_MDR	Medical diary reason
CDR_DCF_MED_DRY_PERM_CD	Diary permanent code
CDR_DCF_MED_MRE_DRY_MDT	Medical reexamination diary month date
CDR_DCF_MED_LISTG_CD	Listing code
CDR_DCF_MED_VORULE	Vocational rule
CDR_DCF_MED_RBSS	Regulation basis code
CDR_DCF_MED_NDDSS_DT	National Disability Determination Services system date
CDR_DCF_MED_INSERT_TS	Inserted time stamp
CDR_DCF_MED_LU_TS	Last updated time stamp

Note: In DAF21, we combined five variables that indicated whether a decision was made at the initial, reconsideration, Administrative Law Judge (ALJ), Appeals Council (AC), and District Court (DC) level into a single categorical variable (CDR_WFAL_Aln) that specifies the highest adjudicative level at which a determination was made.

The occurrences of CDRs derived from each file follow the usual ‘n’ structure for occurrences. The ‘n’ variables should be used *within* source files (either Waterfall or DCF), but not across source files. In other words, **the ‘n’ value for each source file identifies variables from the same CDR in that file (e.g., Waterfall) but does not align with the same ‘n’ from the other source file (e.g., DCF).** The earliest CDR based on the date of initial decision from the Waterfall file will have n=1, then 2, etcetera. All of the Waterfall-sourced CDRs with n=1 apply to the same CDR in the Waterfall file—the first CDR for that beneficiary in the Waterfall File. Similarly, the earliest CDR from the DCF will have n=1, then 2, etcetera. All of the variables for DCF-sourced CDRs with n=1 apply to the same CDR in the DCF—the first CDR for that beneficiary in DCF. The variables from the DCF Medical table do not have an ‘n’ suffix; they are associated with the most recent review in the Review table.

Of course, it is possible that the n=1 variables will index the same CDR from both files, if a person’s first full medical review (recorded in the Waterfall File) also happens to be their first centrally-initiated CDR (recorded in the DCF). It may be possible to link across CDRs using

SSN and the date of the initial decision, but we have not fully vetted the concordance between the two source files.

C. Working with information from the Waterfall File

For each CDR in the file (denoted with the same ‘n’ value), information is populated from the initial mailer through the highest level of adjudicative review. Each CDR will have information available on the initial mailer, used to initiate the CDR process. This is retained in the variables CDR_WFAL_ROMn, CDR_WFAL_TYPEn, and CDR_WFAL_WRMn.

After the mailer information, the population of information in the Waterfall File depends on the level of review required to make an ultimate decision on the CDR. It may be helpful to recall the [levels for receiving a determination](#) from SSA on a disability claim, including CDRs:

1. **Initial.** This level of review is named with a prefix of CDR_WFAL_INIT in the DAF standalone CDR file.
2. **Reconsideration.** This level of review is named with a prefix of CDR_WFAL_RCN in the DAF standalone CDR file.
3. **ALJ.** This level of review is named with a prefix of CDR_WFAL_ALJ in the DAF standalone CDR file.
4. **Appeals council.** This level of review is named with a prefix of CDR_WFAL_AC in the DAF standalone CDR file.
5. **Federal district court.** This level of review is named with a prefix of CDR_WFAL_DC in the DAF standalone CDR file.

For each level in the determination process, the Waterfall File generally has the following measures:

- Date the file was received at that level of review (with a suffix of STRTD or RCPTDT)
- Date that a decision was made at that level (with a suffix of DODEC)
- Result of that decision (with a suffix of RDT), with values of continuance (C, meaning that benefits continue; S for cessation of benefits; and T for termination of benefits)

The appeals process can be followed for a particular CDR (based on the ‘n’ variable) by looking at the highest level of adjudication (CDR_WFAL_ALn) as well as the combination of decision dates (DODEC) and results (RDT). In general, cases that receive a continuance decision (RDT=C) at one level of the determination process will not appeal their claim to the next level, so we would expect their records for the higher decision levels to be entirely missing. In other words, a case that was decided as a continuance of benefits at the initial level would not have data for the RCN, ALJ, AC or DC variables. Conversely, a case with data for the higher-level variable (e.g., DC), will likely have denial decisions for each of the lower-level decisions. Table

VIII.2 shows the set of codes used to populate the variable indicating the highest adjudicative level (CDR_WFAL_ALn).²⁸

Table VIII.2. Values for Adjudicative Level in the Waterfall File

DAF variable name	Values of ALn
CDR_WFAL_ALn	A=Initial B=Reconsideration (pre-hearing) C=Reconsideration, Disability Hearing Unit (DHU) G=Reopening (very few cases have this value; we recommend recoding these cases to B, but users should review for their particular research purpose) D=ALJ H=ALJ, hearings (we believe this may reflect an earlier version of the ALJ indicator that was used through early 2001; we recommend recoding these to D, but users should review for their particular research purpose) E=Appeals Council J= Indicates that the source variable AL_AC was coded "A" (very few cases have this value, and we believe this may reflect a data entry error for A=Appeals Council; we recommend recoding these to E, but users should review for their particular research purpose) F=District Office

^a Reconsiderations (CDR_WFAL_ALn=B) are for beneficiaries who receive a pre-hearing cessation decision from a Disability Examiner at the initial level and ask for their case to be reviewed by a different Disability Examiner. Disability Hearing Units (CDR_WFAL_ALn=C) review cases when beneficiaries request a hearing following a pre-hearing cessation/termination decision, or SSA reopens the case to change the basis for a cessation or to revise a determination following a favorable decision at the initial or pre-hearing level. As such, a reconsideration under the Disability Hearing Units could also have had a reconsideration of a pre-hearing cessation decision by a different Disability Examiner, but this is not necessarily the case. Both types of decisions are part of reconsiderations, and any less than fully favorable decision from both B and C can appeal to the ALJ level.

D. Working with information from the DCF review and event tables

Users should be aware that the DCF information in the DAF is limited to initial decisions, appeals and higher level decisions are not observed in these data. For each CDR in the file (denoted with the same 'n' value), a user may subset to the type of CDR using the review type (CDR_DCF_REV_TYPn) and review type detail (CDR_DCF_REV_DETn) variables. For example, researchers interested in work CDRs for SSDI beneficiaries could subset to reviews with CDR_DCF_REV_TYPn equal to "W." The majority of CDRs have a final event code indicating continuance (011), cessation (019), or deferral (11, 12, 13, 17, 18, or 19). However, other codes are possible. For example, the final event code for a work CDR will be an administrative stop (023, 033) if the reported earnings are [clearly not SGA](#).

²⁸ The one exception here is that an initial level award could be reopened as a reconsideration case under the Disability Hearing Unit to revise the decision to an unfavorable or only partially favorable determinations based on medical factors.

E. How does information in the standalone CDR file relate to other information in the DAF?

Before adding the standalone CDR file, the DAF already had other information from the 831 & 832/833 files that captured information from the initial stage of the medical review. The DAF retains the nearly six dozen variables we have traditionally included from the 831 and 832/833 files. We have not attempted to reconcile possible inconsistencies among the 823/833 and the new waterfall and DCF files. Of particular note, these include:²⁹

- AGE18RETDT, the date of the age 18 redetermination for SSI
- AGE18_CDR, indicating a successful redetermination
- ALXyymm, indicating the highest adjudicative level for the most recent determination in the 831 and 831&833 files
- BFDn, indicating the basis for the determination
- DDn, indicating the medical CDR date
- DOCn, indicating the district office code for the decision
- DODECn, indicating the date of decision
- DPMn, indicating permanent disability code
- DRn, indicating the reason for medical reexamination
- DTn, indicating the type code for medical reexamination
- JUDVLn, indicating the level at which disability decision was made
- PDn, indicating the presumptive disability decision code
- PDDn, indicating the date of each presumptive disability decision, which is awarded before DDS makes a final determination
- RBn, the regulation basis code documenting the applicant meets or not meet the specific program regulations of SSDI and SSI disability programs
- RDTn, the result of the disability determination
- RFCn, the reason for the continuance or cessation of disability benefits
- RIDn, identifies the program (SSDI or SSI) with which the 831 or 832/833 record is associated

We recommend that users who are interested in redeterminations consider using the standalone CDR file instead of the 832/833 data and suggest caution when working with the latter, as those files may not contain complete information. As is the case with data from the Waterfall and DCF files, **the ‘n’ value from the 831 & 831/833 identifies variables from the same application/redetermination in those files but does not necessarily align with the same ‘n’**

²⁹ Several of the variables in this list also have a “FIRST” version, which includes the information from the first populated occurrence.

from the Waterfall File or DCF. Additionally, as described previously, information from the 831 & 832/833 is limited to data collected from the initial and reconsideration levels of each adjudication. This means that information from the 831 & 832/833 files may not contain the final decision, if the claim was appealed and adjudicated by an ALJ or higher; that information would, however, be present in the Waterfall file.

Appendix A

DAF21 Ticket to Work (TTW) Linkable File Extract

Memo

To: Paul O'Leary and Chris Earles
From: Christian Carrillo Paz
Date: 2/23/2023
Subject: Documentation for DAF21 Ticket to Work Extract (TTW Linkable File)

As described in table VI.1 of volume 2, the data for the Ticket To Work Extract was pulled during DAF21 processing of the Ticket Task. This memo describes the extract you requested, how we created it, and the file name, location, and contents.

I. Description of the Extract

This extract is a linkable file of Ticket to Work (TTW) participants merged with the DAF21 component files (DMG, Annual, Ticket, Payments, VRRMS). A shorthand reference for this extract is the “TTW extract” for DAF21.

II. Methodology

We identified TTW participants as those beneficiaries from the TTW Ticket Base File where the Ticket assigned date (TKTASGNDDT) occurred after December 31, 2005. The resulting finder file is located at:

OPDR.TG.PRD.ETTW.FINAL.DAF21E.TTWTKBS

This finder file was then merged to the DAF21 components to create the extract files. The Ticket components were merged such that there is only one Ticket file associated with the TTW extract. Starting in DAF21, due to space limitations, we had to split the merge of the Enrolled Annuals files so that there are two files associated with the TTW extract. The first file contains annuals information for years 1994-2005, and the second file includes years 2006-2021. All beneficiaries on the TTW extract are on the file associated with the regular Annual files unless they did not appear on any regular DAF Annual file.

III. File Name, Location, and Contents

Several files comprise the TTW extract file and are housed on the SSA mainframe. The file names are listed below. All variables contained in the DAF21 are included in the TTW extract and documentation on these variables is available in Volumes 1 through 5 of the DAF documentation, especially Volume 5, “DAF21 Variable Detail Pages.” All variable names in the TTW extract are identical to those in the DAF21.

- SSN extract file – OPDR.TG.PRD.ETTW.FINAL.DAF21E.TTWTKBS

To: Paul O'Leary and Chris Earles
From: Christian Carrillo Paz
Date: 2/23/2023
Page: 3

- contains the SSNs for records included in the TTW extract and also contains all variables from the Ticket Base component
- DMG Extract file – OPDR.TG.PRD.ETTW.FINAL.DAF21E.TTWDMG
 - contains all variables from the DMG component for SSNs in the TTW extract.
- Annual Extract files
 - OPDR.TG.PRD.ETTW.FINAL.DAF21E.TTWANN.PRE06
 - contains all variables from Annual files from 1994 through 2005 for SSNs in the TTW extract.
 - OPDR.TG.PRD.ETTW.FINAL.DAF21E.TTWANN.PST06
 - contains all variables from Annual files from 2006 through 2021 for SSNs in the TTW extract.
- Ticket Extract File – OPDR.TG.PRD.ETTW.FINAL.DAF21E.TTWTKAN
 - contains all variables from all Ticket Annuals files for SSNs in the TTW extract.
- Payments Horizontal Extract File -
OPDR.TG.PRD.ETTW.FINAL.DAF21E.TTWPYH
 - contains all variables from the Payments horizontal file for SSNs in the TTW extract.
- Payments Vertical Extract File -
OPDR.TG.PRD.ETTW.FINAL.DAF21E.TTWPYV
 - contains all variables from the Payments vertical file for SSNs in the TTW extract.
- VRRMS Extract File (through February 2017) -
OPDR.TG.PRD.ETTW.FINAL.DAF21E.TTWVRRMH
 - contains all variables from the VRRMS file from January 1994 through February 2017 for SSNs in the TTW extract.
- VRRMS Extract File (March 2017 forward) -
OPDR.TG.PRD.ETTW.FINAL.DAF21E.TTWVRRM
 - contains all variables from the VRRMS file from March 2017 forward for SSNs in the TTW extract.

A frequency for the latest Ticket assignment date for each beneficiary is included below to show the range of Ticket dates in the extract, as well as the distribution by month and year.

To: Paul O'Leary and Chris Earles
From: Christian Carrillo Paz
Date: 2/23/2023
Page: 4

Mathematica® Inc.

Table 1: Frequency of Latest Ticket Assignment Dates (by Month, by Year)**Total Obs =1,441,012**

Month	2007	2008	2009	2010	2011	2012	2013	2014
January	4,358	5,059	6,058	5,584	5,981	6,369	6,584	6,629
February	4,154	5,085	5,799	4,958	5,879	6,912	10,366	6,506
March	4,342	4,508	5,604	6,990	6,885	8,907	6,187	7,031
April	3,832	4,524	5,183	6,322	6,205	9,200	7,385	8,543
May	4,748	4,680	4,729	6,154	7,074	7,441	6,571	7,193
June	4,657	4,484	5,147	7,525	5,987	8,985	5,850	6,599
July	4,379	5,247	5,198	6,711	5,452	7,654	6,412	10,839
August	5,411	5,320	5,465	6,721	6,784	7,491	7,930	8,883
September	3,510	5,951	5,330	6,073	9,025	6,164	8,034	7,611
October	4,804	5,424	5,618	6,161	5,811	6,934	8,884	7,885
November	4,207	4,420	4,670	5,983	5,314	6,151	6,186	6,481
December	3,868	5,781	5,121	5,419	64,218	5,439	5,814	8,009
Yearly Total	52,270	60,483	63,922	74,601	134,615	87,647	86,203	92,209

Month	2015	2016	2017	2018	2019	2020	2021	2022
January	8,402	8,123	8,601	8,732	9,258	10,326	6,499	7,463
February	8,906	8,905	8,287	8,254	8,538	9,332	6,579	7,413
March	8,392	9,586	9,346	8,983	9,208	7,917	7,521	8,543
April	8,472	8,518	8,377	8,801	9,534	7,640	7,684	6,790
May	8,009	8,978	9,416	8,943	10,073	6,699	7,034	3,365
June	8,158	9,108	9,525	9,125	9,937	7,063	8,153	
July	8,402	8,070	8,385	8,678	9,281	6,448	7,907	
August	9,079	9,584	10,201	10,093	9,691	6,675	8,425	
September	8,002	8,320	8,168	8,163	9,197	6,754	8,287	
October	8,662	8,215	9,116	9,445	10,164	6,710	8,301	
November	7,428	7,980	8,070	7,969	8,336	5,939	7,614	
December	8,082	7,908	7,479	7,418	8,730	6,375	7,016	
Yearly Total	99,994	103,295	104,971	104,604	111,947	87,878	91,020	33,574

Appendix B

DAF21 National Beneficiary Survey (NBS) Linkable File Extract

Memo

To: Paul O'Leary and Chris Earles
From: Christian Carrillo Paz
Date: 2/23/2023
Subject: Documentation for DAF21 NBS Linkable File Extract

We've completed the NBS extract under the DAF21 Assist Partners task. This memo describes the extract you requested, how we created it, and the file name, location, and contents.

I. Description of the Extract

This extract is a mini-DAF of SSNs contained in the NBS finder file provided by SSA. There are extracts from all DAF21 component files (DMG, Annual, Ticket, Payments, VRRMS). A shorthand reference for this file is the "NBS Extract" for DAF21.

II. Methodology

We kept records for beneficiaries from the NBS finder file provided to us by SSA in February 2022. The file is unique on SSN, but each SSN has more than one link ID (the link ID variables described below). Its file location is:

OPDR.TG.PRD.ETTW.#5413.DAF21.NBSFIND.SA.V1

The NBS finder file was then merged to the DAF21 data to create the extract files. The Ticket and Annual files were merged such that there is only one Ticket file and Enrolled Annual files associated with the NBS extract. All beneficiaries on the NBS extract are on the file associated with the regular Annual files unless they did not appear on *any* regular DAF Annual file. Only five beneficiaries in the NBS finder file did not appear on any regular DAF Annual file.

Note that the NBS extract is structured on a one-record-per-SSN basis. Each person has several link ID variables to indicate sample and round of selection.

III. File Name, Location, and Contents

Several files comprise the NBS Extract file and are housed on the SSA mainframe. The file names are listed below. All variables contained in the DAF21 are included in the NBS extract file. Documentation on the variables in the DAF21 is available primarily in Volume 5 of the DAF documentation, "DAF21 Variable Detail Pages" and also in Volumes 1 through 4. All variable names in the NBS Extract file are identical to those in DAF21

- SSN extract file -- OPDR.TG.PRD.ETTW.FINAL.DAF21E.NBSLIST.SA.V1
 - contains a unique set of SSNs for records included in the NBS linkable file.
- DMG Extract file -- OPDR.TG.PRD.ETTW.DAF21E.NBSDMG

To: Paul O'Leary and Chris Earles
From: Christian Carrillo Paz
Date: 2/23/2023
Page: 3

- contains all variables from the DMG component for SSNs in the NBS linkable file.
- Annual Extract file -- OPDR.TG.PRD.ETTW.FINAL.DAF21E.NBSANN
 - contains all variables from all Annual files for SSNs in the NBS linkable file.
- Ticket Extract File -- OPDR.TG.PRD.ETTW.FINAL.DAF21E.NBSTKT
 - contains all variables from all TKT files (Base and Annuals) for SSNs in the NBS linkable file.
- Payments Horizontal Extract File --
 OPDR.TG.PRD.ETTW.FINAL.DAF21E.NBSPMTH
 - contains all variables from the Payments horizontal file for SSNs in the NBS linkable file.
- Payments Vertical Extract File --
 OPDR.TG.PRD.ETTW.FINAL.DAF21E.NBSPMTV
 - contains all variables from the Payments vertical file for SSNs in the NBS linkable file.
- VRRMS Extract File (through February 2017) --
 OPDR.TG.PRD.ETTW.FINAL.DAF21E.NBSVRRMH
 - contains all variables from the VRRMS file from January 1994 through February 2017 for SSNs in the NBS linkable file.
- VRRMS Extract File (March 2017 forward) --
 OPDR.TG.PRD.ETTW.FINAL.DAF21E.NBSVRRM
 - contains all variables from the VRRMS file from March 2017 forward for SSNs in the NBS linkable file.

In addition to variables from the DAF21, the NBS Extract file contains fourteen additional variables indicating the NBS unique identifier, LINKID. Someone can appear multiple times in the NBS with a different LINKID if they were selected for more than one survey round, if they were selected for both the ticket and the representative beneficiary sample (for rounds 1-4, round 5 had no ticket sample), and/or if they were selected for the clustered and unclustered ticket samples (rounds 1-4).

In the following variables, an ending of _2 means it is the LINKID for the representative beneficiary sample and _1_a or _1_b are the LINKIDs for the ticket sample of the given round (if both of these are populated, it means they were selected for both the clustered and unclustered ticket samples). For Round 7, there is a variable ending in _3, meaning that the LINKID is for the Successful Workers Longitudinal Sample. Here is a full list of the LINKID variables that can be found in each file:

LINKID_R5

To: Paul O'Leary and Chris Earles
From: Christian Carrillo Paz
Date: 2/23/2023
Page: 4

LINKID_R1_2
LINKID_R1_1_a
LINKID_R1_1_b
LINKID_R2_2
LINKID_R2_1_a
LINKID_R2_1_b
LINKID_R3_2
LINKID_R3_1_a
LINKID_R3_1_b
LINKID_R4_2
LINKID_R4_1_a
LINKID_R4_1_b
LINKID_R6_2
LINKID_R6_1_a
LINKID_R6_1_b
LINKID_R7_2
LINKID_R7_3
LINKID_R7_1_a
LINKID_R7_1_b

cc: Dawn Phelps
Matt Urato
Brittany Colip
Janine Bologna

Appendix C

DAF21 Survey and Demonstration Projects Linkable File Extract

Memo

To: Paul O'Leary and Chris Earles
From: Christian Carrillo Paz
Date: 2//23/2023
Subject: Documentation for DAF21 SSA Survey and Demonstration Projects Extract (SDP Extract File)

We've completed the SSA survey and demonstration projects extract under the DAF21 Assist Partners task. This memo describes the extract you requested, how we created it, and the file name, location, and contents.

I. Description of the Extract

This extract is a set of linkable files for three survey and demonstration projects (SDP) merged to the DAF21 component files (DMG, Annual, Ticket, Payments, VRRMS). A shorthand reference for this extract is the "SDP extract" for DAF21.

II. Methodology

We combined SSNs for survey and demonstration project beneficiaries from the BEST, NSCF, and YTD lists provided in DAF12 with additional lists provided in DAF17 of AB, BOND, BOPD, HOPE, MHTS, POD, PROMISE, and SED. For DAF21, we also added-in the SSN lists provided for the ODRD and RETAIN demonstration projects. The resulting finder file is located at:

OPDR.TG.PRD.ETTW.FINAL.DAF21E.DPLIST

The finder file for the survey and demonstration project beneficiaries was then merged to the DAF21 components to create the extract files. The Ticket components were merged such that there is only one Ticket file associated with the SDP extract. Starting in DAF21, due to space limitations, we had to split the merge of the Enrolled Annuals files so that there are two files associated with the SDP extract. The first file contains annuals information for years 1994-2005, and the second file includes years 2006-2021. All beneficiaries on the SDP extract are on the file associated with the regular Annual files unless they did not appear on *any* regular DAF Annual file.

III. File Name, Location, and Contents

One SDP finder file (composed of three SSN lists) and seven DAF component files make up the SDP extract file and are housed on the SSA mainframe. The file names are listed below. All variable names are identical to those in DAF21, except where noted below, with documentation

To: Paul O'Leary and Chris Earles
From: Christian Carrillo Paz
Date: 2/23/2023
Page: 3

on these available in Volumes 1 through 5 of the DAF documentation, especially in Volume 2, “Working with the DAF21,” and Volume 5, “DAF21 Variable Detail Pages.”

- SSN Extract file – OPDR.TG.PRD.ETTW.FINAL.DAF21E.DPLIST
 - contains the SSNs for records included in the SDP extract file.
The SSN Extract file comprises four SSN lists:
 - BEST SSN list – OPDR.TG.PRD.ETTW.N4671.DAF11.EXT.BEST.FL.V1
 - NSCF SSN list – OPDR.TG.PRD.ETTW.N4671.DAF11.NSCF.FL.V1
 - YTD SSN list – OPDR.TG.PRD.ETTW.N4671.DAF11.YTD.FL.V1
 - AB, BOND, BOPD, HOPE, MHTS, POD, PROMISE, SED SSN list – OPDR.TG.PRD.ETTW.#6266.DAF17E.DPLNEW.FL.V1
 - ODRD and RETAIN SSN list – OPDR.TG.PRD.ETTW.#7429.DAF20.DPLNEW.FL.V1

DAF Components

- DMG Extract file – OPDR.TG.PRD.ETTW.FINAL.DAF21E.DPDMG
 - contains all variables from the DMG component for SSNs in the demonstration project list.
- Annual Extract file
 - OPDR.TG.PRD.ETTW.FINAL.DAF21E.DPANN.PRE06
 - contains all variables from Annual files from 1994 through 2015 for SSNs in the demonstration project list.
 - OPDR.TG.PRD.ETTW.FINAL.DAF21.DPANN.PST06
 - contains all variables from Annual files from 2006 through 2021 for SSNs in all demonstration project list.
- Ticket Extract File – OPDR.TG.PRD.ETTW.FINAL.DAF21E.DPTICKET
 - contains all variables from all Ticket files (Base and Annuals) for SSNs in the demonstration project list.
- Payments Horizontal Extract File – OPDR.TG.PRD.ETTW.FINAL.DAF21E.DPPAYH
 - contains all variables from the Payments horizontal file for SSNs in the demonstration project list.
- Payments Vertical Extract File – OPDR.TG.PRD.ETTW.FINAL.DAF21E.DPPAYV

To: Paul O'Leary and Chris Earles
From: Christian Carrillo Paz
Date: 2/23/2023
Page: 4

Mathematica® Inc.

- contains all variables from the Payments vertical file for SSNs in the demonstration project list.
- VRRMS Extract File (through February 2017) --
OPDR.TG.PRD.ETTW.FINAL.DAF21E.DPVRRMH
 - contains all variables from the VRRMS file from January 1994 through February 2017 for SSNs in the demonstration project list.
- VRRMS Extract File (March 2017 forward) --
OPDR.TG.PRD.ETTW.FINAL.DAF21E.DPVRRM
 - contains all variables from the VRRMS file from March 2017 forward for SSNs in the demonstration project list.

In addition to variables from the DAF21, the SDP extract file contains additional variables indicating the SDP sample(s) the beneficiary is from. These variables are AB_FLAG, BEST_FLAG, BOND_FLAG, BOPD_FLAG, HOPE_FLAG, MHTS_FLAG, NSCF_FLAG, ODRD_FLAG, POD_FLAG, PROMISE_FLAG, RETAIN_FLAG, SED_FLAG, and YTD_FLAG. Table 1 below shows the frequency of beneficiaries in each SDP sample.

Table 1: Frequencies from Combined SDP Lists

Survey or Demonstration	Flag Variable	Number of Beneficiaries with Indicator
Accelerated Benefits Demonstration	AB_FLAG	2,005
Benefits Entitlement Services Team Demonstration	BEST_FLAG	1,183
Benefit Offset National Demonstration	BOND_FLAG	994,099
Benefit Offset Pilot Demonstration	BOPD_FLAG	1,838
Homeless Outreach Projects and Evaluation Demonstration	HOPE_FLAG	6,941
Mental Health Treatment Study	MHTS_FLAG	2,238
National Survey of SSI Children and Families	NSCF_FLAG	9,242
Ohio Direct Referral Demonstration	ODRD_FLAG	663
Promoting Opportunity Demonstration	POD_FLAG	10,070
Promoting Readiness of Minors in SSI	PROMISE_FLAG	13,444
Retaining Employment and Talent After Injury/Illness Network	RETAIN_FLAG	771
Supported Employment Demonstration	SED_FLAG	2,972
Youth Transition Demonstration	YTD_FLAG	6,177

cc: Dawn Phelps

Matt Urato

Brittany Colip

Rebekka Nickman

Appendix D

DAF21 Master Earnings File (MEF) Documentation and JCL/SAS Code

Documentation for DAF21 Linkable Master Earnings File (MEF)

This document describes the extract, how we created it, the file name, location, and contents.

I. Description of the Extract

Each year Social Security Administration (SSA) receives Internal Revenue Service (IRS) records showing work and earnings information for U.S. workers. SSA stores this earnings information as the Master Earnings File (MEF). This file contains data derived from IRS Form W-2 (1978-to-date), quarterly earnings records (1951-1977), and annual income tax forms (Form 1040 Schedule SE). These data include regular wages and salaries, tips, self-employment income, and deferred compensation (contributions or distributions). SSA uses this information for administrative purposes in determining benefit eligibility and amounts, but SSA also uses this information for policy analysis and research in accordance with IRS use and disclosure rules. MEF data are not available to contractors or grantees and so are stored separately from the other DAF files. This DAF version of the MEF is a summarized version of the Detailed Earnings Record (DER) from the MEF, with variables similar to those found in the Summary Earnings Record (SER). Descriptions of the SER and DER are provided in the MEF section of the SSA Program Analyst Manual (also known as the “RAND” data manual).

II. Methodology

We extract MEF earnings data for beneficiaries in the current DAF demographic file or any RSA participant with a verified SSN in the RSA-911 files from 1998 to the DAF year. We obtain MEF data through the Office of Research Evaluation and Statistics (ORES) using finder processing. ORES then creates initial extract files under their “bucket” processing. This processing uses combinations of the Earnings Type of Employment (EET) and Earning Type of Record (ERT) in the raw MEF data to combine earnings values into various wage, self-employment, deferred compensation, and pension categories. The bucket process from which we derive this data changed as of 7-15-2011. These changes renamed variables and identified Health Savings Account (HSA) amounts separately from the deferred compensation category. We do not use these deferred compensation values in the DAF matched file so the only effect of the new bucket process on our MEF programming was to change the names of some variables.³⁰

The MEF data include all workers with IRS taxable earnings, regardless of whether they are Social Security or Medicare taxable. The best value to use from the MEF differs according to which taxes apply to the earnings. Both Social Security taxable earnings and Medicare taxable earnings provide possible wage and earnings amounts on the MEF, though neither is entirely satisfactory. Social Security earnings are capped at the Social Security annual taxable limit and so do not include earnings above that limit. Further, federal workers hired before January 1, 1984, railroad workers, and many state and local governments are not covered by Social Security. Medicare taxes are not capped, and so Medicare earnings generally present the best amount for wages or earnings.

³⁰ Prior versions of the DAF linkable MEF file did not include FICA tips in the FICA wage value (from box 3), but the current version includes tips.

However, while most workers are covered by Medicare, including most that are not covered by Social Security, some state and local government employees are covered by neither Social Security nor Medicare.³¹ As a result, both Social Security earnings and Medicare earnings will each miss some earnings. In these cases of earnings not covered by Social Security or Medicare, the Total Compensation amount from the W-2 (Box 1) will provide the best source for earnings. Based on 2010 guidance the SSA's Office of Research, Evaluation, and Statistics, we estimate total earnings from self-employment and each employer in a year as the Maximum of the following earnings computations:

1. Total compensation + deferred compensation + Medicare self-employment + Deferred Payments (this is a negative amount that removes payments for the current year that do not represent work in current year)
2. Social Security taxable wages + Social Security taxable self-employment earnings.
3. Medicare taxable wages + Medicare taxable self-employment.

This method only excludes earnings that are not taxable by IRS or Medicare such as Medical Insurance paid from pretax earnings. In cases where the individual is covered by Social Security but not Medicare, it may be capped at the FICA limit.

We operationalized these computations as follows:

$$\text{BOX1_WAGE} = \text{W2_BOX1_WTOTCMP_NOMISS} + \text{W2_DEF_COMP_NOMISS} + \text{PAYMENT_457_NOMISS};$$
$$\text{BEST_WAGE} = \text{MAX}(\text{BOX1_WAGE}, \text{W2_BOX3_WGE_FICA_NOMISS}, \text{W2_BOX5_WAGE_MED_NOMISS});$$
$$\text{BEST_SE} = \text{MAX}(\text{SEI_FICA_NOMISS}, \text{SEI_MED_NOMISS});$$

Where "NOMISS" means missing values have been set to zero. Non-FICA wages and tips are already counted in wage variables for "BEST_WAGE" so they are not counted separately.

We calculate the variable "BEST_WAGE" for each employer in a given year. Different employers are identified by their Federal Employer Identification Number (FEIN).³² We select the largest of these wage values as the primary wage for the year. Next, we add the wage values for all employers for the year to get total wages for the year. We add total wages and "BEST_SE," the maximum for self-employment in the year, to get total earnings for the year.

In 2012 we learned that the MEF data include values listed as earnings that are actually "third party payer" payments. These are insurance, pension, and other benefit payments that are not wages. SSA has developed a list of such third party payer (TPP) EINs and uses this list to remove earnings alerts that show up as unreported earnings through SSA earnings enforcement activities.

³¹ See Olsen, Anya and Russell E. Hudson, "Social Security Administration's Master Earnings File: Background Information." Social Security Bulletin, Vol. 69 No. 3 (released October 2009)

³² A FEIN is a nine-digit number (for example, 12-3456789) assigned to sole proprietors, corporations, partnerships, estates, trusts, and other entities for tax filing and reporting purposes (See IRS form Social Security-4).

We have examined this issue and believe researchers should not count the earnings from any of these TPP EINS. These values are listed on the MEF as both “covered” and “non-covered” earnings. In many cases the first six months of TPP payments are regarded as “Covered” and FICA taxable. We therefore recommend that users assume all “earnings” from TPP EINS are non-wage and exclude those values from any earnings analyses. The current version of the DAF21 MEF link-file excludes TPP “earnings” from the primary wage (PRWAGEYYYY), total wage (TLWAGEYYYY), and all earnings (ALLEARNYYYY) values. If the only “earnings” in a year is from a TPP EIN, the work flag (WORKYYYY), will not indicate work. The following variables provide TPP information:

N_TPPEINYYYY: Number of TPP EINS in year (1987-2021)
TPPFEINYYYY: Federal employment identification number (FEIN) for highest TPP EIN in year (1987-2021)
TPPWAGEYYYY: Earnings from all TPP EINS in year (1987-2021)

Only in rare cases will an individual have more than one TPP value in a year. Researchers can use these variables to assess the TPP “earnings” values and use those values as they see fit. Adding TPPWAGEYYYY to any of the other earnings values in the 2021 MEF link-file will yield the value that would have occurred under DAF-MEF versions prior to 2012.

The current list of TPP EINS has 1,247 EINS.

III. Notes of caution

Taxable Medicare earnings prior to 1991 were capped at the FICA maximum. In 1991, 1992, and 1993, taxable Medicare earnings were capped at \$125,000, \$130,200, and \$135,000 respectively, (more than twice the FICA level which was \$53,400 in 1991), but this still did not reflect all earnings.

Prior to 1991, the MEF only contains self-employment earnings that were less than the difference between FICA cap and any Social Security taxable wages that were earned. This was because only these earnings were Social Security taxable. Thus, individuals with wages near or at the FICA limit would show little or no self-employment earnings. With the higher Medicare cap after 1990, the MEF includes more self-employment, but only up to the higher Medicare cap.

Congress removed that Medicare cap in 1994 so thereafter that all wages and self-employment earnings are taxed, and self-employment earnings are complete. In addition, self-employment earnings in 1991 that were subject only to Medicare were not included in the MEF due to a technical glitch at the time.

Most earnings reports in the MEF come from employers, and SSA validates the SSNs on those records before posting that data to a Social Security record. About 4% of records do not match and go to an Earnings Suspense File (ESF) and these earnings are not initially reflected in the MEF. In recent years, SSA Office of the Chief Actuary estimates that wage and salary worker posting counts are about 97.5 % complete as of one year after the end of the tax year, about 98% after two years, and about 98.5% after three years. (The items remaining in the ESF after this time have in recent years become relatively intractable.) For self-employment worker counts, they are about 95% complete as of one year after the end of the tax year, about 97.5% complete after two years, and

almost complete after three years. Associated posted money amounts are similar. However, the average for the combined wage and salary amount one year after the end of the tax year is virtually the same as it is after three years so the incomplete data soon after the end of a tax year do not introduce obvious bias into the data.

Other errors in the data occur due to entry errors or errors that can occur when a prior error is cancelled with an offsetting negative entry (no records are ever removed—they are offset by new entries instead). This can lead to negative values in some cases. Further, Medicare entries at SSA do not receive substantial scrutiny so cases exist where the FICA amount has been corrected (e.g., \$23,421), but the Medicare amount has not (e.g., 2,345,100). Extreme outlier values are also known to occur so care must be taken to deal with such cases.

IV. File name, location, and contents

The MEF linkable extract includes earnings from the MEF for all members of the DAF21 with MEF earnings at any point between 1987 and 2021.

Libname: OPDR.TG.PRD.RTWR.FINAL.DAF21P.MEF21

COPY: OPDR.TG.PRD.RTWR.FINAL.DAF21C.MEF21

Filename: In each case the file names are the same: MEF_APR2023

The data was extracted from the MEF April 24, 2023. Each file has 37,356,145 observations and 352 variables.

We also create a version of the MEF that aligns with the NBS extract files. It has all of the MEF variables but is limited to those SSA beneficiaries who participated in one of SSA's NBS rounds.

Libname: OPDR.TG.PRD.RTWR.FINAL.DAF21P.MEFNBS21

Filename: MEF21_4NBS_APR2023

The file has 43,465 observations and 373 variables.

The specific variables are provided as follows:

VARIABLE	TYPE	LEN	LABEL
SSN	CHAR	9	SOCIAL SECURITY NUMBER
TLWAGEYYYY	NUM	8	TOTAL WAGES NON-TPP EMPLOYERS IN YEAR (1987-2021)
PRWAGEYYYY	NUM	8	WAGES PRIMARY (HIGHEST\$) NON-TPP EMPLOYER IN YEAR (1987-2021)
FEINYYYY	NUM	8	FEDERAL EMPLOYMENT IDENTIFICATION NUMBER (FEIN) OF PRIMARY (HIGHEST\$) NON-TPP EMPLOYER IN YEAR (1987-2021)
SE_EARNYYYY	NUM	8	SELF-EMPLOYMENT EARNINGS IN THE YEAR (1987-2021)
ALLEARNYYYY	NUM	8	TOTAL NON-TPP EARNINGS (TLWAGES+SELF-EMPL) IN YEAR (1987-2021)
WORKYYYY	NUM	8	BINARY WORK INDICATOR (1=WORKING, 0=NOT WORKING) FOR THE YEAR (1987-2021)
N_EMPLRSYYYY	NUM	8	NUMBER OF NON-TPP EMPLOYERS IN YEAR (1987-2021)
LAST_DER_YR	NUM	8	LAST YEAR OF EARNINGS IN DER DATA

VARIABLE	TYPE	LEN	LABEL
N_TPPEINYYYY	NUM	8	NUMBER OF TPP EINS IN YEAR (2087-2021)
TPPFEINYYYY	NUM	8	FEDERAL EMPLOYMENT IDENTIFICATION NUMBER (FEIN) FOR HIGHEST TPP EIN IN YEAR (1987-2021)
TPPWAGEYYYY	NUM	8	EARNINGS FROM ALL TPP EINS IN YEAR (1987-2021)

Because the returned MEF finder file can include many years of data and many records per year for each of the more than 35 million beneficiaries in DAF, these files are very large. To manage the volume of data, we use macro programming to process the files in many smaller segments of data.

The construction code is as follows:

JCL CODE:

```
//MEF2021 JOB PAUL-OLEARY,MSGCLASS=1,REGION=2047M,NOTIFY=&SYSUID
//* IF NECESSARY USE Read TPP EINS FROM KEVIN FOR MEF DATA
//* PER KEVIN KULZER 5/2/2023
//* EARN20P3.EINLIST IS LATEST TPP DATA
//JOB01 EXEC SAS9,WORK='180000,90000'
//SYSPRINT DD SYSOUT=*
//*****WORKING DATA*****
//TEMP DD DSN=&&TEMP,DISP=(NEW,DELETE,DELETE),
// SPACE=(CYL,(3000,3000),RLSE),VOL=(,,,99)
//TEMP0 DD DSN=&&TEMP,DISP=(NEW,DELETE,DELETE),
// SPACE=(CYL,(3000,3000),RLSE),VOL=(,,,99)
//TEMP1 DD DSN=&&TEMP,DISP=(NEW,DELETE,DELETE),
// SPACE=(CYL,(3000,3000),RLSE),VOL=(,,,99)
//TEMP2 DD DSN=&&TEMP,DISP=(NEW,DELETE,DELETE),
// SPACE=(CYL,(3000,3000),RLSE),VOL=(,,,99)
//TEMP3 DD DSN=&&TEMP,DISP=(NEW,DELETE,DELETE),
// SPACE=(CYL,(3000,3000),RLSE),VOL=(,,,99)
//TEMP4 DD DSN=&&TEMP,DISP=(NEW,DELETE,DELETE),
// SPACE=(CYL,(3000,3000),RLSE),VOL=(,,,99)
//TEMP5 DD DSN=&&TEMP,DISP=(NEW,DELETE,DELETE),
// SPACE=(CYL,(3000,3000),RLSE),VOL=(,,,99)
//WORK01 DD DSN=&&TEMP,DISP=(NEW,DELETE,DELETE),
// SPACE=(CYL,(3000,3000),RLSE),VOL=(,,,99)
//WORK02 DD DSN=&&TEMP,DISP=(NEW,DELETE,DELETE),
// SPACE=(CYL,(3000,3000),RLSE),VOL=(,,,99)
//WORK03 DD DSN=&&TEMP,DISP=(NEW,DELETE,DELETE),
// SPACE=(CYL,(3000,3000),RLSE),VOL=(,,,99)
//WORK04 DD DSN=&&TEMP,DISP=(NEW,DELETE,DELETE),
// SPACE=(CYL,(3000,3000),RLSE),VOL=(,,,99)
//WORK05 DD DSN=&&TEMP,DISP=(NEW,DELETE,DELETE),
// SPACE=(CYL,(3000,3000),RLSE),VOL=(,,,99)
//WORK06 DD DSN=&&TEMP,DISP=(NEW,DELETE,DELETE),
// SPACE=(CYL,(3000,3000),RLSE),VOL=(,,,99)
//WORK07 DD DSN=&&TEMP,DISP=(NEW,DELETE,DELETE),
```



```
//      SPACE=(CYL,(3000,3000),RLSE),VOL=(,,,99)
//WORK08 DD DSN=##TEMP,DISP=(NEW,DELETE,DELETE),
//      SPACE=(CYL,(3000,3000),RLSE),VOL=(,,,99)
//WORK09 DD DSN=##TEMP,DISP=(NEW,DELETE,DELETE),
//      SPACE=(CYL,(3000,3000),RLSE),VOL=(,,,99)
//WORK10 DD DSN=##TEMP,DISP=(NEW,DELETE,DELETE),
//      SPACE=(CYL,(3000,3000),RLSE),VOL=(,,,99)
//WORK11 DD DSN=##TEMP,DISP=(NEW,DELETE,DELETE),
//      SPACE=(CYL,(3000,3000),RLSE),VOL=(,,,99)
//WORK12 DD DSN=##TEMP,DISP=(NEW,DELETE,DELETE),
//      SPACE=(CYL,(3000,3000),RLSE),VOL=(,,,99)
//WORK13 DD DSN=##TEMP,DISP=(NEW,DELETE,DELETE),
//      SPACE=(CYL,(3000,3000),RLSE),VOL=(,,,99)
//WORK14 DD DSN=##TEMP,DISP=(NEW,DELETE,DELETE),
//      SPACE=(CYL,(3000,3000),RLSE),VOL=(,,,99)
//WORK15 DD DSN=##TEMP,DISP=(NEW,DELETE,DELETE),
//      SPACE=(CYL,(3000,3000),RLSE),VOL=(,,,99)
//WORK16 DD DSN=##TEMP,DISP=(NEW,DELETE,DELETE),
//      SPACE=(CYL,(3000,3000),RLSE),VOL=(,,,99)
//WORK17 DD DSN=##TEMP,DISP=(NEW,DELETE,DELETE),
//      SPACE=(CYL,(3000,3000),RLSE),VOL=(,,,99)
//WORK18 DD DSN=##TEMP,DISP=(NEW,DELETE,DELETE),
//      SPACE=(CYL,(3000,3000),RLSE),VOL=(,,,99)
//WORK19 DD DSN=##TEMP,DISP=(NEW,DELETE,DELETE),
//      SPACE=(CYL,(3000,3000),RLSE),VOL=(,,,99)
//WORK20 DD DSN=##TEMP,DISP=(NEW,DELETE,DELETE),
//      SPACE=(CYL,(3000,3000),RLSE),VOL=(,,,99)
//WORK21 DD DSN=##TEMP,DISP=(NEW,DELETE,DELETE),
//      SPACE=(CYL,(3000,3000),RLSE),VOL=(,,,99)
//WORK22 DD DSN=##TEMP,DISP=(NEW,DELETE,DELETE),
//      SPACE=(CYL,(3000,3000),RLSE),VOL=(,,,99)
//WORK23 DD DSN=##TEMP,DISP=(NEW,DELETE,DELETE),
//      SPACE=(CYL,(3000,3000),RLSE),VOL=(,,,99)
//WORK24 DD DSN=##TEMP,DISP=(NEW,DELETE,DELETE),
//      SPACE=(CYL,(3000,3000),RLSE),VOL=(,,,99)
//WORK25 DD DSN=##TEMP,DISP=(NEW,DELETE,DELETE),
//      SPACE=(CYL,(3000,3000),RLSE),VOL=(,,,99)
//WORK26 DD DSN=##TEMP,DISP=(NEW,DELETE,DELETE),
//      SPACE=(CYL,(3000,3000),RLSE),VOL=(,,,99)
//WORK27 DD DSN=##TEMP,DISP=(NEW,DELETE,DELETE),
//      SPACE=(CYL,(3000,3000),RLSE),VOL=(,,,99)
//WORK28 DD DSN=##TEMP,DISP=(NEW,DELETE,DELETE),
//      SPACE=(CYL,(3000,3000),RLSE),VOL=(,,,99)
//WORK29 DD DSN=##TEMP,DISP=(NEW,DELETE,DELETE),
//      SPACE=(CYL,(3000,3000),RLSE),VOL=(,,,99)
//WORK30 DD DSN=##TEMP,DISP=(NEW,DELETE,DELETE),
//      SPACE=(CYL,(3000,3000),RLSE),VOL=(,,,99)
```

```
//WORK31 DD DSN=&&TEMP, DISP=(NEW,DELETE,DELETE),
//          SPACE=(CYL,(3000,3000),RLSE),VOL=(,,,99)
//*****INPUT DATA*****
//*****NOTE: THIS NONWAGE DATA CHANGES ONLY WITH TPP UPDATES
//NONWAGE DD DSN=OPDR.TG.PRD.RTWR.EARN20P1.EINLIST.SASLIB,DISP=SHR
//DERTKT1 DD DSN=OPDR.TG.PRD.RTWR.DAF21P.MRAIBKTS.R230424,DISP=SHR
//ALLSMPL1 DD DSN=OPDR.TG.PRD.RTWR.FINAL.DAF21P.MEF21,
//          DISP=(OLD,CATLG,KEEP),
//          UNIT=TSILO,VOL=(,,,99)
//ALLSMPL2 DD DSN=OPDR.TG.PRD.RTWR.FINAL.DAF21C.MEF21,
//          DISP=(OLD,CATLG,KEEP),
//          UNIT=TSILO,VOL=(,,,99)
//*****BE SURE TO ADD A YEAR OF DATA**
//SYSIN DD *

/**/
*;
```

SAS CODE:

```
OPTIONS MPRINT COMPRESS=YES NOCENTER PAGESIZE=MAX LINESIZE=MAX ;
%let source = NEWBCKT;
%let YR = 2021;

DATA Oqpeinnumber;
  SET NONWAGE.Oqpeinnumber (RENAME=(DET_EIN=EIN));
  PROC SORT; BY EIN;

  %MACRO COMBINE(OBSSTRT,OBSEND,SEGMENT);
*OPTIONS OBS=&OBSEND;
*****
***** PROCESS DER
*****;

DATA TEMP.DER_MEF_PROCESSING;
  SET DERTKT1.&source (FIRSTOBS=&OBSSTRT OBS=&OBSEND);
  PROC SORT ; BY EIN;

DATA TEMP1.DER_MEF_PROCESSING;
  MERGE TEMP.DER_MEF_PROCESSING (IN=wagesse RENAME=(YEAR=C_YEAR))
    Oqpeinnumber (IN=nonwages);
  by EIN;
  if wagesse=1;
  TPP_WAGES=nonwages ;
```

```
*****
*****STRING OUT THE WAGES PER YEAR*****
      THIS VERSION ALSO INCLUDES WAGES NOT COVERED BY MEDICARE
```

FROM RUSS HUDSON (THURSDAY, NOVEMBER 18, 2010 6:53 AM):

There are some state and local government where some or all of their

employees are not covered by Social security nor Medicare .

We produce a table for the Green Book that shows by state the coverage of state and local workers. I will forward this table to you in a separate EMAIL .

You can see by state the estimated number of non covered state and local jobs based on 2007 CWHS.

The estimate of total earnings that I use is the Maximum of three earnings computations.

1. Total compensation + deferred compensation + Medicare self employment + Deferred Payment(this is a negative amount that remove payments for the current year that does not represent work in current year).
2. Social Security taxable wages + self employment Social Security taxable earnings.
3. Medicare taxable wages + Medicare taxable self employment.

This method only excludes earnings that are not taxable by IRS or Medicare such as Medical Insurance paid from pretax earnings.

```
** BECAUSE ADDING A MISSING TO A VALID NUMBER MAKES
   THE VALID NUMBER MISSING, WE RESET MISSINGS TO ZERO FIRST
   IT IS IMPORTANT THAT WE RENAME THESE TO NEW VARIABLES AND
   THAT WE INCLUDE THE ELSE STATEMENT
```

```
*****;
ARRAY WMISSNG (7) WAGE_TIPS_IRS DEF_CNTR PAYMENT_457
                  WAGE_TIPS_SS WAGE_TIPS_MED SEI_SS
                  SEI_MED ;
ARRAY NOMISSNG (7) WAGE_TIPS_IRS_NOMISS DEF_CNTR_NOMISS
                  PAYMENT_457_NOMISS WAGE_TIPS_SS_NOMISS
                  WAGE_TIPS_MED_NOMISS SEI_SS_NOMISS
                  SEI_MED_NOMISS;
DO I=1 TO 7;

      IF WMISSNG(I) = . THEN NOMISSNG(I) = 0;
      ELSE NOMISSNG(I) = WMISSNG(I);
```

```

END;

BOX1_WAGE = WAGE_TIPS_IRS_NOMISS + DEF_CNTR_NOMISS +
PAYMENT_457_NOMISS;

BEST_WAGE = MAX(BOX1_WAGE, WAGE_TIPS_SS_NOMISS,
WAGE_TIPS_MED_NOMISS);
BEST_SE = MAX(SEI_SS_NOMISS, SEI_MED_NOMISS);

        BEST_WAGE_TPP=0;
        if TPP_WAGES=1 THEN BEST_WAGE_TPP=BEST_WAGE;
        if TPP_WAGES=1 THEN BEST_WAGE2=0; else
BEST_WAGE2=BEST_WAGE;
        DROP BEST_WAGE; RENAME BEST_WAGE2=BEST_WAGE;
YEAR =INPUT(C_YEAR,9.);
DROP C_YEAR;
IF 1987<=YEAR<=&YR;

PROC SORT data=temp1.DER_MEF_PROCESSING
out=temp2.DER_MEF_PROCESSING;
  BY SSN;

PROC SORT DATA=TEMP2.DER_MEF_PROCESSING
OUT=TEMP3.DER_MEF_PROCESSING;
  BY SSN YEAR  ;

PROC SORT DATA=TEMP3.DER_MEF_PROCESSING
OUT=TEMP4.DER_MEF_PROCESSING;
  BY SSN YEAR BEST_WAGE  ;

**STEP 1--ADD UP EACH EARNING TYPE PER EIN PER YEAR
        AND THEN STRING IT OUT TO 1 OBS PER PERSON
*****;
** WE OPEN THE SORTED DATA SET USING A BY STATEMENT SO WE CAN
        USE FIRST LAST PROCESSING
*****;

DATA WORK&SEGMENT..DER_MEF_PROCESSING_WIDE_&SEGMENT
(DROP=TYPE WAGE_TIPS_NF PENSION DEF_DSTB DEF_CNTR
WAGE_TIPS_IRS WAGE_TIPS_RR WAGE_SS TIPS_SS TIPS_MED WAGE_TIPS_SS
WAGE_TIPS_MED SEI_SS SEI_MED HSA_CNTR PAYMENT_457 SWP_457 PLAN_401
PLAN_403 PLAN_408 PLAN_457 PLAN_501 TPP_WAGES WAGE_TIPS_IRS_NOMISS
DEF_CNTR_NOMISS PAYMENT_457_NOMISS WAGE_TIPS_SS_NOMISS
WAGE_TIPS_MED_NOMISS SEI_SS_NOMISS SEI_MED_NOMISS BOX1_WAGE
BEST_WAGE_TPP SUM_TPPW BY_YR TPP_COUNT_YR
FIRST_SSN_DER FIRST_BEST_WAGE LAST_SSN_DER LAST_BEST_WAGE

```

```

TL_EARNG EMPLOYMENT );
SET TEMP4.DER_MEF_PROCESSING;
RETAIN SUM_TPPW_BY_YR
  DUPLCT_COUNT_YR tpp_count_YR SUM_WAGE_BY_YR SUM_SELF_BY_YR
  TLWAGE1987- TLWAGE1999      TLWAGE2000-  TLWAGE&YR
  PRWAGE1987- PRWAGE1999      PRWAGE2000-  PRWAGE&YR
  TPPWAGE1987-TPPWAGE1999     TPPWAGE2000-  TPPWAGE&YR
  FEIN1987-    FEIN1999      FEIN2000-    FEIN&YR
  TPPFEIN1987-TPPFEIN1999    TPPFEIN2000-  TPPFEIN&YR
  SE_EARN1987-SE_EARN1999    SE_EARN2000-  SE_EARN&YR
  ALLEARN1987-ALLEARN1999    ALLEARN2000-  ALLEARN&YR
  WORK1987-    WORK1999      WORK2000-    WORK&YR
  N_EMPLRS1987-N_EMPLRS1999  N_EMPLRS2000-N_EMPLRS&YR
  N_TPPEIN1987-N_TPPEIN1999  N_TPPEIN2000-N_TPPEIN&YR;
  BY SSN YEAR BEST_WAGE ;

```

```

FIRST_SSN_DER = FIRST.SSN ;
FIRST_YR      = FIRST.YEAR ;
FIRST_BEST_WAGE = FIRST.BEST_WAGE ;

```

```

LAST_SSN_DER = LAST.SSN ;
LAST_YR      = LAST.YEAR ;
LAST_BEST_WAGE = LAST.BEST_WAGE ;

```

** SINCE WORKERS CAN HAVE MORE THAN ONE WAGE PER YEAR, WE HAVE SORTED BY SSN-YEAR-WAGE TO PUT THE LARGEST IN A YEAR LAST. FOR EACH YEAR WE RESET THE VALUE TO 0, AND THEN ADD ACROSS THE EMPLOYERS IN THAT YEAR

*****;

```

IF FIRST.YEAR THEN DO;
  DUPLCT_COUNT_YR = 0;
  TPP_COUNT_YR = 0;
  SUM_WAGE_BY_YR = 0;
  SUM_TPPW_BY_YR = 0;
  SUM_SELF_BY_YR = 0;

```

END;

```

DUPLCT_COUNT_YR = DUPLCT_COUNT_YR - (TPP_WAGES-1);
****WHEN TPP_WAGES=0, THE COUNTER SUBTRACTS -1 (IE, ADDS +1)****;
TPP_COUNT_YR = TPP_COUNT_YR + TPP_WAGES;

```

```

SUM_WAGE_BY_YR = SUM_WAGE_BY_YR + BEST_WAGE ;
SUM_TPPW_BY_YR = SUM_TPPW_BY_YR + BEST_WAGE_TPP ;
*****
*****

```

* NOTE WAGE_TIPS_NF AND TIPS ARE ALREADY IN BEST_WAGE SO IF WE ADD IT HERE WE DOUBLE COUNT

```

*****
*****;
SUM_SELF_BY_YR = SUM_SELF_BY_YR + BEST_SE;

TL_EARNG=0;
TL_EARNG = TL_EARNG + SUM_WAGE_BY_YR + SUM_SELF_BY_YR;

*** CHANGE FOR 2016;
*** ONLY COUNT WAGES IF THEY ARE NON-TPP;
*** SINCE SE CAN INCLUDE LOSSES, BOTH NEG & POS SE EARNINGS (NON-0)
MEAN WORK;
IF (TPP_WAGES=0 and SUM_WAGE_BY_YR > 1)
  OR SUM_SELF_BY_YR ~= 0 THEN EMPLOYMENT = 1;

ARRAY TLWAGE(1987:&YR) TLWAGE1987- TLWAGE1999 TLWAGE2000-
TLWAGE&YR;
ARRAY PRWAGE(1987:&YR) PRWAGE1987- PRWAGE1999 PRWAGE2000-
PRWAGE&YR;
ARRAY TPPWAGE(1987:&YR) TPPWAGE1987-TPPWAGE1999 TPPWAGE2000-
TPPWAGE&YR;
ARRAY FEIN(1987:&YR) FEIN1987- FEIN1999 FEIN2000-
FEIN&YR;
ARRAY TPPFEIN(1987:&YR) TPPFEIN1987-TPPFEIN1999 TPPFEIN2000-
TPPFEIN&YR;
ARRAY SE_EARN(1987:&YR) SE_EARN1987-SE_EARN1999 SE_EARN2000-
SE_EARN&YR;
ARRAY ALLEARN(1987:&YR) ALLEARN1987-ALLEARN1999 ALLEARN2000-
ALLEARN&YR;
ARRAY WORK(1987:&YR) WORK1987-WORK1999 WORK2000-
WORK&YR;
ARRAY EMPLRS (1987:&YR) N_EMPLRS1987-N_EMPLRS1999 N_EMPLRS2000-
N_EMPLRS&YR;
ARRAY TPPEINS (1987:&YR) N_TPPEIN1987-N_TPPEIN1999 N_TPPEIN2000-
N_TPPEIN&YR;

IF FIRST.SSN THEN DO;
  DO I=1987 TO &YR;
    TLWAGE(I) = 0;
    PRWAGE(I) = 0;
    TPPWAGE(I) = 0;
    FEIN(I) = .;
    TPPFEIN(I) = .;
    SE_EARN(I) = 0;
    ALLEARN(I) = 0;
    WORK(I) = 0;
    EMPLRS(I) = 0;
    TPPEINS(I) = 0;
  
```

```

END;
END;
  TLWAGE (YEAR) = SUM_WAGE_BY_YR;
  PRWAGE (YEAR) = BEST_WAGE;
  TPPWAGE (YEAR) = sum_TPPw_by_YR;
  IF TPP_WAGES=0 THEN FEIN (YEAR) = EIN;
  IF TPP_WAGES=1 THEN TPPFEIN (YEAR)= EIN;
  SE_EARN (YEAR)= SUM_SELF_BY_YR;
  ALLEARN (YEAR) = TL_EARNG;
  WORK (YEAR)= EMPLOYMENT;
  EMPLRS (YEAR)=DUPLCT_COUNT_YR;
  TPPEINS (YEAR)=tpp_count_YR;

  LAST_DER_YR=YEAR;
  LABEL LAST_DER_YR='LAST YEAR OF EARNINGS IN DER DATA';
  LABEL TLWAGE1987='TL WAGES ALL EMPLRS IN YR (1987-&YR)';
  LABEL PRWAGE1987='WAGES PRIMARY (HIGHEST$) EMPLR IN YR (1987-
&YR)';
  LABEL FEIN1987='FEIN OF PRIMARY (HIGHEST$) EMPLR IN YR (1987-
&YR)';
  LABEL SE_EARN1987='SELF-EMPL EARNINGS IN YR (1987-&YR)';
  LABEL ALLEARN1987='TL EARNS (TLWAGES+SELF-EMPL) IN YR (1987-
&YR)';
  LABEL WORK1987='WORK FLAG FOR YR (1=WORK 0=NOWORK 1987-&YR)';
  LABEL N_EMPLRS1987='NUMBER OF EMPLRS IN YR (1987-&YR)';
  LABEL N_TPPEIN1987='NUMBER OF TPP EINS IN YR (1987-&YR)';
IF LAST_SSN_DER;

DROP YEAR EIN FIRST_YR I LAST_YR
  BEST_SE SUM_SELF_BY_YR SUM_WAGE_BY_YR BEST_WAGE
  DUPLCT_COUNT_YR;
  *****
  NOTE: WE DONT NEED THE EIN OR THE YEAR
  BECAUSE THIS IS THE WIDE RECORD WITH EINS IN
  EACH YEAR SO THE EIN JUST GIVES THE LAST EIN
  *****;

  %mend COMBINE;

* NOTE: THESE ARE WITHIN EACH NEWBUCKET FILE SO
  NEED NOT BE IN 4 SECTIONS AND TOP NUMBER IN THE
  LAST SECTION MUST EXCEED THE TOTAL FOR THE LARGEST
  INDIVIDUAL NEWBUCKET FILE (HERE 1,239,355,411);

%COMBINE (      1,  40000000,01);

```

```

%COMBINE ( 40000001, 80000000,02);
%COMBINE ( 80000001, 120000000,03);
%COMBINE ( 120000001, 160000000,04);
%COMBINE ( 160000001, 200000000,05);
%COMBINE ( 200000001, 240000000,06);
%COMBINE ( 240000001, 280000000,07);
%COMBINE ( 280000001, 320000000,08);
%COMBINE ( 320000001, 360000000,09);
%COMBINE ( 360000001, 400000000,10);
%COMBINE ( 400000001, 440000000,11);
%COMBINE ( 440000001, 480000000,12);
%COMBINE ( 480000001, 520000000,13);
%COMBINE ( 520000001, 560000000,14);
%COMBINE ( 560000001, 600000000,15);
%COMBINE ( 600000001, 640000000,16);
%COMBINE ( 640000001, 680000000,17);
%COMBINE ( 680000001, 720000000,18);
%COMBINE ( 720000001, 760000000,19);
%COMBINE ( 760000001, 800000000,20);
%COMBINE ( 800000001, 840000000,21);
%COMBINE ( 840000001, 880000000,22);
%COMBINE ( 880000001, 920000000,23);
%COMBINE ( 920000001, 960000000,24);
%COMBINE ( 960000001,1000000000,25);
%COMBINE(1000000001,1040000000,26);
%COMBINE(1040000001,1080000000,27);
%COMBINE(1080000001,1120000000,28);
%COMBINE(1120000001,1160000000,29);
%COMBINE(1160000001,1200000000,30);
%COMBINE(1200000001,          MAX,31);
RUN;

```

```

DATA ALLSMPL1.MEF_APR2023
      ALLSMPL2.MEF_APR2023;
DATA TEMP5.MEF_APR2023;
SET WORK01.DER_MEF_PROCESSING_WIDE_01
      WORK02.DER_MEF_PROCESSING_WIDE_02
      WORK03.DER_MEF_PROCESSING_WIDE_03
      WORK04.DER_MEF_PROCESSING_WIDE_04
      WORK05.DER_MEF_PROCESSING_WIDE_05
      WORK06.DER_MEF_PROCESSING_WIDE_06
      WORK07.DER_MEF_PROCESSING_WIDE_07
      WORK08.DER_MEF_PROCESSING_WIDE_08
      WORK09.DER_MEF_PROCESSING_WIDE_09
      WORK10.DER_MEF_PROCESSING_WIDE_10
      WORK11.DER_MEF_PROCESSING_WIDE_11

```



```

WORK12.DER_MEF_PROCESSING_WIDE_12
WORK13.DER_MEF_PROCESSING_WIDE_13
WORK14.DER_MEF_PROCESSING_WIDE_14
WORK15.DER_MEF_PROCESSING_WIDE_15
WORK16.DER_MEF_PROCESSING_WIDE_16
WORK17.DER_MEF_PROCESSING_WIDE_17
WORK18.DER_MEF_PROCESSING_WIDE_18
WORK19.DER_MEF_PROCESSING_WIDE_19
WORK20.DER_MEF_PROCESSING_WIDE_20
WORK21.DER_MEF_PROCESSING_WIDE_21
WORK22.DER_MEF_PROCESSING_WIDE_22
WORK23.DER_MEF_PROCESSING_WIDE_23
WORK24.DER_MEF_PROCESSING_WIDE_24
WORK25.DER_MEF_PROCESSING_WIDE_25
WORK26.DER_MEF_PROCESSING_WIDE_26
WORK27.DER_MEF_PROCESSING_WIDE_27
WORK28.DER_MEF_PROCESSING_WIDE_28
WORK29.DER_MEF_PROCESSING_WIDE_29
WORK30.DER_MEF_PROCESSING_WIDE_30
WORK31.DER_MEF_PROCESSING_WIDE_31
;
RUN;
PROC SORT DATA=TEMP5.MEF_APR2023 OUT=ALLSMPL1.MEF_APR2023;
BY SSN;
DATA ALLSMPL2.MEF_APR2023;
SET ALLSMPL1.MEF_APR2023;
RUN;

//MEF4NBS EXEC SAS9,WORK='500000,50000'
//***** INPUT FILES
//ALLSMPL1 DD DSN=OPDR.TG.PRD.RTWR.FINAL.DAF21P.MEF21,DISP=SHR
//NBSSNS DD
DSN=OPDR.TG.PRD.ETTW.FINAL.DAF20E.NBSLIST.SA.V1,DISP=SHR
//***** OUTPUT FILE
//MEFNBS DD DSN=OPDR.TG.PRD.RTWR.FINAL.DAF21P.MEFNBS21,
//      DISP=(OLD,CATLG,KEEP),VOL=(,,99)
//SYSIN DD *

/**/
*;

OPTIONS COMPRESS=YES NOCENTER PAGESIZE=MAX LINESIZE=MAX OBS=MAX;

DATA MEFNBS.MEF21_4NBS_APR2023;
MERGE ALLSMPL1.MEF_APR2023
      NBSSNS.NBSLIST (IN=SSNLIST);

```

```
BY SSN; IF SSNLIST;
```

Mathematica Inc.

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

EDI Global, a Mathematica Company

Operating in Tanzania, Uganda, Kenya, Mozambique, and the United Kingdom

Mathematica, Progress Together, and the “spotlight M” logo are registered trademarks of Mathematica Inc.



mathematica.org [website](#)