Volume 1: Getting Started 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 1: Getting Started with the DAF21]. Retrieved from https://www.ssa.gov/disabilityresearch/daf.html.

See page v for suggested citations for DAF datasets.

Acknowledgements

Building the DAF each year is a complex undertaking that requires skill, patience, and dedication. The DAF21 would not exist without its group of programmers, including (in alphabetical order) McCayla Agnew, Svetlana Bronnikov, Alex Bryce, Christian Carrillo Paz, Miriam Chappelka, Brittany Colip, Serge Lukashanets, Dawn Phelps, and its current project director, Matthew Urato. A documentation team including McCayla Agnew, Michael Anderson, Janine Bologna, Christian Carrillo, Brianna Elgart, Monica Farid, Jody Schimmel Hyde, Rebekka Nickman, and Imani Williams develop the twelve volumes of DAF documentation for DAF21, with input provided by the programming team throughout the process. We extend our gratitude to Stephanie Barna and Cindy Castro who produce clean and well-organized versions of the more than 3,500 pages of DAF documentation each year.

The documentation of the DAF has evolved over time to reflect user needs, but the groundwork was laid by significant contributions in earlier versions by staff at Mathematica. In particular, we acknowledge Miriam Loewenberg, who also wrote much of the code used to construct the DAF, Lesley Hildebrand and Laura Kosar, who each directed this project at Mathematica for many years. Additionally, significant contributions to the programming and documentation have been made by Zachary Steckel, Adele Rizzuto, Benjamin Fischer, Purvi Sevak, Abigail Philip, Michelle Lee, and Cassandra Pickens-Jewell. Their contributions continue to be felt, in some cases many years after their significant involvement. In addition, several Mathematica staff have made valuable contributions to the content and structure of the DAF documentation across the years; we thank David Wittenburg, David Stapleton, and Denise Hoffman for their insightful comments and advice.

SSA has been and continues to be a wonderful partner in developing the DAF and its documentation each year. We are sincerely grateful for the insights and suggestions from Paul O'Leary, Chris Earles, Tomas Padilla, Emily Roessel (who left SSA in 2019), and Debra Tidwell-Peters among others over the years.

Contents

Suggested Citations v
Glossary vii
Overview of DAF Documentation x
I. Introduction 1
A. Overview of DAF documentation 2
B. Organization of this volume 3
II. Working with DAF for the First Time 4
A. Structure of the DAF database 4
Data marts 5
DAF extracts (subpopulation mini files) 6
DAF linkable files 6
DAF21 Public Use File (DAF-PUF) 7
B. Where to find the DAF on the SSA mainframe 7
C. Which version of the DAF should researchers use? 7
D. What's new in the DAF? 7
E. Who is in the DAF? 9
1. Identifying beneficiaries in DAF source data 9
2. Beneficiary selection criteria 9
3. Known problem with auxiliary beneficiary selection methodology 10
F. One-observation-per-beneficiary structure 11
III. Description of DAF Components, DAF-Linkable Files, and Local Economic Data 12
A. DAF DMG component 12
B. DAF Annual files 14
C. DAF Ticket files 15
D. DAF Payments files 15
E. DAF Data Marts and Extracts 16
F. DAF Linkable files 16
1. RSA-911 16
2. MEF 18
G. Local economic data from the LAUS and SAIPE 18
H. The DAF Public Use File (PUF) 18
Appendix A: SAS Code for Beneficiary Selection Criteria A.1

Tables

II.1. Status of variables for SSDI-only, SSI-only, and concurrent beneficiaries, by DAF component 11

Figures

III.1. Diagram of the DAF.....13

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. Introduction

The Disability Analysis File (DAF) is a longitudinal data file housed on the mainframe at the Social Security Administration (SSA) Baltimore data center. It is available to help answer questions about Social Security Disability Insurance (SSDI) and/or Supplemental Security Income (SSI) beneficiaries. The DAF repackages existing SSA administrative data with research and analysis in mind and includes detailed information about SSDI and SSI beneficiaries who were below full retirement age (FRA) who have received disability benefits in any month since 1996. The DAF contains both longitudinal and one-time data about these beneficiaries, with data from SSI and SSDI combined into a single record per beneficiary. For each beneficiary, the DAF contains information about the disability determination process, program participation, and engagement with service providers such as Employment Networks (ENs) through the Ticket to Work (TTW) program as well as State Vocational Rehabilitation (State VR) Agencies.

The data contained in the DAF are extracted from various SSA data files, including the Supplemental Security Record (SSR), the Characteristics Extract Record 100% Field File (CER), the Master Beneficiary Record (MBR), the Payment History Update System (PHUS), the Disabled Beneficiary and Dependents Extract (DBAD), the Disability Control File (DCF), EN Payment Cumulative Payment Report, the Vocational Rehabilitation Reimbursement Management System (VRRMS), 831 & 832/833, Numerical Identification File (NUMIDENT), and SSA's Waterfall File that contains information on continuing disability reviews. More information about these source files is contained in Volume 9 of this documentation.

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

The current version of the database, the DAF21, contains data on the beneficiaries mentioned above through December 2021. One-time variables in the DAF21 include data such as Social Security Number (SSN) and Date of Birth, while longitudinal variables track beneficiaries over time and include data such as state of residence, impairment codes, and benefit payment amounts. The DAF contains data as early as January 1994, and many longitudinal variables are available for the full time period (through December 2021). Some variables are only available for a portion of the full-time range, and information in this documentation will provide users with the time period of availability for each variable.

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

A. Overview of DAF documentation

The complete documentation of the DAF is contained in twelve volumes. Together, these volumes provide comprehensive detail for users of all levels and backgrounds.

- Volume 1: Getting Started with the DAF21
- Volume 2: Working with the DAF21
- Volume 3: Tips for Conducting Analysis with the DAF21
- Volume 4: Lists of DAF21Variables
- Volume 5: DAF21 Variable Detail Pages
- Volume 6: Validating the DAF21 Against Other Sources
- Volume 7: DAF21 Development History and Construction Methods
- Volume 8: DAF21 Construction Workflow Charts and Task Tables
- Volume 9: DAF21 Source File Descriptions
- Volume 10: DAF21 Administrative Source File Documentation
- Volume 11: DAF21 Construction Code
- Volume 12: DAF21 RSA Administrative Source File Documentation

The purpose of this first volume is to provide a summary of the DAF to new and potential users of the DAF. It assists users in understanding the content of the DAF, its structure, and linkages between the DAF and other SSA and non-SSA administrative files.

Building on the content in Volume 1 are Volumes 2 and 3, which provide more detailed guidance on extracting data and naming and format conventions used throughout the DAF, while offering guidance about working with the DAF linked to other administrative files. These volumes also provide details on the DAF extracts and data marts, which are smaller data files designed to help users get going with analysis more quickly by working with pre-defined analysis samples.

Subsequent volumes are intended for users getting started with an analysis or requiring more in-depth knowledge than available in the first three volumes. Volume 4 contains listings of all the variables in the DAF, including by analytic category and data file as described below, while Volume 5 provides detailed definitions and additional information about each of these variables, stored in alphabetical order. Volumes 6-11 provide additional detail on more technical aspects of the DAF including validating summary statistics based on the DAF to statistics based on other SSA administrative data (Volume 6), specifics about DAF construction and its history (Volumes 7 and 8), and information about the source files (Volumes 9 and 10) and code (Volume 11) used in building the DAF.

Volume 12 is distinct from the other volumes in that it focuses solely on data from the Rehabilitation Services Administration (RSA) related to the provision of federal-state Vocational

Rehabilitation services. The volume contains information on how we process those files for DAF, as well as policy directives from RSA and information about the source files.

All DAF documentation is available online at: https://www.ssa.gov/disabilityresearch/daf.html. Users with inquiries about the DAF or its documentation should contact ORDES.DAF@ssa.gov.

B. Organization of this volume

Following this introduction, Chapter II provides an overview of the DAF database, including its location, a brief description of the structure and content of the four core DAF components, and a discussion of which beneficiaries are included in the DAF in a given year. Chapter III provides detailed information about the four main DAF components, linkages to other SSA and non-SSA administrative databases, subpopulation extracts from the full DAF, and clean and accessible files that are immediately available for analysis, known as data marts.

II. Working with DAF for the First Time

A. Structure of the DAF database

As shown in the diagram below, the DAF database is sectioned into four core components, each containing different types of information. The components may be used individually or together by merging on the variable SSN, which is included in each component.

- The **DAF.DMG** component is a beneficiary-level file that contains demographic and one-time information, such as DOB, name, and program application dates.
- The **DAF.Annual** component consists of beneficiary-level files, one for each year, with each file containing monthly variables containing programmatic information such as benefits due and paid; earnings and other income; and benefit suspense and terminations.
- The **DAF.Ticket** component consists of multiple files, each of which contains beneficiary-level data related to the TTW program, such as Ticket mail and assignment dates. The Ticket component is separated into two parts, Ticket Base (one file containing non-date-specific information) and Ticket Annuals (calendar year files containing date-specific and longitudinal information such as the year and month the beneficiary's ticket was active).
- The **DAF.Payments** component contains data on payments made to State VR Agencies under SSA's traditional reimbursement system,[2] as well as data on Milestone and Outcome payments made to ENs or State VR Agencies acting as ENs for services provided to Ticket participants. The data relating to the payments made to ENs and State VR Agencies acting as ENs is available in two files: the Payments Horizontal File (one record per beneficiary) and Payments Vertical File (one record per payment). Data relating to traditional reimbursement to State VR Agencies is contained in separate files.

Additionally, as described in the "What's New" section below and in Volume 2, DAF21files include a standalone file related to medical and work-related continuing disability reviews.

For ease of analysis, allowing researchers to access basic information without having to merge across files, each core component also contains eleven common variables (described in the Variable Detail pages in Volume 5):

- DOBBEST – Best Birth Date (constructed from multiple administrative data sources according to a hierarchy developed with SSA)
- DODBEST – Best Death Date (constructed from multiple administrative data sources according to a hierarchy developed with SSA)
- DOEI – SSDI Initial Entitlement Date
- DXPRIBEST – Best Primary Diagnosis Code

[2] As is described in Section III.D below, data relating to cost-reimbursement payments to State VR Agencies have not received the same extensive level of vetting as other core DAF components. For this reason, we suggest that users proceed with caution when analyzing these data, reporting any potential anomalies to ORDES.DAF@ssa.gov.

- FRADATE – Date the beneficiary reaches FRA; note that this variable was named RETIREDATE in earlier versions of DAF
- SEX – Gender of Beneficiary
- SSDI_ONSET – The earliest date of SSDI disability onset among all periods of SSDI disability
- SSDIyy – Indicates whether a beneficiary was in SSDI current pay status in any month of year yy, with years ranging from 1994 (94) to 2021 (21)
- SSIELIG_FIRST – Indicates the earliest date an individual is SSI eligible
- SSIyy – Indicates whether a beneficiary was in SSI current pay status in any month of year yy, with years ranging from 1994 (94) to 2021 (21)
- TKTyy – Indicates whether a beneficiary had an active Ticket (assigned or in-use) in year yy, with years ranging from 2002 (02) to 2021 (21).

In addition to the core components, which contain records for all beneficiaries meeting the DAF selection criteria, the DAF database contains supplemental files, known as data marts and data extracts, which are designed to facilitate analyses by subpopulation or where subpopulations smaller than the full beneficiary population would suffice. These are derived from the core component files and are described in detail in Volume 2. They may be particularly valuable to researchers getting started with the DAF.

Data marts

DAF data marts are designed to help with special research needs.

- **The 10% Data Mart** is a random sample of 10 percent of the beneficiaries in the DAF, including all of the information from the four core components described above. It can be used for testing programs prior to running them using the complete DAF or for analyses where smaller sample sizes are acceptable. A detailed description of the 10% Data Mart is in Volume 2.
- **The Awardee Data Mart (ADM)** is designed to support cohort analyses for the SSDI and SSI disability programs. It contains beneficiaries who received their first child or adult SSI payment or first SSDI payment as an adult between 1996 and the end of the last year covered by the current DAF; SSI child awardees were added to the ADM in DAF20. The ADM includes all DAF.DMG component variables for those beneficiaries as well as constructed payment- and eligibility-based award variables to identify the start of program participation. Researchers interested in assessing trends in beneficiary cohorts or in following outcomes of beneficiaries from first benefit month onward may be interested in this file. A detailed description of the Awardee Data Mart (ADM) is in Volume 2.

DAF extracts (subpopulation mini files)

DAF extracts contain all variables from the full DAF, but only for subpopulations of SSA beneficiaries.

- **The NBS extract** contains a subset of beneficiaries from the DAF who participated in the National Beneficiary Survey (NBS) and includes all DAF data for those beneficiaries through the current year.
- **The SSA Survey and Demonstration Projects** extract contains a subset of DAF beneficiaries who participated in one or more SSA surveys or demonstration projects: as in earlier versions of the DAF, this includes the Benefits Entitlement Services Team (BEST) Demonstration Project; the National Survey of SSI Children and Families (NSCF); or the Youth Transition Demonstration (YTD). This file now includes participants in other SSA demonstrations, including Accelerated Benefits (AB), Benefit Offset National Demonstration (BOND), Benefit Offset Pilot Demonstration (BOPD), Homeless Outreach Projects and Evaluation (HOPE), Mental Health Treatment Study (MHTS), Promoting Opportunity Demonstration (POD), Promoting Readiness of Minors in SSI (PROMISE), and the Supported Employment Demonstration (SED). It includes flags for participation in each survey and demonstration project and includes all DAF data through the current year for those beneficiaries.
- **The TTW participant extract** contains a subset of the beneficiary records in the DAF, consisting of beneficiaries who have assigned a Ticket to a provider under the provisions of the TTW program in or after December 2005. Each year, the file is updated to include new TTW participants as well as DAF data through the current year for those beneficiaries.

DAF linkable files

The DAF can also be linked to non-SSA files, two of which are available on the SSA mainframe. We describe the files here and in more detail in Chapter III of this volume. While these files may be accessed by researchers with proper permissions, access to them is restricted because they are not SSA administrative files. These files include:

- **RSA-911.** The Rehabilitation Services Administration Case Services Report (RSA-911) files are annual fiscal year files that contain information collected by State VR Agencies about individuals who were served by a state vocational rehabilitation agency. These data underwent substantial revisions in 2017—including the individuals for whom data are tracked—as described in more detail in Volumes 2 and 12. These data are maintained by the Department of Education.
- **MEF.** The Master Earnings File (MEF) is a file containing detailed annual earnings data based on W-2 information collected by the Internal Revenue Service (IRS). The MEF extract for DAF contains data from the MEF for DAF beneficiaries and contains variables representing earnings between 1987 and the most recent DAF year.

DAF21 Public Use File (DAF-PUF)

In addition to these data marts, extract/mini-files, and linkable files, we also produce a public use version of the DAF. The DAF-PUF contains a random 10 percent sample of full DAF. It contains the variables from full DAF that are most used by researchers, with some masking to avoid disclosure. It is large enough to avoid disclosure risk and small enough to keep the file size manageable. The DAF-PUF does not include the DAF.Ticket and DAF.Payment portions of DAF due to small sample sizes. DAF-PUF files and documentation are publicly available for download from SSA (https://www.ssa.gov/disabilityresearch/daf_puf.html) and at https://data.gov/.

B. Where to find the DAF on the SSA mainframe

All DAF components and linkable files are stored on the SSA mainframe. Some files that are linkable to the DAF are only available to select SSA staff, including the MEF. To keep the DAF file names easily accessible, particularly when files are updated, the Dataset names (DSN) of all DAF components and linkable files with date of updates are provided on the SSA mainframe at: OPDR.TG.PRD.ETTW.DAF21.FILES.CURRENT in a "file of filenames." Larger DAF files are stored on tape media and smaller files are stored on disk. Since the media type affects how users process these files on the mainframe, we have added a column this year to the "file of filenames" indicating the media type for each DAF file.

Many DAF components and linkable files also have copies stored on the SSA mainframe; those that do are labeled as such on the file of filenames. The filename of the copy is identical to the filename of the original except that the node "DAF21P" is "DAF21C".

C. Which version of the DAF should researchers use?

Researchers should always use the most current version of the DAF because each version updates, enhances, and corrects the information in its predecessor. If a variable from a previous version has since been dropped or changed, it was determined to be out-of-date, unreliable, or obsolete. Researchers should therefore not use prior versions of the DAF to ensure they are not working with variables that have been dropped or changed. Instead, we suggest consulting the list of new, renamed, and deleted variables in Volume 4 and updating any programming based on previous DAFs accordingly. In addition to lists of variables deleted in recent versions of the DAF, Volume 4 contains a list of the variable history of all variables that have changed since DAF12 for users who may have worked with an earlier version of the file.

D. What's new in the DAF?

Every year, the DAF undergoes revision in both structure and content based on user suggestions and the changing availability of data. We made the following changes in the DAF21:

- We have enhanced the information we report on beneficiary's geographic location. We have done this in two ways. First, we incorporated additional snapshot data on SSDI and SSI beneficiaries beyond what we use to select beneficiary records using a second pull from the

Characteristics Extract Record (CER) and Disabled Beneficiaries and Dependents (DBAD) files. This allows us to populate more months with data from the source files than we had been able to do previously; see Volume 7, Chapter II for more details. Second, we incorporate nine-digit postal ZIP codes, which allow us to construct a more accurate county of residence measure than what we were able to do with five-digit ZIP codes. More information on this change and the new variables that result are contained in Volume 3, Section X. "Geographic Measures in the DAF."

- We have incorporated new information on Continuing Disability Reviews (CDRs) from the Disability Control File (DCF) Medical Table. This information is contained in the standalone CDR file in the DAF and augments information on CDRs from other SSA source files. More information about the new variables is contained in Volume 2, Section VIII, "Using Information on Continuing Disability Reviews in the DAF," as well as in Volumes 4 and 5.
- We revised the standalone CDR file by combining 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 new categorical variable (CDR_WFAL_Aln) that specifies the highest adjudicative level at which a determination was made.
- We have added some new information with advice for users who work with the DAF measures of benefits suspended or terminated for work (STW). Specifically, we offer some additional considerations for how these measures function when beneficiaries receive both SSDI and SSI benefits. This new information can be found in Volume 3, Section VIII.B, "STW indicators."
- We have made a few other updates to the notes fields for variables in Volume 5 to enhance the user experience. Of particular note, we have added notes for the SSDIyy and SSIyy detail pages in Volume 5 to indicate how to use the measures in conjunction with age variables to select a sample of beneficiaries receiving federal disability benefits during the year. The current version of these measures include DI beneficiaries and SSI recipients who have transitioned to retirement benefits and so may capture beneficiaries too broadly for most researchers unless additional age restrictions are implemented by the user; we expect to refine our algorithm for constructing those measures in the next DAF cycle.
- We incorporated new measures on termination from benefits as part of our annual validation activities to confirm that the DAF aligns well with other published statistics. This comparison is available in Tables VI.1, VI.2, and VI.3 of Volume 6.
- We have updated the DAF Code Library to allow users to more easily determine whether a beneficiary is no longer entitled to benefits as a result of work activity within a specified time period." The new code complements the existing code in Section II of the Code Library that allows users to determine the number of months that benefits were suspended or terminated as a result of work activity within a specified time period." See https://www.ssa.gov/disabilityresearch/daf.html#library.

E. Who is in the DAF?

In this section, we describe the selection criteria for SSDI and SSI beneficiaries in DAF21; the Statistical Analysis Software (SAS) code implementing these criteria is available in Appendix A.[3]

1. Identifying beneficiaries in DAF source data

Although all records in the DAF are identified and linked to other files by the beneficiary's own SSN stored in the DAF variable SSN, the source variables for SSNs in the underlying SSA administrative files are known by a variety of names. DAF users should be aware of these source variables' different names, such as Person's Account Number (PAN), Beneficiary's Own Account Number (BOAN), or Client's Own Social Security Number (COSSN) and that not all of these variables are retained in the DAF itself. In addition, there are DAF variables, such as Housed Under Number (HUN) and Claim Account Number (CAN) that contain SSNs that in some cases are not the SSN of the beneficiary whose record is identified by the DAF SSN variable. Volume 9 contains detailed information about the source files for the DAF components and their respective SSN-identifier information.

2. Beneficiary selection criteria

The DAF contains all disabled individuals who received at least one month of SSDI benefits from January 1996 through December 2021 and was between 18 and FRA in that month. It also contains all recipients of SSI from birth to FRA who received at least one month of benefits during that same period.[4] SAS code for selecting beneficiaries in DAF21 is provided in Appendix A of this volume. Specifically, SSDI beneficiaries are selected if they meet the following conditions in any month of the year for which beneficiaries are being added:

- are currently in active or temporary deferred or suspended status for the SSDI program; and
- are disabled and either a primary claimant, dependent adult child, or a widow(er); and
- are between 18 years old and FRA;[5] and
- have a non-blank SSN

SSI beneficiaries are selected if they meet the following conditions in any month of the year for which beneficiaries are being added:

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

[3] Volume 7 contains a detailed history of beneficiary selection procedures over time, reasons for changes, and implications for beneficiaries contained in each iteration of the DAF.
[4] As mentioned in detail in Volume 7, beneficiaries selected for inclusion in earlier DAF iterations were subject to the selection criteria in place at that time.
[5] SSI beneficiaries receiving benefits between 65 and FRA are doing so as old age beneficiaries.

- are under FRA; and
- have non-blank SSN

Note that once a beneficiary is selected for inclusion in the DAF, that beneficiary remains on the DAF regardless of entitlement status, disability status, age in relation to FRA, or death. For example, an SSDI beneficiary who was added to the DAF based on benefits received in 2004 and reached FRA in 2007 remains in the DAF.[6] Some data elements continue to be updated in the administrative data source and therefore continue to be updated in the DAF. However, some data previously collected by SSA is no longer collected once the beneficiary reaches FRA. Earnings, for example, no longer affect beneficiaries who have obtained FRA, are not collected by SSA, and therefore are not contained in the DAF. Payment amounts, however, do continue to be recorded even though they are no longer related to disability status.

3. Known problem with auxiliary beneficiary selection methodology

Although the DAF is designed to contain data on all disability beneficiaries who meet the selection criteria described in the section above, a very small portion of records for SSDI auxiliary beneficiaries could be excluded due to a problem with dependents (mostly disabled adult children, or DACs) not having their own SSN. These auxiliary beneficiaries have a value in the MBR BOAN that is not the beneficiary's own SSN but instead is that of the primary beneficiary's. This is because, in the past, SSA sometimes had difficulty obtaining a child's SSN from a parent for an auxiliary claim. Because an SSN is required to process the claim, the parent's SSN was used in place of the child's SSN. While SSA stopped using a parent's SSN in place of a missing auxiliary SSN many years ago, a small number of these cases may still remain active. As a result of these incorrect BOANs for some auxiliary beneficiaries, all the necessary data matching among the SSA administrative sources cannot be done and these beneficiaries are excluded from the DAF. However, we expect this problem to be minor and shrinking because of enumeration at birth and other SSA efforts to have BOANs for all beneficiaries.

For SSDI beneficiaries, the BOAN indicates the SSN of the person actually receiving SSA benefits while the CAN indicates the SSN of the person whose earnings history is the basis of the SSA benefits. For primary beneficiaries, the CAN will be the same as the BOAN, but for auxiliary beneficiaries the CAN and BOAN should be different. When, as in the case described above, the parent's SSN is incorrectly used for the BOAN on an auxiliary record instead of the child's SSN, the BOAN is incorrectly the same as the CAN. As a result, both the child's and the parent's record in the MBR are identified by the same BOAN, causing potential for data confusion. We attempt to identify and screen out auxiliary beneficiary records with incorrect parental BOANs during DAF construction. However, it is possible that our detection method does not screen out every auxiliary beneficiary record with an incorrect BOAN.

[6] Although the beneficiary is part of the DAF, he/she will not appear on any Annual file after 2007 because disability was no longer a factor of entitlement. More detail about years that beneficiaries are excluded from the Annuals files is contained in Section III.B.

F. One-observation-per-beneficiary structure

Records in the DAF are constructed so that information on a beneficiary's participation in the SSDI and SSI programs is combined into a single record. Table II.1 shows the general status of the variables for SSDI, SSI, and concurrent beneficiaries; while the variables shown for each group are generally populated as listed in the table, there are instances where this may not be the case. When working with DAF variables for the first time, we recommend consulting the variable detail pages in Volume 5 of this documentation.

Table II.1. Status of variables for SSDI-only, SSI-only, and concurrent beneficiaries, by DAF component

Beneficiary group	DMG component	Annuals component	Ticket component	Payments component
SSDI-only beneficiaries (never received SSI)	All SSDI-related variables populated SSI variables: missing values	SSDI variables: populated for months of participation, missing values otherwise SSI variables: missing values	Records available for Ticket participants only, all variables populated	Records available only for recipients of State VR Agencies or EN services who generated a payment, all variables populated
SSI-only beneficiaries (never received SSDI)	All SSI-related variables populated SSDI variables: missing values	SSDI variables: missing values SSI variables: populated for months of participation, missing values otherwise	Records available for Ticket participants only, all variables populated	Records available only for recipients of State VR Agencies or EN services who generated a payment, all variables populated
Concurrent beneficiaries (received both SSDI and SSI, not necessarily in the same month)	All variables populated	SSDI variables: populated for months of participation, missing values otherwise SSI variables: populated for months of participation, missing values otherwise	Records available for Ticket participants only, all variables populated	Records available only for recipients of State VR Agencies or EN services who generated a payment, all variables populated

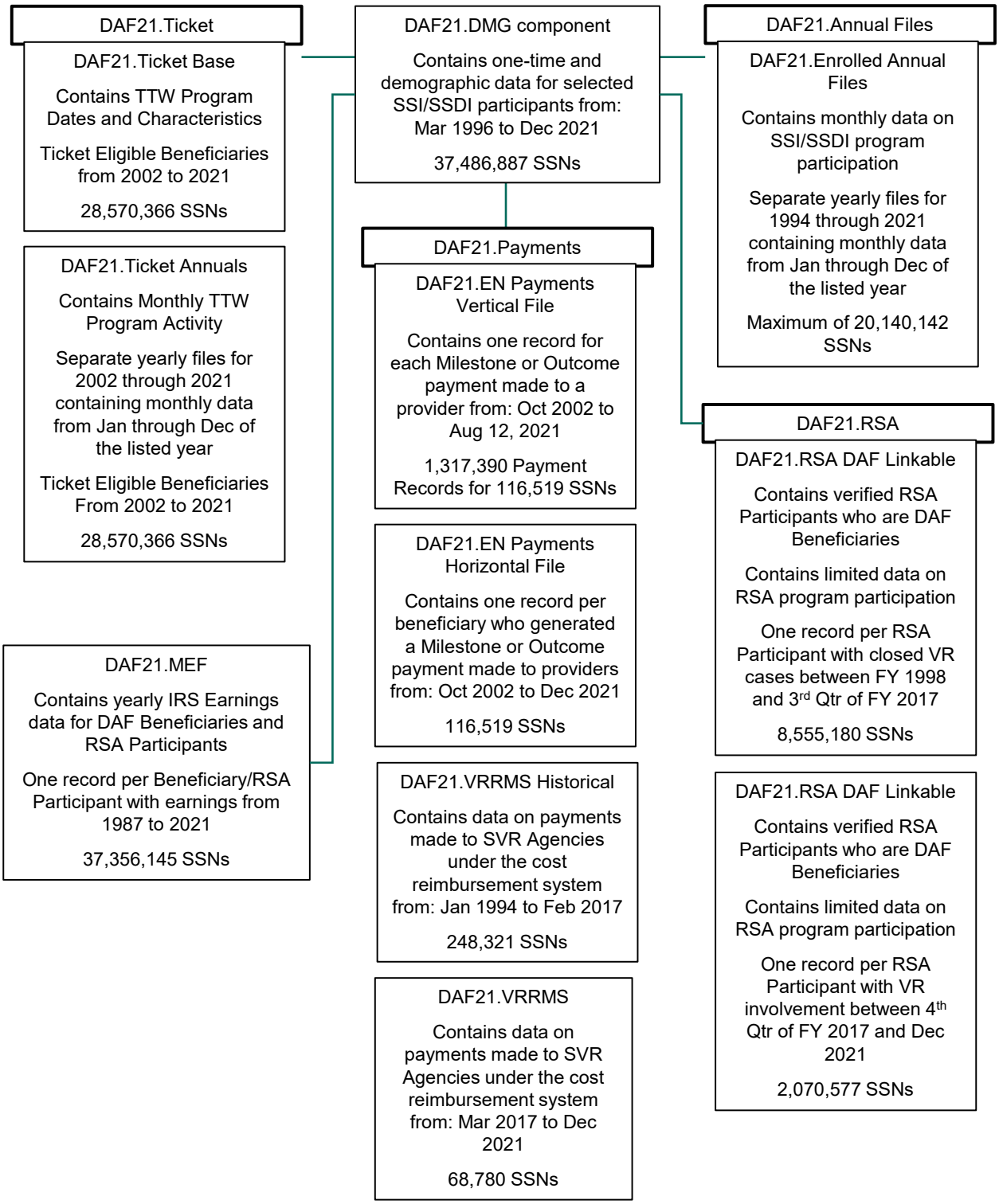
III. Description of DAF Components, DAF-Linkable Files, and Local Economic Data

In this section, we describe the primary components of the DAF and provide information about their structure and the beneficiaries contained in each. Figure III.1 provides a diagram summarizing this information and the linkages across the DAF components. Following this discussion, we offer information about other administrative files that are linkable to the DAF components.

A. DAF DMG component

This component contains primarily demographic data and other permanent characteristics for all beneficiaries who meet the DAF21 selection criteria (see sample size in Figure III.1). Some variables included in this component pertain to beneficiaries' SSI participation and some to their SSDI participation. Typically, data items that are specific to the SSI program are blank or missing if the beneficiary participates only in the SSDI program during all years of the DAF, and vice versa for data items specific to the SSDI program. For example, the SSDI variable TOCn (Type of Claim) is blank for SSI-only beneficiaries, and the SSI variable PDn (Presumptive Disability Decision Code) is blank for SSDI-only beneficiaries. Also included in DAF.DMG component are a limited number of longitudinal data items that are not well suited to inclusion in the DAF.Annual files, such as Primary Insurance Amount (PIA) for SSDI beneficiaries.

Figure III.1. Diagram of the DAF



B. DAF Annual files

The DAF.Annual files contain longitudinal data related to program participation and benefits such as benefit amounts due, zip code where benefits were received, and SSI income exclusions. The component consists of a set of files, one each year, for the years between 1994 and 2021. Each yearly file contains data on beneficiaries who were under FRA at any time during that year and had data in any of ten monthly variables at any point during the year including LAFyymm, PSTAyymm, PAYDyymm, PAYSyymm, STWDIyymm, STWSSIyymm, STWCMyymm, BFWDIyymm, BFWSSIyymm, or BFWCMyymm. These variables were selected because having data populated for at least one of them suggests that the individual may have received benefits in at least one month during the year. Because the Annuals files are limited only to beneficiaries in a given year, the largest Annual file is significantly smaller than the full DMG (see Table II.2)

Beneficiaries who do not have a record in a particular year of the Annuals files can be assumed to have not received benefits in that year because they had not yet begun to receive benefits, because they had died or reached FRA, or because they were terminated from benefits for reasons other than work activity for a full year and not yet reinstated.[7] Versions of the DAF from DAF13 to DAF15 split out the Annuals files in to two files for each year, where the second file contained data on "Non-Enrolled" beneficiaries. The records in that file were for beneficiaries who met the selection criteria for the DAF at some point but did not have any data available on the variables measured above. Because data was missing across those measures, these beneficiaries were determined to not be receiving benefits during the year. Beginning in DAF16, we no longer release the Non-Enrolled Annuals files, as they contain little information of use for research.

Variables in the Annual files are named according to the month and year. For example, the generic variable name DUEDyymm represents "SSDI Benefit Payment Due" while DUED9805 is the specific variable name for "SSDI Benefit Payment Due – 1998 May." If data are not available for a given participant in a given time period, then data are missing for those months. To illustrate, if a beneficiary entered the system in June 1999, all variables for that beneficiary from January 1994 to May 1999 are left blank. As another example, if a beneficiary's benefits are terminated and the beneficiary later becomes re-entitled, the longitudinal variables generally have missing values for the interim months of non-entitlement. Additionally, data availability will affect months in which certain variables are populated; these issues will affect all beneficiaries. For example, EARNyymm, is available only from March 1996 onwards. In Volume 2, we discuss issues related to the source files of variables in the DAF, as the start date of a source file can affect whether a variable is available.

Because the DAF Annual files are rebuilt with each iteration of the database, the values of specific variables may change from one year to the next. For example, a beneficiary in the

[7] Note that beneficiaries whose benefits terminate as a result of work activity have ongoing data in their STW and Benefits forgone due to work (BFW) variables and so remain on the Annual files until their death or attainment of FRA, at which time work activity ceases to be a factor of entitlement.

DAF20 may show a PSTA2009—SSI payment status in September 2020—value of C01, indicating current pay. In the DAF21, however, the same beneficiary's PSTA2009 value may show N01, indicating ineligibility due to excess income. A change like this indicates a change in the underlying administrative data between the time the administrative data was pulled for construction of the DAF19 (in spring 2021) and when it was accessed for construction of DAF21 (spring 2022). SSA performs these retrospective adjustments on a frequent basis as new information affecting beneficiary eligibility and payment is discovered, investigated, and verified.

C. DAF Ticket files

The third component contains data related to the administration of the Ticket program for SSI and SSDI participants who were eligible for that program any time from the program's inception in January 2002 to December 2021. There are, fewer records than on the DAF.DMG component because, by the time Ticket program began in 2002, some beneficiaries present on the DAF were no longer entitled to a disability-based SSA benefit and therefore were not eligible for the Ticket program.

Variables in the DAF.Ticket files contain both one-time and event-based data. One-time data include variables that are reported once such as EVER_ACTIVE, which indicates whether a beneficiary has ever assigned their Ticket. Event-based data include variables like TKTMAILDDTn, which indicates the date of Ticket mailing for up to the current maximum occurrences.[8]

The DAF.Ticket component is split into a base file and a set of annual files. The Ticket Base file includes variables from SSA administrative data related to Ticket program administration such as Ticket mailing and assignment dates. The annual files in the DAF.Ticket component contain annual sets of monthly flags constructed by Mathematica to indicate such actions as whether a beneficiary has assigned his or her Ticket in a specific month. It is important to note that the Ticket Base and Ticket Annual files may contain records for a very small number of beneficiaries that are not included in the Demographic components (DMG) or Annual components because the Ticket files are constructed independently from the rest of the DAF and at a later time so that data may include individuals who were not entitled to benefits at the time the DMG and Annual components are created.

D. DAF Payments files

The fourth component of the DAF contains data related to payments from SSA to State VR Agencies and ENs for employment services provided to beneficiaries. The term "providers" in this context refers to State VR Agencies acting in both Ticket and non-Ticket contexts as well as ENs. DAF.Payments contains two types of payments data: one for traditional cost reimbursement payments to state VR agencies and the second for TTW payments to ENs. The payment system

[8] The maximum number of occurrences for this and other "n" variables is specified on the detail page of each variable in Volume 5.

SSA uses to track traditional cost reimbursements to VR agencies changed in 2017. As a result, the payment data for these VR reimbursements is now in two files in two different formats, one that covers the period through February 2017, while the second file contains data from March 2017 onward. The EN payment file contains information on TTW payments to ENs or State VR Agencies acting as ENs under one of the two TTW payment systems, milestone-outcome (MO) or outcome-only (OO). This second type of payment data has been the same since 2002 when TTW began. Unlike VR reimbursement payments, EN payments are usually paid over many months for the same beneficiary. Because of this, and to assist researchers in using EN payment data, EN Payments are stored both as "horizontal" (with one record per beneficiary) and "vertical" (with one record per payment) files. Volume 2 contains more detailed information about DAF payments data and the differences between the various payments files.

E. DAF Data Marts and Extracts

As briefly described in Section II.A of this volume, DAF data marts and extracts have been made based on the full DAF to facilitate use of the DAF. DAF data marts are files that have been developed to streamline the data access and cleaning processes by preselecting groups of beneficiaries often of interest to researchers and providing additional constructed variables for specific analytic purposes. For example, the ADM identifies beneficiaries based on the date of their first award, allowing users to conduct longitudinal analysis from award onward, without having to identify and extract the subpopulation from the full set of files. Volume 2 provides detailed information about the two data marts in DAF21 (the 10% Data Mart and ADM) and the selection criteria used in each.

Extracts are similar to data marts in that they contain subsets of all beneficiaries on the DAF proper, but are different in that the analysis sample is based on special subpopulations of interest to SSA because of their participation in SSA programs, surveys, or demonstrations. These samples were often, though not exclusively, derived for internal SSA purposes to study demonstration participants. Like the data marts, extracts are available to all interested DAF users. For more information about populations for these files, see the DAF website at: https://www.ssa.gov/disabilityresearch/daf.html.

F. DAF Linkable files

There are two sets of DAF-linkable data files that are also available on the SSA mainframe. These include the DAF-linkable version of the RSA-911 files and the MEF.

1. RSA-911

NOTE: Permission from the Department of Education is required before using the RSA data.

Because State VR Agencies are specifically charged with supporting people with disabilities in their attempts to work, administrative files containing information on VR service receipt are of special interest to many researchers working with the DAF. For this reason, it is possible to merge the DAF with the RSA-911 files. The RSA-911 files contain information collected by

State VR Agencies about individuals with whom they have interacted.[9] The data in the RSA-911 files relate to VR participation and services. Users may be interested in linking this information to DAF.DMG data to see the characteristics of SSA beneficiaries using VR services, to DAF.Annual data to see how VR services relate to work and benefit receipt, or to DAF.Payments to get about the payments made from SSA to State VR Agencies for services provided to beneficiaries.

We build a number of files using the RSA data in combination with the DAF; each meets a slightly different research need though we refer to these files collectively as the DAF-RSA files. We describe these files in more detail in Volume 2 of the DAF documentation. Two file types (DAF-Linkable RSA Summary File and RSA-911 Verified SSN File) contain RSA-911 data for all RSA participants, regardless of whether they are contained in the DAF, but no data from the DAF. The other file type (RSA Mini-DAF) contains DAF data, but is limited to those receiving RSA services. All RSA-911 files that are linkable to the DAF contain data on RSA participants whose SSNs (as contained in the RSA-911) files have been verified using a validation process developed by Mathematica and SSA (described in Volume 2).

There were substantial changes to the RSA-911 files beginning in July 2017. The structure of the DAF-linkable RSA files reflects this, with one set of files corresponding to the earlier period and a separate set of files corresponding to the new data structure. The current RSA-911 files were designed to account for changes in service provision by state VR agencies resulting from the Workforce Innovation and Opportunity Act of 2014 (WIOA). RSA-911 files are now collected quarterly, contain open as well as closed cases, and include records for individuals who received pre- Employment Transition Services (ETS) from a VR agency, even if they did not apply for VR services. Prior to July 2017, the RSA-911 files were annual fiscal year files containing information on individuals whose VR cases had closed. In other words, cases were only recorded when they were closed from receiving VR services, meaning an applicant who sought VR services in 2006 would not generally have a closure record until 2008 or later.

Even before the substantial changes in 2017, the RSA-911 source files with the case closure structure changed slightly over time in the characteristics of the elements included, with major shifts in the elements available occurring between 2001 and 2002, and between 2013 and 2014. As a result, some of the data elements in the pre-2017 files have different characteristics depending on whether they were pre-2002, between 2002 and 2013, or in 2014 or later. To simplify data usage, the DAF structure remains constant across these 2002, 2013, and 2014 changes, but this means data some variables will contain missing data for those periods when the information was not collected by RSA, and the range of DAF-RSA values may be different depending on which RSA version was used. Because the changes to the RSA-911 were so substantial beginning July 1, 2017, the DAF-RSA file, uses a completely new format for this latest data. The source file layouts in Volume 12 contain information that would allow users to identify changes in the RSA file structure over this period.

[9] As described more in what follows, the structure of the RSA-911 data fundamentally changed in 2017. Before that time, RSA-911 data were case closure files, with data only for individuals whose cases had been closed. Starting in 2017, RSA-911 data include open cases (and differ in other important ways from the earlier structure).

2. MEF

NOTE: Use of MEF data is restricted to selected staff at SSA.

The MEF is a file containing detailed annual earnings data based on W-2 information collected by the IRS. The MEF extract for DAF contains data from the MEF for DAF beneficiaries as well as individuals who did not receive disability benefits but who are on the RSA-911 files. The MEF contains variables representing earnings between 1987 and the most recent DAF year. There is one record per SSN. See Volume 10 for the layout of this source file.

G. Local economic data from the LAUS and SAIPE

Beginning in the DAF12, SAS formats containing data from the Local Area Unemployment Statistics (LAUS) and Small Area Income and Poverty Estimates (SAIPE) are available for linking to the DAF. These formats contain county-level annual unemployment rates from the LAUS, and county-level monthly poverty rates and median income from the SAIPE. Because data from the LAUS and SAIPE are geography-specific and not person-level, we have stored this information as SAS formats rather than individual-level variables. More information on these economic data and how to access the SAS formats can be found in Volume 2 of the DAF documentation.

H. The DAF Public Use File (PUF)

SSA is now producing a DAF PUF, containing a limited set of variables from the DMG and Annuals components for a random 10% sample of the DAF. The 10 percent file is large enough to avoid disclosure risk and small enough to keep the file size manageable. The PUF contains a more limited set of variables than the full DAF, and we masked some variables in a variety of ways to avoid disclosure. The PUF is available to researchers without access to the full DAF as well as those who are able to access the DAF but may be interested in testing their code or research idea on a smaller sample. The initial version of the DAF-PUF was developed with DAF15, and it has been updated in each DAF version since DAF17. More information and the files available for download can be found at: https://www.ssa.gov/disabilityresearch/daf_puf.html.

Appendix A:

SAS Code for Beneficiary Selection Criteria

The current selection criteria for SSI beneficiaries are implemented with the following SAS code:

```
IF SSN GT '000000000';

IF  AGE < FRA;

IF (UPCASE(CPST) IN ('C01' 'E01' 'E02' 'M01' 'M02'

                'P01' 'S05' 'S06' 'S07' 'S08' 'S09' 'S21' 'T30'

                'T32' 'T33')) OR

  (DENCDE = ' ' AND     /* NEW IN 2004 TO ELIM DENIALS FOR N CODES */

   UPCASE(CPST) IN ('N01' 'N02' 'N04' 'N05' 'N06'

                 'N09' 'N20' 'N22' 'N33' 'N44' 'N54'));

IF UPCASE(MFT) IN ('DI' 'BI' 'DC' 'BC' 'DS' 'BS');
```

The current selection criteria for SSDI beneficiaries are implemented with the following SAS code:

```
IF 18 <= AGE < FRA;

IF ((SUBSTR(BIC,1,1) = 'A' AND TOC IN ('5' '6')) OR

   (SUBSTR(BIC,1,1) = 'C' AND TOC IN ('3' '4' '7')) OR

   (SUBSTR(BIC,1,1) = 'W'));

IF UPCASE(LAF) IN ('AD' 'AS' 'A9' 'C ' 'D ' 'DP' 'DW' 'E ' 'S ' 'SD'

                'S0' 'S1' 'C2' 'S2' 'S6' 'S8');

IF SSN=" " OR SSN="000000000" THEN DELETE;
```

Mathematica Inc.

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

mathematica.org website

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.