

Benefits and Earnings Public-Use File, 2020

User Guide



Social Security Administration
Baltimore, MD, and Washington, DC
statistics@ssa.gov

June 2025 (version 2)

Table of Contents

Background.....	3
Data Overview	3
Data Dictionary and Field Descriptors	3
BEPUF-2020-Benefits (Beneficiary ID and Benefit Information).....	3
BEPUF-2020-Earnings (Beneficiary Annual Earnings Information)	9
Analytical Considerations	9
Comparison to the BEPUF 2004	10
Additional Resources.....	10
Errata	10

Background

The Old-Age, Survivors, and Disability Insurance (OASDI) program provides monthly benefits to qualified retired and disabled workers and their dependents and to survivors of insured workers. Eligibility and benefit amounts are determined by the worker's contributions to Social Security.

The Benefits and Earnings Public-Use File, 2020 (BEPUF 2020) is a set of fully synthetic microdata records statistically representing a self-weighted 10 percent sample of the current adult OASDI beneficiary population as of December 2020. Each record includes benefit-related variables and the taxable earnings history of the beneficiary. All variable values are generated by randomly selecting values from statistical distributions modeled on real values. Notably, as a fully synthetic data set, no record in the released data set can be matched to any real individual beneficiary.

Data Overview

The BEPUF 2020 consists of two files:

- **BEPUF-2020-Benefits:** A primary file with benefit-related variables for each beneficiary.
- **BEPUF-2020-Earnings:** A secondary file containing taxable earnings amounts at each annual age for each beneficiary.

Each file is available in two formats: comma-delimited text (.csv) and SAS data file (.sas7bdat).

Data Dictionary and Field Descriptors

BEPUF-2020-Benefits (Beneficiary ID and Benefit Information)

- The CSV format contains one beneficiary record per line, with field values in the order specified below, separated by commas.
- The equivalent SAS data file contains the same records with fields named as specified below.

<u>Order</u>	<u>Name</u>	<u>Description</u>
1	ID	Synthetic beneficiary identifier A sequential integer enumerating all records starting at 1 and incrementing up to the total number of records. This field is simply an enumeration and contains no information about the beneficiary.
2	BY	4-digit <u>B</u> irth <u>Y</u> ear of the beneficiary 1920 to 2003, inclusive, for a total of 84 possible birth years. All beneficiaries are defined to be born 12:00 AM on 1 January of the BY , such that a given annual age corresponds exactly with one calendar year.
3	SEX	Sex of the beneficiary M : Male F : Female

<u>Order</u>	<u>Name</u>	<u>Description</u>
4	BT	<p><u>B</u>enefit <u>T</u>ype to which the beneficiary is entitled</p> <p>The BT is based on either the beneficiary's own earnings history or the earnings history of a related beneficiary as follows:</p> <p>A: Worker. Entitled on the beneficiary's own earnings history.</p> <p>B: Spouse. Entitled on the earnings history of a living spouse of the beneficiary, including a divorced living spouse.</p> <p>C: Child. Entitled on the earnings history of a parent of the beneficiary. Only child beneficiaries designated as Disabled Adult Children (DAC) are included in the BEPUF 2020 (see DAC Indicator on page 6).</p> <p>D: Widow(er). Entitled on the earnings history of a deceased spouse of the beneficiary, including a divorced deceased spouse.</p> <p>E: Mother/Father. Entitled on the earnings history of a deceased spouse of the beneficiary, including a divorced deceased spouse, and the beneficiary's current status as caring for a child either under age 16 or disabled.</p> <p>F: Disabled Widow(er). Entitled on the earnings history of a deceased spouse of the beneficiary, including a divorced deceased spouse, and the beneficiary's current status as disabled.</p>
5	IP	<p><u>I</u>nsurance <u>P</u>rogram funding the benefit</p> <p>R: Retirement. Funded by the Old-Age and Survivors Insurance (OASI) program.</p> <p>D: Disability. Funded by the Disability Insurance (DI) program.</p> <p>S: Survivor. Funded by the OASI program.</p>
6	BT2	<p><u>B</u>enefit <u>T</u>ype of the secondary benefit for dually entitled beneficiaries</p> <p>A dually entitled beneficiary is a Worker (BT = A) receiving either a Retirement benefit (IP = R) or a Disability benefit (IP = D) and a secondary Benefit Type (BT2) of B, C, D, E, or F.</p> <p>If a beneficiary is not dually entitled, BT2 is undefined and is indicated in the file with a '-' (hyphen).</p> <p>There is no corresponding IP value for BT2 in the BEPUF 2020.</p>

NOTE: Fields list continues on page 6.

COMPOSITE BENEFIT CLASS (BC) INDICATOR

For the convenience of reference, program rule specification, and validation, the BC is a composite indicator of the full benefit description. It does not itself appear in the data file but is composed of a beneficiary's **BT**, **IP**, and **BT2** values, for which there are 20 valid combinations as follows in the table below.

BC Code	Description	BT	IP	BT2	Type
AR	Retired Worker (not dually entitled)	A	R	-	Worker only
ARB	Retired Worker dually entitled to a Spouse benefit	A	R	B	Dually entitled
ARC	Retired Worker dually entitled to a Child benefit	A	R	C	Dually entitled
ARD	Retired Worker dually entitled to a Widow(er) benefit	A	R	D	Dually entitled
ARE	Retired Worker dually entitled to a Mother/Father benefit	A	R	E	Dually entitled
ARF	Retired Worker dually entitled to a Disabled Widow(er) benefit	A	R	F	Dually entitled
AD	Disabled Worker (not dually entitled)	A	D	-	Worker only
ADB	Disabled Worker dually entitled to a Spouse benefit	A	D	B	Dually entitled
ADC	Disabled Worker dually entitled to a Child benefit	A	D	C	Dually entitled
ADD	Disabled Worker dually entitled to a Widow(er) benefit	A	D	D	Dually entitled
ADE	Disabled Worker dually entitled to a Mother/Father benefit	A	D	E	Dually entitled
ADF	Disabled Worker dually entitled to a Disabled Widow(er) benefit	A	D	F	Dually entitled
BR	Spouse of a Retired Worker	B	R	-	Nonworker
BD	Spouse of a Disabled Worker	B	D	-	Nonworker
D	Widow(er). <i>All are Survivors (IP = S).</i>	D	S	-	Nonworker
E	Mother/Father. <i>All are Survivors (IP = S).</i>	E	S	-	Nonworker
F	Disabled Widow(er). <i>All are Survivors (IP = S).</i>	F	S	-	Nonworker
CR	Child of living Retired Worker	C	R	-	Nonworker
CD	Child of living Disabled Worker	C	D	-	Nonworker
CS	Child of deceased Worker (who was either Retired or Disabled)	C	S	-	Nonworker

Of the above BC designations, AR and AD beneficiaries are *worker only* because their entire benefit is attributable to their own earnings histories. The BCs that are neither worker only nor dually entitled are *nonworker* because the entire benefit is attributable to the earnings history of someone other than the beneficiary.

For convenience, the additional codes AR_A and AD_A specify groupings of respective Worker BCs as follows:

- **AR_A:** all Retired-Worker BCs regardless of dual-entitlement status (AR, ARB, ARC, ARD, ARE, and ARF), and
- **AD_A:** all Disabled-Worker BCs regardless of dual-entitlement status (AD, ADB, ADC, ADD, ADE, and ADF).

NOTE: Fields list continued from page 4.

- The following 6 fields are *Indicators* that have values TTrue (T) or FFalse (F).
- Specific rules apply to their assignment based on **BT**, age, and other factors. For dually entitled beneficiaries, an Indicator can be derived from either the **BT** or the **BT2**, except as noted below.
- Not all Indicators are applicable to every BC. The following descriptions note the BCs to which each indicator applies. For nonapplicable BCs, the Indicator value is always **False (F)**.

<u>Order</u>	<u>Name</u>	<u>Description</u>
7	REDUX	Benefit amount is <i>reduced</i> based on age For dually entitled beneficiaries, this applies only to the BT portion of the benefit, not BT2 (see also MB C and W AX on pages 7 and 8). <i>Applicable BCs: AR_A, AD_A, BR, BD, D</i>
8	CIC	The beneficiary has at least one <u>C</u> hild- <u>i</u> n- <u>C</u> are As a condition of receiving the benefit, the beneficiary is caring for an insured worker's child who is under age 16 or disabled. <i>Applicable BCs: BR, BD, ARB, ADB, E, ARE, ADE</i>
9	DRC	The beneficiary is receiving <u>D</u> elayed <u>R</u> etirement <u>C</u> redits <i>Applicable BC: AR_A</i>
10	DIVORCED	The beneficiary is <i>divorced</i> from a former spouse, living or deceased, on whose earnings history the benefit is based. <i>Applicable BCs: BR, BD, ARB, ADB, D, ARD, ADD, E, ARE, ADE, F, ARF, ADF</i>
11	REMARRIED	The beneficiary is <i>remarried</i> after the death of a previous spouse on whose earnings history the benefit is based. <i>Applicable BCs: D, ARD, ADD</i>
12	DAC	The beneficiary is a <u>D</u> isabled <u>A</u> ddult <u>C</u> hild Child beneficiaries aged 18–19 <i>may</i> be specified as DACs, however beneficiaries older than age 19 receiving a Child benefit of any type are <i>always</i> specified as DACs. Only DAC-designated Child beneficiaries appear in the BEPUF 2020. <i>Applicable BCs: CR, CD, CS, ARC, ADC</i>

Order **Name****Description****13****ACE****Age at Current Entitlement**

The benefit claim age in years. (If there are multiple claims for a beneficiary, this is from the most recent claim.)

The range of valid **ACE** values is highly dependent on **BT**, **BY**, and other factors.

The specified **ACE** ranges in the table below are used in the BEPUF 2020 synthesis and are, for both practical and disclosure reasons, slightly more constrained than actual program claiming rules allow.

TABLE NOTES: FRA is the Full (or Normal) Retirement Age and is solely defined by birth year (see <https://www.ssa.gov/OACT/ProgData/nra.html>). Ranges are shown using interval notation; a parentheses indicates that the range endpoint is not included in the interval whereas a bracket indicates that the range endpoint is included in the interval.

Included BCs	ACE Range
AR_A with Redux and no DRC	[62, FRA)
AR_A with DRC and no Redux	(FRA, 75]
AR_A with both Redux and DRC	[62, 75]
AD_A with no Redux	[18, FRA]
AD_A with a Redux	[55, FRA)
BR and BD with no CIC	[62, 100]
BR and BD with CIC	[15, FRA]
D	[60, 100]
E	[15, FRA]
F	[50, 60]
CR, CD, and CS	[18, 67]

Because the BEPUF 2020 includes synthesized earnings through 31 December 2020, all synthetic beneficiaries are assumed to have survived the entire calendar year 2020 and have reached their current age on 1 January 2021. This allows some practical assumptions for claiming benefits. For example, a Retired Worker cannot claim benefits, and thus become a beneficiary, until age 62. For those Retired Workers born in 1959 who, still aged 61 at the close of 2020, decide to claim their benefits at age 62, allowing for survival past midnight to 1 January 2021 enables them to have effectively claimed their Retired-Worker benefit on their 62nd birthday and to qualify for inclusion in the BEPUF with a valid **ACE**.

14**MBC****Monthly Benefit Credited**

The total benefit amount as of December 2020 that the beneficiary receives each month.

The **MBC** is top-coded at \$3,200, which is the approximate 99th percentile of all **MBC** values in the real 100 percent beneficiary population.

Since all beneficiaries must have some benefit amount (or they would not be beneficiaries), the **MBC** range is (0, \$3,200], or greater than 0 and less than or equal to \$3,200.

<u>Order</u>	<u>Name</u>	<u>Description</u>
15	WAX	<p>Worker-Attributable Fraction of the MBF.</p> <p>For worker-only beneficiaries (AR or AD), this value is 100, indicating 100 percent of the benefit amount is attributable to the beneficiary's own earnings history.</p> <p>By definition, dually entitled beneficiaries have some worker-attributable benefit and some secondary benefit, thus their WAX range is (0, 100), or greater than 0 and less than 100.</p> <p>Similarly, the WAX is 0 for all nonworker beneficiaries.</p>
16	nEarns	<p>Number of ages with earnings for the beneficiary in the Earnings file (<i>see page 9</i>).</p> <p>The nEarns is included only as a convenience since it is calculated directly from the synthesized earnings figures in the Earnings file.</p>
17	AIME	<p><u>A</u>verage <u>I</u>ndexed <u>M</u>onthly <u>E</u>arnings</p> <p>See https://www.ssa.gov/OACT/COLA/Benefits.html#aime for computation details.</p> <p>AIME is included only as a convenience since it is calculated directly from the synthesized earnings figures in the Earnings file (<i>see page 9</i>).</p>
18	PIA	<p>Current <u>P</u>rimary <u>I</u>nsurance <u>A</u>mount</p> <p>This is calculated from the AIME and the 2021 bend points so it can include earnings from 2020.</p> <p>See https://www.ssa.gov/OACT/COLA/Benefits.html#aime for computation details.</p> <p>PIA is included only as a convenience since it is calculated directly from the AIME, which is itself calculated directly from the synthesized earnings figures in the Earnings file (<i>see page 9</i>).</p>

BEPUF-2020-Earnings (Beneficiary Annual Earnings Information)

- This file contains the annual OASDI-taxable earnings amounts for each beneficiary defined in the Benefits file.
- The file contains records only for ages for which the beneficiary had earnings above 0. Therefore, a single beneficiary defined in the Benefits file can have 0, 1, or many related records in the Earnings file.
 - Earnings records are limited to ages 15–84 (inclusive), primarily for disclosure avoidance.
 - Earnings records before 1951 are not included because SSA does not retain annual earnings amounts from those years.
- In the CSV format, each line (or record) contains the nominal taxable earnings amount for a single annual age of a single beneficiary, with field values in the order specified below, separated by commas.
- The equivalent SAS data file contains the same records with fields named as specified below.

<u>Order</u>	<u>Name</u>	<u>Description</u>
1	ID	The beneficiary identifier defined in the Benefits file (<i>see page 3</i>)
2	AGE	Annual age of the beneficiary during which the EARNINGS amount was earned
3	YEAR	Calendar year for which the beneficiary was of the AGE above YEAR is included only as a convenience since it can be easily derived as the beneficiary's BY + AGE .
4	EARNINGS	Synthesized value of taxable earnings the beneficiary earned at AGE , in nominal dollars for the YEAR This value is always top-coded at the taxable maximum for the YEAR . (See https://www.ssa.gov/OACT/COLA/cbb.html for taxable maximum details.)

Analytical Considerations

- The BEPUF data reflect a representative sample of the full 100 percent OASDI beneficiary population.
- Production of synthetic data (synthesis) is essentially the generation of records that are statistically representative of the real population but without containing any real person's actual record. The records are generated (or synthesized) by randomly selecting values for the record variables from statistical models developed with original data. As such, no record in the synthetic BEPUF 2020 population can be traced to, or even began as, any actual record in the SSA administrative data sources.
- This data set can be used as a general population or as various subpopulations for both modeling input and statistical inference.

Comparison to the BEPUF 2004

- The previous version of the BEPUF was released in 2005 and produced from SSA administrative data sources as of the close of 2004. See <https://www.ssa.gov/policy/docs/microdata/earn/index.html> for further information.
- The BEPUF has not been updated since release of the 2004 version until the present 2020 version.
- The 2004 and 2020 BEPUF releases consist mostly of the same information but formatted differently.
- In contrast to the synthetic methods used to produce the BEPUF 2020, the BEPUF 2004 used “anonymization” techniques to limit the risk of disclosure of personally identifiable information. These techniques include rounding and limiting the ranges of the values of sensitive variables, especially earnings and benefit amounts, in actual administrative records. Thus, each resulting anonymized record is derived from the administrative record of a real person, with values altered to conceal identity yet still providing overall statistical value.
- Anonymization techniques are now being replaced by more robust methods including synthesis, which by itself virtually eliminates the potential for disclosure.

Additional Resources

- *Annual Statistical Supplement to the Social Security Bulletin*
<https://www.ssa.gov/policy/docs/statcomps/supplement/index.html>
 - Program descriptions and legislative history
 - Program provisions
 - Glossary
 - Abbreviations
- *Social Security Handbook*
https://www.ssa.gov/OP_Home/handbook/handbook.html

Errata

- June 4, 2025:
 - In the original version of this Guide, we incorrectly used the term “covered earnings” where we should have used “taxable earnings.” The corrected language now appears.
 - Also, in the original Analytical Considerations section, our description of omitted earnings records for years before 1951 was imprecise. The corrected language now appears in the Data Dictionary and Field Descriptors section under BEPUF-2020-Earnings (Beneficiary Annual Earnings Information).