

Chronic Conditions Warehouse

Your source for national CMS Medicare and Medicaid research data



Chronic Conditions Warehouse Virtual Research Data Center

Submission of Researcher Data Finder Files

MARCH 2024 | VERSION 2.4

This page intentionally left blank.

Revision Log

Date	Changed by	Revisions	Version
March 2024	S. Pietzsch	Added a hashing and tokenization section at CMS' request	2.4
July 2023	M. Richardson	Updated address in the "Ship Encrypted Finder Files to:" section	2.3
May 2023	S. Pietzsch	Added Last Name length for the Medicaid data and corrected the Last Name length for the Medicare data.	2.2
February 2023	S. Pietzsch	Enhanced 2.2	2.1
July 2022	S. Pietzsch D. Happe	Updated to new standards and new template	2.0

Table of Contents

1.0 Overview	1
2.0 Finder File Encryption Policy	1
2.1 Encryption Requirements.....	1
2.2 Software Options	1
2.3 Password and Shipping Requirements.....	1
2.4 Ship Encrypted Finder Files to:	1
2.5 Hashing and Tokenization	2
3.0 Finder File Variables	2
4.0 Expected Length and Format of Finder File Variables	3
4.1 Primary Identifiers	3
4.2 Secondary Identifiers	3
4.3 Non-Identifying Variables.....	3
4.4 BENE_ID Finder File.....	3
4.4.1 Finder File Variable Name and Format:	4
4.4.1.1 When submitting a BENE_ID finder file, adhere to the following guidelines:	4
4.5 HICN Finder File	4
4.5.1 Finder File Variable Name and Format:	4
4.5.1.1 When submitting a HICN finder file, adhere to the following guidelines:.....	4
4.6 SSN Finder File	4
4.6.1 Finder File Variable Name and Format:	5
4.6.1.1 When submitting an SSN finder file, adhere to the following guidelines:.....	5
4.7 MBI Finder File.....	5
4.7.1 Finder File Variable Name and Format:	5
4.7.1.1 When submitting an MBI finder file, adhere to the following guidelines:	5
4.8 MSIS_ID/STATE_CD Finder File.....	6
4.8.1 Finder File Variable Name and Format	6
4.8.1.1 When submitting a MSIS_ID/STATE_CD finder file, adhere to the following guidelines:	6
4.9 RES_ID/STATE_CD Finder File.....	6
4.9.1 Finder File Variable Name and Format:	6
4.9.1.1 When submitting a RES_ID/STATE_CD finder file, adhere to the following guidelines:	6
4.10 NPI or TIN Finder File	7
4.10.1 Finder File Variable Name and Format:	7
4.10.1.1 When submitting an NPI or TIN finder file, adhere to the following guidelines:.....	7
4.11 Secondary Identifier Finder File.....	7
4.11.1 Finder File Variable Name and Format:	7
4.11.1.1 When submitting a Secondary Identifier finder file, adhere to the following guidelines:	7
4.12 ICD Code, CPT/HCPCS Code, DRG Code, or NCD Finder Files	8

4.12.1 Finder File Variable Name and Format:	8
4.12.1.1 When submitting an ICD Code, CPT/HCPCS Code, DRG Code, NDC, or PRCDR Code finder file, adhere to the following guidelines:	8
5.0 Where to Get Assistance	9
Appendix A — Tip Sheet for Leading Zero Retention in Microsoft Excel	10
Appendix B — List of Acronyms	14

1.0 Overview

The purpose of this document is to provide requirements for submitting a finder file to the Chronic Conditions Warehouse (CCW) for fulfillment of an approved research data request. A finder file contains individual level Personally Identifiable Information (PII) and is used to search the CCW database for specific data.

2.0 Finder File Encryption Policy

This section describes the minimum standards of data security that all finder files must meet before requesters send them to the CCW.

2.1 Encryption Requirements

To maintain and ensure the security of identifiable information shipped to the CCW, requesters must encrypt all finder files containing Personally Identifiable Information with at least an Advanced Encryption Standard (AES) 256-bit encryption algorithm. This is a block cipher adopted as an encryption standard by the U.S. government. The National Security Agency (NSA) has deemed the use of a 256-bit AES encryption algorithm as secure enough for U.S. government top secret information.

2.2 Software Options

Requesters must encrypt files using software programs capable of AES 256-bit encryption. The CCW team is unable to install all compliant encryption tools; therefore, the finder file must be:

- A self-decrypting archive (SDA) or the file is WinZip (version < 12.0) or 7-Zip compatible, or
- Must decrypt on operating systems available to the CCW Amazon Web Services (AWS) environment, i.e., Red Hat Enterprise Linux (RHEL) or Windows, using a password only. Other versions of Linux (e.g., Ubuntu, Debian, Fedora, SuSE) may be compatible with RHEL but are not guaranteed to be compatible.

The CCW does not have other platforms available. Submitting executables that require installation of software results in delay or denial of the requester's finder file.

2.3 Password and Shipping Requirements

Protect the files with a password that is at least 10 digits in length that must contain a combination of letters and numbers. Individuals must send the encrypted finder file by a courier service with tracking capabilities, such as FedEx, UPS, DHL, or registered mail. Do not include the password along with the data shipment, but instead, email the password and the Data Use Agreement (DUA) number that the finder file belongs to CCW at CMSdata@gdit.com. Inclusion of the password with the data shipment triggers a Centers for Medicare & Medicaid Services (CMS) IT security incident.

2.4 Ship Encrypted Finder Files to:

Shannon Pietzsch

General Dynamics Information Technology
Attn: CCW Research Coordinator
4800 Westown Parkway, Suite 360
West Des Moines, IA 50266

2.5 Hashing and Tokenization

The CCW does not currently allow receipt of hashed or tokenized finder files. CCW is exploring this functionality for availability in the future.

3.0 Finder File Variables

The CCW team can extract Medicare and Medicaid data from the CCW for individual beneficiaries using finder files submitted with data requests approved by CMS. Finder files may consist of the following variables:

- BENE_ID — unique CCW beneficiary ID used to link all CMS data for a single individual
- HICN — Health Insurance Claim number
- SSN — Social Security number
- MBI — Medicare Beneficiary Identifier
- MSIS_ID/STATE_CD — state-assigned unique ID for Medicaid Children’s Health Insurance Program (CHIP) enrolled beneficiaries per state, per year
- RES_ID/STATE_CD — identifies resident in the national repository per state
- NPI — National Provider Identifier
- EIN/TIN — Employer Identification Number/Tax Identification Number
- DOB — date of birth
- Secondary Identifiers* — partial SSN, partial HICN, last name, sex, DOB, ZIP code
- ICD-9/10 code — International Classification of Diseases diagnosis codes (ICD-10 replaced ICD-9 10/01/2015. Individuals can submit ICD-9 codes as finder files for older years of data)
- CPT/HCPCS code — Current Procedural Terminology code/Healthcare Common Procedure Coding System
- DRG Code — diagnosis-related group code
- NDC — National Drug Code
- PRCDR code — procedure code
- STATE_CD — state abbreviation code
- ZIP code — regional postal code (either five- or nine-digit codes are acceptable)

***NOTE:** The CCW team has not found match rates on last name to be reliable search criteria and users should be prepared for a lower match rate. Partial SSNs should be the last four or last six digits of the SSN while the partial HICN is the nine digits minus the alphanumeric identifier. When submitting a partial SSN, partial HICN, or last name finder file, individuals must include at least two other identifiers from DOB, sex, or ZIP code. Should the search result in more than one match, the CCW team discards both from the final beneficiary cohort.

4.0 Expected Length and Format of Finder File Variables

Below is a list of commonly submitted finder file variables and their expected lengths and formats. Complying with the length and format ensures the CCW team can search for the database identifier. Later, this document addresses each finder file individually. Whenever leading zeros appear in a variable, the requester should preserve them (i.e., SSN, partial SSN, DOB, etc.). Reference [Appendix A — Tip Sheet for Leading Zero Retention in Microsoft Excel](#).

4.1 Primary Identifiers

- BENE_ID — 15-character alphanumeric identifier
- SSN — nine-digit identifier
- HICN — maximum 12-character alphanumeric identifier
- MBI — 11-character alphanumeric identifier
- MSIS_ID — maximum 32-character alphanumeric identifier (always paired with state code)
- RES_ID — 10-character alphanumeric identifier (always paired with state code)
- STATE_CD — two-character state abbreviation code
- NPI — 10-digit numeric identifier
- TIN — nine-digit numeric identifier

4.2 Secondary Identifiers

- Partial SSN — four- or six-digit numeric identifier; must be last four or six digits of full SSN
- Partial HICN — nine-digit numeric identifier; must be first nine digits of full HICN
- Last Names — max 24-character identifier for Medicare and 30-character identifier for Medicaid with non-alphanumeric characters removed
- Sex — male/M = 1, female/F = 2, and blank/missing = 0
- DOB — 10-character date variable with format MM/DD/YYYY
- ZIP Code — five- or nine-digit numeric postal code

4.3 Non-Identifying Variables

- ICD-9/10 Code — seven-character alphanumeric diagnosis code
- CPT/HCPCS Code — five-character alphanumeric billing code
- DRG Code — three-digit numeric diagnosis group code
- NDC — 11-digit numeric drug code
- PRCDR — seven-character alphanumeric procedure code

4.4 BENE_ID Finder File

The CCW team includes a unique CCW beneficiary identifier (BENE_ID) in each data file delivered as part of the output package, thus allowing linkage of an individual's data across data sources and types. The assignment of BENE_IDs is unique for each DUA.

To continue studying beneficiaries from previous requests, a submission of a BENE_ID finder file can be an option. A BENE_ID finder file must include one record per BENE_ID. A BENE_ID finder file does not need the sex and DOB identifiers.

4.4.1 Finder File Variable Name and Format:

1. BENE_ID: 15-character alphanumeric identifier of a previously shipped beneficiary

4.4.1.1 When submitting a BENE_ID finder file, adhere to the following guidelines:

- BENE_IDs should be 15-character length and only contain letters and numbers. In this case, the CCW team has previously encrypted the BENE_ID based on the researcher's DUA
- Finder files must be in CSV (ASCII comma delimited) format
- A carriage return must end each line
- Left-justified entries without any leading/trailing/embedded spaces

4.5 HICN Finder File

Submitting a HICN finder file requires a list of beneficiaries' HICNs. The CCW team extracts data for all Medicare beneficiaries identified by the submitted HICNs. The requester may also include an individual's sex and DOB information in the HICN finder file. While submitting these additional variables does not change the process of strictly matching to HICN, the crosswalk returned to the requester includes match flags on HICN, sex, and DOB variables.

The requester does have the option of restricting the linkage of HICN finder file beneficiaries to include only those who strictly match on their sex and/or DOB as well as on HICN. If the requester would like to have these additional selection criteria imposed, they must clearly state this on the RIF Specification Worksheet located on the Research Data Assistance Center's (ResDAC's) website at www.resdac.org.

4.5.1 Finder File Variable Name and Format:

1. HICN: Maximum 12-character alphanumeric identifier for both Railroad Retirement Board (RRB) and Social Security Administration (SSA) formats
2. SEX: Male = 1, Female = 2, unknown = blank/missing entry
3. DOB: 10-character date variable with format MM/DD/YYYY
(SEX and DOB are optional variables)

4.5.1.1 When submitting a HICN finder file, adhere to the following guidelines:

- HICNs containing only letters and numbers with leading zeros intact. Reference [Appendix A — Tip Sheet for Leading Zero Retention in Microsoft Excel](#).
- Exclude all dashes in the HICN
- Sex containing only 1 (male), 2 (female), and blank/missing (unknown/missing)
- DOB in month/day/year format with slashes intact
- Finder files must be in CSV (ASCII comma delimited) format
- A carriage return must end each line
- Left-justified entries without any leading/trailing/embedded spaces

4.6 SSN Finder File

Submitting an SSN finder file requires a listing of beneficiaries' SSNs. The CCW team extracts data for all beneficiaries identified by the submitted SSNs. Prior to searching the data, the CCW team removes invalid SSNs from the finder file but denotes invalid SSNs in the crosswalk as SSN_MATCH = 9. The requester may also include an individual's sex and DOB information in the SSN finder file. While submitting these additional variables does not change the process of strictly matching to SSN, the crosswalk returned to the requester includes match flags on SSN, sex, and DOB variables.

The requester does have the option of restricting the linkage of SSN finder file beneficiaries to include only those who strictly match on their sex and/or DOB, as well as on SSN. If the requester desires to have these additional selection

criteria imposed, they must clearly state this on the RIF Specification Worksheet located on ResDAC's website at www.resdac.org.

The CCW uses the SSA's validation rules found here:

- <https://secure.ssa.gov/poms.nsf/lnx/0110201035>

4.6.1 Finder File Variable Name and Format:

1. SSN: nine-digit character identifier
 2. SEX: male = 1, female = 2, unknown = blank/missing entry
 3. DOB: 10-character date variable with format MM/DD/YYYY
- (SEX and DOB are optional variables)

4.6.1.1 When submitting an SSN finder file, adhere to the following guidelines:

- SSNs with a nine-digit number, formatted as a character variable, with leading zeros intact. Reference [Appendix A — Tip Sheet for Leading Zero Retention in Microsoft Excel](#).
- Exclude all dashes in SSN
- Sex containing only 1 (male), 2 (female), and blank/missing (unknown/missing)
- DOB in month/day/year format with slashes intact
- Finder files must be in CSV (ASCII comma delimited) format
- A carriage return must end each line
- Left-justified entries without any leading/trailing/embedded spaces

4.7 MBI Finder File

Submitting an MBI finder file requires a listing of beneficiaries' MBIs. The CCW team extracts data for all Medicare beneficiaries identified by the submitted MBIs. The requester may also include an individual's sex and DOB information in the MBI finder file. While submitting these additional variables does not change the process of strictly matching to MBI, the crosswalk returned to the requester includes match flags on MBI, sex, and DOB variables.

The requester does have the option of restricting the linkage of MBI finder file beneficiaries to include only those who strictly match on their sex and/or DOB as well as on MBI. If the requester would like to have these additional selection criteria imposed, they must clearly state this on the RIF Specification Worksheet.

4.7.1 Finder File Variable Name and Format:

1. MBI: 11-character alphanumeric identifier
 2. SEX: male = 1, female = 2, unknown = blank/missing entry
 3. DOB: 10-character date variable with format MM/DD/YYYY
- (SEX and DOB are optional variables)

4.7.1.1 When submitting an MBI finder file, adhere to the following guidelines:

- MBIs with 11-character length, only containing letters and numbers, and have leading zeros intact. Reference [Appendix A — Tip Sheet for Leading Zero Retention in Microsoft Excel](#).
- Exclude all dashes in MBI
- Sex containing only 1 (male), 2 (female), and blank/missing (unknown/missing)
- DOB in month/day/year format with slashes intact
- Finder files must be in CSV (ASCII comma delimited) format
- A carriage return must end each line
- Left-justified entries without any leading/trailing/embedded spaces

4.8 MSIS_ID/STATE_CD Finder File

Submitting an MSIS_ID/STATE_CD finder file requires a listing of beneficiaries' MSIS_ID/STATE_CDs. The MSIS_ID is only unique to a beneficiary within a given state, for a given year. Each state is responsible for assigning MSIS_IDs to Medicaid recipients in that state. Therefore, it's possible for MSIS_ID formats to vary in length and number/letter combination by state. A beneficiary's MSIS_ID may change over the years and vary by state. A beneficiary's BENE_ID may apply to multiple MSIS_ID/STATE_CDs.

To continue studying beneficiaries from previous requests or from state issued Medicaid data, a submission of a MSIS_ID/STATE_CD finder file can be an option. The CCW team extracts data for all Medicaid beneficiaries identified by the submitted MSIS_ID/STATE_CD within a given yearly range. Requesters do not need to include sex and DOB for an MSIS_ID/STATE_CD finder file.

4.8.1 Finder File Variable Name and Format

1. Encrypted MSIS_ID*: max 32-character alphanumeric identifier

OR

1. Unencrypted MSIS_ID: max 20-character alphanumeric identifier
2. STATE_CD: two-character state abbreviation

*When submitting an encrypted MSIS_ID finder file, the requester must give the DUA number the CCW team previously extracted the IDs.

4.8.1.1 When submitting a MSIS_ID/STATE_CD finder file, adhere to the following guidelines:

- MSIS_IDs, encrypted and unencrypted, always paired with STATE_CD
- Encrypted MSIS_IDs should be a max 32-character length and only contain letters and numbers
- Unencrypted MSIS_IDs should be a max 20-character length and only contain letters and numbers
- STATE_CD should contain two-character state abbreviation
- Finder files must be in CSV (ASCII comma delimited) format
- A carriage return must end each line
- Left-justified entries without any leading/trailing/embedded spaces

4.9 RES_ID/STATE_CD Finder File

Submitting a RES_ID/STATE_CD finder file requires a list of beneficiaries' RES_ID with their corresponding state codes. The CCW team extracts data for all Medicare beneficiaries identified by the submitted RES_ID/STATE_CD.

4.9.1 Finder File Variable Name and Format:

1. RES_ID: 10-character alphanumeric identifier
2. STATE_CD: two-character state abbreviation

4.9.1.1 When submitting a RES_ID/STATE_CD finder file, adhere to the following guidelines:

- RES_ID with 10-character length and only contain letters and numbers
- STATE_CD containing two-character state abbreviation
- Finder files must be in CSV (ASCII comma delimited) format
- A carriage return must end each line
- Left-justified entries without any leading/trailing/embedded spaces

4.10 NPI or TIN Finder File

Submitting an NPI or TIN finder file only requires a listing of the identifier. The CCW team extracts data for all Medicare beneficiaries identified by the submitted identifier. The CCW team does not return crosswalks for these types of finder files.

4.10.1 Finder File Variable Name and Format:

1. NPI: 10-digit numeric identifier

OR

1. TIN: nine-digit numeric identifier

4.10.1.1 When submitting an NPI or TIN finder file, adhere to the following guidelines:

- No dashes in any identifier
- Finder files must be in CSV (ASCII comma delimited) format
- A carriage return must end each line
- Left-justified entries without any leading/trailing/embedded spaces

4.11 Secondary Identifier Finder File

There are some additional requirements when sending secondary identifiers to ensure the BENE_IDs returned are the beneficiaries approved for the study. In addition to last name, partial SSN, or partial HICN, requesters must submit at least two other identifiers from DOB, sex, or ZIP code. Unlike the other finder files, secondary identifier finder files use a strict match on all the variables submitted. A strict match requires that all submitted identifiers match the CCW records. If after the strict match, more than one BENE_ID results, then the CCW team excludes all BENE_ID matches from the cohort.

4.11.1 Finder File Variable Name and Format:

1. Partial SSN: four- or six-digit character identifier

OR

2. Partial HICN: nine-digit numeric identifier

OR

3. Last Name: maximum 24-character alphanumeric identifier for Medicare beneficiaries
maximum 30-character alphanumeric identifier for Medicaid enrollees

AND (at least two of the three below)

4. SEX: male = 1, female = 2, unknown = blank/missing entry
5. DOB: 10-character date variable with format MM/DD/YYYY
6. ZIP: five- or nine-digit numeric postal code

4.11.1.1 When submitting a Secondary Identifier finder file, adhere to the following guidelines:

- All Partial SSNs should be a four- or six-digit number, formatted as a character variable, from the last four or six digits of the full SSN with leading zeros intact. Reference [Appendix A — Tip Sheet for Leading Zero Retention in Microsoft Excel](#).

- All Partial HICNs should be nine-character length from the first nine characters of full HICN
- Partial HICNs should only contain numbers and keep leading zeros intact. Reference [Appendix A — Tip Sheet for Leading Zero Retention in Microsoft Excel](#).
- Last names can have a max length of 24-characters for Medicare and 30-characters for Medicaid with non-alphanumeric characters removed. For best match results, the CCW team recommends researchers remove accents from names.
- No dashes in any identifier
- Double quote fields containing a comma (i.e., addresses) to avoid SAS treating the comma as a delimiter
- Sex containing only 1 (male), 2 (female), and blank/missing (unknown/missing)
- DOB in month/day/year format
- ZIP codes should be either a five- or nine-digit number with leading zeros intact. [Appendix A — Tip Sheet for Leading Zero Retention in Microsoft Excel](#).
- Finder files must be in CSV (ASCII comma delimited) format
- A carriage return must end each line
- Left-justified entries without any leading/trailing/embedded spaces

4.12 ICD Code, CPT/HCPCS Code, DRG Code, or NCD Finder Files

Submitting an ICD Code, CPT/HCPCS Code, DRG Code, or NCD finder file only requires a listing of the code(s). The CCW team extracts data for all beneficiaries identified by the submitted codes.

4.12.1 Finder File Variable Name and Format:

1. ICD: seven-character alphanumeric diagnosis code

OR

1. CPT/HCPCS: five-character alphanumeric billing code

OR

1. DRG: three-digit numeric diagnosis group code

OR

1. NDC: 11-digit numeric drug code

OR

1. PRCDR: seven-character alphanumeric procedure code

4.12.1.1 When submitting an ICD Code, CPT/HCPCS Code, DRG Code, NDC, or PRCDR Code finder file, adhere to the following guidelines:

- Keep all leading zeros in any code. Reference [Appendix A — Tip Sheet for Leading Zero Retention in Microsoft Excel](#).
- No dashes or periods in any code
- Left-justified entries without any leading/trailing/embedded spaces

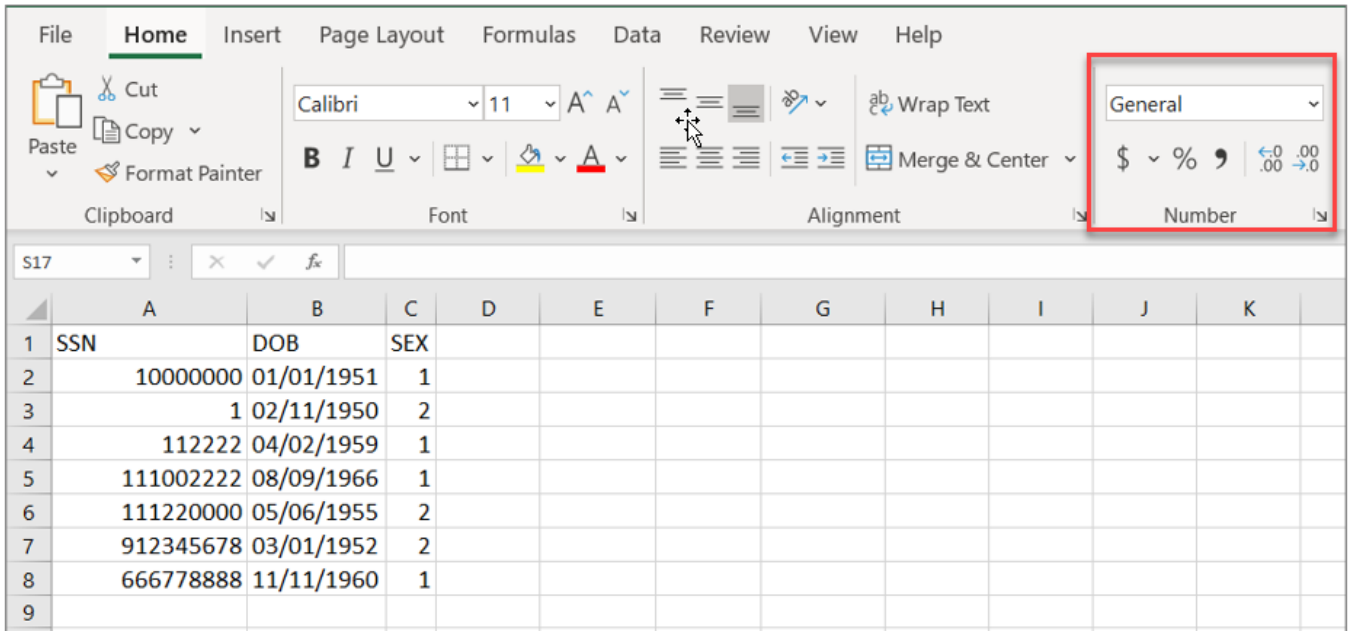
5.0 Where to Get Assistance

The CCW Help Desk staff provide assistance between 8:00 am to 5:00 pm ET, Monday through Friday (excluding most federal holidays). Contact the CCW Help Desk at ccwhelp@ccwdata.org or 1-866-766-1915.

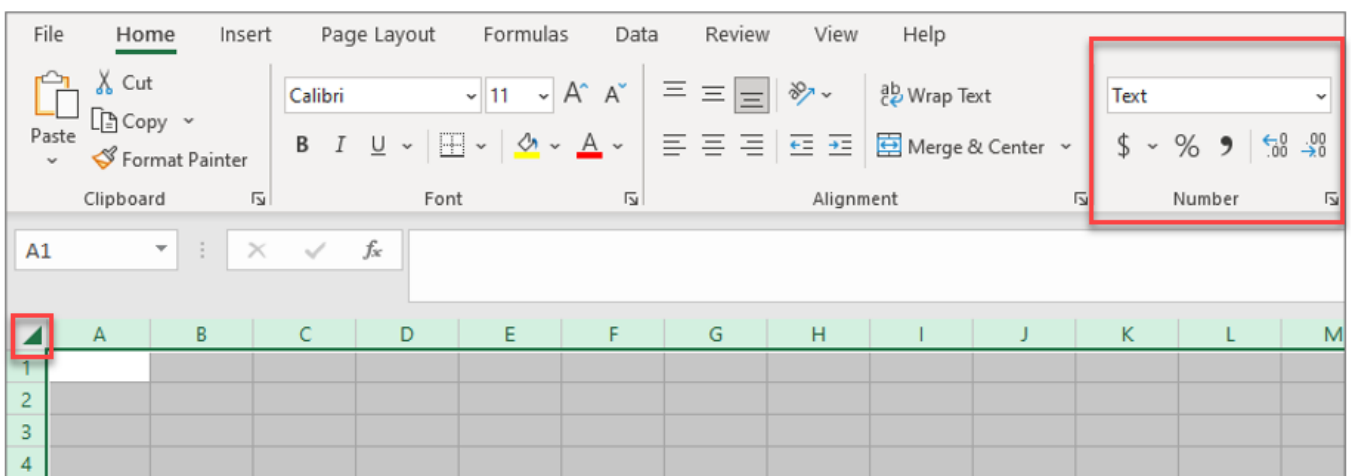
Appendix A — Tip Sheet for Leading Zero Retention in Microsoft Excel

NOTE: The images below use Microsoft 365 and the 2011 version of Excel. All SSN data are invalid.

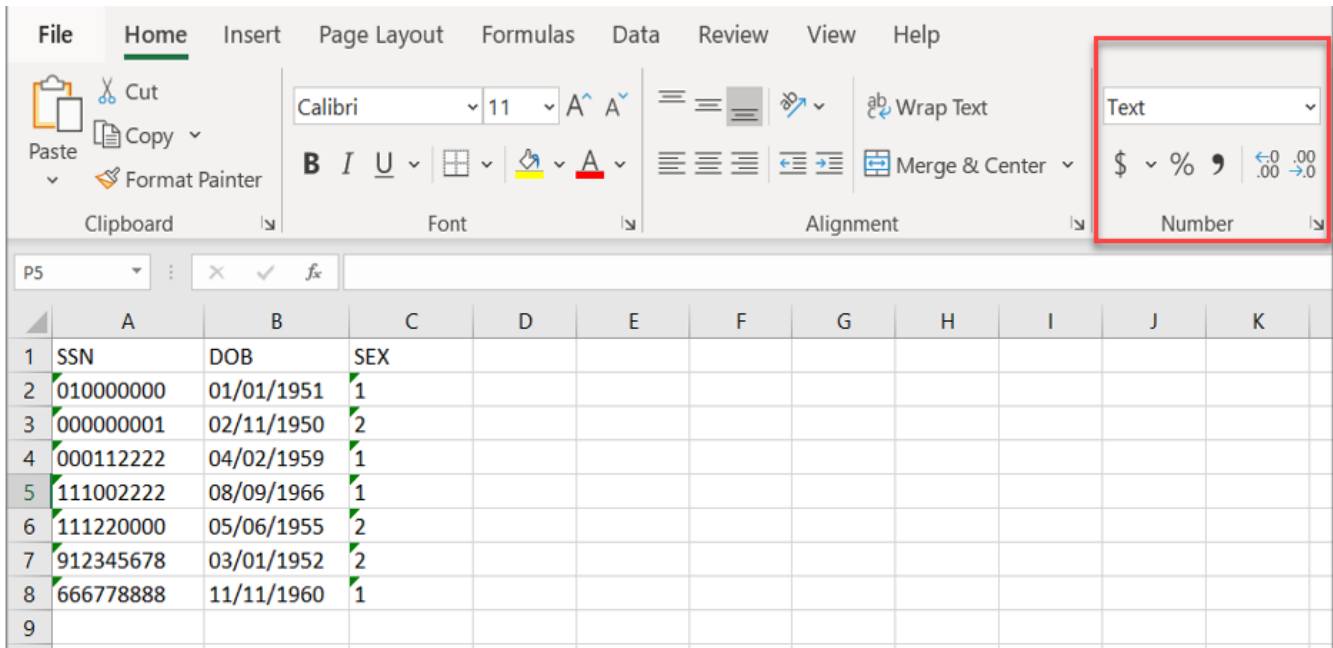
- Excel defaults to the general format. In general format, Excel removes the leading zeros of numbers (i.e., SSN, date of birth, bene_id, unique_id). To retain leading zeros, change the Excel format from General to Text before entering data.



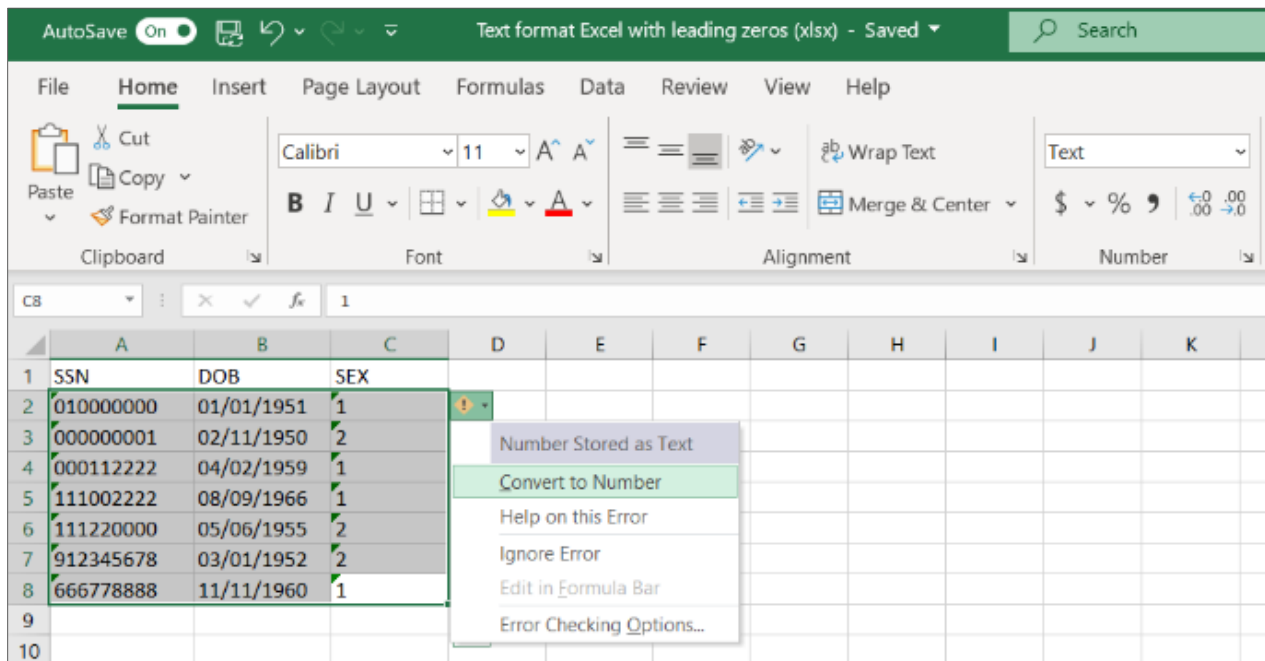
- Users must highlight the target cells before changing the format to Text. For finder files, it's a good idea to submit all variables using the Text format, so Excel keeps the desired formats intact. Select the triangle at the intersection of columns and rows to select all cells.



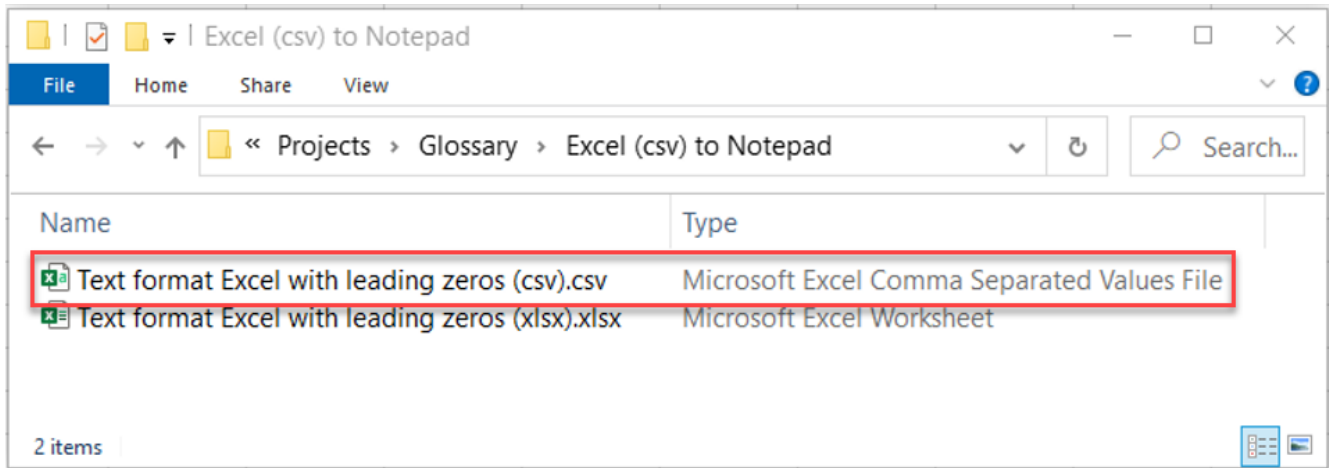
- By changing the format to Text before entering data, Excel keeps leading zeros; however, users cannot perform statistics on data while in the Text format.



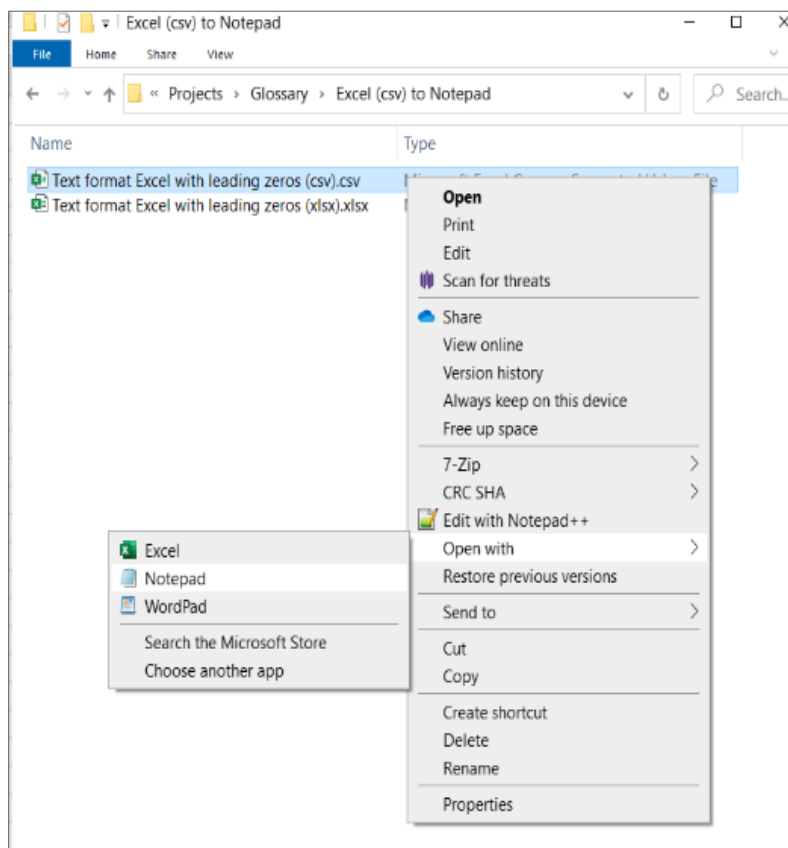
- Researchers can convert text formatted columns back to the default format by highlighting the cells, selecting the yellow warning icon, and choosing the desired conversion.



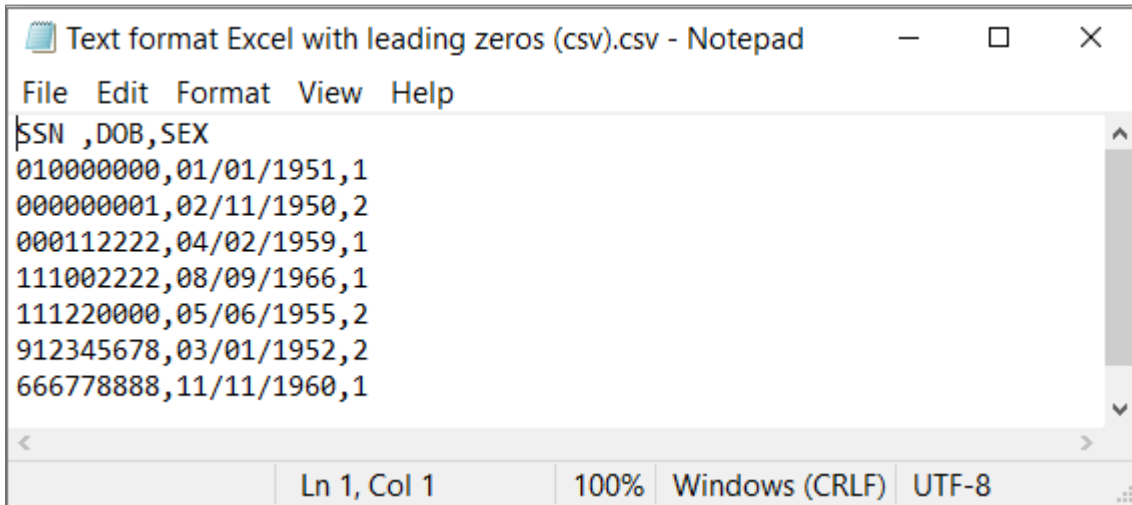
- When finished editing the data, save the Excel file as a comma-separated values (CSV) file. CCW database users should submit the .csv file for the finder files.



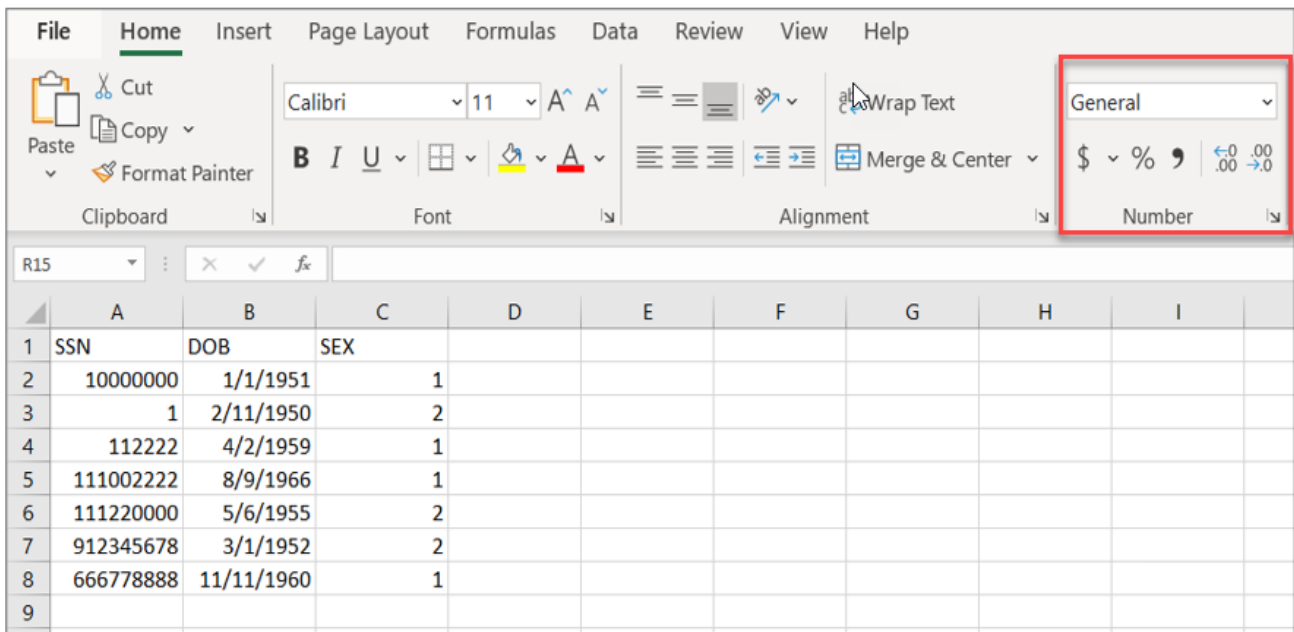
- Open the .csv Excel in Notepad to verify that formats are still intact.



7. This image shows the desired format for researcher submitted finder files.



8. Warning, opening the .csv Excel file for a second time automatically returns the format to General and removes any leading zeros. Do not save the .csv Excel file after opening it a second time, as this saves the format as General. Only save the final Excel version as a .csv file.



Appendix B — List of Acronyms

Acronym	Definition
AES	Advanced Encryption Standard
AWS	Amazon Web Services
BENE_ID	CCW beneficiary identifier
CCW	Chronic Conditions Warehouse
CHIP	Children’s Health Insurance Program
CMS	Centers for Medicare & Medicaid Services
CPT	Current Procedural Terminology
DOB	Date of birth
DRG	Diagnosis-related group
DUA	Data Use Agreement
EIN	Employer Identification Number
HCPCS	Healthcare Common Procedure Coding System
HICn	Health Insurance Claim number
ICD-9 or -10	International Classification of Diseases
MBI	Medicare Beneficiary Identifier
NDC	National Drug Code
NPI	National Provider Identifier
NSA	National Security Agency
PII	Personally Identifiable Information
PRCDR	Procedure
ResDAC	Research Data Assistance Center
RHEL	Red Hat Enterprise Linux
RRB	Railroad Retirement Board
SDA	Self-decrypting archive
SSN	Social Security number
TIN	Tax Identification Number