

Version 08 HHS-HCC Risk Adjustment Modeling Python Software
Documentation for the 2026 Benefit Year
July 31, 2026¹

Section 1343 of the Patient Protection and Affordable Care Act (ACA) provides for a permanent risk adjustment program. To protect against potential effects of adverse selection and help stabilize premiums in the individual and small group (including merged) markets, the risk adjustment program transfers funds from plans with relatively low-risk enrollees to plans with relatively high-risk enrollees. It generally applies to non-grandfathered individual and small group (including merged market) plans inside and outside Exchanges.

The HHS federally-certified risk adjustment methodology was first described in the HHS Notice of Benefit and Payment Parameters for 2014 final rule (78 FR 15410), which appeared in the *Federal Register* on March 11, 2013. Modifications to the HHS risk adjustment methodology for the 2026 benefit year are described in the HHS Notice of Benefit and Payment Parameters for 2026 final rule (90 FR 4424) (2026 Payment Notice final rule), which appeared in the *Federal Register* on January 15, 2025.² As described in the 2026 Payment Notice final rule, HHS has implemented Affiliated Cost Factors (ACFs), which reflect significant costs not already associated with HHS-hierarchical condition categories (HCCs) or Prescription Drug Categories (RXC) in the adult and child risk score models. The 2026 benefit year uses the Version 08 (V08) HSHCC classification, first implemented in the 2025 benefit year, and include interaction variables based on HCC count factors, first implemented in the 2023 benefit year. HHS implemented changes from the previous Version 07 (V07) classification in the V08 HHS-HCC classification to improve the prediction of sickle cell disease costs. As described in the HHS Notice of Benefit and Payment Parameters for 2025 final rule (89 FR 26247) (2025 Payment Notice), the following changes were made in the V08 HHS-HCC classification:

- updated mappings to include sickle cell disease diagnosis codes in the model (within HCC 71),
- ungrouped HCCs 70 and 71 in the adult and child models,
- reassigned HCC 70 and 71 to a higher severity in the infant models, and
- relabeled HCC 70 and HCC 71 to reflect these changes.

The 2026 benefit year risk adjustment model was recalibrated using blended coefficients from the 2020, 2021, and 2022 enrollee-level External Data Gathering Environment (EDGE) data. The high-cost risk pool calculation incorporated into the HHS risk adjustment methodology since the 2018 benefit year continued for the 2026 benefit year.³

¹ This document corresponds to Python software for the HHS risk adjustment models for the 2026 benefit year. This second release of Python software reflects revisions from the final 2025 benefit year SAS software instructions posted on the CCIIO website on March 30, 2026, available at: <https://www.cms.gov/files/document/cy2025-diy-instructions-03-30-2026.pdf>

² See the Patient Protection and Affordable Care Act; HHS Notice of Benefit and Payment Parameters for 2026; Final Rule; 90 FR 4424 (January 15, 2025) (2026 Payment Notice final rule), available at: <https://www.govinfo.gov/content/pkg/FR-2025-01-15/pdf/2025-00640.pdf>.

³ HHS maintained the \$1 million threshold and 60 percent coinsurance rate high-cost risk pool parameters for the 2026 benefit year. See the 2026 Payment Notice final rule, 90 FR 00640.

The federally-certified risk adjustment methodology that HHS will use when operating a risk adjustment program on behalf of a State for the 2026 benefit year⁴ will calculate a plan average risk score for each risk adjustment covered plan⁵ based upon the relative risk of the plan's enrollees and apply a State payment transfer formula to determine risk adjustment payments and charges for plans within a state market risk pool. The methodology addresses three considerations: (1) adverse selection in the individual and small group (including merged) markets; (2) plan metal level differences and permissible rating variation; and (3) the need for risk adjustment transfers that net to zero. The Federally certified risk adjustment methodology developed by HHS applicable for the 2026 benefit year:

- Is developed based on enrollee-level EDGE data, which directly reflects claims data for ACA individual and small group (including merged) market enrollees;
- Employs the HCC grouping logic used in the Medicare risk adjustment program, but with HCCs refined and selected to reflect the expected population in risk adjustment covered plans;
- Includes a selected number of RXCs and RXC interactions in the adult models;
- Includes a new category of factors called ACFs, which reflect significant costs not already associated with HCCs or RXCs;
- Establishes concurrent risk adjustment models, one for each combination of metal level (platinum, gold, silver, bronze, catastrophic) and age group (adult, child, infant);
- Pools catastrophically high-cost enrollees nationally with a portion of the costs funded by a percent of premium charge to issuers of risk adjustment covered plans in each national market;⁶
- Results in State transfers that net to zero within a state market risk pool;
- Adjusts State transfers for plan metal level, geographic rating area, induced demand, premium assistance Medicaid alternative plans or other types of State subsidy plans (if applicable),⁷ and age rating, so that transfers reflect health risk and not other cost differences; and
- Transfers funds between plans within a state market risk pool based on differences in relative actual risk.⁸

Key Revisions for Benefit Year 2026:

- (July 2026 Revisions) Updated Table 1 to add ACFs.
- (July 2026 Revisions) Updated Table 2 to add 2026 CPT/HCPCS codes used for diagnosis filtering, as described in Section II. Table 2 includes review of 2025 quarterly

⁴ HHS will operate risk adjustment for the 2026 benefit year in all 50 states and the District of Columbia.

⁵ See 45 CFR 153.20 for a definition of the term "Risk Adjustment Covered Plan".

⁶ The high-cost risk pool calculations under the HHS federally-certified risk adjustment methodology involve two national markets – one for the individual market (including catastrophic and non-catastrophic plans, and merged market plans), and another for the small group market. See, for example, the Patient Protection and Affordable Care Act; HHS Notice of Benefit and Payment Parameters for 2018; Amendments to Special Enrollment Periods and the Consumer Operated and Oriented Plan Program; Final Rule, 81 FR 94058 at 94080 – 94082 (Dec. 22, 2016).

⁷ In the 2026 benefit year, Arkansas, California, Colorado, Connecticut, Massachusetts, New Mexico, and New York have premium assistance Medicaid Alternative plans or other types of state subsidy plans.

⁸ For the 2026 benefit year, statewide average premiums are reduced (i.e., adjusted) by 14 percent in the state payment transfer formula to account for the proportion of administrative costs that do not vary with claims. See the 2018 Payment Notice, 81 Fed. Reg. at 94099 - 94100.

updates with effective dates as of April 1, 2026. Replaced the 2024 column of code information with 2025 codes (used for historical data purposes). (Table 2 will be updated again in late 2026 to include CPT/HCPCS quarterly updates through October 1, 2026.)

- (July 2026 Revisions) Added definition of an acceptable service for risk adjustment to explanatory text in Section II. For the 2026 benefit year and beyond, HHS determined through its annual review of CPT/HCPCS codes that a subset of previously risk adjustment-eligible CPT/HCPCS codes, including certain telehealth codes, do not meet the definition of an acceptable service, as previously communicated in guidance in March 2026.⁹
- (July 2026 Revisions) Revised Table 3 ICD-10 to V08 HHS-Condition Categories (CC) Crosswalk to remove FY2025 and CY2025 MCE columns. Revised explanatory text in Sections II and V to clarify that FY2026 ICD-10 diagnosis codes and FY2026 MCE edits should be used in 2026 benefit year risk adjustment and that FY2025 code valid information is retained for historical purposes. (Table 3 will be updated again in late 2026 to include FY2027 ICD-10 diagnosis codes and FY2027 MCE code edits.)
- (July 2026 Revisions) Updated Tables 6 and 7 to include ACFs in the 2026 adult and child models.
- (July 2026 Revisions) Updated Table 9 Model Factors for the 2026 benefit year (using 2020, 2021, and 2022 benefit year enrollee-level EDGE data).
- (July 2026 Revisions) Updated Tables 10a and 10b to contain National Drug Codes (NDCs) and HCPCS codes in the National Library of Medicine's RxNorm dataset as of April 1, 2026. (Tables 10a and 10b will be updated again in late 2026 to reflect monthly RxNorm updates of NDCs and HCPCS codes through October 1, 2026).
- (July 2026 Revisions) Added new Table 10c ACF to NDC Crosswalk, that maps NDC codes to the applicable ACF.
- (July 2026 Revisions) Added new Table 10d ACF to HCPCS Crosswalk, that maps HCPCS codes to the applicable ACF.
- (June 2026 Revisions) Revised explanatory text in Section II to clarify that the HCPCS-level file for RXC and ACF assignment uses all medical claims and bill types.

The state payment transfer formula that is part of the HHS federally-certified risk adjustment methodology consists of concurrent risk adjustment models, one for each combination of metal level (platinum, gold, silver, bronze, and catastrophic) and age group (adult, child, infant). This document provides the detailed information needed to calculate risk scores given individual diagnoses.

CMS has created two versions of software (Python software and HHS-developed risk adjustment model algorithm “Do It Yourself [DIY]” software) and software instructions for issuers to use with their enrollment data to simulate their enrollee populations’ 2026 benefit year risk scores within the HHS-HCC risk adjustment models. **This software is being issued only as a supplemental tool for issuers of risk adjustment covered plans to better understand and simulate the calculation of plan liability risk scores for their enrollees.**

⁹ “FAQs on HHS-Operated Risk Adjustment Program Code Filtering for 2026 Benefit Year and Beyond: Telehealth and Other Codes.” March 11, 2026. <https://www.cms.gov/files/document/hhs-ra-telehealth-faq20260311.pdf>

Note: Python software replaces SAS software starting with benefit year 2026.

This software is not a required prerequisite to submitting claims data to the EDGE server for risk adjustment, nor is it a requirement of the HHS-operated risk adjustment program. Furthermore, issuers should not use this software to filter their own claims prior to submitting claims data to the EDGE server. The EDGE server software may have several additional layers of operational rules. This software merely provides a simulation tool for issuers to calculate enrollees' risk scores. Because state risk adjustment transfers are dependent on the data submitted by other issuers within the state market risk pool, an issuer that wishes to use this information to assist with estimating its 2026 benefit year transfer(s) should do so with caution and in combination with other data.

This document describes software for HHS-HCC risk adjustment modeling (Version 08). The Python software requires Python version 3.12 or higher.

This software (V0826 141 E1) is designed to be used only with 2026 dates of service and with ICD-10 diagnosis codes. If the user will be using historical data (i.e., 2025 or earlier service dates), the user should refer to earlier versions of the software for HHS-HCC risk adjustment modeling also posted on the CCIIO website.

Questions: Please direct questions regarding these instructions to CMS at RARIPaymentOperations@cms.hhs.gov.

List of Tables¹⁰

Table 2. CPT/HCPCS Included List for Diagnosis Code Filtering

Table 3. ICD-10 to HHS-Condition Categories (CC) Crosswalk

Table 4. HHS-Hierarchical Condition Categories (HCC) Hierarchies

Table 10a. Prescription Drug Categories (RXC) to National Drug Code (NDC) Crosswalk

Table 10b. Prescription Drug Categories (RXC) to Healthcare Common Procedure Coding System (HCPCS) Crosswalk

Table 10c. Affiliated Cost Factor (ACF) to National Drug Code (NDC) Crosswalk

Table 10d. Affiliated Cost Factor (ACF) to Healthcare Common Procedure Coding System (HCPCS) Crosswalk

Table 11. Prescription Drug Categories (RXC) Hierarchies

Terminology: The abbreviations ICD-10 and ICD-10-CM are used interchangeably in this document to refer to International Classification of Diseases, 10th Revision, Clinical Modification. The abbreviations CC and HCC used in these instructions refer to the HHS-HCC risk adjustment models. These are different HCCs from those used in the CMS-HCC risk adjustment model for Medicare Part C.

I. Software description

¹⁰ HCCs described in the tables that have splits, such as HCC035_1 and HCC035_2, are identified in the EDGE reference data as 351 and 352.

The software reads four user-provided input Comma Separated Values (CSV) files (Section IV); constructs demographic variables for each enrollee; crosswalks ICD-10 diagnoses to Condition Categories (CCs); creates HCCs by imposing hierarchies on the CCs; creates RXCs and ACFs based on NDCs and HCPCS codes, and imposes hierarchies on RXCs.

The software uses the demographic variables, adult enrollment duration variables, HCCs, and RXCs to compute risk scores for three models (adult, child, infant); cost sharing reduction (CSR)-adjusted scores for each model including adjustment for enrollment in premium assistance Medicaid alternative plans; and final scores based on the enrollee's age and plan benefit design. Scores for enrollees without diagnoses, NDCs, or HCPCS codes are computed from person-level variables; i.e., zeros are assigned to all CCs, HCCs, and RXCs.

The software consists of a main folder `HHS_HCC_software_V0826.141.E1` with a `./software` folder where all programs and data are stored. The software's main program (`./software/HHS_HCC/transform.py`) reads user-specified parameters as defined in the config file (`./software/HHS_HCC/config.py`) and user-defined input files located in a designated folder (`./software/HHS_HCC/data/input/user_defined`). From this folder, the transform program reads in the four user-provided input files (Section IV). The software's main program references three additional files:

- `./software/HHS_HCC/config.py` – defines necessary parameters (user and program defined);
- `./software/HHS_HCC/utils.py` – creates functions to calculate risk score variables;
- `./software/common/utils.py` – creates additional functions to format input data and map CCs to HCCs.

The software:

Step 1: Includes external data files; these are most likely to vary among releases.

Step 2: Defines internal variables and functions.

Step 3: Merges the PERSON, NDC, HCPCS, and DIAGNOSES files, and outputs one record for each enrollee record in the PERSON data set. Input records must be fully compliant with validity rules (e.g., SEX must be 1/2), and all data sets must be sorted by the common person identifier variable. The name of the common person identifier variable is ID.

Step 3.1: Reads user-defined input and checks for expected columns, validates variable inputs, and calculates derived columns.

Step 3.2: Appends calibration coefficients for all models.

Step 3.3: Merges the PERSON, NDC, HCPCS and DIAGNOSES data sets by the person identifier variable (ID). Each enrollee must have exactly one PERSON record, and may have zero or more NDC, HCPCS, or DIAGNOSES records.

Step 3.4: Performs tasks when the enrollee's first record is detected.

Step 3.5: If the enrollee has at least one NDC or HCPCS code, this step: creates RXCs and ACFs using the crosswalk functions *create_rxc* and *get_acf_variables*, respectively, specified in `./software/HHS_HCC/utils.py` (see Section II for details regarding the python function library and functions specific to this version of software).

Step 3.6: If the enrollee has at least one diagnosis, this step: creates CCs using the crosswalk function *get_enrollee_diagnosis_hhs_cc_df* specified in `./software/HHS_HCC/utils.py` (see Section II for details regarding the python function library and functions specific to this version of software); performs ICD-10 edits; and creates additional CCs for some ICD-10 diagnoses.

Step 3.7: Creates demographic variables; creates HCCs by applying hierarchy rules to CCs; sets HCCs to zero if the enrollee has no diagnoses; applies hierarchy rules to RXCs; sets RXCs and ACFs to zero if the enrollee has no NDC or HCPCS codes; applies validity filters to various input variables; creates additional model-specific variables (e.g., *RXC*HCC* interactions, HCC groups, severe illness and transplant severe illness indicators, adult HCC-contingent enrollment duration indicators); creates unadjusted and CSR-adjusted scores for each plan level for each enrollee including enrollment in premium assistance Medicaid alternative plans; and defines output formats and labels for variables.

Step 4: The software concludes by writing a csv output file to the `./software/HHS_HCC/data/output` folder, as well as a program log in the `./software/HHS_HCC/logs` folder.

II. Files included with the software

The following programs and files are included and can be found in the HHS HCC software V0826.141.E1/software/ folder:

- **./software/common:**
 - **utils.py** – program containing functions referenced among CMS software packages. The software uses the following top-level functions from this program:
 - *map_ccs_to_hccs* – Creates HCCs by applying hierarchy rules to CCs. Table 4, HHS-Hierarchical Condition Categories (HCC) Hierarchies, summarizes the hierarchy assignments.
- **./software/HHS_HCC:**
 - **config.py** – program containing all user-provided parameters.
 - **utils.py** – program containing the following functions required for risk score calculation:
 - *get_bene_info_df* – verifies required columns are present in the PERSON data set.
 - *get_bene_age_sex_vars* – assigns age/sex categories.
 - *get_enrollee_diagnosis_hhs_cc_df* – Crosswalks ICD-10 codes to CC categories that are transformed to HCC categories, and contains ICD-10 codes used in the risk adjustment models that are valid on the

DIAGNOSIS_SERVICE_DATE, and performs edits on ICD-10 codes based on age and/or sex. The Medicare Code Edits (MCEs) and further specified CC age and sex splits are performed by this function. If the enrollee has an invalid age and/or sex for a particular ICD-10 code, then the ICD-10 code will be ignored. Table 3, ICD-10 to Condition Categories (CC) Crosswalk, summarizes the ICD-10 code edits; it describes the ICD-10 Medicare Code Edits (MCEs) for age, and additional edits for CC age and sex splits.

- *create_rxc* – assigns RXCs.
- *get_acf_variables* – assigns ACFs.
- *create_adult_model_vars* – creates adult model variables.
- *create_child_model_vars* – assigns child model variables.
- *create_infant_model_vars* – assigns infant model variables.
- *create_score_tables* – calculates risk scores.
- **transform.py** – primary program that merges input files, crosswalks NDCs and HCPCS to RXCs and ACFs, crosswalks ICD-10 codes to CCs, creates HCC and risk score variables by calling various external and internal functions. Table 3, ICD-10 to Condition Categories (CC) Crosswalk, summarizes the ICD-10 to CC assignments. Only ICD-10 codes assigned to HCCs in the risk adjustment models are included in this crosswalk. All other ICD-10 codes will be ignored by the software. Table 10a, NDC to RXC Crosswalk, and Table 10b, HCPCS to RXC Crosswalk, summarize the NDC and HCPCS assignments to RXCs. Table 10c, NDC to ACF Crosswalk, and Table 10d, HCPCS to ACF Crosswalk, summarize the NDC and HCPCS assignments to ACFs. NDC and HCPCS not listed in these tables will be ignored by the software.
- **./software/HHS_HCC data/input/internal:**
 - **CSR_TABLE.CSV** – Contains CSR indicators and risk adjustment factors for use in calculating CSR-adjusted risk scores.
 - **HCC_HIERARCHY.CSV** – Contains HHS-Hierarchical Condition Categories (HCC) Hierarchies. Table 4, HHS-Hierarchical Condition Categories (HCC) Hierarchies, summarizes the hierarchy assignments.
 - **ICD10_HHS_CC_mappings_2026.csv** – Crosswalks ICD-10 codes to CC categories. The file includes ICD-10 codes valid in FY2026 and FY2027 as indicated in columns valid_ICD10_2026 and valid_ICD10_2027 FY2027 values will be added in a future release.
 - **rxc_HCPCS_mappings.csv** – Contains Table 10b RXC to Healthcare Common Procedure Coding System (HCPCS) Crosswalk.
 - **rxc_NDC_mappings.csv** – Contains Table 10a RXC to National Drug Code (NDC) Crosswalk.
 - **RXC_hierarchy.csv** – Contains Prescription Drug Categories (RXC) Hierarchies. Table 11. Prescription Drug Categories (RXC) Hierarchies, summarizes the hierarchy assignments.
 - **RXC_interactions.csv** – Contains RXC*HCC interactions created by the software.
 - **acf_HCPCS_mappings.csv** – Contains Table 10d ACF to Healthcare Common Procedure Coding System (HCPCS) Crosswalk.

- **acf_NDC_mappings.csv** – Contains Table 10c ACF to National Drug Code (NDC) Crosswalk.
- **adult_group_mappings.csv** – Contains adult HCC groupings created by the software.
- **adult_model_factors.csv** – Contains relative coefficients for adult regression model, created using CY2020, CY2021, and CY2022 data and corresponding denominators (defined as the weighted average plan liability for the full modeling sample for the given year).
- **bundled_mother_ICD10_codes.csv** – Contains FY2026 completed pregnancy diagnoses for use in detecting mother-infant bundled claims.
- **bundled_infant_ICD10_codes.csv** – Contains FY2026 newborn diagnoses for use in detecting mother-infant bundled claims.
- **child_group_mappings.csv** – Contains child HCC groupings created by the software.
- **child_model_factors.csv** – Contains relative coefficients for child regression model, created using CY2020, CY2021, and CY2022 data and corresponding denominators (defined as the weighted average plan liability for the full modeling sample for the given year).
- **Infant_maturity_mappings.csv** – Crosswalks infant maturity categories to HCCs.
- **infant_model_factors.csv** – Contains relative coefficients for infant regression model, created using CY2020, CY2021, and CY2022 data and corresponding denominators (defined as the weighted average plan liability for the full modeling sample for the given year).
- **infant_severity_mappings.csv** – Crosswalks infant severity categories to HCCs.
- **severe_list.csv** – Contains severe HCCs and HCC groups for adult and child models.
- **transplant_list.csv** – Contains transplant HCCs and HCC groups for adult and child models.

III. Creation of a diagnosis data set, NDC data set, and HCPCS data set

A. Diagnosis-level data set. The diagnosis input CSV file (DIAGNOSES) must include ICD-10-CM diagnosis codes used for risk adjustment, listed in Table 3, ICD-10 to Condition Categories (CC) Crosswalk. The user must evaluate each claim or encounter record to determine whether its diagnoses are included in the DIAGNOSES data set. Encounter records normally report dates, provider or bill types, diagnoses and procedures, and other information, though they may not have payment information.

This section explains how each record is evaluated to determine whether the record's diagnoses are to be used in CC/HCC creation. It is the user's responsibility to create the DIAGNOSES data set according to the filtering logic below. This document provides filtering instructions and a list of the CY2025 (for historical data purposes) and CY2026 CPT/HCPCS codes that define service

or procedure types that identify acceptable sources of diagnoses for risk adjustment.¹¹ However, the user must create the DIAGNOSES data set for input to the risk adjustment algorithm; the data set is not created by the software.

NOTE: Supplemental diagnosis codes may be submitted in certain circumstances. These instructions and the software do not address the addition of supplemental diagnosis codes. Therefore, risk score output from this software will not account for inclusion of supplemental diagnoses.

Only ICD-10-CM diagnosis codes from risk adjustment eligible bill types and/or service codes should be included in the DIAGNOSES data set. ICD-10 codes that are not listed in Table 3 may be included in the DIAGNOSES data set but are ignored by the software.¹² The steps below provide logic to determine which diagnoses are allowable. Note that Steps 1 and 3 refer to Table 2, CPT/HCPCS Included List for Diagnosis Code Filtering, which provides the 2025 (for historical data purposes) and 2026 CPT/HCPCS codes used to define service or procedure types that are acceptable sources of diagnoses for risk adjustment. An acceptable service is any service that involves two-way, real-time interactive communication equivalent to a face-to-face encounter between a patient and a qualified healthcare professional licensed to diagnose on which the healthcare professional can reliably record a diagnosis.

- The CPT/HCPCS codes identifying services with diagnoses allowable for risk adjustment are listed in column A of Table 2.
- Column B lists the short descriptions of the CPT/HCPCS codes.
- Columns C and D, respectively, indicate whether a CPT/HCPCS code is acceptable in 2025 or 2026.
- Notes begin on row 6,384 of the Excel table with the line “Notes:” and should not be imported by any program.

The DIAGNOSES data set should include diagnoses from claims/encounter records with **discharge dates or through dates** within the benefit year. Though the term “claim” is used in the steps below, the steps apply equally to encounter records. For the EDGE server, only claims with discharge diagnoses are used for HHS risk adjustment.

1. Professional source of diagnosis
 - a. For professional records, use diagnoses from records that have at least one line item with an acceptable CPT/HCPCS code (Table 2). If there is at least one acceptable line on the record, use all the header diagnoses. There are three possible values for CPT/HCPCS codes in columns C and D:
 - i. yes = code is acceptable in that calendar year
 - ii. no = code is not acceptable in that calendar year
 - iii. N/A = code is not in existence in that calendar year

¹¹ Definitions taken directly from the Current Procedural Terminology (CPT®) codes and the Healthcare Common Procedure Coding System (HCPCS) code set. Note that although CY2024 codes are provided for historical purposes, this software is designed to be used only with CY2025 data.

¹² If the user conducts fiscal year code validity checks described later in this section before using the software, only codes valid for risk adjustment will be included in the final diagnosis-level file.

- b. For professional records, if a line item has an acceptable CPT/HCPCS code, use all diagnoses from the line item.
 - c. If there are no acceptable service lines on the record, do not use any of the diagnoses for risk adjustment.
- 2. Inpatient facility source of diagnosis
 - a. Use all header diagnoses from records where facility bill type code equals one of the following:
 - i. 111 (inpatient admit through discharge); or
 - ii. 117 (inpatient replacement of prior claim).
 - b. There is no procedure screen for inpatient facility record types.
- 3. Outpatient facility source of diagnosis
 - a. Restrict records to those with facility bill type code equal to:
 - i. 131 (hospital outpatient admit through discharge); or
 - ii. 137 (hospital outpatient replacement of prior claim); or
 - iii. 711 (rural health clinic admit through discharge); or
 - iv. 717 (rural health clinic replacement of prior claim); or
 - v. 731 (clinic – freestanding admit through discharge); or
 - vi. 737 (clinic – freestanding replacement of prior claim); or
 - vii. 761 (community mental health center admit through discharge); or
 - viii. 767 (community mental health center replacement of prior claim); or
 - ix. 771 (federally qualified health center admit through discharge); or
 - x. 777 (federally qualified health center replacement of prior claim).
 - xi. 851 (critical access hospital admit through discharge); or
 - xii. 857 (critical access hospital replacement of prior claim); or
 - xiii. 871 (freestanding non-residential opioid treatment programs [OTPs]); or
 - xiv. 877 (OTPs replacement of prior claim).
 - b. For records with at least one acceptable CPT/HCPCS code (Table 2) on a service line, use all header diagnoses. Otherwise, do not use the diagnoses for risk adjustment.

Fiscal year code validity: Section IV further describes the diagnosis input data set. After creating that data set, the user will have the variables needed to conduct fiscal year validity checks before using the software if desired. Table 3 identifies the fiscal year(s) in which the diagnosis codes used for risk adjustment are valid. The user should check that for a given diagnosis (variable ICD10) and service date (variable DIAGNOSIS_SERVICE_DATE), the diagnosis code has a Y in the corresponding Table 3 Code Valid column. For this preliminary 2026 software, the user should use information from the FY2026 columns. (FY2025 code valid information is included for historical data purposes.) ICD-10 diagnosis codes with service dates of January 1, 2026 – September 30, 2026 should have a Y in the Code Valid in FY2026 column; otherwise, the user should exclude them. The 2026 software will be updated later this year to include FY2027 code information corresponding to service dates of October 1, 2026 – December 31, 2026. As noted, this software can detect that an ICD-10 diagnosis code is not valid for a given fiscal year and will flag the enrollee record in the program log as a ‘warning’ (see Section VIII).

Note on bundled claims for mother and newborn infant: In practice, some hospital claims for childbirth include both the mother’s record and the newborn infant’s record on the same claim (diagnoses and procedure codes). Because there are separate adult, child, and infant risk adjustment models and some of the diagnosis codes may not be distinguishable between mother and infant on bundled claims, **any bundled claims should be redefined as two separate records whenever possible (mother and infant, each with a separate ID, sex, and age) in order for the diagnoses to be appropriately included in the input data set and used for appropriately calculating risk scores.**

The user will need to independently create a program to detect any bundled claims and redefine them as two separate claims (i.e., it is not part of these instructions). For example, a bundled claim detection program would need to identify enrollees with a claim containing the following elements:

Mother is the enrollee:

- AGE_LAST $\geq$ 2 (an age corresponding to the child or adult models; more specifically age should be appropriate for a maternity diagnosis [i.e., 9-64)¹³ and
- ICD-10 diagnoses corresponding to a completed pregnancy HCC (HCC 207 or 208 or 209) and
- ICD-10 diagnoses corresponding to a newborn HCC (HCC 242 or 243 or 244 or 245 or 246 or 247 or 248 or 249).

Infant is the enrollee:

- AGE_LAST = 0 (an age corresponding to the infant model; more specifically age is appropriate for a newborn diagnosis at birth) and
- ICD-10 diagnoses corresponding to a completed pregnancy HCC (HCC 207 or 208 or 209) and
- ICD-10 diagnoses corresponding to a newborn HCC (HCC 242 or 243 or 244 or 245 or 246 or 247 or 248 or 249).

See bundled_mother_ICD10_codes.csv and bundled_infant_ICD10_codes.csv or Table 3, ICD-10 to Condition Category (CC) Crosswalk, for diagnosis codes corresponding to the completed pregnancy and newborn HCCs.

As noted, this software can detect that an enrollee might have bundled claims and will optionally flag the enrollee record in the “Errors/warnings/notes log,” but it cannot redefine them as separate mother/infant claims (see Section VIII.5).

Infants with a record in the person-level file that cannot be matched with a claim or who do not have claims will have no diagnoses in the diagnosis data set. Infants without diagnoses will be assigned to the lowest severity category and the Age 1 maturity category for infants. Age 0 infants with diagnoses but who lack a newborn HCC will be assigned to the corresponding severity category and the Age 1 maturity category for infants. Male infants will also have the

¹³ Section IV of this document identifies the two age variables used in the software and specifies when each is used. MCE edits specifying appropriate age for maternity diagnoses are included in Table 3.

male demographic factor assigned. Age 0 male infants who lack a newborn HCC will have their demographic factor reassigned to Age 1.

B. NDC-level data set. The NDC input data set must include NDCs used for risk adjustment, listed in Table 10a RXC to NDC Crosswalk and Table 10c ACF to NDC Crosswalk. Only pharmacy claims (not medical claims) are the acceptable sources for NDCs. The user must evaluate each claim to determine whether the claim's NDCs are included in the NDC data set.

The NDCs are to be used for RXC and ACF creation. It is the user's responsibility to create the NDC data set for input to the risk adjustment software; the data set is not created by the software. The inclusion of RXCs in the 2026 benefit year HHS operated risk adjustment methodology is limited to the adult risk adjustment models. The inclusion of ACFs in the 2026 benefit year HHS operated risk adjustment methodology is limited to the adult and child risk adjustment models. Users should not include information for infant enrollees in the NDC data set.

The NDC data set should include NDCs from pharmacy claims with **prescription filled dates** within the benefit year. NDC codes should be in the 11-digit, no dashes, HIPAA format to match the format required for EDGE submission.¹⁴ (Note: Tables 10a and 10c in the Excel file contain the NDC codes formatted as text, not numbers, to retain any leading zeroes needed for 11-digit codes.) NDC codes that are not listed in Tables 10a or 10c may be included in the NDC data set but are ignored by the software and are not included in RXCs for the adult risk adjustment models' risk score calculations or for the adult and child ACF assignments for the risk adjustment models' risk score calculations. Section IV further describes the NDC data set.

C. HCPCS-level data set. The HCPCS input data set must include HCPCS codes used for risk adjustment RXCs, listed in Tables 10b RXC to HCPCS Crosswalk and 10d ACF to HCPCS Crosswalk. These RXC-specific and ACF-specific HCPCS codes may be sourced from inpatient, outpatient, and professional medical claims regardless of bill type.

The HCPCS codes in the HCPCS input data set are to be used for RXC and ACF creation. It is the user's responsibility to create the HCPCS data set for input to the risk adjustment software; the data set is not created by the software. The inclusion of RXCs in the 2026 benefit year HHS operated risk adjustment methodology is limited to the adult risk adjustment models. The inclusion of ACFs in the 2026 benefit year HHS operated risk adjustment methodology is limited to the adult and child risk adjustment models. Users should not include information for infant enrollees in the HCPCS data set.

The HCPCS data set should include HCPCS codes from inpatient, outpatient, and professional medical claims with **discharge dates or through dates** within the benefit year for adult enrollees. HCPCS codes that are not listed in Tables 10b or 10d may be included in the HCPCS

¹⁴ The source for the NDC codes is the U.S. Food and Drug Administration's Comprehensive NDC SPL Data Elements File: <https://www.fda.gov/ForIndustry/DataStandards/StructuredProductLabeling/ucm240580.htm>. The NDCs are validated as current prescriptions through the U.S National Library of Medicine's RxNorm data set: <https://www.nlm.nih.gov/research/umls/rxnorm/>. The RxNorm Technical Documentation includes an algorithm the user can access to normalize NDC codes to the 11-digit, no dashes, HIPAA format. The source for the NDC start/end dates is the U.S. Food and Drug Administration's Orange Book: <https://www.accessdata.fda.gov/scripts/cder/ob/index.cfm>.

data set but are ignored by the software and are not included in RXCs for the adult risk adjustment models' risk score calculations or for the adult and child ACF assignments for the risk adjustment models' risk score calculations. Section IV further describes the HCPCS data set.

IV. CSV data sets supplied by the user

This section describes the four input CSV data sets required to create CC and HCC groupings, RXC and RXC*HCC interactions, ACFs, enrollment duration variables, demographic variables, and risk score variables—a person-level data set (PERSON), a diagnosis data set (DIAGNOSES), an NDC data set (NDC), and a HCPCS data set (HCPCS). It is the responsibility of the user to create these input data sets with the variables listed in this section. All input data sets must be ordered in ascending order by the person identifier variable.

Note on CSR_INDICATOR

In operations, cost-sharing reduction (CSR) plan variations will be identified by the Health Insurance Oversight System (HIOS) variant ID or a new variant ID created specifically for EDGE to ensure that premium assistance Medicaid Alternative plans (i.e., private options) or other types of state subsidy plans have the correct CSR RA factor applied.¹⁵ Listed below are the codes that will be used to identify the plan variation.¹⁶ Please note that unlike the risk adjustment software person-level CSR indicator, the HIOS variant ID is a plan-level indicator.

HIOS Variant ID	Cost-Sharing Reduction (CSR) Level	RA Software Person-level CSR Indicator	CSR RA Factor
0	Non-CSR/unknown CSR	1	1
1	Non-CSR/unknown CSR	1	1
4	CSR: 73% Actuarial Value (AV) Silver Plan Variation	1	1
30	CSR: State Subsidy Off-Exchange Plan – Silver	1	1
31	CSR: State Subsidy – Silver	1	1
34	CSR: State Subsidy – Silver	1	1
42	CSR: State Subsidy – Gold	2	1.07
5	CSR: 87% AV Silver Plan Variation	3	1.12
6	CSR: 94% AV Silver Plan Variation	3	1.12
31	CSR: Massachusetts State Subsidy – Silver	3	1.12

¹⁵ See supra note 7.

¹⁶ We note that when we identify unique State-specific plans that have higher plan liability than the standard plan variants, we utilize the corresponding CSR adjustment factor in the plan liability risk score calculation that maps to the plan's AV. For example, we use a CSR adjustment factor of 1.12 for all Massachusetts CSR wrap-around plans with AVs above 94 percent, as discussed in the 2024 Payment Notice final rule, 88 FR at 25772. When appropriate, identified unique State-specific plans may warrant a State-specific CSR factor table. For example, in addition to the CSR variants listed with factors of 1.12 in the table in this document, plan variants of 04 are also 1.12 in Massachusetts. See, e.g., the March 2026 EDGE Server Reference Data Code Updates, available at: regtap.cms.gov/reg_library_openfile.php?id=6249.

HIOS Variant ID	Cost-Sharing Reduction (CSR) Level	RA Software Person-level CSR Indicator	CSR RA Factor
32	CSR: State Subsidy – Silver	3	1.12
34	CSR: Massachusetts State Subsidy – Silver	3	1.12
35	CSR: State Subsidy – Silver	3	1.12
36	CSR: State Subsidy – Silver	3	1.12
2	CSR: Zero Cost Sharing – Bronze	4	1.51
3	CSR: Limited Cost Sharing – Bronze	5	1.19
2	CSR: Zero Cost Sharing – Platinum	6	1.31
3	CSR: Limited Cost Sharing – Platinum	7	1.04
2	CSR: Zero Cost Sharing – Gold	8	1.39
3	CSR: Limited Cost Sharing – Gold	9	1.1
43	CSR: State Subsidy – Gold	9	1.1
2	CSR: Zero Cost Sharing – Silver	10	1.46
3	CSR: Limited Cost Sharing – Silver	11	1.15

Note on Enrollment Duration

The adult models include a series of partial-year enrollment indicators for enrollees with at least one HCC whose enrollment period in an issuer’s plans is 6 months or less. There are two steps involved in creating the enrollment duration indicator variables:

STEP 1: For the PERSON file, the user should create an ENROLDURATION variable for each enrollee with 12 possible values corresponding to 1-12 months based on an enrollee’s total number of days enrolled in the plan in the benefit year as described below. Although ENROLDURATION will only be used to create variables needed for the adult models, this software was designed for ENROLDURATION to be constructed for *all* enrollees to maintain consistency in the variables present in the PERSON file. Thus, enrollees missing ENROLDURATION will receive this Error message: *WARNING:root:Enrollee XX is missing ENROLDURATION. This enrollee may be excluded or have incomplete eligibility/enrollment history.* Once created, the ENROLDURATION variable will be ignored for enrollees in the child or infant models.

STEP 2: The monthly enrollment duration interaction variables for enrollees with at least one HCC (HCC_ED1–HCC_ED6) are created by the software for adult enrollees as specified in Section VI.

The variable names must be spelled as written; Python column names are case-sensitive (i.e., SEX and Sex and sex and SeX designate different columns) and are in upper case.

1. PERSON data set
 - a. ID (Person identification code).
 - i. String or numeric type, any length, not missing

- ii. Unique to an individual, and unique in the data set (i.e., no duplicates).
 - iii. Converted to string by the software.
- b. SEX.
 - i. Integer type, 1 byte, 1=male, 2=female, not missing.
 - ii. Converted to upper case by the software.
- c. DOB.
 - i. String type, in the format specified in config.py, valid calendar date, not missing, provides the enrollee's date of birth.
 - ii. Used to calculate AGE_AT_DIAGNOSIS for MCE diagnosis code age edits.
- d. AGE_LAST (Age as of last day of enrollment in benefit year).
 - i. Integer type, 0 or greater, not missing.
 - ii. Used for all risk adjustment purposes except MCE diagnosis code age edits.
 - iii. For infants born in the previous year but not discharged until the benefit year, users should substitute Age 0 for Age 1 in AGE_LAST.
- e. METAL (Enrollee's plan level – platinum, gold, silver, bronze, catastrophic).
 - i. String type, 1 byte, P/G/S/B/C (only 1 of these values), not missing.¹⁷
 - ii. Converted to upper case by the software.
- f. CSR_INDICATOR (Person-level indicator. Enrollees who qualify for cost-sharing reductions or those enrolled in premium assistance Medicaid alternative plans will be assigned CSR_INDICATOR = 1-11. Non-CSR recipients will be assigned CSR_INDICATOR = 1, which is associated with a factor of 1.00 (no adjustment to risk score). State-specific CSR programs have unique income eligibility criteria, which can differ year to year for the state and across states. For additional information, refer to the state-specific documents, as published by the applicable state. In BY 2026, states offering state-specific CSRs include CO, CT, NM, NY, and MA. AR differs as their use of unique CSR variants represents Medicaid Expansion participants enrolled in Marketplace QHPs, as applicable).
 - i. Integer type, integer, 1-11, not missing.
 - ii. Values are:
 - 1 = Non-CSR recipient, and enrollees with unknown CSR (HIOS Variant ID = 00, 01).
 - 1 = Enrollees in 73% AV Silver Plan Variation (HIOS Variant ID = 04).
 - 1 = Enrollees in approximately 73% AV State Subsidy Off-Exchange Plan of Silver Level QHP (HIOS Variant ID = 30).
 - 1 = Enrollees in 73%-86% AV State Subsidy Plan of Silver Level QHP (HIOS Variant ID = 31).
 - 1 = Enrollees in 73%-86% AV State Subsidy Plan of Silver Level QHP (HIOS Variant ID = 34).
 - 2 = Enrollees in 90%-100% AV State Subsidy Plan of Gold Level QHP (HIOS Variant ID = 42).

¹⁷ Although the user is required to select a single metal level for the enrollee, the software produces score variables for all levels. The final unadjusted and CSR-adjusted score variables correspond to the single metal level selected, as is noted in Section VI.

- 3 = Enrollees in 87% AV Silver Plan Variation (HIOS Variant ID = 05).
- 3 = Enrollees in 94% AV Silver Plan Variation (HIOS Variant ID = 06).
- 3 = Enrollees in 94% AV Massachusetts State Subsidy Plan of Silver Level QHP (HIOS Variant ID = 31).
- 3 = Enrollees in 100% AV State Subsidy Plan of Silver Level QHP (HIOS Variant ID = 32).
- 3 = Enrollees in 94% AV Massachusetts State Subsidy Plan of Silver Level QHP (HIOS Variant ID = 34).
- 3 = Enrollees in 87%-100% AV State Subsidy Plan of Silver Level QHP (HIOS Variant ID = 35).
- 3 = Enrollees in 94%-100% AV State Subsidy Plan of Silver Level QHP (HIOS Variant ID = 36).
- 4 = Enrollees in Zero Cost Sharing Plan Variation of Bronze Level QHP (HIOS Variant ID = 02).
- 5 = Enrollees in Limited Cost Sharing Plan Variation of Bronze Level QHP (HIOS Variant ID = 03).
- 6 = Enrollees in Zero Cost Sharing Plan Variation of Platinum Level QHP (HIOS Variant ID = 02).
- 7 = Enrollees in Limited Cost Sharing Plan Variation of Platinum Level QHP (HIOS Variant ID = 03).
- 8 = Enrollees in Zero Cost Sharing Plan Variation of Gold Level QHP (HIOS Variant ID = 02).
- 9 = Enrollees in Limited Cost Sharing Plan Variation of Gold Level QHP (HIOS Variant ID = 03).
- 9 = Enrollees in 90%-100% AV State Subsidy Plan of Gold Level QHP (HIOS Variant ID = 43).
- 10 = Enrollees in Zero Cost Sharing Plan Variation of Silver Level QHP (HIOS Variant ID = 02).
- 11 = Enrollees in Limited Cost Sharing Plan Variation of Silver Level QHP (HIOS Variant ID = 03).

g. ENROLDURATION

- i. Integer type, integer, 1-12, not missing.
- ii. Person-level enrollment duration variable. Although ENROLDURATION is for use in adult models only, user should create it for all enrollees for consistency in PERSON file preparation. Values will be ignored for enrollees in child or infant models.
- iii. Allowable values are 1-12 based on months enrolled in plan in benefit year as defined by days:
 - 1 = 1–31 days enrolled
 - 2 = 32–62 days enrolled
 - 3 = 63–92 days enrolled
 - 4 = 93–123 days enrolled
 - 5 = 124–153 days enrolled

- 6 = 154–184 days enrolled
- 7 = 185–214 days enrolled
- 8 = 215–245 days enrolled
- 9 = 246–275 days enrolled
- 10 = 276–306 days enrolled
- 11 = 307–335 days enrolled
- 12 = 336–366 days enrolled

2. DIAGNOSES data set

- a. ID (Person identification code).¹⁸
 - i. String with no special characters, any length, not missing.
 - ii. Unique to an individual.
- b. ICD10 (ICD-10-CM diagnosis codes).
 - i. Character type, 7-byte field, no periods or embedded blanks, left justified.
 - ii. Converted to upper case by the software.
 - iii. Codes should be to the greatest level of available specificity.
 - iv. Age and sex edits for diagnoses are performed in `./software/HHS_HCC/transform.py` in the function `get_enrollee_diagnosis_hhs_cc_df` to ensure CCs are assigned appropriately for the age and sex of the enrollee.
 - v. Only diagnoses from allowable sources should be included in the DIAGNOSES data set.
 - vi. Invalid diagnoses are ignored.¹⁹
 - vii. A valid ICD-10 diagnosis must have a valid `DIAGNOSIS_SERVICE_DATE`.
- c. `DIAGNOSIS_SERVICE_DATE`
 - i. String type, in the format specified in `config.py`, valid calendar date, not missing, provides the diagnosis’s service date.²⁰
 - ii. As described in Section III, this variable can be used with ICD10 and Table 3 to precheck that a diagnosis code is valid for a given fiscal year.²¹

3. NDC data set

- a. ID (Person identification code).²²
 - i. String with no special characters, any length, not missing.
 - ii. Unique to an individual.
- b. NDC

¹⁸ Please note that in operation, this information can not include personally identifiable information.

¹⁹ In the context of this software’s instructions, valid refers to “included” in the HHS-HCC risk adjustment model and invalid refers to “not included.”

²⁰ Valid diagnosis service date in this version of software (V0825 141 E3): year is 2025, month is 01-12, and day is 01-31 and appropriate for the given month (i.e., cannot be February 30). The service date cannot occur before the date of birth.

²¹ The software has a fiscal year validity check. If an ICD-10 code is not valid for a given `DIAGNOSIS_SERVICE_DATE` (e.g., a deleted in FY2025 code with a FY2025 service date), the optional software warning message will be *WARNING:root:Diagnosis date YYYY-MM-YY for enrollee ID XXXX is not within benefit year 2025. Diagnosis will be skipped..*

²² Please note that in operation, this information can not include personally identifiable information.

- i. NDC normalized drug code (11-digit, no dashes, HIPAA standard format), 11-character string type, left justified. This format matches the format required for submission to the EDGE server. Only NDCs from pharmacy claims filled in the benefit year or inpatient claims with discharges in the benefit year (Section III B) for adult enrollees should be included in the NDC data set.

4. HCPCS data set

- a. ID (Person identification code).²³
 - i. String with no special characters, any length, not missing.
 - ii. Unique to an individual.

- b. HCPCS

HCPCS drug code, 5-character string type, left justified. HCPCS drug codes from inpatient, outpatient, or professional medical claims with discharge or through dates in the benefit year for adult and child enrollees, regardless of bill type (Section II), should be included in the HCPCS-level file.

AGE_AT_DIAGNOSIS, the age as of the diagnosis service date, is calculated by the software using DOB from the PERSON data set and DIAGNOSIS_SERVICE_DATE from the DIAGNOSES data set. It is used only for MCE diagnosis code age edits.

The four user-provided data sets (PERSON, DIAGNOSES, NDC, HCPC) are illustrated below. These examples are not based on actual data.

- Person-level data set example (PERSON) containing seven variables; we use ID as the person identifier variable to illustrate:

ID	SEX	DOB	AGE_LAST	METAL	CSR_INDICATOR	ENROLDURATION
201	1	19621201	63	P	0	12
202	2	20120315	13	C	0	12
301	2	19700414	55	G	2	7
302	1	19760101	49	B	4	12
304	X	19760132		R	1	3
305	1	19850101	40	S	0	12

- Diagnosis data set example (DIAGNOSES) containing three variables; we use ID as the person identifier variable and ICD-10 diagnoses to illustrate:

ID	ICD10	DIAGNOSIS_SERVICE_DATE
201	E118	20250113
201	H9319	20250113
201	M532X9	20250629
201	M25461	20250630
201	M25569	20250706
201	M25579	20250706
201	209	20250835
202	J4530	20250219

²³ Please note that in operation, this information can not include personally identifiable information.

302	J200	20250317
302	Z430	20250504
303	E890	20250929
304	Z0000	20250617
305	B20	20250302

- NDC data set example (NDC) containing two variables; we use ID as the person identifier variable and NDC normalized drug code, 11-digits, HIPAA standard format, string type, left justified, to illustrate:

ID	NDC
201	00002751001
202	
303	42291018920
304	13411019102
305	0003-1964-1

- HCPCS data set example (HCPCS) containing two variables; we use ID as the person identifier variable and HCPCS code, 5 digit string type, left justified, to illustrate:

ID	HCPCS
302	C9482
303	J1324
304	Q3028
305	J87

- ID 301 has no diagnoses; the other IDs in PERSON have one or more diagnoses.
- ID 303 in DIAGNOSES, NDC, and HCPCS will be ignored because there is no ID 303 in PERSON.
- Missing or invalid information in any PERSON variable will cause that enrollee and all his/her diagnoses, NDCs, and HCPCS codes to be ignored (e.g., ID 304).
- Missing or invalid information in DIAGNOSES will cause that diagnosis to be ignored (e.g., ID 201 DIAG 209).
- Missing or invalid information in NDC will cause that NDC to be ignored (e.g., ID 202, and ID 305).
- Missing or invalid information in HCPCS will cause that HCPCS to be ignored (e.g., ID 305).
- Risk scores for enrollees without diagnoses, NDCs, and HCPCS codes are calculated using only PERSON-level information (e.g., ID 301).

If an enrollee has N different diagnoses, the enrollee will have N records in DIAGNOSES and 1 record in PERSON. If an enrollee has no diagnoses, the enrollee will have zero records in DIAGNOSES and 1 record in PERSON.

V. Parameters supplied by the user

The user must supply the following in the config.py file, located in the ./software/HHS_HCC folder. The file has variables initialized such that the user is only expected to change values within the *run_spec* dictionary, including keys for the following variables:

- **switch_edits** – a Boolean value indicating the use of MCE age criteria when performing the ICD-10 mappings, with a value of True or False
- **date_format** – a string value that indicates the datetime python library string format of the dates provided in the PERSON input file, ie. an example of format "%Y%m%d" would be 19500121, and an example of "%m/%d/%Y" format would be 1/21/1950
- **intermediate_variable_output_switches** - a dictionary indicating the desired intermediate variables to include in the final output; turning these switches on allows users to review the intermediate steps the software takes in preparing model inputs and outputs; True = on, False = off
 - demographic_age_sex: Demographic age/sex variables created by the software.
 - cc_assignments: CCs created by the software (before hierarchies are applied).
 - hcc_assignments: HCCs created by the software (after hierarchies are applied), before model groupings are applied.
 - rxc_assignments: RXCs created by the software (after hierarchies are applied).
 - rxc_hcc_interactions: RXC × HCC interactions created by the software.
 - acf_assignments: ACFs created by the software.
 - hcc_groups_and_interactions: HCC groups and HCC interactions created by the software.
 - hhs_hcc_assignments: HHS HCCs created by the software (after hierarchies are applied and groupings applied).
 - severe_and_transplant_indicators_and_interactions: SEVERE and TRANSPLANT indicators, and interactions with HCC_CNT, created by the software.
 - adult_hcc_contingent_enrollment_indicators: Adult models HCC-contingent enrollment duration indicators (HCC_ED1–HCC_ED6) created by the software.
 - infant_model_assignments_and_interactions: Infant model maturity categories, severity level categories, and maturity by severity level interactions created by the software.

NOTE: All user-supplied parameters should be re-entered by user. The default setting is an example only and should not be used as a default.

VI. Variables output by the software

The software generates a person-level output CSV file. In addition to the originally provided variables, various intermediate variables are included, along with the following variables:

1. Demographic age/sex variables created by the software:

```
AGE0_MALE      AGE1_MALE

MAGE_LAST_2_4  MAGE_LAST_5_9  MAGE_LAST_10_14 MAGE_LAST_15_20
MAGE_LAST_21_24 MAGE_LAST_25_29 MAGE_LAST_30_34 MAGE_LAST_35_39
```

MAGE_LAST_40_44 MAGE_LAST_45_49 MAGE_LAST_50_54 MAGE_LAST_55_59
MAGE_LAST_60_GT

FAGE_LAST_2_4 FAGE_LAST_5_9 FAGE_LAST_10_14 FAGE_LAST_15_20
FAGE_LAST_21_24 FAGE_LAST_25_29 FAGE_LAST_30_34 FAGE_LAST_35_39
FAGE_LAST_40_44 FAGE_LAST_45_49 FAGE_LAST_50_54 FAGE_LAST_55_59
FAGE_LAST_60_GT

2. CCs created by the software (before hierarchies are applied).
3. HCCs created by the software (after hierarchies are applied).
4. RXCs created by the software (after hierarchies are applied).
5. RXC*HCC interactions created by the software.
6. ACFs created by the software.
7. HCC groups and HCC interactions created by the software.
8. SEVERE and TRANSPLANT indicators, and interactions with HCC_CNT, created by the software.
9. Adult models HCC-contingent enrollment duration indicators (HCC_ED1–HCC_ED6) created by the software.
10. Infant model maturity categories, severity level categories, and maturity by severity level interactions created by the software.
11. Score variables created by the software:
 - a. Adult Models
 - i. SCORE_ADULT_PLATINUM
 - ii. SCORE_ADULT_GOLD
 - iii. SCORE_ADULT_SILVER
 - iv. SCORE_ADULT_BRONZE
 - v. SCORE_ADULT_CATASTROPHIC
 - b. Child Models
 - i. SCORE_CHILD_PLATINUM
 - ii. SCORE_CHILD_GOLD
 - iii. SCORE_CHILD_SILVER
 - iv. SCORE_CHILD_BRONZE
 - v. SCORE_CHILD_CATASTROPHIC
 - c. Infant Models
 - i. SCORE_INFANT_PLATINUM
 - ii. SCORE_INFANT_GOLD
 - iii. SCORE_INFANT_SILVER
 - iv. SCORE_INFANT_BRONZE
 - v. SCORE_INFANT_CATASTROPHIC
12. CSR-adjusted score variables:
 - a. Adult model
 - i. CSR_ADJ_SCR_ADULT_PLATINUM
 - ii. CSR_ADJ_SCR_ADULT_GOLD
 - iii. CSR_ADJ_SCR_ADULT_SILVER
 - iv. CSR_ADJ_SCR_ADULT_BRONZE
 - v. CSR_ADJ_SCR_ADULT_CATASTROPHIC
 - b. Child model
 - i. CSR_ADJ_SCR_CHILD_PLATINUM

- ii. CSR_ADJ_SCR_CHILD_GOLD
 - iii. CSR_ADJ_SCR_CHILD_SILVER
 - iv. CSR_ADJ_SCR_CHILD_BRONZE
 - v. CSR_ADJ_SCR_CHILD_CATASTROPHIC
 - c. Infant model
 - i. CSR_ADJ_SCR_INFANT_PLATINUM
 - ii. CSR_ADJ_SCR_INFANT_GOLD
 - iii. CSR_ADJ_SCR_INFANT_SILVER
 - iv. CSR_ADJ_SCR_INFANT_BRONZE
 - v. CSR_ADJ_SCR_INFANT_CATASTROPHIC
- 13. Final unadjusted and CSR-adjusted score variables depending on the enrollee's metal (plan benefit) level and CSR indicator, including enrollment in premium assistance Medicaid alternative plans, created by the software.
 - a. Adult scores
 - i. SCORE_ADULT
 - ii. CSR_ADJ_SCR_ADULT
 - b. Child scores
 - i. SCORE_CHILD
 - ii. CSR_ADJ_SCR_CHILD
 - c. Infant scores
 - i. SCORE_INFANT
 - ii. CSR_ADJ_SCR_INFANT

The user must determine which of the scores is appropriate for the enrollee, depending upon the enrollee's age and plan benefit design of that enrollee.

VII. Computing platforms

The software has been tested using Python version 3.12.

VIII. Steps

1. Ensure that python is installed. Verify that python has been installed properly by checking the version in the terminal or shell window. It is recommended that a more recent version of Python is downloaded, specifically version 3.12 or higher. Otherwise, syntax errors may occur.
2. Install software:
 - Use pip to install the package dependencies in ./requirements.txt, from the base project folder (./) where the requirements.txt file is saved. Run the following command:

```
pip install -r requirements.txt
```

Note, the `requirements.txt` file includes version numbers for pandas and numpy, common Python data manipulation libraries. Depending on the version of Python installed, a different version of pandas may need to be installed. Pandas version 2.3.3, the version listed in the `requirements.txt` file, is generally compatible with Python version 3.12 or higher.

3. Edit user-defined parameters in the `run_spec` variable within the `./software/HHS_HCC/config.py` file. The file has variables initialized such that the user is only expected to change the value of the following variables within the `run_spec` dictionary (See Section V.)
4. Prepare user-provided input data sets:
 - Create PERSON, DIAGNOSES, NDC, and HCPCS data sets using the guidelines in Section III and data set descriptions in Section IV.
 - These data sets are created by the user; they are not created by the risk adjustment modeling software
5. Generate scores:
 - After the setup steps are complete, navigate to the base directory where the software folder is saved and run the transform script using the following python command:

```
python ./software/HHS_HCC/transform.py
```

Review log errors/warnings, notes: When the transform step is complete, locate the log for the program (`./software/HHS_HCC/logs`). The log file will contain print statements for each completed step and any warning or error information. If errors occur, first check that the user-defined input data is in the correct format, as outlined in Section IV. If all steps run without errors, locate the final output data table located in the `./software/HHS_HCC/data/output` folder named `HHS_HCC_V0826.141.E1_scores.csv`. This table includes the final scores along with all intermediate variables.

End of Document