

**Version 07 HHS-HCC Risk Adjustment Modeling “Statistical Analysis System (SAS)”**  
**Software Documentation for the 2021 Benefit Year**  
**April 13, 2022<sup>1</sup>**

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 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 plans inside and outside Exchanges.

The HHS risk adjustment methodology was first described in the HHS Notice of Benefit and Payment Parameters for 2014 final rule (78 FR 15410), which was published in the *Federal Register* on March 11, 2013. Modifications to the HHS risk adjustment methodology for the 2021 benefit year are described in the HHS Notice of Benefit and Payment Parameters for 2021 final rule (85 FR 29164) (2021 Payment Notice final rule), which was published in the *Federal Register* on May 14, 2020.<sup>2</sup> The 2021 benefit year implements a new Version 07 (V07) HHS-HCC classification. Changes to the previous Version 05 (V05) classification, referred to as Version 06a, were initially detailed in a technical paper posted June 17, 2019: Potential Updates to HHS-HCCs for the HHS-operated Risk Adjustment Program.<sup>3</sup> HHS conducted further analyses and applied additional classification changes, referred to as Version 06b, which were described in the HHS Notice of Benefit and Payment Parameters for 2021 proposed rule (85 FR 7088).<sup>4</sup> Following input from the public on proposed changes and internal review and analysis, HHS finalized the V07 HHS-HCC classification with certain modifications from Version 06b in the 2021 Payment Notice final rule. A draft V07 ICD-10 to HHS-HCC Crosswalk for benefit year 2021 was posted May 12, 2020,<sup>5</sup> and the final V07 classification changes were presented in the 2021 Payment Notice final rule.<sup>2</sup> In addition to classification changes, the 2021 benefit year risk adjustment model was recalibrated using blended coefficients from the 2016, 2017, and 2018 enrollee-level External Data Gathering Environment (EDGE) data. As described in the 2021 Payment Notice final rule, the HHS risk adjustment methodology gradually transitioned from using MarketScan® commercial database, and the 2021 benefit year is the first

---

<sup>1</sup> This document corresponds to software for the HHS risk adjustment models for the 2021 benefit year, with revisions from the updated 2021 benefit year software instructions posted on the CCIIO website on January 7, 2022, available at <https://www.cms.gov/files/zip/2hhs-hcc-software-v0721141a2.zip>.

<sup>2</sup> See the Patient Protection and Affordable Care Act; HHS Notice of Benefit and Payment Parameters for 2021; Final Rule; 85 FR 29164 (May 14, 2020), available at: <https://www.govinfo.gov/content/pkg/FR-2020-05-14/pdf/2020-10045.pdf>. The final 2021 benefit year risk adjustment model coefficients are available at: <https://www.cms.gov/CCIIO/Resources/Regulations-and-Guidance/Downloads/Final-2021-Benefit-Year-Final-HHS-Risk-Adjustment-Model-Coefficients.pdf>.

<sup>3</sup> Potential Updates to HHS-HCCs for the HHS-operated Risk Adjustment Program, posted on the CMS website on June 17, 2019, available at: <https://www.cms.gov/CCIIO/Resources/Regulations-and-Guidance/Downloads/Potential-Updates-to-HHS-HCCs-HHS-operated-Risk-Adjustment-Program.pdf>.

<sup>4</sup> See the Patient Protection and Affordable Care Act; HHS Notice of Benefit and Payment Parameters for 2021; Proposed Rule; 85 FR 7088 (February 6, 2020), available at: <https://www.govinfo.gov/content/pkg/FR-2020-02-06/pdf/2020-02021.pdf>.

<sup>5</sup> The draft V07 crosswalk included ICD-10 diagnosis codes through the most recent fiscal year (2020) at time of posting, May 12, 2020, available at: <https://www.cms.gov/files/document/draft-2021-update-icd-10-crosswalk-hhs-hcc-risk-adjustment-model.xlsx>.

year in which no MarketScan® data is used in the blended coefficients.<sup>6</sup> The high-cost risk pool calculation incorporated into the HHS risk adjustment methodology since the 2018 benefit year continued for the 2021 benefit year.

The methodology that HHS will use when operating a risk adjustment program on behalf of a State for the 2021 benefit year<sup>7</sup> will calculate a plan average risk score for each covered plan based upon the relative risk of the plan's enrollees and apply a state payment transfer formula in order to determine risk adjustment payments and charges for plans within a State market risk pool. The HHS risk adjustment 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 for the 2021 benefit year:

- Is developed on enrollee-level EDGE data, which directly reflects claims data for ACA individual and small group (including merged) market enrollees;
- Employs the hierarchical condition category (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<sup>8</sup>;
- Includes a selected number of Prescription Drug Categories (RXC) and RXC interactions in the adult models;
- 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<sup>9</sup>;
- Results in state transfers that net to zero within a State market risk pool;
- Adjusts transfers for plan metal level, geographic rating area, induced demand, premium assistance Medicaid alternative plans, 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.

#### Key Revisions for Benefit Year 2021:

- (August 2021 Revisions) Updated Table 2 to add 2021 CPT/HCPCS codes used for diagnosis filtering, as described in Section III. Table 2 includes review of 2021 quarterly updates with effective dates as of April 1, 2021. Replaced the

---

<sup>6</sup> Updates to data used for risk adjustment model recalibration are described in the 2021 Payment Notice final rule, 85 at 29173, available at: <https://www.govinfo.gov/content/pkg/FR-2020-05-14/pdf/2020-10045.pdf>.

<sup>7</sup> HHS will operate risk adjustment for the 2021 benefit year in all 50 states and the District of Columbia.

<sup>8</sup> See 45 CFR 153.20 for the definition of "Risk Adjustment Covered Plan".

<sup>9</sup> The high-cost risk pool calculations under the HHS 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).

2019 column of code information with 2020 codes (used for historical data purposes).

- (January 2022 Revisions) Updated Table 2 to include review of 2021 quarterly updates of CPT/HCPCS codes with effective dates as of October 1, 2021.<sup>10</sup>
- (August 2021 Revisions) Updated software to include Fiscal Year (FY) 2021 ICD-10 diagnosis code assignments and FY2021 Medicare Code Editor (MCE) edits and to remove FY2020 ICD-10 assignments and FY2020 MCE edits. Revised fiscal year validity checks for ICD-10 diagnosis codes and corresponding service dates.
- (August 2021 Revisions) Revised Table 3 ICD-10 to HHS-Condition Categories (CC) Crosswalk to reflect new V07 classification with updated Fiscal Year (FY) 2021 ICD-10 diagnosis codes effective January 1, 2021. Revised FY2021 Medicare Code Editor (MCE) code edits to correspond to V07. Revised explanatory text in Section III to clarify that FY2021 ICD-10 diagnosis codes and FY2021 MCE edits should be used in 2021 benefit year risk adjustment. Although codes and MCE edits from the previous fiscal year are usually included for historical purposes, this table does not include FY2020 information due to the change to the V07 classification. Calendar Year (CY) 2021 is the initial implementation year for the V07 HHS-HCC risk adjustment model. Removed the Additional CC column and added two new columns, Second CC and Third CC, to account for ICD-10 code mappings to more than 2 HCCs.
- (January 2022 Revisions) Updated software to account for new FY2022 ICD-10 codes, new CY2021 (FY21/FY22) MCE edits, and updated CC assignments. Revised Table 3 ICD-10 to HHS-Condition Categories (CC) Crosswalk to contain FY2021 and FY2022 ICD-10 diagnosis codes and FY2021 and FY2022 MCE age and sex conditions. Updated ICD-10 code labels to reflect changes in FY2022. Updated CC assignments to account for new FY2022 ICD-10 codes. Updated the combined set of MCE age and sex conditions to be used for Calendar Year (CY) 2021 that covers both fiscal years (FY2021 and FY2022). Revised explanatory text in Section III to clarify the use of FY2021 and FY2022 ICD-10 diagnosis codes and MCE edits.
- (August 2021 Revisions) Updated Table 4 to include hierarchies and HCC numbers and labels consistent with changes from the V05 to the V07 HHS-HCC risk adjustment models.
- (August 2021 Revisions) Updated Tables 10a and 10b to contain NDCs and HCPCS codes in the National Library of Medicine's RxNorm dataset as of April 1, 2021.
- (January 2022 Revisions) Updated software to account for most recent 2021 NDC and HCPCS codes used in RXC crosswalks. Updated Tables 10a and 10b to contain NDCs and HCPCS codes in the National Library of Medicine's RxNorm dataset as of October 1, 2021. (Tables 10a and 10b will be updated as EDGE reference data updates and

---

<sup>10</sup> As a result of the Coronavirus (COVID-19) public health emergency (PHE), the CPT/HCPCS list in Table 2 was expanded in Calendar Year (CY) 2020 to include certain telehealth and telephonic service codes that were not previously accepted for HHS-operated risk adjustment. The allowable telehealth services on Table 2 will be accepted for risk adjustment eligible diagnosis filtering for the HHS-operated risk adjustment program applicable for the individual, merged and small group markets for the 2021 benefit year, if the services are otherwise allowable under applicable state law. For more information on the use of telehealth and telephonic services in the HHS-operated risk adjustment program, please refer to the [Risk Adjustment Telehealth and Telephone Services During COVID-19 FAQs](https://www.cms.gov/files/document/RA-Telehealth-FAQ.pdf) (<https://www.cms.gov/files/document/RA-Telehealth-FAQ.pdf>), updated May 6, 2021.

posted on REGTAP in April 2022 to be used as the final set of NDCs and HCPCS codes for the 2021 benefit year.)

- (April 2022 Revisions) Updated software to account for most recent NDC and HCPCS codes used in RXC crosswalks. Updated Tables 10a and 10b to contain NDCs and HCPCS codes in the National Library of Medicine's RxNorm dataset as of December 1, 2021 to be used as the final set of NDC and HCPCS codes for the 2021 benefit year.
- (August 2021 Revisions) Updated coefficients and denominators for the 2021 benefit year using 2016, 2017, and 2018 EDGE data (Sections II and VIII).
- (April 2022 Revisions) Added facility bill type code 87x (freestanding non-residential opioid treatment programs [OTPs]) to the list of acceptable outpatient facility sources of diagnosis.

The state payment transfer formula that is part of the HHS 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. Please direct questions regarding these instructions to HHS HCC Risk Adjustment Models at [hshccraops@cms.hhs.gov](mailto:hshccraops@cms.hhs.gov). This mailbox will be used only to answer questions pertaining to operations of the HHS risk adjustment models. We look forward to assisting with inquiries pertaining to your risk adjustment program operations using the HHS-HCC risk adjustment models for the 2021 benefit year.

CMS has created two versions of software (SAS 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' 2021 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.**

**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 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 2021 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 07). The software requires SAS® version 9.

This software (V0721 141 A3) is designed to be used only with 2021 dates of service and with ICD-10 diagnosis codes. If the user will be using historical data (i.e., 2020 or earlier service dates), the user should refer to earlier versions of the software for HHS-HCC risk adjustment modeling also posted on the CCHIO website.

## **List of Tables**<sup>11</sup>

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 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**

The software reads four user-provided input SAS® data sets (Section IV); constructs demographic variables for each enrollee; crosswalks ICD-10 diagnoses to Condition Categories (CCs) using SAS® formats which are stored in a FORMAT library; creates Hierarchical Condition Categories (HCCs) by imposing hierarchies on the CCs; creates Prescription Drug Categories (RXC) based on National Drug Codes (NDCs) and Healthcare Common Procedure Coding System (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 demographic variables; i.e., zeros are assigned to all CCs, HCCs, and RXCs.

The software's main program (CY21P07C) calls primary macro CY21M07C and passes a set of user-specified parameters (a macro is a subroutine that performs a specific task). Macro CY21M07C calls five external macros (provided as separate files):

- AGESEXV6 – creates age/sex variables;
- I0V07ED1 – performs edits on ICD-10 codes based on age and/or sex;
- V07141L1 – assigns labels to HCCs and RXCs;
- V07141H1 – sets selected HCCs to zero based on hierarchical rules;
- SCOREV4 – calculates risk score variables.

---

<sup>11</sup> 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.

Identical program files with .SAS and .TXT extensions are provided. The .TXT versions are easier to view with some programs. The user must use the files with extension .SAS when installing the software. File names are case sensitive on some computing platforms, so software modules assume that file names are upper case (e.g., I0V07ED1.SAS).

The software:

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

Step 2: Defines internal macro variables, formats, and internal macros.

Step 3: Merges the PERSON, NDC, HCPCS, and DIAGNOSIS SAS® data sets, 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 M/m/F/f/1/2), and all data sets must be sorted by the common person identifier variable. The name of the common person identifier variable is set in the macro variable &IDVAR (e.g., &IDVAR = *ID, or HICNO, or SSN, or EnrolleeID*).

Step 3.1: Declares variable lengths, retained variables, and arrays.

Step 3.2: Appends calibration coefficients for all models.

Step 3.3: Merges the PERSON, NDC, HCPCS and DIAGNOSIS data sets by the person identifier variable named in &IDVAR. Each enrollee must have exactly one PERSON record, and may have zero or more NDC, HCPCS, or DIAGNOSIS 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 using the crosswalk formats specified in parameter &RXCFMTN and &RXCFMTH (see Section II for details regarding the format library and formats specific to this version of software).

Step 3.6: If the enrollee has at least one diagnosis, this step: creates CCs using the crosswalk formats specified in parameter &CCFMT0Y1 and &CCFMT0Y2 (see Section II for details regarding the format library and formats specific to this version of software); performs ICD-10 edits using macro I0V07ED1; and creates additional CCs for some ICD-10 diagnoses.

Step 3.7: When the enrollee's last record is detected, this step: creates demographic variables using macro AGESEXV6; creates HCCs by applying hierarchy rules to CCs using macro V07141H1; sets HCCs to zero if the enrollee has no diagnoses; applies hierarchy rules to RXCs; sets RXCs 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., severe illness indicators, HCC groups, interaction terms, adult enrollment duration indicators, RxC and RxC\*HCC interactions); 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 uses SAS® CONTENTS and PRINT procedures to document the output data set.

## II. Files included with the software

The following programs and files are included:

- **CY21P07C** – main program containing all user-provided parameters (see below for the parameter and variable list). The program calls primary macro CY21M07C.
- **CY21M07C** – primary macro that merges input files, crosswalks NDCs and HCPCS to RXCs, crosswalks ICD-10 codes to CCs, creates HCC and risk score variables by calling various external and internal macros. 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. NDC and HCPCS not listed in the tables will be ignored by the software.
- **AGESEXV6** – creates age/sex variables.
- **I0V07ED1** – 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 macro.<sup>12</sup> 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 sex, and additional edits for CC age and sex splits.
- **V07141L1** – assigns labels to HCCs and RXCs. Table 4, HHS-Hierarchical Condition Categories (HCC) Hierarchies, lists the HCC labels.
- **V07141H1** – copies CCs into HCCs and sets selected HCCs to zero based on hierarchical rules. Table 4, HHS-Hierarchical Condition Categories (HCC) Hierarchies, summarizes the hierarchy assignments.
- **SCOREV4** – calculates risk score variables.
- **CY21F07C\_FY2021\_ICD10.TXT** – is a text version of the format that crosswalks ICD-10 codes to CC categories (and is provided for reference). The format includes ICD-10 codes valid in FY2021.
- **CY21F07C\_FY2022\_ICD10.TXT** – is a text version of the format that crosswalks ICD-10 codes to CC categories (and is provided for reference). The format includes ICD-10 codes valid in FY2022.
- **CY21F07C\_ICD10\_MCE\_AGE.TXT** – is a text version of the format that crosswalks ICD-10 codes to an acceptable age range if MCE edits on ICD-10 codes are to be performed (provided for reference only).

---

<sup>12</sup> The diagnosis-code edits used are based on the Definitions of Medicare Code Edits (MCEs), which are updated and published each year to correspond with ICD-10 code updates. The MCEs detect inconsistencies based on a person's age and diagnosis or sex and diagnosis.

- **CY21F07C\_ICD10\_MCE\_SEX.TXT** – is a text version of the format that crosswalks ICD-10 codes to an acceptable sex value if MCE edits on ICD-10 codes are to be performed (provided for reference only).
- **CY21F07C\_ICD10\_BUNDLED\_MOTHER.TXT** – is a text version of the format that contains FY2021/FY2022 completed pregnancy diagnoses for use in detecting mother-infant bundled claims (provided for reference only).
- **CY21F07C\_ICD10\_BUNDLED\_INFANT.TXT** – is a text version of the format that contains FY2021/FY2022 newborn diagnoses for use in detecting mother-infant bundled claims (provided for reference only).
- **CY21F07C\_NDC7\_4\_21\_12.TXT** – is a text version of the format that contains Table 10a RXC to National Drug Code (NDC) Crosswalk.
- **CY21F07C\_HCPCS7\_4\_21\_12.TXT** – is a text version of the format that contains Table 10b RXC to Healthcare Common Procedure Coding System (HCPCS) Crosswalk.
- **CY21F07C.TRN** – a SAS® transport file containing one format library with all requisite formats. Format name suffixes must be specified as macro parameters in the main program as follows:
  - HHS\_V07FY21Q141C – 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 in FY2021. This suffix must be specified in macro parameter **CCFMT0Y1**.
  - HHS\_V07FY22Q141C – 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 in FY2022. This suffix must be specified in macro parameter **CCFMT0Y2** for this version of the software.
  - NDCV2112\_RXCV7\_4F – crosswalks NDC codes to RXC categories for codes valid in calendar year 2021. This format must be specified in macro parameter **RXCFMTN**.
  - HCPC2112\_RXCV7\_4F – crosswalks HCPCS codes to RXC categories for codes valid in calendar year 2021. This format must be specified in macro parameter **RXCFMTH**.
  - AGECY21MCE – crosswalks ICD-10 codes to an acceptable age range if MCE edits on ICD-10 codes are to be performed. This suffix must be specified in macro parameter **AGEFMT0**.
  - SEXCY21MCE – crosswalks ICD-10 codes to an acceptable sex value if MCE edits on ICD-10 codes are to be performed. This suffix must be specified in macro parameter **SEXFMT0**.
- **CY21C07A.TRN** – a SAS® transport file containing relative coefficients for regression models, created using CY2016, CY2017, and CY2018 data and corresponding denominators (defined as the weighted average plan liability for the full modeling sample for the given year).

The two SAS® transport files (with filename extension .TRN) contain the SAS® format library and model coefficients data set. They may be used on any SAS® version 9 platform after uploading them and converting them using SAS® PROC CIMPORT.



If your computing platform is z/OS, both transport files should be uploaded using the following attributes: RECFM(F or FB) LRECL(80) BLKSIZE(8000).

The two transport files should be converted (imported) as follows:

- **Model coefficients:**

```
FILENAME INC      "user defined location of transport file CY21C07A.TRN";
LIBNAME INCOEF    "user defined location for creation of coefficient file";

proc cimport infile=INC data=INCOEF.Coefficients; run;
```

- **Format library:**

```
FILENAME INF      "user defined location of transport file CY21F07C.TRN";
LIBNAME LIBRARY   "user defined location for creation of format library";

proc cimport infile=INF library=LIBRARY; run;
```

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

A. Diagnosis-level data set. The diagnosis input SAS® data set (DIAGNOSIS) 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 DIAGNOSIS 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 DIAGNOSIS data set according to the filtering logic below. This document provides filtering instructions and a list of the CY2020 (for historical data purposes) and CY2021 CPT/HCPCS codes that define service or procedure types that identify acceptable sources of diagnoses for risk adjustment.<sup>13</sup> However, the user must create the DIAGNOSIS 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 sources allowable for risk adjustment should be included in the DIAGNOSIS data set. ICD-10 codes that are not listed in Table 3 may be included in the DIAGNOSIS data set but are ignored by the software.<sup>14</sup> 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 2020 (for historical data

---

<sup>13</sup> Definitions taken directly from the Current Procedural Terminology (CPT®) codes and the Healthcare Common Procedure Coding System (HCPCS) code set. Note that although CY2020 codes are provided for historical purposes, this software is designed to be used only with CY2021 data.

<sup>14</sup> 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.

purposes) and 2021 CPT/HCPCS codes used to define service or procedure types that are acceptable sources of diagnoses for risk adjustment.

- 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 2020 or 2021.
- Column E identifies applicable footnotes on the CPT/HCPCS codes.
- Notes begin on row 6,601 of the Excel table with the line “Notes:” and should not be imported by any program.

The DIAGNOSIS 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
  - 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. 761 (community mental health center admit through discharge); or
    - vi. 767 (community mental health center replacement of prior claim); or
    - vii. 771 (federally qualified health center admit through discharge); or
    - viii. 777 (federally qualified health center replacement of prior claim).
    - ix. 851 (critical access hospital admit through discharge); or
    - x. 857 (critical access hospital replacement of prior claim); or

- xi. 87x (freestanding non-residential opioid treatment programs [OTPs]).
- 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 DIAG) and service date (variable DIAGNOSIS\_SERVICE\_DATE), the diagnosis code has a Y in the corresponding Table 3 Code Valid column. ICD-10 diagnosis codes with service dates of January 1, 2021 – September 30, 2021 should have a Y in the Code Valid in FY2021 column; otherwise, the user should exclude them. ICD-10 diagnosis codes with service dates of October 1, 2021 – December 31, 2021 should have a Y in the Code Valid in FY 2022 column; otherwise, the user should exclude them. As noted, this software can detect that an ICD-10 diagnosis code is not valid for a given fiscal year and will optionally flag the enrollee record in the “Errors/warnings/notes log” (see Section VIII.5, message 16).

**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)<sup>15</sup> 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

---

<sup>15</sup> 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.

- 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 CY21F07C\_ICD10\_BUNDLED\_MOTHER.TXT and CY21F07C\_ICD10\_BUNDLED\_INFANT.TXT 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, message 25).

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 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 National Drug Code input SAS® data set (NDC) must include NDCs used for risk adjustment, listed in Table 10a RXC 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 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 2021 benefit year HHS operated risk adjustment methodology is limited to the adult risk adjustment models. Users should not include information for child or infant enrollees in the NDC data set.

The NDC data set should include NDCs from pharmacy claims with **prescription filed 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.<sup>16</sup> (Note: Table 10a in the Excel file contains 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 Table 10a 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. Section IV further describes the NDC data set.

---

<sup>16</sup> 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>.

C. HCPCS-level data set. The Healthcare Common Procedure Coding System (HCPCS) input SAS® data set must include HCPCS codes used for risk adjustment RXCs, listed in Table 10b RxC to HCPCS Crosswalk. Inpatient, outpatient, and professional medical claims are acceptable sources for HCPCS codes. Inpatient and outpatient claims should be restricted to the same facility bill type codes used for the diagnosis data set (see Section III. 2a and 3a). HCPCS should only be used for medications when an NDC is not available from a pharmacy claim. The user must evaluate each claim to determine whether the claim’s HCPCS codes are included in the HCPCS data set.

The HCPCS codes in the HCPCS input data set are to be used for RxC 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 2021 benefit year HHS operated risk adjustment methodology is limited to the adult risk adjustment models. Users should not include information for child or 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 Table 10b may be included in the HCPCS data set but are ignored by the software and are not included in RXCs for the adult risk adjustment models’ risk score calculations. Section IV further describes the HCPCS data set.

#### IV. SAS® data sets supplied by the user

This section describes the four input SAS® data sets required to create CC and HCC groupings, RxC and RxC\*HCC interactions, enrollment duration variables, demographic variables, and risk score variables—a person-level data set (PERSON), a diagnosis data set (DIAGNOSIS), 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 and premium assistance Medicaid Alternative plans (i.e., private options) will be identified by the Health Insurance Oversight System (HIOS) variant ID. Listed below are the codes that will be used to identify the plan variation.<sup>17</sup> Please note that unlike the risk adjustment software person-level CSR indicator, the HIOS variant ID is a plan-level indicator.

<b>Cost-Sharing Reduction (CSR) Level</b>	<b>HIOS Variant ID</b>	<b>CSR RA Factor</b>	<b>RA Software Person-level CSR Indicator</b>
CSR: 94% AV Silver Plan Variation	06	1.12	1
CSR: 87% AV Silver Plan Variation	05	1.12	2
CSR: 73% AV Silver Plan Variation	04	1.00	3

<sup>17</sup> We note that Massachusetts CSR variant plans have a state-specific CSR factor table, as discussed in the 2021 Payment Notice final rule, 85 at 29190). In addition to the CSR variants listed above with factors of 1.12, plan variants of 04 are also 1.12 in Massachusetts only.

Cost-Sharing Reduction (CSR) Level	HIOS Variant ID	CSR RA Factor	RA Software Person-level CSR Indicator
CSR: Zero Cost Sharing – Platinum	02	1.00	4
CSR: Zero Cost Sharing – Gold	02	1.07	5
CSR: Zero Cost Sharing – Silver	02	1.12	6
CSR: Zero Cost Sharing – Bronze	02	1.15	7
CSR: Limited Cost Sharing – Platinum	03	1.00	8
CSR: Limited Cost Sharing – Gold	03	1.07	9
CSR: Limited Cost Sharing – Silver	03	1.12	10
CSR: Limited Cost Sharing – Bronze	03	1.15	11
CSR: Premium Assistance Medicaid Alternative Plan w/94% AV Silver Plan	36	1.12	12
CSR: Premium Assistance Medicaid Alternative Plan w/Zero Cost Sharing – Silver	32	1.12	13
Non-CSR/unknown CSR	00	1.00	0
Non-CSR/unknown CSR	01	1.00	0

#### Note on Enrollment Duration

The adult models include enrollment duration factors for months enrolled when an enrollee's enrollment period in an issuer's plans is less than 12 months. 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: [Msg33] Invalid ENROLDURATION, enrollee rejected*. Once created, the ENROLDURATION variable will be ignored for enrollees in the child or infant models.

STEP 2: The monthly enrollment duration indicator variables (ED\_1–ED\_11) are created by the software for adult enrollees as specified in Section VI.

The variable names must be spelled as written; SAS® variable names are case-insensitive (i.e., SEX and Sex and sex and SeX designate the same variable), but are illustrated in upper case.

#### 1. PERSON data set

- a. &IDVAR (Person identification code). As noted, &IDVAR is the name of the common person identifier variable (e.g., ID).
  - i. Character or numeric type, any length, not missing.
  - ii. Unique to an individual, and unique in the data set (i.e., no duplicates).
- b. SEX.
  - i. Character type, 1 byte, 1/M=male, 2/F=female, not missing.

- ii. Converted to upper case by the software.
- c. DOB.
  - i. Numeric type, 8-digit numeric field (YYYYMMDD), 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. Numeric type, integer, 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. Character type, 1 byte, P/G/S/B/C (only 1 of these values), not missing.<sup>18</sup>
  - 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 must be assigned CSR\_INDICATOR = 1-13. Non-CSR recipients must be assigned CSR\_INDICATOR = 0).
  - i. Numeric type, integer, 0-13, not missing.
  - ii. Values are:
    - 1 = Enrollees in 94% AV Silver Plan Variation.
    - 2 = Enrollees in 87% AV Silver Plan Variation.
    - 3 = Enrollees in 73% AV Silver Plan Variation.
    - 4 = Enrollee in Zero Cost Sharing Plan Variation of Platinum Level QHP.
    - 5 = Enrollee in Zero Cost Sharing Plan Variation of Gold Level QHP.
    - 6 = Enrollee in Zero Cost Sharing Plan Variation of Silver Level QHP.
    - 7 = Enrollee in Zero Cost Sharing Plan Variation of Bronze Level QHP.
    - 8 = Enrollee in Limited Cost Sharing Plan Variation of Platinum Level QHP.
    - 9 = Enrollee in Limited Cost Sharing Plan Variation of Gold Level QHP.
    - 10 = Enrollee in Limited Cost Sharing Plan Variation of Silver Level QHP.
    - 11 = Enrollee in Limited Cost Sharing Plan Variation of Bronze Level QHP.
    - 12 = Enrollee in a Premium Assistance Medicaid Alternative Plan with 94% AV Silver Plan Variation.

---

<sup>18</sup> 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.

- 13 = Enrollee in a Premium Assistance Medicaid Alternative Plan with Zero Cost Sharing Plan Variation of Silver Level QHP.
- 0 = Non-CSR recipient, and enrollees with unknown CSR.

g. ENROLDDURATION

- Numeric type, integer, 1-12, not missing.
- Person-level enrollment duration variable. Although ENROLDDURATION 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.
- 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. DIAGNOSIS data set

- &IDVAR (Person identification code). As noted, &IDVAR is the name of the common person identifier variable (e.g., ID).<sup>19</sup>
  - Character or numeric type, any length, not missing.
  - Unique to an individual.
- DIAG (ICD-10-CM diagnosis codes).
  - Character type, 7-byte field, no periods or embedded blanks, left justified.
  - Converted to upper case by the software.
  - Codes should be to the greatest level of available specificity.
  - Age and sex edits for diagnoses are performed in macro I0V07ED1 to ensure diagnoses are appropriate for the age and sex of the enrollee.
  - Only diagnoses from allowable sources should be included in the DIAGNOSIS data set.
  - Invalid diagnoses are ignored; warning messages are optional.<sup>20</sup>
  - A valid ICD-10 diagnosis must have a valid DIAGNOSIS\_SERVICE\_DATE.
- DIAGNOSIS\_SERVICE\_DATE

<sup>19</sup> Please note that in operation, this information can not include personally identifiable information.

<sup>20</sup> 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."



- i. Numeric type, 8-digit numeric field (YYYYMMDD), valid calendar date, not missing, provides the diagnosis's service date.<sup>21</sup>
- ii. As described in Section III, this variable can be used with DIAG and Table 3 to precheck that a diagnosis code is valid for a given fiscal year.<sup>22</sup>

### 3. NDC data set

- a. &IDVAR (Person identification code). As noted, &IDVAR is the name of the common person identifier variable (e.g., ID).<sup>23</sup>
  - i. Character or numeric type, any length, not missing.
  - ii. Unique to an individual.
- b. NDC
  - i. NDC normalized drug code (11-digit, no dashes, HIPAA standard format), 11-character field, 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. &IDVAR (Person identification code). As noted, &IDVAR is the name of the common person identifier variable (e.g., ID).<sup>24</sup>
  - i. Character or numeric type, any length, not missing.
  - ii. Unique to an individual.
- b. HCPCS
  - i. HCPCS drug code, 5-character field, left justified. Only HCPCS drug codes from inpatient, outpatient, or professional medical claims from allowable sources (Section III) with discharge or through dates in the benefit year (Section III C) for adult enrollees should be included in the HCPCS data set.

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 DIAGNOSIS data set. It is used only for MCE diagnosis code age edits.

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

---

<sup>21</sup> Valid diagnosis service date in this version of software (V0721 141 A3): year is 2021, 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.

<sup>22</sup> 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 FY2021 code with a FY2021 service date), the optional software warning message will be Message 16 *Diagnosis lookup failed, diagnosis ignored*.

<sup>23</sup> 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 FY2021 code with a FY2021 service date), the optional software warning message will be Message 16 *Diagnosis lookup failed, diagnosis ignored*.

<sup>24</sup> Please note that in operation, this information can not include personally identifiable information.

- 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	M	19581201	63	P	0	12
202	F	20080315	13	C	0	12
301	F	19660414	55	G	5	7
302	M	19720101	49	B	11	12
304	X	19700132		R	16	3
305	M	19810101	40	S	0	12

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

ID	DIAG	DIAGNOSIS_SERVICE_DATE
201	E118	20210113
201	H9319	20210113
201	M532X9	20210629
201	M25461	20210630
201	M25569	20210706
201	M25579	20210706
201	209	20210835
202	J4530	20210219
302	J200	20210317
302	Z430	20210504
303	E890	20210929
304	Z0000	20210617
305	B20	20210302

- 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, character field, 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 digits, 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 DIAGNOSIS, 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 DIAGNOSIS 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 demographic information (e.g., ID 301).

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

## V. Parameters supplied by the user

The user must set the following parameters when calling macro CY21M07C:

- **INP** – input PERSON SAS® data set name (e.g., *IN1.Person*).
- **IND** – input DIAGNOSIS SAS® data set name (e.g., *IN2.Diagnosis*).
- **INN** – input NDC SAS® data set name (e.g., *IN3.NDC*).
- **INH** – input HCPCS SAS® data set name (e.g., *IN4.HCPCS*).
- **OUTDATA** – output SAS® data set name (e.g., *OUT.OutputScores*).
- **IDVAR** – name of the person identifier variable (e.g., *ID*, or *HICNO*, or *SSN*, or *EnrolleeID*). This variable can be either character or numeric type, and any length.
- **KEEPVAR** – variables written to the output data set. There is a list of KEEP variables in the program, but the user can alter the list (e.g., *DOB*, *AGE\_LAST*, *SEX*, *METAL*, *CSR\_INDICATOR*, *SCORE\_*;, *CSR\_ADJ\_SCR\_*;, or *\_ALL\_*).
- **CCFMT0Y1** – format name suffix for formats that crosswalk ICD-10 codes to HHS-CCs for fiscal year 2021. For this version of the software it is *HHS\_V07FY21Q141C*.
- **CCFMT0Y2** – format name suffix for formats that crosswalk ICD-10 codes to HHS-CCs for fiscal year 2022. For this version of the software it is *HHS\_V07FY22Q141C*.
- **RXCFMTN** – format name for format that crosswalks NDC codes to RXC for calendar year 2021. For this version of software, it is *NDCV2112\_RXCV7\_4F*.
- **RXCFMTH** – format name for format that crosswalks HCPCS codes to RXC for calendar year 2021. For this version of software, it is *HCPC2112\_RXCV7\_4F*.
- **AGEFMT0** – format name suffix for formats that crosswalk ICD-10 codes to an acceptable age range when MCE edits on ICD-10 codes are performed. For this version of the software it is *AGECY21MCE*.
- **SEXFMT0** – format name suffix for formats that crosswalk ICD-10 codes to an acceptable sex value when MCE edits on ICD-10 codes are performed. For this version of the software it is *SEXCY21MCE*.

## VI. Variables output by the software

The software generates a person-level output SAS® data set. As noted, the user can specify variables to KEEP in the **KEEPVAR** parameter of the macro CY21M07C call.

The following variables can be specified:

1. Any person-level variable from the original PERSON data set.
2. 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

3. CCs created by the software (before hierarchies are applied).
4. HCCs created by the software (after hierarchies are applied).
5. HCC groups and HCC interactions created by the software.
6. RXCs created by the software (after hierarchies are applied).
7. RXC\*HCC interactions created by the software.
8. Adult models enrollment duration indicators (ED\_1–ED\_11) created by the software.
9. Infant model maturity categories, severity level categories, and maturity by severity level interactions created by the software.
10. 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

11. 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

12. 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 SAS® v9 on this platform:

- Linux (server)

## VIII. Steps

1. Install software:

- Copy files to the computing platform on which the risk scores will be calculated. If the platform is z/OS, upload the two transport files (.TRN) using RECFM(F or FB) LRECL(80) BLKSIZE(8000).

- Use files with .SAS extensions. Files with .TXT extensions are identical, but might be more easily viewed by the user. File names are case sensitive on some computing platforms; software modules assume that file names are upper case (e.g., I0V07ED1.SAS).
2. Prepare software-provided SAS® input format library and coefficients data set:
    - Convert both .TRN files (containing the SAS® format library and model coefficients data set) using SAS® PROC CIMPORT on the computing platform on which the risk scores will be calculated as described in Section II.
    - The format library and coefficients data set are provided with the software, but must be imported by the user; they are not imported by the risk adjustment modeling software.
  3. Prepare user-provided SAS® input data sets:
    - Create PERSON, DIAGNOSIS, 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.
  4. Generate scores:
    - Set parameters as described in Section V.
    - Execute SAS® program CY21P07C and generate variables described in Section VI.
  5. Review errors/warnings, notes: the software prints messages in the “Errors/warnings/notes log” for various situations. The user may print (or suppress printing) any of them. To print messages of type nn, set macro variable MSGnn to blank; e.g., %let MSG01= ; . To suppress printing messages of type nn, set macro variable MSGnn to \*; e.g., %let MSG01=\*; .

We recommend the following be printed because they indicate possible errors in data sets, variables or variable values:

```

ERROR : [Msg01] Variable --- is not in --- file
ERROR : [Msg02] User-provided variable --- in --- file must be --- type
ERROR : [Msg03] Duplicate IDVARS in PERSON file
ERROR : [Msg04] Program halted due to duplicate IDVARS in PERSON file
OK : [Msg05] PERSON file is free of duplicate IDVARS
ERROR : [Msg06] Program halted due to non-existent variable(s) in PERSON file
OK : [Msg07] PERSON file contains all requisite variables
ERROR : [Msg08] Program halted due to incorrect user-provided variable type(s) in PERSON file
OK : [Msg09] PERSON file's variables have the correct type
ERROR : [Msg10] Program halted due to non-existent variable(s) in DIAG file
OK : [Msg11] DIAG file contains all requisite variables
ERROR : [Msg12] Program halted due to incorrect user-provided variable type(s) in DIAG file
OK : [Msg13] DIAG file's variables have the correct type
WARNING: [Msg14] Diagnosis matches no enrollee, diagnosis ignored
WARNING: [Msg15] Blank diagnosis code, diagnosis ignored
WARNING: [Msg18] Missing IDVAR, enrollee rejected
WARNING: [Msg19] Invalid SEX, enrollee rejected
WARNING: [Msg20] Invalid DOB, enrollee rejected
WARNING: [Msg21] Invalid AGE_LAST, enrollee rejected
WARNING: [Msg22] Invalid METAL, enrollee rejected
WARNING: [Msg23] Invalid CSR_INDICATOR, enrollee rejected
WARNING: [Msg24] Failed HHS HCC filter, enrollee rejected
WARNING: [Msg27] Invalid DIAGNOSIS_SERVICE_DATE, diagnosis ignored
WARNING: [Msg28] Invalid AGE_AT_DIAGNOSIS, diagnosis ignored
WARNING: [Msg29] AGE_AT_DIAGNOSIS > AGE_LAST, diagnosis ignored
ERROR : [Msg30] Program halted, file --- does not exist

```

```

WARNING: [Msg31] AGE_LAST minus AGE_AT_DIAGNOSIS > 1, diagnosis ignored
WARNING: [Msg32] DOB > DIAGNOSIS_SERVICE_DATE, diagnosis ignored
WARNING: [Msg33] Invalid ENROLDURATION, enrollee rejected
ERROR   : [Msg34] Program halted due to non-existent variable(s) in NDC file
OK      : [Msg35] NDC file contains all requisite variables
ERROR   : [Msg36] Program halted due to incorrect user-provided variable type(s) in NDC file
OK      : [Msg37] NDC file`s variables have the correct type
WARNING: [Msg38] NDC matches no enrollee, NDC ignored
WARNING: [Msg39] Blank NDC code, NDC ignored
ERROR   : [Msg41] Program halted due to non-existent variable(s) in HCPCS file
OK      : [Msg42] HCPCS file contains all requisite variables
ERROR   : [Msg43] Program halted due to incorrect user-provided variable type(s) in HCPCS file
OK      : [Msg44] HCPCS file`s variables have the correct type
WARNING: [Msg45] HCPCS matches no enrollee, HCPCS ignored
WARNING: [Msg46] Blank HCPCS code, HCPCS ignored

```

We recommend the following be printed during testing with small data sets. The user may choose to suppress printing the messages during production runs with large data sets as these conditions tend to generate many messages.

```

WARNING: [Msg16] Diagnosis lookup failed, diagnosis ignored
NOTE    : [Msg17] Enrollee has no diagnoses, risk score based on remaining information
WARNING: [Msg25] Possible bundled mother/infant claim(s) -- ---
WARNING: [Msg40] NDC lookup failed, NDC ignored
WARNING: [Msg47] HCPCS lookup failed, HCPCS ignored

```

Suppressing printed output for type nn does not affect whether an enrollee record or diagnosis is rejected. I.e., diagnosis code ZZZZZZ will be ignored by the software even if %let MSG16=\*; is set.

**End of Document**