

Version 03 HHS-HCC Risk Adjustment Modeling Software Documentation

Section 1343 of the Affordable Care Act provides for a permanent risk adjustment program. To protect against potential effects of adverse selection, the risk adjustment program transfers funds from plans with relatively lower-risk enrollees to plans with relatively higher-risk enrollees. It generally applies to non-grandfathered individual and small group plans inside and outside Exchanges.

The HHS risk adjustment methodology is described in the HHS Notice of Benefit and Payment Parameters for 2014, final rule (CMS-9964-F), which was published in the *Federal Register* on March 11, 2013.

The methodology that HHS will use when operating a risk adjustment program on behalf of a State for the 2014 benefit year will calculate a plan average risk score for each covered plan based upon the relative risk of the plan's enrollees, and apply a payment transfer formula in order to determine risk adjustment payments and charges between plans within a risk pool within a market within a State. The risk adjustment methodology addresses three considerations: (1) the newly insured population; (2) plan metal level differences and permissible rating variation; and (3) the need for risk adjustment transfers that net to zero. The risk adjustment methodology developed by HHS for the 2014 benefit year:

- Is developed on commercial claims data for a population similar to the expected population to be risk adjusted;
- 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 risk adjustment population;
- Establishes concurrent risk adjustment models, one for each combination of metal level (platinum, gold, silver, bronze, catastrophic) and age group (adult, child, infant);
- Results in payment transfers that net to zero within a risk pool within a market within a State;
- Adjusts payment 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 risk pool within a market within a State.

This document provides instructions for the HHS risk adjustment models for the 2014 benefit year, with revisions from the software instructions posted on the CCHIO website on May 7, 2013.

Key Revisions:

- Updated Table 1 to add 2014 CPT/HCPCS used for diagnosis filtering, as described in Section III.
- Revised directions for professional source of diagnosis (Section III).
- Removed bill type 110 (inpatient nonpayment/zero claim) from inpatient facility source of diagnosis; removed bill type 130 (hospital outpatient nonpayment/zero claim) and

revised number for bill type from 73x to 77x (federally qualified health center) for outpatient facility source of diagnosis (Section III).

- Revised note on bundled claims for mother and newborn infant (Section III), and added warning message for bundled claim records (V0314L1M).
- Revised directive for assigning cost-sharing indicator to enrollees in premium assistance Medicaid alternative plans (Section IV and SCOREV3).
- Added version code and diagnosis date to the diagnosis-level file and included information on distinguishing between ICD-9-CM and ICD-10-CM diagnosis codes (Section IV).
- Removed AGE_FIRST from person-level file, and added date of birth (DOB) to the person-level file (Section IV).
- Added DIAGNOSIS_SERVICE_DATE to the diagnosis-level file (Section IV), and replaced AGE_FIRST with AGE_AT_DIAGNOSIS for MCE age conditions (computed from DIAGNOSIS_SERVICE_DATE and DOB) (V03EDIT2).
- Updated Table 2 to add FY2014 ICD-9-CM codes and specified that MCE age and sex conditions are for FY2014 (H0314L1O.TRN).

The 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. We note that these instructions and the corresponding software contain only ICD-9-CM codes. The United States has delayed the implementation of ICD-10-CM diagnosis codes until at least October 1, 2015. The preliminary ICD-10 to HHS-Condition Category (CC) crosswalk comparable to Table 2 will be posted in Summer 2014. We also intend to provide combined ICD-9 and ICD-10 instructions when we issue the preliminary ICD-10 crosswalk.

The model instructions are based on the methodology described in the final notice of benefit and payment parameters. Please direct questions regarding the model instructions to HHS HCC Risk Adjustment Models at hhshccraops@cms.hhs.gov. This mailbox will be used only to answer questions pertaining to operations of the HHS risk adjustment model as posted to this site. We look forward to assisting with inquiries pertaining to your risk adjustment program operations using the HHS-HCC risk adjustment models for the 2014 benefit year.

CMS has created two versions of software (SAS software and HHS-developed risk adjustment model algorithm software) and software instructions for issuers to use with their enrollment data to simulate their enrollee populations' risk scores within the risk adjustment model, as finalized in the 2014 notice of benefit and payment parameters. **This software is being issued only as supplemental guidance for issuers to better understand and simulate the calculation of plan liability risk scores for their expected 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 risk adjustment program. Further, 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 information. This software merely provides a simulation to calculate enrollees' risk scores.

This document describes software for HHS-HCC risk adjustment modeling (version 03). The software requires SAS® version 9.

Terminology: In this document, the abbreviations HHS-HCC and HCC are used interchangeably for Health and Human Services Hierarchical Condition Categories. The abbreviations HHS-CC and CC are used interchangeably for HHS Condition Categories. The abbreviations ICD-9 and ICD-9-CM are used interchangeably for International Classification of Diseases, 9th Revision, Clinical Modification. The abbreviations ICD-10 and ICD-10-CM are used interchangeably for International Classification of Diseases, 10th Revision, Clinical Modification.

I. Software description

The software reads two user-provided input SAS® datasets; constructs demographic variables for each enrollee; crosswalks ICD-9 diagnoses to Condition Categories (CCs) using SAS® formats which are stored in a FORMAT library; and creates Hierarchical Condition Categories (HCCs) by imposing hierarchies on the CCs.

The software uses the demographic variables and HHS-HCCs 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 are computed from demographic variables; i.e., zeros are assigned to all HHS-CCs and HHS-HCCs.

The software's main program (V0314L1P) calls primary macro V0314L1M and passes a set of user-specified parameters (a macro is a subroutine that performs a specific task). Macro V0314L1M calls six external macros (provided as separate files):

- AGESEXV6 – creates age/sex variables;
- V03EDIT2 – performs edits on ICD-9 codes based on age and/or sex;
- V03127M2 – assigns selected ICD-9 codes to multiple HHS-CCs;
- V03127L2 – assigns labels to HHS-HCCs;
- V03127H2 – sets selected HHS-HCCs to zero based on hierarchical rules;
- SCOREV3 – calculates risk score variables.

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., V03EDIT2.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; these are least likely to vary among releases.

Step 3: Merges the PERSON and DIAGNOSIS SAS® datasets, and outputs one record for each enrollee record in the PERSON dataset. Input records must be fully compliant with validity rules (e.g., SEX must be M/m/F/f/1/2), and both datasets 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 and DIAGNOSIS datasets by the person identifier variable named in &IDVAR. Each enrollee must have exactly one PERSON record, and may have zero or more 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 diagnosis, this step: creates HHS-CCs using the crosswalk format specified in parameter &FMNAME (see Section II for details regarding the format library and formats specific to this version of software); performs ICD-9 edits using macro V03EDIT2; and creates additional HHS-CCs using macro V03127M2 for some ICD-9 diagnoses.

Step 3.6: When the enrollee's last record is detected, this step: creates demographic variables using macro AGESEXV6; creates HHS-HCCs by applying hierarchy rules to CCs using macro V03127H2; sets HHS-HCCs to zero if the enrollee has no diagnoses; applies validity filters to various input variables; creates additional model-specific variables (e.g., severe illness indicators, HHS-HCC groups, interaction terms); 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 procedure calls to document the output dataset.

II. Files included with the software

The following programs and files are included:

- **V0314L1P** – main program containing all user-provided parameters (see below for the parameter and variable list). The program calls primary macro V0314L1M.
- **V0314L1M** – primary macro that merges input files, crosswalks ICD-9 codes to HHS-CCs, creates HHS-HCC and risk score variables by calling various external and internal macros. Table 2, ICD-9 to HHS-Condition Categories (CC) Crosswalk, summarizes the

ICD-9 to CC assignments. Only ICD-9 codes assigned to HHS-HCCs in the risk adjustment models are included in this crosswalk. All other ICD-9 codes will be ignored by the software.

- **AGESEXV6** – creates age/sex variables.
- **V03EDIT2** – performs edits on ICD-9 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. If the enrollee has invalid age and/or sex for a particular ICD-9 code, then the ICD-9 code will be ignored. Table 2, ICD-9 to HHS-Condition Categories (CC) Crosswalk, summarizes the ICD-9 code edits; it describes the ICD-9 Medicare Code Edits (MCEs) for age and sex, and additional edits for CC age and sex splits.¹
- **V03127M2** – assigns selected ICD-9 diagnosis codes to additional HHS-CCs as required in Table 2.
- **V03127L2** – assigns labels to HHS-HCCs. Table 3, HHS-Hierarchical Condition Categories (HCC) Hierarchies, lists the HHS-HCC labels.
- **V03127H2** – copies HHS-CCs into HHS-HCCs and sets selected HHS-HCCs to zero based on hierarchical rules. Table 3, HHS-Hierarchical Condition Categories (HCC) Hierarchies, summarizes the hierarchy assignments.
- **SCOREV3** – calculates risk score variables.
- **H0314L1O.TXT** – is a text version of the formats that crosswalk ICD-9 codes to HHS-CC categories (and is provided for reference). The formats include ICD-9 codes valid in FY2011 or FY2012 or FY2013 or FY2014.
- **H0314L1O.TRN** – a SAS® transport file containing one format library with all requisite formats. Format names must be specified as macro parameters in the main program as follows:
 - **HHS_I03141Y11Y12Y13Y14OC** – crosswalks ICD-9 codes to HHS-CC categories that are transformed to HHS-HCC categories, and contains ICD-9 codes used in the risk adjustment models that are valid in FY2011 or FY2012 or FY2013 or FY2014. This must be specified in macro parameter **FMNAME**.
 - **HHS_AGEY11Y12Y13Y14OMCE** – crosswalks ICD-9 codes to an acceptable age range if MCE edits on ICD-9 codes are to be performed. This must be specified in macro parameter **AGEFMT**.
 - **HHS_SEXY11Y12Y13Y14OMCE** – crosswalks ICD-9 codes to an acceptable sex value if MCE edits on ICD-9 codes are to be performed. This must be specified in macro parameter **SEXFMT**.
- **C0310L1O.TRN** – a SAS® transport file containing relative coefficients for regression models created using CY2010 data, and a denominator defined as the weighted average plan liability for the full modeling sample.

¹ 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-9 code updates. The MCEs detect inconsistencies based on a person's age and diagnosis or sex and diagnosis. In addition to the standard MCE age 0 edit list, macro V03EDIT2 and Table 2 include these codes as requiring age = 0: V302, V312, V322, V332, V342, V352, V362, V372, and V392. These codes were added because, like the other codes in the V30-V39 range which do appear on the MCE age 0 edit list, they correspond to the infant's birth status and should not appear on the mother's record or on an older child's record.

The two SAS® transport files (with filename extension .TRN) contain the SAS® format library and model coefficients dataset. 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 C0310L10.TRN";  
LIBNAME INCOEF    "user defined location of SAS coefficients dataset";
```

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

- Format library:

```
FILENAME INF      "user defined location of transport file H0314L10.TRN";  
LIBNAME LIBRARY   "user defined location of SAS format library";
```

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

III. Creation of a diagnosis dataset according to sources of diagnoses allowable for risk adjustment

The diagnosis input SAS® dataset (DIAGNOSIS) must include ICD-9-CM diagnosis codes used for risk adjustment. These diagnosis codes are listed in reference Table 2, ICD-9 to HHS-Condition Category (CC) Crosswalk.² The user must evaluate each claim or encounter record to determine whether its diagnoses are included in the DIAGNOSIS dataset. 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 HHS-CC (and HHS-HCC) creation. It is the user's responsibility to create the DIAGNOSIS dataset according to the filtering logic below. This document provides filtering instructions and a list of the 2012 (for historical data purposes) and 2014 CPT/HCPCS codes that identify acceptable sources of diagnoses for risk adjustment.³ However, the user must create the DIAGNOSIS dataset for input to the risk adjustment algorithm; the dataset is not created by the software.

NOTE: CMS stated that 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.

² The United States has delayed implementation of ICD-10-CM diagnosis codes until at least October 1, 2015. The preliminary ICD-10 to HHS-Condition Category (CC) crosswalk comparable to Table 2 will be posted in Summer 2014.

³ Definitions taken directly from the Current Procedural Terminology (CPT®) codes and the Healthcare Common Procedure Coding System (HCPCS) code set.

Only ICD-9-CM diagnosis codes that are from sources allowable for risk adjustment should be included in the DIAGNOSIS dataset. ICD-9 codes that are not listed in Table 2 may be included in the DIAGNOSIS dataset, but are ignored by the software. The steps below provide logic to determine which diagnoses are allowable. Note that Steps 1 and 3 refer to Table 1, CPT/HCPCS Included List, which provides the 2012 and 2014 CPT/HCPCS codes used to define service 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 1. Column B contains the short descriptions of the procedure codes. Columns C and D, respectively, indicate whether a CPT/HCPCS code is acceptable in 2012 or 2014. Column E identifies applicable notes for the CPT/HCPCS codes. Notes begin on row 7059 of the Excel table with the line “Notes:”, and should not be imported by any program.

The diagnosis-level input file 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.

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 1). 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. (blank) = 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. 114 (inpatient interim- last claim); or
 - iii. 116 (inpatient adjustment of prior claim); or
 - iv. 117 (inpatient replacement of prior claim).
 - b. If the facility bill type code =114, then also use all header diagnoses on related records with facility bill type code = 112 (inpatient interim-first claim) or 113 (inpatient interim-continuing claim).
 - c. If the facility bill type code =117 and is being used as a 114 or final interim bill, then also use all header diagnoses on related records with facility bill type code = 117 and discharge status code 30 (still a patient).
 - d. There is no procedure screen for these 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 though discharge); or

- ii. 132 (hospital outpatient interim- first claim); or
 - iii. 133 (hospital outpatient interim- continuing claim); or
 - iv. 134 (hospital outpatient interim-last claim); or
 - v. 136 (hospital outpatient adjustment of prior claim); or
 - vi. 137 (hospital outpatient replacement of prior claim); or
 - vii. 71x (rural health clinic); or
 - viii. 77x (federally qualified health center); or
 - ix. 76x (community mental health center).
- b. For records with at least one acceptable CPT/HCPSC code (Table 1) on a service line, use all header diagnoses. Otherwise, do not use the diagnoses for risk adjustment.

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 dataset and used for appropriately calculating risk scores.

The user will need a non-provided program to detect these bundled claims and redefine them as two separate claims (i.e., this software can detect possible bundled claims, but cannot redefine them). A program that detects and redefines bundled claims would need to identify enrollees with a claim containing these 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)⁴ and
- ICD-9 diagnoses corresponding to a completed pregnancy HCC (HHS-HCC 207 or 208 or 209) and
- ICD-9 diagnoses corresponding to a newborn HCC (HHS-HCC 242 or 243 or 244 or 245 or 246 or 247 or 248 or 249).

Infant is the enrollee:

- AGE_LAST $<$ 2 (an age corresponding to the infant model; more specifically age should be appropriate for a newborn diagnosis) and
- ICD-9 diagnoses corresponding to a completed pregnancy HCC (HHS-HCC 207 or 208 or 209) and
- ICD-9 diagnoses corresponding to a newborn HCC (HHS-HCC 242 or 243 or 244 or 245 or 246 or 247 or 248 or 249).

See Table 2, ICD-9 to HHS-Condition Category (CC) Crosswalk, for diagnosis codes corresponding to the completed pregnancy and newborn HCCs.

⁴ Section IV of this document identifies the two age variables used in the software and specifies when each is used.

If the bundled claim is not detected by the user and redefined as two separate claims (one for the mother and one for the infant) for use in preparing the DIAGNOSIS dataset, the bundled claims should be included. If the enrollee is a female with an age appropriate for a maternity diagnosis, then the diagnoses on the bundled claim are assumed to correspond to the mother's enrollee ID. If the enrollee is less than age 2, then the diagnoses on the bundled claim are assumed to correspond to the infant's enrollee ID. All diagnoses on a bundled claim that could apply to either the mother or the infant (i.e., diagnoses that do not require passing an age/sex edit) are assumed to apply to the enrollee's ID. All HHS-HCCs that are valid for the mother or infant (based on diagnoses that pass or do not require age/sex edits) will be used to compute the enrollee's risk score; and all diagnoses that do not pass age/sex edits (i.e., newborn diagnosis codes for the mother; pregnancy diagnosis codes for the infant) will be ignored.

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-level file. Infants without diagnoses will be assigned to the lowest severity category and the "term" maturity category for infants. Male infants will also have the male demographic factor assigned.

IV. SAS® datasets supplied by the user

This section describes the two input SAS® datasets required to create HHS-CC and HHS-HCC groupings, demographic variables, and risk score variables—a person-level dataset (PERSON) and a diagnosis dataset (DIAGNOSIS). It is the responsibility of the user to create these input datasets with the variables listed in this section. Both input datasets must be ordered in ascending order by the person identifier variable.

Although the ICD-10-CM implementation has been delayed and the ICD-9-CM and ICD-10-CM diagnosis code transition may not occur until calendar year 2015, we are providing instructions in preparation of that transition. As these instructions apply only to ICD-9 diagnoses, the user must identify and remove (or bypass) ICD-10 codes when ICD-9 and ICD-10 codes are comingled in user data files. As we noted above, we will issue ICD-9 and ICD-10 combined instructions in future guidance.

ICD-9 codes are 3-5 bytes and ICD-10 codes are 3-7 bytes, so code length cannot uniquely identify ICD-9 codes in comingled files. A small number of ICD-9 and ICD-10 codes are identical but have clinically different meanings, so a lookup table cannot uniquely identify ICD-9 codes in comingled files.

Users must include a version code to indicate whether a user-provided diagnosis code is ICD-9 or ICD-10, and include a diagnosis service date. The software will use these indicators to confirm ICD-9 diagnoses.

Note on CSR_INDICATOR

In operations, cost-sharing reduction 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. Please note that unlike the risk adjustment software indicator, the HIOS variant ID is a plan-level indicator.

CSR Level	HIOS Variant ID	CSR RA Factor	RA Software Person-level Indicator
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
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 Plan	03	1.00	8
CSR: Premium Assistance Medicaid Alternative Plan w/94% AV Silver Plan	36	1.12	9
CSR: Premium Assistance Medicaid Alternative Plan w/Zero Cost Sharing - Silver	32	1.12	10
Non-CSR/unknown CSR	00	1.00	0

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 dataset

- 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 dataset (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 tasks 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.⁵
 - 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 =1-10. Non-CSR recipients must be assigned CSR_INDICATOR = 0).
 - i. Numeric type, integer, 0-10, 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.
 - 9 = Enrollee in a Premium Assistance Medicaid Alternative Plan with 94% AV Silver Plan Variation
 - 10 = 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.

2. DIAGNOSIS dataset

- a. &IDVAR (Person identification code). As noted, &IDVAR is the name of the common person identifier variable (e.g., ID).⁶
 - i. Character or numeric type, any length, not missing.
 - ii. Unique to an individual.
- b. DIAG (ICD-9-CM diagnosis code).
 - i. Character type, 5 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 macro V03EDIT2 to ensure diagnoses are appropriate for the age and sex of the enrollee.
 - v. Only diagnoses from allowable sources should be included in the DIAGNOSIS dataset.
 - vi. Invalid diagnoses are ignored; warning messages are optional.

⁵ 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.

⁶ Please note that in operation, this information can not include personally identifiable information.

- vii. A valid ICD-9 diagnosis must have a DIAGNOSIS_VERSION_CODE indicating ICD-9 and a valid DIAGNOSIS_SERVICE_DATE.
- c. DIAGNOSIS_VERSION_CODE
 - i. Character type, 1 byte field, indicates whether a diagnosis code is ICD-9-CM or ICD-10-CM.
- d. DIAGNOSIS_SERVICE_DATE
 - i. Numeric type, 8 digit numeric field (YYYYMMDD), valid calendar date, not missing, provides the diagnosis's service date.

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

The two user-provided datasets (PERSON and DIAGNOSIS) are illustrated below. These examples are not based on actual data.

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

ID	SEX	DOB	AGE_LAST	METAL	CSR_INDICATOR
201	M	19521201	61	P	0
202	F	19521115	61	C	6
301	F	19600414	53	G	2
302	M	19660101	47	B	4
304	X	19660132		R	16

Diagnosis dataset example (DIAGNOSIS) containing four variables; we use ID as the person identifier variable to illustrate:

ID	DIAG	DIAGNOSIS_VERSION_CODE	DIAGNOSIS_SERVICE_DATE
201	2440	9	20130113
201	38831	9	20130113
201	71783	9	20130629
201	71906	9	20130630
201	71946	9	20130706
201	71947	9	20130706
201	ABCDE	3	20130835
202	4660	9	20130219
302	4660	9	20130317
302	V550	9	20130504
303	2440	9	20131129

- ID 301 has no diagnoses; the other IDs in PERSON have one or more diagnoses.
- ID 303 in DIAGNOSIS 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 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 ABCDE).
- Risk scores for enrollees without diagnoses 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 V0314L1M:

- **INP** – input PERSON SAS® dataset name (e.g., *IN1.Person*).
- **IND** – input DIAGNOSIS SAS® dataset name (e.g., *IN2.Diagnosis*).
- **OUTDATA** – output SAS® dataset 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 dataset. 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_*).
- **FMNAME** – format that crosswalks ICD-9 codes to HHS-CCs. For this version of the software it is *HHS_I03I4IY1IY12Y13Y14OC*.
- **AGEFMT** – format name that crosswalks ICD-9 codes to an acceptable age range when MCE edits on ICD-9 codes are performed. For this version of the software it is *HHS_AGEY1IY12Y13Y14OMCE*.
- **SEXFMT** – format name that crosswalks ICD-9 codes to an acceptable sex value when MCE edits on ICD-9 codes are performed. For this version of the software it is *HHS_SEXY1IY12Y13Y14OMCE*.

VI. Variables output by the software

The software generates a person-level output SAS® dataset. As noted, the user can specify variables to KEEP in the **KEEPVAR** parameter of the macro V0314L1M call.

The following variables can be specified:

1. Any person-level variable from the original PERSON dataset.
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. HHS-CCs created by the software (before hierarchies are applied).
4. HHS-HCCs created by the software (after hierarchies are applied).
5. HHS-HCC groups and HHS-HCC interactions created by the software.
6. Infant model maturity categories, severity level categories, and maturity by severity level interactions created by the software.
7. 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
8. CSR-adjusted score variables:
 - a. Adult model
 - i. CSR_ADJUSTED_SCORE_ADULT_PLATINUM

- ii. CSR_ADJUSTED_SCORE_ADULT_GOLD
 - iii. CSR_ADJUSTED_SCORE_ADULT_SILVER
 - iv. CSR_ADJUSTED_SCORE_ADULT_BRONZE
 - v. CSR_ADJUSTED_SCORE_ADULT_CATASTROPHIC
 - b. Child model
 - i. CSR_ADJUSTED_SCORE_CHILD_PLATINUM
 - ii. CSR_ADJUSTED_SCORE_CHILD_GOLD
 - iii. CSR_ADJUSTED_SCORE_CHILD_SILVER
 - iv. CSR_ADJUSTED_SCORE_CHILD_BRONZE
 - v. CSR_ADJUSTED_SCORE_CHILD_CATASTROPHIC
 - c. Infant model
 - i. CSR_ADJUSTED_SCORE_INFANT_PLATINUM
 - ii. CSR_ADJUSTED_SCORE_INFANT_GOLD
 - iii. CSR_ADJUSTED_SCORE_INFANT_SILVER
 - iv. CSR_ADJUSTED_SCORE_INFANT_BRONZE
 - v. CSR_ADJUSTED_SCORE_INFANT_CATASTROPHIC
- 9. 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_ADJUSTED_SCORE_ADULT
 - b. Child scores
 - i. SCORE_CHILD
 - ii. CSR_ADJUSTED_SCORE_CHILD
 - c. Infant scores
 - i. SCORE_INFANT
 - ii. CSR_ADJUSTED_SCORE_INFANT

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

VII. Computing platforms

The software has been tested using SAS® v9 on three platforms:

- MS Windows (PC)
- Linux (server)
- z/OS (IBM mainframe).

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., V03EDIT2.SAS).
2. Prepare software-provided SAS® input format library and coefficients dataset:
 - Convert both .TRN files (containing the SAS® format library and model coefficients dataset) 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 dataset 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 datasets:
 - Create PERSON and DIAGNOSIS datasets using the guidelines in Section III and dataset descriptions in Section IV.
 - These datasets 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 V0314L1P 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 datasets, 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: [Msg26] Invalid DIAGNOSIS_VERSION_CODE, diagnosis ignored
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
```

We recommend the following be printed during testing with small datasets. The user may choose to suppress printing the messages during production runs with large datasets 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 demographic information
WARNING: [Msg25] Possible bundled mother/infant claim(s) -- ---
```

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