Methods for calculating risk predictions using RSI 3.0 Models.

Document Revision 1.0
October 20, 2022

Overview
This document describes how to make outcome predictions using RSI 3.0 models described in: Greenwald S, Chamoun GF, Chamoun NG, Clain D, Hong Z, Jordan R, Manberg PJ, Maheshwari K, Sessler DI: *Risk Stratification Index 3.0, a Broad Set of Models for Predicting Adverse Events During and After Hospital Admission*. Anesthesiology, December 2022, vol 137, 673-686. https://doi.org/10.1097/ALN.0000000000004380

Risk predictions are calculated by applying model coefficients to a set of variables derived from the billing claims history in the year prior to the date of hospital admission. The only information used from the day of admission is the admitting diagnosis (rather than the principal diagnosis) along with the principal procedure for planned admissions. The calculations also utilize information regarding patient sex, admission status (elective or not), and age at the time of admission.

Model Endpoints and Abbreviation:
(In-hospital duration is denoted '0000' days, and 90-day post-admission is '0090' days)

Period	Endpoint	Abbreviation (combines 4-digit mnemonic for Endpoint with 4-digit duration in days)
In-Hospital	Length of Stay	LOSX0000
	Mortality	MORX0000
	Discharge to Facility	IPDF0000
90-days Post Admission	Pneumonia	PNEX0090
	Acute Kidney Injury	AKIX0090
	Sepsis	SEPX0090
	Major Cardiovascular Complication	MCEX0090
	Respiratory Failure	REFX0090
	Mortality	MORX0090
	Unplanned Admission	IPAU0090

Variables

Medical history is represented by a set of binary variables indicating the presence (1) or absence (0) of individual and categories of ICD-10 diagnostic and procedure codes.[1] We used a custom procedure to reduce the domain of ICD-10 diagnostic codes to a

representative subset of 4,426 codes by collapsing rare codes into their parent codes to avoid overfitting. (Refer to Supplemental Figure 1.[2]) *Crosswalk_RSI30v1.xlsx* contains the crosswalk between individual ICD-10 diagnostic codes and the set of 4,426 representative codes. ICD-10 diagnostic codes were additionally represented by the diagnostic category as defined by their corresponding default Clinical Classifications Software Refined category.[3] Similarly, ICD-10 procedure codes were represented by their procedural category corresponding to the default Clinical Classifications Software category.[4]

Patient age (if included in the model) is represented by normalized age:

AGE_NORM = age/100.

Temporal information relative to a prediction date is encoded using three sets of these variables representing the presence or absence of relevant codes or categories in the past 90 or 365 days or on the day of admission (only for the principal diagnosis and principal procedure (for elective admissions.)) The variable prefix associated with each of these time intervals is H90 , H365 and IP0, respectively.

Variable Name Nomenclature:
Variable names incorporate both temporal information and code source type. Temporal information is indicated by H90, H365 and IP0 as described above. Code source information is indicated by prefixes D_, C_ and S_ for individual ICD10-CM diagnostic codes, CCSR category of a diagnostic code, and CCS category of a procedural code, respectively.

- Notation for Individual ICD-10 diagnostic codes excludes the decimal used in their formal definition (e.g., K59.09 Other constipation is denoted as K5909.)
- Notation for the CCSR diagnostic category of an ICD-10-CM diagnostic code uses a 3-letter abbreviation preceding a 3-digit numerical category subtype (e.g., DIG001, which is "intestinal infection").
- Notation for the CCS procedural category of an ICD-10-PCS procedural code uses a numerical category (e.g. 152, which is "knee arthroplasty").

Examples:

D_H365_K5909	Presence/absence of ICD-10-CM code K59.09 within 1yr prior to admission
C_H90_DIG001	Presence/absence of CCSR category associated with any of the ICD-10-CM diagnostics codes listed within the "intestinal infection" DIG001 category within 90 days prior to admission

S_IP0_152   Presence/absence of the "knee arthroplasty" CCS category (152) of the ICD-10-PCS Principal Procedure code associated with of the admission (if an elective case)

ICD-10-CM to Representative Diagnostic Code Crosswalk

Crosswalk_RSI30v1.xlsx contains the crosswalk between individual ICD-10 diagnostic codes and the set of 4,426 representative codes (i.e., "DCODEs") used in RSI 3.0 models. The file also contains the associated labels to identify the meaning of the ICD-10 and DCODEs. Note that if a given ICD-10-CM is not included in the crosswalk, it is not directly represented in any model; however, its corresponding CCSR diagnostic category may be in a model.

Coefficient Files:

Model coefficients are stored in Excel files associated with a given endpoint. Each file contains 4 sheets, one sheet of coefficients per model based on patient sex (male or female) and inpatient admission status (planned (IPP) or unplanned (IPU)).

The coefficient sets for a given endpoint and observation period are saved on sheets IPA_EEEEDDDD_S within Excel file IP_EEEEDDDD.xlsx , where :

A   is the admission type (planned (P) or unplanned (U))
EEEE   is the endpoint (see table above)
DDDD   is the number of observation days (where 0000 is in-hospital and 0090 for 90 days post-admission)
S   is patient sex (m for male, f for female).

For each endpoint, the user must apply the appropriate set of coefficients for the patient admission from among the 4 available sets available for the endpoint based on the sex of the patient (male or female) and admission status (planned (i.e., elective) or unplanned (non-elective.))

Prediction calculation:

The predicted outcome (P) for a given patient admission is obtained by applying the appropriate coefficient set to the binary clinical variables and the normalized age using the following expressions:

For Binary Endpoints (i.e., all endpoints except LOS):

$$P = \exp\left(-\exp\left(-(\beta_0 + \sum_{i=1}^{N} \beta_i x_i)\right)\right)$$

Where each of the *N* variables x_i identified in the coefficient file is multiplied by its corresponding **coefficient** β_i and summed with the **intercept** β_0 . This result is then transformed by the exponential functions (exp()) to calculate the probability of outcome.

For Length of Stay (LOS) endpoint:

$$LOS = \beta_0 + \sum_{i=1}^{N} \beta_i x_i$$

Where each of the *N* variables x_i is multiplied by its corresponding coefficient β_i and summed with the intercept β_0 . This result is predicted length of stay.

N.B. The columns "Parameter" and "Estimate" in the coefficient files contain the names of variables and their associated coefficient values. Note that different models have different numbers of variables.

REFERENCES

1.   CMS: International Classification of Diseases, Tenth Revision (ICD-10). Baltimore, MD, Centers for Medicare & Medicaid Services, 2021
2.   Greenwald S, Chamoun GF, Chamoun NG, Clain D, Hong Z, Jordan R, Manberg PJ, Maheshwari K, Sessler DI: Risk Stratification Index 3.0, a Broad Set of Models for Predicting Adverse Events During and After Hospital Admission. Anesthesiology 2022; 137: 673-686
3.   AHRQ: Clinical Classification Software Refined (CCSR), Agency for Healthcare Research and Quality, 2021
4.   AHRQ: Clinical Classification Software ICD-10-PCS (CCS), Agency for Healthcare Research and Quality, 2021