**Introduction**

Programs for Injury Categorization using diagnosis codes of the International Classification of Diseases, Version 9 Clinical Modification (ICD-9-CM) were originally developed using Stata statistical software (Statacorp, College Station, Texas). After the introduction of ICD-10-CM in 2015, an update to accommodate this change was developed using R statistical software (R Project, Vienna, Austria). The context for ICDPIC and ICDPIC-R, along with a general history of injury severity scoring, has been presented in a previous publication.1

The accompanying programs, ICDPIC-R-2021, are a further update in response to numerous inquiries and suggestions. The most important changes are as follows:

ICDPIC and the initial version of ICDPIC-R had been designed to use data coded with ICD-9-CM or ICD-10-CM (US Clinical Modification), which limited its value for international users.2 ICDPIC-R-2021 allows the user to specify whether data are in ICD-10-CM or basic ICD-10 format.

The default “ROCmax” option for calculating Abbreviated Injury Scores3 in ICDPIC-R was based upon mortality data in the American College of Surgeons (ACS) National Trauma Data Bank (NTDB), using an *ad hoc* algorithm to quantify the relative severity of each individual diagnosis code. The ROCmax option in ICDPIC-R-2021 allows the user to use either the ACS Trauma Quality Improvement Project (TQIP, the successor to NTDB) or the Health Care Utilization Project (HCUP) National Inpatient Survey (NIS) as the reference database. Furthermore, the original *ad hoc* algorithm has been replaced by the well-established methodology of ridge regression to estimate the independent effect of each injury diagnosis.

If the ROCmax option is chosen, a prediction of mortality for each subject (Pmort) is provided directly from the ridge regression, as well as the estimated Injury Severity Score4 (ISS). As in the earlier Stata version of ICDPIC, a “New Injury Severity Score”5 (NISS) is now also calculated for all options.

**Programs used to derive ROCmax reference table**

**icdaisA** – Reads in raw data from the 2017 TQIP Research Data Set or the 2016 NIS. Identifies cases with at least one injury diagnosis specified by an ICD-10-CM code, and classifies the diagnoses by body regions required for calculation of the Injury Severity Score (ISS).4  The National Trauma Data Standard used by TQIP considers valid ICD-10-CM injury codes to be those in the ranges S00-S99, T07, T14, T20-T28, and T30-T32, so **icdaisA** recognizes only these codes in the calculation of injury severity. ICDPIC-R also requires that injury codes conclude with the letter “A” (indicating an initial encounter), except for codes indicating a fracture, where codes concluding with the letters “B” or “C” indicate an initial encounter with an open fracture. ICD-10-CM codes that explicitly state that the subject lived or died (S06.\*\*6A, S06.\*\*7A, or S06.\*\*8A) are converted to S06.\*\*9A, which does not specify the outcome. **icdaisA** also creates an additional data set for each data source by truncating ICD-10-CM codes to the underlying basic ICD-10 code (format \*\*\*.\*) and removes basic ICD-10 codes that are duplicated within an individual subject as a result. Prepares data for regression analysis.

**icdaisB** – Reads in each of the data sets prepared by **icdaisA**, transforms them into matrices, and performs logistic ridge regression using R package **glmnet**, which is described in detail in the documentation for that package. For each reference dataset (TQIP or NIS) and each format (ICD-10-CM or basic ICD-10), the logistic ridge regression results in an independent estimate of effect (log odds ratio) for each diagnosis code. These are tabulated and can be combined with the estimated model intercept to estimate the probability of mortality for individual subjects. **icdaisB** also determines the largest effect estimate in each body region for each subject, which will subsequently be stratified into Abbreviated Injury Scores (AIS)3 in order to estimate ISS.

**icdaisC** – Reads in the tabulated effect estimates for each diagnosis code produced by **icdaisB**. For each reference dataset and format, initializes cutpoints categorizing the effect estimates into AIS scores of 1, 2, 3, 4, or 5, and calculates the resulting ISS and NISS.4,5 Uses a “greedy algorithm” to determine the cutpoints for which the c-statistic (area under a Receiver Operator Characteristic curve) for ISS to predict mortality is maximized. For each diagnosis, reference dataset, and format, tabulates the optimal AIS estimates along with the effect estimates and intercepts from ridge regression. These are summarized in the tables **TQIP\_NIS\_ais\_cm.csv** and **TQIP\_NIS\_ais\_base.csv**.

**cat\_trauma** – Reads in user data in the specified format and, depending upon the options selected by the user, either the table **TQIP\_NIS\_ais\_cm.csv** or the table **TQIP\_NIS\_ais\_base.csv**. If the user has specified option “ROCmax”, calculates ISS and NISSfrom the data in these tables. If the user has specified option “GEMmax” or “GEMmin”, converts ICD-10-CM codes to ICD-9-CM codes using a General Equivalence Mapping table published by the US CMS, and then calculates ISS from the table previously used in the Stata version of ICDPIC. Also categorizes ICD-10 codes that specify injury mechanism according to a table published by the US CDC.6

**Comparison of methods using test data**

Four datasets were used to compare the results of different options for cat\_trauma:

A 20% sample of 2017 TQIP data coded using ICD-10-CM

A 20% sample of 2017 TQIP data coded using basic ICD-10

A 20% sample of 2016 NIS data coded using ICD-10-CM

A 20% sample of 2016 NIS data coded using basic ICD-10.

Each of the sample data sets was processed using all relevant options for procedure cat\_trauma. C-statistics (areas under a Receiver-Operator Characteristic curve) were calculated for the prediction of mortality by the approximated ISS (designated RISS) and the approximated NISS. If the ROCmax method was used, a c-statistic was also calculated for the mortality prediction obtained directly from ridge regression (Pmort). For TQIP data, a c-statistic was also calculated for the ISS that had been reported by trauma registrars (designated TISS). The results of these determinations are shown in Table 1.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| TABLE 1 | | | | |
|  | C-statistics for prediction of mortality | | | |
|  | TISS | RISS | NISS | Pmort |
|  |  |  |  |  |
| TQIP-CM |  |  |  |  |
| ROCmax (TQIP) | .840 | .856 | .861 | .886 |
| ROCmax (NIS) | .840 | .813 | .823 | .800 |
| GEMmax | .840 | .760 | .774 | - |
| GEMmin | .840 | .765 | .775 | - |
|  |  |  |  |  |
| TQIP-Base |  |  |  |  |
| ROCmax (TQIP) | .840 | .842 | .847 | .864 |
| ROCmax (NIS) | .840 | .815 | .825 | .806 |
|  |  |  |  |  |
| NIS-CM |  |  |  |  |
| ROCmax (TQIP) | - | .712 | .710 | .747 |
| ROCmax (NIS) | - | .757 | .755 | .815 |
| GEMmax | - | .665 | .668 | - |
| GEMmin | - | .673 | .671 | - |
|  |  |  |  |  |
| NIS-Base |  |  |  |  |
| ROCmax (TQIP) | - | .718 | .717 | .746 |
| ROCmax (NIS) | - | .739 | .739 | .774 |
|  |  |  |  |  |

For each data source and method, ISS was also categorized as recommended by Copes *et al*.,7 and the mortality within each category for different data sources and options was tabulated. For TQIP data, the percentage of cases for which RISS and TISS were in the same or an adjacent category was also tabulated. These results are shown in Table 2.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| TABLE 2 | | | | | | | | |
|  | Mortality for ISS category (%) | | | | | | | Category  near TISS |
|  | 1-8 | 9-15 | 16-24 | 25-40 | 41-49 | 50-75 | Unk |
|  |  |  |  |  |  |  |  |  |
| TQIP-CM |  |  |  |  |  |  |  |  |
| ROCmax (TQIP) | 0.74 | 2.60 | 7.55 | 27.0 | 48.5 | 64.7 | 0 | 93.4% |
| ROCmax (NIS) | 0.77 | 1.46 | 2.69 | 10.1 | 18.2 | 29.9 | 0 | 84.9% |
| GEMmax | 0.94 | 2.43 | 5.63 | 10.6 | 18.4 | 48.0 | 1.79 | 90.5% |
| GEMmin | 1.01 | 2.66 | 8.61 | 17.7 | 26.6 | 40.5 | 1.79 | 93.9% |
| TISS | 0.68 | 1.83 | 5.62 | 24.9 | 37.0 | 60.2 | 5.47 | - |
|  |  |  |  |  |  |  |  |  |
| TQIP-Base |  |  |  |  |  |  |  |  |
| ROCmax (TQIP) | 0.83 | 2.46 | 7.83 | 25.5 | 42.3 | 60.2 | 0 | 93.6% |
| ROCmax (NIS) | 0.76 | 1.50 | 2.08 | 9.1 | 17.7 | 32.2 | 0 | 83.7% |
| TISS | 0.68 | 1.83 | 5.62 | 24.9 | 37.0 | 60.2 | 5.47 | - |
|  |  |  |  |  |  |  |  |  |
| NIS-CM |  |  |  |  |  |  |  |  |
| ROCmax (TQIP) | 1.50 | 3.43 | 8.59 | 16.5 | 29.0 | 33.3 | 0 | - |
| ROCmax (NIS) | 1.07 | 2.22 | 3.92 | 11.3 | 19.4 | 30.8 | 0 | - |
| GEMmax | 1.51 | 2.52 | 6.27 | 5.7 | 10.9 | 17.1 | 0.70 | - |
| GEMmin | 1.62 | 2.57 | 8.25 | 11.0 | 16.7 | 10.7 | 0.70 | - |
|  |  |  |  |  |  |  |  |  |
| NIS-Base |  |  |  |  |  |  |  |  |
| ROCmax (TQIP) | 1.45 | 3.24 | 7.44 | 14.5 | 21.4 | 29.6 | 0 | - |
| ROCmax (NIS) | 1.14 | 2.38 | 3.36 | 8.2 | 14.2 | 22.7 | 0 | - |
|  |  |  |  |  |  |  |  |  |

**Suggested options for different types of data**

The procedure cat\_trauma, which calculates ISS, NISS, and Pmort in ICDPIC-R, will not run unless specific options have been selected. Default values are not provided because, in view of the above findings, results may differ significantly depending upon the kind of data being processed. Some guidelines for which options to specify are given below. Ultimately, the validation of ICDPIC-R will depend upon its performance using other independent data; one of the first of these is the study of Sebastião *et al*,8 who found that the GEMmin option seemed to function better than the GEMmax option for data coded with a mix of ICD-9-CM and ICD-10-CM. Experience from other countries will be of particular interest to see whether TQIP or NIS is a better reference database, and it may vary from one setting to another.

Given the results so far, ICDPIC-R-2021 should function best for the following types of data with the given options:

Data from US trauma registries coded using both ICD-9-CM and ICD-10-CM

icd10=”TRUE”, i10\_iss\_method=”gem\_min”

Data from US trauma registries coded using only ICD-10-CM

icd10=”cm”, i10\_iss\_method=”roc\_max\_TQIP”

Data from US trauma registries coded using both ICD-9-CM and ICD-10-CM

icd10=”TRUE”, i10\_iss\_method=”gem\_min”

Data from US administrative sources coded using only ICD-10-CM

icd10=”cm”, i10\_iss\_method=”roc\_max\_NIS”

Data from non-US sources coded using basic ICD-10

icd10=”base”, i10\_iss\_method=”roc\_max\_TQIP”

OR

icd10=”base”, i10\_iss\_method=”roc\_max\_NIS”

**References**

1. Clark DE, Black AW, Skavdahl DH, Hallagan LD. Open-access programs for injury categorization using ICD-9 or ICD-10. *Injury Epidemiology* 2018; 5:11.
2. Airaksinen NK, Heinänen MT, Handolin LE. The reliability of the ICD-AIS map in identifying serious road traffic injuries from the Helsinki Trauma Registry. *Injury* 2019; 50:1545-1551.
3. Committee on Medical Aspects of Automotive Safety, AMA. Rating the severity of tissue damage. I. The abbreviated scale. *JAMA* 1971; 215:277-280.
4. Baker SP, O’Neill B, Haddon W Jr., Long WB. The injury severity score: A method for describing patients with multiple injuries and evaluating emergency care. *J Trauma* 1974; 14:187-196.
5. Osler T, Baker SP, Long WA. Modification of the injury severity score that both improves accuracy and simplifies scoring. *J Trauma* 1997; 43:922-925.
6. Annest JL, Hedegaard H, Chen L, Warner M, Smalls E. Proposed framework for presenting injury data using ICD-10-CM external cause of injury codes. 2014. <http://www.cdc.gov/injury/wisqars/pdf/ICD-10-CM_External_Cause_Injury_Codes-a.pdf>.
7. Copes WS, Champion HR, Sacco WJ, Lawnick MM, Keast SL, Bain LW. The injury severity score revisited. *J Trauma* 1988; 28:69-77.
8. Sebastião YV, Metzger GA, Chisolm DJ, Xiang H, Cooper JN. Impact of ICD-9-CM to ICD-10-CM coding transition on trauma hospitalization trends among young adults in 12 states. *Inj Epidemiol* 2021, in press.