ICD-9 AND ICD-10 GENERAL EQUIVALENCE MAPPINGS (GEMS)

STATISTICAL ANALYSIS SYSTEM (SAS) PROGRAM SPECIFICATION

Prepared by the Sentinel Operations Center
June 19, 2018
Version: 1.0.0

The Sentinel System is sponsored by the U.S. Food and Drug Administration (FDA) to proactively monitor the safety of FDA-regulated medical products and complements other existing FDA safety surveillance capabilities. The Sentinel System is one piece of FDA's Sentinel Initiative, a long-term, multi-faceted effort to develop a national electronic system. Sentinel Collaborators include Data and Academic Partners that provide access to healthcare data and ongoing scientific, technical, methodological, and organizational expertise. The Sentinel Coordinating Center is funded by the FDA through the Department of Health and Human Services (HHS) Contract number HHSF223201400030I. This project was funded by the FDA through HHS Mini-Sentinel contract number HHSF223200910006I.

Table of Contents

Modification History

Version	Date	Modification	By
1.0.0	06/19/2018	• First published version	Sentinel Operations Center

I. INTRODUCTION

The United States mandated the transition from the 9th revision of the International Classification of Disease (ICD-9) to the 10th revision (ICD-10) on October 1, 2015. To conduct healthcare research and public health surveillance across these two different eras, translation from one coding system to the other is required. The purpose of this SAS program is to map ICD-9 diagnosis or procedure codes to their exact or approximate ICD-10 codes using the Center for Medicare and Medicaid Services' (CMS) General Equivalence Mappings (GEMs).

To facilitate the translation of codes between the two codes systems, CMS has developed two GEMs: forward mapping and backward mapping. The forward GEMs translate ICD-9 codes to ICD-10 codes, and the backward GEMs translate ICD-10 codes to ICD-9 codes. On occasion, the relationship between the two coding systems can be a one-to-many translation or non-existent. Therefore, different applications of GEMs can lead to different translations. However previous work has identified the forward-backwards GEMs has evidence of the highest precision, recall, and F-score metrics.[1]

The purpose of this program is to import GEMs forward and backwards files, downloaded directly from the CMS website, and create output SAS datasets that can then be used to translate ICD-9 codes to approximate or exact ICD-10 codes.

II. REQUIREMENTS

To run this program, users must:

1) Select, download, and extract the GEMS files for a specific fiscal year (FY) from the CMS website at: https://www.cms.gov/Medicare/Coding/ICD10/index.html. For more detailed instruction on downloading these files, please see Appendix A, Instruction for Downloading CMS files.
2) Provide an input Excel worksheet that includes a list of ICD-9 codes to be translated, described in Appendix B, ICD-9 Codes of Interest List (excel sheet).

III. FUNCTIONALITY

This program allows users flexibility in:

1) Creating translation mappings for ICD-9 diagnosis or procedure codes.
2) Selecting a specific FY for translation via selection of the GEMS files from the CMS website.
3) Specifying a limited list of ICD-9 codes for translation.

IV. USER PARAMETERS

The following user-specified parameters will drive the program's functionality:

Parameter Name	Type (Char or Num)/Length	Valid Values	Notes/Description	Example
GEMS_FY	Char(4)	CCYY	Required field representing the fiscal year of the GEMS downloaded from CMS to be processed	GEMS_FY = 2008
indsn_subsetID9	Char(Varies)	Input Excel filename including file extension	Optional input filename for dataset that lists subset of ICD-9 codes to be translated	indsn_subsetID9 = subseti9.xlsx
invar_subsetID9	Char(Varies)	Valid SAS variable name	Optional parameter, required if indsn_subsetID9 populated, that is the variable name of field that represents ICD-9 codes to include	invar_subsetID9 = I9
indsn_gemsID9	Char(Varies)	Input filename, including .txt file extension	Required parameter, input forward mapping GEMS filename. Value must include .txt file extension.	indsn_gemsID9 = 2018_I9gem.txt
indsn_gemsID10	Char(Varies)	Input filename, including .txt file extension	Required parameter, input backward mapping GEMS filename. Value must include .txt file extension.	indsn_gemsID10 = 2018_I10gem.txt
outdsn_gemsF	Char(Varies)	Valid SAS filename, not including file extension	Required parameter, output forward mapping file dataset name	outdsn_gemsF = gemsF_&GEMS_FY.

Parameter Name	Type (Char or Num)/Length	Valid Values	Notes/Description	Example
outdsn_gemsB	Char(Varies)	Valid SAS filename, not including file extension	Required parameter, output backward mapping file dataset name	outdsn_gemsB = gemsB_&GEMS_FY.
outdsn_gemsFB	Char(Varies)	Valid SAS filename, not including file extension	Required parameter, output forward and backward mapping file dataset name	outdsn_gemsFB = gemsFB_&GEMS_FY.
outdsn_ID10F	Char(Varies)	Valid SAS filename, not including file extension	Required parameter, output forward distinct ICD-10 list file dataset name	outdsn_ID10F = gemsID10F_&GEMS_FY.
outdsn_ID10B	Char(Varies)	Valid SAS filename, not including file extension	Required parameter, output backward distinct ICD-10 list file dataset name	outdsn_ID10B = gemsID10B_&GEMS_FY.
outdsn_ID10FB	Char(Varies)	Valid SAS filename, not including file extension	Required parameter, output forward and backward distinct ICD-10 list file dataset name	outdsn_ID10FB = gemsID10FB_&GEMS_FY.
outdsn_descstats	Char(Varies)	Valid SAS filename, not including file extension	Required parameter, output descriptive statistics file dataset name	outdsn_descstats = gemsStats_&GEMS_FY.
inloc_i9xls	Char(Varies)	Valid file path including end of path separator	Optional parameter, required if indsn_subsetID9 populated, that provides location of file represented by parameter indsn_subsetID9. Value must include end of path separator.	inloc_i9xls= C:\dev\test\input\
inloc_gems	Char(Varies)	Valid file path including end of path separator	Required parameter, location of input files	inloc_gems = C:\dev\test\input\

Parameter Name	Type (Char or Num)/Length	Valid Values	Notes/Description	Example
			represented by parameters indsn_gemsID9 and indsn_gemsID10. Value must include end of path separator.	
outloc_map	Char(Varies)	Valid file path including end of path separator	Required parameter, location of output files to be created, represented by parameters outdsn_gemsF, outdsn_gemsB, outdsn_gemsFB, outdsn_ID10F, outdsn_ID10B, outdsn_ID10FB. Value must include end of path separator.	outloc_map = C:\dev\test\output\

V. OUTPUT

A. OUTPUT MAPPING FILES

The mapping files contains variables required to translate ICD-9 codes to ICD-10 codes and a flag variable to represent the type of translation. One file will be output by translation type, that is for forward, backward, and forward-backward mappings. The names of these output datasets will be specified by the User via parameter in the SAS program.

Variable Name	Type (Char or Num)/Length	Format	Valid Values	Notes/Description
ID9	Char(8)	$8.	ICD-9 diagnosis or procedure code	ICD-9 Code(s) of interest, decimal point removed.
ID10	Char(8)	$8.	ICD-10 diagnosis or procedure code	Mapped ICD-10 Code(s), punctuation removed.
FLAG	Char(8)	$8.	00000 - identical translation 10000 - approximate translation 11000 - no mapping Other – requires cluster of codes	This variable describes the crosswalk between the ICD-9 and ICD-10 code.

**Full explanation of clusters is beyond the discussion of this document. Users should consult the CMS webpage (https://www.cms.gov/Medicare/Coding/ICD10/2018-ICD-10-CM-and-GEMs.html).

B. DISTINCT ICD-10 LISTING FILES

The distinct ICD-10 file will list all valid ICD-10 code values represented in each of the output mapping file. As a result, three files will be generated with each run of the SAS program to represent ICD-10 codes based on the forward, backward, and forward-backward mappings. The names of these output datasets will be specified by the user via parameter in the SAS program.

Variable Name	Type (Char or Num)/Length	Format	Valid Values	Notes/Description
ID10	Char(8)	$8.	ICD-10 diagnosis or procedure code	ICD-10 codes ascertained from the forward, backward, or forward-backward mapping algorithms

C. DESCRIPTIVE STATISTICS FILE

A single descriptive dataset will be output with each run of the SAS program. This dataset will quantify the number of records, distinct ICD-9 code counts, and distinct ICD-10 code counts included in the forward, backward, and forward-backward translations. The names of this output dataset will be specified by the user via parameter in the SAS program.

Variable Name	Type (Char or Num)/Length	Format	Valid Values	Notes/Description
GEMS_Type	Char(30)	$30.	GEMS FY CCYY, forward map GEMS FY CCYY, backward map GEMS FY CCYY, combined map	Value of CCYY will represent the fiscal year of the GEMS included in translation
MapCount	Num(8)	8.	>= 1	Count of records in mapping
NoMatchICD-9 Count	Num(8)	8.	. - N/A, ICD-9 look-up not provided >= 0	Count of distinct ICD-9 codes supplied in lookup not included in mapping. Note value will be missing if ICD-9 look-up not provided by user.
DistID9Count	Num(8)	8.	>= 1	Count of distinct ICD-9 codes in mapping
DistID10Count	Num(8)	8.	>= 1	Count of distinct ICD-10 codes in mapping

VI. METHODS

The following logic will be executed by this program:

1) If user-specified an import Excel file that contains a list of ICD-9 codes for translation and load these values into a parameter to limit records in subsequent steps.

2) Separately import forward and backward GEMS text files.
    a. If ICD-9 filtering parameter exists, limit result extracts to records where ICD-9 code corresponds to a value included in parameter.
    b. Deduplicate all GEMS files (remove absolute duplicates across rows of data).
    c. Output forward mapping extract to create final dataset.
        i. Ensure file matches data dictionary
        ii. Save to permanent location, using user-specified parameter for naming file.
    d. Output backward mapping extract to create final dataset.
        i. Ensure file matches data dictionary
        ii. Save to permanent location, using user-specified parameter for naming file.
    e. Union join forward and backward mapping files to the combined mapping final dataset.
        i. Ensure file matches data dictionary
        ii. Save to permanent location, using user-specified parameter for naming file.

3) Identify distinct ICD-10 codes per forward, backward, and forward-backward mapping files to create three final output datasets, one per each type of mapping, as described below.
    a. Select distinct values of ICD-10 from final mapping dataset created in previous step.
        i. Ensure file matches data dictionary
        ii. Save to permanent vacation, using user-specified parameter for naming file.

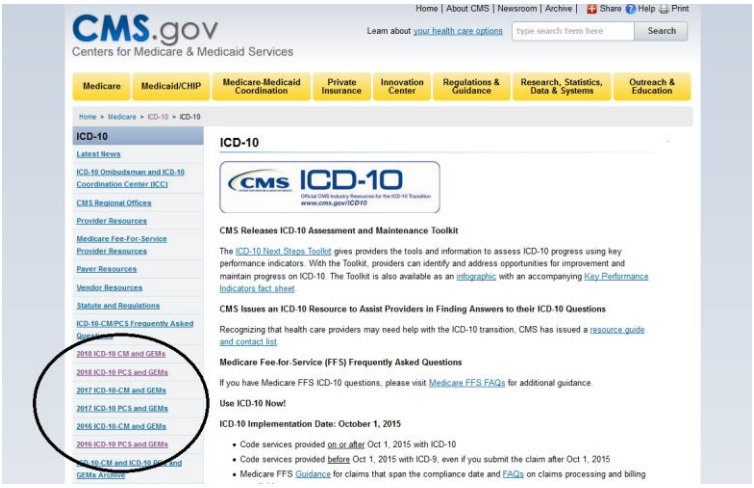
4) Process forward, backward, and forward-backward mapping files to create final, output descriptive dataset as follows:
    a. Per each mapping file, create variables defined in the descriptive file format above.
    b. Combine resulting extracts to create a single file that includes descriptive statistics for all mapping files, that is one row for forward mappings, one row for backward mappings, and another row for the combined mapping.
    c. Ensure file matches data dictionary.
    d. Save to permanent vacation, using user-specified parameter for naming file.

VII. APPENDIX

A. INSTRUCTION FOR DOWNLOADING CMS GEMS FILES

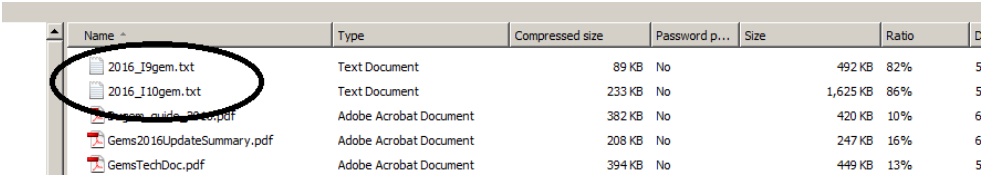
These are instructions for downloading the GEMS files from the CMS website.

1) Navigate to CMS website https://www.cms.gov/Medicare/Coding/ICD10/index.html
2) On website, identify and select link to GEMS file by desired fiscal year



3) Select link to GEMS files from CMS website to download files to local computer

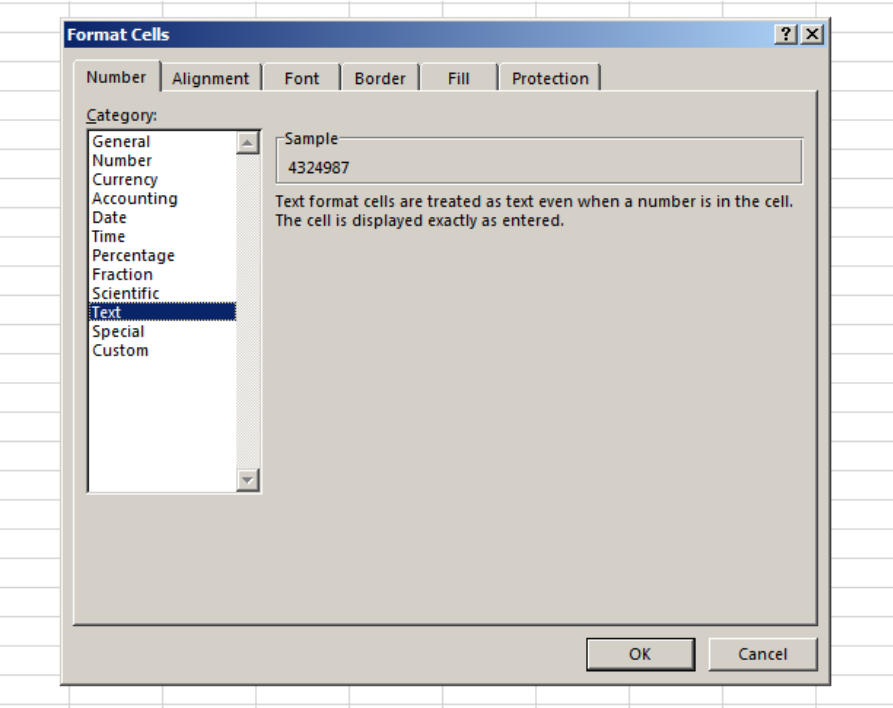
4) Locate local copies of GEMS files, move to desired location and be sure this location and these filenames are referenced in the SAS program. Please note that the forward GEMS mapping filename is ICD-9 Gem.txt and the backward GEMS mapping filename is I10Gem.txt



B. ICD-9 CODES OF INTEREST LIST (EXCEL SHEET)

This is the file format of the input excel sheet of ICD-9 codes that require translation to ICD-10. This file must be an Excel document that includes the following column to list valid ICD-9 codes. Note that the decimal point must be removed from the included ICD-9 codes values. Also note that the field must be character type for the program to process this data without error.

To convert a numeric value to character in Excel, a user would need to format cells by selecting Numbers Category Text per screen shot below.



Variable Name	Type (Char or Num)/Length	Format	Valid Values	Notes/Description
ICD9	Char(8)	$8.	Valid ICD-9 procedure or diagnosis code	ICD-9 Code(s) of interest

Example:

ICD-9
630
631
632
633
63300

VIII. REFERENCES

1. Fung KW, Richesson R, Smerek M, et al. Preparing for the ICD-10-CM Transition: Automated Methods for Translating ICD Codes in Clinical Phenotype Definitions. EGEMS (Washington, DC) 2016; **4**(1):1211 doi: 10.13063/2327-9214.1211[published Online First: Epub Date]|.