Paper 181-2012

Cluster Mapping of Medicare Severity-Diagnosis Related Groups for ICD-9-CM to ICD-10-CM Conversions Using SAS®

Keith A. Jones, Ph.D., Sr. Consultant, Qualimatix Consulting, Dallas, Texas, USA

ABSTRACT

Diagnosis Related Groups (DRG's) have been critical to Medicare procedure-based billing, prospective payment and risk management since it was first implemented. More recently, Medicare developed 'Severity-DRG's (MS-DRG) to weight top factors to optimize cost, time, and clinical outcomes; and new rules for US Medicare claims processing. Now US Medicare provides Generalized Equivalence Map (GEM) clusters to help convert ICD-9-CM to ICD-10-CM in 2013, and better optimize billings in the future. Such mappings can be used by insurers to optimize clinical outcomes and costs, using SAS/STAT® procedures and SAS® Enterprise Miner™, with reporting process automation by SAS Enterprise Guide, and visualization using SAS/GRAPH® and with SAS/JMP®, to build and cluster clinical diagnosis and procedures for cost classification models. Examples are presented to support scaled SAS solutions to better comply with, and benefit from, these new Medicare resources. These examples use Centers for Medicare Services (CMS) GEM files, and Public Use File (PUF) data, which can be applied across all the diverse SAS system platforms.

INTRODUCTION

The origins of Medicare procedure-based billing and prospective payment risk management capitation go back to the mid 1970's to early 1980's, when many well-intentioned bipartisan efforts of US Congress built upon Social Security to also provide affordable health insurance to all US citizens, in a manner sustainable as an entitlement to all US citizens that paid into the Social Security fund. This has since been expected as an entitlement by all US citizens. Despite many challenges, it can be argued no other nation than US has better managed its health benefit systems.

From the beginning of Medicare and Medicaid, it was difficult to fund and sustain, for very many reasons historically documented; and are now more relevant, as similar factors and issues continue today. We currently face expansions of past coverage to even more patients by government mandates, with many new assumptions of payer responsibility pending multiple scenarios of outcomes of these mandates; within a wide range of possible conditional events dependent upon political factors for the next several years. Meanwhile, it is most appropriate that both private and public healthcare organizations anticipate a range of scenarios, equally to expect both extremes of private and public healthcare mandates, as potential outcomes for funding reflect a wide range of scenarios based on future economics.

Central to this are emerging opportunities for both private and public organizations, so we must prepare for all scenarios due to many initiatives and improvements in diagnostic related groupings and procedure based billing standards by US government, especially Centers for Medicare and Medicaid Services and Centers for Disease Control; as well as major US health insurer corporations and industry organizations. Most past initiatives have involved the use of SAS, published SAS source code, and SAS software toolsets; some of which are presented here.

In particular, this paper will focus on SAS related code and software that can be applied by private corporations and insurers to automate and optimize their operations and bottom lines, using SAS language and SAS products to both improve upon profitability and quality of service to healthcare consumers. This includes examples of SAS products and procedures that can improve quality of service in clinical outcomes, as well as reduce costs for consumers and insurers, and to optimize government operations and lower dependence on US tax payers, by using SAS software.

But first it is necessary to understand the concept of Medicare General Equivalence Map (GEM) qualitative methods versus quantitative data clustering, as it is more commonly regarded by SAS statisticians; and also understand how SAS software can help optimize GEM's, to both improve profitability and ROI's of insurer organizations, and reduce Medicare costs to assure future sustainability. The detailed specifications of Medicare GEM files and requirements for their use with ICD-10-CM in future claims processing are not presented here, but are provided on the CMS website and official Medicare publications, several of which are listed in the bibliography at the end of this paper.

We demonstrate here a simple and practical basic process flow which can be implemented using SAS software and tool-sets that apply the emerging new Medicare ICD-10-CM and Medicare Severity-DRG weights, for both common objectives; to optimize current and future government Medicare costs, and to also optimize the profitability of private co-insurance for Medicare populations. But this requires that we must first understand current and past Medicare cost management codes, and then how SAS statistical platforms can best leverage these new Medicare rules. This includes SAS models and analytics for national Medicare ROI, as well as insurer's Per Member Per Month (PMPM).

MEDICARE STANDARDIZED DIAGNOSIS AND PROCEDURE BILLING CODES

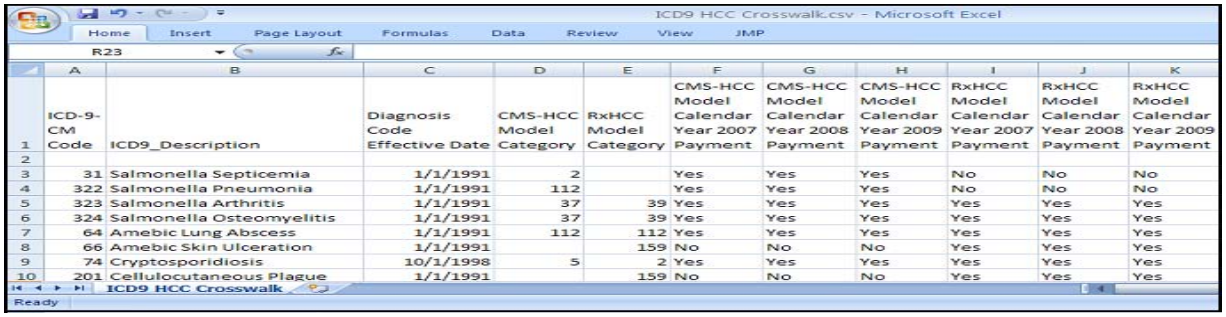
The Centers for Medicare and Medicaid Services (CMS) has mandated that all Medicare and Medicaid billing claims must be converted from ICD-9-CM diagnostic codes to ICD-10-CM no later than October 1, 2013. This is a major undertaking, to change databases and ETL processes alone; however, there are equally as many (or more), complicated processes to be developed and deployed in order to support alignments and rules for stored processes to involve SAS (and other platform) forecasts, models and reporting involving historical data across this timeline, and to maintain statistical precision we must accommodate for adjustment to ROI and forecasts based on both code-sets.

INTERNATIONAL CLASSIFICATION OF DISEASE (ICD) CODES

The International Classification of Diseases was first published by American Public Health Association in 1898, and was maintained by the League of Nations Health Organization (LNHO) from 1920 to 1948. Then it was maintained by World Health Organization (WHO) of the United Nations as versions ICD-6 to ICD-9. In 1978, the US National Centers for Health Statistics (NCHS) published its first expanded version, ICD-9-CM, for use in US Medicare and Medicaid billing. Since then US Centers for Medicare and Medicaid Services (CMS) publishes annual modifications, and in 1983 work began on ICD-10. In 2008, CMS published its first official version of ICD-10-CM, and mandated all Medicare and Medicaid claims to be submitted with ICD-10-CM to fully replace all ICD-9-CM codes by October, 2013. This is a formidable undertaking, as there is no one-to-one correspondence, or consistent one-to-many, or even many-to-one, correspondence between these two sets of codes. Due to lack of federal funds, conversions have been slow; and there have been almost monthly adds and deletes, in both sets of codes. Yet the October, 2013 mandate by CMS is still in effect; despite sweeping new US National Healthcare regulations are scheduled for release in 2014 (and are likely to be influenced by outcome of November, 2012 US national elections). Also, it is also worth noting the UN and WHO have announced they will release another very different new version of ICD in 2015, which will be known as 'ICD-11'. Since we may need to accommodate all sets of ICD codes in the future, it is critical that health care providers, insurers and pharmaceutical organizations are equipped with flexible analytic tools such as SAS.

ICD-9-CM CODES AND MEDICARE CROSSWALK MAPPING FILES

Because ICD-9-CM has been used since 1978 when Medicare was first implemented, much of its maintenance, including IT and data management of updates, has evolved into routine and fairly simple standards and processes. This has involved annual CMS updates with adds and deletes to the current accepted ICD-9-CM diagnosis codes, in what are known as Medicare "Crosswalk" mapping files. This has been fairly easy and economical for IT to maintain, as it basically allows providers to 'look-up' an ICD-9-CM diagnosis code; and if it is allowable in current year, simply submit it(and if not and same claim was submitted previous year, look-up equivalent codes accepted in current year). Medicare Crosswalk mapping files were (and still are) popular with providers. But the US government (and healthcare insurers) are increasingly uncomfortable with the Crosswalk mappings, as they make it easier to file and get paid for complex claims; yet can take advantage of insurers and government, and increase costs to patients and taxpayers.



	A	B	C	D	E	F	G	H	I	J	K
1	ICD-9-CM Code	ICD9_Description	Diagnosis Code Effective Date	CMS-HCC Model Category	RxHCC Model Category	CMS-HCC Model Calendar Year 2007 Payment	CMS-HCC Model Calendar Year 2008 Payment	CMS-HCC Model Calendar Year 2009 Payment	RxHCC Model Calendar Year 2007 Payment	RxHCC Model Calendar Year 2008 Payment	RxHCC Model Calendar Year 2009 Payment
2											
3	31	Salmonella Septicemia	1/1/1991	2		Yes	Yes	Yes	No	No	No
4	322	Salmonella Pneumonia	1/1/1991	112		Yes	Yes	Yes	No	No	No
5	323	Salmonella Arthritis	1/1/1991	37	39	Yes	Yes	Yes	Yes	Yes	Yes
6	324	Salmonella Osteomyelitis	1/1/1991	37	39	Yes	Yes	Yes	Yes	Yes	Yes
7	64	Amebic Lung Abscess	1/1/1991	112	112	Yes	Yes	Yes	Yes	Yes	Yes
8	66	Amebic Skin Ulceration	1/1/1991		159	No	No	No	Yes	Yes	Yes
9	74	Cryptosporidiosis	10/1/1998	5	2	Yes	Yes	Yes	Yes	Yes	Yes
10	201	Cellulocutaneous Plague	1/1/1991		159	No	No	No	Yes	Yes	Yes

Display 1. Example Medicare ICD-9-CM HCC Diagnostic Procedure Rx Billing MS Excel .CSV Crosswalk File

ICD-10-CM CODES AND MEDICARE GENERALIZED EQUIVALENCE MAPPING (GEM) FILES

There are no 'Crosswalk' code mappings for ICD-10-CM (and probably never will be), as it was designed with many practical objectives; but notably two of the most important are: 1) enough specificity so that a single diagnosis code, rather than up to as many as ten, can confirm the clinical procedure and medications prescribed in a claim, for more rapid processing, with quality and cost improvements for Medicare; 2) clinical outcomes and costs can be more efficiently analyzed by emerging IT and analytics technology (such as the SAS analytics platforms), for better financial management of clinical procedures and prescription medications and clinical management for greater benefits from healthcare insurance, in a manner that is flexible and easily updated with new diagnostic codes. Several commercial ICD-10-CM to MS-DRG products, as well as CMS Code Editor, can flag possible problems, but ultimate responsibility for code assignments to meet current ICD-10-CM billing is now fully upon the claimant.

Because ICD-9-CM codes have been used for US medical claims since Medicare was first implemented (over 30 years ago), it is understandable that a completely different set of diagnostic codes are a difficult pill to swallow, both for US healthcare industry and US citizens. However, since ICD-9-CM is over 30 years old, with increasing costs to maintain and validate, it has been argued ICD-10-CM is long overdue. It can thus be understood the reason for far more codes in ICD-10-CM is at least partly due to fact that, in the past 30 years, we have identified new diseases, new syndromes, new complications and co-morbidities, so complexities of diagnoses have more than doubled.

But more importantly, as this has all transpired while we have been still using ICD-9-CM, it has been necessary to submit claims with not just one ICD-9-CM primary diagnosis code, but often many more secondary diagnosis codes. The obvious inefficiency, for clinical as well as financial risk (since reality is most claims payment as well as patient healthcare are often based on only the first primary diagnosis code), has led to a realization that what is needed is a single more specific set of codes (which the basic objective of ICD-10-CM). But until ICD-10-CM is fully implemented, providers and insurers may need to support at least 2, or potentially even 3, parallel sets of ICD codes. Meanwhile, SAS platforms can help prepare for any contingency, for rapid deployments to support new analytic solutions.

MEDICARE-SEVERITY DIAGNOSIS RELATED GROUPING (DRG) CODINGS

The concept of Diagnosis Related Groupings (DRG's) was first developed at Yale in 1970's and implemented in 1980 by US Health Care Finance Administration (HCFA) to support early Medicare's Prospective Payment System (PPS). HCFA was since reorganized as US Centers for Medicare and Medicaid (CMS), and has been the US government foundation for all Medicare and Medicaid healthcare claim cost containment programs and regulations since.

CMS also developed and publishes a series of 'groupings of groupings' of 15,000 ICD codes into 800 DRG's; then group DRG's into under 200 Hierarchical Condition Categories (HCC's) for Medicare age clinical case type analysis. Many private insurers have further stratified these groupings using SAS cluster and principal component analysis for their own ROI profitability. In 2008, Medicare released MS-DRG's which include weights for 5 top cost and clinical outcome factors to additionally weight DRG's. These weights can be very valuable for insurers to build on their ROI and related cost estimation forecasting to apply Per Member Per Month (PMPM) projections for business planning.

CLINICAL PROCEDURAL TERMINOLOGY (CPT) CODES

CPT codes are maintained by the American Medical Association (AMA). New editions are released each year. CPT coding is similar to ICD coding, except that it identifies the services rendered rather than the diagnosis on the claim. The Healthcare Common Procedure Coding System is based on the AMA's Current Procedural Terminology (CPT). These are the primary clinical procedures accepted for Medicare claims. In addition, there are specific and generic Rx medical prescription codes, which have been additionally problematic to Medicare and health insurers, since in the past decade generics are often prescribed by providers and billed at low cost outside of the records of insurers or Medicare (so clinical outcomes and cost analysis by Medicare and insurers must factor to account for data gaps).

MEDICARE CPT BILLING AND NOT-OTHERWISE-CLASSIFIED (NOC) RX CODES

Although CPT codes and Rx codes for prescribed medications were fairly reliable to support modeling of clinical outcomes and costs up until wide use of generics, it is more complicated since an increasing number of patients are using OTC, alternative medications, and vitamins that are not included in patient or provider records. This makes clinical outcome analytics even more complicated, since the use of Rx inputs must be down-weighted based on some kind of capture ratio or co-variant outcome variables to account for gradual or sudden loss of this data dimension.

MEDICARE SEVERITY-DIAGNOSIS RELATED GROUP (MS-DRG) CODE MAPPING WEIGHTS

As an additional quantitative support to statistical clustering and modeling methods for integration with the more qualitative GEM mappings, and Medicare-specific Hierarchical Condition Category (HCC) based case type data used to derive baseline averages for key demographics and costs, in 2008 CMS released Medicare Severity-DRG weights. The weights are developed by CMS and its analytics consultants using exploratory statistical factor analysis, principal components, and other cluster profiling methods to derive coefficient dimensions based 5 categories of factors with most significant impacts on potential cost and clinical complications, for each Medicare DRG/HCC. These MS-DRG weights can be used by health providers and insurers to improve precision of their SAS models and forecasts.

MEDICARE CLINICAL OUTCOME COST OPTIMIZATION INFORMATICS AND ANALYSIS

Once you have a general understanding of the Medicare diagnostic codes, and Medicare clinical procedure and drug prescription groups, as well as new Medicare claims formats, conventions and rules for ICD-9-CM to ICD-10-CM GEM's (and new severity-cost weights); you can begin integrating into SAS Medicare clinical analysis and models. The first step is to download or order Medicare data samples and detail data for baselines and model calibrations.

MEDICARE PUBLIC USE FILE (PUF) FILES

Soon after Medicare was first implemented, the US government began providing detailed files for use by Medicare providers and health insurers to analyze their clinical outcome and cost containment performance. The first of these were Medicare Provider Analysis and Review (MEDPAR) files, which were produced annually from 1985 to 2004.

These files contained 100% of Medicare billings and payments, with encrypted patient identities, and diagnostic, medication and clinical procedures, and with a wide variety of patient demographics. Similar Pubic Usage Files (PUF's) were provided as 5% random samples at no cost up till 2008, when these files were no longer annually produced due to lack of funds. Yet these 2008 files are useful for training and test calibrations new files are available.

The 100% Medicare patient ID-encrypted files are very useful to help establish baselines and variance parameters for comparisons to individual provider and health insurer's internal performance, and for forecast and ROI estimation model calibrations. As a result of public comments and congressional testimony related to the National Healthcare law of 2010, Medicare may resume providing new 5% sample PUF files in the near future. Again, in the meantime, the Medicare PUF files published in 2008 can be useful for training and model building for baseline calibrations.

MEDICARE BASIC STAND ALONE (BSA) AND RELATED PUF FILES

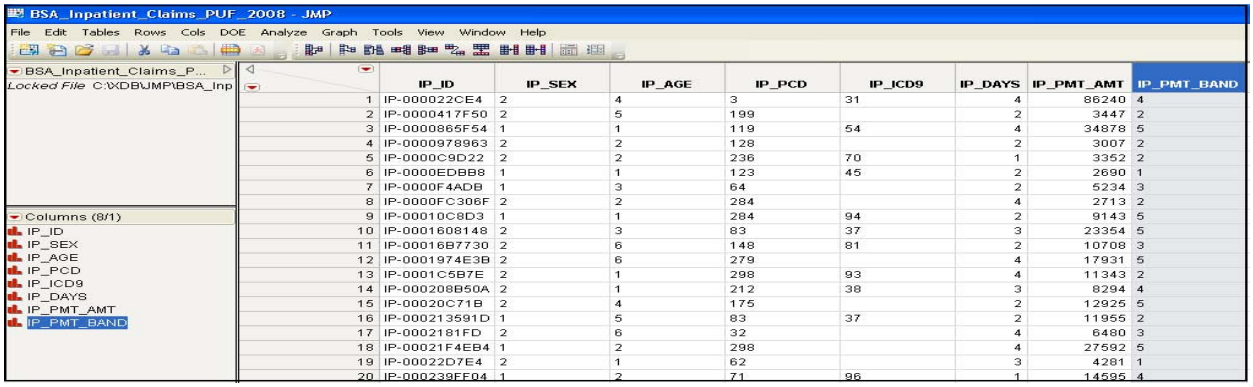
There are many categories of available Medicare BSA and PUF files, including Inpatient and Outpatient Claims, Durable Medical Equipment, Prescription Drug Events, Home Health, Skilled Nursing Facility, Chronic Conditions, and Hospice Claims data. Each is a 5% randomized sample, with ICD-9-CM diagnosis codes (as of 2008), clinical procedures, demographics, entitlement, enrollment status, and more. Again, SAS formats are provided by CMS.

The most widely used (and largest) files relate to Medicare Inpatient and Outpatient claims. While the 100% files can be optionally ordered in SAS file formats, 5% sample files are provided similar to ICD-9-CM Crosswalk files in Excel .CSV formats. However, Medicare provides sample SAS code templates for converting the .CSV files to SAS formats (such as shown in the following SAS code). Users can write their own SAS code if they wish, but it is recommended to use official CMS-Medicare claims processing defined SAS Proc Formats, as documented in CMS Tech User files.

```
/*************************************************************************\

     PROGRAM:                  ICD9_BSA_INPATIENT_CLAIMS_PUF_SAS_INPUT_PREPROCESS
     DESCRIPTION:              Read Excel .csv format inpatient PUF to SAS dataset
     INPUT FILES:                 From '2008_BSA_Inpatient_Claims_PUF.csv'
     OUTPUT FILES:                To 'BSA_Inpatient_Claims_PUF_2008.sas7bdat'
     OTHER OUTPUT:             Proc formats and proc freq descriptive statistics
     SOURCE:                   US GOVT / CENTERS FOR MEDICARE AND MEDICAID SVCS
\*************************************************************************/

options ls=120 ps=42 missing=' ' nocenter validvarname=upcase compress=binary;
filename in "DIRECTORY_NAME\2008_BSA_Inpatient_Claims_PUF.csv" ; * <- In File Path;
libname PUF 'INSERT_OUTPUT_DIRECTORY_NAME'; * <-- Place Output Library Path Here *;

Proc format;

Value SEXF
    1 = 'MALE '
    2 = 'FEMALE';

Value AGEF
    1      = 'Under  65 '
    2      = '65 - 69 '
    3      = '70 - 74 '
    4      = '75 - 79 '
    5      = '80 - 84 '
    6      = '85 & Older';

Value LOSF /* Length Of Stay */
    1 =  '1 day'
    2 =  '2-3 days'
    3 =  '4-7 days'
    4 =  '8 or more days';
```

```
Value DRGF /* Diagnosis Related Grouping Codes*/
1     =  "Heart transplant or implant of heart assist system"
2     =  "ECMO or trach w MV 96+ hrs or PDX exc face, mouth & neck w maj O.R."
3     =  "Trach w MV 96+ hrs or PDX exc face, mouth & neck w/o maj O.R."
4     =  "Liver transplant"
...
308      =  "Extensive O.R. procedure unrelated to principal diagnosis"
309      =  "Prostatic O.R. procedure unrelated to principal diagnosis"
310      =  "Non-extensive O.R. proc unrelated to principal diagnosis"
311      =  "Ungroupable";

Value  $ICDF /* Clinical Procedure Codes */
''   =  'No procedure performed'
'00' =  'Not elsewhere classified'
'01' =  'Incise-excis brain/skull'
'02' =  'Other skull/brain ops'
'03' =  'Spinal cord & canal ops'
'04' =  'Cran & periph nerve ops'
...
'95' =  'Eye & ear dx/treatment'
'96' =  'Non-op intubat & irrigat'
'97' =  'Replace & remov devices'
'98' =  'Nonop remove foreign bod'
'99' =  'Other nonoperative proc';
run;

data PUF.BSA_Inpatient_Claims_PUF_2008(label='INPATIENT CLAIMS PUBLIC USE FILE');
infile in dlm = ',' dsd missover truncover firstobs = 2;
  input IP_CLM_ID              :$19.
        BENE_SEX_IDENT_CD
        BENE_AGE_CAT_CD
        IP_CLM_BASE_DRG_CD
        IP_CLM_ICD9_PRCDR_CD   :$2.
        IP_CLM_DAYS_CD
        IP_DRG_QUINT_PMT_AVG
        IP_DRG_QUINT_PMT_CD;

  format BENE_SEX_IDENT_CD              SEXF.
         BENE_AGE_CAT_CD                AGEF.
         IP_CLM_BASE_DRG_CD             DRGF.
         IP_CLM_ICD9_PRCDR_CD           $ICDF.
         IP_CLM_DAYS_CD                 LOSF. ;

  label IP_CLM_ID           = "Encrypted PUF ID"
    BENE_SEX_               = "Beneficiary gender code "
    BENE_AGE                = "Beneficiary Age category code "
    IP_CLM_BASE_            = "Base DRG code "
    IP_CLM_ICD9_PRCDR_CD    = "ICD9 primary procedure code"
    IP_CLM_DAYS_CD          = "Inpatient days code"
    IP_DRG_QUINT_PMT_AVG    = "DRG quintile average payment amount"
    IP_DRG_QUINT_PMT_CD     = "DRG quintile payment amount code" ;
run;

proc freq data=PUF.BSA_Inpatient_Claims_PUF_2008;
  table BENE_SEX_IDENT_CD
        BENE_AGE_CAT_CD
        IP_CLM_BASE_DRG_CD
        IP_CLM_ICD9_PRCDR_CD
        IP_CLM_DAYS_CD
        IP_DRG_QUINT_PMT_AVG
        IP_DRG_QUINT_PMT_CD  /missing;
run;
```

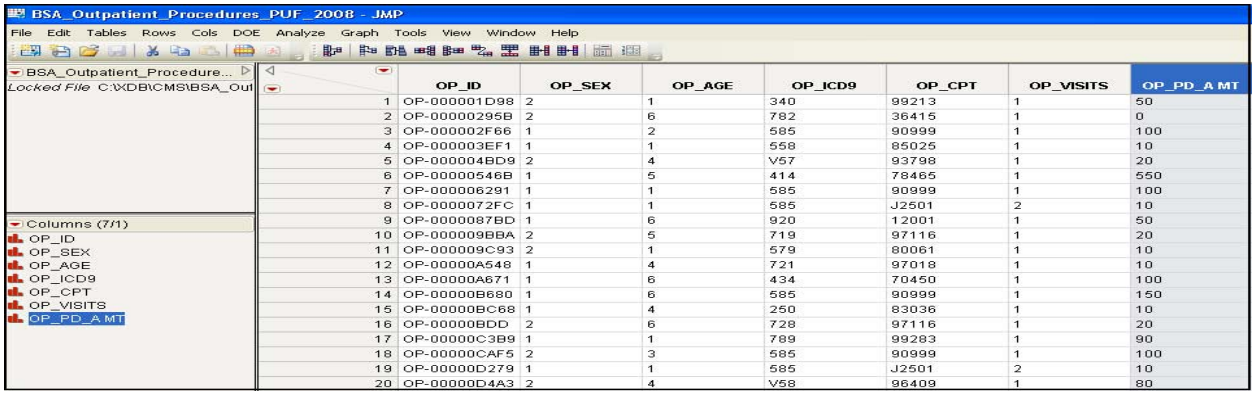
This code has been adapted for the following BSA Inpatient Claims File by a routine load process set up in SAS/JMP (with only minor variable renames to accommodate in-house IT client naming conventions). This display is followed by a corresponding BSA Outpatient Claims file in SAS/JMP as separate files that can be merged or split for analysis.



	IP_ID	IP_SEX	IP_AGE	IP_PCD	IP_ICD9	IP_DAYS	IP_PMT_AMT	IP_PMT_BAND
1	IP-000022CE4	2	4	3	31	4	86240	4
2	IP-0000417F50	2	5	199		2	3447	2
3	IP-0000865F54	1	1	119	54	4	34878	5
4	IP-0000978963	2	2	128		2	3007	2
5	IP-0000C9D22	2	2	236	70	1	3352	2
6	IP-0000EDBB8	1	1	123	45	2	2690	1
7	IP-0000F4ADB	1	3	64		2	5234	3
8	IP-0000FC306F	2	2	284		4	2713	2
9	IP-00010C8D3	1	1	284	94	2	9143	5
10	IP-0001608148	2	3	83	37	3	23354	5
11	IP-00016B7730	2	6	148	81	2	10708	3
12	IP-0001974E3B	2	6	279		4	17931	5
13	IP-0001C5B7E	2	1	298	93	4	11343	2
14	IP-000208B50A	2	1	212	38	3	8294	4
15	IP-00020C71B	2	4	175		2	12925	5
16	IP-000213591D	1	5	83	37	2	11955	2
17	IP-0002181FD	2	6	32		4	6480	3
18	IP-00021F4EB4	1	2	298		4	27592	5
19	IP-00022D7E4	2	1	62		3	4281	1
20	IP-000239FF04	1	2	71	96	1	14595	4

Display 2. SAS/JMP Medicare Inpatient Procedures Patient Demographics by Banded Paid Amounts

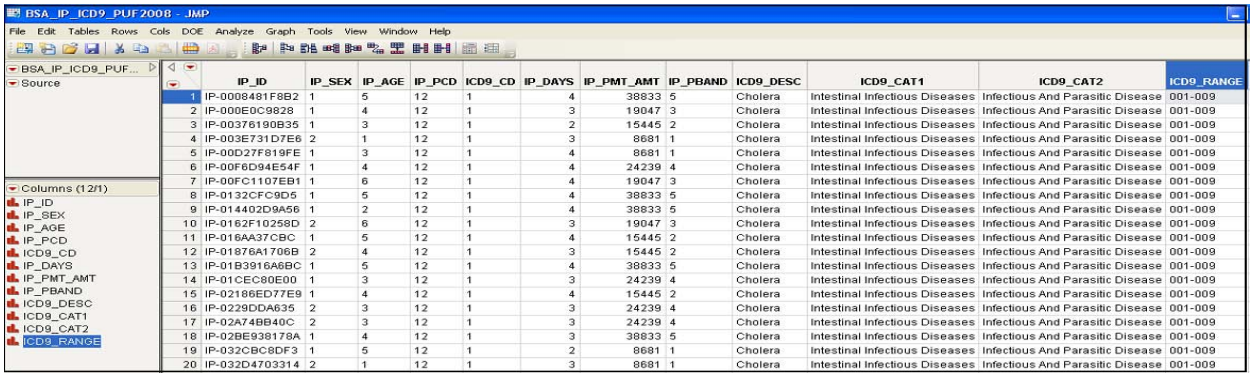
The encrypted Patient ID's were originally designed to align subjects longitudinally across years, and across BSA file types (with first 2 characters for BSA file type and alphanumeric encrypted code after a dash). The files can be appended to a single expanded SAS table file or maintained in individual SAS files or tables depending on objectives.



	OP_ID	OP_SEX	OP_AGE	OP_ICD9	OP_CPT	OP_VISITS	OP_PD_A MT
1	OP-000001D98	2	1	340	99213	1	50
2	OP-00000295B	2	6	782	36415	1	0
3	OP-000002F66	1	2	585	90999	1	100
4	OP-000003EF1	1	1	558	85025	1	10
5	OP-000004BD9	2	4	V57	93798	1	20
6	OP-00000546B	1	5	414	78465	1	550
7	OP-000006291	1	1	585	90999	1	100
8	OP-0000072FC	1	1	585	J2501	2	10
9	OP-0000087BD	1	6	920	12001	1	50
10	OP-000009BBA	2	5	719	97116	1	20
11	OP-000009C93	2	1	579	80061	1	10
12	OP-00000A548	1	4	721	97018	1	10
13	OP-00000A671	1	6	434	70450	1	100
14	OP-00000B680	1	6	585	90999	1	150
15	OP-00000BC68	1	4	250	83036	1	10
16	OP-00000BDD	2	6	728	97116	1	20
17	OP-00000C3B9	1	1	789	99283	1	90
18	OP-00000CAF5	2	3	585	90999	1	100
19	OP-00000D279	1	1	585	J2501	2	10
20	OP-00000D4A3	2	4	V58	96409	1	80

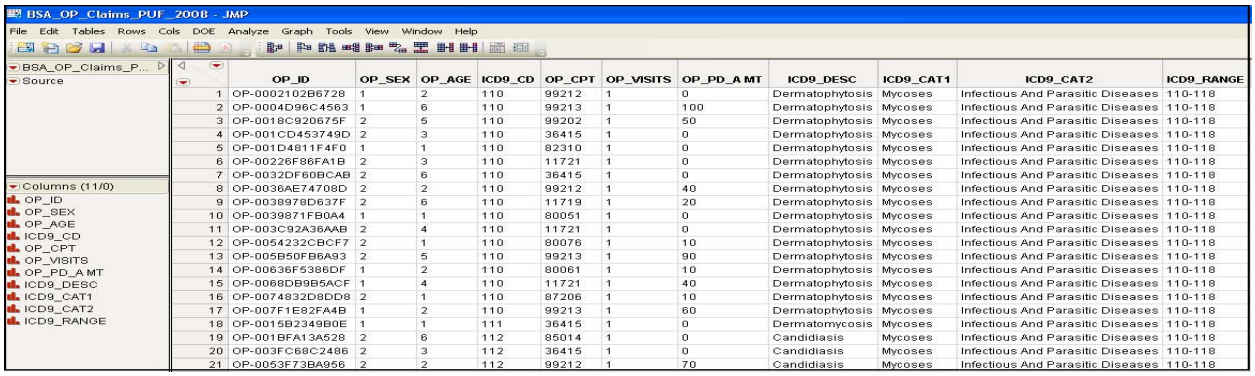
Display 3. SAS/JMP Medicare Outpatient Procedures Patient Demographics by Paid Amounts

These files can also be merged to Medicare Claims diagnostic codes and coding categories for categorical analysis and profiling, using SAS statistical procedures such as clustering, principal components, ANOVA, among others. This can be done at a general to intermediate level using SAS/JMP or SAS/Enterprise Guide, or at a more advanced level using the large number of Analysis Nodes in SAS/Enterprise Miner. The following displays are aligned for the ICD-9-CM Crosswalk descriptor mappings and categories, but can also be aligned to ICD-10-CM GEM mappings by following the rules and specifications in the "GEM Technical Users Guide", which can be found on the CMS website.



	IP_ID	IP_SEX	IP_AGE	IP_PCD	ICD9_CD	IP_DAYS	IP_PMT_AMT	IP_PBAND	ICD9_DESC	ICD9_CAT1	ICD9_CAT2	ICD9_RANGE
1	IP-0008481F8B2	1	5	12	1	4	38833	5	Cholera	Intestinal Infectious Diseases	Infectious And Parasitic Disease	001-009
2	IP-000E0C9828	1	4	12	1	3	19047	3	Cholera	Intestinal Infectious Diseases	Infectious And Parasitic Disease	001-009
3	IP-0037619OB35	1	3	12	1	2	15445	2	Cholera	Intestinal Infectious Diseases	Infectious And Parasitic Disease	001-009
4	IP-003E731D7E6	2	1	12	1	3	8681	1	Cholera	Intestinal Infectious Diseases	Infectious And Parasitic Disease	001-009
5	IP-00D27F819FE	1	3	12	1	4	8681	1	Cholera	Intestinal Infectious Diseases	Infectious And Parasitic Disease	001-009
6	IP-00F6D94E54F	1	4	12	1	4	24239	4	Cholera	Intestinal Infectious Diseases	Infectious And Parasitic Disease	001-009
7	IP-00FC1107EB1	1	6	12	1	4	19047	3	Cholera	Intestinal Infectious Diseases	Infectious And Parasitic Disease	001-009
8	IP-0132CFC9D5	1	5	12	1	4	38833	5	Cholera	Intestinal Infectious Diseases	Infectious And Parasitic Disease	001-009
9	IP-014402D9A56	1	2	12	1	4	38833	5	Cholera	Intestinal Infectious Diseases	Infectious And Parasitic Disease	001-009
10	IP-0162F10258D	2	6	12	1	3	19047	3	Cholera	Intestinal Infectious Diseases	Infectious And Parasitic Disease	001-009
11	IP-016AA37CBC	1	5	12	1	4	15445	2	Cholera	Intestinal Infectious Diseases	Infectious And Parasitic Disease	001-009
12	IP-01876A1706B	2	4	12	1	3	15445	2	Cholera	Intestinal Infectious Diseases	Infectious And Parasitic Disease	001-009
13	IP-01B3916A6BC	1	5	12	1	4	38833	5	Cholera	Intestinal Infectious Diseases	Infectious And Parasitic Disease	001-009
14	IP-01CEC80E00	1	3	12	1	3	24239	4	Cholera	Intestinal Infectious Diseases	Infectious And Parasitic Disease	001-009
15	IP-02186ED77E9	1	4	12	1	4	15445	2	Cholera	Intestinal Infectious Diseases	Infectious And Parasitic Disease	001-009
16	IP-0229DDA635	2	3	12	1	3	24239	4	Cholera	Intestinal Infectious Diseases	Infectious And Parasitic Disease	001-009
17	IP-02A74BB40C	2	3	12	1	3	24239	4	Cholera	Intestinal Infectious Diseases	Infectious And Parasitic Disease	001-009
18	IP-02BE938178A	1	4	12	1	3	38833	5	Cholera	Intestinal Infectious Diseases	Infectious And Parasitic Disease	001-009
19	IP-032CBC8DF3	1	5	12	1	2	8681	1	Cholera	Intestinal Infectious Diseases	Infectious And Parasitic Disease	001-009
20	IP-032D4703314	2	1	12	1	3	8681	1	Cholera	Intestinal Infectious Diseases	Infectious And Parasitic Disease	001-009

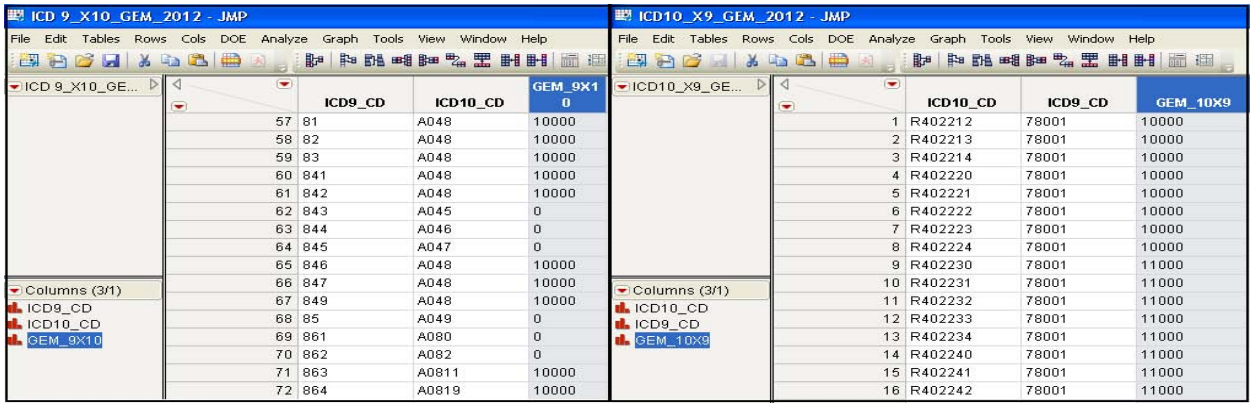
Display 4. SAS/JMP Medicare Inpatient Claims Demographics with ICD-9-CM Categories and Costs



	OP_ID	OP_SEX	OP_AGE	ICD9_CD	OP_CPT	OP_VISITS	OP_PD_AMT	ICD9_DESC	ICD9_CAT1	ICD9_CAT2	ICD9_RANGE
1	OP-0002102B6728	1	2	110	99212	1	0	Dermatophytosis	Mycoses	Infectious And Parasitic Diseases	110-118
2	OP-0004D96C4563	1	6	110	99213	1	100	Dermatophytosis	Mycoses	Infectious And Parasitic Diseases	110-118
3	OP-0018C920675F	2	5	110	99202	1	50	Dermatophytosis	Mycoses	Infectious And Parasitic Diseases	110-118
4	OP-001CD453749D	2	3	110	36415	1	0	Dermatophytosis	Mycoses	Infectious And Parasitic Diseases	110-118
5	OP-001D4811F4F0	1	1	110	82310	1	0	Dermatophytosis	Mycoses	Infectious And Parasitic Diseases	110-118
6	OP-00226F86FA1B	2	3	110	11721	1	0	Dermatophytosis	Mycoses	Infectious And Parasitic Diseases	110-118
7	OP-0032DF60BCAB	2	6	110	36415	1	0	Dermatophytosis	Mycoses	Infectious And Parasitic Diseases	110-118
8	OP-0036AE74708D	2	2	110	99212	1	40	Dermatophytosis	Mycoses	Infectious And Parasitic Diseases	110-118
9	OP-0038978D637F	2	6	110	11719	1	20	Dermatophytosis	Mycoses	Infectious And Parasitic Diseases	110-118
10	OP-0039871FB0A4	1	1	110	80051	1	0	Dermatophytosis	Mycoses	Infectious And Parasitic Diseases	110-118
11	OP-003C92A36AAB	2	4	110	11721	1	0	Dermatophytosis	Mycoses	Infectious And Parasitic Diseases	110-118
12	OP-0054232CBCF7	2	1	110	80076	1	10	Dermatophytosis	Mycoses	Infectious And Parasitic Diseases	110-118
13	OP-005B50FB6A93	2	5	110	99213	1	90	Dermatophytosis	Mycoses	Infectious And Parasitic Diseases	110-118
14	OP-00636F5386DF	1	2	110	80061	1	10	Dermatophytosis	Mycoses	Infectious And Parasitic Diseases	110-118
15	OP-0068DB9B5ACF	1	4	110	11721	1	40	Dermatophytosis	Mycoses	Infectious And Parasitic Diseases	110-118
16	OP-0074832D8DD8	2	1	110	87206	1	10	Dermatophytosis	Mycoses	Infectious And Parasitic Diseases	110-118
17	OP-007F1E82FA4B	1	2	110	99213	1	60	Dermatophytosis	Mycoses	Infectious And Parasitic Diseases	110-118
18	OP-0015B2349B0E	1	1	111	36415	1	0	Dermatomycosis	Mycoses	Infectious And Parasitic Diseases	110-118
19	OP-001BFA13A528	2	6	112	85014	1	0	Candidiasis	Mycoses	Infectious And Parasitic Diseases	110-118
20	OP-003FC68C2486	2	3	112	36415	1	0	Candidiasis	Mycoses	Infectious And Parasitic Diseases	110-118
21	OP-0053F73BA956	2	2	112	99212	1	70	Candidiasis	Mycoses	Infectious And Parasitic Diseases	110-118

Display 5. SAS/JMP Medicare Outpatient Claims Demographics with ICD-9-CM Categories and Costs

In the following display can be seen two side-by-side SAS files for latest Medicare ICD-9-CM to ICD-10-CM GEM's, and to corresponding ICD-10-CM to ICD-9-CM GEM's (these GEM'S will not perfectly align to 2008 Claims data, so require some special imputation rules for precision analysis; but the purpose here is simply to show the two GEM's).

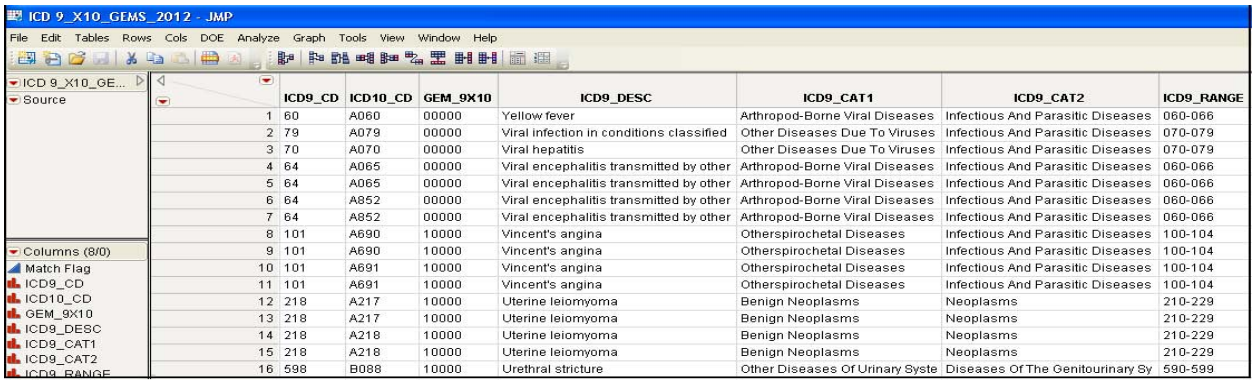


	ICD9_CD	ICD10_CD	GEM_9X10
57	81	A048	10000
58	82	A048	10000
59	83	A048	10000
60	841	A048	10000
61	842	A048	10000
62	843	A045	0
63	844	A046	0
64	845	A047	0
65	846	A048	10000
66	847	A048	10000
67	849	A048	10000
68	85	A049	0
69	861	A080	0
70	862	A082	0
71	863	A0811	10000
72	864	A0819	10000

	ICD10_CD	ICD9_CD	GEM_10X9
1	R402212	78001	10000
2	R402213	78001	10000
3	R402214	78001	10000
4	R402220	78001	10000
5	R402221	78001	10000
6	R402222	78001	10000
7	R402223	78001	10000
8	R402224	78001	10000
9	R402230	78001	11000
10	R402231	78001	11000
11	R402232	78001	11000
12	R402233	78001	11000
13	R402234	78001	11000
14	R402240	78001	11000
15	R402241	78001	11000
16	R402242	78001	11000

Display 6. SAS/JMP Medicare GEM's for ICD-9-CM to ICD-10-CM and ICD-10-CM to ICD-9-CM Reverse Maps

It should be noted that each GEM file has 5 byte maps that specify the correspondence of a directional mapping of the smaller number of ICD-9-CM codes (15,000) to ICD-10-CM codes (68,000+). Each byte is a special category flag that can be coded for SAS code grouping rules or Proc Formats according to CMS "GEM Users Guide". The direction of mapping is critical to align ICD-9-CM to ICD-10-CM during conversion or regression testing of models and reports, or for aligning any forecasting models or year-over-year historical comparison reporting until ICD-9-CM fully rolls off. This may involve running and maintaining parallel SAS files and processes, or unified files or tables to use SAS/SQL or native SQL processes to select and transform the Medicare Claims data for accurate forecasting and reporting.



	ICD9_CD	ICD10_CD	GEM_9X10	ICD9_DESC	ICD9_CAT1	ICD9_CAT2	ICD9_RANGE
1	60	A060	00000	Yellow fever	Arthropod-Borne Viral Diseases	Infectious And Parasitic Diseases	060-066
2	79	A079	00000	Viral infection in conditions classified	Other Diseases Due To Viruses	Infectious And Parasitic Diseases	070-079
3	70	A070	00000	Viral hepatitis	Other Diseases Due To Viruses	Infectious And Parasitic Diseases	070-079
4	64	A065	00000	Viral encephalitis transmitted by other	Arthropod-Borne Viral Diseases	Infectious And Parasitic Diseases	060-066
5	64	A065	00000	Viral encephalitis transmitted by other	Arthropod-Borne Viral Diseases	Infectious And Parasitic Diseases	060-066
6	64	A852	00000	Viral encephalitis transmitted by other	Arthropod-Borne Viral Diseases	Infectious And Parasitic Diseases	060-066
7	64	A852	00000	Viral encephalitis transmitted by other	Arthropod-Borne Viral Diseases	Infectious And Parasitic Diseases	060-066
8	101	A690	10000	Vincent's angina	Otherspirochetal Diseases	Infectious And Parasitic Diseases	100-104
9	101	A690	10000	Vincent's angina	Otherspirochetal Diseases	Infectious And Parasitic Diseases	100-104
10	101	A691	10000	Vincent's angina	Otherspirochetal Diseases	Infectious And Parasitic Diseases	100-104
11	101	A691	10000	Vincent's angina	Otherspirochetal Diseases	Infectious And Parasitic Diseases	100-104
12	218	A217	10000	Uterine leiomyoma	Benign Neoplasms	Neoplasms	210-229
13	218	A217	10000	Uterine leiomyoma	Benign Neoplasms	Neoplasms	210-229
14	218	A218	10000	Uterine leiomyoma	Benign Neoplasms	Neoplasms	210-229
15	218	A218	10000	Uterine leiomyoma	Benign Neoplasms	Neoplasms	210-229
16	598	B088	10000	Urethral stricture	Other Diseases Of Urinary Syste	Diseases Of The Genitourinary Sy	590-599

Display 7. SAS/JMP Medicare GEM ICD-9-CM to ICD-10-CM Cross-Category Descriptions Clinical Review

Whether GEM files are merged or joined forward (ICD-9-CM to ICD-10-CM) or backward (ICD-10-CM to ICD-9-CM), it is critical to follow instruction in the Medicare "GEM Users Guide", and the latest guidelines on the CMS website.

MEDICARE CLINICAL OUTCOMES AND COST OPTIMIZATION INFORMATICS

The statistical procedures and methodologies for Medicare clinical outcomes and cost optimization informatics can be supported by several SAS vertical clinical and pharmaceutical industry products and statistical procedures, and can be integrated across SAS/BI, SAS/DI, SAS/AP and SAS/HPF Analytic Platforms. Large organizations will benefit most from SAS/BI architectures and SAS/Enterprise Miner and SAS/Enterprise Guide, but smaller organizations and researchers can also benefit from SAS On Demand and SAS/JMP Clinical products for Medicare models and reports.

ANALYZING MEDICARE HEALTH INSURANCE PROVIDER AND PATIENT CLINICAL OUTCOME VARIANCE

During the period of transition from ICD-9-CM to ICD-10-CM (which may yet involve several years depending on the organization as well as future Medicare regulations), it may likely be necessary to maintain not only the usual budget and forecast versus actual variance estimation and reporting, but also variances between cost projections based on ICD-9-CM versus ICD-10-CM based assumptions. It has already been reported that ROI variances have been experienced by healthcare providers and insurers, and many have had higher than usual errors in predictive models and forecasts, likely at least partially due to growing complexity and changes in both ICD-9-CM and ICD-10-CM.

Due to the nature of emerging impacts of the National Healthcare law of 2010 upon Medicare, and all other health services delivery regulations, it will be critical to have rapid flexibility to respond to these changes; and to quickly make and validate impact of such changes. SAS/BI and SAS/JMP scalable platforms are ideal to help achieve this.

In many cases, there are a wealth of SAS clinical industry vertical products that will provide standardized support for these emerging Medicare regulations, and the SAS/BI, SAS/JMP and SAS On Demand platforms can also provide for rapid development and deployment of proprietary organizational solutions using SAS programming software tools, including SAS add-on products such as SAS/QC, SAS/OR and SAS/IML, among others. But in order to truly enable your organization to mostly precisely forecast, estimate and predict Medicare costs and ROI, you will most likely need to increase number and accuracy of your SAS models, as well as frameworks to support automated SAS processes.

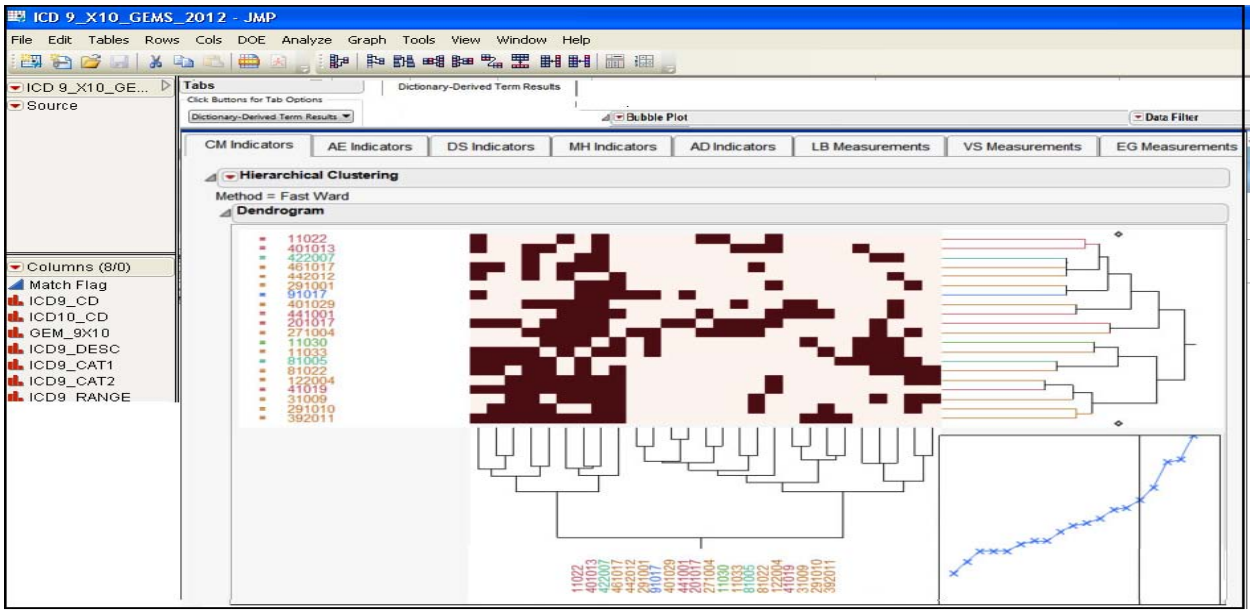
PROFILING MEDICARE PAYMENT PERFORMANCE BY PROVIDER AND PATIENT CLUSTERING MODELS

Profiling and segmentation are as important to Clinical Outcome Informatics, as they are to Customer Relationship Management, and many of the same statistical methods and tools are appropriate for Medicare Claims analysis of provider and patient clinical outcomes for efficacy and cost containment. Depending on your SAS product licenses and configurations, this can include SAS/STAT procedures, such as clustering, principal components, ANOVA and many others, or SAS/Enterprise Miner nodes such as decision trees, batch quantization and self-organizing maps. All of these and more are also supported by SAS/JMP, at a less advanced level yet suitable for smaller organizations.

In the following displays, two of the many advanced SAS/JMP statistical profiling model methods are shown that are among alternatives for fast and flexible solutions for analysis and segmentation as well as comparison of ICD-9 –CM to ICD-10-CM variances and commonalities, to assist in analysis of Medicare population and sub-populations to optimize clinical outcomes and costs. The first demonstrates a cluster analysis plot of ICD-9-CM vs ICD-10-CM diagnostic codes and weighted using claims paid for K-means, as one of many SAS/JMP clustering options.

Display 8. SAS/JMP Medicare GEM File ICD-9-CM to ICD-10-CM Code Cluster Bubble Plot Overlaps

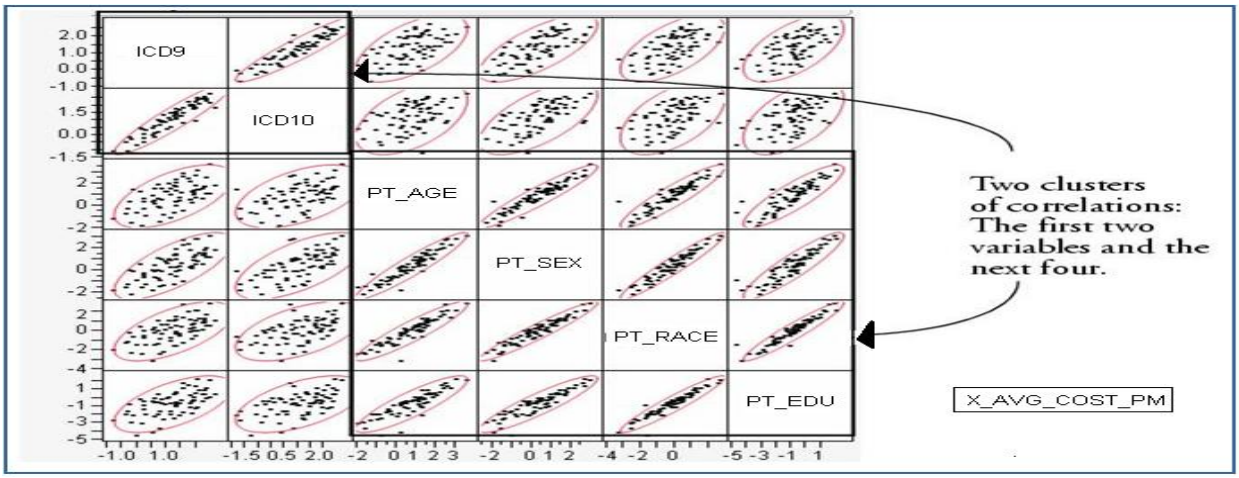
The second example demonstrates a SAS/JMP Hierarchical Clustering Dendrogram to apply GEMS for ICD-9–CM vs ICD-10-CM aggregation using Bivariate Forward-Reverse Cross-Validation methods, which are another option of SAS/JMP. More advanced profiling methods are available with SAS/EM nodes. However, the capabilities of SAS/JMP are acceptable for small sample Medicare claims profiling and clustering analysis. Although on a client based platform SAS/JMP can run a long time for such analysis unless sample size is very small, SAS On Demand is a web based alternative. For most organizations, one is enough; but multi-tiered tiered solution is recommended for large healthcare provider and health insurance organizations to be best positioned for future Medicare changes.



Display 9. SAS/JMP Medicare GEM File ICD-9-CM to ICD-10-CM Bivariate Reverse Cross-Validation Tree

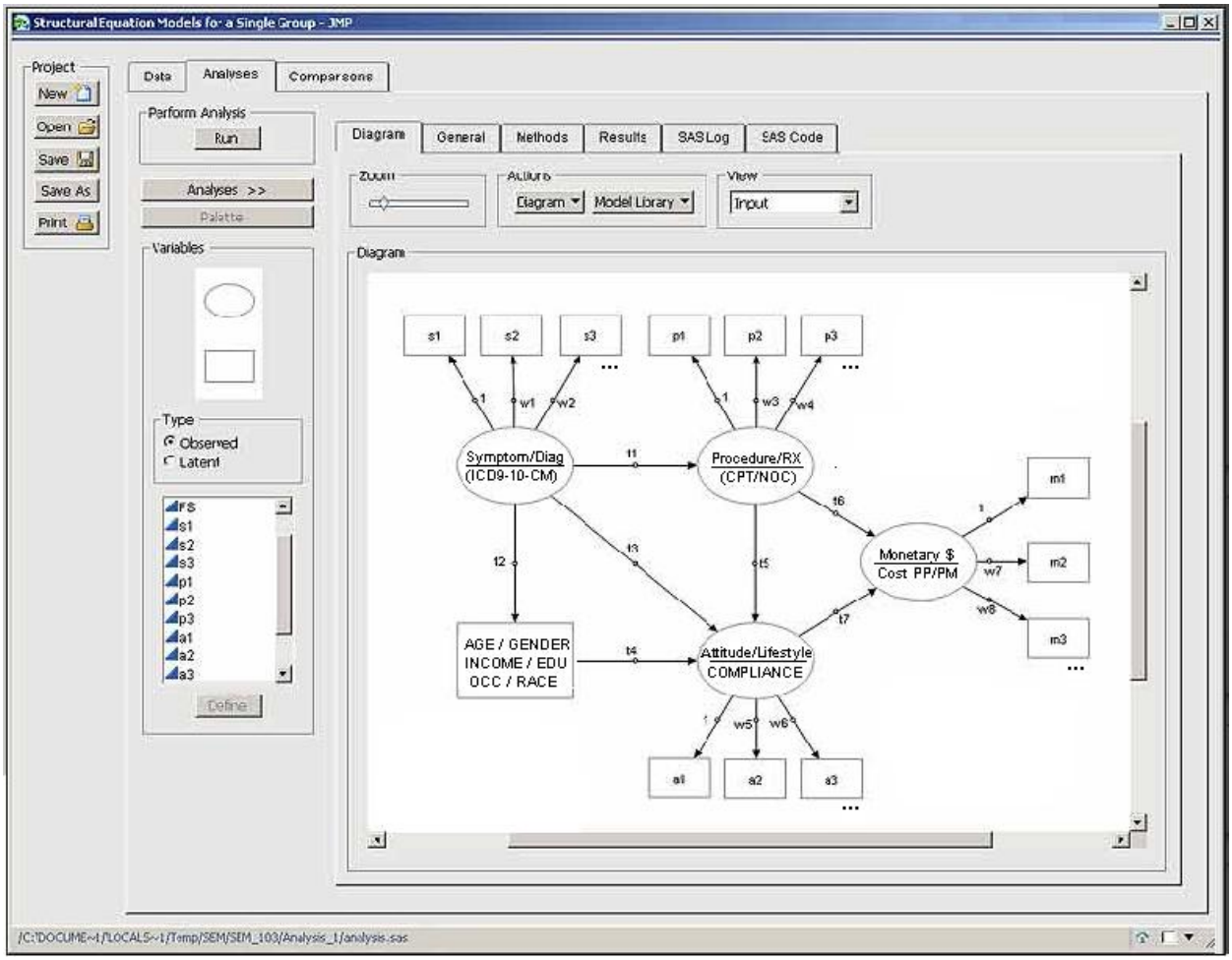
PREDICTING MEDICARE PROVIDER AND PATIENT COST VARIANCE USING MS-DRG WEIGHTED MODELS

The most advanced SAS statistical Medicare clinical outcome and cost optimization models involve predictive models for ROI estimations and forecasting. This normally involves weights and factors based on profiling and segmentation SAS procedures, with baselines and calibration against both overall populations or randomized samples of Medicare Claims data, compared to proprietary sub-population data from an individual provider or healthcare insurer (or company group plans); in order to better forecast and predict costs and payments in order to optimize ROI within a range of risk scenarios, which can be supported by the following cluster matrix for MS-DRG correlated demographics.



Display 10. SAS/JMP Annotated Output for Two-Tier Basic Medicare Profile Cluster Predictive Model Matrix

After all ICD-9-CM to ICD-10-CM conversions are completed, such use of the Medicare Severity–Diagnosis Related Group weights can also be an extremely valuable tool to improve all Medicare clinical cost management models, using SAS/STAT Logistic Regression and other GLM/NLM procedures, as well as SAS/EM and SAS/JMP. These models can range from simple models that integrate from 2-dimensional clustering to complex models, such as use of ICD-9-CM to ICD-10-CM correlation matrix with claim diagnostics and procedures, as well as patient demographics and optionally supplemented by many other available Medicare special surveys and scales; and calibrated to baselines derived from Medicare BSA and PUF file variables in order to build very complex predictive models, including discriminative analysis, neural networks, CART/CHAID decision trees, and Structural Equation Models.



Display 11. SAS/JMP Structural Equation Model for Medicare Cost Optimization (as developed by the author)

In this final display is shown an example of such a Structural Equation Model, originally developed by this author in SAS/STAT Proc CALIS, and on SAS/EM; then ported to SAS/JMP in accordance with CMS clinical outcome analytic guidelines that include all the steps presented previously, from ICD-9-CM and ICD-10-CM data formats, Medicare GEM mappings, MS-DRG weight factoring, and SAS exploratory clustering and predictive modeling components (as used by the author for SAS platform applications and SAS tool customizations). Together these SAS platform analytic and data management tools, and both Medicare and internal organization data, can provide effective and efficient analytics platforms to support all ICD-9 and ICD-10 conversions, or even ICD-11 conversions into the future.

CONCLUSION

The many new changes underway by Medicare and other US and international organizations to improve upon all ICD-9 to ICD-10, and Medicare Severity DRG's, have a great potential to improve upon healthcare costs containment and efficiency into the future. All SAS analytic and data management platforms are positioned and ready for rapid development and deployment of practical and effective solutions that leverage the best of academic and business analytics to optimize resources to support future US national health care at reasonable cost for best quality of care.

ACKNOWLEGEMENTS

Many thanks to AMERIGROUP, BCBS, HUMANA, UHG, and QUALEX Consulting for their review of this article.

REFERENCES

- Centers for Medicare and Medicaid Services (CMS), 2008. Basic Stand Alone (BSA) Inpatient Claims Public Use File (PUF) General Documentation. Washington, DC; US Government.
- Centers for Medicare and Medicaid Services (CMS), 2008. Basic Stand Alone (BSA) Outpatient Claims Public Use File (PUF) General Documentation. Washington, DC; US Government.
- Centers for Medicare and Medicaid Services (CMS), 2008. BSA Inpatient Claims PUF Data Dictionary and 'Codebook. Washington, DC; US Government.
- Centers for Medicare and Medicaid Services (CMS), 2008. BSA Outpatient Claims PUF Data Dictionary and Codebook. Washington, DC; US Government.
- Centers for Medicare and Medicaid Services (CMS), "Converting MS-DRGs to ICD-10-CM and ICD-10-PCS", 2012. Available at https://www.cms.gov/ICD10/17_ICD10_MS_DRG_Conversion_Project .
- Centers for Medicare and Medicaid Services (CMS), "General Equivalence Maps for Technical Users", 2012. Available at http://www.cms.gov/ICD10 .
- Centers for Medicare and Medicaid Services (CMS), "International Classification of Diseases, Tenth Revision, Clinical Modification (ICD-10-CM)", 2012. Available at http://www.cdc.gov/nchs/icd/icd10cm.htm .
- Mayes, Rick. "The Origins, Development, and Passage of Medicare's Revolutionary Prospective Payment System". Journal of History of Medicine and Allied Sciences, Vol. 62 (2007), pp. 21–55.
- Petrou, C. (Univ. of Louisville), "Use of Text Mining to Predict Patient Compliance", SUGI Proceedings, 2008.
- Pope, G., et al. (RTI), "Risk Adjustment of Medicare Capitation Payments Using the CMS-HCC Model". Health Care Financing Review, Vol. 25 (2004), pp. 119-141.
- Schacherer, C. and Gershenwald, J. "Utilizing SAS® as an Integrated Component of the Clinical Research Information System", SUGI Proceedings, 2008.
- Tesfamicael, M. (Univ. of Louisville), "Data Mining to Investigate the Prescribing of Medications Longitudinally", SUGI Proceedings, 2007.
- US Dept. of Commerce / NIST, "CMS Medicare Severity Diagnosis Related Grouper (MS-DRG Medicare Code Editor)", 2012. Available at http://www.ntis.gov/products/grouper.aspx .
- Warner, Christine (SRA), "Automating Predictive Analysis to Predict Medicare Fraud". SUGI Proceedings, 2005.

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Name:	Keith A. Jones, Ph.D.
Enterprise:	Qualimatix Consulting
Address:	350 N. Ervay, Ste. 1901
City, State ZIP:	Dallas, TX - 75201
Work Phone:	214-468-8191
Fax:	866-697-1544
E-mail:	info@qualimatix.com
Web:	www.qualimatix.com