Claims Based Indicators of Cardiac Surgical Site Infection

Profiling hospitals' risks of surgical site infections after cardiac procedures.

Prepared by:

Susan Huang, James Livingston, Richard Platt

Eastern Massachusetts Prevention Epicenter

Department of Ambulatory Care and Prevention,
Harvard Medical School and Harvard Pilgrim Health Care,

Channing Laboratory, Department of Medicine,
Brigham and Women's Hospital, Harvard Medical School
Boston, MA

Profiling hospitals' risks of surgical site infections after cardiac procedures.

OVERVIEW

Goal

Identify hospitals that may have unusually high or low risks of surgical site infection following invasive and non-invasive cardiac procedures.

Rationale

A typical health plan's claims data contain indicators of infections following surgery that can be used to estimate hospital-specific infection rates. These indicators include selected diagnoses, procedures, and prescriptions for antibiotics. These claims-based markers of infection can be used to guide further assessment of hospitals with potentially elevated risks of infection. Hospitals identified as having high percentages of patients with likely infection through this mechanism may not necessarily have high risks; additional evaluation would be required to determine the actual risk, and then to determine whether these risks could be reduced. This use of claims data has been described by Platt et al. (Emerg Infect Dis. 2002;8:1433-41, http://www.cdc.gov/ncidod/EID/vol8no12/02-0039.htm)

Routine assessment of hospitals' infection indicator scores may complement existing patient safety programs. Hospital-based infection control programs typically identify only a minority of infections, since most infections occur after discharge from the hospital and infected patients typically do not return to the hospital for care. In addition, it is difficult to compare hospitals' reported infection rates because hospital-based surveillance programs use different methods to identify infections.

Health plans' benefits and claims systems differ in ways that affect the accuracy (predictive value) of these indicators as predictors of infections. However, because differences between hospitals appear to be relatively consistent across health plans, it is possible to control for plan-to-plan differences. Information can also be combined and compared across health plans.

How these programs work

Every six or twelve months, health plans identify patients who have had a qualifying cardiac procedure since the previous time profiles were created. The initial profile should include all historical data that is readily accessible to the health plan.

Health plans create the following files for each patient who underwent cardiac procedures within the period of interest:

- Claims describing the procedure

- All inpatient and outpatient claims related to medical care within 30 days after the procedure

- All pharmacy claims from at least 6 months before the date of the procedure until 30 days afterward.

The included programs accomplish the following:

- Assign patients to a specific type of coverage, e.g., commercial, PPO, Medicare. This assignment allows the comparison of results within types of coverage. We have found

that the type of coverage affects the number and type of claims, and therefore the analyses control for this.

- Classify patients as having pharmacy benefits or not, based on their having a pharmacy dispensing claim in the six months preceding the procedure date. It is also important to control for this in the final comparisons.

- Assess the completeness of claims for individuals, and for specific types of coverage. Some claims files have insufficient information to allow useful comparisons. We perform several checks to identify and exclude data that is likely to be misleading.

- Identify patients with inpatient or outpatient claims that are indicators of infection during the 30 days after the cardiac procedure.

- Group these patients by hospital in which the cardiac procedure was performed, and by the time period in which they occurred.

- Compare the rates of infection indicators in different hospitals, taking into account the number of procedures performed, and adjusting for the patients' sex, and comorbidity status (assessed via the chronic disease score – which is computed from the drugs dispensed during the six months preceding the procedure). Analyses of infection indicator rates are further stratified by invasive and non-invasive procedures.

How these programs should be used
We recommend running these programs every six or twelve months and comparing data across half-years. As a general guideline, when comparing two half-years, at least 500 total procedures are needed for these analyses to be meaningful. In addition, at least 4 hospitals with at least 40 procedures each are most likely needed. Hospitals identified as possibly having higher than expected risks should be assessed further.

It is likely to be worthwhile for health plans operating in a single region to pool their tabulated results, in order to obtain more precise estimates of hospital-specific rates.

Interpreting and using the results
Remember that these claims based results are surrogate markers of infection. On average, approximately half of the patients with one of these indicators has the infection confirmed when the medical record is reviewed. Given this fact, **these claims-based surrogate infection rates cannot be compared to rates reported by individual hospitals**. Instead, the value of this analysis is in ranking hospitals according to their surrogate infection rates and in using a high threshold to guide further inquiry about the possibility that a hospital's rate is actually unusually high.

A high rate of these surrogate measures does not prove that a problem exists.

Some reasons that a high rate of these surgical site infections might be misleading include:

- an anomaly in the claims data from one or more institutions,

- unusual coding practices by some providers,

- an insufficient number of events to be informative,

- an unusual case mix that includes a disproportionate number of high risk patients.

USING THESE PROGRAMS

Computer environment.
The standard computer environment for all data files and programs is **SAS version 9**.

Users should set up a directory for housing SAS programs and datasets. The name of this directory should be chosen in accordance with local usage; it could be named **SSI_cardiac** to distinguish this study from other projects.

Please place all five supplied SAS programs in the project directory – preferably, in a subdirectory named **program --** or follow local naming convention.

Programs, supplied datasets, sample outputs
This plan contemplates running a set of five SAS programs by users on their own computers. The five programs, along with some ancillary SAS datasets, are provided in the project package.

Programs
1. **episodes.sas**

     Creates cardiac procedure episodes.

2**. qualify.sas**

     Applies user's time horizon to episodes, and marks each one as having prescription history (or not).

 3.   **detection.sas**

     Determines the presence of several types of infection indicators during the 30 days following each procedure.

 4. **cdscore.sas**

     Computes a comorbidity index (chronic disease score) for each episode.

 5. profile.sas

     Runs a series of validation steps to ensure the completeness and consistency of input data and creates profiles for individual hospitals. These validation steps result in an output file named "**validate_{today}.rtf**" which directs the user to key data elements for assessment of appropriateness of content, quantity, and distribution of the data.

     Also creates an analytic summary named "**summary_{today}.rtf**", and at the user's option, an Excel spreadsheet named "**patient_detail_{today}.xls**" with patient details for selected hospitals. The selection is based on hospital's rank in overall infection indicator rate.

Supplied test datasets
The test datasets are: SURGERY, DEMOG, DISP, CLAIMS -- all with .SAS7BDAT suffix.

They are similar to the production datasets that the user will prepare -- as described in the section "Extraction of data from source files", below.

But the test datasets were created artificially, so they do not have the full range of values expected for an actual surgical population.

Program testing with supplied test datasets

1. Place the supplied test datasets in a test directory, to be identified in each of the above programs, using the 'LIBNAME testdata' statement provided (see each program's header section).

2. Place the three ancillary datasets supplied in the same directory where you place the SAS programs. The ancillary datasets are: FORMGEN, MODEL2, CDSNDCS -- all with .SAS7BDAT suffix.

3. Run the five programs in sequence against the test data, and compare your results against the sample output provided in the package.

Sample outputs

The first four SAS programs produce a standard set of output files with suffixes customarily named .log, .lst, and .saserr.

The .log and .saserr files are of interest only to a programmer, who should review them for possible execution errors. Comparison against the test output .logs (included with this package) may be helpful.

If nothing unexpected occurs in running the first four SAS programs, their .saserr output files should be empty.

The .lst output files provide tabulations that help to validate the programs and data. Here are some key tables to examine.

>**episodes.lst** -- surgcode, episodeclass, finlstat.

>**qualify.lst --** sex, age, finlstat, leadinst, loe.

>**detection.lst** -- iorec, indicators at the service level, indicators at the case level, insurance product.

>**cdscore.lst** -- first five patients' progressive CDS, adult, sex, age, nanyrx30, nanyrx_6, clarktc.

Again, comparison against the .lsts (included with this package) will be helpful.

In contrast to the above, the final program in the sequence -- **profile.sas** -- produces a number of outputs in specially-tailored formats which provide the bulk of the analytical and follow-up capability of this package.

The significant outputs of **profile.sas** are these. They each have the run date in their names, so the user can easily distinguish test output from production output, as well as one production run from another.

validate_{today}.rtf

summary_{today}.rtf

logistic_{today}.txt

highlights_{today}.txt

profile_{today}.txt

patient_detail_{today}.xls

Using your own data

Identification of cardiac procedure patients.
- For the initial use of this program package, the time frame for case identification should begin with the earliest conveniently available data, beginning on January 1 or July 1, up through the latest half year (December 31 or June 30) for which data are complete.

- Subsequent analyses can be run at six-month intervals, comparing the most recent results against the initial baseline.

- Identify all members who had at least one of the procedure codes of interest (see SURGCODE below), on a date within the selected time frame, during a hospital stay. We have provided procedure codes which include coronary artery bypass graft (CABG) surgeries (invasive and minimally invasive) as well as non-invasive angioplasty procedures.

Extraction of data from source files.
- Please prepare each of the four datasets shown below.

- Extract them from your source data, converting to the variable names and values shown.

Procedure file
This file consists of one record for each CABG or angioplasty procedure. Multiple records per patient are permitted -- but may not be common.

Variable name	Description	Field Type & Format	Notes
File name: SURGERY			
HMOID	Actual member ID	Char(14)/Num(12)	As used by your HMO or plan.
PLANTYPE	Type of insurance plan which member belonged to on surgdate	Char (14)	Used to segment patient population. [Will also need to specify values for this variable in the PROFILE.SAS program.]
SURGINST	Hospital where procedure done	Char (14)	Any useful identifier
SURGDATE	Date of cardiac procedure	Num (8)	SAS date – if unknown, will use admitdat as surgdate
ADMITDAT	Date of admission to surgical stay	Num (8)	SAS date
DISCHDAT	Date of discharge	Num (8)	SAS date
DSTATUS	Discharge status	Char (1)	H=to home A=transfer to acute care facility E=transfer to extended care facility D=deceased.
SURGCODE	CABG procedure codes Angioplasty procedure codes	Char (5)	*ICD-9*: 36.10-36.19, 36.2 *ICD-9*: 35.00-35.04, 35.96, 36.01-36.07, 36.09, 39.66, 35.10-35.14, 35.20-35.28, 35.33, 39.61

Demographic file
One record for each patient in the SURGERY file.

File name: **DEMOG**			
Variable name	*Description*	*Field Type & Format*	*Notes*
HMOID	Actual member ID	Char(14)/Num(12)	Same as in SURGERY file
SEX	Member gender	Char (1)	M or F
DOB	Member date of birth	Num (8)	SAS date

<u>Pharmacy dispensing file</u>

All pharmacy dispensings for the patients in the SURGERY file, for the period six months prior to their SURGDATE, through at least one month afterward.

Each record represents one dispensing (new, or refill).

File name: DISP			
Variable name	*Description*	*Field Type & Format*	*Notes*
HMOID	Actual member ID	Char(14)/Num(12)	Same as in SURGERY file
GENERIC	Generic drug name	Char (30)	Special attention as indicative of staphylococcal infection: CEPHALEXIN, CEPHRADINE, CLINDAMYCIN, DICLOXACILLIN, VANCOMYCIN (please conform to above spellings).
NDC	National Drug Code	Char (11)	Must be a character field, filled with the eleven-digit NDC. If your pharmacy records use a numeric field, convert as: ndc=put(ndcnum,z11.);
AMT	Dispensed amount	Num (8)	Amount dispensed. Non-positive values will cause this record to be ignored.
RXDATE	Drug dispensing date	Num (8)	SAS date

<u>Utilization file</u>
This file covers both inpatient and outpatient activity (denoted by the flag variable IOREC), for each patient in the SURGERY file, for the period that runs from the admit date for the procedures of interest, to at least one month after the surgeries of interest.

Multiple records are permitted for any hospitalization or other inpatient stay, and for each outpatient encounter.

It is not necessary to purge this file of the claims that defined the surgeries of interest.

File name: CLAIMS			
Variable name	*Description*	*Field Type & Format*	*Notes*
HMOID	Actual member ID	Char(14) / Num(12)	Same as in SURGERY file
IOREC	Inpatient - Outpatient Flag	Char (1)	I=Inpatient stay O=Outpatient encounter
SERVINST	Type of service facility	Char (2)	HI=hospital inpatient HO=hospital outpatient ER=emergency dept OF=office visit EC=extended care HM=home care
SERVDATE	Date of service	Num (8)	SAS date
ADMITDAT	Date of admission (if IOREC=I)	Num (8)	SAS date
DISCHDAT	Date of discharge (if IOREC=I)	Num (8)	SAS date
DSTATUS	Discharge status (if IOREC=I)	Char (1)	H=to home A=transfer to acute care facility E=transfer to extended care facility D=deceased blank=still here.
NDX	Max. number of diagnoses	Num (8)	Number of elements in diagnosis vector
DX1 – DXn	Diagnosis vector	Char (6) ea.	ICD-9: xxx.xx
NPX	Max. number of procedures	Num (8)	Number of elements in procedure vector
PX1 – PXn	Procedure vector	Char (5) ea.	ICD-9: xx.xx CPT-4: xxxxx

Production run with your actual data
Extract datasets as described above.

Place them in the project subdirectory identified with the 'LIBNAME realdata' statement provided in each program header Then, switch from test mode to production mode, by activating the alternate %let run= statement in each program header:

> **As shipped: %let run=test;**
>
> **%*let run=real;**
>
> **Change to: %*let run=test;**
>
> **%let run=real;**

NOTE: By convention, many users segregate SAS programs from SAS datasets in subdirectories. If you wish to do that, merely indicate where the datasets are to be found, in the libname statements shown in each program header.

In each program there will also be one or more parameter-setting statements for you to customize. Each begins with **%let**.

For instance, in each program as supplied, this statement appears near the top:

> **%let user=Your_Health_Plan;**

To facilitate self-documentation of program output, please change the text string on the right of the = sign in the above statement to your own identity -- no blanks permitted.

Other parameters, specific to single programs, are noted below.

Task 1. Place the supplied program episodes.sas in your project directory. Edit its header section to reflect your local computer environment and actual directory names. Then run episodes.sas.

Task 2. Place the supplied program **qualify.sas** in your project directory, and edit its header section as in task 2. Run this program to screen patients' demographic data, to qualify patient episodes.

Two parameters are required to control this program. Their current settings are:

> %let early='01Jan2002'd; *set to first date of cardiac procedures wanted in this analysis ;
>
> %let late= '030Jun2004'd; *set to last date of cardiac procedures wanted in this analysis ;

As the comment suggests, change these date values to correspond to the time frame of the analysis you wish to run. Follow the indicated format while editing – these are, in fact, proper SAS date constants as shown.

Task 3. Edit the program header for the supplied program **detection.sas**, to show actual directory locations. Running this program will screen both the utilization (claims) and pharmacy (dispensing) datasets, for each qualified patient episode.

Place the supplied SAS dataset **formgen** in the data subdirectory created for this project. This is used as a lookup table for NDCs which correspond to anti-staphylococcal drugs, as markers for infection.

All indicators of infection within 30 days after the cardiac procedure date will be collected, summed up and used to compute a 'risk of infection' score[1], However, as a simplification in this application, any single indicator will be sufficient to establish an overall SSI infection indicator for this analysis.

Two parameters are required for this program. Current values:

> %let ndx=4;   * maximum number of diagnoses in any single claims record ;

> %let pdx=5;   * maximum number of procedures in any single claims record;

Please reset these two numeric values if necessary. They are also to be specified in the claims record extract -- see the Data Development Plan, CLAIMS file specification.

Task 4. Edit the program header for the supplied program **cdscore.sas .**

Also, place the supplied SAS datasets **model2** (which is small) and **cdsndcext** (which is fairly large) in the project data directory.

While a number of parameters -- besides %let user -- appear near the top of this program, none of them needs to be changed.

Running **cdscore.sas** will examine the pharmacy dispensings (for most drugs) in the six months prior to each cardiac procedure, to compute a chronic disease score for each patient, according to Clark's total cost model (reference 2).

Task 5. Edit the program header section of the supplied program **profile.sas** .

Run the program to produce cross-tabulations of infection indicators by other key predictors, over the whole time frame as well as by half-year.

These tabulations are intended to be reviewed by the user.

In particular, hospitals will be ranked into quartiles based on their patient infection indicator rate.

Additionally, a set of logistic regression analyses will be performed by profile.sas, to quantify the multivariate significance of predictors of infection.

Finally**, profile.sas** will optionally produce a patient-level Excel spreadsheet, to assist the user with specific follow-up as desired for hospitals in selected quartile(s) based upon the patient infection indicator rate.

---

[1] See Platt et al. (Emerg Infect Dis. 2002;8:1433-41, http://www.cdc.gov/ncidod/EID/vol8no12/02-0039.htm)

Four parameters are needed to control the operation of **profile.sas**. Their current values:

%let ref_prod=Commercial; * replace Commercial with the PLANTYPE value corresponding to your most populous plan membership among surgical (or all) patients. This allows the user to compare two groups of insurer plan types with one another (e.g. Commercial vs All Others). Alternatively, if there is one particular insurer plan of interest, one can select that type in lieu of the most populous plan. For example, the evaluation of cardiac procedure infections among Medicare patients may be of particular interest given their age and risk factors. Select the plan type of interest here. All other plan types will be compared to the plan type selected. If desired, insurer plans with limited enrollment can be excluded from the dataset.

%let mincount=40; * this establishes a threshold number of cardiac procedures during the entire timeframe analyzed, for retaining a hospital in the multivariate analysis. 40 is generally a reasonable setting for major plans.

%let rxthreshold=80; * this is a percentage value which is used to decide, in the program calculations, whether to keep all patients regardless of prescription (Rx) activity present in the DISP file, or to drop those who have no such activity. As set (80%), if more than one-fifth of patients have no Rx history, all will be retained and prescription data will be ignored. But if less than one-fifth lack Rx history, those so lacking will be dropped and prescription data will be retained. It is suggested that this parameter be used as is; but if a different value is needed, express it as a whole percentage -- that is, 75 is valid, while 66.7 is invalid.

%let listquart=4; * select the quartile for detailed patient output; Hospitals are ranked from low to high based upon their overall infection indication rate. 4 is the highest infection quartile. As set, this parameter will allow the program to produce a patient linelist (in Excel spreadsheet format) for all procedures performed at the highest quartile hospitals. Other possible values are: 34 (both quartile 3 and 4), 1234 (all hospitals), or 0 (none).

DATA VALIDATION: PROGRAMMER

1. Procedure dataset.

After creating the SURGERY dataset, run the first program provided in the production stream package: **episodes.sas**

While building procedure episodes, this program also provides ancillary validation statistics.

What to look for:

a. Are there ZERO observations in the SURGERY dataset? If so, check the format of your institution's ICD-9 procedure codes.

Perhaps the decimal point in the codes, by local convention, is not actually present in your claims data. Or, perhaps your procedure codes are defined as numeric instead of character.

See page 1 of the output **episodes.lst** for the actual procedure codes present in your procedure file.

b. Is there a reasonable distribution of the number of cardiac procedures across the SURGINST values?

Perhaps all or most of the procedures appear with a single SURGINST (or with blank SURGINST). Identification of individual hospitals is a critical part of the analysis, so this situation needs to be resolved.

Note that we restricted the length of the SURGINST variable to 18 characters, for economy in producing the output tables. Recode the hospital names, if necessary, to distinguish them within this restriction.

c. Procedure and admit dates.

What to look for: The range of dates is found in the proc univariate [see pp. 2-5 of **episodes.lst** output for the variables SURGDATE and ADMITDAT-- not as formatted date values, but as five-digit numbers which represent internal SAS date values. Look at the extreme values at the end of the univariate output, and compare them with the time horizon you intend to analyze.

The following table lists half-year start and end dates, in SAS date format and unformatted, to facilitate that comparison.

'01Jan2002'd	15341	'30Jun2002'd	15521
'01Jul2002'd	15522	'31Dec2002'd	15705
'01Jan2003'd	15706	'30Jun2003'd	15886

'01Jul2003'd	15887	'31Dec2003'd	16070
'01Jan2004'd	16071	'30Jun2004'd	16252
'01Jul2004'd	16253	'31Dec2004'd	16436
'01Jan2005'd	16437	'30Jun2005'd	16617
'01Jul2005'd	16618	'31Dec2005'd	16801
'01Jan2006'd	16802	'30Jun2006'd	16892
'01Jul2006'd	16983	'31Dec2006'd	17166
'01Jan2007'd	17167	'30Jun2007'd	17347
'01Jul2007'd	17348	'31Dec2007'd	17531
'01Jan2008'd	17532	'30Jun2008'd	17713
'01Jul2008'd	17714	'31Dec2008'd	17897

If these procedure dates are narrower than the time span you expected, you may need to reextract the SURGERY dataset.

However, if they are broader than the timespan you want to analyze, merely set the appropriate parameters in the second program, **qualify.sas --** see the two **%let** statements which apply a date range**.**

If procedure date is missing but admit date is present for any observation, the program episodes.sas will use the admit date as the procedure date -- a reasonable assumption for most cardiac procedures, as we have determined in our earlier work.

Discharge dates are less critical; if not available, the program stream will still work reasonably well.

d. Discharge status.

The first program in our analysis suite, episodes.sas, will exclude patients who do not survive the procedure, as indicated by the value **'D'** in the DSTATUS field. There may be a very few, or none, of these in your claims data, but do check the DSTATUS codes present to see that they conform to the specifications for the SURGERY dataset.

[See the table of discharge codes on page 19 (approximately) of your output **episodes.lst**]

Fix: Recode the discharge status field as intended, in your data re-extraction.

2. Demographic Files.

Patients with unknown gender or date of birth, or who are under age 18, will be dropped at some point in the program stream.

The complete distribution of patient sex, and of age, is shown on pp. 3-4 of the output **qualify.lst**

Those who pass the age screening as adults are shown on page 7 of the output **cdscore.lst**

Claims based indicators of cardiac surgical site infection 03-09-05.doc

What to look for: If too few patients (or none) pass the age screening, check to be sure that you have coded patient date of birth in the SURGERY file as an internal SAS date. Recall that SAS uses the internal value 0 to represent 01Jan1960, so older patients' dates of birth may be carried internally as negative integers.

3. Pharmacy Files.

a. Generic field -- only the five indicated anti-staphylococcal antibiotics are used as indicators of infection. The detection.sas program will recognize these dispensings either by NDC, or

by their generic name if the latter conform to these spellings (all upper case):

CEPHALEXIN, CEPHRADINE, CLINDAMYCIN, DICLOXACILLIN, VANCOMYCIN.

All other generics may be spelled *ad lib*.

b. NDC -- these values are used in a lookup table to compute the chronic disease score for the patient, during the six month pre-operative period. They are also used as a double-check on the presence of anti-staphylococcal drugs in the post-operative period.

What to look for: Be sure the NDC is expressed as an 11-character field, even though the only legitimate characters in the field will be numerals. Also, leading zeros are significant in the NDC field. Be sure that generic names, if used, are in all upper case.

c. AMT -- only prescriptions with positive amounts are considered.

d. RXDATE -- to be useful as indicators of post-op infections the anti-staphylcoccal drugs must be prescribed in the immediate post-procedure period within 30 days following the cardiac procedure.   See the distribution of all RXDATEs (both pre- and post-op) in the univariate output **detection.lst**, on pp 4-5.

4. Utilization files.

a. IOREC -- not absolutely critical, but helps to localize the type of infection indicator found, and to characterize the type of post-procedure follow-up.

b. NDX -- the maximum number of diagnosis codes to be found in any claims record.

c. NPX -- the maximum number of procedure codes to be found in any claims record.

d. SERVDATE -- again, to be useful in detecting indications of post-procedure infection, these dates must closely approximate the procedure dates. Check the distribution of servdate in the univariate output **detection.lst**, on pp 2-3.

DATA VALIDATION: ANALYST

Review your program output to determine whether the hospital-specific profiles are likely to be informative. At a minimum, these eleven characteristics should be checked. Others may also be necessary, depending on the specifics of a health plan's data.

1) Is there a sufficient number of cardiac procedures within the requested time period to warrant an analysis?

2) Is there a sufficient number of invasive and non-invasive cardiac procedures to warrant a stratified analysis?

3) Is there a sufficient number of hospitals and a sufficient number of procedures per hospital?

4) Is there an adequate number of procedures related to each coverage type (e.g. Commercial versus Other)?

5) Are prescription data included in this analysis?

6) Are there reasonable numbers of infection indicators in the dataset?

7) Can I see the data by procedure type, plan type, hospital, and infection indicator?

8) Among hospitals meeting the institutional threshold, are the patients of similar age?

9) Among hospitals meeting the institutional threshold, do patients have similar chronic disease scores?

10) Are the data sufficient to produce multivariate analyses?

11) Can I see a line list of individual procedures and their associated infection indicators for hospitals with higher than expected infection scores?

In the text below, we illustrate this checking procedure in the supplied output files ("validate_test" and "summary_test") which correspond to the test dataset. These output files are produced by **profile.sas.**

1) **Is there a sufficient number of cardiac procedures within the requested time period to warrant an analysis?**

   BEFORE AN INSTITUTIONAL THRESHOLD IS APPLIED

   In the provided example validation file, "validate_test", examine page 3, Table of Yearhalf by CardiacProcedureType. Look at the column marked "Total" to determine if the number of episodes in each half year is adequate for analysis. The number of procedures per half year needed to pursue an adequate analysis will depend on the number of hospitals being evaluated (see #3 below). In general, when comparing two half-years, there should be at least 500 total procedures, with no restriction on the minimum number per half-year. This criterion of 500 total procedures is not met in this test dataset. We provide this test dataset as an illustration of problematic data.

If numbers are adequate, determine if the numbers are similar across half years, and whether the numbers are sufficient for a meaningful analysis. In the provided example, total numbers of procedures appear similar in the first 4 half years. However, the first half year of 2004 (2004H1) has only 55 episodes (see highlighted output below). Some explanations for this sharp decrease in number compared to the prior 4 half years include the fact that the 2004H1 half year may not be complete (data still accruing), health plan coverage for these cardiac procedures may have changed after 2003H2, new modalities in cardiac care after 2003H2 may have reduced the frequency of these procedures, or there is a problem with the data pull. In evaluating the data pull, consider evaluating whether the requested time period is correct and whether any of the ICD9 codes have zero procedures (see Data Validation: Programmer above). After double checking the data pull, one could determine that the time period requested should not include data from 2004H1 since there are insufficient numbers for analysis. The programs would then need to be re-run based upon a more restricted time period.

Table of yearhalf by CardiacProcedureType			
yearhalf	CardiacProcedureType		
Frequency	Invasive	Noninvasive	Total
2002H1	45	132	177
2002H2	49	132	181
2003H1	48	129	177
2003H2	37	144	181
2004H1	18	37	55
Total	197	574	771

AFTER AN INSTITUTIONAL THRESHOLD IS APPLIED

As mentioned in Task 5 of the Program Plan, a minimum number of procedures is required for a hospital to be entered into the multivariate analysis of predictors of infection indicators. Page 9 provides the Table of Yearhalf by CardiacProcedureType after the institutional threshold (default = 40) has been applied. The above validation steps may be repeated for this table.

Table of yearhalf by CardiacProcedureType			
yearhalf	CardiacProcedureType		
Frequency	Invasive	Noninvasive	Total
2002H1	34	106	140
2002H2	26	65	91
2003H1	27	61	88
2003H2	31	120	151
2004H1	14	33	47
Total	132	385	517

2) Is there a sufficient number of invasive and non-invasive cardiac procedures to warrant a stratified analysis?

BEFORE AN INSTITUTIONAL THRESHOLD IS APPLIED

Examine the same table - page 3, Table of Yearhalf by CardiacProcedureType. Look at the columns marked "Invasive" and "Noninvasive" to determine if the number of each type of episodes in each half year is adequate for analysis. Determine if the numbers are similar across half years, and whether the numbers are sufficient for a meaningful analysis. In the example provided, the number of invasive and noninvasive procedures per half year is similar for the first 4 half years. However, 2004H1 remains an outlier for both subtypes with 18 invasive procedures and 37 non-invasive procedures. In addition, the small number of invasive procedures in the first 4 half years (range 37-49) makes it unlikely that an analysis intended to identify hospitals with high infection indicators will be possible. See highlighted output below. Check the adequacy of the data pull (see Data Validation: Programmer above). The number of procedures per half year needed to pursue an adequate analysis will depend on the number of hospitals being evaluated (see #3). In general, there should be at least 250 total invasive and 250 total non-invasive procedures when comparing two half-years in order to justify a stratified analysis. If there is an adequate number for one procedure-type, but not the other, then the programs can be run with selective attention given to results pertaining to that procedure type.

Table of yearhalf by CardiacProcedureType			
yearhalf	CardiacProcedureType		
Frequency	Invasive	Noninvasive	Total
2002H1	45	132	177
2002H2	49	132	181
2003H1	48	129	177
2003H2	37	144	181
2004H1	18	37	55
Total	197	574	771

AFTER AN INSTITUTIONAL THRESHOLD IS APPLIED

Page 9 provides the Table of Yearhalf by CardiacProcedureType after the institutional threshold (default = 40) has been applied. The above validation steps may be repeated for this table.

3) Is there a sufficient number of hospitals and a sufficient number of procedures per hospital?

BEFORE AN INSTITUTIONAL THRESHOLD IS APPLIED

Examine the table on p4, Table of LeadInst by CardiacProcedureType. Look at the "Total" column. In general, there should be at least 10 hospitals with at least 40 total procedures. In the example provided, 10 hospitals (A, B, C, D, J, K, L, M, N, and T) meet this criterion (see highlighted output below). For an adequate subanalysis, there should be at least 10 hospitals with at least 40 procedures of each subtype. In the example provided, neither subtype meets this criterion.

Table of leadinst by CardiacProcedureType			
leadinst(Surgical institution for lead procedure) Frequency	CardiacProcedureType		
	Invasive	Noninvasive	Total
Hospital_A	13	51	64
Hospital_B	25	59	84
Hospital_C	8	37	45
Hospital_D	12	31	43
Hospital_E	14	25	39
Hospital_F	7	24	31
Hospital_G	5	23	28
Hospital_H	11	22	33
Hospital_I	6	14	20
Hospital_J	15	32	47
Hospital_K	9	36	45
Hospital_L	14	46	60
Hospital_M	12	29	41
Hospital_N	10	30	40
Hospital_O	8	22	30
Hospital_P	5	23	28
Hospital_Q	5	20	25
Hospital_S	4	16	20
Hospital_T	14	34	48
Total	197	574	771

AFTER AN INSTITUTIONAL THRESHOLD IS APPLIED

Page 10 provides the Table of LeadInst by CardiacProcedureType after the institutional threshold (default = 40) has been applied.

Table of leadinst by CardiacProcedureType			
leadinst(Surgical institution for lead procedure)	CardiacProcedureType		
Frequency	Invasive	Noninvasive	Total
Hospital_A	13	51	64
Hospital_B	25	59	84
Hospital_C	8	37	45
Hospital_D	12	31	43
Hospital_J	15	32	47
Hospital_K	9	36	45
Hospital_L	14	46	60
Hospital_M	12	29	41
Hospital_N	10	30	40
Hospital_T	14	34	48
Total	132	385	517

4) Is there an adequate number of procedures related to each coverage type (e.g. Commercial versus Other)?

BEFORE AN INSTITUTIONAL THRESHOLD IS APPLIED

Examine page 5, Table of Plantype by CardiacProcedureType. Look at the column marked "Total" and determine if the number of procedures related to patients with "Commercial" insurance or "Other" insurance types is adequate for analysis. In general, there should be at least 100 procedures of each type in order to evaluate this predictor. Similarly, for subanalyses limited to invasive or non-invasive procedures, there should be at least 100 "Commercial" and 100 "Other" procedures within each category. In this example, there appears to be approximately this number of invasive cardiac procedures from "Commercial" (101) and "Other" insurer type (96) patients to enable an assessment of whether plan types predict infection indicators for invasive cardiac procedures.

Table of plantype by CardiacProcedureType			
plantype(Insurance product)	CardiacProcedureType		
Frequency	Invasive	Noninvasive	Total
Commercial	101	285	386
Other	96	289	385
Total	197	574	771

Also examine page 5, Table of Plantype by Yearhalf. Within each yearhalf, look for zero cells, which would be one indicator of an inadequate number of procedures to assess plantype as a predictor. Notice in this example that the number of patients in each plan is very small for 2004H1 (28 Commercial and 27 for Other insurer types). This should warrant an evaluation of the adequacy of the data pull for 2004H1 (see above) and the potential for excluding that half year from the analysis.

Table of plantype by yearhalf						
plantype(Insurance product)	yearhalf					
Frequency	2002H1	2002H2	2003H1	2003H2	2004H1	Total
Commercial	89	89	90	90	28	386
Other	88	92	87	91	27	385
Total	177	181	177	181	55	771

AFTER AN INSTITUTIONAL THRESHOLD IS APPLIED

Page 11 provides the Table of Plantype by CardiacProcedureType and Table of Plantype by Yearhalf after the institutional threshold (default = 40) has been applied. The above validation steps may be repeated for this table

5) **Are prescription data included in this analysis?**
Examine page 6 of the example validation output. The title of the first table provides information to the analyst about whether or not prescription data was included in the analysis. In a previous programming step (see Program Plan, Task 5) a threshold percent of prescription activity is used to determine whether to keep all patients regardless of peri-procedure prescription (rx) activity, or to drop those who have no such activity. As set, if more than 80% of patients have prescription activity, then those lacking prescription activity will be dropped. If less than 80% of patients have prescription history, then all patients will be retained, but prescription data will be ignored).

Look at the title of the first table on page 6. In this example, the table is entitled "Table of Yearhalf by Indicator_Using_Rx" (see highlighted output below). This title tells the analyst that prescription data was included in the analysis and patients lacking prescription activity were dropped. If prescription data were ignored, then the title would read "Table of Yearhalf by Indicator_Ignore_Rx."

Table of Indicator_Using_Rx by plantype			
Indicator_Using_Rx	plantype(Insurance product)		
Frequency	Commercial	Other	Total
no	363	365	728
yes	23	20	43
Total	386	385	771

6) **Are there reasonable numbers of infection indicators in the dataset?**

BEFORE AN INSTITUTIONAL THRESHOLD IS APPLIED

Examine page 6, Table of Yearhalf by Indicator_Using_Rx. Look at the column labeled "yes". This column shows the number of infection indicators identified by claims data within the dataset. Infection indicators are provided by half year. In the test dataset provided, 5 of 55 procedures have an infection indicator in 2004H1. Previous analyses have suggested that, on average, claims data will identify 15-20% of procedures as having an infection indicator. This percentage will decrease if prescription data are not included in the analysis (see #5). In this test dataset, 6% (43/771) of total procedures have an associated infection indicator. Given the lower-than-expected percent of infection indicators, an evaluation of each type of infection indicator is warranted to evaluate whether one of the indicators may have no counts, potentially suggesting a data pull problem.

Table of yearhalf by Indicator_Using_Rx			
yearhalf	Indicator_Using_Rx		
Frequency	no	yes	Total
2002H1	163	14	177
2002H2	177	4	181
2003H1	168	9	177
2003H2	170	11	181
2004H1	50	5	55
Total	728	43	771

To evaluate each type of infection indicator, examine the series of crosstabs following the Table of Yearhalf by Indicator_Using_Rx. Crosstabs include Inpatient Indicator, Indicator During Procedure Stay, Indicator During Rehospitalization, Outpatient Indicator, ED Indicator, Clinic Indicator, Care of Wound, Lab Test for Infection, and Antistaph Abx Dispensed. For example, in the first crosstab provided, 4 procedures have an inpatient infection indicator ("Yes") and 767 do not ("No"). Evaluate the crosstabs for unexpectedly high or low percentages of indicators which may suggest a data pull error. In the example validation output file, there are no infection indicators from the admission involving the procedure or ED visits). If prescription data are not included, then Antistaph Abx Dispensed should not have any infection indicators.

Inpatient Indicator		
inpatient	Frequency	Percent
no	767	99.48
yes	4	0.52

AFTER AN INSTITUTIONAL THRESHOLD IS APPLIED

Page 12-14 provides the above tables after the institutional threshold (default = 40) has been applied. The above validation steps may be repeated for this table

7) **Can I see the data by procedure type, plan type, hospital, and infection indicator?**
Yes, we have provided stratified tables for those who wish to see this level of detail. Tables reflect data after application of the institutional threshold. Note: Non-invasive cardiac procedures involve angioplasty procedures. Invasive cardiac procedures include both invasive and minimally invasive coronary artery bypass graft (CABG) surgeries. Alterations of ICD9 codes can be made programmatically to change the procedures of interest or to narrow the evaluation (e.g. separate minimally invasive CABGs from traditional CABGs). Such changes, while feasible, are not supported by this manual.

P16-17: Commercial patients, procedures with any infection indicator

Episodes with any infection indicator		All hosps.		Hospital_A		Hospital_B	
		N	PctN	N	PctN	N	PctN
CardiacProcedureType	Indicator_Using_Rx						
Invasive	no	63	92.6	7	87.5	11	100.0
	yes	5	7.4	1	12.5	.	0.0
Noninvasive	no	180	94.2	22	91.7	27	84.4
	yes	11	5.8	2	8.3	5	15.6
Both	no	243	93.8	29	90.6	38	88.4
	yes	16	6.2	3	9.4	5	11.6

P18: Commercial patients, procedures with any inpatient indicator
(note, in this test dataset, there are no inpatient indicators.

P19-20: Commercial patients, procedures with any outpatient indicator

P21-22: Other insurer type patients, procedures with any infection indicator

P23-24: Other insurer type patients, procedures with an inpatient indicator

P25-26: Other insurer type patients, procedures with an outpatient indicator

The above tables are also provided by half year

P27-34: Commercial patients, procedures with any infection indicator

P35-42: Commercial patients, procedures with any inpatient indicator

P43-50: Commercial patients, procedures with an outpatient indicator

P51-60: Other insurer type patients, procedures with any infection indicator

P61-70: Other insurer type patients, procedures with an inpatient indicator

P71-80: Other insurer type patients, procedures with an outpatient indicator

8) **Among hospitals meeting the institutional threshold, are the patients of similar age?**

Pages 87-89 provide the mean age of patients with Commercial plans by invasive or non-invasive procedure, hospital, and presence of an infection indicator. For example, there are 8 patients with commercial plans who had invasive cardiac procedures performed at Hospital A. Only one of these patients had an infection indicator identified by claims data. The patients without an infection indicator had a mean age of 63.4 years and the patient with an infection indicator was 63 years old (see highlighted output below). Evaluate whether patients are missing age data which may suggest a data pull problem. Also evaluate whether the mean age of patients undergoing invasive versus non-invasive procedures is similar across hospitals. Large differences in the mean age of patients across hospitals may account for some differences in the percent of patients with infection indicators following cardiac

procedures. In this example dataset, there are insufficient numbers of patients to determine the mean age in several hospitals.

		All hosps.		Hospital_A		Hospital_B	
Mean Age of Patient		Pt. age at surgery		Pt. age at surgery		Pt. age at surgery	
		N	Mean	N	Mean	N	Mean
CardiacProcedureType	Indicator_Using_Rx						
Invasive	no	63	64.2	7	63.4	11	59.4
	yes	5	52.8	1	63.0	.	.
Noninvasive	no	180	60.6	22	60.5	27	61.1
	yes	11	59.0	2	46.0	5	59.8
Both	no	243	61.5	29	61.2	38	60.6
	yes	16	57.1	3	51.7	5	59.8

Pages 90-92 provide similar data for patients with "Other" insurer type plans.

9) Among hospitals meeting the institutional threshold, do patients have similar chronic disease scores?

Pages 93-95 provide the mean chronic disease score (CDS) of patients with Commercial plans by invasive or non-invasive procedure, hospital, and presence of an infection indicator. Evaluate whether hospitals are missing CDS data or have an unusually high or low mean CDS which may suggest a data pull problem. Alternatively, large differences in the mean CDS of patients across hospitals may account for some differences in the percent of patients with infection indicators following cardiac procedures. In this example dataset, there are insufficient numbers of patients to determine the mean CDS in several hospitals.

Pages 96-98 provide similar data for patients with "Other" insurer plans.

Pages 99-100 provide CDS data using a cutpoint of 4500 for patients with Commercial plans.

Pages 101-102 provide CDS data using a cutpoint of 4500 for patients with "Other" insurer plans.

10) Are the data sufficient to produce multivariate analyses?

Examine the final results in the file called "summary_test." There are 3 descriptive tables and 3 multivariate analyses that are applied to data from hospitals meeting the institutional threshold number of procedures. The first includes both invasive and non-invasive procedures (descriptive table p2 (see output below), multivariate analysis p5-7).

leadinst	quartile	nepisodes	indicator	indic_pct	inpt_pct	inhosp_pct
Hospital_N	1	40	0	0.0	0.0	0.0
Hospital_J	1	47	1	2.1	0.0	0.0
Hospital_D	2	43	1	2.3	0.0	0.0
Hospital_T	2	48	2	4.2	0.0	0.0
Hospital_M	2	41	2	4.9	0.0	0.0
Hospital_L	3	60	3	5.0	0.0	0.0
Hospital_B	3	84	8	9.5	1.2	0.0
Hospital_A	3	64	7	10.9	1.6	0.0
Hospital_C	4	45	5	11.1	2.2	0.0
Hospital_K	4	45	5	11.1	0.0	0.0
		517	34			

The second is limited to invasive procedures (descriptive table p3, multivariate analysis p8-10), and third is limited to non-invasive procedures (descriptive table p4, multivariate analysis p11-13).

Look for warning labels that indicate the model did not run. Warning labels will appear in the table labeled "Model Convergence Status" as well as at the top of the page in which the multivariate results appear. In the test dataset, these warning labels are seen in the analyses limited to either invasive or non-invasive procedures. For example, in the analysis of invasive procedures, the following statement appears– "Quasi-complete separation of data points detected," – in the box labeled "Model Convergence Status" on page 9. In addition, the warning label - "The validity of the model fit is questionable" appears at the top of p10.

Your_Health_Plan

Profile Analysis
Logistic Model #1A
Invasive cardiac procedures only
Any infection indicator, including prescription data

Class Level Information			
Class	Value	Design Variables	
	0:Low quartile	0 0 0 0 0 0 0 0 1	
sex	M	1 0	
	F	0 1	

Model Convergence Status
Quasi-complete separation of data points detected.

These warnings suggest that the data are insufficient to generate reliable results. The most common reason for insufficiency is small numbers of procedures, particularly when stratified by the desired predictors. In the case of the example provided, analyses are reliable only for the total dataset (invasive and non-invasive combined). In the event of a model that will not run, re-evaluation of the choice of study period, insurer types, and institutional threshold number of procedures may help to increase the number of procedures being analyzed.

11) Can I see a line list of individual procedures and their associated infection indicators for hospitals with higher than expected infection scores?

Yes. The programs provide the user with the option to generate an Excel file called "patient_detail_(date)." In specifying the contents of this file, the user inputs the desired hospitals to be displayed according to quartiles of infection indicators. For example, "4" selects all hospital with the highest quartile of infection indicators. For relevant hospitals, a line list will be displayed by cardiac episode, with provision of patient identifiers, procedure types and dates, demographic data, and type and date of infection indicator flag. An example line list representing three patients with four infection indicator flags is provided below (see file "patient_detail_test"). These patients all had their procedures performed at Hospital A, a hospital with an infection indicator rate in the third highest quartile. The first line item shows that a 45 year old female patient with Commercial insurance underwent a non-invasive cardiac procedure (angioplasty) at Hospital A on 3-11-2002. Within 30 days from the procedure, this patient had an indicator flag for wound care which was based upon an outpatient claim on 3-26-02. Patients with more than one infection indicator flag appear more than once.

Infection quartile	Procedure institution	Patient ID	Patient gender	Patient age	Patient birthdate	Insurance product
3	Hospital_A	P600069COM	F	45	12/19/1956	Commercial
3	Hospital_A	P600069COM	F	47	12/19/1956	Commercial
3	Hospital_A	P600085COM	F	63	2/9/1939	Commercial
3	Hospital_A	P600086OTH	F	72	10/22/1929	Other

Cardiac procedure type	Main cardiac procedure	Pt. episode #	Admit date for surgery	Procedure date	End of follow-up	Discharge status	Inpt/Outpt claim	Claims source
Noninvasive	39.66	1	3/11/2002	3/11/2002	4/10/2002	to acute care	Outpt.	Claims
Noninvasive	39.66	4	3/10/2004	3/10/2004	4/9/2004	to acute care	Outpt.	Claims
Invasive	36.2	1	3/27/2002	3/27/2002	4/26/2002	to acute care	Outpt.	Claims
Noninvasive	35.27	1	3/28/2002	3/28/2002	4/27/2002	to extended care	Inpt.	Claims

Pre-disch. flag	Rehospitalized	Any follow-up diag.	Inpt. flag	ED flag	Ambulatory flag	Lab test	Wound care	Antistaph abx
0	0	1	0	0	0	0	1	0
0	0	1	0	0	0	0	1	0
0	0	1	0	0	0	0	1	0
0	0	1	0	0	0	0	1	0

Type of facility	Flagged diagnosis	Flagged procedure	Antistaph antibiotic dispensed	Service date	Days after admit	Days after discharge
hosp inpt		13160		3/26/2002	15	8
hosp inpt		13160		3/25/2004	15	8
hosp inpt		11001		4/11/2002	15	7
hosp inpt		11042		4/12/2002	15	7

This linelist enables the user to review specific patients with individual hospitals when this is desired, for instance if a hospital has an unusually large number of patients with one of these indicators. Furthermore, it provides specific data related to episodes with flagged infection indicators to allow a focused investigation on whether these episodes represent true infection. Remember that these claims based results are surrogate markers of infection. On average, approximately half of the patients with one of these indicators has the infection confirmed when the medical record is reviewed.

INTERPRETING AND USING THE RESULTS

These programs allow health plans to identify hospitals whose cardiac procedures have a high association with infection indicators. Final summary results are provided for hospitals meeting the threshold number of procedures (default = 40).

First, a descriptive table is provided which details the percent of cardiac procedures associated with infection indicators in each hospital. Hospitals are ranked into quartiles based upon their percent infection indicators (1= lowest percent infection indicators, and 4 = highest percent infection indicators). In the test dataset provided, hospitals C and K are in the highest quartile for percent infection indicators (Quartile =4) for all cardiac procedures, including both invasive procedures and non-invasive procedures. ("summary_test" file). This ranking provides an indication of which hospitals may have elevated rates of surgical site infections following cardiac procedures. Based upon these rankings, further assessment of Quartile 4 hospitals could reasonably be pursued.

Since quartile rankings do not control for other predictors or confounders of infection indicators, we also provide a multivariate analysis that simultaneously evaluates predictors of infection indicators. Variables assessed include hospitals serving health plan members, along with sex, chronic disease score, plan type (Commercial versus Other insurer types), and procedure type (invasive versus non-invasive). Two additional multivariate analyses are provided which limit the model to invasive and non-invasive procedures.

Each multivariate analysis provides odds ratios for the prediction of infection indicators. For example, in the test dataset, the multivariate analysis on page 7 ("summary_test" file) evaluates the prediction of infection indicators among invasive and non-invasive procedures combined. Hospitals are assigned an odd ratio (OR) describing their risk of infection indicators among plan members undergoing cardiac procedures as compared to hospitals with infection indicator rates in the lowest quartile. Using the column labeled "Point Estimate", Hospital T has an OR of 3.6 while Hospital B has an OR of 5.4. Using the columns labeled "95% Wald Confidence Limits," one can determine which odds ratio estimates are significant (p<0.05) by determining whether their confidence limits do not cross 1.0. Using this criterion, we find that Hospitals K and C have infection indicators which are significantly higher than hospitals in the lowest quartile. Additionally, while gender is not a significant predictor of an infection indicator, a chronic disease score >= 4,500 confers a significantly lower risk of an infection indicator (note: test data does not mimic real data). For both medical and financial reasons, it is of interest why certain hospitals have a significantly higher rate of infection indicators. Since our infection indicators are based upon claims data, they are surrogate, not definitive, markers of infection. Some reasons that a high rate of these surgical site infections might be misleading include:

- an anomaly in the claims data from one or more institutions,

- unusual coding practices by some providers,

- an insufficient number of events to be informative,

- an unusual case mix that includes a disproportionate number of high risk patients.

Given the surrogate nature of our infection indicators, our rates cannot be directly compared to hospital measures of infection rates. On average, approximately half of the patients with one of

these indicators has the infection confirmed when the medical record is reviewed. We thus recommend using a high rate of infection indicators to suggest a high rate of actual infection. In general, we suggest that attention be directed to hospitals which a) meet the threshold number of procedures (default – 40), b) are in the top quartile of percent infection indicators (Quartile = 4 in the descriptive tables), and c) have a significantly higher rate of infection indicators compared to lowest quartile hospitals in the multivariate analysis. Additional inquiry into these hospitals to confirm a high rate of infection can then be undertaken. We provide a line list of cardiac procedures associated with an infection indicator flag based upon claims data such that further inquiry and evaluation at the hospital level can be made.

In addition to identification of hospitals likely to have higher infection rates, these programs allow automated serial evaluation of infection indicators in these hospitals can provide evidence of persistently high infection rates or evidence of improvement over time.