Sifting through Messy Data with SAS®: Creating Kaiser Permanente's Non-Formulary Report

Beth Worrall, Kaiser Permanente, Denver, CO

ABSTRACT

Analysts are constantly challenged to present easily understood results stemming from a variety of distinct raw data sources. The Pharmacy Department at Kaiser Permanente was interested in a report which evaluated costs and prescription information for medications not included on the department's formulary. The challenge was to link variables from text and Excel files using SAS® datasets, analyze cost and prescription information, and convert analysis results into a usable format for providers. This presentation will first illustrate how to convert data from text and Excel files to SAS® datasets. It will then show how the resulting datasets were analyzed and formatted using PROC MEANS, arrays, and the null data step to produce a final report for providers.

BACKGROUND

The pharmacy department at Kaiser Permanente issues a formulary specifying medications which are covered for a co-payment. Medications not on this formulary may be prescribed, but in order for them to be covered, a non-formulary process must be followed. When a non-formulary prescription is submitted to the pharmacy, an exception code must be filled indicating the reason for the prescription. Examples of exception codes are:

- form filed—a form is filed by the prescribing physician and approved by the facility's physician-in-charge
- specialty initiated—the prescribed medication is allowed to be prescribed by certain specialty physicians
- patient pays cash—the patient has elected to pay cash for a medication not covered under the formulary.

Due to the organization's desire to uncover variation in physician practices and pharmacy processes, interest grew in the generation of a new report which explored non-formulary medication utilization within the organization. Data existed in two locations: a text file pulled from the Pharmacy Information Management System (PIMS) containing all prescriptions and an Excel file specifying formulary medications to be excluded from the report. These files were converted to SAS® datasets, merged, and analyzed. The resulting report included, for each medication, the total cost, total number of prescriptions, and number of prescriptions for each exception code. The information was grouped by pharmacy location.

INPUTTING TEXT AND EXCEL FILES

Data were converted to SAS® datasets from a text file and an Excel file. The datasets were then merged to exclude certain medications from subsequent analysis. Appendix 1 shows 5 observations from the merged dataset MEDS1.

- Text file

```
FILENAME TEXT
  'C:\MYDOCU~1\SASDATA\NONKC.TXT';
DATA TEXT;
LENGTH  PROD_NM $25 PHARMACY $16
            USR_COST 8 FORMEX  8;
  INFILE TEXT DLM=';';
  INPUT PROD_NM $ PHARMACY $ USR_COST
        FORMEX
  ;
  PROC SORT;
    BY PROD_NM;
  RUN;
```

- Excel file

```
FILENAME EXCLUDE
  DDE 'EXCEL|PRODUCTS!R3C1:R460C1';
DATA EXCLUDE;
  LENGTH PROD_NM $25;
  INFILE EXCLUDE
    MISSOVER DLM='09'X NOTAB DSD;
  INPUT PROD_NM $
  ;
PROC SORT;
  BY PROD_NM;
RUN;
```

- Merged dataset

```
DATA MEDS1;
  MERGE TEXT(IN=IN1) EXCLUDE(IN=IN2);
  BY PROD_NM;
  IF (IN1=1) AND (IN2=0);
RUN;
```

SUMMARIZING THE DATA

Total cost and number of prescriptions were then calculated, by medication and pharmacy. The utilization of a macro simplified this calculation. The dataset COST_MED totaled cost and number of prescriptions by medication while the dataset COST_PH totaled the same by pharmacy. The two datasets were then merged to create a dataset containing cost and number of prescriptions by pharmacy alone and by medication name within each pharmacy. The first 5 observations for each dataset are shown in Appendix 2.

```
%MACRO COSTS(VAR1,VAR2,DATA1,
                    COST,COUNT);
  PROC SORT DATA=MEDS1;
    BY &VAR1 &VAR2;
  PROC MEANS DATA=MEDS1 SUM NOPRINT;
    VAR USR_COST;
    BY &VAR1 &VAR2;
    OUTPUT OUT=&DATA1
              SUM=&COST N=&COUNT;
  PROC SORT;
    BY PHARMACY;
  RUN;
%MEND COSTS;
%COSTS(PROD_NM,PHARMACY,COST_MED,
           COST_MED, COUNTMED)
%COSTS(PHARMACY,PHARMACY,COST_PH,
           COST_PH,COUNT_PH)

DATA MEDS2;
  MERGE COST_MED(IN=IN1) COST_PH (IN=IN2);
   BY PHARMACY;
  IF (IN1=IN2=1);
RUN;
```

USING ARRAYS

A more detailed breakdown of the number of prescriptions was then added to the dataset. The breakdown specified the number of prescriptions for each exception code, by medication and pharmacy. The following steps lead to the detailed breakdown.

- The number of prescriptions for each medication by pharmacy and exception code was calculated using PROC FREQ. Ten observations from the resulting dataset TRANSMED are shown in Appendix 3(a).

```
PROC SORT DATA=MEDS1;
  BY PHARMACY FORMEX;
PROC FREQ DATA=MEDS1 NOPRINT;
  TABLES PROD_NM/OUT=TRANSMED;
  BY PHARMACY FORMEX;
PROC SORT;
  BY PHARMACY PROD_NM;
RUN;
```

- Exception codes and their counts were transposed so there was one observation per medication within each pharmacy. The dataset BOTH1 contains transposed variables *formex* which indicates the exception code and *count* which equals the number of prescriptions for that exception code. Appendix 3(b) displays 5 observations from the dataset BOTH1.

```
%MACRO TRANS(VAR1);
  PROC TRANSPOSE DATA=TRANSMED
        OUT=&VAR1 PREFIX=&VAR1;
    VAR &VAR1;
    BY PHARMACY PROD_NM;
%MEND TRANS;
%TRANS(FORMEX)
%TRANS(COUNT)

DATA BOTH1;
  MERGE FORMEX COUNT;
  BY PHARMACY PROD_NM;
RUN;
```

- More variables were then created so all counts for exception 1 codes had the same name (C1), all counts for exception 2 codes had the same name (C2), and so on. This was accomplished with arrays. First, two arrays were created: FORMEX specified the exception code and COUNT contained the count for that exception code. For each exception code (FORMEX1-FORMEX5) and corresponding count (COUNT1-COUNT5), the corresponding count for a "1" exception code was to be mapped to C1, the count for a "2" exception code was to be mapped to C2, etc. In order to establish these mappings, another array, C1-C10 (there are 10 possible exception codes) was created.

  Example: If FORMEX1=5 and COUNT1=15 then C5=15. This is shown in Appendix 3(c), which contains 10 observations from BOTH2.

```
DATA BOTH2;
  SET BOTH1;
  BY PHARMACY PROD_NM;
  ARRAY FORMEX FORMEX1-FORMEX5;
  ARRAY COUNT COUNT1-COUNT5;
  DO I=1 TO 5;
    ARRAY C C1-C10;
    DO J=1 TO 10;
      IF FORMEX(I)=J THEN DO;
          C(J)=COUNT(I);
     END;
    END;
  END;
  KEEP C1 C3 C4 C5 C7 C8 C10 PROD_NM PHARMACY;
RUN;
```

- Finally, the dataset BOTH2. containing counts for all exception codes, was merged with the dataset MEDS2, containing the number of prescriptions for each medication and pharmacy. Five observations are shown in Appendix 3(d).

```
DATA PHARMACY;
  MERGE MEDS2(IN=IN1) EXCEP2;
  BY PHARMACY PROD_NM;
  IF (IN1=1);
  IF FIRST.PROD_NM;
RUN;
```

EXPORTING DATA TO EXCEL

The data was then exported to an Excel file for final formatting. An example of a page from the final report is shown on Appendix 4.

```
DATA _NULL_ ;
  SET PHARMACY;
  BY PHARMACY PROD_NM;
  FILE "C:\MYDOCU~1\EXCEL\NONFORM.XLS";
  IF FIRST.PHARMACY THEN DO;
    PUT(PHARMACY PROD_NM COUNTPH SUM
        C1 C3 C4 C5 C7 C8 C10) (';');
  END;
  ELSE IF NOT (LAST.PHARMACY THEN DO;
    PUT ';' (PROD_NM COUNTPH SUM
          C1 C3 C4 C5 C7 C8 C10) (';');
  END;
  ELSE IF LAST.PHARMACY THEN DO;
          PUT (PHARMACY) (';') 'TOTAL'  ';' (TOTCOUNT
        TOTCOST) (';') ;
        PUT ' ';
        END;
RUN;
```

CONCLUSIONS

When faced with messy data, remember the following tools to help transform the data into a usable report.

- Use SAS® to read, merge, and analyze data from many sources.
  - Arrays are helpful in transforming data to a usable format.
  - The null datastep is an easy way to convert SAS® data back to Excel in order to finalize reports.

REFERENCES

SAS Institute Inc., *SAS® Language, 1996 Edition*, Cary, NC

SAS Institute Inc., *SAS® Procedures Guide*, Version 6, Third Edition, Cary, NC

SAS Institute Inc., *SAS® Guide to Macro Processing*, Version 6, Second Edition, Cary, NC

For More Information, Contact:
Beth Worrall, Pharmacy Analyst
Pharmacy Administration
Kaiser Permanente
16601 E. Centretech Pkwy
Aurora, CO 80011
303-739-3661
E-mail: elizabeth.e.worrall@kp.org

APPENDIX 1

DATASET MEDS1

OBS	PROD_NM	PHARMACY	USR_COST	FORMEX
1	MEDICATION A	K-P LAKEWOOD	82.66	1
2	MEDICATION A	K-P EMERGENCY	41.33	3
3	MEDICATION A	K-P EAST DENVER	41.33	1
4	MEDICATION A	K-P EAST DENVER	41.33	1
5	MEDICATION A	K-P EAST DENVER	41.33	1

APPENDIX 2

DATASET COST_MED

OBS	PROD_NM	PHARMACY	COST_MED	COUNTMED
1	MEDICATION A	K-P ARAPAHOE	123.99	3
2	MEDICATION A	K-P BASELINE	165.32	2
3	MEDICATION A	K-P BOULDER	165.32	2
4	MEDICATION A	K-P BROOMFIELD	123.99	2
5	MEDICATION A	K-P DIRECT RX	165.32	3

DATASET COST_PH

OBS	PHARMACY	COST_PH	COUNT_PH
1	DENVER WAREHOUSE	15.25	11
2	K-P ARAPAHOE	12316.61	166
3	K-P AURORA CTR	4868.22	60
4	K-P BASELINE	2496.00	36
5	K-P BOULDER	6796.82	120

DATASET MEDS2

OBS	PROD_NM	PHARMACY	COUNTMED	COST_MED	COUNT_PH	COST_PH
1	MEDICATION A	K-P ARAPAHOE	3	123.99	166	12316.61
2	MEDICATION AM	K-P ARAPAHOE	1	25.01	166	12316.61
3	MEDICATION AN	K-P ARAPAHOE	1	13.29	166	12316.61
4	MEDICATION AR	K-P ARAPAHOE	2	51.50	166	12316.61
5	MEDICATION AT	K-P ARAPAHOE	3	100.35	166	12316.61

APPENDIX 3

(A) DATASET TRANSMED

OBS	PHARMACY	PROD_NM	FORMEX	COUNT
746	K-P LAKEWOOD	MEDICATION A	1	1
747	K-P LAKEWOOD	MEDICATION B	1	1
748	K-P LAKEWOOD	MEDICATION Z	1	4
749	K-P LAKEWOOD	MEDICATION T	3	1
750	K-P LAKEWOOD	MEDICATION A	4	2

(B) DATASET BOTH1

OBS	PHARMACY	PROD_NM	FORMEX 1	COUNT 1	FORMEX 2	COUNT 2	FORMEX 3	COUNT 3
1	DENVER WAREHOUSE	MEDICATION F	1	9	.	.	.	.
2	DENVER WAREHOUSE	MEDICATION M	5	15	.	.	.	.
3	DENVER WAREHOUSE	MEDICATION P	1	1	.	.	.	.
4	K-P ARAPAHOE	MEDICATION A	1	2	4	1	.	.
5	K-P ARAPAHOE	MEDICATION AM	4	1	.	.	.	.

(C) DATASET BOTH2

OBS	PHARMACY	PROD_NM	C1	C3	C4	C5	C7	C8	C10
1	DENVER WAREHOUSE	MEDICATION F	9	.	.	.	.	.	.
2	DENVER WAREHOUSE	MEDICATION M	.	.	.	15	.	.	.
3	DENVER WAREHOUSE	MEDICATIOB P	1	.	.	.	.	.	.
4	K-P ARAPAHOE	MEDICATION A	2	.	1	.	.	.	.
5	K-P ARAPAHOE	MEDICATION AM	.	.	1	.	.	.	.

(D) DATASET PHARMACY

OBS	PHARMACY	PROD_NM	C1	C3	C4	C5	C7	C8	C10	COUNT_MED	COST_MED	COUNTPH	COST_PH
1	DENVER WAREHOUSE	MEDICATION F	9	.	.	.	.	.	.	9	7.02	25	15.25
2	DENVER WAREHOUSE	MEDICATION M	.	.	.	15	.	.	.	15	3.98	25	15.25
3	DENVER WAREHOUSE	MEDICATIOB P	1	.	.	.	.	.	.	1	5.25	25	15.25
4	K-P ARAPAHOE	MEDICATION A	2	.	1	.	.	.	.	3	123.99	166	12316.61
5	K-P ARAPAHOE	MEDICATION AM	.	.	1	.	.	.	.	1	25.01	166	12316.61

APPENDIX 4

PRODUCT NAME (BRAND)	GENERIC NAME	# NF RX	COST	#NF RX FORM FILED	#NF RX HOSPITAL/ER
MEDICATION X	MEDICATION X	9	$ 0.18	9	.
MEDICATION Y	MEDICATION Y	1	$ 0.20	1	.
MEDICATION Z	MEDICATION Z	1	$ 14.87	1	.
PHARMACY ZIP		11	$ 15.25	11	0