Paper 14-26

The Metamorphosis of a Study Design

Marge Scerbo, CHPDM/UMBC
Craig Dickstein, Intellicisions Data Inc.

Abstract

In a perfect world, there would be perfect data, perfect analysts, and perfect programmers creating perfect outcomes to every possible study. Unfortunately, one, two, or all of these factors are usually imperfect. Data are, especially data in large volumes, rarely flawless. Researchers and analysts designing studies may have great ideas of studies to undertake, but may have little idea of whether it can be done or how to do it. Programmers may be incredibly facile with the software, but rarely comprehend all the intricacies needed to complete a study. Thus, study designs are not often etched in stone. Most likely, they are the outcome of a long and tedious process of checks and balances.

This paper will take the reader through the process of developing a study design, using SAS software to provide results on which to base outcomes. A health care policy issue will be used as the basis for the discussion, but the ideas should carry across many industries.

Introduction

A good programmer analyst must work with a variety of methodologies within a single project. The 'programmer' portion of the brain is organized, methodical, and logical. The 'analyst' is quite a different story; patience, foresight, and a depth of understanding of both the data and the outcome, beyond merely understanding code structures, is required. In a sense, the analyst must be a mind reader and a magician.

Health care data are a world unto itself. There are vast amounts of administrative (billing) data produced daily. Except for the payment and/or collection of bills, these data are largely underused or misused.

Note that the study discussed in this paper is fictional, and that none of the data can be associated with any state or institution.

Background

Health care billing data are in three primary formats: HCFA-1500, Pharmacy, and UB-92.

HCFA-1500 records contain professional fees for those services provided by an 'individual' practitioner. The place of service for these claims and encounters can encompass many venues, including doctor's offices, laboratories, and hospitals.

Pharmacy data are the cleanest, most efficient, and easiest to manipulate. Most pharmacy data are collected at the 'point of sale' (POS), right in the drug store. These records contain information about the drug, the prescription, the provider, and the patient.

On the other hand although they represent the largest percent of health care dollars, UB-92 data are not easy to use. These files are uniquely produced by facilities: acute care hospitals, hospices, nursing homes, emergency rooms and outpatient clinics. These data are far more complicated when used for analysis.

Initial Study Design Proposal

HMOs (Health Maintenance Organizations) and MCOs (Managed Care Organizations) are based on the premise that by providing good preventive measures and case management, fewer facility charges will be incurred, and overall cost will drop. In this vein, HMO and MCO management is always looking at the bottom line for possible savings.

Under a particular contract of interest to this study, an insurance organization is responsible to pay for the first 31 days in a nursing facility, either an ICF (Intermediate Care Facility) or SNF (Skilled Nursing Facility). An HMO administrator had the idea that money could be saved by moving patients from acute care hospitals with a stay of over 5 days into intermediate care facilities (ICF). It is unclear whether the administrator requested any clinical input to this initial premise.

The first step proposed in the study design is to count the number of patients at the close of calendar year 2000 who remain in the hospital and the number of patients who have been transferred to an ICF during calendar year 2000.

The initial programmatic request includes the following:

- Determination of the frequency of discharge status in the UB-92 2000 year data
- Selection of those patients who are still hospitalized

- Selection of patients who have been hospitalized over 5 days
- Calculation of their cost per day
- Calculation of the cost per day for patients who are in an ICF
- Calculation of the difference in costs

A report of the above information is requested to be delivered within one week of the initial idea.

Overview of the Data

UB-92 data are both complicated and extensive. These data files are not comparable to a hospital medical record, which contains notations on every drug, laboratory test, physician visit, procedure, etc. that was incurred during a patient stay. Rather, these files contain billing data. Charges are collapsed into revenue codes with units of service attached. For example, multiple laboratory services may be grouped under revenue code 300, 'Laboratory, General Classification'. In addition, one inpatient hospital stay may in fact be defined across several UB-92 records, some with charges and others with adjustments, some across a particular date range and others across the final date range leading to discharge.

Within this HMO, in order to better utilize UB-92 files, programs have been written to create discharge summaries, where all possible records associated with a patient stay are built into one large record. Charges, units, and days are totaled with this philosophy. These data sets have been validated for quality and are sorted by the recipient identifier (RECIPID) to allow for ease in merging processes. In this situation, the discharge summary for acute care inpatient discharges (SAS data set ACUTE00) and a separate discharge summary for nursing home facilities (SAS data set LTC00) are used.

For analysis purposes, the demographic information on each patient (AGE, GENDER), the length of stay (LOS) field containing the total number of days, and the payment (PAYMENT) field containing the total cost of the stay are very useful variables.

Preliminary Analysis

In order to satisfy the first request, the programmer runs a frequency of the discharge status (DISCHSTATUS) on the acute care discharge summary file (ACUTE00). There exists a format for the discharge status (STATUS.) that is used to produce more readable results:

```
proc freq data = acute00;
   tables dischstatus;
   format dischstatus status.;
run;
```

The results are shown in Table A. This frequency analysis identifies that 1,021 patients were transferred to an ICF in calendar year 2000 while 844 patients were still in the hospital at the end of the year. Code 30 is defined as *'still in the hospital'* and code 5, '*transferred to ICF*'.

The analyst, stepping beyond the exact specifications, next runs a frequency on the length of stay field found in the same file. Rather than producing a report that showed the count of each length of stay, the programmer grouped the days to provide a more concise report:

```
proc format;
   value days
          0-5     = '0-5'
          6-10    = '6-10'
          11-15   = '11-15'
          16-20   = '16-20'
          21-31   = '21-31'
          32-high = 'high';
run;
proc freq data = acute00;
   tables los;
   format los days.;
   where dischstatus = 30;
run;
```

The output of this report is shown in Table B. It clearly displays that of the 844 patients still in the hospital, only 32.7% of the patients (267 people) had stays over 5 days. This table serves several purposes, the most important being a checkpoint for the files to be created in the next steps of the study.

The analyst then proceeds to create the requested data set containing only those patients still in the hospital and with a stay greater than 5 days:

```
data stillin;
   set acute00 (where = (los gt 5
        and dischstatus = 30));
run;

The LOG reports:

NOTE: The data set WORK.STILLIN has 276
      observations and 59 variables.
```

This data step allows the analyst a double check that 276 patients fell into this category.

The average cost per day of these 276 patients is then calculated using PROC SUMMARY. (Note that PROC MEANS also provides similar output.)

```
proc summary data = stillin;
     output out= inptcost
            sum= inptdol inptdays;
     var payments los;
run;
```

Output of this procedure demonstrates the following:

- A total cost (INPTDOL) of $15,779,690
- A total number of days (INPTDAYS) of 6,053
- Thus, the average cost per day is $2,606 (INPTDOL/INPTDAYS)

A similar process is then used to calculate the number of people in ICFs. Since there are several types of long term care facilities contained in this file, a provider type (PROVTYPE) code of 57 is used to identify ICFs:

```
data nursinghome;
    set ltc00  (where= (provtype = 57));
run;

NOTE: The data set WORK.NURSINGHOME has
      5983 observations and 43 variables.
```

The total costs are calculated thus:

```
proc summary data = nursinghome;
     output out= nhcost
            sum= nhdol nhdays;
     var payments los;
run;
```

Output of this procedure shows:

- A total cost (NHDOL) of $876,121,178
- A total number of days (NHDAYS) of 1,143,242
- Thus an average cost per day is $767 (NHDOL/NHDAYS).

Thus, the difference of cost per day between the inpatient hospital and the nursing home is $1,839 ($2,606 - $767).

Although the programmer had followed the initial specifications, certain inaccuracies were apparent when this information was presented to an HMO committee:

- There are 1,021 patients identified in Table A as transferred from an inpatient hospital to an ICF. Why then are 5,983 patients identified in nursing homes?
- Although the HMO is required to pay up to 31 days in a nursing home facility, there is no note of that in the design

Therefore, the next phase of the study begins.

Revised Study Design

The study design is now revised to include additional subset criteria:

- Include only patients who can be identified as having transferred from a hospital to an ICF (discharge status of 30)
- Of those 1,021 patients, include only those with a length of stay of 31 days or less

Additional Analysis

The first step to be completed is the selection of those patients who were transferred from a hospital to an ICF:

```
data icf;
     merge stillin (in = hosp)
           nursinghome (in = ltc
               where = (los le 31));
     by recipid;
     if in hosp and in ltc;
 run;

NOTE: The data set WORK.ICF has 1021
      observations and 43 variables.
```

The next step is to summarize the payments for this group of patients:

```
proc summary data = icf;
     output out= icfcost
            sum= icfdol icfdays;
     var payments los;
run;
```

The output data set contains this information:

- A total cost (ICFDOL) of $859,230
- A total number of days (ICFDAYS) of 9,524
- Thus, an average cost per day of $1,639 (ICFDOL/ICFDAYS).

Consequently, the difference in cost per day between the inpatient hospital and the nursing home is $967 ($2,606 - $1,639).

So there are clearly cost savings identified. At this point, the HMO management realized that there had been no clinical input. The physician on staff was brought in to review the study to date.

Additional studies were requested that included:

- A report of the 15 top primary diagnoses associated with those patients still hospitalized

- A report of the 15 top primary diagnoses associated with those patients in ICFs
- A list of the top five costliest hospitals for those patients still in the hospital
- A list of the top five costliest nursing homes to which hospital patients were transferred

The results of these studies are displayed in Tables C to F. The clinical advisor on this project will comment on the diagnoses shown and determine the appropriateness of any transfer from hospital to ICF. This portion of the report is beyond the scope of the analyst's outcomes.

Final Study Design

The final study design includes all those pieces in the earlier analysis with the appropriate edits:

- Determination of the frequency of discharge status in the UB-92 2000 year data
- Selection of those patients hospitalized over 5 days who are still hospitalized
- Calculation of their cost per day
- Selection of those ICF patients who can be identified as having been transferred from a hospital (discharge status of 30)
- These records should be further subset to select only those patients with a length of stay of 31 days or less in the nursing home
- Calculation of the cost per day for ICF patients
- Calculation of the difference in costs
- A report of the 15 top primary diagnoses associated with those patients still hospitalized
- A report of the 15 top primary diagnoses associated with those patients in ICFs
- A list of the top five hospitals identified by costs to those patients still in the hospital
- A list of the top five nursing homes identified by costs to those patients transferred from the hospital

In addition, new tables are requested which show an age (children and adult) breakdown by gender of both populations studied. These new results are shown in Tables G and H.

Conclusion - Outcome of Study

As shown, this study was implemented as simple frequencies and iteratively enhanced, as each resulting table was available. Since this analysis might have direct impact on patients' treatments, it was important that clinical input was requested.

There are several questions still pending. Is this truly the final study or is this all that is available in the time alloted? Can any requirements be placed on physicians concerning length of stay that are clinically sound?

Should additional analysis take place? For example:

- Do the type of condition and the status of the patient control the outcome?
- What other factors may affect overall length of stay?
- Do patients recover more quickly in hospitals?

Of all the information collected and reported for this study, the HMO administrator was able to act immediately on only one factor. It is clearly shown in Table E that one specific hospital far exceeded payments than all other institutions. The director met with hospital staff to discuss the high volume of patients who were hospitalized for over 5 days. Suggestions were made to begin actual medical record analysis to assess this possible problem.

For more information on study designs, check the case study discussed in *Health Care Data and the SAS System*.

References

Scerbo, M., Dickstein, C., and Wilson, A. (2001). Health Care Data and the SAS System, Cary, NC: SAS Institute, Inc.

Contact Information

For more information contact:

Marge Scerbo
CHPDM/UMBC
1000 Hilltop Circle
Social Science Room 309
Baltimore, MD 21250
Email: scerbo@chpdm.umbc.edu

Craig Dickstein
Intellicisions Data Inc.
P.O. Box 502
Weare, NH 03281
Email: cdickstein@att.net

Table A - Frequency of Patient Status

Status	Frequency	Percent	Cum. Frequency	Cum. Percent
Disch/Trans to Home or Self Care	87,219	85.4	87,219	85.4
Disch/Trans to Other Hospital	1,688	1.6	88,907	87.0
Disch/Trans to SNF	5,619	5.5	94,526	92.5
Disch/Trans to ICF	1,021	1.0	95,547	93.5
Disch/Trans to Other Institution	2,420	2.3	97,967	95.9
Left Against Medical Advice	1,397	1.3	99,364	97.2
Patient Died	1,901	1.8	101,265	99.1
Still A Patient	844	0.8	102,109	100.0

Table B - Formatted Frequency of Total Days - Still in Hospital

Totaldays	Frequency	Percent	Cum. Frequency	Cum. Percent
0-5	568	67.3	568	67.4
6-10	81	9.6	649	77.0
11-15	42	5.0	691	82.0
16-20	35	4.2	726	86.1
21-31	64	7.6	790	93.7
high	54	6.3	844	100.0

Table C– Top 15 Diagnoses for 'Still in Hospital'

Primary Diagnosis	Count
Schizoaffective Disorder	32
Congestive Heart Failure	28
HIV Aids	23
Pneumonia	23
Respiratory Distress	23
Hypovolemia	16
Septicemia	16
Rehabilitation Procedure	15
Depress Psychosis	12
Extreme Immaturity	12
Food/Vomiting	12
Paranoid Schizophrenia	12
Staphyloccal Pneumonia	10
Bipolar Affective Disorder	9
Decubitus Ulcers	9

Table D– Top 15 Diagnoses for ICF

Primary Diagnosis	Count
CVA	98
Senile Dementia	40
Cardiovascular Disease	36
Alzheimers	32
Cerebrovascular Disorder	32
Hip Replacement	26
Hypertension	24
Psychosis	18
Paralysis Agitans	17
Multiple Sclerosis	14
Diabetes	14
Decubitus Ulcers	9
Presenile Dementia	7
Depressive Disorder	7
Neoplasm	5

Table E - Top 5 Hospitals

Hospital	Total Payments
University Center	$2,187,634
Ben Franklin Hospital	$1,506,123
Union Square Hospital	$414,168
Childrens Center	$361,421
St. Johns	$310,448

Table F - Top 5 ICFs

Hospital	Total Payments
Freetown Rehabilitation Center	$96,940
Morristown Center	$49,461
Main Eldercare	$38,821
Northeast Convalescent Home	$35,503
St. Michaels Nursing Center	$23,606

Table G - Demographic Identifiers of 'Still in Hospital'

Female		Male	
Children	Adult	Children	Adult
27	99	43	108

Table H - Demographic Identifiers of 'Transferred to ICF'

Female		Male	
Children	Adult	Children	Adult
33	352	25	258