UNI-PRO: A SYSTEM TO SCREEN LARGE HEALTH CARE DATA SETS USING SAS

William J. McDonald  J. Jon Veloski
Harper Consulting Group

ABSTRACT

Government health insurance programs and private insurance companies routinely screen masses of patient claims data to find clues to what appear to be inefficient use of hospital resources, or care of substandard quality. Using the capabilities of SAS, techniques were developed to analyze data from health care billing systems. The UNI-PRO system uses specifications embedded in SAS FORMAT libraries, tables stored in sequential files, as well as more complex hard-coded criteria. It is possible to locate unusual patterns of care based on patient demographics, disease/procedure (ICD-9-CM) coding and other combinations of data on hospitalizations that help to guide subsequent manual review of detailed medical records. Algorithms using RETAIN statements and SET statement pointers make it possible to identify the possibility of premature discharge from hospitals, unexpected readmissions to hospitals and inappropriate transfers among health care providers. When multiple claim records are identified they can be reduced to a manageable quantity by simple or stratified random sampling. Components of this system have been implemented in six different settings: three statewide/regional Medicare review systems, a Medicaid research system, a health maintenance organization (HMO) review system and an ambulatory surgery review system.

INTRODUCTION

The Federal government and the insurance industry continue to show interest in finding new and more effective ways to monitor the cost and quality of health care. The Health Care Financing Administration (HCFA) began in 1974 to monitor care delivered to Medicare and Medicaid beneficiaries by means of a network of Professional Standards Review Organizations (PSRO). Nine years later, the PSRO system was replaced by the Peer Review Organization (PRO) program. Also established was a pre-paid capitation system for the health care of Medicare enrollees joining health maintenance organizations (HMO) and competitive medical plans (CMP). The quality assurance function of peer review expanded and became more definitive by requiring PROs to assess the quality of care provided to Medicare beneficiaries enrolled in HMO/CMPs. More recently, the Omnibus Budget Reconciliation Act of 1986 extended the utilization and quality assurance responsibility of the PROs to include outpatient surgical claims.

All of these activities have contributed to the collection of data in the health care industry, increasing the demand for statistically sound screening methods and reporting capabilities. This paper describes a system written entirely in SAS that has been used to meet the requirements of representative governmental and private groups.
SYSTEM DESIGN/OBJECTIVES

When the UNI-PRO system was designed in 1983, six goals were established.

Ability to Respond to Frequent Change

Three major factors brought about a need for a system that would be flexible and could adapt to change quickly and economically. First, any organization that deals with Federal contracts subject to periodic renewal must be prepared to respond to changes dictated by new contract specifications. These changes are somewhat frequent, often significant and always beyond the control of the contractor. A second factor is related to the dynamic nature of medicine. Improvements in diagnosis and treatment lead to changes in the coding of diseases and medical procedures using the International Classification of Diseases (ICD-9-CM). Finally, the technology of reviewing the quality of health care has been changing rapidly over the past decade.

Flexible, Complex Edits

Rigorous yet efficient editing procedures are crucial when fiscal intermediaries and governmental agencies combine the data sets that are created by many diverse providers of health care. Since errors can be introduced as a result of variation in data collection policies and differing standards of accuracy, we chose to perform virtually every possible edit on each relevant field from input such as the standard HCFA 1450 (UB-82) form that is used for all inpatient Medicare claims and is also used by many insurance companies. The UNI-PRO edit makes use of table-driven edits as well as more complex error checking written in Boolean logic.

The larger edit tables are stored in SAS FORMAT libraries. Each entry in the FORMAT contains the value as well as a string of encoded information that relates that value to other data in an input record.

For example, certain ICD-9-CM diagnostic codes are permissible only for a particular gender group, only for a certain age group, or may require the presence of other ICD-9-CM codes. During the editing process each of the six diagnostic codes (dx1…dx6) in a UB-82 record is examined by the following SAS statement:

```
DXSTRING=PUT (dx_n, $DXFMT.);
```

When the value of dx_n is present in the FORMAT library a vector of codes pertaining to that ICD-9-CM value is placed in the variable DXSTRING. Examples of typical codes contained in this string include the following:

<table>
<thead>
<tr>
<th>Code</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>1,2</td>
<td>Sex-specific diagnosis</td>
</tr>
<tr>
<td>N</td>
<td>Diagnosis found in newborns</td>
</tr>
<tr>
<td>C</td>
<td>Diagnosis found in children</td>
</tr>
<tr>
<td>M</td>
<td>Manifestation code</td>
</tr>
<tr>
<td>O</td>
<td>Obstetrical diagnosis</td>
</tr>
<tr>
<td>G</td>
<td>Diagnosis found in elderly</td>
</tr>
</tbody>
</table>

Thus each diagnosis can be edited in relation to other characteristics of each patient as follows:

```
IF INDEXC (DXSTRING,'1') GT 0
AND SEX NE 'M'
THEN .. (error) .. ;
```

When no formatted string for any dx_n can be found in the FORMAT, it is invalid. Similar concepts can be used to edit ICD-9-CM procedure codes, Current Procedural Terminology (CPT) codes, provider codes and their characteristics and Diagnosis Related Groups (DRGs).

More complex logical edits are written using IF, THEN, DO, ELSE statements to ensure clarity and flexibility. For example, comparisons can be made between dates of admission, dates of discharge from the
hospital and dates that procedures were performed.

Any editing exception found in each record is encoded in either of two binary arrays, ERROR or WARNING. The location of the error flag in the array identifies the specific reason for rejection of a record. At the completion of editing each record these two arrays are examined to determine whether the record is to be accepted for master file processing. Certain combinations of errors and warnings may result in the rejection of a record. The values in arrays can be used to report errors and to accumulate counts of each type of error.

Flexible Data Management

The major source of input is the 2000 byte UNIBILL tape submitted on a routine batch processing cycle. The power and flexibility of SAS INPUT and DATA step statements facilitate the reading of this complex tape file. Included are multiple record types, and arrays which are used to pack variable amounts of data into fixed length records.

Records having errors identified in the edit process are put into a separate cumulative error file which can be corrected interactively using the SAS procedure, FSEDIT. After corrections are made the entire cumulative error file is recycled for editing.

The "clean" records are passed on for addition to the master file. Although most claim records are complete summaries of a patient's hospital stay, one unusual problem is that of interim bills. These multiple claims require a separate SAS subsystem that combines partial patient claim records into a single bill reflecting the patient's entire hospitalization. All file maintenance procedures are simplified by using the SAS UPDATE procedure. The updating process is carefully managed by using separate names for variables in the input record and master record. By doing so it is possible to edit any adjustments made to master file records and then to take action when certain events occur. For example, adjustments to Medicare claim records initiated by fiscal intermediaries are processed differently from those initiated by a PRO or HCFA.

The data management system also maintains a complete audit trail of transactions which have been introduced to the system. There are cases where the extent of errors in the input tapes may cause rejection and return of entire tapes. If a tape is deemed acceptable, claims with errors are to be maintained in separate error files and complete and accurate counts maintained of receipt and disposition of record count totals.

Magnetic output from the system include adjustment tapes and Federal reporting in the form of the PRO Hospital Discharge Data Set (PHDDS) which can be written with SAS PUT and FILE statements.

Accurate Selection of Cases for Review

Many of the PRO review activities have explicit sampling requirements defining how claims are to be selected for more detailed review of actual medical records. HCFA has issued detailed sampling process criteria. The viability of PRO activities and the acceptability of data reported to the public are contingent upon the selection of samples in accordance with accepted practices. This ensures that the review findings are truly representative of the population (universe) subject to review.

To make proper use of sample statistics for projections and estimates, it is essential to provide total universe counts of all records eligible for selection in the various review categories. Some patient claim records may be eligible for inclusion in more than one universe and thus be eligible for selection in more than one review category. Membership in a universe for each record is encoded in an array, REVTYP. The value
and location of each element of the array is set when a record is a part of a universe from which a sample was drawn.

Some of the most common sample universes and related sampling percentages are:

<table>
<thead>
<tr>
<th>Sample Universe</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hospital specific sample</td>
<td>3%</td>
</tr>
<tr>
<td>Specialty hospital sample</td>
<td>15%</td>
</tr>
<tr>
<td>Readmission 0-7 days</td>
<td>50%</td>
</tr>
<tr>
<td>Readmission 16-31 days</td>
<td>25%</td>
</tr>
<tr>
<td>Rehab unit transfers</td>
<td>25%</td>
</tr>
</tbody>
</table>

In the interest of efficiency, a subset of the master file is temporarily created to include only those patient claim records meeting certain criteria. With this subset of claims, the system is able to identify both simple universes and the more complex universes requiring inter-record comparison.

**Inter-Record Comparisons** - The UNI-PRO system uses as its primary sort key the patient's identification number (HICNO), admission date (ADMDATE), claim thru date (THRUDATE), and provider number (PROVNO). By tracking sequences of admissions, discharges and transfers in the records of patients, patterns of care are identified. Since a patient may receive health care at different points in time in the same facility, or in different facilities, it is important to be able to examine the relationships among related records.

Although the identification of these readmissions and transfers could be accomplished by using special indices or complex data structures, we chose the following approach:

```plaintext
DATA SAMPLING;
   RETAIN ... ;
   SET SAMPLING(FIRSTOBS=2) ... ;
   HICNO2=HICNO;
   ADMDATE2=ADMDATE;
   SET SAMPLING;
```

This use of the SET statement pointer allows one to look forward in a data set to find out whether the current record is the first of a related pair of records. When the current record is marked as the first of a pair, the RETAIN statement transfers appropriate data to the next record. In looking forward in the data set it is also possible to create variables in the current record which reflect the next or last admission date related to the present record.

**Statistical Sampling** - For some samples, it is acceptable to use simple random selection. In these situations, we used the UNIFORM function:

```plaintext
IF UNIFORM( ) LE .1 THEN ......;
```

Some procedures require that the size of the universe be known before the sample can be drawn in order to satisfy criteria of a minimum or maximum number of items. In this situation, output from PROC SUMMARY is merged with the current subset and samples drawn accordingly.

Most review requirements, however, are based on hospital-specific interval samples. The intervals are computed from the percentage of cases to be reviewed and initialized in an array, INTRV. Continuity is assured from one batch to the next, and among processing cycles by using a residual interval counter (an array, TAKEME) indicating the next record to be selected.

Other arrays are established for report preparation, as follow:

```plaintext
DO OVER REVTYP;
   IF REVTYP (is in universe);
   THEN COUNT+1;
   IF COUNT EQ TAKEME THEN DO;
      REVTYP= {is in sample);
      TAKEME=TAKEME+INTRV;
   END;
END;
```
Thus, for each record this approach examines each of the review types in the array, REVTYP. When the value of REVTYP indicates that the record is in the universe, the universe count is incremented. When the value of the universe count reaches the value of TAKEME, the REVTYP is changed from a code indicating membership in a universe to one indicating that the record is in the sample. Finally the value of TAKEME is incremented by the sampling interval to the number of the next member of the universe to be sampled.

Simplified Report Writing

The opportunity to make use of the intermediate output from procedures such as MEANS and FREQ, combined with PUT statements made SAS a natural choice. The capabilities of SAS in this area are well-documented in previous proceedings of SUGI in many areas of application, including health care1-5.

End-User Retrievals

Because methods of reviewing health care are changing and improving rapidly, it was essential to use a system that would enable users to prepare ad-hoc retrieval and reports. Again, the capabilities of SAS in this area are well-documented1-5.

IMPLEMENTATION

The approach described has been used in six different settings between 1983 and 1988.

Regional PSRO The system was developed to process and edit approximately 100,000 discharges per year and to produce quarterly PHDDS tapes for Federal reporting.

Statewide Medicaid research system Components of the UNI-PRO system were used to process approximately 2.5 million records from a Medicaid billing system including inpatient, outpatient, home health, physicians' offices and pharmacy records.

Small PRO Priorities in this installation included intensive editing, preparation of PHDDS tapes, and the development of reports written in SAS by end-users (30,000 discharges per year).

Large PRO (250,000 discharges per year) In addition to editing and complete data base management, this application of the UNI-PRO system requires the use of simple and stratified random samples.

Special Periodic Review: HMO hospitalizations (2,000 discharges per year) Hospitalizations for patients enrolled in HMO programs in one state are processed and reviewed monthly.

Special Periodic Review: Ambulatory Surgery Data on claims submitted by hospital outpatient departments and ambulatory surgical centers in one state will soon be processed routinely and screened quarterly in order to select claims which must be reviewed more carefully using actual medical records.
CONCLUSIONS

To be effective, any methods for screening large health care data sets must address three basic issues. Careful editing is essential to assure that results will be as accurate as possible. It is recognized that the errors introduced into large data sets at various levels of processing by providers and fiscal intermediaries, if ignored, will invariably lead to inaccurate screening. Elaborate inter-record comparisons are needed in order to identify repeated hospitalizations, premature discharges, and transfers to other health care facilities. Finally, complex approaches to statistical sampling are required when large numbers of selected records must be reduced to manageable quantities. SAS provides the flexibility needed to solve these three problems, as evidenced by the use of the UNI-PRO approach in six diverse settings.

The results of developing this system have confirmed that SAS can help one to solve these three problems, in addition to supporting data base management, report writing and ease-of-use by a wide base of individuals. The availability of PC/SAS® will open new opportunities to improve the capabilities of systems such as the one described here, by making it possible to download samples for local processing. In fact, by using many of the approaches described herein, it would be possible to process limited volumes of data on a smaller computer using PC/SAS.

REFERENCES


CONTACT AUTHOR

William J. McDonald
Harper Consulting Group
1001 Harper Avenue
Drexel Hill, PA 19026
USA

*SAS and PC/SAS are the registered trademarks of SAS Institute, Inc., Cary, NC, USA.