

Dealing with Health Care Data using the SAS® System

Alan C. Wilson, UMDNJ-RWJMS, New Brunswick, NJ

Marge Scerbo, CHPDM, UMBC, Baltimore, MD

ABSTRACT

Analysis of health care data without an awareness of their peculiarities and eccentricities is risky business. You need to know of possible variability and discontinuities in medical coding practices, and in personal, demographic, time and place variables. Above all, you need to know your data. A clear understanding of the quality and completeness of the data is essential. This paper provides an overview of different problems that might be encountered, as well as some possible solutions using SAS software.

INTRODUCTION

In a perfect world, accurate, standardized health care information would flow freely among patients, providers and payers. Health care administrators, planners, researchers and reviewers would have quick and easy access to the data necessary for their activities. In the world we live in however, health care data are collected and delivered in diverse formats, many with no common identifiers.

A whole new industry has sprung up to provide uniform sets of utilization and price data to insurers, federal and state agencies, employers, and data organizations. Data vendors develop custom tools to capture clinical, financial, and other health-related data in collection hubs from hundreds of sources in a variety of formats. They standardize the data, perform edit and verification checks, provide data correction tools to submitting organizations and work with them to improve accuracy. The data are then compiled into databases to provide custom reports and make the data available for epidemiological and bio-statistical analysis.

With the aid of SAS software, you can deal with health care data yourself. This paper shows you some of the techniques we use daily, and gives you some tips on avoiding the pitfalls we have come across.

ORIGINS OF DIVERSITY

From the time you are born until you die, data are being collected about your health care. Births and deaths have been registered centrally for many

years in a standard format. The health care claims data we will be discussing are a different matter.

Over the life-time of a typical individual, health care billing claims will arise from the conditions serviced by providers listed in Table 1.

Table 1. Typical health care areas and providers

Areas	
Maternity/ Prenatal care	Children's health
Prevention/ well visits	Acute illness
Chronic/ long-term care	Hospice care.
Providers	
Dentists	Pharmacists
Physicians	Other licensed health professionals
Ambulatory care centers	Acute care hospitals
Rehabilitation/chronic care centers	Psychiatric hospitals

Most of these will likely have a different storage and/or reporting formats. Add onto this the different claims and utilization reporting requirements among federal and state departments, and third party payers, and you can see how the problem has grown.

The situation is not completely hopeless though. Many of the common variables needed for billing are included in each reporting format and some of these can be used to standardize and link the data.

Table 2. Common variables

Visit/Patient Identifiers	
Patient identifier	Provider identifier
Date of visit	Place of service
Diagnosis (one or more)	Patient age
Date of Birth	Sex
ZIP code (residence)	Source of admission
Discharge date	Discharge status
Services/Procedures and Insurance	
Laboratory or Radiology service	Medical or Surgical service
Pharmacy	Dates of services or procedures
Expected source of payment	Billed charges for each service

In some cases, racial and ethnic category, names, addresses, and other personal identifiers (social security or Medicare number) may be omitted from the database.

The claims files described above are the ones we have most experience with, but are just some of the available sources of information. The others include the beneficiary file (information obtained at enrollment or registration) and the provider file (containing information identifying and describing the health care providers).

The setting and provider of the care as well as the particular disease or condition of interest, e.g. cancer, cardiovascular disease, accidents, surgical procedures etc., will determine the appropriate data source.

THE RECORD LAYOUT

It is important to know where the data originated. The completeness and quality of the information will depend upon the source (see later). The first step is to acquire and study the record layout. Be aware of when periodic changes in record layout take place, and adjust your data input statements accordingly.

Hospital Discharge Data

In contrast to birth and death data, statewide hospital discharge data are not currently maintained by all states. Most of the 36 states that do use one of two standard formats: the UB-92 (Uniform Billing) reporting format with the UB-92 claim form, or the UHDDS (Uniform Hospital Discharge Data Set) and the HCFA-1500 form. The Department of Health, Education and Welfare initiated the UHDDS in 1974 to improve the uniformity of hospital discharge data for patients enrolled in the Medicaid and Medicare programs. Other states, like Maryland, use unique systems such as HSCRC (Health Services Cost Review Commission). Virginia has no statewide system currently in use.

Table 3. Hospital Discharge Format, by State

NESUG state	System type
Connecticut	UB-82
Delaware	UB-82
Maine	UHDDS
Maryland	HSCRC
Massachusetts	UB-82
New Hampshire	UHDDS
New Jersey	UB-82
New York	Unique
Pennsylvania	UB-82
Rhode Island	UHDDS
Vermont	UHDDS
(Virginia)	Not statewide

MAKING A SAS DATASET (an example)

With the raw data (UBDATA95.DAT) and the record layout in hand, you can read the external file and create a new SAS dataset (CLAIMS95.ssd01) in the RESEARCH library on the UNIX computer.

Briefly, SAS code is submitted to identify the location and name of the source file, to specify a SAS data library, to assign a name to the SAS dataset, and to point to the location of the data element in the raw record. Once the SAS data set has been created, the data are accessible to further SAS DATA steps or PROC steps. Here are the steps, one at a time:

A **FILENAME** statement references the external file location with a 'fileref' - SUGI24 here:

```
filename SUGI24 '/users/UBDATA95.DAT';
```

A **LIBNAME** statement references a 'libref', the physical location where the SAS data library is stored. To save the newly created SAS dataset permanently, so it is not deleted at the end of the SAS session, use the libref along with the data set name in the **DATA** statement to form a two-level name.

```
libname RESEARCH '/users2/awilson/research';
```

```
data research.claims95;
  infile SUGI24 ..... obs=10;
  input @26 admdat mmdyy6.
        @3 medred $9.
        ..... ;
run;
```

The **INFILE** statement above specifies which file contains the raw data, along with information about the records physical nature (fixed block, logical record length etc.).

The **INPUT** statement assigns the three essential parts defining a SAS variable: 1) a name 2) a type (character or numeric) and 3) a length. This example uses column format to input a SAS date informat and a character informat. The DATA step is closed with a **RUN** statement.

We recommend the following steps. Include the **OBS=** option in the INFILE statement. Submit the DATA step and check the LOG for messages. List the test dataset to the OUTPUT window using the PRINT procedure, or view it interactively with the FSVIEW or FSBROWSE procedure:

```
proc print data=research.claims95;
or
proc fsview data=research.claims95;
proc fsbrowse data=research.claims95;
run;
```

If you are satisfied with the test results, remove the OBS= option and resubmit the DATA step. Remember to check the LOG again, and view the dataset. Get to know your data.

VARIABLE CODING

In addition to the record layout, the coding practices current at the time that the data was input must be known. This applies to most of the variables listed in Table 2, and particularly to racial or ethnic grouping, which frequently differs between databases and within databases over time.

ICD (International Classification of Diseases)

Throughout the U.S. and most of the world, diseases, injuries and causes of death are classified according to the World Health Organization (WHO) codes. These have been revised periodically over the years as seen in Table 4.

Table 4. ICD Revisions of Coronary Heart Disease (CHD) and acute myocardial infarction (AMI) codes.

ICD Revision	Dates	CHD	AMI
ICD/6	1949-57	420	No code
ICD/7	1958-67	420	No code
ICDA/8	1968-78	410-413	410
ICD/9	1979-87	410-413	410

ICD-9-CM Codes- Diagnoses and Procedures

The Clinical Modification of the Ninth Revision codes (ICD-9-CM) Volumes 1 & 2 provide additional fourth and fifth digits to the diagnosis rubric to allow more detailed reporting. Institutional (in-hospital) charges for procedures are coded with ICD-9-CM (Volume 3) four digit codes to distinguish them from the physician's portion of the charges for procedures, which are billed with CPT codes (see later).

ICD-9-CM diagnosis codes are 3-6 digit codes in the format **nnn.ab** where:

nnn specifies a specific condition or disease code. The first digit can be the number 0 through 9, a 'V' or an 'E'.

a identifies more specific information concerning location or degree. A '9' in this position normally defines NOS or Not Otherwise Specified.

b is a further categorization of the condition.

SAS software is incredibly powerful in handling these codes. Most ICD-9 diagnosis fields are stored with an implicit decimal as 5 digit character fields. If the data come from various sources, some stored

with a period and others without, the SAS function **COMPRESS** is extremely useful. This function will remove the period from the field so:

```
length newdiag $5;
newdiag = compress(diag, '.');
```

creating the value of newdiag 2950 from the value of diag 295.0

If you have to add the period to those fields that do not include the period, this code can be used:

```
length newdiag $6;
newdiag = substr(diag,1,3)||'.'||
substr(diag,4,2);
```

which will create the value of newdiag 295.0 from the value of diag 2950.

CPT codes

Physicians billing for outpatient and in-hospital services or procedures use the American Medical Association's Current Procedural Terminology (CPT) five digit codes, which are revised each year. Laboratory charges are also coded with CPT codes. CPT coders for each department (ob/gyn, surgery, pediatrics, medicine etc.) are trained in that specialty, and use a translation program to produce HCPCS (pronounced Hicpiks) codes for submission to Blue Cross/Shield, Medicare and state Medicaid payers.

HCPCS

HCFA (Health Care Financing Administration) uses HCPCS National Level II Codes (HCFA Common Procedure Coding System) for Medicare claims. HCPCS is used increasingly by commercial insurance for pharmaceuticals and durable medical equipment (DME) charges.

DRG Codes

HCFA does not pay by disease, but rather by a coding system. For inpatient hospital care, diagnosis-related groups (DRG) cover costs. The DRG coding is a system of risk-adjustment (see later) using diagnosis and procedure codes.

Dental billing and pharmacy charges have different coding system, which we will not cover here. Revenue codes are also stored electronically in different forms.

WARNING - DISCHARGE DATA

The principal diagnosis is the condition thought to be present (*after study*) at the time of hospital

admission. Other diagnoses may include additional conditions present on admission and those arising during and after the episode of care; these are represented as secondary diagnoses. One complication of administrative discharge data is that it is sometimes difficult for a coder to distinguish these from the medical record. In such cases, the coding follows administrative rules which may not accurately reflect the clinical record.

RECODING DATA

It may be necessary to recode variables derived from different databases if there is a discrepancy, e.g. change from a character code to a numeric one. Another reason for recoding is to achieve a particular goal in collection, aggregation, and analysis of the data.

Composite Disease Variables

Combined ranges of ICD-9-CM codes, e.g. diabetes 250.00-250.90, hypertension 401.00-405.99, and schizophrenia 295.00-295.99 can be used to indicate the presence or absence of various disease entities.

The SAS **SUBSTR** function is useful in analyzing such conditions or diseases as in:

```
if substr(diag,1,3) = '295' ...
```

allowing analysis of all schizophrenic disorders without having to define each diagnostic code falling into that condition.

Another approach is to use the **COLON (:)** modifier with a SAS operator which will select those values beginning with 295, thus as above:

```
if diag =: '295' ...
```

or for a range to locate ischemic heart disease other than AMI:

```
if '411' <=: diag <=: '414' ....
```

or with the **IN** operator for discrete AMI codes:

```
if diag in: ('4101', '4106') ....
```

Beware of using the **INDEX** function in studying ICD-9 codes. The **INDEX** function will search a character string for a set of characters. For example:

```
if index(diag,'295') ...
```

will return the correct search for Schizophrenic Disorders but also codes such as 729.5 which defines a soft tissue pain in limb, an incorrect result.

DVS Cause of Death Codes

The NCHS (National Center for Health Statistics) which is part of the Centers for Disease Control (CDC), within the Department of Health and Human Services (DHHS) has several lists of cause-of-death codes that collapse the ICD-9 into more manageable categories, ranging from 16 to 282 diseases. The coding from the various revisions of the ICD is adjusted to avoid discontinuities.

RISK ADJUSTMENT

Constructing comorbidity or severity scores from several diagnoses, procedures and other variables can enhance the usefulness of the hospital discharge record. For example, in acute myocardial infarction (MI), the following have been correlated with a poor long-term outcome: recurrent MI, new onset heart failure or cardiogenic shock, and left ventricular dysfunction. The length of stay or the number of diagnoses and/or the number of procedures may be considered. The number and type of diagnostic tests employed may also be an indicator of the number of possible diagnoses, and may indicate the number and types of problems addressed during the encounter.

In comparisons of different providers for example, case-mix evaluation is frequently required to adjust for variations in clinical status. The components to be considered include severity of primary diagnosis, functional status, age, sex, race, insurance, and socioeconomic status, and comorbidity. This last factor can be derived from specific ICD-9 codes. A limitation of the administrative discharge record is that it lacks information on the results of clinical and lab tests performed on the patient.

A common comorbidity index is the Charlson score, which was published in 1987. Several groups using administrative data have used stepwise multiple logistic regression analysis of disease-specific data to obtain improved weighting factors. Receiver operating characteristic (ROC) curves can be used to evaluate the new models. Care should be taken to properly account for confounding, bias and statistical power. The SAS **LOGISTIC** procedure can be used to perform these analyses.

Take note of a paradox here, that some ICD-9 codes, which clinically are associated with worse outcome, for example prior (old) AMI, arthritis, chronic angina, and essential hypertension, are sometimes unexpectedly associated with *reduced* likelihood of mortality. This may arise as the result of a combination of selective coding, and the limit of

the number of diagnoses that can be recorded on the discharge summary.

Stratification or Aggregation

Subgroups of interest can be investigated by stratifying analyses by sex, race, age, county or other variables. Too small a number of events will lead to a diminished reliability of the estimates of rates and other measures. Conversely, to increase the number of events, data across several strata or years can be combined or aggregated. However this may mask trends in subgroups. There should be evidence that rates or trends are similar between groups before they are combined.

SAS software can accomplish stratification through the use of **PROC FORMAT** to create a user-defined format (as long as these categories are based on a single variable). For example, the field AGE can be formatted to produce different classifications for analysis and reporting:

```
PROC FORMAT
  value agecat
    0-14      = '<15'
    15-44     = '15-44'
    45-64     = '45-64'
    65-high   = '65 or older';
  value agegroup
    0-21      = 'Children'
    22-high   = 'Adult';
RUN;
```

which will produce two different age categories for analysis. For example, the National Center for Health Statistics (NCHS) uses two grouping systems to stratify by age. A four-age strata system and a ten-age strata system are shown below.

Table 5. NCHS age-groups

<15	0-4
15-44	5-14
45-64	15-24
65 or older	25-34
	35-44
	45-54
	55-64
	65-74
	75-84
	85 or older

Character to Numeric Recoding

This technique or vice-versa is often necessary when using data from different sources. For

example, a DRG is a 3 digit value, which may be stored as numeric or character. In order to analyze across sources, a conversion to one field type may be necessary. The SAS functions, **PUT** and **INPUT** can accomplish this:

```
*gender is a numeric field in this data set
and $sex is a character format;
data new;
  set old;
  sex = put(gender,$sex.);
run;

*gender is a character field in this data
set with values 1 and 2;
data new2;
  set old2;
  sex = input(gender,3.);
run;
```

Creating new fields

Another a useful technique to view codes and descriptions on-line is to create a format using the **PROC FORMAT CNTLIN=** option. This process uses a data set containing both the codes and the definitions to create a new format. Here is the SAS code:

```
*create a temporary data set of ICD-9 codes
and descriptions from a text file;

data diags;
  infile 'ICD9.txt';
  input @1 icd9 $5. @6 descript $40.;
  length diag $46;
  diag = icd9 || ' ' || descript;
  keep diag icd9;
run;

*create a new data set;
data fmt (rename =(icd9=start diag=label));
  set diags ;
  fmtname = '$Diag';
run;

*create a format;
proc format cntlin = fmt;
run;

*create a new field that has both code and
description;
data diags2;
  set diags;
  length longdiag $46;
  longdiag = put(icd9,$diag.);
run;
```

CHANGING THE STRUCTURE

The data you first receive may not be in the correct structure. In this example you received a .dbf file with one record containing the patient identifier (mrn), date of service (admdate) and length of stay information, followed by a varying number of records containing the diagnosis and/or procedure codes associated with that visit.

The approach used in this example is to populate each of the records with the information identifying the visit, and then to collapse them into a single line record for that visit.

This code uses the SAS DBF procedure to import the data followed by 'by group' processing:

```
filename in 'c:\.....\visits.dbf';
proc dbf db3=in out=one;
run;

data pt_list; *the patient file;
  set one;
  medrec=mrn;
  admdat=admdate;
  if los = . then delete;
run;

proc sort data=pt_list;
  by medrec admdat;
run;

data dxtable; *adding the mrn ;
  retain medrec;
  set one;
  if los ne . then medrec=mrn;
run;

data dxtable; * adding the admdat;
  retain admdat;
  set dxtable;
  if los ne . then admdat=admdate;
run;

proc sort data=dxtable; * add line numbers;
  by medrec admdat;
run;

data dxtable;
  set dxtable;
  by medrec admdat;
  if first.admdat then line=0;
  line + 1;
run;

proc sort data=dxtable *make one line ;
  by medrec admdat line;
run;

data line;
  set dxtable;
  by medrec admdat line;
  format diag1-diag40 $5.;
  format proc1-proc40 $5.;
  array diag[*] diag1-diag40;
  array proc[*] proc1-proc40;
  retain diag1-diag40;
  retain proc1-proc40;
  if first.admdat then do i=1 to 40;
    diag[i]=.;
    proc[i]=.;
  end;
  diag[line]=diagcode;
  proc[line]=proccode;
```

```
if last.admdat then output;
run;

data newvisit;
  update line pt_list;
  by medrec admdat;
run;
```

DATA QUALITY

A good system for entering claims data will have an up-to-date code dictionary – reflecting revisions and inactive codes and using the fifth digit in ICD-9-CM. This should be supplemented by exploratory data analysis by summarizing (**PROC MEANS**) and range checking (**PROC UNIVARIATE**). Note that absolute perfection is highly suspect, and that a particular collection of data represents one of a series of possible sets of data.

Variability in Diagnosis, Personal and Time Variables

Some variables appear to be more or less reliable than others. Study of the reproducibility of the race/ethnicity fields have shown that over 20% of black /non-hispanic patients will appear in a different category when re-admitted. Some states, such as New Hampshire and Oklahoma do not have a hispanic ethnicity category.

To validate the diagnosis of acute myocardial infarction for a pilot registry program, and to audit the reliability of some of the other UB fields, a random sample of 669 patient charts was selected in New Jersey. The diagnosis was accurate (definite MI) in 76% of the cases (81% in men, 58% in women). On the other hand it was wrong (no MI) in only 6% of the cases (5% men, 9% women). Classification accuracy these and some other variables is shown in the next table:

Table 6. Chart review results from NJ

Variable	Rate/669	Percent Accuracy
Sex	664	99.3
Age	667	99.7
Acute MI (men)	325	81
Acute MI (women)	155	58
Race	661	98.8
Length of Stay	664	99.3
Vital Status	661	98.8
Procedure (cardiac cath)	668	99.9
Admit date	667	99.7

INTEGRATING INFORMATION BY RECORD LINKAGE.

Matching people within or across files in order to link records, to study re-admissions or to assemble an episode of care, is not an easy task in the absence of unique identifiers such as Social Security number or Medicare number. These are frequently omitted to maintain the confidentiality of the patient record.

Data fields that contain useful linking information are first and last name, month, day and year of birth, ZIP code, city or county code, race and sex.

In general the more value levels a variable has, the more useful it is in the linking procedure. A unique identifier of course has the most levels, while sex, with only 2 levels, contributes less than month of birth, with 12 levels. Different weights need to be given according the expected reliability of each value. For example, the date value 01/01/YY is frequently used if information is lacking on the exact month and day.

Exact, or deterministic, matching using 'key' variables, by 'fuzzy' matching, or probabilistic matching can link records. The latter generates 'scores' for each possible match, sets thresholds for acceptance and for clerical checking of 'possible' matches. In the NJ experience the sensitivity and specificity of the probabilistic algorithm matching UB-82 records to death records were 98.16% and 98.17% respectively.

STATISTICS

Once the SAS data set has been created, and you are satisfied with its quality, there are a variety of statistical procedures in SAS software (Table7) that are useful in analyzing, exploring or manipulating health care data:

Table 7. Common uses for SAS Procedures

PROC FREQ	Rates Frequencies Chi-square test Confidence intervals
PROC CORR	Correlation
PROC LOGISTIC	Logistic regression
PROC PHREG	Cox regression for survival

DATA PRESENTATION

Graphical displays of results from data analyses emphasize points of interest. The following SAS procedures produce the types of figures listed below:

Table 8. Graphical presentation with SAS software

PROC CHART or GCHART	- Horizontal bar chart for comparisons - vertical bar chart for time trends - clustered bar chart - stacked bar chart - pie chart for proportions
PROC PLOT or GPLOT	- line graph for time trends - scatter plots
PROC GMAP	- maps (convert ZIP codes to latitude/longitude)

CONCLUSION

Limitations in Data

Always remember that this is **billing** data and is regulated by the requirements of the third party payer. The accuracy of the information entered by the billing clerks depends on what the clinical provider writes in the patient chart. The admitting diagnosis and discharge diagnoses determine the outcome of the payment claim. That is the outcome of interest, not the health care outcome of the patient. Diagnostic tests, repeated hospitalizations, and procedures may be limited by the regulations of payer. Whether or not they occurred may be unknown from the claims record.

When working with health care data, you must be aware that the records are the results of the sometimes-conflicting needs of administrative regulations, health economics, social sciences, medical records and clinical medicine. Diagnostic and coding practices are always subject to change. Always evaluate how these needs will impact on the results of their own analyses and research. A more accurate estimate may often be obtained by using multiple data sources in combination.

The Future

Information about electronic media claims (EMC) using the electronic UB-92 form (HCFA-1450) can be found at www.hcfa.gov/medicare/edi/edi5.htm

under the category "Health Claims." Apart from the clinical note, the claim is essentially paper-free.

HCFA is developing the following:

- A guide for states to use in developing, implementing, and using encounter data systems.
- The Tape-to-Tape project
- The State Medicaid Uniform Research Files (or SMRF) project
- The Hospital Cost and Utilization Projects (HCUP-1, 2, 3) for the Agency for Health Care Policy and Research.
- A prototype ICD-10 procedural coding system
- A supplemental chapter for ICD-9-CM diagnosis coding (H-Codes) for vital signs, findings, laboratory results and functional health status.
- An episodic grouper defining episodes of care which will allow care to be evaluated over time and across of full scope of services.

Information sources:

- Your state health department, library or data center.
- The CDC (www.cdc.gov), and the NCHS (www.cdc.gov/nchswww) can be accessed through their web sites
- You can reach the NCHS at (301) 436-7035. The number for the NCHS library is (301) 436-6147

- State population estimates can be accessed at the Census Bureau (www.census.gov/population/www/estimates/statepop.html)

- National Death Index: (301) 436-8951

- WONDER (Wide Ranging Online Data for Epidemiology Research) at the CDC, provides a wealth of data. Contact (404) 332-4569 or the CDC web site for account information

- National Association of Health Data Organizations, web site www.nahdo.org

- Agency for Health Care Policy and Research (AHCPR) web page www.ahcpr.gov/data/ provides links to many sites including HCUP (Healthcare Cost and Utilization Project) which took discharge data from 17 states and organized it into 1 format

REFERENCES

Centers for Disease Control. "Using Chronic Disease Data: a Handbook for Public Health Practitioners" Atlanta: Centers for Disease Control, 1992.

Steinwachs DM, Weiner JP, Shapiro S. "Management Information Systems and Quality" in *Providing Quality Care: Future Challenges*, N Goldfield, DB Nash (eds.) Health Administration Press, 1995

SAS is a registered trademark or trademark of SAS Institute Inc. in the USA and other countries. ® indicates USA registration.

CONTACT INFORMATION

Alan Wilson (732) 235-7857
awilson@umdnj.edu

Marge Scerbo (410) 455-6807
scerbo@chpdm.umbc.edu

A GLOSSARY OF HEALTH CARE ACRONYMS

Acronyms	Meaning
ACG	Ambulatory Care Groups
AHCPR	Agency for Health Care Policy and Research
AP-DRG	All Patient DRG
APG	Ambulatory Patient Groups
APR-DRP	All Patient Refined DRG
CDC	Centers for Disease Control
CPT	Current Procedural Terminology
DHHS	Department of Health and Human Services
DME	Durable medical Equipment
DRG	Diagnosis Related Group
EDI	Electronic Data Interchange enrollment
EMC	Electronic Media claims
HCFA-1450 (UB-92)	A Claims Form used by institutional providers for Medicare, Part A
HCFA-1500	A Claims Form used by non-institutional providers for Medicare,
HCFA	Health Care Financing Administration
HCQIP	HCFA Health Care Quality Improvement Program
HCUP	Hospital Cost and Utilization Projects
HEDIS	Health Plan Employer Data and Information Set
HIPAA	Health Insurance Portability and Accountability Act of 1996
HISSB	CDC Health Information and Surveillance Systems Board
HL7	A data standard – Health Level Seven
HMO	Health Care Management Organization
HCPCS	HCFA Common Procedure Coding System
HSCRC	Health Services Cost Review Commission
ICD-9	International Classification of Diseases
MCO	Managed Care Organization
NAHDO	National Associate of Health Data Organizations
NCHS	National Center for Health Statistics
NCI	National Cancer Institute
NCQA	National Committee for Quality Assurance
PRO	Peer Review Commission
SDO	Standards Development Organization (HL7)
SEER	Surveillance Epidemiology and End Results
SMRF	State Medicaid Uniform Research Files
SNOMED	A vocabulary
UB-92 (82)	Uniform Billing
UHDDS	Uniform Hospital Discharge Data Set
WHO	World Health Organization
X12	A data standard (ANSI accredited standards committee X12)