

Control Card File

Table 1: Sample Control Card File

```
*****  
* CONTROLFILESAMPLE.TXT  
  
* FILE DEFINITIONS  
INPUT1=sample1.txt  
INPUT2=  
OUTPUT=sample_acg.txt  
PRINT=sample_prn.txt  
EDC=sample_edc.txt  
  
* ACG BRANCHING SPECIFICATIONS  
DELIVERED  
  
* INPUT FILE LAYOUT  
ID=1,11  
AGE=12,2  
SEX=14,1  
PCOST=23,5  
POP=28,1  
ICD=29,5,3  
  
* OUTPUT FILE LAYOUT  
OUTREC=ID,ACG,HRCI,CWT
```

The format of the input and output records is controlled through the use of control cards. Table 1 provides an example of a “control card file.” Although, strictly speaking, the order of the control cards does not matter, it may be helpful to think of this file as having four main

components. The first part, **FILE DEFINITIONS*, specifies the location of the input and output files. The second part, **ACG BRANCHING SPECIFICATIONS*, provides specifications on optional branching algorithms (e.g., DELIVERED, LOWWEIGHT discussed in more detail later). The third part, **INPUT FILE LAYOUT*, provides additional detail about the positioning of data in the input files. The final component of the control card file, **OUTPUT FILE LAYOUT*, specifies the output(s) produced by the software.

A few simple rules apply to all control cards. Most have the same general form⁹:

Keyword=parameter1,parameter2,parameter3

Each control card or keyword must start with a valid keyword and must be contained entirely on one line. Control card keywords are not case sensitive, and the order of the control cards in the control file does not affect the program. Comment lines, starting with either an asterisk (*), a forward slash followed by an asterisk (/*), or two forward slashes (//) are ignored.

The parameters of the card depend on which keyword is being used, and all three parameters are not always required. The number required depends on the keyword. For example, on the ID control card, only two parameters are allowed: the starting column and the length. On the ICD card, three parameters are allowed: the starting column, the length, and the number of occurrences. On the DELIVERED card, no parameters are required. Most parameters are simply integers defining the starting column and length of an item. Some parameters, such as on the FEMALE card, must be enclosed in quotation marks. Table 2 presents all valid control card keywords followed by a detailed description of each.

Control cards may also be used to control certain user-defined options. Specifically, EDC reporting, the splitting of pregnancy into “delivered” and “not delivered” categories, and the optional bifurcation of newborns into “low” and “normal” birth weight categories are all controlled by the appropriate control card(s).

⁹ OUTREC=field, field, ... is one exception.

Control Card Keywords

Table 2: Control Card Keywords

(1) <i>Filenames</i>	(2) Record Layout	(3) Optional Fields	(4) Miscellaneous
EDC	AGE	DELIVERED	LRECL1
INPUT1	DOB	DELIVERED2	LRECL2
INPUT2	FEMALE	FIELD	NOCRLF
NONMATCH	ICD	LOWWEIGHT	NOWARN
OUTPUT	ID	LOWWEIGHT2	
PRINT	ID2	POP	
	OUTREC	PCOST	
	SEX	PREGNANT	
		PREGNANT2	

(1) Filenames

(1) EDC

This card invokes the optional Dino-Cluster algorithm and instructs the system to output a Dino-Cluster file. The format of this card is as follows:

EDC=*filename*

where *filename* is the complete path and filename for the Dino-Cluster output file. The use of this card is optional. In addition to providing a file of individual level EDC assignment, inclusion of this card produces a series of additional descriptive tables in the print file.

(2) INPUT1

This file is used to specify the location of the input file for the software as follows:

INPUT1=*filename*

where *filename* is the complete path and filename for the input record. The input data must consist of records that contain a person's ID, age or date of birth, sex, and ICD-9 code(s) of the patient. The user may also provide additional information about pregnancy, delivery status, birth weight, population grouping variable (e.g., PCP assignment) and pharmacy cost. General considerations of the components, structure, and format of this file are discussed in more detail

in the next section of this chapter. Control card keywords are used to indicate the starting and ending columns of the various pieces of data required in the input file.

(3) INPUT2

The control card for INPUT2 is functionally identical to INPUT1 and is invoked as follows:

INPUT2=*filename*

where *filename* is the complete path and filename for the second input file. The second input file, while similar to the first, is typically limited to diagnosis information only (but occasionally is used to carry user-defined flags for pregnancy, delivery, and low birth weight status).

(4) NONMATCH

The NONMATCH card is used to specify the file to which unclassifiable diagnoses are to be written. Analysis of this file, consisting of unique combinations of member ID and diagnosis codes not found in the ICD-9 to ADG mapping, is often useful for evaluating the quality of the input diagnosis and/or effectiveness of the assignment algorithm. The format of this control card is as follows:

NONMATCH=*filename*

where *filename* is the complete path and filename for the nonmatch file.

(5) OUTPUT

The OUTPUT card is used to specify the file into which ACG assignments (and, at the user's discretion, ADG and other assignments) are to be written and is of the form:

OUTPUT=*filename*

where *filename* is the full path/directory and name into which this file is to be written.

(6) PRINT

The general summaries of the run, warnings, and error messages are written to the designated print file. The format of this control card is as follows:

PRINT=*filename*

where *filename* is the complete path and filename. Discussed in more detail in a subsequent section, the print file includes summary statistics and simple statistical reports on the ADG, ACG, and EDC distributions of this population.

(2) Record Layout

(1) AGE

This card is used to specify the position and length of the age on the INPUT file (discussed in more detail later in this document) as follows:

AGE=*start,length*

where *start* and *length* indicate starting columns and number of columns to read from the input. The age must be right justified or incorrect results may occur. Note that either an AGE or a DOB card can be specified, but not both. If two input files are used, the AGE code must appear on the first file.

(2) DOB

This card is used as an alternative to the AGE card. It specifies that the date of birth is to be used in the age calculation as follows:

DOB=*start,format,agedate*

where *start* indicates the starting column from the input, and *format* indicates the format of the date as follows:

- 1=YYMMDD
- 2=CCYYMMDD (CCYY=century & year, e.g., 1990)
- 3=MMDDYY
- 4=MM/DD/YY
- 5=MM/DD/CCYY
- 6=YYDDD (Julian date)
- 7=YYMM (a day of 01 is assumed)
- 8=CCYYDDD (Julian date)

If the DOB card is used then an *agedate*, indicating the date from which the age is to be computed, must be provided and the format must be CCYYMMDD. Note: Either an AGE or a DOB card can be specified, but not both. If two input files are used, the DOB code must appear on the first file.

(3) FEMALE

This card is used to recode the sex for females as follows:

FEMALE="code"

The optional *code* indicates the code on the input to consider female. This card is used to override the default assumption of 2 or F for female. All other codes are considered male. To recode blanks, insert the proper number of blanks between the two quotation marks. There may be more than one FEMALE card to supply additional codes. This code is found on the first file, if two input files are used.

(4) ICD

This card is used to control the position and length of the ICD-9 codes as follows:

ICD=*start,length,number*

where *start* is the starting column of the input, *length* is the length of each field, and *number* is the number of consecutive codes. If *number* is not specified, then the software assumes there is only one diagnosis code per line. The user can specify more than one ICD card, a useful option if the diagnosis codes are in different positions on the same record. If two files are used, ICD codes should always be on the second file.

(5) ID

This card is used to specify the position and length of the person ID as follows:

ID=*start,length*

where *start* and *length* indicate starting columns and number of columns to pick up from the input. Note that the ID fields can exist in multiple columns and do not have to be contiguous on a record. Multiple ID cards would be used to define the position and lengths of the parts of a noncontiguous ID.

(6) ID2

This card is used to specify the position and length of the person ID on input file 2 as follows:

ID2=*start,length*

where *start* and *length* indicate starting columns and number of columns to pick up from the input. Note that the ID does not have to be contiguous on a record. If only one input file is used, then the ID2 card is not needed. If an ID2 card is specified, but no second input file is specified, then a warning message is given. If a second input file is specified but no ID2 card is included, then the ID position in file 2 is assumed to be the same as the ID position in file 1.

(7) OUTREC

This card is used to control the format of the OUTPUT record as follows:

OUTREC=*field,field,field,...*

where the *fields* are used to specify the order of the output. If no OUTREC control card is used, the default is to write out only the ID and ACG. Any of 14 predefined fields or any user-defined field (see the FIELD control card) may be specified in any order. The predefined field names are: ACG, ADG, AGE, CWT, DEL, DOB, HOS, HRCI, ICD, ID, LOW, POP, PCOST PRE, RUB, and SEX. Except for the ICD-9 codes, the fields will be written out in the order specified. If ICD-9 codes are indicated for output, they will always be written at the end of the output record regardless of where they are specified on the OUTREC card. Please see the separate section subsequent in this document entitled "Output Files" for additional details on what is written to the output files produced by the software.

(8) SEX

This card controls the position and length of the sex field as follows:

SEX=*start,length*

where *start* and *length* indicate starting columns and number of columns. Note: Sex codes are always found in the first input file.

(3) Optional Fields

(9) DELIVERED

This card instructs the system to classify patients who are associated with a pregnancy according to whether they delivered during the period of observation. If the DELIVERED card is used, ACGs associated with pregnancy (those beginning with '17') will end with '1' (delivered) or '2' (no delivery). The format is:

DELIVERED

If the DELIVERED card is not used, the software will not assign an ACG that specifies delivery status, and the ACGs associated with pregnancy (those beginning with '17') will end in '0'.

The user may also use this control card to specify the location of a user-supplied flag on the first input file (INPUT1) indicating that the woman has delivered a baby during the period of observation. The format is:

DELIVERED=*start*, "*code*"

where *start* is the starting column number on the input record of the data used to indicate a delivery, and "*code*" is the value of the data that indicates a delivery. For example, DELIVERED=106,"1" indicates that if column 106 in the input data contains "1", the patient delivered a baby. There is no limit to the number of DELIVERED cards that can be used.

Note: If the user-supplied flags indicate a woman delivered, but there is no other indication that she was pregnant, then a warning message is printed, and the woman is automatically considered to have been pregnant.

(10) DELIVERED2

This card is identical to the DELIVERED control card, except that it is used to identify values in data file 2, when two input data files are used.

(11) FIELD

This card specifies a user-defined field as follows:

FIELD=*field name*,*start*,*length*

where *field name* is a user-defined name for this field. The user-defined name cannot be one of the predefined names (see OUTREC card description). *Start* and *length* indicate the starting column and number of columns for the field. A field name can be defined only once. User-defined fields are picked up from the INPUT1 file only. Field names may then be used on the OUTREC card to write those fields to the output record. Examples of using the field card include picking up and carrying through to the output record a population stratifier variable or some resource measure, such as total charges, thereby enabling further analysis of the output file without having to re-merge these data. Note a few restrictions on the field name: 1) it must be 10

characters or fewer, 2) it must start with an alphabetic letter, and 3) it can contain only letters and numbers.

(12) LOWWEIGHT

This control card instructs the system to classify infant patients according to whether they had low birth weight, i.e., less than 2,500 grams. To do this, the user must specify the location of a user-supplied flag on the first input data file that indicates that a newborn had a low birth weight. This flag should only be used for patients who are infants (1 year or younger) during the observation period. The format is:

LOWWEIGHT=*start*,"*code*"

where *start* is the starting column number on the input record and "*code*" is the value of the data used to indicate low birth weight. For example, LOWWEIGHT=107,"1" indicates that if column 107 is "1", the infant should be considered to have had low birth weight. There is no limit to the number of possible LOWWEIGHT cards. Please note that the use of the LOWWEIGHT card is optional. If the LOWWEIGHT card is not used, the software will not assign an ACG that specifies birth weight status.

(13) LOWWEIGHT2

This control card is identical to the LOWWEIGHT control card, except that it is used to identify the low birth weight flags on the second input file, when two input data files are used.

(14) POP

This card is used to specify the position and length of the population stratifier field as follows:

POP=*start,length*

where *start* and *length* indicate starting columns and length. This field is treated as a character field and may contain any data.

(15) PCOST

This card is used to specify the position and length of pharmacy cost field as follows:

PCOST=*start,length*

where *start* and *length* indicate starting columns and number of columns. The PCOST must be all numeric (leading blanks are allowed). It is assumed to be a whole dollar amount (no cents). If it does not meet these conditions, it is set to zero and a WARNING09 will appear in the warning message section of the print file.

(16) PREGNANT

This control card specifies the location of a flag on the first input file indicating that a woman is pregnant during the period of observation. This flag would be necessary only if comprehensive pregnancy diagnosis data were not available during the period of analysis. For example, some health plans using a global fee for all perinatal services may not have pregnancy diagnoses in their claims data until after delivery has occurred. The format is:

PREGNANT=*start*, "*code*"

where *start* represents the starting column number on the input record of the data used to indicate a pregnancy and "*code*" is the value of the data that indicates pregnancy. For example, PREGNANT=105,"1" specifies that if column 105 is "1", the woman should be considered pregnant. Any other entry indicates that the woman is not pregnant. There is no limit to the number of PREGNANT= cards that can be used at one time. If this card is not used, and no PREGNANT2= card is used (see below), then the software will, by default, use the ICD-9 codes alone to classify patients according to pregnancy status.

(17) PREGNANT2

This control card is identical to the PREGNANT= control card, except that it is used to identify the pregnancy values from data file 2, when two input data files are used.

(4) Miscellaneous

(18) LRECL1

This card is used to specify the record length for input file 1 in the case of fixed-length input records. It is needed only if the input is a fixed length with no end-of-record mark—normally a carriage return and line feed—at the end of each record. It is useful for inputting data directly from database files, which usually store data as fixed length records, or after a header with no end-of-record mark. The format of this card is:

LRECL1=*length,skip*

where *length* is the length of each record, and *skip* is the number of bytes to skip at the start of the file. At the end of the file, if a complete record cannot be read, then the trailing bytes are ignored.

(19) LRECL2

This card is used to specify the record length for input file 2 in the case of fixed-length input records. Its format is identical to that of the LRECL1 card described above.

(20) NOCRLF

This card is used to specify that output records be written without an end-of-record mark. The format for this card is:

NOCRLF

(21) NOWARN

This card is used to adjust the warning threshold. Normally, the software will abort execution if warnings are produced for more than 20 consecutive people in the first half of the input file. If a NOWARN card is included in the control card file, then although warnings will still be generated, execution will not abort if this threshold is reached. Its format is:

NOWARN

d. Required Components of the Input File

The following are general considerations:

- The software accepts only flat ASCII (or .txt) files.
- A record may contain more than one diagnosis code. Blank diagnosis codes are simply ignored.
- If only one input file is specified, then it is assumed to contain the ID, age or DOB, sex, and ICD-9 codes. This file may contain multiple records per ID. The age and sex are picked up from the first record for a particular ID. When two input files are specified, the first must contain the ID, age or DOB, and sex, and the second must contain the ICD-9 codes and an ID. If two input files are used, then the first should

contain only one record per unique ID, whereas the second can contain multiple records per ID.

- The system can use either the age or the date of birth. If the date of birth is supplied, then a date (the AGEDATE) from which to compute the age MUST be supplied on the DOB control card.
- The age/DOB and sex are extracted from the first record for each individual in the data set. If a person has more than one record, the age/DOB and sex on subsequent records are ignored. Preliminary validation of age, or the month and day of a date of birth code, should be performed. The software will stop execution if more than 20 consecutive people are identified as having no demographic information or being "older" than 107 years of age in the first half of the input data file.
- Special sex codes can be indicated on a control card.

i. Augmenting Diagnosis Data

Certain aspects of the ACG sorting algorithm are optional and are intended to help the user make a better assessment of the case-mix patterns of the study population. These options are initiated by user-supplied flags for pregnancy, delivery, and/or low birth weight, which serve to augment the diagnosis information typically found in claims or encounter data. Please note that the use of user-supplied data for classifying pregnant women and infants is *optional*. The discussion here complements that outlined in the *Version 5.0 Documentation and Application Manual*, Chapter 7, "Basic Data Requirements for ACG Categorization and Analysis," and discusses the technical 'how to' of this process.

ii. Pregnancy and Optional User-supplied Information on Pregnancy Status

By default, the ACG software searches for a defined set of ICD-9 codes to determine pregnancy status. The diagnosis codes used to indicate a pregnancy include: 640xx-677xx, V22x, V23x, V24x, V27x, and V28x. Individuals with a diagnosis in this range, and who are not otherwise excluded on the basis of age or sex, are assigned an ACG beginning with '17.' If an individual's diagnoses indicate pregnancy but she is younger than 10 years or older than 55, a warning is issued that the patient is of a suspicious age to be pregnant. Similarly, if an individual's diagnoses indicate pregnancy but that person is coded as male, a warning is issued of suspicious sex to be pregnant. Those who are coded as male and those who are younger than 5

years or older than 60 (regardless of sex) are assumed to have received a pregnancy diagnosis in error and will not be assigned an ACG beginning with '17.'

Alternatively, the user may decide that available ICD-9 code data are insufficient indicators of pregnancy status. That is, in some cases patients may actually be pregnant when no ICD-9 code is present. User-supplied 'flags' can be defined by using control cards (designated PREGNANT or PREGNANT2 and discussed later in this chapter) that serve to augment the ICD-9 diagnosis data. If a user-supplied 'flag' indicates pregnancy on the input file, the ACG software will classify such a patient as pregnant regardless of any other indication of pregnancy (including typical exclusions due to inappropriate age and sex). Warnings of suspicious age and/or sex may be issued but the user-defined information will be assumed to be correct. If the ICD-9 data indicate pregnancy when a user-supplied flag does not, the individual will also be assumed to be pregnant as long as she meets the default criteria for age and sex (i.e., female between the ages of 5 and 55). In other words, a user-supplied indication of pregnancy serves as a supplement to the diagnostic data, rather than as a strict replacement.

iii. Delivery Status and Optional User-supplied Information on Delivery Status

Because of the significant increase in resources used in association with delivery, the user may want to specify that the ACG system categorize pregnant women as either having delivered or not having delivered during the period of observation. Unless otherwise instructed by the user, the ACG system will not automatically attempt to classify pregnant women by delivery status. However, two control cards, DELIVERED and DELIVERED2, can be used to instruct the software to search for either ICD-9 diagnosis codes or a user-supplied flag indicating delivery status. As with the classification of pregnancy status, the user-supplied data serves as a supplement to ICD-9 coding indicating delivery. The ICD-9 codes that the software uses to subdivide these ACG categories include: 664.0, 664.01, 664.11, 664.21, 664.31, 650, 654.21, 656.31, 658.11, 658.21, 658.31, 661.01, 661.11, 661.21, 661.31, 663.11, 663.31, 669.5, 669.7, 669.70, V27* (where * = 0-9).

iv. User-supplied Information on Birth Weight Status (Optional)

When adequate information is available, the user may also want to specify that the ACG software categorize infants as having either low or normal birth weight, in order to enhance the case-mix adjustment process. Unless otherwise instructed by the user, the ACG system will not

automatically attempt to classify infants by birth weight status. However, either one of two control cards, LOWWEIGHT and LOWWEIGHT2, can be used to instruct the software to search for a user-supplied flag indicating birth weight status. Unlike the DELIVERED control card discussed above, the software will not classify patients as low or normal birth weight on the basis of ICD-9 codes.

v. ICD-9 Codes and Decimal Points

The recoding and handling of explicit or implicit decimal points is automatic and cannot be modified by the user. If *no decimal point* is detected in the ICD-9 code, then the software assumes that a) the decimal point is implied after the third character, b) the first three characters of the diagnosis code are right justified, and c) the fourth and fifth characters are left justified. For example, "4011 " is assumed to be 401.1, "0112 " is assumed to be 011.2, and "112 " is assumed to be 112. If the input ICD-9 code *does contain a decimal point*, then the software will take up to three characters to the left of the decimal point and right justify them with leading zeros. It will then take up to two characters to the right of the decimal point and left justify them. For example, "401.1" will be recoded to 4011, "11.2" will be recoded as 0112, and "112." will be recoded as 112.

e. Output Files

There are potentially four output files produced by the software, including 1) a print file containing summary statistics of the run; 2) an output file containing (at a minimum) individual level ACG assignment; 3) an optional Expanded Diagnosis Clusters output file containing each unique combination of member IDs and EDC assignment; and 4) a nonmatched file containing unique combinations of member IDs and diagnosis codes not identified by the software.

i. Print File

The print file of the software provides the following:

- A copy of the control cards used for that run.
- A listing of the input and output files defined for that run.
- Selected summaries of input and output data including:
 - the number of input records on the first input file;
 - the number of input records on the second input file;
 - the number of output records (this is equivalent to the number of persons

- appearing in either file one or file two);
 - the sum of unique diagnosis codes across all input records;
 - the sum of unique diagnosis codes across member IDs;
 - the sum of unique non-grouped diagnosis codes across member IDs;
 - non-grouped diagnosis code percentage; and,
 - number of people with non-grouped diagnosis codes.
- Frequency and percentage distributions of ACGs and ADGs.
- Selected summary statistics of the distribution of ADGs including:
 - the number and percent distribution of people with numbers of ADGs ranging from zero to 10 or more per person;
 - the number and percent distribution of people with numbers of ACGs ranging from 1 to 4 or more per person;
 - the average number of ADGs per person;
 - the average number of major ADGs per person;
 - the average number of ADGs for those who have an ADG;
 - the average number of major ADGs;
 - the average number of ADGs for those with an ADG;
 - the average number of major ADGs for those with an ADG;
 - the average number of ADGs for those with a major ADG; and
 - the average number of major ADGs for those with a major ADG.
- Selected summary statistics and distribution of unique diagnoses including:
 - the average number of unique diagnosis codes per person;
 - the average number of unique diagnosis codes per person for those with diagnoses; and
 - the number and percent distribution of people with unique diagnosis codes ranging from zero to 10 or more per person;
- Age/Sex Distribution.
- Frequency distribution of EDCs and MEDCs and prevalence per 1,000 persons.
- Average numbers of EDCs per person overall and for those with EDCs.
- Percentage distribution of EDC and MEDC in 5 resource utilization bands (RUBs)¹⁰.
- Estimated concurrent resource use by EDC and MEDC in 5 RUBs.
- Observed and age/sex-adjusted prevalence, standardized morbidity ratios (SMRs), and SMR confidence intervals by MEDC.
- Frequency and percent distribution of acgPM probability scores.
- Number of cases and associated predicted acgPM relative resource use by

¹⁰ Note: the non-user RUB is excluded.

alternative acgPM risk probability thresholds for selected chronic conditions.

Warning Messages

-- Warnings associated with each individual who generates a warning are listed on one line per person at the end of the print file. The warning lines include the ID, ACG assignment, age, sex, date of birth (if entered as input), and a list of warnings numbers. The warning numbers can be compared to the text for each warning printed above the individual-level warning data.

ii. Output File

The output record produced by ACGGROUP is controlled through the OUTREC control card (discussed previously). The fields will be written out only if specified, and in the order specified. The only exception to this rule of order is that ICD-9 codes will be written at the end of the record. The field lengths are as follows:

<u>Name</u>	<u>Field</u>	<u>Length</u>
ACG	ACG code	4
ADG	ADG vector	34 one-character fields for ADG 1–34 (where a 1=the person had this ADG; 0=otherwise) (Please note: Because they are no longer in use, the values for ADG 15 and ADG 19 will always be 0.)
AGE	Age	3 (right justified, leading zeros)
CWT	Concurrent weight	Based on a nationally representative database. Written to the output file as ##.###.
DEL	Delivery status	1=person delivered; 2=person did not deliver; 9=optional branching turned off
DOB	Date of birth	8 in CCYYMMDD format if possible
ICD	ICD-9 codes	6 characters each (where the first 5 are the ICD-9 code, and a mismatch flag is in the sixth position [1=no ADG code and blank=successfully matched])
ID	ID code	Same as designated for input
HRCI	acgPM scores	All written to the output file as ###.### in the following order: 1) <i>Total cost predicted resource index</i> – an estimate for Year 2 total expenditures (including pharmacy charges) expressed as a relative weight;

- 2) *Pharmacy cost predicted resource index* – an estimate for Year 2 predicted pharmacy expenditures also expressed as a relative weight;
- 3) *Probability of being in the high total cost cohort* – a probability score with values between zero and one indicating the likelihood that a person will have high cost in the subsequent time period; and
- 4) *Probability of being in the high pharmacy cost cohort* – a probability score indicating the likelihood that a person will have high pharmacy cost in the subsequent time period.

HOS	Hospital Dominant	A Boolean indicator (e.g., a value of zero or one) for the presence of conditions likely to lead to a hospitalization.
LOW	Low birth weight	1=low birth weight; 2=normal birth weight; 9=optional branching turned off
PRE	Pregnancy status	1=person is pregnant; 2=person is not pregnant;
RUB	Resource utilization band	0 = nonusers 1 = healthy users 2 = low morbidity 3 = moderate morbidity 4 = high morbidity 5 = very high morbidity
SEX	Sex code	1 (F=female, M=male)

User-defined fields specified on the FIELD card can also be specified for output. The length of user-defined fields is the same as specified by the user on the FIELD cards. If no OUTREC card is included, then the software automatically writes the unique member ID and ACG to the output file.

Because the software is not limited to a maximum number of unique ICD-9 codes per person, the number of codes on the output will depend on the number of unique ICD-9 codes for any particular patient. Only unique codes are written to the output. Even if codes are duplicated on the input, only one occurrence of each code is written to the output in the order read. If ICD-9 codes are indicated for output, they will always be written at the end of the output record regardless of where they are specified on the OUTREC card. There is one output record produced for each unique ID.

iii. Differentiating Nonusers from Those with Only One Input File

If only one input file is used to enter both demographic and diagnostic data, all patients without associated diagnoses or without groupable diagnoses are assigned to ACG 5100. To more cleanly separate nonusers from those who used services but had no classifiable diagnoses, it is recommended that individuals classified into ACG 5100 be subdivided into two groups according to the following criteria. Nonusers, defined as individuals with no claims or encounters, should be reassigned to ACG 5200. This will leave only those individuals in ACG 5100 (which can be renamed 5110) who, although they are users of services, have no diagnosis and/or *only* unclassified diagnoses.

iv. The EDC File

The Dino-Cluster output file is optional and will be produced by the software only if the EDC control card is present. The format of this file is as follows:

<u>Columns</u>	<u>Field</u>
1-5	EDC code
6-11	space
12-n	Person ID where n depends on the length of the ID n=11+length of ID

v. Mismatched ICD-9 List

All input ICD-9 codes without a corresponding ADG code will be written to an output file called the mismatched ICD-9 file. Because all currently defined ICD-9-CM codes are now included in the ICD to ADG mapping, this mismatch file will typically be limited to: 1) 'E' codes that, because they typically indicate the cause of an injury rather than an underlying morbidity, are excluded from the mapping; 2) local coding that is not otherwise included; and 3) genuine errors in coding. Mismatched codes are written out one time for each person who has that code. In this way, the file can be used to get a listing of codes and/or a listing of people who have mismatched codes. It can also be used to count the number of patients with each mismatched code. The format of this file is as follows:

<u>Columns</u>	<u>Field</u>
1-5	ICD-9 code
6-n	Person ID where n depends on the length of the ID n=5+length of ID

This file should be analyzed carefully. Briefly, using this file to create a frequency distribution of non-grouped ICD-9 codes can help identify diagnosis codes that are not in standard format. Historically, the five most frequently occurring codes will account for more than 50% of all codes not identified by the software. How the codes are justified and/or padded and whether decimal places are explicit or implied are the two leading causes of problems pertaining to mismatched codes. For a more theoretical discussion of ICD-9 coding considerations, see Chapter 7, "Basic Data Requirements for ACG Categorization and Analysis," of the *Version 5.0 Documentation and Application Manual*.

The software will assign ADGs to "legal" ICD-9-CM and ICD-9 (non-CM) diagnostic codes. In certain circumstances, ADGs will also be assigned to "illegal" codes, where the intent of the codes was clear (for example, a three-digit code that is unambiguous, but where only a four- or five-digit code is considered valid according to official ICD-9 publication vendors).

vi. Error Messages

Error messages generated by ACGGROUP are either written to the print file or displayed on the screen. Errors are basically of two types: those that involve the control cards and those that involve the data. Each message is numbered. The following sections describe the error messages in order by message number for each of the two types of errors/warnings. Most messages are self-explanatory and can be resolved by checking the documentation for the specific control card in error. Further clarification is supplied below, as appropriate.

vii. Control Card Errors and Warnings

The control cards are written to the print file in a numbered list. Control card error messages will, many times, contain a number that refers to the specific control card in that list. While many error messages are generic to multiple control cards, by using the control card number and referring to the list of control cards in the print file, you can easily identify the specific control card in error. For example, given the following control cards:

```
id=x,11  
age=12,3  
sex=15,1  
icd=21,6
```

The print file would list the control cards as follows:

```
1: id=x,11
```

2: age=12,3
3: sex=15,1
4: icd=21,6

The following error message would also be printed:

ERROR02: Parameter 1 is missing or non-numeric on control card: 1

This message indicates that on control card #1, which is the ID card, the first parameter, which is an "x," is non-numeric. It should indicate the starting column of the patient ID.

The following is a complete list of the control card error messages

1. ERROR01: Two control cards were found and only one is allowed: #
Certain control cards are allowed to be used only once per run. The software found more than one of these cards. The number given in the error message points to the control card in error.
2. ERROR02: Parameter 1 is missing or non-numeric on control card: #
The control card referenced has an invalid first parameter, which is usually the starting column number.
3. ERROR03: Parameter 1 was not specified on control card: #
4. ERROR04: Parameter 2 is missing or non-numeric on control card: #
The control card referenced has an invalid second parameter, which is usually the length.
5. ERROR05: Parameter 2 was not specified or is invalid on control card: #
6. ERROR06: Value must be enclosed in quotes on control card: #
7. ERROR07: Could not obtain memory to hold control cards.
Your system does not have sufficient free RAM to read in all of the control cards. Eliminating other programs that are running at the same time as ACGGROUP or adding more system memory will clear up this problem.
8. ERROR08: Not enough memory to continue program:
Your system does not have sufficient free RAM. Eliminating other programs that are running at the same time as ACGGROUP or adding more system memory will clear up this problem.
9. ERROR09: Both AGE and DOB control cards were found and only one is allowed:
10. ERROR10: Parameter 3 is non-numeric or not specified on control card: #
11. ERROR11: The date format code specified on the DOB card is invalid: #
12. ERROR12: Age date is non-numeric on control card: #