

USING SAS SOFTWARE IN PHARMACOEPIDEMIOLOGIC RESEARCH: IDENTIFYING EPISODES OF DRUG USE AND DETERMINING AVERAGE DAILY DOSE

Charlotte Corelle, Kaiser Permanente Center for Health Research
Marjorie Bennett, Kaiser Permanente Center for Health Research
Richard E. Johnson, Kaiser Permanente Center for Health Research
Bentson McFarland, Kaiser Permanente Center for Health Research

ABSTRACT

This paper describes how we used base SAS software to transform pharmacy records of all dispensings of a particular drug to records of all episodes of use of that drug. The challenge was to develop a sound and efficient process that would correctly link dispensings into episodes. This paper presents a flexible algorithm for constructing episode records from dispensing records and displays the code for implementing the algorithm. The algorithm uses variables common to many pharmacy databases (i.e., user number, product number, dispensing date, dispensed quantity, strength, and day's supply) to set the boundaries of a user's dose and duration for each period of exposure to a particular drug. The flexibility of the algorithm comes from its use of some function of either quantity and/or day's supply and/or milligrams of drug dispensed to decide whether or not to link one dispensing to another as part of the same drug use episode.

INTRODUCTION

This paper describes a SAS software application that transforms a pharmacy dispensing data set to a data set containing one record for each period of drug exposure for a drug user and calculates the average daily dose during each exposure period. The resulting episode-based file is important because it can be used to produce tables describing utilization patterns of a particular drug. In addition, after selecting appropriate controls, the file can be used to study the duration and degree of exposure to the drug as they relate to potential drug effect, whether adverse or beneficial.

RESEARCH ENVIRONMENT

We developed the algorithm and code at the Center for Health Research (CHR) of Kaiser Permanente, Northwest Region (KPNW). The work was done in conjunction with CHR's participation in a Cooperative Agreement with the Food and Drug Administration to conduct studies on adverse effects of marketed drugs.

The CHR uses KPNW as a research laboratory. KPNW is a health maintenance organization (HMO) with 378,000 members located in the Portland, Oregon and Vancouver, Washington metropolitan areas. The HMO has had an

automated outpatient pharmacy system since 1986. The database records more than two million outpatient prescription annually, providing extensive information on each dispensing. The database, in turn, can be linked to an automated hospital admission and discharge database, an outpatient utilization and morbidity database, a tumor registry, and a KPNW membership eligibility database. Across all databases a single unique identifier is assigned to each patient. We can select all dispensing records for drugs of particular interest to the FDA. Then, using the unique identifier from the dispensing record, we can map to the other databases to get demographic, morbidity, and health care data for the drug users.

RESEARCH OBJECTIVE

Our work with the FDA required us to develop a method for processing pharmacy dispensing data (one record per user per dispensing per date) so as to identify a drug user's periods of continuous exposure to any specified drug. A single period of drug use is called an episode of use, and an episode can encompass multiple dispensings of the drug if the dispensings are sufficiently close. Each dispensing record contains unique information (such as quantity of product units dispensed, or strength of drug product, or day's supply of drug dispensed) that suggests a window of days during which the drug is being used. If a second dispensing occurs within that window, it is likely that the dispensings reflect a period of continuous use.

We needed an algorithm to compare the number of days between dispensings to Δ , such that Δ could represent an arbitrary number (e. g., 10 days), or it could represent some function of the value of one or more variables on the dispensing record. After correctly grouping dispensings into episodes, our goal was to determine the average daily dose during each of a user's episodes.

METHOD

This paper describes only how to identify episodes and determine average daily dose with a data set containing all dispensings of the particular drug(s) of interest. Figure 1 shows the variables to include in your initial dispensing data set.

Figure 1

patient id# (unique identifier of a person)
dispense date
prescription number
refill number
quantity dispensed
strength of drug product
day's supply (#days this quantity should last)

Before beginning to identify episodes you will need to prepare your dispensing data set. Running frequencies will indicate missing or out-of-range values; then recode variables and/or adjust the episode algorithm if necessary. For example, perhaps day's supply in the data is zero or missing in a few cases and quantity dispensed is always a positive, non-zero integer. If so, you may decide to recode day's supply (when zero or missing) to the value of quantity dispensed. Also, for each dispensing record, create a variable that calculates the amount of drug delivered (multiply the quantity of product units dispensed by the product unit strength). Collect all of a user's dispensings and sort them in ascending date order.

The integrity of the episode identification method relies to some extent on there being no more than one dispensing record per patient on a given dispensing date. Sometimes drug users pick up more than one prescription fill on the same date, however. Because this generates multiple dispensing records, collapse multiple dispensings on the same date into a single dispensing that summarizes the amount of drug delivered across the multiple dispensings. In preparing our data, we summarized the amount of drug delivered, the quantity dispensed, and the day's supply when multiple dispensings on a given date are refills of the same prescription number. For multiple dispensings of different prescription numbers we summarized amount and quantity but used the highest day's supply value across the multiple dispensings.

After multiple dispensing records for the same person on the same date are eliminated, the dispensing file still should remain sorted by patient identifier and dispensing date. Figure 2 contains the set of variables you need to begin the process of grouping prepared dispensing records into episodes.

Figure 2

patid (patient id#)
dsdate (dispense date)
deliv (amount of drug delivered)
dayssup (day's supply)
dispquan (quantity dispensed)

Figure 3 displays example code for the important next step which implements the algorithm for linking a drug user's dispensings into episodes. In this DATA step we create an episode incrementing variable called ADDEP and give it a value of one to identify a dispensing record that begins a

new episode or a value of 0 to identify a dispensing record that is part of a continuing episode or is a user's first dispensing record. The incrementing variable gets its value by resolving a compound expression. In this example, the expression subtracts the dispensing date on the previous record from dispensing date on the current record and compares the result to the sum of day's supply on the previous record added to one-third the quantity dispensed on the previous record. If the expression is true, its value resolves to one and indicates the start of a new episode. Otherwise, its value resolves to zero. You can change the expression to change the criteria for linking dispensings into episodes.

In the same DATA step, we create an episode counting variable called EPNUM and give it a value of one on a user's first dispensing record. The final line of code allows episode number to be retained from record to record of a user, incrementing only according to the value of the episode incrementing variable during each iteration of the DATA step.

Figure 3

```
data step1;  
  set dispensings  
  (keep=patid dsdate deliv dispquan dayssup);  
  by patid dsdate;  
  
  addep=  
    ((dsdate-lag(dsdate)) gt  
     (lag(dayssup)+(lag(dispquan/3))));  
  
  if first.patid then do;  
    addep=0;  
    epnum=1;  
  end;  
  
  epnum+addep;  
run;
```

Our final programming step generates a file containing only one record for each of a drug user's episodes. This step requires us to establish a beginning and ending date for each episode and to summarize the total amount of drug delivered across all dispensings of an episode. When those tasks are done, we can calculate average daily dose for each episode. The variables held in the resulting episode-based file are displayed in Figure 4.

Figure 4

patid
epnum (episode number)
epbdate (episode begin date)
epedate (episode end date)
epdeliv (drug amount delivered over the episode)
epavdose (average daily dose during episode)

Figure 5 shows an example of code used to implement the final step of the conversion process from a dispensing-based file to an episode-based file. To begin this step the data must be sorted by patient identifier, episode number, and dispensing date. To set the episode end date, retain the first dispensing date for that episode number as the episode begin date. To set the episode end date in this example, we added the value of day's supply on the final record of an episode to the final dispensing date. You may wish to change the criteria for determining when an episode ends. For example, you might add the greater of day's supply or one-third of quantity dispensed to the final dispensing date in order to establish the episode end date. On the other hand, you might want to implement a rule that says all episodes end on some fixed number of days after the last dispensing.

Figure 5

```

data episodes (keep=patid epbdate epdate epdeliv
epnum eplength epavdose);
set step1(keep=patid epnum dsdate deliv dispquan
dayssup);
    by
        patid epnum dsdate;

retain epbdate epdeliv;

if first.epnum then do;
    epbdate=dsdate;
    epdeliv=deliv;
end;

else epdeliv=epdeliv+deliv;

if last.epnum then do;
    epdate=dsdate+dayssup;
    eplength=epdate-epbdate;
    epavdose=round((epdeliv/eplength),.01);
    output episodes;
end;

```

CONCLUSION

An episode-based file is of value because it allows for investigation of effects associated with dose and length of use of a drug. An episode-based file such as the one produced by our method can be joined with demographic data about the users to produce descriptive tables or to select controls to test for potential drug effects. The benefit of the method described is its flexibility with an algorithm that can be modified according to particular research questions and drugs of interest.

ACKNOWLEDGMENTS

SAS is a registered trademark of SAS Institute Inc. in the USA and other countries. " Indicates USA registration.

Other brand and product names are registered trademarks or trademarks of their respective companies.

AUTHOR CONTACT

Charlotte Corelle
 Kaiser Permanente Center for Health Research
 3800 N. Kaiser Center Drive
 Portland, Oregon 97227
 (503)335-6740