

Paper 214-27

SPAN SPAN AWAY! Creating One Unique Record for Overlapping Admissions and Discharges from Multiple Inpatient Hospital Stays

Shabnam Mehra, University of South Florida, Tampa, FL

Robyn Spittle, University of South Florida, Tampa, FL

Becky Larsen, University of South Florida, Tampa, FL

Diane Haynes, University of South Florida, Tampa, FL

Kathy Dailey, University of South Florida, Tampa, FL

ABSTRACT

The Policy and Services Research Data Center (PSRDC) in the Department of Mental Health Law & Policy at The Louis de la Parte Florida Mental Health Institute performs research with a variety of secondary mental health datasets. The data format is often manipulated to accommodate our research objectives. One example of this manipulation is "span coding" using SAS®. This code is specifically designed to transform overlapping inpatient hospital stays into one unique record by means of the admission and discharge dates. The techniques in this paper apply to health care agencies and the pharmaceutical industry but may easily be applied to other disciplines. Obtaining a unique record for inpatient hospital stays using multiple admission and discharge dates can be easily accomplished by using IF..THEN..ELSE..DO statements with retain statements. It is a simple and swift method to manipulate secondary data to a desirable format for answering research questions.

INTRODUCTION

The Policy and Services Research Data Center (PSRDC) in the Department of Mental Health Law & Policy at The Louis de la Parte Institute Florida Mental Health Institute performs research on a variety of datasets related to services provided to persons with mental health concerns. The PSRDC receives secondary data from many agencies and often the layouts of the data from these agencies are different. The analyst must have in depth knowledge of the data layout and the purpose for which the data was primarily collected. The intent of secondary analysis is to extract pertinent mental health policy directives from data that was collected for other objectives. For example, mental health claims data that was collected for accounting purposes can be used judiciously to answer mental health policy research questions. Often the data layout must be manipulated to accommodate the research objectives. For instance, each claim in the datasets' original form may represent a billing period rather than a treatment period, which would be more useful for research purposes. Close management of variables such as dates, cost, and services is mandatory when using secondary data to answer research questions. For example, overlapping inpatient hospital stays will need to be modified to one record for research intentions. This can be easily attained by using an IF..THEN..ELSE..DO statement to determine if an observation meets specified criteria.

THE CHALLENGE

The code described here is performed on secondary data containing inpatient hospital stays. The code uses admission and discharge dates to determine service utilization within the mental health system. We are specifically interested in utilities such as length and cost of the inpatient stay. If only one date variable is provided per record in a table and there are multiple records, then an array code would be appropriate. Array coding would also be useful if admission dates and discharge dates are in two separate tables each having multiple records per person. However, if two date variables are available in the same table, such as admission and discharge dates, then span coding is more suitable. A span

can be defined as the length between unique admissions and discharges to assist in the production of a unique record per person that replaces overlapping admissions and discharges in the dataset. The techniques in this paper apply to health care agencies and pharmaceutical industry but may easily be applied to other disciplines.

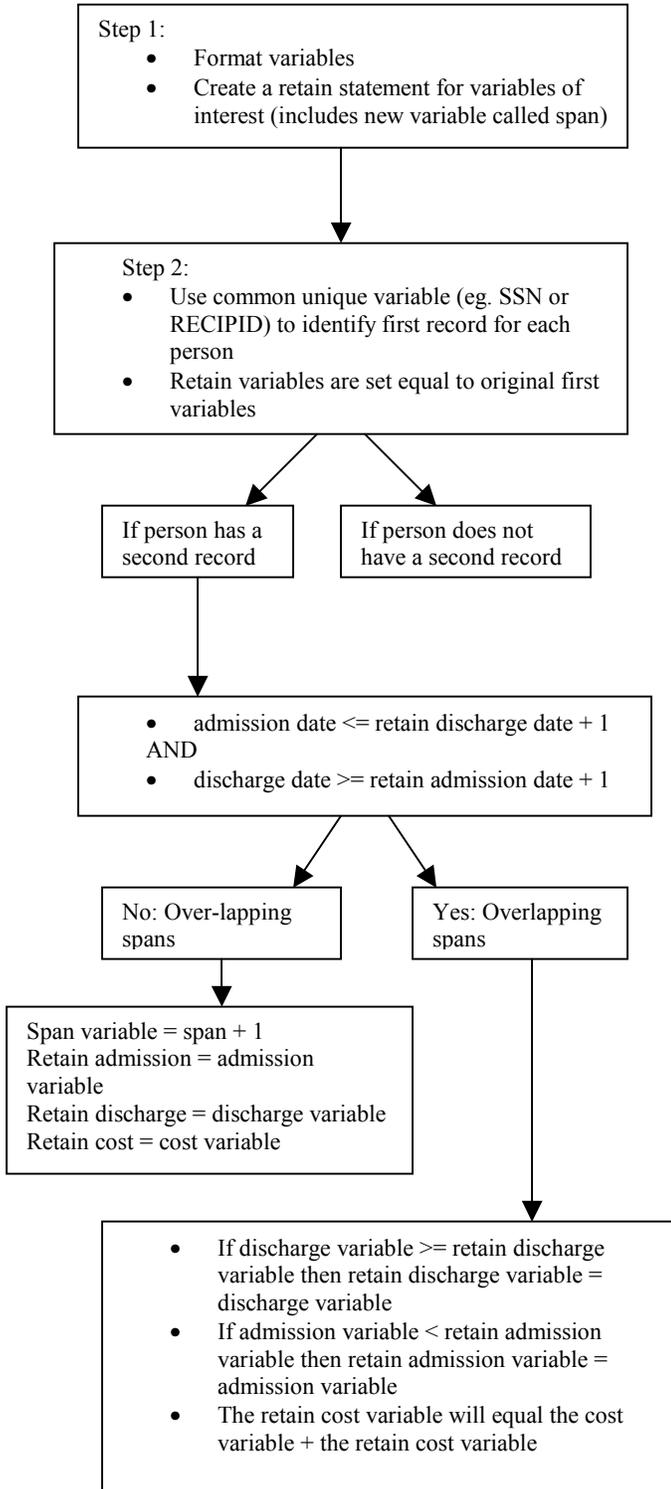
SPAN CODING

The challenge is to attain a unique record that engulfs the overlapping admission and discharge dates in a hospital stay. Once this challenge is completed, length of stay and cost of the hospital stay or any other variable included can easily be calculated. If ..then .. else..do coding with retain statements can be easily employed to create spans. The first step in the code is to format the retained admission and discharge dates. A retain statement is used to keep the previous dates (called INPBEG and INPEND), and cost (INPPAY) and to create a variable called span. The next step is to set the first span to start with the first record, in this case, using the recipient identification variable (ID). The variables for each recipient's first observation (span, admission, discharge, and cost) are initialized and are set equal to the original variable values. Other variables that pertain to the unique records could also be initialized at this point.

Now, to the span coding...if the recipient's observation is not the first and fulfills the requirement that the admission date (FDATE) is less than or equal to the retain discharge date (INPEND) plus one and the discharge date (LDATE) is greater than or equal to the retain admission date (FDATE) is less than or equal to the retain discharge date (INPEND) plus one and the discharge date (LDATE) is greater than or equal to the retain admission date (INPBEG), then we proceed to the coding for overlapping spans.

Plus one is incorporated to take care of continuous spans. It facilitates in the creation of one record for all continuous spans, which are spans where the discharge date on the first record is one day less than the admission date on the second or following record. The overlapping span coding states that if the discharge date (LDATE) is greater than the retain discharge date (INPEND) then the retain discharge date is set to equal the discharge date. Conversely, if the admission date (FDATE) is less than the retain admission date (INPBEG), then the retain admission date is set to equal the admission date. The retain variable cost (INPPAY) will equal the cost (MEDPAY) plus the retain variable cost. If the recipient's observation is not the first and does not fulfill the above requirements, that is, there is no span overlapping, then the variable span equals span plus one, the retain admission (INPBEG) equals the admission date (FDATE), the retain discharge date (INPEND) equals the discharge date (LDATE) and the retain cost (INPPAY) equals the cost variable (MEDPAY).

Flowchart of Span Coding :



Shown here is the original dataset as it was attained from the data collector:

ID	FDATE	LDATE	MEDPAY
123	7/03/01	7/10/01	\$120.00
123	7/05/01	7/12/01	\$153.00
123	7/15/01	7/20/01	\$898.00
123	7/20/01	7/25/01	\$328.00
123	7/25/01	7/30/01	\$1050.00
456	5/24/01	6/01/01	\$290.00
456	5/27/01	6/03/01	\$576.00

This coding process will create a dataset like the example below:

ID	FDATE	LDATE	MEDPAY	INPBEG	INPEND	SPAN	INPPAY
123	7/03/01	7/10/01	\$120.00	7/03/01	7/10/01	1	\$120
123	7/05/01	7/12/01	\$153.00	7/03/01	7/10/01	1	\$273
123	7/15/01	7/20/01	\$898.00	7/15/01	7/20/01	2	\$898
123	7/20/01	7/25/01	\$328.00	7/15/01	7/25/01	2	\$1126
123	7/25/01	7/30/01	\$1050.00	7/15/01	7/30/01	2	\$2276
456	5/24/01	6/01/01	\$290.00	5/24/01	6/01/01	1	\$2900
456	5/27/01	6/03/01	\$576.00	5/24/01	6/03/01	1	\$8660

At this point, the dataset is sorted by ID, SPAN, INPBEG, and INPEND in descending order to attain the longest begin and end period for each span for each ID. Concatenating the ID and the SPAN variables creates a new variable, called RECIPJ. The RECIPJ is sorted and the longest span is placed as the first record for each person. Finally, the first RECIPJ is selected and placed in a new dataset thus keeping one record with the longest stay for each recipient.

Result of span coding:

ID	FDATE	LDATE	MEDPAY
123	7/03/01	7/12/01	\$273.00
123	7/15/01	7/30/01	\$2276.00
456	5/24/01	6/03/01	\$866.00

Span Code:

```

Data fspan;
  set vis_day;
  by ID;
  format inpend mmddyy10. inpbeg mmddyy10.;
  retain inpbeg inpend span inppay;
  if first.ID then do;
    do;
      span=.;
      inpbeg=.;
      inpend=.;
      inppay=.;
    end;
    inpbeg=fdate;
    inpend=ldate;
    inppay=medpay;
    span=1;
  end;

  else do;
    if fdate<=inpend+1 and ldate >=inpbeg then do;
      if ldate>inpend then inpend=ldate;
      if fdate< inpbeg then inpbeg=fdate;
      inppay=medpay+inppay;
    end;
    else do;
      span=span+1;
      inpbeg=fdate;
      inpend=ldate;
      inppay=medpay;
    end;
  end;
run;

Proc Sort data = fspan;
  by ID span inpbeg descending inpend;
run;

Data tspan;
  set fspan;
  format j $3.;
  j=put(span, 3.);
  recipj=ID||left(j);
run;

Proc Sort data = tspan;
  by recipj inpbeg descending inpend;
run;

Data fspan2;
  set tspan;
  by recipj;
  if first.recipj then output;
run;

```

Until this code was developed at the data center we used arrays to complete the spans coding. Span coding and array coding were run on the same data producing these results: arrays coding had 203 lines of code and 250.7 CPU time (seconds) and 450.11 real time (seconds) to produce the desired results on the same sample dataset. Where as, span coding had 73 lines of code and 0.74 CPU time (seconds) and 1.36 real time (seconds) to produce

the same results. However, it should be noted that array coding is always more useful when there are multiple records per person and only one date variable.

CONCLUSION

Obtaining a unique record for inpatient hospital stays using admission and discharge dates can be easily accomplished using the span coding. It is a simple and swift method for manipulating secondary data to a desirable format for answering research questions.

TRADEMARKS

SAS is a registered trademark of SAS Institute Inc., Cary, NC, USA and other countries.

® Indicates USA registration.

REFERENCES

SAS Version 8.2, Cary , NC : SAS Institute Inc.

SAS Institute Inc. (1990) , "SAS Language Reference, Reference Version 6, Frist Edition , " Cary , NC : SAS Institute Inc.

SAS Institute Inc. (1990) , "SAS Procedures Guide, Version 6, Third Edition , " Cary , NC : SAS Institute Inc.

Delwiche L.D.and Slaughter S.J.(2000) "The Little SAS Book," Second Edition, Cary , NC : SAS Institute Inc.

Worden J. and Shinaberry L.(2001) "Hip Hip Array... Expanding and Modifying Records with an Array." *Proceedings of Twenty Sixth Annual SAS Users Group International Conference*, 227-26.

ACKNOWLEDGEMENT

Special thanks to Mary Rose Murrin at our Data Center for her spans logic. A big thanks to Paul Stiles for putting up with numerous requests for rereads and input. And many thanks to the rest of the PSRDC team for their encouragement and assistance.

CONTACT INFORMATION**Policy and Services Research Data Center (PSRDC)**

Mental Health Law and Policy
 Louis de la Parte Florida Mental Health Inst
 University of South Florida
 13301 Bruce B. Downs Blvd.
 Tampa, Florida 33612

Shabnam Mehra

Phone: 813-974-9315

e-mail: smehra@fmhi.usf.edu

Robyn Spittle

Phone: 813-974-8209

e-mail: spittle@fmhi.usf.edu

Becky Larsen

Phone: 813-974-8209

e-mail: rlarsen@fmhi.usf.edu

Diane Haynes

Phone: 813-974-9244

e-mail: haynes@fmhi.usf.edu

Kathy Dailey

Phone: 813-974-8209

e-mail: kdailey@fmhi.usf.edu