ABSTRACT
It is common in clinical research to experience missing observations in a longitudinal trial. To analyze the data as an intent to treat sample, it is necessary to impute data for the missing observations or use techniques that will handle missing data. One common imputation procedure is Last Observation Carried Forward (LOCF) where the last observed value for each case is carried out and replaces the missing observations. An alternative approach is to apply a maximum likelihood estimation procedure that can be accomplished with a mixed model that does not require any imputation. It is thought that the LOCF method is biased such that it is related to early drop outs of the less preferable treatment. A macro is used to apply the LOCF method. The results are compared to that generated by PROC MIXED.

OVERVIEW
Intent to treat (IT) analyses require that all complete and incomplete cases be included. Outcome measures of incomplete cases are often imputed for the period after the drop out. A common practice is to carry the last observation forward (LOCF,1) on the outcome measures. More recently, IT methods using a mixed repeated measures model (2) has been introduced and can easily be analyzed using SAS® PROC MIXED (3).

These techniques are now common practice and work well if the missing data is at random (MAR). This assumption is often true when the missing data occurs because a subject drops out or terminates based upon the observed data. When the missing data depends upon the unobserved values of the missing data itself, it is referred to as nonignorable missing (4). In these instances, it is most desirable to retrieve as much data as possible even after drop out that can improve imputation.

EXAMPLE
In a randomized clinical trial (5), subjects were treated under double blind with either nortriptyline which is an antidepressant or placebo for up to 3 years. Only 1 year of data is presented here. Subjects remain under double blind until a recurrence of depression or drop out occurs. Hamilton rating scores (HRS) which measure depression and General Life Functioning scores (GLF) were collected monthly. Comparison of symptoms and functioning are compared between 3 different imputation approaches.

1) no imputation
2) last value carried forward
3) retrieved scores after dropout or recurrence

MACRO
A SAS® macro was written to replace the missing data after drop out with the last observed data. The macro parameters include the data set (DATA), subject identification (ID), repeated variable (TIME), outcome variable (VAR), and maximum number of repeated observations (R).

%MACRO
LOCF (DATA=_LAST_, ID=ID, TIME=TIME, VAR=, R=)
;
proc sort;
   by &id;
proc means noprint data=&data;
   by &id;
   var &time;
   output out=lasttime max=maxtime;
data gen;
   set lasttime;
   do i=1 to &r;
      &time=i;
      output;
   end;
proc sort data=gen;
   by &id &time;
proc sort data=&data;
   by &id &time;
data all;
   merge gen &data;
   by &id &time;
data lastobs(drop=&var);
   set all;
   if maxtime eq &time; lastobs=&var;
proc sort data=lastobs;
   by &id;
proc sort data=all;
   by &id;
data new;
   merge lastobs all;
   by &id;
   if &var eq . and &time gt maxtime then &var=lastobs;
%MEND LOCF;

RESULTS
Repeated measures anovas using PROC MIXED were run on the raw data using a mixed model approach, on LOCF imputed and on retrieved imputed data. Significance for factors of drug, time and drug by time interactions can be seen in table 1. Least square means are presented in figure 1 and 2. It is evident from the
figures that the pattern of scores from the LOCF approach are quite different than either the no imputed or the retrieved imputed data.

**FIGURE 1: Hamilton Depression Score**

![Graphs showing Hamilton Depression Score over time for different conditions.]

**FIGURE 2: General Life Functioning**

![Graphs showing General Life Functioning over time for different conditions.]

### Table 1-Significance Values from PROC MIXED

<table>
<thead>
<tr>
<th></th>
<th>NONE</th>
<th>LOCF</th>
<th>RETRIEVED</th>
</tr>
</thead>
<tbody>
<tr>
<td>HRS-Drug</td>
<td>.0001</td>
<td>.0001</td>
<td>.0001</td>
</tr>
<tr>
<td>HRS-Time</td>
<td>.0060</td>
<td>.0001</td>
<td>.0321</td>
</tr>
<tr>
<td>HRS-Drug*Time</td>
<td>.0443</td>
<td>.6109</td>
<td>.1243</td>
</tr>
<tr>
<td>GLF-Drug</td>
<td>.0061</td>
<td>.0001</td>
<td>.0001</td>
</tr>
<tr>
<td>GLF-Time</td>
<td>.4526</td>
<td>.1299</td>
<td>.6686</td>
</tr>
<tr>
<td>GLF-Drug*Time</td>
<td>.5698</td>
<td>.6718</td>
<td>.8031</td>
</tr>
</tbody>
</table>

### CONCLUSION

A SAS® Macro was written and utilized to compute LOCF data for missing observations. The results of this approach was compared to that using only data available and to retrieved data. It is decided that the LOCF results in a much different pattern of scores than the other 2 approaches which is explained by a poor score that results in a recurrence. These values are then carried out beyond the drop out point.

It is seen that the LOCF data is biased to maximize the difference between the treatment groups. The retrieved data is biased to minimize the difference between the treatment groups. It is impossible to determine the correct scores that would have been seen if none of the subjects recurred or dropped out, but new imputation techniques are being explored and hopefully will one day replace the LOCF method.

### REFERENCES


Supported in part by MH52247 and MH43832.

Key words: Missing value imputation, LOCF, Mixed models, Longitudinal data

### CONTACT INFORMATION

Patricia R. Houck  
Senior Statistical Coordinator  
UPMC Health System, E-1128 WPIC  
3811 O’Hara Street, Pgh., Pa. 15213  
e-mail HOUCKPR@MSX.UPMC.EDU