LOCF Method and Application in Clinical Data Analysis

Huijuan Xu, Biogenidec, Inc.

ABSTRACT

The Last Observation Carried Forward (LOCF) imputation method can be used when data are longitudinal (i.e. repeated measures have been taken per subject by time point). The last observed non-missing value is used to fill in missing values at a later point in the study. Therefore one makes the assumption that the response remains constant at the last observed value. There are several ways of achieving this goal. Before we use this imputation method, we need to know several things: what the data structure is like, at what time point the observed value can be carried forward, and what is the last time point for non-missing value to be carried forward in the study. In this paper, a couple of data merging methods will be discussed and applied to a simple case and a more complicated case.

INTRODUCTION

In the analysis of clinical trial results, dropouts need to be addressed during the trial. Missing data due to dropouts occur for two major classes of reasons*: noninformative* reasons (essentially random, e.g., a study subject moves to another city with his or her family) and *informative* reasons (essentially nonrandom, and possibly related to treatment assignment or outcomes, e.g., the study subject suffers a relapse or develops intolerable side effects and is removed from the trial by the investigators). (1)

Because one can never be fully certain whether data are noninformatively or informatively missing, it is considered good practice not to ignore dropouts. Last observation carried forward (LOCF) is a commonly used way of imputing data with dropouts. LOCF uses the last value observed before dropout, regardless of when it occurred. (1) The FDA has traditionally viewed LOCF as the preferred method of analysis, considering it likely (but not certain) to be conservative and clearly better than using observed cases, where only the data observed are used.

The Last Observation Carried Forward (LOCF) imputation method can be used when the data are longitudinal (i.e. repeated measures have been taken per subject by time point). The last observed value (non-missing value) is used to fill in missing values at a later point in the study. Therefore one makes the assumption that the response remains constant at the last observed value. (2)

METHOD AND APPLICATION

One simple example and one complex example are given below for illustration of the LOCF imputation method and application.

In the following original data from SAS® code, there are values in some visits:

```
data val1;
 input visit $10. value ;
 cards;
 Week  2     1.5
 Week  4     2.3
 Week  8     3.2
run;
```

After using the LOCF imputation method, the final data should be like:

```
  VISIT       VALUE
  Week  2       1.5
  Week  4       2.3
  Week  6       2.3
  Week  8       3.2
  Week 10       3.2
  Week 12       3.2
```

To fulfill this, a data shell needs to be built up using the following code for scheduled visit to merge with:

```
data vis1;
  input visit $10.;
  cards;
  Week  2
  Week  4
  week  6
  Week  8
  week 10
  Week 12
  ;
  run;
```

According to Roland Berry (3), the following three data steps will create the final data:

```
data vis1_;
  retain _seq 0;
  merge vis1 val1(in=val keep=visit);
  by visit;
  if val then _seq=_seq+1;
run;

data _val1;
  set val1;
  _seq=_n_;
run;

data final1;
  merge vis1_ _val1;
  by _seq;
  drop _seq;
run;
```

Berry also provided a macro %LOCF for this purpose (4).

Here the author presents another way to do LOCF which combines all three data steps into one data step by using a temporary variable Tempvar in a RETAIN statement;

```
data final1(drop=tempvar);
  retain tempvar 0;
  merge vis1 val1;
  by visit;
  if value=. then value=tempvar;
  else tempvar=value;
run;
```

This example has more complex data:

```
data val2;
 input subjid $3. visit $10. value stdy;
 cards;
 101   screening   1.3   -1
 101   Week  4     2.3   28
 101   Week  8     3.2   56
 101   Week 14     4.5   98
 101   Week 18     3.8  126
 102   Week  2     1.2   14
 102   Week  4     2.4   28
 102   Week 14     5.2   56
;
run;
```

Assume the final data look like the following:

```
Subjid  visit      value  stdy
 101   screening   1.3   -1
 101   Week  0     1.3    1
 101   Week  2
 101   Week  4     2.3   28
 101   Week  6     2.3
 101   Week  8     3.2   56
 101   Week 10     3.2
 101   Week 12     3.2
 101   Week 14     4.5   98
 101   Week 18     3.8  126
 102   Week  0
 102   Week  2     1.2   14
 102   Week  4     2.4   28
 102   Week  6     2.4
 102   Week  8     5.2
 102   Week 10     5.2
 102   Week 12     5.2
 102   Week 14     5.2   56
```

Note that the last pre-dose value (baseline - Week 0) can not be carried forward to the dosing visits (from Week 2 to Week 14), and the last value in dosing visits can not carried forward to the follow-up visit (after Week 14). For the Subjid=102, the study day (stdy) is 56 for the last visit Week 14. Therefore it is necessary to re-map the visit for this

record based on the study day (stdy). The actual visit for this record is Week 8 after re-mapping.

First, a shell (Vis2) should be built up using the following code. Week 0 is the first time point, and Week 14 is the last time point (endpoint) to which the LOCF will be done (a variable Visnum is added for ordering purpose):

```
data vis2(drop=num);
  length visit $10. subjid $3.;

  do subjid='101', '102';
    visnum= 0;
    visit='screening';
    output;
    visnum= 10;
    visit='Week 0';
    output;
    do num= 0, 2, 4, 6, 8, 10, 14;
      visnum=num*10;
      visit = 'Week '||put(num,2.);
      output;
    end;
  end;
run;
```

Then merge the shell with the original data by subjid and visnum:

```
proc sort data=val2;
  by subjid visnum;
run;
proc sort data=vis2;
  by subjid visnum;
run;
data final2; *(drop=val_b val_p);
  retain val_b val_p ;
  merge vis2 val2;
  by subjid visnum;
  if visnum<20 then do;
    if value=. then value=val_b;
    else val_b=value;
  end;
  else if 20<=visnum<160 then do;
    if value=. then value=val_p;
    else val_p=value;
  end;
run;
```

In summary, an appropriate shell is key to make a correct LOCF. All strata and/or sorting variables such as subject ID, test code and time variables such as visit number should be considered when building up the shell. In the shell, the start and end point should be decided. More than one temporary variable may be used if the data includes different periods of time points (e.g. pre-dose, dosing, and follow-up). For LOCF, the observed value is carried forward to a later time point. However, an exception exists when the last

non-missing value needs to be re-mapped to the true (actual) visit by the study day window. In this case, it looks like that the observed value is carried backward to an earlier time point.

CONCLUSION

In the analysis of clinical trial results, the last observation carried forward (LOCF) is a commonly used way of imputing data with dropouts. The last observed value (non-missing value) is used to fill in missing values at a later point in the study. An appropriate shell is key to make a correct LOCF. All sorting variables such as subject ID, test code and time variables should be considered when building up the shell. In Roland Berry's macro, there are three data steps to do LOCF while in this paper, only one data step is needed for LOCF.

ACKNOWLEDGEMENT

Thanks to my colleague John Hotaling for his review and comments on this paper.

REFERENCES

(1) Last Observation Carried Forward Versus Mixed Models in the Analysis of Psychiatric Clinical Trials by Robert M. Hamer, et al**,** Am J Psychiatry 166:639-641, June 2009.
(2) Statitical Solutions, SOLAS, for Missing Data Analysis, 2008.
(3) LOCF processing by Roland Berry, 2004.
(4) Macro %LOCF web link: http://www.datasavantconsulting.com/roland/locf.sas, Roland Berry.

CONTACT INFORMATION

Huijuan Xu, MPH
Staff Analyst, Statistical Programming
biogen idec, Inc.
14 Cambridge Center, Cambridge, MA 02142
Tel: (617) 679-2373
Fax: (617) 679-3280
E-mail: hjx01803@yahoo.com

DISCLAIMER