

Macro to replace missing values

Kamrunnisa & Mokhfur Chowdhury, Shafi Consultancy Bangladesh

1. Introduction

To capture all data from all patients is the objective in clinical trials. However, this does not happen often, resulting in missing values appearing in the data. Of course values are also excluded from analysis because rescue medication was used or other events had taken place. It should be noted that just ignoring these missing and excluded data is not an acceptable option when planning, conducting or interpreting the analysis of a confirmatory clinical trial. Fortunately, when there is missing data, some commonly used methods are available to replace missing values. This paper will show how programmers can save time, improve efficiency and consistency by the use of macro to replace missing values.

2. Macro 'Imputation' aim

The main purpose of the macro is to replace the missing values using one of three methods specified by the user in the macro call. The missing values can be imputed using last observation carried forward/backward, linear interpolation or by using a summary statistic, such as mean, median, mode, minimum or maximum values.

3. Methods to replace missing values

LOCF/LOCB : LOCF or LOCB - the last measured observation before the missing value is forwarded. This method works best if the observations are expected to remain at same level or if there are only a few missing values.

Sample macro code for LOCF

```
*LOCF;
%IF %UPCASE(&imp_mthd.)=LOCF AND %LENGTH(&bslncond.)=0 %THEN %DO;
DATA valby_&imp_mthd.1;
RETAIN &prevars. ;
SET indtasrt ;
BY &grp_by.;

ARRAY origvars[*] &imp_var. ;
ARRAY prevar[*] &prevars. ;
ARRAY _rval[*] $50. &rvar.;

IF FIRST.&within. THEN DO i=1 TO DIM(origvars);
firstvis=1;
prevar[i]=.;
END;
DO I=1 TO DIM(origvars);
IF origvars(i) EQ . AND prevar(i) NE . THEN DO;
origvars(i)=prevar(i);
_rval(i)="Replaced By %UPCASE(&imp_mthd.)";
END;
END;
END;
OUTPUT valby_&imp_mthd.1;
DO i=1 TO DIM(origvars);
IF (origvars[i] NE .) THEN DO;
prevar[i]=origvars[i];
END;
END;
RUN;
```

```
imputation(indata = pft,
imp_var = wt fev fvc pef,
grp_by = study ptno visit pt,
within = visit,
imp_mthd= locf,
outdata = imp_bylocf);
```

Missing values for wt,fev,fvc and pef variables are being replaced by last non missing values within visit

```
bslncond = visit EQ 2 and ptm EQ ^-0:05'T,
tknpretrt= y,
```

As an additional option user can replace missing by using baseline condition or carried forward values from off-treatment period

SUM : This method is used to replace missing values with statistical summaries.

LINT : Applying linear interpolation procedures to irregularly-sampled raw data can obtain time series with equidistant sampling intervals. The application of approximation methods to the time series produces function we can use to fill in gaps in the data under stationary conditions.

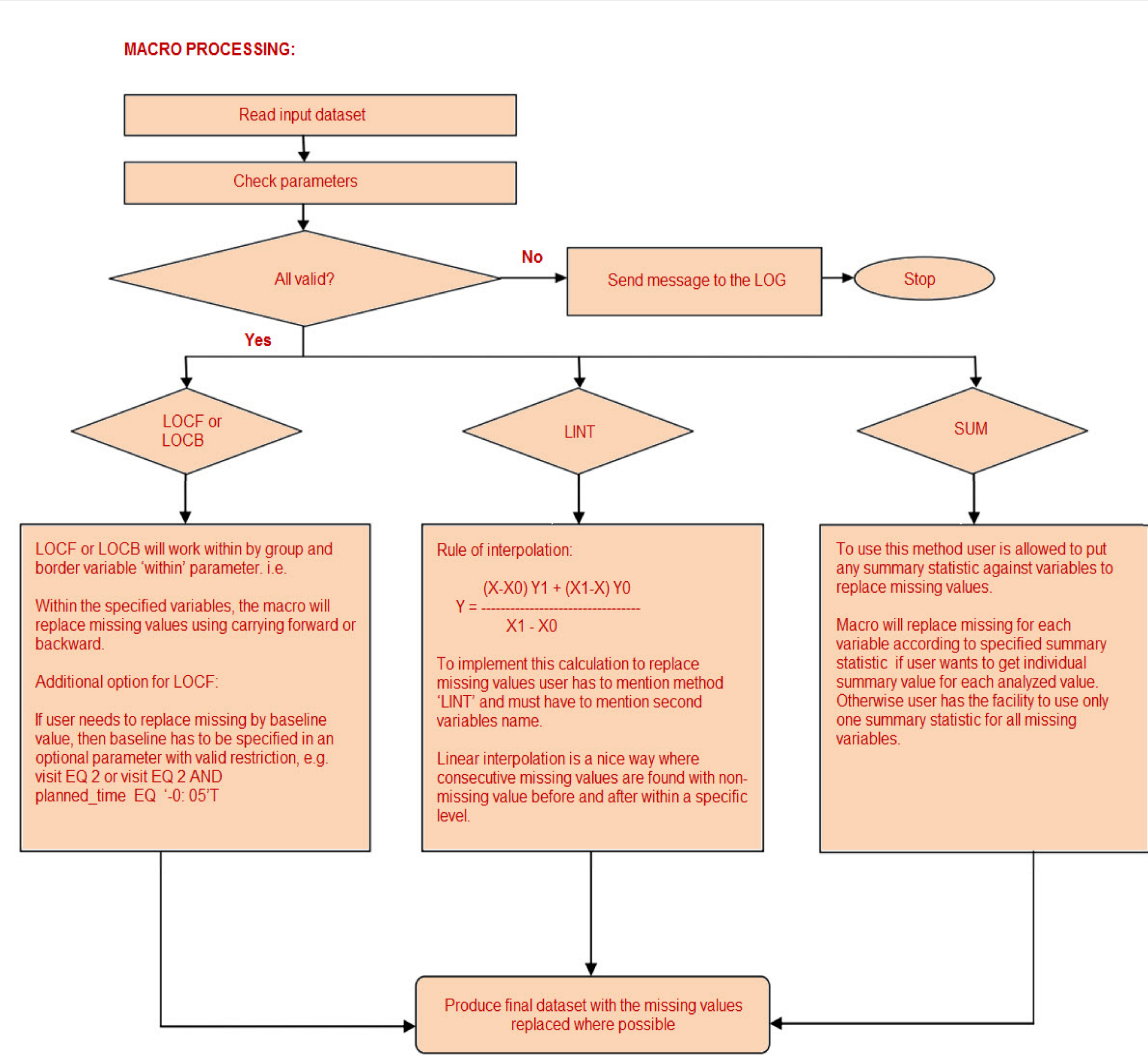
4. Problems encountered and how they were resolved

There were two main aims of the macro, to be easy to use for replacing missing values, and to be flexible enough for different types of data. To ensure the macro is easy to use meant reducing the number of macro parameters to the bare essentials. This means the users are clear what each parameter they are specifying do, making them more in control and feeling less like they are using a black box.

Macro parameters can be used to define which method is used to replace which variable, but to make this easy to define for users is a challenge. Making the macro call more like a procedure is the solution. As users are used to calling procedures, making the macro call more like a procedure helps to make the call more user friendly, easy to remember and more importantly, easy to use.

5. Features of macro 'Imputation':

- The Macro allows the user to define Input datasets with the flexibility to apply restrictions such as: Where Visit IN(2,4,6). Also allows the user to define output datasets.
- Allows the user to be able to replace missing values with different summary calculations (i.e. Weight by mean values, pulse rate by max values within the same by group) in the same macro calling.



- Users have the flexibility to use baseline values with restrictions to replace missing values. (e.g. baseline condition may be as such : visit EQ 2 or visit EQ 2 AND planned_time EQ ^-0:05'T).
- Users are able to use LOCF for replacing missing values with pre-treatment values. Users can specifically mention this, so a value can be replaced by a pre-treatment value if there is no available value to be carried forward. This also can be done by using LOCB if pre-treatment values are not allowed for a specific study.
- LINT(Linear interpolation) allows user s to replace consecutive missing with non-missing before and after value in any level, user must specify one valid single variable in the right place. User s must specify the second variable with all known values (current, previous and next values) to calculate interpolation. Valid value can be any one those are being used in group variables.
- Users can order and group variables in order to get their desired output.
- One or multiple variables can be replaced by the macro at a time.
- Checks have been put in place to restrict unexpected values from bring entered into parameters . For example the user will not be able to pass invalid datasets, variables or unspecified imputation methods. The Macro will send a message to the Log when such an instance occurs, advising the user on how to resolve these issues.

6. Limitations

The structure of the data is fixed, so updates to the macro will be required if the data structure changes. It means that one macro to cover all studies is not possible, but it is possible to use one macro for a project containing many studies.

7. Conclusion

The macro is easy to use, ensures consistency within trials, and when used on a project level, it ensures consistency across studies within a project. For a small organization, it is versatile enough for use to use with various clients with no or minimal change, thus saving a great deal of time. Having one macro which does this also means that everyone is familiar with the macro, and is therefore comfortable to use it.

Please feel free to contact us for further information.

www.ShafiConsultancy.com

