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 recording these missing and censored 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 - 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

```
/*LISP*/
#define IMP_VAR 1
#define IMP_BY 0
#define WITHIN 1

Macro "Imputation" aim: 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.

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.

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.

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 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