PhUSE 2006

**Paper PO11**

# Partial Dates; decisions and implications of handling partially missing dates

Rachel Bowman, GlaxoSmithKline, Harlow, UK

## ABSTRACT

A common occurrence with many Clinical Trials data is the collection of partial dates, when the full date is unknown. Partial dates make summarising the data successfully and accurately more difficult and may impact the validity of the trial if not handled correctly. This paper looks at several different common situations where partial dates occur and some considerations to be taken into account as to how and why they should be handled.

## INTRODUCTION

The FDA states "Missing values represent a potential source of bias in a clinical trial. Hence every effort should be undertaken to fulfill the requirements of the protocol concerning the collection and management of data. In reality, however, there will almost always be some missing data. A trial may be regarded as valid, nonetheless, provided the methods of dealing with missing values are sensible, particularly if those methods are predefined in the protocol [1]". This guidance goes on to state "no universally applicable methods of handling missing values can be recommended."

Dates are an integral part of the data collected within clinical trials. Visits, investigational product doses, adverse events and medical history are just a few common data types where dates are often used. However, alongside the obvious advantages of date collection is the common occurrence of missing and partial dates, where the complete date is not known. As stated above by the FDA Guidance, these situations need to be handled in a robust manner and documented accordingly. As there is no foolproof method for the handling of partial dates, this paper will highlight several common situations and discuss different solutions to the problem of partial dates.

## WHAT IS A PARTIAL DATE?

A partial date is simply any date where the date is incomplete, but not wholly missing. This may be the day or month or year or any combination of two of these. More commonly in clinical trials, the day and/or month are missing. For example:

- -JUN2006
- - - - -2006
15 - - -2006

Partial dates can occur in any date variable, but are most common in variables where the date is historical information, for example prior medication data. For current dates, there are several measures which can be put in place at the point of data collection to minimise their occurrence. This is especially the case with the increased use of electronic Case Report Forms where the data is entered directly into the database by the subject or investigator. Point of entry checks can be put in place to reduce the number of partial dates. However, historical data is notoriously more problematic where the subject, often understandably, does not have the full date available to them. For example, an accurate start date for a medication they have been taking for a number of years.

Where it has not been possible to eliminate the occurrence of partial dates, it is imperative that they are handled correctly to retain the integrity of the data. In the example above where a subject does not know the date they began taking a medication because they began taking it a number of years ago, it is unlikely to impact the data integrity if a date is imputed which is a few weeks or even months away from the actual date; but it does still need to be handled correctly. What if the missing date is for the end date of a prior medication which has ended in the weeks prior to the study start? If this drug had a long half-life, it could interfere with the efficacy of the study drug. In this situation, making the most appropriate approximation for the missing information is more important.

So if a variable contains a partial date – what options are available to you? The simplest option is to do nothing. This may seem like a casual option but in some situations this may be the best choice. The decision on whether it is acceptable to do nothing and leaving the date as partially completed depends on how the data is to be used. If the

date is merely to be listed – what further understanding will be gained by the reviewer if a date is imputed? In such circumstances, leaving the date as it was collected may be the optimal course of action.

Therefore in what circumstances would it be best to impute a date? This depends on how the date is to be used. As stated above, if the date is merely to be reported, it preferable to leave the date as it is. However, if the date is to be used, for example, to calculate the duration of an adverse event or the duration of a drug, then a partial date will prevent the duration being calculated. In this circumstance, a decision would need to be made as to whether a missing duration is acceptable to the reporting of the study or whether it would be beneficial to impute a date in order to calculate the duration. Once the decision has been made to impute the date, how this is done will depend on the data being collected by the partial date. For example, a partial start date will be handled in a different manner from a partial end date. Which part of the date is missing will also need to be taken into consideration, as missing days, months and years will need to be handled differently.

A partial start date with a missing start day, e.g. - -JUN2006, will often be set to the first of the month, 01JUN2006. If the month is missing e.g. 15- - -2006, the month will often be set to January for that year, 15JAN2006. The reverse is often true for a partial end date, with the missing day set to the end of the month and the missing month set to December. Consideration needs to be taken with partial end dates with missing days, to ensure that months having differing number of days are taken into account. Special care also needs to be applied with February because of the extra day included in leap years. If there is only a year and both the day and month are missing, then both are imputed, giving 1st January for a start date or 31st December for a stop date.

| Partial Start Date | Imputed Start Date | Partial End Date | Imputed End Date |
|---|---|---|---|
| - -JUN2006 | 01JUN2006 | - -JUN2006 | 30JUN2006 |
| 15- - -2006 | 15JAN2006 | 15- - -2006 | 15DEC2006 |
| - -FEB2004 | 01FEB2004 | - -FEB2004 | 29FEB2004 |
| - -FEB2005 | 01FEB2005 | - -FEB2005 | 28FEB2005 |
| - - - - -2006 | 01JAN2006 | - - - - -2006 | 31DEC2006 |

Table 1: Example of imputed start and stop dates

This is simple enough if the first or last day of the month or first or last month of the year will suffice as an imputed date but this may not always be the case. In some circumstances it may be preferable to set the imputed date to the first or last study contact date or the first or last date of dosing. How do you decide what is the best approach to take? Below, this idea is discussed further through some examples relating to concomitant medications and adverse events.

## THE CONSERVATIVE APPROACH

The FDA quote stated in the introduction to this paper talks about a clinical trial being valid as long as the methods used to handle missing data in general are sensible. What would be regarded as "sensible"? This is an ambiguous term which is open to many interpretations. Industry generally supports the position of taking a conservative response to partial and missing dates, but what does that actually mean? Below are outlined several different situations which could be handled in different ways to give different results depending on what is deemed the most cautious and appropriate.

### CONCOMITANT MEDICATIONS

Situation A – subject has a partial concomitant medication start date of "--Apr2006" (see figure 1). As discussed above, missing start dates may be set to the first of the month, which is shown under option 1. However, this then pushes the concomitant medication to starting before the first dose of the Study Medication (15Apr2006) and would suggest that the Study Medication had no involvement with the concomitant medication being taken. Is this really the most conservative approach? If not is there an alternative? The missing concomitant medication start date could be set to equal the first dose of Study Medication, option 2. This option allows the concomitant medication to be classed as an on-treatment medication, and is therefore the most conservative.

### ADVERSE EVENTS

Situation B (figure 2)– a subject has two adverse events with partial start and stop dates:

| | Start date | Stop date |
|---|---|---|
| AE1 | - -JUN2006 | - -JUN2006 |
| AE2 | - -JUL2006 | - - - - -2006 |

Let's start with AE1. As the partial start date of AE1 has the same month and year as the study medication start

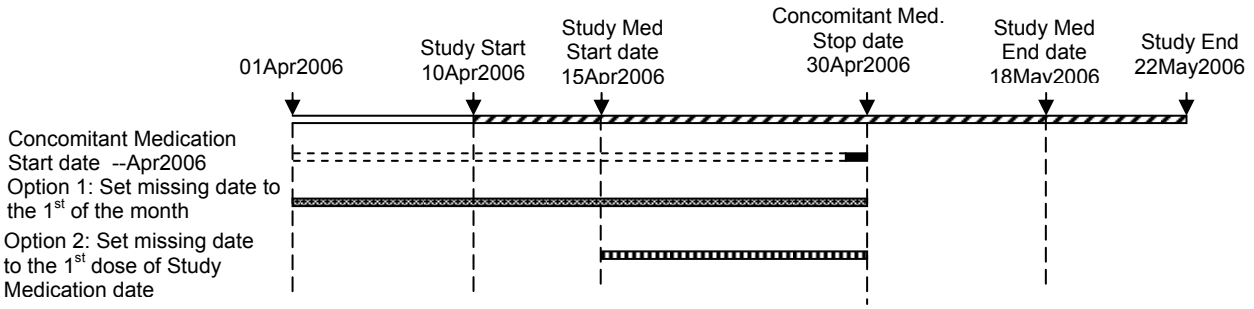


Figure 1: Missing Concomitant Medication Start Date

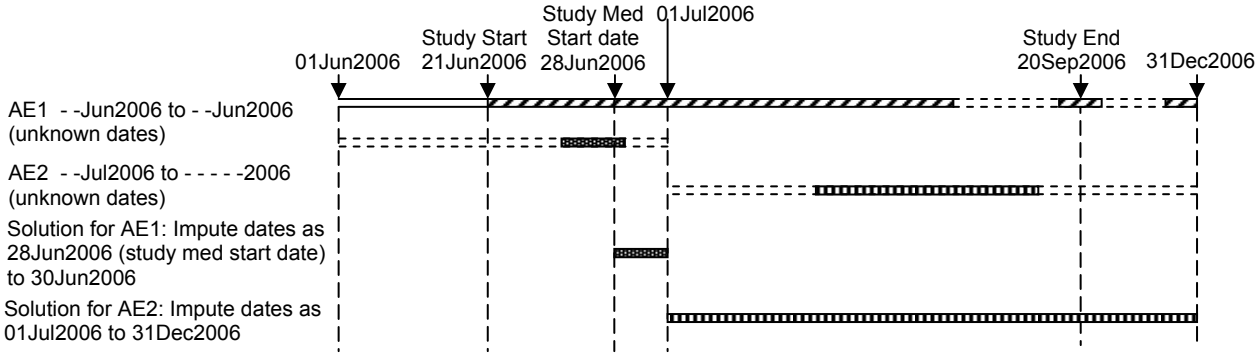


Figure 2: Missing Adverse Event Start and Stop Dates

date, there are two options available. The partial start date may be set to the first of the month or to equal the study medication start date. As previously discussed, the first option would indicate the adverse event began prior to the study medication. However, the second option, setting the start date of AE1 to the study medication start date will suggest the adverse event had a short duration, as the adverse event end date is also defined as June 2006, but began during the treatment period of the study drug. Although the second option is not ideal, as AE1 may have had a longer duration, it is more conservative to associate the adverse event start with a date during study medication. Another solution to consider is not to impute a date at all but merely to assign a study phase to the start of the adverse event. In this example, a phase of "treatment" could be allocated to the start of the adverse event, which would ensure it was classed most conservatively, without defining an actual date to the start of the adverse event.

All we know about AE2 is that it began in July 2006 and ended sometime later in 2006! Imputing a start date is simple enough – setting the start date of AE2 to 01Jul2006 is the earliest the AE may have started and is during the treatment phase of the study medication so is a conservative estimation. The end date has a couple of options as it could be imputed to be 31Dec2006 or to equal the study end date, in this case 20Sep2006. Which is preferable? Firstly, as there is very little information included in the partial end date for AE2, any imputation is little more than an "educated estimation", so there is little to choose between the two options. However, it could be argued that an adverse event may continue for a long period after the study has ended, therefore setting the end date for AE2 to the end of the year would be the most conservative option.

As suggested with AE1, it can be argued that the actual date is less important than the stage within a study where an event occurs. This is especially true for adverse events where treatment emergent events are often reported, making the allocation of the start of an adverse event the priority. Therefore leaving the partial dates as they are with missing information and allocating a phase or stage to the adverse event may be more suitable in many cases.

## CONSIDERATIONS OF STUDY TYPE

The examples given above all relate to parallel group studies. Crossover studies introduce a new dimension and new

considerations to be taken into account. Let's look at an example of a crossover study (figure 3). This example shows an adverse event with a missing start date (- - Jul2006). The adverse event may have begun at anytime in July prior to 27 July, which means it may have begun whilst the subject was on either Drug A, Drug B or during the washout period between Drug A and Drug B. The decision, on which option is most conservative, depends on the study and what drugs A and B are.

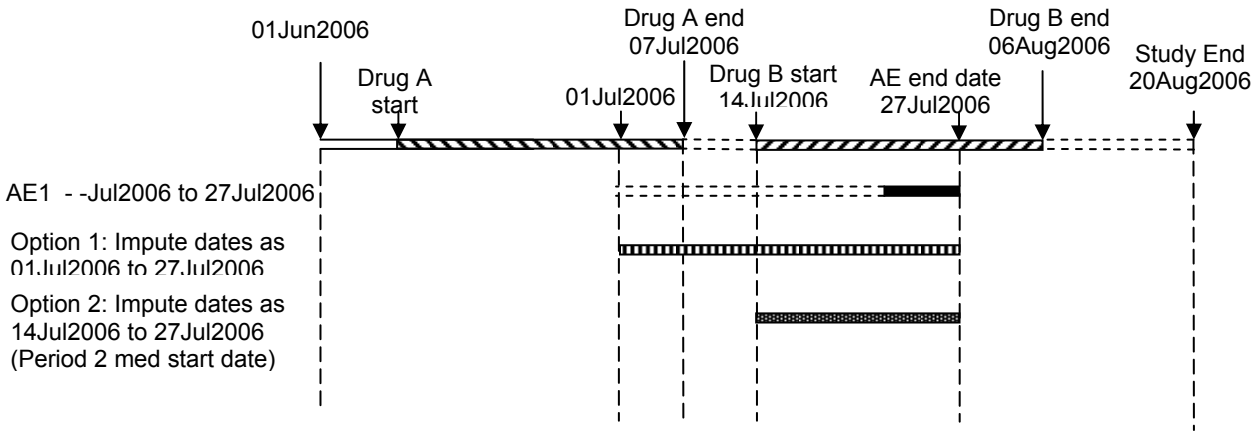


Figure 3: Crossover study example

Here are some scenarios:

1. Drug A is the study drug and Drug B is an active comparator. The most conservative option would be for the adverse event to have started during the study drug period; therefore option 1 would be selected
2. Drug A is a placebo and Drug B is the study drug. Option 2 should be selected as the most conservative option, so again the adverse event starts during the study drug period.
3. Drug A is a 10mg dose of study drug, Drug B is a 15mg dose of study drug. As both drugs are the study drug, it may suffice to set the adverse event to begin during the earliest dose of study drug, therefore Option 1. However, as Drug B is a higher dose of the study drug, it may be better to associate the adverse event with the Drug B, option 2. It could, however, be argued that associating a drug to the lowest dose of study drug is more conservative, depending on the study and the type of adverse event.

These three scenarios are only a few of the possible situations which may occur in cross-over studies and as shown by scenario three, the decision on how to handle the partial date may not be clear cut. Where there is no obvious most conservative option, the main consideration is that the rules should be sensible and clearly defined in advance of study reporting.

## CODING PARTIAL DATES

Once a clear definition for handling partial dates has been defined, the SAS code can be developed. Partial dates can be easily imputed using a relatively simple piece of code, as highlighted below, which sets missing start days and months to the first of the month and January respectively and missing end days and months to the appropriate date 28th/29th/30th/31st and December, respectively. Note how partial dates need to be captured as a character field whereas the imputed date is a numeric.

#### EXAMPLE INPUT DATASET

```
#    Variable    Type    Len    Format    Informat    Label
-------------------------------------------------------------------------
1    startdt     Num       8    DATE9.    DATE9.      Start Date
2    startd_     Char      9                          Start Date - Character
3    enddt       Num       8    DATE9.    DATE9.      End Date
4    endd_       Char      9                          End Date - Character

  startdt     startd_         enddt      endd_

15DEC2002    15DEC2002    14FEB2003    14FEB2003
        .    --DEC2002            .    --FEB2003
        .    -----2004            .    --FEB2004
        .                         .    -----2005
```

**EXAMPLE CODE FOR IMPUTING PARTIAL DATES**

```
* Partial start date *;

  if startdt = . and startd_ ne '' then do;

    * Month and year entered, default to start of month *;
    if substr(startd_,1,2) = '--' and substr(startd_,3,3) ne '---' then
      istartdt=input(trim('01')||substr(startd_,3),date9.);

    * Year entered, default to start of year *;
    else if substr(startd_,1,5) = '-----' then
      istartdt=input(trim('01JAN')||substr(startd_,6),date9.);

  end;
  else
    istartdt = startdt;

  * Partial end date *;

  if enddt = . and endd_ ne '' then do;

    * Month and year entered, default to end of month *;
    * Use INTNX function to advance date by 1 month *;
    * Then subtract 1 day to get last day of previous month *;
    if substr(endd_,1,2) = '--' and substr(endd_,3,3) ne '---' then
      ienddt=intnx('month',input(trim('01')||substr(endd_,3),date9.),1)-1;

    * Year entered, default to end of year *;
    else if substr(endd_,1,5) = '-----' then
      ienddt=input(trim('31DEC')||left(substr(endd_,6)),date9.);

  end;
  else
    ienddt = enddt;
```

**EXAMPLE OUTPUT DATASET**

```
#    Variable    Type    Len    Format    Informat    Label
-------------------------------------------------------------------------
1    startdt     Num       8    DATE9.    DATE9.      Start Date
2    startd_     Char      9                          Start Date - Character
3    enddt       Num       8    DATE9.    DATE9.      End Date
4    endd_       Char      9                          End Date - Character
5    istartdt    Num       8    DATE9.                Imputed Start Date
6    ienddt      Num       8    DATE9.                Imputed End Date

  startdt     startd_         enddt      endd_       istartdt       ienddt

15DEC2002    15DEC2002    14FEB2003    14FEB2003    15DEC2002    14FEB2003
        .    --DEC2002            .    --FEB2003    01DEC2002    28FEB2003
        .    -----2004            .    --FEB2004    01JAN2004    29FEB2004
        .                         .    -----2005            .    31DEC2005
```

## CONCLUSION

The main consideration to handling partial dates is to ensure that any manipulation to the date is sensible and non-biased. Whatever method is opted for, it is most important that the method is defined in advance of reporting the data in the protocol or analysis plan.

One final consideration; however the partial date is dealt with, the original partial date should be maintained within the dataset as although the imputed date may be required for certain tables and graphs, it is usual to include the date as it is collected in data listings.

## REFERENCES

[1] Food and Drug administration, International Conference on Harmonisation; Guidance on Statistical Principles for Clinical Trials; Availability http://www.fda.gov/CbER/gdlns/ichclinical.pdf

## ACKNOWLEDGMENTS

Many thanks to Anthony Cooper for kindly donating his SAS code.

## CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Rachel Bowman
GlaxoSmithKline
NFSP
Third Avenue
Harlow
UK
rachel.l.bowman@gsk.com