

How to process data from clinical trials and their open label extensions

Thomas Grupe, Bayer Vital GmbH, Leverkusen, Germany
Stephanie Bartsch, Bayer Vital GmbH, Leverkusen, Germany

ABSTRACT

Data handling of pivotal clinical trials and their open label extensions is always a challenging task. Using the latest methods of Electronic Data Capture (EDC) correctly combined with the powers of SAS® we can meet this challenge successfully. An open-label extension can accept patients from more than one trial of the same study drug. All the trials leading to the same open-label extension are considered "main phase". The open-label extension can be called "extension phase". An interim analysis is done on the main phase, and safety analyses are done on the extension phase taking into account the 'main phase' data as well. One of the biggest challenges is to decide how to use EDC correctly in such a scenario. Before collecting or combining data there is a choice to make. It is how to set up EDC. One option is to include both the main and extension phase together in one database and split them later. The other option is to create two databases, one for each phase, and then reconcile them later.

INTRODUCTION

Patients participating in double-blind controlled pivotal trials are often transferred into long-term open label extension trials after completion of the main part of the trial. These trials are often requested to offer patients benefiting from a new drug treatment the possibility to receive the drug until market launch and to generate long-term safety data.

Although these long-term extension trials cannot be regarded as entirely independent from the double-blind trials, there are good reasons to separate them under another study number. One reason is that patients that were part of the placebo or comparator group in the original trial may take the investigational drug in the extension trial.

From a data management point of view, different strategies can be applied to handle a controlled trial followed by long-term extension trials. The purpose of the paper is to show what could be gained and lost if such trials are set up with an EDC system – either as two EDC databases or as one EDC database for both trials with a later split in two SAS DBs with appropriate study numbers. The latter solution will be presented as the actually preferred one.

INITIAL SITUATION

Extension trials following pivotal trials can be planned in several different ways. You can find them combined within one study protocol under one study number as well as separated in two study protocols with an extra study number for each. Another kind of extension planning is to have one extension trial as a pool study of several different main trials.

Thinking of how to collect and transfer the data of such trials into the desired database structures leads to two main options. One option includes the set up of the pivotal and the extension trial in an EDC system under one study number. These studies are divided into two parts consisting of a main phase ending up in an interim analysis and an open-label extension phase focusing on the collection of safety data. Another option is to set up the two phases as two studies separately in the EDC system. The first option we call the SAS approach as SAS would be used to split the data into two databases, the second we call the EDC approach because separation of data would already take place in the EDC system (see figure 1).

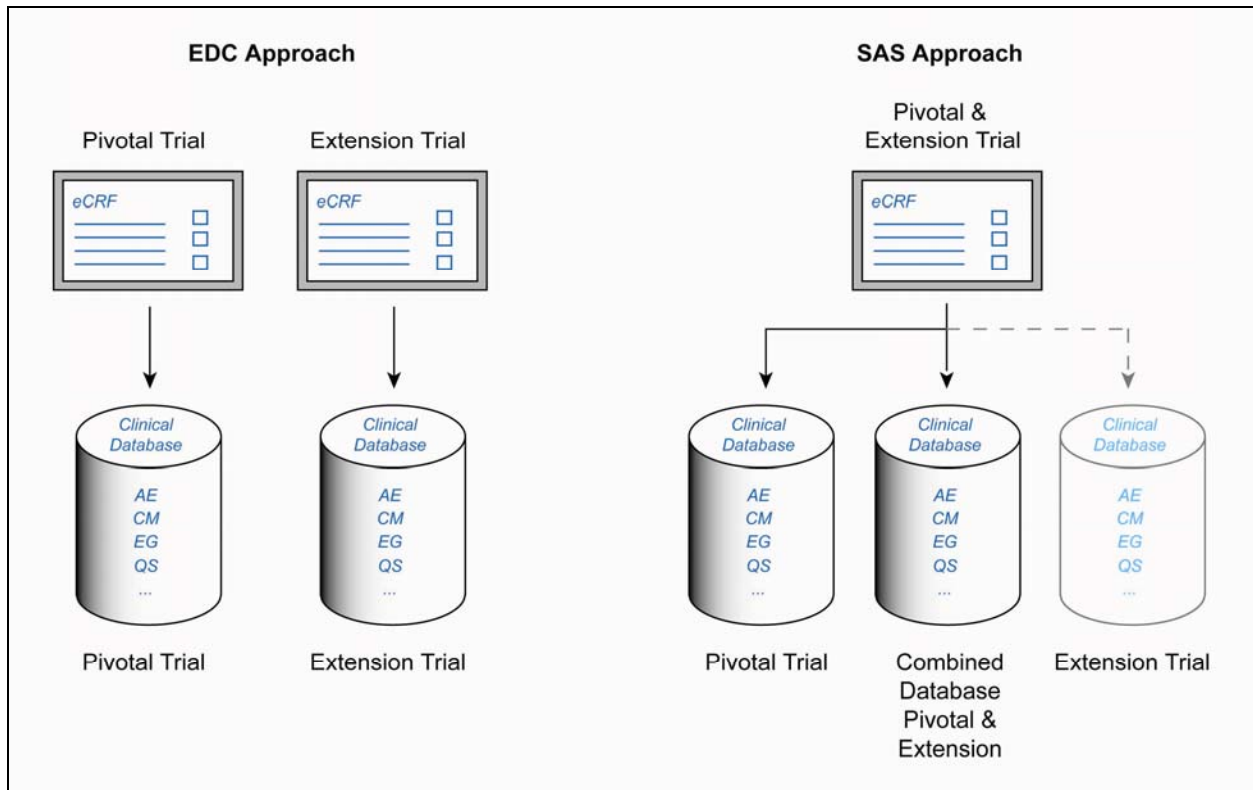


Figure 1: The EDC Approach and the SAS Approach

The current process of clinical data handling at Bayer gives the opportunity to think about either approach. For most studies we have a setup in an EDC system with a nightly data transfer followed by mapping to the clinical database in SAS. In one of our current projects we faced the situation of having the pivotal trial and its extension with separate protocols and therefore separate study numbers.

First, we will present the pros and cons of the EDC approach. Second, we focus on the SAS approach. Then, we describe the way we implemented the SAS approach, what problems we had and how we solved them.

THE EDC APPROACH

PROS

If setting up pivotal and extension trial in the EDC system separately, one of the major advantages would be to have a clear differentiation between the two studies and their databases – during data entry in the EDC system itself as well as in SAS. As a consequence from that the database lock and freeze would be easy, because there would be no link between the two studies. They could be handled as completely independent trials, and after any detected inconsistencies would have been cleaned for the pivotal main trial, it could be closed like any other study.

A third pro could be that other systems drawing information from the patient data entered in the Electronic Case Report Form (eCRF) would allocate this information to the appropriate study. For us an example is our system for reporting and tracking of recruitment (IMPACT).

CONS

The big drawback to using the EDC approach is that data have to be transferred from the pivotal to the extension study database. In the worst case it would have to be entered again into the eCRF of the extension. In any case, one would not get rid of transferring demographics, ongoing adverse events and ongoing concomitant medication. Other data like medical history are likely to be discussed if it is necessary to copy them into the extension database as well.

As a result of this data transfer, there would be much reconciliation work to be done – for site personnel, monitors and data managers. Additionally, in case these studies are pooled, one would have problems with duplicate adverse events.

Another con, especially for the sites, would be that queries affecting data in the pivotal trial would have to be updated in the extension if the data were previously transferred.

THE SAS APPROACH

PROS

When having just one eCRF for both studies, it would be much easier to use for the sites and monitors. Depending on the decision of the patient at the end of the pivotal trial, additional entry screens would be added to the eCRF, and clinical sites and monitors would just go on using the visit independent data entry screens as before. Neither data would have to be transferred nor reconciled.

Another pro would be that one is more flexible in generating different databases which could either consist of data from the pivotal trial only or consist of data from both trials as a combined database. If needed, generation of a database with data from just the extension would be no big deal as well.

CONS

On the downside the clear differentiation of both studies would be lost. Consequently, concomitant medications and adverse events which are ongoing at the time when a patient decides to proceed with the open label extension could be updated during the extension and the status of these data at the end of the pivotal trial would then be lost. Tracking of changes also would not be easy to program. Consequences for statistical analysis should be discussed with the statistical members of the study team to cover their needs when choosing the SAS approach.

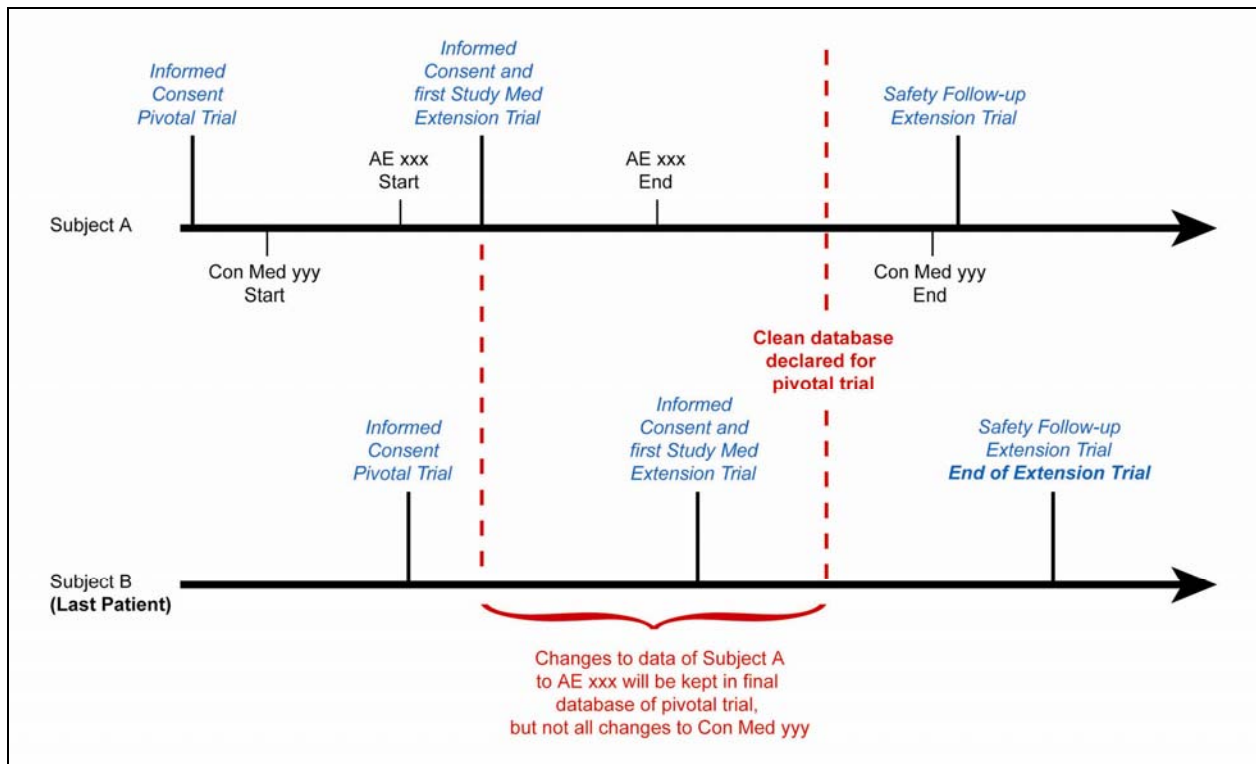


Figure 2: Possible Changes on AE and/or Concomitant Medication Data from Pivotal Trial

Secondly, if the studies have separate study numbers and protocols, a setup of both trials in one EDC database would affect possible connections to other systems as both studies would run under the study number of the pivotal trial. Thus, for example metrics about recruitment or "first patient first visit" of the extension could not automatically be retrieved.

Thirdly, EDC systems mostly offer only one possibility to close databases of studies at the end. But at the time of database closure of the pivotal trial this would be needed although the eCRF would be further used to enter data for the extension that is still running.

IMPLEMENTATION OF THE SAS APPROACH

After balancing all reasons for the two approaches it was decided to split the open-label extension study

PhUSE 2010

programmatically using SAS. Thus, data of both parts are retrieved from one EDC database. In order to meet all demands in splitting the trial into two parts further discussions with the study team, especially with statistical members, were necessary. Following points were specified:

DATABASES

Two databases should be created for statistical needs. On the one hand a database including the pivotal trial only is required for submission or evaluation analyses. This database consists of efficacy as well as safety data. On the other hand another database with pivotal and extension data will be generated for long term data collection and particularly safety analysis of both parts. However, a third database containing only extension data was requested by the Data Management function in order to facilitate the query management during the trial.

SPLIT OF STUDY DATASETS

In a study there are three kinds of dataset structures resulting in different ways of splitting:

- visit independent datasets, e.g. Adverse Events or Concomitant Medication
- visit dependent datasets, e.g. ECGs or Vital Signs
- datasets without visit or start dates, e.g. Demographics or Inclusion/Exclusion Criteria

The separation of *visit independent data* requires a good data quality as the first study medication date and time of the extension trial as well as the start date and time of the visit independent event are used as basis for the splitting. All events starting before the first study medication intake of the extension study will be assigned to the pivotal part. If the events occur at the cut-off day, the time of event and study medication intake is taken into consideration. A problem at this point is missing time information. A query should be raised then to the investigator in order to ensure the right assignment.

The pivotal part should be analysed before the closure of the extension part. Unfortunately a tracking of all changes after the subjects go into the extension part is impossible as the EDC data is updated every night. This especially applies to the visit independent forms like adverse events. A decision was made to save the status at the closure of the pivotal part regardless of the previous database status of each single subject when changing over to the extension trial. Due to the fact that visit independent forms include safety data only, the statistical members will not analyse the extension part on its own.

Visit dependent datasets contain visit information in each record. The first visit of the long term extension part will be used as cut-off point. Unscheduled visits will be integrated by incrementing the chronologically previous visit number. Thus, a clear differentiation between pivotal and extension part can be applied to every dataset with visit dependent data. If the visit information is missing the dataset will be split with the visit independent approach on the basis of the measurement date.

Some datasets contain neither visit information nor any start dates. Therefore single programs are created for splitting these datasets into two parts.

FURTHER IMPLEMENTATION

As a result of the splitting some variables have to be re-structured. Corresponding variables are relative days calculated to the study medication start and end. The calculation of relative days depends on the structure of the extension set up. If the pivotal and extension part are summarised under one study number, the relative days are calculated across both parts and saved in the combined database. On the other hand a differentiation in two study numbers will set the relative days by study number in the combined database.

The concept of relative days has no influence on the pivotal database as there is only one study number for this database in any case. Relative days of this database are assigned during the separation of the study. All further relative days that are needed will be set in the analysis database created by the statistical department.

Another variable which will be changed in many datasets is the sequence number, which is often used to generate unique records by subject or other applicable key variables. This change is only applied to databases where two different study numbers for pivotal and extension study are applicable. The numeration of the extension study will start at "1" again in the extension trial.

PROBLEMS AND SOLUTIONS

After all definitions were specified, a SAS macro was programmed. The macro is called at every EDC data transfer to the final database structure which is basis for further statistical processing. One output of this macro is a MS Excel[®]

PhUSE 2010

table which classifies the splitting of every dataset into the three categories specified above. Other output Excel files contain information about records of visit independent forms with missing time variables. As a clear assignment of these records to one of the parts is impossible, the Data Manager raises queries to the site.

Having only one database in the EDC tool makes it difficult to track or transfer information from the extension trial to other connected systems, e.g. the already mentioned IMPACT. To circumvent this lack of information, reports are programmed on the SAS data of the extension phase and are manually entered to the affected systems.

Another important feature is the implementation of various checks which test the consistency of the databases before and after splitting. For example when generating a database with extension data additionally, summing up records from both main and extension should give the same number of records before the split. Another check confirms that subjects who did not enter the extension have no records in the extension database and vice versa that subjects who did enter the extension have records in the extension database. This information ensures the quality of the program and can be used for validation.

Especially the splitting of visit independent datasets requires a high data quality. A missing study medication start date of extension causes an incorrect separation of both parts. This influences listings and tables produced during study conduct by Data Management or Statistics. Therefore a permanent up-to-date query management has to be assured.

CONCLUSION

The main and open-label phases of double-blind studies and their extensions can be combined into one EDC database or kept in two separate databases. Each has advantages and disadvantages. The advantages of the one-database solution outweighed its disadvantages. Avoiding data reconciliation and facilitating data entry are considered the primary reasons for selecting this solution. Taking the tight timelines of today's drug development into account, this way of data handling seems to be the right way forward.

CONTACT INFORMATION

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

Thomas Grupe / Stephanie Bartsch

Bayer Vital GmbH

Building K9

51638 Leverkusen

Work Phone: +49 214 30 51605 / +49 214 30 51816

Fax: +49 214 30 51285

Email: until end of 2010: thomas.grupe@bayerhealthcare.com, stephanie.bartsch@bayerhealthcare.com
from 2011 ongoing: thomas.grupe@bayer.com, stephanie.bartsch@bayer.com

Brand and product names are trademarks of their respective companies.