

PhUSE 2008

Paper P008

How big? Oh my... it's time to prepare!

Julie Oates, GlaxoSmithKline PLC, Harlow, UK

ABSTRACT

What factors do you need to consider when setting up a large study (10,000+ patients).

INTRODUCTION

When a large study (10,000+ patients) is initiated, careful consideration needs to be taken to ensure the:

- Complexity of the study is kept to a minimum
- Systems can cope with the vast amount of data
- Individuals involved in the study understand the importance of their role

RESEARCH

It is important not to reinvent the wheel. Utilise the experience from previously reported large studies within your company and talk to external contacts

If your company has **previously reported** a large study, investigate the learning's, both good and bad, by contacting members of the team involved. They have had to live with decisions they made at the study set-up phase, find out what they would do differently.

Note the issues they may have encountered whilst setting up the study and particularly any issues raised whilst the study was ongoing and after the study ended.

The three words I repeatedly heard from colleagues were **organisation, communication** and **consistency**.

Collaborations with **Academic Research Organisations (AROs)** provide invaluable specialist advice based on previous experience and should be sought wherever possible.

Obviously not forgetting the **internet** which is a minefield of useful information,

STUDY DESIGN

Clearly identified study objectives in the protocol will help focus the design of the data capture element.

Large studies bring their own problems; every minute issue which may arise is magnified. Identify a team (i.e. Data Management, Statistics & Programming, Clinical to name just a few) who will each bring their own expertise to the table. Keep the **complexity of the data capture to a minimum by maximising the efficiency of the CRF design**.

Remember clever programming can combine and report data from different areas of the Case Report Form without having to capture data to the finite level

PhUSE 2008

SYSTEMS

Can the current systems cope?

Upfront risk assessments of the end-to-end systems are imperative. This should include any external company systems which are used as part of your reporting process. Also assess the stability of the systems over time for studies with long study duration. e.g. Will the systems be upgraded during this time?

These will be looked at individually and are grouped into the following areas:

DATA CAPTURE

Electronic Case Report Forms (eCRF) is usually the main method of data capture. There are a number of risk assessment questions to address

- **Access to system** - How many sites/countries need to access the system?
Previous experience with smaller studies using eCRFs to capture data has few problems. Potentially a large number of users trying to access the system at once could slow down the system or at worse crash it.
- **Maintenance work** – When will this be performed?
If your data is collected in numerous countries, these will cross many time zones, making it difficult to identify a suitable window for any essential maintenance work.
- **Extracting data** – How will this be performed?
Extracting large quantities of data could be quite time consuming, schedule enough time to be able to meet appropriate deadlines
- **Post study start eCRF changes.**
This should be avoided if at all possible. Deciding to collect additional information on a study which has, say, been active for X number of patients can cause many issues. Do you go back in time to try to collect the new data for the patients already in the study? This can be a nightmare for Data Management. Encourage your Study Team (ST) to plan carefully up front when the eCRF is being designed to prevent this from occurring.

External data capture is often used to collect additional data not entered onto the eCRF e.g. LAB data. Make sure a process has been identified and tested with the external vendor ahead of any reporting. This ensures the vendor provides adequate reassurance their systems can cope with the amount of data and the transferral process is problem free. Test these processes with your vendors.

DATA CLEANING

The study team define the **data validation system (DVS)** checks whilst setting up the study. These will be automatically activated at the time data is entered on the system, e.g. Systolic BP value limits flag a potential data query.

Data management (DM) will define a process for raising queries and chasing these up with sites. With a larger study this potentially could be more problematic. Ensure specific data areas of interest (i.e. Primary/Secondary endpoints) are fully checked but scale down the areas to query wherever possible.

Assessing the limitations of the software is crucial, e.g. handling of the amount of data and time taken to run checks

DATA REPORTING

As with any of the systems discussed so far, the reporting tools need to be assessed for robustness. The reporting aspects split up into two main areas, the creation of the:

- **Analysis & reporting (A&R) data sets**
- **A&R displays (Tables, Figures & Listings)**

To test the reporting system, **simulate** a large study to see how it is handled. Select a previously reported study which displays the same type of “events schedule” as the one you would desire, then “balloon” the data to replicate the subject count and visit schedule you require.

PhUSE 2008

Run a selection of programs which require a large degree of manipulation (e.g. LAB data), also test submit these in batch.

Assess the **time taken to run the program** and the **percentage central processing unit (%CPU)**.

Table 1: Example table of test results from the creation of reporting data sets from the original raw data sets

Data set	Size of raw data set (KB)	Max % CPU required to run program	Time taken to run (mins)	Comments
Adverse events	16,040	11.2	1	Ran ok
Lab data	1,294,504	12.5	97	Program fell over with error "Work file XXX is full or may be damaged". This error occurred whilst running a standard LAB transformation macro. Re-ran with MEMSIZE=300 and it fell over after 97 mins. The program fell over with error "Work file XXX is full or may be damaged". This error occurred whilst running an imbedded study specify macro
Vital signs	12,408	12.3	2	Ran ok

By assessing the % CPU taken of the server, will indicate how running your program will impact other people using the system on the same server. This may highlight problems and potentially crash the system especially if multiple programs run at the same time (batch submitting).

Note down all aspects of the testing as this can be used in the official risk assessment (if required)

Will you need to combine data for submissions? With one or more large studies combined this could just escalate all the issues found to date. Consider your submission requirements when testing your systems well ahead of time.

Good programming practice should be followed to minimise the amount of time a program takes to run. Only keep the variables you require and delete temporary data sets throughout your program if no longer required.

DATA ARCHIVING

At some point the study will need to be archived. The system used will need to be assessed, to understand its limitations

- **Is there a restriction on the size of individual files?** This might specifically apply to data sets, LAB data being a prime example; in this case it may need to be split up into more manageable data sets. i.e. LAB_CHEM and LAB_HAEM.
- **Is there an overall restriction for a study to be archived?** Sometimes "folders" have size limitations; check these out ahead of time.
- **Will users be affected by the response times when accessing the study?** Large files could affect the amount of time a user may require to view the contents of the file.

TEAMWORK

As with any study, **communication** is imperative, maybe more so when working on a large study to alleviate problems arising. Hand in hand with this are the organisational skills required to ensure every aspect of the study stays on track.

Documenting issues as they arise and sharing this information across appropriate members of the team will enable the study to run more smoothly. Take into account the issues which were collated when the system(s) were tested and ensure these are considered when defining timelines/responsibilities.

When reporting a large study which may not complete for a number of years, **consistency** in staff working on the study is preferred but unlikely in today's climate. This only reinforces the importance of the communication and organisation skills. With these, it will enable a **consistent** approach taken to in the conduct of the study.

PhUSE 2008

ONGOING REVIEWS OF THE DATA

With large studies, ongoing reviews of the data are usually required as this will be a good opportunity to see how a large number of patients react to the drug. Obviously to the company, the study treatment groups will be blinded but an Independent Data Monitoring Committee (IDMC) will review the data unblinded to assess for safety issues.

As the study team need to remain blinded, it is common to employ a Statistical Data Analysis Centre (SDAC), who will be responsible for providing the IDMC members with the unblinded displays (Tables, Graphics, and Listings as required) on behalf of the trial sponsor. It is imperative to set up a process with the SDAC to ensure all roles are clearly defined and there are no opportunities of potentially unblinding trial staff to the data unintentionally.

The following questions come to mind:

- How much of the work will the SDAC be performing, i.e. writing programs from scratch or using standard programs?
- How will the SDAC receive the file to unblind the treatment groups?
- Will GSK receive blinded versions of the displays which are sent to the IDMC?
- What process will the SDAC approach regarding updating GSK after the IDMC meetings?

Also consider the other various types of periodic blinded reporting the team will be required to perform during the collection phase of the study e.g. Data looks and general safety signal detection.

CONCLUSION

This poster only covers certain considerations when designing/reporting large studies, but by no means have all avenues been covered. Within your own company you can use the information as a starting point to get those cogs turning for your own requirements

In summary, please remember:

- Issues are SCALED up due to the size of the study
- Don't dive straight into designing the study without first researching other large studies, identifying the pit falls along the way
- Risk assessments should definitely be performed on all the systems you intend to use to capture/clean/report/archive the study data
- Having a team who communicate well, are highly organised and remain consist for the duration of the study is preferred
- Always use the LAB data set as your example!
- For IDMC reporting, ensure a clear defined process is in place to ensure the blind is maintained.
- Plan ahead!

ACKNOWLEDGMENTS

I would like to thank the following people who have helped to tease out the information required to create this poster

- LuAnn Jahnke (DM)
- Julie Anderson (Statistics)
- Diane Miller (Statistics)
- Everybody who reviewed this poster

CONTACT INFORMATION

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

Author Name: Julie Oates
Company: GlaxoSmithKline PLC
Address: New Frontiers Science Park South (H82/4/201 A09)
Third Avenue
City: Harlow
Postcode : CM19 5AW
Email: Julie_A_Oates@gsk.com