Paper DH04

From Data Capture to Decision Making
Innovation through Standardization

Michaela Jahn, Stephan Laage-Witt
F. Hoffmann-La Roche Ltd, Basel, Switzerland

ABSTRACT

Almost everybody involved in the evaluation of clinical studies is experiencing requests to be innovative and at the same time do the daily job of exploring data, detecting clinical signals, ensuring the safety of the patients in clinical trials, potentially seeing signs of early efficacy of the treatment under investigation and making early decisions related to the clinical studies.

Since the evaluation of data is very often done with sub-optimal tools and processes requiring a lot of manual work, time and space that is needed to become creative and develop new ideas is lost.

Standardization and the use of efficient processes is one way of granting time and space to become innovative.

We will show where standardization is implemented at Roche throughout the data flow up to early decision making. The time gained from standardization and efficient processes can be spent as quality time for data exploration.

INTRODUCTION

Clinical study designs are increasingly complex. A growing number of studies are using adaptive designs and require decisions during the conduct of the study. At the same time, the amount of data, the variety of data types and the time pressure for decision making is growing. During study conduct, scientists are under high time pressure and need to manage a multitude of tasks, such as medical data review, signal detection on study and project level, conducting preliminary analysis, preparing for database closure and working on publications and presentations.

Above all, the clinical scientists are expected to drive the innovation in pharmaceutical drug development with new study designs, new assays and new ways to look at data. Innovative thinking requires time and a mind at ease, a contradicting requirement in the busy world of clinical studies.

The key topic of this paper is to describe how an effective and streamlined data flow from data capture to decision making can support the scientists in their intrinsic key responsibility to innovate drug development. The specific deliverables of the revised data flow are focusing on two aspects:

1. Early and speedy access to quality data during study conduct including integrated data displays and the ability to pool data across studies and projects.
2. Flexibility to manage changing study designs and incorporate changes to studies during setup and conduct.

Such improvements need to be achieved on the back of a high economic pressures for further improved operational efficiency and continuously high levels of data quality and regulatory compliance.

A COMPREHENSIVE APPROACH TO REDESIGN THE DATA FLOW

Addressing the scientists needs according to these requirements is a tall order. It requires a comprehensive approach looking at the systems, data standards and business processes in a combined fashion. Standardization is the underlying common characteristics to all of these because it offers re-usability and reduced time and effort.

Specifically, there are 4 topics that require consideration:

1. *Simplifying the Data Flow and Tools For Clinical Studies:* The data flow and the involved tools need to be redesigned for seamless data transfers between systems and across functions.

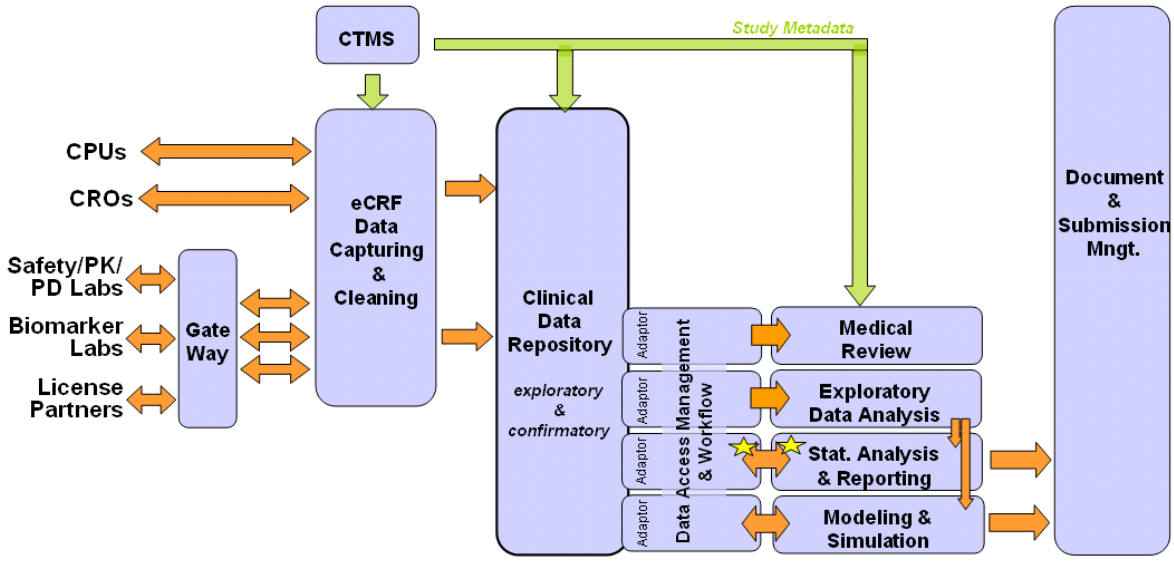
2. *Providing Speedy Access To Study Data:* For each ongoing study, early access to quality study data is required. In addition, the data flow needs to allow for the speedy implementation of study amendments at any time during the study.
3. *Standardizing Data Formats and Displays:* On a project level, the key requirement is to implement integrated data views across multiple studies with minimal manual effort.
4. *Clarifying Cross-Functional Business Processes:* Finally, all functions involved need to be absolutely clear on their contribution and responsibilities across the entire data flow. In addition, there needs to be a clear distinction between mandatory process steps and deliverables versus areas where flexibility is possible and welcome.

SIMPLIFYING THE DATA FLOW AND TOOLS FOR CLINICAL STUDIES

In the department of exploratory development at Roche, the starting point for a redesigned system environment was the analysis of the existing toolset and the data flow. This activity took place 2 years ago and revealed an environment which was historically grown and driven by functional needs. The analysis showed that a multitude of tools and platforms was in use with many interfaces between systems. In addition, a variety of data flows was supported, for example paper based data collection, electronic data captured with off-line laptops at the investigator site and web-based online data capture.

The key design principles for the future system landscape was to minimize the number of tools and databases, eliminate redundant data storage where possible, and use the same tools or platforms across functions. The different options for the data flows was reduced to one preferred way of working: on-line EDC data capture. Access to clinical data via a graphical data review tool was part of the package. The output from this initiative (Figure 1) was a program of system implementation projects to be implemented over the course of 2 years.

Figure 1 Planned System Landscape



The organization took a number of key decisions with far reaching consequences:

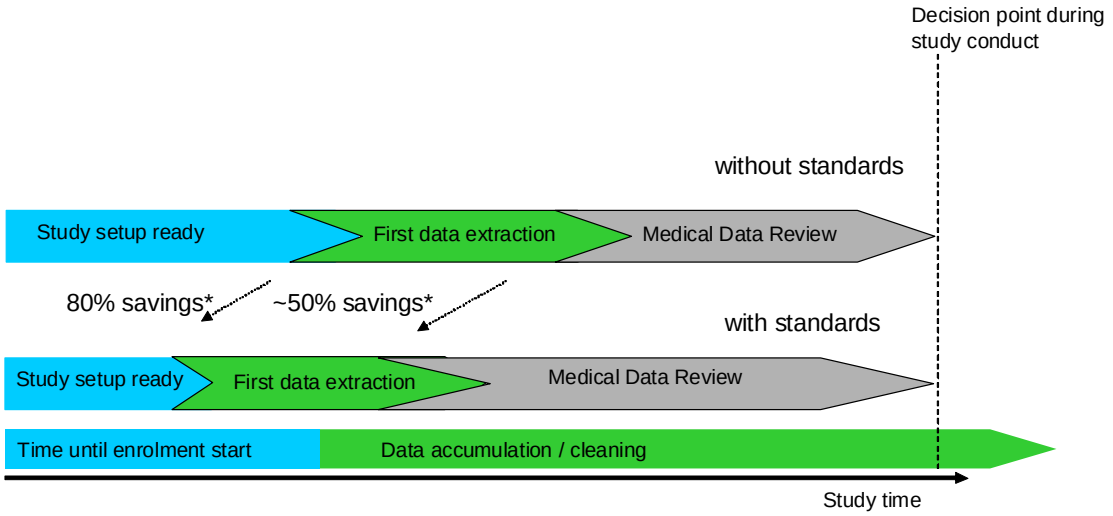
- Tools and Platforms:
  - Medidata Rave® was selected as the data management tool for all clinical trials in exploratory development.
  - An existing platform to store clinical data in a SAS® format was defined as the single, cross-functional repository and is used for all clinical data.
  - Data extraction from Medidata Rave® and upload into the repository is managed via SAS® programs with shared responsibilities between Data Management and Statistics.
  - Tibco Spotfire® was selected as interactive data review tool during study conduct and for scientific decision making.

- Data flow:
    - Web-based EDC via Medidata Rave® was defined as the single and only data flow for all studies in exploratory development.
    - Data are uploaded continuously from Medidata Rave® into the data repository, starting with the first subject enrolled until database closure. There is no difference in the data flow during study conduct and after database closure – except for breaking the blind in the data repository.
    - Clinical scientists are offered access to the SDTM datasets during study conduct.
- Data Standards
    - CDISC/CDASH was implemented as standard for data capture.
    - CDISC/SDTM was implemented as standard for data extraction.

PROVIDING SPEEDY ACCESS TO STUDY DATA

A key requirement for clinical scientists is early and speedy access to study data. This can be greatly supported by the use of global data standards. A Gartner report from 2009 showed that CDISC data standards can reduce the time for study setup by up to 80% and the time for data extraction into a usable format by up to 50% (Figure 2). Such time savings translate directly into the thinking time and space for scientists for decision making.

Figure 2 Saving with Standards



The redesigned data flow offers a variety of components for early and speedy access to study data.

- All studies are handled in the same way with respect to tools, processes and data.
- The use of a pre-defined global library in Medidata Rave® enables faster eCRF and database design. This leads to significantly reduced study start up times.
- Data extraction programs and graphical displays are frontloaded and developed prior to first subject enrollment. When the first subjects arrives the study specific machinery is ready to go so that data arrives quickly in the repository and is available via Tibco Spotfire® displays.

STANDARDIZING DATA FORMATS AND DISPLAYS

Data standardization supports the fast database setup and enables a fast data flow during study conduct. Beyond that, standards are extremely valuable when it comes to integrated analysis reaching across studies. Finally, standards are a strong enabler for presenting data in an interpretable fashion. Downstream tools need to find variable names and types based on standardized names, and scientists are becoming used to this nomenclature.

For the redesigned data flow, CDISC data standards play a key role:

- The study specific databases are built from standardized e-Forms according to the CDASH definitions.

- Data is extracted into a standardized data model (SDTM) which serves all downstream users of the data.
- A global data model captures not only the variable names and types but also hosts descriptions and other metadata helpful for the correct usage of the data.

Re-usability does not only apply to the standardized e-Forms for data collection. It also facilitates the implementation of standardized extraction scripts in SAS® in order to turn the clinical views from Medidata Rave® into the SDTM compliant output files in SAS®.

CLARIFYING CROSS-FUNCTIONAL BUSINESS PROCESSES

The importance of clear accountabilities between the functions can not be underestimated. Unclear roles and responsibilities lead to duplication of work and extra effort for checking the deliverables. This adds to the time and the overall effort and must be avoided wherever possible. Therefore, clear definition of the cross-functional business processes was an essential component of the revised data flow.

To do this most effectively, we implemented a database driven approach for the design of business processes using a tool called "System Architect®" (Figure 3). Initially established for the cross-functional dataflow processes, System Architect is now used for a broad variety of business processes in exploratory drug development.

System Architect® allows to define roles, process steps and deliverables in a relational database avoiding any redundancies. The database content can be turned into multiple output formats ranging from graphical representations via cross-referenced HTML output and PDF documents to custom queries, such as a listing of all tasks and deliverables of a selected role.

Figure 3 Documentation of Business Processes

For the redesigned data flow, we mapped all process steps from study setup through conduct and study closure into System Architect. The roles and responsibilities for any activity were discussed and agreed with all involved functions. The systematic overview of the roles and responsibilities of data management versus statistical programming and other functions facilitated the interface definition of cross-functional boundaries. The HTML and PDF based process maps became very helpful documents for the work of the teams.

NEW RESPONSIBILITIES FOR CLINICAL SCIENCE

Early and speedy access to clinical data during study conduct is a privilege which comes with responsibilities.

In order to work with the data, the clinical scientists need to acquaint themselves with the concept of data models. As a pre-requisite to data exploration, the meaning and interpretation of the variables in data sets needs to be understood.

When receiving data early during study conduct, it is inherent that the data are not clean. This should not cause friction in a team but should be understood by all parties involved.

Clinical scientists need to apply the concept of data exploration: first comes a question, then the data are explored using an adequate tool to get an answer to the question. Following the concept of question and answer should help to look at data in a structured manner, not to get lost in the jungle of data.

Accessing the study data and performing data exploration comes as well with the responsibility to protect the integrity of the study.

SUMMARY OF SUCCESS

The revised data flow including all changes outlined above were implemented over the course of 2 years. Since the beginning of 2010, all clinical studies in exploratory development at Roche are using the defined data flow, streamlined system environment and the CDISC data standards. We now see some of the expected benefits materializing for the clinical studies.

- Study setup times in Data Management are reduced significantly. eCRF build and database are consistently not on the critical path. Depending on the study type, the time between protocol synopsis finalization and eCRF deployment to the sites are reduced substantially.
- The speed of access to data was improved dramatically. Depending on the needs of the study, data are available via Tibco Spotfire® displays in very fast turn-around times. With increasing automation of the process, the times will shrink even further. The limiting factor is no longer the processing time at the sponsor. With all such improvements, we now see that other factors are becoming more important, specifically the time when data is entered at the investigational site or the shipment time of samples to central laboratories.
- Tailored graphical displays for scientists are available for most of the studies. The science feedback to the Tibco Spotfire® displays is very positive. The templates are providing data in a well accepted graphical format. The interactive functionality of Tibco Spotfire® was well received by the scientists.
- In the meantime, there is also quite some experience with the ability to change study designs during study setup or study conduct. A significant portion of exploratory studies require changes or study amendments. In many case, those could be implemented with very fast turn-around times. Key factors for this were the streamlined toolset with less system interfaces and the clarification of roles and responsibilities for this tasks.
- Finally, the partnership across all involved functions was improved significantly which had direct impact on the speed of study setup, data availability and analysis and reporting. The most helpful contributions were the agreement on the same technical platform, SAS®, for data extraction and reporting and, again, the clarification of accountabilities across the functions.

CONCLUSIONS & LEARNINGS

The key elements to enable scientific innovation in drug development are

- Early and speedy access to study data in a useable format and
- Time and space for scientists to work with the data.

The daily business in drug development, however, frequently does not provide room for both, data availability and thinking time. Correspondingly, the objectives of the comprehensively redesigned data flow at Roche Exploratory Drug Development was to achieve this while maintaining the overall efficiency and regulatory compliance.

The clinical data flow is using a complex machinery. With the approach outlined above we could show that significant improvements are possible if all components of the data flow are taken into account and a comprehensive improvement program is implemented. We also observed that sophistication does not necessarily lead to improvements. Rather the opposite: simplification and standardization are enablers for data availability and thinking time.

Finally, anyone who has access to study data during conduct is required to use this carefully. Obviously, any information that potentially impacts the blinded status need to be managed with very tight scrutiny. Moreover, requests to improve the data quality or change data formats can be addressed with very limited resources only. Scientists have to accept the data as they are. However, once they buy in to this constraint, there are many opportunities to explore the data and apply innovative thinking for the benefit of the drug development project.