CLINICAL INVESTIGATION

Pathway to improved data management in clinical trials

Wolfgang Kuchinke*,[1]
& Töresin Karakoyun[2]
[1]Heinrich-Heine University Düsseldorf, Clinical Centre, Moorenstr. 5, 40225 Düsseldorf, Germany
[2]Coordination Centre for Clinical Trials, CRI, Moorenstr. 5, 40225 Düsseldorf, Germany
*Author for correspondence: wolfgang.kuchinke@uni-duesseldorf.de

3rd Annual Data Management in Clinical Trials
Vienna, Austria, 19–20 February 2014

Data management in clinical trials is changing rapidly, moving towards eClinical Trials. eSource data are gaining in importance and primary care data, as well as hospital data, are being used to support patient recruitment and data collection. Here, the two EU projects TRANSFoRm and EHR4CR offer software solutions. Risk-based and remote monitoring is complementing the data collection process. Increasingly operational data are collected in addition to clinical data to improve quality control and trial supervision. But the most important aspect is the comprehensive implementation of standards. Here, the conference provided strategies to deal with the introduction of standardization, especially for the integration of diverse data sources into the clinical trials process.

At the 3rd Annual Data Management in Clinical Trials 2014, presenters from academia, hospitals, clinical research organizations (CROs) and pharmaceutical industry presented their experiences dealing with the challenges of new developments in clinical trials data management. Three developments stood out that may change the way data management is done: first, increased importance of eSource data and data from the patient care domain; second, integration of risk-based monitoring and remote source data validation; and third, implementation of standards and usage of operational data for comprehensive quality control, trial supervision and intelligent outsourcing. Two EU funded projects, TRANSFoRm and EHR4CR, introduced new ways to integrate care data into the research process. The conference thus provided valuable insights into the present problems and future developments of data management in clinical trials.

Data management in clinical trials is the process of collecting, cleaning, managing, analyzing and archiving of trial subject data in compliance with standards and regulations. The primary aim of clinical data management is to provide high-quality data by keeping the number of errors and missing data as low as possible. For this purpose, a number of best practices are available to ensure that data are complete, robust and processed correctly and that patient safety is ensured. Computer systems to support clinical data management have to maintain an audit trail and provide means for data cleaning and resolution of data discrepancies. Although, software for clinical data management has become sophisticated and is able to handle data even of large and complex clinical trials, redundancies and inefficiencies still slow down the trial process. This problem has resulted in the development of new concepts for data management [1], such as the ones for eClinical Trials and eSource data integration. The 3rd Annual Data Management in Clinical Trials congress turned its focus on these new concepts and provided opportunities to discuss the current state of implementation.

The ever-changing landscape of data management

Jonathan Andrus (SCDM and BioClinica; CA, USA) opened the conference with an

FUTURE SCIENCE part of fsg

excellent overview of the current problems and recent advances in data management and discussed risk-based monitoring, Electronic Data Capture (EDC), standardization, outsourcing of data management operations and regulatory challenges. US FDA [2,3] and EMA (European Medicines Agency) [4] support eSource data collection and the integration of different types of data sources. The clinical data management landscape can be seen as a 'garden', where technologies and processes can grow, are replaced or may disappear. Several parts of this clinical data management garden have grown recently. First, the standards landscape (e.g., ISO, CDISC, IHE) has increased in importance. CDASH [5] is increasingly succeeding as data standard for acquiring data by limiting the amount of data that must be collected. An important consideration is that CDASH should be mapped to SDTM [6] to ensure less effort in post-export transformation of data. Recently, SHARE [7], a repository of metadata standards, is being continuously extended. Second, growth in the garden of industry solutions is being hampered by the continuing uncertainty regarding sufficient funding for clinical trials and instability created in the software solutions area by vendor acquisitions/mergers. Third, in the electronic solutions landscape the Bring Your Own Device (BYOD) model is increasingly accepted and for example, smart phones as patient diaries for patient reported outcome (PRO) are being employed. In addition, shared site portals able to distribute data and documents between study participants are gaining in importance. Fourth, standardization in site qualification and training as well as risk-based monitoring grow in importance, with the aim to limit the number of necessary site audits. Risk-based monitoring has been supported for example by a recent FDA white paper [2,3].

Standards are applied to eSource data collection to allow direct data capture and to prevent the occurrence of errors. In the acceptance of eSource data, the Clinical Trials Transformation Initiative [8] has played a role. In summary, the impact of data management in the clinical trials enterprise is changing and now begins to adopt more and more the role of a data broker [9]. Such a data broker manages the secure exchange of data from multiple sources, takes care of semantic interoperability, standards and data privacy requirements. Furthermore, the data broker watches over different data collection systems used at a site and creates integration points.

Data quality as result of good data management

Data quality is the central theme of data management in clinical trials. Erik Merwitz (AbbVie; IL, USA) described the role of data management organizations in support of risk-based monitoring. AbbVie was a late adaptor of EDC technology; because the employment of eClinical trials made changes in the entire management structure necessary. The aim was to develop data management into a hub for all types of study data, for clinical data but also for operational trial data. On the data reporting side, a dashboard should be used as a communication tool showing predictive analysis results and visualizations to further study transparency. The inclusion of operational data in trial analysis makes it possible to focus on the small number of sites that are not compliant.

Albrecht de Vries (Janssen R&D; NJ, USA) discussed risk-based trial oversight. Increasing costs of on-site monitoring visits resulted in considering source data validation as an inefficient procedure to guarantee data quality. In fact, monitors may only find a fraction of existing errors at a site; and often significant findings are missed. With the concept of quality-by-design, parameters for risks can be defined and risk assessments can drive trial processes and site activities. To improve remote monitoring one should bring together all separate data sources, such as data repositories, study database and audit finding database. Using risk-based monitoring the monitor becomes able to focus on sites that show the highest risk and the worst compliance. The presentation induced a discussion to what degree in future on-site monitoring will still be necessary. In fact, only with on-site monitoring it may be possible to verify that an enrolled patient has actually existed. With remote monitoring, on-site monitoring activities will shift from data verification to relationship building and training. However, the definition of risk is still vague and issues, such as slow recruitment, bad data quality, protocol deviations, severe adverse events, all may present risks to the conduct of a clinical trial.

Expansion of opportunities for data collection & source data

A consistent theme of presentations in this segment was eSource data collection. It illustrates the need to move beyond the conventional collection of trial data and use additional sources. Such integration of data sources can improve the conduct of clinical trials; but the integration is far from simple. Two new solutions were presented that are the results of EU FP7 projects: TRANSFoRm [10] and EHR4CR [11]. Whereas TRANSFoRm extends the reach of data management to primary care data collected by family doctors, EHR4CR exploits hospital information system (HIS) data.

Wolfgang Kuchinke (Heinrich-Heine University Düsseldorf; Düsseldorf, Germany) presented the TRANSFoRm project [10] that deals with clinical trial

data management as part of the Learning Healthcare System [12]. TRANSFoRm is developing an infrastructure that facilitates the reuse of primary care data from electronic health records (EHR) to improve both patient safety and the conduct of clinical research in Europe [13]. Data are collected by several means: EHR, functional case report form (CRF) and a mobile solution for PRO. Additional tools, such as Query Workbench, Decision Support System and Quality Tool, complement the functionalities provided by TRANSFoRm. EHR enabled clinical research is supported in four use cases: identification of populations of patients based on predefined eligibility criteria; identification of patients for recruitment; extraction of a set of clinical data for a given patient; and, extraction of data sets for a given population identified by predefined sets of clinical data. A central role in query creation and data collection is played by a terminology service ensuring semantic interoperability of heterogenic data sources [14]. TRANSFoRm is unique in that it couples knowledge generation with knowledge exploitation (Learning Healthcare System) and provides services for clinicians (decision support), not just using them as trial site personal.

Yiannis Karageorgos (Bristol-Myers Squibb R&D; Brussels, Belgium) described investigator site eSource concept and inclusion of eSource data in the eCRF. The main issue here is good clinical practice (GCP) compliance at hospital sites, a place where requirements of standards collide with life's reality. The question is, how can the sponsor control data collection at the site and ensure a high data quality? The first step should be an integrated risk assessment of the quality of the site. Mapping of minimal quality requirements for system, site, staff and so forth, may allow the mitigation of risks created by emerging problems, such as staff changes, system failures or non-performance of sites. The eSRA project 'Common Investigator Site eSource Readiness Assessment Tool' offers a tool developed by the eClinical Forum that provides a list of quality requirements, including ones for regulatory aspects, such as system vendor, audit trail, data access, privacy, back-up and system maintenance. These requirements, expressed as questions, can be used to evaluate sites and identify compliance issues to support risk mediation.

Ulrike Schwarz-Boeger (Hospital of the Technical University Munich; Munich, Germany) showed an eSource approach for eCRF data collection that had been realized together with HIS vendor Siemens some time ago. In hospitals, data sources often still exist mainly as paper records. Although paper records are easy to use for physicians, the resulting EHR consists of digitalized paper documents with limited usability for electronic processing. Electronic data sources in a hospital are usually the HIS (billing, pathology, radiology) and other systems, such as tumor boards and laboratory systems. Problems may be generated by the habit of account sharing and the missing activation of an audit trail in HIS. In future, the pre-filling of eCRF and the direct data entry with hospital data will become more popular employing standards such as HL7 and CDISC.

Kjell Pennert (The Royal Marsden NHS Foundation Trust; London, UK) raised the issue of the integration of non-CRF data. He presented a personal view from the Marsden Hospital in London with an output of about 450 studies per year. Associated with the conduct of clinical trials are GCP awareness, standard operating procedures, governance, awareness of EU directives and national laws. The hospital wants to move from data management to information management, which will include eSource data and the ability to provide trials data in a form suitable for statistics. At the Marsden Hospital, no paper records exist anymore, an EHR has been built in-house. The EHR can be used for patient recruitment, triggering the existence of an eligible patient and assigning a study number. Recently, the decision was made to replace the clinical research system to use it more efficiently for clinical trials.

In the following panel discussion, the audience was interested in the overlap between the two EU projects – EHR4CR and TRANSFoRm. Controversial opinions existed about the equivalence of differences in the appearance of user interfaces (buttons, scales, and so forth) between web-based and android-based applications for PRO. It was recognized that with eSource and the integration of EHR data, compliance criteria, quality management and system validating becomes more complex and difficult. The EHR is not compliant with GCP and normally clinical trial applications cannot access the EHR because of data security and privacy protection reasons. The question came up, if an eSource repository can become fully validated at all?

Standardization & implementation

Adela Pau (Almirall; Barcelona, Spain) described the move from data management to data science. In this transition, standardization, redefinition of roles, implementation of eClinical trials technologies and smart outsourcing all play a role. In general, in eClinical projects standardization is the most important aspect. In the case of SDTM, most companies implement an extended form of SDTM (SDTM plus) that includes extensions for specific domains. Links have to be built to controlled terminologies and standard annotations. To be serious with eClinical trials means that one has to enforce data integration across different data collec-

tion systems, use global standards and employ metadata repositories. MyTrial, the eClinical solution at Almirall, will include operational data to support integration with CROs. The aim is to build a partnership with CROs to manage the trial risks and to maintain the same level of quality in all associated CROs. This is done by knowledge transfer, the implementation of target indicators and risk indicators. Any deviation from a standard should be subject of an approval process that considers the implications any change can cause.

Pantaleo Nacci (Novartis Vaccines and Diagnostics; NC, USA) stressed the importance of the right choice of data management platform for the implementation of standards. It was decided to use the SAS drug development platform and to implement SDTM 3.2, a version that allows storing trial metadata. A governance board for standards was set up and it was decided that all new clinical studies will be pushed into a clinical data repository to allow data pooling (EDC, paper CRFs [legacy], clinical trial management system, severe adverse events). The transfer of legacy data from old clinical trials into the repository turned out to be a major challenge. Besides considering the use of data standards, original/new coding, missing metadata and the reconstruction of old trials had to be considered for this legacy data transfer.

Richard Perkins (eClinical Forum; Friesenheim, France) and Töresin Karakoyun (Coordination Centre for Clinical Trials; Düsseldorf, Germany) demonstrated EHR4CR [15,16] an EU-funded project that develops a platform to support clinical trials by using patient data from hospitals. Increased investments in R&D have become necessary because of a growing number of endpoints and necessary patients that leads to more complex clinical trials. EHR4CR tools and services were designed to support protocol feasibility, patient identification/recruitment and the pre-population of eCRFs with patient data. Töresin Karakoyun is the head of the clinical research informatics (CRI) group [17] and led the requirements engineering step of the project establishing usage scenarios, use cases and requirements specifications. To identify patients or patient populations, queries are created in a central platform and distributed to local hospital sites. At these sites an approval by a physician is necessary to run the query at the local data warehouse. In case the local search results in potential candidates, the physician receives query results and can notify and contact the patient for recruitment. Anonymous feedback is returned to the central platform and shown in a dashboard as available sites that have identified potential candidates meeting the defined inclusion criteria. The following discussion focused on the role of the hospital using the EHR4CR platform. Hospitals participating in EHR4CR will have a competitive advantage by enabling easier ways of research while ensuring privacy protection. The tool can help investigators, but personal control by physicians still exists. For service provision and sustainability, it is planned to found an institute that may charge a reduced price for academic trials in contrast to commercial trials.

How to implement a standard-based electronic data collecting platform from scratch was shown by Rinkey Prasad (British American Tobacco; London, UK). eCRF data in combination with PRO data for quality-of-life data and device data of smoking behavior have to be combined. Smoking behavior data are captured automatically with optical devices. The first aim was to adopt all relevant standards (CDISC, ePRO, defineXML, LOINC). But besides applying new standards and standard terminologies, the coping with deviations and resistance to standards also plays an important role. A consequent standardization may result in resistance; clinical data managers often assign unique names and have well-tried processes that they do not want to change.

Conclusion

The 3rd Annual Data Management in Clinical Trials congress provided an opportunity to hear from a diverse group of professionals working in clinical trials data management about how they cope with the challenges of new developments in the field. Conducting eClinical trials has become increasingly complex [18] and careful planning, comprehensive employment of standards, a risk-based approach to simplify source data validation and the inclusion of operational data and data visualization tools to monitor trial conduct are successful approaches to make eClinical trials work.

Financial & competing interests disclosure

The authors have no relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript. This includes employment, consultancies, honoraria, stock ownership or options, expert testimony, grants or patents received or pending, or royalties.

No writing assistance was utilized in the production of this manuscript.

References

1 Kush R, Alschuler L, Rucceri R *et al.* Implementing single source: The STARBRITE proof-of-concept study. *J. Am. Med. Inform. Assoc.* 14(5), 662–673 (2007).

2 US FDA. Guidance for industry: oversight of clinical investigations – a risk-based approach to monitoring (2013). www.fda.gov/downloads/Drugs/ /Guidances/UCM269919.pdf

3 US FDA. Guidance for Industry: Electronic Source Data in Clinical Investigations (2013). www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM328691.pdf

4 European Medicines Agency. Reflection paper on expectations for electronic source data and data transcribed to electronic data collection tools in clinical trials (2010). www.ema.europa.eu/docs/en_GB/document_library/Regulatory_and_procedural_guideline/2010/08/WC500095754.pdf

5 Standards and Innovations. Clinical Data Acquisition Standards. www.cdisc.org/cdash

6 Standards and Innovations. Study Data Tabulation Model. www.cdisc.org/SDTM

7 CDISC Shared Health and Clinical Research Electronic Library. www.cdisc.org/cdisc-share

8 Grignolo A. The Clinical Trials Transformation Initiative (CTTI). *Ann. Ist. Super. Sanita.* 47(1), 14–18 (2011).

9 Limaye N. Data management redefined. *Perspect. Clin. Res.* 1(3), 110–112 (2010).

10 Project news. TRANSFoRm project agreement signed by European Commission. www.transformproject.eu

11 Project news. EHR4CR executive summary. www.ehr4cr.eu/docs/EHR4CR%20Executive%20Summary%20Nov%2005,%202011.pdf

12 National Research Council. *The Learning Healthcare System: Workshop Summary (IOM Roundtable on Evidence-Based Medicine)*. The National Academies Press, DC, USA (2007).

13 Wallace P, Delaney B, Sullivan F. Unlocking the research potential of the GP electronic care record. *Br. J. Gen. Pract.* 63(611), 284–285 (2013).

14 Ethier JF, Dameron O, Curcin V *et al.* A unified structural/terminological interoperability framework based on LexEVS: application to TRANSFoRm. *J. Am. Med. Inform. Assoc.* 20(5), 986–994 (2013).

15 Ouagne D, Hussain S, Sadou E, Jaulent MC, Daniel C. The Electronic Healthcare Record for Clinical Research (EHR4CR) information model and terminology. *Stud. Health. Technol. Inform.* 180, 534–538 (2012).

16 Kalra D, Schmidt A, Potts HWW, Dupont D, Sundgren M, De Moor G. EHR4CR Research Consortium, Case report from the EHR4CR project – a European survey on electronic health records systems for clinical research. *iHealth Connections* 1(2), 108–113 (2011).

17 CRI news. Clinical Research Informatics. What we do. www.uniklinik-duesseldorf.de/de/unternehmen/institute/kks/clinical-research-informatics/what-we-do/

18 Ohmann C, Kuchinke W. Future developments of medical informatics from the viewpoint of networked clinical research. Interoperability and integration. *Methods Inf. Med.* 48(1), 45–54 (2009).