Clinical trials information in drug development and regulation

Valkenhoef, Gert van; Tervonen, Tommi; Brock, Bert de; Hillege, Hans

university of groningen
faculty of economics and business

12003-Other

Clinical trials information in drug development and regulation: existing systems and standards

Gert van Valkenhoef
Tommi Tervonen
Bert de Brock
Hans Hillege

SOM is the the research institute of the Faculty of Economics & Business at the University of Groningen. SOM has six programmes:

- Economics, Econometrics and Finance
- Global Economics & Management
- Human Resource Management & Organizational Behaviour
- Innovation & Organization
- Marketing
- Operations Management & Operations Research

Research Institute SOM
Faculty of Economics & Business
University of Groningen

Visiting address:
Nettelbosje 2
9747 AE Groningen
The Netherlands

Postal address:
P.O. Box 800
9700 AV Groningen
The Netherlands

T +31 50 363 7068/3815

www.rug.nl/som

Clinical trials information in drug development and regulation: existing systems and standards

Gert van Valkenhoef
University of Groningen and University Medical Center Groningen
g.h.m.van.valkenhoef@rug.nl

Tommi Tervonen
Erasmus University Rotterdam

Bert de Brock
University of Groningen

Hans Hillege
University Medical Center Groningen

Clinical trials information in drug development and regulation: existing systems and standards

Gert van Valkenhoef*[1,2], Tommi Tervonen[3], Bert de Brock[2], and Hans Hillege[1]

g.h.m.van.valkenhoef@rug.nl, tervonen@ese.eur.nl, e.o.de.brock@rug.nl, h.hillege@umcg.nl
[1]*Department of Epidemiology, University Medical Center Groningen, The Netherlands*
[2]*Faculty of Economics and Business, University of Groningen, The Netherlands*
[3]*Econometric Institute, Erasmus University Rotterdam, The Netherlands*

March 16, 2012

Abstract

Clinical trials provide pivotal evidence on drug efficacy and safety. The evidence, information from clinical trials, is currently used by regulatory decision makers in marketing authorization decisions, but only in an implicit manner. For clinical trials information to be used in a transparent and accountable way, it must be available in a format enabling efficient access and further processing, so that decisions can be explicitly linked to the underlying evidence. Thus, processing and management of clinical trials information plays a critical role in enabling transparent decisions.

With the aim of identifying bottlenecks that prevent transparent decision making based on clinical trials evidence, we review the information systems and data standards that process clinical trials data in drug development and regulation. We find that while systems and standards for the management of single clinical trials are relatively mature, the transfer of information to the public and to decision makers is still an ad-hoc and text-based affair, and the integration of data from multiple studies remains difficult.

1 Background

The drug lifecycle consists of the discovery, clinical development, market authorization and marketing phases [45]. In discovery, promising candidate drugs (leads) are identified (often using computational methods) and evaluated in several phases using different preclinical methods. If a lead is likely to be both efficacious (i.e. it likely has the capacity to produce a therapeutic effect) and safe in humans, it may enter the clinical phase. In clinical development, the drug is evaluated in humans through clinical trials, first primarily for safety (phase I), and then for efficacy (phase II and III). The phase III clinical trials are confirmatory trials that demonstrate the efficacy and adverse event profile of the drug in comparison to placebo, another drug, or both. In the market authorization phase, the evidence from discovery and development presented by the pharmaceutical company is assessed by a team of experts assembled by the regulatory authority (e.g. the Food and Drug Administration (FDA) in the US or the European Medicines Agency (EMA) in the EU). In the end, the regulators evaluate whether the drug has a favorable benefit-risk profile (i.e. the favorable effects are likely to outweigh the unfavorable effects). The pivotal evidence in this assessment is provided by the phase III trials. If the drug is approved by the regulatory authority, it enters the marketing phase. Clinical trials are also performed in the marketing phase, either by the pharmaceutical company or by others, often in the setting of a risk management plan. The outcome of these trials may trigger a re-evaluation by the regulatory authorities, that can lead to suspension of withdrawal of a drug from the market.

The development of new pharmaceuticals is often hampered by failure in their late-development phase, market authorization or subsequent clinical use. For example, in 2009 40% of the drugs submitted for approval in Europe received a negative opinion or were withdrawn before receiving an opinion [11]. In many

*Corresponding author

cases, failure may be prevented, or development may be stopped earlier, through better communication of the requirements for clinical evidence and the methods used to assess benefit-risk [30, 9, 26]. The accessibility of key information on past decisions and clinical trials could facilitate this process greatly.

However, presently there are clear information gaps with respect to the efficiency, transparency and reproducibility of the current drug development, approval, and regulation processes, including the secrecy surrounding data [46], the lack of consistency, transparency, and reproducibility of the methods used to reach conclusions about benefit-risk [16] and, finally, the insufficient communication of important information to patients and professionals [28].

Since the results of clinical trials serve as the main sources of information regarding new medicines, a comprehensive overview of the various information systems that store and process this information could identify the gaps in the information transfer that have led to the problems mentioned above. With this aim, we provide an overview of the existing systems and standards supporting the management and transfer of information from clinical trials. For a brief overview of drug information systems that considers the entire drug information life cycle, we refer to [51].

2 Clinical trial information systems

First, in Sections 2.1 and 2.2, we take the industry perspective and focus on the management of trial information within the pharmaceutical company. Section 2.3 introduces trial registration and Section 2.4 describes regulatory submission and assessment. In Section 2.5 we concentrate on the systems that provide the product information for approved drugs, such as package inserts. Standards and data models for clinical trials information are discussed in Section 2.6 and controlled terminologies in Section 2.7.

2.1 Operational management

Operational management refers to the administrative and data-gathering activities surrounding a single trial. The operational management of clinical trials can be automated by using Electronic Data Capture (EDC), previously also called Remote Data Entry/Capture [2]. EDC can be defined as a computerized replacement for (paper-based) Case Report Forms (CRFs), in which information is entered into a database through a computer entry form [12]. Normally data are validated before being accepted into a database to catch possible data entry errors early on (validation and cleaning), while an audit trail is provided for all data entries and modifications [12]. Clinical Data Management Systems (CDMSs) offer a comprehensive solution to trial data management by including data from multiple sites into a single database. This may be done by means of EDC systems, but also by entering data from paper CRFs, usually through double data entry to prevent transcription errors. Clinical Trial Management Systems (CTMSs) are an extension of CDMSs and provide integrated solutions for the management of clinical trials, including advanced features such as subject recruitment tracking and randomization as well as medication inventory tracking. Real-world systems may overlap the distinctions between EDC system, CDMS and CTMS, and the terms are often used interchangeably. The term EDC, for example, may generally refer to all or most of the functionality described above [12].

Until recently, the management and data collection of the vast majority of clinical trials were paper-based activities. In 2000, only 12% of trials were using an EDC system [5]. However, this number has significantly increased over the last years, to 20% in 2004 [5] and 41% in 2007 [12]. It seems likely that this trend will continue in the foreseeable future, especially given the fact that in 2007 96% of the pharmaceutical companies and 71% of the Clinical Research Organizations (CROs) were to some extent using EDC technology [54]. Because of the increasingly international character of many trials, it is not surprising that the advantages of web-based EDC are becoming widely recognized [42].

Therefore there is a clear need for interoperability of the different operational management systems. In the area of electronic source data (see e.g. [15, 33]), several technologies have gained momentum, such as Electronic Case Report Forms (eCRFs), Electronic Patient Reported Outcomes (ePROs), and Electronic Laboratory Data (eLab) [54]. Now that data are being collected more and more in electronic form, there are efforts underway to automatically enter data from Electronic Health Record (EHR) systems in EDC systems in order to reduce the amount of information that has to be entered manually [44, 8]. It is believed that by reducing the necessity of frustrating time-consuming tasks such as double data entry, the reluctance of sites to participate in clinical trials will decrease [8].

Furthermore, while statistical analyses and report generation of clinical trials have been computer-aided for a long time now, they usually require programming of statistical routines. Analyses are performed using general-purpose statistics programs such as SAS, SPSS, R and SPlus, and are therefore not integrated with CTMSs. Thus, to extract data from the CTMS in a format suitable for statistics programs, intermediate processing is required. However, this approach leads to a loss of the traceability of the summary statistics back to the underlying source data.

Market share information about CTMSs is hard to come by, but a 2001 study indicated that Clinsoft Corporation (now acquired by Phase Forward) and Oracle Corporation (`http://www.oracle.com`) dominate the market. Oracle has recently acquired Phase Forward [41]. These commercial packages have been criticized for only being concerned with the delivery of valid and accurate data that conform to the Good Clinical Practice (GCP) guidelines, and neglecting end-to-end processes as well as the usability of the interface and interoperability with other systems [40]. Moreover, CTMSs can be prohibitively expensive, require considerable expertise to set up, and need specialized in-house IT support to operate [40]. Independent groups or organizations based in developing countries may not be able to afford CTMSs for these reasons, and there is little to no good information about the available commercial offerings [22]. One solution that has been proposed is the development of an open source CTMS through a collaborative effort of research organizations and funders [22]. OpenClinica (`http://www.openclinica.org`) provides an open source solution for web-based clinical data management and is compliant with the regulatory requirements. Similar functionality is provided by the caBIG Clinical Trials Suite (`http://cabig.nci.nih.gov/adopt/CTCF`) of the National Cancer Institute (NCI), although it is primarily focused on the cancer domain. There are several academic clinical data management systems, e.g., TrialDB [35, 4, 36], COATI [40], and OpenSDE [32].

The field of CTMSs has gradually started with the transition from data-centric single study systems toward more interoperable, comprehensive and standards-based systems. This development was initially spurred by a push from the FDA for electronic submissions (see Section 2.6), but has now taken on a life of its own, with the industry and CROs increasingly acknowledging the benefits of information technology [54].

2.2 Data warehousing

A data warehouse for clinical research should enable analyses between trials and across compounds. Clinical trials data warehousing offers unique challenges owing to the diversity inherent in the domain. These challenges include establishing a shared vocabulary of the clinical domain to ensure consistent reporting and capturing the clinical domain in a data model that is sufficiently rich to enable data mining, synthesis, and the analysis of data across different trials. Several academic clinical data management systems have incorporated data warehousing features (e.g. TrialDB, COATI, OpenSDE). Phase Forward, the company behind ClinTrial, offers a data warehousing solution under the name 'Clinical Data Repository' [43], which integrates data from different sources through a consistent audit trail, adopting a meta-data based approach to validation. Although Oracle Clinical facilitates the management of multiple studies in a single database, it is primarily aimed at reducing maintenance overhead rather than at enabling analysis across trials. Other data warehousing projects include the academic METABASE [50], Organon D3W [55, 56], the Cancer Biomedical Informatics Grid (caBIG) CTODS/Cactus, and the Human Studies Database (HSDB) project [49]. ISO has recently published a standard for the deployment of a clinical data warehouse [29], but it focusses on health care rather than research.

2.3 Trial registration

When judging the merits of a treatment, it is critical that all relevant existing clinical trials can be efficiently identified. However, until recently, the scientific literature was the only public source of clinical trial results. This caused difficulty in finding relevant trials and led to insufficient or inaccurate trial reporting and publication bias [10]. In the late 1990s and early 2000s, many countries world wide adopted legislation that requires the design (research plan) of clinical trials to be registered *before* participants are recruited. The prospective registration of clinical trials ensures that the existence of clinical trials is known, even if their results are not published in the peer-reviewed scientific literature. This enables publication bias to be detected more easily. In the US, investigators are additionally required

to register the trial results in the ClinicalTrials.gov registry [59]. So far, other countries do not require the registration of results.

The registration of clinical trials is now a well-established practice and has become a key tool in addressing some of Evidence-Based Medicine (EBM)'s challenges [60]. However, the current registries contain only text-based or semi-structured information and there is, for example, no common vocabulary for labeling interventions. The amount of protocol information registered is often insufficient to judge the validity of reported results and the problem of identifying all relevant studies has not yet been solved [60]. In addition, the information publicly available may be incomplete or even "largely incomprehensible" [59].

2.4 Regulatory assessment

After drug development, the pharmaceutical company compiles the evidence collected from the discovery and development processes into a dossier that is submitted to the regulators who decide upon its market authorization. This is a critical step, which can lead to the disqualification of a compound due to rejection or withdrawal [11]. Both the EMA and the FDA approve approximately 20 – 30 drugs per year [11, 27, 34], and with a 30% failure rate, this amounts to 30 – 45 submissions per year. Submissions to the EMA and the national medicines boards in Europe are mainly text-based, containing aggregate-level results of clinical trials based on the applicant's statistical analysis. Since June 2009 the FDA has additionally required the electronic submission of individual patient data to be able to perform independent analyses [21]. The dossier, especially the clinical trial results, forms the basis on which regulators assess the benefit-risk profile of a new drug. In some cases, the regulatory authorities may give market approval on the condition that additional studies (phase IV trials) are conducted by the company. Such trials are most common in western Europe [53].

In the context of the FDA Critical Path initiative, the FDA and the NCI initiated the JANUS project to build a standards-based clinical data repository specifically meant for the meaningful integration of data [6]. It is argued that robust, machine-readable meta-data are required to achieve the full potential of such a repository. The construction of a suitable meta-data model is a monumental task, especially given the modeling of the numerous decisions made during a statistical analysis. Furthermore, this infrastructure would eventually need to be shared between the regulators and the industry, requiring a complex and sustained cooperation effort. However, the JANUS project is increasingly enabled by standards established by the Clinical Data Interchange Standards Consortium (CDISC) and Health Level 7 (HL7). It is built around an open source data model also called JANUS [23]. The latest released version of the data model is from 2005, but the JANUS project is still ongoing, and was approved as an agency-wide initiative in 2008 [39]. Eventually, JANUS should offer FDA reviewers easy access to both raw and derived data and facilitate re-analyses [6]. The wider standards-based efforts should result in interoperable tools to be used by both the regulators and the industry [6].

Since 2004, the EMA has established clinical trial registration in accordance with the EU Directive 2001/20/EC through the EudraCT system. EudraCT was opened to the public only recently, on 22 March 2011 [20], and the records are being released in a staggered fashion. Due to its limited functionality, EudraCT should be considered as a trials registry rather than a database supporting evidence-based regulatory assessment. However, the EMA does publish the European Public Assessment Reports (EPARs) of all centrally approved or refused medicines on its website. Note that this does not include all applications submitted to the EMA, as they can be withdrawn before a decision is reached [11]. The EPAR contains information on all trials, but is completely textual without a semantic structure. Moreover, its information is directly derived from the submission by the applicant, while there is no standardization concerning what information should be provided, or in which format.

2.5 Medicinal product information

After the regulatory assessment, the information provided by the Summary of Product Characteristics (SmPC) is made available to professionals and patients via the drug label and package inserts. The SmPC is a text document containing important information on the approved medicinal product, such as recommended dosage, contra-indications, possible interactions with other medicines, and side effects. The information is initially stored in the annex for marketing approval as governed by the regulatory authorities [14]. In Europe, the SmPC belongs to the EPAR. Although most of the data contained within the SmPC originate from Phase I-III trials, the label might be changed on the basis of new information

obtained in the marketing phase. The results of pharmacovigilance processes, in the EU summarized in Periodic Safety Update Reports [13], or the outcome of Phase IV clinical trials can lead to such changes. This especially applies to drug profiling, which may result in different safety instructions for patient subgroups, such as children. It must be noted that information does not automatically go from clinical trial reports to the SmPC, but the SmPC is the result of a dialogue between the pharmaceutical company and the regulators, that is mainly based on the results obtained in clinical trials.

Both the EMA and the FDA have proposed initiatives for a more structured SmPC. The EMA has introduced the Quality Review of Documents (QRD) and the Product Information Management (PIM) standards. The QRD annotated template [19] provides a loose verbal structure that should be followed by the SmPCs [17]. PIM is a standard for submitting data in a structure defined by a Document Type Definition [18]. Both QRD and the more advanced PIM are designed for transferring information in a structured format that facilitates translating the product information into the official languages of all EU member states. The FDA has a Structured Product Labeling (SPL) standard similar to PIM [24] and provides label information in SPL format; one can browse through the labels in a user-friendly format on the DailyMed site (`http://dailymed.nlm.nih.gov/dailymed/`) of the National Library of Medicine. However, QRD, PIM, and SPL do not impose semantics on pharmacokinetic and pharmacodynamic properties, the main quantitative clinical data visible in the SmPC. Some non-profit organizations provide condition-specific drug labeling and/or trial information, for example, the Saskatchewan Lung Association for lung diseases (`http://www.sk.lung.ca/drugs`) and the NCI for different types of cancer (`http://www.cancer.gov/drugdictionary`).

Although the efforts to realize publicly accessible SmPC information seem to have paid off and both the EMA and the FDA have created very good SmPC databases, they are not linked to any clinical trial results databases. Such functionality would be preferable as it would enable one to trace the scientific evidence from a drug on the market back to the original clinical trials. Moreover, the drug compendia merely replicate the SmPC information and have been shown to lack consistency in drug-to-drug interactions due to the insufficient standardization of the terminology used [58].

2.6 Standards and data models

The realization that a common standard for clinical trial data would be beneficial for the semantic interoperability of information systems gave birth to the Clinical Data Interchange Standards Consortium (CDISC) in 1997. Currently, CDISC is a large global non-profit organization with representatives from industry, regulatory authorities, and academia, all dedicated to the development of vendor-neutral, platform-independent and freely available standards. One of the first standards established by CDISC was the Study Data Tabulation Model (SDTM) standard. SDTM is a content standard that describes the core variables and domains to be used when composing a clinical trial dataset to be submitted to the FDA. Other standards are the Trial Design Model (TDM) to represent trials' design and help interpret SDTM data sets, the Protocol Representation Model (PRM) to standardize the content of trial protocols, the Laboratory Data Model (LAB) for the exchange of clinical laboratory data, the Analysis Data Model (ADaM) for an efficient generation, replication, and review of statistical analysis results, the Clinical Data Acquisition Standards Harmonization (CDASH) for identifying a basic dataset of elements that should be captured in a CRF, and the Standards for Exchange of Nonclinical Data (SEND) for data collected from preclinical toxicology studies. Finally, the Operational Data Model (ODM) standard defines the content and structure of CRFs and clinical databases in XML. It facilitates the acquisition, exchange, and archiving of operational data from several sources during the course of a clinical trial. The importance of the CDISC standards partly lies in the fact that the FDA has indicated that clinical trial data should be presented to the agency in the SDTM and ADaM formats. The operational standards appear to be robust and well adopted, and the CDISC is expanding its activities to establishing interoperability with EHR systems, to developing models that support specific therapeutic areas, and to enabling cross-study analysis through the SHARE initiative. Figure 1 illustrates the role of the CDISC standards in the operational management of clinical trials and in their regulatory submission. The CDISC website (`http://www.cdisc.org/`) offers extensive documentation on the standards.

CDISC has engaged in several collaborations, of which the one with HL7 is of special importance. HL7 has been developing standards for the electronic exchange of medical, financial, and administrative data between health care information systems since 1987. The foundation of HL7 standards development work is the Reference Information Model (RIM), a high level object model of the health care domain. Several

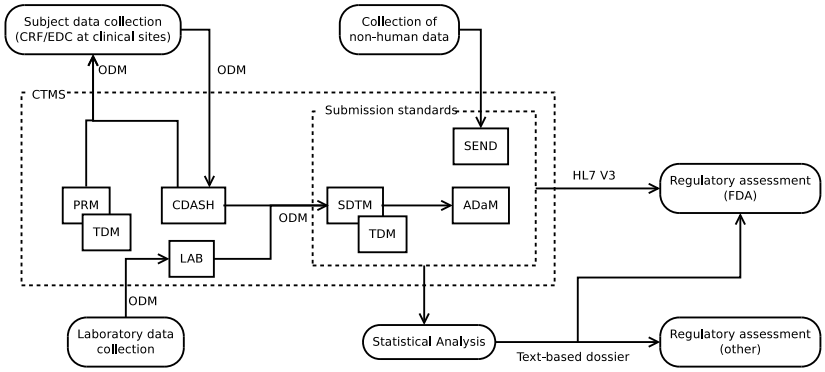

Figure 1: Standards in the operational management of clinical trials and their regulatory submission. Abbreviations: Analysis Data Model (ADaM), Clinical Data Acquisition Standards Harmonization (CDASH), Clinical Trial Management System (CTMS), Health Level 7 (HL7), Laboratory Data Model (LAB), Operational Data Model (ODM), Protocol Representation Model (PRM), Trial Design Model (TDM), Standards for Exchange of Nonclinical Data (SEND).

standards are derived from the RIM, such as V3 Messages for the meaningful interchange of data between health care systems, GELLO for rule-based decision support, and the Clinical Document Architecture for semantically structured documents. CDISC has adopted the HL7 V3 Messaging standard for the exchange of clinical trial data. For the FDA, this standard will replace the antiquated SAS transport file submission format. By adopting the HL7 messaging standard, CDISC ensures that both the clinical trial data and their electronic exchange are standardized. Moreover, through this collaboration, CDISC and other participating parties hope to align the current CDISC standards for clinical trial data with the HL7 RIM standards for healthcare. To achieve this, an overarching domain analysis model is being developed, called the Biomedical Research Integrated Domain Group (BRIDG) model, which is intended to bridge the gap between clinical research and healthcare [25]. Bridging this gap would offer new possibilities for the development of true translational medicine, which means that data obtained in healthcare could be more easily used in clinical research and vice versa.

The BRIDG project is a collaboration between the CDISC, HL7, the NCI and the FDA that aims at bringing together the common elements of their various standards to a shared view of semantics of the domain of protocol-driven research and its associated regulatory artifacts [1]. The model is intended to be implementation independent in the sense that it models the problem domain, and not any specific solution. For example, unlike some other CDISC standards it does not specify the format in which to submit data to the FDA. BRIDG relies on external vocabularies and ontologies, but the specific terminology used is up to the implementer. Due to the increasing complexity of the BRIDG model, several sub-domain views are now delivered as part of the model. These are the protocol representation, study conduct, adverse event and regulatory perspectives. While the operational aspects of clinical trials are well covered by these perspectives, a data analysis perspective is currently missing as there is no adequate standard for statistical analysis. An 'ontological perspective' is planned in the form of a Web Ontology Language (OWL) representation of the BRIDG model, which would enable more formal validation, for example against the RIM. Sophisticated, highly formalized ontologies can be used in advanced applications, such as computer-aided reasoning [47], and thus might enable even broader use of the BRIDG model.

2.7 Controlled terminologies

Controlled terminologies (synonymously: controlled vocabularies, coding systems) of clinical terms are an important first step in the application of information technology to medicine [7]. Controlled terminologies predate information technology, e.g. the International Classification of Diseases (ICD) was already introduced in 1893. The ICD formally codes diseases and enables (for example) the assessment of disease incidence from medical records. Other terminologies fill other niches, for example the Medical Subject Headings (MeSH) [38] is used to index the medical literature (e.g. PubMed meta data is coded in MeSH), and the Medical Dictionary for Regulatory Activities (MedDRA) is used for coding safety

data (e.g. adverse events). Many of these specialized terminologies are organized into a strict hierarchy, which means that some specific terms may fit in multiple places [7]. The Systematized Nomenclature of Medicine, Clinical Terms (SNOMED CT) terminology is an important attempt to create a clinical terminology with comprehensive coverage [48]. It currently contains around 311,000 concepts and 800,000 terms [52]. It also goes beyond a simple hierarchical structure and provides the logical relationships that hold between terms; over 1.3 million such relationships are currently modeled [48, 52]. Due to this complex logical structure, SNOMED CT could itself be viewed as an ontology.

The Unified Medical Language System (UMLS) [31, 3] is essentially a collection of over 60 biomedical terminologies and a coding of the relationships between them through the 'Metathesaurus'. The ICD, SNOMED CT and MeSH are among the terminologies integrated by the UMLS. Like SNOMED CT, concepts in the UMLS are linked through a complex system of relationships. Some of these relationships originate directly from the source terminologies, while others are generated specifically for the Metathesaurus [3]. However, the mappings between terminologies in the UMLS are far from complete [37] and mapping between terminologies, especially to SNOMED CT, is an active area of research, e.g. [37, 57].

Thus, there are many controlled terminologies for medicine (in fact, most were not mentioned), but unfortunately there is as yet no standardization of which ones should be used, and mapping between them is an open problem. For example, in clinical research MedDRA is used to code Adverse Drug Events (ADEs), while the healthcare area prefers the SNOMED CT dictionary. This hinders the interoperability of the various information systems being used.

3 Discussion

Over the last decades, several standardization bodies (notably CDISC and HL7) and CTMS vendors have put great effort into automating the information-intensive aspects of drug development. In this area, the focus is shifting from core data management to electronic sourcing, such as linking to the EHR, and to using increasingly advanced and standardized information flows. However, these systems and standards are still largely oriented toward the operation of single studies, while the issue of storing the data of multiple studies in a structured and meaningful way remains largely unsolved. Although progress has been made, there are no known large, successful, and publicly available data warehouses, nor any standards that would enable cross-study analyses of aggregate level results.

A process view on drug development and regulation (Figure 2) shows that in spite of the largely successful efforts to create an electronic infrastructure for information management during the execution and regulatory submission of clinical trials (top-left of the diagonal), both the dissemination and integration of the resulting evidence to the public and the scientific community remain inefficient and ad-hoc processes (bottom-right). The flow of information from the CTMS to the FDA in the US is standards-based and largely automated, but in other countries it is usually a text-based transfer of aggregated results that does not support the independent verification or re-analysis of the submitted data.

Although the operational systems that help manage individual studies during the development phase are mature and standardized, the subsequent transfer of evidence to scientific journals, public registries, regulators, and (eventually) clinical practice is a largely ad hoc and text-based affair. Consequently, a lot of effort is put into making the results of trials public, but the current systems do not facilitate optimal use of that information. Whereas the current system is centered on single studies, most decision makers need a system that enables the integration of evidence across studies. We believe that this issue is one of the root causes of the lack of transparency in the processes of drug development and regulation.

Acknowledgements and Funding

This study was performed in the context of the Escher project (T6-202) of Top Institute Pharma.

We would like to thank Morris Swertz of the Groningen Bioinformatics Centre and Marcel Hekking and Hans van Leeuwen of Merck, Sharp & Dohme for their insightful comments.

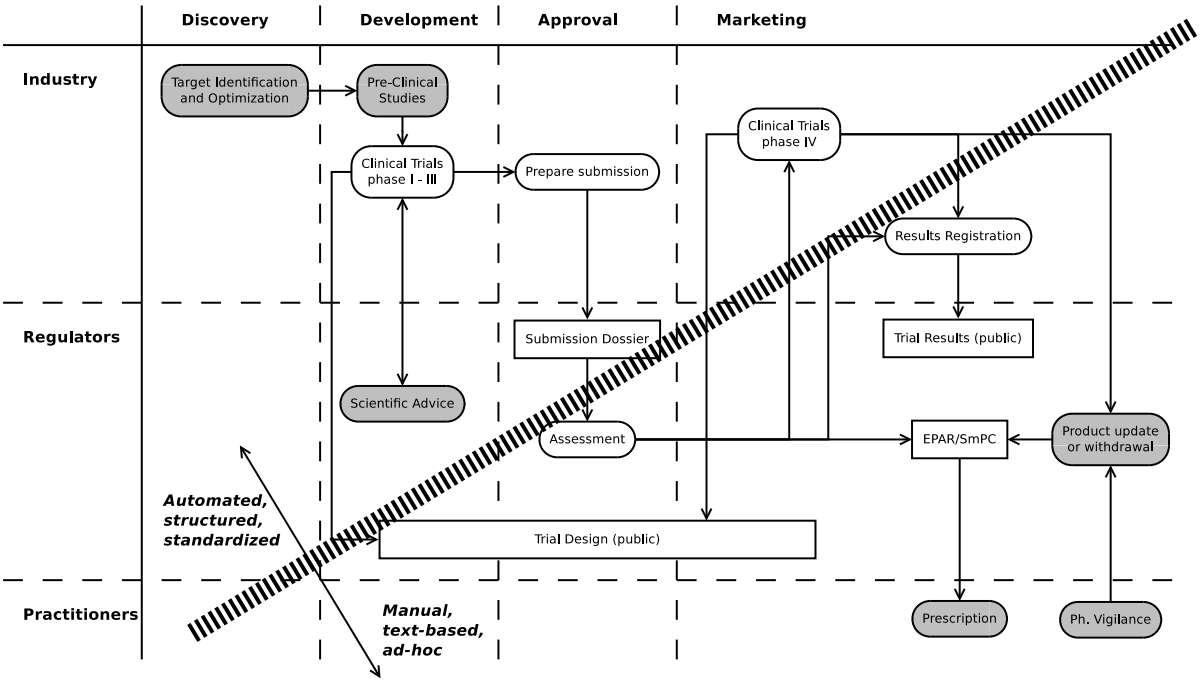

Figure 2: A process view on drug development and regulation. The boxes with round corners represent processes, and those with straight corners information products. The arrows indicate the transfer of evidence. The gray boxes are not discussed in detail, but complete the picture of drug development and regulation. Abbreviations: European Public Assessment Report (EPAR), Summary of Product Characteristics (SmPC).

References

[1] Biomedical Research Integrated Domain Group (BRIDG). *BRIDG Model Release 3.0.3 Users Guide*, 2010. URL `http://gforge.nci.nih.gov/frs/?group_id=342`.

[2] P. Bleicher. Clinical trial technology: at the inflection point. *Biosilico*, 1(5):163–168, 2003. doi: 10.1016/S1478-5382(03)02373-4.

[3] O. Bodenreider. The Unified Medical Language System (UMLS): integrating biomedical terminology. *Nucleic Acids Res*, 32:D267–270, 2004. doi: 10.1093/nar/gkh061.

[4] C. Brandt, P. Nadkarni, L. Marenco, B. Karras, C. Lu, L. Schacter, J. Fisk, and P. Miller. Reengineering a database for clinical trials management: Lessons for system architects. *Control Clin Trials*, 21:440–461, 2000.

[5] CDISC. CDISC 2004 research project on attitudes, adoption, and usage of data collection technologies and data interchange standards; executive summary, Sept 2005.

[6] CDISC and FDA. Walking down the critical path: The application of data standards to FDA submissions. A Discussion Paper by CDISC and FDA, February 2005.

[7] J. J. Cimino. Review paper: coding systems in health care. *Methods Inf Med*, 35:273–284, 1996.

[8] ClinPage. Trial-friendly EHR systems, February 2009. URL `http://www.clinpage.com/article/trial-friendly_ehr_systems/C9`. Archived at `http://www.webcitation.org/5mjBWgITQ`.

[9] P. M. Coplan, R. A. Noel, B. S. Levitan, J. Ferguson, and F. Mussen. Development of a framework for enhancing the transparency, reproducibility and communication of the benefit-risk balance of medicines. *Clin Pharmacol Ther*, 89(2):312–315, 2011. doi: 10.1038/clpt.2010.291.

[10] K. Dickersin and D. Rennie. Registering clinical trials. *J Am Med Inform Assoc*, 290(4):516–523, 2003. doi: 10.1001/jama.290.4.516.

[11] H.-G. Eichler, B. Aronsson, E. Abadie, and T. Salmonson. New drug approval success rate in europe in 2009. *Nat Rev Drug Discov*, 9:355–356, 2010. doi: 10.1038/nrd3169.

[12] K. El Emam, E. Jonker, M. Sampson, K. Krleza-Jeric, and A. Neisa. The use of electronic data capture tools in clinical trials: Web-survey of 259 canadian trials. *J Med Internet Res*, 11(1):e8, 2009. doi: 10.2196/jmir.1120.

[13] EMEA. ICH Topic E 2 C (R1), clinical safety data management: Periodic safety update reports for marketed drugs, step 5: European Medicines Agency (EMEA), 1997.

[14] EMEA. A guideline on summary of product characteristics: The rules governing medicinal products in the European Union. EudraLex, Vol. 2C, 2005.

[15] EMEA. Reflection paper on expectations for electronic source documents used in clinical trials (draft). European Medicines Agency (EMEA). Report No.: EMEA/505620/2007, 2007.

[16] EMEA. Reflection paper on benefit-risk assessment methods in the context of the evaluation of marketing authorization applications of medicinal products for human use. Committee for medicinal products for human use, European Medicines Agency (CHMP-EMEA). Report No.: EMEA/CHMP/15404/2007, 2007. URL `http://www.ema.europa.eu/docs/en_GB/document_library/Regulatory_and_procedural_guideline/2010/01/WC500069634.pdf`.

[17] European Commision. A guideline on summary of product characteristics (SmPC), September 2009. URL `http://ec.europa.eu/enterprise/sectors/pharmaceuticals/files/eudralex/vol-2/c/smpc_guideline_rev2_en.pdf`.

[18] European Medicines Agency. PIM: data exchange standard, 2010. URL `http://pim.emea.europa.eu/des/index.html`. Archived at `http://www.webcitation.org/5nlDWKtab`.

[19] European Medicines Agency. Human medicines - quality review of documents (QRD), 2010. URL `http://www.emea.europa.eu/htms/human/qrd/qrdtemplate.htm`. Archived at `http://www.webcitation.org/5nlDmjKRR`.

[20] European Medicines Agency. About EU clinical trials register, 2011. URL `https://www.clinicaltrialsregister.eu/about.html`. Archived at `http://www.webcitation.org/5yNkGPZZX`.

[21] FDA. Guidance for industry: Providing regulatory submissions in electronic format drug establishment registration and drug listing. US Food and Drug Administration (FDA). OMB Control No. 0910-0045, May 2009.

[22] G. Fegan and T. Lang. Could an open-source clinical trial data-management system be what we have all been looking for? *PLoS Med*, 5(3):347–349, 2008. doi: 10.1371/journal.pmed.0050006.

[23] Food and Drug Administration. Janus operational pilot, 2010. URL `http://www.fda.gov/ForIndustry/DataStandards/StudyDataStandards/ucm155327.htm`. Archived at `http://www.webcitation.org/5nmLXVuZE`.

[24] Food and Drug Administration. Structured product labeling resources, 2010. URL `http://www.fda.gov/ForIndustry/DataStandards/StructuredProductLabeling/default.htm`. Archived at `http://www.webcitation.org/5nlE9bM4Y`.

[25] D. Fridsma, J. Evans, S. Hastak, and C. Mead. The BRIDG project: A technical report. *J Med Inform Assoc*, 15(2):130–137, 2008.

[26] J. J. Guo, S. Pandey, J. Doyle, B. Bian, Y. Lis, and D. W. Raisch. A Review of Quantitative Risk-Benefit Methodologies for Assessing Drug Safety and Efficacy-Report of the ISPOR Risk-Benefit Management Working Group. *Value Health*, 13(5):657–666, 2010. doi: 10.1111/j.1524-4733.2010.00725.x.

[27] B. Hughes. 2009 FDA drug approvals. *Nat Rev Drug Discov*, 9(2):89–92, 2010. doi: 10.1038/nrd3101.

[28] A. Irs, T. Janse de Hoog, and L. Rägo. Development of marketing authorisation procedures for pharmaceuticals. In N. Freemantle and S. Hill, editors, *Evaluating Pharmaceuticals for Health Policy and Reimbursement*, pages 3–23. Blackwell Science Ltd, 2004. doi: 10.1002/9780470994719.ch2.

[29] ISO TC 215 (Health Informatics). ISO/PRF TS 29585: Deployment of a clinical data warehouse, 2011. URL `http://www.iso.org/iso/catalogue_detail.htm?csnumber=45582`.

[30] L. Liberti, A. Breckenridge, H. G. Eichler, R. Peterson, N. McAuslane, and S. Walker. Expediting patients' access to medicines by improving the predictability of drug development and the regulatory approval process. *Clin Pharmacol Ther*, 87:27–31, Jan 2010. doi: 10.1038/clpt.2009.179.

[31] D. A. B. Lindberg, B. L. Humphreys, and A. T. McCray. The unified medical language system. *Methods Inf Med*, 32:281–291, 1993.

[32] R. Los, A. van Ginneken, and J. van der Lei. Extracting data recorded with OpenSDE: Possibilities and limitations. *Int J Med Inform*, 74:473–480, 2005.

[33] R. Marks. Validating electronic source data in clinical trials. *Control Clin Trials*, 25(5):437–446, 2004. doi: 10.1016/j.cct.2004.07.001.

[34] A. Mullard. 2010 fda drug approvals. *Nat Rev Drug Discov*, 10(2):82–85, 2011. doi: 10.1038/nrd3370.

[35] P. Nadkarni, C. Brandt, S. Frawley, F. Sayward, R. Einbinder, D. Zelterman, L. Schacter, and P. Miller. Managing attribute-value clinical trials data using the ACT/DB client-server database system. *J Am Med Inform Assoc*, 5(2):139–151, 1998.

[36] P. Nadkarni, C. Brandt, and L. Marenco. TrialDB: a clinical studies data management system, 2010. URL `http://ycmi.med.yale.edu/trialdb/`. Archived at `http://www.webcitation.org/5mm4G1K1S`.

[37] P. M. Nadkarni and J. D. Darer. Determining correspondences between high-frequency MedDRA concepts and SNOMED: a case study. *BMC Med Inform Decis Mak*, 10:66, 2010. doi: 10.1186/1472-6947-10-66.

[38] S. Nelson, M. Schopen, A. Savage, J.-L. Schulman, and N. Arluk. The MeSH translation maintenance system: Structure, interface design, and implementation. In *Proceedings of the 11th World Congress on Medical Informatics*, pages 67–69, San Francisco, 2004.

[39] A. Oliva. Janus update, 2009. URL `http://gforge.nci.nih.gov/docman/view.php/142/17571/Oliva%2020090310%20DIA%20Janus%20Update.pdf`. Presentation.

[40] A. Oliveira and N. Salgado. Design aspects of a distributed clinical trials information system. *Clin Trials*, 3(4):385–396, 2006. doi: 10.1177/174077450609156.

[41] Oracle Corp. Oracle buys phase forward, 2010. URL `http://www.oracle.com/us/corporate/press/068204`.

[42] J. Paul, R. Seib, and T. Prescott. The internet and clinical trials: Background, online resources, examples and issues. *J Med Internet Res*, 7(1):e5, 2005. doi: 10.2196/jmir.7.1.e5.

[43] Phase Forward, Inc. Clinical data repository, 2010. URL `http://www.phaseforward.com/products/cdc/cdr/`. Archived at `http://www.webcitation.org/5p8AxFMw0`.

[44] H. Prokosch and T. Ganslandt. Perspectives for medical informatics reusing the electronic medical record for clinical research. *Methods Inf Med*, 48(1):38–44, 2009. doi: 10.3414/ME9132.

[45] H. Rang, editor. *Drug Discovery and Development: Technology In Transition*. Churchill Livingstone, 2005.

[46] I. Roberts, A. Po, and L. Chalmers. Intellectual property, drug licensing, freedom of information, and public health. *Lancet*, 352(9129):726–729, 1998.

[47] D. L. Rubin, N. H. Shah, and N. F. Noy. Biomedical ontologies: a functional perspective. *Brief Bioinform*, 9(1):75–90, 2008. doi: 10.1093/bib/bbm059.

[48] S. Schulz, S. Hanser, U. Hahn, and J. Rogers. The semantics of procedures and diseases in SNOMED CT. *Methods Inf Med*, 45:354–358, 2006.

[49] I. Sim. CTSA human studies database project. Abstract for All Hands Meeting, October 2008.

[50] M. Swertz, R. Oostergo, and B. de Brock. How to design a platform for the uniform integration, extraction, and querying of clinical and biobank data sets? *IEEE Trans Soft Eng*, (Submitted Manuscript), 2010.

[51] T. Tervonen, B. de Brock, P. A. de Graeff, and H. L. Hillege. Current status and future perspectives on drug information systems. In *Proceedings of the 18th European Conference on Information Systems, ECIS 2010*, Pretoria, South Africa, 2010.

[52] The International Health Terminology Standards Development Organisation. SNOMED CT components, 2011. URL `http://www.ihtsdo.org/snomed-ct/snomed-ct0/snomed-ct-components/`. Archived at `http://www.webcitation.org/5yNmLpPkv`.

[53] F. Thiers, A. Sinskey, and E. Berndt. Trends in the globalization of clinical trials. *Nat Rev Drug Discov*, 7(1):13–14, 2008.

[54] Tufts, CSDD and CDISC. Study on the adoption and attitudes of electronic clinical research technology solutions and standards; summary of results, 2007. URL `http://www.cdisc.org/stuff/contentmgr/files/0/e35818eab8d8cc7b9d159ffeba5cdda5/misc/tuftstop30rdkjan08.pdf`.

[55] P. Vervuren. Visions of a drug development data warehouse. In *Proceedings of the First Conference of the Pharmaceutical Users Software Exchange (PhUSE 2005)*, 2005.

[56] P. Vervuren and F. Dietvorst. Contours of a drug development data warehouse. In *Proceedings of the Second Conference of the Pharmaceutical Users Software Exchange (PhUSE 2006)*, 2006.

[57] A. Vikstrom, Y. Skaner, L. E. Strender, and G. H. Nilsson. Mapping the categories of the swedish primary health care version of ICD-10 to SNOMED CT concepts: rule development and intercoder reliability in a mapping trial. *BMC Med Inform Decis Mak*, 7:9, 2007. doi: 10.1186/1472-6947-7-9.

[58] A. Vitry. Comparative assessment of four drug interaction compendia. *Br J Clin Pharmacol*, 63(6): 709–714, 2006.

[59] A. J. J. Wood. Progress and deficiencies in the registration of clinical trials. *N Engl J Med*, 360(8): 824–830, 2009. doi: 10.1056/NEJMsr0806582.

[60] D. A. Zarin, N. C. Ide, T. Tse, W. R. Harlan, J. C. West, and D. A. B. Lindberg. Issues in the registration of clinical trials. *JAMA*, 297(19):2112–2120, 2007. doi: 10.1001/jama.297.19.2112.

List of research reports

12001-HRM&OB: Veltrop, D.B., C.L.M. Hermes, T.J.B.M. Postma and J. de Haan, A Tale of Two Factions: Exploring the Relationship between Factional Faultlines and Conflict Management in Pension Fund Boards

12002-EEF: Angelini, V. and J.O. Mierau, Social and Economic Aspects of Childhood Health: Evidence from Western-Europe

12003-Other: Valkenhoef, G.H.M. van, T. Tervonen, E.O. de Brock and H. Hillege, Clinical trials information in drug development and regulation: existing systems and standards


www.rug.nl/feb