Clinical trial management and quality assurance is a complex activity that, when manually executed, is prone to errors and delays, and organizations involved in the conduct of clinical drug trials must rely on database systems to ensure adequate data integrity and timely reporting. We report the design and implementation of an integrated computer system for the management and control of multiple phase II to IV clinical drug trials, and for automated generation of monitoring and statistical analysis reports that are fully compliant with international guidelines. This Windows-based system incorporates a number of third-party software tools and applications, and its major components are COATI (Control, Assessment and Tracking of Therapeutic Investigations), a client-server database application; DART (Data Analysis and Reporting Tool) for automated data abstraction and reporting; and PANDA (Data Analysis Package) for automated statistical analysis. The system is in production for two years and was used in 15 clinical trials in a diversity of medical conditions and study designs.

INTRODUCTION

Although most people are aware that clinical trials are a data-intensive activity that deals with dozens of variables collected repeatedly in hundreds of patients, it is some times overlooked the fact that a considerable amount of information from heterogeneous sources is also part of a clinical trial. For example, it is necessary to store and process information from various things such as investigators, institutional review boards (IRB), medication batch numbers, monitoring visits, normal ranges of laboratory values, discrepancy resolution forms (DRF) issued and answered, communications log, just to mention a few.

Efficient clinical trial monitoring requires continuous surveillance of enrolment status, investigator performance and protocol compliance in addition to tracking all the administrative procedures and documentation required by GCP guidelines. Detailed reports of study status must be regularly submitted to the sponsor and every query must be answered promptly. These reports contain a considerable amount of information, typically presented in elaborate tables and listings, and are expected to be accurate. Consequently, monitoring is a complex and time-consuming activity requiring highly trained and specialized personnel, and reporting activities represent a substantial portion of total monitoring costs.

Statistical analysis of the study data is also a complex task. Present ICH guidelines require a considerable amount of tables and listings that are difficult to construct. To add to the difficulty, these tables frequently integrate information from different parts of the case report form and often combine events occurring at different times in the study. Other difficulties encountered during the statistical analysis include, for example, the identification of patients and observations eligible for the primary and secondary study populations, or the verification of model assumptions in each of the many analysis that are necessary.

Although research has been active in the field of clinical trial data management, most reported systems have addressed the needs of a single trial or a single type of pathology and commercial systems present a number of limitations in customization, interface type and design.

To address these difficulties, we developed and implemented an integrated environment for tracking and monitoring multiple clinical trials, and for statistical analysis and reporting. This system has been in operation for two years at our organization and was used in 15 clinical trials covering a variety of medical conditions.

DESIGN OBJECTIVES

We defined as follows the main objectives for an application to be used in the management and control of Phase II-IV clinical drug trials and as a tool for the statistical analysis of those trials. It should implement applicable laws and regulations and the international (ICH) guidelines on Good Clinical Practice, on the methodological and statistical aspects of clinical trials and on the content and format of statistical reports, and should be able to produce reports fully compliant with those guidelines. The system should be flexible enough to
accommodate the specificity of each research project and be easily configurable for a particular trial. The interface should be form-based, with data-entry from dynamic lists for ease of use, although the system was intended to be used by clinical study monitors or someone familiar with clinical trials. Finally, data and applications should be platform-independent, and extensive use should be done of existing software tools and packages.

DATA MODEL

We approached the problem by developing initially a concept model of the clinical drug trial paradigm. The final model had the features of a semantic net, and is summarily presented in Figure 1 in the form of a top-level entity-relationship diagram using standard nomenclature. At the implementation level, the model was translated into a relational model.

Although only the main entities are presented in Figure 1, it is readily apparent that a major feature of the model is that it is centered on a patient’s study visit, which seems appropriate considering that clinical trial data is typically organized around study visits. On the right side of the figure, the characteristic structure of a clinical trial’s case report form emerges. The model attempts to capture the whole scenario around a clinical trial and to be neither restricted to the clinical, nor to the administrative or the financial aspects.

The model is also time-oriented, a feature that is not evident from the diagram but which is absolutely necessary for reporting and analysis purposes. For example, in the display of adverse events by patients it is necessary to report which concomitant medication, along with their dosage, had been administered during the interval between the first and last day of the adverse event, and which associated diseases were present during that period.

Because of diverging characteristics across study protocols and among sponsors’ procedures, it was necessary to develop a structure capable of considerable flexibility. For example, the procedure used to allocate patients to therapies varies substantially across studies: the relationship between the study groups and therapies may be known, or there may be only a supply code for each patient; the

---

**Figure 1.** Top-level Entity-Relationship diagram of the data model of a multiple clinical trials management system (COATI). In the E-R diagram, entities are represented as rectangles and associations as arcs connecting entities. Entities may be forced to exist in an association (represented by a single dash near the entity) or not (represented by a circle); the symbol nearer the end of the arc represents the degree of the association: one-to-one (a single dash) or one-to-many (a three branched arc).
allocation table may be done by us, supplied by the sponsor, or unknown to us. It is therefore necessary to model the methodologies of treatment allocation as the only means of ensuring the necessary flexibility to meet all possible combinations of allocation modalities. Figure 2 shows the part of the structure that models treatment allocation. This model satisfies all the above requirements and also provides adequate recording of uncommon situations, such as changes or errors in the administration of the study medication.

Since an important objective of this system was to be also used as an analytical tool, the model also includes the definition of the protocol in terms of statistical analysis, for example the study populations used in the statistical analyses (the safety population and four types of efficacy populations: intention-to-treat, modified intention-to-treat, all patients treated and per protocol), the study limits or the scale of measurement of baseline and efficacy variables (nominal, multiple choice, ordinal, interval and survival).

ICH guidelines were implemented at the data model level by including design details extracted from the guideline; at the reporting level by creating tables and graphs as directed in the guideline; and at the interface level by including the relevant parts of the guidelines into the help system.

ARCHITECTURE

The complete system includes a clinical trials management and monitoring system (COATI, Control, Assessment and Tracking of Investigations), a library of the ontologies necessary for coding patient findings (ICD for concomitant illnesses, ICD for concomitant medications, WHOART for adverse events), a data dictionary, and a report generator (DART, Data Analysis and Reporting Tool) for producing the tables and listing necessary for the control of the trial and the creation of status reports, and the tables and listings required for ICH compliant statistical reports. A statistical analysis program (PANDA, DAta ANalysis Package) was developed to perform, under supervision, the statistical analysis of the study baseline and efficacy variables. This program implements the ICH guidelines on methodological aspects of clinical trials and outputs the tables required by these guidelines.

IMPLEMENTATION

The system was developed in Delphi (Inprise Corporation, USA), a RAD (Rapid Application Development) object-oriented tool for the development of multi-tier database systems for Windows. The present implementation is two-tier (client-server) and the current DBMS is Interbase (Interbase Corporation, USA), a relational DBMS with super-server architecture. However, the system is DBMS-independent.

Reports are produced with Business Objects (Business Objects, France), a decision support system with data mining and data warehousing capabilities. PANDA is an application written in STATA programming language. STATA (STATA Corporation, USA) was selected as the base statistical package because of its rich library of statistical functions. Since the output of STATA is text only, a Microsoft Word Visual Basic (Microsoft Corporation, USA) program was developed for formatting and final display of the results.

FUNCTIONALITY

The system is menu-driven, form-based and uses a standard Windows GUI. The interface is oriented by study and patient. In Figure 3 is displayed an example of the screen used for protocol definition and study parameterization. The sample form shown is used to capture the overall study plan. Since we defined a core of common data elements for all clinical trials (e.g., adverse event reporting), configuration of a new trial is simply a matter of defining the overall study plan, eligibility criteria, study medications and baseline and efficacy variables. This task requires an average of 30-45 minutes, most of the time being spent in the definition of eligibility criteria.
In Figure 4 is shown the patient form. On-study patient data is entered into a single form that follows the traditional structure of the case report forms, a feature that increases user’s (study monitors) adherence to the system.

The user is also allowed to maintain and inspect the metadata required by the system that includes, for example, the drug, illness and adverse event ontologies, conversion tables and reference values for laboratory parameters, or a table of documentation required for clinical trials. However, the maintenance can be done within the Patient form, avoiding thus the swapping between screens. Monitoring procedures are entered in a different section, and include visit reports, problems found, etc. Finally, tables, listings and graphs required for trial control and reporting are automatically printed when selected from a table. Presently, a total of 54 different types of monitoring reports and graphs have been defined.

Among other features of COATI are computerized checking of eligibility criteria, centralized patient registration and randomization, and trial site management. COATI also supports a large number of important needs on communications such as direct patient registration and randomization through OCR technology, and remote data-entry. Whenever the technology is available at the sponsor or at the study centers, study reports may be made available in real-time through the Internet.

For the production of a statistical report, the system creates temporary files that contain the data required for the efficacy analysis. One file contains, in tabular format, the patient numbers, study center, study group, study stratum and randomization date, followed by the values of demographic and baseline variables. Another file, also in tabular format, displays in each row the patient number and visit number and date followed by the values of all efficacy variables recorded at that visit. A third file contains the description of all baseline variables, including the name, type of scale of measurement, and values of nominal, categorical and multiple choice variables. If a study contains data of the survival type, another file is created with the patient number and, for each end-point, the time to the end-point and whether the end-point was reached. Then, PANDA is launched and, inside this application, the user is guided in the parameterization of the analysis. The application uses the information on the scale of measurement of the variables and a decisional tree of clinical trials statistical analysis methodology. For example, the application suggests the best method for dealing with missing values or the best transformation towards normality or, when selecting baseline variables to be used as regressors in secondary analyses, such as post-stratification or subgroup analysis, the system checks for violations of the assumptions of multiple linear regression when
the dependent variable is interval, of logistic regression when the dependent variable is binary, or of Cox regression when the dependent variable is of survival type. The output consists of tables of summary statistics, results of hypothesis testing, estimates of the difference between treatments and, when appropriate, results of the tests for the treatment by center interaction. Additional tables, listings and graphs, as required by the ICH guideline, are automatically produced. There are presently 78 different types of reports.

EXPERIENCE

In the past two years since coming into production this system has been used in 15 clinical trials, covering many different clinical conditions and trial designs: diabetes mellitus, Alzheimer’s disease, vascular dementia, bronchial asthma, chronic obstructive pulmonary disease, colorectal cancer, pyelonephritis, osteoarthritis, seborrheic dermatitis, arterial hypertension and hyperlipidemia. Although a formal evaluation has not yet been done, we estimate that, after the introduction of the system, time savings in the creation of monitoring and statistical reports are in the order of, respectively, 70% and 85%. The reports are much more comprehensive now, and it is quite evident that the error rate is much lower than previously.

Our ultimate goal is the creation of an integrated environment for clinical practice and research. Wherever physicians are using a computer patient record for patient management, COATI may verify the data in the database to identify patients suitable for ongoing clinical trials. The data analysis capability of PANDA will assist investigators in conducting research on stored data; study reports and patient data may be made available to other investigators, physicians or to a study sponsor via the World Wide Web. Such an integrate system could easily boost the research and trial conducting capabilities of a research-oriented institution.

References
2. The European Agency for the Evaluation of Medicinal Products. Note for guidance on structure and content of clinical study reports. 1996.