End-to-End Management of Clinical Trials Data
A Revolutionary Step Toward Supporting Clinical Trials Analysis Over the Next Decades of Clinical Research
Table of Contents

Introduction ......................................................... 1
Capturing Clinical Data at the Source .................. 2
Clinical Standards: An Integrated Technology Approach ........ 2
Visual and Statistical Analysis of Clinical Data ............ 4
Features of an End-to-End Clinical Analytics Environment .......... 5
Benefits and Predicted Outcomes ............................... 6
Closing Thoughts ..................................................... 7
Introduction

The introduction of electronic data capture (EDC) systems in clinical trial data management and analysis triggered the promise of better-managed clinical trials – with greater efficiency, faster access to higher-quality clinical data and better mapping of data flows to business processes. Over time, however, as many organizations transitioned from paper-based to electronic systems, they ended up with multiple EDC systems, legacy clinical data management systems and additional systems for managing different aspects of clinical data management and analysis. As the result of years of dealing with disparate systems, many pharmaceutical organizations are unable to execute timely queries against historical or ongoing clinical trials – or even retrieve details about one particular trial beyond generating some standard reports. The cost and effort to update, maintain and keep these numerous systems in a validated state is prohibitive for many organizations. In addition, some organizations have neglected to dedicate enough detail to robust integration routines that bring diverse sources of clinical data together repeatedly for different trials or therapeutic areas.

Today, the outcome of this electronic Rube Goldberg machine is a set of large libraries of ad hoc integration code that is unusable beyond the specific trial for which it was initially produced. Increasing regulatory scrutiny, combined with pressure from public opinion and patient advocacy groups, is putting pressure on pharmaceutical sponsors to demonstrate the safety and efficacy of new compounds both before approval and long after a compound has been released to the market.

In order to do so, however, pharmaceutical companies will need to do more than make incremental changes to their business processes or information technology environments. Conducting clinical trials with the right balance of time, costs, quality and processes, and ensuring that clinical trials are supported by the right technology environment, will require a revolutionary change – one that replaces silo-based systems with end-to-end clinical data management.

End-to-end clinical data management involves setting up specifications for clinical trial data once during the lifetime of the clinical trial, and being able to reuse the data structure and produce standards-based data – from the point when patient data is captured in the clinic all the way to when the analysis report is in the hands of regulatory authorities and the pharmaceutical sponsor.

End-to-end clinical data management involves setting up specifications for clinical trial data once during the lifetime of the clinical trial, and being able to reuse the data structure and produce standards-based data – from the point when patient data is captured in the clinic all the way to when the analysis report is in the hands of regulatory authorities and the pharmaceutical sponsor.
Capturing Clinical Data at the Source

Life sciences companies conduct 12 years of clinical research, on average, before a new human therapy or device is approved for commercialization. Industry experts estimate that it costs companies between $175 million and $8 billion to bring a new drug to market. About half of the cost is incurred in the development phase, where the drug is tested in humans during the gold-standard of clinical research – the blinded, placebo-controlled clinical trial. Understanding the clinical trial protocol and the statistical approach is essential for managing a clinical trial. Therefore, a flexible trial design – one that guarantees the structure and quality of the incoming clinical data, and the rapid conversion of data coming from disparate sources into analysis-ready data sets – is key for a successful trial.

Over the last 20 years, clinical trials have become more complex and have been taking longer to complete. Factors driving increased trial complexity and duration include the intricacy of the therapeutic area, the increased biological understanding of the disease and the more stringent inclusion/exclusion criteria for trial subjects. Reconciling clinical trial data – including safety and efficacy data – from different trials conducted by different organizations in a pharmaceutical partnership, or different departments within a single organization that may have run over multiple years, is a herculean task. Too many compounds and therapies are still failing in late-stage trials, which means more investment dollars lost and more opportunities missed for viable therapies that could make a difference to the patient.

The adoption of EDC systems in clinical trials has changed the way clinical trials can be defined. The increased efficiency made possible by EDC systems capturing clinical data at the source has helped offset some of the burden created by greater trial complexity. EDC systems have improved the speed and quality of clinical trials data capture and have made a positive impact on the downstream clinical data flow and analysis.

Clinical Standards: An Integrated Technology Approach

End-to-end clinical data management cannot be achieved without a documented mechanism for reusing the electronic data elements that must be defined at different stages of the clinical data flow. As it progresses through the flow, clinical data is reshaped from a per-subject data capture format to an all-subjects tabular format – and finally into the “one proc away” analysis-ready clinical domain table and report. A common dictionary of data elements is required to align these three different data models and to avoid unnecessary conversions and overhead resulting from the integration of clinical data. In addition, data around the data – e.g., dictionaries, controlled terminologies and metadata resulting from the transition from a horizontal to a vertical format – must also be managed and mapped.
The standards developed by the Clinical Data Interchange Standards Consortium (CDISC) have gained increasing industry and regulatory acceptance. The Food and Drug Administration’s preferred standards for the electronic Common Technical Document Guidance used for submission are the Study Data Tabulation Model (SDTM) and the Analysis Data Model (ADaM) standards – respectively, the tabulation and analysis standards. The Clinical Data Acquisition Standards Harmonization (CDASH) standards model resembles SDTM, but it has been adapted for clinical data capture in electronic systems. Many organizations have used SDTM as a basis for the exchange and management of data definitions for both standard and therapeutic areas, either internally to guide the flow of clinical data from the EDC system into the analytical environment or as a standards “contract” with the contract research organization (CRO) to agree on the expected clinical data format. In addition, SDTM is now used as a base data model for collecting data from different trials and therapeutic areas or combining a pharmaceutical company’s legacy clinical data into a single clinical repository.

In order to exploit these standards, a pharmaceutical company’s technology must be able to keep track of evolving CDISC or organization-specific clinical standard data models, and link them into the different studies that use these models. In addition, the ability to reuse the clinical standards utilized when setting up the clinical trial is important. Finally, when these data standards are used, it is important to pre-define the target data models, such as the tabulation or analysis-ready data formats. The pre-definition of the standards allows for standardization of the integration and analysis code, which further decreases the time needed to produce analysis reports after the important “last patient-last visit” trial milestone.

Data points gathered during a clinical trial have diverse sources: the clinic where the investigator team enters the data directly into an EDC system and the randomization IVRS/IWRS systems; the laboratory for analysis of biological tissue; remote measurement systems, such as blood glucose sensors or blood pressure meters; pharmacokinetics groups; and other specialized laboratories and CROs. Most clinical trials are now conducted using EDC systems instead of paper-based systems, and this leads to higher-quality clinical data and the ability to track the origin of data and potential problems more easily.

The act of bringing together all these data types during the trial or just before submission puts a heavy burden on clinical data management groups. Software that generates robust, traceable and human-readable data integration code before the trial has started is of enormous benefit, especially if the software also respects the organization’s clinical data standards. The software should also allow integration procedures to be changed when the actual trial data begins to appear in the clinical systems. For example, in one study, the complexity and burden of clinical trial execution grew at the slowest rate for protocols in Phase III, as companies, looking to contain costs, gathered more data in the early phases of clinical research. Because not all data will actually be used for submission, a data custodian is needed to keep track of all electronic data elements at a metadata (i.e., definition) level to ensure the validity, traceability, confidentiality and compliance of the trial data in agreement with regulations and the clinical trial protocol.

Visual and Statistical Analysis of Clinical Data

In 2010, the FDA released a Draft Guidance for Industry called *Adaptive Design Clinical Trials for Drugs and Biologics*. This guidance described possibilities for changing protocol elements after a trial has started, as long as the conditions that govern the changes have been addressed in the trial protocol. Adaptive trials are suited for some therapeutic areas, and although the bias introduced into statistical analysis by this methodology has received mixed reviews, it is clear that there are new opportunities for reducing the length of clinical trials, or at least optimizing their overall duration and length. Adaptive trials also introduce complexities into the statistical analysis (e.g., Bayesian statistics instead of traditional or “frequentist” approaches), trial logistics (e.g., supply chain for delivery of sufficient drugs or devices to clinical centers) and, generally, the forecasting of resources (people, sites, investigators, subjects, etc.) needed to complete the trial. In addition, the technology environment must be adapted during the trial to support interim analysis (e.g., by the data monitoring committee) that respects blinding of patient data and the required workflow. The identified requirement in this case is the ability to integrate and combine exploratory and advanced analytics with the clinical data management function.

As data sets are created and readied for analysis, it is important to keep the traceability of the data in mind. In order to validate a derived variable's meaning and importance, you must be able to trace variables from the analysis data set all the way back to where they originated in the capture phase. The ability to provide metadata-level traceability in the electronic file that describes the contents of the data, from the CRF to the submission electronic file, is an essential characteristic of modern clinical IT solutions.

All stakeholders involved in bringing together clinical data – the CRO, clinical data manager, statistical programmer and clinician – are responsible for ensuring its quality. In fact, in an effort to raise data quality, many pharmaceutical companies are intentionally blurring the lines between programmers and data managers to ensure more interaction between the two functions. Ideally, a clinical data manager would plot a distribution graph of essential patient variables, and the statistician would provide the clinician with decisional support and a wide range of graphical views of safety and efficacy data.

Increasingly, pharmaceutical companies are adopting collaboration models with their outsourcing partners. In addition, they are using CDISC standards to streamline downstream data handling. These changes have a tremendous impact on clinical development groups. The ability to actively manage data and accompanying data standards model for a study, therapeutic area or compound goes hand-in-hand with the integrated analysis and visualization. This makes it possible to test programs for cleaning and analyzing data on versioned test or production data without having to extract from external database silos – or without having to obtain new data batches from the CRO. The production-ready versions of these program libraries can then be tagged and locked down for the final near-submission execution run.

Ideally, a clinical data manager would plot a distribution graph of essential patient variables, and the statistician would provide the clinician with decisional support and a wide range of graphical views of safety and efficacy data.
The final objective of any clinical trial is determining the safety and efficacy of the new compound, device or biological under development. The gathered, standardized, cleansed and summarized data is ready to be studied from a clinical perspective. Analysis has both exploration and confirmation phases. In the exploration phase, a clinician should be able to shift quickly from patient population plots and distribution graphs to single patient profiles, while aligning different events and findings on a time axis. The value in providing standard graphs for visualization is the ability to rebuild them with statistical rigor in a validated environment later, should this be necessary. In the confirmation stage, the exploration needs to be rerun by the statistician, and the program code should be easily modifiable to create a scientific analytical report.

Features of an End-to-End Clinical Analytics Environment

The objective of end-to-end clinical data management is to spend less time on operational data management activities and more time on exploring, monitoring data quality and executing advanced analytics and statistics. This clinical framework enables pharmaceutical companies to rapidly assess the effect of a new therapy, as well as follow up on trial progress by creating a trial metrics reporting engine that provides management oversight on financial and performance metrics for all ongoing trials.

As clinical data flows from data capture to analysis, it is frequently subjected to a workflow that enriches the data, unblinds it or allows CROs to add data elements to the complete set of trial data. These workflows involve manual steps or complex processing steps that can use programmatic actions. A good example is the ability to check CRO-delivered data against a library of data specifications or standards as it is being uploaded to a sponsor’s environment. Each step should be controlled, should fit in the business process logic and should make use of automated control mechanisms embedded in the analytical repository.
The technology environment that supports end-to-end clinical data management includes, as a key feature, a clinical repository that:

- Contains actual data extracted on a timely basis from the EDC system.
- Is both secure and centrally accessible over the Internet.
- Delivers data and information to the clinical consumer via a simple, intuitive interface.

This clinical repository should make it easy to run analysis programs, upload new data and make reports widely available to global users. In addition, it should make the data available long after the date of submission – for example, to combine with post-marketing trials or perform comparative outcomes analysis. In summary, it should be a true knowledge hub, not a technological conundrum that hesitates between storing data and making it available for analysis.

Benefits and Predicted Outcomes

Spending less time on operational data activities and more time on analysis of the data is an initial – but not the only – projected outcome of an end-to-end clinical data management system. With the implementation of an end-to-end clinical data management model, time and resources in the clinical development organization are freed up to start using advanced analytics to derive real insight from clinical data. Some examples include:

- Applying predictive sciences, such as modeling and simulation, to better predict which compounds are likely to fail.
• Combining clinical data with data from the clinical trial management system to optimize the conduct of the trial, such as investigator selection, investigator fraud detection, patient recruitment and trial resource optimization.
• Creating a historical warehouse and enabling the execution of cross-trial safety queries.
• Enabling faster FDA data review cycles.
• Enabling traceability between SDTM and ADaM.
• Combining high-volume genomics technology data with patient data to create better patient stratification, aid in the design of the next trial and lead to a biological understanding of the effect of the drug or device.
• Visualizing clinical data before submission using integrated safety and summary analysis.

Closing Thoughts

The pharmaceutical industry today is redefining clinical data operations, data management and submission-driven activities, including collaborating with external organizations. Many pharmaceutical companies are responding to this new environment by setting up collaboration models with external clinical data management partners based on strong data standards governance processes. An end-to-end clinical data management approach can underpin this strategy by reusing integrated standards over the different components of the clinical data flow, including data capture, creating data tabulations and final analysis of clinical data. Each of the different modules can then be outsourced or split up along departmental lines according to the business processes in place, the objectives of the company and the therapeutic areas under investigation.

End-to-end data management using advanced data standards management, along with an integrated analytics and advanced in-memory visualization environment, is a revolutionary step toward supporting clinical trials analysis over the next decades of clinical research. Bringing together granular, well-documented clinical data will lead to shorter trials, will improve the modeling of new trials before the therapy is tested on the patient, and will lead to higher-quality drug submissions to authorities. This, in turn, will support the industry’s drive to consistently explore new domains and provide therapies for the unmet medical needs of smaller patient subgroups in a specialist market, while getting them to market faster. ²

After all, clinical data analysis should not be restricted by software limitations or huge data volumes; rather, it should be stimulated by the innovative drug models that researchers and clinicians can devise to improve the patient’s health.

About SAS

SAS is the leader in business analytics software and services, and the largest independent vendor in the business intelligence market. Through innovative solutions, SAS helps customers at more than 55,000 sites improve performance and deliver value by making better decisions faster. Since 1976, SAS has been giving customers around the world THE POWER TO KNOW®. For more information on SAS® Business Analytics software and services, visit sas.com.