Cytel
Shaping the Future
of Drug Development

Data Management Fundamentals for Your Next Clinical Trial

TABLE OF CONTENTS

SUMMARY HIGHLIGHTS

Data is the most crucial asset in any clinical trial and is used to ultimately drive the decision-making process related to the development candidate. Therefore, for any sponsor, paying close attention to the data management aspects of clinical operations should be paramount. The principles of data management are simple and well-founded. However, the application of these principles needs careful consideration, depending on various scenarios and the size of the organization. When implementing data management for your trial, it is critical to plan ahead and fully understand all the steps and activities involved.

Fortunately, both strategic and tactical opportunities are available to help sponsors successfully implement a data management strategy, and ensure quality and simplicity in data collection to enable subsequent analysis. In this ebook, we outline some considerations to help sponsors navigate key decisions that need to be made throughout data management implementation.

We discuss the importance of establishing clear data standards for both operational efficiency and regulatory purposes; the practicalities of selecting an electronic data capture system; how to ensure your data is processed quickly and to a high level of quality; and managing a close alignment between data management and statistics functions.

by Paul Fardy

About Paul Fardy

Paul Fardy is Executive Director of Data Management at Cytel and leads the company's clinical data management operations in the USA, Europe, and India. Prior to joining Cytel, he has held leadership positions within CROs and pharmaceutical organizations including Eisai, Ipsen, and GlaxoSmithKline.

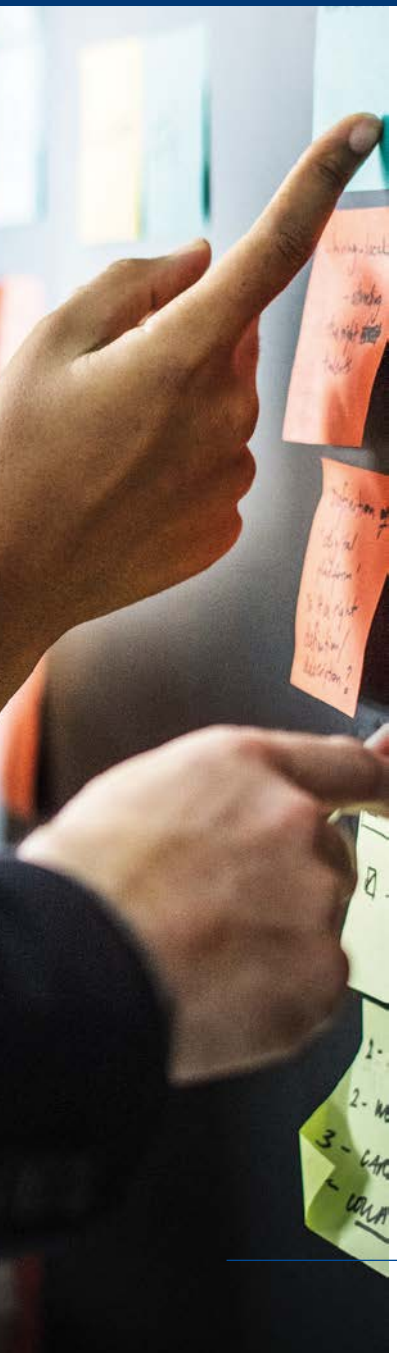
SET UP CLINICAL DATA STANDARDS AS THE FOUNDATION OF SUCCESS

Cytel

In clinical data collection and analysis, traceability is the process in which we lay a clear path to ensure that the results we have obtained can be reproduced. Essentially, traceability is a two-way process. A regulatory reviewer wants to be able to trace a clear path from the analysis results to the ADaM, to the SDTM, and back to the Case Report Form (CRF). Conversely, during up-front planning within the trial, we want to create a clear and traceable path forward from the CRF, then to the SDTM and ADaM, and then on to the Analysis results.

> "The function of CDASH is to define the naming conventions for the clinical database and to outline how variables are mapped to SDTM."

Clinical Data Acquisition Standards Harmonization (CDASH), developed with participation from all three ICH regions (US, Europe, and Japan), provides recommended data collection fields for 16 domains, including DEMOG and AE. CDASH also includes implementation guidelines, best practice recommendations, and regulatory references. Although it is sometimes thought that CDASH defines the layout of the CRF and eCRF, this is a misconception. The function of CDASH is to define the naming conventions for the clinical database and to outline how variables are mapped to SDTM. It defines how questions should be formulated for data collection within the CRF and eCRF through the use of standard CDISC-controlled terminology.



One key benefit of implementing CDASH standards is improvement in the quality of outputs. With the development of more structured information during the study setup, the design of eCRFs can become easier owing to the ability to reuse and retrieve information. The time for tasks can be considerably reduced, thus resulting in a potentially beneficial effect on timelines. Beyond the immediate benefits during study setup, data review is improved through the use of standards and data transfers are simplified. A standard approach across studies serves to support subsequent data analysis. Importantly, from a regulatory submission point of view, CDASH represents a starting point to ensure that we begin thinking about standards at the outset of the clinical data collection and analysis data flow and it supports traceability

At Cytel, we have implemented CDASH by creating a library of standard CDASH forms within our EDC systems. The library serves as a robust starting point when setting up a new study and is flexible enough to accommodate customizations such as specific sponsor preferences by incorporating a range of versions of the same form (e.g., for laboratory data).

By using the standard library, we are able to streamline the process from eCRF design, through data collection, and to SDTM dataset creation. Through this standardized approach, once data is extracted from the EDC system, it can be efficiently migrated to SDTM by using a standard set of transformation scripts. The migration is mainly driven by metadata, allowing us to streamline the generation of submission-ready documentation such as define.xml and Reviewer's Guide, as well as the SDTM datasets themselves.

67%

of CDASH version 2.0 maps directly to SDTM variables

A recent poster presentation (Howard and Labout, 2017)[1] noted that 67% of CDASH version 2.0 maps directly to SDTM variables (CDASH version 2.0 includes mapping), and 86% of CDASH maps directly with standard mapping included.

DETERMINE YOUR EDC STRATEGY

Cytel

We know that the options in the EDC world are many and varied, with the opportunity to spend a lot of money and not get exactly what you wanted or expected. In this environment, sponsors expect their internal and external data management teams to be equipped to understand the different options and to be prepared to evaluate them according to the needs of all parties.

When selecting an EDC system (or any other clinical technology for that matter), it is important to set up an evaluation criteria matrix in which the key factors relevant to the specific circumstances of the trial are listed. These factors will not only be system-specific but also take into consideration the company setup, stability (both financially and in terms of time, i.e., in the mid- to long-term), and other factors. For each EDC system under review, a score should be attributed to each category so that an overall final score may be calculated. It is sometimes useful to have a weighted scoring system that allows the most important factors to have a greater influence on the overall score. After all the systems have been evaluated, a decision can be made about the most appropriate one to adopt.

Cytel

Evaluation Criteria

Pricing Models
Subscriptions
Licensing
Pay-As-You-Go

Operational Features
Interim Analyses
Reporting
Data Standardization

Technical/Help Desk Support
Modifications
Response Time
Help Desk

Pricing Models:

Many different pricing models are available, with familiar options such as Licensing, Subscriptions, Pay-As-You-Go, and Pay-Per-Protocol.

Operational Features:

We need to evaluate whether the EDC system satisfies the requirements of the protocol under consideration.
For example,

- Does the system support interim analyses?
- What reporting functionality is available?
- What is the interface for setting up the database and how well does the EDC library handle reusable components for data standardization?

Technical/Help Desk Support:

This is an important area to research from both the development as well as the end-user perspectives. Is help desk support available in all the regions where your trial will be running? How are modifications made to the product? What is the average time to obtain a response from the support team?

Different Stakeholder Needs

In addition to the above points, we must remember that different stakeholders have different priorities and pain points.

What Does the Clinical Site Want?

User-friendliness is of vital importance to clinical sites. A clinical site may be using multiple systems and would need to be able to address the data requirements around seeing their patients. If the EDC system becomes a barrier to their progress, it will be hard to get them to stay up-to-date with their entry.

What Does the Clinical Data Manager Want?

- User-friendly, easy-to-develop systems
- Enough functionality in a system (but, more than enough is not necessarily better)
- An EDC system that is easy to train on and intuitive to sites and clinical monitors
- Robust reporting in standard reports as well as the ability to easily create custom reports
- Systems that allow for easy integration of changes while maintaining the integrity of earlier versions

What Does the Biostatistician Want?

- Reliable data extracts in the desired formats
- Robust reporting
- Data extracts that have not been adversely affected by protocol change
- Incorporate Technologies Beyond the EDC System

The EDC system may be the lynchpin of data collection for the clinical trial, but many platforms go beyond the capture of CRF data. It is important to consider how important an integrated approach is to your study. An up-front discussion is invaluable in discovering whether features such as integrated ePRO, IVRS, or coding tools are key factors in the protocol's success. The costs involved in integration may be well spent when compared with the cost and process of additional manual reconciliations.

Electronic Patient-Reported Outcomes (ePro) and eSource Considerations

As the scope and scale of technology increase, so do the opportunities to use it in many different ways. The utilization of electronic patient-reported outcomes has led to more accurate, reliable data being collected, and is used to demonstrate both safety and efficacy. Various reminder options allow the subject to input the required information at the scheduled time, while built-in audit trails confirm whether this has happened at the required time. More direct input is achieved through eSource, whereby the data is captured directly from the subject to the database.

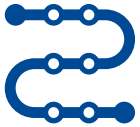
Patient-centricity is now an extremely important consideration in clinical development, and so we must also ensure that data capture devices suit the needs of patients. For example, tablets may be unsuitable for patients with conditions that can affect dexterity.

The right technology should be used for a specific set of patient needs and should be set up so patients and investigators can use it effectively.

PROACTIVELY PROJECT MANAGE THROUGH TO DATABASE LOCK

Database lock is a significant milestone in the clinical trial, on which further data analysis and reporting timelines depend. At Cytel, we have defined 6 key steps to guide project data management activities and ensure that the database is locked on time, every time.

1. Planning Steps



2. Creating and Managing a Realistic Timeline

3. Obtaining Timely Sign-off From Principal Investigators (PIs)

4. Engagement of Sites

5. Involving the Biostatistics Team

6. Tackling Cohorts of Data

1. Planning Steps

As with many aspects of clinical data management, and indeed clinical trials management in general, meticulous up-front planning is a must. Several basic guidelines must be observed and adhered to throughout the study as a matter of course:

- Creation of a solid database by ensuring proper data is being collected
- Creation of solid data quality checks, ensuring they are written in a meaningful fashion
- Creation of a comprehensive Data Management Plan, obtaining input from key stakeholders
- Cleaning the data by addressing issues and queries on an ongoing basis
- Engaging on a regular basis with all stakeholders - clinical sites, biostatistics departments, sponsor, and clinical operations.

2. Creating and Managing a Realistic Timeline

A realistic timeline needs to be created to ensure that the tasks involved are properly sequenced and concluded. With the database lock relying on the input of a number of stakeholders, the timeline needs to include specific tasks and deadlines that are addressed to the specific groups involved. At Cytel, we would typically create a database close-out checklist that incorporates all these items. Input to and sign-off to the checklist and timeline should be obtained from the clinical operations group and the sponsor.

At this point, the main task involves efficiently collecting and cleaning all the data that belongs in the electronic data capture system and also conducting any reconciliation of data, such as laboratory data, that has been collected outside the EDC system.

3. Obtaining Timely Sign-off From Principal Investigators (PIs)

Getting PI sign-off is a critical step and also a potential delaying factor in achieving database lock on time. While PIs are trained in the appropriate EDC at the outset of the study, typically, they do not work regularly within the EDC system. This time lapse between the initial training and sign-off/study close-out may mean that familiarity with the system is lost. This issue can be addressed by allocating sufficient time in the timeline to support the PIs in this activity.

4. Engagement of Sites

One potential issue can arise when a clinical site recruits all its patients at the beginning of the study. If the study continues for an extended time and the site has already completed their patient enrolment, the site may have little further interaction; by the time database lock occurs, the site may have become somewhat disengaged from the study. Generally, and especially in such situations, it is advisable to address any data issues as early as possible while the site is still engaged. Keeping sites engaged and keeping up with data issues will ensure a much smoother process in the final stages before database lock.

5. Involving the Biostatistics Team

The interaction between data management and biostatistics is important in ensuring a streamlined and efficient approach. This is also true for all processes leading up to and including database lock. It is important to ensure that the biostatistics group has been involved in creating the data cleaning plan as well as in addressing any data issues identified. By working together, the clinical data management team can make sure that the requirements of the biostatistics group are addressed. We will discuss this important point in more detail on page 18.

6. Tackling Cohorts of Data

One way in which database lock can be smoothed is to create cohorts of data - perhaps based on sets of subjects who were enrolled or randomized by a certain date. The clinical monitors can be provided with sets of data to be addressed, and once done, the team can move on to the next cohort of data. This is an efficient way to handle large amounts of data and streamlines the process at the end.

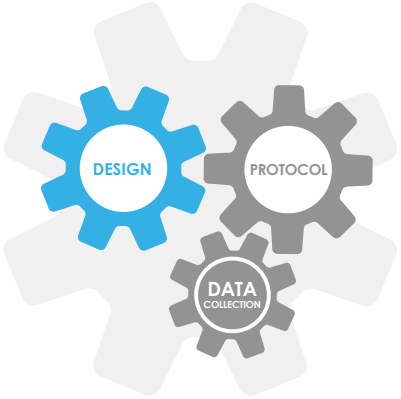
OVERCOME FUNCTIONAL SILOS

When paper studies were the norm, a strong functional alignment was required between clinical research associates and data managers. The increased use of EDC has transformed these operational interactions, allowing the site staff to concentrate on entering the data, while the data manager investigates the data on an ongoing basis and relays back to site staff any issues with quality. In many respects, in clinical research today, the interaction between clinical research associates and clinical data managers operates by default, in line with the ongoing activities of the clinical trial.

Therefore, it is often more important to ensure that the clinical trial operations process is designed to support a close interaction between the data manager and the biostatistics and programming team; this working relationship is instrumental in ensuring data quality and avoiding duplication of work. Unfortunately, operational silos within some organizations can represent a barrier to this collaboration, leading to individual functions losing visibility over the "big picture" of the data collection and analysis process; this may prevent risks, issues, and trends from being identified.

In particular, as centralized statistical monitoring techniques become more widely used, the data manager and biostatistician work closely together to apply these, and ensure that the process of data cleaning is more efficient and that valuable clinical resources are used more effectively. Any data anomalies that emerge throughout the study can be described to the biostatistician (without the risk of unblinding any of the team members) and so, can provide greater insights when preparing the statistical analysis.

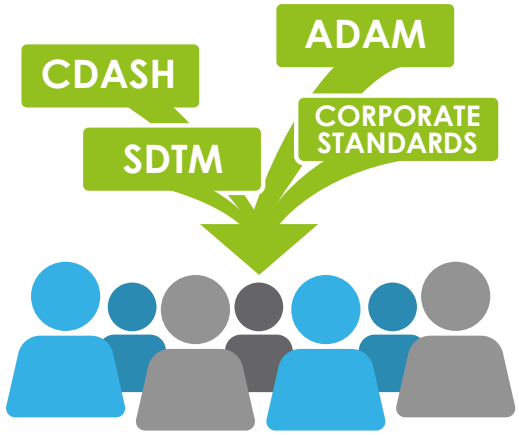
If data is collected and cleaned without the input of biostatistics, the assumptions made may not be adequate, thus resulting in additional work and compromised timelines. Cytel has identified 5 important interactions between data management and biostatistics; these interactions should be accounted for in operational implementation during the course of a clinical trial.



1 Study and Protocol Design

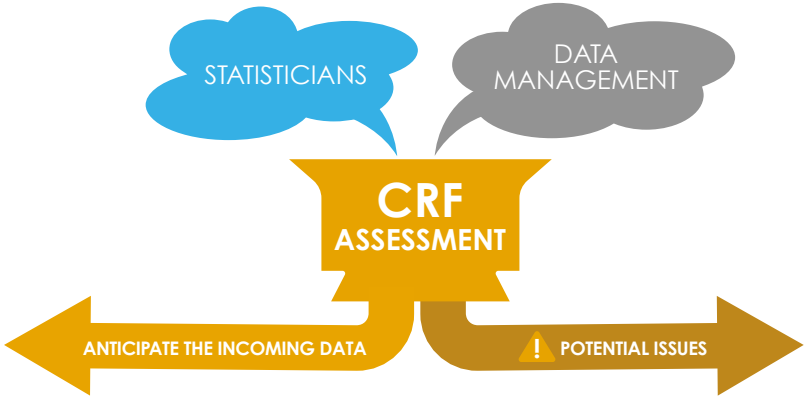
While biostatisticians and other stakeholders are designing the study, data management can add value by reviewing the protocol drafts and assisting with the table of events.

Additionally, data managers have a key role to play in providing advice about what is possible in terms of data collection. A seasoned data manager is tuned in to what works from the sites' perspective and can make recommendations to help ensure a protocol is successful.



2 Standards

We discussed the importance of standards earlier, and an integrated data management and biostatistics team should use standards developed and reviewed by an interdisciplinary group. During standards development, the best possible input has been collected and applied. When data management and biostatistics collaborate, the workflow from CDASH, SDTM, and ADaM, as well as corporate standards are all implemented, resulting in a seamless and efficient process.

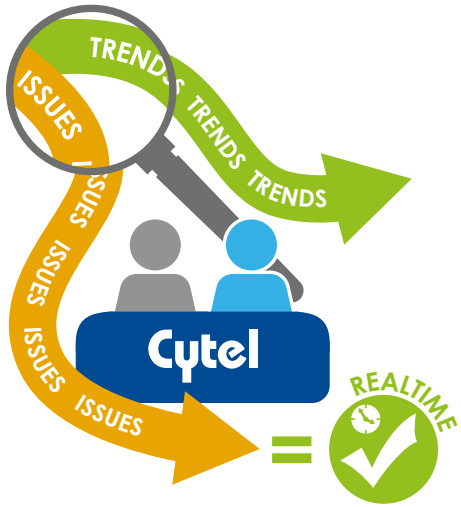


3 CRF Design and CRF Assessment

The biostatistics group should be involved early on in the review of the CRF and should be a part of the sign-off process. On completion of the CRF, an assessment is required to ensure all critical variables are being collected at the appropriate time points. A side-by-side review involving both data management and biostatistics will enable each group to identify any potential issues with CRF design and will help both teams better anticipate the data that will be entered. The statistical analysis plan should contain a number of direct data handling rules, including any derived variables and the algorithms and methods for handling missing data.

4 CRF Guidelines

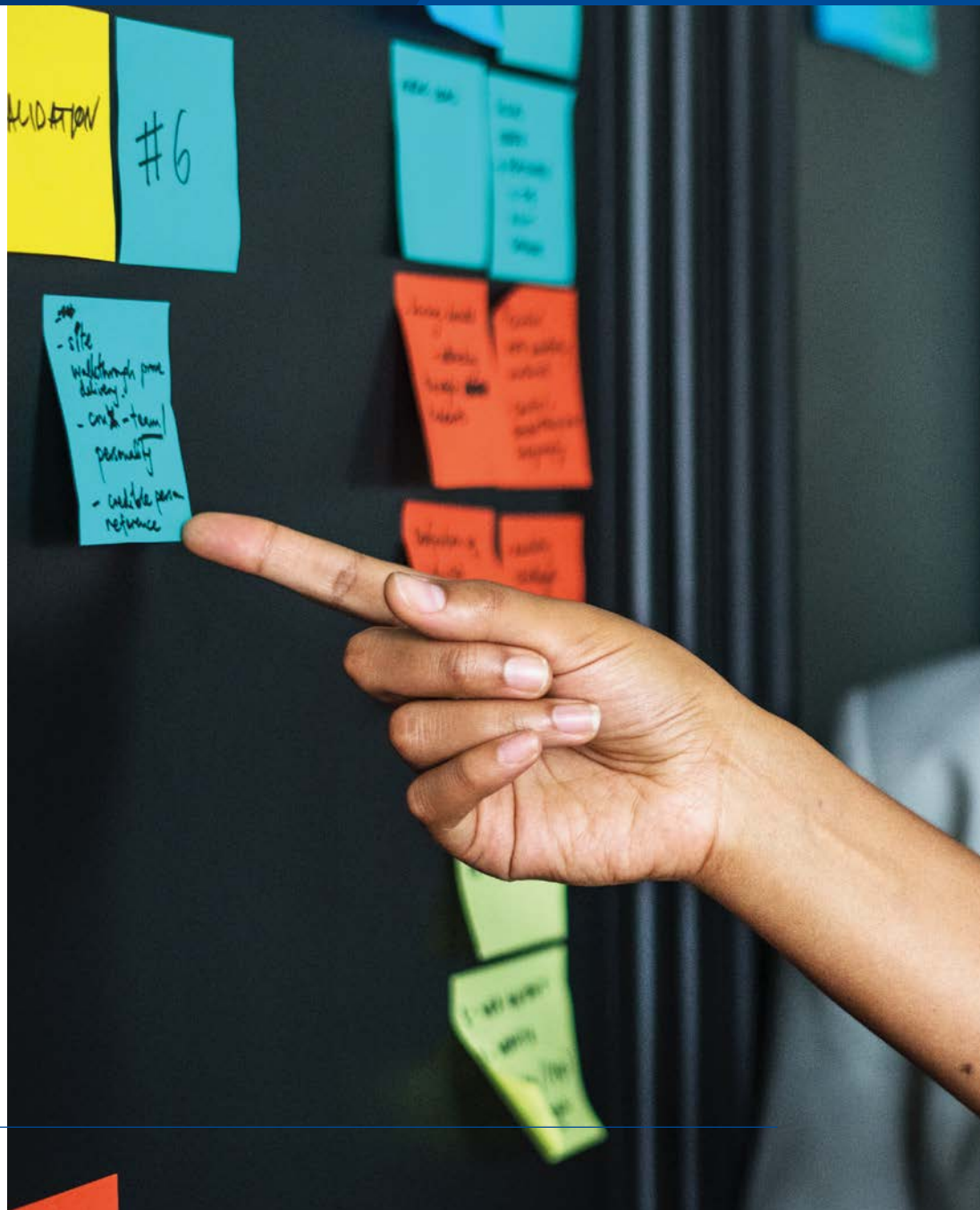
Once the CRF is finalized, the data management team develops case report form guidelines and a data cleaning plan. The biostatistics group needs to be involved in the review and any decisions about these documents. It is efficient to have both groups on the same page regarding the state of the data when it is ready for analysis because the two teams would have set the expectations jointly.



5 Study Conduct

While biostatisticians and other stakeholders are designing the study, data management can add value by reviewing the protocol drafts and assisting with the table of events.

Additionally, data managers have a key role to play in providing advice about what is possible in terms of data collection. A seasoned data manager is tuned in to what works from the sites' perspective and can make recommendations to help ensure a protocol is successful.



CONCLUSION

It is often stated that running a clinical trial is a complex process, and when viewed holistically, this is undoubtedly true. However, many of the individual activities that underpin the process of a clinical trial are actually relatively simple when properly handled. A well-defined approach to the management of your clinical data ensures that your most valuable asset is successfully transformed into information that is able to drive robust drug development decision-making.

References

1) Howard, K. and Labout, S. (2017). SDTM vs CDASH: Why You Need Both! [Poster] PhUSE 2017. [online] [Accessed 23 Sep. 2018] Available at: www.cdisc.org/system/files/all/CDISC_SDTM_and_CDASH_You_Need_Both.pdf

About Cytel

Cytel is shaping the future of drug development. Our solutions bring together technology, data science, data analytics, and clinical research services to drive superior outcomes and results for clinical trial sponsors. With operations across North America, Europe, and India, Cytel employs 900 professionals, with strong talent in biostatistics, programming, and data management.

Our global data management team members have an average of 10 years' experience each, and we provide seamless biometrics outsourcing services working in close collaboration with Cytel's biostatistics and statistical programming teams.

For more information about Cytel visit:

www.cytel.com

GLOSSSARY

ADaM (Analysis Data Model)
defines dataset and metadata standards that support efficient generation, replication, and review of clinical trial statistical analyses, and traceability among analysis results, analysis data, and data represented in the Study Data Tabulation Model (SDTM).It is one of the required standards for data submission to FDA (U.S.) and PMDA (Japan).

CDASH (Clinical Data Acquisition Standards Harmonization)
establishes a standard way to collect data in a similar way across studies and sponsors so that data collection formats and structures provide clear traceability of submission data into the Study Data Tabulation Model (SDTM), delivering more transparency to regulators and others who conduct data review.

CDISC (Clinical Data Interchange Standards Consortium)
is an open, multidisciplinary, neutral, 501(c)(3) non-profit standards developing organization (SDO) that has been working through productive, consensus-based collaborative teams, since its formation in 1997, to develop global standards and innovations to streamline medical research and ensure a link with healthcare. The CDISC mission is "to develop and support global, platform-independent data standards that enable information system interoperability to improve medical research and related areas of healthcare."

SDTM (Study Data Tabulation Model)
provides a standard for organizing and formatting data to streamline processes in collection, management, analysis, and reporting. It is one of the required standards for data submission to FDA (U.S.) and PMDA (Japan).

Reference: www.cdisc.org/standards/foundational

www.cytel.com

Cytel