Harnessing the power of clinical data
Harnessing the power of clinical data

Managing the data burden

Life sciences companies are rich in data and gaining exponentially more clinical data all the time. New scientific breakthroughs, the growth in research partnerships and digitization that brings new insights and new ways of gathering information — all feed the data expansion. This trend, along with the need to improve the review process and safeguard patients, has led to the development of the Clinical Data Interchange Standards Consortium (CDISC) open data standards.

Standardization can improve the quality of medical research by making it simpler to transmit and share complex data, minimizing resource-intensive programming, and making the data submitted to regulatory authorities consistent. Data inconsistency remains a huge challenge for companies, resulting in numerous questions from reviewers, who must clarify submissions and ensure that products on the market are proven safe and effective.

For some years, regulatory authorities have urged companies to use data standards in their submissions. Then, in December 2014, the U.S. Food and Drug Administration (FDA) published binding guidance on study data standards, making the CDISC standards mandatory as of December 17, 2016. Also, starting in October 2016, Japan’s Pharmaceuticals and Medical Devices Agency (PMDA) is requiring companies to submit in CDISC format.

While organizations must make responding to these mandates a priority, they must also understand the broader business value of standardization. Standardization helps improve operational productivity, makes data sharing seamless internally and externally, ensures cross-functional consistency, and enables organizations to effectively leverage big data.
Biometrics: turning data into insights

The ability to process clinical data in a standardized format is just the start. Companies also need to be able to analyze study data through various statistical methods to determine the impact of a product on an individual. This is referred to as biometrics.

Data is collected in various ways: at clinical investigative centers, per the clinical trial protocol and through case report forms (CRFs). Information in the CRFs, including adverse events, comorbidities, medical history, labs, etc., is gathered and coded using medical and drug dictionary codes, such as those found in the Medical Dictionary for Regulatory Activities (MedDRA) and the World Health Organization (WHO) Medical Dictionary. The data must then be analyzed to determine the impact of a drug or biologic on a subject in terms of safety and efficacy.

This is a complex set of activities that requires data experts focused on cleaning, articulating, standardizing, summarizing and analyzing data. For regulatory departments, these activities are required to prepare integrated versions of efficacy and safety reports of drug activity. Trial data summaries are presented in a reviewer’s guide to assist in the review process. Every piece of data has to be built into the submission to support the analysis of the product, and the data needs to be clearly and consistently formatted with thorough explanations and within the limits of regulatory guidelines.

In the age of personalized medicine, the ability of a company to provide data insights and to demonstrate the value of its product is critical, and biometrics plays an integral role in winning product approval and providing true differentiation.

CDISC proficiency: value beyond compliance

While clinical and regulatory leads will generally have some experience with CDISC and biometrics, for organizations to achieve compliance and follow all the rules and processes, they must take a systematic approach and commit to the spirit of CDISC. They need clear, consistent and comprehensive datasets to improve the clinical research process, reduce the programming time, enhance the review cycle and ultimately speed the process of bringing products to market.

The CDISC mandate has far-reaching implications for companies. For example, it sets standard database specifications for dataset development and a submission data definition specification for the development of define.XML. Datasets need to be described and submitted in accordance with the guidelines.

Just as important, the entire process — from entering and processing the data to standardizing and summarizing it — needs to be understood so that companies can respond to questions from regulators.

CDISC standards ease submission for oncology drug

An immuno-oncology drug for a prominent sponsor received six approvals from the FDA after it was submitted using standardized clinical data. The use of Clinical Data Interchange Standards Consortium (CDISC) standards for the submission ensured rational and objective explanations of the safety and efficacy findings to the satisfaction of the regulatory reviewers.

This example demonstrates the important role CDISC clinical standards play in not only reducing drug development costs (a CDISC survey found an 18 percent reduction in the cost of data transfer through standardization) but also in improving the sponsor’s confidence levels with consistent, comprehensive, transparent and traceable data analysis.

Standardization helps improve operational productivity, makes data sharing seamless internally and externally, ensures cross-functional consistency, and enables organizations to effectively leverage big data.

The ramifications of noncompliance with CDISC standards are extensive. Being unable to quickly respond to queries from regulators about data inconsistencies and deficiencies will require companies to commit additional resources to investigate those issues. This could, in turn, result in losing market opportunities to competitors or might, potentially, yield a refusal to file (RTF) or loss of marketing authorization.

While nonconformity can have negative implications, the reverse is also true. For example, more recently, several immuno-oncology therapy applications have submitted standardized datasets and realized fast approval times from the FDA, compared with products that had gaps in data rationalization and explanation. According to recent surveys conducted by CDISC and other industry-based organizations, CDISC standards improve data exchange by 52 percent, study efficiency by 46 percent, data quality by 41 percent and — perhaps most crucial for the industry — speed of approval by 32 percent.\(^2\)

While many factors contribute to vastly improved rates of FDA approval for new drugs — Forbes noted in 2015 that the approval rate for new drugs had reached 96 percent\(^3\) — standardization certainly helps the process.

DXC Technology Life Sciences recommends that organizations take a proactive approach to CDISC standards development from the point of gathering raw clinical data. We recommend that they build partnerships with contract research organizations that have been involved in clinical trial assignments involving raw clinical data and partners that have a deeper understanding of the broad business needs and objectives in a complex regulatory environment.

CDISC compliance will require a multifaceted approach that unites science, software, intelligence and automation, cloud, digital, and big data analytics tools, along with experts in biometrics and CDISC.

**CDISC standards: consider the advantages**

- **Improve**
  - Data Exchange (52%)
  - Study Efficiency (46%)
  - Data Quality (41%)
  - Speed of Approval (32%)

- **Reduce**
  - Burden of Regulatory Submissions (38%)
  - Time Spent Mapping Legacy Data (31%)
  - Cost of Data Transfer (18%)

Clinical Research Organizations (CROs) and Technology Service Organizations (TSOs) are consistently adopting the standards more readily than life sciences and pharmaceutical companies.

---

1. Ibid.
Steps to CDISC excellence

Several factors contribute to CDISC excellence:

- Mapping the data
- Regulatory insights
- Statistical programming
- Statistical analysis
- Quality assurance/control

Room for interpretation

A common challenge arises with mapping unscheduled visits and unscheduled assessments. Since the Study Data Tabulation Model (SDTM) allows for interpretation on how to deal with these situations, an assessment must be made based on the given situation and how the company wants to represent the data.

For example, a similar situation arose at two pharmaceutical companies with regard to mapping unscheduled lab tests during a scheduled visit. In one case, the sponsor wanted the visit to show as scheduled. This required SDTM insight to find a variable that would allow the visit to show as scheduled while the assessment was defined as unscheduled. In the second case, based on the needs of the client with regard to capturing the data, the visit was defined as unscheduled.

Mapping the data

Life sciences biometrics teams need to start with an understanding of the specific needs for the company and the submission. Existing data needs to be assessed and articulated in line with CDISC standards and guidelines through specifications and programming methods.

Then organizations must determine where the gaps lie. For example, how is data managed for proposed patient visits, assessments and interventions that are planned in the protocol versus unplanned visits and data deviations? These deviations might arise from collection at the source level; subject screen failures; lack of adverse event notifications or reconciliation of serious adverse events; and incomplete assessment of laboratory, drug accountability and pharmacokinetic data.

While mapping data, it’s important to take into consideration the way reviewers use CDISC to understand applications. A common mapping challenge for companies is that the case report forms (CRFs) or electronic data capture systems often don’t allow organizations to collect data in Study Data Tabulation Model (SDTM) format. Changing the values on the data to ensure they are SDTM-compliant at the point of collection — through the Clinical Data Acquisition Standards Harmonization (CDASH) data standard — can alleviate those inconsistencies and reduce the mapping time.

An example would be to have controlled terminology presented in CDISC format. Managing the mapping at the front end will improve traceability by ensuring that SDTM data submitted to the regulators can be linked to the raw data. In fact, a CDISC survey found CDISC standardization reduces time spent mapping legacy data to analysis data by 31 percent.*

A future outlook: CDISC in 2020

CDISC Standards

- Fully harmonized, from protocol through data collection, analysis and reporting/submission, including controlled terminology
- Augmented to facilitate data sharing for all major disease areas to fully support global clinical research studies
- Used for all types of clinical research (academic, regulated, outcomes, public health) worldwide

CDISC Healthcare Link initiative will support a future where the clinical researcher will take full advantage of data from the EHR, devices, wearables and registries; data will be created “research-ready” by designing in CDISC standards from the beginning; seamless flow of data among researchers and practitioners will become the foundation for a learning healthcare system.

Stay current

A pharmaceutical company recently prepared SDTM mapping specifications for multiple studies that included several questionnaires. Each questionnaire, containing hundreds of questions, was mapped individually, and the process was quite time-consuming.

Upon review by DXC, it was found that the company was unaware of the CDISC standards in place for many industry-standard questionnaires. DXC corrected the original mapping effort to fit into the CDISC standard, which caused rework, as well as a loss of revenue from the original work done by the sponsor.

Ultimately, DXC was able to remap all the questionnaire data, which allowed it to pass validation and make the datasets submission-ready.

Regulatory insights

While CDISC standardization has benefits for both regulators and the industry, a lack of understanding or expertise with CDISC creates problems for companies. Furthermore, updates to CDISC standards, as well as new regulatory developments, affect CDISC and other requirements for analyzing study data. The standards are constantly changing as CDISC works with regulatory and industry organizations worldwide to improve.

Achieving CDISC excellence means staying abreast of developments. While mapping challenges can be addressed on the front end — starting with protocol standardization and achieving Analysis Data Model (ADaM) datasets — changes will also need to be consistent with the most recent standards developments. For example, agencies are currently keenly focused on standardizing oncology data, which is one of the largest and arguably most prominent therapeutic areas in clinical research. They are fine-tuning the standards with terms, terminologies, techniques, measurements and assessments.

To stay ahead, the biometrics team needs to establish a relationship with the CDISC community and stay current on developments from relevant committees. Team members will need to attend conferences, take part in peer discussions and workshops, deliver talks, write articles, read publications and interact with customers, clients and working groups.

Statistical programming

In addition to biometrics and CDISC expertise, companies need to leverage clinical programming capabilities to ensure datasets are compliant with SDTM, Standard for Exchange of Nonclinical Data (SEND) and ADaM.

Often, data needs to be transformed from raw data that stores information in one row (horizontal) to multiple rows (vertical). Controlled terminology requirements also need to be applied to test names, codes and other variables. This requires the assistance of both clinical standard analysts and programmers.
Applying derived variables to clinical data is a very detailed process handled by statistical programmers. A dosing schedule needs to be derived for each patient, and then event data needs to be fed into this schedule to see where it occurred. Determining rules to make the schedule and finding events can be difficult enough, but then rules need to be developed to break ties between the start/end dates of the phase versus the start/end dates of events.

These rules can range from fairly straightforward to extremely complex. A company with a single-dose trial may decide that every tie should fall back to the previous phase, yielding a simple, consistent approach. However, studies that use multiple doses, different drugs, washout phases, etc., could require more complex rules. For instance, a company may want adverse event (AE) data phased differently than lab data. In addition, the tiebreak rule may change throughout the patient’s course in the trial. The event may be moved back or ahead if there is a tie between the study’s drug and washout phases, so the event shows in the study’s drug phase.

Clinical research uses several programming tools, in particular SAS and PL/SQL, to manage the data. SAS requires clear, methodical programming to ensure that each step is followed and the results are clear for quality control and, ultimately, for the regulatory reviewers. Some organizations have business tools for the conversion of clinical data into CDISC-standard format based on PL/SQL.

**Statistical analysis**

With vast amounts of data to manage in a standardized format, organizations must also work to determine what it all means.

The SDTM datasets are used to build analysis specifications in compliance with the ADaM guidelines. ADaM covers, in standard format, baseline flags, indicators of timing, treatment, analysis population, analysis day, and visits of safety and efficacy parameters. Many of these datasets require complex algorithms to derive endpoints of the study and have to be in line with the SAP, which helps statisticians design the ADaM datasets, and with the needs of statisticians and clinical staff. An experienced programmer should coordinate and interpret these complex programming needs.
Phase III studies generally involve about 5,000 to 10,000 people, which can result in millions of records from AEs, labs, patient history, concomitant medications and so forth. By the time companies get to Phase IV studies, the number of subjects can easily top 100,000 with future potential for millions in certain therapeutic areas such as vaccines, diabetes, HCV, gene therapy, etc. The number of records is overwhelming in such large studies. Therefore, it is necessary to deploy big data analytics to manage programming complexities and achieve study insights into patient profiles and AE reports.

Big data analytics holds great promise in clinical data standardization. The tool can be used in personalized or precision medicine, targeted and gene therapy procedures and also in translational sciences, especially as biomarker and biospecimen analysis gains prominence. By using big data analytics procedures to pool data in SDTM from multiple studies, companies can generate new insights and improve the effectiveness of patient treatment. This patient-focused approach will be needed so that companies can meet demands for personalized medicine and better health outcomes.

Quality assurance/control

Before starting the mapping process, companies must put quality assurance (QA) in place to ensure that every step is validated, and a QA person should be designated to oversee each step. Standard operating procedures, macros and new versions of SAS need to be validated every 2 years, so processes must be established to ensure that these updates are properly managed and meet global regulatory guidelines concerning clinical data. QA is therefore both a first and a last step in ensuring CDISC compliance.

Integral to QA is having the full picture of data-standardization insights, along with the broader business goals and needs. This ensures the integrity of information across the product life cycle. CDISC standardization and biometrics require complex processes — from mapping to programming to statistical analysis — plus constant and consistent QA oversight.

Achieving this consistency and data veracity requires the involvement of an integrated team of experts across each process, combined with business and regulatory knowledge that can support the product on the market.

Life sciences companies understand the importance of standardization but often struggle to achieve it because they may lack resources, have the wrong tools and, crucially, do not have the necessary CDISC knowledge and expertise. CDISC standardization needs to begin early in the clinical trial process and have broad application.

At a higher level, CDISC standardization must support the business case, validate the tools used for standards development and guide expertise from inception and submission to approval and beyond. Part and parcel of that expertise is knowledge of the changing regulatory environment and rapidly evolving standards and how these affect specific business objectives and outcomes.
## CDISC Standardization for clinical data

<table>
<thead>
<tr>
<th>SDTM</th>
<th>ADaM</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Standardized, predefined submission metadata and terminology for</td>
<td>• Sourced and derived data to support statistical analysis</td>
</tr>
<tr>
<td>domains and variables</td>
<td>• Replacement for sponsor-defined analysis datasets</td>
</tr>
<tr>
<td>• Standard classes and rules allowing model extensions</td>
<td>• Uses SDTM as source to enable traceability</td>
</tr>
<tr>
<td>• Use of supplemental qualifier data allows for the submission of</td>
<td></td>
</tr>
<tr>
<td>data not standardized by the model</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>SEND</th>
<th>CDASH</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Nonclinical tabulation data equivalent of SDTM</td>
<td>• Standardized basic data collection (CRF) for many domains</td>
</tr>
<tr>
<td>• Standardized domains for preclinical data</td>
<td>• Consensus-based standards with input from three ICH regions</td>
</tr>
<tr>
<td>• Growing trend for FDA to request preclinical data in submission</td>
<td>• SDTM data from the collection point</td>
</tr>
<tr>
<td>packages</td>
<td>• Reduced data mapping</td>
</tr>
</tbody>
</table>

By working with partners or developing internal expertise to become CDISC proficient, companies will ensure consistent, standardized data, which means cleaner, more comprehensive clinical trial data for submissions, the ability to respond rapidly to queries from reviewers, and fewer requests for additional analysis.

For organizations grappling with disparate clinical data challenges, leveraging CDISC is a powerful vehicle for driving business value beyond compliance. An improvement in the submission process can lead to a shortened review time and, ultimately, faster time to market. At the same time, a commitment to CDISC excellence supports the commercial objectives of an organization, leading to improved sharing of information and greater insights through big data analytics that can enhance the performance of the portfolio.

### DXC: The CDISC differentiator

Successfully navigating the complex regulatory process is no easy task. For many companies, it requires partnering with a knowledgeable, experienced and trusted provider. DXC’s team of CDISC and biometrics experts provides support that is scalable, flexible and client-focused, meeting the specific needs of the always-changing regulatory environment.

This deep domain expertise is combined with comprehensive cloud, big data analytics and consulting capabilities, along with industry-leading electronic document management and collaboration solutions, allowing companies to make use of the highest level of expertise across the business.

At DXC Life Sciences, we are committed to helping our clients by leveraging, partnering and investing in a progressive, digital platform approach that drives capital-efficient and agile innovation.

Learn more at www.dxc.technology/life_sciences.

About DXC

DXC Technology (NYSE: DXC) is the world’s leading independent, end-to-end IT services company, helping clients harness the power of innovation to thrive on change. Created by the merger of CSC and the Enterprise Services business of Hewlett Packard Enterprise, DXC Technology serves nearly 6,000 private and public sector clients across 70 countries. The company’s technology independence, global talent and extensive partner alliance combine to deliver powerful next-generation IT services and solutions. DXC Technology is recognized among the best corporate citizens globally. For more information, visit www.dxc.technology.