Biomarker Use Facilitated by CDISC Standards

JOYCE HERNANDEZ, CLINICAL STANDARDS MANAGER AT ELIASSEN GROUP
Introduction

Within the last decade, there have been significant advances in genome sequencing and new technologies that inform the identification and development of biomarkers. It has been known for some time that the “one size fits all” model for prescribing therapy does not benefit all patients. Some people will respond well to a certain medication, whereas others will have a reduced response, if any, and a portion may even experience adverse reactions.

Value of Biomarkers

The value of biomarkers to develop patient groups based on clinical response to treatment has been increasingly accepted. Molecular profiling is especially helpful within the following areas:

- Increased understanding of diseases at their molecular level
- Improved insight on the impact of pharmaceuticals on molecular processes

Personalizing prescribed therapies based upon a patient’s specific set of biomarkers is also known as precision medicine. President Barack Obama recently launched the Presidential Precision Medicine Initiative to collect molecular biomarker and other health data on a large patient population.

Dependencies for Success

To maximize the successful use of biomarkers as part of precision medicine, several dependencies need to be addressed:

- Consistency in the definition of disease states and their related phenotypic results such as risk or degree of medical efficacy for a given therapy.
- Standard representations of complex molecular data in a format that facilitates analysis in combination with clinical data.
- Management of large volumes of data.
- Efficient and accurate qualification processes for biomarkers.
- Availability of supportive data or services such as specialized tests, biological resources (e.g., biobanks), management of samples, etc.
- Flexibility to accommodate advances in technology and processes.
- Data acquisition/management best practices to assist in aligning patients/subjects to biomarkers/expected outcomes/treatment options.
- Partnerships across all interested parties (e.g., patient advocacy groups, healthcare providers, clinical diagnostic laboratories, payers, academia, pharmaceutical/biotech companies, standards organizations, government agencies).
- Consistent approach to performing analysis and reporting.
Overcoming Key Issues

A lot of the issues revolve around data acquisition, integration, quality and consistency. CDISC standards, such as its pharmacogenomics standard, can contribute to overcoming barriers that impede the overall qualification and research processes. Implementation of robust standards can help overcome inconsistency in reporting disease identification/stages, phenotypic results, complex molecular data, analysis and reporting, specimen collection and handling, and clinical data collection.

DISEASE STATE IDENTIFICATION AND DEFINITION (PHENOTYPIC DATA)

One of the critical areas that needs to be addressed is the number of different healthcare and biomedical disease definition standards (e.g., ICD-10, SNOMED CT, OMIM), which can vary in granularity of the disease state definition and identification and in some cases are not harmonized to each other. While many disease classification terminologies exist, they were created for different purposes, and none of them do a complete job. ICD-9 and ICD-10 largely support accurate representation of medical conditions for billing claims. MedDRA, on the other hand, was developed to support reporting of adverse events and not disease, per se. There is also a common issue related to the definition of testing and intervention procedures.

To help overcome these challenges, CDISC has partnered with the National Cancer Institute (NCI) to develop new terminologies where none already exist and harmonize with existing terminologies wherever possible. The recommended path to solve this issue is for all interested parties across government, research, clinical practices and academia to select the most robust “preferred” terminology and to jointly extend it to cover all disease areas of interest at the appropriate granular level.

STANDARDIZED MOLECULAR AND CLINICAL DATA

The ability to combine complex microbiology and molecular data with clinical information in order to perform biomarker qualification and medical research is a key requirement. CDISC recognized this need and invested resources in extending the Study Data Tabulation Model (SDTM) to accommodate molecular data. Version 1.0 of the Study Data Tabulation Model Implementation Guide: Pharmacogenomics/Genetics (SDTMIGPGx v1.0) contains many examples of genetic variation, genotyping and gene expression data formatted using the SDTM model for regulatory submission purposes.

Since sample management is a key area of concern, the model accommodates the collection of biospecimen and handling data. The SDTM model is also used to represent clinical trial data such as laboratory, ECG, adverse events, interventional therapies, medical history, imaging results, pathology for human trials and toxicology results for pre-clinical in tabulation datasets. Recently the model was also enhanced to collect information about medical devices.
MANAGING LARGE REPOSITORIES OF DATA

Some molecular tests tend to generate vast amounts of data. For many trials, only a subset is germane to the disease being studied. The SDTM model does have the capability to represent this subset while still pointing to the data source containing the complete set of data. It also supports reporting variations at the codon level, which includes only a very small snippet of molecular information, in order to reduce dataset size. Codon-level reporting also facilitates associating amino acid changes in proteins, which unlike DNA actually perform miscellaneous activities and functions within a cell, to the underlying genetic variations.

DATA CONTENT STANDARDIZATION

The focus of CDISC standards has been to employ standardized dataset formats and terminologies to ensure consistent data content. CDISC terminologies help to standardize clinical and molecular information in a variety of areas, some of which are listed below:

- Topics used for treatment, testing and classification of events
- Units of measure
- Position of a patient during a test
- Testing methods
- Anatomical locations involved during testing

ADAPTIVE TRIALS

Because of constant innovation of the technologies used for molecular testing, it is important to be able to accommodate adaptive trials. Some studies tend to run for multiple years and would benefit if they were able to take advantage of changes in technology or best clinical practices. These changes may need to be reflected within protocol amendments. CDISC developed the Study Design Model in XML (SDM-XML) in order to represent key aspects of a trial’s design using a machine readable format. This enables an operational system to make use of key protocol components to assist or automate some aspects of trial management. It also enhances the capability of the sponsor or regulatory agency to conduct analysis utilizing the protocol data.

DATA ACQUISITION GUIDELINES

Identifying which information is needed to perform clinical studies can sometimes be a challenge. Fortunately, when you review information typically collected to conduct medical research, it becomes readily apparent there is a significant overlap even across multiple therapeutic areas. The CDISC standards organization recognized this and developed the Clinical Data Acquisition Standards Harmonization (CDASH) as a set of best practices to guide the collection of data. CDASH’s primary focus is the effective design of electronic case report forms (or eCRFS) so that studies can be initiated more rapidly and the downstream production of tables and analysis datasets can be streamlined.
To augment the CDASH core data elements, CDISC has been partnering with the National Cancer Institute (NCI), TransCelerate, FDA, EMA, PMDA, Innovative Medicines Initiative (IMI), World Health Organization (WHO) and Clinical Path Institute (CPATH) to develop standards specific to various therapeutic areas (TA). CDISC TA standards are now available for over 25 diseases including cancer, respiratory diseases such as Asthma and COPD, Diabetes, Parkinson’s disease, Multiple Sclerosis and others. As part of the data collection modules, CDISC also developed a model to collect data directly from laboratories known as the Lab Model.

**ANALYTICAL MODELS AND GUIDELINES**

As researchers, we tend to spend a lot of time collecting data and then performing analyses based on the collected data. As a general convention, analysis datasets (ADs) are created from the collected data and are then used to perform the analyses, which are summarized in tables and figures. The CDISC Study Data Tabulation Model (SDTM) standard is used for the standardization of the collected data and the CDISC Analysis Data Model (ADaM) standard is used for the standardization of the analysis data.

The key features of ADaM include:

- A standard subject-level dataset (ADSL) that contains important variables that are used for multiple analyses. Examples include demographics, treatment assignments, trial disposition, and dates of important events.
- A standard structure, Basic Data Structure (BDS) that contains one or more rows for derived concepts that are analyzed. The analyzed concepts are specific to the individual trial. For example, a trial for HIV might create concepts to analyze the log of a viral load, the time to progression, or the total score of a pain questionnaire.
- A standard structure, the Occurrence Data Structure (OCCDS) that is used for the analyses of events. Examples of OCCDS might include data sets to support the analysis of adverse events, medical history, or concomitant medications.
- Emphasis on creation of full and informative metadata that accompanies the analysis dataset. This metadata describes the logic used to derive variables or observations based on the collected data and is important because it provides the user of the AD with the ability to understand how the collected data was used to create the derived data.

**BASING STANDARDS ON A MEDICAL RESEARCH UML MODEL**

One of the main objectives for CDISC has been to develop standards that can interface easily with healthcare by providing the semantic basis for harmonization between standards. Another goal was to have consistency across the entire medical research discipline to enhance the traceability of data from the inception of a protocol through submission of its data to regulators. It became apparent that a conceptual model representing the entire research process from study design to product approval and pharmacovigilance was needed. CDISC built a conceptual model called Biomedical Research Integrated Domain Group (BRIDG) which supports key processes for medical research:
To augment the CDASH core data elements, CDISC has been partnering with the National Cancer Institute (NCI), TransCelerate, FDA, EMA, PMDA, Innovative Medicines Initiative (IMI), World Health Organization (WHO) and Clinical Path Institute (CPATH) to develop standards specific to various therapeutic areas (TA). CDISC TA standards are now available for over 25 diseases including cancer, respiratory diseases such as Asthma and COPD, Diabetes, Parkinson's disease, Multiple Sclerosis and others. As part of the data collection modules, CDISC also developed a model to collect data directly from laboratories known as the Lab Model.

ANALYTICAL MODELS AND GUIDELINES

As researchers, we tend to spend a lot of time collecting data and then performing analyses based on the collected data. As a general convention, analysis datasets (ADs) are created from the collected data and are then used to perform the analyses, which are summarized in tables and figures. The CDISC Study Data Tabulation Model (SDTM) standard is used for the standardization of the collected data and the CDISC Analysis Data Model (ADaM) standard is used for the standardization of the analysis data.

The key features of ADaM include:

- A standard subject-level dataset (ADSL) that contains important variables that are used for multiple analyses. Examples include demographics, treatment assignments, trial disposition, and dates of important events.
- A standard structure, Basic Data Structure (BDS) that contains one or more rows for derived concepts that are analyzed. The analyzed concepts are specific to the individual trial. For example, a trial for HIV might create concepts to analyze the log of a viral load, the time to progression, or the total score of a pain questionnaire.
- A standard structure, the Occurrence Data Structure (OCCDS) that is used for the analyses of events. Examples of OCCDS might include data sets to support the analysis of adverse events, medical history, or concomitant medications.
- Emphasis on creation of full and informative metadata that accompanies the analysis dataset. This metadata describes the logic used to derive variables or observations based on the collected data and is important because it provides the user of the AD with the ability to understand how the collected data was used to create the derived data.

BASING STANDARDS ON A MEDICAL RESEARCH UML MODEL

One of the main objectives for CDISC has been to develop standards that can interface easily with healthcare by providing the semantic basis for harmonization between standards. Another goal was to have consistency across the entire medical research discipline to enhance the traceability of data from the inception of a protocol through submission of its data to regulators. It became apparent that a conceptual model representing the entire research process from study design to product approval and pharmacovigilance was needed. CDISC built a conceptual model called Biomedical Research Integrated Domain Group (BRIDG) which supports key processes for medical research:

- Study Design and Statistical Analysis Plan
- Data Collection
- Clinical Trial Conduct
- Integration of Clinical, Biospecimen and Molecular data
- Analysis and Reporting
- Electronic Submissions
- Safety and Pharmacovigilance

SUMMARY

The diagram below depicts the key CDISC standards (within the oblong boxes) at a glance and how they map to the typical clinical study processes (located at the top of the diagram). The BRIDG model and Controlled Terminology have a much broader scope, so they span all the processes. The emergence of the Electronic Health Record (EHR) has provided an opportunity to integrate processes from clinical research and healthcare. The CDISC Healthcare Link project seeks to exchange data between EHRs and clinical research systems to reuse healthcare data wherever possible for clinical research and to improve quality and patient safety.

As you can see, CDISC has strived to provide an end-to-end solution which further enhances the data traceability. It provides a rich foundation that can be expanded further to fully support the identification, qualification and use of biomarkers in research and, by extension, healthcare.

Data Traceability for Clinical Trials

- **Design**
  - Protocol Model
  - SDM-XML

- **Data Collection**
  - CDASH LAB
  - ODM-XML

- **Data Tabulation**
  - SDTM, SEND, MDEV
  - Define-XML Dataset-XML

- **Reporting & Analysis**
  - ADaM
  - Define-XML Dataset-XML

BRIDG
Controlled Terminology
CDISC-IHE Healthcare Link
About Eliassen Group

Eliassen Group has been driving the success of our clients, consultants, and employees within Technology Staffing and Consulting Services since 1989. For over 25 years, we have been at the forefront of innovation by uniting talented people and quality solutions within successful companies to help them attain their business objectives. Our solution oriented approach to understanding your needs ensures that we identify and represent the best possible talent available in the market, personally selected and matched to your requirements by our highly-tenured, extraordinarily capable recruiting team. Our consulting divisions include expertise focused on Agile, Big Data, Biometrics & Data Solutions, Creative Services, Government Services, Life Sciences and Workforce Management. Eliassen Group is privately held and headquartered in Wakefield, MA. For more information visit www.eliassen.com.