PharmaSUG 2016 - Paper DS15

Transforming Biomarker Data into an SDTM based Dataset

Kiran Cherukuri, Seattle Genetics, Inc., Bothell, WA

ABSTRACT

Biomarkers play an increasingly important role in drug discovery and development. They are used as a tool for understanding the mechanism of action of a drug, investigating efficacy and toxicity signals at an early stage of pharmaceutical development, and in identifying patients likely to respond to a treatment.

This paper provides an Introduction to the implementation of SDTM standards for data that defines a genetic biomarker and data about genetic observations. The draft CDISC SDTM Pharmacogenomics/Genetics Implementation Guidance will be referenced and rationale for using specific aspects of the draft guidance or suggesting a modification will be explained. The variables used, considerations taken and the process for setting up the pharmacogenomics/ genetics biomarker domains will be described.

INTRODUCTION

Study Data Tabulation Model (SDTM) provides a general framework for describing the organization of data collected during clinical trials. This framework is described in SDTM Implementation Guide (SDTMIG) and includes details about the common types of data for each of the basic structures defined in SDTM. One of the reasons to convert legacy data or data in non-standard format to CDISC SDTM format is to provide Case Report Tabulation (CRT) data to a regulatory agency, such as the FDA. This standardized format used with available software tools will allow efficient access and correct interpretation of the data submitted. This will also help in efficiently preparing for Integrated Summaries of Safety and Efficacy (ISS/ ISE).

Incorporating Pharmacogenomics and pharmacogenetics data that is obtained in clinical trials by testing biological samples collected from the patient and the results of which may have implications for the subject or for a drug is not addressed in the approved SDTMIG. Recently, CDISC PGx team came up with draft guidance SDTMIG for Pharmacogenomics/ Genetics (SDTMIG-PGx) that provides some direction to sponsors.

This paper mainly discusses,

- What a Biomarker is,
- Significance of Biomarkers in clinical trials,
- And, introduces the domains defined in SDTMIG- PGx and describes how the data collected fits into the
 defined domains and what challenges were faced to incorporate certain biomarker data into SDTM format.

What is a Biomarker?

Biological Markers, commonly called Biomarkers are any signs of biological process that can be objectively measured. A biological process can be something that is normally happening in the body or during a development of a disease or in response to drug in a patient undergoing treatment. Biomarkers indicate when and to what extent a physiological or biological process has taken place in the body of a person. Thus, they can be used for both in diagnosis and for monitoring a success of a therapy.

The National Institutes of Health Biomarkers Definitions Working Group defined a biomarker as a 'characteristic that is objectively measured and evaluated as an indicator of normal biological processes, pathogenic processes, or pharmacologic responses to a therapeutic intervention'. World Health Organization (WHO) in coordination with the United Nations and the International Labor Organization, has defined a biomarker as "any substance, structure, or process that can be measured in the body or its products and influence or predict the incidence of outcome or disease". A biomarker can be a gene, molecule, enzyme, specific cell or a protein. Examples of biomarkers include everything from pulse and blood pressure, an antibody, a presence of a certain genetic mutation in the DNA of the cancer cell, electrocardiographic testing, and image information such as computed tomography (CT), magnetic resonance imaging (MRI) and positron-emission tomography (PET).

Significance of Biomarkers in Clinical Trials

Prior to the introduction of biomarkers, treatment choices were traditionally made based on a patient's medical history and pathology. Possible influences on drug response that are usually considered when making treatment decisions or drug adjustments include age, sex, disease, environmental factors, diet, and drug interactions. However, even when these factors are taken into account, drug response often varies among patients, ranging from positive outcomes to

fatal adverse reactions. The reason being, most of the therapies are developed to treat all patients with the same clinical diagnosis; "One size fits all". Many therapies only work in a fraction of the patients for which they are prescribed. Even though these therapies work, a significant progress can be made by getting the right treatment to the right person. With the help of discovering, confirming and clinically validating biomarkers this can be achieved.

In drug discovery and development, biomarkers of all types play a big role; the use of biomarkers in clinical development delivers information on the efficacy and safety of the treatment and ideally indicates which patients benefit from a potential drug. Biomarkers play a critical role in improving the drug development process both in Early and Late phase clinical trials. The purpose of biomarker data in early phase clinical trials is to support proof of concept, dose selection and as go/no-go decision-making tool. In late phase clinical trials, disease biomarkers or pharmacogenomics biomarkers help provide better understanding of inter-patient variability in response.

In oncology, biomarkers are used to provide information of a person's risk of developing cancer and to determine their prognosis once a cancer is diagnosed and, in some cases to predict how a patient responds to a particular medications. In relation to the treatment of cancer, biomarkers fall into two main classes:

- Prognostic Biomarkers: These biomarkers provide information about how the disease is likely to progresses.
 Including how long the patient is expected to live. This determines how aggressively the patient needs to be treated.
- Predictive Biomarkers: These biomarkers help to determine which patients are more appropriate candidates to receive a treatment; this lets the physicians to choose one treatment over another for the patient.
- Pharmacodynamic Biomarkers: These biomarkers could be used as a parameter of the drug activity to demonstrate proof of principle and be used to optimize the dosing schedule of the drug during the earlier of the drug development program.

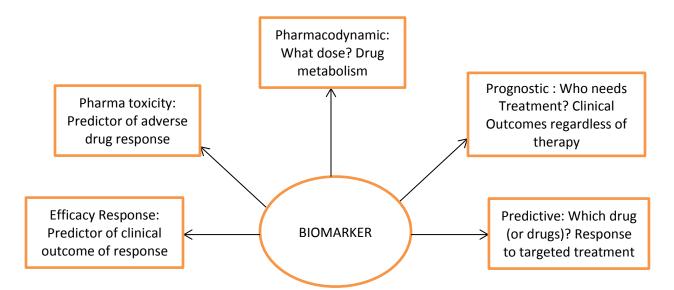


Figure 1: Application of Biomarkers in Clinical trials

STUDY DATA TABULATION MODEL IMPLEMENTATION GUIDE FOR PHARMACOGENOMICS AND PHARMACOGENETICS (SDTMIG-PGx)

Pharmacogenomics and pharmacogenetics both refer to the study of how the particulars of an individual's genetic sequences affect the response of the individual to drugs. The term pharmacogenetics is regarded as the study of genetic variation that gives rise to differing responses to drugs; and the term pharmacogenomics, is the genome-wide analysis of genetic determinants of drug-metabolizing enzymes, receptors, transporters, and targets that influence therapeutic efficacy and safety and drug-related phenotypes.

The purpose of the SDTMIG-PGx is to provide guidance on the implementation of the SDTM for biospecimen- and genetics-related data, including but not limited to: biospecimen collection and handling, genetic mutation, genotyping, gene expression, cytogenetics, viral genetics, and proteomics.

The domains introduced in this document are intended to hold data that fall into one of three general categories: data about biospecimens, data about genetic observations, and data that define a genetic biomarker or assign it to a subject.

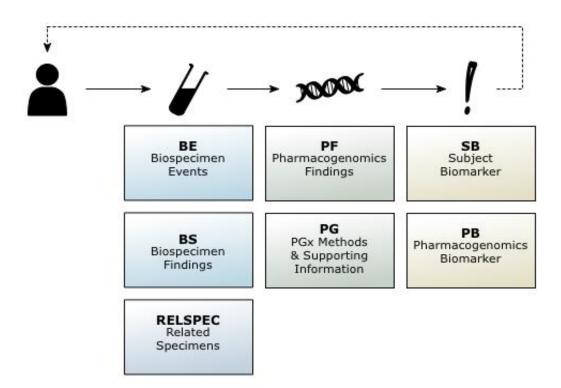


Figure 2: SDTMIG-PGx Domains and Datasets, Study Data Tabulation Model Implementation Guide: Pharmacogenomics/Genetics Version 1.0 (Draft)

BE - Biospecimen Events Domain

- This Events class domain is used to capture information about actions taken that affect a specimen or alter its status. Data may include what the action taken was (e.g., transportation, freezing, thawing), when the action occurred (the date/time associated with it), and who or what party became accountable for the specimen (e.g., site, laboratory).
- The dataset is structured as one record per biospecimen event per specimen collected per subject.
- Key variables are STUDYID, USUBJID, BEREFID, BETERM and BESTDTC

BS – Biospecimen Findings Domain

- This Findings class domain contains the details regarding the characteristics of biospecimens and extracted samples (e.g., RNA, DNA) such as specimen volume, quantity of extracted sample, specimen condition and the integrity of the DNA or RNA samples.
- The dataset is structured as one record per biospecimen finding per specimen collected per subject.
- Key variables in this dataset are STUDYID, USUBJID, BSREFID, BSTESTCD and BSDTC

RELSPEC - Related Specimens Domain

- This dataset holds the hierarchy of specimen relationships. This is a Special-Purpose Dataset and is structured as one record per specimen.
- Key variables in this dataset are STUDYID, USUBJID and REFID.

PF - Pharmacogenomics/Genetics Findings Domain

- This Finding class domain contains the results of genetic variation and gene expression tests. For genetic
 variation tests, the results may include portions of the genetic sequence and comparisons with reference
 gene sequences.
- This dataset is structured as one record per method/ setup observation per specimen collected per date of test per subject.
- Key variables in this dataset are STUDYID, USUBJID, PFREFID, PFTESTCD, PFGENRI, PFREFSEQ, PFNSPCES, PFNSTRN and PFDTC

PG – Pharmacogenomics/Genetics Methods and Supporting Information Domain

- This Findings class domain contains additional data that may affect interpretation of PF data, such as setup processes, test or sequencing parameters, and QC observations.
- This dataset is structured as one record per method/setup observation per specimen collected per run per subject.
- Key variables in this dataset are STUDYID, USUBJID, PGREFID, PGTESTCD, PGGENRI, PGNSPCES, PGNSTRN and PGDTC

SB - Subject Biomarker Domain

- SB domain is a special purpose domain that holds data about any genetic biomarkers (as defined in PB) that a subject may have.
- This dataset is structured as one record per subject per observed biomarker in the study
- Key variables in this dataset are STUDYID, USUBJID, SBREFID, SBMRKRID, SBGENRI, SBGENTYP, SBNSPCES and SBNSTRN

PB - Pharmacogenomics/Genetics Biomarker Domain

- PB domain is a special-purpose domain that documents known associations between observed variations and medical conclusions (e.g., disease diagnosis, resistance of a virus to a particular drug).
- This dataset is structured as one record per biomarker used in the study.
- Key variables in this dataset are STUDYID, USUBJID, PBMRKRID, PBGENRI, PBGENTYP, PBNSPCES, PBNSTRN and PBMRKR.

MAPPING PROCESS

Transforming raw biomarker data files to SDTM format is illustrated in this section.

1) The example here shows data of CD30 mRNA expression and Cell of Origin testing. Gene expression in this study is obtained to detect a disease subtype, so these tests are done only at baseline.

The raw data file shown below contains Assessments of Gene expression for a cell of origin which are collected as a tests performed (LBTEST), and corresponding test Reference ID's (LBREFID), Visit (VISIT), Cell of Origin Classification (COCLASS) and Results (LBORRES) apart from standard variables like STUDYID, USUBJID, etc.

VISIT	LBREFID	COCLASS	CYCLE	DAY	LBTEST	LBORRES
BASELINE	YF122516	GCB	0	1	ACTB	5785
BASELINE	YF122516	GCB	0	1	ANT	77
BASELINE	YF122516	GCB	0	1	CD163	79
BASELINE	YF122516	GCB	0	1	CD30	118
BASELINE	YF122516	GCB	0	1	CD3E	669

Table 1: Raw Biomarker data, example 1

Based on the data collected, Gene Expressions with findings, this data can be mapped to Pharmacogenomics/Genetics Findings (PF) domain mentioned in the SDTMIG-PGx. Table 2 shows the mapped PF SDTM dataset; Some of the variables generated in this dataset; apart from standard variables, like USUBJID, STUDYID, DOMAIN, VISIT, etc. are,

PFTEST/ PFTESTCD (Name of Pharmacogenomics Lab Test): Type of Test Performed

PFCAT (Category for Pharmacogenomics Lab Test): Gene Expression/ Gene Variation.

PFGENRI (Genetic Region of Interest): Cell of Origin Classification, values from COCLASS variable in raw dataset

PFTSDTL (Pharmacogenomics Lab Test Detail): Assessments from Gene Expressions, values from LBTEST variable raw dataset.

PFORRES (Results or Findings): Digital count value, values from LBORRES variable in raw dataset

PFSEQ	VISIT	PFSPID	PFTESTCD	PFTEST	PFCAT	PFGENRI	PFTSDTL	PFORRES	PFSTRESC
1	BASE LINE	YF12251 6	NUC	Nucleotide	GENE EXPRES SION	GCB	ACTB	5785	5785
2	BASE LINE	YF12251 6	NUC	Nucleotide	GENE EXPRES SION	GCB	ANT	77	77
3	BASE LINE	YF12251 6	NUC	Nucleotide	GENE EXPRES SION	GCB	CD163	79	79
4	BASE LINE	YF12251 6	NUC	Nucleotide	GENE EXPRES SION	GCB	CD30	118	118
5	BASE LINE	YF12251 6	NUC	Nucleotide	GENE EXPRES SION	GCB	CD3E	669	669

Table 2: SDTM PF Domain, generated from Table 1 raw data file.

There were a couple of challenges faced while setting up this dataset.

- One of the main assumptions of PF domain is PFTESTCD and PFTEST should not include gene names and symbols. Whether collecting the complete detail for a variation or mutation or a subset, the variable PFGENRI will be used to collect the gene of interest. LBTEST variable in the raw data, in table 1, carries the assessments drawn from the gene expression; so this variable cannot be mapped to PFTEST (Name of Pharmacogenomics Lab Test) variable in PF domain. Instead the values in LBTEST variable should be mapped to PFTSTDL.
- Similarly, before new variables were added to the Implementation guide, COCLASS (Cell of Origin Classification) variable cannot be directly mapped to any variable in PF domain; the new variable PFGENRI can be used to capture these values.
- 2) The example here describes the data collected performing flow cytometry procedure for CD33 expression, ADC binding and target saturation and DNA damage. Several tests are performed under this method to determine and discover biomarkers for an early phase clinical trial.

Since SDTMIG-PGx primarily focuses on genetic data, the raw data shown in this example does not get mapped to any of the domains mentioned in the implementation guide. So, domain ZZ is used in this example to generate a SDTM format dataset which was generated based on the finding domains mentioned in the current draft guidance. Future guidance may include new domains that can help to transform this data to a defined SDTM domain.

The raw data file contains Specimen Type (LBSPEC), Specimen ID (LBREFID), Test Name (LBTEST), Abbreviated test name (LBTESTA), Results or findings in original units (LBORRES), Original units (LBORRESU), Sample collection date (LBDTC), Parent Procedure (LBSCAT) and Planned time point collection (LBTPT) apart from few standard variables like STUDYID, USUBJID, VISIT, etc.

LBSPEC	LBREFID	LBTPT	LBTEST	LBTESTA	LBORRES	LBORRESU	LBSCAT
BM			%IsotypePE+(1	117PPCTH	1.09	%	CD33
Aspirate-			17+Blsts)Tube4				Total
Transfix	10324724						ExpandS
	1						at BM
Whole		Pre-	%IsotypePE+(1	117PPCTH	1.05	%	CD33
Blood-		dose of	17+Blsts)Tube4				Total
Transfix	10377659	SGN-					Exp and
	3	CD33A					Sat
Whole			%IsotypePE+(1	117PPCTH	0.61	%	CD33
Blood-			17+Blsts)Tube4				Total
Transfix	10237252						Exp and
	5						Sat
BM			%IsotypePE+(1	117PPCTH	0.90	%	CD33
Aspirate-			17+Blsts)Tube4				Total
Transfix	10355869						ExpandS
	4						at BM

Table 3: Raw Biomarker Flow data, example 2

The flow data collected is mapped into ZZ domain based on standard findings SDTM format. Table 4, shows the variables generated in SDTM format. Some of the variables generated in the dataset; apart from few standard variables, like USUBJID, STUDYID, DOMAIN, VISIT, etc.

ZZTEST/ ZZTESTCD (Name of Pharmacogenomics Lab Test): Laboratory test name and abbreviated laboratory test name.

ZZCAT (Category for Pharmacogenomics Lab Test): Defines the type of specimen used for a measurement (examples: SERUM, PLASMA, URINE, WHOLE BLOOD).

ZZORRES (Results or Findings): The original result of the laboratory test.

ZZMETHOD (Method of Test): Parent Procedure

ZZTPT (Planned Timepoint Name): Text Description of the time when the specimen was taken

ZZTPTNUM: Numeric value of Timepoint variable

ZZTESTCD	ZZTEST	ZZCAT	ZZORRES	ZZORRESU	ZZMETHOD	ZZTPT	ZZTPTNUM
117PPCTH	%IsotypePE+(1 17+Blsts)Tube4	BM Aspirate- Transfix	1.09	%	CD33 Total ExpandSat BM		
117PPCTH	%IsotypePE+(1 17+BIsts)Tube4	Whole Blood- Transfix	1.05	%		Pre-dose of SGN- CD33A	1
117PPCTH	%IsotypePE+(1 17+Blsts)Tube4	Whole Blood- Transfix	0.61	%	CD33 Total Exp and Sat		
117PPCTH	%IsotypePE+(1 17+Blsts)Tube4	BM Aspirate- Transfix	0.90	%	CD33 Total ExpandSat BM		

Table 4: SDTM ZZ Domain, generated from raw data file data collected performing flow cytometry procedure for CD33 expression, ADC binding and target saturation and DNA damage

CONCLUSION

The role of biomarkers data in clinical trials is rapidly growing; with this, the generation of a submission compliant dataset becomes very important. The draft CDISC SDTMIG-PGx defines a standard for mapping gene related biomarker data. Working towards implementing these standards will help in standardizing the data and collaborative efforts with the industry clinical and data-standards experts will result in new SDTMIG-PGx domains supporting various biomarkers data. Above all, just like mapping any other data we are more familiar with, understanding the biomarkers data is crucial to allow proper capture of the values collected into SDTM domain.

ACKNOWLEDGEMENTS

I would like to thank Rajeev Karanam, Assoc. Director, Clinical Programming, and my Manager Jay Gadhiya for encouraging me to write this paper and for reviewing and Ken Wood, Director, Diagnostics and Biomarkers and David Taft, Scientist for their guidance.

REFERENCES

The following links are current at the time of publication.

- Clinical Data Interchange Standards Consortium, Inc. (CDISC). "Study Data Tabulation Model Implementation Guide: Pharmacogenomics/Genetics Version 1.0 (Draft)". Available at: http://www.cdisc.org/pharmacogenomics-genetics
- http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3078627/

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Kiran Cherukuri Seattle Genetics, Inc. 21823 - 30th Drive S.E. Bothell, WA 98021 425-527-4868 kcherukuri@seagen.com

SAS® and all other SAS® Institute Inc. product or service names are registered trademarks or trademarks of SAS® Institute Inc. in the USA and other countries. ® indicates USA registration.