A Proposed SDTM Implementation of Response Data for Solid Tumor Trials in Oncology

Mei Dey, Merck & Co., Inc
Lisa Pyle, Merck & Co., Inc

Introduction

In recent years, many pharmaceutical companies are increasingly shifting their research focus to areas, such as cancer, where patient needs aren't met by existing treatments and where the companies have expertise and good chance for scientific and commercial success. Cancer research is seen to have high potential for offering treatment improvements.

Clinical Data Interchange Standards Consortium (CDISC) Study Data Tabulation Model (SDTM) is an emerging new industry standard for submitting tabulation data to the U.S. Food and Drug Administration (FDA). While it creates opportunities to standardize data structure, transforming various clinical data using CDISC SDTM poses significant challenges.

This paper explores an innovative method to map response data in solid tumor cancer trials to a SDTM formatted Response Domain (RS). The proposed domain is an attempt to aid data transfer between vendors and sponsors and facilitate efficacy data analysis.

Response Data

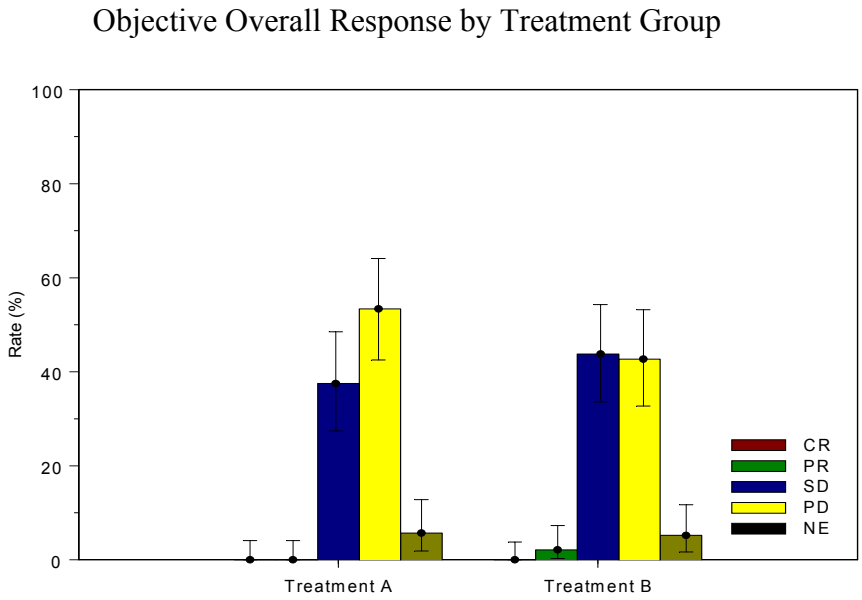
Response data is one of the key efficacy measurements for oncology trials. There are generally two types of efficacy analysis for oncology trials that require response endpoint data: response analysis and time-to-event analysis.

Response endpoint is typically an overall patient assessment performed by the investigator and/or independent radiologist. It is typically a categorical variable indicating the assessment of disease status usually based on certain well established clinical method, for example, Response Evaluation in Solid Tumors (RECIST) method.

According to NCI dictionary of cancer terms, the 4 categories of response for solid tumors are:

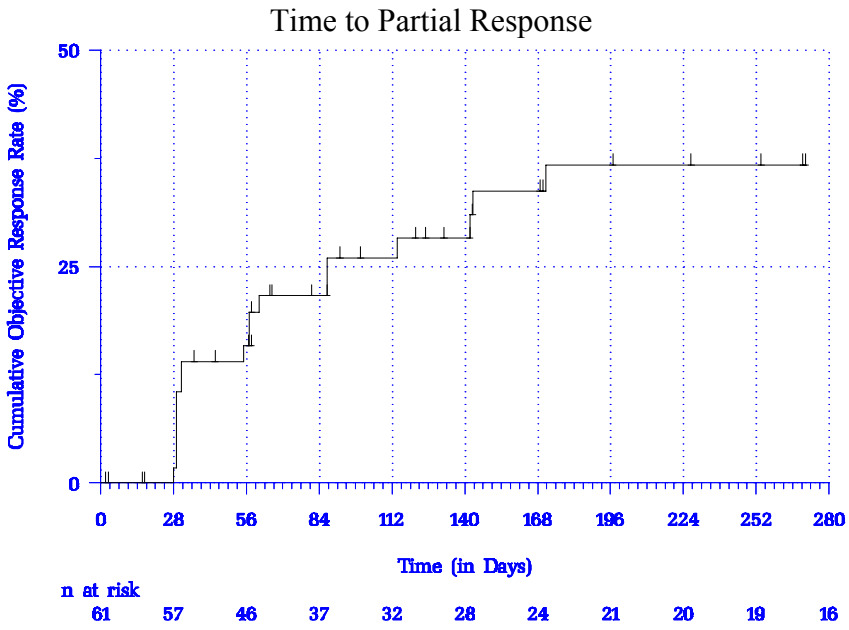
Complete Response (CR)	The disappearance of all signs of cancer in response to treatment. This does not always mean the cancer has been cured. Also called complete remission.
Partial Response (PR)	A decrease in the size of a tumor or in the extent of cancer in the body, in response to treatment. Also called partial remission.
Stable disease (SD)	Cancer that is neither decreasing nor increasing in extent or severity.
Disease Progression (PD)	Cancer that continues to grow or spread.

Here is an example graph showing overall response rates by treatment group:



In addition to summarizing the response rate, estimating the time to events is also important. Time-to-event analyses are analyses of longitudinal data on the occurrence of events where an event is a qualitative change that can be situated in time. In oncology, changes in disease status are considered events where patients transition from one specific disease state to another. For example, a patient can transition from stable disease to partial response if the test therapy is effective and so estimating the time to partial response for a therapy is of primary interest (See example below). Other variations of time to events based on disease status may be explored depending on the therapy and objectives of the trial (e.g. Progression-Free Survival, Time to Progression).

Here is an example graph showing time to partial response:



In many late stage oncology trials, the independent radiologist assessment of disease status is obtained by expert medical imaging vendors and in addition to the investigator's assessment is the source for the primary efficacy endpoint analyses just described. Different imaging vendors provide their own unique data transfer file formats (See attachment).

Due to the non-standard nature of these files formats, sponsors are challenged with importing data from different sources and integrating the data for their clinical trials. This conversion effort will increase programming and validation effort which will impact analysis cycle time.

CDISC SDTM Background

SDTM developed by CDISC SDS (Submission Data Standards) team is an evolving new industry standard structure for submitting tabulations to the FDA. In this new model, data is submitted in SDTM format.

In SDTM, data is stored in domains which are similar to tables or views or SAS datasets. A domain, distinguished by a unique two-character domain code (e.g., DM), is a collection of observations with a common topic. There are 3 types of domains – intervention, events and findings. Each has a well-defined structure and variables. SDTM specifies which variables may be present in each type of domain. Some are mandatory and others are optional. SDTM is compatible with SAS version 5 transport file format. This puts restrictions on variable names and length of character variables.

The CDISC V3.1.1 SDTM Implementation Guide (SDTM-IG) is the current industry standard that describes the details of mapping study data to SDTM. It is anticipated to be mandatory that submissions to the FDA follow SDTM format in the future after a multi-year transition period. While pharmaceutical companies, contract research organizations and the FDA rush to embrace SDTM, the process of transforming collected clinical data (source data) from various proprietary database structures to the open, industry standard model poses significant challenges.

One of the advantages of the SDTM model is that it defines data structure and is not dependent on individual vendor's system. Therefore moving forward, SDTM provides an opportunity to facilitate data exchange between vendor and sponsor using a single standard. Further, standard submission data provides many benefits to the regulatory reviewers. FDA reviewers can use standard tools to load, validate, store and analyze electronic data. Faster, more efficient, comprehensive and accurate data review can lead to improvement in patient safety.

CDISC Analysis Data Model (ADaM) Background

Submissions to regulatory agencies, such as the FDA include the derived data used in safety and efficacy analysis. In addition to having an understanding of the CDISC SDTM model foundation, it is also of interest to understand the CDISC Analysis Data Model (ADaM). Since the SDTM datasets are not intended to support statistical analyses, the ADaM model further describes the organization, format and structure for analysis data sets that are used for statistical modeling, efficacy and safety reporting.

The ADaM Implementation Guide outlines 4 principles to consider when developing an efficacy analysis dataset: traceability to the raw data to facilitate clear and unambiguous communication, be useable by currently available software tools, be linked to metadata and be analysis ready.

Since data collected by the medical imaging vendor ultimately needs to be transformed into an ADaM format that facilitates statistical reporting, having a framework for the requirements of the response analysis data set is vital. It is important to determine what collected vendor data should be retained and mapped to the proposed CDISC SDTM response domain.

The two principles of the ADaM model that were most important for determining what response data to map were the components of *analysis ready* and *traceability*. The final analysis data set for statistical modeling needs to be analyzed with little or no additional pre-programming to prepare data. The response analysis dataset needs to be ready or in a format to submit to a SAS procedure to produce results from statistical methods. For the response criteria, the analysis dataset should support response rate and time-to-event analyses.

Since most analysis datasets are derived from SDTM datasets, it is expected that there be some level of traceability between SDTM dataset(s) and analysis dataset(s). In general, the CDISC ADaM guidelines recommends to include as much supporting data in the traceability records that is needed except in instances where it is not practical to do so.

Proposed Response Domain (RS)

There have been many efforts to model response data per CDISC SDTM implementation by regional CDISC subgroups, yet no formal oncology therapeutic area domains have been accepted or adopted by CDISC.

We have developed a findings domain to collect response data. Choosing to use a findings domain may seem counterintuitive as the analyses performed on this data are related to *events*. However, to consider the event you need to know when the change occurred from one state to the next by assessing the findings from the imaging assessments. Since disease assessments are exactly findings obtained from planned medical image assessments, a findings domain class is the most logical SDTM class to hold response assessments. We have mapped the necessary collected information into the RS domain that will ultimately be relevant to the statistical analyses that follow.

Understanding the ADaM requirements of the final analysis data set and taking into consideration the SDTM modeling framework discussed in the previous sections, a model of the response findings RS domain is depicted in the following table.

RESPONSE Domain Define – RS

One record per response observation per time point per visit per subject.

Variable Name	Variable Label	Type	Origin	Role	Core
STUDYID	Study Identifier	Char	CRF	Identifier	Req
DOMAIN	Domain Abbreviation	Char	Derived	Identifier	Req
USUBJID	Unique Subject Identifier	Char	Sponsor Defined	Identifier	Req
RSSEQ	Sequence Number	Num	CRF or Derived	Identifier	Req
RSGRPID	Group Id	Char	Sponsor Defined	Identifier	Perm
RSREFID	Reference Id	Char	Sponsor Defined or Derived	Identifier	Perm

RSSPID	Sponsor Id	Char	Sponsor Defined	Identifier	Perm
RSTESTCD	Short Name of Test/Response	Char		Topic	Req
RSTEST	Response Assessment/Test	Char	CRF or Derived	Synonym Qualifier	Req
RSCAT	Category for the Response	Char	CRF or Derived	Grouping Qualifier	Perm
RSSCAT	Subcategory for the Response	Char	CRF or Derived	Grouping Qualifier	Perm
RSORRES	Result or Finding in Original Units	Char	CRF or Derived	Result Qualifier	Exp
RSORRESU	Original Units	Char	CRF or Derived	Variable Qualifier	Perm
RSSTRESN	Numeric Result/Finding in Standard Units	Num	Derived	Result Qualifier	Perm
RSSTRESC	Character Result/Finding in Std Format	Char	Derived	Result Qualifier	Exp
RSSTRESU	Standard Units	Char	CRF or Derived	Variable Qualifier	Perm
RSMETHOD	Method Used to Assess Response	Char	CRF	Record Qualifier	Exp
RSOBJ	Object or Therapy of the Observation or Response	Char	Sponsor	Plus	Perm
RSNAM	Vendor Name	Char	CRF	Record Qualifier	Exp
RSEVAL	Radiologist or Reader	Char	CRF	Record Qualifier	Exp
RSEVALID	Radiologist or Reader ID	Char	CRF	Plus	Perm
RSBLFL	Baseline Flag	Char	CRF or Derived	Record Qualifier	Exp
RSDRVFL	Derived Flag	Char	Derived	Record Qualifier	Perm
VISIT	Visit Name	Char	CRF or Derived	Timing	Perm
VISITNUM	Visit Number	Num	CRF or Derived	Timing	Req
VISITDY	Planned Study Day of Visit	Num	CRF or Derived	Timing	Perm
RSDTC	Date/Time Medical Image taken	Char	CRF or Derived	Timing	Exp
RSDY	Study Day of the Image	Num	Derived	Timing	Perm
RSTPTNUM	Planned Time Point Number	Num	CRF or Derived	Timing	Perm
RSTPT	Planned Time Point Name	Char	CRF or Derived	Timing	Perm
RSELTM	Elapsed Time from Reference Point	Char	Derived	Timing	Perm
RSTPTREF	Time Point Reference	Char	Sponsor Defined	Timing	Perm

RESPONSE Results – RS Notes (cont'd)

Variable Name	Variable Label	Notes
STUDYID	Study Identifier	Unique identifier for a study within the submission.
DOMAIN	Domain Abbreviation	Two-character abbreviation for the domain most relevant to the observation.

USUBJID	Unique Subject Identifier	Unique subject identifier within the submission.
RSSEQ	Sequence Number	Sequence number to ensure uniqueness within a dataset for a subject.
RSGRPID	Group Id	Used to tie together a block of related records in a single domain.
RSREFID	Reference Id	Could hold additional image id information if available??
RSSPID	Sponsor Id	Optional Sponsor-defined reference number.
RSTESTCD	Short Name of Test/Response	Short name of the measurement, test, or examination described in RSTEST. Limited to 8 characters.
RSTEST	Response Assessment/Test	Verbatim name o the test or examination used to obtain the measurement or finding. Limited to 40 characters.
RSCAT	Category for the Response	Used to categorize Response observations.
RSSCAT	Subcategory for the Response	A further categorization of Response.
RSORRES	Result or Finding in Original Units	Result of the Response of assessment of disease.
RSORRESU	Original Units	Original Units in which the data were collected.
RSSTRESN	Numeric Result/Finding in Standard Units	Used for continuous or numeric results or findings in standard format.
RSSTRESC	Character Result/Finding in Std Format	Contains the result value for all findings, copied or derived from RSORRES in a standard format or standard units.
RSSTRESU	Standard Units	Standardized unit used for RSSTRESC or RSTRESN.
RSMETHOD	Method Used to Assess Response	Disease assessment method, e.g. RECIST.
RSOBJ	Object of the Observation or Response	Used to identify or link what the response is made on. For example, the response can be due to the trial treatment regimen or it can be the response to a prior regimen in which case, the prior regimen name would be the object. Examples) Treatment Regimen, Regimen X, Treatment IV, etc.
RSNAM	Vendor Name	Medical Imaging Company Name
RSEVAL	Radiologist or Reader	Evaluator Name, in case of RECIST, it's the reader's name or the adjudicator name, i.e. Joe Smith
RSBLFL	Baseline Flag	
RSDRVFL	Derived Flag	
VISIT	Visit Name	Protocol-defined description of the visit
VISITNUM	Visit Number	Numeric version of VISIT
VISITDY	Planned Study Day of Visit	
RSDTC	Date/Time Medical Image taken	This is the date of the response. This date should correspond to the date the image was obtained from the patient not the date the image was examined by the radiologist or the reader. This ensures that the date of the response is real-time.
RSDY	Study Day of the Image	Study day of the Response, measured as integer days. Algorithm for calculations must be relative to the sponsor-defined RFSTDTC variable in the Demographics domain.
RSTPTNUM	Planned Time Point Number	Numeric version of RSTPT to aid in sorting
RSTPT	Planned Time Point Name	Text Description of time when measurement should be taken
RSELTM	Elapsed Time from Reference Point	Elapsed time relative to a planned fixed reference (RSTPTREF)

RSTPTREF	Time Point Reference	Name of the fixed reference point referred to by RSELTM.

Mapping Considerations for Imaging Data

1. Analysis endpoints

Our effort involves evaluating the solid tumor data submitted by the vendor and focusing on the data points required for data analysis and submission.

In our trial, imaging vendors follow a charter that describes the process the radiologist must follow to evaluate and read scans per the RECIST method. In summary, their procedures involve reading patients scans at baseline and follow up visits. During these reads, the radiologist assesses imaging quality; they identify target and non-target lesions and any new lesions. They take detailed measurements of all target lesions and assess patient disease status at each visit. Lastly, they provide an overall response assessment at the end of the trial. All of data specific to the entire imaging read process is transmitted by vendor to the sponsor whether or not they are used for the analysis.

The large quantity of collected data and the lack of standards leave imaging vendors with little direction and open to divide the data into any format they choose. One vendor may divide the data into the following 6 datasets whereas another vendor may choose to divide the same collected data into 4 datasets in yet a completely different structure.

The following six data topics are typically captured by the radiologists and submitted by vendor. For this discussion, the vendor data below will be used to describe the process for narrowing down to the key data points used in the trial.

Vendor Source Data Set	Description
RECIST	Contains data from the assessment of the patients' lesions according to the RECIST criteria. This dataset contains separate rows/observations for each subject and visit.
TARGETLESION	Contains only target/index lesion information, with data for each lesion found on separate rows.
NONTARGETLESION	Contains only non-target/non-index lesion information, with data for each lesion found on separate rows.
NEWLESION	Contains information about new lesions developed during the study, each new lesion is illustrated on a separate row.
EXAM	Contains information about the type of exam and the date the radiographic information was taken at the site
RESPONSE	Contains combined data from the Radiology and Oncology Evaluation assessments

For our solid tumor trial, the focus is on obtaining the response using RECIST criteria because it's the primary endpoint for the trial. This enabled us to immediately narrow our focus from the 6 datasets above to 3 datasets of interest: RECIST, EXAM and RESPONSE. RECIST contains RECIST assessment per patient per visit or the findings; EXAM contains date of

radiographic exam which is equivalent of date of response or the timing variables that are needed for analysis; RESPONSE contains cumulative global assessment or the collected result for each patient.

Even in these 3 datasets not all data points are needed for the primary analysis. For example, even though the Overall Target, non Target Lesion responses are important for radiologist to make assessment, it's the final integrated response by reader for each patient that needs to be mapped.

2. Adjudicated data

Adjudication of data usually occurs by a neutral third party when there are conflicting reading results among readers. They usually review all of the data and make a final determination of response. Since response data at each time point can be reviewed by multiple reviewers, the reader's name is recorded and mapped into RSEVAL. However, the best overall response data per patient is the final adjudicated response and is mapped to the TEST = "Best Response (Adjudicated)" without a reader designation in RSEVAL which implies final reading.

3. Mapping Investigator Data

In addition to external imaging reading, sponsors sometimes choose to collect response data directly on the eCRFs per the investigator review at pre-planned time points and this data can be directly mapped to the RS domain with little or no additional effort. We recommend using the same mapping (e.g. RSTESTCD, RSTEST, RSORRES) whenever feasible for easy comparison with external imaging data should a need arise.

4. Other SDTM implementation considerations

Although the complete set of imaging data is not mapped to the response domain some of the measurement data and procedure information may need to be retained by the sponsor. Since the method for determining response by imaging vendors is generally a well-established medically accepted approach with much documentation supporting the derivation, it seems counterproductive to map the measurement details within the same SDTM domain as the response data. Two recently proposed Oncology Domains in the CDISC community are the Tumor Results(TR) and Tumor Identification(TU) domains. These comprehensive domains contain the tumor lesion measurements and the details on the target and non-target lesion identification. The additional tumor data should be in the TR and TU domain and linked back to this response domain RS by RELREC methods, sequencing IDs (e.g. xxSEQ), or some other relational method.

This will enable information like modality, lesion location, lesion measurement to reside in at least one other domain (e.g. LS) or more domains. The expectation is that RS domain can be related to LS domain (e.g. via xxSEQ, RELREC, etc).

Imaging companies have programs that take the lesion measurement inputs and provide the derivation results much like central lab vendors, our only concern for mapping data to a response domain is to map the response criteria at each time point or each time a scan is read and evaluated per patient.

Response Domain (RS) Mapping Details and Examples

With these considerations in mind, we provide a RS domain mockup with sample data illustrating how we envision response data to be mapped. The first example describes how the

radiologist assessments would be mapped and the second provides a related mockup for how the investigator's assessments would be mapped.

Example 1 Subject 0001 Mapped Response Data from Medical Imaging Vendor

	RSSEQ	SUBJID	RSGRPID	RSDTC
Row 1	2345678913	0001	RAD	28-Jan-2008
Row 2	2345678914	0001	RAD	5-Mar-2008
Row 3	2345678915	0001	RAD	11-Mar-2008
Row 4	2345678916	0001	RAD	22-Apr-2008
Row 5	2345678917	0001	RAD	2-May-2008
Row 6	2345678918	0001	RAD	11-Mar-2008
Row 7	2345678919	0001	RAD	2-May-2008

	RSTESTCD	RSTEST	RSOBJ
Row 1 (cont)	OVERALL	Overall Response	TREATMENT REGIMEN
Row 2 (cont)	OVERALL	Overall Response	TREATMENT REGIMEN
Row 3 (cont)	OVERALL	Overall Response	TREATMENT REGIMEN
Row 4 (cont)	OVERALL	Overall Response	TREATMENT REGIMEN
Row 5 (cont)	OVERALL	Overall Response	TREATMENT REGIMEN
Row 6 (cont)	BEST	Best Response	TREATMENT REGIMEN
Row 7 (cont)	FIRSTPD	First Documented Progressive Disease	TREATMENT REGIMEN

	RSORRES	RSMETHOD	RSNAM	RSEVAL
Row 1 (cont)	PARTIAL RESPONSE	RECIST	XHumanImaging Co.	Reader 1
Row 2 (cont)	STABLE DISEASE	RECIST	XHumanImaging Co.	Reader 1
Row 3 (cont)	COMPLETE RESPONSE	RECIST	XHumanImaging Co.	Reader 1
Row 4 (cont)	NOT EVALUABLE	RECIST	XHumanImaging Co.	Reader 1
Row 5 (cont)	PROGRESSIVE DISEASE	RECIST	XHumanImaging Co.	Reader 1
Row 6 (cont)	NOT ASSESSED	RECIST	XHumanImaging Co.	NA
Row 7 (cont)	PROGRESSIVE DISEASE	RECIST	XHumanImaging Co.	Reader 2

1. RSDTC is exactly the date the image was taken. Since date of response is crucial to be associated with each response for time-to-response analysis, we request our vendor to provide the exam date with each assessment record. This should not be the date the image is read by the radiologist as it can take days even weeks in some cases to transfer the images to the reader for assessment. All subsequent programming for analyses should be based on this date

2. RSOBJ is the only supplemental qualifier defined in this findings domain. In this example, it is hard-coded to "TREATMENT REGIMEN" since the response on the observation is the response or outcome that results from the regimen administered in the trial during treatment. In practice this is directly related to the treatment EXTRT in the exposure (EX) domain for the trial and can be the generic drug name.

3. RSTESTCD is limited to 8 characters in length to be SDTM compliant. At this time, there is no CDISC dictionary of test codes for response, so they are sponsor defined. In this trial, OVERALL identifies the Overall Response at each time point and BEST identifies the Adjudicated Best Response. RSTEST is the full description of the test and is a one-to-one relationship with the RSTESTCD.

4. RSGRPID and RSNAM in tandem contain the name of the imaging vendor provider. The variables are one-to-one like RSTESTCD and RSTEST; they are the short description and full name of the vendor.

5. RSORRES is the result of the response of assessment of disease. Example shows NCI dictionary of cancer terms. RSORRES of 'NOT EVALUABLE (NE)' are patients whose disease status can't be evaluated (e.g. missing scans, poor quality images).

6. RSMETHOD is hard-coded to RECIST since this is the method used to obtain the disease assessment or response in this trial. This does not indicate the method of the image, such as CT, X-ray or MRI, as the scan or the image is not the topic of this domain.

7. RSEVAL is the image evaluator or reader. It holds the reader's actual name. In some instances it may be of interest to map the reader's associated identification code RSEVALID but at this time there is no accommodation for this information formally in the RS Domain define. If there is a need, RSEVALID would be an additional supplemental qualifier to this domain.

Example 2 - Subject 0001 Response Data from the Investigator

	RSSEQ	SUBJID	RSGRPID	RSDTC
Row 1	2345678925	0001	OPA	15-Jan-2008
Row 2	2345678926	0001	OPA	28-Mar-2008
Row 3	2345678927	0001	OPA	10-Apr-2008
Row 4	2345678928	0001	OPA	15-May-2008
Row 5	2345678929	0001	OPA	1-Jun-2008

	RSTESTCD	RSTEST	RSOBJ
Row 1 (cont)	OVERALL	Overall Response	TREATMENT REGIMEN
Row 2 (cont)	OVERALL	Overall Response	TREATMENT REGIMEN
Row 3 (cont)	OVERALL	Overall Response	TREATMENT REGIMEN
Row 4 (cont)	OVERALL	Overall Response	TREATMENT REGIMEN
Row 5 (cont)	OVERALL	Overall Response	TREATMENT REGIMEN

	RSORRES	RSMETHOD	RSEVAL
Row 1 (cont)	STABLE DISEASE	RECIST	INVESTIGATOR
Row 2 (cont)	STABLE DISEASE	RECIST	INVESTIGATOR
Row 3 (cont)	STABLE DISEASE	RECIST	INVESTIGATOR
Row 4 (cont)	PROGRESSIVE DISEASE	RECIST	INVESTIGATOR
Row 5 (cont)	PROGRESSIVE DISEASE	RECIST	INVESTIGATOR

There is no mapping complication and the data can readily be distinguished from the external imaging data by the variable RSGRPID in the RS domain.

Mapping Benefits

- This proposed RS domain allows sponsors to integrate response data easily from all sources: external and site collected investigator data.
- There is minimal post processing required for further analysis since the format is nearly analysis ready and very similar to the final ADaM dataset format. In this model, the traceability is achieved by augmenting the analysis data with the rows of data from which the derived overall response is determined and adding variables to indicate the domain, variable and sequence number of the source SDTM data.
- This RS domain focuses on the basic SDTM structure and avoids the use of supplemental qualifiers in an attempt to streamline the external data transfer from vendor to sponsor.
- This proposal of a standard response domain eliminates the need for sponsors to remap data from each vendor and provides a transfer standard to be adopted by industry.

Operational challenges

There are many challenges to streamline the analysis of response data in oncology trials. Obtaining buy-in from both the sponsor and the vendor to accept the SDTM approach as the method for transferring data is a start. However, vendors are probably not apt to agree to SDTM as they are short of resources and have limited SDTM knowledge. Similarly, key responsible people in sponsor organizations might not be SDTM knowledgeable either. Sponsors probably don't want to invest the time to remap data from every different vendor based on their own unique data formats either. Therefore, understanding long-term benefit of adopting industry standard and identifying the responsible party to convert imaging data or response data to SDTM will be the key to streamline the process.

It is worthwhile to note that people from both sponsor and vendor sides who understand data analysis issues need to be involved. The mapping process requires SDTM knowledge, clinical knowledge as well as database, ADaM and analysis expertise.

Code lists for variables like RSTESTCD and RSMETHOD currently are sponsor-derived but moving forward some industry standard for codes would be more beneficial in defining the RS mapping.

Even if a RS domain is adopted, issues regarding linking the response data to other potential domains still need to be ironed out. Sponsors may require that additional imaging data be mapped and retained so the relationship between the additional data and the response data would need to be defined and mapped as well.

Conclusion

Mapping response data using SDTM is an innovation solution taking consideration of analysis needs, internal/external data requirements and ADaM. We believe that this proposed domain would benefit oncology trials and we feel that this proposed domain could be expanded to other indications besides solid tumor trials with ease. Adopting this mapping as an oncology standard would streamline analysis and reporting within the therapeutic area. In addition, we try to explore using it as a data transfer model to facilitate data exchange between vendor and sponsor.

CDISC SDTM model is the new industry standard that poses significant challenges. Using single standard facilitates data collection, analysis and data transfer, which will ultimately translate into shorter development time within solid tumor trials of the oncology therapeutic area and beyond.

References

Analysis Data Model, Prepared by the CDISC ADaM Team, Version 2.1

ADaM Implementation Guide, Prepared by the CDISC ADaM Team, Version 1.0

Allison, Paul D., "Survival Analysis Using The SAS System A Practical Guide", 1995, The SAS Institute

New Guidelines to Evaluate the Response to Treatment in Solid Tumors, Journal of the National Cancer Institute, Vol. 92, No. 3, February 2, 2000

New response evaluation criteria in solid tumours: Revised RECIST guideline (version 1.1), European Journal of Cancer 45 (2009)

Study Data Tabulation Model, Prepared by the CDISC Submission Data Standards Team, Version 1.1

SDTM Implementation Guide, Prepared by the CDISC Submission Data Standards Team, Version 3.1.1

SDTM Implementation Guide, Prepared by the CDISC Submission Data Standards Team, Version 3.1.2

ACKNOWLEDGEMENTS
The author would like to thank Mary Varughese, Jason Clark, Cong Chen and Linda Sun for their careful review and helpful comments. The author also thanks the management team for their review of this paper and help to go through the company publication process.

CONTACT INFORMATION
Your comments and questions are valued and encouraged. Contact the authors at:

Mei Dey
Merck & Co., Inc.
P.O. Box 1000
351 North Sumneytown Pike
North Wales, PA 19454
Phone: (267) 305-6814
Fax: (267) 305-6474
Email: mei_dey@merck.com

Lisa Pyle
Merck & Co., Inc.
P.O. Box 1000
351 North Sumneytown Pike
North Wales, PA 19454
Phone: (267) 305-6887
Fax: (267) 305-6474
Email: lisa_pyle@merck.com

Attachment - Sample Vendor Define Documentation

RECIST Dataset

Dataset	Field Label	Format	Length	Description	Plausible Values
RECIST	LAB_ID	character	8	Contains external lab name	(a) Equals: "ABC LAB" (b) Field cannot be NULL
RECIST	PROTOCOL_ID	character	30	Protocol ID	(a) Equals: "MK-999-001" (b) Field cannot be NULL
RECIST	SITE_ID	character	20	The site id for the site to which the subject is enrolled.	(a) Permitted value is a 3 digit number (b) Values range from 001-999 (c) Field cannot be NULL
RECIST	SUBJECT_ID	character	20	The id assigned to the subject by the sponsor.	(a) Permitted value is a 4 digit number (b) Values range from 0001-9999 (c) Field cannot be NULL
RECIST	TIMEPOINT_NAME	character	50	The sponsor-defined timepoint name for this record.	(a) Permitted values are listed in Appendix 5 (b) Field cannot be NULL
RECIST	READER_ID	character	50	The user ID for the reader.	(a) e.g. RFeldman or R Feldman (b) Field cannot be NULL
RECIST	READER_TYPE	character	20	The description of the reader	(a) Permitted values are listed in Appendix 2 (b) Field cannot be NULL
RECIST	SLD	character	10	Sum of the Longest Diameters	(a) Sum of the longest lesion diameters measured in mm (b) Plausible range values would be 0 – 2000 or "UE" or "ND" (c) If no target lesions are recorded then "NA" should be recorded (d) If at least one lesion has an LD of "UE" or "ND", then "UE" or "ND" should be recorded for the SLD UNLESS the lesions that have measurements available confirms progression, then the SLD will be recorded using the measurements recorded (e) Decimal places are not allowed, round as 0 – 4: round down, 5 – 9: round up. (f) Field cannot be NULL.

RECIST	PCB_SLD	character	10	Percent Change from Baseline for SLD.	(a) Derived value as: 100x[SLD(current visit) – SLD (SCR visit)]/SLD(SCR visit) (b) Can be missing if either visit above is missing (c) Will be missing if 'current visit' is Screening (d) Shrinkage in the lesion LD (length) will result in a negative % change (e) Decimal values are not permitted, round as 0 – 4: round down, 5 – 9: round up (f) Values of "UE", "ND" and "NA" are acceptable
RECIST	PCN_SLD	character	10	Percent Change from Nadir for SLD.	(a) Nadir = the smallest non-missing SLD since treatment began to the visit **prior** to the current visit = ♣ (b) Derived as: 100x{[SLD(current visit) - ♣]/♣} (c) Can be missing if current visit is missing (d) Will be missing if 'current visit' is Screening (e) Decimal values are not permitted, round as 0 – 4: round down, 5 – 9: round up (f) Values of "UE", "ND", and "NA" are acceptable (g) Shrinkage in the lesion LD (length) will result in a negative % change.
RECIST	RAD_T_RSP	character	16	Overall Target Lesion response by radiologist	(a) Permitted values are: "CR", "PR", "SD", "PD", "UE", and "NA". (b) Field will only be NULL for Screening (c) Field cannot be NULL for any visit after Screening
RECIST	RAD_NT_RSP	character	16	Overall Non-Target Lesion response by radiologist	(a) Permitted values are: "CR", "SD", "PD", "NA", and "UE". (b) Field will only be NULL for Screening (c) Field cannot be NULL for any visit after Screening
RECIST	NEW_LESIONS	character	1	True if new lesions (since baseline) were noted for this timepoint.	(a) Permitted values are 'T' or 'F'. (b) No missing values are allowed. (c) Value of 'T' indicates that New lesions are recorded.
RECIST	INT_RSP	character	16	Integrated Response by reader for this timepoint.	(a) Permitted values are: "CR", "PR", "SD", "PD", "UE", "ND", and "NA" (b) Field will only be NULL for Screening (c) Field cannot be NULL for any visit after Screening
RECIST	INT_RSP_DT	date	11	Date of INT_RSP.	(a) Must be in YYYYMMMDD format (b) No partial dates (c) Date field will only be recorded if INT_RSP is present
RECIST	BEST_RSP	character	16	Best response per subject per timepoint	(a) Permitted values are: "CR", "PR", "SD", "PD", "UE", and "NA" (b) Field will be NULL at Screening (c) Field cannot be NULL for any visit after Screening

RECIST	BEST_DT	Date	11	Date of BEST_RSP	(a) Must be in YYYYMMMDD format (b) No partial dates (c) Date is only expected when BEST_RSP is completed
RECIST	PROGRESSION_DT	Date	11	Date of Progression	(a) Must be in YYYYMMMDD format (b) No partial dates (c) Field can be NULL if subject did not have PD as noted in the INT_RSP (d) Must be NULL if Visit is 'Screening'
RECIST	OVER_BEST_RSP	character	16	Overall Best Response	(a) Permitted values are: "CR", "PR", "SD", "PD", and "UE"
RECIST	OVER_BEST_RSP_DT	date	11	Overall Best Response date	(a) Must be in: YYYYMMMDD format (b) No partial dates allowed

EXAM Dataset

Dataset	Field Label	Format	Length	Description	Plausible Values
EXAM	LAB_ID	character	8	Contains external lab name	(a) Equals: "ABC LAB" (b) Field cannot be NULL
EXAM	PROTOCOL_ID	character	30	Protocol ID	(a) Equals: "MK-999-001" (b) Field cannot be NULL
EXAM	SITE_ID	character	20	The site id for the site to which the subject is enrolled.	(a) Permitted value is a 3 digit number (b) Values range from 001-999 (c) Field cannot be NULL
EXAM	SUBJECT_ID	character	20	The id assigned to the subject by the sponsor.	(a) Field cannot be NULL (b) Permitted value is a 4 digit number (c) Values range from 0001-9999
EXAM	TIMEPOINT_NAME	character	50	The sponsor-defined visit name for this record.	(a) Permitted values are listed in Appendix 5 (b) Field cannot be NULL
EXAM	EXAM_NO	character	10	The exam sequence number	(a) Assigned by the ABC LAB project team. Typically

					begins with 1, 2, 3, etc. (b) Value cannot be “ND”, “UE”, or “NA” (c) Field cannot be NULL
EXAM	EXAM_DATE	date	11	Date that the radiographic exam was performed on the subject at the site.	(a) DDMMMYYYY format required (b) No partial dates are acceptable (c) Date cannot be NULL if exam was performed (d) Values of “ND”, “UE”, and “NA” are not acceptable
EXAM	EXAM_TYPE	character	30	Description of the type of the exam.	(a) Acceptable Exam Types are listed in Appendix 4. (b) If “OTHER”, the specific exam name will be concatenated with the exam type of “OTHER”. (c) Field cannot be NULL
EXAM	EXAM_OTHER	Character	50	Further description of the exam	(a) Free text field to further describe the exam. (b) Must be present when EXAM_TYPE is “OTHER”, “X-RAY”, or “PHOTO”.
EXAM	RI	character	10	The reconstruction interval (mm).	(a) Acceptable values are: 0 – 20 (b) Decimal points are acceptable (c) Field may be “NA” depending on Exam type, NULL field is not acceptable

RESPONSE Dataset

Dataset	Field Label	Format	Length	Description	Plausible Values
RESPONSE	LAB_ID	character	8	Contains external lab name	(a) Equals: “ABC LAB” (b) Field cannot be NULL
RESPONSE	PROTOCOL_ID	character	30	Protocol ID	(a) Equals: “MK-999-001” (b) Field cannot be NULL
RESPONSE	SITE_ID	character	20	The site id for the site to which the subject is enrolled.	(a) Permitted value is a 3 digit number (b) Values range from 001-999 (c) Field cannot be NULL
RESPONSE	SUBJECT_ID	character	20	The id assigned to the subject by the sponsor.	(a) Field cannot be NULL (b) Permitted value is a 4 digit number (c) Values range from 0001-9999
RESPONSE	OVER_BST_RSP	character	2	Overall Best Response per patient	(a) Permitted values are: “CR”, “PR”, “SD”, “PD”, “UE” (b) Will only be captured once per subject. (c) Field will be NULL if only the SCREENING

					timepoint was received.
RESPONSE	OVER_BST_RSP_DT	Date	9	Overall Best Response date	(a) Must be in DDMMMYYYY format. (b) No partial dates. (c) Field will be NULL if only the SCREENING timepoint was received.
RESPONSE	PROGRESSION_DT	Date	9	Date of Progression	(a) Must be in DDMMMYYYY format. (b) No partial dates. (c) Field cannot be NULL if a timepoint response is PD.
RESPONSE	FIRST_BST_RSP_DT	Date	9	Date of First Response	(a) Must be in DDMMMYYYY format. (b) No partial dates. (c) Field will only be populated when there is a “CR” or “PR” for the BEST_RSP.