

Table of Contents

1	INTRODUCTION TO CDISC	3
1.1	Background	3
1.2	The Scope of CDISC Activities.....	3
2	THE LAB TEAM	4
2.1	Team Members	4
2.2	Purpose of the Lab Team.....	4
3	THE LAB REFERENCE RANGE MODEL	5
3.1	Defining the problem.....	5
3.2	Designing a New Model.....	5
3.2.1	Assumptions Made	5
3.2.2	Approach Taken	6
3.3	Contents of the Model	6
3.3.1	Data Fields.....	6
3.3.2	Code Lists.....	7
3.3.3	Date and Time Format Conventions.....	8
3.4	Using the LAB Reference Range Model.....	8
3.4.1	General	8
3.4.2	Good Transmission Practice.....	9
3.4.3	Study.....	9
3.4.4	Base Battery	10
3.4.5	Base Test	10
3.4.6	Subject Characteristics	11
3.4.7	Unit of Measure (UOM).....	12
3.4.8	Normal Definition	12
3.4.9	Delta Definition	13
3.4.10	Exclusion Definition.....	14
3.4.11	Alert Definition	14
3.5	Comments on Practical Application	15
3.5.1	Use of Model Fields	15
3.5.2	Data Compression	15
3.5.3	Transmission Agreements	16
4	REVIEW PROCESS	16
5	DIRECTIONS FOR SUBMITTING COMMENTS.....	16

CDISC LAB MODEL REFERENCE RANGE FORMAT VERSION 1.0.1

1 Introduction to CDISC

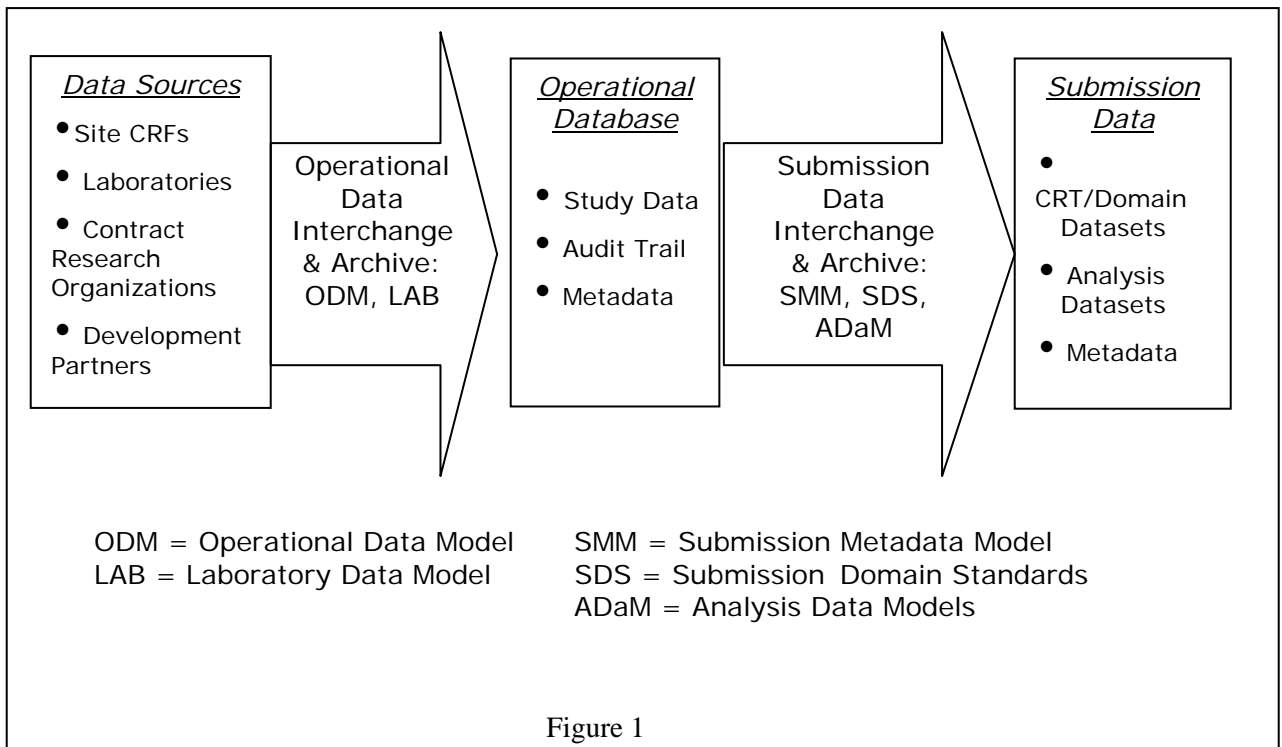
1.1 Background

CDISC is an open, multidisciplinary, non-profit organization committed to the development of industry standards to support the electronic acquisition, exchange, submission, and archiving of clinical trials data and metadata for medical and biopharmaceutical product development.

The mission of CDISC is to lead the development of global, vendor-neutral, platform-independent standards to improve data quality and accelerate product development in our industry.

1.2 The Scope of CDISC Activities

CDISC seeks to achieve its mission through the development of standard data models designed to support the end-to-end data flow of clinical trials from the data sources into an operational database and through to analysis and submission, as shown in Figure 1 below:



2 The LAB Team

2.1 Team Members

The members of the LAB team represent a broad cross-section of stakeholders from within the biopharmaceutical industry who have an interest in clinical laboratory data. The team is currently comprised of representatives from 5 pharmaceutical companies, two technology companies, one CRO and 4 central laboratories. The Lab Team has had representation from the following companies and groups during the development of this release of the LAB Reference Range Model:

BARC NV
Covance CLS
Covidence GmbH
CRL Medinet
Duke Clinical Research Institute
Eli Lilly and Company
GlaxoSmithKline
HL7
IBM Life Sciences
LabCorp
Merck
Physician's Reference Laboratory
Pfizer
Procter and Gamble
Quintiles CRO
Quintiles Lab
Regenstrief Institute
Schering AG

The membership of the team has incorporated expertise from a variety of clinical, technical and statistical disciplines and a variety of different perspectives including academic, commercial and European.

2.2 Purpose of the Lab Team

One of the largest components of clinical trial data is laboratory data, making the development of models to support it an important CDISC initiative.

Formed in 2000, the Laboratory Data Team (LAB) has as its mission the development of standard models for the acquisition and interchange of laboratory data.

The mission of the Lab Team is to:

- Define requirements for improving operational laboratory data interchange.
- Develop standard content models for the acquisition and interchange of clinical trials laboratory data
- Test these models with complex real laboratory data to assure their functionality
- Explore other opportunities to improve laboratory data processing with standards

3 The LAB Reference Range Model

3.1 Defining the problem

Standard models for the interchange of laboratory data do exist already but they are very seldom used within the biopharmaceutical industry. Examples of such standards are ACDM, ASTM, HL7 and X12. None of these standards specifically address the need to transmit the Reference range based data for a clinical trial protocol.

In the absence of acceptable industry standards, each CRO and biopharmaceutical company has developed their own format, specifically designed for their own particular needs. Furthermore, these formats rarely apply to all clinical trials and instead usually tend to be developed on a per-study basis.

3.2 Designing a New Model

3.2.1 Assumptions Made

The assumptions made in the design of the LAB reference range model can be summarized as follows:

- A separate model for the interchange of reference range and flag definitions should be made and should be based upon the laboratory result data model.
- A successful model should be designed specifically to handle the transfer of all reference ranges and flags defined for a clinical trial.
- The structure and contents of the model should be intuitive and clearly understandable to industry stakeholders familiar with clinical trial data and should have straightforward and easy to follow rules.
- The model should be sufficiently flexible that it could be applied to any clinical trial and keep pace with industry changes but not sacrifice simplicity in an attempt to cope with extreme cases.
- The first release of the model should be as comprehensive as possible to avoid the need for continual updates and revisions in the future other than those related to changes within the industry.
- The model should not be limited to any one specific implementation and so risk rejection by industry stakeholders who would otherwise be willing to embrace it but for whom the implementation chosen is incompatible with their technical infrastructures.
- The development of the model should concentrate on content first and implementation second.
- The model should accommodate as much as possible the different practices of the many central laboratories, CROs and biopharmaceutical companies within the industry.
- The model should support data interchange between any organizations in the industry and not just the classic flow from lab to CRO to sponsor. For example: reference laboratory to central laboratory to CRO to biopharmaceutical company to partner biopharmaceutical company.

- The model should use existing standards and draw upon the work of existing standards organizations wherever appropriate (for example, code lists from HL7, ISO, LOINC, etc.).
- The model should allow some controlled flexibility in the way that some data is represented to support differing preferences within the industry.
- The model should support both incremental and cumulative data interchange.
- The model should support the interchange of data from either one or more studies within a single file.

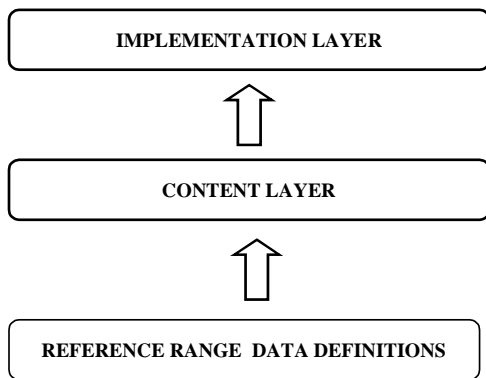
3.2.2 Approach Taken

The first step was to define exactly what the industry means by ‘clinical laboratory reference range data’ in order to build a ‘superset’ of data items that meet the need of all satisfaction of all stakeholders involved. This superset of fields constitutes the content of the model.

Next, the structural definitions of those data items were defined in terms of data type, length, default values, standards of representation, code lists and whether or not they should be optional or required. Wherever appropriate existing standards were employed. For example: ISO 8601 date and time representations and some HL7 code lists are used.

From there the notion of a ‘multi-layer’ model was developed whereby the first layer would be the content layer and above that would be an implementation layer, the idea being that the content would not change but the implementation could. The advantage of this approach is that it offers flexibility but retains control: it doesn’t make the use of the model dependent upon any one implementation and if different implementations are used the content remains the same so the standard still applies.

The design of the model is thus as follows:



3.3 Contents of the Model

3.3.1 Data Fields

The superset of data fields is separated into 7 logical levels as follows:

1. Good Transmission Practice
2. Study
3. Base Battery
4. Base Test
5. Subject Characteristics
6. Units of Measure
7. Definition

The Definition level contains the following blocks for defining different types of normative ranges and flags. A specific transmission can use one, several or all of these blocks.

- Normal Reference Range Definition
- Delta Definition
- Exclusion Definition
- Alert Definition

3.3.2 Code Lists

Since the purpose of the model is to improve the interchange of laboratory data it is important to consider not only the structure and the contents of the model in terms of the fields it contains and how they are organized but also the population of those fields. Accordingly, for some fields, code lists have been suggested. The purpose of these code lists is to offer a higher degree of standardization and so further improve the reliability and accuracy of data interchange.

Some of the suggested code lists are as follows:

Variable	Code List
LOINC Code – The corresponding LOINC code for the Test or Lab Test ID.	LOINC
Units (Reported, Conventional, SI)	HL7 Common ISO derived units and ISO+ extensions
Subject Gender	HL7 Gender Vocabulary Domain
Subject Race	HL7 Race Vocabulary Domain
Subject Medical Condition	SNOMED or ICD

A full list of recommended codes can be found in the CDISC Laboratory Data Interchange Reference Range Transmission Data Fields document at WWW.CDISC.ORG.

THE RECOMMENDED HL7 CODE TABLES CAN BE ACCESSED ON THE CDISC WEBSITE AT WWW.CDISC.ORG. USE OF THE HL7 CODE MNEMONIC IS RECOMMENDED.

The use of LOINC[®] (Logical Observation Identifier Names and Codes) codes to supplement internal test codes is recommended. A more standardized test code system will permit higher-

level analysis of data and will be of assistance in the future should regulatory authorities mandate such a standardized coding system. Note that on March 21, 2003 the U.S. Department of Health and Human Services announced that all federal agencies that deal with health care data will adopt laboratory LOINC codes to standardize electronic interchange of clinical laboratory results. Some pharmaceutical companies have already made the switch from proprietary internal codes to LOINC. Although many companies will not be able to use LOINC exclusively, supplementing internal codes with LOINC should be advantageous. For further discuss about LOINC codes, see the CDISC LAB Base Model Definition document.

The World Wide Web URL <http://www.loinc.org> provides downloads of all LOINC files and documentation.

3.3.3 Date and Time Format Conventions

Times are clock readings within a 24-hour period. The hour (hh) ranges from 00 to 23. The minutes (mm) and the seconds (ss) range from 00 to 59. The optional fractional part (.nnn) expresses milliseconds. Dates, times, and datetimes are to be interpreted as local clock readings at the place the data was collected. In a datetime, the +hh:mm (or -hh:mm) is the offset in hours and minutes to add to (or subtract from) Coordinated Universal Time to get the local clock reading at the time the data was collected. A special offset of -99:99 means that the relationship between the local clock and Universal Time is not known.

Examples:

- 3:14 pm on 3 January 2001 in Chicago (6 time zones west of Greenwich, standard time) would be represented as "2001-01-03T15:14:00-06:00".
- 3.5 seconds after midnight on the morning of July 20th 2001 in Chicago (daylight time) would be represented as "2001-07-20T00:00:03.500-05:00".

Note: The above formats for dates, times, and datetimes (YYYY-MM-DDThh:mm:ss±hh:mm) are compatible with ISO 8601 except for the use of the -99:99 offset.

In the CDISC SAS data set implementation all date/time fields should be defined as a character variable of length 25, and contain the date, time and UTC offset as specified in the recommended ISO 8601 format. The data recipient can parse the date and/or time from such a text field into a true SAS date and/or time field (which is numeric) and apply the UTC offset as needed for any calculations.

3.4 Using the LAB Reference Range Model

3.4.1 General

The field lengths given in the model are maximum field lengths.

The default implementation of the LAB Model is bar delimited ASCII. Therefore, for ASCII implementations, the fields should be separated by the vertical bar or pipe character '|'.
|

All fields must be present for the SAS and ASCII implementations. This means that all of the fields must be represented in the transmission whether it has a value or not. In the XML implementation, optional fields that would take a NULL value should be defaulted completely out of the file, with no empty name tag present, since the schema provides the information on the fields location and that the field is optional .

The SAS Variable Names provided on the Reference Range Version 1.0.1 spreadsheet have been matched to the current draft Version 3.0 of the CDISC SDS model.

3.4.2 Good Transmission Practice

Good Transmission Practice data provides information about either the transmission as a whole or a particular record within it but not the study for which the transmission is being made.

Note that the Transmission Source ID identifies the organization that is the source of the transmission but not necessarily the source of the data. For example, in a transmission from a central laboratory to a CRO this would identify the central laboratory as the source of the transmission and in a transmission from a CRO to a biopharmaceutical company this would identify the CRO as the source of the transmission.

Model Version, File Creation Date and Time and Transmission Source ID must always be populated.

Variable	Description
Model Version	The version of the CDISC Laboratory Data Interchange Standard model indicating by definition which fields are contained within it. This field is formatted as ##-#-## where “#” is an integer. The sequence indicates major-intermediate-minor releases. The current version “01-0-01” is the first minor enhancement release of the first major release.
File Creation Date and Time	The local date and time the data file was created. This includes a Universal Time Offset plus/minus hours and minutes.
Transmission Source ID	The ID of the organization that is the source of the data transmission. This normally is an ID defined by the sponsor.
Transmission Source Name	The Name of the organization that is the source of the data transmission.

3.4.3 Study

Study data describes the study and the type of transmission being made for it. Data for more than one study can be in a data file.

Study ID and Transmission Type must always be populated.

Variable	Description
Study ID or Number	The ID of the study. This normally is an ID defined by the sponsor.
Study Name	The name of the study. This normally is a short name defined by the

	sponsor.
Transmission Type	This indicates what type of transmission the data transmission is. There are two transmission types: C – Cumulative, I - Incremental.

3.4.4 Base Battery

Battery data describes the battery, panel or group of tests whose reference ranges are being defined. The Battery level is placed here so that in a hierarchical model (XML), all reference ranges of a related type (e.g. Serum Chemistry) are grouped together. The Battery ID field is required. If a Battery, Panel or Group ID does not exist, use the Test ID from the next level.

Variable	Description
Battery ID	The ID of the battery or panel to which the test belongs. If a Battery ID does not exist, use the Test ID from the next level. . This normally is an ID defined by the transmitting party (e.g. centr lab).
Battery Name	The name of battery or panel to which the test belongs.

3.4.5 Base Test

Test data identifies the test, the laboratory that is to perform the test and the authority who has defined the reference range information for that test.

The test identification fields are available for both the data recipient and data provider test. This is done to accommodate those situations where a common or standardized test code is not available or agreed upon between the data recipient or data provider.

Note that the Performing Laboratory ID identifies the laboratory that will actually perform the test and will thus be the source of the *data*. This might be either a reference laboratory or the central laboratory.

The Lab Test ID must always be populated.

Variable	Description
Lab Test ID	The ID of the test performed as defined by the data provider.
Lab Test Name	The name of the test performed as defined by the data provider.
Test ID	The ID of the test performed as defined by the data recipient.
Test Name	The name of the test performed as defined by the data recipient.
LOINC Code	The LOINC code ID for the test performed.
LOINC Code List ID	If utilized, the code list identifier and version number for the LOINC code.
Additional Test Description	Additional Test Description to characterize additional aspects associated with the test. This information may be important for interpretation for the study. For example: instrument model number, lot numbers, etc

Variable	Description
Performing Laboratory ID	The ID of the laboratory that performed the test. This normally is an ID defined by the sponsor.
Performing Laboratory Name	The name of the laboratory that performed the test.
Reference Range Defining Entity	A coded value that indicates the entity that defined the reference range set. C - Central Lab R - Referral Lab S - Sponsor

3.4.6 Subject Characteristics

Subject characteristic data describes the demographic factors that will be used to define sub-populations with their own reference range values. The Subject Characteristics level follows the test level so that within a test, all reference range definitions for a sub-population are grouped together in a hierarchical implementation (XML)

The model assumes that at least one subject age bracket must be defined (this may be 0 to 999) and thus the five fields that define an age cohort (Subject Age Boundary Type, Subject Age Lower Limit, Subject Age Lower Limit Units, Subject Age Upper Limit and Subject Age Upper Limit Units) must always be populated.

Note that the lower and upper age boundaries may have different units – an age cohort may be defined as 6 months to 3 years.

For both subject sex and subject race the use of the appropriate HL7 table is recommended. If sex and or race are not used to define a sub-population, these fields should be left blank (defaulted out in the XML implementation).

Variable	Description
Subject Sex	The sex of the reference range sub-population.
Subject Sex Code List ID	If utilized, the code list identifier and version number for the Subject Sex Code.
Subject Race	The biological race of the reference range sub-population
Subject Race Code List ID	If utilized, the code list identifier and version number for the Subject Race code.
Subject Age Boundary Type	The inclusiveness of the low and high boundaries for the age bracket. There are four boundary types: B - Both boundaries are inclusive L - The low boundary is inclusive N - Neither boundary is inclusive U - The upper boundary is inclusive
Subject Age Lower Limit	The low boundary for the age range bracket that defines the sub-population.
Subject Age Lower Limit Units	The units in which the Subject Age Lower Boundary is expressed.

	There are three subject age units: Y - Years M - Months D - Days
Subject Age Upper Limit	The high boundary for the age range bracket that defines the sub-population.
Subject Age Upper Limit Units	The units in which the Subject Age Upper Boundary is expressed. There are three subject age units: Y - Years M - Months D - Days
Medical Condition	The medical condition in which the reference range applies.
Medical Condition Code List ID	If utilized, the code list identifier and version number for the Medical Condition code.

3.4.7 Unit of Measure (UOM)

The UOM level defines the unit of measure for the reference range and/or flag definition(s) that follow it.

The Unit System must always be populated. The actual UOM is an optional field so that a flag based on a text value (e.g. "Present") which has no UOM may be defined. If a reference range or flag is based on a test with numeric results, the UOM field should always be filled in.

The HL7 extension of the ISO 1000 standard is recommended as the basis for units of measure.

Variable	Description
Units System.	The UOM system whose reference ranges are being defined. There are three result classes: C - Conventional R - Reported SI - Standard International
UOM	The units of measure in which the reference range or flag is expressed.
UOM Code List ID	If utilized, the code list identifier and version number for units of measure.

The definitional blocks that follow (or are nested under, in XML) the Unit of Measure level are all optional. One or more definitions may be associated with each combination of Test, Subject Characteristics and Unit of Measure.

3.4.8 Normal Definition

The Normal Definition block defines the normal **range** for a specified test, sub-population and unit of measure when a test produces numeric results, or the normal **value** when a test produces text results. Thus, when this block is used, either the Normal Low and High fields, or the Normal Value field should have value, but not all three.

If a normal range or value is defined, the Normal Range Start Date and Time must be populated. The Normal Range End Date and Time need be populated only if the normal reference range has been edited and new values have taken effect.

Variable	Description
Normal Range Start Date and Time	The date and time at which the normal range definition that follows took effect. This date and time stamp is considered to be inclusive.
Normal Range End Date and Time	The date and time at which the normal range definition that follows was replaced and hence is no longer valid. This date and time stamp is considered to be inclusive.
Normal Comment	A comment that can further define the normal range or value being defined.
Normal Low	The value that defines the normal range low point for the test and subject characteristics.
Normal High	The value that defines the normal range high point for the test and subject characteristics.
Normal Value	The value that is the normal value for the test and subject characteristics. Used with text result tests when a specific text result is considered normal. If multiple normal values are used, a comma should be used to separate the values.

3.4.9 Delta Definition

The Delta Definition block defines a flag based upon the change that a current value represents from another specified value. This other value is defined within the Delta Base Value field. The “Custom” value type is for use when the comparison value is not a single Baseline or Prior Value (e.g. the “Baseline” is the average of two screening visits).

Several Delta definitions may be present for a specific test and set of subject characteristics – e.g. a test may have a change from baseline and a change from prior flag. In the ASCII or SAS implementations this requires a separate row for each definition; in XML all such definitions may be nested as repeated Delta Definition elements under the UOM element. The model allows a delta to be defined in either absolute or relative mode. A single delta definition (ASCII/SAS row or XML element) should contain only one mode.

If the Delta Definition block is used, the Delta Start Date and Time must be populated.

Variable	Description
Delta Start Date and Time	The date and time at which the delta definition that follows took effect.
Delta Comment	A comment that can further define the delta value being defined.
Delta Base Value	The value that defines the base from which the change is to be computed. There are three base types: B - Baseline C – Custom (Sender and receiver must have defined the Custom Base Value in a document outside this transmission format)

	(An example of a custom base value is when the Delta is defined as a percentage of the Low Normal range value) P – Prior (The most recent previous value)
Delta Minus Absolute	The absolute value to subtract from the base value to find the negative change (D-) flag point.
Delta Minus Relative	The percentage to use to compute the absolute value to subtract from the base value to find the negative change (D-) flag point.
Delta Plus Absolute	The absolute value to add to the base value to find the positive change (D+) flag point.
Delta Plus Relative	The percentage to use to compute the absolute value to add to the base value to find the positive change (D+) flag point.

3.4.10 Exclusion Definition

The Exclusion Definition block defines the exclusion **high and/or low value(s)** for a specified test, sub-population and unit of measure when a test produces numeric results, or the exclusion **value** when a test produces text results. Thus, when this block is used, either the Exclusion Low and High fields, or the Exclusion Value field should have value, but not all three.

If an exclusion flag is defined, the Exclusion Range Start Date and Time must be populated.

Variable	Description
Exclusion Start Date and Time	The date and time at which the exclusion definition that follows took effect.
Exclusion Comment	A comment that can further define the exclusion value being defined.
Exclusion Low	The value that defines the point at which a low exclusion flag (LX) shall be set.
Exclusion High	The value that defines the point at which a high exclusion flag (HX) shall be set.
Exclusion Value	The value that defines when an exclusion flag (EX) shall be set.

3.4.11 Alert Definition

The Alert Definition block defines the value at which a test produces a specified alert flag.. The Panic, Telephone and Reference values are intended for use with numeric results. The Abnormal value is intended for tests with a text result. Thus, when this block is used, either the Panic, Telephone and Reference fields, or the Abnormal field should have value.

If an alert flag is defined, the Alert Start Date and Time must be populated.

In the ASCII and SAS implementations, the Transaction Type resides as the last field in the record and must be populated to reflect the type of transaction for the record. In the XML implementation, the Transaction Type resides in each definitional block (Normal, Delta, Exclusion and Alert) and must have a value for each definitional block that is used.

Variable	Description
Alert Start Date and Time	The date and time at which the alert definition that follows took effect.
Alert Comment	A comment that can further define the alert flag set being defined.
Panic Low	The value that defines the point at which a low panic alert flag (LP) shall be set.
Telephone Low	The value that defines the point at which a low telephone alert flag (LT) shall be set.
Reference Low	The value that defines the point at which a low normal alert flag (LN) shall be set.
Reference High	The value that defines the point at which a high normal alert flag (HN) shall be set.
Telephone High	The value that defines the point at which a high telephone alert flag (HT) shall be set.
Panic High	The value that defines the point at which a high panic alert flag (HP) shall be set.
Abnormal	The value that defines when an Abnormal alert flag (AB) shall be set.
Transaction Type	This indicates what type of record the data record is and consequently how it should be processed when it is imported into the study database. There are 4 transaction types: M - Remove (existing record) I - Insert (new record) R - Retransmit (existing record without changes) U - Update (or revise existing record at result record)

The CDISC Lab team recognizes that data recipients may wish to confirm that a file has not been corrupted or truncated during electronic transmission. Various technologies (such as check sums) exist to perform this task and do not require fields or rows be added to the data model for this purpose. The CDISC Lab team recommends the use of these technologies, and has not specified any End of File markers or row counters in its flat file implementations.

3.5 Comments on Practical Application

3.5.1 Use of Model Fields

The model is designed to support the requirements of many different organizations within the industry and many different types of clinical trials. Accordingly, while many of the fields it contains apply to all clinical trials there are some, which do not because they have special uses. This means that for any given study it should be expected that, while most of the fields certainly will be used, some certainly would not be.

The same is true of the various definitional blocks. Any particular study may define only one or two types of ranges or flags, and hence will not include the others.

3.5.2 Data Compression

For the purposes of physically transferring and storing data files made using the model, tests have shown that typically compression rates of approximately 95% can be achieved.

3.5.3 Transmission Agreements

The CDISC LAB Model is applicable for use for data interchanges between sender and receiver but is not a “plug-in” that covers all data requirements. Implementation of the LAB Reference Range Model requires adoption of a TRANSMISSION AGREEMENT between sender and receiver.

A transmission agreement would cover the following types of information:

- Implementation type (e.g., bar delimited ASCII, SAS, XML)
- Use of standardized text fields
- Code lists
- Local test codes
- Interpretation of reference ranges (e.g., inclusive or exclusive of upper and lower limits)
- How to represent subject age brackets (e.g., years, months or days)

4 Review Process

All CDISC models follow the CDISC Standards Development Process, which includes multiple levels of review by internal CDISC teams, focused external groups, and the public. For more details of this process, see the CDISC Standards Development Process document (CDISC-COP-001). This model is currently at its public review stage.

5 Directions for Submitting Comments

Comments about this model can be submitted by using the Public [Discussion Forum on the CDISC website](#).

We invite those interested in the model, implementing the model for the first time or using the model in production to post their comments, questions or suggestions. We hope to create an active forum used to foster discussion, to provide assistance and to share experiences. Posted comments will be reviewed on an ongoing basis by members of the LAB team. The discussions will not only assist model users but will guide the LAB team in updating the model as needed and in creating training materials and educational programs.