PhUSE 2014

Paper RG04

Submitting Summary Level Site Data: Overview, Implementation, and Case Studies

Rob Woolson, Rho Inc., Chapel Hill, NC USA
Jeffrey Abolafia, Rho Inc., Chapel Hill, NC USA
Ben Vaughn, Rho Inc., Chapel Hill, NC USA

ABSTRACT

In December 2012 the US FDA issued draft guidance for submitting Bioresearch Monitoring (BIMO) Summary Level Site Data for the Center for Drug Evaluation and Research's (CDER) Inspection Planning. The guidance calls for submitting aggregate results-level data for each site enrolling subjects in pivotal trials in a clinical development program, in addition to the subject-level data we currently submit. The FDA is beginning to request this data and the time and resources required to create them are not trivial (3). In the past year our organization has received several requests from the FDA for Summary Level Site Data (BIMO dataset).

This paper presents an overview of the BIMO dataset and implementation issues that arise from its creation. The paper divides the discussion into four major sections. The first section provides an overview of the guidance pertaining to BIMO data. The second section describes the input sources and processes that are required to create the BIMO dataset. The third section presents two case studies outlining our experiences and lessons learned. The last section provides a few recommendations related to clarifying some elements of the BIMO dataset and supporting information.

INTRODUCTION

In 2012 the FDA Safety and Innovation Act (FDASIA) and the fifth reauthorization of the Prescription Drug User Fee Act (PDUFA V) were authorized by the US Congress. One of the key performance goals of PDUFA V was enhancing benefit-risk assessment for marketing applications. As a result, CDER recently initiated a pilot program to evaluate a risk-based modeling approach to selecting clinical investigator sites for inspection. This approach requires that CDER have sufficient data to select sites for inspection. In December 2012 the US FDA issued two draft guidances that facilitate providing CDER the data required for the timely selection of inspection sites. The first guidance, titled "Providing Submissions in Electronic Format – Summary Level Clinical Site Data for CDER's Inspection Planning" (1) outlines the FDA's recommendation that sponsors submit summary level clinical site data in standardized format. The second guidance, titled "Specifications for Preparing and Submitting Summary Level Clinical Site Data for CDER's Inspection Planning" (2) provides the technical specifications and a data model for creating the BIMO dataset. These guidances apply to new drug applications (NDAs), biologics licensing applications (BLAs), and NDA and BLA supplemental applications containing new clinical study reports that are submitted to CDER.

This data model is unique in that it supports the presentation of aggregate results-level data for each site in a clinical development program rather than subject-level data. Constructing this dataset can add significant time and cost to preparing the data package that accompanies a regulatory submission (3). In the following sections we look at the structure and content of the BIMO dataset, the inputs and processes to create this dataset, case studies and lessons learned, shortcomings of the data model, and recommendations for improving the model.

BIORESEARCH MONITORING (BIMO) SUMMARY LEVEL CLINICAL SITE DATA

DESCRIPTION

The technical guidance (2) calls for submitting a single dataset that includes data from all pivotal studies used to support safety and efficacy in an application. The dataset should be named "clinsite.xpt" and should be submitted in SAS® transport file format. The structure of the dataset is one record per pivotal study, clinical site, treatment arm, and primary endpoint for the intent-to-treat (ITT) population. The technical guidance provides a standard data model for constructing this dataset. It includes a list of variables which must be included, their attributes, controlled terminology, and the metadata required to document each variable. The specification contains variables to 1) identify and characterize each study; 2) identify the sponsor; 3) document the NDA or BLA number; 4) describe individual clinical investigator sites, including financial disclosure information; 5) identify treatment arm; 6) document endpoint types; and 7) report primary endpoint results by site and treatment arm.

SUBMISSION CONSIDERATIONS

For applications submitted in the eCTD format, the BIMO dataset should be placed in Module 5 as specified below:

It should be accompanied with a data definition file or "define file" in pdf format. The technical guidance includes a sample BIMO dataset as well as a sample define file.

Since the impetus for submitting the BIMO dataset is a more timely approach to site selection, sponsors should carefully consider several options for when to submit this data. The optimal strategy would be to provide the BIMO dataset for a pre-NDA or pre-BLA meeting, and ask the FDA if it agrees with the format and content. If this isn't feasible, a second option would be to submit the BIMO dataset as soon as possible after the pre-NDA or pre-BLA meeting. Lastly, one may wait until the time of the submission to provide the dataset. In all cases, the BIMO dataset should be included with the NDA/BLA submission at the path noted above. If any of the studies have data that present issues, we strongly recommend discussing the BIMO dataset with the FDA in a pre-submission meeting.

INPUT SOURCES AND PROCESSES

ONE DATASET – MANY GOALS, MANY INPUT SOURCES

The risk-based model mentioned above uses the BIMO dataset as input. The BIMO dataset is intended to describe and summarize the characteristics and outcomes of clinical investigations at the individual study site level across a clinical development program. More specifically:

> The summary level clinical site dataset is intended to characterize the individual clinical investigator sites, describe aspects of the studies with which those clinical investigator sites are associated, and present the characteristics and outcomes of the study at the site level. (1)

Given the wide array of data that are presented in a single location, it is not surprising that many sources of data are used to create the dataset. The sources of data include, but may not be limited to:

- SDTM (or raw) datasets from multiple studies
- ADaM (or legacy analysis) datasets from multiple studies
- clinical trials management system (CTMS) output
- financial disclosure forms and databases, and
- regulatory affairs databases (2,3).

The information requested in the BIMO dataset and the likely origin of each are summarized in Table 1.

TABLE 1: INFORMATION REQUESTED IN BIMO DATASET AND THE (LIKELY) DATA SOURCE/ORIGIN

Data Source/Origin	Information	Comment(s)
Study-level metadata	Study Number; Title, Number of Sponsors; Sponsor Name(s); IND Number; Study Conducted Under IND?	
Regulatory affairs database	NDA Number; BLA Number; Supplement Number	
Raw/SDTM/Legacy Analysis/ADaM dataset(s)	Site Identifier; Treatment; Number of Subjects Enrolled; Subjects Screened; Subjects Discontinued; Treatment Efficacy Endpoint Estimate; Treatment Efficacy Endpoint; Site-specific Treatment Effect (and variation); Number of Censored Observations; Number of non-SAEs Number of SAEs; Deaths	Data are most likely contained in ADaM efficacy and safety datasets. However, there are scenarios in which the site-specific analysis results requested may not have been conducted as a part of the planned statistical analysis (see Case Study 1, below).
Protocol/Statistical Analysis Plan	Text description of Primary Endpoint(s); Primary Endpoint Variable Type.	
None – hardcode per guidance (2)	Domain Abbreviation	Guidance requests that 'DE' be used as the domain abbreviation.
CTMS (or other separately maintained database)	Number of Protocol Violations	Protocol violation information is frequently captured in the SDTM DV dataset. However, for older legacy studies, that may not be the case.
FDA 1572 forms	Maximum financial disclosure amount by any single investigator by site; total financial disclosure by site (for all principal and sub-investigators); investigator names, phone numbers, fax numbers, email addresses, and postal address information (country, state, province, postal code, street, etc.)	Data may be contained in a regulatory affairs database, but often it is not.

Not only are there several data sources, but it is also likely that:

- the data sources listed above are maintained by different functional groups (e.g., regulatory, biostatistics, clinical monitoring);
- the different functional groups have different standards for data accuracy, validation, and quality control;
- the functional groups are a part of different organizations (e.g., independent consultants, CROs, sponsor,); and/or
- multiple organizations may have overlapping responsibilities within the same clinical development program.

A proposed process for assembling and summarizing this information is detailed in the following sections.

PLAN AHEAD – THIS IS NOT YOUR 'STANDARD' STANDARDIZED DATASET

As noted above, there are numerous input sources that are potentially stored and maintained in varying formats, with varying levels of accuracy, by several functional groups. As such, preparing, assembling, summarizing, and verifying the accuracy of the cadre of information required to create the BIMO dataset is a process that should be undertaken as early as possible in marketing application development.

Based on our recent experiences, we think some essential steps to successfully creating a submission-ready BIMO dataset are:

1. Start the data assembly process as early as possible.

   Regardless of the steps one chooses to take in creating the BIMO dataset, start the process as early as possible. This is particularly true in large, multi-center, and/or international clinical development programs.

Starting the process at the time of final site selection in the pivotal trial(s) is best. Much of the essential regulatory information is not only available at this time, but can also be consolidated into a 'digestible' format for later programing. In addition, this increases the likelihood that the BIMO dataset will be available for the pre-NDA or pre-BLA meeting.

2. Identify and select a BIMO dataset coordinator.

   Given the numerous functional areas (and companies) that may need to contribute information to the BIMO dataset, there should be a single BIMO dataset coordinator.

3. Create and maintain a data/information source inventory and action log.

   There are more than 30 data elements included in the BIMO dataset. Creating and maintaining a data/information source inventory log can be useful.

   For each data element and each study, collecting the following information is recommended:

   - Requisite source data

     As noted above, input data can come from many sources and many functional areas. Take an inventory of the data that are needed to create the BIMO dataset elements for each study. For a small clinical development program, the list of requisite data may be short and may be maintained within a single organization. However, this is often not the case.

   - The functional area and owner of the data

     Identify the group (and individual) responsible for creating and maintaining the data source(s). In some cases, the group responsible may not be part of a functional area accustomed to creating, maintaining, and verifying the accuracy of data to be submitted to the FDA. Establishing contact early can help get a jump on understanding and, if necessary, updating the current source data.

   - Source data status

     After listing all of the source data required to create the BIMO dataset, the BIMO dataset coordinator should evaluate the status of each source data element. In general, the source data elements fall into one of the following categories:

     - the source data are validated and are well-suited for inclusion in the BIMO dataset (e.g., adverse event data by site) and no additional validation or quality control (QC) activities are required;

     - the source data are validated and are well-suited for inclusion in the BIMO dataset, however, the analyses requested in the BIMO dataset may not have been previously performed (e.g., site-level analysis of the primary endpoint when the primary analysis does not support by-site treatment estimates);

     - the source data are in a non-validated database (e.g., a regulatory affairs database using Excel or Access); or

     - the source data do not exist.

   After collecting the above information, establishing the following elements of a data format and/or analysis validation and QC action plan is recommended:

   - Planned data format

     Select an acceptable data format for the data element (e.g., Access, Excel, SAS, etc.). Discussions with the individual(s) responsible for programming the BIMO dataset can be useful.

   - Required additional analyses

     Identify analyses that were not conducted as a part of the original, planned analyses.

- Production, validation, and QC plan

  A production, validation, and/or QC plan may need to be established for data elements that require a new data format, are housed in a non-validated database, or have not previously been entered into a database.

- Production, validation, and QC plan owner

  Identify the individual responsible for overseeing the completion of production, validation, and/or QC activities of the source data element.

A sample of two BIMO data elements from a data/information source inventory and action log is shown in Table 2:

TABLE 2: SAMPLE OF DATA ELEMENTS FROM DATA/INFORMATION SOURCE INVENTORYAND ACTION LOG

Variable Name	PROTVIOL		FINLDISC	
Variable Label	Number of Protocol Violations		Disclosure Amount	
Description	Number of protocol violations at a given site by treatment arm. This value includes multiple violations per subject and all violation types (i.e., not limited to only significant deviations).		Total financial disclosure amount (USD) by site calculated as the sum of disclosures for the principal investigator and all sub-investigators to include all required parities. Under the applicable regulations (21 CFR Parts 54, 312, 314, 320, 330, 601, 807, 812, 814, and 860).	
Study	xx-03, xx-04, xx-05	xx-01, xx-02	xx-03, xx-04, xx-05	xx-01, xx-02
Functional Area (Owner)	Biostatistics (Person 1 – Company x)	Clinical Monitoring (Person 2 - Company y)	Regulatory Affairs (Person 3 – Company z)	Regulatory Affairs (Person 4 - Sponsor)
Current Data Format	Protocol deviation data for studies xx-03 and xx-04 are in the SDTM DV domains.	Sponsor only collected and reported protocol deviation data for legacy studies xx-01 and xx-02 in a CTMS database.	Financial disclosure information has not been consolidated into a database. Information is contained in scanned .pdf FDA 1572 forms.	Financial disclosure information was keyed into the Excel spreadsheet located here (file path).
Planned Data Format	SDTM DV data from studies xx-03, xx-04, and xx-05 will be used as input for the BIMO dataset. No change.	SAS dataset	Excel spreadsheet	No update; Excel spreadsheet format can be used as input into BIMO dataset
Production / Validation / QC Plan (Owner)	Validated via applicable SOPs. (Person 1 – Company x)	An interim SAS dataset will be created. CTMS will be merged with existing legacy datasets to obtain key subject and site information. Production and validation will be	Financial disclosure information by site, PI, and sub-PI will be keyed into the Excel spreadsheet template located here	An independent QC of the keyed values will be conducted. (Person 5 – Company z)

TABLE 2: SAMPLE OF DATA ELEMENTS FROM DATA/INFORMATION SOURCE INVENTORYAND ACTION LOG

Variable Name	PROTVIOL		FINLDISC	
		conducted by statistical programming group. Specifications for interim protocol deviation dataset are found here (file path). (Person 1 – Company x).	(file path). The keyed values will be independently QC'd based on the abbreviated QC plan located here (file path). (Person 3 – Company z)	
Completion Date	N/A	mm/dd/yyyy	mm/dd/yyyy	mm/dd/yyyy
Status	Complete	Validation in progress.	Data entry in progress. Check in mm/dd/yyyy.	Independent comparison of spreadsheet to master site list, PI list, and sub-PI list and to FDA 1572 forms ongoing. Expected completion date mm/dd/yyyy.

4. Create BIMO dataset programming specifications

   Specifications should be created that describe the programmatic steps required to pull together all of the data sources identified in the inventory and action log.

5. Program, validate, and QC the BIMO dataset

   With the previous steps completed, the BIMO dataset can be programmed and independently validated. Although it is our experience that the programming and independent validation process is similar to that of most other 'standard' datasets, the QC process is more involved. Some points of emphasis in the BIMO dataset QC process are as follows:

   - Verification of the format of the dataset per the guidance (2) – are the appropriate number of records included in the dataset (e.g., Number of Studies x Number of Sites x Number of Primary Endpoints x Number of Treatment Arms) and are the appropriate fields populated?

   - Confirmation that records are associated correctly – the numerous data sources, data source formats, and functional areas/groups responsible for the creation of the source data can easily produce programming mistakes. Are the right study/site/endpoint records from each data source correctly connected?

   - Validation that efficacy and safety-related statistics are calculated correctly – are the statistics contained in the study-level CSRs (if previously computed) the same as the statistics contained in the BIMO dataset? If the statistics in the BIMO dataset have not previously been computed, are they accurate?

   - Regulatory-related QC – are all investigators included, is the contact information correct, are the financial disclosure amounts correct, etc.?

6. Prepare for additional FDA requests

   The FDA has consistently requested supportive by-study, by-site patient level listings and other supportive data summaries. Sponsors should anticipate additional FDA requests and allocate resources to comply.

RECENT EXPERIENCES AND CASE STUDIES

The following sections describe two recent experiences related to BIMO dataset construction, including challenges and a brief discussion.

CASE STUDY 1

Clinical Development Program description: This was a multi-national, multi-center development program including three phase II studies and four phase III studies. Data were submitted as a part of a 505(b)(1) marketing application.

Data source status: Source data included raw data, legacy analysis data, SDTM datasets, ADaM datasets, and a regulatory affairs database. In some cases, data had not been assembled into a database suitable for use in programming.

Challenges: Challenges encountered while creating the BIMO dataset included the following:

- Four different CROs were responsible for executing the seven phase II and phase III studies.
- Data were in several different formats. No regulatory database existed for two of the studies.
- The primary analysis of the primary endpoint utilized an analysis method which did not support by-site estimates.
- Although the guidance states that only 'pivotal' studies be included in the BIMO dataset (2), subsequent FDA communication requested that phase II and phase III studies be included.
- A request was made by the FDA to also include 10 different by-study, by-site patient level listings with the BIMO dataset and documentation.

Discussion: The primary challenges in this example were: 1) assembling all required data and identifying missing data; 2) incorporating primary analyses results which did not support by-site analyses; and 3) complying with updated FDA requests.

Data needed to create the BIMO dataset were held by various CROs, the sponsor, and independent consultants. Locating the requisite data and identifying data that were missing was a time consuming process, in particular since some of the studies were completed several years earlier. Some data were stored in paper files and required manual entry and verification.

The primary analyses of some of the studies included in the BIMO dataset did not support by-site treatment estimates. A related, though different, analysis approach was used to create by-site treatment and treatment difference estimates. The different approach was described in the associated BIMO dataset documentation.

The FDA-requested BIMO patient level listings required additional programming, validation, and QC time. New listings were created as the listings created to support the study-level CSRs did not match the format and content of the FDA-requested BIMO listings.

CASE STUDY 2

Clinical Development Program description: This was a multi-center development program including four phase III studies. Data were submitted as a part of a 505(b)(2) marketing application.

Data source status: Source data included SDTM datasets, ADaM datasets, and a regulatory affairs database.

Challenges: Challenges encountered while creating the BIMO dataset included the following:

- Two different CROs were responsible for executing the phase III studies.
- The primary analysis of the primary endpoint was non-parametric.

- A request was made by the FDA to also include 10 different by-study, by-site patient level listings with the BIMO dataset and documentation.

Discussion: The two primary challenges in this example were: 1) mapping non-parametric analysis results into the BIMO dataset specifications; and 2) creating by-study, by-site patient level listings.

At present, the BIMO dataset specifications assume parametric results, including means, rates, and associated measures of variability. In an effort to provide measures of central tendency related to treatment, median values were provided. Similarly, in an effort to characterize treatment difference, the Hodges-Lehman-Sen shift statistic was provided. Because measures of variability were not easily applicable to the shift statistic, the associated p-value was reported instead. These deviations from the BIMO dataset specifications were included in the associated documentation.

Unlike Case Study 1, the patient level listings used to create the study-level CSRs were already created in a format that could easily be adapted to the format requested by the FDA. Though in a similar format, an extensive table of contents had to be created for each study, site, and listing type in order to comply with the FDA's request.

A FEW RECOMMENDATIONS

Although the BIMO dataset specifications and guidelines generally provide a clear roadmap for its creation, the following are a few recommendations related to clarifying some elements of the BIMO dataset and supporting information:

- Guidelines should be created to accommodate primary analyses that are non-parametric.
- If required, supportive by-study, by-site patient level listings should be included as an element in the relevant guidance documents.
- Additional specifications should be created to clarify expectations for:
    - missing or non-missing values (e.g., some data elements are site specific, but not treatment specific);
    - placeholder records for arms that do not occur at a given site
    - studies with multiple phases and/or multiple treatment comparisons; and
    - characterizing safety and efficacy data in cases where the safety and efficacy populations are not the same (e.g., subject data presented 'as randomized' in efficacy data elements, but presented 'as treated' in safety data elements).
- BIMO dataset metadata should comply with standardized metadata guidelines.

CONCLUSION

It is likely that the BIMO dataset will continue to be a part of FDA marketing applications for the foreseeable future. Construction of the site-level BIMO dataset for a clinical development program can be a time consuming process that potentially involves numerous service providers, functional areas, and input sources. CROs and other organizations responsible for its assembly should begin preparing the dataset as early as possible in the marketing application process. Successfully executing all of the activities involved in BIMO dataset submission, from identifying and assembling all of the relevant information (or perhaps confirming its absence) through BIMO dataset production and QC, require a well-organized plan.

REFERENCES

(1) Guidance for Industry: Providing Submissions in Electronic Format – Summary Level Clinical Site Data for CDER's Inspection Planning (FDA DRAFT Guidance: December 2012).
(2) Specifications for Preparing and Submitting Summary Level Clinical Site Data for CDER's Inspection Planning (FDA DRAFT Guidance: Technical Specifications, November 2012).
(3) Johnson, James, "Evolving Regulatory Guidance on Submission of Standardized Data". 2013. Annual PhUSE Conference.

ACKNOWLEDGMENTS

We would like to thank our colleagues Carol Baker and Frank Dilorio for their valuable input during preparation of this paper.

CONTACT INFORMATION

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

Rob Woolson
Rho, Inc.
Rob_woolson@rhoworld.com

Jeffrey Abolafia
Rho, Inc.
Jeff_abolafia@rhoworld.com

Ben Vaughn
Rho, Inc.
Ben_vaughn@rhoworld.com