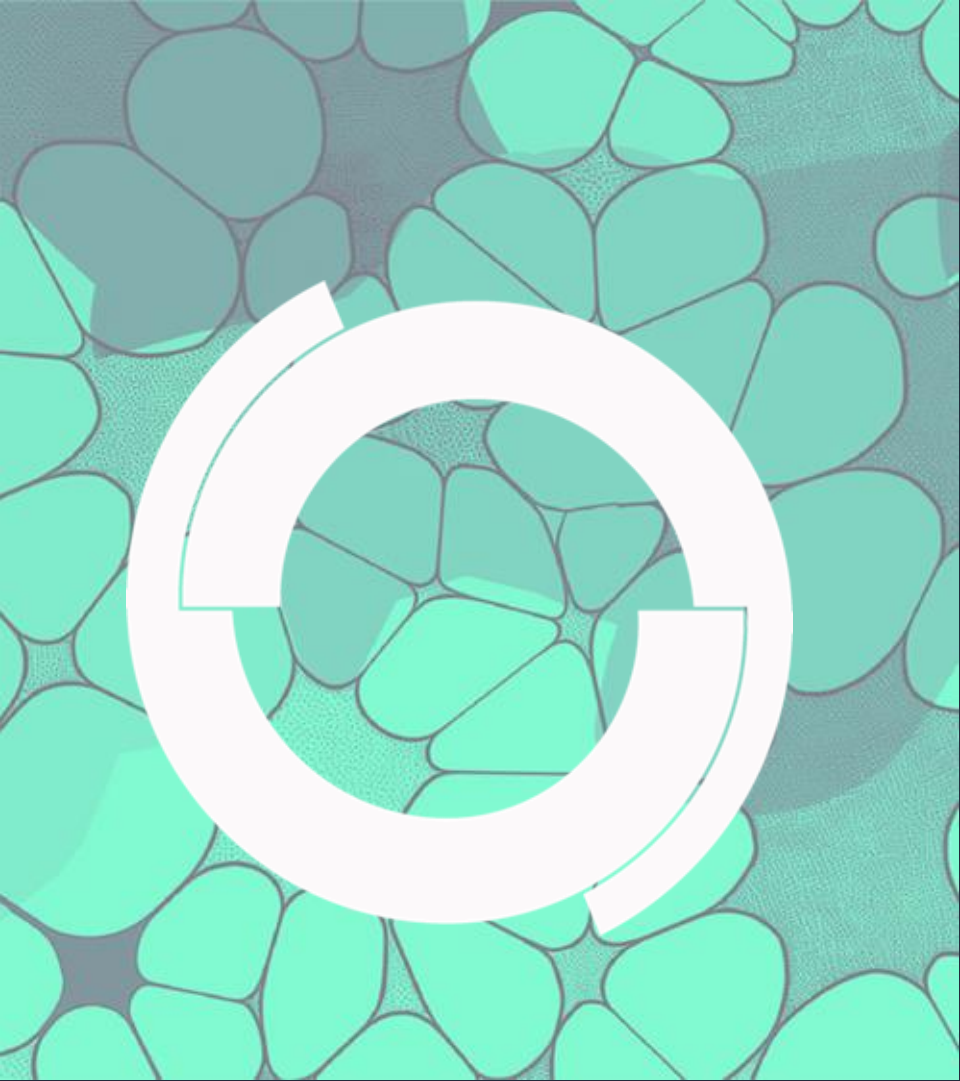


Common Issues in BIMO Clinical Site Dataset Packages



Michael Beers is the Director of Consultative Services at Certara, where he spends his days helping clients use Pinnacle 21 Enterprise, managing the validation rules and versions of standards, and working as a Data Fitness Analyst at the FDA supporting their use of Pinnacle 21 Enterprise (DataFit). He has more than 15 years of experience working with clinical trial data, and as a CDISC Consultant, at both pharmaceutical companies and CROs.



Common Issues in BIMO Clinical Site Dataset Packages

Michael Beers | March 2026

Agenda

- Introduction
- Missing Information
- Unexpected Location in eCTD
- Incorrect Usage of Variables
- Traceability Issues
- Conclusion

Introduction

Introduction

- What is BIMO and Why is it Important?

Bioresearch Monitoring Program Information

Program Information

FDA's Bioresearch Monitoring (BIMO) program is a comprehensive program of on-site inspections, data audits, and remote regulatory assessments designed to monitor all aspects of the conduct and reporting of FDA regulated research. The BIMO program was established to assure the quality and integrity of data submitted to the agency in support of new product approvals and marketing applications, as well as, to provide for protection of the rights and welfare of the thousands of human subjects and animals involved in FDA regulated research. It has become a cornerstone of the FDA preapproval process for new medicines, medical devices, food and color additives, veterinary products and, tobacco products introduced to the U.S. consumer.

Introduction

- What is a BIMO CLINSITE dataset?

III. SUMMARY-LEVEL CLINICAL SITE DATASET

A. Organization of the Site-Level Dataset

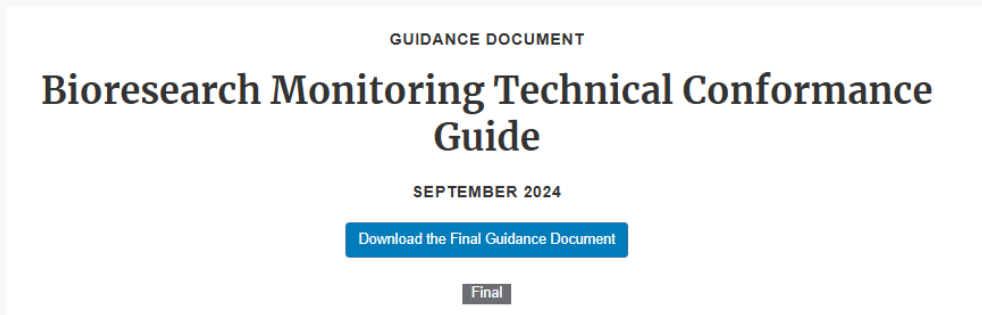
A single summary-level clinical site dataset that contains data from all major (i.e., pivotal) studies used to support safety and efficacy in the application, including studies with different treatment indications, should be provided.

For each major (i.e., pivotal) study used to support safety and efficacy, data by clinical site and treatment arm for the safety population (SAFPOP) and primary efficacy population (EFFPOP) should be provided.

For clinical investigator sites involved in multiple studies in support of an application, the site data should be reported independently for each study within the dataset.

Introduction

- BIMO CLINSITE Guidance:



- BIMO TCG Contains:

- Guidance for critical variables
- Expected eCTD folder structure
- Expected file formats
- Specification for CLINSITE variables
- Examples of (fictional) CLINSITE datasets

B. Variables and Variable Names for Site-Specific Efficacy Results

C. eCTD Folder Structure for Summary-Level Clinical Site Dataset

D. File Format

APPENDIX 3: CLINICAL SITE DATA ELEMENTS SUMMARY LISTING

APPENDIX 4: EXAMPLES

Missing Information

Missing define.xml

- A define.xml is required, per the BIMO TCG:

The summary-level clinical site dataset should be accompanied by a data definition file. The contents of the define file for a dataset and fictional examples are presented in Appendix 3 and Appendix 4 of this Guide.

- A define.xml is necessary to describe the data. Example:

13	SAFPOP	Number of Subjects in Safety Population	Num	Integer	Total number of subjects in safety population at a given site by treatment arm. When a subject has transferred from one site to another, the applicant should handle reporting of such subjects consistently across sites and include in the define file the reporting convention used. The applicant may opt to further explain the reasons subjects transferred between sites in the BIMO Data Reviewer's Guide, if a guide will be provided.	20
----	--------	---	-----	---------	---	----

Incomplete define.xml Documentation

- The define.xml should provide clear transparency into variable origins, comments, and methods.
 - Comments/Derivations should be specific, not general (general = mostly-useless)
- An example is a study where all records in the CLINSITE dataset had SAFPOP = 0, which seems unusual.
 - The define.xml uses a very general comment, which does not provide clarity
 - Note: No reviewers guide was provided either

Variable	Label / Description	Type	Length or Display Format	Controlled Terms or ISO Format	Origin / Source / Method / Comment
SAFPOP	Number of Subjects in Safety Population	integer	8		Assigned Total number of subjects in safety population at a given site by treatment arm.

Missing Reviewers Guide

- The BIMO TCG does mention using it to further explain and document
 - Not a requirement though, as it states “...*the BIMO Data Review Guide when one is provided...*”
- The BIMO TCG provides a recommendation for a template
 - PHUSE (BIMO) Bio-research Monitoring Data Reviewer's Guide
 - <https://advance.hub.phuse.global/wiki/spaces/WEL/pages/26804785/BIMO+Bio-research+Monitoring+Data+Reviewer+s+Guide>
 - Includes:
 - PHUSE Project_BDRG_Introduction and Overview
 - BDRG_Completion Guidelines
 - BDRG_Template
 - 3 examples
- The BDRG provides a section to explain any conformance issues
 - Best practice: explain any conformance issues in the reviewer's guide preemptively

Missing Variables

- Not Applicable Variables
 - Examples: CENSOR1 (Censored Observations in SAFPOP) and CENSOR2 (Censored Observations in EFFPOP)
 - Only to be populated when ENDPTYPE (Primary Endpoint Type) equals “time to event”. BIMO TCG states: *If not applicable, leave blank*
- Critical Variables
 - Examples: ENDPOINT (Primary Endpoint), ENDPTYPE, TRTEFFR1 (Treatment Efficacy Result for SAFPOP), TRTEFFR2 (Treatment Efficacy Result for EFFPOP), etc.
 - Its important to ensure these aren't dropped
- Recommendation
 - Always include all of the variables specified in the BIMO TCG
 - If a variable is not applicable, it should still be included, and guidance should be followed (such as leaving it blank)
 - The define.xml should include the reason why the variable is blank for all records

Missing Data Values

- ARM Variable with Null Values
 - Usually due to there only being screen failure subjects at a site
 - In this case, ARM is typically null but SCREEN (Number of Subjects Screened) is populated
 - Recommendation: ARM should not be left null in this case
 - BIMO TCG states:

Variable Index	Variable Name	Variable Label	Type	Controlled Terms or Format	Notes or Description	Sample Value
11	ARM	Description of Planned Treatment Arm	Char	String	Plain-text label for the name given to an arm or treatment group as referenced in the clinical study report (limit 200 characters). When no arm or treatment group is available due to only screen failure subjects at site, use label "Screen Failure."	Active name and dose (e.g., "Active 25mg"), Comparator product name (e.g., "Drug x"), Placebo, Screen Failure

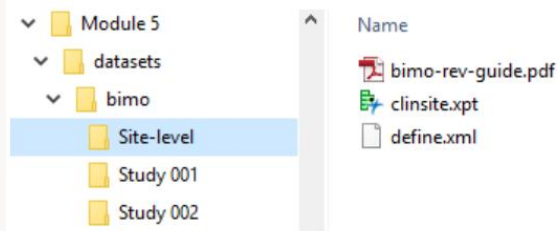
Unexpected Location in eCTD

Missing Data Unexpected Location in eCTD

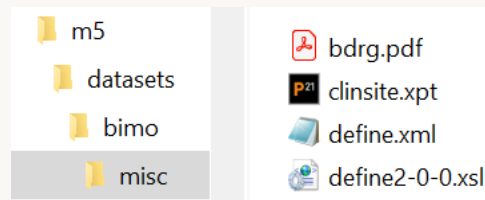
- Expected Location:
 - m5/datasets/bimo/site-level

Within the eCTD folder structure, place the site-level dataset define file and BIMO Data Reviewer's Guide, if it is being submitted, in the M5 folder as follows:

Figure 2: Place the Site-Level Dataset Define File and BIMO Data Reviewer's Guide in the M5 Folder

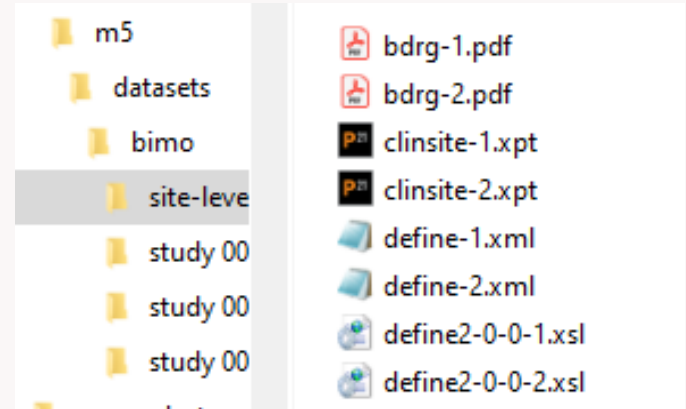


- Common Unexpected Location:
 - m5/datasets/bimo/misc



Missing Data Unexpected Location in eCTD

- Other Common Unexpected Folder Structure:
 - m5/datasets/bimo/misc/site-level
 - m5/datasets/bimo/misc/*study*
 - separate CLINSITE datasets/documentation for different studies in separate folders
 - m5/datasets/bimo/listings
 - Same folder, multiple versions of dataset/documentation:



- Recommendation: use exact folder structure listed in the BIMO TCG

Incorrect Usage of Variables

Using Old Versions of BIMO Guidance

Changes from BIMO 1.0 to BIMO 2.0:

- Variables Removed:
 - PROTVIOL (Number of Protocol Violations)
 - SITEEFFE (Site-Specific Treatment Effect)
 - SITEEFFS (Site-Specific Treatment Effect Standard Deviation)
- Variables Added:
 - COHORT (Description of Planned Cohort)
 - IMPDEV (Number of Important Protocol Deviations)
 - NOIMPDEV (Number of Non-Important Protocol Deviations)
- Variables Changed:
 - STUDYTL (Study Title) changed to TITLE
 - SPONNAME (Sponsor Name) changed to SPONSOR

Changes from BIMO 2.0 to BIMO 3.0:

- Variables Removed:
 - CENSOR (Number of Censored Observations)
 - TRTEFFR (Treatment Efficacy Result)
 - TRTEFFS (Treatment Efficacy Result STD)
- Variables Added:
 - EFFPOP (Number of Subjects in Efficacy Population)
 - TRTEFFR1 (Treatment Efficacy Result for SAFPOP)
 - TRTEFFR2 (Treatment Efficacy Result for EFFPOP)
 - CENSOR1 (Censored Observations in SAFPOP)
 - CENSOR2 (Censored Observations in EFFPOP)
- Variables Changed:
 - COUNTRY variable now uses GENC

Recommendation: always use the latest version of BIMO TCG

Invalid Data Values

- STATE Variable – Common Scenarios of Invalid Values
 - COUNTRY = USA, but STATE is missing or populated with “NA”
 - Example: USA = COUNTRY, STATE = NA, CITY = null, POSTAL is populated

COUNTRY	STATE	CITY	POSTAL
USA	NA		07960
USA	NA		65212
USA	NA		14225
USA	NA		11530
USA	NA		11735
USA	NA		21287
USA	NA		21076

- STATE values not using GENC codelist values

STATE	CITY	POSTAL
NY	New York	10065
FL	Tampa	33612

Invalid Data Values

- STATE Variable – Common Scenarios of Invalid Values
 - Incorrect values in the STATE variable
 - Examples: city mistakenly used instead, misspellings, or leading spaces

STATE	CITY	POSTAL
Inidiana	Indianapolis	46202
Orlando	Portland	97210

- STATE incorrectly populated with Puerto Rico
 - COUNTRY should be PRI and STATE should be left null.

NCIt Concept Code	NCIt Preferred Term	GENC Name (FDA Standard)	GENC 2 Letter Code	GENC 3 Letter Code (FDA Standard)	GENC Number	NCIt Subset Code	NCIt Subset Name
C17043	Puerto Rico	PUERTO RICO	PR	PRI	630	C124085	Geopolitical Entities, Names and Codes Terminology

Invalid Data Values

- SCREEN Variable
 - How its supposed to be implemented:
 - BIMO TCG states: *Total number of subjects screened (and consented) at a given site (overall number per site as subjects have not yet been assigned to treatment arm).*

Table C: Example for Clinical Site Data Elements Summary Listing with One Endpoint

STUDYID	TITLE	SPONCNT	SPONSOR	IND	UNDER-IND	NDA	BLA	SUPP-NUM	SITEID	ARM	COHORT	SAFPOP	EFFPOP	SCREEN	DISCSTUD	DISCRT
ABC-123	Double blind...	1	DrugCo, Inc.	000001	Y	200001	-	-	001	Active	*	26	54	61	3	2
ABC-123	Double blind...	1	DrugCo, Inc.	000001	Y	200001	-	-	001	Placebo	*	25	54	61	4	1
ABC-123	Double blind...	1	DrugCo, Inc.	000001	Y	200001	-	-	002	Active	*	23	44	54	2	1
ABC-123	Double blind...	1	DrugCo, Inc.	000001	Y	200001	-	-	002	Placebo	*	25	43	54	4	3
ABC-123	Double blind...	1	DrugCo, Inc.	000001	Y	200001	-	-	003	Active	*	27	55	62	3	0
ABC-123	Double blind...	1	DrugCo, Inc.	000001	Y	200001	-	-	003	Placebo	*	26	56	62	5	3

Invalid Data Values

- SCREEN Variable
 - Common incorrect implementations:
 - Values in SCREEN differ for records for the same SITEID
 - Typically due to using number of screened subjects for each ARM instead
 - Screen failure records are not included in the count for SCREEN.

SITEID	ARM	COHORT	SAFPOP	EFFPOP	SCREEN
01	Screen Failure		•	•	1
11	ABC123		5	5	5
11	DEF456		4	4	4
12	ABC123		1	1	1
13	ABC123		2	1	2
13	DEF456		1	1	1
14	ABC123		6	5	6
14	DEF456		4	4	4

Invalid Data Values

- TRTEFFR1 & TRTEFFR2 populated when SAFPOP/EFFPOP = 0
 - TRTEFFR1 (Treatment Efficacy Result for SAFPOP) & TRTEFFR2 (Treatment Efficacy Result for EFFPOP)
 - Typically due to values of 0 for TRTEFFR1/TRTEFFR2 (when SAFPOP/EFFPOP = 0)

SITEID	ARM	COHORT	SAFPOP	EFFPOP	SCREEN	DISCSTUD	DISCRT	ENDPOINT	ENDPTYPE	TRTEFFR1	TRTEFFR2
100	GHI654	Cohort 1	1	1	1	0	0	Rate of episodes ...	Continuous	0	0
101	LMN987	Cohort 3	0	0	1	0	0	Rate of episodes ...	Continuous	0	0
124	GHI654	Cohort 1	2	2	3	0	0	Rate of episodes ...	Continuous	0.89	0.89
124	GHI654	Cohort 4	2	2	3	0	0	Rate of episodes ...	Continuous	0.89	0.89
124	Screen Failure		0	0	3	0	0	Rate of episodes ...	Continuous	0	0
150	GHI654	Cohort 4	1	1	1	0	0	Rate of episodes ...	Continuous	0	0

- What do these values represent...an actual result, or a lack of a result?

Traceability Issues

Missing Sites

- All sites from each study must be listed in the CLINSITE dataset
- Incorrect Implementation: Occasionally it is seen that sites are missing from CLINSITE
 - Likely due to no randomized subjects from those sites for the study
- Recommendation: These sites should still be listed in CLINSITE
 - Likely would have just one record for that SITEID
 - For the ARM variable, the BIMO TCG states: *When no arm or treatment group is available due to only screen failure subjects at site, use label "Screen Failure."*

Counts Not Matching

- Variables need to be populated with correct counts from the SDTM or ADaM datasets
- Documentation should clearly explain where information comes from
- Example: NSAE (Number of Non-Serious Adverse Events) variable
 - If value isn't a straightforward count from the AE or ADAE dataset it should be clearly defined in the define.xml and possibly explained BIMO reviewers guide
- Many variables have somewhat obvious predecessors in SDTM/ADaM
 - Examples include:
 - Number of Subjects Discont. Study (DISCSTUD)
 - End of Study Status (EOSSTT) in ADSL
 - Number Subjects Discont. Study Treatment (DISCTRT) variables.
 - End of Treatment Status (EOTSTT) in ADSL

Conclusion

Conclusion

- CLINSITE dataset is used by FDA's clinical investigator site selection tool
 - Strict adherence to specifications and guidance is necessary in order to:
 - Avoid issues with loading and using the dataset
 - Avoid hindrances to agency's safeguarding of trial participants
 - Reduce the possibility of delays in submission timelines
- Preparers of CLINSITE datasets should do the following:
 - Develop a standard process for creating these submission deliverables
 - Strictly adhere to specifications, guidance, expectations from the FDA
 - Verify that these deliverables are of high quality with no issues
 - **Note: Pinnacle 21 Enterprise supports validation of CLINSITE datasets against specifications listed in BIMO TCG**

Contact Information

- Michael Beers
- Certara
- michael.beers@certara.com