

基于 CDISC 标准的生物等效性试验 基本数据集构建与应用

Development and application of a core data set for bioequivalence trials based on CDISC standards

王思琪¹, 李钊合², 张敏广²,
谢娜¹, 朱亚河¹, 武英¹,
刘亨辉²

(1. 华北理工大学 公共卫生学院, 河北 唐山 063210; 2. 北京裕荣医学数据科技有限公司, 北京 100176)

WANG Si - qi¹, LI Zhao - he²,
ZHANG Min - guang², XIE Na¹,
ZHU Ya - he¹, WU Ying¹,
LIU Heng - hui²

(1. School of Public Health, North China University of Science and Technology, Tangshan 063210, Hebei Province, China; 2. Beijing Yurong Medical Data Technology Co., Ltd., Beijing 100176, China)

作者简介: 王思琪(2000 -), 女, 硕士研究生, 主要从事数据标准化相关研究

通信作者: 武英, 副教授, 研究生导师

MP: 15133915086

E - mail: tsydyd@sina.com

刘亨辉

MP: 18618242428

E - mail: liuhenghui@your - data.cn

摘要:目的 为解决生物等效性试验数据在遵循临床数据交换标准协会标准递交过程中存在的效率低、一致性差等问题, 构建一套可复用的标准化基本数据集, 以提升数据准备效率与监管递交质量。**方法** 系统回顾多项已完成 BE 试验项目, 涵盖不同试验设计、剂型及研究类型, 通过业务分析提炼共性数据结构与映射规则, 并严格依据 CDISC 标准构建覆盖 SDTM 与 ADaM 两层的标准化数据集模板。**结果** 本研究共归纳出适用于生物等效性试验的 SDTM 数据元专用属性 274 项及 ADaM 数据元专用属性 338 项, 形成了一套适用于多数 BE 试验场景的标准化基本数据集, 并在一项两制剂、两周期、两交叉评价某凝胶贴膏贴剂的 BE 研究中成功应用。**结论** 本研究所构建的标准化数据集能为 BE 试验数据标准化提供了初步框架, 待进一步大规模应用验证, 明确其对提升 BE 试验数据处理的规范化水平与工作效率的适用性与稳健性。对保障数据质量、提高递交成功率具有重要实践价值, 并为行业内的基本数据集构建提供了可借鉴的范例。

关键词: 临床数据交换标准协会; 生物等效性试验; 基本数据集; 研究数据制表模型; 分析数据模型; 数据融合

DOI: 10.13699/j.cnki.1001-6821.2026.02.023

中图分类号: R969 文献标志码: A

文章编号: 1001-6821(2026)02-0291-06

Abstract: Objective To address issues such as low efficiency and poor consistency in the submission of bioequivalence trial data under the Clinical Data Interchange Standards Consortium standards, this study aimed to develop a set of reusable standardized core data sets to improve data preparation efficiency and regulatory submission quality. **Methods** A systematic review of multiple approved BE trial projects was conducted, covering different trial designs, dosage forms, and study types. Common data structures and mapping rules were extracted through business analysis, and standardized data set templates covering both SDTM and ADaM layers were developed in strict accordance with CDISC standards. **Results** This study summarized 274 SDTM and 338 ADaM data element - specific attributes applicable to bioequivalence trials. A set of standardized core data sets suitable for most BE trial scenarios was established and successfully applied in a two - formulation, two - period, two - sequence crossover BE study evaluating a gel patch. **Conclusion** The standardized dataset constructed in this study provides a preliminary framework for the standardization of BE trial data, which requires further

validation through large-scale applications to clarify its applicability and robustness in enhancing the standardization level and work efficiency of BE trial data processing. It holds significant practical value for ensuring data quality and improving the success rate of submissions, while also serving as a referenceable example for the development of foundational datasets within the industry.

Key words: CDISC; bioequivalence trial; core data set; SDTM; ADaM; data harmonization

2024年,国家药品监督管理局药品审评中心登记的4900项临床试验中,2368项为生物等效性试验,占有所有临床试验的48.2%^[1],这一数据凸显了BE试验在当前我国药物临床试验中的核心地位。随着仿制药一致性评价工作的持续推进,BE试验作为验证仿制药与原研药体内代谢一致性的关键环节,其数量持续维持在较高水平。生物等效性试验具有业务内容较为固定、数据结构和分析方法重复性高等特点,有利于通过构建统一的数据标准来提升数据质量和管理效率^[2]。基本数据集是在系统建设中定义的具有主题的、可标识的、能被计算机处理的最小数据集,收纳最基础、最核心的数据项,用以指导数据的标准化采集储存、提供的基本数据集分类和内容^[3],从源头实施数据治理,为信息系统间数据交换、共享以及大数据应用奠定基础。本研究在系统归纳整理生物等效性临床试验业务需求的基础上^[4],结合临床数据交换标准协会(Clinical Data Inter-change Standards Consortium, CDISC)发布的系列数据标准^[5-7],以有利于提升数据质量和统计分析效率为基本原则,明确数据采集、映射与分析全流程的规范化要求,系统构建生物等效性试验的基本数据集,为今后实现生物等效性试验数据分析的自动化和提高评审效率奠定基础。

资料与方法

1 研究材料

本研究材料来源于业内某专业数据统计公司已完成的48项生物等效性试验项目,涵盖2019年至2024年间完成并已提交至国家药品监督管理局药品审评中心并完成审评的项目。其中,两制剂单次给药交叉设计32项、两制剂单次给药平行设计12项、重复设计4项;涉及口服固体制剂、口服液体制剂、透皮贴剂、栓剂、吸入制剂等多种剂型;部分试验还包括食物效应研究等内容。

为系统构建符合标准的数据集^[8],本研究对上述各试验的统计分析报告进行了全面梳理与综合分析。同时,深入研读了CDISC标准相关指南,包括但不限

于研究数据制表模型(Study Data Tabulation Model, SDTM)、分析数据模型(Analysis Data Model, ADaM)及其对应治疗领域实施指南(Implementation Guide, IG),并以《GB/T 18391 信息技术元数据注册系统》《药品监管信息化标准体系》《药物临床试验质量管理规范》为主,《卫生健康信息数据元标准化规则》《卫生健康信息数据元目录》^[9-10]等卫生健康信息标准作为补充,为生物等效性试验数据集的标准化构建提供理论依据与技术参考。

2 数据标准化规则

2.1 数据集分类

在BE试验中,数据集的分类对于高效检索与管理至关重要。合理的分类不仅有助于研究人员和监管人员快速、精准定位目标数据,也为海量试验数据的系统性管理、质量评估与交叉比对提供支持。将递交数据标准与分析数据标准划分为人口学及社会学特征、健康史、体格检查、实验室检查、临床辅助检查、药品与器械、主诉与症状、计划与干预八大类,以此构建规范化、可互操作的数据管理框架。如表1所示。

2.2 数据元标准化格式

基本数据集数据元属性统一采用标识类、定义类、关系类、表示类、管理类5类12项属性,并按照通用性程度区分为数据元公用属性(6项)和数据元专用属性(6项)^[11],具体见表2。

2.3 数据元提取及标准化

采取“自下而上”的数据整理方法^[12],首先从生物等效性试验基本数据需求框架中抽象提取出基本数据集及变量,继而进行系统性的汇总、整理与分析。在此过程中,剔除个别项目的特异性数据项,对语义相近的数据项进行合并与整合,最终依据标准规范对数据元专用属性进行统一化描述。

2.3.1 充分引用现有标准作为基础

在对生物等效性试验数据需求框架进行分析时,应确保与CDISC标准以及我国现行的国家与卫生健康标准保持完整引用和高度一致。CDISC标准已构建起全面覆盖临床研究数据要素的规范化体系,能够为试验数据的采集、交换与归档提供完整的技术支

表 1 生物等效性试验数据标准数据元分类

Table 1 Classification of standard data elements for bioequivalence trial data

Dataset Category	Submission Standard Datasets	Analysis Standard Datasets
Demographic and Socioeconomic Characteristics	Demographics (DM)	Subject – Level Analysis Dataset (ADSL)
Health History	Medical History (MH)	Medical History Analysis Dataset (ADMH)
Physical Examination	Vital Signs (VS), Physical Examination (PE)	Vital Signs Analysis Dataset (ADVS), Physical Examination Analysis Dataset (ADPE)
Laboratory Tests	Laboratory Tests (LB), Microbiology (MB)	Laboratory Analysis Dataset (ADLB)
Clinical Auxiliary Examination	Electrocardiogram (EG)	Electrocardiogram Analysis Dataset (ADEG)
Medications and Devices	Pharmacokinetic Concentrations (PC), Pharmacokinetic Parameters (PP)	Pharmacokinetic Concentration Analysis Dataset (ADPC/ADPP)
Complaints and Symptoms	Adverse Events (AE)	Adverse Events Analysis Dataset (ADAE)
Plans and Interventions	Exposure (EX), Trial Arms (TA), Trial Elements (TE), Trial Inclusion/Exclusion Criteria (TI), Trial Visits (TV)	Trial Design Analysis Dataset (ADSL/ADEX)

表 2 标准体系数据元公用属性和专用属性

Table 2 Common and specific attributes of data elements in the standard framework

No.	Attribute Category	Data Element Attribute Name	Constraint	Remarks
1	Identification	Data Element Identifier	Mandatory	Specific Attribute
2		Data Element Name	Mandatory	Specific Attribute
3		Version	Mandatory	Common Attribute
4		Registration Authority	Mandatory	Common Attribute
5		Context	Mandatory	Common Attribute
6	Definition	Definition	Mandatory	Specific Attribute
7	Relation	Classification Scheme	Mandatory	Common Attribute
8	Representation	Data Type of Data Element Values	Mandatory	Specific Attribute
9		Representation Format	Mandatory	Specific Attribute
10		Permitted Values for Data Element	Mandatory	Specific Attribute
11	Administration	Steward Organization	Mandatory	Common Attribute
12		Registration Status	Mandatory	Common Attribute

撑。以“出生日期/时间”这一数据元为例,在 CDISC 标准中,其变量名定义为“BRITHDTC”,变量类型为字符型,并采用 ISO 8601 国际标准格式进行表示,体现了临床数据标准化在结构、语义与格式层面的统一性要求。

2.3.2 现有标准存在差异进行整合

在对生物等效性试验数据标准进行构建时,常需对 CDISC 标准与我国目前现有标准中均作出规定的变量进行整合与取舍。以人口学数据中的“性别”变量为例:CDISC 标准采用受控术语,其值域定义为“M (男性)”“F (女性)”及“U (未知)”;而我国现行国家标准《GB/T 2261. 2 – 2003 个人基本信息分类与代码人的性别代码》中,性别的允许值则包括“1 (男性)”“2 (女性)”“0 (未知的性别)”及“9 (未说明的性别)”。

两者在编码体系与类别上存在差异。

针对此类情况,本研究构建的生物等效性试验数据标准遵循以下原则:在满足我国监管申报与数据质量管理要求的前提下,优先采纳并映射至国际广泛接受的 CDISC 标准,以确保数据的全球可交换性与分析一致性;同时,通过建立明确的编码转换规则,保障与我国国家标准在逻辑层面的兼容性与可追溯性。具体而言,在值域映射过程中,可将国标中的“1”与“2”分别对应 CDISC 的“M”与“F”,“0”和“9”则可统一映射至“U”。这一处理方式既顺应临床试验数据标准化的国际趋势,也能在国内法规框架下实现数据的规范整合,从而支持生物等效性试验数据在多监管环境下的高效使用与合规提交。

结 果

通过对相关标准规范的全面梳理与辩证分析,本研究共归纳出用于生物等效性试验的递交标准基本数据集数据元专用属性 274 项、分析标准基本数据集数据元专用属性 338 项。这些属性均可在 CDISC 标准体系中追溯其来源^[13]。

表 3 药代动力学浓度(PC)数据集数据元专用属性(部分)

Table 3 Specific attributes of data elements in the pharmacokinetic concentration (PC) dataset (partial)

Data Element Identifier	CDISC Variable Name	Data Element Name	Definition	Data Type	Display Format	Permitted Values
DE15.02.001.00	USUBJID	Unique Subject Identifier	Identifier used to uniquely identify a subject across all studies for all applications or submissions involving the product.	Char	AN. . 40	As per coding rules defined in the study protocol
DE15.01.001.00	PCTESTCD	Pharmacokinetic Test Short Name	Short name of the analyte or specimen characteristic.	Char	AN. . 8	As per CDISC standards
DE15.01.002.00	PCTEST	Pharmacokinetic Test Name	Name of the analyte or specimen characteristic.	Char	AN. . 40	As per CDISC standards
DE15.03.001.00	PCORRES	Result or Finding in Original Units	Result of the measurement or finding as originally received or collected.	Num	N. . 10. 4	Actual measured concentration value
DE15.03.002.00	PCORRESU	Original Units	Original units in which the data were collected. The unit for PCORRES.	Char	AN. . 10	Unit corresponding to the measured value
DE15.03.003.00	VISIT	Visit Name	Protocol – defined description of clinical encounter.	Char	AN. . 40	As per the visit schedule defined in the study protocol
DE15.04.002.00	PCDTC	Date/Time of Specimen Collection	Date/time of specimen collection represented in ISO 8601 character format.	DateTime	DTM	ISO 8601 standard (YYYY – MM – DDThh:mm)
DE15.05.001.00	PCSTAT	Completion Status	Used to indicate a result was not obtained.	Char	AN1	C = Completed; N = Not Done
DE15.06.001.00	PCLLOQ	Lower Limit of Quantitation	Indicates the lower limit of quantitation for an assay.	Num	N. . 10. 4	Value determined during assay validation

2 分析标准基本数据集

ADaM 模型以 SDTM 命名规则为基础,并具有其他附加元数据。其构建源于 SDTM 数据集,但与 SDTM 相比,ADaM 数据集具备可分析性、可追溯性和可支持性的基本属性。虽然 SDTM 和 ADaM 标准均旨在协助申办者向监管机构进行数据递交,但两者的侧重点不同,SDTM 主要用于呈现原始数据,而 ADaM 更侧重于存储数据分析的信息。鉴于 ADaM 变量多从 SDTM 继承而来,本文的研究重点将聚焦于在此基础上需要新衍生的变量。在衍生过程中变量的设计

1 递交标准基本数据集

在标准化过程中,若《药品监管信息化标准体系》已涵盖相关数据元,则予以直接沿用;若数据元来源于 CDISC 标准,则依据数据元描述规范,对其标识符、数值类型、表示格式及允许值进行重新定义。为清晰说明其标准化处理方式,本文以药代动力学浓度(PC)数据集中典型的变量为例,展示其专用属性的具体构成与定义,详见表 3。

遵循以下原则:其一,分析需求驱动原则,即所有衍生变量均需直接服务于既定的统计分析目的,例如为计算药代动力学参数或进行假设检验而构建;其二,可追溯性原则,要求每个衍生变量均能通过明确的算法或元数据链路回溯至一个或多个 SDTM 源变量,确保数据分析结果的真实性与可复核性;其三,标准化与一致性原则,对于同类分析场景,应采用统一的衍生逻辑和命名规范,以支持数据的跨试验比对与整合。兹以 ADPC 数据集为例,详细阐述其衍生变量的数据元信息,见表 4。

表 4 药代动力学浓度(ADPC)数据集数据元专用属性(部分)

Table 4 Specific Attributes of Data Elements in the Pharmacokinetic Concentration (ADPC) Dataset (Partial)

Data Element Identifier	CDISC Variable Name	Data Element Name	Definition	Data Type	Display Format	Permitted Values
DE15.07.001.00	TRTA	Actual Treatment	Description of the treatment drug or regimen a subject actually received during the analysis period.	Char	AN..40	As defined in the study protocol, e.g., "Test Drug T", "Reference Drug R"
DE15.07.002.00	TRTAN	Actual Treatment (N)	Numeric code for the actual treatment.	Num	N2	As defined in the study protocol, e.g., 1 = Test Drug T, 2 = Reference Drug R
DE15.04.002.00	TRTSDT	Date of First Exposure to Treatment	Date the subject first received the actual treatment.	DateTime	DTM	ISO 8601 standard (YYYY-MM-DD)
DE15.04.003.00	TRTSTM	Time of First Exposure to Treatment	Time the subject first received the actual treatment.	DateTime	DTM	ISO 8601 standard (hh:mm)
DE15.04.004.00	APERIOD	Period	Numeric identifier for the treatment cycle used in pharmacokinetic analysis.	Num	N2	e.g., 1 = Period 01, 2 = Period 02
DE15.04.005.00	APERIODC	Period (C)	Description of the treatment cycle used in pharmacokinetic analysis.	Char	AN..20	e.g., "Period 01", "Period 02"
DE15.04.006.00	ARELTM	Analysis Relative Time	Time point of blood sampling relative to the actual dosing time.	Num	N..5.2	Measured time point, e.g., 0.00 (pre-dose), 0.50, 1.00, 2.00

讨 论

本研究结合 CDISC 标准与我国国家药品监督管理局信息化标准,系统性构建了生物等效性试验 SDTM 与 ADaM 数据元的专用属性集,形成了覆盖数据定义、表示与值域规范的标准化框架。标准化过程中遵循“引用优先、映射为本、适度调适”的原则,在保障数据国际可交换性的同时,兼顾国内监管与数据管理实际需要。以下结合研究过程与结果,从标准融合路径、实施优势与研究创新性三方面展开论述。

1 CDISC 标准与中国数据标准体系的融合路径

本研究在构建数据元专用属性时,并未简单照搬 CDISC 或国内标准,而是在系统梳理与对比的基础上,探索出一条“以 CDISC 为骨架、以我国国家药品监督管理局信息化标准为补充”的融合路径。首先,在结构性属性上充分承接 CDISC 框架。SDTM 与 ADaM 的数据域划分、变量命名规则及基本数据类型均严格遵循 CDISC 标准,确保数据在结构层面具备与国际监管机构及行业通用工具的无缝对接能力。其次,在语义与值域层面实施精细化映射。对于 CDISC 与国内标准同时涵盖且存在差异的数据元,并未强行统一编码体系,而是通过建立清晰的映射规则,在提交数据时优先采用 CDISC 受控术语,同时保留与国家标准之间的逻辑对应关系,实现“一套数据、双重合规”。数

据收集端允许研究机构依据各中心现行操作规范或国内标准要求采集数据(如性别可记录为“男/女”或“M/F”),保持原始数据的完整性与灵活性;数据递交端则根据清晰的术语映射规则,在数据递交前统一转换为 CDISC 受控术语,并完整保留与国家标准的逻辑对应关系;监管审评端建议审评人员以 CDISC 标准为主线审阅数据,必要时可通过映射关系追溯至原始采集记录,兼顾国际规范性与国内可追溯性。此种策略既避免了强制统一编码对现有研究流程的冲击,又为监管机构保留了核查路径,在实操层面平衡了国际递交效率与本土合规需求。这种处理既维护了数据在全球范围内的语义一致性,也满足了我国数据归档与二次利用的本土化要求。最后,针对我国特有数据需求进行属性扩展。对于国内监管或分析特别关注、而在 CDISC 中未充分定义的数据项,本研究在遵循相同元数据规范的前提下,依据相关标准予以补充定义,确保标准体系在覆盖全面性上不存在盲区。

2 基于审评需求驱动的 BE 试验基本数据集构建路径

BE 试验基本数据集指的是在国家药品监管层面针对 BE 试验必须收集的最小数据集,而收集数据的根本目的是为了提供规范化试验信息,满足 BE 试验审评决策、药品一致性评价及临床用药安全评估的需要^[10]。基于此,本研究确定需求分析作为 BE 试验基本数据集研究技术路线的核心环节。具体而言,首

先从所要递交的 BE 试验统计分析报告需求分析开始,从分析统计分析报告所需数据、要提供的何种审评决策信息着手,进行 BE 试验全流程业务需求分析,开展试验设计、数据采集、统计分析等核心业务的用例分析,在此基础上确定要收集的最基本数据,最后规范性描述 BE 试验基本数据集、核心数据元及标准代码集。

3 基本数据标准研究创新性不足

为适配全球医药行业发展与严苛监管需求,生物等效性试验数据的质量、完整性及可追溯性愈发关键。本研究通过需求驱动构建数据标准,建立监管规范与数据落地的映射机制,形成“监管要求 - 数据标准 - 分析应用”闭环,既保障科研规范性,又保留技术实施灵活性,为后续高质量自动化流程奠定基础^[14]。需说明的是,该标准体系虽经一项未上市生物等效性试验初步验证,但其适用性与稳健性仍需大规模实际应用进一步验证。此外,本研究聚焦数据集的业务定义与结构化描述,数据元标准化仅围绕“专用属性”展开,未系统阐述版本、注册机构、提交状态等服务于数据元注册、发布、检索及生命周期管理的“公用属性”。鉴于数据集的业务含义、结构与内容是保障数据质量及分析可比性的核心基础,故为本研究重点;而公用属性的标准化应用,需依托更高层级的机构或监管层面数据治理框架统筹实现。

在 CDISC 标准框架下,深度融合中国卫生健康标准,构建我国临床试验数据递交标准,是实现我国数据质量管理现代化和审评效率提升的战略举措。这不仅优化技术流程,更是监管科学进步和产业高质量发展的关键基础设施,确保数据兼具国际可比性与适用性,为提升仿制药在全球创新地位提供数据支撑。当前,以生物等效性试验为代表的试验类型,其数据集、数据元及代码集亦需持续修订。未来需加强国际协作及跨卫生领域合作,共同制定、维护标准,以实现真正的数据交换与共享。本研究产出不仅为行业提供了可直接参考或采纳的技术规范,其“从实践中提炼,并回归实践验证”的研究方法论,也为其他治疗领域或数据类型的临床试验数据标准化工作提供了可资借鉴的思路与范例。

参考文献:

- [1] 国家药品监督管理局药品审评中心,2024 年度药品审评报告 [EB/OL]. 2025-03-18 [2025-12-27]. <https://www.cde.org.cn/main/news/viewInfoCommon/54538c67b7e764fc51666567fc620241>.
- [2] 许明哲,崔琛琛,等. 我国仿制药质量关键影响因素分析 [J]. 中国药事,2020,34(10):1130—1144.
- [3] 国家药品监督管理局,国家药监局关于发布《医疗器械生产许可与备案管理基本数据集》等 2 个信息化标准的公告 (2025 年第 75 号) [EB/OL]. 2025-08-01 [2025-11-25]. <https://www.nmpa.gov.cn/xxgk/ggtg/ylqxggtg/ylqxhybzhgg/20250801145013147.html>.
- [4] 汪旻晖,江思艳,胡骅,等. 生物等效性研究数据管理质量控制体系的建立及应用 [J]. 中国新药杂志,2018,27(22):2602—2606.
- [5] 王泽,陈浩,陈佳齐,等. 临床数据交换标准协会体系及其应用现状 [J]. 中国公共生,2024,40(02):227—233.
- [6] FDA. Providing Regulatory Submissions Electronic Format Certain Human Pharmaceutical Product Applications [EB/OL]. 2023-12-12 [2024-12-14]. <https://www.fda.gov/regulatory-information/search-fda-guidance-documents/providing-regulatory-submissions-electronic-format-certain-human-pharmaceutical-product-applications>.
- [7] PMDA. New Drug Review with Electronic Data | Pharmaceuticals and Medical Devices Agency [EB/OL]. 2023-10-03 [2024-12-14]. <http://www.pmda.go.jp/english/review-services/reviews/0002.html>.
- [8] 李丹彤,梁会营,刘广建. 临床科研数据库建设中的数据标准化问题探讨 [J]. 中国数字医学,2021,16(01):29—34.
- [9] 国家药品监督管理局,药物临床试验质量管理规范:2020 年局令第 57 号 [EB/OL]. (2020-04-26) [2025-11-23]. <https://www.nmpa.gov.cn/zhuanli/ypzhcglibf/ypzhcglibfzhewj/2020042616-2401243.html>.
- [10] 中华人民共和国卫生部. 卫生信息数据元目录 第 1 部分:总则:WS 363.1-2011 [S]. 北京:人民卫生出版社,2011:2—3.
- [11] 王骏,韩景静,黄钦. 从药品技术审评谈临床试验的数据标准化 [J]. 中国新药杂志,2014,23(19):2228—2230.
- [12] 娄苗苗,杨喆,刘丹红,等. 卫生数据标准化方法研究 [J]. 中国卫生信息管理杂志,2013,10(5):440—443.
- [13] SUI Y, HE J, FAN M, *et al.* An end-to-end digital platform for TFL automation — From metadata to TFL outputs. PharmaSUG China 2024. (2024). AA-10057.
- [14] 王崇旭. 临床试验数据管理电子化和 CDISC 数据自动化的实现 [D]. 东南大学,2018.

(本文编辑:冀希炜 收稿日期 2025-12-27)