

# 真实世界数据相关标准体系研究与应用进展\*

印冠锦 张梦阳

吴惠庶

(神州医疗科技股份有限公司 北京 100080)

(中国政法大学中加法律研究中心 北京 100088)

洪 娜

弓孟春

(神州医疗科技股份有限公司 北京 100080)

(南方医科大学卫生与健康管理局 广州 510515)

〔摘要〕 阐述真实世界数据标准体系和研究进展,分析国内真实世界数据标准面临的挑战并提出相关建议,包括在行业内整体全面规划医学语义标准和数据模型、进一步细化个人数据使用和调用的管理方式等,为我国真实世界数据相关标准制定提供有力依据。

〔关键词〕 真实世界数据;真实世界研究;真实世界证据;数据标准;术语标准

〔中图分类号〕 R-058 〔文献标识码〕 A 〔DOI〕 10.3969/j.issn.1673-6036.2022.06.006

**Research and Application Progress of Related Standards for Real World Data** YIN Guanjin, ZHANG Mengyang, Digital Health China Technologies Co. Ltd., Beijing 100080, China; WU Huishu, Sino-Canadian Law Research Center, China University of Political Science and Law, Beijing 100088, China; HONG Na, Digital Health China Technologies Co. Ltd., Beijing 100080, China; GONG Mengchun, Health Management Institute, Southern Medical University, Guangzhou 510515, China

〔Abstract〕 The paper expounds the standard system and research progress of Real World Data (RWD), analyzes the challenges of RWD standards in China, and puts forward relevant suggestions, including the overall planning of medical semantic standards and data models in the industry, further refinement of the management mode of the usage and invocation of personal data, etc., so as to provide a strong basis for the formulation of standards related to RWD in China.

〔Keywords〕 Real World Data (RWD); Real World Study (RWS); Real World Evidence (RWE); data standard; terminology standard

〔修回日期〕 2022-03-17

〔作者简介〕 印冠锦, 硕士, 医学数据研究员; 通信作者: 弓孟春, 博士, 特聘研究员。

〔基金项目〕 国家重点研发计划“健康管理综合服务应用示范”(项目编号: 2020YFC2006400)。

## 1 引言

### 1.1 真实世界数据概念和范畴

为了指导和规范申办者利用真实世界数据生成真实世界证据从而支持药物研发, 国家药品监督管理局食品药品审核查验中心于 2021 年 4 月 15 日颁

布《用于产生真实世界证据的真实世界数据指导原则（试行）》（以下简称《指导原则》）<sup>[1]</sup>。《指导原则》对真实世界数据（Real World Data, RWD）给出定义，即来源于日常所收集的各种与患者健康状况和/或诊疗及保健有关的数据。真实世界证据（Real World Evidence, RWE）则是在 RWD 满足适用性条件下，经过科学、充分的分析产生的。目前我国真实世界数据主要分为医院信息系统数据、死亡登记数据、医保数据、药品安全性主动监测数据、组学数据、登记研究数据、患者报告数据、自然人群队列数据、个体健康监测数据以及其他特定功能数据等。

## 1.2 制定真实世界数据相关标准的意义

建立统一的 RWD 标准将有利于解决数据来源相对独立、封闭，数据系统种类繁多，存储分散、标准不一，数据整合、交换困难，数据碎片化、信息孤岛现象等问题，因此推动 RWD 在更大范围和更深层次上传播、共享和使用具有重要意义<sup>[2]</sup>，这一过程主要包括制定实施医学术语定义、使用指导性规范以及建立术语规范化体系，如术语表、叙词表、分类表、编码、本体等。

## 1.3 真实世界数据标准国际研究现状

目前各国政府机构，如中国国家药品监督管理局（National Medical Products Administration, NMPA），欧洲药品管理局（European Medicines Agency, EMA），美国食品药品监督管理局（Food and Drug Administration, FDA），都已经发布 RWD 标准框架相关文件，旨在评估 RWD 适应性，使 RWD 转化为真实可用的 RWE 相关标准，为医疗器械审批和监管决策提供指导和建议。例如美国国会公布的《21世纪治愈法案》，美国 FDA 的《美国联邦食品、药品和化妆品法案》以及《用于药品和生物制品提交的真实世界数据标准行业指南》，EMA 的 MINERVA 项目定义和收集真实世界数据源的元数据等。

## 1.4 RWD 应用存在的问题

从数据来源看，相较于传统的随机对照试验

（Randomized Controlled Trial, RCT），RWD 在记录、采集以及数据存储环节尚缺乏强有力的质量控制，易导致数据缺失、记录不准确等问题，这将在很大程度上影响后续的数据治理、应用以及可追溯性，研究者也难以发现问题并进行核对和修正。我国系统性 RWD 研究相较于欧美国家起步较晚，目前我国系统性 RWD 研究尚无成熟标准和规范借鉴。由于缺乏统一标准，RWD 研究存在数据来源相对独立、封闭，数据系统种类繁多，存储分散、标准不一，数据整合、交换困难，数据碎片化、信息孤岛现象等问题。电子病历数据还可能因记录者对文字类型的主观性差异，影响对临床结局的客观评价<sup>[1]</sup>。

## 2 真实世界数据的标准体系和研究进展

### 2.1 数据标准体系框架（表 1）

表 1 常用数据标准

数据标准名称	发布机构	对象
CDASH	CDISC	临床试验数据
OMOP CDM	OHDSI	观察性数据
FHIR	HL7	有关患者和动物参与 医疗服务的各方面数据

2.1.1 CDISC CDASH 美国临床数据交换标准协会（Clinical Data Interchange Standards Consortium, CDISC）就数据收集、数据类型、数据提交建立以下 5 套标准<sup>[3-4]</sup>：基础标准（Foundational Standards）、数据交换标准（Data Exchange Standards）、治疗领域标准（Therapeutic Areas）、术语集（Terminology）和 SHARE（Shared Health and Research Electronic Library）。CDISC 在组织 RWD 方面的调研和研究时发现，RWD 标准框架构建将有助于加速数据共享、提升数据治理能力、提高数据一致性、加快新疗法的开发、强化隐私保护<sup>[5]</sup>。2018 年 CDISC 制定临床数据获取协调标准（Clinical Data Acquisition Standards Harmonization, CDASH），为简化和规范临床试验的数据收集和病例报告表（Case Report Form, CRF）提供设计标准<sup>[6-7]</sup>。定义一组基本数据采集字段并且这些字段会出现在 CRF 需要

采集的信息上。基于 CDASH 设计 CRF 字段能够保持数据标准的统一与数据的可追溯性,可以直接从 CRF 中抽取大量符合研究数据表格模型的数据集,从而减少后期繁杂的映射转换。近期美国 FDA 发布《用于药品和生物制品提交的真实世界数据标准行业指南》,建议在提交包含真实世界数据的文件时应使用恰当的 CDISC 标准以确保完成数据提交<sup>[8]</sup>。

2.1.1.2 OHDSI /OMOP CDM 观察性健康医疗数据科学与信息学 (Observational Health Data Sciences and Informatics, OHDSI) 项目是由美国哥伦比亚大学领导的全球非营利研究联盟,主要研究医学大数据分析解决方案,通过各组织间合作收集和分析数据结果,进而促进更好的决策和医疗。拥有超过 6 亿人口的临床数据规模,累计协作研究发表了上百篇论文<sup>[1,9]</sup>。OHDSI 采用观察性医疗结果合作组织 (The Observational Medical Outcomes Partnership, OMOP) 的通用数据模型 (Common Data Model, CDM)。CDM 是一套具有统一标准的数据模型,将观察性数据的格式和内容规范化,目的是将不同数据库包含的数据转换为通用格式,然后使用标准化数据库进行系统分析<sup>[10-11]</sup>。

2.1.1.3 FHIR 快速医疗互操作性资源 (Fast Health Interoperable Resources, FHIR) 概念于 2011 年提出,其建立在一系列被称为“资源”的模块化组件之上,可以任意组配用于对应工作系统。其特点包括注重实施、代码库丰富、即开即用的互操作性等<sup>[12-13]</sup>。

## 2.2 医学术语标准

2.2.1 概述 将医学术语标准化,可以解决表述与理解不一致的问题,继而对临床诊疗与医学科研中所用到的语言进行统一命名起到重要作用,协助不同医疗信息系统之间互联互通及数据共享。目前应用较多的医学术语标准包括综合性临床医学术语、疾病分类与编码、观测指标等,见表 2。

表 2 常用术语标准

术语标准名称	内容	适用领域
SNOMED CT	综合性临床医学术语	概念表、描述表、关系表
ICD	疾病分类与编码	类目表、指导手册、字母顺序索引
LOINC	观测指标	LOINC 数据库、对照程序

2.2.2 SNOMED CT 医学系统命名法 - 临床术语 (Systematized Nomenclature of Medicine Clinical Terms, SNOMED CT), 是当前国际上使用范围最广的临床医学术语标准,包含 19 个层级体系,概念表、描述表和语义关系表是其最基本的组成单元。SNOMED CT 目前已应用于 80 多个国家和地区,在美国、加拿大、新西兰、英国和澳大利亚等国家 SNOMED CT 被认可为国家标准<sup>[2,14]</sup>。

2.2.3 ICD 国际疾病分类 (International Classification of Diseases, ICD) 是疾病分类统计术语集合。主要用于死亡和重要疾病的归类统计,以便于健康系统资源规划。国际上 70% 的健康资源都是通过该系统的统计结果加以分配使用的。目前全球通用的是第 10 次修订本,统称为 ICD - 10。ICD - 11 中文版于 2018 年编写完成,自 2019 年 3 月起实施<sup>[15]</sup>。

2.2.4 LOINC 观测指标标识符逻辑命名与编码系统 (Logical Observation Identifiers Names and Codes, LOINC) 是一套包含实验室结果和影像结果的检验代码系统。涵盖血清学、毒物、微生物、放射医学报告等指标。成份与分析物名称、时间、检查与检验属性类型、标尺与检测方法、体系/样本类型是目前 LOINC 的重要组成部分<sup>[4,13]</sup>。

## 2.3 质量控制标准

RWD 质量控制的关键是确保数据的完整性、准确性和透明性。多个国家的监管机构在发布的指导原则中提到将数据完整性标准作为临床试验数据质量的国际通行原则。此外临床试验数据管理应严格遵循国家药监部门发布的《临床试验数据管理工

作技术指南》《临床试验的电子数据采集技术指导原则》《药品临床试验质量管理规范》，其均为传统药物或医疗器械的临床试验遵循的标准，对于RWD质量控制具有借鉴意义。我国在RWD质量控制方面出台了一系列指导性文件，包括《真实世界数据用于医疗器械临床评价技术指导原则》《用于产生真实世界证据的真实世界数据指导原则（试行）》《2018真实世界研究指南》《真实世界证据支持药物研发与审评的指导原则（试行）》，其中均提出一些指导性原则。

## 2.4 隐私安全及隐私保护标准

2.4.1 国外标准 患者个人隐私与健康数据在真实世界数据的处理与应用中需要被重点关注与保护。国外数据安全与隐私保护所依托的标准主要有《健康保险流通与责任法案》（Health Insurance Portability and Accountability Act, HIPAA）涉及的隐私控制方面的内容，以及欧洲《通用数据保护条例》（General Data Protection Regulation, GDPR）。其中HIPAA对于数据安全的要求可以分为技术层面、物理层面、行政层面和基本安全基线4个层次，对于个人信息的保护则是基于对患者隐私权利来设计的<sup>[7]</sup>。不同于欧洲的顶层制度设计，GDPR所认为的个人信息权属于人权的一部分，因此对可携带权、最小目的限制等方面均做出详细规制。

2.4.2 国内标准 近年来随着法律体系的逐步完善，我国对于数据保护的监管体系向纵深发展并逐步细化。随着《网络安全法》《数据安全法》《个人信息保护法》相继出台，数据分类分级、关键基础设施认定等一系列政策法规及制度体系落地，各数据权利主体的权益有了法律层面的有效保障。

## 2.5 注册及报告标准

真实世界研究（Real World Study, RWS）是临床研究的一种，目前在我国遵循一般临床试验的注册要求。中国临床试验注册中心（Chinese Clinical Trial Registry, ChiCTR）加入了世界卫生组织国际临床试验注册中心平台。平台规定了所有的在人体中和采用取自人体的标本进行的研究，包括各种干

预措施疗效和安全性的有对照或无对照试验（如RCT、病例-对照研究、队列研究及非对照研究）、预后研究、病因学研究和诊断技术、试剂、设备等的诊断性试验，均需注册并公告。关于数据报告，要求注册表中提供临床试验计划使用的数据收集和管理系统信息。中国临床试验注册中心于2016年3月14日起要求在填注册申请表时填入公开原始数据计划<sup>[6,11]</sup>。

## 3 国内RWD标准面临的挑战与建议

### 3.1 语义标准和数据模型缺失

通用数据模型（CDASH、OMOP CDM、FHIR）和语义标准（SNOMED CT、ICD、LOINC）均来源于国外，由于其自身特点和应用场景不同，国内各类机构中仍呈现出多种语义标准和数据模型共存的现象。借鉴国外现有成果与先进经验，建设具有自主知识产权的中文医学数据模型与术语体系，推动中文医学数据标准化任重而道远。

### 3.2 个人信息保护不完善

尽管近年来我国对于数据保护的监管体系逐步向纵深发展细化，但是对于个人数据的保护仍处于前期探索阶段，对于个人数据的使用、调用和相关管理仍基于管理框架和控制点等比较粗犷的管理方式，没有细化到字段层面<sup>[16]</sup>。同时对于脱敏技术、隐私计算技术等新技术的使用尚缺乏相关法律法规标准，给实践过程的合规性带来很大挑战。如“匿名化”是法律概念，不能仅通过技术上的判断得到认定，再识别风险仍然存在<sup>[17]</sup>；如果合规架构建立在对特定匿名化技术的信任之上，则需结合现有案例并综合考虑法律上所有客观场景因素，对技术是否足以在法律上认定为匿名化做出论证。此外“透明化原则到底需要做到什么程度算是法律所接受的”等问题都需要在真实世界研究中进一步加以考虑以切实保护患者隐私。

### 3.3 质量控制及报告体系欠缺

针对质量控制及报告体系我国已出台相关指导

性文件。中国针灸学会编制了《真实世界临床研究数据管理规范（征求意见稿）》，较为详细地阐述了数据管理相关人员资质、培训、数据管理流程等具体要求。但截至目前尚未形成明确的评价工具用以报告真实世界研究中 RWD 质量。尽管目前 RWS 注册仍按照传统临床试验要求进行，但由于其研究设计等方面具有独特性，因此有必要制定单独的注册标准。

### 3.4 我国 RWD 标准建设建议

3.4.1 针对语义标准和数据模型缺失问题 建议行业内整体全面规划医学语义标准和数据模型，例如建立以 SNOMED CT 为核心的医学语义标准，中文数据模型的建立参照 OMOP CDM；创造良好标准和数据实施环境，加强院内数据治理，打通系统间的信息“孤岛”，促进科研合作、共享；加强与 OHDSI、CDISC 等国际机构的合作；加强行业间交流合作与创新。

3.4.2 针对个人信息保护粗放问题 建议进一步细化个人数据使用和调用的管理方式，例如数据方对数据进行加密、脱敏，在物理层面和系统层面对访问权限进行管控，计算方数据计算时使用联邦学习等隐私保护技术；政策制定方应完善新技术相关法律法规标准，积极建立统一完备的行业标准，在实践中对个人信息保护技术进行充分论证；政策监管方应与申请方及时沟通交流，保证研究符合相关法律法规标准，保障个人隐私安全。

3.4.3 针对质量控制及报告体系欠缺问题 建议完善 RWD 质量管理体系；制定 RWD 标准操作流程；定期培训专业人员；学习国外先进注册标准，结合国情制定我国注册标准。例如建立 RWS 透明化制度，在研究方案预注册发表，统计分析计划（Statistical Analysis Plan, SAP）制定，数据清洗方案等方面接受监督。积极融入中国真实世界数据与研究联盟（China Real World Data and Studies Alliance, ChinaREAL），通过科学且严密的顶层设计建立系列 RWD 与 RWS 技术规范体系，促进高质量 RWE 的产生。

## 4 结语

总体来看我国 RWD 标准研究体系尚缺乏整体规划，存在条块分割问题。一些标准在一定范围、地域、系统内适用，但不足以成为行业认可的公共准则。目前尚未有涵盖临床症状、实验室检查、诊断、药物、疾病分类、临床实践的术语标准，现有中文临床术语存在术语范围局限、体系不完整、结构不合理等问题<sup>[18]</sup>。在未来 RWD 支持的 RWE 研究框架下，安全机制下的多方合作、基于高通量多组学 RWD 的 RWE 研究成为必然趋势。来自政府、学术界、产业界等多方合作的 RWD 标准化框架有助于临床队列数据的有效整合与共享，在多方共识的基础上基于大规模真实世界数据开展高质量研究<sup>[19]</sup>。此外全自动高通量表型化技术与临床表型组建设、基于多组学 RWD 的 RWE 研究，可对生命过程、疾病等集合大量影响因子的复杂生命现象进行高层次的分析、解读，将大力推动新型疾病相关生物标志物的发现以及精准医学的高速发展。

## 参考文献

- 1 国家药品监督管理局药品审评中心. 用于产生真实世界证据的真实世界数据指导原则（试行）. [EB/OL]. [2021-04-15]. <https://www.cde.org.cn/zdylz/domesticinfopage?zdyzIdCODE=7d2e46cea0e459358257760383526e9d>.
- 2 钱庆, 吴思竹. 国外医学术语标准化发展对我国的启示 [J]. 医学信息学杂志, 2013, 34 (5): 42-46, 51.
- 3 CDISC eSDI Group. Leveraging the CDISC Standards to Facilitate the use of Electronic Source Data within Clinical Trials [EB/OL]. [2020-09-16]. <http://citeseerx.ist.psu.edu/viewdoc/download;jsessionid=B83A10847B9D1FB6D4DF68C1254AB655?doi=10.1.1.116.390&rep=rep1&type=pdf>.
- 4 Rhoda F, Barbara J W. Report of Qualitative Delphi Survey [EB/OL]. [2021-07-27]. [https://www.cdisc.org/sites/default/files/2021-08/CDISC\\_RWD\\_Connect\\_Report\\_07-27-2021.pdf](https://www.cdisc.org/sites/default/files/2021-08/CDISC_RWD_Connect_Report_07-27-2021.pdf).
- 5 Facile R, Muhlbradt E E, Gong M, et al. Use of Clinical Data Interchange Standards Consortium (CDISC) Standards for Real-world Data: Expert Perspectives From a Qualitative Delphi Survey [J]. JMIR Med Inform, 2022, 10

- (1): e30363.
- 6 CDISC. CDASH [EB/OL]. [2021-06-17]. <https://www.cdisc.org/standards/foundational/cdash>.
  - 7 Regan P M, Johnson D G. Privacy and Trust in Socio-technical Systems of Accountability [EB/OL]. [2021-06-17]. [https://doi.org/10.1057/9781137032225\\_7](https://doi.org/10.1057/9781137032225_7).
  - 8 Food and Drug Administration. Data Standards for Drug and Biological Product Submissions Containing Real-World Data Guidance for Industry [EB/OL]. [2021-06-17]. <https://www.fda.gov/media/153341/download>.
  - 9 OHDSI. Who We Are [EB/OL]. [2021-06-17]. <https://ohdsi.org/who-we-are/>.
  - 10 OHDSI. OMOP CDM Quiries Introduction [EB/OL]. [2021-06-17]. <http://cdmqueries.omop.org/>.
  - 11 中国临床试验注册中心. 中国临床试验注册中心注册指南 [EB/OL]. [2021-06-17]. <https://www.chictr.org.cn/registry.aspx#One>.
  - 12 HL7. FHIR Home [EB/OL]. [2021-06-17]. <http://build.fhir.org/>.
  - 13 LOINC. LOINC Home [EB/OL]. [2021-06-17]. <https://loinc.org/>.
  - 14 SNOMED. SNOMED CT Home [EB/OL]. [2021-06-17]. <https://www.snomed.org/>.
  - 15 国家卫生健康委. 关于印发国际疾病分类第十一次修订本 (ICD-11) 中文版的通知 [EB/OL]. [2018-12-21]. <http://wjw.yl.gov.cn/index.php?m=Article&a=show&id=14913>.
  - 16 Stein P. Legal Evolution the Story of an Idea [M]. Cambridge: Cambridge University Press, 1979.
  - 17 Gratton é. If personal information is privacy's gatekeeper, then risk of harm is the key: A proposed method for determining what counts as personal information [J]. Albany Law Journal of Science & Technology, 2014 (24): 19210.
  - 18 谢雪娇, 张黎黎, 奈存剑, 等. 国外医学术语标准开发方法及对我国的启示 [J]. 中华医学图书情报杂志, 2019, 28 (11): 16-21.
  - 19 Food and Drug Administration. Real-World Data: Assessing Electronic Health Records and Medical Claims Data To Support Regulatory Decision-Making for Drug and Biological Products [EB/OL]. [2021-06-17]. <https://www.fda.gov/regulatory-information/search-fda-guidance-documents/real-world-data-assessing-electronic-health-records-and-medical-claims-data-support-regulatory>.

(上接第 18 页)

运营管理体系建立显得尤为重要。基于 DAMA 数据管理体系, 通过剖析定位数据分布业务场景和系统, 构建具有普适性的医院运营管理数据中心核心指标体系、数据标准和管理体系, 推动医院核心业务工作与运营管理工作深度融合, 是缓解公立医院经济运行压力、提升内部资源配置效率和运营管理效益的重要手段。目前公立医院运营管理体系构建普遍处于初级探索阶段, 未来需进一步深耕拓展, 推进其管理模式和运行方式加快转变, 以推动公立医院高质量发展。

## 参考文献

- 1 刘智勇, 叶鸿志, 谭辉艳, 等. 我国医疗保障信息化建设历程与发展趋势 [J]. 医学信息学杂志, 2021, 42 (8): 2-7.
- 2 王雪艳, 相悦丽, 朱旭东, 等. 医疗大数据在公立医院绩效考核中应用的探讨 [J]. 医学信息学杂志, 2022, 43 (1): 45-48, 62.
- 3 国家卫生健康委员会, 国家中医药管理局. 关于加强公立医院运营管理的指导意见 (国卫财务发 [2020] 27 号) [EB/OL]. [2020-12-26]. [http://www.gov.cn/zhengce/zhengceku/2020-12/26/content\\_5573493.htm](http://www.gov.cn/zhengce/zhengceku/2020-12/26/content_5573493.htm).
- 4 梁立. 公立医院运营管理体系构建研究 [J]. 经济师, 2021 (10): 255-256.
- 5 张华峻, 张婉涛, 吴先威. 信息化助推新型医院运营管理 [J]. 中国数字医学, 2019, 14 (5): 76-78.
- 6 王西雯. 医院运营管理体系构建创新路径探究 [J]. 经济管理文摘, 2021 (24): 70-71.
- 7 王韬. 医院绩效及运营管理信息化发展现状分析 [J]. 中国数字医学, 2021, 16 (10): 1-4.
- 8 陈勇, 沈东平, 王国光, 等. 医院成本控制及绩效评价信息管理平台构建 [J]. 医学信息学杂志, 2018, 39 (5): 34-38.
- 9 左铭, 李春伟, 徐琪, 等. 大型医院运营数据中心的应用与实践 [J]. 中国数字医学, 2021, 16 (10): 5-9.
- 10 赵玉虹. 大数据治理助推公立医院精益运营 [J]. 中国卫生人才, 2021 (10): 12-14.
- 11 贺谦, 肖辉, 彭乔立. 基于医院运营数据中心的智能报表系统设计与实现 [J]. 中国卫生信息管理杂志, 2021, 18 (4): 509-513.
- 12 李楠, 王觅也, 郑涛, 等. 医院管理指标数据质量监测体系建设与实践 [J]. 中国数字医学, 2021, 16 (1): 56-59, 116.