中 华 中 医 药 学 会 团 体 标 准

T/CACM XXX—2018

中医真实世界数据
采集操作规范

Standard for collection operation in real world data of Chinese Medicine

（稿件类型：公开征求意见稿）
（本稿完成时间：2018 年 11 月 16 日）

201×-××-××发布　　201×-××-××实施

中华中医药学会 发布

目 次

前言 I
引言 II
1. 范围 1
2. 规范性引用文件 1
3. 术语及定义 1
4. 数据采集原则 2
4.1. 真实性 2
4.2. 完整性 2
4.3. 规范性 2
4.4. 可追溯性 2
5. 构建数据库 2
5.1. 数据库设计原则 2
5.2. 数据库设计 3
5.3. 数据库实施 3
5.4. 数据库测试 3
6. 选择数据源 3
6.1. 选择原则 3
6.1.1. 关键变量可及 3
6.1.2. 观察周期完整 3
6.1.3. 样本量充分 4
6.1.4. 隐私保护恰当 4
6.2. 选择范围 4
6.3. 采集质量控制 4
7. 采集数据 4
7.1. 回顾性数据库采集 4
7.2. 前瞻性数据采集 5
7.3. 健康监护设备数据采集 5
8. 数据汇交 5
9. 数据描述 5
附录 A 研究文件归档与保存 6
参考文献 7

前言

《中医真实世界研究数据采集操作规范》（以下简称“本规范”）按照 GB/T 1.1—2009《标准化工作导则第 1 部分：标准的结构和编写》规定的规则起草。

本规范由中国中医科学院提出。

本规范由中华中医药学会归口。

本规范起草单位：中国中医科学院中医临床基础医学研究所、中国中医科学院、安徽中医药大学第一附属医院。

本规范起草组负责人：谢琪、刘春、刘保延。

主要执笔人：王斌、周洪伟、忻凌、汪涛、曹馨宇、史华新、郭玉峰、张妮楠。

参与指导专家：刘保延、孙塑伦、毛树松、刘建平、温泽淮、刘健、吴大嵘、高蕊、李金学、魏军平、宋军、唐晓颊、边永君、吴圣贤、陶英。

引言

《T/CACM 1051-2017 中医真实世界研究技术规范通则》于 2018 年发布实施，主要内容包括临床研究方案设计、伦理审查、研究注册、质量管理、医学术语规范化应用、模板设计、数据采集、数据预处理、数据分析、研究报告。规范的数据采集是形成可靠临床证据的关键步骤，在真实世界研究的过程中，直接影响临床研究方案的实施效果及数据分析质量。本规范针对中医真实世界研究的特点和我国的实际情况，参照国际真实世界研究的模式、设计理念与方法，在我国相关法律、法规指导下，融合多学科知识，提出中医诊疗实际环境下开展真实世界研究，进行数据采集的操作规程与技术要求，以期有利于促进中医真实世界研究质量的提高，保证研究结果的真实性和科学性。

中医真实世界数据采集操作规范

1. 范围

本规范提出了中医真实世界研究中数据采集的方法、步骤和技术要求，包括：数据采集原则、构建数据库、确定数据源、采集数据、数据汇交、数据描述。可供从事中医真实世界研究的人员使用。

2. 规范性引用文件

下列文件对于本规范的应用是必不可少的。凡是注明日期的引用文件，仅所注明日期的版本适用于本规范。凡是不注明日期的引用文件，其最新版本（包括所有的修改版本）适用于本规范。

GB/T 1.1—2009 标准化工作导则 第 1 部分：标准的结构和编写

WS/T 445—2014 电子病历基本数据集编制规范

《临床试验数据管理工作技术指南》（国家食品药品监督管理总局 2016 年第 112 号通告）

《临床试验的电子数据采集技术指导原则》（国家食品药品监督管理总局 2016 年第 114 号通告）

《T/CACM 1051-2017 中医真实世界研究技术规范通则》

3. 术语及定义

下列术语和定义适用于本规范。

3.1 真实世界数据（Real-world Data ，RWD）

指来自真实医疗环境的患者健康状态相关数据和（或）不同来源的日常医疗健康保健数据。

3.2 真实世界证据（Real-world Evidence，RWE）

指真实世界数据分析得到的干预措施的使用和潜在获益或风险相关的临床证据。

3.3 回顾性数据库（Retrospective Database, RD）

指在研究开始前已经存在的、基于医疗和决策管理目的所形成的数据库，并非针对特定研究问题收集数据而形成。

3.4 电子数据采集（Electronic Data Capture, EDC）

指一种基于计算机网络的用于临床试验数据采集的技术，通过软件、硬件、标准操作程序和人员配置的有机结合，以电子化的形式直接采集和传递临床数据。

4. 数据采集原则

4.1. 真实性

数据采集应符合中医真实世界研究的要求，根据实际情况，记录研究对象的所有真实信息，形成真实世界证据。

4.2. 完整性

是指完整地采集研究所需的全部数据，应保证对研究对象所有诊疗信息及其他相关信息的完整记录。

4.3. 规范性

数据采集者应遵循统一的采集方法和采集流程，确保数据以规范的形式和流程获取。

4.4. 可追溯性

是指所采集数据可以追溯到原始数据，与原始数据保持一致。

5. 构建数据库

5.1. 数据库设计原则

临床研究方案设计具有多样性，每个研究项目的数据收集依赖于临床研究方案。临床研究数据库应保证完整性，并尽量依从数据库范式与约束要求。就特定的研究项目来说，数据库的建立应当以该项目的数据采集表为依据，数据集名称、变量名称、变量类型和变量规则等都应反映在注释数据采集表上。

5.2. 数据库设计

数据库设计至少包括 3 个部分，即：引言、外部设计和结构设计。**引言包括：**编写说明、背景、修订审批记录。**外部设计包括：**标识符和状态、使用该数据库的程序、数据库中表（Table）的命名约定、数据库表中字段（表的域 Column）命名约定、数据库中存储过程（Store Procedure）命名约定、数据库中触发器（Trigger）命名约定、数据库中创建的视图（View）命名约定、数据库中函数（FUNCTION）命名约定、数据库中数据库链接（DATABASE LINK）的命名约定、输入参数及输出参数命名约定、变量命名约定。**结构设计包括：**概念结构设计、逻辑结构设计、表目录、表结构、物理结构设计、设计数据的存取路径、设计数据的存放位置、完整性设计、安全保密设计、用户身份的标识、存取控制。

5.3. 数据库实施

运用数据库管理软件提供的数据语言，根据逻辑设计和物理设计的结果建立数据库，编制与调试应用程序，组织数据入库，并进行试运行。

5.4. 数据库测试

测试数据库之前需制定测试计划书，测试系统功能是否与前期设计及说明书一致。测试内容包括：浏览及录入页面设计，各个访视顺序、访视中的录入表格顺序及每个数据点的顺序；不同用户浏览权限的准确性等。

6. 选择数据源

6.1. 选择原则

6.1.1. 关键变量可及

在开始数据采集前，首先要依据研究方案确定的基线变量、关键变量，选择合适的数据源。所选数据源须包含足够详细的变量，以支撑基线的比较和关键变量的分析。当某一部分的关键变量缺失，就需要考虑从其它数据源中补充数据。

6.1.2. 观察周期完整

所选数据源的时间跨度足够长，应能够覆盖病例完整治疗期与随访期。

6.1.3. 样本量充分

针对真实世界数据混杂性特点，需要足够多的样本量支撑降低混杂因素的处理方法（如倾向性评分等）。

6.1.4. 隐私保护恰当

数据源获取方法及数据的检索策略必须通过伦理审查委员会的审核，对于没有足够的匿名化处理方法达到隐私保护要求的数据源，应舍弃。

6.2. 选择范围

除了研究者设计数据库用于收集研究数据之外，真实世界研究的数据还可来源于电子健康记录（electronic health records, EHR）、社会保障或医疗保险数据库、疾病登记数据库、家庭监护设备数据、健康移动设备的监护数据以及健康相关的纸质记录等。中医真实世界研究中的数据主要来源于以电子病历（Electronic Medical Records, EMR）为核心的医疗卫生服务记录，以门诊病历或者住院病历的形式存储于医院信息系统的数据库中，不仅包括一般人口学信息、诊断记录、处方、医嘱、检查检验报告以及健康移动设备的监护数据等；还包括中医的四诊信息，中医证候等。

6.3. 采集质量控制

完整性，诊疗数据应有生成时间，实验室检查结果应附正常参考值，涉及的干预措施应有相应鉴别信息，诸如中成药记录应附生产厂家及批号。**可溯源性，**记录临床数据的原始值，需附数据源、产生时间及操作者，对数据的任何修改，需附日期和时间、修改原因、操作者。**一致性，**针对多个数据源的数据采集，要保证数据源、过滤条件、纳排标准、术语编码、暴露的测量方法的一致性。

7. 采集数据

7.1. 回顾性数据库采集

医院、体检机构的电子健康记录（EHR），社会保障或医疗保险数据库、疾病登记数据库等数据源的采集采用回顾性数据库的方法。首先要提交回顾性研究方案，并通过伦理审查；其次是提交数据使用申请；然后才能通过专用的检索、查询模块，根据纳入、排除

标准，设置约束条件，利用导出功能，批量导出数据。

7.2. 前瞻性数据采集

前瞻性数据主要包括如临床试验的补充数据，实用性随机对照试验，注册登记研究（Registry），健康调查，公共健康监测等。该类数据在收集之前已有明确的研究目的，制定好了数据采集模板，数据采集一般通过电子数据采集系统来完成，因此数据质量较高。

7.3. 健康监护设备数据采集

数据源包括健康移动监护设备、家庭监护设备等。通常以日志文件或者系统自带的数据库的形式存储相关数据，导出后需要进行加工处理，需要设定唯一标识字段，以便于与其他来源数据关联。

8. 数据汇交

对于不同来源数据库的汇交。首先，建立统一的数据标准，将不同数据库的数据结构进行标准化处理。然后，确定各数据库之间的关联关系，保证采集数据的每一条记录都应有唯一标识码。推荐使用医疗保险手册号或者身份证号作为各数据的唯一标识。

9. 数据描述

对已经采集完成的数据需要进行必要的描述，包括：数据来源、数据量、获取方式等数据特征描述；数据表的定义规则、变量定义规则、隐私保护需要、共享要求等数据库描述；文件格式、存储方式、备份策略、保存时限、数据库说明文档等数据文件描述；数据采集者、质量控制人员、数据分析人员等数据用户描述：。

附录 A 研究文件归档与保存

应保存的档案		研究者	数据所有权单位
1	临床研究单位及主要研究者基本情况	保存	保存
2	数据使用协议	保存（原件）	保存（原件）
3	数据采集表（样表）	保存	保存
4	数据描述报告	保存	保存
5	数据提交报告	保存	保存
6	数据库及说明文档等	保存	保存

参考文献

[1] Junhua Zhang，Boli Zhang. Clinical research of traditional Chinese medicine in big data era [J]. Frontiers of Medicine，2014，8(3)：321-327.

[2] Food and Drug Administration. Use of Real-World Evidence to Support Regulatory Decision-Making for Medical Devices：Guidance for Industry and Food and Drug Administration Staff[EB/OL].[2017-8-31]. https://www.fda.gov/downloads/medicaldevices/deviceregulationandguidance/guidancedocuments/ucm513027.pdf.

[3] 刘保延. 真实世界的中医临床科研范式. 中医杂志[J]. 2013，54(6)：451-455.

[4] 卜擎燕, 熊宁宁, 邹建东等. 从临床研究数据管理角度设计病例报告表[J]. 中国新药杂志，2007，16(5)：339-343.

[5] 孙鑫,谭婧, 唐立等. 重新认识真实世界研究[J]. 中国循证医学杂志，2017，2(17)：126-130.

[6] 谢琪, 江丽杰, 刘保延,等. 开展真实世界中医药效果比较研究的关键问题及对策的探讨[J]. 世界中医药, 2014(1):28-31.

[7] Sherman, R.E., et al., Real-World Evidence - What Is It and What Can It Tell Us? New England Journal of Medicine, 2016. 375(23): 2293.

[8] 詹思延效果比较研究

[9] GavinPowell. 数据库设计入门经典[M]. 清华大学出版社, 2007.

[10]王雯, 刘艳梅, 谭婧,等. 回顾性数据库研究的概念、策划与研究数据库构建[J]. 中国循证医学杂志, 2018(2): 230-237.