

团体标准

T/CRH 0009—2021

生殖健康科学数据质量控制规范

Quality control Specification for Reproductive health science data

2021 - 11 - 22 发布

2022 - 01 - 01 实施

中国生殖健康产业协会

发布

目 次

前言 II

1 范围 1

2 规范性引用文件 1

3 术语和定义 1

4 质量控制要求 2

 4.1 概述 2

 4.2 数据质量管理体系 2

 4.3 数据采集过程 2

 4.4 数据准入 2

 4.5 数据控制 2

 4.6 从运行维护到共享的反馈机制 3

 4.7 全程记录控制 3

 4.8 软硬件配置 3

 4.9 质量评价 3

参考文献 4

前 言

本文件按照GB/T 1.1-2020《标准化工作导则 第1部分：标准化文件的结构和起草规则》的规定起草。

请注意本文件的某些内容可能涉及专利。本文件的发布机构不承担识别专利的责任。

本文件由中国生殖健康产业协会提出并归口。

本文件起草单位：国家卫生健康委科学技术研究所、中国标准化研究院、北京大学第一医院、北京大学第三医院、北京大学公共卫生学院、四川大学华西第二医院、东南大学、首都医科大学附属北京妇产医院、首都医科大学附属北京儿童医院、重庆医科大学附属第一医院、北京理工大学、河南省人口和计划生育科学技术研究院、中国生殖健康产业协会。

本文件起草人：马旭、赵君、杨雪莹、张宏光、齐大逊、代巧云、张月、任冠华、杨慧霞、马京梅、甄秀梅、周虹、李小洪、王棵、王蓓、李光辉、郝婵娟、张俊辉、闫怀志、蒋丽芳、刘秋玲。

本文件由中国生殖健康产业协会标准化技术委员会负责解释。

生殖健康科学数据质量控制规范

1 范围

本文件规定了生殖健康科学数据的质量控制要求。
本文件适用于生殖健康科学数据信息的管理与服务。

2 规范性引用文件

下列文件中的内容通过文中的规范性引用而构成本文件必不可少的条款。其中，注日期的引用文件，仅该日期对应的版本适用于本文件；不注日期的引用文件，其最新版本（包括所有的修改单）适用于本文件。

GB/T 30524-2014 科技平台 元数据注册与管理

3 术语和定义

GB/T 30524-2014界定的及下列术语和定义适用于本文件。

3.1

生殖健康 reproductive health

在生命每个阶段中个体的生殖系统、生殖过程和生殖功能的状况，包括与生殖相关的身体健康和心理健康。

3.2

生殖健康科学数据 scientific data of reproductive health

遵循知情同意原则下，涉及生殖健康的一切数据。

3.3

数据集 dataset

具有同一特征属性的数据集合。

3.4

元数据 metadata

关于数据的数据。

[GB/T 30524-2014，定义3.2]

3.5

质量控制 quality control

质量管理的一部分，致力于满足质量要求。

3.6

数据源 data source

生殖健康科学数据中未经数据采集方标准化整理的原始数据，以及元数据、数据集描述信息、全字段说明文件、全字段示例数据和共性数据等。

4 质量控制要求

4.1 概述

生殖健康科学数据包括标记信息、基础信息、数据特征信息、管理信息等，为保证生殖健康科学数据的完整、精准、有效、安全，质量控制应分环节实施，建立完整的、适宜的质量评价体系，进行识别、度量、监控、预警等一系列管理活动。

4.2 数据质量管理体系

数据质量管理体系是通过有效的数据质量控制手段，进行数据的管理和控制，消除数据质量问题，保障数据的完整性、准确性、及时性和一致性。其实现过程包括建立数据质量控制工作流程、形成规范性文件、依照规范文件实施、并持续改进。数据质量管理体系不是固定的，应随着实际应用逐渐完善。

4.3 数据采集过程

4.3.1 采集手段

根据数据质量要求，选择合理的采集手段。

4.3.2 采集进程跟踪

利用计算机程序在数据采集过程中，通过数据之间应当具有的各种关系（几何关系、数据取值范围、相联关系等等）来限制约束数据可能出现的质量问题，如，将数据质量控制规范与数据库操作规范建立起联动关系，有效保证数据的准确性。

4.4 数据准入

合格和基本合格的数据源获取准入资格，但基本合格的数据源应寻找问题并及时改善数据质量。不合格的数据源可经数据采集方参考原始数据进行核实和完善后，再次进行质量评价，如果质量符合要求，给予准入资格；如仍不符合要求，则不具有准入资格，拒绝进入，并做出相应地标识。

4.5 数据控制

4.5.1 控制要求

对不合格的数据进行标识和改善；对于无法改善的不合格数据，予以剔除；对于改善之后且再一次质量评价为合格的数据，予以准入；对上述的整个过程进行详细的记录。

4.5.2 纠正措施

针对不合格数据找到具体原因。将纠正过程编制形成文件，文件内容应包含数据不合格的原因；确定和实施纠正的措施；采取措施的结果；避免不合格数据再次发生的措施。

4.5.3 改善措施

4.5.3.1 由数据采集方参考原始数据记录进行改善和补充，更改的内容应有记录。

4.5.3.2 由计算机系统校验出的数据问题即时改善，无法改善的数据予以剔除。

4.6 从运行维护到共享的反馈机制

生殖健康科学数据的管理方、采集方和使用方应有专人负责科学数据的管理，并有具体的联系方式，保证完整畅通的信息反馈渠道。

4.7 全程记录控制

与数据质量控制有关的所有过程应进行全程记录，以提供符合要求和质量控制体系的证据。记录应保持清晰、易于识别和检索。记录文件应包含记录的标识、保护、检索、保存期限和处置所需的控制。

4.8 软硬件配置

软、硬件性能和技术指标应具备数据采集、转换、预处理、上传过程中技术要求，使用的软件应具备检验校对的功能。

4.9 质量评价

质量评价方法包括：直接评价法、间接评价法和综合评价法，评价实施流程说明见表1。

表1 数据质量评价实施流程说明

质量评价方法	质量评价实施流程说明
直接评价法	步骤 1：确定质量评价方法，即抽样检查或完全检查； 步骤 2：如运用随机抽样检查法，要确定数据集各要素的重要性，确定抽样比率（如选用完全检查法此步骤可省略）； 步骤 3：确定数据质量评价指标及检测方法； 步骤 4：检测程序的开发； 步骤 5：计算机检测、人工检测或二者相结合进行检测； 步骤 6：数据质量评价； 步骤 7：填写评价报告。 注：经过前5个步骤后，如果质量不合格，应改善后重复上述检查步骤。
间接评价法	步骤 1：收集用于评价数据集的有关资料； 步骤 2：分析评价质量定量元素与非定量元素； 步骤 3：完成评价报告。
综合评价法	步骤 1：收集用于评价数据集的有关资料； 步骤 2：用综合法分析质量定量元素与定性元素； 步骤 3：完成质量评价报告。

参 考 文 献

- [1] TR-REC-063 数据质量管理规范
-