

团 体 标 准

T/CHIA 60-2025

脑肿瘤临床研究大数据采集与质量控制 技术规范

Technical specification of big data collection and quality control for
brain tumor clinical research

2025-10-24 发布

2025-12-01 实施

中国卫生信息与健康医疗大数据学会 发布

目 次

前 言.....	I
1 范围.....	1
2 规范性引用文件.....	1
3 术语和定义.....	1
4 数据采集的方式和频率.....	3
5 数据采集流程.....	4
6 数据质量控制流程.....	7
参考文献.....	12
附录 A.....	13

全国团体标准信息平台

前 言

本标准按照 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。

本标准由首都医科大学附属北京天坛医院提出，归口于中国卫生信息与健康医疗大数据学会。

本标准起草单位：首都医科大学附属北京天坛医院、中国医学科学院北京协和医学院、首都医科大学三博脑科医院、空军军医大学第二附属医院、郑州大学第一附属医院、空军军医大学第一附属医院、华中科技大学同济医学院附属协和医院、山东大学齐鲁医院、广东三九脑科医院、深圳市第二人民医院、东部战区总医院、山西省人民医院、浙江大学医学院附属第一医院、重庆市人民医院、苏州大学附属第一医院、昆明医科大学第一附属医院、宁夏医科大学总医院、复旦大学附属上海华山医院、南昌大学第二附属医院、重庆医科大学附属第一医院、首都医科大学附属北京儿童医院。

本标准主要起草人：张力伟、贾旺、李德岭、季楠、肖丹、单广良、耿静瑶、赵旭洋、胡子豪、施世琦、杨慧珍、邢浩然、席天舒、张宏伟、屈延、闫东明、蒋晓帆、姜晓兵、李刚、朱丹、李维平、马驰原、吉宏明、詹仁雅、吴南、陈罡、杨智勇、马辉、王镛斐、祝新根、孙晓川、葛明、杨亚坤、赵天智、刘献志、罗鹏、曾亦斌、倪石磊、李海南、黄国栋、潘灏、王春红、徐庆生、汪攀、孙亮、龙江、霍显浩、沈明、郭华、杨刚、王佳。

脑肿瘤临床研究大数据采集与质量控制技术规范

1 范围

本标准规定了脑肿瘤临床研究大数据采集与质量控制的基本原则，包括数据采集方式和流程、数据质量评价指标以及数据质量控制的工作流程。

本标准旨在对脑肿瘤临床研究大数据的采集和质量控制进行规范化要求，适用于拟开展脑肿瘤临床研究大数据采集和质量控制的各类机构和组织。本标准也可为其他病种的临床研究大数据采集和质量控制提供参考。目前，全国已有 50 余家神经外科技术优势单位应用本标准进行脑肿瘤临床数据的采集与质量控制工作。

2 规范性引用文件

下列文件对于本标准的应用是必不可少的。凡是注日期的引用文件，仅注日期的版本适用于本标准；凡是不注日期的引用文件，其最新版本（包括所有修改单）适用于本标准。

GB/T 36344-2018 信息技术 数据质量评价指标

3 术语和定义

下列术语和定义适用于本标准。

3.1

脑肿瘤 brain tumor

又称颅内肿瘤，是指发生于颅腔内的神经系统肿瘤，按照其原发部位可分为两类：起源于颅内组织的肿瘤称为原发性颅内肿瘤，从身体远隔部位转移或由邻近部位延伸至颅内的肿瘤称为继发性颅内肿瘤。

3.2

临床研究 clinical research

指医疗卫生机构开展的，以人（个体或群体）为研究对象，不以药品、医疗器械（含体外诊断试剂）等产品注册为目的，研究疾病的病因、诊断、治疗、康复、预后、预防、控制及健康维护等的活动。

3.3

大数据采集 big data collection

从数据源中得到大量多源异构的原始数据，通过标准化处理流程将其转化为满足数据使用与共享需求的过程。

3.4

质量控制 quality control

是指在临床研究过程中，为确保临床研究所有相关活动符合质量要求而实施的技术和活动。

3.5

病例报告表 case report form, CRF

是指按照研究方案要求设计，记录患者相关信息的纸质或者电子文件。

3.6

电子数据采集系统 electronic data capture (EDC) system

是一种基于计算机网络的用于临床研究数据采集的技术，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研究数据。

3.7

前置机 front-end server

是指位于网络架构中的前端，负责处理用户请求、接收数据、转发流量和提供服务的计算机或服务器。在临床研究领域，其通常指的是一种用于数据采集和预处理的计算机或服务器。

3.8

采集频率 collection frequency

在一定时间内进行信息或数据采集（3.3）的次数或频率。

3.9

数据字典 data dictionary

是临床研究使用的所有数据元素的定义的集合。通过定义数据项、数据结构、数据流、数据存储、处理逻辑等，对临床研究中使用的数据进行详细的说明。

3.10

数据采集工具 data collection tool

实现数据采集（3.3）并对采集到的数据进行整理、转换和存储的软件或硬件设备。

3.11

数据分析 data analysis

通过对采集到的数据进行处理、整理和特征提取，从中提取有价值的信息并将其合成为知识的过程。

3.12

质询 query

在临床研究的数据质量控制中，质询是指针对数据质量相关的事项，通过书面或口头的方式提出问题、获取信息、解决问题的行为或过程。

3.13

数据质量评价指标 data quality evaluation indicator

是指对数据规范性、完整性、准确性、一致性等方面进行评价和衡量的标准化指标。

4 数据采集的方式和频率

4.1 数据采集方式

基于数据采集工具，数据采集方式可分为纸版 CRF 采集、EDC 系统采集、前置机采集等；基于数据采集的实时性，数据采集方式可分为在线采集和离线采集。不同的采集方式对应不同的准备工作，同时也存在各自的优缺点和适用场景，在数据采集过程中应根据实际情况灵活选择合适的采集方式。

4.1.1 基于数据采集工具分类

4.1.1.1 纸版 CRF 采集

根据研究目的和数据需求，设计 CRF 表的各个字段名称、内容，经过必要的质量控制与排版后，打印为纸版 CRF。例如，采集脑肿瘤数据时，需设计字段以涵盖肿瘤特征、神经系统查体、治疗信息等关键内容，且需保证表述通俗易懂、清晰易读。

由患者或调查员手工填写纸版 CRF，之后再由专人将纸版 CRF 中的数据录入电子数据库。

为便于实际操作中规范执行上述 CRF 表的设计与数据采集要求，本标准提供脑肿瘤患者“基本情况”部分的数据采集 CRF 模板供参考，详见规范性附录。

4.1.1.2 EDC 系统采集

根据研究目的和数据需求，首先确定 EDC 系统的功能和特性，随后进行系统开发，包括数据库设计、界面开发、系统逻辑编写和电子化调查表植入等步骤，完成系统开发后，还应进行系统功能与安全性测试。在 EDC 系统编写字段逻辑时，需考虑脑肿瘤字段间的跳转逻辑，例如，患者在表单中选择存在失语状况，系统应激活“失语类型”和“失语程度”字段以供填写。

由研究人员或调查员登录 EDC 系统，通过填写电子 CRF 输入数据。EDC 系统对数据进行逻辑核查与验证后，可自动将其存储到数据库中。

4.1.1.3 前置机采集

根据设计方案开发前置机系统，包括程序代码的编写和软硬件环境的配置，后进行系统的测试与调试，合格后前置机系统便可投入使用。

对前置机系统完成网络连接、数据采集与解析规则、存储路径等参数设置后，即可与数据源进行连接并开启数据采集功能，将采集到的数据传输至数据库中进行存储。

4.1.2 基于数据采集的实时性分类

4.1.2.1 在线采集

在线采集指数据采集时通过互联网接入数据库，例如在线 EDC 系统、在线前置机系统等。

4.1.2.2 离线采集

离线采集指首先通过离线数据采集设备采集数据，采集设备包括纸版 CRF、离线软件、离线前置机等，后通过人工操作将采集到的数据上传到计算机或数据库中，实现电子化存储。

4.2 数据采集频率

研究人员应根据研究方案和不同的数据采集方式，合理设定采集数据的频率。例如，基于前置机采集数据，可设定在每天下班后从医院的信息系统推送数据到前置机，以减轻工作时间医疗机构的网络传输压力。

5 数据采集流程

5.1 编制数据字典

数据采集实施前,应基于研究方案和数据统计分析的需求编制数据字典,数据字典应满足以下要求:

5.1.1 明确定义

数据字典应对每个数据元素进行明确的定义,包括指标名称、数据类型、数据结构、数据长度、取值范围、格式要求、处理逻辑等。对于脑肿瘤相关数据元素,需要细化并明确其定义,如肿瘤类型、肿瘤大小、位置、分级等关键信息。

5.1.2 统一命名

数据字典应使用统一的命名规则,以确保数据的一致性和可读性。对于脑肿瘤相关变量,例如肿瘤类型,建议参考国际通用肿瘤分类标准,包括《国际疾病分类肿瘤学专辑第三版》(ICD-O-3)和《WHO 2016 中枢神经系统肿瘤分类》。

5.1.3 准确性

数据字典应准确反映数据的含义和属性,避免模糊或冲突的定义。

5.1.4 标准化

数据字典应符合现行或通用的卫生行业的数据集标准和规范,确保数据的一致性和互操作性。

5.1.5 跨科室数据考量

数据字典可考虑采集其他科室的相关数据,例如肿瘤科、放射科等。

5.2 开发数据采集工具

在进行数据采集前,需要开发相应的数据采集工具,主要包括纸版 CRF 表、EDC 系统和前置机系统这三种采集工具,可以根据实际情况进行灵活选择。下面是各采集工具的设计要点:

5.2.1 纸版 CRF

——纸版 CRF 的设计应符合数据采集的需求和标准,应覆盖临床研究方案涉及的所有指标,且指标设置应满足统计分析的要求。

——纸版 CRF 的设计应考虑填写者的背景和知识水平,尽量使用易于理解的语言,避免使用过于专业化或技术化的词汇。

——纸版 CRF 的问题设计应符合逻辑顺序,例如先填写基本信息,再填写病历资料,最后填写辅助检查结果等。

——纸版 CRF 问题设计时应避免诱导性或矛盾性的问题,且应符合相关法规要求,如伦理委员会审批方案或修正案等。

——纸版 CRF 中的问题设计时,应尽量避免设置过多的跳转关系,如确实需要设置跳转关系,应确保逻辑清晰,简化跳转路径。

——纸版 CRF 问题设计时,应尽量将复合问句拆分成系列单句,单次只问一个问题,便于填写者理解问题的核心。

——纸版 CRF 中敏感问题(如个人隐私、心理状况等)应尽量置后,以避免影响填写者的情绪或对填写者造成困扰。

——纸版 CRF 中当使用数据编码时,相同类型问题的数据编码次序应一致,如是非题的答案选项统一设置为“1=是,2=否”,而不应有的指标设置为“1=是,2=否”、有的指标设置为“1=是,0=否”。

——纸版 CRF 应允许填写者进行修改和更正，同时设置记录修改时间和原始数据的位置。

5.2.2 EDC 系统

EDC 系统的设计除满足纸版 CRF 设计的要点之外，还应关注以下设计要点：

——EDC 系统应具有生成符合临床试验方案的电子 CRF 的功能，最好能够支持上传、下载和在线编辑的电子 CRF 文件。

——EDC 系统应具备易于使用和操作的用户界面，减少用户的学习成本。

——EDC 系统应提供方便的数据录入界面，并能进行数据实时验证和自动逻辑核查，例如数据值的范围、逻辑关系等，防止用户录入错误或不合理的数据。

——EDC 系统应配置功能模块，用于产生和发布临床研究数据质疑，以便数据管理员或临床监察员能够对数据填写员提出质询并及时解决问题。

——EDC 系统对于不同的账户应设置不同的权限控制功能，以明确各个账户的任务分工。

——EDC 系统应具备数据备份和恢复的机制，能够定期备份数据，并具备可靠的数据恢复机制，确保数据的安全性和可靠性。

——EDC 系统应具备数据加密、权限管理和访问控制等安全机制，确保数据的安全性和完整性。

——EDC 系统应能够记录数据的每一次修改、更新、审核、稽查和签名等操作，并且可以通过授权的方式查看。

——EDC 系统应具有定期更新机制，以灵活适应临床试验方案的调整和技术的发展。例如，在开展垂体瘤相关的临床研究时，系统应能新增泌乳素等内分泌化验指标，以满足垂体瘤临床研究的数据采集需求。

5.2.3 前置机系统

——前置机系统应能接收和处理从不同信息系统传输的数据，包括数据的接收、解析、预处理和转换等。

——前置机系统应能进行数据校验和逻辑性检查，确保数据的准确性和完整性。

——前置机系统应能安全地传输并存储采集到的数据，并具备定期备份和恢复的机制，确保数据的安全性和可靠性。

——前置机系统应该具备权限管理和访问控制机制，确保只有授权人员可以访问和修改数据，避免数据的不合规使用。

——前置机系统应建立灵活的更新策略。例如，在需要的情况下，前置机系统可新增接口，以应对不同来源信息系统的格式。

5.3 采集数据

数据采集的方式主要包括人工采集和前置机自动采集两种，各种采集方式的要点如下：

5.3.1 人工采集

——采集人员应当经过培训，了解数据采集的流程，遵循数据采集的要求和规范。

——采集人员应熟悉采集数据的内容和要求，最好具备相关领域的专业知识。对于脑肿瘤的人工采集，鉴于其专业性和复杂性，最好是由具有丰富经验和专业知识的外科医生来执行。

——采集人员应具备良好的数据录入和计算机操作能力。

——采集人员应具备良好的沟通能力，能够与患者进行有效的沟通。

——采集过程应有相应的监督和审核机制，能够及时发现和解决问题。

——应确保数据采集过程的可追溯性，方便后续数据的溯源。

5.3.2 前置机自动采集

- 信息系统应具备数据推送功能，并能确保推送数据的准确性和完整性。
- 信息系统应确保推送至前置机的数据与源数据一致。
- 信息传输应具备一定的安全性，确保数据在传输过程中不被篡改或泄露。
- 数据应及时推送至前置机，以便前置机及时接收数据并进行后续操作。
- 数据推送过程应做到定期化、自动化。
- 数据推送过程应有相应的日志记录和异常处理机制，例如发现、记录以及及时处理网络中断、系统故障等问题。
- 推送的数据应具有可追溯性，即能够追溯到数据的来源和推送的全过程。

5.4 数据质量核查

5.4.1 系统自动核查

EDC 系统或前置机系统通过逻辑一致性检验等方式对采集到的数据进行质量核查，如格式核查、异常值核查、逻辑核查等。

5.4.2 人工纠错及补录

在自动核查的基础上，对不便于开展自动核查和纠错的情况，例如数据形式为开放文本，杂乱无序，无法通过程序进行统一判断的非结构化字段，可开展人工核查、纠错和数据补录。

5.5 数据存储

数据质量核查完成后，需要对采集数据进行存储，主要有硬件存储和软件存储两种方式。用户可以在综合考虑数据量大小、访问频率、共享需求、安全要求及使用场景等因素的基础上，根据实际情况灵活选择存储方式。

5.5.1 硬件存储

硬件存储是指使用物理设备来存储数据的方法，依赖具体的物理介质来承载数据。常见的硬件存储设备包括：单机电脑存储、移动介质存储（U 盘、光盘、硬盘）、服务器存储等。

当仅小规模个人使用，且数据量不大、无需多用户共享时，用户可选择单机电脑存储，即通过个人电脑的内部存储设备来存储数据。用户可直接在文件资源管理器中创建文件夹或文件，并将数据保存到这些位置。

当存在离线保存数据或存在偶尔转移数据的需求时，用户可选择移动介质存储，即通过物理设备来存储数据。对于像脑肿瘤数据这类包含庞大基因序列、医学影像等信息的大容量数据，用户可选择稳定、可靠、大容量的移动硬盘；对于小容量数据的临时传输，U 盘、光盘等移动介质则更为合适，以 U 盘存储为例，用户可以将数据文件拖拽至插入计算机的 U 盘中，或者使用操作系统提供的“复制”和“粘贴”功能将数据保存到 U 盘。

当有多用户在固定网络环境下工作，需要集中管理和共享数据时，用户可选择服务器存储，即利用专门设计的服务器设备来集中存储数据。数据可通过网络传输到服务器并存储在其硬盘阵列中，便于集中维护和访问。

5.5.2 软件存储

软件存储涉及使用软件解决方案来管理数据存储，侧重于通过程序和网络技术实现数据的组织与维护。常见的软件存储方法包括：数据库管理系统、云存储等。

当面临数据量较大、结构复杂，且需要进行高效管理、频繁查询以及多用户并发操作的场景时，用户可选择数据库管理系统，即通过提供结构化的数据存储方式来存储数据。用户首先需要创建数据库和

表结构,然后将数据填入到表单中并存储。常见的数据库管理系统包括 Oracle、Access、MySQL 等。

当需要跨地点访问数据、方便数据分享,且不想投入过多精力维护硬件设备时,用户可基于互联网进行数据存储,即云存储。用户可以通过云存储服务提供商提供的客户端软件将数据上传到云端服务器存储。

5.6 数据安全

5.6.1 数据采集安全

- 确保数据来源的合法性,严格审核数据采集的法律和伦理依据。
- 在数据采集过程中获取数据主体的知情同意,明确告知数据使用目的和范围。
- 严格遵守数据采集过程中的合规性要求,确保采集行为符合相关法律法规。

5.6.2 数据存储安全

- 建立数据分类分级管理制度,根据数据敏感程度制定差异化存储策略。
- 实施严格的访问控制机制,限制数据访问权限,防止未经授权的数据操作。
- 对敏感数据采用加密存储技术,如对称加密、非对称加密、哈希加密等,保护数据的机密性和完整性。
- 建立数据备份与恢复机制,确保数据的灾难恢复能力。

5.6.3 数据传输安全

- 采用加密传输技术,如端到端加密等,保护数据在传输过程中不被非法获取和篡改。
- 实施数据完整性校验机制,确保传输数据的准确性和一致性。
- 遵守安全通信协议,选择经过验证的安全传输方案。
- 严格遵守跨境数据传输合规要求,确保数据跨境传输的合法性。

5.6.4 数据使用安全

- 建立严格的数据使用权限管理体系,精确控制数据访问和使用范围。
- 对敏感数据采用脱敏技术,保护个人隐私和敏感信息。
- 实施数据使用痕迹追踪机制,记录和监控数据使用行为。
- 部署数据防泄露解决方案,防止未经授权的数据泄露风险。

5.6.5 数据销毁安全

- 制定安全删除程序,确保数据彻底从系统中删除且不可恢复。
- 对存储介质采用物理销毁方法,彻底消除数据残留。
- 清除数据销毁过程中的所有痕迹,防止数据重建。

6 数据质量控制流程

6.1 制定计划

在数据质量控制工作具体实施前,应制定相应的工作计划,计划应满足如下要求:

6.1.1 目的明确

应明确数据质量控制的目的和目标,确保数据质量控制工作能够达到预期效果。

6.1.2 方式合理

应根据临床研究数据采集与管理的具体情况，选择合适的数据质量控制方式和方法。本标准提供三种数据质量控制的方式。

6.1.2.1 电子质控

通过在 EDC 系统中设置数据必填项、填写值范围、填写字数限制等方式进行数据质量控制。

6.1.2.2 单位质控

由数据质控员在数据填报单位开展，通过核查填报数据质量、督促单位回复质询并整改数据等工作进行的数据质量控制。

6.1.2.3 中心质控

由临床研究的数据管理员或数据质控员，对数据质量进行中心化的统计分析、质量评估，必要时可撰写数据质量评估报告并反馈。

通常情况下多采用电子质控、单位质控、中心质控相结合的方式开展临床研究数据的质量控制工作。

6.1.3 周期合适

应根据数据采集的特点，确定合理的数据质量控制工作周期、频率和时机，确保及时发现和纠正数据质量问题。本标准提供三种数据质量控制的工作周期。

实时质控：在数据采集、处理和使用过程中实时监测和控制数据质量。通过设定数据必填项、填写值范围等对数据进行实时监测。

定期评估：定期对数据质量进行评估和分析，以了解数据质量的整体状况和存在的问题，并制定相应的改进措施。定期评估可以通过抽样调查、数据抽取和统计分析等方法，对数据质量进行量化评估和分析。

基于需求制定周期：根据数据采集的特点，制定相应的数据质量控制周期。例如，对于每天上传数据的前置机，需每日进行电子质控，以确保数据的完整性和准确性；对于每周上传数据的前置机，则可每周进行一次电子质控。

6.1.4 指标可衡量

应制定可衡量的数据质量指标，用于评估数据质量的水平和改进的效果，确保对数据质量进行有效监控和管理。

6.1.5 人员分工明确

应明确相关人员的职责和分工，确保每个环节的责任和任务清晰明确、数据质量控制工作能否顺利协调和推进。

6.2 数据质量分析

数据质量分析主要是通过计算数据质量评价指标，分析、查找可能存在的问题，并基于数据整改的结果，进行数据质量的再评估。因此，数据分析过程可能会进行多轮数据质量评估、整改、再评估，而不是单线条的一次数据质量评估。

6.2.1 选择并计算数据质量评价指标

数据质量评价指标是数据质量的量化体现，应满足以下要求：

——基于数据的特点，选择合适的指标，确保计算结果能够准确地描述问题的本质。

——指标的计算公式应清晰、明确，不因数据分析人员的变更而产生歧义。

——指标的计算方法应可重复，在多轮数据质量评估时应保持一致。

对于脑肿瘤的临床研究大数据采集，可从多个维度评价数据的质量，主要包括：数据采集的连续性、波动率、准确率、及时性、指标填写率、填报率、随访率等。在实际工作中，可根据数据采集与质量控制的实际需求，灵活选择包括但不限于上述若干指标的组合作为相应的数据质量评价指标体系。

6.2.1.1 连续性

连续性用于评估数据采集的连续状态。操作上，应先依据实际情况定义每个时间单位（例如每天、每周、每月等）的最小数据采集量，然后可采用图表的方式分析、展示每个时间单位实际采集的数据量，也可采用二分类指标（“是/否”）评价数据采集是否连续。

每个时间单位内数据采集量应大于 0。

6.2.1.2 波动率

波动率用于量化数据采集量的变化幅度。波动率的计算首先应选取过去 N 年（通常 $N \geq 3$ ）同一时期（如相同月份、周次、日期等）的数据采集量，计算其算数平均值；再用当前数据采集量与历史均值的差值进行波动率评价。

$$\text{波动率} = (\text{当前数据量} - \text{历史同期均值}) / \text{历史同期均值} \times 100\%$$

不同年份同期数据采集量的波动率增幅应控制在 20% 以内，降幅应不超过 10%。

6.2.1.3 准确率

准确率用于评价采集数据与实际数据的一致程度。准确率的计算应选取采集数据集与实际数据集进行比对，以两者相符的数据量占总数据量的比例作为评价依据。

$$\text{准确率} = (\text{一致的数据量} / \text{所有数据量}) \times 100\%$$

评价脑肿瘤患者诊断的准确率时，可将采集到的诊断名称与患者的病理报告进行比对，计算二者相符的患者数占所有患者数的比例。脑肿瘤患者的病理诊断准确性应大于 95%。

6.2.1.4 及时性

及时性用于评价数据采集的及时程度，包含延迟时间与延迟率两个维度。计算时，应先确定每个病例信息采集的延迟时间（即实际信息采集日期减去应采集日期）；再计算延迟时间的最小值、最大值、平均数、标准差、中位数、四分位数等指标，也可通过设定延迟时间阈值来界定是否延迟，进而评价延迟率。

$$\text{信息采集延迟时间} = \text{实际信息采集时间} - \text{应采集时间}$$

$$\text{延迟率} = (\text{延迟的患者数量} / \text{所有患者数量}) \times 100\%$$

患者信息填报的延迟时间一般不应超过 14 天。

6.2.1.5 指标填写率

指标填写率用于量化数据采集的完整程度，通过实际填写字段占应填写字段的比例进行评价。结合实际情况，可分别计算指标总填写率、必填项填写率等。

$$\text{指标总填写率} = (\text{已填写字段总数} / \text{应填写字段总数}) \times 100\%$$

$$\text{必填项填写率} = (\text{已填写的必填项字段总数} / \text{应填写的必填项字段总数}) \times 100\%$$

要求每个病例必填项填写率应达到 100%，非必填项尽可能填写完整。

6.2.1.6 填报率

填报率用于量化数据采集的漏报情况。填报率的计算应先明确应填报的数据总量，再统计实际采集的数据量，两者的比值即为填报率。

填报率=（实际填报的数据量/应填报的数据量）×100%

病例填报率应大于 95%。

6.2.1.7 随访率

随访率用于评估对需随访的患者有效跟踪情况。随访率的计算应统计已完成随访的患者数量及需要随访的患者总数，两者的比值即为随访率。

随访率=（已随访患者的数量/需要随访的患者数量）×100%

可根据研究方案中对随访时间点的定义，分别计算不同时点的随访率，如患者手术后 3 个月的首次随访率、1 年随访率等。要求首次随访率应大于 90%。

6.2.2 撰写数据质量评估报告

数据质量评估报告是数据质量评价指标的集合，用于从多个维度展示数据质量。报告应满足以下要求：

- 报告应该清晰地表达结果和问题，避免模糊和含糊不清的表述。
- 报告应该有合理的结构和逻辑，避免结构异常和逻辑混乱。
- 报告应该言简意赅，避免冗长和重复的内容，突出重点和关键信息。
- 报告应该符合规范的格式和排版要求，包括字体、字号、段落和图表的使用等。

6.3 对数据质量提出质询

基于数据质量分析的结果，对可能存在质量问题的数据进行质询，促进及时整改、提高数据质量。质询应遵循以下要点：

- 质询应明确目的，即为了解数据质量问题的根本原因、寻找解决方案或改进措施。
- 质询应提出明确的问题，以确保能够获得有用的信息和反馈。
- 质询应客观中立，不带有个人偏见或主观判断。
- 质询应及时进行，以便尽快采取相应措施进行纠正和改进。

6.4 数据整改

根据数据质量评估结果和质询，应及时查找问题、分析原因，提出并实施数据质量改进方案。数据整改应遵循以下要点：

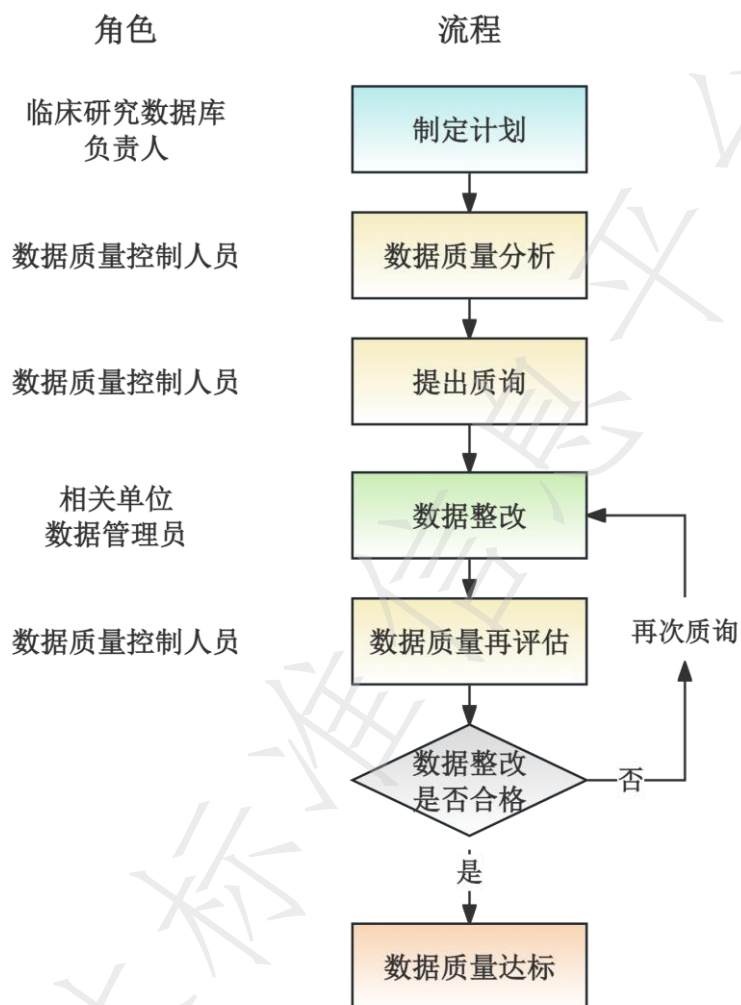
- 数据整改应确保数据的准确性，即确保数据的真实、完整和正确。
- 数据整改应及时进行，尽快纠正数据质量问题，以避免对后续数据使用产生不良影响。
- 数据整改应全面覆盖受影响的数据，不仅仅修正表面上的错误，还应找出根本原因并进行改进。
- 数据整改应具有可追溯性，即能够追溯到问题的发生原因、整改措施的实施过程和结果，以便进行监督和评估。
- 数据整改后应保留修改时间与签名，并保证原内容清晰可辨。

6.5 数据质量再评估

对于整改后数据，应再次进行数据质量评估，以确保数据整改达到既定的效果。数据质量再评估应遵循以下要点：

- 再评估应明确评估的目标和范围，确保评估的内容和重点与数据质量控制的目标一致。
- 再评估应全面覆盖数据质量控制的各个方面，包括数据收集、存储、处理、传输等环节，以全面了解数据质量问题的修复情况。
- 再评估应准确、客观地评估数据质量控制的效果和成果，确保评估结果真实可信。
- 再评估的结果应能够为数据质量控制工作的持续改进提供指导和建议，包括识别问题、找到改进措施等。

数据质量控制各环节角色与流程示意图如下所示。



参考文献

- [1] 王忠诚.王忠诚神经外科学.第2版[M].湖北科学技术出版社,2015.
- [2] 中华人民共和国国家卫生健康委员会.关于印发医疗卫生机构开展研究者发起的临床研究管理办法的通知[Z].国卫科教发〔2024〕32号,2024-09-26.
- [3] GB/T 36625.3-2021 智慧城市 数据融合 第3部分:数据采集规范
- [4] 药物临床试验质量管理规范(国家食品药品监督管理局令第3号)
- [5] 王珊,萨师煊.数据库系统概论.第5版[M].高等教育出版社,2014.
- [6] GB/T 35295-2017 信息技术 大数据 术语
- [7] 刘萍,王炜,李红星.行政管理学[M].经济科学出版社,2008.
- [8] GB/T 36344-2018 信息技术 数据质量评价指标

附录 A

(规范性附录)

脑肿瘤患者基本情况 GRF 表模板

1.患者类型:

1-门诊患者 2-住院患者 3-急诊患者 4-其他

1.1 如果为“门诊患者”，则门诊号：_____

1.2 如果为“门诊患者”，则门诊就诊时间：□□□□年□□月□□日

1.3 如果为“门诊患者”，则门诊决策医师为：_____

1.4 如果为“住院患者”，则住院号：_____

1.5 如果为“住院患者”，则住院次数：□□（次）

1.6 如果为“住院患者”，则本次就诊是否手术 1-是 2-否

1.7 如果为“急诊患者”，则本次就诊是否手术 1-是 2-否

2.医疗机构名称：_____

3.姓名：_____

4.性别:

0-未知的性别 1-男性 2-女性 9-未说明的性别

5.出生日期：□□□□年□□月□□日

6.身份证件号码：□□□□□□□□□□□□□□□□□□

7.年龄：□□□（岁）

8.婚姻状况:

10-未婚 20-已婚 21-初婚 22-再婚 23-复婚 30-丧偶 40-离婚

90-未说明的婚姻状况

9.电话号码：_____

10.民族:

1-汉族 99-其他

11.职业类别:

1-国家机关、党群组织、企事业单位负责人 2-各类专业技术人员

3-办事人员及有关人员 4-商业及服务业人员 5-农、林、牧、渔、水利业生产人员

6-生产、运输人员及有关人员 7-军人 8-不便分类的其他人员 9-无业者

10-离/退休 11-学生 98-不详

12.学历:

- 1-大学本、专科或以上 2-高中（包括中专）3-初中 4-小学 5-文盲
98-不详

13.籍贯-省（自治区、直辖市）：_____

14.籍贯-市（地区、州）：_____

15.常住地址-省（自治区、直辖市）：_____

16.常住地址-市（地区、州）：_____

17.户籍地址-省（自治区、直辖市）：_____

18.户籍地址-市（地区、州）：_____

19.入院途径:

- 1-门诊入院 2-急诊入院 3-其他科室或其他医疗机构转入 4-其他

20.入院科室：_____

21.入院日期：□□□□年□□月□□日

22.出院日期：□□□□年□□月□□日

23.主管医师姓名：_____

24.住院期间是否死亡:

- 1-是 2-否

24.1 如果“是”，则死亡日期为：□□□□年□□月□□日

24.2 如果“是”，则根本死因为：_____

25.是否药物过敏:

- 1-是 2-否

25.1 如果“是”，则过敏药物为：_____

26.ABO 血型代码:

- 1-A 2-B 3-O 4-AB 5-不详

27.Rh 血型代码:

- 1-Rh 阴性 2-Rh 阳性 3-不详

28.是否输血:

- 1-是 2-否

28.1 如果“是”，则输血方式为：_____

28.2 如果“是”，则输血情况为：_____

红细胞：_____ (ml)

血小板：_____ (ml)

血浆：_____ (ml)

全血：_____ (ml)

其他：_____ (ml)

29.入院日常生活能力评定量表 (Barthel) 得分：□□□

30.出院日常生活能力评定量表 (Barthel) 得分：□□□

31.住院费用金额：_____ (元)

32.医疗费用来源类别：

1-城镇职工基本医疗保险 2-城镇居民基本医疗保险 3-新型农村合作医疗保险

4-贫困救助 5-商业医疗保险 6-全公费 7-全自费 8-基金救助

9-其他社会保障 10-其他

33.脑肿瘤诊断名称：_____

34.脑肿瘤侧别：

0-不详 1-右侧 2-左侧 3-一侧受累但无法明确 4-双侧 5-中线结构

35.脑肿瘤受累部位：_____

36.治疗结果：

1-治愈 2-好转 3-稳定 4-恶化 5-死亡 9-其他

37.病理号：_____

38.院内感染诊断：

1-无 2-中枢神经系统感染 3-呼吸系统感染 4-泌尿系统感染 5-消化系统感染

6-其他系统感染

39.其他诊断描述：_____

40.入院病情：

1-有 2-临床未确定 3-情况不明 4-无