**团体标准编制说明**

《队列数据管理 数据采集质控要求》

(阶段：征求意见稿)

一、团体标准任务来源、编制背景、目的和意义

（一）任务来源

本标准根据2024年10月12日北京慢性病防治与健康教育研究会下达的《北京慢性病防治与健康教育研究会团体标准立项公告》，严格按照《医学队列平台数据技术》要求，由北京大学第六医院负责此项标准的起草制定工作。

（二）编制背景

《队列数据管理 数据采集质控要求》的编制背景主要是由于国内外在队列研究标准方面普遍缺乏适用于跨病种队列数据融合的质量标准体系。目前的研究环境中，缺乏针对跨病种队列数据融合前各单病种数据源的质量评价，也缺少对融合后形成的数据完整性以及应用的科学性等方面的系统评价。这导致各类专病及自然人群队列间的融合交叉研究在质量评价工作中面临困难。因此，编制这一标准旨在建立一个科学、系统的质量标准体系，以提升跨病种队列研究的整体质量和可信度。

（三）编制目的

《队列数据管理 数据采集质控要求》编制的目的是建立一套科学、系统的质量标准体系，以规范跨病种队列数据的采集和质量控制，确保数据的完整性、准确性和可重复性。

（三）编制意义

《队列数据管理 数据采集质控要求》的编制意义在于为跨病种队列研究提供明确的质量标准和规范，确保数据采集和处理过程的科学性和严谨性。这一标准的实施将有助于提高数据的可靠性和有效性，支持临床研究和公共卫生领域的科学决策。同时，通过建立统一的质量控制要求，可以减少研究中的不确定性和误差，提高不同队列之间数据融合的可比性与适用性。

二、工作简况

（一）起草单位及主要起草人

本标准由北京大学第六医院主要负责起草，天津市安定医院、北京大学、中国医学科学院肿瘤医院、山东大学齐鲁医院、北京大学第一医院、中国电子技术标准化研究院、中国疾病预防控制研究中心参与，专家团队涵盖了精神、神经、肿瘤、消化、心血管、疾病监测以及自然人群等领域。

|  |  |  |  |
| --- | --- | --- | --- |
| 姓名 | 专业 | 职称 | 工作单位 |
| 刘肇瑞 | 精神障碍流行病学 | 研究员 | 北京大学第六医院 |
| 徐广明 | 社会精神病学 | 主任医师 | 天津市安定医院 |
| 尹慧芳 | 精神病与精神卫生学 | 副主任医师 | 天津市安定医院 |
| 黄雨 | 计算机软件与理论 | 研究员 | 北京大学 |
| 黄悦勤 | 社会精神病学 | 教授 | 北京大学第六医院 |
| 邓咏妍 | 社会精神病学 | 无 | 北京大学第六医院 |
| 李明慧 | 社会精神病学 | 助理研究员 | 天津市安定医院 |
| 魏文强 | 流行病学 | 研究员 | 中国医学科学院肿瘤医院 |
| 孙可欣 | 流行病学 | 副研究员 | 中国医学科学院肿瘤医院 |
| 吕明 | 流行病学 | 教授 | 山东大学齐鲁医院 |
| 李航 | 皮肤病与图像学 | 主任医师、教授 | 北京大学第一医院 |
| 李瑞琪 | 控制科学与工程 | 高级工程师 | 中国电子技术标准化研究院 |
| 陈园生 | 公共卫生 | 研究员 | 中国疾病预防控制中心 |
| 杨孝荣 | 循证医学 | 副研究员 | 山东大学齐鲁医院 |
| 陈浩 | 流行病学 | 助理研究员 | 山东大学齐鲁医院 |
| 白倩倩 | 皮肤病与性病学 | 主治医师 | 北京大学第一医院 |
| 王悦 | 公共卫生 | 助理研究员 | 中国疾病预防控制中心 |

（二）主要编制过程

2024年6月，北京大学第六医院同天津市安定医院、北京大学、中国医学科学院肿瘤医院、山东大学齐鲁医院、北京大学第一医院、中国电子技术标准化研究院、中国疾病预防控制研究中心等有关起草单位启动了标准前期研究项目“跨队列研究数据融合质量控制标准研发”,对跨队列研究数据融合工作的流程和质控环节展开调研，并撰写了调研报告，明确了跨队列研究数据融合工作的质量控制应基于工作流程开展，包括原始队列数据采集、原始队列数据元定义、队列数据融合三个步骤，同时对于各环节的质量控制报告应有统一要求。

2024年7月，北京大学第六医院会同其他起草单位开展准备团体标准立项相关材料工作，一同针对制定《队列数据管理 数据采集质控要求》团体标准的具体工作进行了部署，确定了总体工作方案，并组建了标准起草小组。

2023年8月-2024年9月，标准编写小组按照标准的修订程序进行标准编制，形成标准初稿，并在起草单位内部多次讨论和修改，形成了第一版的标准初稿，提交北京慢性病防治与健康教育研究会进行立项审查。

2024年9月，根据北京慢性病防治与健康教育研究会发布的“北京慢性病防治与健康教育研究会团体标准立项公告”，北京大学第六医院同其他起草单位向北京慢性病防治与健康教育研究会提交申请，明确了《队列数据管理 数据采集质控要求》编写工作组的具体成员，并确定了参编单位和参编人员排序。

2024年10月，北京慢性病防治与健康教育研究会同意该标准立项。立项下达后，北京大学第六医院同其他起草单位进一步对第一版《队列数据管理 数据采集质控要求》标准初稿进行讨论和修改，同时修订的还包括北京慢性病防治与健康教育研究会关于队列数据管理中的其他三项标准。最终形成了标准初稿，并提交北京慢性病防治与健康教育研究会进行初稿审核。

三、标准编制原则

1.科学规范。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》中的规范性要求进行编写，对本文件规定了跨队列研究实施过程中采集内容、采集流程和质量控制要求等内容，进行规范性阐述。

2.系统全面。跨队列研究实施过程中采集内容、采集流程和质量控制的全过程进行系统梳理，标准内容涵盖对数据采集的准备、过程和采集后的质量控制处理等。对跨队列研究队列数据管理的数据融合质量控制进行全面规范，编制高水平的团体标准。

3.实用好用。尽可能从实用的角度出发，充分考虑标准的可实施性和实施效果，吸引了各方工作人员参与，多方面征集标准相关部门以及其他相关方的意见，完善标准内容。

4.协调一致。标准编制过程中，对相关的国家标准以及政策文件进行了充分的分析和考虑，保证本标准与国家、行业标准及政策文件的协调。

四、主要技术内容的论据

本文件规定了在开展跨队列研究前，原始队列数据的采集内容以及质控要求等内容。本文件适用于指导拟开展跨队列研究的原始队列数据采集工作，包括但不限于社区人群队列、区域性人群队列、针对某一疾病种类或基于特殊机构建立的人群队列。本文件核心技术内容包括：采集内容及方式和质控要求等章节。

术语和定义：规定了队列、跨队列、研究对象的定义，结合实际经验，重点参考了《Connecting Cohorts to Diminish Alzheimer’s Disease (CONCORD-AD): A Report of an International Research Collaboration Network》和《Using a cross-cohort comparison design to test the role of maternal smoking in pregnancy in child mental health and learning: evidence from two UK cohorts born four decades apart》两篇文章中的相关论述。规定了病例报告表、标准操作程序的定义，结合实际经验，参考了《药品临床试验管理规范》。

采集内容及方式：采集内容通常包括采集调查对象的社会人口学数据、疾病或健康数据等。采集方式通常包括自评问卷、他评问卷、体格检查、实验室检查、影像学检查以及生物样本检测。数据采集的内容应符合WS/T 306 卫生信息数据集元数据规范、WS 363卫生信息数据元目录（系列标准）、WS/T 370 卫生信息基本数据集编制规范、WS 372-2012 疾病管理基本数据集、WS 375 疾病控制基本数据集（系列标准）。

质控要求：由三个部分组成。对于信息采集前质控要求，根据队列数据采集实践、《临床队列遗传及共享平台建立研究手册》，明确了标准操作程序、病例报告表、人员培训、现场检查等方面的质量管理要求。对于信息采集中质控要求，根据管理实际，参考中华预防医学会T/CPMA 001-2018《大型人群队列研究数据处理技术规范》以及《中国精神卫生调查现场执行及质量控制》，规定了数据采集控制要求。对于信息采集后质控要求，根据管理实际，参考中华预防医学会T/CPMA 001-2018《大型人群队列研究数据处理技术规范》、《中国精神卫生调查现场执行及质量控制》和《临床队列遗传及共享平台建立研究手册》，规定了数据清理、数据库管理、文件管理以及生物样本的保存要求。

五、主要试验、验证及试行结果

该项标准依托科技部国家重点研发计划《大型队列间联合研究柔性化大数据云平台支撑系统研发》项目课题一《跨队列研究及数据共享标准、机制及质量控制方法研究》以及2023年中国疾控中心公共卫生领域卫生健康标准化前期研究项目《跨队列研究数据融合质量控制标准研发前期研究项目》,对制定的标准进行试验方法验证。

在拥有存储10万余例重大疾病队列的结构化临床、随访和影像数据的临床科研数据中心以及近4万的社区人群队列进行了广泛的使用，并进行了大量的验证试验和处置工作，在此基础上对标准草案进行了论证、补充、细化和完善。目前已形成了一套稳定的数据采集流程和数据质量控制方法，基本上能够保证数据质量，促进数据融合和分析。

六、采用国际标准或国外先进标准的情况

美国食品药品监督管理局（FDA）发布的“STUDY DATA TECHNICAL CONFORMANCE GUIDE”（研究数据结构化技术指南），该指南对提交至FDA动物和人体研究数据提出了标准化规范，以便于FDA对于科研项目的评审。但没有对数据采集控制有明确规定的标准。

七、与现行法律法规、强制性标准和其他有关标准的关系

1、法律法规依据

《国家健康医疗大数据标准、安全和服务管理办法（试行）》第九条 国家卫生健康委员会鼓励医疗卫生机构、科研教育单位、相关企业或行业协会、社会团体等参与健康医疗大数据标准制定工作。公民、法人或者其他组织可提出制修订健康医疗大数据标准的立项建议，并提交相应标准项目建议书。第三十三条 责任单位应当结合服务和管理工作需要，及时更新、甄别、优化和维护健康医疗大数据，确保信息处于最新、连续、有效、优质和安全状态。

2、相关标准

WS/T 306 卫生信息数据集元数据规范

WS 363卫生信息数据元目录（系列标准）

WS 364 卫生信息数据元值域代码（系列标准）

WS 365 城乡居民健康档案基本数据集

WS/T 370 卫生信息基本数据集编制规范

WS 372-2012 疾病管理基本数据集上

WS 375 疾病控制基本数据集（系列标准）

八、重大意见分歧的处理经过和依据

无

九、作为推荐性标准的建议

无

十、贯彻该标准的要求、措施建议及预期效果

确保跨队列研究质量的重要前提是原始队列数据在采集时达到质量要求，可保证开展的跨队列研究才具备学术价值。因此，制定《队列数据管理 数据采集质控要求》标准将为提升原始队列数据质量带来显著的改进和效益。

首先，在规范数据采集方面，标准将统一队列数据的采集内容和方式，确保数据类型和适用性的准确性和一致性。同时，明确数据采集流程，提升数据采集过程的透明度和可追溯性，从而提高数据的真实性和可信度。

其次，在提升数据质量方面，标准将提出详细的数据质量控制措施，包括数据收集过程的评估、唯一性和完整性检查，以及合法性和内部一致性检查等。这些措施将确保数据的准确性和可靠性，通过对数据质量的系统评估，减少错误数据和重复数据，提高数据录入的准确性和完整性。此外，通过实施数据质量保障措施，如培训、标准化流程和定期质量检查，持续提升数据质量管理水平。

此外，制定标准将促进数据共享和应用。规范的队列数据管理和采集质控要求将有助于不同队列之间的数据共享，促进跨队列研究中的数据融合和应用，为未来的跨病种队列数据分析提供可靠的数据基础，扩大数据的应用范围，提升研究成果的社会和学术价值。

总体而言，制定《队列数据管理 数据采集质控要求》标准将有效提升跨队列研究的数据质量和管理水平，促进数据共享和应用，推动科学研究的创新与发展。这将为跨病种队列数据分析提供坚实保障，使不同时期采集的队列数据充分体现其价值。

十一、废止或代替现行相关标准的建议

无

十二、其他应予说明的事项

无