**团体标准编制说明**

《队列数据管理 数据融合质控要求》

(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 数据融合质控要求》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 数据融合质控要求》的编写目的在于建立一套系统化的标准，以确保不同队列数据在融合过程中的质量控制。这一标准旨在解决当前跨队列研究中存在的数据质量不一致、缺乏标准化评估和管理等问题。通过明确数据质控的要求和流程，确保在数据融合过程中，各种数据源的准确性、完整性和一致性，从而提高研究结果的可信度和科学性。此外，该标准还旨在促进数据共享与交流，为研究人员提供清晰的指导，推动跨病种研究的协作与创新，最终提升科学研究的整体质量和效率。

（三）编制意义

《队列数据管理 数据融合质控要求》的编写意义在于为跨队列数据融合提供系统化的质控标准，确保数据在整合过程中的准确性和一致性，从而提升研究结果的可靠性和科学性。该标准促进数据共享与合作，提高数据管理效率，减少错误与重复工作，增强研究的可重复性和可比较性，满足法规与伦理要求。总体而言，这一标准对推动科学研究的创新与发展、提升数据管理水平具有重要意义，为跨病种研究提供坚实基础。

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，北京大学第六医院、北京大学、中国医学科学院肿瘤医院、中国疾病预防控制研究中心、山东大学齐鲁医院、北京大学第一医院、天津市安定医院、中国电子技术标准化研究院，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 | 副研究员 | 北京大学第六医院 |
| 黄雨 | 计算机软件与理论 | 研究员 | 北京大学 |
| 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 |
| 丁若溪 | 社会医学与卫生事业管理 | 副研究员 | 北京大学第六医院 |
| 张婷婷 | 社会精神病学 | 主管技师 | 北京大学第六医院 |
| 陈园生 | 公共卫生 | 研究员 | 中国疾病预防控制中心 |
| 吕明 | 流行病学 | 教授 | 山东大学齐鲁医院 |
| 张媛 | 流行病与卫生统计学 | 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 |
| 徐广明 | 社会精神病学 | 主任医师 | 天津市安定医院 |
| 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 |
| 王悦 | 公共卫生 | 助理研究员 | 中国疾病预防控制中心 |
| 张同超 | 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 |
| 潘鹏 | 精神病与精神卫生学 | 主治医师 | 天津市安定医院 |
| 颜国利 | 精神病与精神卫生学 | 副主任医师 | 天津市安定医院 |

（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 数据融合质控要求》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 数据融合质控要求》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 数据融合质控要求》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两项标准。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

（一）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数据融合的质量控制要求等内容，进行规范性阐述。

2.系统全面。对跨队列研究队列数据管理的数据融合质量控制的全过程进行系统梳理，标准内容涵盖对数据融合前、融合中以及融合后的质量控制流程和细则等。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

本文件规定了跨队列研究实施过程中数据融合的质控要求等内容。本文件适用于指导拟开展跨队列研究的数据融合工作，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：融合前需明确的内容、融合中的规范、融合后的要求等章节。

术语和定义：规定了队列、跨队列定义，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。规定了数据融合和变量的定义，参考了《Data Fusion: Concepts and Ideas》和《Data Fusion: Mathematical Principles and Applications》两本专著。

融合前需明确的内容：根据数据融合实践，参考《Data Fusion: Concepts and Ideas》，明确了融合的目的、队列数据集基本信息的具体要求、队列数据可用性评价内容，以及获得队列数据的许可的要求。

融合中的规范：根据数据融合实践，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》、《Data Fusion: Mathematical Principles and Applications》和《Data Fusion Mathematics: Theory and Practice》，明确了拟融合的原始队列数据的清理要求、数据融合的规则、定量评分以及融合结果验证的内容。

融合后的要求：根据数据融合实践，参考《Data Fusion: Concepts and Ideas》，规定了文档记录过程管理、融合过程可追溯性和透明性要点以及定期评估和更新融合后的数据集的要求。

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。但没有对队列数据的质控数据元进行定义及核查的标准。

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

WS 375 疾病控制基本数据集（系列标准）

GB/T 41814 信息技术 生物特征识别校准、增强和融合数据（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

在现代医学和公共卫生研究中，跨队列研究已成为重要工具，但数据复杂性和多样性使标准化管理和融合要求显得尤为重要。因此，贯彻该标准的要求包括：统一数据采集方法和格式，确保数据一致性和可比性；建立全面的数据质量控制标准，以避免数据缺失、错误和偏差；明确数据融合过程中的处理步骤，确保科学性和规范性；提高研究过程的透明度和可追溯性，以便于后续审查和验证。

为实现这些目标，建议采取以下措施：一是制定详细的实施指南，明确数据采集、清理、处理和融合的具体流程；二是提供培训，帮助研究人员理解和应用标准；三是建立质量监控机制，定期评估数据质量；四是推动数据共享平台建设，促进不同研究团队之间的合作与交流。

预期效果方面，贯彻该标准将显著提升数据融合的质量和研究的可靠性，通过规范化的流程和工具，提高研究效率，减少重复工作。同时，统一的数据标准将促进数据共享与再利用，推动跨学科、跨机构的合作研究。此外，标准的实施还将增强数据融合的科学性和规范性，确保数据处理的准确性，最终推动科学研究的深入和发展。

总体而言，制定《队列数据管理 数据融合质控要求》标准不仅能够显著提升跨队列研究的数据管理和融合水平，还能为相关领域的研究提供可靠的技术支持和规范指导，推动科学研究的深入和发展。

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