**团体标准编制说明**

《胃癌专病队列数据集建设规范要求》

(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

本项团体标准的制定工作遵循北京市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》，严格按照国家重点研发计划课题《跨队列研究及数据共享标准、机制及质量控制方法研究》的相关要求，起草与制定工作由标准牵头单位山东大学齐鲁医院负责执行。

（二）编制背景

我国是胃癌高发国家，患者数量在全球占有较大比例。然而，在胃癌研究领域，目前存在以下问题：各医院电子病历系统差异较大，数据格式缺乏一致性，影响多中心数据的整合和共享；数据标准的缺失导致“数据孤岛”现象严重，宝贵的临床信息未能得到有效利用；人口流动性大，随访数据采集难度高，且缺乏规范的随访计划，数据丢失问题突出。因此，当前胃癌专病队列数据面临的主要问题在于缺乏统一的管理规范以及数据共享机制。如何促进胃癌专病队列数据间的开放，避免数据孤岛的形成，避免信息资源的浪费，是亟需解决的问题。

（三）编制目的

构建一个规范化、标准化的胃癌专病队列数据集，以统一数据结构和内容为核心，推动全国范围的多中心合作，实现数据的开放与共享。此规范力求通过全面系统地整合医疗数据资源，提供涵盖胃癌分类、诊断、治疗进展、临床决策及预后情况等多方面的必要信息。进一步地，基于深度处理与分析的数据库，我们期望能够实现对胃癌的早期发现与控制，构建疾病预后模型，为患者提供精准的治疗指导，并构建胃癌精准医疗知识网络。最终，该规范旨在提升我国胃癌患者的健康水平，增强疾病风险分析、预测及防控能力，为医学科研与医疗服务的质量与效率提供强有力的数据支持。

（三）编制意义

为胃癌领域的数据管理提供一套全面而系统的标准框架，该框架不仅规范了数据建模、数据采集、数据治理、数据存储及更新，还强调了数据安全的重要性。此规范的核心价值在于，它确保了胃癌相关医疗数据在整合过程中的准确性和一致性，从而极大提升了基于这些数据所得出的研究结果的可靠性和科学性。通过实施这一规范，能够有效促进跨机构、跨平台的数据共享与合作，提高数据管理的效率，显著减少因数据不一致或错误所导致的重复工作。同时，该规范还增强了研究的可重复性和可比较性，使得不同研究之间的结果能够相互验证，进一步推动胃癌研究的深入发展。

二、工作简况

（一）起草单位及主要起草人

本标准的起草工作由山东大学齐鲁医院提出，与复旦大学、北京大学肿瘤医院、复旦大学附属中山医院、北京协和医院、北京大学第三医院、山东大学临床研究中心、山东省医学会临床流行病学与循证医学分会共同完成。起草团队涵盖胃肠外科、肿瘤流行病学、生物统计学及信息技术等领域的专家。

|  |  |  |  |
| --- | --- | --- | --- |
| 姓名 | 专业领域 | 职称 | 工作单位 |
| 吕明 | 肿瘤流行病学 | 主任医师 | 山东大学齐鲁医院 |
| 于文滨 | 胃肠外科 | 主任医师 | 山东大学齐鲁医院 |
| 陈兴栋 | 肿瘤流行病学 | 研究员 | 复旦大学 |
| 李文庆 | 肿瘤流行病学 | 研究员 | 北京大学肿瘤医院 |
| 陈世耀 | 消化内科 | 主任医师 | 复旦大学附属中山医院 |
| 杨红 | 消化内科 | 主任医师 | 北京协和医院 |
| 丁士刚 | 消化内科 | 主任医师 | 北京大学第三医院 |
| 刘联 | 肿瘤内科 | 主任医师 | 山东大学齐鲁医院 |
| 杨孝荣 | 肿瘤流行病学 | 副研究员 | 山东大学齐鲁医院 |
| 张同超 | 肿瘤流行病学 | 助理研究员 | 山东大学齐鲁医院 |
| 张媛 | 肿瘤流行病学 | 副主任医师 | 山东大学齐鲁医院 |
| 陈浩 | 肿瘤流行病学 | 副研究员 | 山东大学齐鲁医院 |
| 闫治波 | 胃肠外科 | 副主任医师 | 山东大学齐鲁医院 |
| 魏猛 | 胃肠外科 | 主治医师 | 山东大学齐鲁医院 |
| 殷晓霖 | 流行病与卫生统计学 | 助理研究员 | 山东大学齐鲁医院 |
| 白浩 | 流行病与卫生统计学 | 助理研究员 | 山东大学齐鲁医院 |
| 张雪凝 | 流行病与卫生统计学 | 博士生 | 山东大学 |
| 李真 | 流行病与卫生统计学 | 博士生 | 山东大学 |
| 崔晓晗 | 胃肠外科 | 主治医师 | 山东大学齐鲁医院 |
| 李紫琳 | 流行病与卫生统计学 | 博士生 | 山东大学 |

（二）主要编制过程

2024年7月，确定起草组成员并获取知情同意书，同时参考现有标准、指南共识及相关文献，通过专家访谈/调研，撰写相关报告。

2024年8月，根据研究和调研结果起草《胃癌专病队列数据库建设规范要求》初稿，并在起草单位内部多次讨论和修改，形成标准初稿，提交北京慢性病防治与健康教育研究会审核并通过立项。

2024年9月，根据评审专家反馈意见继续修改标准初稿，向北京慢性病防治与健康教育研究会提交审核并明确编写工作组成员。

2024年10月，召开团体标准起草专家组讨论会，对标准初稿的框架体系和内容继续研讨，形成标准征求意见稿，并提交审核。

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，对本文件规定了胃癌专病队列数据库建设规范要求等内容，进行规范性阐述。

2.系统全面。该标准通过全面系统整合医疗数据，提供胃癌分类、诊断、治疗发展、临床决策、预后情况等所需信息；基于数据库的深度处理和分析，实现对胃癌的早期发现和控制，建立疾病预后模型，给予患者治疗指导，构建精准医疗知识网络，全面提升医学科研、医疗服务的质量和效率。

3.可操作性强。本标准注重实际应用中的可操作性，确保各项规定和流程在实际操作中易于理解和执行。通过简化复杂流程、明确操作步骤、提供实施指南等方式，增强标准的实用性和适用性，便于医疗工作者、科研人员及管理人员等各方有效应用。

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

本文件规定了可支撑胃癌标准数据集构建方法，包括数据建模、数据采集、数据治理、数据存储和数据安全。本标准适用于医疗机构、研究机构、企业等胃癌数据集设计、研发和管理，其他相关领域可参考使用。本标准的术语和定义包括数据、特征、数据质量、数据集、数据建模、数据清洗、数据治理、数据采集、数据脱敏、数据标注。

胃癌专病队列数据集建设标准规范了数据建模、数据采集、数据治理、数据存储和数据安全五个方面：

一、数据建模需建立胃癌专病库字段数据集标准，制定数据模型之前，应召集专病数据集使用人员、管理人员、专家与数据建模工程师一起讨论，结合专病数据集的建库目标，共同确定数据范围。数据标准的建立将充分借鉴行业标准，结合医疗健康的行业规范及专病对数据的实际应用要求。数据标准规范建议遵守国家、行业标准代码规范，遵守卫生行业相关标准规范，遵守院内信息化相关标准规范。胃癌数据集包含人口学信息，就诊记录，病史，体格检查，专科检查，诊断，检验，检查，手术治疗，术后恢复和并发症，病理，治疗，随访，不良反应等。

二、数据采集从收集和储存手术切除的肿瘤组织、血液等生物样本，收集和整合人口统计学、检查结果、诊断、手术、治疗等临床数据，随访患者术后信息等方面进行规范。数据采集包括回顾性数据采集和前瞻性数据采集两方面。回顾性数据采集包括采集对象、数据采集方式、数据采集技术和数据采集结果四个部分。前瞻性数据采集要求数据采集技师具备一定的工作经验，并接受相关数据采集流程、设备操作和数据安全的培训。采集过程需确保被采集对象知情同意，采集员应详细介绍采集过程和要求，以确保数据的真实性、准确性和有效性。

三、数据治理对数据预处理、主数据管理和数据质控等进行规范。数据预处理通过数据抽取、过滤、转换、加载和清洗，确保数据的准确性、完整性和一致性。主数据管理则整合各部门的核心数据，进行统一清洗和丰富，并将这些数据提供给操作和分析应用系统。数据质控通过建立评价模型和管理机制，对数据的完整性、一致性和正确性进行验收和持续监控，确保其符合质量标准并能安全有效地使用。

四、数据存储中需对不同类型数据库的存储管理技术进行规范，其主要有文件管理系统和数据库管理系统。分布式文件系统可以用于存储专病数据集的影像资料、检查结果、出院小结等非结构化数据。数据库管理系统根据应用场景分为分析型数据库、关系型数据库、图数据库、NoSQL数据库等四大类。

五、数据安全中需对数据访问、数据加密、数据脱敏、数据备份进行规范。采用多级授权管理模式，依据成员的级别、职责及对数据集的贡献大小，规定用户可以使用的系统功能和资源。专病数据集在存储前要加密以防止重要数据丢失或盗用，传输过程中对数据进行加密传输同样重要。在进行数据导出操作时，患者的身份识别等敏感信息可以通过数据脱敏机制进行隐藏，避免信息泄露。

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》,对制定的标准进行试验方法验证。在拥有胃癌专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进行了试用，并在此基础上对标准草案进行了论证、补充、细化和完善。目前已构建形成一套科学的胃癌专病队列数据集建设标准规范要求，促进数据融合和共享。

六、采用国际标准或国外先进标准的情况

目前，国外尚无胃癌专病队列数据库建设的相关标准，但已有比较成熟的大型肿瘤数据库，如美国监测、流行病学和结果（Surveillance，Epidemiology，and End Results，SEER）数据库是1973年由美国国家癌症研究所建立，用于监测肿瘤发病情况、分期、治疗以及预后信息的数据库；美国国家癌症数据库（National Cancer Database，NCDB）始于1988年，由美国外科医师学会与美国癌症协会建立，是以医院为基础的癌症登记系统；日本癌症登记数据库是由日本国家癌症中心负责建立，分为基于人群和基于医院的癌症登记；日本的国家临床数据库（National Clinical Database，NCD）是一个全国性的外科手术数据录入电子系统，于2010年在日本外科协会和日本胃肠外科协会等的支持下启动。这些数据库的发展、演变、运行可以为我国胃癌大数据平台的建设提供经验。

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）》

2、相关标准

WS 445-2014 电子病历基本数据集

WS/T 671-2020 国家卫生与人口信息数据字典

ISBN 978-7-117-28858-3 胃癌标准数据集（2019版）

T/BISSC 01-2022 专科疾病标准数据集建设规范

T/CMDA 002-2020 肝胆疾病标准数据规范：肝癌 CT/MRI 影像标注和质控标准

NCCN 临床实践指南：胃癌指南（2018.V2）

中国临床肿瘤学会（CSCO）胃癌诊疗指南 2019

ICD-9-CM-3 世界卫生组织国际疾病分类手术码

GB/T 25000.12-2017 系统与软件工程 系统与软件质量要求和评价(SQuaRE) 第 12 部分：数据质量模型

GB/T 35295-2017 信息技术 大数据 术语

GB/T 34960.5-2018 信息技术服务 治理 第 5 部分：数据治理规范

GB/T 39725-2020 信息安全技术 健康医疗数据安全指南

GB/T 5271.28-2001 信息技术 词汇 第 28 部分：人工智能 基本概念与专家系统

GB/T 42755-2023 人工智能 面向机器学习的数据标注规程

DB3209/T 1259.1-2023 公共数据平台 数据治理规范 第一部分:总则

DB3206/T 1076-2024 医疗卫生公共信用平台建设技术规范

GB/T 36344-2018 信息技术 数据质量评价指标

GB/T22239-2019 信息安全技术 网络安全等级保护基本要求

GB/T35273-2020 信息安全技术 个人信息安全规范

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

本标准建议作为推荐性标准执行。

十、贯彻该标准的要求、措施建议及预期效果

贯彻该标准的要求包括：构建规范化的胃癌专病队列数据集，统一数据结构和内容；推动全国范围内的多中心合作，实现数据开放和共享；提升疾病风险分析、预测和防控能力，为患者精准医疗和科研进展提供数据支持。

措施建议包括：制定详细的实施指南，明确数据采集、数据治理、数据存储和数据安全的具体流程。明确采集流程和标准，完善数据清洗与质量控制方法，优化数据存储结构和格式，确保数据隐私和共享安全性。

预期效果方面：以期通过科学规范的专病数据集建设，整合资源建立大数据平台，提高我国胃癌研究和治疗水平，实现疾病防控与患者健康的全面提升。

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