

ICS 11.020

C 07

# 团体标准

T/BRACDCHE 003-2025

## 跨队列研究数据共享分发要求

Data Sharing and Distribution Requirements for Cross-cohort  
Study

2025-05-26 发布

2025-05-26 实施

北京慢性病防治与健康教育研究会 发布

## 目 次

前 言 .....	I
引 言 .....	II
1 范围 .....	1
2 规范性引用文件 .....	1
3 术语和定义 .....	1
4 数据共享的组织架构及职责 .....	2
5 数据共享申请及审批程序 .....	3
6 数据共享流程 .....	4
7 数据共享后的服务与监管 .....	5
7.1 提供数据使用帮助 .....	5
7.2 数据安全性与成果追踪 .....	5
附录 A .....	6

## 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。

本文件由北京大学第一医院提出。

本文件由慢性病防治与健康教育研究会归口。

本文件起草单位：北京大学第一医院、北京大学人民医院、北京大学第三医院、北京大学深圳医院、北京大学第一医院太原医院、北京大学第六医院、山东大学齐鲁医院、中国电子技术标准化研究院、北京大学、北京六元空间信息科技有限责任公司。

本文件主要起草人：李航、杨之辉、白倩倩、周城、慕彰磊、李琨、王文慧、王艺萌、于波、吴瑕、沈长兵、牛旭平、李俊琴、侯瑞霞、刘肇瑞、张婷婷、丁若溪、吕明、张媛、李瑞琪、黄雨、朱文伟、孙国锋、黄文昊

# 引 言

在现代科学研究中，从海量的医疗数据中进行分析挖掘，可获得大量有意义的信息应用于临床决策支持、疾病预防。然而，目前医疗信息出于隐私保护及利益等问题，多存储于各自医疗机构，互不流通；如何安全有效地实现数据在各医疗机构中传输并高效利用海量数据输出研究成果，以便数据共建共享具有重要意义。

本文件从数据共享的组织架构、数据共享的原则、流程、及共享后服务与监管四方面入手，明确了对负责数据共享的数据审批管理委员会的组织架构要求及责任分工；规定了数据共享前对数据申请者的资质、数据申请目的与用途（即课题计划书）的考核要求，以保证数据被用于符合伦理、有价值的课题研究中，保证数据对相应课题的支撑力度、研究者的科研能力，确保数据在分发后能得到有效利用；规定了共享数据的质量、数据提取与发放过程中的精准性与安全性要求；描述了精准提取需求数据的方法和人工审核过程，给出了数据发放后的技术支持与监管要求，以便确保数据充分、正确且安全的使用，避免发放后数据的不正当使用或浪费。

本文件为大型专病队列研究数据共享分发提供清晰、明确的操作指南，促使其按照统一的要求和流程进行数据共享，减少数据提取与传输过程的数据泄露风险，确保数据分享后合规、高效的应用，实现大型专病队列研究数据库共建、共享、共赢的运转模式。

# 跨队列研究数据共享分发要求

## 1 范围

本文件规定了跨队列研究数据共享分发的术语和定义、数据共享的组织架构及职责、数据共享原则、数据共享流程和数据共享后的服务与监管。

本文件适用于医学队列研究中大型专病队列研究。

## 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。

YD/T 4680-2024 电信网和互联网数据安全管控平台技术要求和测试方法

T/CPMA 001-2018 大型人群队列研究数据安全技术规范

T/CPMA 001-2019 大型人群队列现场调查管理技术规范

T/CPMA 002-2019 大型人群队列终点事件长期随访技术规范

T/CPMA 003-2023 成人慢性病行为危险因素监测基本数据集标准

YY/T1833 人工智能医疗器械质量要求和评价 第2部分:数据集通用要求

## 3 术语和定义

下列术语和定义适用于本文件。

### 3.1

**专病队列研究** specific disease cohort study

通过对一组特定疾病患者的追踪观察以评估疾病危险因素、预后因素和治疗效果等方面问题的研究方法。包含前瞻性、回顾性及双向队列研究等。

**注1:** 主要特点是对患有该疾病的患者进行长期追踪随访，观察其治疗预后、并发症的发病率和整体死亡率等结果，从而深入揭示疾病的演变规律，为临床诊疗和疾病防控提供有力支撑。

### 3.2

**数据共享** data sharing

数据监管机构通过公开、免费或付费的形式，将医疗大数据资源提供给其他机构、研究团队或个人使用的一种高效且规范的数据管理模式。

**注1:** 旨在促进数据价值最大化、推动医学研究和临床实践。

**注2:** 制定合理的数据共享规范至关重要，这其中需要充分考虑数据共享的前提要求、安全保障和监管等多个方面，通过设计科学的数据共享准则，兼顾数据的开放性与安全性，确保数据的安全共享和高效应用。

### 3.3

#### 数据分发 data distribution

依据预设的规则与流程将数据精确地发放给指定的接收方。

注1：这一过程需确保数据准确无误和及时有效的传递，同时维护数据安全，防止未授权访问或泄露。

### 3.4

#### 数据脱敏 data masking

数据匿名化

为保护个人隐私和数据安全，通过技术手段去除或隐藏数据中的敏感信息的过程。

注1：常见脱敏技术包括替换、重排、加密、截断和掩码，旨在保留数据的实用价值，同时消除可能引发身份识别的风险。

注2：这一过程对于保障数据主体的隐私权益，促进数据的合法合规使用至关重要。

### 3.5

#### 制式化数据 standardized data

制式化数据是指经过标准化处理的数据，这种处理通常遵循特定的格式或规范，以确保数据的一致性和可比性。制式化过程可能包括数据清理、格式转换、单位统一、编码标准化等步骤。

## 4 数据共享的组织架构及职责

### 4.1 组织架构

负责数据共享的组织架构包括数据审批管理委员会、数据管理委员会、学术委员会。

### 4.2 数据审批管理委员会

4.2.1 宜在构建专病队列研究数据平台时同时成立，宜由牵头单位组织建立，由牵头单位及下设单位，以及负责数据监管的单位共同组成。

4.2.2 应由数据管理委员会与学术委员会构成，承担数据的组织生产、数据共享的审批与分发职责。

4.2.3 应在数据产生前与各数据提供单位签订协议，确认数据审批管理委员会具有数据分发及使用权利；而数据提供单位享有数据共享优先申请的权益。

### 4.3 数据管理委员会

4.3.1 应由负责数据存储与安全保障的数据监管机构兼任；

4.3.2 在数据共享中应至少负责数据申请书的可行性评估、数据的质量控制、提取、脱敏与传递、数据共享前后的风险评估。

### 4.4 专病学术委员会

4.4.1 应由专病队列相应临床专业内的资深学者组成，以保障数据使用的科学性与安全性。

4.4.2 在数据共享中应负责数据申请书的科学性评估、数据共享后的成果追踪。

## 5 数据共享申请及审批程序

### 5.1 数据共享的申请

5.1.1 数据申请单位应向数据审批管理委员会提交制式化数据申请书，说明申请数据所服务的课题研究，经学术委员会评估课题的价值与意义、数据管理委员会审查课题可行性、申请单位伦理委员会进行课题伦理审查、申请单位在国家医学研究登记备案信息系统完成备案后，方能获批数据共享资格。

5.1.2 申请书内容应至少包括：研究背景、研究目的、研究方法、时间安排、伦理批件、申请人研究背景及用于接收数据的申请人邮箱。其中，研究方法为申请书主体部分，应包括研究设计、研究对象、研究变量定义、样本量估计及统计方法。申请书应按照附录1的格式要求附制式化数据需求表格。

5.1.3 制式化数据需求表格内容应由数据审批管理委员会制定，内容为专病队列数据库所能提供的所有原始变量、变量类型（如连续变量、分类变量等）、变量取值范围，供数据申请单位在申请时从中选择：

- a) 数据分析所需变量，以确保所提取数据能满足研究课题所需的必要信息；
- b) 主要入组标准所对应的变量及其取值范围，以确保后期数据的精准提取与分发。

5.1.4 数据监管中心宜开发数据电子化申请平台，通过统一在线模板和表单以规范数据申请书的填写，便捷化数据申请书的提交、审核和提交后的数据匹配，减少手动操作误差，提升审核的效率。

### 5.2 数据管理委员会审批

5.2.1 数据管理委员会应根据数据申请书中的制式化数据需求表格，与数据库进行匹配，并输出可行性报告。

5.2.2 可行性报告内容包括数据库中符合入组标准的患者样本量及各所需变量的缺失比例，供数据管理委员会及学术委员会进行申请书审批时参考。

### 5.3 学术委员会审批

5.3.1 数据监管机构应邀请学术委员会专家对申请书的研究意义、创新性、可行性、伦理进行全面评定，决定是否具备获得数据的资质。

5.3.2 受邀参与评定的学术委员会专家应为相应临床专业内的副高级及以上职称的学者、具备5年及以上主持或参与该专业领域临床研究的经验，且以第一作者或通讯作者发表临床研究至少10篇。

### 5.4 伦理审批及临床研究注册

5.4.1 数据申请单位应在提交数据申请书前在各单位伦理委员会完成伦理审批，并将伦理批件作为数据申请书的必要附件材料。

5.4.2 数据申请单位应在伦理审核通过后，在国家全民健康保障信息平台医学研究登记备案信息系统完成备案，以确保数据在临床研究中的规范化分析使用。

### 5.5 签订数据保密协议

5.5.1 数据发放前，申请单位应与数据审批管理委员会签订保密协议。

5.5.2 保密协议中，申请单位应承诺数据仅限于所申请课题使用，不得泄露或转交第三方，确保数据安全与合规使用。

## 5.6 数据共享透明度

5.6.1 数据管理委员会宜定期公开数据共享情况，包含数据共享申请的数量、批准率、共享数据的应用领域等信息，通过增加数据共享的透明度增强数据共享工作的公信力。

5.6.2 数据管理委员会可创建公众平台，通过平台定期发送数据共享报告，以提高各单位对数据使用申请的参与意愿。

## 6 数据共享流程

### 6.1 数据质量控制

6.1.1 数据平台应建立统一的数据标准和格式，确保不同来源的数据具有一致性和可比较性，便于数据共享后的整合和分析，数据质量控制具体方法可遵循“YY/T1833”的第2部分要求执行。

6.1.2 数据管理委员会宜成立数据维护小组，定期对质控不达标的数据进行质询、修改与清洗，确保数据的质量、时效性和可应用性。

### 6.2 数据提取及脱敏

6.2.1 数据监管中心后台应根据制式化数据需求表格，从数据库中精准匹配并提取满足核心入组条件的病例中所需的关键变量信息，并对数据文件实施严格加密措施。可遵循T/CPMA 001-2019、T/CPMA 002-2019及T/CPMA 003-2023 执行。

6.2.2 参照《大模型人工智能技术在医疗服务领域应用的专家共识》，数据监管中心的后台系统可具备一定自动化、智能化能力，实现制式化数据需求表格与数据库之间的智能化匹配，减少人员手动操作带来的数据泄露风险，确保数据传输的安全性，电子数据安全问题可遵循“YD/T 4680-2024”执行。

6.2.3 数据监管中心后台应对数据进行脱敏处理，隐藏患者姓名、电话号码、身份证号码、住址等个人敏感信息，隐私信息脱敏及加密的方法可按“T/CPMA001-2018”执行。同时为每位患者分配唯一的标识码用于后续的患者识别，充分保护患者隐私安全。

### 6.3 人工审核

6.3.1 数据提取后，应再次审核如下方面：数据文件是否已加密处理；患者个人信息是否已脱敏；提取的病例是否满足申请书中规定的入组标准；提取的变量是否与课题需求匹配；数据发放的邮箱地址是否与申请者的邮箱一致。

6.3.2 数据共享前，应再次进行共享风险评估，包括：评估数据泄露、误用或滥用的风险，并制定应对措施。

6.3.3 数据的审核及共享风险评估应由数据管理委员会和独立的第三方安全机构共同完成，确保客观性和全面性。

6.3.4 负责数据审核及共享风险评估的数据管理委员会成员及第三方机构人员在访问数据前，应经

过严格的数据安全的专业训练及考核。

**6.3.5 人工审核环节应遵循以下原则：**1) 实行物理隔离，即使用专用于数据审核的计算机设备，该设备应不具备数据下载和导出功能；2) 实施网络隔离，即审核工作只能通过内部网络进行，严禁与外部网络连接。具体措施要求宜参考“T/CPMA001-2018”。

## **6.4 数据传递**

**6.4.1** 应在数据审核确认无误后，由数据监管中心后台向申请者的数据连接地址发放数据。

**6.4.2** 传递的数据文件应设定有限的下载次数和使用期限，确保数据仅限于本次研究项目使用，避免非授权传播。

## **7 数据共享后的服务与监管**

### **7.1 提供数据使用帮助**

**7.1.1** 数据审批管理委员会应向用户全面提供详细的数据说明字典，阐明数据结构、字段含义及使用规则。

**7.1.2** 数据审批管理委员会宜组织专业的数据分析培训课程，内容包括数据结构解读、脱敏数据的分析处理以及高级数据分析技术等，确保各申请单位能够准确理解和高效使用数据，促进各申请单位高质量研究成果的产出。

**7.1.3** 数据审批管理委员会应提供邮件对话通道，由委员会内兼具临床经验和数据分析专长的研究者负责，及时响应参与单位的数据相关疑问，协助处理数据问题，提供针对性的技术指导与解决方案。

### **7.2 数据安全性与成果追踪**

**7.2.1** 数据管理委员会宜建立实时追踪和审计机制，通过追踪数据的访问日志、使用频次等方式，评估数据的合规使用情况，以及时发现并防止数据滥用行为。

**7.2.2** 学术委员会应定期与各数据申请使用单位进行沟通，追踪课题进展，以确保课题按计划顺利推进。

**7.2.3** 学术委员会宜要求数据申请使用单位在项目结束后递交成果报告，详细描述数据的应用情况和研究成果，以评估数据使用的实际效益，为后续的数据共享策略提供参考依据。

**7.2.4** 学术委员会可建立奖励机制，对研究成果突出的数据使用单位进行嘉奖，鼓励更多单位参与数据共享，促进数据的不断更新及扩展。

附录 A  
(规范性)  
制式化数据需求表格格式

表 A.1 规定了制式化表格的格式和内容。

表 A.1 制式化数据需求表格样式

数据库涵盖变量	变量类别	变量取值范围	申请变量 (多选)	主要入组条件	
				对应变量 (多选)	变量取值要求
性别	分类	男/女	√		
年龄	连续	0-100	√	√	>18
婚姻	分类	未婚/已婚/离异			
xx 专病的疾病分型	分类	A/B/C/D		√	=A
xx 专病的疾病严重度 xx 评分	连续	0-30	√	√	>10
...					