

T/GDFDTAEC

团 体 标 准

T/GDFDTAEC 16—2025

增龄健康状况队列数据和生物样本采集 处理指南

Guidelines for collection and processing of aging health status cohort data and biological samples

2025 – 12 – 19 发布

2025 – 12 – 19 实施

广东省食品药品审评认证技术协会 发 布

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有可能涉及专利，本文件的发布机构不应承担识别这些专利的责任。本文件由广东省食品药品审评认证技术协会提出并归口。

本文件由国家重点研发计划项目（2023YFC3606300）提供支持。

本文件起草单位：中山大学，华中科技大学，四川大学，北京大学，广西医科大学，武汉大学，复旦大学泰州健康科学研究院，开滦总医院，国药东风总医院，深圳市宝安区公共卫生服务中心、广东食标科技有限公司。

本文件主要起草人：柳雁，夏敏，潘安，李佳圆，庞元捷，姜侠，杨晓波，周小涛，高景利，荣爽，袁子宇，耿婷婷，杨坤，熊田甜，梁友，张竞雯，穆青。

引 言

增龄健康状态队列是指对同一研究人群在较长时期内进行多次、系统性的随访和数据收集，以追踪研究对象在年龄增长过程中生理、心理、社会功能等广泛健康状态的动态演变规律及其影响因素。

当前，我国在大跨度年龄健康研究领域仍存在空白，对增龄人群健康状态的动态演变过程的探索尚显不足。因此，建立和维护一个具有全国代表性、大规模、高质量的增龄健康状态队列，对于推动生物医学研究创新、实现健康增龄、为政府公共卫生决策制定提供科学依据具有重要的战略意义。它不仅能为自然科学领域发展做出实质性贡献，更有助于在全球生物医学研究中展现中国智慧和方案。

然而，大型人群队列研究具有随访周期长、任务重、覆盖区域广的特点，导致其数据来源广泛、内容复杂。规范、准确的数据是保障队列研究高质量和可持续性的关键前提。

本标准旨在解决上述挑战，针对增龄健康状态队列研究中的关键数据管理环节建立规范，具体包括：

- 生物样本的采集、暂存、检测与保存；

- 问卷数据的采集、处理与储存；

- 多类型来源数据的整合与治理；

- 数据和样本的共享。

本标准将为不同队列研究的多源异构数据制定数据融合治理的标准化规范，以保证不同地区和研究条件下实现方法的一致性，也为新建队列的组织实施和数据管理提供重要的参考依据。

增龄健康状态队列数据和生物样本采集处理指南

1 范围

本文件规定了增龄健康状态队列研究中数据采集、生物样本采集、数据与生物样本处理、以及数据和样本共享的基本要求和方

本文件适用于增龄健康人群队列研究的数据和生物样本的采集与处理。

2 规范性引用文件

下列文件对于本文件的应用是必不可少的。凡注日期的引用文件，其后所列的所有版本均适用于本文件；凡不注日期的引用文件，其最新版本（包括所有修订版）适用于本文件。

GB/T 37864 生物样本库质量和能力通用要求

GB/T 38576 人类血液样本采集与处理

GB/T 38735 人类尿液样本采集与处理

GB/T 41908 人类粪便样本采集与处理

GB/T 45633 人类唾液样本采集与处理

ISO/IEC 27001:2013 信息技术 安全技术 信息安全管理体

3 术语和定义

3.1

增龄健康状态 aging health status

随着年龄增长，个体在生理、心理和社会功能等方面的健康状况综合体现。

3.2

增龄健康状态队列 aging health status cohort

以中老年自然人群为研究对象，基于生命历程理论建立的前瞻性纵向研究队列。通过在较长的时间跨度内进行多频次、标准化、系统性的追踪随访，全方位收集生物学、心理学及社会环境数据，旨在揭示人群在增龄过程中健康稳态的演变轨迹，以追踪研究对象在年龄增长过程中生理、心理、社会功能等广泛健康状态的动态演变规律及其影响因素。

3.3

问卷信息采集 questionnaire information collection

通过结构化调查问卷收集研究对象人口学信息、健康史、生活方式、心理状况等相关信息的过程。

3.4

数据脱敏 data desensitization

对包含个人敏感信息的数据进行处理，屏蔽或替换能识别个人身份的信息。

3.5

元数据 metadata

描述数据采集时间、地点、方法及处理记录等信息的数据说明文件。

4 方案设计

4.1 方案设计整体原则

- 4.1.1 标准化原则：统一操作流程与质量标准。
- 4.1.2 安全性原则：保障采样对象安全，数据和生物样本安全。
- 4.1.3 可追溯原则：建立完整的样本标识与记录系统。
- 4.1.4 质量控制原则：实施全过程质量监控。
- 4.1.5 多中心原则：多中心并行开展调查时，应建立统一协调机制，各中心严格遵循统一的数据采集标准和质量控制流程，提前根据标准进行人员培训，确保调查工作协同一致。具体题目和调查项可根据各自具体项目细化，但应注意确保不同研究中心和随访轮次之间内容一致。
- 4.1.6 生理安全性与伦理关怀并行原则：首先需考量老年人机能衰退，通过优化采样体位与时长确保过程的高耐受性；其次应考虑将生物样本与身体功能评估动态关联，并建立危急值反馈机制以保障受试者权益。

4.2 数据和生物样本采集处理方案流程

增龄健康状态队列数据和生物样本采集处理方案流程图如下图所示。

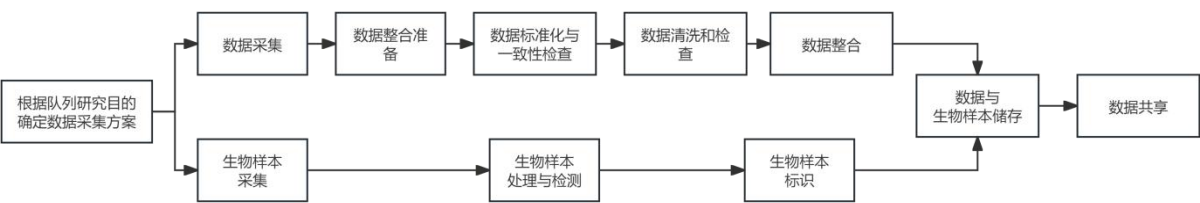


图1、增龄健康状态队列数据和生物样本采集处理方案流程图

5 数据采集

5.1 数据来源

- 5.1.1 涵盖问卷调查数据、体格/临床检查数据、生物样本检测数据三大来源。
- 5.1.2 数据采集内容包括一般人口学信息、问卷调查信息、体格检查信息、实验室检查测和影像检测信息、其他信息等。

5.2 基本要求

数据和生物样本采集需遵循《涉及人的生命科学和医学研究伦理审查办法》的规定，应满足《人类遗传资源管理条例》和《人类遗传资源管理条例实施细则》中相关要求。

5.3 问卷调查数据

- 5.3.1 问卷调查数据可按照一般人口学信息、躯体功能、心理精神状态、社会适应能力进行整合。躯体功能包括个人疾病史、家族病史、健康相关行为、自评健康状况、女性生育史；心理精神状态包括心理状况、认知状况；社会适应能力包括社会与文化背景、社会经济状况、社会支持等。
- 5.3.2 问卷调查由经过专业培训的人员实施，调查前需核对受试者身份并取得知情同意。
- 5.3.3 可采用纸质或电子问卷形式，均需使用经验证的标准化调查工具。
- 5.3.4 在采集过程中，可根据问卷设计的跳转逻辑提示受试者，如遇受试者无法回答的问题可予记录 并说明原因。

5.4 体格/临床检查数据

5.4.1 体格检查数据主要包括身高、体重、腰围、臀围、握力、收缩压、舒张压、心率等基础数据，以及由以上数据计算得到的衍生数据。

5.4.2 体格检查数据需要注意变量值间的逻辑核查，识别冲突值。对于异常数据应分析出现原因，通过通过核查原始问卷与数据集，重新计算衍生指标等方式修正错误。

5.4.3 体格检查数据采集应在统一标准操作程序（SOP）下由经培训并通过一致性考核的人员，使用经定期校准且符合计量认证的设备，于预设时间窗口以标准化流程完成。

5.5 生物样本实验室检测数据及其他数据

5.5.1 实验室检查数据以躯体功能维度为主，包括血常规、尿常规、肝功能、肾功能、血脂、尿酸、血糖、白蛋白等及各种衍生指标。

5.5.2 影像检查主要包括B超、CT等。

5.5.3 其他数据指那些难以归类于传统指标、但对研究结果具有潜在干扰或补充解释作用的数据，比如由智能手机或手环记录的每日步数、地理空间移动轨迹、久坐提醒触发频率、APP的使用时间和频率等

6 生物样本采集、处理、检测

6.1 生物样本采集

6.1.1 可供采集的生物样本包括血液、尿液、粪便和唾液等。

6.1.2 采集前详细记录受试者信息并取得知情同意，采集人员需经过培训并使用符合标准的容器和耗材。

6.1.3 血液样本采集：需满足GB/T 38576的要求，所有采血管室温静置30分钟后方可预处理，需避免溶血和过度震荡；记录采集时间及异常情况（如采血困难等）。

6.1.4 尿液样本采集：需满足GB/T 38735的要求，清晨第一次中段尿，使用无菌尿杯收集，采集后30分钟内送检或处理。

6.1.5 粪便采集：需满足GB/T 41908的要求。受试者在无菌铺垫上排便，使用采样勺取约1 g粪便置于含有保存液的管中，并将剩余部分另存于采集杯，采集后立即低温保存。

6.1.6 唾液采集：需满足GB/T 45633的要求。用口咽拭子重点擦拭双侧扁桃体区域充足次数（如6次），并分别置入DNA保存管和RNA保存管，留置一支备用作为阴性对照。

6.1.7 采集过程中避免污染，记录采样时间、环境条件和操作人员信息。关键步骤可进行双人复核并保留操作记录。

6.2 生物样本处理

6.2.1 生物样本处理与保存需满足GB/T 37864的要求。

6.2.2 采集后应尽快对样本进行预处理和保存，生物样本应在通风橱内低温操作。

6.2.3 血液样本处理与保存应满足GB/T 38576的要求，对因溶血、脂血或样本量不足等异常情况应记录并特殊标记。

6.2.4 粪便样本处理与保存需满足GB/T 41908的要求，应置于核酸保存管中，立即混匀并低温保存；分装操作需使用无菌器械，并挑取中段粪便以避免食物残渣。

6.2.5 口咽拭子采集后需满足GB/T 45633的要求，应短暂低温保存（如4℃），使用封口膜或密封袋防漏。

6.2.6 所有样本分装后应即时扫码入库，并在1小时内转移至-80℃冷冻箱或液氮罐长期保存。

6.2.7 样本入库应使用统一的标识系统，确保每个管都有对应的记录。

6.3 生物样本检测

对样本进行检测时，宜尽量采用现有国家标准或卫生行业标准，若使用其他方法进行测试时，需要对方法学进行验证，保证测试结果的稳定性以及可靠性。

7 数据处理与整合

7.1 数据整合准备工作

7.1.1 备份原始数据

在数据处理和整合前，除保存原始数据外，还应对每个处理阶段的数据文件进行备份，以确保数据的安全性和可追溯性。

7.1.2 文档管理

在数据整合过程中，必须详细记录每一步的数据处理过程，包括处理依据、方法和结果，便于后续核查和查询，同时保障数据处理的透明性和可重复性。

7.2 数据标准化与一致性检查

7.2.1 数据类型与格式统一

7.2.1.1 数据类型检查

应确保每列数据的类型符合预期（如数值型、字符型、日期型等），并统一格式，避免因格式不统一而导致的分析问题。

7.2.1.2 时间格式统一

时间型数据格式统一为“hh:mm”，确保小时数不超过24，分钟和秒数不超过60，并核对日期型数据的合理性（如确诊时间应小于等于当前年龄）。

7.2.1.3 单位标准化

所有数值数据应采用统一的计量单位，例如年龄以“年”为单位，体格检查、实验室检测指标采用国际单位制。

7.2.2 编码与映射

根据变量的数学属性保留其逻辑信息：

- 1) 对于分类变量，应采用独热编码以确保类别间相互独立且无虚假大小关系，将每个类别转换为一个独立的二进制列（0或1）；
- 2) 对于有序变量，需映射为等距或按比例数值以保留其等级序列（如：低=1，中=2，高=3）；
- 3) 对于连续变量，则需通过标准化或归一化消除量纲差异，确保数值在同一数量级下具有可比性。

7.2.3 逻辑性检查

核查数据的逻辑合理性，包括时间顺序（如出生日期应早于调查日期）、变量间的逻辑关系（如“从未吸烟”状态下不应报告“戒烟时长”），确保数据符合预期逻辑。

7.3 数据清洗和检查

7.3.1 数据空值处理

应注意识别数据中的空值，区分缺失值与因题目跳转产生的空值，并将缺失值统一编码为“999”。

7.3.2 异常值检测和处理

7.3.2.1 异常值识别

对数值型数据，根据统计方法（如3倍标准差）识别异常值，并结合生理常识判断其合理性。例如年龄应在纳入标准范围内；身高、体重和BMI等生理数据应处于常识范围内。确保选择题未出现无效选项（如选项编码为1-3，但填报为4）；单选题答案是否为单选等。

7.3.2.2 异常值处理

处理异常值应遵循先识别成因后精准施策的原则：对于超出合理范围的数据，进一步核查是否为录入错误，对于确证的错误数据应予修正或删除；而对于真实存在的极端值，应通过稳健统计或数据变换等手段减小其影响。最关键的是保持过程透明，必须在报告中对比处理前后的结果以确保研究结论的客观性与稳健性。

7.3.3 重复值检测

7.3.3.1 重复行处理

查找并处理重复行，判断是否删除或合并。

7.3.3.2 唯一标识符验证

为每个原始队列重新设计统一研究编号，并确保每个研究对象由独特的编号标识，避免重复。

7.3.4 衍生变量计算与验证

核查衍生变量的计算准确性，例如根据身高和体重计算BMI，确保计算过程和结果的准确性。

7.4 数据整合

问卷数据、体格检查数据以及实验室检查数据应分别进行数据整合。

8 数据和生物样本存储

8.1 数据存储

8.1.1 数据命名

所有数据集、文件和变量命名遵循统一规则，采用规范的命名格式以便识别和管理。例如，可采用“基线/随访_队列类型_疾病代号_组学类型_样本ID”的命名方式。

8.1.2 数据档案管理

所有数据集附带标准化的元数据文件，包括采集时间、地点、方法、处理记录等信息。

8.1.3 数据脱敏与隐私保护

8.1.3.1 研究数据中涉及个人身份的敏感信息（如姓名、身份证号、出生日期等）进行脱敏处理是至关重要的。

8.1.3.2 可对个人敏感字段应进行遮蔽或加密处理。

8.1.3.3 原始敏感数据需单独加密存储，仅供特定授权人员访问。

8.1.3.4 脱敏过程应由专人负责并建立复核机制，操作过程应记录在案。

8.1.4 数据存储与备份

8.1.4.1 所有电子数据在安全可靠的环境中存储是至关重要的。

8.1.4.2 可采用冷热三级存储体系（热存储为SSD、温存储为硬盘阵列、冷存储为磁带库），并部署在符合ISO 27001等信息安全管理标准的数据中心。

8.1.4.3 数据存储系统采用基于角色的访问控制和双因素身份验证机制，存储设备应启用生物识别+密码等双重访问认证。

8.1.4.4 备份策略应遵循“3-2-1”原则：至少保留三份数据副本、使用两种不同介质存储、并将其中一份备份存放于异地。

8.1.4.5 所有备份介质加密是至关重要的，并由专人管理密钥。

8.1.5 数据访问控制与安全管理

8.1.5.1 可采用最小权限原则，按需分配数据访问权限。

8.1.5.2 所有数据管理人员签署保密协议是至关重要的，访问系统时通过双重认证等安全手段。

8.1.5.3 实施日志审计机制，记录用户操作并定期审查，以防止未授权访问或数据泄露。

8.2 生物样本存储

8.2.1 样本标识

8.2.1.1 所有生物样本采用统一的标识体系，可使用条码或二维码标签，标签信息包含受试者研究编号、样本类型、采集日期和分装编号等，且在样本管身和管盖均进行标注。

8.2.1.2 可采用三级编码规则（地区代码+队列类型+个体ID+样本类型+分装序号）实现样本唯一标识。

8.2.2 样品入库

8.2.2.1 样本入库前应进行外观检查（如是否溶血、受污染等）并进行质量评估（例如随机抽样检测核酸/蛋白完整性）。

8.2.2.2 样本入库时应对照样本标识一致性、样本完整性进行检查。

8.2.2.3 所有样本即时生成入库记录，并与相应的研究数据在数据库中关联保存，以便追溯。

8.2.3 存储条件

8.2.3.1 短期存储若使用4-6℃环境储存，储存时间不宜不超过24小时，-80℃储存不宜超过5年。

8.2.3.2 长期存储可采用液氮气相。

8.2.3.3 储存生物样本使用的相应容器、标签应满足储存年限、条件的需求。

8.2.4 存储管理

8.2.4.1 可采用24小时温度监控与报警系统，进行双人双锁管理制度，每月库存盘点，定期质量抽检，重点检查存储条件符合性、温度报警记录、样本完整性等。

8.2.4.2 对于发现的异常样本应及时隔离并记录，评估对后续实验的影响；严重质量问题应启动追溯机制并可能召回相关样本。

8.2.5 样本运输

8.2.5.1 运输容器宜使用认证的低温运输箱，干冰量应保证至少5天维持温度。

8.2.5.2 运输要求需全程进行温度监控与记录，交接时检查样本状态并记录。

9 数据共享

9.1 共享原则

9.1.1 分级共享

9.1.1.1 数据共享实行分级管理，包括：

a) 完全开放数据：脱敏处理、无个人身份信息的数据，可在伦理合规前提下对科研机构开放；

b) 受控访问数据：含有限敏感信息或经组合可能识别个人的数据，通过申请、伦理审查和协议授权，在受控环境限获批研究者使用；

c) 限制访问数据：高度敏感数据（如原始基因组），仅项目指定的数据管理机构和经严格审批合规的研究团队可在专用环境访问。

9.1.1.2 各级数据的访问条件、使用范围、保存期限、审查流程应明确规定，并与信息安全管理程度保持一致。

9.1.2 使用条款

9.1.2.1 共享数据和样本仅限于批准的科研或公共卫生目的；未经书面许可，禁止商业用途。

9.1.2.2 使用共享数据或样本取得的论文、报告、专利等成果，应在适当位置致谢增龄健康状态队列项目及样本来源单位；致谢内容应符合数据使用协议。

9.1.2.3 使用者应在合理期限内向队列数据库回馈整理后的研究成果（如衍生变量、分析结果、代码或算法工具等）；具体内容、形式和时限要求应在数据使用或材料转移协议中约定。

9.2 跨中心协作

9.2.1 数据交换

9.2.1.1 多中心数据交换应建立安全传输通道，采用加密、认证、完整性校验等措施；防止传输过程中数据被篡改、泄露或丢失。

9.2.1.2 数据交换过程应采用区块链等不可篡改技术记录日志；日志至少包含发起单位、接收单位、数据类型范围、时间戳、操作人、审批信息等，确保全程可追溯。

9.2.1.3 数据交换应同步提供完整元数据（数据字典、采集方案、质量控制信息、版本号、处理记录等）；确保数据一致、可解释、可复现。

9.2.2 样本转运

9.2.2.1 各中心和参与单位间的样本转运应制定统一的运输标准和流程；明确包装、温度、时限、交接、安全防护等要求。

9.2.2.2 样本转运由指定专人负责，全程进行温度监控记录；出现温度异常、延误或意外情况时，及时报告并采取补救措施。

9.2.2.3 样本运抵后应立即检查外观、数量，并进行必要的质量检验；检查结果录入样本管理系统，对不合格样本予以标记和处置。

10 质量控制

10.1 问卷调查质量控制

10.1.1 承担问卷调查工作的机构应制定调查质量控制方案，明确调查员培训与质控负责人职责，确保调查人员资质合格、操作流程规范统一。

10.1.2 问卷调查全过程应实施质量监控：现场调查时检查问卷完整性和一致性，纠正遗漏或错误信息；调查后由数据管理人员进行逻辑审核和复核，及时发现并更正问题。

10.1.3 应建立问卷调查质控记录和反馈机制，对质控发现的问题及时整改并形成闭环改进；保存质控活动记录，并可追溯至相关责任人员。

10.1.4 调查员在现场对填写结果进行初步检查，提醒受试者纠正明显遗漏或矛盾。

10.1.5 在数据整理阶段，对问卷数据进行逻辑审核和一致性校验，及时发现并处理错漏。

10.1.6 对于因跳题而产生的空值，注明为“逻辑跳转”并保留空白；对非跳题产生的缺失值，采用统一编码（例如“999”）表示。

10.2 体格检查质量控制要求

10.2.1 体格检查应由经过专业培训并通过一致性考核的人员实施，并由质控负责人监督执行，确保检查操作规范、结果可靠。

10.2.2 所用测量仪器设备须按国家计量规程定期检定或校准，确保其性能符合要求，并对关键设备使用前后进行质量校验。

10.2.3 应建立体格检查质量监测机制，对重要指标实行复测或抽查验证，对异常测量结果及时复核确认，并记录质控结果用于持续改进检查方法。

10.3 生物样本质量控制

10.3.1 生物样本库质量控制应参考GB/T 37864的要求。

10.3.2 实验室应建立生物样本检测质量管理体系，定期开展室内质量控制和室间质量评价（能力验证），并由专人审核质控结果，持续改进检测方法。

10.3.3 生物样本的接收、运输、处理与保存须严格按照标准操作规程（SOP）执行，防止标本识别错误、污染或降解，并记录环境条件和时间信息，确保样本质量。

10.3.4 检测仪器应按计量规范定期检定或校准，检测方法须经过验证并标准化；检测结果应实行双人复核或信息系统自动审核，确保结果准确可靠。

10.4 数据处理质量控制要求

-
- 10.4.1 研究单位应建立数据处理质量控制制度，明确数据清理、整合、分析等环节的操作规范、权限和责任，确保数据处理过程受控可追溯。
- 10.4.2 数据处理所用软件程序应经过验证，重要算法和分析步骤需由第二人独立复核或同行评估，关键结果应重复计算验证，保证数据处理的准确性和可靠性。
- 10.4.3 应建立数据处理变更记录和质量审计机制，对数据清洗、转换、分析等操作进行记录和审计追踪，并定期形成质量报告用于改进数据处理流程。
- 10.4.4 对所有数据应进行完整性校验，可使用 MD5 校验等技术定期扫描数据损坏情况并自动恢复。
- 10.4.5 需识别并处理空值、异常值和重复记录：对非跳题的缺失值应统一编码（例如采用“999”编码）。
- 10.4.6 使用统计方法（如3倍标准差法）和专业知识发现异常值并加以处理。
- 10.4.7 检查重复记录并确保每位受试者具有唯一研究编号。
- 10.4.8 所有清洗、修改和质量检查过程应有完整记录，并定期生成质量控制报告。
-