ICS 67.040
CCS 53

T/GDFDTAE

团　体　标　准

T/GDFDTAEC XXXX—2025

增龄健康状态队列数据和样本采集处理指南

Guidelines for data collection in aging health status queue research

（征求意见稿）

在提交反馈意见时，请将您知道的相关专利连同支持性文件一并附上。

2025－XX－XX 发布　　2025－XX－XX 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（）　提供支持。

本文件起草单位：

本文件主要起草人：

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中国方案。

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。

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、生物样品采集、数据与样本处理的基本要求和方法。

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。

ISO/IEC 27001:2013 信息技术 安全技术 信息安全管理体系 要求

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

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。

3.3

生物样本 biological sample

在研究中从受试者采集的用于检测分析的生物材料，如血液、尿液、粪便、口咽拭子等。

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

样本标识系统 sample identification system

对生物样本采用统一编码规则进行唯一标识和管理的系统。

3.6

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 数据、生物样本采集和管理方案设计

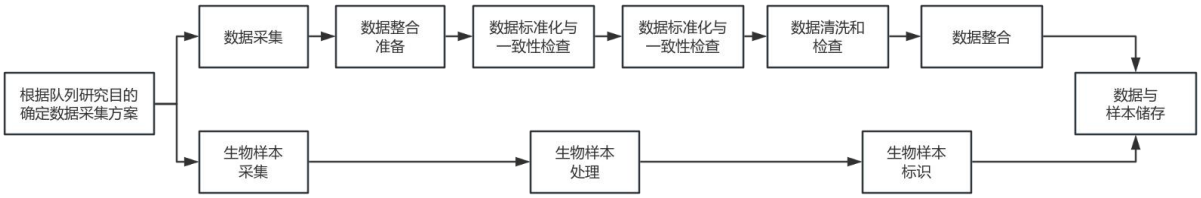
4.1 数据质量管理原则

增龄健康队列研究在进行数据和样本采集前，应根据研究目标完成数据和样本管理设计,方案设计宜涵盖在对增龄人群进行持续随访过程中，综合采集问卷调查、生物样本检测等多源数据的质量管控。

4.2 数据采集、处理、存储整体原则

4.2.1 标准化原则：统一操作流程与质量标准；
4.2.2 安全性原则：保障采样对象安全，数据和生物样本安全；
4.2.3 可追溯原则：建立完整的样本标识与记录系统；
4.2.4 质量控制原则：实施全过程质量监控。

4.3 数据和生物样本采集、处理、储存整体流程



5 数据信息采集

5.1 采集方式

5.1.1 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。
5.1.2 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。
5.1.3 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。
5.1.4 体格检查数据采集应在统一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。

5.2 采集内容

5.2.1 问卷信息包括问卷调查信息以及体格检查信息，其中问卷调查信息可覆盖躯体生理功能、心理精神状态和社会适应能力等多个维度。
5.2.2 躯体生理功能包括一般人口学信息、个人及家族疾病史、健康行为、 自评健康状况、女性生育史等。
5.2.3 心理精神状态包括心理状况和认知状况。
5.2.4 社会适应能力包括社会经济状况、文化背景和社会支持等。
5.2.5 体格检查主要包括身高、体重、腰围、臀围、握力、收缩压、舒张压、心率等身体测量指标。
5.2.6 具体题目和调查项可根据研究目的细化，确保不同研究中心和随访轮次之间内容一致。

5.3 质量控制

5.3.1 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。
5.3.2 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。
5.3.3 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。

6 数据整合准备工作

6.1 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。

6.2 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 数据标准化与一致性检查

7.1 数据类型与格式统一

7.1.1 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。

7.1.2 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）。

7.1.3 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 编码与映射

为分类数据制定统一的编码规则，如统一编码“1”为“是”，“2”为“否”，“3”为“不清楚”。

7.3 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。

8 数据清洗和检查

8.1 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”。

8.2 异常值检测和处理

8.2.1 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30-100岁之间；身高、体重和BMI等生理数据应处于合理范围内。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。

8.2.2 异常值处理

对于超出合理范围的数据，进一步核查是否为录入错误以决定是否保留、删除或替换异常值。如果保留，考虑标记为特殊值以便后续分析。

8.3 重复值检测

8.3.1 重复行处理

查找并处理重复行，判断是否删除或合并。

8.3.2 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。

8.4 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。

9 数据整合

9.1 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。

9.1.1 问卷数据整合

问卷调查数据可按照一般人口学信息、躯体生理功能、心理精神状态、社会适应能力进行整合。躯体生理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。

9.1.2 体格检查数据

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。

体格检查数据以躯体生理功能维度为主，数据包括基础指标和衍生指标。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。

9.1.3 实验室检查数据整合

实验室检查数据以躯体生理功能维度为主，包括基础指标、衍生指标。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。

10 数据质量控制

10.1 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。
10.2 对问卷调查数据进行逻辑一致性检查，例如验证时间顺序（如出生日期必须早于调查日期）和变量间关系（如“从未吸烟 ”状态下不应记录戒烟时长）。
10.3 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 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。
10.5 检查重复记录并确保每位受试者具有唯一研究编号。
10.6 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。

11 生物样本采集

11.1 采集样本类型

11.1.1 可供采集的生物样本包括血液、尿液、粪便和口咽拭子等。
11.1.2 血液样本采集：受试者空腹 10 小时以上，使用抗凝管和非抗凝管分别采集约12 mL 静脉血，采集后轻轻颠倒混匀并置于室温静置后进行离心。
11.1.3 尿液样本采集：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。
11.1.4 粪便采集：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。
11.1.5 口咽拭子采集：用拭子重点擦拭双侧扁桃体区域各 6 次以上，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。

11.2 样本处理与保存

11.2.1 采集后应尽快对样本进行预处理和保存。
11.2.2 血液样本采集后于 4℃ 、3000 rpm 离心 15 分钟，分离血浆或血清、红细胞、白细胞等生物样本。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。

11.2.3 粪便样本应置于核酸保存管中，立即混匀并低温保存；分装操作需使用紫外消毒器械，并挑取中段粪便以避免食物残渣。
11.2.4 口咽拭子采集后应短暂 4℃保存，使用封口膜或密封袋防漏。
11.2.5 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。
11.2.6 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。

11.3 样本标识与标签

11.3.1 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。
11.3.2 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。
11.3.3 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。

11.4 样本采集注意事项

11.4.1 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。
11.4.2 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
11.4.3 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）
11.4.4 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。

12 数据与样本储存要求

12.1 数据的储存要求

12.1.1 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。

12.1.2 元数据管理

12.1.2.1 分类变量应采用统一编码规则（如“ 1 ”代表“是 ”，“2 ”代表“否 ”，“3 ”代表“不清楚 ”）。
12.1.2.2 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。

12.1.3 数据脱敏与隐私保护

12.1.3.1 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。
12.1.3.2 可对个人敏感字段应进行遮蔽或加密处理。
12.1.3.3 原始敏感数据需单独加密存储，仅供特定授权人员访问。
12.1.3.4 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。

12.1.4 数据存储与备份

12.1.4.1 所有电子数据在安全可靠的环境中存储是至关重要的。
12.1.4.2 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。
12.1.4.3 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。
12.1.4.4 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。
12.1.4.5 所有备份介质加密是至关重要的，并由专人管理密钥。

12.1.5 数据访问控制与安全管理

12.1.5.1 可采用最小权限原则，按需分配数据访问权限。
12.1.5.2 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。
12.1.5.3 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。

12.2 样本的存储与运输

12.2.1 存储条件

12.2.1.1 短期存储若使用4℃环境储存，储存时间不宜不超过24小时，-80℃存储不宜超过5年。
12.2.1.2 长期存储可采用液氮气相长期保存。

12.2.2 存储管理

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，每月库存盘点与质量抽检。

12.2.3 样本运输

12.2.3.1 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。
12.2.3.2 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。