标准·方案·指南

大型人群队列研究数据安全技术规范（T/CPMA 002–2018）

中华预防医学会
通信作者：李立明，Email：lmleeph@vip.163.com
基金项目：国家重点研发计划（2016YFC0900500，2016YFC0900504）
DOI：10.3760/cma.j.issn.0254–6450.2019.01.004

Technical specification of data security for large population-based cohort study（T/CPMA 002–2018）

Chinese Preventive Medicine Association
Corresponding author: Li Liming, Email: lmleeph@vip.163.com
Fund programs: National Key Research and Development Program (2016YFC0900500, 2016YFC0900504)
DOI：10.3760/cma.j.issn.0254–6450.2019.01.004

前　言

本标准按照GB/T 1.1–2009给出的规则起草。

本标准由北京大学提出，归口中华预防医学会。

主要起草单位：北京大学、中国医学科学院、北京理工大学。

本标准主要起草人：李立明、余灿清、吕筠、卞铮、谭云龙、刘亚宁、郭彧、汤海京、杨旭。

本标准为首次发布。

引　言

大型队列研究涉及人数往往超十万人，势必需要收集大量的个人信息，包括个人身份识别信息、个人生活习惯信息、身体状况及疾病信息等，除个人信息外还会收集社会、经济、环境等与健康相关领域的其他数据。巨量数据一旦泄露，将对研究对象个体及研究工作造成不可估量的影响及危害，因此大型人群队列的数据隐私保护的重要性不言而喻。

大型人群队列数据库作为重要信息的承载主体，存储着各种隐私数据、业务数据，其安全性稳定性直接关系着工作是否能正常运行，队列人群个人数据是否得到保护，因此保证数据库的安全稳定运行是十分重要的。大型人群队列由于涉及调查地域广阔、人数众多，往往需要使用调查应用系统接入数据库，因此，数据库作为后台开发网络应用系统，面临的风险更大。本标准将就大型人群队列的数据隐私保护及数据库安全稳定管理进行规范。

大型人群队列研究数据安全技术规范

1　范围

本标准规定了大型队列研究实施过程中现场调查、数据处理、数据分析中的数据隐私安全保护要求及大型队列研究应用数据库管理时要求的安全原则。

本标准对不同来源、不同类型的队列数据隐私性及数据库安全管理进行规范化要求，适用于已建立或拟开展大型人群队列研究的机构，包括但不限于大型自然人群队列、区域性人群队列、针对某一特殊疾病或基于特殊机构开展的人群队列。本标准还可供规模相对较小的人群队列研究参考。

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。凡是注日期的引用文件，仅注日期的版本适用于本文件；凡是不注日期的引用文件，其最新版本（包括所有修改单）适用于本文件。

GB/T 20269信息安全技术信息系统安全管理要求

GB/T 20271 信息安全技术信息系统通用安全技术要求

3　术语和定义

下列术语和定义适用于本文件。

3.1　个人信息 personal information

是指以电子或者其他方式记录的能够单独或者与其他信息结合识别自然人个人身份的各种信息，包括但不限于自然人的姓名、出生日期、身份证件号码、个人生物识别信息、住址、电话号码等。

3.2　直接个人信息 direct personal information

可以单独识别本人的个人信息，如姓名、家庭住址、身份证号码、基因等。

3.3　间接个人信息 indirect personal information

不能单独识别本人，但和其他信息结合可以识别本人的个人信息，如身高、体重、个人生活习惯、疾病史等。

3.4　数据隐私保护 data privacy protection

指对单位和个人敏感的数据进行保护的措施。

3.5 数据控制者 data controller

决定数据处理目的和方式的单位和个人，涉及采集、处理、存储隐私数据的全过程。

3.6 数据处理者 data processor

代表控制者处理数据的单位和个人。

3.7 大数据 big data

规模在获取、存储、管理、分析方面远远超出传统数据库软件工具能力范围的数据集合，具有海量数据规模、快速数据流转、多样的数据类型和价值密度低的四大特征。

4 大型队列研究数据隐私保护

4.1 一般要求

队列数据隐私保护是对队列研究中采集、产生、储存和利用等覆盖数据生存周期的全过程的数据信息实施符合安全等级的保护原则，包括：

——落实安全管理制度，制定安全规划；

——评估数据隐私级别及暴露风险；

——制定、选择并落实安全策略；

——保证维护支持；

——人员安全教育及管理等。

4.2 数据隐私的类型

队列研究主办机构作为数据信息保护的责任主体，应全面评估收集信息的安全层次，本着“收集者即负责者”的基本原则，对不同层面的数据其隐私保护要执行差别化管理，对直接个人信息、个人隐私信息等敏感数据应进行加密处理。

a）队列研究收集的隐私数据包括：直接个人信息，如姓名、籍贯、性别、婚姻、身体、出生日期、民族等信息及间接个人信息如身高、体重、个人生活习惯、疾病史、收入、生理心理状态、宗教信仰等，均应受到隐私保护；

b）两个及以上个人信息相关联时的信息安全保护等级应高于任何单一信息；

c）应注意妥善保管队列研究中收集的各项纸质材料，如知情同意书、登记表、各类疾病报告卡片、纸质问卷等，避免数据隐私泄露；

d）应采用多种技术手段保护电子调查资料数据的安全性。

4.3 数据隐私参与角色

队列研究的数据保护者根据角色分工可分为数据控制者及数据处理者，需区分其不同角色的工作性质，制定针对性隐私保护策略制定相应的工作方案：

4.3.1 数据控制者

队列研究的数据控制者应该包括研究的设计者及管理者，该角色的职责包括：

a）依据国家相应法律法规，讨论并决定预期安全保护目标，确定适宜的安全保护等级，实施适当的技术方案和组织措施，包括研究设计的整理构架对于数据隐私保护的影响，以确定处理手段和处理隐私数据策略；

b）秉承数据保护原则的目的，选择必要的保障措施，如数据加密、数据匿名化、数据最小化、数据脱敏、分布式隐私保护等，以符合法律要求，最大限度的保护数据主体及队列人群的权利。

4.3.2 数据处理者

队列研究的数据处理者指参与研究且存在接触隐私数据机会的工作人员、合作方人员等，对该角色的要求包括：

a）未经控制者授权同意，不得对数据进行处理、转移；

b）使用数据进行分析时只能在指定的电脑上进行，或通过个人账号密码登陆服务器进行，不应使用个人电脑进行数据处理；

c）及时发现数据泄露的风险，并向数据控制者报告，协助数据控制者进行相应的补救措施。

4.4 数据隐私环节

队列研究主办机构应根据伦理学的要求和现场工作的实践，通过加密和其他安全措施，保护受试者的基本利益。应根据其数据隐私保护工作环节可分为研究设计、现场调查和数据处理三个阶段：

4.4.1 研究设计

不同的研究设计决定了采用不同的隐私保护技术，数据控制者应全面考虑数据的储存、使用和管理方式，制定合理的数据隐私保护策略和具体实施方案。为其合理选择风险防范措施提供真实可靠的依据：

a）研究设计应采取措施提高纸质资料的保密性、降低单机版数据录入程序及数据库在多终端的泄露风险；

b）研究设计应将直接个人信息与间接个人信息分离，调查问卷不得呈现直接个人信息，可使用研究编码进行链接，避免因问卷遗失造成的个人信息泄露。

4.4.2 现场调查

4.4.2.1 知情同意授权

数据采集人员应向研究对象提供其接受调查必须的所有信息，通过完整充分的说明和介绍，对研究对象的有关询问进行全面必要的回答和解释，使研究对象全面了解需调查的内容及隐私数据安全性保证。

数据采集应在研究对象在填写知情同意书，信息收集合法化后开始执行。知情同意书保存期限不应短于研究开展时限。

4.4.2.2 纸质调查形式

数据采集人员需谨守职业道德，不应对外泄露研究对象隐私，纸质调查形式还应遵循以下数据隐私保护原则：

a）妥善保管纸质调查问卷和各种记录表格，调

查完成后及时回收保存,不得造成信息泄露;

b)纸质调查问卷需匿名化处理;

c)不应对纸质问卷进行复印翻拍;

d)保存完毕后根据相应保存制度及保密程度进行妥善销毁。

4.4.2.3 电子化调查形式

a)应确保信息采集的电子终端设备为授权设备且仅用于调查工作,不得使用非授权设备进行调查;

b)终端设备硬盘应经过软件进行全盘加密及相应权限设置,避免设备遗失及误操作造成数据损失;

c)终端设备应启用防火墙和防毒软件,并进行硬盘加密,以防设备丢失导致数据泄露;

d)终端设备如需安装非调查用的第三方软件和连接网络,应经专业人士评定其风险性;

e)数据经过移动终端安装的调查数据采集软件录入后,应为加密格式保存,不得在终端设备上以未加密的形式保存隐私数据。

4.4.3 数据处理

4.4.3.1 数据处理活动的记录

每一位接触数据的人员,应当依其职责保持处理活动的记录。具体记录应包括以下所有信息:

a)控制者以及联合控制者、控制者代理人和数据保护员的姓名和联系信息;

b)处理的目的;

c)数据主体的类别和个人数据的分类的描述;

d)申请的变量记录表。

4.4.3.2 隐私数据管理策略

数据控制者、处理者应当执行合适的技术措施和有组织性的措施来保证合理应对风险的安全水平,制定符合本队列研究的特性的隐私数据存储、使用、交换及发布制定相关操作规程要求,包括以下方面:

a)数据控制者应根据其不同的数据性质按照信息安全等级制定相应的保护定级策略,隐私性数据安全保护等级原则上不应低于第三级;

b)数据处理者应根据其不同的工作内容得到差异化的授权;

c)隐私数据的储存应由专人管理,数据文件形式及数据库形式都应根据其相应的存储特性进行处理,非电子数据形式应保证其纸质资料的安全性。

d)个人隐私数据仅限于本研究项目使用,进行交换及发布时应经过匿名化和加密处理;

e)应识别用于收集数据的动态数据库及用于研究分析的静态数据库的不同权限并进行相应管理;

f)应保持数据库系统持续的保密性、完整性、可用性以及弹性的能力;

g)在发生自然事故或者技术事故的情况下,保证存储有用信息以及及时获取个人信息的能力;

h)定期对测试、访问、评估技术性措施以及组织性措施的有效性进行处理,力求确保处理过程的安全性;

i)安全账户的等级评估应当尤其重视处理过程中的风险问题,特别是抵御意外和非法销毁、损失、变更、未经授权披露或者是个人数据的传送、存储和处理过程中的风险;

j)考虑通过去中心化的分布式节点储存方式代替中心数据库以便提高安全保护等级。

4.4.3.3 隐私数据保护技术

数据控制者、处理者宜使用以下技术,对数据进行隐私保护:

a)基于数据失真的技术:通过添加噪音、随机化、阻塞与凝聚、差分隐私保护等方法,使敏感数据失真但同时保持某些数据或数据属性不变,仍然可以保持某些统计方面的性质;

b)基于数据加密的技术:采用安全多方计算SMC、分布式匿名化等加密技术在数据挖掘过程隐藏敏感数据;

c)基于限制发布的技术:有选择地发布原始数据、不发布或者发布精度较低的敏感数据,实现隐私保护。

4.4.4 数据分析

数据完成处理后,交由数据分析者进行科研分析,为保证数据分析阶段数据隐私保护的安全性,应做到以下几方面:

a)做好数据分析活动的记录;

b)数据分析者应签署相关保密协议以确保数据安全及不试图进行研究对象的身份确认;

c)数据使用者不应直接接触隐私数据,数据处理者负责向其提供相应的数据;

d)数据应去除个人隐私数据的相关变量,所有研究对象的ID号应该进行数据脱敏以匿名化;

e)提供给数据分析者的所有数据应经过安全渠道进行传递;

f)分析数据库应存放在安全介质,数据分析者对其全权负责。

5 大型队列研究数据库安全稳定性管理

5.1 数据库安全的原则

a)全面覆盖原则:从信息采集生成、存储备份、分析处理、共享使用、传输发布,到销毁清除等数据生命周期中的不同阶段,有针对性的提出安全管理规范和部署技术措施;

b)分级保护原则:不同的数据其来源、内容、用途存在很大差异,数据保护的需求也有所不同。对

不同级别和类型的数据，在数据存储、数据共享、数据加密、数据销毁的环节应采取不同的措施；

c）审计追责原则：对数据的全部操作和访问操作都应该记录操作员和访问者的身份信息，安全措施对数据的访问行为进行审计，任何对数据操作和访问行为都应该可以追溯到个人；

d）守法合规原则：数据库安全防护应严格遵守网络安全法以及相关的法律法规。

5.2 数据库安全的策略

大型人群的队列研究应将数据库安全工作处于信息安全防护体系的核心位置，避免受到外部攻击者攻击。数据库自身应具备充分的安全措施，能够抵御并发现入侵者。为了保证数据库安全，宜从物理安全、网络安全、服务器安全和数据库安全四个层次进行以下操作：

5.2.1 物理安全

数据库物理安全的基本要求包括：

a）应控制数据、电脑、媒介或拷贝材料的建筑、房间、橱柜的使用权；

b）在仓库中记录删除、访问的媒介或拷贝材料；

c）保证存储媒介的安全，存储安全包括物理安全、网络安全和计算机系统和文件的安全，以防止未经授权的访问或不需要的数据更改、信息的泄露或销毁。存储介质的质量和相关的数据读取设备的可用性需保证数据的可访问性，应保证存储介质质量可靠；

d）仅在特殊情况下传输敏感数据，向计算机制造商提供包含敏感数据的故障硬盘可能会导致安全问题；

e）数据文件应设定适宜的频率复制到新存储媒介上；

f）任何存储数据，即使是短期项目，都应包含至少两种不同的存储形式，例如硬盘驱动器和DVD光盘，应定期检查数据完整性；

g）存储数据的地区和房间应经过严格考察，无论是储存数码或非数码资料、光或磁存储介质，应保证存储的物理环境微气候适宜且无发生自然灾害的危险；印刷的材料和照片易受阳光和酸的影响。应选取优质材料，如无酸纸、不生锈的回形针文件夹和盒子等。

5.2.2 网络安全

数据库网络安全应根据的基本要求包括：

a）不应在服务器或连接到外部网络的计算机上存储包含个人信息的敏感数据，特别是主机服务器；

b）应利用数据库漏洞扫描系统扫描数据库，给出数据库的安全评估结果，暴露当前数据库系统的安全问题；

c）应利用专业的安全软件扫描应用系统，发现应用漏洞，及时堵住；

d）管理者宜模拟攻击者攻击，对数据库进行探测性分析，重点检查对象权限是否越权等，并收集应用系统漏洞和数据库漏洞；

e）应检查端口是否安全、访问协议是否安全；

f）应采用信任IP访问，通过设置入站规则或防火墙来限制数据库访问的信任；

g）应加强防火墙保护和与安全相关的升级和补丁操作系统，避免病毒和恶意代码。

5.2.3 服务器的安全

服务器安全是指服务器上的操作系统安全及对服务器安全保护采用的安装防火墙安全软件的安全，应确保数据库服务器及应用服务器相对独立，且由于服务器硬件损坏导致的系统崩溃磁盘损坏的几率较大，宜建立自有机房，或租用专业机房、云服务器。以下为服务器应采用的各方面安全策略，同时凡是接入局域网的终端设备也应该采用相同的安全策略：

5.2.3.1 操作系统的安全

操作系统包括服务器系统及所有纳入服务器局域网络内可以与服务器进行交互的终端操作系统，具体应做到以下安全策略：

a）应安装正版的操作软件，包括数据库操作系统，如SQL Server；

b）应对补丁进行定期升级更新；

c）操作系统宜采取最少应用软件安装原则；

d）操作系统的账户管理应采取以下策略：禁用超级用户，停用访客Guest账户，禁止远程访问；去除所有测试账户、共享账户和普通账户；对用户组策略设置相应的权限，并且经常检查系统账户，删除不用账户；

e）应减少使用管理员账户登录的频率，以免被某些软件窥探到；

f）应制定密码策略，设置密码的最小值、使用期限，重命名管理员账户，并为其设置高强度密码，包括大小写英文字符、数字、特殊字符等，并定期更换；

g）应安装防火墙，启用Windows系统自带的防火墙或者安装第三方专业防火墙，不应直接进行外网连接；

h）应安装正版杀毒软件，并保持一定频率的病毒库升级和全盘查杀。

5.2.3.2 文件的安全

对于服务器及终端机保存的文件，应采用以下方式进行安全保护：

a）通过线路-交互式不间断电源（UPS）系统保护服务器；

b）实现对数据文件的密码保护和控制访问，例如“禁止访问”、“只读”、“读写”或“管理员权限”；

c）对文件、文件夹或整个硬盘加密的访问控制；

d）将共享文件的权限设置为授权用户，避免任

何有权进入网络的用户能够访问这些共享文件；

e）在未加密前不通过电子邮件或其他文件传输方式发送个人或机密数据；

f）在需要时以统一的方式销毁数据：删除文件和重新格式化硬盘驱动器；

g）对机密数据的管理人员或用户实施保密协议；

h）对于包含个人或敏感信息的数据，应尽量避免云存储。

5.2.4 数据库安全

5.2.4.1 数据库备份

数据库备份可对数据库或事务日志进行复制，当系统、磁盘或数据库文件损坏时，可以使用备份文件进行恢复，防止数据丢失，常见的数据备份分为完整备份、差异备份和事务日志备份，应根据数据库需求应用不同策略。数据库备份执行时应采用以下策略：

a）宜备份整个系统而不是指定文件，内容包括用户表、系统表、索引、视图和存储过程的所有数据库对象；

b）每次更改数据之后应备份或定期进行备份。可使用自动备份程序来备份频繁使用的和关键的数据文件；

c）主拷贝的备份应为适合长期数字保存的文件格式，即开放或标准格式，而不是私有格式；

d）包含个人信息的数据，不宜过多备份，可保留主文件和一个备份副本，并对数据进行加密；

e）备份频率：原则是尽可能的减少数据损失，对于普通数据，如一天一次，重要数据应提高备份频率，对于大型数据库，宜采用差异备份；

f）备份文件存储媒介的选择取决于文件的数量、数据类型和备份方法，宜选用合适的安全介质进行存储。

5.2.4.2 数据库加密

数据库库内加密可以保障数据库安全和数据安全，加密不应影响数据库性能，可对特定列进行列级加密，或对整库进行数据库级加密。加密应满足以下要求：

a）加密可用于安全存储和发送文件；

b）为保证数据安全，任何包含敏感信息和数据的数字文件或文件夹都应加密，如队列人群的身份证号、姓名、家庭住址、联系方式等。直接个人信息可从数据文件中删除，并在更严格的安全措施下单独存储；

c）加密类型和级别与受保护的数据的敏感程度相对应；

d）密钥应储存安全，密钥管理机制应方便可靠。

5.2.4.3 数据库分类

大型人群队列数据库由于实时更新，数据库还应做以下分类设置：

a）科研分析前应对实时数据库进行清理，生成不同的数据库（实时数据库、镜像分析数据库等）以实现不同的使用目的；

b）可用于识别个人身份的信息（姓名、身份证、住址等）不可对研究者开放；

c）研究编码作为个体在研究项目的唯一标识，应进行匿名化处理。

5.3 安全管理员要求

安全管理员需要对网络及数据库的软硬件进行安全管理，应满足以下要求：

a）配备专人作为网络及数据库管理员，做好网络的日常维护与网络及数据库管理，保证所维护管理的系统正常运转；

b）安全管理员的具体工作内容包括网络基础设置管理、网络操作系统管理、网络应用系统管理、网络用户管理、网络安全保密管理、网络信息存储备份和网络机房管理等；

c）网络管理员的职业道德是管理员从事该工作的核心，应定期接受安全及职业培训，提高责任心、业务水平；

d）网络管理员及使用者应该具备良好的职业道德，在职或离职后的约定时间段，都应做到不更改、不攻击、不泄露数据，不得恶意操作造成数据内容泄露。

5.4 安全审计员要求

数据库安全审计管理的人员要求，应满足以下要求：

a）宜配备专人建立独立审计，确保风险管理的有效性；

b）根据不同安全等级的不同要求，制定相应的安全审计管理规定；

c）对安全审计的产生的不同数据进行记录、存储、分析、查阅，进行提供完整的数据库审计分析、泄密轨迹分析、数据库访问关系可视、数据库攻击威胁分析等；

d）汇总审计问题书面通知所有有关部门和人员，以便进行相应的调整。

参 考 文 献

[1] GB 17859—1999 计算机信息系统安全保护等级划分准则.
[2] GB/T 20269—2006 信息安全技术 信息系统安全管理要求.
[3] GB/T 20271—2006 信息安全技术 信息系统通用安全技术要求.
[4] GB/T 20273—2006 信息安全技术 数据库管理系统安全技术要求.

（收稿日期：2018-12-13）
（本文编辑：李银鸽）