• 医学信息研究 •

基于数据安全的健康医疗科学数据分级指南研究*

张胜发 马玉环 张敬晨 王嘉阳 孙婧雯 张 月 张晓宇 周 伟

（中国医学科学院国家人口健康科学数据中心 北京 100730）

〔摘要〕 **目的/意义** 完善健康医疗科学数据分级保护，促进数据安全共享、充分利用，提升数据资源价值。**方法/过程** 采用文献分析法，并开展多轮专题专家小组讨论，研讨健康医疗数据分级的原则、规则和实施步骤。**结果/结论** 基于健康医疗科学数据安全相关法律法规或政策要求，构建健康医疗科学数据分级指南，提出健康医疗科学数据分级的具体原则、评级规则与实施步骤，为健康医疗科学数据生产、管理及应用活动中的数据分级提供参考。

〔关键词〕 数据分级；健康医疗数据；科学数据；数据安全

〔中图分类号〕R－058 〔文献标识码〕A 〔DOI〕10.3969/j.issn.1673－6036.2023.08.004

Study on the Guidelines for Data Grading of Health and Medical Science Data from the Perspective of Data Security

ZHANG Shengfa, MA Yuhuan, ZHANG Jingchen, WANG Jiayang, SUN Jingwen, ZHANG Yue, ZHANG Xiaoyu, ZHOU Wei

Chinese Academy of Medical Sciences, National Population Health Data Center, Beijing 100730, China

〔**Abstract**〕 **Purpose/Significance** To improve the graded protection of health and medical science data, to promote data security sharing and full utilization, and to enhance the value of data resources. **Method/Process** The literature review method is adopted in this study, followed by multiple rounds of thematic expert group discussions to discuss the principles, rules and implementation steps of health and medical science data grading. **Result/Conclusion** Based on the requirements of laws, regulations, or policies related to health and medical science data, this study constructs a guideline for data classification of health and medical science data. It proposes specific principles, rating rules, and implementation steps for health and medical science data grading, providing references for data classification in the production, management and application of health and medical science data.

〔**Keywords**〕 data grading; health and medical data; scientific data; data security

1 引言

近年来我国逐步加强科学数据管理和开放共享。2018 年《科学数据管理办法》提出要加强和规范科学数据管理，保障科学数据安全，提高开放共享水平[1]。目前健康医疗科学数据尚无统一定义。

〔修回日期〕 2023－03－16

〔作者简介〕 张胜发，博士，发表论文 26 篇；通信作者：周伟，高级工程师。

〔基金项目〕 国家科技基础条件平台中心委托任务课题（项目编号：2022WT07）。

参考《科学数据管理办法》对科学数据的定义，健康医疗科学数据是指在健康医疗领域，通过基础研究、应用研究、试验开发等产生的数据，以及通过观测监测、考察调查、检验检测等方式取得并用于科学研究活动的原始数据及其衍生数据”[2]。健康医疗科学数据范围非常广，不仅包括基础医学、临床诊疗、公共卫生、药物研发、健康管理等数据，还包括与生命健康相关的基因组学、微生物学等数据[3]。我国健康医疗领域科学数据规模骤增，年数据总量已接近全球的三分之一，增速居世界之首，成为世界最大健康医疗科学数据生产国[4]。在我国大力构建数据安全与个人信息保护体系的总体趋势下，完善健康医疗科学数据分级保护制度不仅是健康医疗科学数据使用管理和安全防护的基础和核心，也是促进数据安全共享、充分利用、提升数据价值的前提条件。近年来我国相继出台多部健康医疗科学数据安全相关法律法规，要求建立数据分类分级保护制度，如《中华人民共和国数据安全法》《科学数据管理办法》。然而，当前我国数据分级保护工作和研究还处于探索阶段，尚未出台适用于健康医疗科学数据方面的分级指南或标准。

2 国外健康医疗科学数据分级指南研究现状

从世界范围来看，各国对数据的分级保护都是基于其法律法规或政策对数据安全的要求开展的。然而，健康医疗科学数据范围广、类型复杂，包括个人层面的个人信息、临床数据、组学数据等，机构层面的医疗卫生服务、运营管理等数据，以及国家层面的人口基础数据、统计数据等，国内外尚未形成专门针对健康医疗科学数据安全管理的法律法规。目前健康医疗科学数据分级管理或研究基本围绕个人信息安全、生物信息安全、数据安全、国家安全等方面相关法律法规要求开展，个人信息、个人敏感信息、重要数据、人类遗传资源信息等都是健康医疗科学数据分级管理的基础数据要素。

2.1 欧盟

欧盟采用统一立法模式对数据治理过程中的数据收集、存储、应用、流转等各领域进行规范，为数据分级分类提供法律基础。如欧盟在个人数据隐私保护领域发布的《通用数据保护条例》（General Data Protection Regulation，GDPR）对重要、敏感数据等做出特别规定[5]。其第 2 章第 9 条提到“对揭示种族或民族出身、政治观点、宗教或哲学信仰、工会成员的个人数据，以及以唯一识别自然人身份为目的的基因数据、生物特征数据，健康、自然人的性生活或性取向的数据处理应当被禁止”[6]。基于 GDPR 等相关法律法规要求，欧洲数据保护委员会发布《数据保护影响评估指南》等系列指南，提出从影响层面、风险发生可能性等维度对数据进行安全分级，为数据分级具体研究或实践提供较系统的指导[7]。欧盟目前尚未形成系统的健康医疗科学数据分级管理规范和实践。欧盟数据分级相关规定主要包括《通用数据保护条例》等，均为普遍和通用性的数据分级保护相关标准，并未基于健康医疗科学数据特点开展分级管理实践。从研究情况来看，有学者[8]从医学伦理、患者隐私等角度提出数据分级应遵循的原则、安全要求等，并对数据分级保护与科学数据共享的关系进行论述，但并未发表具体可行的分级管理研究成果。

2.2 美国

美国对健康医疗科学数据的分级保护工作主要从个人隐私保护、个人健康医疗数据保护、国家信息安全等方面展开。20 世纪 90 年代末，美国发布《健康保险携带和责任法案》（Health Insurance Portability and Accountability Act，HIPAA），明确涉及个人健康数据处理、传输、公开时要遵守的具体规定[9]。21 世纪以来，《个人可识别健康信息的隐私标准》确保公民的个人医疗健康信息受到合理保护。其后《经济和临床健康信息科技法案》对 HIPAA 再次进行扩展补充，形成相对完善的个人健康医疗数据保护制度体系[10-11]。同时，在信息安全方面出台分级管理要求相关规定，如 2010 年《受控非密信息》要求建立和实施受控非密信息管理登记备案及标识管理制度，重点关注信息的保密性、完整性、可用性，主要特点是统一分类、精细

管理[12-13]。同时，美国在科学数据共享、数据管理以及数据安全方面发布政策声明，要求对科学数据实行以“完全与开放”为原则的开放共享政策[14]。

在健康医疗数据安全相关法规或政策要求下，美国学者对数据分级政策、分级标准、分级方法等方面开展研究或实践。美国国家标准与技术研究院发布的《数据分类分级实践：促进以数据为中心的安全管理》提出应从标准、技术、流程等方面开展数据分级实践[15]。部分数据管理企业对数据分级政策、标准、方法和流程进行探索和实践，如 Satori 数据安全平台根据数据敏感性将其分为公开数据、内部数据、机密数据和限制数据[16]。美国部分高校或科研机构对其拥有的健康医疗相关科学数据进行分级划分。如加州大学伯克利分校提出数据应按照 HIPAA 等法规或制度要求，结合数据字段内容、是否影响组织或个人、是否包含人类基因组信息等进行分级评定[17]。

3 我国健康医疗科学数据分级指南研究现状

3.1 健康医疗科学数据分级的法律依据

目前我国没有专门针对健康医疗科学数据安全的法律法规，主要遵循个人信息安全、数据安全、网络安全等方面法律法规规定。个人信息安全方面，《中华人民共和国刑法》《中华人民共和国个人信息保护法》明确个人信息的范围和安全保护要求。数据或信息安全方面，《中华人民共和国网络安全法》《中华人民共和国数据安全法》《网络数据安全管理条例（征求意见稿）》等对我国网络数据安全与个人信息保护进行细化和补充。2018 年《国家健康医疗大数据标准、安全和服务管理办法（试行）》明确保障健康医疗大数据安全的要求[18]。2019 年《中华人民共和国人类遗传资源管理条例》明确人类遗传资源信息安全管理相关要求[19]。2021 年施行的《中华人民共和国生物安全法》从维护国家生物安全的角度提出应当根据风险监测的数据、资料等信息，定期组织开展生物安全风险调查评估，并对生物安全信息的共享、发布等内容提出明确要求[20]。

3.2 已发布的数据分级指南

根据已有法律法规相关要求，目前我国在数据分级方面已发布一些指南或标准，涉及网络安全、政务数据、网络数据等。2017 年《信息安全技术 网络安全等级保护定级指南》提出网络安全等级保护应按照受侵害客体和受侵害程度分为 5 个等级，实施网络信息安全分级管理。2021 年《网络安全标准实践指南——网络数据分类分级指引》提出考虑影响对象、影响程度两个要素对网络数据进行分级。针对政务数据分级分类已发布许多地方标准或指南，如浙江省《数字化改革 政务数据分类分级》提出按照敏感程度和影响程度进行分级。从行业来看，工业、金融、电信等行业已开展数据分级分类探索，发布《证券期货业数据分类分级指引》《工业数据分类分级指南（试行）》等相关标准或政策。随着 2018 年《国家健康医疗大数据标准、安全和服务管理办法（试行）》的发布，健康医疗大数据分级相关研究陆续开展，《信息安全技术 健康医疗数据安全指南》提出根据数据重要程度、风险级别以及对个人健康医疗数据主体可能造成的损害和影响级别进行分级。部分地区开展健康医疗大数据分级相关规范或指南编制，如 2020 年《四川省健康医疗大数据共享应用指南》根据数据重要程度和风险级别对健康医疗数据进行分级。在科学数据方面，目前仅发布一项国家标准，即 2011 年《机械科学数据 第 1 部分：分级分类方法》，其提出 7 级数据分级方法，但是该分级方法的发布早于《中华人民共和国数据安全法》等法律法规，其分级方法未能体现我国现有法律规范对科学数据安全管理的要求。

3.3 国内健康医疗科学数据安全分级指南相关研究

目前数据分级研究主要聚焦于政府数据、行业数据、工业、金融等方面。张峰等[21]对数据安全分类分级发展情况进行介绍，最后提出数据安全分类分级推进思路及发展建议。王真[22]对我国数据分级分类制度现状进行描述，阐述政府、金融、证券、工业、网络及健康医疗等领域数据分级分类的探索

与实践情况。侯利阳等[23]概述我国法律法规对数据分级保护的规定，提出建立统一强制标识制度、统一登记备案系统和执行机构以及统一分级制度的建议。陈兴跃[24-25]则认为数据分级分类管理是实施数据全生命周期安全保护的重要基础，提出数据分级分类工作的主要实施步骤。金涛等[26-27]结合国内数据安全相关标准和实践要求，提出根据受侵害对象及其受侵害程度进行数据分级的方法。

总体来看，我国当前数据分级分类保护工作和研究尚处于探索阶段，各项法律法规中对数据分级分类的有关规定不具体、不完善，政府相关部门、科研机构和科学数据管理机构无法根据相关法律法规开展健康医疗科学数据具体工作。健康医疗科学数据相关技术规范或标准等仅是概括性地提出数据分级分类保护基本原则和基本要求，并没有出台具有可操作性的分级分类保护规范或标准。已有分级规范或指南编制目的及侧重点差异较大，缺乏统一性，且无法满足当前健康医疗科学数据管理需求。

4 健康医疗科学数据分级指南构建

基于文献研究发现，数据分级指南主要包括数据分级原则、分级规则、分级实施等主要内容，笔者对健康医疗科学数据分级的原则、规则、实施步骤等内容进行文献研究，结合健康医疗科学数据安全管理特点和要求，提出适用于健康医疗科学数据的分级原则、规则和实施步骤。

4.1 健康医疗科学数据分级原则

基于对已有数据分级指南和学术研究成果中提及数据分级原则的梳理，结合健康医疗科学数据特点，提出健康医疗科学数据的分级原则，主要包括以下 5 方面。一是依法依规原则。数据分级必须符合法律法规、相关部门规定要求，尽可能遵循相关标准、指南、规范的相关规定。二是多维度评级原则。数据分级根据数据泄露或公开后对国家安全、公共利益、个人或组织合法权益的危害程度，综合数据隐私性、保密性等维度。三是就高从严原则。对包含多个维度的数据分级时，如维度级别不同，采用就高不就低原则。四是可操作原则。数据分级应符合数据安全管理要求，不应过于复杂或过于粗犷，各级别界限明确。五是时效性原则。数据分级具有一定时效性，数据分级结果符合评级时间的安全要求，时间变化、法律法规与政策要求变化等因素可能导致原有数据分级不再适用，需要对数据进行重新评级。

4.2 健康医疗科学数据分级规则

健康医疗科学数据分级应考虑将影响对象和影响程度作为分级规则的制定依据，结合相关法律法规要求确定分级规则。按照《中华人民共和国网络安全法》《中华人民共和国数据安全法》《中华人民共和国个人信息保护法》等要求，提出应根据数据遭到篡改、破坏、泄露或者非法获取、非法利用，对国家安全、公共利益、个人或组织合法权益造成的危害程度，将健康医疗科学数据划分为 5 个等级，见表 1。

表 1 健康医疗科学数据分级规则

级别	危害程度
第 1 级	对国家安全、公共利益、个人或组织合法权益无危害
第 2 级	对个人、组织合法权益造成轻微危害，对国家安全、公共利益无危害
第 3 级	对个人、组织合法权益造成一般危害，对公共利益造成一般危害，对国家安全无危害
第 4 级	对个人、组织合法权益或公共利益造成严重危害，或者对国家安全造成一般危害
第 5 级	对国家安全造成严重危害

4.3 健康医疗科学数据分级实施步骤

根据已有研究成果，结合专家小组讨论，提出健康医疗科学数据分级工作的 6 个步骤。

第 1 步：数据分类。健康医疗科学数据应按照不同类型涉及的数据安全相关法律法规、指南标准、政策或行业要求等规定进行分类。如涉及个人，应依据《中华人民共和国个人信息保护法》等进行分类，可分为个人数据和匿名化数据；如涉及国家安全或社会利益，应依据《中华人民共和国数

据安全法》《信息安全技术 重要数据识别指南》等进行分类，可以划分为核心数据、重要数据和一般数据；如涉及人类遗传资源信息，则应依据《中华人民共和国人类遗传资源管理条例》进行分类，可以划分为人类遗传资源信息和非人类遗传资源信息。

第 2 步：明确分级对象。数据分类后，在进行数据分级时，首要考虑的问题是明确分级对象颗粒度。从健康医疗科学数据资产分类来看，可以对数据文件定级，也可以对数据文件包含的数据内容（字段、记录等）进行定级，还可以对两者同时定级，因此，需要设置合理的分级颗粒度，既达到差异化保护的目的，又不影响具体分级实践。例如，一个包含生理健康信息的数据文件，如果其研究对象为国家一级运动员，应根据国家对运动员生理参数信息管理要求对该数据进行重点保护，在数据内容（字段、记录等）评价基础上，在数据文件层面根据研究对象特征进行数据分级评价。

第 3 步：识别数据定级影响因素。在明确分级对象后，应依据数据对应类别，分析其是否符合各类数据相关法律法规要求，是否符合各类数据相关指南或标准，是否符合各类数据相关政策要求、行业规定（如个人隐私、医学伦理、人类遗传资源信息、公共卫生信息、生理健康信息等），并对法律法规、政策、指南规范等未覆盖的数据情形进行综合分析，识别数据定级影响因素。如《中国人类遗传资源采集行政审批许可事项服务指南》中规定“重要遗传家系是指患有遗传性疾病或具有遗传性特殊体质或生理特征的有血缘关系的群体，患病家系或具有遗传性特殊体质或生理特征成员五人以上，涉及三代”，则相关数据应从患有遗传性疾病、患病家系人数、患病家系代际数等方面进行分级评级。

第 4 步：确定分级标准。在数据分类的基础上，根据涉及不同类型数据的有关法律法规或政策要求不同，确定数据分级标准。如包含个人信息的数据，需要区分该信息是直接包含个人标识信息，还是可重新识别个人标识信息，重要数据则应综合评估其对国家或社会的安全影响再进行分级。对已有具体政策或法律法规明确规定的数据（如人类遗传资源信息），则应根据相关法律法规对有明确规定的特定数据分级，并依据相关规定进行重点保护。同时，需要考虑所有数据的数量、形式等可能对数据分级产生的影响。

第 5 步：制定评价工具。在识别定级因素和分级标准的基础上，应采用定性或定量方法进行判定，定性判定需要明确判断依据、原则、标准等内容，定量判定应明确判定依据、标准、方法等。数据分级遵循就高从严原则，如果存在多维度数据且各维度级别不同、数据集中的多个数据级别不同等情况，数据分级应取各维度评级中最高的级别作为该数据的分级级别。

第 6 步：分级变更。当健康医疗科学数据安全要求、数据内容等发生变化时，需要及时进行分级变更。数据等级变更后，原有数据级别及其针对性的安全管理方式不再适用。

5 结语

在我国大力构建数据安全与个人信息保护体系的总体趋势下，完善健康医疗科学数据分级保护，不仅是数据使用管理和安全防护的基础和核心，也是促进数据安全共享、充分利用、提升数据资源价值的前提条件。本研究按照《中华人民共和国数据安全法》《中华人民共和国个人信息保护法》等法律法规要求，结合已有数据分级相关研究成果，提出健康医疗科学数据分级应遵循依法依规、多维度评级、就高从严等原则，按照影响对象、影响程度等维度进行评级，其实施包括数据分类、明确分级对象、识别数据定级影响因素等步骤，可以为政府相关部门、科研机构和科研人员开展健康医疗科学数据安全管理提供理论指导，为健康医疗科学数据生产、管理及应用活动中的数据分级提供参考。

参考文献

1 张贵兰，王健，潘尧，等．科学数据共享服务模式及其演化研究［J］．情报理论与实践，2022，45（2）：70－77.

2 国务院办公厅．关于印发科学数据管理办法的通知［J］．中华人民共和国国务院公报，2018（11）：10－13.

3 杨朝晖，王心，徐香兰．医疗健康大数据分类及问题探讨［J］．卫生经济研究，2019，36（3）：29－31.
4 吴思竹，王安然，修晓蕾，等．欧美生物医学科学数据中心建设及启示［J］．数字图书馆论坛，2022（4）：2－10.
5 盛小平，杨绍彬．GDPR 对科学数据开放共享个人数据保护的适用性与作用分析［J］．图书情报工作，2020，64（22）：48－57.
6 STAUNTON C，SLOKENBERGA S，PARZIALE A，et al. Appropriate safeguards and article 89 of the GDPR：considerations for biobank，databank and genetic research［J］. Frontiers in genetics，2022，13（2）：719317.
7 肖君拥，张雪亭．欧盟数据保护影响评估制度及其镜鉴［J］．电子科技大学学报（社科版），2022，24（5）：18－29.
8 KALKMAN S，MOSTERT M，GERLINGER C，et al. Responsible data sharing in international health research：a systematic review of principles and norms［J］. BMC medical ethics，2019，20（1）：21.
9 Glenn T，Monteith S. Privacy in the digital world：medical and health data outside of HIPAA protections［J］. Current psychiatry reports，2014，16（11）：494.
10 KADAKIA K T，HOWELL M D，DESALVO K B. Modernizing public health data systems：lessons from the Health Information Technology for Economic and Clinical Health（HITECH）Act［J］. Journal of American medical association，2021，326（5）：385－386.
11 MENNEMEYER S T，MENACHEMI N，RAHURKAR S，et al. Impact of the HITECH act on physicians' adoption of electronic health records［J］. Journal of the American medical informatics association，2016，23（2）：375－379.
12 周亚超，左晓栋．美国受控非密信息分类与安全控制解析［J］．网络空间安全，2020，11（3）：12－17.
13 魏波，周荣增．美国信息安全立法及其启示与分析［J］．网络空间安全，2019，10（5）：1－6.
14 储节旺，汪敏．美国科学数据开放共享策略及对我国的启示［J］．情报理论与实践，2019，42（8）：153－158.
15 SCARFONE K，SOUPPAYA M. Data classification practices：facilitating data－centric security management［EB/OL］.［2022－07－27］. https：//csrc. nist. gov/publications/detail/white－paper/2021/05/19/data－classification－practices－data－centric－security－management/draft.
16 SATORI. Guide：data classification［EB/OL］.［2022－12－31］. https：//satoricyber. com/data－classification/data－classification－types－criteria－levels－methods－and－more/.
17 Office Berkeley Information Security. Data classification guideline［EB/OL］.［2023－03－07］. https：//security. berkeley. edu/data－classification－guideline.
18 国家卫生健康委员会．关于印发国家健康医疗大数据标准、安全和服务管理办法（试行）的通知［J］．中华人民共和国国家卫生健康委员会公报，2018（7）：7－11.
19 中华人民共和国人类遗传资源管理条例［J］．中华人民共和国国务院公报，2019（18）：29－35.
20 中华人民共和国生物安全法［J］．中华人民共和国全国人民代表大会常务委员会公报，2020（5）：733－743.
21 张峰，于乐，马禹昇，等．数据安全分类分级研究与实践［J］．信息通信技术与政策，2021，47（8）：45－50.
22 王真．数据分级分类研究［D］．北京：北京外国语大学，2021.
23 侯利阳，贺斯迈．如何对数据进行分级分类保护［J］．检察风云，2020（19）：14－15.
24 陈兴跃．数据分级分类正式入法具有重大实践指导意义［J］．信息安全研究，2020，6（10）：949－952.
25 陈兴跃．《中华人民共和国数据安全法（草案）》公开征求意见：数据分级分类正式入法［J］．中国信息化，2020（7）：9－10.
26 金涛，王建民．GB/T 39725—2020《信息安全技术健康医疗数据安全指南》［J］．标准生活，2022（3）：46－51.
27 金涛．数据安全分级划分［J］．信息安全研究，2021，7（10）：969－972.