



美国数据库禁止中国学者访问？

张建军

首都医科大学学报编辑部，北京 100069

摘要

2025 年 4 月 4 日起，美国国家卫生研究院（NIH）禁止中国、俄罗斯、伊朗等国家的学者访问 SEER、db-GaP、TCGA 和 GDC 等关键生物医学数据库。这一政策对中国医学研究，尤其是癌症、基因组学和精准医学领域产生了极大影响。本文基于 PubMed 数据，分析了中国学者利用以上 4 个数据库的发文趋势，探讨了访问限制对中国科研尤其是医学研究的潜在影响，并提出了应对策略。笔者认为，尽管短期内中国医学研究可能面临数据获取障碍，但长期来看，这一挑战可能推动中国科研自主化进程，并促进更广泛的国际科技合作。

关键词：数据库访问限制；中国医学研究；科研政策；数据自主化；国际合作。

引言

近年来，中国在全球科研产出中的贡献显著增长，中国的总体科研产出正在慢慢超过美国，这一领先优势在医学领域尤为显著^[1]。在医学研究中，医学数据库是医学研究的核心支撑，其整合了临床、基因组和影像等数据，促进了全球协作与资源共享；同时，医学数据库可提升研究效率，支持大数据分析和 AI 挖掘，推动精准医学和个性化治疗的发展；此外，医学数据库可辅助临床决策（如循证医学、疫情监测）、加速药物研发（靶点发现、不良反应追踪），并助力医学教育。总之，医学数据库是现代医学研究的重要工具，从基础科研到临床转化各环节均依赖其支撑。然而，2025 年 4 月 2 日美国国立卫生研究院（National Institutes of Health, NIH）发布了一

条题为“Implementation Update: Enhancing Security Measures for NIH Controlled-Access Data”的声明（Notice Number: NOT-OD-25-083）^[2]，该声明核心内容为“遵照之前美政府发布的 EO 14117 和 28 CFR Part 202 两文件关于‘防止相关国家或相关人员访问美国敏感个人数据和政府相关数据’的要求，从 2025 年 4 月 4 日起，NIH 禁止相关国家的机构访问 NIH 的受控访问数据库和相关数据。这些国家包括中国（包括香港和澳门）、俄罗斯、伊朗、朝鲜、古巴和委内瑞拉”。这条限制声明直接影响了中国医学研究从业人员对 SEER（The Surveillance, Epidemiology, and End Results：美国国立癌症研究所建立的癌症数据库）、dbGaP（The Database of Genotypes and Phenotypes：美国国家生物技术信息中心建立的用于归档、精选和发布由调查基因型和表型间相互作用的研究所产生的信息的数据仓库）、TCGA（The Cancer Genome Atlas：美国于 2005 年发起，旨在通

DOI: 10.14218/MRP.2025.06031

通讯作者：张建军 Email: syxuebao@ccmu.edu.cn

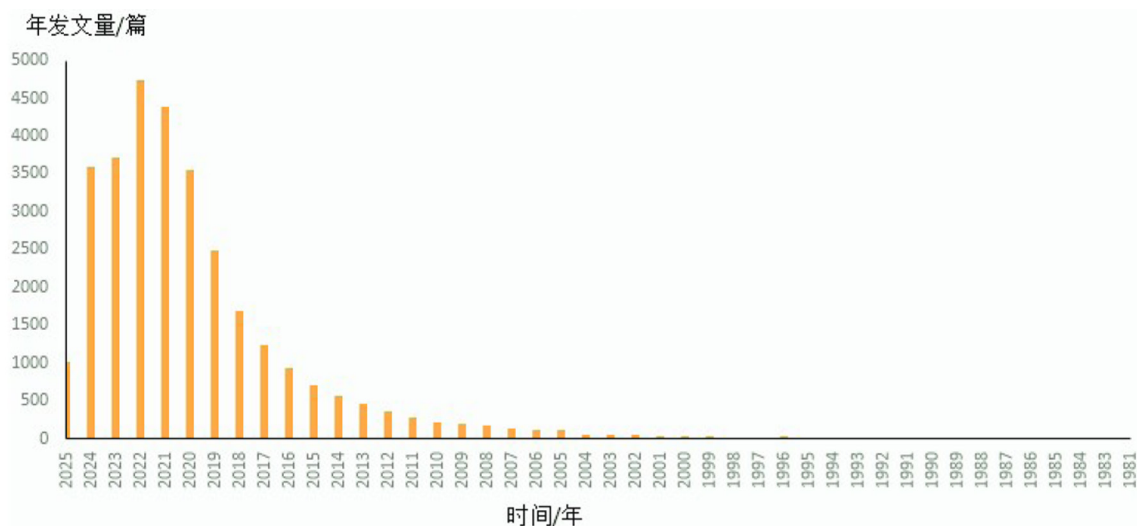


图1 全球学者利用SEER、dbGaP、TCGA、GDC数据库年发文量。

过基因组分析技术研究癌症中的基因组变化)和GDC (Genomic Data Commons: 一个癌症知识网络, 支持来自癌症研究项目的基因组、临床和生物样本数据的托管、标准化和分析)等关键生物医学数据资源的获取, 而这些数据库在癌症研究、基因组学和临床医学中具有不可替代的作用。本文结合中国学者利用这4个医学数据库的发文情况, 分析这一政策对中国医学研究的影响, 并提出可行的应对策略, 以促进中国科学界特别是医学界的可持续发展。

数据与方法

本文数据来源于PubMed, 利用其高级搜索功能调查全球学者利用上文提及的四大数据库(SEER、dbGaP、TCGA、GDC)的发文量, 对比全球与中国在四大数据库中的研究贡献, 分析数据访问受限的影响。在PubMed进行搜索的时间为2025年4月10日。另外, 作为补充数据, 国内期刊利用这4个数据库的发文情况使用万方数据平台检索。

结果

在PubMed上以“China [Affiliation]”为搜索条

件, 结果显示截至2025年4月10日, 中国学者从1858年至今在医学领域共发文3,722,690篇, 近百年年均发文量约39,603篇。1949年中华人民共和国成立至今76年, 中国学者共发文3,772,584篇, 年均发文量约为49,369篇。以年发文量为指标, 中国医学领域发文趋势大概可分为5个阶段: 第一阶段为1986年之前, 年发文量不超过百篇; 第二阶段为1987–1989这3年, 年发文量都上了百篇; 第三阶段为1990–2000年这10年, 年发文量都以千篇计; 第四阶段为2001–2012年, 年发文量都以万篇计; 第五阶段为2013–至今, 年发文量以10万篇计, 2024年年发文量达到432,677篇。可以看出, 自1990年开始, 大概以10年为一个阶段, 中国在医学领域发文以10倍速度增长。

以“SEER(The Surveillance, Epidemiology, and End Results)[Text Word]”OR “dbGaP(The Database of Genotypes and Phenotypes) [Text Word]” OR “TCGA(The Cancer Genome Atlas) [Text Word]” OR “GDC(Genomic Data Commons)[Text Word]”为搜索条件, PubMed显示自1981年第1篇发文至今, 全球利用这4个数据库(之一或以上)发文量共计31,079篇, 43年来年均发文量约为706篇, 2022年发文最多为4745篇(图1)。在以上

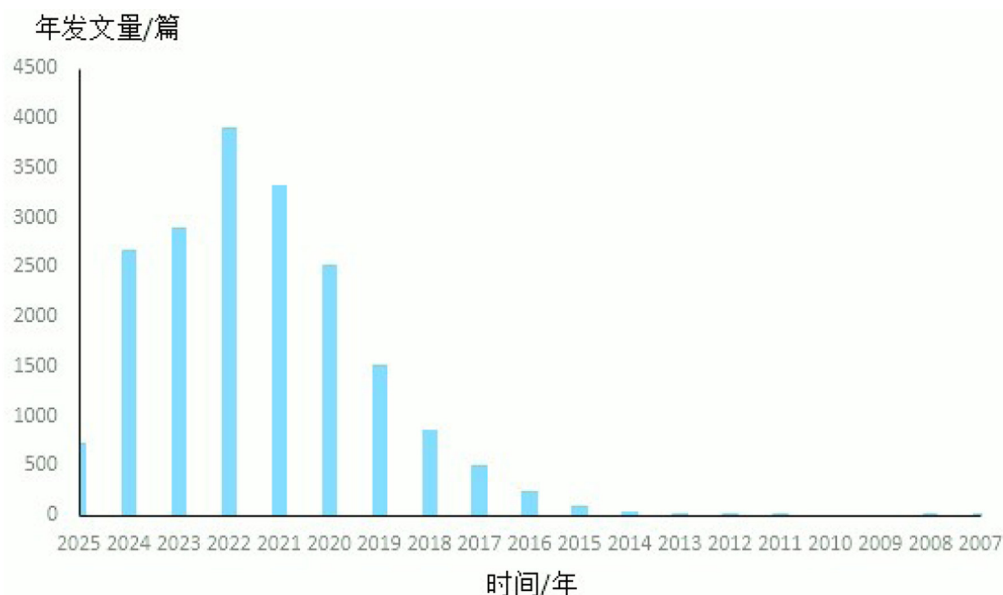


图2 中国学者利用SEER、dbGaP、TCGA、GDC数据库年发文量。

搜索条件基础上合并“China [Affiliation]”，结果显示中国利用这4个数据库的第1篇发文出现在2007年，至今共计19,358篇（图2），占全球发文量的62.3%（19358/31079），近20年年均发文量为1019篇，发文最多为同为2022年，共计3898篇；如果以2007年为起始年计算，则占全球发文量的63.6%（19358/30442），占中国总发文量的0.5%（19358/3556719）。

另外，据万方数据平台显示的中国学者利用这4个数据库在国内期刊的发文情况，从2016年发文2篇至今，共计65篇相关文献，年均发文约7篇，因数据量过小，不做进一步分析。

讨论

从结果可以看出，自2007年以来，尤其是2010年之后，中国学者利用SEER、dbGaP、TCGA、GDC数据库年发文量逐年快速增长。虽然占中国自2007年医学领域总发文量的比重（0.5%）不算太高，但和全球学者利用这4个医学数据库发文量相比，其占比非常高（63.6%），同为2022年发文最高峰，中国学者的发文量占比高达82.1%（3898/4745）。

另外，从图1和图2的发文趋势也可以看出，全球学者利用这4个数据库发文量变化趋势和中国学者的发文量变化趋势非常拟合。说明SEER、dbGaP、TCGA、GDC这4个医学数据库对中国学者的重要性，也说明中国学者利用这4个医学数据库为医学研究做出了大量科研产出。

因此，NIH的这条限制政策无疑会对中国学者的医学研究造成比较严重的影响：首先，中国学者对这4个数据库的访问受限，会造成中国学者在医学研究上，尤其是癌症和基因组学研究上关键数据缺失，阻碍了中国学者的相关研究进展；其次，虽然限制政策提出如果需要和中国机构进行国际合作，可以提交申请，但相关程序的不透明和美国政府的刻意打压，大概率中国学者会被排除在基于这些数据库的国际合作项目之外，中国学者的国际合作愈发困难；最后，如果这些政策长期延续执行，中国又没有找到应对之策，可能会导致中国在精准医学、生物信息学等领域落后，导致中国科研，尤其是医学领域的竞争力下降。

作为对这条限制政策的应对策略，本文建议：（1）利用中国“集中力量办大事”的优势，尽快建立中国本土数据库，尤其要加快中国自主医学数据库

的建设。可喜的是,近10年来,中国对科研数据工作越来越重视,在2018年出台了国家层面的《科学数据管理办法》^[3],之后又出台了《中华人民共和国人类遗传资源管理条例》^[4]、《中华人民共和国生物安全法》^[5]和《中华人民共和国个人信息保护法》^[6],用于专业数据的管理和保护。并建成了以20个国家科学数据中心和31个国家资源库为重要基础的科技资源共享平台,如国家基因库生命大数据平台(China National GeneBank DataBase, CNGBdb)、国家基因组科学数据中心(National Genomics Data Center, NGDC)等。但这些数据库在数据规模、标准化程度与国际领先数据库仍有差距^[7]。(2)积极主动拓展国际合作,与欧洲、亚洲等非美国主导的科研机构合作,减少对美数据的依赖。(3)适时调整研究方向,聚焦临床医学、中医药等受数据限制较小的领域。中国具有世界上最大的临床规模,中医药又是中国传统文化的瑰宝,聚焦这两方面研究,假以时日,必将建成规模最大最全的临床数据库和中医药数据库。(4)中国学者应积极主动呼吁国际科学界应倡导数据共享的开放性原则,避免科学合作过度政治化。

小结

美国对四大数据库的访问限制在短期内确实会对中国医学研究构成挑战,但长期来看,可能会在推动中国科研自主化进程中起到积极作用。通过建立本土数据库、加强国际合作、优化科研资源配置,中国科学界可以有效降低外部政策风险,并且继续保持自身的全球科研竞争力。

参考文献

- [1] 中国在学术研究上的巨大进步被认为缩小了美国在医学上的领先优势. 南华早报, 2024-09-22.
- [2] Implementation Update: Enhancing Security Measures for NIH Controlled-Access Data. <https://grants.nih.gov/grants/guide/notice-files/NOT-OD-25-083.html>.
- [3] 科学数据管理办法. https://www.most.gov.cn/xxgk/xinxifelnlei/fdzdgknr/fgzc/gfxwj/gfxwj2018/201804/t20180404_139023.html.
- [4] 中华人民共和国人类遗传资源管理条例. https://www.gov.cn/zhengce/zhengceku/2019-06/10/content_5398829.htm.
- [5] 中华人民共和国生物安全法. https://www.cae.cn/cae/html/main/col2/2023-04/06/20230406145037595807725_1.html.
- [6] 中华人民共和国个人信息保护法. http://www.npc.gov.cn/c2/c30834/202108/t20210820_313088.html.
- [7] 许哲平NIH数据访问限制政策对我国科学数据工作的影响研究. <https://mp.weixin.qq.com/s/iB3Lf9hnNof5mIs-9JZXPcA>.