

收录 12 万余篇案例—— 这个数据库成为临床医生评价新“标尺”

● 本报记者 高雅丽

2022 年一个寻常的工作日，西安医学院第一附属医院消化内科副主任医师王佳收到了一封邮件。邮件的附件是一张盖有中华医学会红色印章的电子证书，她撰写的一篇关于“内脏反转”罕见病例的报告被中国临床案例成果数据库正式收录。她当时并不知道，这张证书将在 2024 年成为她职称晋升材料中“最具分量的部分”。

从业多年，王佳和无数基层医生一样，曾陷入“为发论文而搞科研”的困境。中国临床案例成果数据库的出现，让评价回归临床，让真才实学取代“帽子”，打开了她的职业上升通道。

近日，中国科协举行的中国临床案例成果数据库通气会公布的数据显示，截至 2025 年底，中国临床案例成果数据库已收录来自全国 5000 多家医疗机构的规范化病例报告 12.2 万篇，总阅读量突破 3100 万人次，覆盖 80 多个国家和地区的研究者。

“数论文”不再是唯一答案

“医生的成长没有捷径，是在一个具体病例的诊治中摸爬滚打出来的。评价一位医生，最重要的是看他‘会不会看病、能不能治好病’。”中国临床案例成果数据库学术委员会主任委员、空

军军医大学西京医院教授韩英的这句话，道出了无数临床医生的心声。

长期以来，中国医疗系统形成了以科研论文为核心的职称评价体系。韩英表示，“论文”这把“标尺”不能准确丈量一些擅长临床却不擅写文章的医生的真实水平。

2021 年，人力资源社会保障部等部委联合发布的《关于深化卫生专业技术人员职称制度改革的指导意见》提出，要突出实践能力业绩导向，鼓励卫生专业技术人员扎根防病治病一线，实行成果代表作制度，临床病案、手术视频等均可作为业绩成果代表作。

政策已经指明方向，如何落地？早在 2019 年，中国科协委托中华医学会，启动建设中国临床案例成果数据库，把体现医生真才实学的病例和经验转化为可被规范记录、科学评价、广泛共享的学术成果。

中华医学会杂志社社长魏均民指出，每一篇案例报告都需遵循病例报告规范，并经过“三审三校”同行评议流程，“由 3000 余名各专业评审专家来确保每一篇入库案例都经得起检验”。

陕西省是全国首批中国临床案例成果数据库共建省份。陕西省卫生行业学会服务中心主任张磊举例介绍，在

市级医院正高级资格评审中，2 份入选案例库病案等同于 1 篇 SCI(科学引文索引)论文或 2 篇专业论文，此举将逐步淡化论文在评审中的比重。

在 2025 年陕西省卫生高级职称评审初审中，已有 117 名卫生专业技术人员凭借优质临床案例申报高级职称。作为试点的空军军医大学，使用临床案例成果参评人数占总人数的 24%，通过率 27%。

王佳坦言，在改革前，临床医生普遍存在科研焦虑和畏难情绪——既怕做不好基础科研，又怕处理疑难杂症。而现在，“大家遇到复杂病例时，第一反应是‘这个值得深挖，可以写成一篇好案例’”。

“现在，一篇诊疗规范、逻辑清晰、具有借鉴意义的高质量病例报告，同样能作为代表其临床能力的‘硬通货’，在职称晋升中获得认可。这让医生能心无旁骛地回归临床、聚焦患者。”韩英说。

帮助更多医生和患者

如果说评价改革是中国临床案例成果数据库的“牵引力”，那么临床实践的深度赋能则是持续发展的“内生动力”。

河北医科大学第二医院副院长潘

文森介绍，该院胆胰内镜外科的青年医师团队通过对大量内镜病例的深度总结与前沿技术探索，在国内外期刊连续发表案例研究成果。

基层医生面对疑难病例时，往往因经验不足难以精准判断。中国临床案例成果数据库中涵盖的常见病、多发病诊治经验，能帮助他们强化临床思维；而罕见病、复杂病例的分享，则让他们开阔眼界，提出更具有针对性的转诊建议。

王佳把临床工作比作“玩狼人杀”——初始线索寥寥，需要像侦探一样顺藤摸瓜，不断鉴别、排除，最终抓住疾病“真凶”。

她以被收录的一篇病例报告介绍，在整理一位患者历经两次淋巴结活检才确诊的病例报告时，她深入查阅文献发现该病若采用常用的细针穿刺，很难取得足够样本。“现在我通过中国临床案例成果数据库分享了该病例，或许能帮到其他医生和患者。”

这种赋能还延伸到国际舞台。通过标准化病例格式和双语内容建设，中国临床案例成果数据库已吸引 80 多个国家的研究者访问。2024 年以来，美国、澳大利亚、日本等国的研究者频繁查阅库中病例资源。中国医学界的实践智慧正进入全球知识治理，国际影响力持续提升。

医疗场景下大模型应用效果回顾性评测专家共识发布

本报讯 近日，《医疗场景下大模型应用效果回顾性评测专家共识(2025 版)》(以下简称共识)中英文版本在《数字医学与健康》与《智慧医学(英文)》发布。

据悉，该共识由国家新闻出版署医学期刊知识挖掘与服务重点实验室牵头，联合《数字医学与健康》编辑委员会、《智慧医学(英文)》编辑委员会、中华医学会杂志社指南与标准研究中心、医疗人工智能研究及应用安徽省重点实验室共同制定，汇聚了医学、人工智能、伦理、法学、统计学等多领域专家力量，全面构建了我国医疗大模型回顾性评测的系统化技术框架。共识

面向大语言模型在医疗场景实际落地前的评估需求，旨在为模型评测机构、研发机构和临床应用方提供统一、规范、可复用的技术指南，推动人工智能在医疗行业的高质量发展。

共识从评测流程、指标体系、团队建设、数据集设计、反馈更新机制和报告规范等方面构建了完整评测框架。其中，评测流程强调科学客观、真实全面与伦理合规；指标体系涵盖结构化与生成式任务的定量与定性评估；团队建设强调医学专家、工程技术人员、伦理与法律专家的协同；数据集构建突出临床真实性、全面代表性、公正性及动态扩展能

力；反馈与更新机制确保评测体系的长期迭代；报告模板则规范评测结果披露与版本管理。共识最终形成 6 条核心推荐意见，均获得超过 80% 的专家同意并达成共识。

值得关注的是，基于国家卫生健康委办公厅、国家中医药局综合司、国家疾控局综合司联合发布的《卫生健康行业人工智能应用场景参考指引》，共识明确提出医疗大模型评测的六大核心能力维度，包括医疗知识问答、医疗复杂语言理解、医疗诊断与治疗推荐、医疗专业文书生成、医疗多轮对话以及医疗多模态交互，从而为不同类型模型的评测

提供了清晰的应用场景定位和指标选择依据。同时，共识围绕患者隐私保护、数据脱敏、算法公平性、输出安全性等关键问题提出严格要求，有助于开辟医疗人工智能安全、可控、可解释的发展路径。

据悉，未来该工作将继续推动前瞻性研究、真实世界验证与动态评测机制的深化，持续完善评测标准体系，共同促进人工智能在医疗健康领域的安全应用和高质量发展。

(张思玮)

相关论文信息：<https://doi.org/10.3760/cma.j.cn101909-20250924-00177>

<https://doi.org/10.1016/j.jimed.2025.09.001>