



復旦大學 大健康信息研究中心
BIG HEALTH INFORMATICS RESEARCH CENTER
FUDAN UNIVERSITY

泌医智汇 信息快报

2025 年 05 月版

目 录

1 文献速递	1
1.1 射波刀治疗前列腺癌专家共识	1
1.2 老年前列腺癌临床诊治专家共识	4
1.3 Core GRADE：让证据评价更高效	8
1.4 Core GRADE 1：方法概述	11
1.5 Core GRADE 2：确信度分级目标选择与不精确性评估	14
1.6 Core GRADE 3：评估证据不一致性	17
1.7 Core GRADE 4：评估偏倚风险与发表偏倚	21
1.8 Core GRADE 5：评估间接性	23
1.9 四大医学顶刊同发：SPIRIT 2025 随机试验方案指南	26
1.10 CCR 最新：新辅助免疫联合 PARPi 具有治疗 MIBC 的潜力	29
1.11 Eur Urol 最新：机器人辅助单对接技术为 III 级 IVC 癌栓切除术带来新选择	32
1.12 JAMA 最新：混合性尿失禁治疗如何选？多中心 RCT 给出答案	35
1.13 Eur Urol 最新：肺癌筛查联合腹部 CT 可筛查肾癌	38
1.14 Eur Urol 最新：MRI 靶向活检或重塑前列腺癌诊断范式	41
1.15 Nat Med 最新：ctDNA 或可指导 LuPSMA 个体化治疗	45
2 期刊小知识	48
2.1 期刊介绍 European Journal of Surgical Oncology	48
2.2 期刊介绍 Annals of Surgical Oncology	50
2.3 期刊介绍 World Journal of Surgical Oncology	53
2.4 影响因子新变革：不再计入撤稿论文的被引用次数	55
2.5 Nature 独家：21 世纪被引用次数最多的论文	56
2.6 Nature 报道：数百篇论文未披露 AI 使用	60
2.7 Nature 发文：AI 生成的文献综述可能威胁科学进步	63

1 文献速递

1.1 射波刀治疗前列腺癌专家共识

标题：射波刀治疗前列腺癌专家共识

编制者：中华医学会放射肿瘤治疗学分会，中国核学会近距离与智慧放疗分会，上海医学会肿瘤放射治疗分会，北京市放射治疗质量控制和改进中心.

出处：中华转移性肿瘤杂志, 2025, 8(2): 150-157.

放疗是前列腺癌根治性治疗手段之一，与前列腺癌切除术有着相当的肿瘤控制率。由于其特殊的生物学特性（ $\alpha/\beta < 2 \text{ Gy}$ ），采用大分割放疗可提高肿瘤控制率。射波刀作为精准放疗的代表，利用金标追踪技术将治疗精度达到亚毫米级别，且 5 次左右的治疗频次对患者更为方便。本共识希望可以为前列腺癌的射波刀规范治疗提供建议，助力前列腺癌的安全有效治疗。

射波刀治疗前列腺癌的适应证与禁忌证

目前，射波刀在前列腺癌治疗中的应用主要在 2 个方面，即局限性前列腺癌的根治性放疗和前列腺癌寡转移灶的放疗。

（一）适应证

1.局限性前列腺癌，包括低危、中危、部分高危和极高危组患者（推荐级别 B，证据等级 2a）。

2.无法耐受手术或拒绝手术的局限性前列腺癌（推荐级别 B，证据等级 2a）。

3.有保留性功能需求的局限性前列腺癌（推荐级别 B，证据等级 2a）。

4.寡转移性前列腺癌转移灶的放疗，包括区域或非区域淋巴结、骨、内脏转移灶（推荐级别 B，证据等级 2a）。

5.低瘤负荷的转移性前列腺癌（M1）原发灶的放疗（推荐级别 B，证据等级 2a）。

（二）禁忌证

1.既往拟治疗部位的放疗史。

2.处于急性感染期或重度感染的患者。

3.病灶内或周围存在高密度金属植入物干扰靶区勾画和金标追踪。

4.需金标追踪者因解剖限制或出血倾向无法进行穿刺植入金标。

5.合并严重或未控的基础疾病，预期无法耐受。

6.精神病患者或依从性差或其他原因无法配合治疗。

治疗前准备

治疗前需要病情评估、多学科讨论、签署知情同意书、金标植入。

射波刀治疗前列腺癌需要有精确的影像引导技术来保证放疗计划的准确实施。射波刀的金标追踪方式通过曝光采集图像和图像对比来确定机械臂需要校准的数值和度数，进行实时校正。因此，金标植入是前列腺癌射波刀治疗中非常重要的一步。临床上常采用超声引导下经会阴前列腺穿刺行金标植入术。植入 4 颗金标是同时兼顾降低患者不适和增加定位精度的一个适宜选择。建议同一病灶植入的每 2 颗金标间距>2 cm，均匀分布在前列腺内；4 颗金标不能位于同一直线或同一平面上，需构成一个立体结构，且植入的金标不能在 45°正交 X 线获取正交二维影像有重合。

靶区勾画及处方剂量

前列腺癌射波刀治疗的处方剂量	
治疗部位	处方剂量
前列腺	9.5 Gy× 4次、7~8 Gy× 5次、6 Gy× 6次（低瘤负荷的M 1期）、6.1 Gy× 7次
转移灶	最大径<3 cm为18~20 Gy× 1次，最大径≥3 cm为12 Gy × 2次、8~10 Gy× 3次、6~8 Gy× 5次

前列腺癌立体定向放疗的计划要求和危及器官限量	
参数	靶区要求和危及器官限量
计划靶体积	$V_{36.25\text{ Gy}} \geq 95\%$
临床靶体积（前列腺）	$V_{40\text{ Gy}} \geq 95\%$
临床靶体积-计划靶体积外扩范围	5 mm，向后外扩3 mm
直肠	$V_{18.1\text{ Gy}} < 50\%$ ， $V_{29\text{ Gy}} < 20\%$ ， $V_{36\text{ Gy}} < 1\text{ cm}^3$
膀胱	$V_{18.1\text{ Gy}} < 40\%$ ， $V_{37\text{ Gy}} < 10\text{ cm}^3$ （最好 $V_{37\text{ Gy}} < 5\text{ cm}^3$ ）
前列腺尿道（如可见）	$V_{42\text{ Gy}} < 50\%$ （最好，非强制）
股骨头	$V_{14.5\text{ Gy}} < 5\%$
尿道球	$V_{29.5\text{ Gy}} < 50\%$
小肠	$V_{18.1\text{ Gy}} < 5\text{ cm}^3$ ， $V_{30\text{ Gy}} < 1\text{ cm}^3$

治疗实施

计划实施前应进行剂量验证，同时治疗时应确保摆位、固定方式与放疗定位时保持一致。采用射波刀自带的 45° 正交 X 线获取正交二维影像，同时依据计划设计时的匹配方式进行位置验证。当位置精度满足临床需求时方可实施治疗。

1.2 老年前列腺癌临床诊治专家共识

标题：老年前列腺癌临床诊治专家共识(2024 版)

编制者：逢城, 钟秋子, 朱刚, 刘明, 中华医学会老年医学分会.

出处：中华老年医学杂志, 2025, 44(2): 107-116.

前列腺癌主要发生于老年男性，其发病率随年龄增加而逐渐升高。我国前列腺癌平均发病年龄为 72.35 岁，患者主要集中在 65 岁以上。老年前列腺癌患者通常合并症较多，使得前列腺癌的治疗更加复杂。临床研究通常很少纳入高龄老年患者，因此老年前列腺癌群体的新型药物或治疗手段的循证医学证据有限。

目前我国暂时没有针对老年前列腺癌患者的诊疗指南。由国内泌尿外科、老年医学科、放疗科、肿瘤内科等领域专家组成的共识专家工作组，参考国内外有关规则、标准或规范，结合国际研究以及我国老年前列腺癌特点制订本共识，

为老年前列腺癌患者提供更多临床可行的诊疗选择。本共识参考世界卫生组织老年人定义,将年龄 ≥ 65 岁的患者定义为老年前列腺癌患者,并将其作为本共识的目标人群。

一般健康状况评估

共识意见 1: 前列腺癌患者治疗决策选择应基于患者的健康状况,而不是以实际年龄为指导。对于有老年医学专业团队合作的单位,推荐采用标准的老年综合评估流程评价健康状况以帮助治疗选择;老年综合评估的初始筛查推荐采用 G8 筛查工具,得分 ≤ 14 分建议进行进一步 CGA;对于缺乏老年医学专业团队合作的单位,建议使用 ECOG 评分评估患者健康状况。根据 CGA 结果可将老年前列腺癌患者分为 3 种类型:①体健;②体弱;③衰弱。其中体弱或衰弱型患者是老年干预的目标人群。

局限性前列腺癌的治疗

共识意见 2: 局限性前列腺癌推荐采用 CUA 指南中风险分组进行危险度划分以进行治疗决策。前列腺癌根治手术没有硬性的年龄上限,接受根治性前列腺切除术的中/低危局限性患者预期寿命应大于 10 年,接受根治性前列腺切除术的高危/局部进展性前列腺癌患者预期寿命应大于 5 年。

共识意见 3: 合并症较多的高龄患者,尤其伴有排尿障

碍者应谨慎选择放射治疗。放射治疗联合 ADT 是中高危前列腺癌患者的标准治疗，但是高龄患者放射治疗联合 ADT 的最佳 ADT 治疗时机和持续时间目前尚有争议。

共识意见 4：预期寿命>10 年的低危前列腺癌患者可考虑行主动监测。预期寿命<5~10 年且 CGA 分组为体弱或衰弱型的前列腺癌患者可考虑行观察等待。

共识意见 5：局灶治疗主要适用于低危患者及部分中危患者。局灶治疗的适应证选择、病灶定位、随访方案等还需要长期随访研究结果。

共识意见 6：不推荐将单纯 ADT 作为局限性前列腺癌患者的标准治疗方案。

晚期前列腺癌的治疗

共识意见 7：对于新诊断的老年 mCSPC 患者，ADT 是基础治疗。其中，新诊断的体健型老年 mCSPC 患者，可在 ADT 的基础上联合新型内分泌治疗药物或多西他赛作为一线治疗。体弱或衰弱型 mCSPC 患者，可在 ADT 的基础上联合新型内分泌治疗药物作为一线治疗，而体弱或衰弱型高转移负荷 mCSPC 患者建议健康状况改善后再考虑多西他赛治疗。另外，对于新诊断的低转移负荷 mCSPC 患者，在全身性药物治疗基础上建议联合前列腺局部放射治疗。

共识意见 8: (1)PSA 倍增时间 ≤ 10 个月同时 PSA >2 ng/ml 的 nmCRPC 患者建议阿帕他胺、恩扎卢胺或达罗他胺治疗; (2)体健型老年 mCRPC 患者可接受与年轻患者相似的治疗, 包括多西他赛化疗、阿比特龙、恩扎卢胺、阿帕他胺、奥拉帕尼和 Ra-223。选择治疗方法和确定治疗顺序时, 应考虑患者的整体临床状态; (3)体弱型的 mCRPC 患者在治疗顺序同等的前提下建议新型内分泌治疗药物优先于化疗药物; (4)体弱型的 mCRPC 患者如果身体状况可以纠正, 可以在状况改善后采用多西他赛标准化疗方案, 使用时建议从较低剂量开始治疗; (5)衰弱型 CRPC 患者可使用新型内分泌治疗药物治疗以及对症治疗。

支持治疗

共识意见 9: (1)前列腺癌患者在接受 ADT 治疗前和治疗过程中建议常规评估心血管基线风险, 并进行常规监测; (2)前列腺癌患者 ADT 治疗过程中建议常规监测血糖、血脂及骨密度; (3)前列腺癌患者内分泌治疗中建议补充维生素 D; 如果膳食中钙摄入量低, 建议补充钙剂; (4)mCSPC 患者不推荐常规应用地诺单抗或双膦酸盐等骨保护治疗; (5)转移性 CRPC 患者, 地诺单抗或双膦酸盐可显著降低骨相关事件的发生率; (6)建议前列腺癌患者接受 ADT 治疗前以及在治疗

过程中定期进行认知评估。

1.3 Core GRADE：让证据评价更高效

标题:Why Core GRADE is needed: introduction to a new series in
The BMJ

作者: Guyatt G, Hultcrantz M, Agoritsas T, et al

出处:BMJ. 2025; 389: e081902.

医学研究方法在过去 40 年间取得了长足进步，证据的合成、获取和在临床指南中的应用也有了显著改善。其中一项重要进展就是 GRADE（推荐等级的评估、制订与评价）方法，它已成为全球评价证据质量及制定推荐强度的主流体系。然而，随着 GRADE 方法的不断扩展与细化，其复杂性也日益增加，给用户带来了挑战。为此，来自全球 14 个国家的 GRADE 专家联合制定了 Core GRADE，旨在为用户提供一个简洁、有序的核心指南。

GRADE 方法的现状与挑战

自 2004 年推出以来，GRADE 方法在系统评价和临床实践指南制定方面产生了重大影响。Cochrane 协作网的所有系统评价都需要遵循 GRADE 方法。GRADE 还得到全球 120 多个组织的认可和使用，包括世界卫生组织（WHO）、

美国内科医师学会、UpToDate 和 The BMJ。

GRADE 方法被广泛应用的原因包括其实用性和用户友好性，以及严谨的方法学和详细的指导。然而，随着 GRADE 方法的不断发展，一些问题也逐渐显现。首先，GRADE 变得越来越复杂，以至于新用户和现有用户都可能会觉得它很笨拙。其次，跟上不断演变的 GRADE 指导很困难，用户需要查找和处理持续出现的新指导，而这些内容往往零散且难以整合。

Core GRADE：回归本质的解决方案

为了解决上述问题，GRADE 专家团队制定了 Core GRADE，将 GRADE 方法的核心要素进行简明而系统的整理，确保用户能够对证据质量和推荐强度做出最佳判断。

Core GRADE 侧重于在临床决策背景下单一干预措施与单一对照之间的比较。它对 GRADE 证据质量评价及从证据转化为推荐的原则进行了系统梳理、重新表述、延伸拓展和改进。Core GRADE 在与以往 GRADE 文献保持基本一致的基础上，纳入了现已成为其核心方法的关键进展。

Core GRADE 的基本原则

1. 灵活性：在评价证据质量和从证据转向推荐时，通常存在多个最佳判断。因此，Core GRADE 提供的是指导而非

规则。

2. 实用性：注重实际应用。
3. 概念简化：将复杂概念简化，使其易于理解和操作。
4. 选择性纳入：如果某项指导对于对证据质量以及推荐方向和强度做出合理判断并非必需，则不纳入 Core GRADE。

Core GRADE 系列论文

Core GRADE 系列论文共包括七篇：

1. 第一篇概述 Core GRADE，涉及证据质量评价和从证据到推荐或决策的转化。
2. 第二篇聚焦证据质量评价的关键问题：质量评价的目标、研究设计和因不精确性降级。
3. 第三篇探讨结果不一致性问题：各研究结果差异显著。
4. 第四篇涉及偏倚风险、因大效应量和剂量-反应梯度提高证据等级的可能性，以及发表偏倚的问题。
5. 第五篇关注间接性相关问题，包括适用性和替代结局。
6. 第六篇介绍 Core GRADE 证据总结表的创建，该表是呈现干预措施对患者重要结局影响的关键单一来源。
7. 第七篇讨论指南制定者和卫生技术评估从业者在从

证据转化为推荐或决策时面临的诸多问题。

Core GRADE 的推出应该能够使 GRADE 方法充分发挥其潜力，通过在系统评价和卫生技术评估指导中提供可信的证据质量判断，以及在临床实践指南中提供推荐意见，从而优化临床决策。

1.4 Core GRADE 1: 方法概述

标题: Core GRADE 1: overview of the Core GRADE approach

作者: Guyatt G, Agoritsas T, Brignardello-Petersen R, et al.

出处: BMJ. 2025; 389: e081903.

Core GRADE 方法概述

Core GRADE 方法主要针对单一干预措施与单一对照措施的比较，主要从个体患者角度出发，而非人群或公共卫生视角，但基本原则也适用于从人群或公共卫生角度出发的证据和决策。

Core GRADE 包括两个基本过程：①对证据的确信度进行评价，并在证据总结表中呈现；②从证据转化为推荐意见，明确推荐的方向（支持或反对）和强度（强或条件性/弱）。

Core GRADE 应用始于明确的临床问题，通过 PICO（人群、干预措施、对照措施、结局）结构构建问题，然后进行

证据整合与评价，最后形成推荐意见。

构建临床问题

精准构建 PICO 是 Core GRADE 方法的基础步骤。这要求对目标人群进行精确界定，包括年龄、疾病严重程度等特征。同时，需全面识别对患者重要的结局指标，并依据系统评价或患者焦点小组意见对其重要性进行分级，分为关键结局、重要但非关键结局和不重要结局。

在构建 PICO 时，还需考虑不同患者群体可能存在的治疗效应差异。例如，某些患者因年龄、基础疾病等不同，可能对同一干预措施产生不同反应。Core GRADE 用户应基于有限且有方向性的假设，对这些差异进行深入分析，以确保问题构建的准确性和全面性。

证据确信度评估

证据确信度评估是 Core GRADE 方法的核心环节。GRADE 使用四级评价证据确信度：高、中、低、极低。评价不是针对单个研究，而是针对解决特定结局的整体证据。这一评价基于五个关键领域：精确性、一致性、偏倚风险、间接性和发表偏倚。例如，对于随机对照试验，证据质量通常从“高”开始评价；而对于非随机对照试验，则通常从“低”开始。如果在上述任一个领域存在问题，证据质量等级将被

相应下调。

从证据到推荐

在完成证据总结与确信度评估后，需将证据转化为推荐意见。决策小组需综合考虑证据确信度、结局的重要性、干预措施的利弊平衡，以及患者价值与偏好。此外，还需考虑成本、可行性、可接受性和公平性等次要标准，以确保推荐意见的全面性和适用性。例如，在评估某种新药的疗效时，不仅要考虑其降低死亡率的效果，还要权衡其可能带来的副作用和经济成本，同时考虑患者对不同结局的重视程度。

Core GRADE 方法的优势与局限性

Core GRADE 方法的优势在于提供了一个结构化、透明且灵活的框架，使决策者能够在全面考虑多种因素后制定科学合理的推荐意见。它强调基于证据的决策，同时允许在证据不确定性情况下进行专业判断。

然而，该方法也存在一定的局限性。证据确信度和推荐意见制定过程中的主观判断不可避免。Core GRADE 方法通过其严谨的框架和透明的流程，尽可能地减少了主观因素的影响，提高了决策的科学性和可靠性。

1.5 Core GRADE 2: 确信度分级目标选择与不精确性评估

标题: Core GRADE 2: choosing the target of certainty rating and assessing imprecision

作者: Guyatt G, Zeng L, Brignardello-Petersen R, et al.

出处: BMJ. 2025; 389: e081904.

在临床研究和循证医学领域,准确评估证据的确信度对于制定可靠的指南与决策至关重要。Core GRADE 方法作为一套系统的评估体系,为证据确信度分级及推荐强度提供了有力工具。本文将深入探讨 Core GRADE 方法的第二部分,即如何选择确信度分级的目标以及如何因不精确性而降低证据等级。

确信度分级的目标选择

在评估干预措施效果时,关键在于确定其是否优于标准治疗或其他现有疗法。Core GRADE 方法聚焦于两个关键问题:一是与替代方案相比是否存在效果,这对应于使用无效假设阈值;二是效果是否足够大以对患者重要,这与使用最小重要差异 (MID) 相关。

选择确信度分级目标可分为三个步骤:

1. 确定关注点:首先要明确是关注效果是否重要,还是

关注是否存在与替代方案相比的真实潜在治疗效果。例如，在某些情况下，我们可能更关心某种新药相较于现有药物是否能带来显著的临床获益，这时就需要关注效果是否重要；而在另一些情况下，我们可能更关注是否存在任何真实的效果，无论其大小，这时则需关注是否存在真实潜在治疗效果。

2. 建立效应估计值：通常通过进行 meta 分析来获得最佳估计值，并选择是使用固定效应模型还是随机效应模型进行分析。固定效应模型假设所有研究的真实效应是一致的，而随机效应模型则允许研究间存在真实效应的变异。一般而言，在 Core GRADE 方法中，更倾向于考虑绝对效应而非相对效应，因为绝对效应更能直接反映患者实际可能获得的临床收益或风险。

3. 根据点估计值选择目标：点估计值是 meta 分析得到的效应估计值。若使用 MID 阈值，当点估计值大于 MID 时，评估的重点是对重要效果的确信度；当点估计值小于 MID 时，则评估对不重要效果的确信度。

若使用无效假设阈值，则根据点估计值是否接近无效假设，评估对存在真实益处或危害的确信度。若点估计值接近无效假设，则评估对不重要效果的确信度。

不精确性评估

不精确性定义：干预研究旨在估计真实的潜在治疗效果，meta 分析提供点估计值，置信区间（CI）给出了真实效应可能存在的范围。95% CI 是最常用的置信区间，其宽度反映随机误差对证据确信度的影响程度。

判断是否因不精确性降级以及降级幅度：

1. 是否跨越阈值：这是判断是否因不精确性降级的关键。若 CI 不跨越所选阈值，则通常不因不精确性降级；若 CI 跨越阈值，则需考虑降级。

2. 降级幅度：随着 CI 变宽，Core GRADE 用户对效果真实性的确定性逐渐降低，此时可考虑因不精确性降一级或两级。

若 CI 跨越多个阈值，如同时跨越重要益处和重要危害的阈值时，考虑降两级。

另外，根据简单语言描述结果的不确定性程度，考虑降两级，以更准确地传达不确定性。

大效应量与样本量有限时的不精确性评估

当 CI 不跨越阈值但效应量很大时，需格外关注。因为大效应量在随机对照试验中较为罕见，早期研究结果的重复性往往不佳。此时，Core GRADE 用户应考虑样本量和事件数是否足够。若样本量和事件数未达到最佳信息量（OIS）标

准，则因不精确性降级。OIS 是基于单个随机对照试验的常规样本量计算得出的，计算时需考虑控制组事件发生率、相对风险降低幅度等参数。

对于二分类结果，若 meta 分析中所有研究的总样本量超过 OIS，则不降级；若低于 OIS，则因不精确性降级。对于连续变量，首先规定想避免错过的干预组与对照组间的最小差异，然后利用现有研究中的一项标准差来计算 OIS。

浅显易懂语言表述的作用

用浅显易懂的语言向临床医生和患者清晰传达研究结果至关重要。GRADE 提供了浅显易懂语言表述的指导，这些表述可帮助 Core GRADE 用户决定是否因不精确性降级以及降级幅度。这些表述将在本系列的第六篇论文中进一步详细讨论。

1.6 Core GRADE 3: 评估证据不一致性

标题：Core GRADE 3: rating certainty of evidence — assessing inconsistency

作者：Guyatt G, Schandelmaier S, Brignardello-Petersen R, et al.

出处：BMJ. 2025; 389: e081905.

本文将介绍 Core GRADE 系列的第三篇论文核心内容，

探讨如何判断和处理研究结果之间的一致性，帮助系统评价作者和指南制定者判断是否因证据的不一致性而降低证据确信度分级。

什么是不一致性？

不一致性 (inconsistency) 指的是不同研究结果之间在无法解释的差异。这种差异可能源于随机误差，也可能源于研究人群、干预措施、对照设置或结局测量的不同。当这种差异足够大，以至于不同研究结果可能导致截然不同的临床实践推断时，我们尤其需要关注。

评估不一致性的关键步骤

1. 选择合适的效应指标

- 二分类结局：应关注相对效应（如风险比、比值比），而非绝对效应。原因是相对效应在不同患者亚组之间通常保持一致，而绝对效应则会由于这些患者群体的基线风险不同而变化。

- 连续性结局：应关注绝对效应，如均值差。

2. 为可能出现的不一致性做准备

在进行系统评价之前，研究者应：

- 构建少量（理想情况下不超过 3 个）有方向性的先验假设，以解释可能出现的不一致性。

- 这些假设应基于前期证据或对生物学的深入理解。
- 明确指出不一致性的预期方向，即预期在某个亚组中效应会更大或更小。

3. 判断不一致性的严重程度

核心判断标准包括三个视觉标准和一个统计标准：

视觉标准（通过森林图评估）：

- 点估计值的差异程度：点估计值之间差异越大，越需要考虑降低证据等级。
- 置信区间的重叠程度：置信区间重叠越少，越需要考虑降低证据等级。
- 点估计值与确信度分级阈值的关系：如果大多数研究的点估计值位于阈值（无效假设或最小重要差异）的同一侧，则不太需要降低证据等级。

统计标准：

- I^2 统计量：描述效应估计值的差异中归因于异质性而非随机误差的百分比。
- I^2 越接近 100%，随机误差解释观察到的差异的可能性越小。
- 但需谨慎解释，特别是当点估计值大多位于确信度分级阈值的同一侧时。

4. 评估亚组效应的可信度

判断亚组效应可信度的关键标准包括：

- 交互作用检验的 P 值：P 值越小，越不可能由偶然因素导致亚组间效应差异，亚组效应越可信。

- 与先验假设的一致性：若亚组效应基于先前证据或生物学机制的深入理解提出，并包含效应方向的假设，则可信度更高。

- 是否基于研究内比较：研究内比较（如同一研究中不同剂量组比较）比研究间比较更具说服力。

如果判定亚组效应可信度为中等或高，应为每个亚组提供单独的证据总结。

如果判定亚组效应可信度低，则仅提供所有研究的综合结果，并对整个人群的不一致性进行降级。

总结

Core GRADE 方法为评估证据不一致性提供了系统的框架。当用户构建广泛的 PICO 框架时，需提前准备应对可能出现的不一致性结果。通过构建先验假设，评估森林图（包括点估计差异、置信区间重叠程度、点估计值与确信度分级阈值的关系），谨慎参考 I^2 统计量，以及评估亚组效应的可信度，可以判断研究结果不一致性的严重程度。

如果发现显著不一致性，则需确定先验假设是否能解释这种不一致性。若亚组效应可信，应为每个亚组提供单独的证据总结；若不可信，则需对所有符合条件的研究评估不一致性，并适当降级证据确信度。

1.7 Core GRADE 4: 评估偏倚风险与发表偏倚

标题：Core GRADE 4: rating certainty of evidence—risk of bias, publication bias, and reasons for rating up certainty

作者：Guyatt G, Wang Y, Eachempati P, et al.

出处：BMJ. 2025; 389: e083864.

在 Core GRADE 的高、中、低和极低这四类证据确信度等级体系中，随机对照试验（RCT）初始证据确信度为高，而非随机干预研究（NRSI，等同于观察性研究）初始证据确信度为低。本文讨论了 Core GRADE 如何处理因偏倚风险、发表偏倚导致证据确信度降级，以及在 NRSI 中将证据确信度升级的原因。

偏倚风险的评估与处理

评估 RCT 偏倚风险时，研究者需使用合适的工具对单个研究进行分类，判断其总体偏倚风险为低或高。偏倚风险评估工具包括 Cochrane 偏倚评估工具 RoB 2 和最新开发的

ROBUST-RCT 等。这些工具覆盖了随机序列生成、分配隐藏、盲法、结果数据缺失等多个评估项目。

评估 NRSI 偏倚风险时，研究者需考虑队列研究、病例对照研究等不同设计类型的特点。偏倚风险评估工具包括 ROBINS-I 和 Newcastle-Ottawa 量表等。关键评估项目包括：参与者的选择偏倚、干预措施的测量偏倚、结局指标的测量偏倚，以及混杂因素的控制等。

当高偏倚风险研究在证据总体中占主导地位时，Core GRADE 建议降低证据确信度等级。然而，如果低偏倚风险研究提供了重要且不同的结果，则应仅使用低偏倚风险研究的结果，并不降低证据确信度等级。反之，如果低偏倚风险研究与高偏倚风险研究结果相似，则应使用所有研究的结果，并保持证据确信度等级不变。

发表偏倚的评估与处理

发表偏倚是由于研究结果的性质（通常为阴性结果）而导致研究未被发表，从而影响证据的完整性。Core GRADE 方法在处理发表偏倚时，强调了对小样本量研究和商业资助研究的警惕性。当大部分或所有研究均为小样本量且由商业资助时，应高度怀疑发表偏倚的存在，并相应降低证据确信度等级。

评估发表偏倚的常用方法包括漏斗图分析和统计学测试。然而，这些方法均存在局限性，因此 Core GRADE 建议使用“未检测到”和“高度怀疑”来描述发表偏倚的情况，以增加评估的透明度和准确性。

提升 NRSI 证据确信度的原因

尽管 NRSI 通常从低质量证据开始，但在某些情况下，其证据质量可以被提升。这主要基于两个原因：一是效应量的大小，二是剂量-反应梯度。当 NRSI 显示干预效果非常显著（相对危险度 >5.0 或 <0.2 ），可以考虑将证据确信度等级提升两级。此外，如果存在明确的剂量-反应梯度，即干预措施的剂量或强度与效应之间存在正相关或负相关关系，也可以作为提升证据确信度等级的依据。但需要注意的是，在考虑剂量-反应梯度时，应排除潜在的混杂因素，以确保因果关系的真实性。

1.8 Core GRADE 5: 评估间接性

标题:Core GRADE 5: rating certainty of evidence-assessing indirectness 作者:Guyatt G, Iorio A, De Beer H, et al.

出处:BMJ.2025;389:e083865.

本文是 Core GRADE 系列论文的第五篇，主要探讨证据

的间接性问题。

什么是证据的间接性？

证据的间接性指的是研究证据与临床问题之间的不匹配。具体而言，当研究中的人群、干预措施、对照措施或结局指标 (PICO) 与临床问题中的目标 PICO 存在显著差异时，就会产生间接性。这种不匹配可能降低证据的适用性，从而影响其确信度。

证据间接性的两种类型

1. 间接比较:这种间接性主要出现在网络 Meta 分析中，当需要比较两种干预措施的效果，但缺乏直接比较的证据时，只能通过它们与第三种干预措施的比较来间接推断。Core GRADE 主要关注直接比较，因此本文不深入讨论间接比较。

2. 与 PICO 相关的间接性：这是本文讨论的重点。在实际的临床问题中，我们有明确的目标 PICO，而研究证据中的 PICO 与目标 PICO 之间存在差异。例如，研究人群的年龄、疾病特征可能与目标人群存在差异；干预措施的剂量、疗程或给药途径可能与预期不符；结局指标的测量方式也可能存在差异。

指南制定者和卫生技术评估从业者必须仔细界定目标

PICO，并评估现有证据与目标 PICO 的匹配程度。当两者存在重大差异时，研究提供的就是间接证据，可能需要对证据确信度进行降级。

两种场景下的间接性处理

1. 在检索直接证据时的间接性

在检索直接证据的过程中，可能会发现一些研究在 PICO 的一个或多个要素上与目标 PICO 存在差异。例如，研究人群的年龄、种族背景、合并症等与目标人群有所不同。然而，在大多数情况下，这些差异并不足以导致证据确信度的降级。原因在于真正的亚组效应相对罕见，基线风险差异（如合并症差异）通常不会导致相对效果的显著不同。

但在某些特殊情况下，例如干预措施依从性低、主要关注替代结局而非患者重要结局、或对照组存在问题，间接性就会变得较为严重，需要考虑对证据确信度进行降级。

2. 在特意检索间接证据时的间接性

当无法获得直接证据或直接证据质量极低时，Core GRADE 用户可能会选择与其目标 PICO 存在较大差异的间接证据。这种情况下，用户必须认真评估间接性，可能需要降低证据确信度。

间接证据的忽视问题

有时指南制定者会错误地认为没有任何证据，而忽视了可用的低质量证据，如临床经验。制定指南和卫生技术评估时，在预期直接证据缺乏的情况下，应当系统地检索可获得的间接证据，并评估其确信度。即使间接证据质量较低，也要优于完全缺乏依据而做出决策。

总结

在制定指南和进行卫生技术评估时，间接性是导致证据确信度降级的常见原因。当直接证据缺乏或质量极低时，Core GRADE 用户应当系统地检索可获得的间接证据，并评估其确信度。对于证据中的每一个 PICO 要素，当其与目标 PICO 存在差异时，都必须评估这种差异是否可能导致干预效果的重要变化，进而决定是否降低证据确信度。

1.9 四大医学顶刊同发：SPIRIT 2025 随机试验方案指南

The image shows four screenshots of medical journals' websites, all announcing the SPIRIT 2025 Statement. The top-left screenshot is from JAMA, showing a search bar and the title "SPIRIT 2025 Statement: Updated Guideline for Protocols of Randomized Trials" dated April 28, 2025. The top-right screenshot is from THE LANCET, showing the title "SPIRIT 2025 statement: updated guideline for protocols of randomised trials" dated April 28, 2025. The bottom-left screenshot is from BMJ, showing the title "SPIRIT 2025 statement: updated guideline for protocols of randomised trials" dated April 28, 2025, with a DOI link and citation information. The bottom-right screenshot is from naturemedicine, showing the title "SPIRIT 2025 statement: updated guideline for protocols of randomized trials" dated April 29, 2025, with navigation links and a DOI link.

临床试验方案是试验全过程的基础，其质量直接影响着

试验的设计、实施和结果解释。2025 年 4 月底，四大医学顶刊 JAMA、Lancet、BMJ、Nature Medicine 同时发布了 SPIRIT 2025 随机试验方案指南 “SPIRIT 2025 statement: updated guideline for protocols of randomized trials”，旨在进一步提高临床试验方案的完整性和透明度，为研究人员、伦理委员会、期刊编辑、决策者等各方利益相关者提供更好的指引。

SPIRIT (Standard Protocol Items: Recommendations for Interventional Trials) 声明于 2013 年首次发布。SPIRIT 声明与 CONSORT 声明相呼应，旨在制定标准项目清单来提升试验方案的完整性。

SPIRIT 2025 主要变化

与 SPIRIT 2013 相比，SPIRIT 2025 主要变化如下：

1. 新增清单条目

- 条目 11：增加了关于患者和公众如何参与试验设计、实施和报告的条目。

- 条目 29：增加了关于试验监查的条目（替代了之前的审计条目）。

2. 修订的清单条目内容

- 条目 4：修订条目包含了试验注册日期。

- 条目 5: 修订为除了试验方案外, 还需说明如何获取统计分析计划。

- 条目 7b: 修订条目包含了指导委员会成员的经济利益和其他利益冲突。

- 条目 24a/24b: 将条目分为两个子条目, 分别规定 (a) 谁将被设盲和 (b) 如何设盲, 并包含了对干预措施相似性的描述。

- 条目 27d: 修订为明确提及敏感性分析。

3. 删除/合并的清单条目

- 删除了关于审计的条目 (由试验监查条目替代) 。

- 删除了附录条目:

- 知情同意材料
- 生物样本

- 将试验研究者获取数据的条目与数据管理条目合并 (条目 26) 。

- 将作者资格指南和使用专业写作者的条目与传播政策条目合并 (条目 8) 。

4. 整合来自 CONSORT Harms 2022 和 TIDieR 的清单条目

- 修订条目以强调不良事件的报告 (条目 10、17、27a),

并要求提供有关干预措施和对照措施的更多细节（条目 15a）。

5. 清单条目的结构和组织

- 创建了一个关于开放科学的新章节,包括试验注册(条目 4)、获取试验方案和统计分析计划(条目 5)、去识别化参与者数据的共享计划(条目 6)、资金来源和利益冲突披露(条目 7)以及试验结果的传播计划(条目 8)。

- 将条目 14a/b 拆分为两个子条目,涵盖 (a) 参与者和 (b) 研究中心及工作人员的资格标准。

- 将项目 27b/c 拆分为两个子条目,涵盖分析人群和处理缺失数据的方法。

- 使 SPIRIT 清单条目的措辞与 CONSORT 清单条目保持一致。

- 阐明和简化了某些条目的措辞。

1.10 CCR 最新：新辅助免疫联合 PARPi 具有治疗 MIBC 的潜力

标题：Treatment Efficacy and Molecular Dynamics of Neoadjuvant Durvalumab and Olaparib in Resectable Urothelial Bladder Cancer: the NEODURVARIB Trial

作者：Rodríguez-Moreno JF, de Velasco G, Álvarez-Fernández C, et al.

出处：Clin Cancer Res. 2025; 31(9): 1644-1656.

肌层浸润性膀胱癌（MIBC）的治疗策略一直在不断探索与优化中。近年来，新辅助治疗在 MIBC 中的应用逐渐增多，旨在通过术前治疗缩小肿瘤体积、降低手术难度，并提高患者的生存率。度伐利尤单抗（durvalumab）是一种 PD-L1 免疫抑制剂，奥拉帕利（olaparib）是一种 PARP 抑制剂。联合应用这两种药物有望通过免疫激活与 DNA 损伤修复抑制的协同作用，提高治疗效果。本研究（NEODURVARIB 试验）旨在评估新辅助度伐利尤单抗联合奥拉帕利在可切除 MIBC 中的疗效及分子动力学变化。

NEODURVARIB 是一项多中心、II 期临床试验，共纳入 29 名可接受根治性膀胱切除术且适合接受新辅助治疗的 T2-T4a 期膀胱癌患者。该研究主要排除了在接受免疫抑制药物治疗及对免疫治疗有禁忌证的患者。纳入研究的患者接受两个周期的治疗。具体方案包括：度伐利尤单抗以 1500mg 的固定剂量通过静脉滴注给药，每 4 周一次，奥拉帕利口服，每次 300mg，每日两次，持续 8 周，根治性膀胱切除术计划在治疗结束后进行，但可根据研究者判断提前至第 6 周后进

行，奥拉帕利需在膀胱切除术前至少 7 天停药。研究在治疗前后分别收集肿瘤样本，用于全外显子测序、RNA 测序及免疫组化检测，以评估肿瘤的基因突变谱、肿瘤突变负荷、同源重组缺陷（HRD）、免疫细胞浸润等分子特征，并与临床疗效进行相关性分析。同时，研究还在治疗的不同时间点收集外周血样本，通过流式细胞术分析循环免疫细胞亚群的变化。主要研究终点是评估新辅助治疗对 MIBC 分子特征的影响，通过比较治疗前后肿瘤的突变谱和基因表达模式来实现。次要研究终点包括病理反应和放射学反应的评估，以及毒性反应的观察。

该研究的主要结果显示，在 29 名接受联合治疗的患者中，13 名（44.8%）达到了病理完全缓解（pCR）。在 25 名可评估放射学反应的患者中，7 名（28%）获得部分缓解，15 名（60%）病情稳定，3 名（12%）发生疾病进展。93.1% 的患者完成了两个周期的新辅助治疗。26 例（90%）患者接受了根治性膀胱切除术。未发生与治疗相关的严重不良事件。全外显子测序结果显示，治疗前后肿瘤的基因突变谱、肿瘤突变负荷和 HRD 保持稳定，且与疗效无关。RNA 测序分析发现，治疗后肿瘤样本中富集了与上皮-间充质转化（EMT）、免疫炎症反应相关的基因，尤其是在疗效欠佳的

患者中更为明显。免疫组化分析显示，治疗后 PDL1 和 FAP 在肿瘤细胞中的表达显著增加。流式细胞术分析未发现循环免疫细胞亚群在应答者和非应答者之间的显著差异，但观察到应答者中 CD4+CD7-CD8-T 细胞比例在治疗期间有增加趋势。

该研究证实，度伐利尤单抗联合奥拉帕利的新辅助治疗在 MIBC 中显示出较好的临床疗效。基因组特征在治疗期间保持稳定，但转录组水平的显著变化，特别是 EMT 和免疫炎症反应的激活，可能是影响治疗效果的关键因素。此外，该治疗组合的安全性良好，未观察到严重不良事件。

1.11 Eur Urol 最新：机器人辅助单对接技术为 III 级 IVC 癌栓切除术带来新选择

标题：Robot-assisted Single-docking Approach for Level III Inferior Vena Cava Tumor Thrombectomy: Surgical Technique and Outcomes

作者：Wang S, Pokhrel G, Cui J, et al.

出处：Eur Urol. 2025; S0302-2838(25)00198-8.

肾细胞癌中合并下腔静脉（IVC）癌栓的发生率约为 4-10%，预后较差，需要采取复杂的手术治疗方案。III 级癌

栓是最为复杂的一类，其特点是肿瘤血栓延伸至 IVC 的肝后段或膈肌下段。这类手术操作难度极大，传统开放手术方式创伤大、并发症风险高。近年来，机器人辅助手术技术在治疗 RCC IVC 癌栓中逐渐显现优势，但在 III 级癌栓手术中仍存在诸多挑战，如对接策略、血管控制、暴露技术等方面的问题。本研究提出了一种用于 III 级 IVC 癌栓的新型单对接机器人辅助癌栓切除术（RATT）技术，并通过回顾性分析评估了该技术的安全性、可行性和临床结局。

该研究的研究对象主要是 2019 年至 2023 年间，接受单对接 RATT 治疗的 RCC 合并 III 级 IVC 癌栓患者。癌栓根据 Mayo 分类进行分类，并发症使用 Clavien-Dindo 系统进行分级。手术使用达芬奇 Si 手术系统，采用单对接六孔布局（三个机器人臂孔和三个辅助孔）。手术关键步骤包括：（1）手术开始离断肾动脉；（2）优先结扎 IVC 属支，依次将生殖静脉、腰静脉、肾上腺中央静脉、肝短静脉游离后结扎；（3）根据癌栓累及范围采用不同的静脉临时阻断策略。IIIa 癌栓，只需阻断 IVC 远心端、左肾静脉及肝后 IVC（策略 A），对于 IIIb、c、d 癌栓，需要阻断 IVC 远心端、左肾静脉、第一肝门及 IVC 近心端，取出癌栓后将肝后 IVC 阻断，快速松开第一肝门及 IVC 近心端（策略 B）；（4）切除癌栓后，适

当调整 IVC 临时阻断位置；（5）用连续缝合技术修复 IVC 壁。术后监测患者血流动力学和肾功能，以及指导早期活动。预防性使用肝素抗凝，出院后抗凝 1 个月。定期随访，包括伤口评估、并发症监测、功能评估、肾功能检查、肿瘤标志物、影像学检查等。

研究结果发现，15 例手术均成功采用单对接技术完成。中位手术时间为 225 分钟，中位出血量为 400 毫升，12 例（80%）术中接受输血。中位 IVC 阻断时间为 14 分钟。7 例患者行节段性 IVC 切除，其中 2 例接受血管成形术重建。11 例（73.3%）患者出现并发症，包括 3 例 Clavien-Dindo IIIb 级事件。中位随访时间为 24 个月，期间 5 例患者死亡，1 例出现肝转移。

研究结论显示，III 级 IVC 癌栓的单对接 RATT 是一种可行且可重复的技术，为传统方法提供了一种安全有效的替代方案。该技术通过消除多次对接和患者体位调整的需要来简化手术流程，并解决了实现最佳手术暴露和有效血管控制方面的挑战。然而，该技术仍需在更大样本量和更长随访时间的多中心研究中进一步验证其长期疗效。

1.12 JAMA 最新：混合性尿失禁治疗如何选？多中心 RCT 给出答案

标题：Midurethral Sling vs Onabotulinumtoxin in Females with Urinary Incontinence: the MUSA Randomized Clinical Trial

作者：Harvie HS, Menefee SA, Richter HE, et al.

出处：JAMA. 2025: e254682.

良性前列腺增生（BPH）是中老年男性群体中常见且高混合性尿失禁（MUI）是指同时存在压力性尿失禁（SUI）和急迫性尿失禁（UUI）的情况，发病率随年龄增长而升高，严重影响生活质量。患者通常认为 UUI 比 SUI 更令人烦恼，而两者的组合也比单一症状更具困扰性。对于保守治疗（如行为疗法、盆底肌锻炼和药物治疗）失败的患者，临床上对于高级膀胱治疗（如手术/侵入性操作）的获益及治疗顺序一直存在不确定性。先用 A 型肉毒毒素或骶神经调控治疗 UUI 的方法并不能保证 SUI 症状得到改善。关于先用尿道中段吊带术治疗 SUI 是否能显著改善 UUI 的数据有限。本研究（MUSA 随机临床试验）旨在比较 A 型肉毒毒素和尿道中段吊带术治疗女性 MUI 的疗效，为临床治疗决策提供依据。

MUSA 试验是一项多中心、随机、优效性临床试验，由美国 7 个临床中心参与。纳入标准：年龄 ≥ 21 岁女性；中度

或重度 SUI 和 UUI 症状困扰至少 3 个月；咳嗽压力诱发试验阳性；3 天膀胱日记记录至少 4 次 UUI 发作；保守治疗和口服药物治疗无效。参与者按 1:1 被随机分配至 A 型肉毒毒素组或尿道中段吊带术组。A 型肉毒毒素组在门诊或手术室接受膀胱内注射 100 单位 A 型肉毒毒素；尿道中段吊带术组在手术室接受合成聚丙烯网片尿道中段吊带术，可采用经耻骨后和经闭孔两种技术。A 型肉毒毒素组可在 3-6 个月期间接受额外一次注射。两组患者在 6 个月内除了分配的干预措施外，不接受其他尿失禁治疗。6 个月后，所有患者可交叉接受另一种治疗或接受研究外的其他治疗。主要结局：6 个月时 MUI 症状变化，使用泌尿生殖障碍量表（UDI）测量。次要结局：3 个月时 UDI 总分变化、6 个月时 UDI 压力和刺激子量表分数变化。

研究发现，共有 150 名女性被随机分配，其中 140 名接受治疗（A 型肉毒毒素组 71 名，尿道中段吊带术组 69 名）。137 名有基线后数据，134 名完成 6 个月随访，130 名完成 12 个月随访。参与者平均年龄为 59.0 岁。基于 UDI 总分，MUI 严重程度平均为 184.3，平均每日尿失禁发作次数为 7.2。82.5% 的女性具有压力性和急迫性平衡的 MUI。两组在 6 个月时 UDI 总分均有改善，组间无显著差异。A 型

肉毒毒素组均值下降 66.8 分，尿道中段吊带术组均值下降 84.9 分，组间均值差异 18.1 分 ($P=0.12$)，表明 A 型肉毒毒素注射并不优于尿道中段吊带术。3 个月时，两组 UDI 总分差异无统计学意义 (4.7 分, $P=0.69$)。6 个月时，两组 UDI 刺激和压力子量表分数均有改善。两组 UDI 刺激分数相比，差异无统计学意义 ($P=0.27$)。尿道中段吊带术组 UDI 压力分数改善程度显著高于 A 型肉毒毒素组，均值差异 20.1 分 ($P < 0.001$)。两种治疗相关的并发症较少。尿道中段吊带术组有较高比例的患者出院时带导尿管 (11.8% vs 0%)，但两组在治疗 2 周后需要间歇性自导尿的比例相似。12 个月内，尿道中段吊带术组复发性尿路感染发生率高于 A 型肉毒毒素组 (17.6% vs 6.9%)。尿道中段吊带术组中，阴道网片暴露发生率为 2.9%，1.5% 需要手术修正。

研究结论显示，在经保守治疗无效的中重度 MUI 女性患者中，A 型肉毒毒素注射与尿道中段吊带术在 6 个月时对 MUI 症状改善方面未观察到显著差异。中重度 SUI 和 UUI 症状难以通过单一治疗方式得到有效控制。这些研究结果有助于基于患者个体偏好与临床医生建议来制定治疗决策。

1.13 Eur Urol 最新：肺癌筛查联合腹部 CT 可筛查肾癌

标题：Abdominal Noncontrast Computed Tomography Scanning to Screen for Kidney Cancer and Other Abdominal Pathology Within Community-based Computed Tomography Screening for Lung Cancer: Results of the Yorkshire Kidney Screening Trial

作者：Stewart GD, Godoy A, Farquhar F, et al.

出处：Eur Urol. 2025; 87(5): 561-570.

癌症的早期发现可以改善患者预后。目前，美国和英国已经针对高危吸烟人群建立了肺癌筛查项目。而在肺癌筛查的基础上增加其他疾病的筛查，为那些单独筛查成本效益较低的疾病提供了新的筛查途径。

吸烟也是肾细胞癌（RCC）的危险因素。RCC 早期通常无症状，导致诊断较晚；然而，如果及早发现，其 10 年无病生存率可达 95%。鉴于此，研究者们提出了在基于 CT 的肺癌筛查中同时增加腹部非增强 CT 扫描（NCCT），以筛查肾癌及其他腹部病变。本研究，即 Yorkshire 肾脏筛查试验（YKST），旨在评估这一联合筛查方法的可行性、接受度及筛查结果。

YKST 是一项前瞻性诊断研究，其研究对象来自 Yorkshire 肺部筛查试验（YLST）的参与者，这些参与者在

2021年5月至2022年10月期间参加了YLST第二次研究访视（T2）。YLST的入选标准包括：年龄55-80岁，有吸烟史，无痴呆、身体虚弱，过去5年内未患肺癌，无转移性癌症病史，过去12个月内未做胸部CT，未纳入姑息治疗登记。

YKST的纳入标准是参加YLST的T2访视并能提供知情同意，排除标准包括过去6个月内做过腹部CT、不能接受腹部NCCT和既往有RCC病史。符合条件的YLST T2参与者在参加YLST访视时进行腹部NCCT。扫描范围从膈肌至肾脏底部，与胸部CT分开进行。研究人员记录了联合筛查的每个环节所花费的额外时间。扫描结果正常的参与者自动收到“正常扫描”信函，无需进一步的检查或随访。对于图像存在异常的参与者，则会被纳入影像审查会议进行进一步讨论和评估。根据讨论结果，参与者可能会被给予以下几种处理建议之一：

对于明确无需进一步处理的良性病变，给予安慰信；

对于已知或相关问题，建议重新与相应诊所联系；

对于非泌尿科问题，转诊至相应的非泌尿科；

对于泌尿科问题，转诊至泌尿科；

对于需要进一步CT扫描及多学科团队评估的情况，安

排相应的检查和会诊。对于所有腹部 CT 异常的参与者，在 6 个月时审查电子健康记录，并根据最终诊断将其分为严重或非严重类型。此外，研究还对 500 名参与者进行了 3 个月问卷调查，包括所有腹部 CT 扫描异常的参与者和正常扫描参与者的随机样本，评估参与者对研究过程的满意度和决策后悔程度。该研究的主要结局包括量化腹部 NCCT 在肺癌筛查中的接受度、评估参与者对联合筛查的接受度，以及评估 NCCT 在肺癌筛查中的可行性和流程。

研究结果发现，共有 4019 名参与者同意参加 YKST。年龄较大者、女性和目前仍在吸烟者参与筛查的意愿较低。在 3 个月问卷调查中，93% 的参与者对提供的信息“满意”或“非常满意”，98% 认为他们“可能”或“肯定”有足够的时间决定参与，96% 对结果给予方式“满意”或“非常满意”。97% “同意”或“强烈同意”他们参与的决定是正确的，只有 3.9% “同意”或“强烈同意”他们后悔参与 YKST 的决定。实施联合筛查所需的平均额外时间为 13.3 分钟。腹部 NCCT 的辐射剂量被逐步降低至最终中位剂量 4 mGy，低于国家和地方参考水平。所有扫描剂量迭代的图像质量均被评为“优秀”。

在 4019 名参与者中，2586 名（64%）腹部 NCCT 结果

正常,而 787 名需要腹部 NCCT 复查但无需进一步处理,611 名需要进一步评估(检查和/或门诊随访)。共有 211 名(5.3%)参与者发现了新的严重病变,包括 25 名肾肿块/复杂囊肿患者,其中 10 名经组织学证实为肾癌;10 名其他癌症患者;以及 60 名腹主动脉瘤患者。共有 25 名(0.62%)参与者接受了以治愈为目的的治疗。1017 名(25%)参与者有非严重性病变,最常见的是良性肾囊肿(727 名),而仅有 259 名需要进一步检查。

研究结论包括,在肺癌筛查的胸部 CT 检查中增加腹部 NCCT 的联合筛查方法具有较高的接受度和可行性,参与者满意度高,扫描质量优异,且额外所需时间合理。筛查发现的严重病变、癌症和腹主动脉瘤的患病率与现有筛查项目相当。未来的工作应侧重于筛查严重腹部异常,而非仅限于肾癌。在临床实施之前,需要开展研究来评估该筛查策略的临床影响,如疾病分期的改变、疾病特异性生存率的提高以及成本效益分析等。

1.14 Eur Urol 最新: MRI 靶向活检或重塑前列腺癌诊断范式

标题: VISION: An Individual Patient Data Meta-analysis of Randomised Trials Comparing Magnetic Resonance Imaging

Targeted Biopsy with Standard Transrectal Ultrasound Guided Biopsy in the Detection of Prostate Cancer

作者: Kasivisvanathan V, Wai-Shun Chan V, Clement KD, et al.

出处: Eur Urol. 2025; 87(5): 512-523.

过去三十年来,对于有前列腺癌可疑症状 (PSA 升高、直肠指诊异常和/或前列腺癌家族史) 的患者,传统诊断方法是系统性经直肠超声 (TRUS) 引导下的前列腺活检。然而,TRUS 对临床显著性前列腺癌 (csPCa) 的敏感性较低,仅为 48%,导致许多患者漏诊或需要多次重复活检。

近五年来,活检前磁共振成像 (MRI) 及 MRI 靶向活检 (MRI±TB) 被提倡作为诊断局限性前列腺癌的替代标准。在该路径中,临床疑似局限性前列腺癌的患者先接受前列腺 MRI 检查,MRI 可疑者进行 MRI±TB 以确定癌症存在和组织学特征,而 MRI 无可疑病变者可避免活检。

PRECISION 和 PRECISE 试验是两项随机对照试验,设计目的是对比 MRI±TB 与 TRUS 活检在诊断前列腺癌中的效果。PRECISION 研究证明 MRI±TB 优于 TRUS 活检,而 PRECISE 研究则证明 MRI±TB 不劣于 TRUS 活检。本研究 (VISION 研究) 是一项计划中的单个病例数据 Meta 分析 (IPDMA),旨在比较 MRI±TB 与 TRUS 活检在 csPCa 诊

断中的效果，为前列腺癌诊断提供更高级别的证据。

该研究的相关研究人员于 2023 年 11 月 12 日在 MEDLINE、EMBASE、Web of Science、Cochrane Central of Registered Trials 和 ClinicalTrials.gov 数据库中进行了系统检索。纳入标准包括：（1）研究对象为未接受过活检且有局限性前列腺癌临床可疑症状的男性；（2）干预措施为 MRI 土 TB，MRI 可疑病例进行靶向活检，不可疑病例避免活检；（3）对照措施为标准 10-12 针芯 TRUS 活检；（4）仅纳入随机对照试验。

主要结局指标是诊断为 csPCa（Gleason 评分 $\geq 3+4$ ）的患者比例。次要结局指标包括诊断为临床不显著性前列腺癌（cisPCa）的患者比例、MRI 后避免活检的患者比例、活检阳性比例、接受根治性治疗的比例、不良事件发生率和健康相关生活质量（QOL）评分等。

数据分析采用意图治疗（ITT）原则，使用三水平广义线性混合模型对所有研究同时进行分析，以考虑患者在不同研究中心内的聚类效应。此外，还进行了敏感性分析和亚组分析，以评估不同因素对主要结局的影响。

使用 QUADAS-2、QUADAS-C 和 ROB 2.0 工具对纳入研究的偏倚风险进行评估，并使用 START 标准和 STARD 清单

对研究报告的质量进行评估。

研究结果发现,检索到的 76 项研究中,仅有 PRECISION 和 PRECISE 试验符合纳入标准,共纳入 953 名患者进行 IPDMA 分析。评估后发现这两项研究的偏倚风险较低。

主要结局方面, MRI±TB 检测 csPCa 的比例显著高于 TRUS 活检 (36.3% vs 27.6%; 差异 8.7 个百分点, $P=0.004$)。这表明 MRI±TB 不仅不劣于而且优于 TRUS 活检。

次要结局方面, MRI±TB 组 cisPCa 检出率显著低于 TRUS 组 (9.6% vs 21.9%; 差异-12.3 个百分点, $P<0.001$)。MRI±TB 组活检阳性率显著高于 TRUS 组 (48.8% vs 20.4%; 差异 29.3 个百分点, $P<0.001$)。不良事件发生率在 MRI±TB 组为 2.9%, TRUS 组为 5.1%, 差异无统计学意义。TRUS 组最常见的不良事件是脓毒症,有 7 例 (1.5%), 而 MRI±TB 组仅 1 例 (0.2%)。

该研究提供了 1a 级证据,对于局限性前列腺癌, MRI±TB 诊断路径优于传统的 TRUS 活检,因为它能更好地检测出 csPCa,同时避免检测到 cisPCa。MRI 应被纳入前列腺癌诊断的标准路径。

1.15 Nat Med 最新：ctDNA 或可指导 LuPSMA 个体化治疗

标题：Lutetium-177-PSMA-617 or cabazitaxel in metastatic prostate cancer: circulating tumor DNA analysis of the randomized phase 2 TheraP trial

作者：Kwan EM, Ng SWS, Tolmeijer SH, et al.

出处：Nat Med. 2025.

在转移性去势抵抗性前列腺癌（mCRPC）的治疗中，靶向前列腺特异性膜抗原（PSMA）的放射性配体¹⁷⁷Lu-PSMA-617（简称 LuPSMA）已成为一种新的标准治疗方案。然而，目前尚缺乏能够预测其治疗效果的基因组生物标志物。TheraP 试验是一项随机对照 2 期临床试验，比较了 LuPSMA 与卡巴他赛在既往接受过多西他赛化疗的 mCRPC 患者中的疗效。本研究对 TheraP 试验中 180 名患者的 290 份血浆游离 DNA（cfDNA）样本进行了详细分析，旨在寻找能够指导 LuPSMA 合理使用的预测性基因组生物标志物。

该研究基于 TheraP 试验的血液样本进行相关性分析。从参与者中收集了 290 份血浆 cfDNA 样本，并进行深度靶向测序。测序使用一个验证过的定制前列腺癌研究试剂盒，能够分辨体细胞突变、染色体非整倍体、局灶性拷贝数改变和复杂结构重排。同时，对匹配的白细胞 DNA 进行同步测

序，以区分肿瘤特异性和种系变异。

通过基因组的非整倍体和体细胞突变测量循环肿瘤 DNA 分数 (ctDNA%)。基于 ctDNA% 将患者分为预定义的预后类别：低或未检测到 (<2%)、中 (2-30%) 和高 (>30%)。此外，还分析了关键前列腺癌驱动基因的基因组改变，包括 TP53、PTEN 和 AR 等。通过 PET 检查评估 PSMA 肿瘤摄取和代谢肿瘤体积，这些参数与 ctDNA% 相关联，以评估其作为生物标志物的潜力。

该研究结果显示，低 ctDNA% 预测更好的生化反应和无进展生存期 (PFS)：治疗前 ctDNA% 低的患者在 LuPSMA 治疗中表现出更高的 PSA50 生化反应率 (100% vs 58%， $P=0.0067$) 和更长的 PFS (中位 14.7 个月 vs 6.0 个月； $HR=0.12$ ， $P=2.5 \times 10^{-4}$)。但在总生存期 (OS) 方面，这种益处并未延续。ctDNA% 作为预测和预后生物标志物潜力：在不同的 ctDNA% 类别中，LuPSMA 与卡巴他赛的治疗效果存在差异，表明 ctDNA% 有望成为区分 LuPSMA 和卡巴他赛治疗效果的候选预测和预后生物标志物。

PTEN 变异：PTEN 有害变异与卡巴他赛治疗的更差 PFS 和 OS 相关，但在 LuPSMA 治疗中，PTEN 变异患者 PSA50 反应率更高，且 PFS 有所改善。ATM 缺失：部分 ATM 缺失

患者在 LuPSMA 治疗中获得有利结果，表明 ATM 缺失可能与 LuPSMA 疗效相关。TP53 变异：TP53 变异与较差的 PFS 和 OS 相关，但与特定治疗方案无关。AR 变异：AR 变异与两种治疗方案的疗效无显著关联。

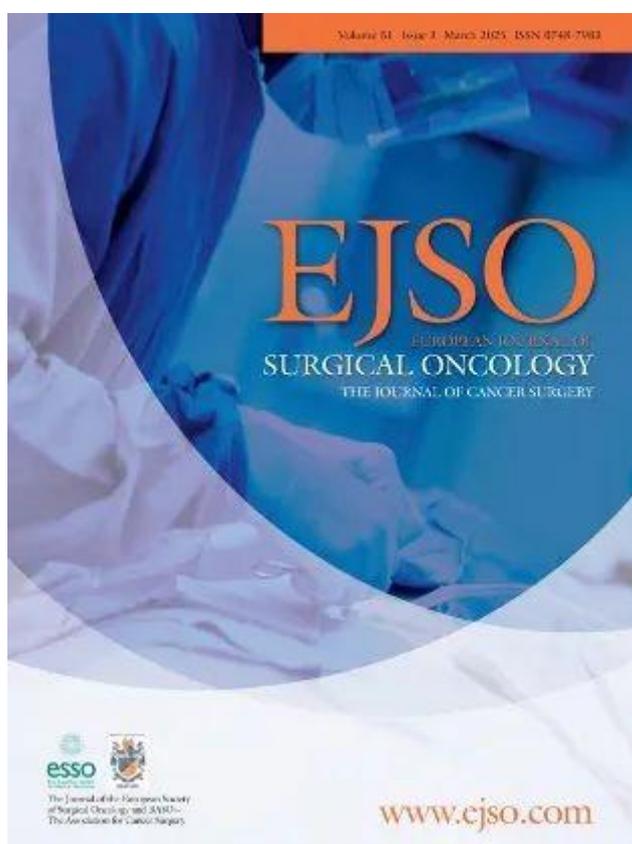
研究未发现单个 mCRPC 基因变异是 LuPSMA 或卡巴他赛获得性耐药的主要原因。尽管有些患者在疾病进展时出现了新的 TP53、RB1、PTEN 和 AR 突变，但这些突变大多是亚克隆。

该研究结论显示，ctDNA% 是一个潜在的预测和预后生物标志物，能够帮助区分 LuPSMA 和卡巴他赛在 mCRPC 患者中的治疗效果。该研究为 LuPSMA 的选择提供了新的候选生物标志物，并扩展了 mCRPC 预测生物标志物的范围，为进一步发展 PSMA 靶向放射性药物治疗的个体化预后预测和治疗选择策略提供了依据。

2 期刊小知识

2.1 期刊介绍 | European Journal of Surgical Oncology

European Journal of Surgical Oncology 是欧洲外科肿瘤学会官方期刊，创刊于 1975 年，2006 年获得首个影响因子。目前 SCIE Q1 区，中科院医学大类 2 区。



期刊基本信息

期刊全称：European Journal of Surgical Oncology

期刊缩写：Eur J Surg Oncol

ISSN：0748-7983

eISSN：1532-2157

出版周期：月刊

出版者：Elsevier

期刊官网：<https://www.ejso.com/>

投稿网址：

[https://www.editorialmanager.com/ejso/default.as](https://www.editorialmanager.com/ejso/default.aspx)

px

影响因子及分区

目前最新影响因子（2023 IF）为 3.5。

JCR 分区：Q1，Surgery 31/292；Q2，Oncology 107/322。

中科院分区：医学大类 2 区；外科学小类 2 区；肿瘤学小类 2 区。

收稿范围

European Journal of Surgical Oncology 旨在推动外科肿瘤学的研究和实践，编辑欢迎关于临床研究和外科肿瘤学其他所有方面的投稿，包括手术质量控制、流行病学、预防，以及与外科肿瘤学实践相关的转化研究。该期刊设有原始论文、综述、社论等栏目。

审稿周期

从投稿到初审决定的中位时间为 13 天，总审稿的中位

时间为 57 天，从投稿到收到接受决定的中位时间为 74 天。

(审稿周期仅供参考，请以稿件实际处理周期为准)

发文情况

2021 年至 2024 年期间,每年平均刊载研究论文 323 篇、综述 51 篇。2021—2023 年发文量 (含参与发表、含社论) 排在前 3 位的国家或地区分别是荷兰、中国大陆和意大利。2021—2023 年发文量 (含参与发表、含社论) 排在前 3 位的机构分别是阿姆斯特丹大学、鹿特丹伊拉斯姆斯大学和拉德堡德大学。

论文处理费

期刊提供传统订阅模式 (不收论文处理费) 或者开放获取 (open access) 模式。目前开放获取模式的论文处理费为 3810 美元。

2.2 期刊介绍 | Annals of Surgical Oncology

Annals of Surgical Oncology 是美国外科肿瘤学会官方期刊, 创刊于 1994 年, 1997 年获得首个影响因子。目前 SCIE Q1 区, 中科院医学大类 2 区。



期刊基本信息

期刊全称：Annals of Surgical Oncology

期刊缩写：Ann Surg Oncol

ISSN：1068-9265

eISSN：1534-4681

出版周期：月刊

出版者：Springer

期刊官网：<https://link.springer.com/journal/10434>

投稿网址：

<https://mc.manuscriptcentral.com/aso>

影响因子及分区

目前最新影响因子（2023 IF）为 3.4。

JCR 分区：Q1，Surgery 35/292；Q2，Oncology

110/322。

中科院分区：医学大类 2 区；外科学小类 2 区；肿瘤学小类 3 区。

收稿范围

Annals of Surgical Oncology 为各专业肿瘤外科医生提供肿瘤学最新进展的研究论文，致力于传播高质量的临床和转化性肿瘤研究成果，推动全球外科肿瘤学专业的发展。该期刊设有原始论文、综述、社论等栏目。

审稿周期

我们参考已发表的论文进行推算，从投稿到被期刊接受的平均时间大约是 2 个月。（审稿周期仅供参考，请以稿件实际处理周期为准）

发文情况

2021 年至 2024 年期间，每年平均刊载研究论文 686 篇、综述 49 篇。每年的美国外科肿瘤学会年会的会议摘要都以增刊形式发表在该期刊上。2021—2023 年发文量（含参与发表、含会议摘要）排在前 3 位的国家或地区分别是美国、日本和中国大陆。2021—2023 年发文量（含参与发表、含会议摘要）排在前 3 位的机构分别是得克萨斯大学、俄亥俄大学和纪念斯隆-凯特琳癌症中心。

论文处理费

期刊提供传统订阅模式（不收论文处理费）或者开放获取（open access）模式。目前开放获取模式的论文处理费为 4590 美元。

2.3 期刊介绍 | World Journal of Surgical Oncology

World Journal of Surgical Oncology 创刊于 2003 年，2010 年获得首个影响因子。目前 SCIE Q1 区，中科院医学大类 3 区。

期刊基本信息

期刊全称：World Journal of Surgical Oncology

期刊缩写：World J Surg Oncol

eISSN：1477-7819

出版周期：月刊

出版者：BMC

期刊官网：<https://wjso.biomedcentral.com/>

投稿指南：

<https://wjso.biomedcentral.com/submission-guidelines>

影响因子及分区

目前最新影响因子（2023 IF）为 2.5。

JCR 分区：Q1，Surgery 68/292；Q3，Oncology 183/322。

中科院分区：医学大类 3 区；外科学小类 3 区；肿瘤学小类 3 区。

收稿范围

World Journal of Surgical Oncology 发表与外科肿瘤学及其相关学科的论文，例如流行病学、癌症研究、生物标志物、预防、病理学、放射学、癌症治疗、临床试验、多模式治疗和分子生物学等。该期刊重点发表原始研究论文。

审稿周期

我们参考已发表的论文进行推算，从投稿到被期刊接受的平均时间大约是 3 个月。（审稿周期仅供参考，请以稿件实际处理周期为准）

发文情况

2021 年至 2024 年期间，每年平均刊载研究论文 320 篇、综述 45 篇。2021—2023 年发文量（含参与发表）排在前 3 位的国家或地区分别是中国大陆、日本和美国。2021—2023 年发文量（含参与发表）排在前 3 位的机构分别是福建医科大学、浙江大学和南京医科大学。

论文处理费

期刊采用开放获取（open access）出版模式，目前论文处理费为 2990 美元。

2.4 影响因子新变革：不再计入撤稿论文的被引用次数

近日，科睿唯安宣布，从 2025 年《期刊引证报告》（JCR）开始，计算期刊影响因子（IF）时将不再计入撤稿论文的被引用次数，但撤稿论文仍将计入期刊论文总数。这一决定可能会对学术界广泛关注的期刊评价指标产生深远影响。

Clarivate to stop counting citations to retracted articles in journals' impact factors

Clarivate will no longer include citations to and from retracted papers when calculating journal impact factors, the company



政策变更的背景

虽然撤稿论文在科睿唯安 Web of Science 索引中仅占 0.04%，但近年来撤稿论文数量有所增加，总体撤稿率已上升至约 0.2%。这一上升趋势，加上撤稿所需时间缩短，促使科睿唯安做出政策调整。Web of Science 主编表示，此举是为了“先发制人地防范撤稿论文的引用可能对期刊影响因

子造成广泛扭曲的情况”。

IF 新计算公式

IF 是期刊过去两年发表的文献在当年被引用的总次数与该期刊过去两年发表的文献总数的比值。IF 的计算仅涉及原始研究论文 (Article) 和综述 (Review) 两类文献。根据新政策，以 2024 年的某一期刊 IF 为例，IF 的计算公式将变为：

IF (2024) = 该期刊 2022—2023 年发表文献在 2024 年被引用总次数 (剔除撤稿引用) / 该期刊 2022—2023 年发表的文献总数 (包含撤稿论文)

分子部分：一旦论文被撤稿，该论文此后的被引用次数将不再计入 IF，撤稿前已获得的被引用次数将保留。目的是确保撤稿论文的引用不会影响期刊 IF 的数值。

分母部分：撤稿论文仍将计入期刊论文总数，以保持透明度和问责制。

2.5 Nature 独家：21 世纪被引用次数最多的论文

近期，Nature 杂志评选出了 21 世纪被引用次数最多的 25 篇论文，分析数据来自五个数据库 (Web of Science、Scopus、OpenAlex、Dimensions、Google Scholar)，这

些数据库涵盖了 21 世纪发表的数千万篇论文。Nature 杂志新闻团队对这 25 篇论文分析后发现，mRNA 疫苗、CRISPR 基因编辑技术、希格斯玻色子的发现和引力波的首次测量等重大科学突破均未进入榜单，被引用次数最多的论文主要集中在人工智能（AI）、研究方法改进和研究软件等领域。Nature 杂志新闻专题版面刊发了新闻团队的报道“Exclusive: the most-cited papers of the twenty-first century”。

[nature](#) > [news feature](#) > [article](#)

NEWS FEATURE | 15 April 2025 | Clarification [22 April 2025](#)

Exclusive: the most-cited papers of the twenty-first century

A Nature analysis reveals the 25 highest-cited papers published this century and explores why they are breaking records.

密歇根大学社会学家 Misha Teplitskiy 指出：“科学家们重视方法、理论和实证发现，但在实践中，方法类论文获得的引用更多”。21 世纪被引用次数最多的十篇论文，请见如下截图。

TOP TEN MOST-CITED WORKS OF THE TWENTY-FIRST CENTURY

Rank (median)	Citation	Times cited (range across databases)
1	Deep residual learning for image recognition (2016, preprint 2015)	103,756–254,074
2	Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\Delta\Delta C_t}$ method (2001)	149,953–185,480
3	Using thematic analysis in psychology (2006)	100,327–230,391
4	<i>Diagnostic and Statistical Manual of Mental Disorders, DSM-5</i> (2013)	98,312–367,800
5	A short history of SHELX (2007)	76,523–99,470
6	Random forests (2001)	31,809–146,508
7	Attention is all you need (2017)	56,201–150,832
8	ImageNet classification with deep convolutional neural networks (2017)	46,860–137,997
9	Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries (2020)	75,634–99,390
10	Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries (2016)	66,844–93,433

AI 论文引领潮流

21 世纪被引用次数最高的论文是微软研究人员 2016 年发表的关于“深度残差学习”的研究。该论文提出了 ResNet 架构，解决了深度神经网络中信号在多层传递时衰减的问题，为后来的 AlphaGo、AlphaFold 和 ChatGPT 等 AI 工具奠定了基础。

在 21 世纪被引用次数最多的十篇论文中，有 4 篇与 AI 相关。除 ResNet 外，还包括：2012 年引发深度学习革命的 AlexNet 论文；2017 年提出 Transformer 架构的论文，该技术成为大语言模型的核心；2001 年提出的机器学习算法“随机森林”，因其开源、免费、易用性被广泛引用。

研究软件和实验方法备受青睐

排名第二的是 2001 年发表的一篇利用 qPCR 分析相关

基因表达数据的方法学论文。该论文提供了一种简单的方法让生物学家计算不同条件下的基因活性变化，例如在药物治疗前后。

其他高被引的软件和方法论文包括：

- 关于 SHELX 晶体结构分析软件的论文（第 5 位）
- 描述 scikit-learn 机器学习库的论文（第 15 位）

改进研究质量的工具受欢迎

排在第三位的是心理学家 Virginia Braun 和 Victoria Clarke 发表的关于“主题分析法”的论文，该论文提供了一个可靠的定性研究方法框架。

同样，2009 年发表的 PRISMA 声明（系统综述和元分析首选报告项目）也跻身榜单（第 11 位）。PRISMA 为研究人员提供了 27 项清单，指导如何规范报告系统综述和元分析研究。

癌症与健康研究

癌症与健康领域的研究论文在榜单上占据重要位置。世界卫生组织 GLOBOCAN 项目发布的全球癌症统计报告（2018 年和 2020 年版）分别排在第 9 和第 10 位，这些报告为研究人员、倡导者和政策制定者提供了癌症发病率和死亡率的关键数据。

被称为“精神病学圣经”的《精神疾病诊断与统计手册》第五版（DSM-5）则位居第 4 位。它描述了包括成瘾和抑郁症在内的精神障碍分类和诊断的标准，被世界各地的研究人员和卫生专业人员广泛使用。

2.6 Nature 报道：数百篇论文未披露 AI 使用

近日，科学记者 Diana Kwon 在 Nature 杂志官网新闻版面撰文“Science sleuths flag hundreds of papers that use AI without disclosing it”，报道了令人担忧的学术现象：数百篇学术论文在使用人工智能（AI）工具的情况下，未按规定进行披露。一些期刊悄无声息地删除了这些 AI 使用痕迹，未留下任何修改记录。

[nature](#) > [news](#) > [article](#)

NEWS | 24 April 2025

Science sleuths flag hundreds of papers that use AI without disclosing it

Telltale signs of chatbot use are scattered through the scholarly literature – and, in some cases, have disappeared without a trace.

By [Diana Kwon](#)

AI 写作痕迹无所遁形

“重新生成回答” “作为一个 AI 语言模型” ——这些是

研究人员使用 AI 的明显迹象,已被科学侦探们在学术文献中发现。

ChatGPT 等生成式 AI 工具已经迅速改变了学术出版生态。科学家们越来越多地使用这些工具来撰写稿件,出版商也纷纷制定规范其伦理使用的指南。虽然政策各异,但许多出版商要求作者披露在科学论文撰写过程中使用 AI 的情况。

然而,科学侦探们已识别出数百篇疑似使用了 AI 工具却未披露的论文。某些论文被悄悄修正,即标志性的 AI 短语被删除,却未作任何说明。一些研究人员认为,这种不透明的修改对科研诚信构成了潜在威胁。

“我是一个 AI 语言模型”

自 2023 年以来,一些专家已标记出明显存在未披露 AI 使用迹象的论文,比如包含“重新生成回答”短语的论文。当用户希望得到新答案时,一些基于大语言模型的聊天机器人会生成这样的短语。

肯塔基州路易斯维尔大学的研究者 Alex Glynn 发现了数百篇带有 AI 使用特征的论文。他创建了 Academ-AI 在线追踪系统来记录这些案例,目前已列出 700 多篇论文。在对首批 500 篇被标记论文的分析中,Glynn 发现其中 13%的论

文发表在 Elsevier、Springer Nature 和 MDPI 等大型出版商的期刊上。

波兰卡特维兹经济大学的研究员 Artur Strzelecki 也收集了论文中未披露的 AI 使用案例，重点关注知名期刊。他发现了 64 篇论文，这些论文发表在被 Scopus 数据库分类为其所在领域前 25% 的期刊上。

出版商的回应

Nature 杂志新闻团队联系了几家被 Glynn 和 Strzelecki 标记的出版商。所有出版商均表示，被标记的论文正在调查中。他们还指出了各自的 AI 政策——在某些情况下，这些政策不要求披露 AI 使用情况，或者仅要求披露某些特定用途。例如，Springer 表示，AI 辅助文本编辑，包括针对可读性、风格以及语法或拼写错误所做的修改，无需披露。

英国 IOP 出版社指出，虽然该出版社在 2023 年要求作者必须声明 AI 使用，但去年已修改政策：鼓励作者披露 AI 使用，但不再强制要求。IOP 出版社正专注于通过自动化与人工检查相结合来确保内容的准确性，而不是完全禁用 AI。IOP 政策明确禁止使用 AI 创建、更改或操纵研究数据或结果。

“隐形修正”的问题

Glynn 表示，对于未披露 AI 使用的论文，出版商通常不

会发布正式修正或撤稿。删除 AI 特征的短语不一定需要勘误表，但应该有说明提及这一变更。Glynn 认为，不说明这些变更会带来问题，因为 AI 短语的存在表明论文中可能存在其他问题。

Glynn 与法国图卢兹大学计算机科学家 Guillaume Cabanac 及其同事追踪了隐形修正，他们表示这些修正威胁到科研诚信，因为透明的记录保存对确保已发表内容的可信度至关重要。Cabanac 表示，出版商需要迅速采取行动来解决不诚实使用 AI 的问题，既要剔除现有的相关论文，也要采取措施来筛查新投稿的论文。

2.7 Nature 发文：AI 生成的文献综述可能威胁科学进步

在科研领域，人工智能（AI）的应用正以前所未有的速度改变着研究方式。像 OpenAI 的“deep research”这类工具，为研究人员提供了快速编撰文献综述的可能性。然而，这一看似便利的技术背后，却潜藏着可能威胁科学进步的隐患。近日，Nature 杂志在通信栏目发文“AI-generated literature reviews threaten scientific progress”，概述了 AI 生成文献综述的双重隐患。

[nature](#) > [correspondence](#) > [article](#)

CORRESPONDENCE | 20 May 2025

AI-generated literature reviews threaten scientific progress

By [Xusen Cheng](#) & [Lulu Zhang](#)

通讯作者：程絮森，中国人民大学信息学院教授。研究方向：生成式人工智能与数字经济、AI 与人机团队协作、大数据与 AI 行为、信息管理与数字治理等。

AI 文献综述的双重隐患

AI 生成的文献综述至少存在两大隐患。首先，若缺少真正理解领域复杂性所需的认知努力，科学家们可能会丧失对自己学科内部“地图”的掌握能力。文献综述的过程涉及组织和分析前人研究成果，这包括发现研究中的矛盾之处、识别方法论的缺陷，以及追踪思想的演变。这一过程虽然耗时且充满挑战，但深化了研究者对学科的理解。如果这一环节被 AI 取代，尤其对早期职业科学家而言，他们可能会被剥夺关键的训练机会。

其次，AI 生成的文献综述往往呈现出精美且完整的外观。表面上看似全面的 AI 综述可能会让研究者产生“一切都已了解”的错觉，从而减少质疑和进一步探究的动力。

人类认知的不可替代性

如果文献综述工作被外包给机器，我们可能会培养出一批能够高效检索信息但难以突破既有框架进行思考的研究者。真正的创新需要的不仅仅是数据汇编，更需要杂乱、耗时的认知工作，从而将信息转化为原创见解。

在科学研究中，研究者需要通过深入思考来发现新问题、提出新假设、构建新理论。这种能力建立在对现有知识的批判性思考和创造性重组基础上，而非简单的信息检索和整合。当研究者亲自梳理文献时，他们不仅是在收集信息，更是在与前人的思想对话，这个过程本身就可能激发新的见解。

科学研究依赖于人类的这种认知能力，目前没有任何算法可以复制这种能力。尽管 AI 工具在处理大量信息方面表现出色，但它们缺乏真正的理解力和创造力。AI 可以分析模式、总结信息，但无法理解研究背后的动机、直觉和灵感。

在拥抱新技术的同时，我们更应该思考如何在保持人类独特认知优势的基础上，合理利用 AI 工具的辅助功能。这一平衡需要我们对 AI 工具的作用和局限性保持清醒认识。

信息 快报

INFORMATION EXPRESS

  <http://bighealth.fudan.edu.cn/>
 021-54237485
 上海市徐汇区东安路130号



扫码关注
随时掌握泌尿外科专业资讯