

# 提高动物实验系统评价/Meta 分析的证据确定性： GRADE 方法的实证研究

李腾飞<sup>1,2,3</sup>, 郑卿勇<sup>1,2</sup>, 许建国<sup>1,2</sup>, 李艺羿<sup>1,2,3</sup>, 周泳佳<sup>1,2,3</sup>, 徐彩花<sup>1,2</sup>, 张明悦<sup>1,2</sup>, 田杰祥<sup>4</sup>, 王 钢<sup>4</sup>, 田金徽<sup>1,2</sup>

(1. 兰州大学循证医学中心, 兰州大学基础医学院, 兰州 730000; 2. 甘肃省循证医学重点实验室, 兰州 730000; 3. 甘肃中医药大学护理学院, 兰州 730000; 4. 甘肃中医药大学附属医院风湿免疫科, 兰州 730000)

**[摘要]** 动物实验是生物医学研究中的重要手段, 是连接基础研究与临床试验的桥梁。动物实验的系统评价/Meta 分析 (systematic review/meta-analysis, SRs/MAs) 是整合动物实验证据的重要手段, 能够促进成果向临床研究转化, 降低转化风险, 并推动基础研究的资源整合。随着证据推荐分级的评估、制订与评价 (grading of recommendations assessment, development and evaluation, GRADE) 方法的不断发展, 其在动物实验 SRs/MAs 中的应用受到了越来越多的关注。本文首先阐述了 GRADE 方法在动物实验 SRs/MAs 中的应用原理及具体应用类型, 包括定性描述的系统评价、Meta 分析及网状 Meta 分析; 接着深入分析了 GRADE 方法在实际应用中的误用情况, 主要包括未正确进行证据体分级、证据体分级不当、误用于定性系统评价、升降级过程记录与结果不一致, 以及误用于提供推荐意见; 此外, 还全面探讨了 GRADE 方法在动物实验 SRs/MAs 中的证据确信心升降级因素, 包括偏倚风险、间接性、不一致性、不精确性和发表偏倚对证据降级的影响, 以及大效应量和跨物种一致性对证据升级的作用; 最后, 针对上述问题提出了改进策略, 包括进一步研究与优化 GRADE 方法在动物实验 SRs/MAs 中的应用细节、制定符合动物实验研究特点的 SRs/MAs 报告规范, 以及加强研究人员在 GRADE 方法上的专业培训等。本文旨在通过提升动物实验 SRs/MAs 的证据质量, 增强其在临床决策中的可靠性, 促进动物实验研究成果更高效地转化为临床实践。

**[关键词]** 动物实验; 系统评价; Meta 分析; 推荐分级的评估、制订与评价; 证据确信心; 证据分级

**[中图分类号]** Q95-33; R-332 **[文献标志码]** A **[文章编号]** 1674-5817(2025)01-0101-11



## Improving the Certainty of Evidence in Animal Experiment Systematic Review/Meta-Analysis: An Empirical Study of the GRADE Method

LI Tengfei<sup>1,2,3</sup>, ZHENG Qingyong<sup>1,2</sup>, XU Jianguo<sup>1,2</sup>, LI Yiyi<sup>1,2,3</sup>, ZHOU Yongjia<sup>1,2,3</sup>, XU Caihua<sup>1,2</sup>, ZHANG Mingyue<sup>1,2</sup>, TIAN Jiexiang<sup>4</sup>, WANG Gang<sup>4</sup>, TIAN Jinhui<sup>1,2</sup>

(1. Evidence-based Medicine Center, School of Basic Medical Sciences, Lanzhou University, Lanzhou 730000, China; 2. Key Laboratory of Evidence-based Medicine of Gansu Province, Lanzhou 730000, China; 3. School of Nursing, Gansu University of Chinese Medicine, Lanzhou 730000, China; 4. Department of Rheumatology, The Affiliated Hospital of Gansu University of Chinese Medicine, Lanzhou 730000, China)

Correspondence to: WANG Gang, E-mail: 1256935844@qq.com;

TIAN Jinhui (ORCID: 0000-0002-3859-9587), E-mail: tjh996@163.com

**[基金项目]** 2023 年中央财政转移支付地方项目“中医药循证能力提升建设”(2101704)

**[第一作者]** 李腾飞(1998—), 男, 硕士研究生, 研究方向: 循证医学与医学信息学。E-mail: ltf980102@163.com。ORCID: 0009-0005-1168-3075;  
郑卿勇(1998—), 男, 博士研究生, 研究方向: 循证医学与医学信息学。E-mail: easonzz@foxmail.com。ORCID: 0000-0002-9480-0169

**[通信作者]** 王 钢(1965—), 男, 博士, 主任医师, 博士生导师, 研究方向: 中西医结合治疗自身免疫性疾病的临床与实验研究。E-mail: 1256935844@qq.com;  
田金徽(1978—), 男, 博士, 教授, 研究方向: 网状 Meta 分析与卫生技术评估。E-mail: tjh996@163.com。ORCID: 0000-0002-3859-9587

**[ABSTRACT]** Animal experiments are essential tools in biomedical research, serving as a bridge between basic research and clinical trials. Systematic reviews and meta-analyses (SRs/MAs) of animal experiments are crucial methods for integrating evidence from animal experiment, which can facilitate the translation of findings into clinical research, reduce translational risks, and promote resource integration in basic research. With the continuous development of the Grading of Recommendations, Assessment, Development, and Evaluation (GRADE) methodology, its application in SRs/MAs of animal experiments has gained increasing attention. This article first outlines the principles and specific applications of the GRADE methodology in SRs/MAs of animal experiments, including qualitative descriptive systematic reviews, meta-analyses, and network meta-analyses. It then deeply analyzes the misuse of the GRADE methodology in practice, including incorrect evidence grading, improper classification of evidence, misapplication in qualitative systematic reviews, inconsistencies between the documentation of the upgrading and downgrading process and results, and inappropriate use for making recommendations. Furthermore, this article comprehensively discusses the factors influencing the grading of evidence certainty in SRs/MAs of animal experiments, including the impact of bias risk, indirectness, inconsistency, imprecision, and publication bias on evidence downgrading, as well as the role of large effect sizes and cross-species consistency in evidence upgrading. Finally, in response to the issues discussed, improvement strategies are proposed, including further research and optimization of the GRADE methodology for SRs/MAs of animal experiments, the development of reporting guidelines tailored to the characteristics of SRs/MAs in animal experiment research, and enhanced professional training for researchers in the GRADE methodology. This article aims to improve the quality of evidence in SRs/MAs of animal experiments, strengthen their reliability in clinical decision-making, and promote the more efficient translation of findings from animal experiment research into clinical practice.

**[Key words]** Animal experiment; Systematic reviews; Meta-analyses; Grading of recommendations assessment, development and evaluation; Certainty of evidence; Evidence grading

动物实验是医学和生物学领域中知识探索和问题解决的重要工具<sup>[1]</sup>，即使用动物在受控的实验室条件下进行科学研究。作为实验动物科学的组成部分，动物实验对于临床试验和医学证据的转化具有不可替代的作用<sup>[2]</sup>。在医疗实践中，动物实验的系统评价/Meta分析（systematic reviews/meta-analyses, SRs/MAs）发挥着重要作用，它们不仅能够加速动物实验研究发现的临床应用，还能有效避免时间浪费<sup>[3-4]</sup>。特别是在缺乏人类研究证据的情况下，动物实验的SRs/MAs能够为临床决策和卫生政策制订提供支持<sup>[2]</sup>。然而，证据体的质量即确信度（certainty of evidence）对于医疗决策至关重要，因此对动物实验SRs/MAs的证据体质量进行分级评价是必要的。

证据推荐分级的评估、制订与评价（grading of recommendations assessment, development and evaluation, GRADE）方法是一种评估证据体质量和推荐意见强度的重要工具<sup>[5]</sup>。在动物实验SRs/MAs中应用GRADE方法，可以使得证据体分级更加合理，从而提高SRs/MAs结果的可靠性<sup>[6]</sup>。尽管已有研究对动物实验SRs/MAs中GRADE方法的应用进行了理论阐述，并提供了

实例分析<sup>[6-8]</sup>，但目前GRADE方法在动物实验SRs/MAs中的实际应用情况尚不清楚。

因此，本研究全面收集国内外已发表的使用GRADE方法进行证据体质量评估的动物实验SRs/MAs文献，评估GRADE方法在动物实验SRs/MAs中的应用现状、存在的问题以及证据体质量分级情况，以期为未来的动物实验研究提供有价值的参考。

## 1 GRADE方法的分级原理

GRADE方法提供了一个清晰的框架来评估证据体的质量，将证据体分为高、中、低、极低共4个等级。这种方法首先根据研究设计来评估证据体的初始质量，通常将无明显缺陷的随机对照试验（randomized controlled trial, RCT）的证据体质量定为高等级。证据体的初始质量受到研究的偏倚风险、结果的不一致性、不精确性、间接性以及发表偏倚的影响而降低<sup>[9-13]</sup>。将观察性研究的初始证据体质量通常定为低等级，大效应量、存在剂量-效应关系和负偏倚会提升其证据体质量<sup>[14]</sup>。图1呈现了使用GRADE方法进行证据体质量分级的过程。

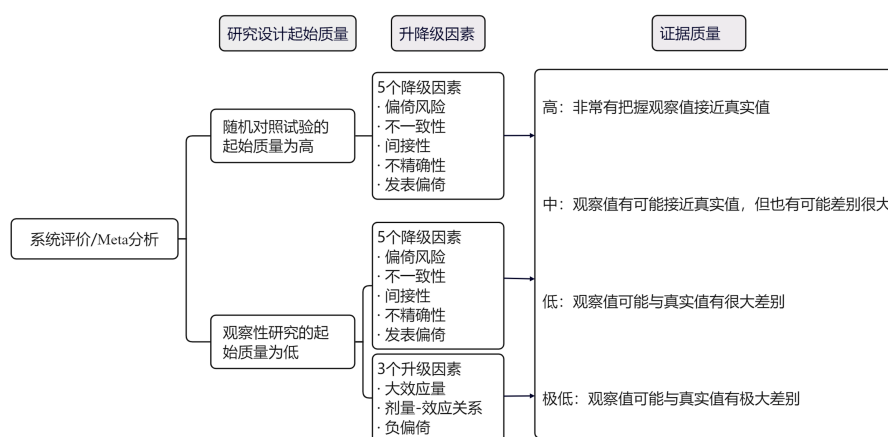


图1 使用GRADE方法进行证据体确信度分级

Figure 1 Using the GRADE methodology to grade the certainty of evidence

此外, GRADE方法还定义了推荐强度, 反映了推荐意见对目标人群潜在影响的不确定性。推荐强度分为强推荐和弱推荐<sup>[15]</sup>, 这一分类不仅基于证据体的质量, 还综合考虑了患者的偏好与价值观、利弊平衡和资源的可利用性<sup>[16-17]</sup>。GRADE方法综合考虑了研究设计、研究质量、结果一致性和其他相关因素, 以确定证据体的质量和推荐强度, 从而为临床实践和决策提供指导。

## 2 动物实验SRs/MAs中GRADE方法的应用类型

在动物实验SRs/MAs中, 数据合并策略通常分为3种主要类型: 定性系统评价、Meta分析以及网状Meta分析。对于不同数据分析方法的GRADE分级见图2。

### 2.1 基于定性描述的动物实验系统评价

使用评估定性系统评价证据可信度的工具(confidence in the evidence from reviews of qualitative research, CERQual)来评价证据质量。它从方法学局限性、相关性、一致性和数据充分性4个维度进行评价。方法学局限性关注原始研究的设计和和实施; 相关性评估研究目的和对象与系统评价问题的匹配度; 一致性检查合并结果与原始研究结果的一致性; 数据充分性评价支持系统评价结果的数据量和丰富性。基于这4个方面的评估, CERQual为动物实验SRs的每个结果提供高、中、低、极低的信度分级。

**实例分析:** Zhou等<sup>[18]</sup>的研究使用CERQual对定性证据进行了质量评估。结果显示, 由于纳入研究设计的局限性、结果的不一致性以及证据的不精确性,

定性证据的质量普遍较低, 影响了证据的可靠性和对临床实践的推荐强度。

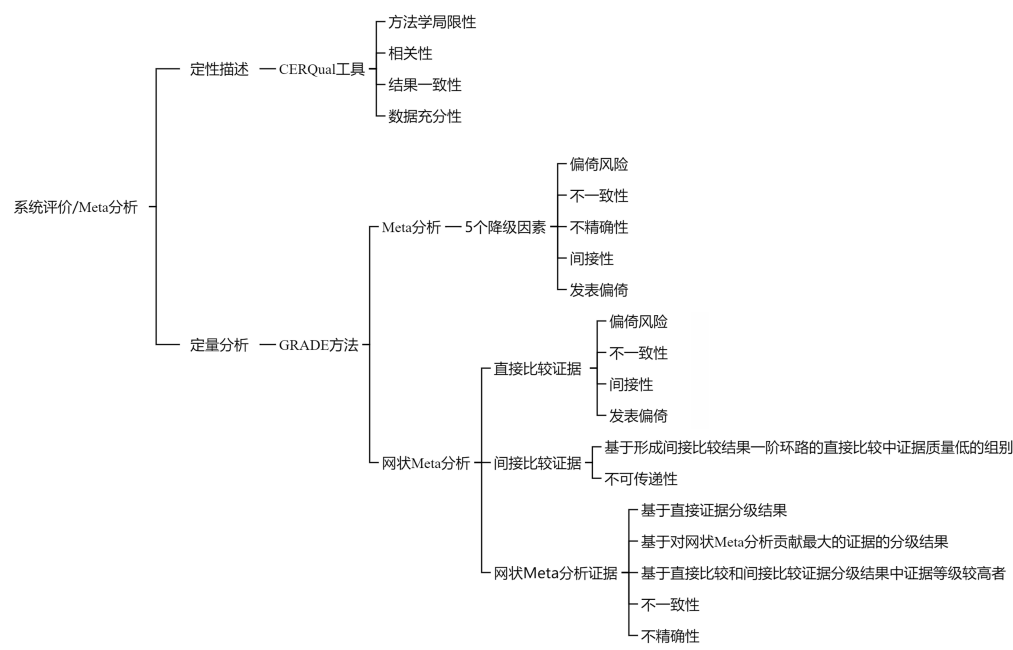
### 2.2 基于Meta分析的动物实验SRs/MAs

动物RCT的起始证据级别为“高”, 主要考虑偏倚风险、间接性、不一致性、不精确性和发表偏倚5个降级因素。使用荷兰动物实验系统评价研究中心(systematic review center for laboratory animal experimentation, SYRCLE)发布的SYRCLE动物实验偏倚风险评估工具评估动物研究的偏倚风险; 间接性从研究对象、干预措施、比较组、结果的角度进行评估; 不一致性通过考虑置信区间(confidence interval, CI)的重叠程度、各个纳入研究效应量的大小和方向、异质性检验的 $P$ 值和 $I^2$ 值进行评估; 精确性评估考虑样本是否达到最优信息样本量(optimal information size, OIS)以及可信区间的宽窄程度; 发表偏倚结合定性和定量方法全面考虑。动物观察性研究的起始证据级别为低, 但确定升级因素的方法仍需进一步研究。

**实例分析:** 何冰等<sup>[19]</sup>的研究使用GRADE方法对黄连制剂治疗大鼠代谢综合征的传统Meta分析结果进行了证据质量评级。结果显示, 由于纳入研究的方法学质量问题、样本量不足以及显著的统计学异质性, 大多数结果的证据质量被评为极低, Lee's指数的改善效果证据质量被评为低。这表明, 虽然黄连制剂显示出对代谢综合征大鼠模型的潜在疗效, 但由于证据质量的限制, 其结论的可靠性和临床应用的推荐强度受到限制, 需要更多高质量的研究来验证这些发现。

### 2.3 基于网状Meta分析的动物实验SRs/MAs

与Meta分析方法不同, 网状Meta分析方法可同时



注：CERQual，定性系统评价证据可信度；GRADE，推荐分级的评估、制订与评价。  
Note: CERQual, confidence in the evidence from reviews of qualitative research; GRADE, grading of recommendations assessment, development and evaluation.

图2 不同数据分析方法的GRADE分级  
Figure 2 GRADE for different data analysis methods

纳入直接比较和间接比较证据，使得分级过程相对复杂。GRADE工作组建议分四步对其进行证据质量分级：第一步，分别呈现直接比较和间接比较的效应量和可信区间。第二步，不考虑不精确性，对每一组直接比较和间接比较的证据质量分别进行分级。如果直接证据等级为高且对网状Meta分析结果的贡献大于等于间接证据，则网状Meta分析证据质量基于直接证据的质量等级；否则仍需对间接证据质量进行分级。第三步，基于形成间接证据一阶环路的直接证据质量等级，采取就低原则决定间接证据质量，并考虑不可传递性。第四步，基于直接证据和（或）间接证据等级，考虑不一致性和不精确性，最终确定网状Meta分析的证据质量。

**实例分析：**Zhou等<sup>[18]</sup>的研究使用GRADE方法对定量证据进行了质量评估。结果显示，由于大多数研究存在结果不一致、不精确性以及可能的选择偏倚和报告偏倚，整体证据质量被评为低。此外，大多数研究未能清晰报告基线特征，进一步影响了证据的可靠性。因此，尽管研究提供了关于不同可降解膜在引导骨/组织再生中的疗效的证据，但由于上述局限性，证据的整体质量较低，限制了对研究结果的解释和应用。

### 3 动物实验SRs/MAs中应用GRADE方法的文献特征分析

检索中国知网、万方、中国生物医学文献数据库、PubMed、Embase和Web of Science数据库，检索时间从各数据库建立至2024年4月1日。检索策略结合主题词（MeSH/Emtree）和关键词。中文检索词包括：“动物实验”“鼠”“兔”“狗”“犬”“蛙”“蟾蜍”“猴”“猿”“猪”“猩猩”“鱼”“果蝇”“系统评价”“Meta分析”等。英文检索词包括：“animal experimentation”“animal model”“mice”“rat”“rabbit”“dog”“frog”“toad”“monkey”“ape”“pig”“swine”“paniscus”“bonobo”“chimpanzee”“gorilla”“pongo”“fish”“fruit fly”“systematic review”“meta-analysis”等。经过筛选，共得到42篇使用GRADE方法的动物实验SRs/MAs的文献，其中英文文献39篇<sup>[18,20-57]</sup>，中文文献3篇<sup>[19,58-59]</sup>，卡方检验显示，中英文文献在应用GRADE方法学频率方面差异无统计学意义（ $\chi^2=0.23$ ， $P=0.63$ ）。

纳入动物实验SRs/MAs的基本特征：鼠、犬和猫是最常用的动物模型。根据国际疾病分类第十一次修



订本 (international classification of diseases-11, ICD-11), 消化系统疾病 (9篇, 21.43%) 和肌肉骨骼系统疾病 (7篇, 16.67%) 是 GRADE 方法应用较多的领域; 药物 (20篇, 47.62%) 是最常见的干预措施, 其次是手术 (7篇, 16.67%) 和饮食干预 (5篇, 11.90%); 暴露因素主要是化学物质和环境暴露 (3篇, 7.14%)。27篇 (64.29%) 文献采用了 Meta 分析方法, 2篇 (4.76%) 文献同时采用了 Meta 分析和定性描述, 13篇 (30.95%) 文献采用了定性描述; 14篇 (33.33%) 文献纳入了 1~10 个原始研究, 18篇 (42.86%) 文献纳入了 11~30 个, 10篇 (23.81%) 文献纳入了大于 30 个。

纳入动物实验 SRs/MAs 的证据确定性评级特点: 首次采用 GRADE 方法评估证据体确定性的文献发表于 2012 年。38篇 (90.48%) 文献采用了 GRADE 方法, 2篇 (4.76%) 文献采用了 GRADE 方法的改编版本, 2篇 (4.76%) 文献同时使用了 GRADE 和 CERQual 两种方法; 41篇 (97.62%) 文献将证据体质量分为高、中、低、极低共 4 个等级, 1篇 (2.38%) 文献将证据体质量分为高、中、低共 3 个等级; 19篇 (45.24%) 文献采用了证据概要表 (evidence profiles, EP), 8篇 (19.05%) 文献使用了结果总结表 (summary of findings, SoF), 15篇 (35.71%) 文献未使用这两种标准化模板。

## 4 动物实验 SRs/MAs 中误用 GRADE 方法的主要类型

### 4.1 未对证据体进行正确分级

8篇文献错误地将纳入的原始研究视为一个评分单元。根据 GRADE 方法, 证据体质量分级应针对每个结局指标单独进行, 这是其与传统证据分级系统的主要区别。即使 SRs/MAs 只纳入了一项研究但报告了多个不同的结局指标, 也应对每个结局指标分别进行质量分级。当存在多个结局指标时, GRADE 建议根据其临床重要性对这些指标进行优先级排序, 区分为“至关重要”和“重要”的指标。在这种情况下, 整体的证据体质量评级应依据“至关重要”结局中评级最低的证据来确定<sup>[60]</sup>。

### 4.2 证据体级别分类不当

1篇文献错误地将证据体分为高、中、低共 3 个等级。这与 GRADE 方法所采用的 4 个等级 (高、中、低、极低) 不符, 分类错误可能会对证据体质量和推荐强度的准确评估产生不利影响。

### 4.3 误用于定性系统评价

13篇文献错误地将 GRADE 方法应用于定性系统评价的证据质量评估。目前, GRADE 方法主要适用于定量 SRs/MAs 的分级, 并不适用于定性系统评价。CERQual 工具最早开发于 2010 年, 是由多个国际机构共同制定的定性系统评价分级系统<sup>[61-62]</sup>。CERQual 工具通过评估方法学上的局限性、相关性、结果的一致性以及数据的充分性 4 个维度来对定性系统评价进行证据质量评级。与 GRADE 方法相似, 二者都旨在衡量证据的可信度, 并且使用高、中、低、极低共 4 个等级来表示证据质量。未来研究需要进一步评估 CERQual 工具在动物实验 SRs/MAs 中的适用性和局限性。

### 4.4 升降级过程记录与结果不一致

3篇文献在升降级过程中的记录与其最终的评级结果不匹配, 这可能反映出评级过程缺乏透明度和详细记录。EP 和 SoF 是标准化的模板, 用于结构化地呈现证据分级结果<sup>[63]</sup>。为了提高证据质量结果的呈现质量, 建议采用 EP 和 SoF。此外, 研究者应基于研究设计设定的起点和升降级因素进行仔细审查, 以确保最终等级的准确性。

### 4.5 误用于提供推荐意见

2篇文献错误地将 GRADE 方法用作直接提供推荐意见的工具。虽然 GRADE 方法在多个研究领域都有应用, 但其在不同领域的应用方式和目的存在差异。在临床实践指南的制定中, GRADE 方法不仅评估证据的质量, 还进一步对基于这些证据的推荐意见进行强度分级以指导临床决策。在卫生技术评估中, GRADE 方法的应用重点在于证据质量的评估, 是否提供具体推荐意见取决于评估的具体目的和需求<sup>[64-65]</sup>。在动物实验 SRs/MAs 中, GRADE 方法主要用于证据质量分级, 而不是直接提供推荐意见。

## 5 动物实验 SRs/MAs 中 GRADE 方法的升降级因素

在排除了 19篇 (45.24%) GRADE 方法使用不当的文献后, 对剩余的 23篇 (54.76%) 文献进行深入分析以探讨升降级因素的特征。23篇文献共评估了 232 个证据体, 其中, 7个 (3.02%) 证据体为高质量; 36个 (15.52%) 证据体为中等质量; 71个 (30.60%) 证据体为低质量; 118个 (50.86%) 证据体为极低质量。在这些 SRs/MAs 中, 共记录了 4 个升级因素和 437 个降级因素。主要的升级因素包括大效应量 (2 个,

50.00%) 和跨物种一致性 (2 个, 50.00%)。主要的降级因素包括严重的偏倚风险 (161 个, 36.8%)、结果不精确 (127 个, 29.1%)、不一致性 (108 个, 24.7%) 和其他 (41 个, 9.4%)。

### 5.1 升级因素

在 232 个证据体中, 4 个 (1.72%) 证据体因特定的正面因素而被上调, 其中, 2 个 (50.00%) 证据体由于显示出大效应量; 2 个 (50.00%) 证据体因在多个物种中观察到一致的研究结果, 这些因素均增强了研究结论的普适性和可信度。

在 GRADE 方法中, 虽然观察性研究的初始证据确定性等级通常较低。但在发现大效应量或者跨物种动物实验显示出一致的效应方向和大小, 证据确定性等级可以得到提升。然而, 不同动物物种间结果的一致性如何被评估 (是作为升级因素还是作为不一致性或间接性的指标) 仍需在未来的研究中进一步探讨。

### 5.2 降级因素

#### 5.2.1 偏倚风险

在 232 个证据体中, 161 个 (69.40%) 证据体因存在严重的偏倚风险而被降级。而这 161 个证据体中, 61 个 (37.89%) 证据体因 SYRCLE 评分低于 8, 表明这些研究在设计上有严重缺陷; 18 个 (11.18%) 证据体因缺乏分配隐藏和盲法; 15 个 (9.32%) 证据体未透明报告随机序列产生和 (或) 盲法的实施; 12 个 (7.45%) 证据体因纳入的大多数研究存在偏倚; 12 个 (7.45%) 证据体因无随机或系统抽样、样本量选择依据不充分、选择性报告研究结果等原因; 43 个 (26.71%) 证据体未具体说明导致偏倚风险的原因。

SYRCLE 偏倚风险评估工具是专为动物实验设计的, 旨在评估研究的内在真实性。该工具包括了选择性偏倚、实施偏倚、测量偏倚、失访偏倚和报告偏倚多个维度<sup>[66-67]</sup>。SYRCLE 工具对每个条目的偏倚风险进行单独评估, 将风险分为低、高或不明确共 3 个等级, 而不是提供一个总体评分<sup>[66,68]</sup>。这种评估方法避免了为工具内各个要素之间分配固定“权重”, “权重”分配往往缺乏合理标准, 并且可能因研究结果和评审而异<sup>[66]</sup>。

选择性偏倚主要涉及“分配序列”“基线特征”“隐藏分组”3 个条目。由于动物实验通常样本量较小且随机分配并非动物实验研究中的标准实践模式, 基线特征的差异可能对实验结果产生显著影响<sup>[69]</sup>。因此, 未来的研究需要特别重视基线特征的均衡性, 尤

其是在未实施随机分配或样本量较小的实验中。

此外, 研究者应当更注重随机化方法和隐藏分组策略的应用, 以降低选择性偏倚。在动物实验中, 科研人员应加强对随机化安置和盲法实施的认知。饲养环境因素 (温度、湿度、光照和水质) 对实验结果有显著影响, 因此, 必须采用标准化程序, 确保实验动物饲养在统一的环境中, 以减少由环境引起的偏倚。盲法的实施能够有效减少干预和结果测量过程中的偏倚, 在评估主观性指标时, 盲法的实施尤为重要<sup>[69-70]</sup>。

#### 5.2.2 间接性

在 232 个证据体中, 29 个 (12.50%) 证据体因存在严重的间接性而被降级。间接性主要来源于研究对象选择、干预措施的限制和结果测量。其中, 人群外推 (37.93%) 是导致间接性的最常见原因; 其次是评估方法的差异性 (17.24%) 和干预措施的变异性 (6.90%); 其余的证据体 (37.93%) 未具体说明导致间接性降级的原因。

在评估动物实验证据的间接性时, 需要考虑 4 个主要问题: 首先是研究对象, 需确保使用的动物模型与所研究的疾病紧密相关, 并考虑不同模型因地理和资源条件的差异; 其次是干预或对照措施, 研究中的干预应与研究兴趣相符, 且生物或社会因素相似, 以减少间接性; 再次是结果的类型, 如果研究仅报告替代结果而非直接的临床结果, 那么, 其在临床实践中的应用可能受限; 最后是间接比较, 当缺乏直接比较时, 需通过共同对照进行间接推断<sup>[8]</sup>。

在评估研究对象的间接性时, 需考虑两种特殊情况: 当 SRs/MAs 的目标是为规划临床试验提供信息时, 如果研究包含了低级动物模型, 必须评估其从动物到人类的外推性差异, 在这种情况下, 使用非人类灵长类动物模型的研究通常不考虑间接性问题; 第二, 当研究结果直接用于指导临床决策或制定卫生政策时, 以低级动物模型为基础的研究被认为具有非常严重的间接性, 而以非人类灵长类动物为模型的研究则被认为具有严重的间接性<sup>[6]</sup> (图 3)。

#### 5.2.3 不一致性

在 232 个证据体中, 108 个证据体 (46.55%) 因不一致性而被降级, 主要原因是效果的不明确性 (69.44%), 表现为结果虽然倾向于支持或反对某种治疗, 但它们跨越了多个有效性类别或由于试验数量不足而无法评估一致性。此外, 15.74% 的证据体因高异

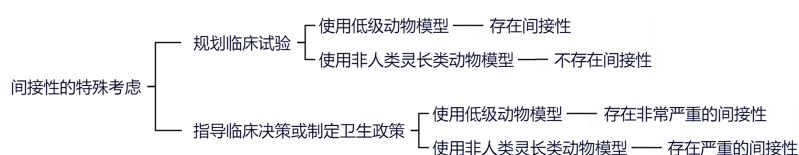


图3 研究对象间接性的特殊情况考虑

Figure 3 Special considerations for indirectness in the study population

质性或  $I^2 > 50\%$  而被降级；1.85% 的证据体因为结果的不一致性；0.93% 的证据体则因为 95%CI 较宽；13 个（12.04%）证据体未具体说明导致不一致性降级的原因。

评估动物实验中的不一致性包括分析 CI 的重叠程度、效应量的大小和方向、异质性检验的  $P$  值和  $I^2$  值。当研究存在无法合理解释的异质性时，建议降低证据的确定性等级；如果异质性可以通过考虑动物的种类、干预措施、比较措施或研究的偏倚风险等因素得到解释，研究者应该在 SRs/MAs 中提供或实施适当的亚组分析。未来的研究应深入分析和解释异质性并通过实施针对性的亚组分析来探究其潜在来源<sup>[8]</sup>。

### 5.2.4 不精确性

在 232 个证据体中，127 个（54.74%）证据体因不一致性而被降级，其中，57 个（44.88%）证据体显示出有限的一致性，表现为效果既不显著也不确定，甚至可能有负面效果；26 个（20.47%）证据体因小样本量和（或）95%CI 穿过零效应线；9 个（7.09%）证据体由于 95%CI 过宽；5 个（3.94%）证据体由于纳入研究的高异质性；30 个（23.62%）证据体未具体说明导致不精确性降级的原因。

证据体的精确性评估主要关注样本量是否达到 OIS 以及 CI 的宽窄程度。在动物实验中，样本量的计算可以参考 Charan 等<sup>[71]</sup> 在 2013 年提出的动物实验样本量计算公式。连续型数据的样本量计算公式为： $2 \times SD^2 \times (Z_{\alpha/2} + Z_{\beta})^2 / d^2$ ，式中， $SD$  为之前同类研究或预实验的标准差； $Z_{\alpha/2} = Z_{0.025} = 1.96$ ； $Z_{\beta} = Z_{0.20} = 0.842$ ； $d$  为两组均值的差值。二分类数据的样本量计算公式为： $2 \times (Z_{\alpha/2} + Z_{\beta})^2 \times P \times (1-P) / (p_1 - p_2)^2$ ，式中， $Z_{\alpha/2} = Z_{0.025} = 1.96$ ； $Z_{\beta} = Z_{0.20} = 0.842$ ； $p_1 - p_2$  为两组事件发生率的差值； $P$  为合并的发生率，即（干预组发生率+对照组发生率）/2。在解释动物实验结果时，效应量的方向通常比其大小更为重要。在评估精确性时，主要基于 CI 是否包含了无效值。此外，如果 CI 包含了两个或多个级

别，则可考虑降级<sup>[6]</sup>。当前面临的挑战包括开发和验证将不同实验单位纳入 OIS 计算的方法，探索如何将动物实验的效应量转化为临床上的决策阈值，以及对于未进行 Meta 分析的动物实验系统评价，如何更准确地描述和评价其结果的精确性。

### 5.2.5 发表偏倚

在 232 个证据体中，12 个（5.17%）证据体因发表偏倚而被降级，其中，6 个（50.00%）证据体是由于漏斗图的不对称性和 Egger 检验的显著性（ $P < 0.05$ ）；4 个（33.33%）证据体是由于小样本量；2 个（16.67%）证据体是由于漏斗图的不对称性。

发表偏倚在证据降级中出现的频率最低，其评估主要基于漏斗图和 Egger 检验。然而，这两种方法都可能会低估或高估发表偏倚<sup>[72]</sup>。因此，需要全面评估发表偏倚的风险，包括细致分析研究结果的分布以识别是否存在阳性结果的过度报告，尤其是小样本研究；审视资金来源和利益冲突的透明度，特别是由制药行业资助的研究；以及评估 SRs/MAs 检索策略的全面性。

## 6 总结及建议

尽管动物实验的 SRs/MAs 结果不是临床决策或卫生政策制定的最佳证据，但当人类面临研究数据不足的公共卫生紧急情况时，决策者可能会依赖动物研究的证据来制定政策。因此，探索如何在动物研究领域有效应用 GRADE 方法，对于提高动物研究证据的质量和适用性至关重要。未来研究应关注以下几个领域：需要进一步研究和优化 GRADE 方法在临床前动物实验 SRs/MAs 中的应用，包括确定 OIS 的计算方法，制定和规范间接性的标准和定义，以及确立升级标准的准则；开发和制定一套符合动物实验研究特性的 SRs/MAs 报告规范，并推动规范在期刊中的采纳<sup>[73]</sup>；确保研究人员在应用 GRADE 方法进行证据质量评级前接受方法论方面的专业培训，以提高评级的专业性和分级结果的透明度；建议推广 GRADE 方法的知识，并在动物实验



SRs/MAAs中加强其使用。

GRADE方法虽然为评估证据质量提供了一个结构化的框架,但在动物实验SRs/MAAs中的应用尚未广泛普及,并且在实施过程中也出现了一些误用情况。当前,采用GRADE方法进行评级的动物实验SRs/MAAs普遍显示出较低的证据质量,这可能限制了它们在临床决策中的应用。为了提升证据的质量和可靠性,未来的研究需要进行更严格的研究设计;相关机构应鼓励进一步的研究以强化GRADE方法的应用;加强研究者对GRADE方法的接受度和理解;在动物实验证据合成中建立和推广GRADE方法的知识体系。

#### [作者贡献 Author Contribution]

李腾飞负责撰写和修改论文;  
郑卿勇、许建国和李艺羿进行文献检索与数据提取;  
周泳佳、徐彩花、张明悦和田杰祥进行数据整理与图表绘制;  
王钢和田金徽构建论文框架,并指导文稿撰写及修改。

#### [利益声明 Declaration of Interest]

本文所有作者均声明不存在利益冲突。

#### [参考文献 References]

- [1] 赵冰. 中医药领域动物实验系统评价报告指南的构建研究[D]. 兰州: 兰州大学, 2022. DOI: 10.27204/d.cnki.glzhu.2022.001639.
- ZHAO B. Research on the establishment of reporting guidelines for systematic review of animal experiments in the field of traditional Chinese medicine[D]. Lanzhou: Lanzhou University, 2022. DOI: 10.27204/d.cnki.glzhu.2022.001639.
- [2] ROBERTS I, KWAN I, EVANS P, et al. Does animal experimentation inform human healthcare? Observations from a systematic review of international animal experiments on fluid resuscitation[J]. *BMJ*, 2002, 324(7335): 474-476. DOI: 10.1136/bmj.324.7335.474.
- [3] SANDERCOCK P, ROBERTS I. Systematic reviews of animal experiments[J]. *Lancet*, 2002, 360(9333): 586. DOI: 10.1016/S0140-6736(2)009812-4.
- [4] HOOIJMANS C R, RITSKES-HOITINGA M. Progress in using systematic reviews of animal studies to improve translational research[J]. *PLoS Med*, 2013, 10(7): e1001482. DOI: 10.1371/journal.pmed.1001482.
- [5] ATKINS D, ECCLES M, FLOTTORP S, et al. Systems for grading the quality of evidence and the strength of recommendations I: critical appraisal of existing approaches The GRADE Working Group[J]. *BMC Health Serv Res*, 2004, 4(1):38. DOI:10.1186/1472-6963-4-38.
- [6] WEI D, TANG K, WANG Q, et al. The use of GRADE approach in systematic reviews of animal studies[J]. *J Evid Based Med*, 2016, 9(2):98-104. DOI:10.1111/jebm.12198.
- [7] 赵冰, 姜彦彪, 尚志忠, 等. 动物实验系统评价中应用GRADE系统的实例解读[J]. *中国循证心血管医学杂志*, 2019, 11(11):1285-1287. DOI: 10.3969/j.issn.1674-4055.2019.11.02.
- ZHAO B, JIANG Y B, SHANG Z Z, et al. Interpretation on the application of GRADE approach in systematic review of animal studies[J]. *Chin J Evid Based Cardiovasc Med*, 2019, 11(11):1285-1287. DOI: 10.3969/j.issn.1674-4055.2019.11.02.
- [8] 张婷, 王欢, 邢丹, 等. GRADE在动物实验系统评价中的应用与挑战[J]. *中国循证医学杂志*, 2019, 19(2):248-252. DOI: 10.7507/1672-2531.201809068.
- ZHANG T, WANG H, XING D, et al. The application and challenge of GRADE in systematic reviews of animal studies[J]. *Chin J Evid Based Med*, 2019, 19(2):248-252. DOI: 10.7507/1672-2531.201809068.
- [9] GUYATT G H, OXMAN A D, VIST G, et al. GRADE guidelines: 4. Rating the quality of evidence—study limitations (risk of bias)[J]. *J Clin Epidemiol*, 2011, 64(4): 407-415. DOI: 10.1016/j.jclinepi.2010.07.017.
- [10] GUYATT G H, OXMAN A D, MONTORI V, et al. GRADE guidelines: 5. Rating the quality of evidence—publication bias[J]. *J Clin Epidemiol*, 2011, 64(12): 1277-1282. DOI: 10.1016/j.jclinepi.2011.01.011.
- [11] GUYATT G H, OXMAN A D, KUNZ R, et al. GRADE guidelines 6. Rating the quality of evidence—imprecision[J]. *J Clin Epidemiol*, 2011, 64(12): 1283-1293. DOI: 10.1016/j.jclinepi. 2011. 01.012.
- [12] GUYATT G H, OXMAN A D, KUNZ R, et al. GRADE guidelines: 7. Rating the quality of evidence—inconsistency[J]. *J Clin Epidemiol*, 2011, 64(12): 1294-1302. DOI: 10.1016/j.jclinepi. 2011. 03.017.
- [13] GUYATT G H, OXMAN A D, KUNZ R, et al. GRADE guidelines: 8. Rating the quality of evidence—indirectness[J]. *J Clin Epidemiol*, 2011, 64(12): 1303-1310. DOI: 10.1016/j.jclinepi. 2011. 04.014.
- [14] GUYATT G H, OXMAN A D, SULTAN S, et al. GRADE guidelines: 9. Rating up the quality of evidence[J]. *J Clin Epidemiol*, 2011, 64(12): 1311-1316. DOI: 10.1016/j.jclinepi. 2011. 06.004.
- [15] ANDREWS J C, SCHÜNEMANN H J, OXMAN A D, et al. GRADE guidelines: 15. Going from evidence to recommendation—determinants of a recommendation's direction and strength[J]. *J Clin Epidemiol*, 2013, 66(7): 726-735. DOI: 10.1016/j.jclinepi.2013.02.003.
- [16] ZHANG Y, COELLO P A, GUYATT G H, et al. GRADE guidelines: 20. Assessing the certainty of evidence in the importance of outcomes or values and preferences—inconsistency, imprecision, and other domains [J]. *J Clin Epidemiol*, 2019, 111: 83-93. DOI: 10.1016/j.jclinepi.2018.05.011.
- [17] ZHANG Y, ALONSO-COELLO P, GUYATT G H, et al. GRADE Guidelines: 19. Assessing the certainty of evidence in the importance of outcomes or values and preferences—Risk of bias and indirectness[J]. *J Clin Epidemiol*, 2019, 111: 94-104. DOI:10.1016/j.jclinepi.2018.01.013.
- [18] ZHOU Z L, YUN J H, LI J, et al. Comparison of the efficacy of different biodegradable membranes in guided bone/tissue regeneration: a systematic review and network meta-analysis [J]. *Biomed Mater*, 2023, 18(3). 1088/1748-605X/acc99a DOI: 10.1088/1748-605X/acc99a.
- [19] 何冰, 林建鹏, 陈冠霖, 等. 黄连制剂治疗大鼠代谢综合征的



- Meta 分析[J]. 中国医学创新, 2021, 18(18): 165-171. DOI: 10.3969/j.issn.1674-4985.2021.18.041.
- HE B, LIN J P, CHEN G L, et al. A meta-analysis of Huanglian preparations in the treatment of metabolic syndrome in rats [J]. Med Innov China, 2021, 18(18):165-171. DOI: 10.3969/j.issn.1674-4985.2021.18.041.
- [20] CORDELLI E, ARDOINO L, BENASSI B, et al. Effects of Radiofrequency Electromagnetic Field (RF-EMF) exposure on pregnancy and birth outcomes: A systematic review of experimental studies on non-human mammals[J]. Environ Int, 2023, 180: 108178. DOI: 10.1016/j.envint.2023.108178.
- [21] KHATIB M N, SHANKAR A, KIRUBAKARAN R, et al. Effect of ghrelin on mortality and cardiovascular outcomes in experimental rat and mice models of heart failure: a systematic review and meta-analysis[J]. PLoS One, 2015, 10 (5): e0126697. DOI: 10.1371/journal.pone.0126697.
- [22] LIN F, CHEN J R, CHEN M M, et al. Protective effect and possible mechanisms of resveratrol in animal models of osteoporosis: A preclinical systematic review and meta-analysis[J]. Phytother Res, 2023, 37(11): 5223-5242. DOI: 10.1002/ptr.7954.
- [23] PARCIANELLO R G, AMERIO E, GINER TARRIDA L, et al. Local hormones and growth factors to enhance orthodontic tooth movement: A systematic review of animal studies[J]. Orthod Craniofac Res, 2022, 25(3): 281-303. DOI: 10.1111/ocr.12544.
- [24] PATIKORN C, NERAPUSEE O, SOONTORNVIPART K, et al. Efficacy and safety of cannabidiol for the treatment of canine osteoarthritis: a systematic review and meta-analysis of animal intervention studies[J]. Front Vet Sci, 2023, 10: 1248417. DOI: 10.3389/fvets.2023.1248417.
- [25] QIN C, GUO Y, YANG D G, et al. Induced pluripotent stem cell transplantation improves locomotor recovery in rat models of spinal cord injury: a systematic review and meta-analysis of randomized controlled trials[J]. Cell Physiol Biochem, 2018, 47(5):1835-1852. DOI: 10.1159/000491064.
- [26] SCAHILL K, JESSEN L R, PRIOR C, et al. Efficacy of antimicrobial and nutraceutical treatment for canine acute diarrhoea: A systematic review and meta-analysis for European Network for Optimization of Antimicrobial Therapy (ENOVAT) guidelines[J]. Vet J, 2024, 303: 106054. DOI: 10.1016/j.tvjl.2023.106054.
- [27] SILVEIRA F M, SCHUCH L F, SCHIMIDT T R, et al. Potentially carcinogenic effects of hydrogen peroxide for tooth bleaching on the oral mucosa: A systematic review and meta-analysis[J]. J Prosthet Dent, 2024, 131(3):375-383. DOI: 10.1016/j.prosdent.2022.02.001.
- [28] XIANG N Y, LIN Y F, SU X Y, et al. Assessing the application of barbed sutures in comparison to conventional sutures for surgical applications: a global systematic review and meta-analysis of preclinical animal studies[J]. Int J Surg, 2024, 110 (5): 3060-3071. DOI: 10.1097/JS9.0000000000001230.
- [29] YE X J, GU Y, BAI Y J, et al. Does low-magnitude high-frequency vibration (LMHFV) worth for clinical trial on dental implant? A systematic review and meta-analysis on animal studies[J]. Front Bioeng Biotechnol, 2021, 9: 626892. DOI: 10.3389/fbioe.2021.626892.
- [30] ZHAO Y, ZHANG Y Z, MENG S J, et al. Effects of S-adenosylmethionine on cognition in animals and humans: a systematic review and meta-analysis of randomized controlled trials[J]. J Alzheimers Dis, 2023, 94(s1): S267-S287. DOI: 10.3233/JAD-221076.
- [31] WILHELM B, RAJIĆ A, PARKER S, et al. Assessment of the efficacy and quality of evidence for five on-farm interventions for *Salmonella* reduction in grow-finish swine: a systematic review and meta-analysis[J]. Prev Vet Med, 2012, 107(1-2): 1-20. DOI: 10.1016/j.prevetmed.2012.07.011.
- [32] YANG J W, KANG Y H, ZHAO W L, et al. Evaluation of patches for rotator cuff repair: A systematic review and meta-analysis based on animal studies[J]. Bioact Mater, 2022, 10: 474-491. DOI: 10.1016/j.bioactmat.2021.08.016.
- [33] SHOJAEI-ZARGHANI S, FATTAHI M R, KAZEMI A, et al. Effects of garlic and its major bioactive components on non-alcoholic fatty liver disease: A systematic review and meta-analysis of animal studies[J]. J Funct Foods, 2022, 96: 105206. DOI: 10.1016/j.jff.2022.105206.
- [34] ARAGOSA F, CATERINO C, DELLA VALLE G, et al. Tibial tuberosity advancement techniques (TTAT): A systematic review[J]. Animals, 2022, 12(16): 2114. DOI: 10.3390/ani12162114.
- [35] BUDSBERG S C, TORRES B T, SANDBERG G S. Efficacy of postoperative antibiotic use after tibial plateau leveling osteotomy in dogs: A systematic review[J]. Vet Surg, 2021, 50 (4): 729-739. DOI: 10.1111/vsu.13603.
- [36] DOLAN E, SWINTON P A, PAINELLI V S, et al. A systematic risk assessment and meta-analysis on the use of oral  $\beta$ -alanine supplementation[J]. Adv Nutr, 2019, 10(3): 452-463. DOI:10.1093/advances/nmy115.
- [37] NGUYEN T, PURCELL E, SMITH M J, et al. Umbilical cord blood-derived cell therapy for perinatal brain injury: A systematic review & meta-analysis of preclinical studies[J]. Int J Mol Sci, 2023, 24(5): 4351. DOI: 10.3390/ijms24054351.
- [38] PEIXOTO H M, DE OLIVEIRA M R F, ROMERO G A S. Serological diagnosis of canine visceral leishmaniasis in Brazil: systematic review and meta-analysis[J]. Trop Med Int Health, 2015, 20(3): 334-352. DOI: 10.1111/tmi.12429.
- [39] REEVES L A, ANDERSON K M, LUTHER J K, et al. Treatment of idiopathic chylothorax in dogs and cats: a systematic review[J]. Vet Surg, 2020, 49(1):70-79. DOI:10.1111/vsu.13322.
- [40] ROSSETTI B R, LEMOS A F, SÓ B B, et al. Effects of estrogen deficiency on the progression of apical periodontitis. A systematic review of preclinical studies[J]. Arch Oral Biol, 2022, 142: 105496. DOI: 10.1016/j.archoralbio.2022.105496.
- [41] SCHEUERMANN L M, GORDON-EVANS W J, NAULT A J. Systematic review of the treatment options for pericardial effusions in dogs[J]. Vet Surg, 2021, 50(1): 20-28. DOI: 10.1111/vsu.13475.
- [42] SELMIC L E, RUPLE A. A systematic review of surgical margins utilized for removal of cutaneous mast cell tumors in dogs[J]. BMC Vet Res, 2020, 16(1): 5. DOI: 10.1186/s12917-019-2227-8.
- [43] CHISARI E, WOUTHUYZEN-BAKKER M, FRIEDRICH A W, et al. The relation between the gut microbiome and osteoarthritis:

- A systematic review of literature[J]. PLoS One, 2021, 16(12): e0261353. DOI: 10.1371/journal.pone.0261353.
- [44] CHANG T Y, WANG H A, WANG Y P, et al. Efficacy of traditional Chinese medicine on animal model of IgA nephropathy: A systematic review and meta-analysis[J]. Evid Based Complement Alternat Med, 2022, 2022: 6106993. DOI: 10.1155/2022/6106993.
- [45] DONATI P, TARDUCCI A, ZANATTA R, et al. Angiotensin-converting enzyme inhibitors in preclinical myxomatous mitral valve disease in dogs: systematic review and meta-analysis[J]. J Small Anim Pract, 2022, 63(5): 362-371. DOI: 10.1111/jsap.13461.
- [46] DONATI P A, TARRAGONA L, FRANCO J V A, et al. Efficacy of tramadol for postoperative pain management in dogs: systematic review and meta-analysis[J]. Vet Anaesth Analg, 2021, 48(3): 283-296. DOI: 10.1016/j.vaa.2021.01.003.
- [47] JABBAR-LOPEZ Z K, UNG C Y, ALEXANDER H, et al. The effect of water hardness on atopic eczema, skin barrier function: A systematic review, meta-analysis[J]. Clin Exp Allergy, 2021, 51(3): 430-451. DOI: 10.1111/cea.13797.
- [48] TANG S Q, RUAN Z, MA A X, et al. Effect of vitamin K on wound healing: A systematic review and meta-analysis based on preclinical studies[J]. Front Pharmacol, 2022, 13: 1063349. DOI: 10.3389/fphar.2022.1063349.
- [49] DOMÍNGUEZ-OLIVA A, MOTA-ROJAS D, SEMENDRICH I, et al. The impact of vegan diets on indicators of health in dogs and cats: A systematic review[J]. Vet Sci, 2023, 10(1): 52. DOI: 10.3390/vetsci10010052.
- [50] ELMA Ö, LEBUF E, MARNEF A Q, et al. Diet can exert both analgesic and pronociceptive effects in acute and chronic pain models: A systematic review of preclinical studies[J]. Nutr Neurosci, 2022, 25(10): 2195-2217. DOI: 10.1080/1028415X.2021.1934956.
- [51] WHITTAKER A L, BARKER T H. The impact of common recovery blood sampling methods, in mice (mus Musculus), on well-being and sample quality: a systematic review[J]. Animals, 2020, 10(6): 989. DOI: 10.3390/ani10060989.
- [52] PASKAL W, GOTOWIEC M, STACHURA A, et al. VEGF and other gene therapies improve flap survival-a systematic review and meta-analysis of preclinical studies[J]. Int J Mol Sci, 2024, 25(5): 2622. DOI: 10.3390/ijms25052622.
- [53] DINI C, NAGAY B E, MAGNO M B, et al. Photofunctionalization as a suitable approach to improve the osseointegration of implants in animal models-a systematic review and meta-analysis[J]. Clin Oral Implants Res, 2020, 31(9): 785-802. DOI: 10.1111/clr.13627.
- [54] KATSAMAKIS S, SLOT D E, VAN DER SLUIS L W M, et al. Histological responses of the periodontium to MTA: a systematic review[J]. J Clin Periodontol, 2013, 40(4): 334-344. DOI: 10.1111/jcpe.12058.
- [55] MATTHEWS J J, DOLAN E, SWINTON P A, et al. Effect of carnosine or  $\beta$ -alanine supplementation on markers of glycemic control and insulin resistance in humans and animals: A systematic review and meta-analysis[J]. Adv Nutr, 2021, 12(6): 2216-2231. DOI: 10.1093/advances/nmab087.
- [56] PEREIRA M B, SYDOR B G, MEMARE K G, et al. *In vivo* efficacy of meglumine antimoniate-loaded nanoparticles for cutaneous leishmaniasis: a systematic review[J]. Nanomedicine, 2021, 16(17): 1505-1518. DOI: 10.2217/nnm-2021-0119.
- [57] WEISSHEIMER T, BISCHOFF K F, MICHEL C H T, et al. Does photobiomodulation on the root surface decrease the occurrence of root resorption in reimplanted teeth? A systematic review of animal studies[J]. Restor Dent Endod, 2023, 48(3): e24. DOI: 10.5395/rde.2023.48.e24.
- [58] 吴倩妮, 田宏亮, 张丽茜, 等. 丹参对大鼠肝再生过程影响的 Meta 分析[J]. 中国中药杂志, 2012, 37(17): 2630-2634. DOI: 10.4268/cjcmm20121726.
- WU Q N, TIAN H L, ZHANG L H, et al. Meta-analysis on impact of Danshen on liver regeneration in rats[J]. China J Chin Mater Med, 2012, 37(17): 2630-2634. DOI: 10.4268/cjcmm20121726.
- [59] 马红, 钟宏婧. 高脂饮食与小鼠/大鼠肠道微生态结构改变相关性的 Meta 分析[J]. 循证医学, 2019, 19(2): 92-101. DOI: 10.12019/j.issn.1671-5144.2019.02.010.
- MA H, ZHONG H J. The meta analysis of the relationship between high fat diet and the changes of intestinal microecology in mice/rats[J]. J Evid Based Med, 2019, 19(2): 92-101. DOI: 10.12019/j.issn.1671-5144.2019.02.010.
- [60] 陈耀龙, 姚亮, NORRIS S, 等. GRADE 在系统评价中应用的必要性及注意事项[J]. 中国循证医学杂志, 2013, 13(12): 1401-1404. DOI: 10.7507/1672-2531.20130240.
- CHEN Y L, YAO L, NORRIS S, et al. Application of GRADE in systematic reviews: necessity, frequently-asked questions and concerns[J]. Chin J Evid Based Med, 2013, 13(12): 1401-1404. DOI: 10.7507/1672-2531.20130240.
- [61] LEWIN S, GLENTON C, MUNTHER-KAAS H, et al. Using qualitative evidence in decision making for health and social interventions: an approach to assess confidence in findings from qualitative evidence syntheses (GRADE-CERQual) [J]. PLoS Med, 2015, 12(10): e1001895. DOI: 10.1371/journal.pmed.1001895.
- [62] 拜争刚, 刘少堃, 黄崇斐, 等. 定性系统评价证据分级工具: CERQual 简介[J]. 中国循证医学杂志, 2015, 15(12): 1465-1470. DOI: 10.7507/1672-2531.20150238.
- BAI Z G, LIU S K, HUANG C F, et al. An introduction of quality classification tool for qualitative evidence: CERQual[J]. Chin J Evid Based Med, 2015, 15(12): 1465-1470. DOI: 10.7507/1672-2531.20150238.
- [63] GUYATT G H, THORLUND K, OXMAN A D, et al. GRADE guidelines: 13. Preparing summary of findings tables and evidence profiles-continuous outcomes[J]. J Clin Epidemiol, 2013, 66(2): 173-183. DOI: 10.1016/j.jclinepi.2012.08.001.
- [64] 杨珂璐, 葛龙, 刘明, 等. GRADE 在卫生技术评估中的应用[J]. 中国药物评价, 2020, 37(6): 411-416. DOI: 10.3969/j.issn.2095-3593.2020.06.003.
- YANG K L, GE L, LIU M, et al. Application of GRADE in health technology assessment[J]. Chin J Drug Eval, 2020, 37(6): 411-416. DOI: 10.3969/j.issn.2095-3593.2020.06.003.
- [65] 杨珂璐, 胡志伟, 刘明, 等. GRADE 在卫生技术评估中应用的实例解析[J]. 中国药物评价, 2021, 38(1): 8-13. DOI: 10.3969/j.issn.2095-3593.2021.01.002.

- YANG K L, HU Z W, LIU M, et al. Application of GRADE in health technology assessment: a case analysis[J]. Chin J Drug Eval, 2021, 38(1):8-13. DOI: 10.3969/j.issn.2095-3593.2021.01.002.
- [66] HOOIJMANS C R, ROVERS M M, DE VRIES R B M, et al. SYRCLE's risk of bias tool for animal studies[J]. BMC Med Res Methodol, 2014, 14: 43. DOI: 10.1186/1471-2288-14-43.
- [67] 陶功财, 张楠, 尚志忠, 等. 评估动物实验偏倚风险的SYRCLE工具实例解读[J]. 中国循证心血管医学杂志, 2019, 11(3):292-295, 300. DOI: 10.3969/j.issn.1674-4055.2019.03.06.
- TAO G C, ZHANG N, SHANG Z Z, et al. Interpretation on examples of SYRCLE' tool for interviewing risk of bias in animal experimentation[J]. Chin J Evid Based Cardiovasc Med, 2019, 11(3): 292-295, 300. DOI: 10.3969/j.issn.1674-4055.2019.03.06.
- [68] 陈匡阳, 马彬, 王亚楠, 等. SYRCLE 动物实验偏倚风险评估工具简介[J]. 中国循证医学杂志, 2014, 14(10): 1281-1285. DOI: 10.7507/1672-2531.20140206.
- CHEN K Y, MA B, WANG Y N, et al. SYRCLE's risk of bias tool for animal studies[J]. Chin J Evid Based Med, 2014, 14(10): 1281-1285. DOI: 10.7507/1672-2531.20140206.
- [69] 张维益, 李艳飞, 戴岩瑞, 等. SYRCLE 偏倚风险评估工具发布是否提高动物实验方法质量?[J]. 中国循证心血管医学杂志, 2019, 11(10):1165-1168, 1173. DOI: 10.3969/j.issn.1674-4055.2019.10.04.
- ZHANG W Y, LI Y F, DAI Y R, et al. Does the publication of SYRCLE's risk of bias tool improve the methodological quality of animal experiments? [J]. Chin J Evid Based Cardiovasc Med, 2019, 11(10): 1165-1168, 1173. DOI: 10.3969/j.issn.1674-4055.2019.10.04.
- [70] 牛牧町, 赵跃, 易少威, 等. 动物实验的研究计划指南——PREPARE 解读[J]. 兰州大学学报(医学版), 2024, 50(2):55-59. DOI: 10.13885/j.issn.1000-2812.2024.02.009.
- NIU M T, ZHAO Y, YI S W, et al. Guidelines for planning animal experiments for research: a PREPARE interpretation[J]. J Lanzhou Univ Med Sci, 2024, 50(2): 55-59. DOI: 10.13885/j.issn.1000-2812.2024.02.009.
- [71] CHARAN J, BISWAS T. How to calculate sample size for different study designs in medical research? [J]. Indian J Psychol Med, 2013, 35(2): 121-126. DOI: 10.4103/0253-7176.116232.
- [72] KOREVAAR D A, HOOFT L, TER RIET G. Systematic reviews and meta-analyses of preclinical studies: publication bias in laboratory animal experiments[J]. Lab Anim, 2011, 45(4): 225-230. DOI: 10.1258/la.2011.010121.
- [73] 陈耀龙. 医学研究报告规范的发展与简介[J]. 兰州大学学报(医学版), 2022, 48(1): 1-4. DOI: 10.13885/j.issn.1000-2812.2022.01.001.
- CHEN Y L. J Lanzhou Univ Med Sci, 2022, 48(1): 1-4. DOI: 10.13885/j.issn.1000-2812.2022.01.001.
- (收稿日期:2024-07-26 修回日期:2024-12-03 )  
(本文责任编辑:丁宇菁)
- [引用本文]**  
李腾飞, 郑卿勇, 许建国, 等. 提高动物实验系统评价/Meta 分析的证据确定性:GRADE 方法的实证研究[J]. 实验动物与比较医学, 2025, 45(1): 101-111. DOI: 10.12300/j.issn.1674-5817.2024.109.
- LI T F, ZHENG Q Y, XU J G, et al. Improving the certainty of evidence in animal experiment systematic review/Meta-analysis: An empirical study of the GRADE method[J]. Lab Anim Comp Med, 2025, 45(1): 101-111. DOI: 10.12300/j.issn.1674-5817.2024.109.

\*\*\*\*\*

## 《实验动物与比较医学》有关实验动物福利伦理内容的说明

本刊是我国实验动物科学与比较医学领域的一本专业学术期刊, 严格遵守国家实验动物相关法律、法规和标准, 包括但不限于《实验动物管理条例》(2017年3月1日修订版)和《实验动物 福利伦理审查指南》(GB/T 35892—2018)等, 同时参考借鉴国际生物医学期刊关于动物实验研究报告的相关指南共识(如ARRIVE 2.0、IGP 2012、IAVE Guidelines 2010等)。因此, 本刊对所有涉及动物实验的来稿均需审查实验动物福利与伦理相关内容。现将一些具体要求说明如下:

1. 涉及动物实验的来稿, 需提供实验动物生产许可证和质量合格证, 以及动物实验场所的实验动物使用许可证。以上证明须与使用动物种类及动物实验单位名称相匹配, 并在正文中列出其对应的许可证编号。

2. 涉及动物实验的来稿, 需在考虑3R(替代、减少和优化)原则的基础上设计动物实验, 并提供作者单位实验动物福利伦理委员会(或相关机构)出具的实验动物福利伦理审查批件。批件中所列内容须与投稿文章相吻合, 并在正文中列出对应的批准编号。

3. 实验动物的用药, 尤其是麻醉镇痛药物必须优先使用药用级麻醉剂, 特别是当涉及存活手术的动物实验时。鉴于无法确定非药用级麻醉剂(如三溴乙醇、水合氯醛等)的相关性状及对实验动物的影响, 从而不能保障实验动物福利及研究结果的可靠性, 而且目前已有更优的市售麻醉药剂可供选择, 因此本刊不建议使用上述试剂。如确需使用, 请提供充足理由说明及相应的批准文件。

4. 涉及肿瘤动物模型的研究, 本刊参考国内及国际通用准则, 建议单个肿瘤体直径不超过20 mm(小鼠)或40 mm(大鼠)且不出明显的肿瘤溃疡。如投稿文章中有超出上述标准的研究内容, 需提交作者单位相关肿瘤动物模型研究的指导原则文件, 以及从科学角度判断肿瘤体积合理性的依据性材料。

《实验动物与比较医学》编辑部