



中文体育类核心期刊元分析的选题和方法学问题

徐 恺, 殷明越, 王 然*

(上海体育大学 竞技运动学院, 上海 200438)

摘 要:目的:探讨中文体育类核心期刊中元分析的选题和方法学问题,进行评价和举例分析,提出相应建议。方法:以中国知网数据库中16本中文体育类核心期刊发表的元分析研究为对象,依据Cochrane手册和经典元分析文献探讨体育科学领域元分析存在的选题问题,并评估所选元分析研究的方法学问题。方法学问题的主要评估指标为:1)是否存在异常值(效应量 >3);2)是否将标准误换算成标准差;3)是否忽略了研究内的相关性;4)是否忽略了研究内的方差;5)是否关注研究内而非研究间;6)报告多个效应量的研究是否仅提取某一个效应量。结果:在选题方面,71.1%的中文元分析的研究主题与既有英文元分析相似,其中27.3%的中文元分析题目名称与英文元分析几乎相同。在方法学方面,82.2%的元分析至少存在1个方法学问题,这些问题可能导致43.8%的元分析结果发生变化。结论:元分析应以解决具有实际意义的研究问题为出发点。在进行元分析时,研究者应仔细检查原始数据的报告形式,并采用多种效应模型、分析方法、合并策略、元分析软件及展示方式,通过团队合作确保元分析结果的稳定性。更重要的是,需在方法部分详细介绍和解释分析过程,以提高研究的透明度和可重复性。

关键词:元分析;选题;方法学;效应量;体育

中图分类号:G812.6

文献标识码:A

1952年,心理学家Hans Jürgen Eysenck发表的一篇论文质疑心理治疗的效果,认为患者在治疗期间的病情改善主要源于自身原因,与治疗方法关系不大,甚至暗示心理治疗可能妨碍患者康复。此观点对心理治疗的信誉造成重大打击。直至1977年,Smith等(1977)开发了元分析技术,综合分析了375项心理治疗研究,涵盖超过4 000名参与者,通过整合跨研究的标准化均数差(standardized mean difference, SMD),最终得出综合效应量达到了0.68。他们根据量化结果有力证明了心理治疗的益处,并有效反驳了Eysenck的观点。因此,元分析是一种解决学术争论的有效统计方法。

在体育科学领域,时间与资源的限制可能导致实验研究存在偏差,如小样本量引起的结果不精确和统计效力不足(张连成等, 2023)。彭凡等(2023)研究认为,约46%的体育科学实验研究样本量不足,结果易受到I型和II型统计错误及抽样误差的影响,依赖于偶然发现的效应,且置信区间较宽。元分析通过整合相似实验研究,不仅增强了统计效力,还能提供更精确的效应量估计和较窄的置信区间(李国春等, 2013; Harrer et al., 2021)。此外,众多实验研究可能得出不一致的结论,体育科研人员和从业者很难从中快速获取有效信息(Higgins et al.,

2019)。作为推动体育科学研究的重要方法,元分析通过明确的排除和纳入标准,最大限度地控制偏差风险,对异质性来源和影响因素进行分析,为理论和实践提供客观、公正的总结与建议(罗杰等, 2013; Higgins et al., 2019)。

然而,元分析的实际意义高度依赖于研究问题的选择以及方法学的严谨性。在查阅体育科学领域元分析研究时发现,部分研究在选题上存在缺陷,如对特定研究主题的重复探讨而并未更新结论、过度依赖国外高被引元分析文章的选题,以及选题的2个衍生问题,即原始研究纳入不全面或题目不精确、忽视原始研究的质量和数量。元分析的结果受原始研究的质量和数量影响。在元分析重复探讨特定的研究主题时,如果纳入完全相同或相似的原始研究,往常只会得到几乎相同的研究结论。这对研究领域的发展不会起到显著作用,反而会造成研究资源的浪费(Higgins et al., 2019)。此外, Kadlec等(2023)评估体能训练领域被引量排名前20的元分析时发现诸多方

收稿日期:2023-10-26; 修订日期:2023-12-20

基金项目:上海市科委“地方院校能力建设计划”项目(21010503500)。

第一作者简介:徐恺(1998-),男,在读硕士研究生,主要研究方向为运动训练,E-mail:2221152066@sus.edu.cn。

*通信作者简介:王然(1987-),男,教授,博士,博士研究生导师,主要研究方向为竞技运动科学,E-mail:wangran@sus.edu.cn。

法学问题,包括存在异常值、未将标准误转换为标准差、未考虑研究内的相关性、忽略了研究内的方差、更关注组内而非组间结果。因此,本研究以我国16本中文体育类核心期刊发表的元分析文章为研究对象,依据Cochrane手册和经典元分析文献探讨我国体育科学领域元分析存在的选题与方法学问题,直观呈现上述方法学问题对元

分析结果的影响,并提出相应建议。

1 选题和方法学问题概况

在中国知网数据库中检索16本中文体育类核心期刊2005—2023年发表的元分析学术论文,经筛选,共128篇文章纳入研究(图1)。

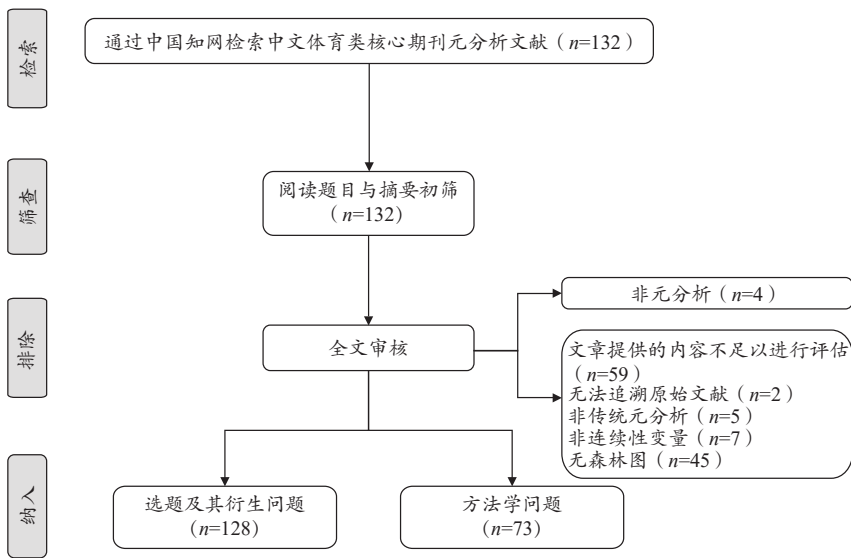


图1 文献检索流程图
Figure 1. Meta-analysis Literature Screening Process and Results

1.1 选题及其衍生问题

如表1所示,有29篇(22.7%)中文元分析重复探讨了相似的研究问题,涉及运动锻炼影响轻度认知障碍或阿尔兹海默病老年人认知功能、高强度间歇训练与中强度持续训练影响超重/肥胖人群减重效果、核心力量训练益处、不同运动锻炼影响老年人骨密度/适应性/平衡能力、复合式/对比训练影响下肢爆发力、运动对毒品成瘾或青少年网络成瘾的干预效果等研究领域。此外,提取128篇中文元分析的关键词并在谷歌学术检索后发现,35篇(27.3%)中文元分析的题目名称与先前发表的英文元分析题目高度相似;91篇(71.1%)中文元分析发表前已存在相似主题的英文元分析;仅有21篇(16.4%)中文元分析在发表前没有相似主题的英文元分析,7篇(5.5%)仅包含中文原始研究。此外,有40篇(31.3%)元分析选题范围较广,但纳入的原始研究数量有限,20篇(15.6%)纳入文章数量少于较早发表的相应英文元分析。

关于原始研究的质量评价,通过检查元分析中的偏倚风险/质量评价进行判断。研究发现,21篇(16.4%)元分析未进行原始研究的偏倚风险/质量评价。在进行了评价的107篇元分析中,有16篇(15.0%)纳入了高风险偏倚或低质量的原始研究;此外,40篇(74.1%)元分析错误地将偏倚风险评价作为对研究质量的评价(详细方法和结果见<https://osf.io/wxfpm/>)。

1.2 方法学问题

经初步检索和筛选,共纳入73篇元分析进行方法学问题评价(表2)。依据Kadlec等(2023)的评价方法,发现存在异常值(效应量>3)和未将标准误转换为标准差的元分析均为27篇(37.0%);32篇(43.8%)文章忽略了研究内的相关性;3篇(4.1%)进行元回归时未考虑原始文献的方差;5篇(4.1%)关注组内结果;35篇(47.9%)提取了多个结局指标中的一个。共有60篇(82.2%)元分析文章存在方法学问题,32篇(43.8%)的方法学问题可能导致元分析结果发生改变。此外,35篇(47.9%)文章仅以SMD形式报告效应量;32篇(43.8%)以加权均数差(weighted mean difference, WMD)形式报告效应量;6篇(8.2%)报道了多个效应量形式。

2 选题及其衍生问题的评价与举例

元分析旨在通过结构化、透明且可复制的方法解决研究问题,以减少偏倚风险,客观公正地回答研究问题,有利于从业者迅速将结论应用于实践(Borenstein et al., 2009; Higgins et al., 2019)。然而,选题上的问题会限制其在体育科学研究中的作用发挥。

2.1 重复探讨特定研究主题

本研究发现,2017—2019年,有4篇元分析重复探讨了高强度间歇训练与中强度持续训练对超重和肥胖人群

的影响,其中 3 篇得出高强度间歇训练优于中强度持续训练的结论。而 *Sports Medicine* 早在 2015 年就已刊文,对此主题进行了全面的元分析,并得出高强度间歇比中强度持续训练更有效的明确结论(Ramos et al., 2015)。此外,2020—2022 年,有 4 篇元分析探讨了运动对轻度认知障碍或阿尔兹海默病老年人/痴呆患者认知能力的影响,这些研究均发现不同形式的运动有助于改善目标群体的认知能力。而 *British Journal of Sports Medicine* 等国际权威期刊在 2020 年之前已发表过此主题的元分析,验证了运动对不同人群认知功能的有效性(Kramer et al., 2018; Wu et al., 2013; Zheng et al., 2016)。提示体育领域一些元分析未能在现有研究的基础上提供新的见解,只是复现了国际上的研究成果。

表 1 对元分析选题及其衍生问题的评价结果		
Table 1 Evaluation Results of the Topic Selection and Its Derivative Problems in Meta-analysis n=128		
项目	数量/篇	占比/%
是否对特定研究主题的重复探讨		
重复探讨特定研究主题	29	22.7
题目与既有英文元分析题目雷同	35	27.3
过度依赖国外高被引元分析		
发表前存在相似主题的英文元分析*	91	71.1
发表前无相似主题的英文元分析	21	16.4
纳入的原始研究均为中文研究	7	5.5
纳入原始研究是否全面/元分析题目是否精确		
过大的选题范围却仅纳入少量原始研究	40	31.3
纳入原始研究数量明显少于英文元分析	20	15.6
是否忽略原始研究的质量和数量		
纳入研究数量的总体范围	4~148	—
纳入研究数量的平均数	25.1	—
纳入研究数量的中位数	17.5	—
纳入研究数量≤10 篇	20	—
统计功效不足	6	4.7
对纳入研究进行偏倚风险/质量评价	107	83.6
未对纳入研究进行偏倚风险/质量评价	21	16.4
纳入原始研究存在低质量/高风险的元分析	16/107	15.0 [#]
将偏倚风险评价认为是研究质量评价	40/54	74.1 ^{&}

注:*元-元分析和系统评价再评价均在国外有类似的“umbrella”综述,因此也在评价范围之内;#n=107;&n=54;“题目与既有英文元分析题目雷同”是指中文元分析的题目名称与英语题目完全一致,仅仅是对英文元分析题目进行了翻译;“发表前存在类似主题的英文元分析”是指中文元分析题目在关键信息上与英语元分析题目重合,研究主题基本一致,不同的是改变了干预方式、结局指标或受试者人群。

进一步的统计分析显示,有 35 篇(27.3%)中文元分析的题目与之前发表的英文元分析题目高度相似。由于新的原始研究会随时间推移不断涌现,可以对特定主题重新进行元分析,纳入更多的原始研究以更新结论和观点,从而推动科学发展(Higgins et al., 2019)。然而,本研究发现有 20 篇(15.6%)与英文元分析题目高度类似的中文

元分析,纳入的原始研究数量却少于此前发表的英文元分析。虽然这些中文元分析可能采用了更严格的排除和纳入标准,但纳入的原始研究数量过少可能不足以更新和深化研究结论。

表 2 对元分析方法学问题的评价结果		
Table 2 Evaluation Results of Methodological Issues in Meta-analysis n=73		
项目	数量/篇	占比/%
方法学问题		
存在效应量>3 的异常值	27	37.0
未将标准误换算成标准差	27	37.0
忽视研究内的相关性	32	43.8
忽略研究内的方差	3	4.1
关注组内结果而非组间结果	5	6.8
仅提取多个结局指标中的一个	35	47.9
存在至少 1 个方法学问题	60	82.2
方法学问题可能影响元分析结果	32	43.8
效应量的报道方式		
以 SMD 形式报道效应量	35	47.9
以 WMD 形式报道效应量	32	43.8
报道 1 个以上效应量	6	8.2

2.2 过度依赖国外高被引元分析

分析结果显示,91 篇(71.1%)中文元分析在选题上过度依赖于高被引的英文元分析。无疑,参考已发表的高质量元分析是快速掌握研究现状和确定元分析选题的关键步骤。然而,这种过度依赖已发表元分析,仅通过调整干预措施、研究对象和结果指标形成的元分析选题可能在创新性和独立性上存在缺陷。例如,有 2 篇分别发表于 2020 年和 2021 年的中文元分析探讨了复合训练/对比训练对下肢爆发力的影响,而 *Strength & Conditioning Journal* 在 2019 年已经发表过此主题的元分析(Pagaduan et al., 2019)。尽管这两篇中文元分析将结果指标从垂直跳跃表现扩大为下肢爆发力,或将受试者限制为青少年,但其结论与先前发表的研究大致相同。

此外,还有 2 篇中文元分析研究了激活后表现提升(postactivation performance enhancement, PAPE)的效果,其中一篇的题目几乎完全复制了 2013 年 *Journal of Sports Sciences* 上发表的英文元分析文章,另一篇的研究思路也可能在一定程度上借鉴了 2016 年 *Sports Medicine* 发表的英文元分析研究。在所有调查的中文元分析中,仅有 21 篇(16.4%)文章在发表前没有相似主题的英文元分析,7 篇(占比 5.5%)仅包含了中文原始研究。这意味着,真正创新的元分析选题占比非常少。为促进学科领域内的知识进步和独立思考,中文体育科学领域的元分析研究需要更多地关注于探索新的研究问题,同时也应当注重对原有研究成果的深入分析和合理利用。

2.3 纳入的原始研究不全面/元分析题目不精确

统计结果表明,有40篇(31.3%)中文元分析纳入的原始研究数量较少。以2023年发表的一项关于运动锻炼对老年人睡眠质量影响的元分析为例,尽管研究主题的实践价值很高,但该研究仅纳入了16篇原始研究,其结论仅强调了运动锻炼对老年人睡眠质量的积极影响,而未指明最优的锻炼方式。相比之下,2022年发表于*Sleep Medicine Review*的一篇同主题网状元分析则纳入了35篇原始研究,并明确指出肌肉耐力训练结合步行是改善老年人睡眠质量的最佳方法(Hasan et al., 2022)。此外,2019年发表的一篇关于抗阻训练对老年人骨密度影响的元分析仅纳入了17篇原始研究,而早在2012年于*Age*期刊上发表的一篇相同主题的英文元分析则纳入了19篇原始研究(Marques et al., 2012)。尽管不能完全排除上述元分析可能设定了更严格的纳入标准,但当元分析的题目设置宽泛,却仅纳入少量原始研究时,可能存在纳入研究不够全面或题目设置不够精确的问题。这种情况可能导致元分析的结论缺乏足够的证据支持,从而影响研究结果的可靠性和实用性。因此,元分析研究应当致力于确保原始研究的全面纳入,同时精确定义研究题目,以便更准确地反映研究的目的和范围。这不仅有助于提高研究的质量和可信度,也能确保研究结果对实践应用具有更高的价值和指导意义。

2.4 忽略原始研究的质量和数量

统计显示,中文元分析纳入的原始研究数量介于4~148篇,平均纳入数量为25.1篇,中位数为17.5篇,且有20篇中文元分析的纳入研究数量不超过10篇。特别值得关注的是,6篇中文元分析可能存在统计效力不足的问题,这主要是因为它们合并了过少的效应量或原始研究数量不足。一般而言,元分析的统计效力可以达到99%以上,这是其相对于单项原始研究的显著优势之一。然而,当原始研究数量不足或合并的效应量过少时,元分析无法发挥其提升统计效力的优势,也可能因此更易受到I型和II型统计错误、抽样误差以及发表偏倚的影响,从而降低合并结果的证据等级(Afonso et al., 2023)。Cochrane手册推荐使用GRADE方法来评估结果的证据等级(Higgins et al., 2019),但调查发现,只有少数中文元分析采用了此方法。

对于原始研究的质量评估(包括偏倚风险),结果表明107篇(83.6%)中文元分析报道了对原始研究的偏倚风险或质量评价,其中16篇(占比15.0%)纳入了被认为质量低或风险高的原始研究。然而,这些元分析并未深入探讨这些低质量或高风险研究可能对合并结果造成的影响,而是在评价后直接进行了数据合并。Cochrane手册建议,通过亚组分析、元回归、敏感性分析等方法探讨这些研究对结果稳定性的可能影响(Higgins et al., 2019)。此外,调查还发现40篇(74.1%)中文元分析将偏倚风险直接

等同于研究质量,这是一个常见的误解。高偏倚风险并不直接等同于研究质量低,偏倚风险更多关联于研究设计、实施、分析或报告的局限性,而研究质量则涉及研究按照尽可能高的方法学标准执行的程度(Büttner et al., 2020)。因此,偏倚风险评价更多指的是研究的方法学质量,而研究质量则涵盖了原始研究报告的完整性、伦理认可、统计功效、精度以及内外部有效性等方面。简言之,偏倚通常是研究者难以完全控制的系统性问题,而研究质量则是研究者可以通过高标准的研究设计和执行来保证的。

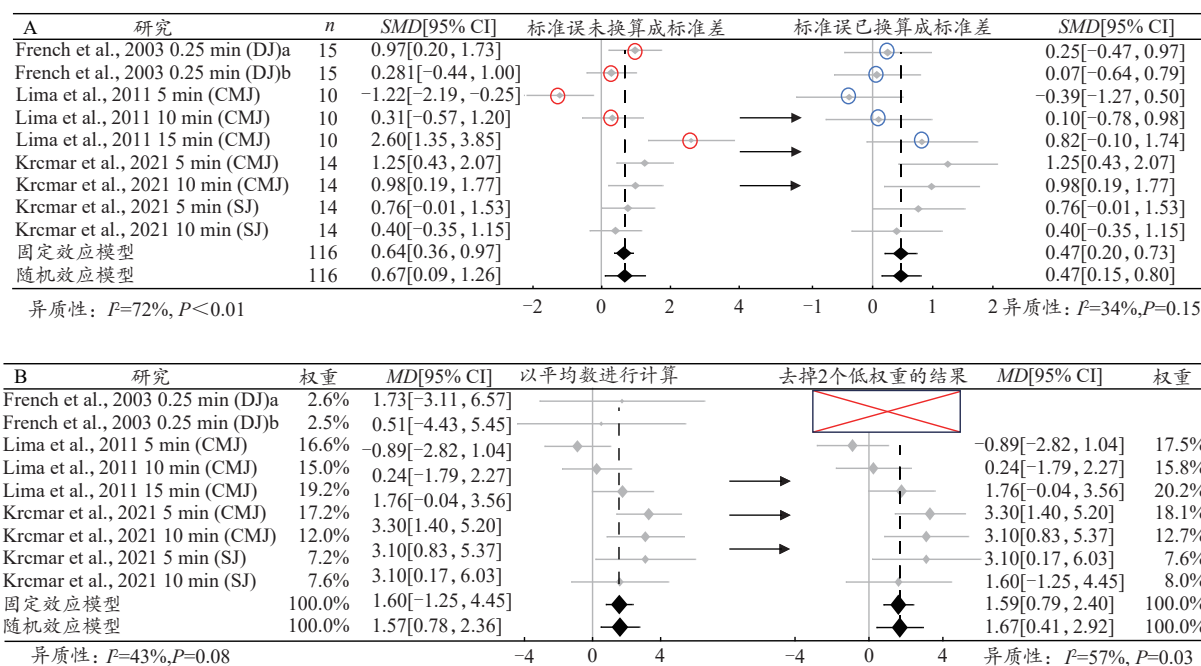
3 元分析方法学问题的评价与举例

在确定选题后,元分析的目标是将所有符合研究问题的相关研究在尽量减少偏倚风险的前提下进行合并,以获得更准确的平均效应量,并通过亚组分析、元回归等方式寻找异质性的来源和相关的影响因素(Borenstein et al., 2009; Higgins et al., 2019)。然而,错误的输入、选择性的数据合并和结果报道等均会导致有缺陷的元分析结果(Kadlec et al., 2023)。如表1所示,本研究发现纳入的73篇元分析中有82.2%存在方法学问题,这些问题可能会导致43.8%的元分析结果发生变化。这意味着,接近一半的研究在修正这些方法学问题后,其结果可能从非常显著($P<0.01$)变为显著($P<0.05$)或无显著差异($P>0.05$)。

3.1 存在异常值和未换算标准误

在对73篇中文元分析进行评估后发现了2个主要的方法学问题:异常值(效应量 >3)和未将标准误换算成标准差,这两个问题的出现频率均为37.0%。尽管效应量 >3 的异常值并不普遍,但在进行元分析时,需要对这类异常值保持高度警惕(Kadlec et al., 2023)。实际上,识别这些异常值相对容易,而将未转换为标准差的标准误直接输入森林图会导致以SMD形式呈现的效应量被异常夸大。例如,某研究中关于冲刺能力急性提升的效应量高达9.61,这在实际情况中是不现实的。这意味着一个100 m跑平均成绩为12 s的人在做准备活动时进行简单的深蹲后就能够跑进11 s。如图2A所示,换算后的效应量明显下降($SMD=0.47$)。

此外,标准误的直接输入还可能导致以WMD形式呈现的权重占比异常大。例如,在某元分析中,有的效应量甚至达到了99%的权重占比。这意味着即使删除其他所有研究,合并结果也几乎不会受到影响。正如图2B所示,在以WMD形式报道效应量时应格外注意权重占比,即使报道结果的单位相同,标准差的大小不同也可能影响合并后的结果。建议在元分析中同时使用WMD和SMD进行对比,以确认结果的稳定性。如果两者存在显著差异,研究者应在元分析报告中提及并进行解释。这一步骤对于保证元分析结果的准确性和可靠性至关重要,有助于避免由于方法学问题导致的研究结论误差。



注:A.未换算标准误对结果的影响;B.平均值差异中低权重对结果的影响;DJ. 跌落跳(drop jump),CMJ. 反向纵跳(countermovement jump),SJ. 下蹲跳(squat jump),后同。

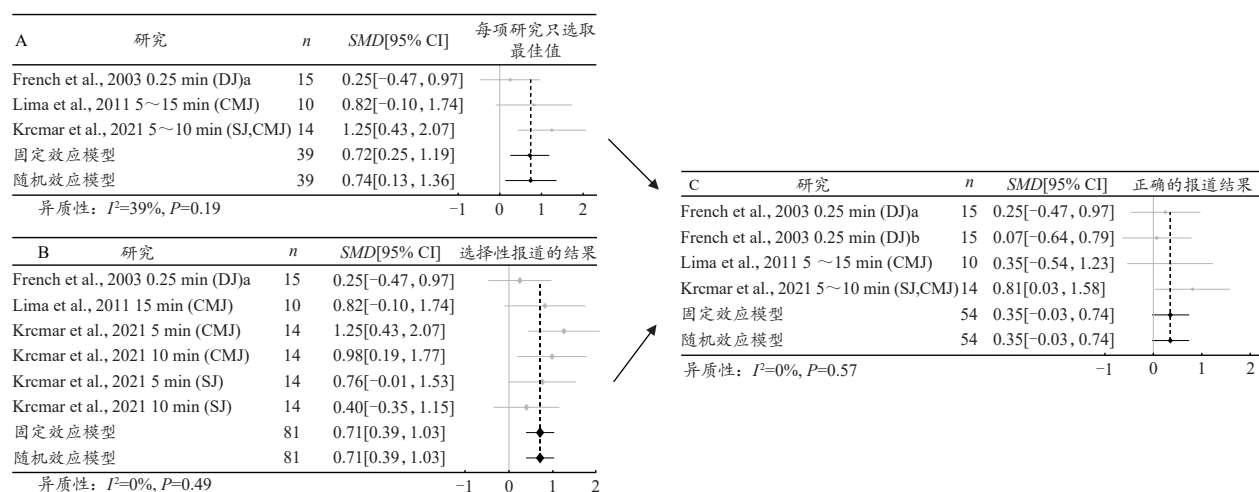
图2 未换算标准误与平均值差异中低权重对结果的影响

Figure 2. Impact of Unconverted Standard Errors and Low-Medium Weights in Mean Differences on Analysis Results

3.2 忽略研究内效应量之间的相关性和选择性报道

在进行元分析时,忽略研究内效应量之间的相关性和仅提取多个结局指标中的最佳值是2个常见的方法学问题。如图3A和3B所示,忽略研究内的相关性可能导致结果过于理想。对此,Cochrane手册推荐在可能的情况下合并相关的结果(Higgins et al., 2019)。如图3C所示,与直接忽视研究内的相关性结果相比($SMD=0.47$, $P<0.01$),合并后的效应量变小($SMD=0.35$),且未达到显著水平($P=0.07$)。另一方面,如图3A所示,如果研究者仅提取多个

结局指标中的最佳值进行报告,可能会导致结果严重倚倚,效应量虚高(固定效应模型下的 $SMD=0.72$, $P<0.01$)。对于这种情况,合并多个效应量过于保守,而直接选取一个最佳值又过于激进。三水平(多层次)元分析提供了一种同时考虑研究内效应量相关性的方法,通过纳入所有可能的效应量来提高分析的透明度和统计效力(Assink et al., 2016)。另一方面,每项研究会纳入所有可能的效应量以增加分析过程的透明度。因此,在面对原始研究中多个具有依赖性的效应量时可以选择使用三水平进行元分析。



注:A.每项研究仅选取最佳效应量的合并结果;B.选择最佳方式进行合并的结果;C.根据Cochrane手册的建议,将具有依赖关系的效应量进行合并[这里假设了French等(2003)的2种干预方式是独立的]。

图3 可能的效应量合并结果

Figure 3. Combined Results of Selecting Effect Sizes

此外,当一项研究报告了多个干预组与一个对照组的比较时,Cochrane手册推荐合并这些干预组(Higgins et al., 2019)。传统元分析是基于效应量独立的假设进行的,然而实际情况可能并非如此,由于相同的实验环境、主试、结局指标等因素,单项研究中的多个干预组之间可能存在一定的相关性^①,且比不同研究干预组之间的相关性更强(Cheung, 2013; Van den Noortgate et al., 2013)。例如,不同原始研究中的深蹲训练组之间和同一研究中的深蹲训练组与超等长训练组之间,可能后者中各组间的相关性更高。Cochrane手册中提到可以将深蹲训练组和超等长训练组都纳入元分析,但这会导致与对照组进行两次对比,因此应将对照组的人数除以对比次数以控制权重(Higgins et al., 2019)。除了推荐通过网状元分析消除单一研究中的多次对比带来的偏倚影响外,Cochrane手册也提到可以仅分析一种干预措施(Higgins et al., 2019),但本研究对此持不同意见。如图3B所示,元分析作者可能会选择地报道最佳值($SMD=0.71, P<0.01$),且异质性非常低($I^2=0\%, P=0.49$)。这样的结果虽然在统计上看似有显著意义,但实际上可能大幅偏离了真实效应量($SMD=0.35, P=0.07$),因而在实际应用中价值有限。因此,在面对多种干预措施和多个测试结果时,如果仅选取一个效应量,且未在方法部分明确解释选取该效应量的原因,元分析作者便有了“选择的余地”。换言之,结果的导向并非由“科学”定义,而是由主观意愿引导形成的。

3.3 忽略研究内的方差

在进行元分析的元回归分析时,忽略研究内的方差是一个常见的问题,但在本研究评估的73篇元分析中,只有3篇(4.1%)在元回归中未考虑原始文献的方差。与传统的线性回归不同,元回归考虑了原始研究中的抽样误差和效应量大小的变异性,从而提供了更为精确的效应量估计。如图4C和4D所示,即使简单线性回归图中体现了每个原始文献的方差($P=0.044$),结果也与元回归大相径庭($P=0.260$)。因此,建议使用元分析软件(如R、Stata等)进行元回归分析(李国春等, 2013; Harrer et al., 2021)。此外,元回归是观察性的,即使在结果中观察到显著的回归关系,也要注意其他可能的混杂因素对结果的影响。如图4B所示,当结果是非线性关系时,可通过三次多项式等模型进行拟合,确定效应量和影响因素的非线性关系(Harrer et al., 2021)。

3.4 忽略组间对比时的相关性系数和基线可比性

在中文体育类核心期刊的元分析文章中,绝大多数研究关注的是组间对比,仅有5篇研究进行了组内对比,占比6.8%,即93.2%的国内研究进行了组间对比,优于国外组间对比的元分析占比(55%)。组间对比主要有2种方法:基于基线变化的对比和直接对比后测结果,这两种方法各有利弊(图5A)。Cochrane手册推荐在理想情况下使

用基于基线变化的对比方法,以减少个体差异的影响(Higgins et al., 2019)。然而,这种方法要求对结果进行2次测量,且测量误差有时可能超过个体间的差异。大多数研究并不直接报告基于基线的变化值,而是通过相关性公式来计算这些变化值。遗憾的是,报告组内相关系数 r 的研究相对较少,大多数研究根据Cochrane手册的建议,保守地假设 r 值为0.5(Harrer et al., 2021; Higgins et al., 2019)。然而,根据PAPE研究的原始数据,得到的相关系数 r 通常高于0.6(Chaouachi et al., 2011; Maroto-Izquierdo et al., 2020)。这表明在使用基于基线变化的对比时,应当考虑使用不同的相关系数,并与直接对比后测的结果进行比较,以选择更可靠的结果(r 的具体取值取决于对应领域中对结局指标相关性系数的报告和敏感性分析的结果,如图5B所示)。

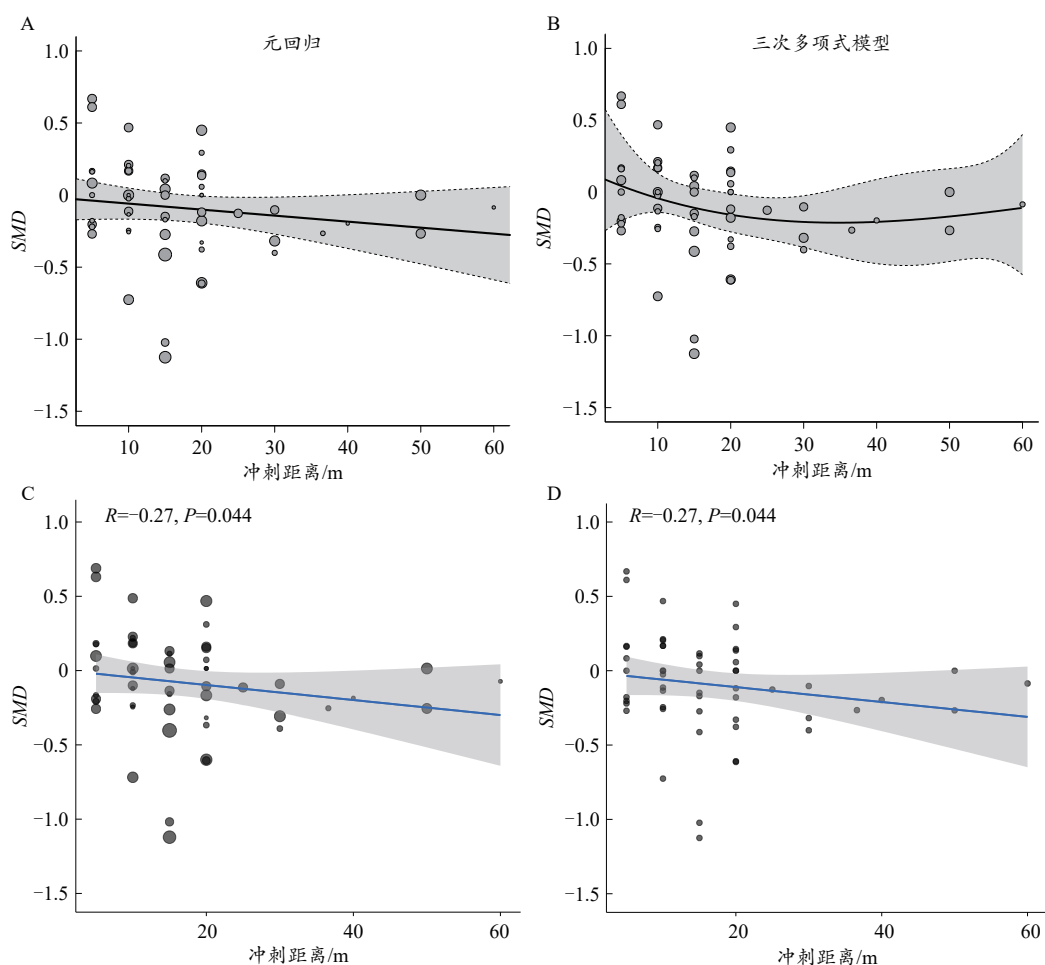
另一方面,直接对比后测结果的主要问题在于基线可能不相似。在评估过程中发现,某些元分析在基线不相似的情况下直接进行了后测对比。例如,某元分析干预组半蹲1RM的前测与后测值分别为102.0 kg和127.8 kg,而对照组的前测与后测值分别为83.8 kg和90.0 kg,对照组12周训练后的后测值仍低于干预组的前测值,这种情况下直接对比后测值会导致严重的偏倚。因此,为了确保元分析结果的稳定性和准确性,建议在可能的情况下综合使用基于基线变化和直接对比后测这2种方法。此外,合理选择和报告相关系数 r 值,以及在基线不相似的情况下采取适当的统计措施,对于提高元分析的质量至关重要。

4 选题和方法学问题的改进建议

4.1 选题方面

1)元分析应致力于解答原始研究未能明确回答的问题,这些问题不仅应具有理论意义,还应具有现实应用价值。确定研究主题后,参考相关英文元分析文献是一种有效的做法。但应在深入理解相关领域研究现状的基础上,选择能够反映该领域学术争议和局限的研究主题。例如,围绕运动前是否应加入静态拉伸的争论已经存在多年,有研究支持静态拉伸能提高运动表现和降低受伤风险(Smith, 1994)。然而,一项涵盖104篇静态拉伸研究的元分析发现,静态拉伸实际上会显著降低力量、功率输出和爆发力表现,且没有证据表明其能降低运动受伤风险(Simic et al., 2013)。这项元分析不仅解决了长期的学术争议,而且提供了实际应用建议,即教练和运动员应避免在热身时进行静态拉伸。这类研究问题非常适合进行元分析。然而,若进一步探索研究问题的子集,如静态拉伸对青少年跳跃表现的影响。由于其结果已然明确,进行元分析将缺乏意义,既不能解决新的学术争议,又缺乏实际应用价值(Ioannidis, 2016)。
























^①但这种相关性并不总是存在的,是由研究设计决定的,如随机交叉实验所产生的效应量之间的相关性会比随机对照实验的更强。



注:A.使用“metafor”包得到的元回归结果;B.使用三次多项式模型对非线性结果进行的拟合;C.使用“ggplot2”包得到的元回归结果;D:元回归未显示出每项研究效应量的方差。

图4 混杂因素对结果的影响

Figure 4. Impact of Confounding Factors on Results

A	研究	n	SMD[95% CI]	仅后测直接对比	平均值变化对比	SMD[95% CI]		
	French et al., 2003 0.25 min (DJ)a	15	0.27[-0.45, 0.99]			0.40[-0.32, 1.13]		
	French et al., 2003 0.25 min (DJ)b	15	0.16[-0.55, 0.88]			0.22[-0.49, 0.94]		
	Lima et al., 2011 5 min (CMJ)	10	-0.22[-1.09, 0.66]			-0.17[-1.04, 0.71]		
	Lima et al., 2011 10 min (CMJ)	10	0.47[-0.43, 1.36]			0.54[-0.35, 1.44]		
	Lima et al., 2011 15 min (CMJ)	10	0.95[0.01, 1.88]			1.03[0.08, 1.97]		
	Krcmar et al., 2021 5 min (CMJ)	14	1.86[0.95, 2.76]			1.81[0.91, 2.71]		
	Krcmar et al., 2021 10 min (CMJ)	14	1.33[0.50, 2.16]			1.42[0.57, 2.26]		
	Krcmar et al., 2021 5 min (SJ)	14	1.82[0.92, 2.72]			1.06[0.26, 1.86]		
	Krcmar et al., 2021 10 min (SJ)	14	1.37[0.53, 2.20]			0.54[-0.21, 1.30]		
	固定效应模型	116	0.82[0.54, 1.10]			0.72[0.45, 0.99]		
	随机效应模型	116	0.87[0.38, 1.36]			0.74[0.35, 1.13]		
异质性: $I^2=67\%$, $P<0.01$								
			异质性: $I^2=51\%$, $P=0.04$					
B	相关性取值	效应模型	SMD	95% CI	I^2	Tau^2	P异质性	P显著性
$r=0.6$	固定	0.79	[0.511 0, 1.062 0]	59%	0.26	0.01	<0.001	
	随机	0.83	[0.388 9, 1.261 2]	59%	0.26	0.01	<0.001	
$r=0.7$	固定	0.88	[0.601 0, 1.161 4]	67%	0.40	<0.01	<0.001	
	随机	0.95	[0.445 3, 1.448 5]	67%	0.40	<0.01	<0.001	
$r=0.8$	固定	1.02	[0.736 5, 1.313 1]	75%	0.66	<0.01	<0.001	
	随机	1.15	[0.539 4, 1.758 1]	75%	0.66	<0.01	<0.001	
$r=0.9$	固定	1.29	[0.985 1, 1.598 7]	84%	1.41	<0.01	<0.001	
	随机	1.29	[0.747 7, 2.429 6]	84%	1.41	<0.01	<0.001	

注:A.直接后测对比的结果和基于基线变化对比的结果;B.不同的相关系数 r 对基于基线变化对比结果的影响。

图5 两种组间对比结果

Figure 5. Comparison Results between Two Groups

2)元分析的质量依赖于所纳入的原始研究的质量和数量。若研究问题仅基于少量原始研究,可能因发表偏

倚和统计功效不足而导致误导性结论。原始研究的低质量可能引起“垃圾入,垃圾出”的问题,合并不可靠的研究

结果同样会得到不可信的结论(Borenstein et al., 2009)。Cochrane 手册推荐使用 GRADE 等级系统评估证据强度,以直观地判断结果的可靠性,避免因原始研究数量少或存在偏倚而得到不准确的结论。此外,作者应全面检索所有相关原始研究,或更精确地定义元分析的研究题目。例如,探讨抗阻训练对下肢爆发力影响的研究题目过于宽泛,需要明确抗阻训练的类型、下肢爆发力的具体指标及目标受试者群体。如此宽泛的题目可能涵盖数百项原始研究,传统元分析难以应对,网状元分析可能是更佳选择。如果元分析作者在广泛的题目下仅纳入少数研究,应进一步细化研究题目,明确干预措施、受试对象和结果指标,以提供给读者更清晰的信息。

3)元分析作者应明确区分偏倚风险和研究质量,并在可能的情况下同时使用偏倚风险评估工具和研究质量评估工具。另一方面,低质量或高风险并不是排除原始研究的理由。Cochrane 手册建议在符合纳入标准的前提下尽量纳入所有研究,通过亚组分析、元回归、敏感性分析等方法探究这些研究可能对综合结果的影响,并为未来研究提出建议(Higgins et al., 2019)。

4)Cochrane 手册指出,元分析的研究目的可以是:比较特定人群中不同干预措施的效果差异、比较特定干预措施在不同人群中的效果、探究影响干预效果的因素。高质量的元分析可能同时包含这3种研究目的。例如,探索高强度间歇训练与中等强度持续训练对青少年心肺适能的影响,旨在确定哪种训练更有效,评估训练效果是否受人群特征的影响,探讨训练效果是否受训练变量(如干预周期)的影响。在处理多种干预措施和结果指标时,可采用多层次网状元分析(Doosti-Irani et al., 2021)和三水平元分析(Harrer et al., 2021),以更准确地反映实际效应量。

4.2 方法学方面

1)在纳入研究数据时,必须细致核对原始研究报告的数据是使用标准误(standard error, SE)还是标准差(standard deviation, SD)。这一步骤对于确保数据准确性至关重要。

2)对于以 *WMD* 形式报告效应量的研究,应谨慎对待权重占比过大的情况,并检查原始文献的数据是否存在问题。若数据无误,可以考虑将 *WMD* 转换为 *SMD*,以确保结果的稳定性,因为权重过大的研究会直接影响合并结果。

3)对于以 *SMD* 形式报告效应量的研究,应特别注意 $SMD > 3$ 的情况。异常高的效应量可能会使结果产生偏差。

4)在研究包含多种干预措施时,根据 Cochrane 手册的指导,可以合并不同的干预方式,或调整多次比较中控制组的人数以平衡权重,或者采用网状元分析。应避免将

多种干预措施看作独立的效应量,元分析的作者应在方法学部分详细说明采取该策略的理由,并通过比较平均前后的结果,检验结果的稳定性。

5)若原始研究报告了不同时间点或使用不同测量方式得到的具有相关性的多个结果,应依据 Cochrane 手册的指导合并这些结果或采用三水平元分析。如果仍选择纳入单一结果,方法学部分应详细阐述原因。

6)执行元回归分析时,应注意考虑每项研究的方差,利用元分析软件执行多种元回归模型和预测模型拟合,以验证结果的稳定性。

7)仅比较后测结果时,应确保考虑到基线的可比性。比较平均值变化时,应报告相关系数 r 值,建议使用不同的 r 值(如 0.3~0.9)进行计算,并进行敏感性分析,以评估结果的稳定性。

8)进行元分析时,应采用开放的思维方式。建议使用多种模型(包括固定效应和随机效应模型)、多种方法(如传统元分析、三水平元分析、网状元分析、贝叶斯方法)、多种分析策略(包括亚组分析、元回归分析、非线性回归分析、敏感性分析等)、多种结果呈现形式(*WMD*、*SMD* 等)和多种软件工具(如 Review Manager、Stata、R 等),并与多位研究人员合作,以确保得到严谨、稳定且可复制的研究结果。

5 结论

元分析的核心价值在于其所提出研究问题的质量。通过对中文体育类核心期刊发表的元分析文章进行选题评估发现,近 3/4 的研究主题与既有英文元分析相似,更有约 1/4 的中文元分析题目名称与英文元分析几乎相同。有效的元分析应当致力于解答原始研究难以单独解决的问题,同时确保这些问题具有实际应用的意义,更具体地说,应聚焦体育科学实践中存在争议的问题。此外,进一步对中文元分析进行方法学评估发现,近一半的元分析存在方法学缺陷,这些缺陷可能会导致合并结果发生偏差。为提升元分析的科学性和实用价值,研究者在未来的工作中应选择具有深远意义的研究问题,并在方法学部分提供尽可能详尽的介绍和解释,以提高研究的透明度和可重复性。

参考文献:

- 李国春, 吴勉华, 余小金, 2013. *Meta 分析导论*[M]. 北京: 科学出版社.
- 罗杰, 冷卫东, 2013. *系统评价/Meta 分析理论与实践*[M]. 北京: 军事医学科学出版社.
- 彭凡, 张力为, 周财亮, 2023. 体育科学实验研究如何确定适宜的样本量[J]. 上海体育学院学报, 47(2): 26-36.
- 张连成, 刘洁, 吕嘉玉, 等, 2023. 从重复原则视角审视体育科学实验设计中的样本量问题[J]. 体育科学, 43(1): 90-97.
- AFONSO J, RAMIREZ-CAMPILLO R, CLEMENTE F M, et al., 2023. The perils of misinterpreting and misusing "publication bias"

- in meta-analyses: An education review on funnel plot-based methods [J/OL]. *Sports Med* [2023-10-11]. <https://doi.org/10.1007/s40279-023-01927-9>.
- ASSINK M, WIBBELINK C, 2016. Fitting three-level Meta-analytic models in R: A step-by-step tutorial[J]. *Quant Meth Psychol*, 12(3): 154-174.
- BORENSTEIN M, HEDGES L V, HIGGINS J P, et al., 2009. Introduction to Meta-Analysis[M]. New York: John Wiley & Sons.
- BÜTTNER F, WINTERS M, DELAHUNT E, et al., 2020. Identifying the ‘incredible’! Part 1: Assessing the risk of bias in outcomes included in systematic reviews [J]. *Br J Sports Med*, 54(13): 798-800.
- CHAOUACHI A, POULOS N, ABED F, et al., 2011. Volume, intensity, and timing of muscle power potentiation are variable [J]. *Appl Physiol Nutr Metab*, 36(5): 736-747.
- CHEUNG M W L, 2013. Modeling dependent effect sizes with three-level meta-analyses: A structural equation modeling approach [J]. *Psychol Methods*, 19(2):211-229.
- DOOSTI-IRANI A, NAZEMPOUR M, MANSOURNIA M A, 2021. What are network meta-analyses (NMAs)? A primer with four tips for clinicians who read NMAs and who perform them (methods matter series)[J]. *Br J Sports Med*, 55(10): 520-521.
- FRENCH D N, KRAEMER W J, COOKE C B, 2003. Changes in dynamic exercise performance following a sequence of preconditioning isometric muscle actions [J]. *J Strength Cond Res*, 17(4): 678-685.
- HARRER M, CUIJPERS P, FURUKAWA T A, et al., 2021. Doing Meta-Analysis with R: A Hands-On Guide [M]. Boca Raton, FL and London: Chapman & Hall/CRC Press.
- HASAN F, TU Y K, LIN C M, et al., 2022. Comparative efficacy of exercise regimens on sleep quality in older adults: A systematic review and network meta-analysis[J/OL]. *Sleep Med Rev*, 27(65): 1-7[2023-10-11]. <https://doi.org/10.1016/j.smrv.2022.101673>.
- HIGGINS J, THOMAS J, CHANDLER J, et al., 2019. Cochrane Handbook for Systematic Reviews of Interventions [R]. London: Cochrane.
- IOANNIDIS J P, 2016. The mass production of redundant, misleading, and conflicted systematic reviews and meta-analyses [J]. *Milbank Q*, 94(3): 485-514.
- KADLEC D, SAINANI K L, NIMPHIUS S, 2023. With great power comes great responsibility: Common errors in Meta-analyses and meta-regressions in strength & conditioning research [J]. *Sports Med*, 53(2): 313-325.
- KRAMER A F, COLCOMBE S, 2018. Fitness effects on the cognitive function of older adults: A Meta-analytic study—revisited [J]. *Perspect Psychol Sci*, 13(2): 213-217.
- KRCMÁR M, KRCMÁROVÁ B, BAKALÁR I, et al., 2021. Acute performance enhancement following squats combined with elastic bands on short sprint and vertical jump height in female athletes [J]. *J Strength Cond Res*, 35(2): 318-324.
- LIMA J B, MARIN D P, BARQUILHA G, et al., 2011. Acute effects of drop jump potentiation protocol on sprint and countermovement vertical jump performance [J]. *Hum Mov*, 12(4): 324-330.
- MAROTO-IZQUIERDO S, BAUTISTA I J, MARTÍN RIVERA F, 2020. Post-activation performance enhancement (PAPE) after a single bout of high-intensity flywheel resistance training [J]. *Biol Sport*, 37(4): 343-350.
- MARQUES E A, MOTA J, CARVALHO J, 2012. Exercise effects on bone mineral density in older adults: A meta-analysis of randomized controlled trials [J]. *Age*, 34(6): 1493-1515.
- PAGADUAN J, SCHOENFELD B J, POJSKIĆ H, 2019. Systematic review and Meta-analysis on the effect of contrast training on vertical jump performance [J]. *Strength Cond J*, 41(3):63-78.
- SIMIC L, SARABON N, MARKOVIC G, 2013. Does pre-exercise static stretching inhibit maximal muscular performance? A Meta-analytical review [J]. *Scand J Med Sci Sports*, 23(2): 131-148.
- SMITH C A, 1994. The warm-up procedure: To stretch or not to stretch. A brief review [J]. *J Orthop Sports Phys Ther*, 19(1): 12-17.
- SMITH M L, GLASS G V, 1977. Meta-analysis of psychotherapy outcome studies [J]. *Am psychol*, 32(9): 752-760.
- VAN DEN NOORTGATE W, LÓPEZ-LÓPEZ J A, MARÍN-MARTÍNEZ F, et al., 2013. Three-level meta-analysis of dependent effect sizes [J]. *Behav Res Methods*, 45(2): 576-594.
- WU Y, WANG Y, BURGESS E O, et al., 2013. The effects of Tai Chi exercise on cognitive function in older adults: A Meta-analysis [J]. *J Sport Health Sci*, 2(4): 193-203.
- ZHENG G, XIA R, ZHOU W, et al., 2016. Aerobic exercise ameliorates cognitive function in older adults with mild cognitive impairment: A systematic review and meta-analysis of randomised controlled trials [J]. *Br J Sports Med*, 50(23): 1443-1450.

Topics Selection and Methodological Issues of Meta-analysis in Chinese Sports Core Journals

XU Kai, YIN Mingyue, WANG Ran*

School of Athletic Performance, Shanghai University of Sport, Shanghai 200438, China

Abstract: Objective: To explore the topic selection and methodological issues in meta-analyses which published in Chinese sports core journals, and to provide recommendations for the improvement of research quality. Methods: Meta-analyses published in 16 Chinese sports core journals were retrieved from the CNKI database. Based on the Cochrane Handbook and classic meta-analysis literature, this study discussed the topic selection issues and evaluated the methodological issues in the selected meta-analyses, the items including: 1) The presence of outliers (effect size>3); 2) conversion of standard errors to standard deviations; 3) consideration

of within-study correlations; 4) oversight of within-study variances; 5) focus on within-study rather than between-study differences; 6) extraction of only one effect size from multiple effect sizes in one study. Results: In terms of topic selection, 71.1% of Chinese meta-analysis research topics were similar to existing English meta-analysis, with 27.3% of them were nearly identical to their English counterparts. In terms of methodological problem, 82.2% of the studies had at least one methodological issues, which could potentially alter 43.8% of the meta-analyses results. Conclusions: The aim of meta-analyses should focus on the resolution of major research problems with practical significance. When performing a meta-analysis, researchers should carefully check the reporting form of original data and employ various effect models, analysis methods, combining strategies, meta-analysis software, and presentation methods, and collaborating with other researchers to ensure the stability of meta-analysis results. More importantly, the author need to detail and explain the analysis process in the methods section to enhance the transparency and reproducibility of the study.

Keywords: *meta-analysis; topic selection; methodology; effect size; sport*

(上接第 58 页)

Artificial Intelligence Empowering Doping Control: Value Implications, Practical Challenges and Development Strategies

XU Weikang

School of Law, China University of Political Science and Law, Beijing 102299, China

Abstract: The field of anti-doping exhibits a natural sensitivity to technological advancements, and with the rapid development of artificial intelligence (AI) technology, doping control is gradually transitioning towards intelligence-driven approaches. This study comprehensively explores the value implications, practical challenges, and development strategies of AI-enabled doping control, drawing upon literature review, logical analysis, and other research methods. The research suggests that AI-based integration and examination of anti-doping intelligence, sample testing and analysis, and intelligent management bring about transformative model changes in the doping control landscape, enabling intelligence integration and testing into one, multidimensional targeting, and comprehensive support throughout the entire chain. However, the application of AI in doping control also faces technical challenges such as model errors and unconscious bias, legitimacy concerns regarding infringements of rights and blurred responsibilities, as well as ethical risks including algorithmic discrimination, black box effects, and technological overreach. To better promote the effective utilization of AI technology in doping control, efforts should focus on data quality control, algorithm-driven transformations, and the implementation of robust technical safeguards. Additionally, the legal framework should prioritize athlete rights protection, shape regulations governing data and algorithms, and optimize oversight and accountability mechanisms. Ethical alignment can be achieved through positioning technology as an aid, prioritizing digital human rights, and fostering multi-stakeholder participation in governance.

Keywords: *artificial intelligence; doping control; data analysis; risk regulation; rights protection*