

临床意义与统计学意义结合的临床试验结果分类方法及评价研究

王杨 郎欣月 朱熠冰 刘小云 赵延延 李思冬 李卫

中国医学科学院北京协和医学院国家心血管病中心/阜外医院医学统计部, 北京 100037

通信作者: 李卫, Email: liwei@mrbc-nccd.com

【摘要】目的 统计学意义在临床试验结果解读中发挥重要作用,但在获得有统计学意义结果时,临床意义的评估常被忽略,本研究尝试提出一种简单且明确的试验结果新分类方法,重点针对已获得统计学阳性的试验,评估其结果是否具有临床意义。**方法** 以2019年美国心脏病学会(ACC)和欧洲心脏病学会(ESC)大会公布结果的临床试验为研究对象,通过荟萃流行病学方法,提取研究水平的特征变量。主要评价指标包括试验设计阶段假设的目标效应值及试验结果所得的观察效应值,基于两者间的差异,对已获得统计学阳性的试验进行细分,识别出临床意义可能不充分的研究;并基于检验把握度的理论提出阈值,作为对该类问题进行识别的基础。**结果** 最终纳入分析的12项临床试验多发表于顶尖专业期刊、具有较高的研究设计与报告质量;观察效应值与目标效应值间有一定相关性($r=0.892$),但基于两者间的差异,并聚焦获得了有统计学意义结果的7项试验,其中被分类为临床意义不充分的共2项,按ACC和ESC公布的试验进行划分,有统计学意义却可能无临床意义的研究各1项(1/3和1/4)。**结论** 已获得统计学阳性结果的试验,仍有必要对其临床意义的充分性进行评估,本研究提出一种新的将临床意义与统计学意义结合的分类标准及在其基础上对临床试验结果可靠性进行评估的方法,辅助研究者识别因临床意义不充分导致的潜在风险,为临床研究结果的合理解读提供一定的参考和帮助。

【关键词】 目标效应值; 观察效应值; 荟萃流行病学; 阳性结果发表偏倚

Integration of clinical significance and statistical significance on clinical study results categorization: a Meta-epidemiology study

Wang Yang, Lang Xinyue, Zhu Yibing, Liu Xiaoyun, Zhao Yanyan, Li Sidong, Li Wei

Medical Research and Biometrics Center, National Center for Cardiovascular Diseases, Fuwai Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing 100037, China

Corresponding author: Li Wei, Email: liwei@mrbc-nccd.com

【Abstract】Objective Statistical significance plays an important role in the interpretation of clinical trial results. However, on the basis of obtaining statistical significance, the assessment of clinical significance is often neglected. This study attempted to propose a simple and unambiguous new classification method for study results, focusing on studies with statistical positive findings to evaluate whether the results have clinical significance. **Methods** Our study subjects were the clinical studies in 2019 ACC and ESC annual meetings. Meta-epidemiology methods were used to extract the characteristic variable from each study. The primary evaluation indicators included target effect-size and observed effect-size. Based on the difference between the two indicators, the studies that had statistical significance were subdivided to identify studies with possible insufficient clinical significance; Furthermore, the theoretical threshold based on power analysis was proposed, which was used as the basis for the interpretation of study results. **Results** There were 12 clinical studies included in the final analysis. All of them were published on top journals. Those studies had

DOI: 10.3760/cma.j.cn112338-20201015-01235

收稿日期 2020-10-15 本文编辑 万玉立

引用本文: 王杨, 郎欣月, 朱熠冰, 等. 临床意义与统计学意义结合的临床试验结果分类方法及评价研究[J]. 中华流行病学杂志, 2021, 42(7): 1280-1285. DOI: 10.3760/cma.j.cn112338-20201015-01235.



relative high quality on both study design and reporting. The correlation coefficient between the observed and target effect-size was 0.892. Among the 7 studies with statistical significance, two of them were classified as insufficient clinical significance. The counts was 1 (1/3) and 1 (1/4) for the studies reported in ACC and ESC respectively. **Conclusions** The achievement of clinical significance is critical even in the study with positive results. This paper proposes a new classification standard that combines clinical significance with statistical significance and further suggests a method to evaluate the reliability of clinical study results in order to assist researchers in identifying potential risks caused by insufficient clinical significance, and provide some reference and help for the reasonable interpretation of clinical study results.

[Key words] Target effect-size; Observed effect-size; Meta-epidemiology; Publication bias among positive study

在循证医学框架下,临床指南多基于高质量临床试验的结果做出更新和推荐,实现临床研究向临床实践的转化。已发表的临床试验文章是指南重要的证据来源,试验结果有统计学意义即所谓阳性结果的研究,相比阴性结果试验有更大概率被接收发表,该现象被称作“发表偏倚(publication bias)”^[1]。在 Meta 分析中,传统发表偏倚可通过经典的漏斗图、Begg's rank test 或 Egger's regression test 等方法进行识别。多项方法学研究证明,由于潜在的发表偏倚,导致在特定治疗领域某些干预方法的效果存在被高估的可能^[2-4],在临床指南的形成与临床实践的开展过程中,需要对这一风险予以关注。

更为关键的是,在大规模临床试验中,由于样本量较大、试验结果更易有统计学意义;甚至可能出现有统计学意义却临床意义不足的情况,该现象将导致对干预效果的错误或不合理评估,并造成临床有效性不佳的治疗方法被用于临床实践,给患者治疗带来潜在的伦理风险。针对大规模临床试验中可能出现的有统计学意义但临床意义不足的问题,目前尚无公认的认识方法和评估手段。对此我们尝试提出“临床试验阳性结果发表偏倚”的概念,并以心血管领域的临床研究为实例,展示在已获得阳性结果的试验中仍可能存在的“无临床意义”的问题,给出临床意义与统计学意义相结合的临床试验结果分类方法,并探讨以此分类作为阳性结果发表偏倚识别基础的可能性,以期从方法学角度对临床研究的试验设计与结果评价提供更细致的参考。

对象与方法

1. 研究对象:以 2019 年美国心脏病学会(ACC)和欧洲心脏病学会(ESC)会议大会公布的临床试验为对象,通过 PubMed、Ovid 及 Web of Science 检索对应的已发表期刊文章,检索截止时

间为 2020 年 2 月 29 日。研究入选标准:①随机对照;②试验为优效性假设;③正式发表的文章中含主要终点结果。排除标准:①针对亚组开展的分析;②期中分析结果或提前终止的试验;③对远期随访结果的分析;④采用了复杂的试验设计方式(如贝叶斯设计、两阶段设计、研究分组 ≥ 3 组);⑤在文章中缺少本研究所需的关键方法学信息。

2. 信息提取:首先针对研究目的设计标准化信息提取表格,正式提取数据前,通过小规模预试验对拟采集指标的可行性进行评估,并根据预试验结果完善采集表格;由经过培训的固定研究人员对所有文章的信息进行提取,避免研究者间的异质性;信息提取过程中如果遇到不确定的问题,则由第二名研究人员进行独立复核,研究者间独立提取的信息存在差异时,经相互讨论后最终统一。

3. 变量处理:由于本研究主要关注原始研究的方法学相关特征,在对研究水平的特征变量(作者、发表年、期刊、研究设计类型、研究假设、主要指标、样本量及样本量计算过程使用的参数、研究结果等)进行提取的基础上,主要研究变量为试验设计阶段研究者假设的目标效应值(target effect-size, Tgt ES)以及试验结果所获得的观察效应值(observed effect-size, Obs ES)估计水平,效应值为干预性研究中试验组与对照组间主要评价指标的差异(相对或绝对)。根据 Tgt ES 和 Obs ES 间的差异可将试验结果分为 3 大类:

情形一,未能拒绝原假设,即传统意义上的阴性结果,包括统计检验 $P \geq 0.05$ 或者获得了试验组显著劣效于对照的情形,此时, $Obs ES < Tgt ES$;

情形二,备则假设成立 ($P < 0.05$),且 $Obs ES \geq Tgt ES$,认为同时有统计学意义及临床意义;

情形三,备则假设成立 ($P < 0.05$),但 $Obs ES < Tgt ES$,此情形下说明虽获得有统计学意义的结果,但实际干预效果的临床意义可能不充分。

在对纳入分析的研究进行上述分类判定的基础上,在研究水平具有充分样本量的前提下,可进一步在所有获得有统计学意义结果的试验中,计算情形三所占的比例(符合情形三的试验数量占情形二和情形三试验总数的比例);由统计学假设检验原理可知,在备则假设成立时,通过其对应的概率密度函数可知,Obs ES 低于 Tgt ES 理论的概率水平为预设把握度减 50%(所扣除的 50% 为 Obs ES 达到或超过 Tgt ES 的概率),考虑临床试验中常用的把握度(power)设定为 80% 或 90%,对应的阳性结果发表偏倚识别理论阈值(有统计学意义但临床意义不足的研究占比)可以通过下述方法计算获得:在预设把握度为 80% 时 = 37.5% [(0.8-0.5)/0.8=0.3/0.8=0.375],在预设把握度为 90% 时 = 44.4% [(0.9-0.5)/0.9=0.4/0.9=0.444];如果情形三所占比例(在有统计学意义的试验中)超过上述阈值,则提示可能有潜在的“阳性结果发表偏倚”。由于本研究最终纳入实证分析的研究数量较少,故仅提出理论阈值的计算方法,而不对研究结果与阈值进行直接的对比。

4. 统计学分析:首先对所提取的研究水平特征变量进行描述性分析,定量指标用 $\bar{x} \pm s$ 描述,定性指标因样本量问题仅报告频数信息。在开展特定的组间比较时,根据变量分布特征,对定量指标用独立样本 *t* 检验或 Wilcoxon 秩和检验、定性指标则采用 χ^2 或 Fisher 精确概率检验。在对比 Tgt ES 和 Obs ES 差异时,首先将效应值统一为相对差异的维度,并以高优指标形式(统一用 $HR > 1$ 代表试验组疗效更优)为准,如果原文报告效应值为低优,即 $HR < 1$ 说明试验组更优时,在进一步处理前先将其取倒数,转变成成为 HR ,取值越大说明试验组更优的高优形式;之后对相对风险维度的效应值结果(HR 值或从组间绝对差异转换的组间相对差异)进行自然对数为底的转换,再通过带一致性参考线(斜率=1)的散点图对经上述标化后的 Tgt ES 与 Obs ES 间的差异程度进行比较;按方法部分给出的 3 种情形对纳入研究的分类结果进行描述,并重点给出在获得有统计学意义的研究中情形三的出现情况。采用 SAS 9.4 及 R 4.0.0 软件进行统计学分析,检验水准取双侧 $\alpha = 0.05$ 。

结 果

1. 基本特征:在 2019 年 ACC 和 ESC 大会公布

结果的 48 项临床研究中,通过检索共获得 44 项试验的原始研究文献。按预先设定的纳入排除标准筛选后,最终纳入本研究的试验文章数量为 12 篇^[5-16],检索流程见图 1,纳入研究的基本信息见表 1。从方法学角度,研究对象(特定的临床试验)普遍具有较高的设计及报告质量,在 12 项研究中,有 8 项试验采用临床复合终点事件作为主要评价指标,有 10 项试验采用了事件驱动(event-driven)的设计方式^[17],有 4 项试验设置了 90% 或更高的检验把握度(power)水平,研究的样本量超过 1 000 例的试验共有 8 项,另外还有 4 项试验预先指明开展期中分析并给出了对应的 Alpha 校正策略,对纳入本研究的原始研究方法学质量特征描述性分析结果见表 2。

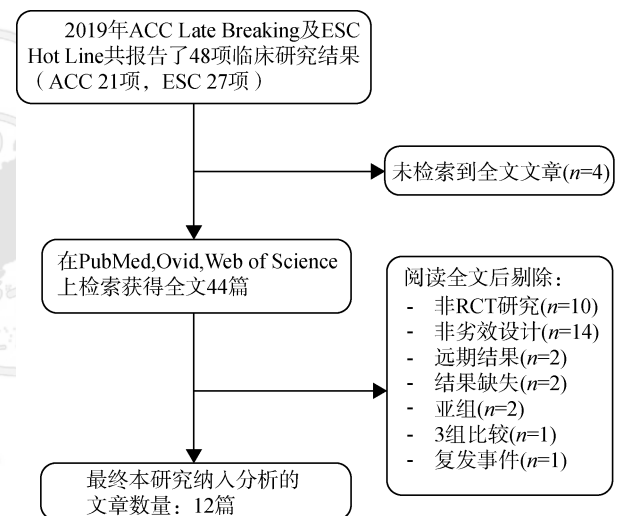


图 1 文献筛选及纳入流程

表 1 纳入研究文献的基本信息

试验名称	来源	终点类型	样本量(例)
WRAP-IT ^[5]	ACC	事件驱动	6 988
CLEAR Wisdom ^[6]	ACC	固定时点事件率	750
REDUCE-IT ^[7]	ACC	事件驱动	7 990
SAFARI-STEMI ^[8]	ACC	事件驱动	4 884
COACT ^[9]	ACC	事件驱动	552
THEMIS ^[10]	ESC	事件驱动	19 000
COMPLETE ^[11]	ESC	事件驱动	4 000
DAPA-HF ^[12]	ESC	事件驱动	4 500
ISAR-REACT 5 ^[13]	ESC	事件驱动	4 000
GALACTIC ^[14]	ESC	事件驱动	785
HOPE 4 ^[15]	ESC	固定时点事件率	1 200
MITRA.FR ^[16]	ESC	事件驱动	288

注:ACC:美国心脏病学会(American College of Cardiology)年会公布结果的研究;ESC:欧洲心脏病学会(European Society of Cardiology)年会公布结果的研究

表 2 纳入研究的方法学特征描述性分析

变量	ACC(n=5)	ESC(n=7)	P 值 ^a
样本量(例)			0.217
<1 000	2	2	
1 000~	1	4	
5 000~	2	0	
>10 000	0	1	
检验侧别			0.523
1 侧	0	0	
2 侧	3	6	
未指明	2	1	
是否引入 Alpha 校正			0.576
校正	1	3	
未校正	4	4	
把握度			0.558
80%~	3	5	
90%~	2	2	
主要终点类型			0.364
复合临床终点事件	2	6	
单一临床终点事件	2	0	
非临床终点事件	1	1	
效应值报告方式			1.000
相对效应值	4	6	
绝对效应值	1	1	

注：^a运用 Fisher's exact test 进行检验；表格中报告结果均为对应分类的出现频数，因样本量较小故未报告百分比

2. Tgt ES 与 Obs ES 的一致性：确证性临床试验，在试验设计阶段设置明确的统计学假设检验，并且基于 Tgt ES 的假设，进行样本量估算。但在按照方案设计完成试验时，Obs ES 受各种因素影响往往与假设的水平存在差异。图 2 结果对该差异进

行了直观的展示，虽然标准化的 Obs ES 与标化 Tgt ES 间具有一定的相关性 (Pearson 相关系数 $r=0.892$)，但同样可以发现 Obs ES 与 Tgt ES 间的差异普遍存在，甚至像 ISAR-REACT 5 试验^[13]，获得了与研究假设完全相反的结果。基于这一现象，我们对纳入分析的试验结果做进一步分类处理。

3. 有统计学意义但临床意义不足的风险识别：对没有提前终止的确证性试验，其结果理论上可归类到本研究方法部分给出的 3 种情形，对 $P \geq 0.05$ 或者获得了试验组显著劣效于对照组的试验，Obs ES 水平会远低于预期 (甚至反向)，即传统意义上的“阴性结果”，以本研究数据为基础，在纳入分析的所有研究中该分类出现频次为 5；如试验结果中 Obs ES 达到并超过试验设计阶段 Tgt ES 水平，对应的统计检验为 $P < 0.05$ ，相当于同时获得统计学及临床意义，在纳入本研究的所有 12 项试验中，有 5 项属于这一分类；最为特殊的一类情况是，试验虽获得了 $P < 0.05$ 的统计检验结果，但其 Obs ES 却没能达到试验设计时的假设水平，这一类试验存在仅获得了统计学意义但可能缺乏临床意义时的潜在风险，被归为“情形三”的试验在此次的研究对象中共出现 2 次。如果以获得了有统计学意义的 7 项试验为准，“阳性结果发表偏倚”的出现频次 (属于情形三的试验在情形二和情形三试验总体中的出现情况) 为 2，按 ACC 和 ESC 公布研究结果进行分层，这一问题的出现频次同为 1 次 (ACC: 1/3, ESC: 1/4)。用于识别在情形三的出现程度是否超过理论阈值的示意结果见图 3。

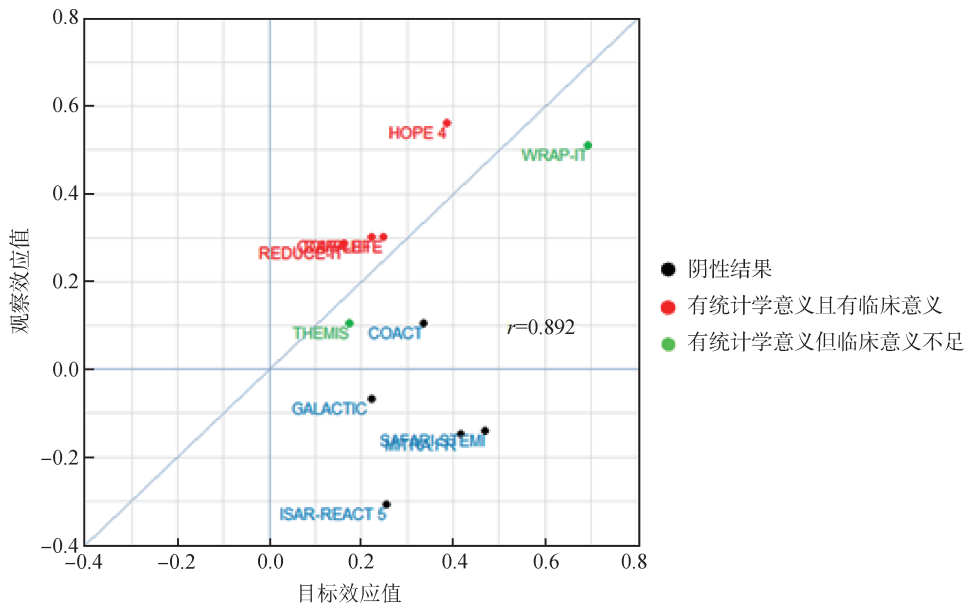


图 2 观察效应值与目标效应值间一致性对比

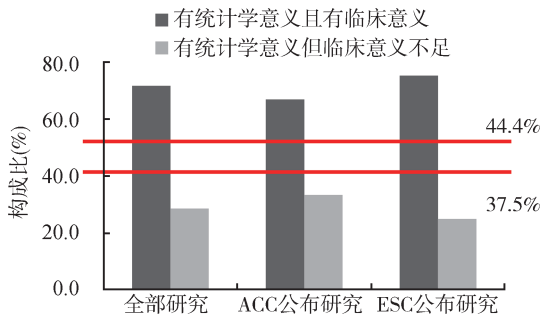


图 3 潜在阳性结果发表偏倚的识别示意图

讨 论

本研究尝试提出一种新的、基于临床意义与统计学意义相结合的临床试验结果分类方法,以发表的大规模临床试验结果为例,通过量化分析展示统计学意义与临床意义间可能存在的差异,特别需要注意的是当试验获得 $P < 0.05$ 的结果时,Obs ES(干预组与对照组间的差异)可能低于试验设计时的假设水平,而在试验设计阶段 Tgt ES 水平,往往是有临床意义的最小差异程度(minimum clinical important difference)。本研究提出了用临床意义与统计学意义相结合的方法对试验结果进行分类,并基于该分类结果进一步对潜在的“阳性结果发表偏倚”进行识别,并以真实临床试验案例开展分析和评价。

在对临床试验结果解读时,综合考虑统计学意义和临床意义有一定必要性,鉴于越来越多的学者提出,应避免在临床研究结果报告时使用传统 P 值,尤其要避免用检验 P 值作研究结果推断的唯一标准。重要的原因之一是统计学意义和临床意义间存在区别,虽然有其他的替代性方法被提出,但目前尚无公认的解决方案^[18-19]。统计检验及统计推断方法仍被普遍用于临床研究的数据分析,基于零假设(H_0 : null hypothesis)的统计检验结果通常被研究者作为“定性”结论的依据,如:当统计检验 $P < 0.05$ 时,认为差异有统计学意义或试验结果为阳性。但本研究结果提示,对传统统计学阳性的研究,仍有做进一步区分的必要,在判定存在仅获得有统计学意义的结果但临床意义不足的一类研究后,如果其所占比例过多(超过其合理阈值)时,则有理由认为在该类临床试验中有潜在的“阳性结果发表偏倚”,该类研究对应的干预措施的效力(efficacy)可能无法传递到临床实践中真正的效果(effectiveness)。

统计学意义与临床意义在很多情况下并不统

一,例如研究样本量过低,导致有临床意义的结果无统计学意义,该情形多出现在探索性研究;对于确证性研究,却往往由于样本量较大,更易得到差异有统计学意义的结果,但差异程度却缺少临床意义,这一类结果的合理解读和评价未被足够关注。事后把握度(post-hoc power)的计算能够一定程度对应到本研究结果,对达到有统计学意义但 Obs ES 小于 Tgt ES 的试验,其计算得到的事后把握度水平会低于预设值。但受限于单项试验结果的机会效应问题,仅通过事后把握度计算无法给出定性结论,而本研究则尝试通过荟萃流行病学(Meta-epidemiology)的研究方法,以临床和统计学意义相结合的方式先对试验结果进行明确分类,然后从多项研究汇总层面,提出“阳性结果发表偏倚”的概念及可能的识别方法,相当于在获得了有统计学意义结果的研究中,评估没有达到临床意义的试验出现的程度,该程度从假设检验原理上具有理论阈值,对预设把握度为 80% 或 90% 的试验,该阈值分别为 37.5% (30%/80%) 和 44.4% (40%/90%),通过评价结果和阈值的对比,对可能存在的阳性结果发表偏倚进行初步识别和判断。

本研究提出的阳性结果发表偏倚与传统发表偏倚概念不同^[20-21],本研究所纳入的文章中阴性结果接近半数(5/12),提示传统发表偏倚问题已经受到了足够的重视。但“阳性结果发表偏倚”却未被关注,若将评价对象限定在获得了统计学意义的研究后,再根据 Obs ES 与 Tgt ES 之差,可将试验结果进一步做分类,如果在这些有统计学意义的试验中,缺少临床意义的试验数量过多(超过理论阈值),将直接导致循证医学指南的不合理推荐,甚至给临床实践带来更为严重的影响。该现象构成了所谓的阳性结果发表偏倚,如何识别或避免过多的试验仅因有统计学意义(缺乏临床意义)被发表,这一问题应该被研究者、方法学人员等予以充分的重视。目前,本研究仅提出了阳性结果发表偏倚的概念及初步识别方式,针对其开展进一步的理论研究具有必要性,如提出针对阳性结果发表偏倚进行检验的统计量,并对其分析性质加以证明,同时可开展扩大规模的荟萃流行病学研究,分析并发现潜在的影响因素,对临床意义不充分风险进行可能的早期识别,以提高临床试验结果的可靠性。

本研究存在局限性。首先研究纳入的临床试验为单年度在 ACC 和 ESC 大会公布结果的研究,其代表性和样本量均有限,但本研究的主要目的在

于提出临床意义与统计学意义相结合的试验结果分类方法,及在分类基础上对潜在阳性结果发表偏倚的识别方式,并辅以实证研究(本研究实证研究部分样本量有限,其结果不作为任何统计推断的基础),重点提出新概念和方法;再者,本研究分析时纳入的效应值多来自生存分析获得的 *HR* 值,结果对其他相对效应值(*RR* 值或 *OR* 值)的适用性尚需进一步做评价,但从统计学原理可知,不同维度效应值间可以进行转化,或可以使用标准化组间差异进行统一,本研究中包括 2 项试验结果就是从绝对效应值向相对效应值进行了转换。最后,本研究仅纳入了优效性假设的试验,结果在向其他比较类型的试验进行外推时可能存在问题,但以非劣效试验为例,检验统计量同样可以通过计算进行相互转换,原则上适用于更广泛的试验设计类型评价。

在获得了有统计学意义结果的研究中,仍然可能存在临床意义不充分的可能,本研究拟提出一种新的临床试验结果分类方法及基于特定分类出现程度而衍生的新概念,即:“阳性结果发表偏倚”。重点在于通过对设计阶段 *Tgt ES* 与结果获得的 *Obs ES* 进行对比,在传统试验结论基础上做进一步细分,对仅获得有统计学意义而临床意义不足的试验予以充分重视,为临床试验结果合理解读提供一定的依据与支持。

利益冲突 所有作者均声明不存在利益冲突

参 考 文 献

- [1] Djulbegovic B, Guyatt GH. Progress in evidence-based medicine: a quarter century on[J]. *Lancet*, 2017, 390(10092):415-423. DOI:10.1016/S0140-6736(16)31592-6.
- [2] Begg CB, Mazumdar M. Operating characteristics of a rank correlation test for publication bias[J]. *Biometrics*, 1994, 50(4):1088-1101. DOI:10.2307/2533446.
- [3] Egger M, Smith GD, Schneider M, et al. Bias in Meta-analysis detected by a simple, graphical test[J]. *BMJ*, 1997, 315(7109):629-634. DOI:10.1136/bmj.315.7109.629.
- [4] Sutton AJ, Duval SJ, Tweedie RL, et al. Empirical assessment of effect of publication bias on Meta-analyses [J]. *BMJ*, 2000, 320(7249): 1574-1577. DOI: 10.1136/bmj.320.7249.1574.
- [5] Tarakji KG, Mittal S, Kennergren C, et al. Antibacterial envelope to prevent cardiac implantable device infection [J]. *N Engl J Med*, 2019, 380(20):1895-1905. DOI:10.1056/NEJMoa1901111.
- [6] Goldberg AC, Leiter LA, Stroes ESG, et al. Effect of bempedoic acid vs placebo added to maximally tolerated statins on low-density lipoprotein cholesterol in patients at high risk for cardiovascular disease:the CLEAR wisdom randomized clinical trial[J]. *JAMA*, 2019, 322(18): 1780-1788. DOI:10.1001/jama.2019.16585.
- [7] Bhatt DL, Steg PG, Miller M, et al. Cardiovascular risk reduction with icosapent ethyl for hypertriglyceridemia [J]. *N Engl J Med*, 2019, 380(1): 11-22. DOI: 10.1056/NEJMoa1812792.
- [8] Le May M, Wells G, So D, et al. Safety and efficacy of femoral access vs radial access in ST-segment elevation myocardial infarction: the SAFARI-STEMI randomized clinical trial[J]. *JAMA Cardiol*, 2020, 5(2): 126-134. DOI: 10.1001/jamacardio.2019.4852.
- [9] Lemkes JS, Janssens GN, van der Hoeven NW, et al. Coronary angiography after cardiac arrest without ST-segment elevation[J]. *N Engl J Med*, 2019, 380(15): 1397-1407. DOI:10.1056/NEJMoa1816897.
- [10] Steg PG, Bhatt DL, Simon T, et al. Ticagrelor in patients with stable coronary disease and diabetes[J]. *N Engl J Med*, 2019, 381(14):1309-1320. DOI:10.1056/NEJMoa1908077.
- [11] Mehta SR, Wood DA, Storey RF, et al. Complete revascularization with multivessel PCI for myocardial infarction[J]. *N Engl J Med*, 2019, 381(15): 1411-1421. DOI:10.1056/NEJMoa1907775.
- [12] McMurray JJV, Solomon SD, Inzucchi SE, et al. Dapagliflozin in patients with heart failure and reduced ejection fraction[J]. *N Engl J Med*, 2019, 381(21): 1995-2008. DOI:10.1056/NEJMoa1911303.
- [13] Schupke S, Neumann FJ, Menichelli M, et al. Ticagrelor or prasugrel in patients with acute coronary Syndromes[J]. *N Engl J Med*, 2019, 381(16):1524-1534. DOI: 10.1056/NEJMoa1908973.
- [14] Kozuharov N, Goudev A, Flores D, et al. Effect of a strategy of comprehensive vasodilation vs usual care on mortality and heart failure rehospitalization among patients with acute heart failure: the GALACTIC randomized clinical trial[J]. *JAMA*, 2019, 322(23): 2292-2302. DOI:10.1001/jama.2019.18598.
- [15] Schwalm JD, McCready T, Lopez-Jaramillo P, et al. A community-based comprehensive intervention to reduce cardiovascular risk in hypertension (HOPE 4): a cluster-randomised controlled trial[J]. *Lancet*, 2019, 394(10205): 1231-1242. DOI: 10.1016/S0140-6736(19) 31949-X.
- [16] Obadia JF, Messika-Zeitoun D, Leurent G, et al. Percutaneous repair or medical treatment for secondary mitral regurgitation[J]. *N Engl J Med*, 2018, 379(24): 2297-2306. DOI:10.1056/NEJMoa1805374.
- [17] Wang JM, Ke CL, Jiang Q, et al. Predicting analysis time in event-driven clinical trials with event-reporting lag[J]. *Stat Med*, 2012, 31(9):801-811. DOI:10.1002/sim.4506.
- [18] Wasserstein RL, Lazar NA. The ASA statement on *P*-values: context, process, and purpose[J]. *Am Stat*, 2016, 70(2): 129-133. DOI:10.1080/00031305.2016.1154108.
- [19] Ioannidis JPA. The proposal to lower *P* value thresholds to 0.005[J]. *JAMA*, 2018, 319(14):1429-1430. DOI:10.1001/jama.2018.1536.
- [20] Hopewell S, Loudon K, Clarke MJ, et al. Publication bias in clinical trials due to statistical significance or direction of trial results[J]. *Cochrane Database Syst Rev*, 2009(1): MR000006. DOI:10.1002/14651858.MR000006.pub3.
- [21] Murad MH, Chu HT, Lin LE, et al. The effect of publication bias magnitude and direction on the certainty in evidence [J]. *BMJ Evid Based Med*, 2018, 23(3):84-86. DOI:10.1136/bmjebm-2018-110891.