

· 思路与方法学 ·

上市后中药临床安全性循证证据体评价研究

廖 星^{1,2} 谢雁鸣¹

摘要 早期循证医学所倡导的证据分类和分级体系多基于不同研究设计类型针对干预效力的评价。这些体系给临床研究,给医学实践带来了巨大影响。随着时间推移,这些体系的应用逐渐出现了弊端。特别是在引入到现今真实世界研究和中药上市后安全性评价研究领域时,遇到了许多无法兼容的问题。国内外有关上市后药品安全性证据分级和分类研究尚不多见,研究者们发现系统地收集和评价有关上市后药品安全性证据有重要意义。借鉴循证医学有效性证据分类、分级的成功经验,探索对上市后中药安全性证据评价、分类和分级标准并形成安全性证据评价的基本框架,从点、线、面等多个角度出发,综合不同证据源,构建上市后中药安全性评价证据体,为今后该领域证据体系的研究提供方向和技术支持。

关键词 上市后中药再评价;安全性;证据评价;循证医学

Evaluating the Body of Evidence on Postmarketing Clinical Safety of Chinese Herbs LIAO Xing^{1,2} and XIE Yan-ming¹ 1 Institute of Basic Research in Clinical Medicine, China Academy of Chinese Medical Sciences, Beijing (100700); 2 Postdoctoral Research Center, China Academy of Chinese Medical Sciences, Beijing (100700)

ABSTRACT The traditional evidence evaluating hierarchy developed at the early stage in evidence-based medicine (EBM), founded on an efficacy evaluation model of intervention, is generalized to whole medical research and clinical practice. However, such a hierarchy system were increasingly abused and misunderstood, which has become an obstacle for clinical decision making process. The hierarchical model is valid for limited questions of efficacy and inadequate for the evaluation of interventions in real world, especially safety evidence for postmarketing Chinese medicines. There are very few studies reporting the evidence grading and classification of postmarketing herbs' safety at home and abroad. Therefore, systematical collection and evaluation of evidence of postmarketing herbs' safety are of important significance. By referring to successful experiences in evidence grading and classification of drug effectiveness in EBM, authors studied the establishment of standards for evaluation, classification and grading of safety evidence of postmarketing Chinese herbs, and the formation of basic frameworks of safety evidence evaluation, and comprehensively analyzed different sources of evidence from the perspectives of point, line, and surface, thus establishing the evidence body for safety evaluation of postmarketing Chinese herbs, and providing methodology and technical supports for future studies on evidence system in this field.

KEYWORDS postmarketing re-evaluation of Chinese herbs; safety; evidence evaluation; evidence-based medicine

随着循证医学 (evidence-based medicine, EBM) 的迅速发展以及不断渗入到医学研究和临床实

践领域,根据研究质量,对各种不同类型的研究进行分级评判形成不同级别的证据已经被广泛应用于医学领

基金项目:国家自然科学基金青年基金资助项目 (No. 81202776); 中国中医科学院客座研究员联合创新研究项目 (No. ZZ070817); 第九批中国中医科学院自主选题研究项目课题 (No. Z0406); 2014 年第七批博士后特别资助项目 (No. 2017T70202)

作者单位:1. 中国中医科学院中医临床基础医学研究所 (北京 100700); 2. 中国中医科学院博士后流动站 (北京 100700)

通讯作者: 谢雁鸣, Tel: 010-64093302, E-mail: datamining5288@163.com

DOI: 10. 7661/CJIM. 2017. 01. 0109

域,而且使广大医学工作者从中受益。但是,随之而来也出现了不当使用“证据分级”,比如用于干预措施效果评价的证据分级被滥用、误用到病因、诊断、预防或者药品不良事件/反应(adverse drug event/reaction, ADE/ADR)评价等^[1]。正如有研究者指出,证据分级不仅仅针对于临床干预疗效的评价,临床实际应用中除了疗效还有其他诸多方面的问题,如病因、诊断、药品不良反应等^[2,3]。再比如早期证据评价体系将所有的随机对照试验(randomized controlled trial, RCT)都置于观察性研究之上,作为证据最高级^[4]。作为金标准的 RCT 无可厚非地用于评价干预措施的疗效,但是对于疗效以外的评价,RCT 有很大的局限性,譬如很难进行药品不良反应的研究^[5]。根据传统的证据分级体系,临床医生可以很快从最高级别的证据,综合多个 RCT 的系统综述获得有关某种干预措施的疗效信息,但是对于非疗效相关的临床问题,如干预措施的风险,则无法全面地从以疗效评价为目的的这种证据评价体系中获得最佳证据,往往要转向观察性研究设计^[6]。观察性研究也可以提供“最佳”证据,甚至如病例报告、病例叙事(anecdotes)也能提供客观证据^[7,8]。虽然病例报告不是最佳的证据源,但是在有关报告药物 ADR 特别是罕见 ADR 时,有着不可忽视的作用^[9]。此外,定性研究也可以作为证据的参考信息^[10],比如来自专家的焦点组访谈或者共识讨论会。

1 国内外证据分级发展史简介

1963 年,证据分级(hierarchy of evidence)概念由两位美国社会科学家首先提出^[11]。第一次在医疗卫生研究领域正式大量使用证据分级,是 1979 年由加拿大定期体检特别工作组(Canadian Task Force on the Periodic Health Examination, CTF-PHE)提出^[12]。到 2004 年,才由“推荐等级的评估、制定与评价”(The Grading of Recommendations Assessment, Development and Evaluation, GRADE)工作组(<http://www.gradeworkinggroup.org/index.htm>)推出国际统一的证据分级和推荐意见标准,向全世界推广应用^[13]。

在国内,也有学者尝试在证据分级研究领域进行探索和创新,如有学者在传统医学证据体的构成及证据分级上就提出了比较可行的建议,如提及“证据体(evidence body)的连续性和多源性”^[14]。也有学者对证据分级评价系统及其在临床决策分析中的应用进行了探讨,如提及通过对“需要治疗的患者数(number needed to treat, NNT)”的比较分析获得

某种干预措施的疗效、毒副作用与费用三者之间的权衡关系,以此获得证据分级的客观依据^[15]。此外,还有学者确定了对于政府决策而探索决策与管理领域的证据分级^[16]。

2 证据评价在不断发展和适应新的需求

虽然这些主流证据评价体系在世界医学研究领域里起到了很大的影响,但是不可否认的一个事实是最强的推荐有可能来自于低级别的证据,最高质量的证据也不意味着可以预示最强的推荐。证据的分级和评价应该根据研究问题以及研究设计类型的质量不同而不同。临床实际中的问题,虽然以治疗/干预疗效评价为核心问题,但是还有许多其他方面的问题,比如诊断、治疗(除了有效性,还有安全性和经济学评价)、预防、病因,因此证据的评价不仅仅只是关注干预措施的疗效,而应该拓展其他问题证据的评价^[17,3]。在治疗领域里诞生的金字塔证据分级,将 RCT 视为最高证据级别的划分,就不宜用于预防和病因学研究领域,后两者更多时候以队列(观察性)研究为最高证据级别。那么如何区分不同问题来评价不同的证据呢?正如有研究者提出以下的评价规则^[18],如图 1。

根据图 1 内容,可见应该依据不同的研究问题选择不同类型的研究证据,比如 RCT 可作为回答治疗或干预措施疗效的最佳证据,病因学、伤害等问题多倾向于队列或病例对照研究,诊断学和疾病筛查等问题则适合用横断面设计,以及疾病自然史和预后等问题则适合运用队列设计。而对于医学证据缺乏的领域则往往可以依靠历史回顾性研究,甚至是日常经验。针对不同的研究问题,选择不同类型的证据以类分也是医学证据研究领域一种可行的做法。

证据分级本为医学研究者的一种主观科研行为,目的是为了获得客观的、可行的、适用的证据,为临床实践服务,因此这个过程或者说体系是在不断发展和变更的。所以即使是来自牛津大学循证医学中心如此权威机构发布的证据分级体系,自 1998 年诞生以来,也一直在接受挑战和争议,同时也随着临床实际应用中发现问题不断增多,如最新 2011 版的证据分级 II (Levels of Evidence 2),就是最好的说明^[19]。该更新的证据分级最鲜明的特点在于,证据的分级基于 7 个分类问题来划分,把临床问题分成治疗,预防,病因,诊断,预后,危害,经济学 7 个方面,更具有针对性和适用性。

随着 EBM 未来 5 个发展方向的改变:EBM 信息资源、临床循证指南、循证实践、探索最佳途径、确保临床决策符合患者价值观和喜好、在卫生政策领域引入循

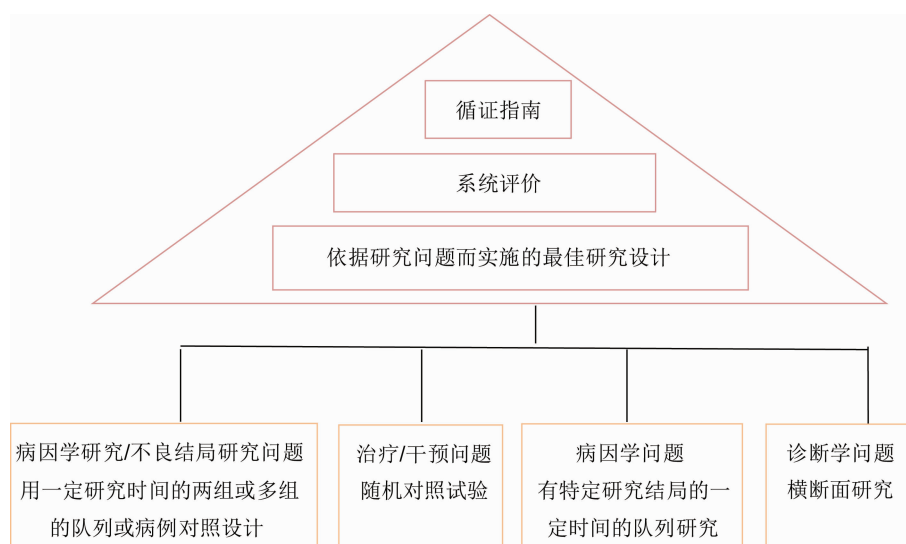


图 1 证据级别的评价规则

证原则^[20]。证据评价发展方向也会随之有所改变。正如牛津大学循证医学中心在研发证据评价体系时提出的一个核心问题:“在所有可获得不同治疗措施中,哪一种措施能够提供最佳的利/弊平衡?”^[21]。基于这样一个“理想化”的问题,如何获得关于这个问题的证据,将是下一个阶段 EBM 研究者朝向奋斗的一个目标。而有关“不同治疗措施的利弊权衡”正是当下医学领域研究热点“比较效益研究(comparative effectiveness research, CER)”关注的核心所在^[22]。

3 安全性证据评价的现状

有关“安全性”的英文表述词有:safety、risk、harm、adverse event、adverse effect、ADE/R。概而言之,就是医疗实践中产生了非预期的医学事件(有害结果),即“伤害”。在 EBM 领域里,不难发现很少有专门的 RCT 或 RCT 的系统评价专门用来评价安全性问题。笔者认为原因如下:RCT 无法有针对性地报告 ADE/ADR,或者因为干预疗程不够长,无法发现非常见的 ADE/ADR,或样本量太少,不能发现 ADE/ADR 有差异性研究结果。有关如何评价干预措施的“伤害”或安全性(harms/safety)?一直以来也是让医学研究者特别是 EBM 研究者所困扰的一个研究主题。所遇到的评价困难也很多,如一些罕见的难以发现的“伤害”并不能通过 RCT 来获得,反而是一些大样本、长期的观察性研究才是其发现的唯一证据^[23]。而一些特殊的身心伤害更是需要一些特别的其他研究设计才能获得证据^[24]。

另外,安全性评价的结果也有诸多不确定性,一旦结论不恰当,就会导致对患者的伤害。而且有关干预的不良结果通常来自于所谓的“低级证据形式”,如病

例报告、数据库回顾性分析、观察性研究等。面对这样一种现状,使得研究者在评价安全性证据时举足维艰。这也迫使研究者不得不重新思考评价安全性证据时理应区别于有效性证据的评价,而其评价标准和体系应当符合自身特点和规律。

目前针对如何获得并评价安全性证据,美国研究者已开始着手这方面的研究。他们认为:RCT 只有基于预先假设才能获得最有效因果推断的研究结果,但一定研究时间段里的用药并不能真正揭示实际用药人群的安全性问题,因此有关安全性结局评价的证据往往级别不高,即使是产生于一些 RCT 的结果评价也多是因“机遇”而产生^[25]。如何去获得高级别的有关安全性结局评价的证据?笔者认为可以针对报告了安全性结局的 RCT 进行系统评价,同时也可以针对其他非 RCT 报告的研究设计进行系统评价。值得注意的一点是,在对比安全性结局的研究中不仅应关注统计学上差异性,还应该关注在同一个可信区间里的临床差异性,这样也可以获得有关安全性结局评价的有用信息,而不致遗漏。另外,美国 FDA 就当下热门的注册登记研究中如何评价干预措施安全性提出了建议:研究结果是否真实?各种偏倚是否导致干预措施看起来更安全还是使其看起来可能会造成更大的伤害,而实际这些伤害却不是由于研究中的干预措施导致的^[26]?美国医疗保健研究与质量局也在其《效果和比较效益方法学指南综述》一书中正式撰写了“对比不同干预措施时如何评价伤害”一章(2009 年)^[27],该章节也随后被发表在国际临床流行病学杂志上^[28]。

由于大多数上市后药品 ADR(尤其严重 ADR)比较罕见,属于小概率事件,RCT 常不足以有效发现药

品安全事件(样本含量较小、观察时间较短、研究对象严格挑选等原因导致)。因此,国内有学者提出安全性评价证据分级标准与有效性评价存在差异,其最佳证据往往来自队列研究或病例对照研究,为此,拟定了一个供参考的有关药品安全性评价证据级别,并提到该评价标准的层次也只是相对的,当前的,并非永久不变^[29]。

4 中医药研究领域安全性证据评价的现状

过去 20 多年里,随着 EBM 的应用和推广,中医药临床有效性研究已经从系统评价和 RCT 中获得了重要的研究证据。然而,有关上市后中药安全性研究的证据整合和分级研究比较少见。

此外,对于目前中医药研究领域里大量涌现的系统评价,也有研究者报道其并没有给中医药临床实践带来太多有利影响,反而引发了一连串对于中医药开展 RCT 的反思^[30]。系统评价处于证据体系的高级别,但是多针对于干预措施的有效性,很少有单独针对安全性结局评价的系统评价,这或许是因为在 RCT 等干预性研究中有关安全性的信息比较少报道^[31,32],更或许是 RCT 并不是用来评价“伤害效应”的最佳设计^[33]。早在 2004 年英国约克大学研究者就提出,在进行系统评价的时候,不仅仅对评价干预措施的疗效结局时应该标准化,而对于如何去评价干预措施的“有害”结局时也应该系统标准化^[34]。研究者在进行评价“Harm”的系统评价时,应该首先确定好研究问题,比如评价长期效应,药物交互作用,某种对患者重要但不严重的伤害,不用仅仅局限于某种特殊的不良事件/反应。然而,如果研究的目的在于获得之前可疑的不良事件/反应,那么对于原始的监测数据的分析往往更加优于一个系统评价^[35]。换言之,有关 ADR 的证据,抛开 EBM 传统的研究设计类型,往往来自于其他渠道,如医院管理信息系统(Hospital Information System, HIS)或国家自发不良反应系统(spontaneous reporting system, SRS)数据也可以成为证据源。近年来,真实世界研究、比较效益研究、实效研究等新理念备受研究者们重视,其共同特点在于倡导开展更为符合日常医疗实践的各种真实世界的研究,形成真实世界证据,打破既往金字塔证据分级的束缚^[36]。为此,研究制定符合循证医学思想,满足上市后中药安全性研究领域和实践需要的高质量证据分类分级标准有着重要意义。

笔者所在研究团队基于既往研究工作,针对上市后中药安全性证据评价提出构建“安全性证据体”的理念^[37],借鉴循证医学有效性证据分类分级的成功经

验,探索对该研究领域证据整合和分级。根据当前可得证据,将前瞻性大样本长期的注册登记的医院集中监测研究视为最高级证据,将来自系统综述或 RCT 中报告的 ADR/ADE 视为第二级证据, HIS 数据回顾性分析或来自 SRS 中的分析视为第三级证据,医院临床实际中 ADR 个案病例讨论报告和文献中 ADR 个案报告以及其他研究类型报告的 ADR/ADE 视为第四级证据、专家意见和共识以及政府部门颁布的相关规范和标准视为第五级证据。如图 2。

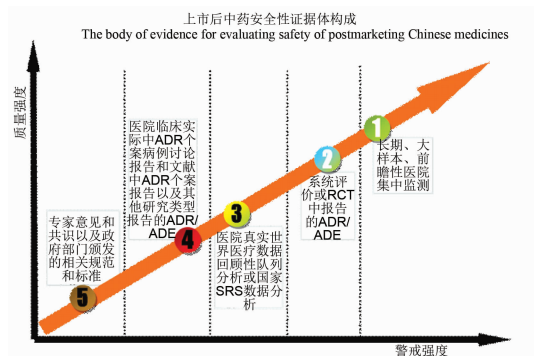


图2 上市后中药安全性证据体的构成

有关上市后中药安全性证据评价体系的构建需要从点、线、面相结合构成“证据体”。从点的角度来看,五种不同来源的证据可以分级进行评价,依据研究设计类型以及研究质量强度来判断警戒信息强度。从线的角度来看,凡是具备5种证据源中两种及其以上者则根据证据源之间结果是否一致获得警戒信息强度进行评价。例如:当长期、大样本、前瞻性的注册登记医院集中监测研究结果和来自国家药品不良反应中心 SRS 数据分析结果一致时可列为最高级证据;当系统评价和大样本 RCT 中报告的 ADR/ADE 结果一致时列为次高级证据,而单独来自两者的 ADR/ADE 报告级别则低于两者结果一致时的级别;医院真实世界医疗数据回顾性队列分析结果和来自国家药品不良反应中心 SRS 数据分析结果一致时可列为中级证据,而单独来自两者的分析结果级别则低于两者结果一致时的级别;多个医院临床实际中 ADR 个案病例讨论报告和文献中 ADR 个案报告以及其他研究类型报告的 ADR/ADE 一致时,可视为较低级别证据,而单独来自各个报告结果级别则低于结果一致时的级别;专家意见和共识以及政府部门颁布的相关规范和标准则为低级别证据。如表 1。

从面的角度来看,5种证据源可以形成一体化证据源,即如果不具备高质量的临床研究时,只要多个源头的研究结果都具备一致性,则所形成的证据具有高

表 1 多源证据证据分级列表

证据级别	定义
I	当长期、大样本、前瞻性的注册登记医院集中监测研究结果和来自国家药品不良反应中心 SRS 数据的结果一致时
II	当系统评价和大样本 RCT 中报告的 ADR/ADE 一致时
III	医院真实世界医疗数据回顾性队列分析结果和来自国家药品不良反应中心 SRS 数据分析结果一致时
IV	多个医院临床实际中 ADR 个案病例讨论报告和文献中 ADR 个案报告以及其他研究类型报告的 ADR/ADE 一致时
V	专家意见和共识以及政府部门颁布的相关规范和标准

度可信性,可以形成强的警戒信息。由于当前 EBM 领域有关药品安全性证据评价体系不能满足现实需求的现状,本研究建议基于以“证据体”,以点、线、面相结合的形式,进行灵活多元化的证据评价模式。从中药药物警戒研究角度出发,多角度、多层次发掘上市后中药安全性评价的证据,以全面指导临床实践。

总之,20 世纪 90 年代以来,循证医学的迅猛发展,证据分级的逐渐成熟,标志着一个以证据为基础的新医学时代的到来。临床证据的分类分级和临床推荐意见强度系统的逐渐成熟,但其内容复杂、应用局限、标准各异,对指导全球范围内各级医疗机构的循证实践并非万能。经典的证据评价等级自创制以来,一直在更新和发展,先后经历了“老五级”、“新五级”、“新九级”和“GRADE”等多个阶段。但其仍以强调临床有效性证据为主,没有包括生物医学领域的全部证据,可看作是对其他领域证据分级和推荐强度的示范标准。近年来,对上市后中药安全性评价问题越来越关注,引起了国家政府和中医药界的高度重视。研究者们越来越发现系统地收集和评价有关上市后中药安全性的研究证据有着时间的紧迫性和时代重要意义。因此,当前为了使应用者能够全面获得有关中药上市后安全性评价证据,特别是为政府药品监管与医疗主管部门开展医改、基本药物目录遴选、药品定价等提供多源证据,同时使得上市后中药的临床使用和研究信息在全世界范围内进行传播,系统、完整地对上市后中药相关的既有研究证据进行整合并获得一个证据评价框架体系,运用循证医学的理念进行全面研究非常必要。

参 考 文 献

[1] Glasziou P, Vandenbroucke JP, Chalmers I. Assessing the quality of research[J]. BMJ, 2004, 328(7430): 39-41.

[2] Sackett DL, Wennberg JE. Choosing the best research design for each question[J]. BMJ, 1997,

315(7123): 1636.

[3] Guyatt GH, Haynes RB, Jaeschke RZ, et al. Users' guides to the medical literature: X X V. Evidence-based medicine: principles for applying the users' guides to patient care[J]. JAMA, 2000, 284(10): 1290-1296.

[4] SUNY Downstate Medical Center. Medical Research Library of Brooklyn. Evidence Based Medicine Course. A Guide to Research Methods: The Evidence Pyramid [EB/OL]. [2013-10-10] <http://servers.medlib.hscbklyn.edu/ebm/2100.htm>.

[5] Howick J. The phylosophy of evidence-based medicine[M]. Oxford: Wiley-Blackwell, 2011:P36.

[6] Rosendaal FR. Bridging case-control studies and randomized trials[J]. Curr Control Trials Cardiovasc Med, 2001, 2(1): 109-110.

[7] Glasziou P, Chalmers I, Rawlins M, et al. When are randomised trials unnecessary? Picking signal from noise[J]. BMJ, 2007, 334(7589): 349-351.

[8] Aronson JK, Hauben M. Anecdotes that provide definitive evidence[J]. BMJ, 2006, 333(7581): 1267-1269.

[9] Vandenbroucke JP. In defense of case reports and case series[J]. Ann Intern Med, 2001, 134(4): 330-334.

[10] Van Schoor NM, Deville WL, Bouter LM, et al. Acceptance and compliance with external hip protectors: a systematic review of the literature[J]. Osteoporosis Int, 2002, 12(13): 917-924.

[11] Campbell DT, Stanley JC. Experimental and quasi-experimental designs for research[M]. Chicago, IL: Rand McNally College, 1963:P5-6.

[12] Canadian Task Force on the Periodic Health Examination. The periodic health examination[J]. CMAJ, 1979, 121(9): 1193-1254.

[13] Atkins D, Best D, Briss PA, et al. Grading quality of evidence and strength of recommendations[J]. BMJ, 2004, 328(7454): 1490-1494.

[14] 刘建平. 传统医学证据体的构成及证据分级的建议[J]. 中国中西医结合杂志, 2007, 27(12): 1061-1065.

[15] 康德英,王家良,洪旗,等. 证据分级评价系统及其在临床决策分析中的应用[J]. 中华流行病学杂志, 2004, 25(4): 351-354.

[16] 陈耀龙,李幼平,杜亮,等. 医学研究中证据分级和推荐强度的演进[J]. 中国循证医学杂志, 2008, 8(2): 127-133.

[17] Sackett DL, Wennberg JE. Choosing the best research design for each question[J]. BMJ, 1997,

- 315(7123): 1636.
- [18] Evidence Analysis Manual: steps in the academy evidence analysis process [OL]. http://www.and-eal.org/vault/2440/web/files/2012_Aug_EA_Manual.pdf.
 - [19] OCEBM Levels of Evidence Working Group. "The Oxford 2011 Levels of Evidence". Oxford Centre for Evidence-Based Medicine [EB/OL]. [2013 - 11 - 11]. <http://www.cebm.net/index.aspx?o=5653>.
 - [20] Gordon H. Guyatt. Five directions for evidence-based medicine [J]. Chin J Evid-Based Med, 2006, 6(3): 157 - 161.
 - [21] Howick J, Chalmers I, Glasziou P, et al. "Explanation of the 2011 Oxford Centre for Evidence-Based Medicine (OCEBM) Levels of Evidence (Background Document)". Oxford Centre for Evidence-Based Medicine[EB/OL]. [2013 - 12 - 12]. <http://www.cebm.net/index.aspx?o=5653>.
 - [22] 谢雁鸣,廖星,王永炎. 将比较效益研究理念引入中医临床评价研究[J]. 中西医结合学报, 2011, 9(8): 813 - 818.
 - [23] Eikelboom JW, Mehta SR, Pogue J, et al. Safety outcomes in meta-analyses of phase 2 vs phase 3 randomized trials: Intracranial hemorrhage in trials of bolus thrombolytic therapy [J]. JAMA, 2001, 285(4): 444 - 450.
 - [24] Scaf-Klomp W, Sanderma R, Wiel HB, et al. Distressed or relieved? Psychological side effects of breast cancer screening in The Netherlands[J]. J Epidemiol Community Health, 1997, 51(6): 705 - 710.
 - [25] Delfini Group, 2010. Delfini Appraisal Tool.Safety Issues & Considerations [EB/OL]. [2013 - 12 - 11]. http://www.delfini.org/Delfini_Pearls_Basics_Safety.pdf.
 - [26] Delfini Group, 2010. Delfini Appraisal Tool. Registry Evaluation Checklist: Safety of Interventions [EB/OL]. [2013 - 10 - 11]. http://www.delfini.org/Delfini_Tool_RegistryAppraisalChecklist_Safety-Interventions.pdf.
 - [27] Chou R, Aronson N, Atkins D, et al. Assessing harms when comparing medical interventions. In: Agency for Healthcare Research and Quality. Methods Guide for Comparative Effectiveness Reviews [EB/OL]. <http://www.effectivehealthcare.ahrq.gov/index.cfm/search-for-guides-reviews-and-reports/?pageaction=displayproduct&productid=318>.
 - [28] Chou R, Aronson N, Atkins D, et al. AHRQ Series Paper 4: Assessing harms when comparing medical interventions: AHRQ and the Effective Health Care Program[J]. J Clin Epidemiol, 2010, 63(5): 502 - 512.
 - [29] 李幼平,文进,王莉. 药品风险管理:概念、原则、研究方法与实践[J]. 中国循证医学杂志, 2007, 7(12): 843 - 884.
 - [30] 罗辉,刘建平.从系统综述看中国随机对照试验的质量[J]. 中西医结合学报, 2011, 9(7): 697 - 701.
 - [31] Ioannidis JPA, Lau J. Completeness of safety reporting in randomized trials: an evaluation of 7 medical areas[J]. JAMA, 2001, 285(4): 437 - 443.
 - [32] Loke YK, Derry S. Reporting of adverse drug reactions in randomized controlled trials—a systematic survey [J]. BMC Clin Pharmacol, 2001, 1: 3.
 - [33] Cuervo GL, Clarke M. Balancing benefits and harms in health care [J]. BMJ, 2003, 327(7406): 65 - 66.
 - [34] McIntosh HM, Woolacott NF, Bagnall AM. Assessing harmful effects in systematic reviews [J]. BMC Med Res Methodol, 2004, 4(1): 1 - 6.
 - [35] Medawar C, Herxheimer A. A comparison of adverse drug reaction reports from professionals and users, relating to risk of dependence and suicidal behaviour with paroxetine[J]. Int J Risk Saf Med, 2004, 16: 5 - 19.
 - [36] Velentgas P, Dreyer NA, Nourjah P, et al. Developing a protocol for observational comparative effectiveness research: a user's guide [M]. Rockville, MD: Agency for Healthcare Research and Quality, 2013: 1 - 163.
 - [37] Liao X, Robinson N. Methodological approaches to developing and establishing the body of evidence on post-marketing Chinese medicine safety[J]. Chin J Integr Med, 2013, 19(7): 494 - 497.

(收稿:2013 - 12 - 29 修回:2016 - 06 - 01)