

识别诊断试验准确性系统综述的方法学异质性

李志霞 杨智荣 项骁 高培 舒正 黄元升 曹宇 孙凤 詹思延

100191 北京大学公共卫生学院流行病与卫生统计学系

通信作者:詹思延, Email:siyan-zhan@bjmu.edu.cn

DOI: 10.3760/cma.j.issn.0254-6450.2016.02.027

【摘要】 目的 分析诊断试验系统综述的方法学异质性来源。方法 通过检索2008年1月1日至2012年12月31日发表在Medline、Embase、Cochrane(CDSR和DARE)数据库中关于诊断试验准确性(DTA)的Meta分析文献,纳入了至少包含10篇原始研究的Meta分析。两名人员独立地对研究特点以及原始研究的数据进行提取,使用混合线性模型对14种潜在的方法学异质性因素来源进行探讨,进而计算出14种异质性因素的诊断比值比(DOR)之比(RDOR)值及其95%CI,从而判断其异质性大小和方向。结果 最终纳入了23篇DTA的系统综述,涵盖550篇原始研究。纳入的550篇原始研究的质量良好。单因素混合线性模型分析显示,“金标准是否足够准确”和“金标准与待检试验是否相互独立”是DTA系统综述的异质性来源。多因素混合线性模型分析显示,金标准不准确的诊断试验,其DOR合并值低于金标准足够准确的诊断试验,RDOR=0.018 6(95%CI:0.001 0~0.358 5);金标准与待检试验不独立的诊断试验,其DOR合并值高于金标准与待检试验相互独立的诊断试验,RDOR=2.396 6(95%CI:1.242 8~4.622 7)。结论 对于诊断试验系统综述,原始研究的金标准是否足够准确、金标准与待检试验是否相互独立为其方法学异质性的来源。

【关键词】 诊断试验准确性;系统综述;异质性

基金项目:高等学校博士学科点专项科研基金(20120001110015)

Methodological bias and variation of systematic reviews on diagnostic test accuracy Li Zhixia, Yang Zhirong, Xiang Xiao, Gao Pei, Shu Zheng, Huang Yuansheng, Cao Yu, Sun Feng, Zhan Siyan
Department of Epidemiology and Bio-statistics, Peking University Health Science Center, Beijing 100191, China

Corresponding author: Zhan Siyan, Email: siyan-zhan@bjmu.edu.cn

【Abstract】 Objective To analyze methodological bias and variation of systematic reviews on diagnostic test accuracy (DTA). **Methods** Meta-analyses on DTA were identified through an electronic search through databases as Medline, Embase and Cochrane between 1 January 2008 and 31 December 2012. Results from Meta-analyses on 10 primary studies were included. Pairs of reviewers worked independently to extract the related data of interest, together with those original data of the primary studies. Mixed linear model was used to investigate the direction and strength of the association among the 14 studies, featuring on estimates of the diagnostic accuracy. **Results** A total of 23 papers on Meta-analyses with 550 primary studies were included. Results from mixed linear model showed that significant low estimates of diagnostic accuracy in studies unsatisfying “the reference standard would likely to correctly classify the target condition” [relative diagnostic odds ratio (RDOR)=0.018 6, 95% CI: 0.001 0–0.358 5]. Studies whose reference standard were not independent of the index test produced significantly higher estimates of diagnostic accuracy (RDOR=2.396 6, 95% CI:1.242 8–4.622 7). **Conclusion** Messages as “Is the reference standard likely to correctly classify the target condition?” and “Was the reference standard independent of the index test”, were the origin of the methodological bias and variation of systematic reviews on diagnostic test accuracy.

【Key words】 Diagnostic test accuracy; Systematic reviews; Bias and variation

Fund program: Fund for the Doctoral Program of Higher Education (20120001110015)

在诊断试验准确性(diagnostic test accuracy, DTA)系统综述和Meta分析(合称DTA系统综述)

中,除了对各原始研究的效应值进行定量合并以外,探讨其异质性来源也非常重要。有研究表明^[1],一

些已发表的诊断试验原始研究,由于研究设计的质量缺陷,其质量控制关键点出现漏洞。这一方面可能会引入偏倚,高估或低估诊断试验的准确性,另一方面会使研究之间出现方法学异质性,导致研究间不同的结果。Rutjes等^[2]2006年的一项研究结果显示,连续纳入患者以及回顾性收集数据会高估试验结果,依据待检试验纳入患者会低估试验结果。然而迄今为止,针对DTA系统综述的方法学异质性进行探讨的研究仍然匮乏。因此,本文采用两水平多变量的混合线性模型,同时兼顾系统综述层面和原始研究层面的效应,拟探讨QUADAS质量评估工具的14个条目是否为DTA系统综述方法学的异质性因素^[3]。

资料与方法

1. 检索策略:2013年12月在Medline、Embase、Cochrane(CDSR和DARE)3个数据库中,系统检索2008年1月1日至2012年12月31日发表的DTA系统综述,语言限制为英文。检索策略由两组检索词构成,一组是关于DTA研究的检索词,另一组是关于系统综述、Meta分析的检索词,两组检索词以“AND”连接,检索流程见图1。

2. 纳入排除标准:纳入标准:进行Meta分析的DTA系统综述;纳入原始研究的数目≥10;提供了原始研究的四格表数据;提供了QUADAS质量评估量表的14个条目评价结果的数据。其中,如果同一篇DTA系统综述包含了多个Meta分析,则只纳入原始研究数量最多的Meta分析。排除标准:重复发表文献;方法学研究;会议摘要、通信、评论及无法获得完整信息的其他文献;研究方案。

3. 文献信息摘录及质量评价:通过EpiData 3.1软件制作信息提取表,提取文献的基本信息(作者、年份、Meta分析中纳入原始研究的数量、目标疾病、待检试验、金标准)及Meta分析中所纳入的原始研究信息[四格表数据(真阳性、假阳性、真阴性、假阴性)和QUADAS质量评估的14个条目信息(14个条目均以“是”、“否”、“不确定”进行回答,其中“是”记1分,“否”和“不确定”记0分,每个条目最高得分为1分)]。以上过程均由两人独立平行完成,意见不一致时由第三人裁决。

4. 统计学分析:用SAS 9.3软件拟合广义混合线性模型,即从原始研究水平建立分层合并受试者工作特征(ROC)随机效应的Meta回归模型^[4],对纳入研究的系统综述中相关原始研究结果进行拟合,

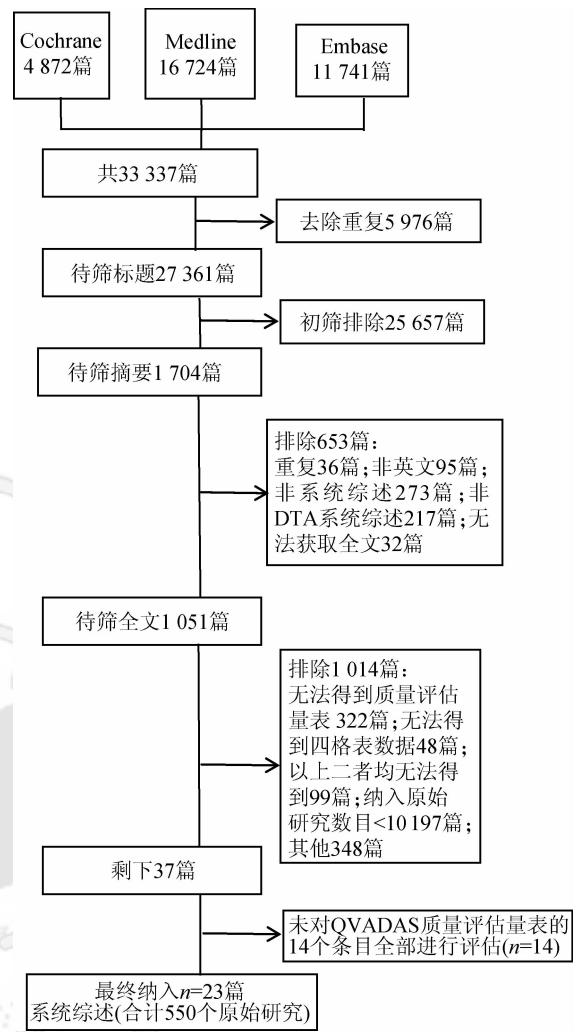


图1 文献纳入排除流程

应变量为各原始研究的诊断比值比(DOR)^[5]的对数logDOR,协变量为纳入的质量评估工具的条目(即异质性因素m)。模型说明:

$$\log DOR_{ij} = \alpha_i + \beta_j S_{ij} + \sum (r_m + v_{im}) X_{ijm} + \eta_{ij} + \varepsilon_{ij}$$

其中, v_{im} 、 η_{ij} 、 ε_{ij} 为随机效应项,均服从正态分布; $Est = 1/a_{ij} + 1/b_{ij} + 1/c_{ij} + 1/d_{ij}$ (其中a, b, c, d为四格表数据, Est代表每个原始研究的标准误)logDOR_{ij}表示第i个系统综述第j个原始研究的logDOR; α_i 为常数项,表示第i个系统综述中总的准确度; β_j 为系数项,表示第i个系统综述中S的变化; $S_{ij} = \text{logit}(SEN) + \text{logit}(1 - SEP)$; r_m 代表第m个异质性因素的平均效应值; v_{im} 代表第m个异质性因素在第i个系统综述中的残差; X_{ijm} 表示第i个系统综述第j个原始研究中第m个异质性因素的取值;最后,通过Excel 2010软件计算出 r_m 及其95%CI的反对数,进而得到DOR之比^[5](RDOR,其意义为具有某种偏倚风险因素研究的DOR合并值是不具有某种相应因素的研究的多少倍)及其95%CI。若某项研究的四格表中含有0,则

对该研究的 tp , fp , fn , tn (即 a, b, c, d) 分别加 0.5 进行校正。

结 果

1. 纳入研究基本特征: 本研究为高等学校博士学位点专项科研基金课题《诊断试验系统综述中识别和处理方法学异质性的研究》中的一个子课题, 该基金课题共检索到相关文献 33 337 篇, 本项子课题在其检出文献的基础上, 根据纳入和排除标准最终纳入分析的有 23 篇, 合计 550 篇原始研究。

23 篇文献的基本特征见表 1。各个 Meta 分析纳入原始研究的数量, 最小为 10^[6-7], 最大为 115^[28]; 评估的目标疾病涵盖了冠状动脉疾病、关节韧带疾病、曲霉病、肺部疾病、癌症、糖尿病、神经系统疾病和败血病; 待检试验包括影像学检查、实验室检测和

患者的主观感受测评。

纳入的 550 篇原始研究, 其 QUADAS 质量评估的均值为 8.79。满足条目 1~14 的原始研究数量分别为 376、318、529、244、450、452、515、327、273、343、238、255、247 和 265, 其中符合条目 3 和条目 7 的研究最多, 比例达 96.18% 和 93.64%, 由此可见大部分原始研究选择的金标准较为准确, 见表 2。

2. 混合线性模型分析: 用 SAS 9.3 软件的 Proc Mixed 模块, 对纳入的 23 篇 Meta 分析, 合计 550 篇原始研究, 进行混合线性模型分析, 建立分层合并 ROC 随机效应的 Meta 回归模型 (表 3)。

(1) 单因素分析: 混合线性模型单因素分析显示, 金标准不准确的诊断试验, 其 *DOR* 合并值是金标准足够准确的诊断试验的 0.026 8 (95% *CI*: 0.001 7~0.427 3) 倍。金标准与待检试验不独立的

表 1 23 篇识别诊断试验准确性系统综述的基本特征

第一作者	发表年份	Meta 分析 纳入原始 文献数量	目标“疾病”	待检试验	金标准
Yin ^[6]	2012	10	术前负荷反应能力	马西莫脉搏血氧计测压	检查心脏指数、心输出量的改变以及中风指标
van der Windt ^[7]	2010	10	神经根病	任何角度的腿部疼痛检查	外科手术
Wu ^[8]	2012	11	败血病	等离子体 sTREM-1 检测	微生物培养
Gargiulo ^[9]	2011	11	冠状动脉疾病	心肌灌注显像/心肌灌注显像	冠状动脉血管造影
Smith ^[10]	2012	12	三角纤维软骨损伤	X 射线检查	关节摄影术
Ringer ^[11]	2012	12	疼痛综合征	骨显像检查	检测 C 反应蛋白
Neto ^[12]	2012	12	精神错乱	利用谵妄筛查工具进行评价: 重症监护患者的评估 (CAM-ICU), 重症监护谵妄筛查表 (ICDSC), 护理谵妄筛查量表 (NuDESC), 谵妄检测评分 (DDS), Neelon 和 Champagne 混乱量表 (NEECHAM)	精神病学专家、老年医学专家或神经学家依据精神障碍进行诊断和判断
Jellema ^[13]	2010	12	结肠直肠癌	腹痛检查	结肠镜检查, 钡灌肠, 临床随访
Santin ^[14]	2012	12	肺结核和结核感染	γ -干扰素释放化验	培养
Chen ^[15]	2012	13	肝转移癌	磁共振成像检查	组织病理学分析 (手术, 活组织切片检查)、术中照射 (人工触诊, 术中超声), 随访
Ye ^[16]	2012	15	糖尿病	口服葡萄糖耐量试验 (OGTT)	2 h 口服葡萄糖耐量试验 + FPG
Wang ^[17]	2011	17	胃癌	影像检查	手术或组织病理学检查
Ngamrueng-phong ^[18]	2010	17	食道癌新辅助治疗的反应	正电子发射计算机断层成像 (FDG-PET)	病理检查
Lucassen ^[19]	2011	19	肺栓塞	D 二聚体检测	肺扫描、计算机断层扫描、肺血管造影或尸检, 深静脉血栓形成的诊断
Ottenheijm ^[20]	2010	22	肩关节障碍	超声检查	手术、磁共振成像和/或摄影
Smith ^[21]	2012	26	肩袖撕裂	磁共振成像 (MRI)	外科手术
Xia ^[22]	2010	26	膀胱癌	反转录 PCR	病理检查
Steingart ^[23]	2011	28	活性肺结核	ELISA	培养或涂片镜检
Smith ^[24]	2012	30	关节窝上唇损伤	磁共振血管造影 (MRA)	关节镜检查
Leefflang ^[25]	2008	30	曲霉病	半乳甘露聚糖检测	欧洲癌症研究和治疗组织/侵袭性真菌感染协作组和美国国立变态反应和感染病研究院真菌病研究组共识标准 (EORTC/MSG)
Smith ^[26]	2011	37	肩袖撕裂	超声检查	关节镜或开放手术
Smith ^[27]	2012	53	前交叉韧带破裂	MRI	关节镜下直接可视化的破裂
Parker ^[28]	2012	115	冠状动脉狭窄	单光子发射计算机断层成像术 (SPECT)	冠状动脉造影

表2 550个原始研究的QUADAS 14个条目质量评估

条目	内容	满足条件的原始研究数量	满足条件的原始研究比例(%)
1	患者是否有代表性	376	68.36
2	纳入标准是否准确	318	57.82
3	金标准是否足够准确	529	96.18
4	待检试验和金标准执行的时间间隔是否合理	244	44.36
5	是否所有的患者都接受了金标准的验证	450	81.82
6	接受的金标准是否为同一个	452	82.18
7	金标准与待检试验是否相互独立	515	93.64
8	待检试验描述是否足够清楚可重复	327	59.45
9	金标准的描述是否足够清楚可重复	273	49.64
10	解读待检试验的结果时,不知道金标准的结果吗	343	62.36
11	解读金标准的结果时,不知道待检试验的结果吗	238	43.27
12	结果是否具有临床可推广性	255	46.36
13	是否报告了不可解释的/中间结局	247	44.91
14	是否解释了退出病例	265	48.18

注:14个条目均以“是”、“否”、“不确定”进行回答,其中“是”记1分,“否”和“不确定”记0分,每个条目最高得分为1分

表3 DTA系统综述14种潜在方法学异质性因素来源的混合线性模型分析

条目	平均效应值(r_m)	$RDOR_m$
单因素分析		
1	-11.749 5(-2.004 4 ~ 3.042 3)	1.216 5(0.469 1 ~ 3.154 3)
2	5.238 3(-2.501 3 ~ 1.279 6)	0.644 3(0.165 4 ~ 2.510 7)
3	0.636 3(2.707 8 ~ 0.360 5)	0.026 8(0.001 7 ~ 0.427 3)
4	10.817 8(-2.537 4 ~ 1.727 1)	0.808 3(0.263 6 ~ 2.478 0)
5	-4.207 0(-1.257 6 ~ 3.127 0)	1.728 6(0.478 9 ~ 6.240 2)
6	-5.030 2(-1.238 1 ~ 2.437 8)	1.580 5(0.388 9 ~ 6.422 4)
7	-2.844 1(-53.676 9 ~ -1.460 7)	2.247 0(1.043 8 ~ 4.837 3)
8	3.252 0(-1.425 5 ~ 0.759 6)	0.492 6(0.048 3 ~ 5.029 2)
9	-38.819 9(-4.127 1 ~ 5.241 1)	1.061 1(0.644 5 ~ 1.747 0)
10	2.145 9(-4.512 6 ~ 0.866 9)	0.342 0(0.070 2 ~ 1.665 7)
11	50.075 1(-2.231 2 ~ 2.048 8)	0.955 1(0.35 0 ~ 2.806 7)
12	16.917 6(-1.765 9 ~ 1.460 9)	0.872 8(0.206 8 ~ 3.683 8)
13	-3.718 9(-1.401 9 ~ 5.698 0)	1.857 4(0.667 6 ~ 5.167 7)
14	9.293 7(-2.462 5 ~ 1.609 5)	0.780 5(0.239 2 ~ 2.547 4)
多因素分析		
3	0.578 2(0.331 8 ~ 2.244 7)	0.018 6(0.001 0 ~ 0.358 5)
7	-2.634 4(-10.594 3 ~ -1.504 0)	2.396 6(1.242 8 ~ 4.622 7)

注:以符合条目的作对照

诊断试验,其DOR合并值是金标准与待检试验相互独立的诊断试验的2.247 0(95%CI:1.043 8 ~ 4.837 3)倍。条目2、4、8、10、11、12、14的RDOR值均<1,条目1、5、6、9、13的RDOR值均>1,但差异无统计学意义(95%CI均包含1)。

(2)多因素分析:根据单因素分析结果,将具有统计学意义的条目(条目3和条目7)共同纳入混合线性模型,进行多因素分析,金标准与待检试验不独立的诊断试验,其DOR合并值是金标准与待检试验相互独立的诊断试验的2.396 6(95%CI:1.242 8 ~

4.622 7)倍;金标准不准确的诊断试验,其DOR合并值是金标准足够准确的诊断试验的0.018 6(0.001 0 ~ 0.358 5)倍。

讨论

本研究中混合线性模型单因素分析和多因素分析的结果一致,结果均表明,金标准不准确会低估DTA系统综述的准确性,多因素分析RDOR=0.018 6(95%CI:0.001 0 ~ 0.358 5);金标准与待检试验不独立会高估DTA系统综述的准确性,多因素分析RDOR=2.396 6(95%CI:1.242 8 ~ 4.622 7)。这一结论与Whiting等^[29]2013年的一项研究结果一致,其研究认为金标准是否合适会对DTA的结果产生影响。此外,Lijmer等^[30]1999年的研究结果发现,未清楚描述金标准的诊断试验与清楚描述金标准的诊断试验相比,RDOR=0.7(95%CI:0.6 ~ 0.9),也同本研究的结果相一致。从理论层面进行分析,以上两种情况的发生均会导致错分偏倚,其中金标准不准确使得真阳性和真阴性的个数降低,从而低估了结果的准确性;而金标准与待检试验不独立则使得假阳性和假阴性的个数大大降低,从而高估了结果的准确性。

此外,本研究中混合线性模型单因素分析的结果提示,“纳入标准不足够准确”、“待检试验和金标准执行的时间间隔不合理”、“待检试验描述不足够清楚不可重复”、“解读待检试验的结果时,知道金标准的结果”、“解读金标准的结果时,知道待检试验的结果”、“结果不具有临床可推广性”和“未解释了退出病例”,会低估DTA系统综述的准确性;“患者没有代表性”、“部分患者接受了金标准的验证”、“接受的金标准不是同一个”、“金标准的描述不足够清楚不可重复”和“未报告了不可解释的/中间结局”,会高估DTA系统综述的准确性。

与既往研究相比,本研究采用混合线性模型探讨了DTA系统综述的14种异质性来源,该模型是两水平多变量模型,同时兼顾了系统综述层面和原始研究层面的效应,在模型参数的处理方面,计算也更为精确。然而,本研究同时存在很多局限性。本研究中所探讨的只是有关DTA方法学方面的14种设计因素,随着DTA系统综述的不断规范化以及文章数量的增多,有更多的异质性因素值得我们去探究。此外,由于本研究所纳入的文献样本量较少,而

混合线性模型对样本量的要求较高,随机效应因子的个数有限,因此只能将单因素有意义的结果放入模型进行进一步验证;而且本研究涉及的诊断领域较多(涵盖了影像学检查、实验室检测和患者的主观感受测评等多个领域),一方面导致了条目 3 和条目 7 结果的 95%CI 较宽,另一方面使得大部分条目的结果无统计学意义,这些异质性因素是否会真正导致结局的变异,还有待于今后纳入更多的 DTA 系统综述进一步去验证。

利益冲突 无

参 考 文 献

- [1] Reid MC, Lachs MS, Feinstein AR. Use of methodological standards in diagnostic test research. Getting better but still not good [J]. *JAMA*, 1995, 274 (8) : 645-651. DOI: 10.1001/jama.1995.03530080061042.
- [2] Rutjes AWS, Reitsma JB, Di Nisio M, et al. Evidence of bias and variation in diagnostic accuracy studies [J]. *CMAJ*, 2006, 174 (4) : 469-476. DOI: 10.1503/cmaj.050090.
- [3] Whiting P, Rutjes AW, Reitsma JB, et al. The development of QUADAS: a tool for the quality assessment of studies of diagnostic accuracy included in systematic reviews [J]. *BMC Med Res Methodol*, 2003, 3 : 25. DOI: 10.1186/1471-2288-3-25.
- [4] Harbord RM, Whiting P, Sterne JAC, et al. An empirical comparison of methods for meta-analysis of diagnostic accuracy showed hierarchical models are necessary [J]. *J Clin Epidemiol*, 2008, 61 (11) : 1095-1103. DOI: 10.1016/j.jclinepi.2007.09.013.
- [5] The Cochrane Collaboration. *Cochrane handbook for systematic reviews of diagnostic test accuracy* [M]. London: The Cochrane Collaboration, 2013.
- [6] Yin JY, Ho KM. Use of plethysmographic variability index derived from the Massimo® pulse oximeter to predict fluid or preload responsiveness: A systematic review and meta-analysis [J]. *Anaesthesia*, 2012, 67 (7) : 777-783. DOI: 10.1111/j.1365-2044.2012.07117.x.
- [7] van der Windt DA, Simons E, Riphagen II, et al. Physical examination for lumbar radiculopathy due to disc herniation in patients with low-back pain [J]. *Cochrane Database Syst Rev*, 2010, 17 (2) : CD007431. DOI: 10.1002/14651858.CD007431.
- [8] Wu YP, Wang F, Fan XH, et al. Accuracy of plasma sTREM-1 for sepsis diagnosis in systemic inflammatory patients: a systematic review and meta-analysis [J]. *Critical Care*, 2012, 16 (6) : R229. DOI: 10.1186/cc11884.
- [9] Gargiulo P, Petretta M, Bruzzese D, et al. Myocardial perfusion scintigraphy and echocardiography for detecting coronary artery disease in hypertensive patients: A meta-analysis [J]. *Eur J Nucl Med Mol Imaging*, 2011, 38 (11) : 2040-2049. DOI: 10.1007/s00259-011-1891-0.
- [10] Smith TO, Drew BT, Toms AP, et al. The diagnostic accuracy of X-ray arthrography for triangular fibrocartilaginous complex injury: a systematic review and meta-analysis [J]. *J Hand Surg Eur Vol*, 2012, 37 (9) : 879-887. DOI: 10.1177/1753193411402762.
- [11] Ringer R, Wertli M, Bachmann LM, et al. Concordance of qualitative bone scintigraphy results with presence of clinical complex regional pain syndrome I: Meta-analysis of test accuracy studies [J]. *Eur J Pain*, 2012, 16 (10) : 1347-1356. DOI: 10.1002/j.1532-2149.2012.00137.x.
- [12] Neto AS, Nassar AP Jr, Cardoso SO, et al. Delirium screening in critically ill patients: A systematic review and meta-analysis [J]. *Crit Care Med*, 2012, 40 (6) : 1946-1951. DOI: 10.1097/CCM.0b013e31824e16c9.
- [13] Jellema P, van Der Windt DA, Bruinvels DJ, et al. Value of symptoms and additional diagnostic tests for colorectal cancer in primary care: Systematic review and meta-analysis [J]. *BMJ*, 2010, 340 : c1269. DOI: 10.1136/bmj.c1269.
- [14] Santin M, Munoz L, Rigau D. Interferon- γ release assays for the diagnosis of tuberculosis and tuberculosis infection in HIV-infected adults: a systematic review and meta-analysis [J]. *PLoS One*, 2012, 7 (3) : e32482. DOI: 10.1371/journal.pone.0032482.
- [15] Chen L, Zhang J, Zhang L, et al. Meta-analysis of gadoteric acid disodium (Gd-EOB-DTPA)-enhanced magnetic resonance imaging for the detection of liver metastases [J]. *PLoS One*, 2012, 7 (11) : e48681. DOI: 10.1371/journal.pone.0048681.
- [16] Ye YC, Xie HZ, Zhao XL, et al. The oral glucose tolerance test for the diagnosis of diabetes mellitus in patients during acute coronary syndrome hospitalization: a meta-analysis of diagnostic test accuracy [J]. *Cardiovasc Diabetol*, 2012, 11 : 155. DOI: 10.1186/1475-2840-11-155.
- [17] Wang Z, Chen JQ. Imaging in assessing hepatic and peritoneal metastases of gastric cancer: a systematic review [J]. *BMC gastroenterol*, 2011, 11 : 19. DOI: 10.1186/1471-230X-11-19.
- [18] Ngamruengphong S, Sharma VK, Nguyen B, et al. Assessment of response to neoadjuvant therapy in esophageal cancer: An updated systematic review of diagnostic accuracy of endoscopic ultrasonography and fluorodeoxyglucose positron emission tomography [J]. *Dis Esophagus*, 2010, 23 (3) : 216-231. DOI: 10.1111/j.1442-2050.2009.00989.x.
- [19] Lucassen W, Geersing GJ, Erkens PMG, et al. Clinical decision rules for excluding pulmonary embolism: a meta-analysis [J]. *Ann Intern Med*, 2011, 155 (7) : 448-460. DOI: 10.7326/0003-4819-155-7-201110040-00007.
- [20] Ottenheijm RP, Jansen MJ, Staal JB, et al. Accuracy of diagnostic ultrasound in patients with suspected subacromial disorders: a systematic review and meta-analysis [J]. *Arch Phys Med Rehabil*, 2010, 91 (10) : 1616-1625. DOI: 10.1016/j.apmr.2010.07.017.
- [21] Smith TO, Daniell H, Geere JA, et al. The diagnostic accuracy of MRI for the detection of partial- and full-thickness rotator cuff tears in adults [J]. *Magn Reson Imaging*, 2012, 30 (3) : 336-346. DOI: 10.1016/j.mri.2011.12.008.
- [22] Xia Y, Liu YL, Yang KH, et al. The diagnostic value of urine-based survivin mRNA test using reverse transcription-polymerase chain reaction for bladder cancer: A systematic review [J]. *Chin J Cancer*, 2010, 29 (4) : 441-446.
- [23] Steingart KR, Flores LL, Dendukuri N, et al. Commercial Serological tests for the diagnosis of active pulmonary and extrapulmonary tuberculosis: an updated systematic review and meta-analysis [J]. *PLoS Med*, 2011, 8 (8) : e1001062. DOI: 10.1371/journal.pmed.1001062.
- [24] Smith TO, Drew BT, Toms AP. A meta-analysis of the diagnostic test accuracy of MRA and MRI for the detection of glenoid labral injury [J]. *Arch Orthop Trauma Surg*, 2012, 132 (7) : 905-919. DOI: 10.1007/s00402-012-1493-8.
- [25] Leeftang MM, Debets-Ossenkopp YJ, Visser CE, et al. Galactomannan detection for invasive aspergillosis in immunocompromised patients [J]. *Cochrane Database Syst Rev*, 2008 (4) : CD007394. DOI: 10.1002/14651858.CD007394.
- [26] Smith TO, Back T, Toms AP, et al. Diagnostic accuracy of ultrasound for rotator cuff tears in adults: a systematic review and meta-analysis [J]. *Clin Radiol*, 2011, 66 (11) : 1036-1048. DOI: 10.1016/j.crad.2011.05.007.
- [27] Smith TO, Lewis M, Song F, et al. The diagnostic accuracy of anterior cruciate ligament rupture using magnetic resonance imaging: a meta-analysis [J]. *Eur J Orthop Surg Traumatol*, 2012, 22 (4) : 315-326. DOI: 10.1007/s00590-011-0829-3.
- [28] Parker MW, Iskandar A, Limone B, et al. Diagnostic accuracy of cardiac positron emission tomography versus single photon emission computed tomography for coronary artery disease: a bivariate meta-analysis [J]. *Circ Cardiovasc Imaging*, 2012, 5 (6) : 700-707. DOI: 10.1161/CIRCIMAGING.112.978270.
- [29] Whiting PF, Rutjes AW, Westwood ME, et al. A systematic review classifies sources of bias and variation in diagnostic test accuracy studies [J]. *J Clin Epidemiol*, 2013, 66 (10) : 1093-1104. DOI: 10.1016/j.jclinepi.2013.05.014.
- [30] Lijmer JG, Mol BW, Heisterkamp S, et al. Empirical evidence of design-related bias in studies of diagnostic tests [J]. *JAMA*, 1999, 282 (11) : 1061-1066. DOI: 10.1001/jama.282.11.1061.

(收稿日期: 2015-07-09)

(本文编辑: 万玉立)