



团 体 标 准

T/GDPA 1—2021

超说明书用药循证评价规范

The specification of evidence-base pharmaceutical evaluation
method for off-label drug use

2021-05-21 发布

2021-05-21 实施

广东省药学会 发布

前 言

本文件按照GB/T1.1—2020《标准化工作导则 第1部分：标准化文件的结构和起草规则》的规定起草。

本文件由广东省药学会团体标准委员会归口。

本文件负责起草单位：中山大学孙逸仙纪念医院。

本文件主要起草人：伍俊妍、邱凯锋、陈泽鹏、何志超、朱建红。

本文件指导专家：郑志华、杨敏。

请注意本文件的某些内容可能涉及专利。本文件的发布机构不承担识别专利的责任。

超说明书用药循证评价规范

1 范围

本文件确立了运用循证药学的方法，结合超说明书用药的特点，制定了超说明书用药循证评价的程序。

本文件适用于药品超说明书用法的循证评价，包括对药品使用中的适应证、给药剂量、适用人群或给药途径不在药品监督管理部门批准的说明书之内的用法进行有效性、安全性及证据质量的评估。

本文件不适用于药品说明书内用法。

2 规范性引用文件

下列文件中的内容通过文中的规范性引用而构成本文件必不可少的条款。其中，注日期的引用文件，仅该日期对应的版本适用于本文件；不注日期的引用文件，其最新版本（包括所有的修改单）适用于本文件。

粤药会[2010]8号 药品未注册用法专家共识

粤药会[2014]72号 医疗机构超药品说明书用药管理专家共识

3 术语和定义

下列术语和定义适用于本文件。

3.1

药品说明书 dispensatory

药品说明书是药品生产企业经国家药品监督管理部门批准，用以指导安全、合理使用药品的技术性资料。

3.2

超说明书用药 off-label use

又称药品未注册用法，是指药品使用的适应证、给药剂量、适用人群或给药途径等不在药品监督管理部门批准的说明书之内的用法。

3.3

循证药物评价 evidence-based drug evaluation

指全面、系统收集药物临床研究与使用证据，严格综合评价药物用于疾病预防、治疗实际过程中的安全性、有效性、经济性和适应性。

4 服务对象

医师、药师、护师、政策制定者、医药卫生管理者等在工作中接触超说明书用药的相关人员。

5 设备要求

药品相关信息数据库：包括但不限于国家药品监督管理局官网、各大循证医学数据库、临床实践指南、临床研究数据库。

6 操作规程（见图1）

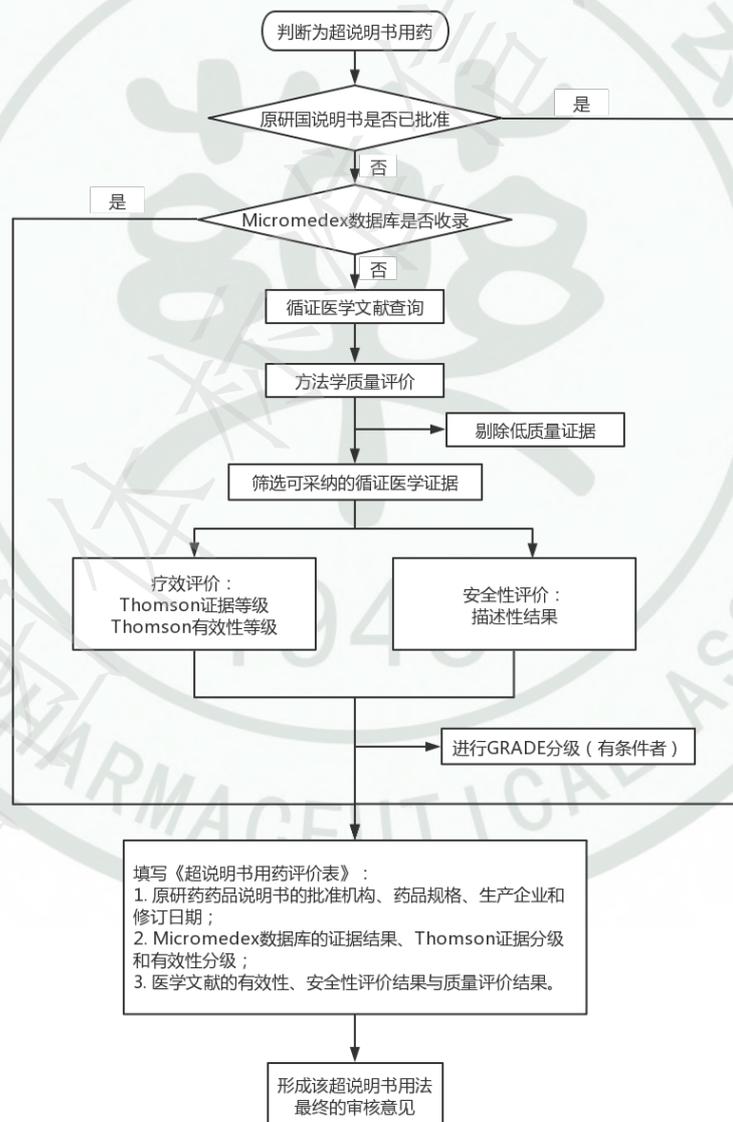


图1 超说明书循证评价流程图

6.1 超说明书的判断

查询国内相关的具体药品说明书，判定是否为超说明书用药，并确定超说明书的类型。

超说明书用药类型主要包括：超适应证用药、超剂量用药、超适用人群用药、超给药途径用药等类型。

6.2 药品原研国说明书的查证

查询相关药物的原研药品说明书中是否有批准相关的内容。包括美国食品药品监督管理局、欧洲药物管理局和日本药品与医疗器械管理局等全球政府机构批准的药品说明书，主要通过该国的官方网站进行查询。若已列入原研药品说明书则直接填写《超说明书用药循证评价表》（见附录A）中“原研国药品说明书循证审核”的对应内容。

原研国的说明书可查询的主要网站链接有：美国FDA官网：<http://www.fda.gov>；欧洲EMA官网：<http://www.ema.europa.eu/ema>；日本PDMA官网：<http://www.pmda.go.jp>等。

6.3 循证数据库的查证

若原研药品说明书未批准相关内容，则查询 Micromedex 数据库，确定其有效性等级、推荐等级、证据等级。对于 Micromedex 数据库已经收录的超说明书用药，可直接引用 Micromedex 的有效性等级、推荐等级、证据等级。而对于原研药品说明书已收录但中国说明书中未有的超说明书用药，可不再进行证据等级与有效性等级判定。

填写《超说明书用药循证评价表》（见附录A）中“Micromedex 数据库循证审核”的对应内容。

6.4 医学文献的查证

对于 Micromedex 数据库没有收录的以及原研药品说明书亦未增加的超说明书用药，则查询相关的医学文献。

医学文献检索顺序按照证据级别从高到低的原则依次为：临床指南、系统评价/Meta分析、随机对照试验、非随机试验、队列研究和病例对照研究。检索范围应覆盖专业的数据库和指南网，包括但不限于 Cochrane 协作网、EMBASE、PubMed 等大型的医学文献数据库，以及美国国家指南文库、加拿大医学会临床实践指南、英国国家卫生与临床优化研究所和苏格兰校际指南网、中国循证医学网，中国学术期刊全文数据库、中国生物医学文献数据库、万方数据库等。

填写《超说明书用药循证评价表》（见附录A）中“医学文献的查证”对应内容，具体方法见 6.5。

6.5 医学文献的质量评价

6.5.1 评价临床指南/专家共识的 AGREE II 工具

AGREE II 评估系统包括 6 个领域 23 个条目及 2 个总体评估条目。每个领域针对指南质量评价的一个特定问题（见表 B.1）。

a) 领域 1，范围和目的（条目 1-3）：涉及指南的总目的，特定卫生问题和目标人群。

b) 领域 2, 参与人员 (条目 4-6): 涉及指南开发小组成员组成的合理程度, 并能代表目标使用人群的观点。

c) 领域 3, 严谨性 (条目 7-14): 涉及证据的收集和综合过程、陈述和更新推荐建议的方法。

d) 领域 4, 清晰性 (条目 15-17): 涉及指南的语言、结构及表现形式。

e) 领域 5, 应用性 (条目 18-21): 涉及指南实施过程中的有利条件和潜在不利因素及其改进策略, 以及应用指南涉及的相关资源问题。

f) 领域 6, 独立性 (条目 22-23): 涉及指南推荐建议的产生不受相关利益竞争的影响和左右。

AGREE II 所有条目均按 7 分等级量表进行评分。如果没有与 AGREE II 条目相关的信息或者报告的概念非常差, 则为 1 分; 如果报告的质量很高, 满足用户手册要求的所有标准和条件, 则为 7 分; 如果条目报道不能满足全部标准或条件, 则根据不同情况给予 2~6 分。分值分配取决于报告的完整性和质量。当更多的标准被满足且理由更充分时, 则分值增加。AGREE II 分别计算各领域的质量分值。6 个领域的分值是独立的, 不能将其合并为一个单一的质量分值。每个领域得分等于该领域中每一个条目分数的总和, 并标准化为该领域可能的最高分数的百分比。

示例: 4 位评价者给领域 1 的评分, 则最大可能得分 = 7 (很同意) × 3 (条目) × 4 (评价者) = 84; 最小可能得分 = 1 (很不同意) × 3 (条目) × 4 (评价者) = 12。领域分值为: (实际得分-最小可能得分) / (最大可能得分-最小可能得分) × 100%。如果评价过程中没有包括某些条目, 则需要适当修改最大可能得分和最小可能得分的计算方法。

6.5.2 评价系统评价/Meta 分析的 AMSTAR 工具

AMSTAR 用于衡量系统评价/Meta 分析避免或减少偏倚的程度, 即方法学质量的一种量表, 是国际强烈推荐的方法学质量评价量表。其评分内容: 量表共 11 个条目, 每个条目下包含若干子条目, 各条目用“完整报告”“部分报告”和“未报告”描述, 分别计 1 分、0.5 分、0 分, 共 11 分; 评价结果: 0~4 分为低质量, 5~8 分为中等质量, 9~11 分为高质量 (见表 B.2)。

6.5.3 评价随机对照试验的 Jadad 量表

Jadad 量表又称为 Jadad 评分或牛津评分系统, 是评价临床试验方法学的质量工具, 其定义的临床试验质量只包括内部有效性, 即临床试验本身的质量。其评价内容: 量表从随机序列的产生、随机化隐藏、盲法、退出与失访等四个方面进行评价, 每方面用“恰当”“不清楚”和“不恰当”描述, 分别计 2 分、1 分、0 分, 共 7 分; 评价结果: 1~3 分为低质量, 4~7 分为高质量 (见表 B.3)。

6.5.4 评价非随机对照干预性研究的 MINORS 工具

MINORS 评价指标适用于非随机对照干预性研究质量的评价。评价指标共 12 条, 每一条分为 0~2 分。前 8 条针对无对照组的研究, 最高分为 16 分; 后 4 条与前 8 条一起针对有对照组的研究, 最高分共 24 分, 总分 0~12 分为低质量文献, 13~18 为中等质量, 19~24 为高质量。0 分表示未报道; 1 分表示报道了但信息不充分; 2 分表示报道了且提供了充分的信息 (见表 B.4)。

6.5.5 评价非随机对照观察性研究的 NOS 量表

NOS 量表适用于评价病例对照研究和队列研究。它通过三大块共 8 个条目的方法评价队列研究 (见表 B.5) 和病例对照研究 (见表 B.6), 具体包括研究人群选择 (selection)、可比性 (comparability)、

暴露 (exposure) 评价或结果 (outcome) 评价。NOS 对文献质量的评价采用了星级系统的半量化原则, 满分为 10 颗星。

6.6 医学研究证据的有效性与安全性评价

对医学文献进行方法学质量评价后, 根据所纳入临床研究进行疗效与安全性评价。

有效性评价根据 Thomson 分级定义进行有效性分级与证据分级 (见表 B.7)。

安全性评价暂时没有权威统一的评价分析方法, 因此以描述性语言对药物安全性结果进行总结, 重点介绍超说明书用法直接相关的内容。

6.7 对审核结果进行 GRADE 分级评价

建议有条件的单位对审核结果进行 GRADE 评价。GRADE 系统是由 2000 年建立的 GRADE 工作组创立的一套证据评级系统, 于 2004 年正式推出。GRADE 系统使用易于理解的方式评价证据质量和推荐等级, 目前已被 WHO、Cochrane 协作网等 100 多个国际组织或协会采用 (见表 B.8)。

与目前存在的其他众多分级标准相比, GRADE 系统具有以下优势: ①由一个具有广泛代表性的国际指南制定小组制定; ②明确界定了证据质量和推荐强度; ③清楚评价了不同治疗方案的重要结局; ④对不同级别证据的升级与降级有明确、综合的标准; ⑤从证据评级到推荐意见强度全过程透明; ⑥明确承认患者价值观和意愿; ⑦就推荐意见的强弱, 分别从临床医生、患者、政策制定者角度做了明确实用的诠释; ⑧适用于制作系统评价、卫生技术评估及指南。

GRADE 系统将证据质量分为“高、中、低和极低”四个等级, 将推荐强度分为“强推荐和弱推荐”两个等级, 并提供了用以描述的符号、字母或数字。

6.8 超说明书用药评价结果的填写

将相关文献资料和评价结果填入《超说明书用药循证评价表》(见附录 A), 并于表中结果部分填写审核结论。提交上级审核部门 (如医院药事管理与药物治疗学委员会) 进行审核。

审核结论应对超说明书的查找材料进行汇总, 包括原研国说明书收录情况、Micromedex 的分级结果、医学文献的有效性、安全性评价结果与质量评价结果, 并结合以上资料对该超说明书用药给出推荐意见。

医学文献循证评价表			
医学文献循证审核类型			
医学文献类别	纳入循证审核数量	疗效评价结论	安全性评价结论
临床指南/专家共识			
系统评价			
随机对照试验			
非随机对照试验			
队列研究			
病例对照研究			
个案报道			
医学书籍			
其它: _____			
<p>医学文献循证查询结论汇总:</p> <p>疗效评价结论汇总意见:</p> <p>安全性评价结论汇总意见:</p>			
<p>超说明书用药循证评价综合推荐意见:</p> <p><input type="checkbox"/> 强推荐 <input type="checkbox"/> 弱推荐</p>			
评价人		复核人	
评价时间		复核时间	

附 录 B
(资料性附录)

表 B.1 评价临床指南/专家共识的 AGREE II 工具

评价领域	条 目	评价内容	评审者 1 评分	评审者 2 评分	评审者 3 评分	评审者 4 评分
范围和目的	1	明确阐述指南的总目的				
	2	明确阐述指南所涵盖的卫生问题				
	3	明确阐述指南所要应用的人群				
指南制定参与人员	4	指南制定组包括所有相关专业的人员				
	5	考虑到目标人群（患者、公众等）的观点和选择				
	6	指南的适用者已经明确规定				
	7	用系统的方法检索证据				
制定严谨性	8	明确阐述了选择证据的标准				
	9	清楚地描述证据群的优势和不足				
	10	明确阐述形成推荐意见的方法				
	11	在形成推荐意见时考虑了对健康的效益、副作用以及风险				
	12	推荐意见和支持证据之间有明确的联系				
	13	指南在发表前经过专家的外部评审				
	14	提供指南更新的过程				
明晰陈述与表达	15	推荐意见明确不含糊				
	16	明确列出针对某一情况或卫生问题不同的选择				
	17	主要的推荐意见清晰易辨				
适用性	18	指南中描述了应用过程中的促进和阻碍因素				
	19	指南提供了推荐意见如何应用于实践的建议和（或）配套工具				
	20	指南考虑了应用推荐建议时潜在的资源投入问题				
	21	指南提供了监控和（或）审计的标准				
编撰的独立性	22	赞助单位的观点不影响指南的内容				
	23	指南记录并考虑了制定小组成员的利益冲突				
合 计						

注：AGREE II 所有条目均按7分等级量表进行评分。如果没有与AGREE II 条目相关的信息或者报告的概念非常差，则为1分；如果报告的质量很高，满足用户手册要求的所有标准和条件，则为7分；如果条目报道不能满足全部标准或条件，则根据不同情况给予2~6分。分值分配取决于报告的完整性和质量。当更多的标准被满足且理由更充分时，则分值增加。AGREE II 分别计算各领域的质量分值。6个领域的分值是独立的，不能将其合并为一个单一的质量分值。每个领域得分等于该领域中每一个条目分数的总和，并标准化为该领域可能的最高分数的百分比。

表 B.2 评价系统评价/Meta 分析的 AMSTAR 工具

评价条目	描述及说明
1	是否提供了前期设计方案? 在系统评价开展以前, 应该确定研究问题及纳入排除标准
2	纳入研究的选择和数据提取是否具有可重复性? 至少要有两名独立数据提取员, 而且采用合理的不同意见达成一致的方法过程
3	是否实施广泛全面的文献检索? 至少检索 2 种电子数据库。检索报告必须包括年份以及数据库, 如 Central、EMbase 和 MEDLINE。必须说明采用的关键词/主题词, 如果可能应提供检索策略 应咨询最新信息的目录、综述、教科书、专业注册库或特定领域的专家, 进行额外检索, 同时还可检索文献后的参考文献
4	发表情况是否已考虑在纳入标准中, 如灰色文献? 应该说明评价者的检索是不受发表类型的限制 应该说明评价者是否根据文献的发表情况排除文献, 如语言
5	是否提供了纳入和排除的研究文献清单? 应该提供纳入和排除的研究文献清单
6	是否描述纳入研究的特征? 原始研究提取的数据应包括受试者、干预措施和结局指标等信息, 并以诸如表格的形成进行总结 应该报告纳入研究的一系列特征, 如年龄、种族、性别、相关社会经济学数据、疾病情况、病程、严重程度等
7	是否评价和报道纳入研究的科学性? 应提供预先设计的评价方法, 如治疗性研究, 评价者是否把随机、双盲、安慰剂对照、分配隐藏作为评价标准, 其它类型研究的相关标准条目一样要交代
8	纳入研究的科学性是否恰当地运用在结论的推导上? 在分析结果和推导结论中, 应考虑方法学的严格性和科学性。在形成推荐意见时, 同样需要明确说明
9	合成纳入研究结果的方法是否恰当? 对于合成结果, 应采用一定的统计检验方法确定纳入研究是可合并的, 以及评估它们的异质性性 (如 Chi-squared test)。如果存在异质性, 应采用随机效应模型, 和/或考虑合成结果的临床适宜程度, 如合并结果是否敏感?
10	是否评估了发表偏倚的可能性? 发表偏倚评估应含有某一种图表的辅助, 如漏斗图、以及其他可行的检测方法和/ 或统计学检验方法, 如 Egger' s 或 Begger' s 回归
11	是否说明相关利益冲突? 应清楚交待系统评价及纳入研究中潜在的资助来源

注: 各条目用“完整报告”“部分报告”和“未报告”描述, 分别计1分、0.5分、0分, 共11分; 评价结果: 0~4分为低质量, 5~8分为中等质量, 9~11分为高质量。

表 B.3 评价随机对照试验的 Jadad 量表

评价项目	分值与内容	描述
随机序列的产生	2分：计算机产生的随机数字或类似方法	恰当
	1分：随机试验但未描述随机分配的方法	不清楚
	0分：采用交替分配的方法如单双号	不恰当
分配隐藏	2分：中心或药房控制分配方案、或用序列编号一致的容器、现场计算机控制、密封不透光的信封或其他使临床医生和受试者无法预知分配序列的方法	恰当
	1分：只表明使用随机数字表或其他随机分配方案	不清楚
	0分：交替分配、病例号、星期日数、开放式随机号码表、系列编码信封以及任何不能防止分组的可预测性的措施、未使用	不恰当
盲法	2分：采用了完全一致的安慰剂片或类似方法	恰当
	1分：试验陈述为盲法，但未描述方法	不清楚
	0分：未采用双盲或盲的方法不恰当，如片剂和注射剂比较	不恰当
退出与失访	1分：描述了退出或失访的数目和理由	恰当
	0分：未描述退出或失访数目或理由	不清楚
		不恰当

注：量表从随机序列的产生、随机化隐藏、盲法、退出与失访等四个方面进行评价，每方面用“恰当”“不清楚”和“不恰当”描述，分别计2分、1分、0分，共7分；评价结果：1~3分为低质量，4~7分为高质量。

表 B.4 评价非随机对照干预性研究的 MINORS 工具

序号	条目	提示
1	明确的给出了研究目的	所定义的问题应该是精确的且与可获得文献有关
2	纳入患者的连贯性	所有具有潜在可能的患者（满足纳入标准）都在研究期间被纳入了（无排除或给出了排除的理由）
3	预期数据的收集	收集了根据研究开始前制定的研究方案中设定的数据
4	终点指标能恰当的反映研究目的	明确的解释用来评价与所定义的问题一致的结局指标标准。同时应在意向性治疗分析的基础上对终点指标进行评估
5	终点指标评估的客观性	对客观终点指标的评价采用评价者单盲法，主观终点指标采用双盲法。否则，应给出未行盲法的理由
6	随访时间是否充足	随访时间应足够长，以使得能对终点指标及可能的不良事件进行分析
7	失访率低于 5%	应对所有的患者进行随访。失访比例不得超过反映主要终点指标的患 者比例
8	是否估算了样本量	根据预期结局事件的发生率，计算了可测出不同研究结局的样本量及其 95%可信区间；且提供的信息能够从显著统计学差异及估算把握水平对预期结果与实验结果进行比较
9~12 条 用于评价有对照组研究的附加标准		
9	对照组的选择是否恰当	对于诊断性试验，应为诊断的“金标准”；对于治疗干预试验，应是能从已发表研究中
10	对照组是否同步	对照组与实验组应该是同期进行的（非历史对照）
11	组间基线是否可比	不同于研究终点，对照组与试验组起点的基线标准应该具有相似性。无可能导致使结果解释产生偏倚的混杂因素
12	统计分析是否恰当	用于计算可信区间或相对危险度的统计资料是否与研究类型相匹配

注：评价指标共12条，每一条分为0~2分，最高分共24分，总分0~12分为低质量文献，13~18为中等质量，19~24为高质量。0分表示未报道；1分表示报道了但信息不充分；2分表示报道了且提供了充分的信息。

表 B.5 评价队列研究的 NOS 量表

栏目	条目	评价标准
研究人群选择	暴露组的代表性如何 (1分)	①真正代表人群中暴露组的特征*；②一定程度上代表了人群中暴露组的特征*；③选择某类人群，如护士、志愿者；④未描述暴露组来源情况
	非暴露组的选择方法 (1分)	①与暴露组来自同一人群*；②与暴露组来自不同人群；③未描述非暴露组来源情况
	暴露因素的确定方法 (1分)	①固定的档案记录 (如外科手术记录)*；②采用结构式访谈*；③研究对象自己写的报告；④未描述
	确定研究起始时尚无要观察的结局指标 (1分)	①是*；②否
组间可比性	设计和统计分析时考虑暴露组和未暴露组的可比性 (2分)	①研究控制了最重要的混杂因素*；②研究控制了任何其他混杂因素* (此条可以进行修改用以说明特定控制第二重要因素)
结果测量	结果测量研究对于结果的评价是否充分 (1分)	①盲法独立评价*；②有档案记录*；③自我报告；④未描述
	结果发生后随访是否足够长 (1分)	①是 (评价前规定恰当的随访时间)*；②否
	暴露组和非暴露组的随访是否充分 (1分)	①随访完整*；②有少量研究对象失访但不至于引入偏倚 (规定失访率或描述失访情况)*；③有失访 (规定失访率) 但未行描述；④未描述随访情况

注：*给分点；从队列的选择、可比性和结果测量3个方面进行评价，评价后星数 (☆) 越多质量越好。

表 B. 6 评价病例对照研究的 NOS 量表

栏目	条目	评价标准
研究人群选择	病例确定是否恰当 (1 分)	①恰当, 有独立的确定方法或人员*; ②恰当, 如基于档案记录或自我报告; ③未描述
	病例的代表性 (1 分)	①连续或有代表性的系列病例*; ②有潜在选择偏倚或未描述
	对照的选择 (1 分)	①与病例同一人群的对照*; ②与病例同一人群的住院人员为对照; ③未描述
	对照的确定 (1 分)	①无目标疾病史 (端点)*; ②未描述来源
组间可比性	设计和统计分析时考虑病例和对照的可比性 (2 分)	①研究控制了最重要的混杂因素*; ②研究控制了任何其他混杂因素* (此条可以进行修改用以说明特定控制第二重要因素)
暴露因素的测量	暴露因素的确定 (1 分)	①固定的档案记录 (如外科手术记录)*; ②采用结构式访谈且不知访谈者是病例或对照*; ③采用未实施盲法的访谈 (即知道病例或对照的情况); ④未描述
	采用相同的方法确定病例和对照组暴露因素 (1 分)	①是*; ②否
	无应答率 (1 分)	①病例和对照组无应答率相同*; ②描述了无应答者的情况; ③病例和对照组无应答率不同且未描述

注: *给分点; 该标准包括3个方面的评价: 病例组与对照组选择方法、病例组与对照组的可比性、接触暴露评估方法, 评价后星数 (☆) 越多质量越好。

表 B. 7 医学文献有效性的 Thomson 分级

有效等级		证据等级	
Class I	治疗有效	A 类	随机对照试验的荟萃分析; 多个、设计良好、大规模的随机临床试验;
Class IIa	证据支持有效	B 类	结论冲突的随机对照试验的荟萃分析; 小规模或研究方法有显著缺陷的随机对照试验; 非随机试验; 非随机研究
Class IIb	有效性具有争议	C 类	专家意见或共识; 个案报道或系列案例
Class III	治疗无效	无证据	没有证据

表 B.8 GRADE 分级系统

证据级别	具体描述	研究类型	总分	表达符号/字母
高级证据	我们非常确信真实的效应值接近效应估计	RCT; 质量升高二级的观察性研究	≥0 分	⊕⊕⊕⊕/A
中级证据	对效应估计值我们有中等程度的信心：真实值有可能接近估计值，但仍存在二者大不相同的可能性	质量降低一级的 RCT; 质量升高一级的观察性研究	- 1 分	⊕⊕⊕○/B
低级证据	我们对效应估计值的确信程度有限：真实值可能与估计值大不相同	质量降低二级的 RCT; 观察性研究	- 2 分	⊕⊕○○/C
极低级证据	我们对效应估计值几乎没有信心：真实值很可能与估计值大不相同	质量降低三级的 RCT; 质量降低一级的观察性研究; 系列病例观察	≤ - 3 分	⊕○○○/D

推荐强度	具体描述	表达符号/数字
支持使用某项干预措施的强推荐	评价者确信干预措施利大于弊	↑↑/1
支持使用某项干预措施的弱推荐	利弊不确定或无论高低质量的证据均显示利弊相当	↑?/2
反对使用某项干预措施的弱推荐		↓?/2
反对使用某项干预措施的强推荐	评价者确信干预措施弊大于利	↓↓/1

参 考 文 献

- [1] American Society of Hospital Pharmacists. ASHP statement on the use of medications for unlabeled uses[J]. *Am J Hosp Pharm*, 1992, 49(8):2006.
- [2] 广东省药学会. 药品未注册用法专家专识[Z]. 2010.
- [3] Fairman K A, Curtiss F R. Regulatory actions on the off-label use of prescription drugs: ongoing controversy and contradiction in 2009 and 2010[J]. *JMCP*, 2010, 16(8): 629.
- [4] Radley D C, Finkelstein S N, Stafford R S. Off-label prescribing among office-based physicians[J]. *Archives of internal medicine*, 2006, 166(9): 1021-1026.
- [5] Haynes B. Of studies, syntheses, synopses, summaries, and systems: the “5S” evolution of information services for evidence-based healthcare decisions[J]. *Evidence Based Nursing*, 2007, 10(1): 6-7.
- [6] Shea B J, Grimshaw J M, Wells G A, *et al.* Development of AMSTAR: a measurement tool to assess the methodological quality of systematic reviews[J]. *BMC medical research methodology*, 2007, 7(1): 10
- [7] Shea B J, Reeves B C, Wells G, *et al.* AMSTAR 2: a critical appraisal tool for systematic reviews that include randomised or non-randomised studies of healthcare interventions, or both[J]. *BMJ*, 2017, 21, 358: j4008.
- [8] Jadad A R, Moore R A, Carroll D, *et al.* Assessing the quality of reports of randomized clinical trials: is blinding necessary?[J]. *Controlled clinical trials*, 1996, 17(1): 1-12.
- [9] GRADE 主页. <http://www.gradeworkinggroup.org>[Z/OL].
- [10] Wells G A, Shea B, O’Connell D, *et al.* (2011) The Newcastle -Ottawa Scale (NOS) for assessing the quality if nonrandomized studies in meta-analyses[OL]. Available: http://www.ohri.ca/programs/clinical_epidemiology/oxford.asp.
-