



普通高等教育“十一五”国家级规划教材



卫生部“十一五”规划教材

全国高等医药教材建设研究会规划教材

全国高等学校教材  
供8年制及7年制临床医学等专业用



第2版

# 循证医学

Evidence Based Medicine

主 编 王家良

副主编 詹思延 许能锋 康德英



人民卫生出版社

PEOPLE'S MEDICAL PUBLISHING HOUSE

全国高等医药教材建设研究会规划教材

全国高等学校教材

供8年制及7年制临床医学等专用

# 循证医学

Evidence Based Medicine

- \* 1. 《细胞生物学》第2版(含光盘)
- \* 2. 《系统解剖学》第2版(含光盘)
- \* 3. 《局部解剖学》第2版(含光盘)
- \* 4. 《组织学与胚胎学》第2版(含光盘)
- \* 5. 《生物化学与分子生物学》第2版(含光盘)
- \* 6. 《生理学》第2版(含光盘)
- \* 7. 《医学微生物学》第2版(含光盘)
- \* 8. 《人体寄生虫学》第2版(含光盘)
- \* 9. 《医学遗传学》第2版(含光盘)
- \* 10. 《医学免疫学》第2版
- \* 11. 《病理学》第2版(含光盘)
- \* 12. 《病理生理学》第2版(含光盘)
- \* 13. 《药理学》第2版(含光盘)
- \* 14. 《临床诊断学》第2版(含光盘)
- \* 15. 《实验诊断学》第2版(含光盘)
- \* 16. 《医学影像学》第2版(含光盘)
- \* 17. 《内科学》第2版(含光盘)
- \* 18. 《外科学》第2版(含光盘)
- \* 19. 《妇产科学》第2版(含光盘)
- \* 20. 《儿科学》第2版(含光盘)
- \* 21. 《感染病学》第2版(含光盘)
- \* 22. 《神经病学》第2版(含光盘)
- \* 23. 《精神病学》第2版(含光盘)
- \* 24. 《眼科学》第2版(含光盘)
- \* 25. 《耳鼻咽喉头颈外科学》第2版
- \* 26. 《核医学》第2版(含光盘)
- \* 27. 《预防医学》第2版(含光盘)
- \* 28. 《医学心理学》第2版(含光盘)
- 29. 《医学统计学》第2版(含光盘)
- \* 30. 《循证医学》第2版(含光盘)
- \* 31. 《医学文献信息检索》第2版(含光盘)
- 32. 《临床流行病学》(含光盘)
- 33. 《肿瘤学》
- 34. 《生物信息学》(含光盘)
- 35. 《实验动物学》(含光盘)
- 36. 《医学科学研究导论》
- 37. 《医学伦理学》(含光盘)

注: 全套书均为卫生部“十一五”规划教材,  
画\*者为普通高等教育“十一五”国家级规划教材

策划编辑 皮雪花  
责任编辑 皮雪花 孙雪冰  
封面设计 郭 森  
版式设计 盖 伟

人民卫生出版社网站:

门户网: www.pmph.com 出版物查询、网上书店  
卫人网: www.ipmph.com 护士、医师、药师、中医师、卫生资格考试培训



ISBN 978-7-117-12916-9



定价(含光盘): 28.00元

- 普通高等教育“十一五”国家级规划教材
- 卫生部“十一五”规划教材
- 全国高等医药教材建设研究会规划教材
- 全国高等学校教材
- 供8年制及7年制临床医学等专业用



# 循证医学

## Evidence Based Medicine

主 编 王家良

副主编 詹思延 许能锋 康德英

编 者 (以姓氏笔画为序)

王吉耀 (复旦大学)

王觉生 (四川大学)

王家良 (四川大学)

刘金来 (中山大学)

许能锋 (福建医科大学)

李 静 (四川大学)

李晓枫 (大连医科大学)

李亚斐 (第三军医大学)

时景璞 (中国医科大学)

邸阜生 (南开大学)


赵亚双 (哈尔滨医科大学)

黄亚玲 (华中科技大学)

康德英 (四川大学)

詹思延 (北京大学)

秘 书 康德英 洪 旗 (四川大学)

 人民卫生出版社

**图书在版编目(CIP)数据**

循证医学/王家良主编. —2版. —北京:人民  
卫生出版社, 2010. 7

ISBN 978-7-117-12916-9

I. ①循… II. ①王… III. ①临床医学-医学院校-  
教材 IV. ①R4

中国版本图书馆 CIP 数据核字(2010)第 093453 号

门户网: <a href="http://www.pmph.com">www.pmph.com</a> 出版物查询、网上书店
卫人网: <a href="http://www.ipmph.com">www.ipmph.com</a> 护士、医师、药师、中医师、卫生资格考试培训

**版权所有, 侵权必究!**

本书本印次封底贴有防伪标。请注意识别。

**循证医学**  
第 2 版

**主 编:** 王家良

**出版发行:** 人民卫生出版社(中继线 010-59780011)

**地 址:** 北京市朝阳区潘家园南里 19 号

**邮 编:** 100021

**E - mail:** [pmph@pmph.com](mailto:pmph@pmph.com)

**购书热线:** 010-67605754 010-65264830

010-59787586 010-59787592

**印 刷:** 北京市卫顺印刷厂

**经 销:** 新华书店

**开 本:** 850×1168 1/16 **印张:** 15 **插页:** 1

**字 数:** 446 千字

**版 次:** 2005 年 8 月第 1 版 2010 年 7 月第 2 版第 4 次印刷

**标准书号:** ISBN 978-7-117-12916-9/R·12917

**定价(含光盘):** 28.00 元

**打击盗版举报电话:** 010-59787491 **E-mail:** [WQ@pmph.com](mailto:WQ@pmph.com)

(凡属印装质量问题请与本社销售中心联系退换)

## 第二版出版说明

全国高等学校八年制临床医学专业规划教材自2005年出版以来,得到了教育部、卫生部等主管部门的认可,以及医学院校广大师生的好评。为了进一步满足教学改革与实践不断推进,以及医学科学不断发展的需要,全国高等医药教材建设研究会和卫生部教材办公室在吴阶平、裘法祖、吴孟超、陈灏珠和刘德培院士等的亲切关怀和支持下于2009年启动了该套教材第二轮的修订工作。

第二轮修订过程中仍坚持“精品战略,质量第一的原则,从精英教育的特点、医学模式的转变、信息社会的发展、国内外教材的对比等角度出发,在注重‘三基’、‘五性’的基础上,从内容到形式都‘更新’、‘更深’、‘更精’,为培养高素质、高水平、富有临床实践和科学创新能力的医学博士服务”的编写宗旨,并根据使用过程中的反馈意见与建议,在第一轮的基础上力求做到:学科体系更加完善,增加了《临床流行病学》、《肿瘤学》、《生物信息学》、《实验动物学》、《医学科学研究导论》和《医学伦理学》;相关学科的交叉与协调更为完善,比如《生物化学》与《医学分子生物学》合并为《生物化学与分子生物学》;内容的选材与框架体系的设计更加注重启发性,强调学生创新能力的培养,并适当给学生留下了思维分析、判断、探索的空间;教材的配套更加健全;装帧设计更为精美。

该套书在修订过程中,得到了广大医学院校的大力支持,作者均来自各学科临床、科研、教学第一线,具有丰富临床、教学、科研和写作经验的优秀专家,作者队伍覆盖了目前国内所有开办临床医学专业八年制及七年制的院校。

修订后的第二版仍以全国高等学校临床医学专业八年制及七年制师生为主要目标读者,并可作为研究生、住院医师等相关人员的参考用书。

全套教材共37种,其中36种于2010年8月出版,1种将于2010年年底出版。

### 全国高等学校八年制临床医学专业卫生部规划教材 编写委员会

顾 问 吴阶平 裘法祖 吴孟超 陈灏珠

主任委员 刘德培

委 员 (按姓氏笔画排序)

丰有吉	孔维佳	王卫平	王吉耀	王宇明	王怀经
王明旭	王家良	王鸿利	冯作化	田勇泉	孙贵范
江开达	何 维	吴 江	张永学	张绍祥	李玉林
李甘地	李立明	李 和	李桂源	李 霞	杨世杰
杨宝峰	杨 恬	步 宏	沈 铿	陈孝平	陈 杰
陈 竺	欧阳钦	罗爱静	金征宇	姚 泰	姜乾金
柏树令	赵仲堂	郝希山	秦 川	贾文祥	贾弘禔
高英茂	黄 钢	葛 坚	詹启敏	詹希美	颜 虹
薛辛东	魏于全				

## 八年制教材目录

*1. 《细胞生物学》 第2版(含光盘)	主 编	杨 恬
	副主编	左 伋 刘艳平
*2. 《系统解剖学》 第2版(含光盘)	主 编	柏树令
	副主编	应大君 丁文龙 崔益群
*3. 《局部解剖学》 第2版(含光盘)	主 编	王怀经 张绍祥
	副主编	张雅芳 胡海涛
*4. 《组织学与胚胎学》 第2版(含光盘)	主 编	高英茂 李 和
	副主编	李继承 陈晓蓉
*5. 《生物化学与分子生物学》 第2版(含光盘)	主 编	贾弘提 冯作化
	副主编	屈 伸 药立波 方定志 冯 涛
*6. 《生理学》 第2版(含光盘)	主 编	姚 泰
	副主编	曹济民 樊小力 王庭槐
*7. 《医学微生物学》 第2版(含光盘)	主 编	贾文祥
	副主编	陈锦英 江丽芳 黄 敏
*8. 《人体寄生虫学》 第2版(含光盘)	主 编	詹希美
	副主编	诸欣平 刘佩梅
*9. 《医学遗传学》 第2版(含光盘)	主 编	陈 竺
	副主编	陆振虞 傅松滨
*10. 《医学免疫学》 第2版	主 编	何 维
	副主编	曹雪涛 熊思东
*11. 《病理学》 第2版(含光盘)	主 编	陈 杰 李甘地
	副主编	文继舫 来茂德 孙保存
*12. 《病理生理学》 第2版(含光盘)	主 编	李桂源
	副主编	吴伟康 欧阳静萍
*13. 《药理学》 第2版(含光盘)	主 编	杨世杰
	副主编	杨宝峰 颜光美 臧伟进
*14. 《临床诊断学》 第2版(含光盘)	主 编	欧阳钦
	副主编	吴汉妮 刘成玉
*15. 《实验诊断学》 第2版(含光盘)	主 编	王鸿利
	副主编	尚 红 王兰兰
*16. 《医学影像学》 第2版(含光盘)	主 编	金征宇
	副主编	冯敢生 冯晓源
*17. 《内科学》 第2版(含光盘)	主 编	王吉耀
	副主编	廖二元 黄从新 华 琦
*18. 《外科学》 第2版(含光盘)	主 编	陈孝平
	副主编	石应康 邱贵兴 杨连粤

*19.《妇产科学》 第2版(含光盘)	主 编	丰有吉 沈 铿
	副主编	马 丁 孔北华 李 力
*20.《儿科学》 第2版(含光盘)	主 编	薛辛东
	副主编	杜立中 毛 萌
*21.《感染病学》 第2版(含光盘)	主 编	王宇明
	副主编	施光峰 宁 琴 李 刚
*22.《神经病学》 第2版(含光盘)	主 编	吴 江
	副主编	贾建平 崔丽英
*23.《精神病学》 第2版(含光盘)	主 编	江开达
	副主编	于 欣 李凌江 王高华
*24.《眼科学》 第2版(含光盘)	主 编	葛 坚
	副主编	赵家良 黎晓新
*25.《耳鼻咽喉头颈外科学》 第2版	主 编	孔维佳
	副主编	周 梁 许 庚 王斌全 唐安洲
*26.《核医学》 第2版(含光盘)	主 编	张永学 黄 钢
	副主编	匡安仁 李亚明
*27.《预防医学》 第2版(含光盘)	主 编	孙贵范
	副主编	凌文华 孙志伟 姚 华
*28.《医学心理学》 第2版(含光盘)	主 编	姜乾金
	副主编	马 辛 林大熙 张 宁
29.《医学统计学》 第2版(含光盘)	主 编	颜 虹
	副主编	徐勇勇 赵耐青
*30.《循证医学》 第2版(含光盘)	主 编	王家良
	副主编	詹思延 许能锋 康德英
*31.《医学文献信息检索》 第2版(含光盘)	主 编	罗爱静
	副主编	马 路 于双成
32.《临床流行病学》 (含光盘)	主 编	李立明
	副主编	詹思延 谭红专
33.《肿瘤学》	主 编	郝希山 魏于全
	副主编	赫 捷 周云峰
34.《生物信息学》 (含光盘)	主 编	李 霞
	副主编	李亦学 廖 飞
35.《实验动物学》 (含光盘)	主 编	秦 川
	副主编	张连峰 魏 泓 顾为望 王 钜
36.《医学科学研究导论》	主 编	詹启敏 赵仲堂
	副主编	刘 佳 刘 强
37.《医学伦理学》 (含光盘)	主 编	王明旭
	副主编	尹 梅 严金海

注：全套书均为卫生部“十一五”规划教材，画\*者为普通高等教育“十一五”国家级规划教材

## 八年制教材再版序言

五年来，在大家的热情呵护下，我们共同见证了八年制临床医学教材——这个新生命的诞生与茁壮成长。如今，第二版教材与大家见面，怀纳第一版之精华而不张扬，吞吐众学者之智慧而不狂放，正如医学精英人才所应具备的气质与神韵。在继承中发展，新生才能越发耀眼；切时代之脉搏，思维才能永领潮头。第二版教材已然跨入新的成长阶段，心中唯觉欣喜和慰藉。

回想第一版教材面世之后，得到了各方众多好评，这充分说明了：这套教材将生命科学信息化、网络化以及学科高度交叉、渗透的特点融于一身，同时切合了环境-社会-心理-工程-生物医学模式的转变，诠释了以人为本、协调发展的战略思想。另外，编委构成的权威性和代表性、内容选择、编排体系、印刷装帧质量等，令广大师生耳目一新，爱不释卷。诚然，第一版教材也并非十全十美，比如有的学科仍以介绍知识为主，启发性不强，对学生难以起到点石成金、抛砖引玉的作用，不利于学生创新思维能力的培养；有的学科、章节之间有重复现象，略显冗余，不够干练。另外，随着学科的进展，部分疾病的临床分类、治疗等内容已略显滞后，亟待最新的研究成果加入其中，充实完善。

鉴此，第一版教材的修订工作便提上日程。此次修订，比当初第一版的编纂过程更为艰辛和严谨，从编者的谨慎遴选到教材内容的反复推敲、字斟句酌，可谓精益求精、力臻完美，经过数轮探讨、分析、总结、归纳、整理，第二版教材终于更富于内涵、更具有生命力地与广大师生们见面了。

“精英出精品，精品育精英”是第二版教材在修订之初就一直恪守的理念。主编、副主编与编委们均是各领域内的医学知名专家学者，不仅著作立身，更是德高为范。在教材的编写过程中，他们将从医执教中积累的宝贵经验、体会以及医学精英的特质潜移默化地融入到教材当中。同时，在主编负责制的前提下，主编、副主编负责全书的系统规划，编委会构成团结战斗的团队，各位专家群策群力、扬长补短、集思广益、查漏补缺，为教材的高标准、高质量的修订出版打下了坚实的基础。

注重医学学科内涵的延伸与发展，同时兼顾学科的交叉与融合是第二版教材的一大亮点。此次修订不仅在第一版的基础上增加了《临床流行病学》、《肿瘤学》、《生物信息学》、《实验动物学》、《医学科学研究导论》和《医学伦理学》，同时还合并了《生物化学》与《医学分子生物学》。通过主编顶层设计，相邻学科主编、副主编协调与磋商，互审编写提纲，以及交叉互审稿件等措施，相当程度上实现了突出中心、合理交叉、避免简单重复的要求。

强调启发性以及创新意识、创新思维和创新能力的培养是第二版教材的另一大特色。除了坚持“三基（基础理论、基本知识和基本技能）和五性（思想性、科学性、先进性、启发性和适用性）”，更注重激发学生的思维，让他们成为自己头脑的主人，批判地看待事物，辩证地对待知识，创造性地预见未来。同时，这版教材也特别注重与五年制教材、研究生教材、专科医师培训教材以及参考书的区别与联系。

以吴阶平、裘法祖、吴孟超、陈灏珠为代表的德高望重的老前辈对第二版教材寄予了殷切期望和悉心指导，教育部、卫生部、国家中医药管理局、国家食品药品监督管理局的各位领导的支持是这版教材不断完善的动力之源。在这里，衷心感谢所有关心这套教材的人们！正是你们的关注，广大师生手中才会捧上这样一本融贯中西、汇纳百家的精品。



八年制医学教材的第一版是我国医学教育史上的重要创举，相信修订后的第二版将不负我国医学教育改革的使命和重任，为培养高层次的具有综合素质和发展潜能的医药卫生人才做出更大的贡献。诚然，修订过程虽然力求完美，但纰漏与瑕疵在所难免，冀望各位领导、同道及师生不吝赐教，以便于这套教材能够与时俱进，不断完善。

是为序。

中国工程院院士  
中国医学科学院院长  
北京协和医学院院长

刘德培

于庚寅端午佳节

二〇一〇年六月十六日

## 第2版前言

在全国高等医药教材建设研究会和卫生部教材办公室的领导与帮助下,为适应全国高等医学院校长学制教学形势发展的需要,本教材第2版的编写扩大了参编单位,增加了新的编委,特别是加强了预防医学专业的力量,因而,使得我国的循证医学教材在临床医学实践的基础上,更加突出了临床医学与预防医学相结合、临床医学实践的防治结合、临床医学循证实践与临床流行病学、流行病学和医学统计学等多学科相结合的三大特色!

第2版教材的内容,继续强调“三基”(基础理论、基本知识、基本技能)“五性”(思想性、科学性、先进性、启发性、适用性),并根据第1版教材应用实践的总结与反馈信息,经过编委会的认真研讨,认为第1版内容设计的“循证医学的方法学基础、循证医学证据的分析与评价、循证医学在临床实践中的应用”三大部分是科学合理的,体现了创新性与先进性。在这个良好的基础之上,我们又汲取了国内外的最新进展,并结合我国的循证医学的发展实际,进一步对内容作了调整和完善。在第一篇“循证医学的基本知识与方法”中,进一步强调了循证医学临床实践的正确概念与方法,最佳证据的检索、收集、综合分析及其质量评价的要素与方法,以及循证医学临床实践的个体化原则和方法;并且在面对公共卫生事件,特别是突发性公共卫生事件如何做到科学的循证防治处理也进行了适当的探讨。在第二篇“循证医学证据的分析与评价”中,除了强调临床循证医学实践外,还有机地强化了预防医学的相关内容,有利于通过教学实践,促进医学生较为全面地掌握对疾病发病的病因、危险因素、早期正确诊断、科学地防治、正确评估疾病的预后、循证医学的科学决策等基本理论、知识和技能,从而提高分析、评价与解决问题的创新能力,无疑会有效提高临床医学的教学质量与水平。在第三篇“循证医学实践的临床范例”中,我们选择了糖尿病、急性心肌梗死、肿瘤、结核病、乙型病毒性肝炎防治中的有关重要问题,作了示范性循证医学实践的探讨,这种个案性的或以问题为基础的循证医学实践(problem based learning and practice, PBL)的尝试,希望能对我国的临床循证医学实践起到“抛砖引玉”的作用。

此外,为了更好地配合循证医学(evidence based medicine, EBM)教学实践,我们还配套地制作了EBM光盘教学资料,其中包括主教材的有关思考题和参考答案;示范性的有关医学文献的分析与评价;PBL个案资料、专业名词解释以及有关文献的分析与评价标准卡。光盘的编辑由康德英老师和洪旗老师承担完成。

尽管全体编委都以极大的热忱和科学敬业的精神,努力完成了第2版教材的编写任务,但难免有谬误之处,殷切希望读者予以批评指正。我们满怀信心地期望读者能将本教材与国内外有关循证医学的图书版本进行比较、分析与评价,更期望通过教学与临床实践,能证明本教材已达到国内外的先进水平。

对给予本教材编写以极大支持和帮助的全中国高等医药教材建设研究会、卫生部教材办公室和四川大学华西临床医学院/华西医院的有关领导致以衷心感谢!

向认真地承担资料录入、编辑、排版、精制图表等工作的康德英老师和洪旗老师致以诚挚的谢意!

王家良

四川大学华西医院  
2010年3月于成都

循证医学是遵循现代最佳医学研究的证据(成果),将其应用于临床对患者进行科学诊治的一门学问。目的在于不断地提高临床医疗质量和医学人才的素质,并促进临床医学的发展,从而更有效地为患者服务并保障人民的健康。

循证医学与传统临床医学最重要的区别在于它所应用的临床证据,是采用科学的标准,进行了严格的分析评价,从而被确认是真实的有临床重要的意义,是实用于具体临床实践的当代科学证据,其重要性还在于随着科学证据的进步而不断地更新,永居前沿。此外,它还充分地体现以人为本的原则,使患者在接受临床诊治过程中,充分地体现自己的价值取向和愿望,构建良好的医患和谐的互相依从的关系,从而使这种循证医学实践的科学决策得以实现并可望获得最佳的结局。

循证医学是要应用现已存在的最佳证据于临床实践,去解决患者目前存在的具体临床“问题”,是“用水来救火”的实践,而不是已经“失火”了,现去生产“水”。所以不能误解循证医学等于临床医学科研,后者是创造最佳证据,是为临床医疗实践或循证医学的临床实践提供“用证”资源。毫无疑问,没有最好的临床研究成果(证据),也就没有循证医学的产生和发展。可见,加强临床医学科学研究,不断地提高研究质量和产生最佳研究证据是根本。

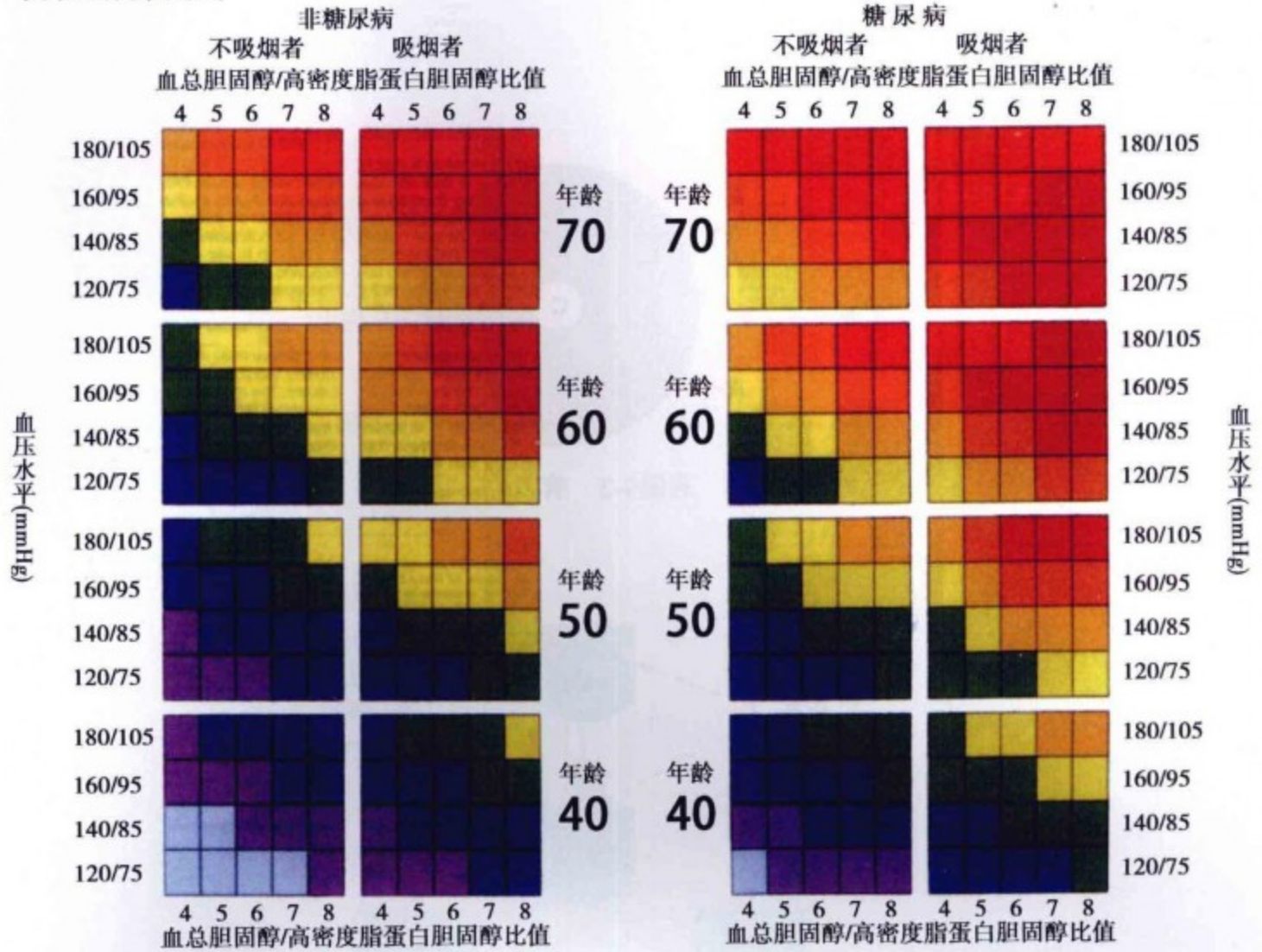
由于当代信息科学、生物医学科学以及经济的高速发展,在经济全球化的当今,必然要使人类的知识“更新换代”,而循证医学的产生与发展就是在临床医学方面的体现。

根据“教育必须面向世界、面向现代化、面向未来”的精神,在教育部和卫生部以及全国医药高等教材建设指导委员会的指导与帮助下,将《循证医学》列入我国长学制医学教育的课程确属高瞻远瞩,并要求组织编写好这本循证医学教材,使之达到培养我国医学精品人才水平,并与国际“接轨”,确属重任!

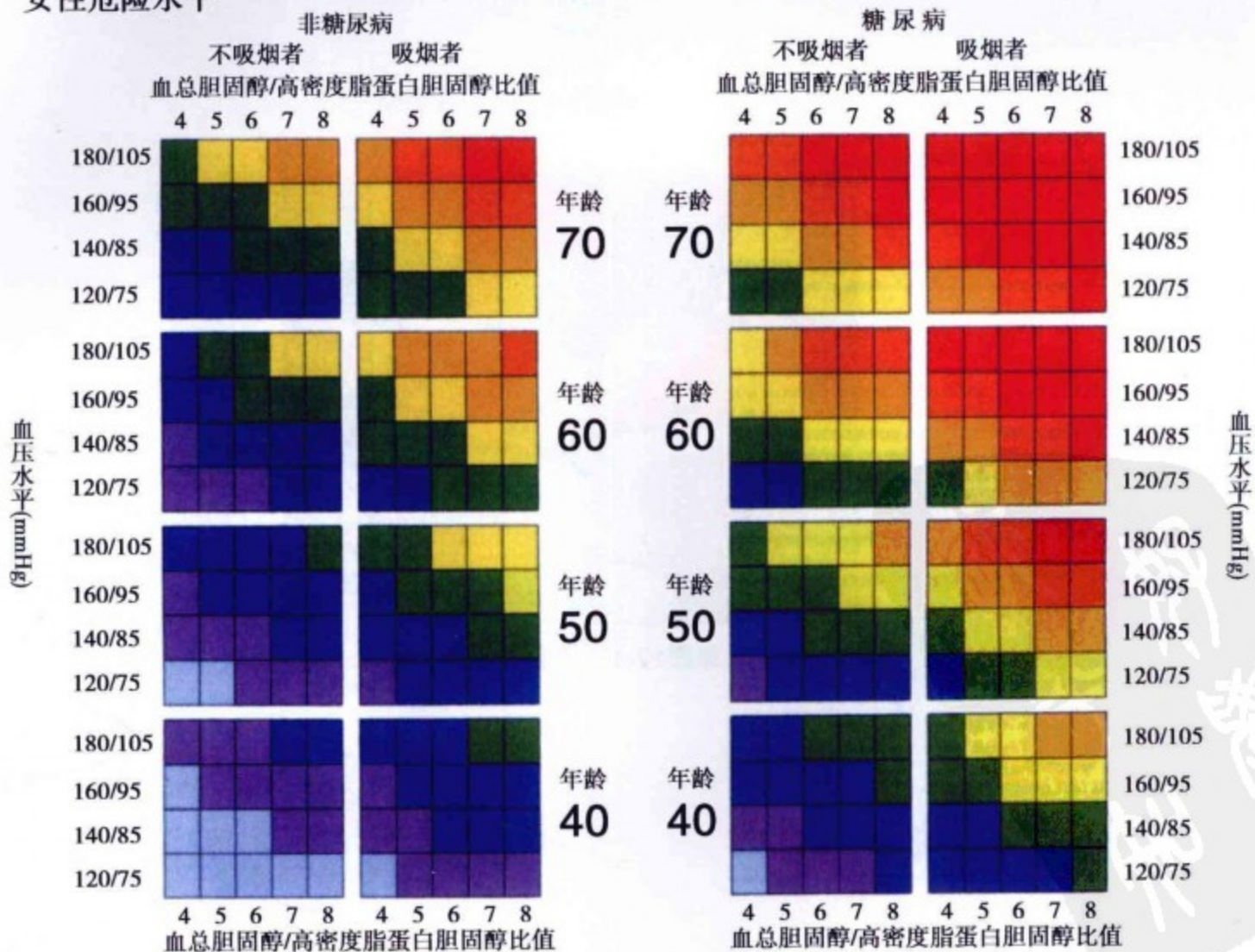
按照上述要求,结合国际和我国20余年来临床流行病学及新近循证医学现况,联系我们自己的教学、研究与临床经验,规划本教材为3篇共22章(包括绪论)。第一篇为循证医学实践基本方法学部分,主要涉及如何发掘患者的临床问题,如何根据问题从各种产生证据的资源中发掘最佳证据,以及分析和整理证据的统计学方法,还有应用证据时要考虑的个体化特点与原则,这是实践循证医学的基本功。第二篇主要涉及对病因、危险因素、诊断、治疗、预后以及临床经济学和卫生技术等证据及其质量分析、评价以及可应用的原则与标准,在此基础上,学习临床决策的方法。这部分是联系临床应用最佳证据的基本功和方法,要深刻地了解与正确应用证据,是要有一定的临床流行病学知识的。第三篇选择了我国人民疾病负担颇重的疾病,以高血压、急性心肌梗死、脑卒中、肺癌以及乙型病毒性肝炎为代表写成了循证医学范例,通过这些范例使读者了解循证医学实践的过程及其密切的临床实践性,从而有利于对循证医学的正确理解。

全体编者都深知自己肩负着编写好本教材以培养精品医学人才的重任。虽然在国际上“有据可循”,但是绝不能生搬硬套,我们的国情、社会经济、人文科学和文化传统乃至临床医学的实际与国外的情况是有很大差异的,因此,应有自己的特色。我们是在学习理解国际先进经验的基础上,联系我国的上述特点,努力地取其精华并且作了一些创新、期望通过循证医学的教学与培养,学习了《循证医学》后,人们也能联系实际而创新。

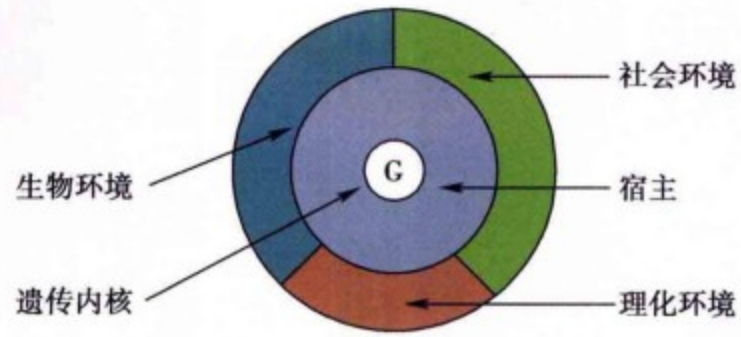
男性的危险水平



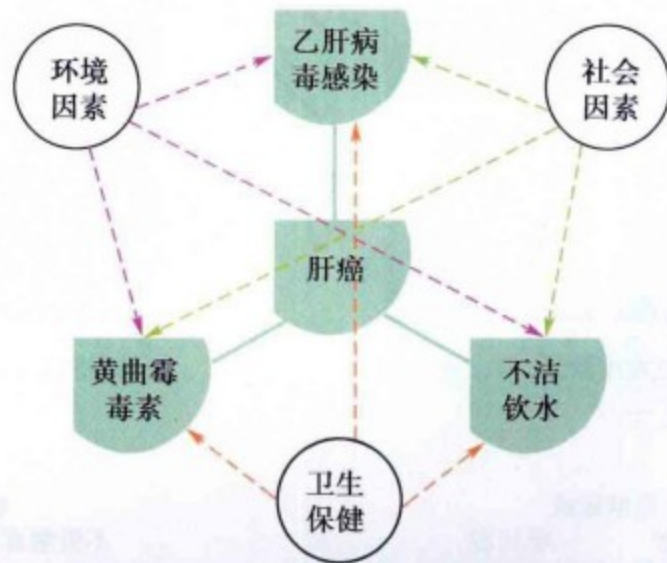
女性危险水平



彩图7-2 心血管发病危险因素与治疗效益的估价



彩图9-3 病因的轮状模型



彩图9-4 病因网模型



彩图12-1 “海豹婴儿”

循证医学  
PDG

# 目 录

## 第一篇 循证医学的基础与方法

<b>第一章 绪论</b>	<b>1</b>
第一节 循证医学的基本概念	1
一、循证医学的概念	1
二、循证医学的临床实践基础	1
第二节 循证医学发展的简史	2
第三节 循证医学实践的类别	4
第四节 循证医学实践的方法	4
第五节 突发性公共卫生事件循证实践的方法	6
第六节 循证医学实践的目的与意义	7
一、循证医学实践的目的	7
二、循证医学实践对临床医学与预防医学的意义	8
<b>第二章 构建临床和公共卫生的循证问题</b>	<b>9</b>
第一节 概述	9
一、找出临床问题的重要性	9
二、找准临床问题应具备的条件	10
三、提出公共卫生事件循证问题的重要性	10
第二节 如何构建临床循证问题	10
一、临床问题的类型	10
二、提出临床问题的形式和方法	11
三、构建临床循证问题的模式	13
第三节 如何提出公共卫生领域的循证问题	14
一、问题的提出及排序	14
二、问题的构建	16
<b>第三章 检索与收集需求证据的方法</b>	<b>17</b>
第一节 循证医学证据资源	17
一、循证医学证据资源发展	17
二、循证医学证据资源分类	17

三、选择证据资源的标准 .....	18
四、常用循证医学证据资源 .....	19
<b>第二节 证据检索和收集的基本步骤 .....</b>	<b>22</b>
一、确定临床问题类型和构建临床问题 .....	22
二、选择合适数据库 .....	23
三、制订检索词和检索策略 .....	23
四、判断检索结果 .....	24
<b>第三节 证据检索实例 .....</b>	<b>24</b>
<b>第四章 证据评价的基本原则与方法 .....</b>	<b>28</b>
<b>第一节 证据的分类与分级 .....</b>	<b>28</b>
一、证据的分类 .....	28
二、证据的分级 .....	28
<b>第二节 证据评价的意义与基本要素 .....</b>	<b>30</b>
一、证据评价的意义 .....	30
二、证据评价的基本要素 .....	30
<b>第三节 证据评价的基本内容与方法 .....</b>	<b>30</b>
一、证据评价的基本内容 .....	30
二、证据评价的基本方法 .....	31
三、证据评价的注意事项 .....	32
<b>第五章 循证医学实践中常用的统计学方法 .....</b>	<b>33</b>
<b>第一节 临床证据的数据资料类型 .....</b>	<b>33</b>
一、分类变量资料 .....	33
二、数值变量资料 .....	34
三、等级变量资料 .....	34
<b>第二节 证据资料的方法学质量评价 .....</b>	<b>34</b>
一、单个研究证据的质量评价 .....	34
二、多个研究证据的质量评价 .....	36
<b>第三节 统计学方法的正确抉择 .....</b>	<b>37</b>
一、单个研究及其统计方法的正确抉择 .....	38
二、多个研究及其统计方法的正确抉择 .....	43
<b>第四节 证据的临床意义与统计学意义 .....</b>	<b>44</b>
<b>第六章 医学文献的系统评价与meta分析 .....</b>	<b>46</b>
<b>第一节 系统评价概述 .....</b>	<b>46</b>
一、基本概念 .....	46
二、为什么要进行系统评价 .....	47
三、系统评价与叙述性文献综述的区别与联系 .....	48
<b>第二节 系统评价的方法 .....</b>	<b>48</b>
一、确立题目、制订系统评价计划书 .....	49
二、检索文献 .....	50

三、选择文献 .....	50
四、评价纳入研究的偏倚风险 .....	50
五、收集数据 .....	52
六、分析资料和报告结果 .....	53
七、解释系统评价的结果 .....	53
八、更新系统评价 .....	54
<b>第三节 meta分析 .....</b>	<b>54</b>
一、概述 .....	54
二、基本过程 .....	54
三、固定效应模型与随机效应模型的选择 .....	57
四、发表性偏倚的识别与分析 .....	60
五、meta回归及其他方法学进展 .....	61
六、慎重应用与评价meta分析的结果 .....	62
<b>第四节 系统评价的评价原则 .....</b>	<b>63</b>
一、系统评价的结果是否真实 .....	63
二、系统评价的结果是否重要 .....	63
三、系统评价的结果是否能应用于我的病人 .....	64
<b>第五节 系统评价的应用 .....</b>	<b>64</b>
一、临床医疗的需要 .....	64
二、科研工作的需要 .....	64
三、反映学科新动态 .....	65
四、医学教育的需要 .....	65
五、卫生决策的需要 .....	65

## **第七章 循证医学实践的个体化原则与方法 66**

<b>第一节 最佳临床证据的特性 .....</b>	<b>66</b>
一、真实性 .....	66
二、重要性 .....	66
三、实用性 .....	66
四、物美价廉 .....	67
<b>第二节 生物学依据 .....</b>	<b>67</b>
<b>第三节 病理生理学依据 .....</b>	<b>67</b>
<b>第四节 社会-心理及经济特点 .....</b>	<b>68</b>
<b>第五节 证据的应用要权衡利弊 .....</b>	<b>68</b>
<b>第六节 个体化干预的效果预测 .....</b>	<b>69</b>

## **第八章 循证医学实践的决策分析 74**

<b>第一节 概述 .....</b>	<b>74</b>
一、决策及分类 .....	74
二、决策分析和临床决策分析 .....	74
三、决策分析的步骤和常用方法 .....	75
<b>第二节 决策分析的实施 .....</b>	<b>75</b>



一、问题的确认 .....	75
二、问题的结构化 .....	75
三、搜集相关信息 .....	76
四、确定最终结局的效用值 .....	77
五、决策树的分析 .....	79
六、敏感性分析 .....	79
<b>第三节 复杂决策问题和MARKOV模型 .....</b>	<b>80</b>
一、复杂决策问题 .....	80
二、Markov模型 .....	80
三、左颈动脉硬化是否手术治疗远期效果评价的Markov模型 .....	82
<b>第四节 决策分析质量评估 .....</b>	<b>83</b>

## 第二篇 循证医学证据的分析与评价

### 第九章 疾病的病因与危险因素的循证分析与评价 85

第一节 临床实践：探求病因 .....	85
第二节 可供探讨与验证病因或危险因素的模型 .....	86
一、流行病学三角模型 .....	87
二、轮状模型 .....	87
三、病因网模型 .....	87
第三节 病因与危险因素所致疾病的因果效应 .....	87
一、单病因(原)的因果效应 .....	88
二、单病因致多病的因果效应 .....	88
三、多病因致单病的因果效应 .....	88
第四节 病因与危险因素的分析评价 .....	88
一、真实性 .....	88
二、重要性 .....	93
三、适用性 .....	93

### 第十章 诊断性试验循证分析与评价 95

第一节 诊断性试验在临床医学中的重要性 .....	95
一、诊断性试验在循证医学实践中的临床意义 .....	95
二、诊断性试验研究现状与发展方向 .....	96
第二节 对诊断性试验研究评估的基本要求 .....	97
一、确定金标准 .....	97
二、选择研究对象 .....	97
三、盲法对比试验结果 .....	97
第三节 循证医学中诊断性试验常用的指标 .....	98
一、敏感度 .....	98
二、特异度 .....	98
三、准确度 .....	98

四、阳性预测值 .....	98
五、诊断比值比 .....	99
六、阳性似然比 .....	99
<b>第四节 诊断性试验在循证医学临床实践中的应用 .....</b>	<b>99</b>
一、ROC曲线的应用 .....	100
二、似然比的临床应用 .....	101
三、提高诊断性试验敏感度或特异度的方法 .....	104
<b>第五节 诊断性试验证据的评估 .....</b>	<b>105</b>
一、诊断性试验的真实性 .....	106
二、诊断性试验的重要性 .....	106
三、诊断性试验的实用性 .....	107
<b>第十一章 疾病治疗的循证分析评价 .....</b>	<b>108</b>
<b>第一节 循证医学治疗实践应找出挑战性的问题 .....</b>	<b>108</b>
<b>第二节 查阅与运用最佳证据 .....</b>	<b>108</b>
一、关注治疗目的 .....	108
二、关注患者的具体状况 .....	109
三、关注治疗性文献的质量 .....	109
<b>第三节 单个临床治疗性证据的评价 .....</b>	<b>109</b>
一、真实性的分析与评价 .....	109
二、治疗性证据的重要性分析与评价 .....	111
三、治疗性证据适用性的分析与评价 .....	113
<b>第四节 系统评价或meta分析证据的评价 .....</b>	<b>115</b>
一、系统评价的真实性评价 .....	115
二、系统评价的重要性评价 .....	116
三、系统评价的适用性评价 .....	117
<b>第十二章 药物不良反应的循证分析与评价 .....</b>	<b>118</b>
<b>第一节 概述 .....</b>	<b>118</b>
一、药物不良反应的概念与分类 .....	118
二、药物研发与安全性研究 .....	119
三、药物不良反应监测发展 .....	120
<b>第二节 药物不良反应诊断方法 .....</b>	<b>120</b>
一、药物不良反应诊断的意义 .....	120
二、药物不良反应相关因素 .....	121
三、药物不良反应的诊断方法 .....	121
<b>第三节 药物不良事件证据的查询与评价 .....</b>	<b>122</b>
一、带着问题查询 .....	122
二、资料检索与证据来源 .....	122
三、证据的评价 .....	123
四、临床应用 .....	124

**第十三章 疾病预后证据的循证分析与评价 126**

第一节 预后证据分析与评价的意义 .....	126
第二节 疾病预后的时态特点 .....	126
一、疾病的自然病史 .....	127
二、疾病的临床病程 .....	127
第三节 掌握影响预后证据的有关因素 .....	127
一、疾病本身与环境的因素 .....	127
二、影响预后证据的质量因素 .....	128
三、控制偏倚因素影响预后证据的方法 .....	129
第四节 评价疾病预后的指标与方法 .....	129
一、常用的指标 .....	129
二、生存率分析 .....	130
第五节 疾病预后证据的评价 .....	131
一、预后证据的真实性评价 .....	131
二、真实预后证据的重要性评价 .....	133
三、真实、重要的预后证据可否用于我们的病人 .....	134

**第十四章 卫生经济学证据的评价与应用 136**

第一节 卫生经济学及其评价的基本概念 .....	136
一、卫生经济学 .....	136
二、卫生经济学评价 .....	136
第二节 卫生经济学评价的基本要素 .....	136
一、成本 .....	136
二、效果 .....	137
第三节 卫生经济学评价的类型 .....	138
一、卫生经济学部分评价 .....	138
二、卫生经济学全面评价 .....	138
第四节 卫生经济学评价方案的设计步骤 .....	142
一、明确所要评价的项目及要进行比较的措施 .....	142
二、确定经济学评价的分析角度 .....	142
三、根据研究目的确定合适的经济评价方法 .....	142
四、确定资料收集的方法 .....	142
五、成本和效果的确定和测量方法 .....	142
六、处理时间对成本和效益的影响 .....	142
七、增值分析 .....	143
八、敏感性分析 .....	143
第五节 卫生经济学评价研究证据的评价原则 .....	143
一、结果是否真实 .....	143
二、研究结果的重要性 .....	144
三、研究结果的适用性 .....	144

**第十五章 卫生技术评估与应用 145**

第一节 卫生技术与卫生技术评估概述 .....	145
一、卫生技术的定义 .....	145
二、卫生技术评估的定义 .....	145
三、卫生技术评估的历史沿革 .....	145
四、卫生技术评估的意义 .....	146
五、卫生技术评估的内容 .....	147
六、循证医学在卫生技术评估中的作用 .....	147
第二节 卫生技术评估的步骤与方法 .....	148
一、确定评估项目 .....	148
二、明确评估问题 .....	148
三、确定评估者 .....	149
四、收集资料 .....	149
五、评价证据 .....	150
六、合成资料 .....	150
七、形成结果 .....	151
八、传播结果 .....	151
九、监测评估结果的影响 .....	151
第三节 卫生技术评估的应用 .....	152
一、应用于医疗技术管理 .....	152
二、应用于药品管理 .....	152
三、应用于医疗设备器械管理 .....	152

**第十六章 临床实践指南的循证评价与应用 154**

第一节 概述 .....	154
一、临床实践指南的概念与发展 .....	154
二、临床实践指南的作用 .....	155
三、指南与其他证据的关系 .....	156
四、循证CPG的网络资源 .....	156
第二节 临床实践指南制订的流程与方法 .....	157
第三节 临床实践指南的评价 .....	159
一、对临床实践指南的评价 .....	160
二、临床实践指南评价工具 .....	160
第四节 临床实践指南应用的原则和方法 .....	163
一、临床应用指南的原则 .....	163
二、指南的临床应用方法 .....	164

**第十七章 循证医学实践的自我评价 165**

第一节 循证实践能力的自我评价 .....	165
一、评价“提出问题”的能力 .....	165
二、评价“寻找最佳外部证据”的能力 .....	166

三、评价“严格评估证据质量”的能力 .....	166
四、评价“整合外部证据与病人价值”的能力 .....	167
第二节 循证实践的效果评价 .....	167
一、临床实践质量是否得到了改善 .....	167
二、多少临床实践有证可循 .....	167

### 第三篇 循证医学实践的临床范例

#### 第十八章 NSAIDs预防乳腺癌的循证决策 169

一、疾病案例 .....	169
二、提出问题 .....	169
三、证据检索与评价 .....	170
四、实践决策 .....	173
五、小结 .....	174

#### 第十九章 急性心肌梗死的循证治疗 175

一、疾病案例 .....	175
二、提出问题 .....	175
三、证据检索 .....	176
四、证据评价 .....	177
五、实施决策 .....	178
六、后效评估 .....	178
七、小结 .....	180

#### 第二十章 2型糖尿病合并严重冠心病时罗格列酮的使用 181

第一节 临床情景 .....	181
一、病历摘要 .....	181
二、查房讨论 .....	181
三、实践循证医学基本步骤 .....	182
第二节 循证临床实践过程 .....	182
一、提出问题 .....	182
二、获取证据 .....	183
三、评价证据 .....	184
四、应用证据 .....	186
五、经验总结 .....	187

#### 第二十一章 慢性乙型肝炎抗病毒的循证治疗 188

一、疾病案例 .....	188
二、提出问题 .....	189
三、证据的检索 .....	189
四、证据评价 .....	192

五、实施决策 ..... 193  
 六、后效预测 ..... 194  
 七、小结 ..... 194

**第二十二章 结核化疗中预防应用保肝药的循证实践 195**

一、疾病案例 ..... 195  
 二、提出问题 ..... 195  
 三、证据检索和评价 ..... 196  
 四、实践决策 ..... 200  
 五、后效评估 ..... 201  
 六、小结 ..... 201

**主要参考文献 203**

**英汉名词索引 209**

**汉英名词索引 214**

# 第一篇 循证医学的基础与方法

## 第一章 绪 论

循证医学(evidence based medicine, EBM)是现代临床医疗诊治决策的科学方法学。旨在针对病人具体的临床问题所作出的有关诊治措施,要建立在最新、最佳的科学证据基础之上。因此,这种医疗水平毫无疑问是超出了传统的临床医学水平,且为与时俱进的一流水平。所以备受国内外医学界的广泛关注并赋予了很大的热情。

既然循证医学称为“临床诊治决策的科学方法学”,那么,它必然适用于临床各个学科的广泛应用。其不同之点,则在于各个学科的具体临床循证实践罢了。因而,有些学科往往在其名称之前,冠以“循证”二字以示其“新意”!

### 第一节 循证医学的基本概念

#### 一、循证医学的概念

循证医学指的是临床医生面对着具体的病人,在收集病史、体检以及必要的试验和有关检查资料的基础上,应用自己的理论知识与临床技能,分析与找出病人的主要临床问题(病因、诊断、治疗、预后以及康复等),并进一步检索、评价当前最新的相关研究成果,取其最佳证据、结合病人的实际临床问题与临床医疗的具体环境作出科学、适用的诊治决策,在病人的配合下付诸实施,最后分析与评价效果。这样做,既能有效地解决病人的临床问题,促进病人康复,同时也会推动临床医疗水平的提高和临床医学的进步。可见,循证医学对病人的诊治决策是建立在当前最新、最佳的证据基础之上,以追求最佳诊治效果,故称之为 evidence based medicine。这样就超出了传统的临床医学的水平。

#### 二、循证医学的临床实践基础

##### (一) 医生

系循证医学实践的主体。作为实践循证医学的医生,首先要具备良好医学理论基础与过硬的临床经验和技能,同时还要有不断进取和革新精神以及全心全意为病人服务的精神,这样,才可能去发现病人的临床问题,并充分地应用自己的智慧和能力去解决病人的问题,促进循证医学实践、提高自己的临床学术水平。

##### (二) 病人

系循证医学实践服务的主体。实践循证医学,务必要取得病人的合作,具备对诊疗措施良好的依从性。为此,应关心体贴病人,构建良好的医患关系,否则,任何有效的方法与措施,如无病人的合作,都难以成功。

### (三) 最佳证据

最佳证据乃为实践循证医学的“武器”，是解决病人临床问题的手段。当然是来源于现代临床医学的研究成果，而这种证据的获取，则依赖于应用科学的方法去检索、分析与评价，结合具体的临床问题择优采用。

### (四) 医疗环境

循证医学实践都要在具体的医疗环境下推行。不同级别的医院，设备条件和医务人员的水平各异，即使某一最佳措施和方法对某疾病疗效很好，可是当医疗环境条件受限时，也是难以实现的。因此，循证医学实践应结合具体的医疗环境，讲求实事求是。

在我国，随着医疗制度的深化改革，国家对人民卫生事业的关注与资源投入，各级卫生医疗机构的设备物资条件正越来越显著的改善。因此，医疗环境的改进与提高为实践临床循证医学创造了很好基础，关键是如何利用良好的医疗条件去全心全意地真正为病人服务。

上述四个部分是循证医学的基础，缺一不可，是一个临床病人科学诊治的复杂系统工程(图1-1)。这里还必须强调的是：要真正地实践循证医学，不掌握必要的临床流行病学的知识、理论与方法学，想要真实地认识、分析和评价最佳证据，有时往往有些困难。因为循证医学的理论和标准及方法学源于临床流行病学，且其为该学科在临床实践中的具体应用。

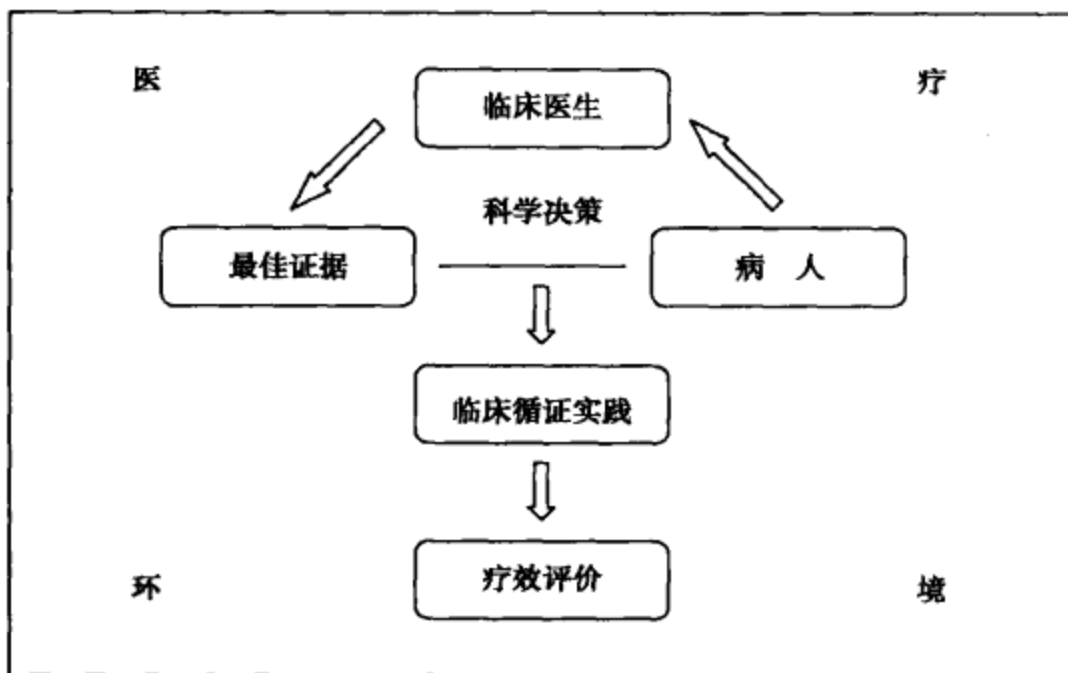


图1-1 循证临床实践示意图

在我国，由于有的人对EBM十分热情，在尚未完全了解EBM的情况下，往往出现过某些不利于正确发展EBM的概念。例如，将EBM等同于Cochrane项目或系统评价(systematic review)或大型多中心随机对照试验，或称EBM为临床科研方法学等。这些概念，有的难免误导，值得吸取经验和教训。

## 第二节 循证医学发展的简史

严格地讲，循证医学并非是在现今才有的，凡是接受过正规医学教育的临床医生，都具备现代生物学、人体解剖学、生理学、病理学、免疫学、临床医学等基本理论知识，他(她)们对患者的诊治，也是从临床实际出发，根据患者的临床特征，结合自己掌握的理论知识和临床的经验得出相应的诊治决策。在一定程度上，当然也是“循证”的，只不过在即时采用最新和最佳的证据方面，也许有所不足。因此，对于现阶段人们应用的临床医疗决策过程，不应都认为是“临床经验医学”。

然而，就临床医学的实质而言，它是一门实用的科学，总是随着自然科学和临床科学的发展，以及人们认识的深化而不断发展和丰富的。因此，临床医生要使自己的临床工作做得更好，就必须不断



地更新自己的知识,学习、掌握和应用先进的技能和理论以指导自己的临床实践。据美国哈佛大学医学院原院长 Sydney Burwell 称:“在大学里教给学生的知识,在10年后约有50%是错的,而教师往往不知道错的是哪一半”。这说明我们不断学习更新自己的知识是何等重要。

当前,据统计在国际范围内已拥有生物医学杂志25 000余种,每年发表的论著达200万余篇,加上不完整统计达千余家的网络系统,发表的信息资料难以计数!如此浩瀚的文献,任何人要想全部阅读是不现实的。而且其中良莠并存,精华与糟粕互混也是必然的。临床医生的工作又是十分繁忙,读书时间十分有限,但是为了汲取当代医学研究的精华和提高医疗质量,却又必须去发掘有价值的研究文献及研究成果并应用于临床实践。

什么样的临床研究文献(成果)是“最佳证据”呢?其真实性(validity)和可靠性(reliability)如何呢?临床重要程度(importance)及实用性(applicability)又怎样呢?这是临床医生阅读和引用“最佳证据”文献的时候必须回答的问题。在国际医学领域里从20世纪70年代后期开始,日益发展和完善的临床流行病学(clinical epidemiology)以其先进的临床科研方法学(clinical research methodology),强调临床科研设计(design),测量(measurement)和评价(evaluation)的科学性,于是,推动了临床科学研究,产生了日益丰富的高质量的研究成果,同时又总结出了一系列严格评价(critical appraisal)的方法和标准,这些又都被国际临床医学界所接受和应用,于是就大大地促进了临床医学信息科学的发展和循证医学实践。

20世纪80年代初期,国际临床流行病学的发源地之一的 McMaster 大学,以临床流行病学创始人之一、国际著名的内科学专家 David L. Sackett 为首的一批临床流行病学家,在该医学中心的临床流行病学系和内科系率先对年轻的住院医师举办了“如何阅读医学文献的学习班”(how to read clinical literature),他(她)们联系患者的临床实际问题,检索与评价医学文献,并应用所获得的新近成果于自己的临床实践,如此在学习应用临床流行病学原理与方法的基础上,进行循证医学培训,取得了很好的效果。经过了反复实践,于1992年起相继在 JAMA 等杂志上,作为对临床医生一种新型培训措施——将这种临床医学实践的方法,正式冠以“循证医学”(evidence based medicine),而发表了系列总结性文献,受到了临床医学界的广泛关注。又经多年实践,由 Haynes 和 Sackett 发起,由美国内科医师学院(American College of Physicians)组织了一个杂志俱乐部(Journal Club),即 ACPJC。为了促进循证医学的发展,从1991年起,对国际上著名的30多种医学杂志发表的论著,由临床流行病学、临床有关学科及方法学专家,有选择性地、系统地进行分析与评价,将最佳的研究论文进行精练的摘要加专家评述的形式,作为 Annals of Internal Medicine 的副刊发表,向临床医生推荐,供循证医学实践之用。1995年 Sackett 医生受聘于英国牛津大学,斯时在英国建立了循证医学中心(Evidence-Based Medicine Center),相继出版了循证医学专著及由英国医学杂志和美国内科医师学院联合主办的循证医学杂志。为了全面地推荐国际上经过严格评价的最佳研究证据,自1999年起,他们还整理编辑并出版了 Clinical evidence 专集,每年两期并公开发行,以推荐临床医生将经过专家筛选、严格评价及评论后的最佳研究成果应用于临床医疗实践。此外,1993年国际上还成立了 Cochrane 协作网(Cochrane Collaboration),广泛地收集临床随机对照试验(RCT)的研究结果,在严格的质量评价的基础上,进行系统评价(systematic review)以及 meta 分析(meta-analysis),将有价值的研究结果推荐给临床医生以及相关专业的实践者,以帮助实践循证医学。

我国于1996年在卫生部领导与支持下,正式成立了中国循证医学中心及 Cochrane 中心,组织了对全国临床医生和相关专业的人员培训,开展了广泛的国际国内合作。出版了两种全国性的循证医学杂志。在医学院校开设了循证医学课程,编辑出版了循证医学专著以及国家级5~8年制循证医学高级教材,推动临床与预防医学实践,提高医学水平等均产生了良好效果。本学科正在全国迅速和健康的发展,这无疑会推动医学的进步与繁荣。

总之,人们对循证医学投以极大的关注,随着时代步伐的前进,它将日臻完善,为临床决策的科学性和临床医学的现代化作出更大贡献。

### 第三节 循证医学实践的类型

循证医学实践的类型,可分为两种类型:循证医学最佳证据的提供者(doer)和最佳证据的应用者(user)。

最佳证据的提供者,是由一批颇具学术造诣的临床流行病学家、各专业的临床学家、临床统计学家、卫生统计学家和社会医学家以及医学科学信息工作者共同协作,根据临床医学实践中存在的某些问题,从全球年逾200余万篇生物医学文献中收集、分析、评价以及综合最佳的研究成果(证据),为临床医生实践循证医学提供证据(当前最佳临床证据的资源可数BMJ编辑出版的Clinical evidence, ACPJC, EBMJ, Cochrane Library)。因此,证据提供者是循证医学的组成部分,没有他(她)们的辛勤劳动就不可能做到循证医学实践。

这些专家提供了最佳证据并不就此完事,他们还有如何将这些优秀成果(证据)推广到临床循证医学实践中去应用的艰巨任务。这就涉及对医学生的循证医学教育,以及对临床医生进行循证医学实践的培训和宣传。只有将最好的研究成果最大程度地转化为广大患者的医疗及防病治病服务,只有广大的临床医生能掌握与应用循证医学临床实践的理论与方法,并能进入到主动性与创造性相结合的自我教育和提高的良性循环,那么才能达到循证医学的真正目的。

最佳证据的应用者,为从事临床医学的医务人员,包括医疗管理和卫生政策的决策者,为了对患者诊治决策以及卫生管理和政策决策的科学化,都应联系各自的实际问题,去寻找、认识、理解和应用最佳最新的科学证据,理论与实践相结合,方能取得最好的结果。

无论证据的提供者和应用者,除了都具有临床的业务基础之外,也要具有相关学科的知识 and 学术基础,只是要求的程度有所不同(表1-1)。当然,证据的提供者本身也可以是应用者;而应用者本身的深化发展,又可以成为证据提供者。

表1-1 循证医学实践的类型

	证据提供者 (doer)	证据应用者 (user)
确定临床问题	+++	+++
任务	<ul style="list-style-type: none"> <li>• 收集与评价文献</li> <li>• 提供最佳证据</li> </ul>	<ul style="list-style-type: none"> <li>• 正确地应用证据</li> </ul>
专业基础与技能	临床实践 +++ 临床流行病学研究方法学 +++ 临床统计学 ++ 卫生统济学 ++ 社会医学 ++ 计算机技能 +++	+++ + + + + +
技术力量	团队力量	个体

### 第四节 循证医学实践的方法

根据国外实践循证医学的教学培训与临床经验,归纳为“五步曲”的循证医学实践方法,见图1-2,其中每个步骤都具有丰富的内涵和科学的方法,它们之间是互相联系的一个完整的整体,如果在任何方面存在着缺陷或不足,都会影响循证医学实践的质量。

#### (一) 找准患者存在的且应解决的临床重要问题

在循证医学的临床实践中,首先应该找准自己的患者究竟存在着什么重要的临床问题?用

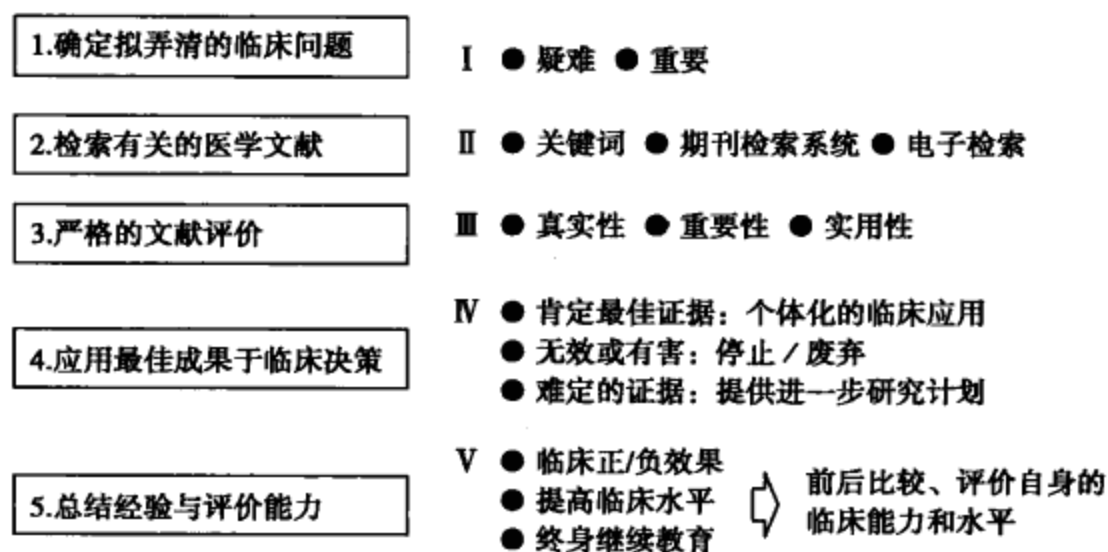


图1-2 实践循证医学“五步曲”图示

现有的理论知识和临床技能是否可以有效地解决? 如果棘手, 这就是循证医学应该回答与解决的问题了。

找准患者存在的需要回答和解决的临床问题, 是实践循证医学的首要关键环节, 如果找不准或者根本不是什么问题, 那么就会造成误导, 或者本身就on不是什么问题, 这就像一个临床科研选题的误差, 必然会造成研究的结果毫无价值一样。

为了找准重要的临床问题, 应该强调的是临床医生必须准确地采集病史、查体及收集有关试验结果, 占有可靠的一手资料, 充分应用自己的理论、临床技能和经验、思维性以及判断力, 经过仔细分析论证后, 方可准确地找出临床存在而需解决的且必须回答的疑难问题。

### (二) 检索有关医学文献

根据第一步提出的临床问题, 确定有关“关键词”、应用电子检索系统和期刊检索系统, 检索相关文献, 从这些文献中找出与拟弄清和回答的, 与临床问题关系密切的资料, 作为分析评价之用。

### (三) 严格评价文献

将收集的有关文献, 应用临床流行病学及 EBM 质量评价的标准, 从证据的真实性、重要性以及实用性作出具体的评价, 并得出确切的结论。这里将有三种结果: ①质量不高的文献, 或质量可靠但属无益或有害的干预证据者, 当弃之勿用; ②研究的证据尚难定论, 当作参考或待进一步研究和探讨; ③属最佳证据, 则可根据临床的具体情况, 解决患者的问题, 用以指导临床决策。如果收集的合格文献有多篇的话, 则可以作系统评价(systematic review, SR)和 meta 分析(meta-analysis)。这样的评价结论则更为可靠。

### (四) 应用最佳证据指导临床决策

将经过严格评价的文献, 从中获得的真实可靠并有重要的临床应用价值之最佳证据, 用于指导临床决策, 服务临床。反之, 对于经过严格评价为无效甚至有害的治疗措施则予以否定; 对于尚难定论并有期望的治疗措施, 则可为进一步的研究提供信息。

将最佳证据用于对自己的患者作相关决策时, 务必遵循个体化原则, 要对具体的情况作具体分析, 切忌生搬硬套。此外, 还要有涉及患者接受相关诊治决策的价值取向和具体的医疗环境及条件, 只有三者统一, 才可能使最佳决策得以实施, 见图1-3。

### (五) 总结经验与评价能力

通过对患者的循证医学临床实践, 必然会有成功或不成功的经验和教训, 临床医生应进行具体的分析和评价, 认真总结以从中获益, 达到提高认识、促进学术水平和提高医疗质量的目的; 此为自身进行继续

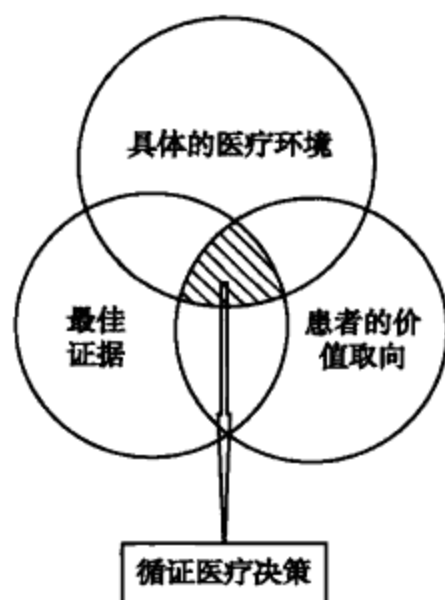


图1-3 循证医疗决策

教育和提高自我临床水平的过程。对于尚未或难于解决的问题,会为进一步研究提供方向。国外通过随机对照试验证明了EBM自我继续教育方式远优于传统的继续教育,进而成为培训临床专科医生的重要手段。

## 第五节 突发性公共卫生事件循证实践的方法

随着我国改革开放的日益发展,国际交往越来越多,加之国内社会主义经济建设的高度发展,城市化巨变和人口流动频繁、自然生态环境以及人为环境污染相应也发生着显著的变化。因此,疾病谱、有关的病原生物体必然也会发生某种变异,其致病性以及致病力也会发生相应改变,这些情况往往可能引起某种“突发性公共卫生事件”,造成地区性、全国性乃至国际性的威胁,像SARS和当前的甲型H1N1流感就是生动的例子。当然,这种事件可以发生在任何一个国家,如防治处理不当,则造成的危害往往无法评估!因此,面对突发性公共卫生事件的威胁,如何做到循证防治是个十分重要的问题。作为医学生和临床医生,学习与掌握这方面的基本知识与方法有重要的意义。因为当一个社会性群体突发卫生事件发生的时候,临床医生是不可能不参与人群的防治工作的。

根据循证医学实践的规律,如发生人群中的公共卫生突发事件,必然有群体性疾病发生,首获报告者必然是疾病预防与控制中心等有关部门。他们获悉后,必然要组成多专业的专家赶赴现场去进行调查诊治病人,调查发病因素与事件的性质,进一步应用知识分析判断并进行有关试验研究,以获取事件的确切证据,并要采取针对性的科学决策措施予以有效防治。

### (一) 事件现场循证防治的要素

在“事件”调查研究、取证和科学决策防治过程中,有关循证防治的重要因素是必须明确的:

1. 明确“事件”的性质,传染性或非传染性,亦即生物性或非生物性,如具传染性,其流行的可能性与程度。
  2. “事件”的危害性及其程度的估计。
  3. “事件”的传播方式与途径,如飞沫-呼吸道、消化道、接触、血源性等。
  4. “事件”波及危险群体有哪些,亦即危险人群的确定,如老年人、婴幼儿、流动人口等。
  5. “事件”易感危险因素有哪些,如年老体弱、营养不良、卫生习惯、居住环境、人群聚集交往等。
  6. 选定对“事件”科学、实用、有效的干预措施得出应急决策。
  7. 政府的行政干预以及干预的程度与范围、国家的法律和法规管理等。
  8. 对干预结局的科学分析与评价;并吸取经验与教训,不断地提高“事件”的防治干预水平。
- 如果面对突发性公共卫生事件,能将上述关键因素落实于具体的疾病、进一步得出相应的科学决策予以循证干预,相信是有助于增强防治效果和提高预防医学的水平。

### (二) “事件”的循证干预原则

鉴于生物性致病原体(细菌与病毒)所致疾病的传染性及流行性特点,危害性很大,程度也较严重。因此,就构成循证干预的重要内容。这里,无论病因是否明确,从宏观角度要遵循处理传染性疾病的“三大科学”原则:

1. 控制与消灭传染源。
2. 切断疾病的传播途径。
3. 增强人体的抗病力。

这三大原则要针对具体的疾病,而采取具体有针对性的、有效而可行的措施,方可达到理想的目的。对于非生物性-传染性的事件,应针对具体的性质与问题予以相应循证处理。现将突发性公共卫生事件的循证干预环节,归纳为图1-4,供参考。

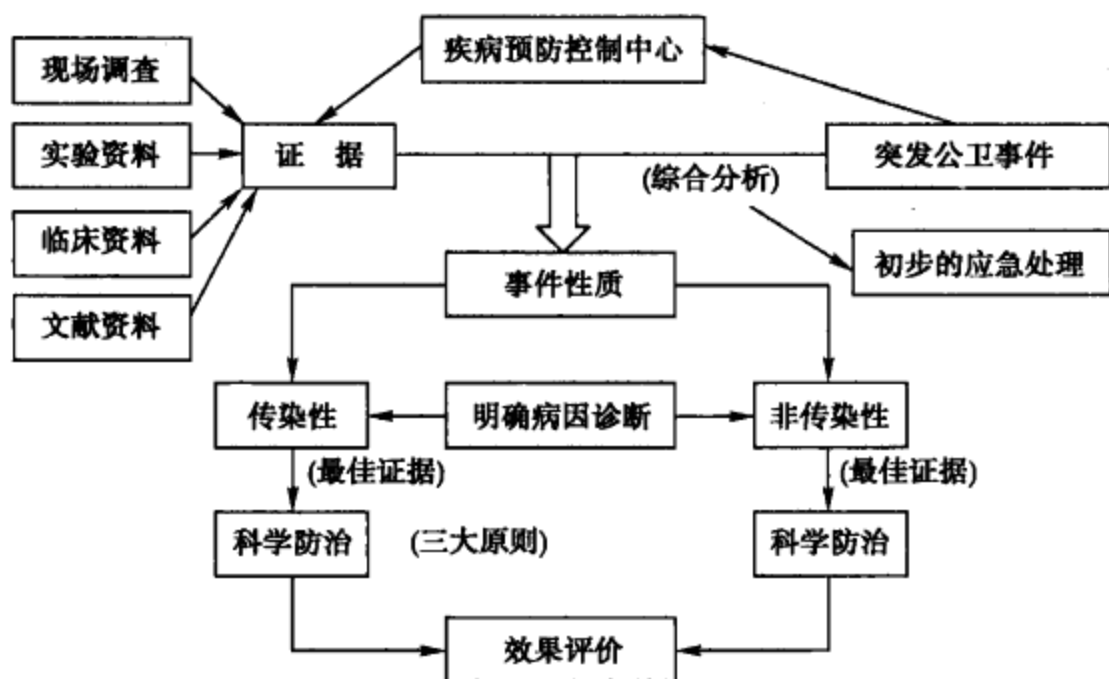


图1-4 突发性公卫事件循证实践示意图

## 第六节 循证医学实践的目的与意义

### 一、循证医学实践的目的

循证医学实践有着强烈的临床性,是为了解决临床医疗实践中的难题,充分地应用医学研究的最佳成果,指导临床医疗实践,促使永葆国际一流水平,以最有效地服务于患者,保障人民的健康,同时也培养高素质的临床医务人员,促进临床医学发展等为其根本的目的。

由于循证医学的概念被人们热情地日趋泛化,似乎包含了医疗卫生各个学科领域,甚至超出了它们的本身而成为当今“震荡世界的伟大思想之一”。毫无疑问:循证医学实践,在使用最现代化的科技信息手段,发掘与评价当今医学研究产出的最佳人类知识,遵循科学的客观规律,做到将先进的理论有机地联系实际,解决具体的临床问题,从而使人们的认识提高到一个新的水平。实际上这是人类本身实践着的科学发展观和认识世界的一个客观过程,只不过是在当今信息科学、人类生物科学、医学等领域知识爆炸和经济全球化的条件下,使得人们认识和改造世界的水平达到了一个新的高度而已,任何不尊重知识、凭经验或感觉,不按事物发展客观规律决策办事,导致临床医疗的失误实在是太多了。但是把循证医学当成一个“伟大思想”是不恰当的。从实践循证医学的本身,其目的归纳如下:

1. 加强临床医生的临床训练,提高专业能力,紧跟先进水平。循证医学要求临床医生要具有过硬的临床能力、敬业和创新上进精神,同时要有高尚的道德情操,并以患者为中心和尊重患者本身价值取向的服务热情。通过具体的EBM实践,提高医学教育水平并培训高素质的临床医生。

2. 弄清疾病的病因和发病的危险因素,弄清了有关疾病的病因或危险因素的证据,有利于指导健康者预防发病的一级预防;对于已经发病而无并发症的患者,也有利于作好预防并发症的二级预防;对于有并发症的患者,也有利于指导三级预防达到降低病死率或病残率的目的。

3. 提高疾病早期的正确诊断率,循证医学的特点,是要对严重危害人类健康的或预后较差的疾病,掌握与综合应用诊断性试验的证据,要力争作出早期正确的诊断,为有效地治疗决策提供可靠的诊断依据。

4. 帮助临床医生为患者选择最真实、可靠、具有临床价值并且实用的治疗措施;此外,还能指导临床合理用药,避免药物的不良反应。

5. 改善患者预后,分析和应用改善患者预后的有利因素,有效地控制和消除不利于预后的因素,以改善患者预后和提高其生存质量。

6. 促进卫生管理决策,应用最佳的研究证据于卫生管理,可促进管理决策的科学化,循证医学的发展对未临床医学的影响无疑是巨大的。

## 二、循证医学实践对临床医学与预防医学的意义

循证医学实践对临床医学以及预防医学的影响可大致概括为以下几个方面:

1. 促进医疗决策科学化避免乱防乱治,浪费资源,因而可提高临床医疗及预防医学水平,促进临床医学与预防医学的发展。
2. 促进临床与预防医学教学培训水平的提高,培训素质良好的人才,紧跟科学发展水平。
3. 发掘临床与预防医学难题,促进临床医学、预防医学与临床流行病学科学研究。
4. 提供可靠的科学信息,有利于卫生政策决策科学化。
5. 有利于患者本身的信息检索,监督医疗,保障自身权益。

最后,引用国际临床流行病学及循证医学创始人 Sackett 对循证医学实践者的四项要求作为本章的结束语:①必须有踏实的临床基本训练,正确收集病史、查体和检验,掌握患者的真实情况,方能发掘临床问题;②必须将循证医学作为终身自我继续教育,不断丰富和更新知识;③保持谦虚谨慎,戒骄戒躁;④要有高度的热情和进取精神,否则就要成为临床医学队伍中的落伍者。

(王家良)

## 第二章 构建临床和公共卫生的循证问题

临床医师对病人的诊治过程是一个不断提出问题、寻找方法、最后解决问题的过程。对一个患者实践循证医学(evidence based medicine, EBM)的第一步就是找出临床问题,构建一个需要回答的问题。能否找准患者急需解决的问题,对于循证医学的临床实践至关重要。而在遇到公共卫生事件时,为了找到该事件的原因以及相关因素,构建恰当的需要回答的问题更为重要。

### 第一节 概述

#### 一、找出临床问题的重要性

##### (一) 实施循证医学的第一步

临床医师每天面对患者时,应该善于在临床实践中观察,发现问题和提出问题。只有提出问题,才有可能带着问题去寻找证据,再根据可信度最强的证据并结合自己的临床经验和病人意愿最后解决临床问题,使患者获益。因此找出问题是循证医学临床实践的起点,找不准问题,就不能提出合适的问题,第一步走不好,必将影响循证医学后面几步的实施。构建一个可以回答的问题可帮助临床医师更好地制订收集证据的策略,便于回答和解决临床问题。当收集不到科学性强的证据时,临床医师可以根据此问题,提出进一步研究计划,作为研究者,通过研究,提供证据。

##### (二) 医学发展的需要

没有问题,不经过思考、总结、实践,医学就不可能进步,患者也不可能得到最好的诊断和治疗。临床医师不要认为自己在医学院学到的知识就足以回答每一个临床问题而不需要继续学习。对于某一问题的答案随着医学发展是会发生改变的,对一个临床问题认识的不断升华才能使之逐渐接近真实。

例如在诊断学教科书中一直将黄疸加上无痛性胆囊肿大即 Courvoisier 征作为胰头癌病人的重要体征介绍给学生,这是否是一成不变的定律呢?该体征用来诊断胰头癌是否敏感而且特异呢?这是在诊断学教科书上所没有学到的。提出这样的问题也是对1890年瑞士外科医生 Courvoisier 提出的此体征的重新评价。临床观察发现此体征在早期胰头癌中并不常见,进一步研究发现此体征用于鉴别良恶性胆道梗阻的价值并不高(阳性似然比2.6),部分胆石症病人也可出现此征。同时,此体征在诊断肝外胆道梗阻性黄疸时的敏感度也较低(37%),因此 Courvoisier 征阴性时并不能排除胰头癌的诊断。从目前临床情况看,此体征对于胰头癌诊断的临床价值较100年前有所降低。这是100多年来医学诊断技术进步的结果,因为随着影像学技术的发展,B超、CT、MR、EFRCP 和超声内镜的应用,在患者出现 courvoisier 征前,就已经被诊断出胰头癌。由此可见,终生学习是必不可少的,只有不断提出问题,寻找答案,才能使医学发展进步。

##### (三) 循证医学所赋予的任务

EBM 实践应以解决患者所患疾病存在的重要临床问题为中心。为此,无论是临床医生或 EBM 的学习者务必抓住患者的临床关键难题,而这些难题关系患者的安危,对 EBM 实践者而言,则往往是乏知的。因此,EBM 的第一关键是找准患者存在的、而医务人员必须回答的临床难题。

## 二、找准临床问题应具备的条件

### (一) 对患者的责任心

对患者有责任感,关心患者,同情患者的医生,会以患者为中心去考虑问题,也会在与患者的交谈和观察中发现更多的临床问题。

### (二) 要有丰富的医学基础知识和临床医学知识

人体无论哪一系统的疾病都有其规律,不了解病因、发病机制和临床表现,不熟悉各种诊断试验和辅助检查的特性、适应证,不了解各种药物的治疗机制,其药理作用及可能发生的不良反应,碰到一个具体的患者,就不可能提出适当的问题。因此具备系统扎实的医学知识是找准临床问题的必要基础。

### (三) 具有一定的人文科学及社会、心理学知识

随着医学模式的改变,许多患者疾病的发生与心理、精神因素有关。也有一些疾病的发病虽然与此关系不大,如慢性肝病、肿瘤,但患者在患病后对疾病的认识和心态会影响其病情及预后。因此,要在这方面去发现问题,了解患者对此病的想法、期望及忧虑。还要了解患者的社会经济状况及家庭负担等。具有一定的人文科学、社会 and 心理学知识,才能与不同性格的患者顺利沟通,交流思想,从而发现患者在心理上存在的问题,并帮助解决,这本身也是治病的一部分。

### (四) 扎实的临床基本技能

包括如何接触病人,采集病史,全面的体格检查和对诊断试验选择与鉴别能力。对病人务必弄清病史,要认真查体,了解入院时情况,如疾病的严重度,掌握重要的阳性体征和阴性体征。了解与疾病有关的实验室和辅助检查资料结果。在此前提下才可能找出病人迫切需要解决的问题。

### (五) 临床综合分析的思维和判断能力

应用已掌握的医学理论知识和临床经验,结合患者临床资料进行综合分析、逻辑推理,从错综复杂的线索中去伪存真、去粗取精,找出主要矛盾,并加以解决的临床思维过程,也是找准临床问题,做出决策的必备条件。

上述五点是寻找和提出临床问题的重要必备条件,任何一点不具备,均不利于找准病人的临床问题。

## 三、提出公共卫生事件循证问题的重要性

突发的公共卫生事件例如 SARS、甲型 N1H1 流感、奶粉中毒等,刚发生时,对其发生原因以及相关因素都不甚了解,要做出正确的决策有很多困难。此时,如能应用已掌握的公共卫生知识和临床经验,提出公共卫生事件的循证问题,找到证据,将为做好循证决策起重要指导作用。

## 第二节 如何构建临床循证问题

### 一、临床问题的类型

由于 EBM 实践者可以是医学生直至高年资临床医生,鉴于层次与阅历不一,在临床实践中即使面临同一患者,由于视角与水平不一,发现和提出的临床问题( clinical question )会大不一样,这些问题可大致概括为以下三个方面:

#### (一) 一般性的问题

主要由 EBM 初学实践者提出,除了具备基础医学知识外,往往需要有关人文科学、社会、心理学的知识。

1. 涉及患者的一般知识性问题 如患者性别、年龄等。
2. 涉及有关所患疾病的基本问题 如某个具体的患者,存在什么临床问题,在什么地方、何种环



境下发病,何时发病,如何发病,有关因素是什么等;此外,患者的主要临床表现又是什么。初学者有兴趣也会提出一些未知的欲求解答的问题。

## (二) 特殊的临床问题

这是临床医生对患者的诊治过程中,在充分掌握了患者病史、临床体征、有关检查资料之后,通过临床综合分析,从专业角度所找到的问题。

1. 患者存在的特殊问题 这些问题不解决则必影响对病人的正确临床处理。例如一个肝硬化患者,近期腹水明显增多,对于这个患者,提出“其腹水有无感染”就是一个十分重要的临床问题,不能确定其是否合并自发性腹膜炎,就无法对其进行正确的治疗。

2. 干预 如何进行相应干预,这往往涉及病因、危险因素的暴露干预、诊治、预后、病人的依从与理解等。例如对一例消化性溃疡患者进行治疗时,必须先对病因提出问题,患者有无幽门螺杆菌感染,有无服用非甾体抗炎药病史,有无其他应激状态等,这些都影响到治疗方案的选择。对于慢性活动性病毒性肝炎病人的治疗,患者的依从性与理解就更显得重要。

3. 干预措施的选择 干预措施也有许多种,每一种措施都有其利和弊,这就存在如何比较抉择的问题。如对恶性肿瘤患者采取手术还是化疗,还是介入性治疗或放疗,不仅要根据疾病病情的分析,解决关键问题,将各种措施的利与弊罗列出来进行比较,还要考虑到患者经济能力与家属沟通进行决策。

4. 干预的最后结局问题 这是作为 EBM 实践者追求最佳结局所感兴趣的问题。结局可以是症状体征改善或者是生存率、死亡率和致残率,使用不同的结局指标,找出的问题也不尽相同。

总之,以上这四个环节是一个有机整体,作为 EBM 实践者在发现临床问题时,一定要牢牢掌握。

## (三) 患者所关心的问题

应结合患者的具体情况提出问题。例如同一疾病不同年龄段的患者所关心的问题是不同的。一项1012名乳腺癌妇女的研究发现,不同年龄段妇女关心的治疗结局是不同的。70岁以上的妇女最关心的是癌症治愈和转移的可能性;小于50岁的妇女关心的是治疗对其性功能的影响;有阳性家族史的妇女最关心的是该病是否有遗传性。因此应针对不同患者的不同情况提出临床需要解决的问题。

## 二、提出临床问题的形式和方法

### (一) 提出临床问题的形式

现在许多学校已经开展以问题为中心的学习,即 problem-based learning,这里的“问题”(problem)是指患者存在的一种症状或体征,例如有无黄疸或其他情况。而本文中所指的“临床问题”是一个可以回答(answerable)的问题(question)。例如“对这位黄疸患者在选择辅助检查时,B超和CT哪一项更好?”。因此,尽管“problem”和“question”均可翻译成“问题”,容易混淆,但两者内涵完全不同。下面讨论提出临床问题(question)的形式。

1. 一般性临床问题 一般性问题是与患者或患者所患疾病有关,即一般知识性问题,由以下两部分构成:

(1) 由问题的词根(谁、什么、何处、何时、怎么样、为什么)加上动词构成:这些常常在患者入院时通过询问病史和体格检查得到。例如对每一项主诉都应包括症状发生的部位、严重度、数量(如出血量)、起病情况(急性还是慢性、持续性还是进展性)、在什么情况下发生、加重和缓解因素、相关的其他症状等;了解以往是否发生过与主诉相同的情况;曾经做过哪些检查;是否曾经有过治疗及如何治疗;对其预后有意义或对主诉疾病治疗有影响的过去史情况;这些相关疾病的治疗情况等。例如呕血作为一个动词,就必须弄清谁呕血(患者的性别、年龄特征),呕血的性质(颜色、量、次数)、何时、何地发生呕血、呕血时病人有无其他症状及什么是发生呕血的主因和诱因及其基本病变等。

(2) 一种疾病或疾病的某一方面:例如“什么原因引起发热?”“急性胰腺炎通常在何时发生并发病?”等。

2. 特殊性的临床问题 在临床实践中,患者与医生均会在诊断、治疗、预后、预防、病因等各个方面提出许多需要解决的临床问题。例如患者常常会问医生“我患的是什么病?”(诊断问题)、“我为什么会患这个病?”(病因问题)、“这个病应该用什么方法进行治疗”(治疗问题)、“这个病对我健康有多大影响,会不会影响我的寿命?”(预后问题)。医生在诊治不同疾病以及同一疾病的不同患者时,提出的问题可能各不相同,归纳起来包括以下几个方面:

(1) 病人本次入院或来门诊求诊需要解决的问题以及在入院后由于病情变化产生的新问题。医生可以对患者发生的每一项症状或体征提出问题。例如对于上述的呕血患者,在求诊时急需解决的主要问题是止血及弄清呕血原因;在出血停止后,患者又出现了计算能力下降,昼夜颠倒,扑翼样震颤,此时患者需要解决的紧要问题就是弄清是否出现了肝性脑病,并对此采取措施。也可以在正确获取和合理解释病史和体检有关新发现的过程中,提出问题。例如一位中年男性,因发现黄疸而求诊。在体检时扪及胆囊肿大而无压痛(Courvoisier征),提出的问题应为“此征对于梗阻性黄疸和肝细胞性黄疸的鉴别诊断是否有意义”,进一步可提问其对于鉴别肝外梗阻的原因,即结石引起还是肿瘤引起有否帮助。

(2) 诊断方面的问题:对于初学者在诊断方面常常提出的问题是某个体征、症状或某项实验室和辅助检查对于该病的诊断效率,即提出有关诊断试验的敏感度、特异度和似然比等问题;而对于有多年临床工作经验的医生常常提出的问题是某项检查对于鉴别诊断方面的意义。通过病史询问和体检,医生会有一个诊断假设。为了证实该假设,医生可能会进行一些实验室或辅助检查来肯定或排除此诊断假设,此时针对诊断试验指标如敏感度、特异度、似然比等可提出问题,对其正确性、可靠性、可接受性、费用及安全性方面也可提出问题。例如对上述黄疸病人的体征(Courvoisier征阳性)及鉴别诊断所采用的B超、CT、MRCP等均可作为诊断试验并提出相应的问题。进一步找出最适应于患者的检查方法。

例如对一位呕血患者,为了寻找出血部位和原因,是否应作急诊胃镜检查?仅凭这一点就可以找出许多临床问题,如“急诊胃镜检查对诊断上消化道出血的敏感度和特异度如何?”“急诊胃镜检查对此患者带来的风险有多大?”“对肝硬化患者和非肝硬化患者带来的利和弊有无差别?”“急诊胃镜检查的诊断结果是否会影响医生对治疗方案的选择?”“有无其他可供选择的诊断措施?”

在选择诊断试验前,还应对患者的验前概率,即患者在没有进行此项诊断检查前患病的可能性大小提出问题,对于上述呕血患者在没有做急诊胃镜检查前,应判断出由食管静脉破裂出血引起呕血的概率有多大?这就与患者的基本情况有关,如果这位患者有肝硬化病史,则食管静脉破裂出血的可能性较大,如果以前有过类似出血史,而且胃镜已证实是由食管静脉曲张引起,则其验前概率就更高。根据验前概率,就可以提出问题,即“做急诊胃镜检查的结果是否影响对此患者所采取的治疗措施?”相反,如果该患者为一老年患者,无肝硬化病史,长期服阿司匹林,据此判断其因食管静脉出血的概率就较小,而急诊胃镜的结果可用来证实或排除食管静脉出血,其诊断结果对其治疗方案选择的影响较大。此外,如果已经对患者作出了初步诊断,还可以提出问题即“能否通过某项诊断试验确定该病的严重程度,又用何种指标来测量随访病人在治疗后的改善情况”等。

(3) 治疗方面的问题:如何选择利大于弊的治疗手段?如何从效果和成本的经济学角度选择治疗方案?特别是如何对目前的常规疗法提出质疑,提出的问题包括根据患者目前病情可以采用什么治疗方法,该治疗方法的有效性如何?有什么不良反应?还有什么替代治疗手段?哪一种方法更有效而花费最少?该治疗对患者的生存质量有何影响?治疗后对患者的预后影响如何?患者对治疗手段的依从性和可接受性如何?例如上述患者诊断为胰腺癌,则根据疾病的严重度选用外科手术或姑息疗法。如果采用后者,为了减轻黄疸危害和提高生存质量,可以通过内镜放置内支架引流减轻胆道梗阻,使黄疸缓解。提出问题是“安置金属支架,还是塑料支架,何者为佳?”“为加强支持治疗对病人是否要给予静脉内营养”。又如在公共卫生方面,“采用中医药对人群中防治甲型H1N1流感是否有效?”等。

而对于食管静脉出血的肝硬化患者,在其出血停止后,为了预防再次出血,可供选择的方案有外科分流或断流手术治疗,内镜下圈套或注射硬化剂,口服 $\beta$ -受体阻滞剂、介入疗法等。此时必须根据患者情况将这些措施预防再出血的效果、风险、后遗症、疗程以及对生存率的影响、费用等逐项列出进行比较选择。找到证据后,医生应结合患者病情提出建议,并在征求患者意见的基础上作出决策。

(4) 病因方面的问题: 包括怎样识别疾病的原因及危险因素? 其发病机制是什么? 例如对于胰腺癌患者提出病因问题包括: 发病的原因是什么? 有无遗传因素? 发生胰腺癌的危险因素是什么? 是否与喝咖啡或与饮酒有关。又如甲型 H1N1 流感患者的致死率有多大? 致死的病因或危险因素是什么? 弄清这些问题对有效防治是很重要的。

(5) 预后方面的问题: 如何来估计临床病程和预测可能发生的并发症和最后结局。针对不同的结局测定指标可以提出不同的预后问题。例如上述预防食管静脉再出血的干预措施,对“再出血的发生率”和“患者的生存率”两种预后效果是否不同?

### (二) 提出临床问题的参考方法

当临床实践中遇到患者存在的难题, EBM 实践者要解决这些问题却存在知识能力不足时, 这就要找准问题并记录下来, 然后通过自己的临床思维, 进行整理, 将其排序, 先抓好关键问题, 并做出如何解决这个(些)问题的策略计划, 有的放矢去查阅文献, 然后进行文献评价, 选择最佳证据, 以解决患者的问题。这里在找临床问题的方法上, 要掌握的是: ①涉及的问题一定是与患者的诊治处理和对病人健康恢复最相关的; ②涉及的问题一定是与实践 EBM 提高医疗水平最为相关的; ③涉及的问题一定是临床上最感兴趣的、最有用的; ④涉及的问题往往也是实践 EBM 中最为常见的。

如果 EBM 实践者在临床医疗日常工作中, 对各种不同患者的难题做到日积月累, 并不断用最佳证据予以解决, 终身坚持, 必成名家, 并对 EBM 作出更大贡献。

### (三) 针对病人实际情况提出问题

上述这两种类型的问题几乎包括了所有需要提出的临床问题(question)。对于学生来讲前一类一般性的背景问题可能多一些; 对于有经验的医生来讲后一类问题多一些。有些问题不需要进一步查资料就可以回答, 但就教学而言, 有必要提出, 让学生了解如何针对病人实际情况提出可回答问题。

为了进一步了解病人的情况(一般性问题)和对病人进行进一步处理(特异性问题), 在上述工作基础上可以书写教育处方(educational prescription), 即提出患者急需解决的临床问题(question), 以进一步寻找答案。教育处方提出的可回答的问题必须十分具体。例如一位重症胰腺炎病人, 在讨论治疗措施时, 不能提类似“重症胰腺炎病人如何治疗?” 因为这样的问题范围太宽, 如果据此去检索文献, 会有上千篇文章, 最终无法归纳总结来回答此问题。因此, 提出的问题必须具体到某一项措施, 例如有人提出“对重症胰腺炎患者是否需要用静脉内营养或肠内营养”的问题, 结合患者实际就可以这样提出问题: “全胃肠外营养和肠内营养对于急性重症胰腺炎在降低感染发生率、减少并发症发生率、缩短住院时间和降低死亡率方面, 哪一种方法较好?” 为回答这个问题, 可以用检索词“parenteral nutrition or TPN and enteral nutrition and acute severe pancreatitis”检索文献, 寻找答案。由此可见, 构建的问题必须包括对象(某种疾病、症状或患者)、需要比较的措施, 这样查找出来的结果, 才能对临床医师决策有所帮助。

### (四) 为临床科研提出问题

临床实践是临床科研选题的丰富源泉, 日常医疗实践中, 无时无刻不面临许多上述诊断、治疗、病因、预后等问题, 不少诊断方法和治疗手段有待于进一步的科学评价。从临床需要出发提出问题, 用可靠的方法进行研究, 以得到可靠证据回答所提出的问题, 解决临床问题, 再用于指导他人的临床实践。

## 三、构建临床循证问题的模式

在构建一个具体的临床问题时, 可采用国际上常用的 PICO 格式。P 指特定的患病的人群

(population/participants), I 指干预(intervention/exposure), C 指对照组或另一种可用于比较的干预措施(comparator/control), O 为结局(outcome)。每个临床问题均应由 PICO 四部分构成。图2-1显示了3个临床问题的组成方式。①对于慢性肾衰竭尿毒症病人肾脏移植与血液透析相比,在生存率和生存质量上哪种方法好? ② ACEI 与 CCB 合用与单用 CCB 相比,在保护肾功能、降低血压和尿蛋白方面是否有更多的作用? ③对于频发的尿路感染,长期小剂量应用抗生素是否能预防复发? 根据 PICO 中的关键词,便于进行检索。

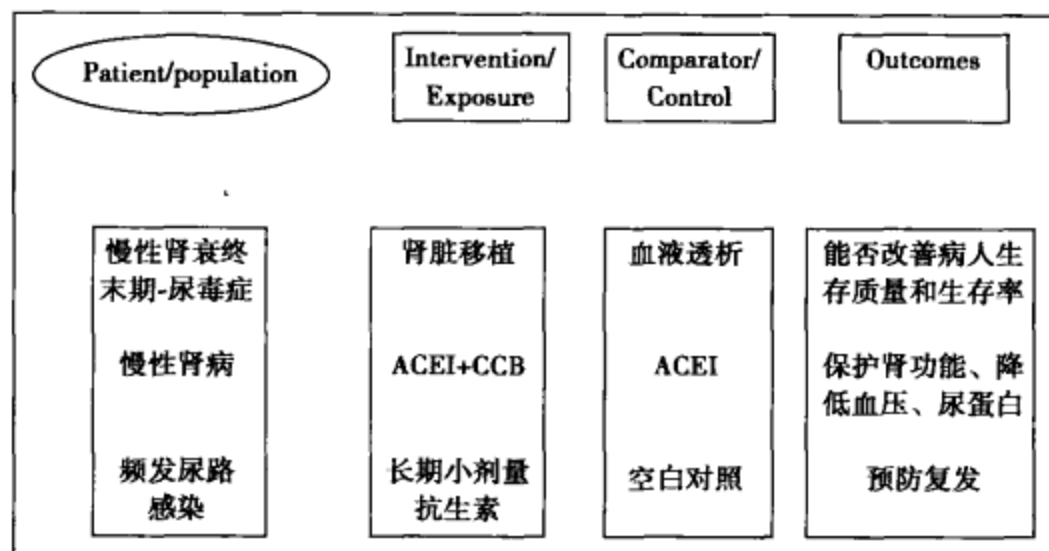


图2-1 临床问题的组成 PICO

总之,要提出一个好的临床问题,需要具备系统扎实的基础与临床专业知识和技能,深入临床实践,善于思考,跟踪本专业研究进展,学会以患者的角度考虑,易提出构建良好的问题。

### 第三节 如何提出公共卫生领域的循证问题

与临床实践以病人个体为中心不同,公共卫生领域的实践对象是人群。在该领域主要围绕三个方面提出问题:一是“what”类问题,即在实践之前需要知道卫生需求以及卫生资源的大小和分布,以制定相应政策,计划干预;二是“how”类问题,即对正在进行的实践项目进行评价,围绕其卫生需求范围、目标人群、质量、成本以及效果或影响等,评估项目的进展以及判定是否需要进行调整;三是“why”类问题,即确定在实践过程中发生的问题,分析其可能的原因,并找到解决的办法。

#### 一、问题的提出及排序

##### (一) 问题的提出

在公共卫生领域中,若遇到一般性问题,不需研究就可以解决。例如,在卫生项目中用于基础设施建设的材料(如水泥)常供应不上,影响了进度,此时需要想办法保证水泥的供应即可,没必要当作问题开展研究。但是若出现:①实际与理论设计脱节;②原因不明问题;③需要在多种解决方案中作出抉择时,则需要当作问题进一步研究。

例如,某地区人口约145 000,总体卫生状况较差,仅有5%的住户家中有厕所,与卫生条件有关的诸如肝炎、肠胃炎等疾病很常见。为此,卫生部门准备实施改水改厕计划,要求改厕户数每年增加15%。然而,两年后目标只实现了不到一半。出现这种情况,可能与两个方面的因素有关:①服务相关因素,如对当地群众的动员力度不够,材料供应不及时,培训效果差,以及卫生工作人员的效率低等。②人群相关因素,如社区成员缺乏对卫生条件差会致病的认识等。对此,提出的问题是:“卫生部门改厕项目不达标的主要原因是什么?”



## 二、问题的构建

### (一) 确定问题的范围

1. 从不同角度提出问题 对于同一问题,卫生领域的研究者、管理者、一般工作者以及社区成员常会从不同的角度来理解,有着不同的看法。收集这些相关人员所关注的重点是非常必要的,可以保证所提问题的全面性和客观性。管理者和卫生工作者一般会从宏观层面上表达他们关心的问题。例如“我们需要调查一下糖尿病患者的自我管理情况”。然而这样的问题比较模糊、无从下手。需要将问题进一步具体为:①糖尿病患者及其亲属对糖尿病和自我管理的措施认识不够;②用以长期随访护理的辅助设备不足;③糖尿病患者的再住院率很高;④在糖尿病患者中对并发症的管理不恰当;⑤糖尿病并发症的发病率很高;⑥病人对治疗的依从性较差等。

2. 确定问题的核心和范围 第二步就是要确定问题的核心,并将其定量化。如上例中可确定两个核心问题:①糖尿病患者的再住院率很高;②糖尿病并发症的发病率很高。

围绕核心问题,进一步明确其内涵与外延,如:①问题的性质:可以从再住院和(或)依从性两个方面来说明“实际情况是……”与“理想的情况是……”之间存在的差别;②问题的分布:谁受到了影响、何时、何地;③问题的大小和强度:这个问题是否很普遍,有多么严重,结局是什么(如残疾、死亡和资源的浪费)。

3. 问题的归类 在确定了核心问题后,还需要进一步将问题归类。

(1) 确定可能引起问题的因素:包括:①社会文化因素,如个体因素(如年龄、性别、教育、职业、家庭组成以及来自家庭的支持等)和社区因素(如整个社区的成员对疾病和治疗的认知水平,社区中一些治疗的可得性,社区成员对某些治疗的偏爱等);②与服务有关的因素:如某些治疗、护理难以实现(包括治疗的成本),医院的管理较差(如治疗方法不当、对病人的咨询服务不够)等;③与疾病有关的因素:如患者开始治疗时疾病的严重程度,个体对治疗的反应情况(病情控制的难易程度)等。

(2) 阐明问题和影响因素之间的关系:这里需要注意的一点是,若对影响问题不同因素的性质、作用、权重和多种因素之间的关系不完全确定时,最好不对这些因素进行分类,建议只选择其中一部分。当然这会有一定的危险性,有可能遗漏重点问题。

### (二) 问题的描述

问题的清楚描述很重要,首先是进一步制订研究计划(研究目的、方法、工作计划、预算等)的基础;其次能够更容易发现类似的研究报告或信息,并从中获益;再者可以明确研究目的、预期结果或达到目标。这一点很重要,将有助于社区成员、卫生工作者、有关部门和资助机构及时了解计划安排,以寻求支持或配合。

问题描述应包括以下信息:①简要描述社会经济和文化特点,概述国家或地区内与问题有关的健康状况和卫生保健系统的情况。如果能够得到相关的统计数字将有助于问题的描述。②简明扼要描述问题的性质(“事情是……”和“事情应该是……”之间的差别),及其范围、分布和严重性(谁受到影响、哪里、自何时起,结果是什么?)。③分析问题的可能影响因素,应说明现有的知识还不足以解决问题。④简要的描述过去曾采用过的解决方案,其效果如何,为什么需要进一步的研究。⑤描述期望得到什么类型的信息,以及如何利用这些信息来帮助解决问题。⑥如果必要,可以将问题描述中涉及的一些重要概念、定义以目录形式列出。

(王吉耀 吕筠)

## 第三章 检索与收集需求证据的方法

哈佛大学医学院的院长 Sydney Burwell 博士曾说过：“医学生在学校接受的知识，10年后其中一半可能是错误的，而可悲的是没有人能预测哪一半是错误的”。因此，为了不被过时、错误的知识所误导，我们需要针对临床工作中遇到的问题，不断查寻新的文献资料，掌握学科发展的历史、现状和发展趋势，更新知识结构。而繁忙的临床医师如何在众多医学刊物和数据库中快速、有效地查寻相关领域最佳信息，是我们共同面临的挑战。

### 第一节 循证医学证据资源

临床医生面临临床问题时常查寻的信息资源包括教科书、专著、中国生物医学文献数据库、Medline 等或咨询专家。全世界有许多医学文献数据库，每个专业或亚专业均有其独特的数据库资源或检索工具。但并不是每个数据库提供的信息资源都是真实、可靠且能方便、快捷地获取。

#### 一、循证医学证据资源发展

循证医学证据资源经历了漫长的发展历程。20世纪90年代前，临床医生主要依靠手工检索工具和光盘先获取题录，再查阅全文。这种方式耗时且不一定能获得最新证据信息。随着计算机和网络技术发展，信息的储存、传输和利用进入了一个新的里程碑。人们逐步摆脱耗时、费力、检索效果差的手工检索。快速、高效的联机检索克服了时空障碍，极大提高了获得最新信息的能力。大存贮量的光盘和联机检索极大降低了计算机检索的费用，使我们更及时、准确、全面利用人类科研的成果。在此基础上，信息专家、方法学家和临床专家们逐步考虑如何从浩如烟海的信息资源中去粗取精、去伪存真，为繁忙的临床医生提取真实、可靠且有临床实用价值的信息。从1991年美国内科医师杂志俱乐部(ACP Journal Club)创刊、1996年 Iain Chalmers 等创立 Cochrane 协作网和建立 Cochrane 图书馆、1999年 BMJ 推出 Clinical Evidence，越来越多的循证医学信息资源供临床医生选择。这些资源或从方法学质量和临床价值方面严格评估原始临床研究，或采用系统评价和 meta 分析方法整合针对同一临床问题的高质量原始研究，或针对临床主题总结治疗某一种疾病的所有证据，帮助临床医生明确哪些干预措施有效、哪些无效甚至有害，极大方便了临床医生开展循证临床实践。随后，各大医学信息数据库提供商相继推出 PIER(Physician's Information Education and Resources)、DynaMed、UpToDate 等数据库，集前几类资源优势于一体，在总结证据基础上，结合专家经验、患者价值观给出推荐意见，并根据证据的质量给出推荐强度。至此，临床医生不再需要自己花费大量时间从 PubMed、EMbase 等原始文献数据库中去检索、获取全文、评价和总结临床研究证据。这类资源的不断完善，使实践循证医学成为可能。

国内目前还没有类似的中文数据资源，使用这些英文资源，国内医生将面临语言、医疗环境差异和费用等难题。

#### 二、循证医学证据资源分类

Haynes 等于2001年和2006年分别提出了循证医学资源的“4S”和“5S”模型，但“5S”模型是

比较理想化的模型,因此我们根据“4S”模型将信息资源分为4类,即证据系统(system)、证据摘要(synopses)、系统评价(syntheses)和原始研究(studies)。

### (一) 证据系统

证据系统即计算机决策支持系统(computerized decision support system, CDSS),是指针对某个临床问题,概括总结所有相关和重要的研究证据,并通过电子病例系统与特定患者的情况自动联系起来,为医生提供决策信息。现有的数据库尚不能达到如此高智能化程度,但已有一些循证医学数据库具有部分功能。如Clinical Evidence (<http://www.clinicalevidence.com>)、由美国医师学院提供的PIER(Physician's Information and Education Resource, <http://pier.acponline.org/index.html>)、UpToDate(<http://www.uptodate.com>)、Harrison's Principles of Internal Medicine(<http://www.harrisonsmed.com/>)。

### (二) 证据摘要

证据摘要即循证杂志摘要(evidence-based journal abstracts)。为了帮助繁忙的临床医师快速、有效地查寻文献,方法学家和临床专家共同组织起来,制定严格的评价标准,对主要医学期刊上发表的原始研究和二次研究证据从方法学和临床重要性两方面进行评价,筛选出高质量的论著以结构式摘要的形式再次出版,并附有专家推荐意见。如美国内科医师学院杂志俱乐部(American College of Physician Journal Club, ACP Journal Club, <http://www.acpjc.org>)和InfoPOEMs(<http://www.infopoems.com/>)。

### (三) 系统评价

系统评价是针对某一具体临床问题(如疾病的病因、诊断、治疗、预后)系统、全面收集全世界所有已发表或未发表的临床研究,严格评价纳入文献的偏倚风险,筛选出符合质量标准的文献,进行定性或定量合成(meta-analysis, 荟萃分析),得出可靠的综合结论。相对于单个原始临床研究,系统评价对精力、时间有限的临床医生来说更实用。系统评价分为Cochrane系统评价和非Cochrane系统评价,前者由Cochrane协作网(<http://www.cochrane.org/index.htm>)的作者制作并发表在Cochrane图书馆,后者发表在杂志上。

### (四) 原始研究

发表在杂志和综合文献数据库、未经专家评估的文献资料。临床医生在检索和应用此类文献,需要自己进行评估研究结果的真实性、临床重要性和适用性后方可应用,否则可能误导。

## 三、选择证据资源的标准

针对某一临床问题,要快捷、高效检索相关信息,正确选择信息资源或数据库十分重要。如何选择信息资源呢? McKibbon 等列出了4条标准(表3-1)。

表3-1 选择或评估信息资源的标准

选择标准	标准内容描述
循证方法的严谨性	✓ 推论是否严格遵循证据?
	✓ 提出推荐意见时是否给出支持其结论的证据强度?
	✓ 是否为读者提供了证据链接以方便阅读?
内容的全面性和特异性	✓ 是否充分覆盖了我的专业领域或内容范围?
	✓ 是否覆盖了我提出的问题类型(治疗、诊断、预后、病因或不良反应)?
	✓ 是否针对我临床实践的具体专业领域?
易用性	✓ 能否快速、始终如一地提供我需要的信息?
可及性	✓ 是否在我需要使用的任何场所均能方便获取?
	✓ 是否能支付其费用?



### (一) 循证方法的严谨性

循证医学信息资源需要针对临床问题提供具有代表性的高质量证据。在总结证据资源时,应精心构建临床问题、系统全面检索相关证据、严格评价单个研究的真实性、恰当总结和合成结果。循证医学信息资源在提出推荐意见时要充分应用已有的系统评价或自己制作系统评价,针对患者重要的结局指标(patient-important outcomes),为不同干预措施的疗效和安全性提供最佳估计,并在充分考虑患者的价值观和选择基础上,采用恰当的分级系统对推荐意见进行分级。

### (二) 内容的全面性和特异性

理想的循证医学信息资源应该为临床实践中可能遇到的所有问题提供相关证据。但针对某一专业领域的信息资源可能会更有效地帮助查寻需要的证据。如要紧跟某些临床专业的最新进展,可查寻某些循证的摘要如 Evidence-based Cardiovascular Medicine, Evidence-based Mental Health 和 Evidence-based Oncology 等。

某些信息资源是专门针对某一类临床问题,如 Clinical Evidence 和 Cochrane 系统评价数据库,目前仅涉及治疗性临床问题,且后者仅包括临床对照试验的系统评价。

### (三) 易用性

某些信息资源查寻起来方便快捷,如美国内科医师学会杂志俱乐部(ACP Journal Club)为小型数据库,收集了内科领域约140种杂志上发表的与临床最相关的高质量研究摘要,其优秀的检索引擎确保读者能很容易查寻任何该领域的信息,包括疾病病因、诊断、治疗和预后等。

MEDLINE 是检索世界生物医学文献资源最主要的数据库之一, MEDLINE 的检索可采用联机检索、光盘检索,近年来由于计算机互联网的应用,90% 国际联机检索系统都已进入互联网。通过隶属美国国立医学图书馆的国家生物技术中心开发的免费的 PubMed(<http://www.ncbi.nlm.nih.gov/pubmed/>)网络检索系统,可检索 MEDLINE 数据库。由于该数据库容量庞大,2008年初已包含约1700万篇文献,每年以70万条速度增加,其检索速度较慢。PubMed 是检索 MEDLINE 较容易的途径,其专门为临床医生设计的“Clinical Queries”检索方式可将检索结果最大限度局限于与临床决策相关的文献资料。

Cochrane 系统评价尽管容易检索,但每一篇系统评价包括内容较多、统计分析结果较复杂、报告质量各异,临床医生需要花很多时间阅读和理解方可应用于临床。

### (四) 可及性

最可靠和最有效的信息资源价格往往非常昂贵。临床医生通常利用所在医院或医学院校图书馆检索网上资源,不太可能个人订阅昂贵的期刊或数据库。中低收入发展中国家的临床医生可通过学术机构的网络进入世界卫生组织的健康网络研究启动项目(Health InterNetwork Access to Research Initiative, HINARI)获取免费信息资源。但某些信息资源是向全世界免费开放的,如 PubMed, Canadian Medical Association Journal(<http://www.cmaj.ca/>)和大多数 BioMed Central 杂志(<http://www.biomedcentral.com/>)。还有一些杂志在发表6~12月后可免费检索或在发表当时部分免费检索,如 BMJ、JAMA、the Mayo Proceedings 等。

## 四、常用循证医学证据资源

根据“4S”模型,下面简单介绍常用的各类数据库资源。

### (一) 证据系统(计算机决策支持系统)

1. Clinical Evidence (<http://www.clinicalevidence.com/ceweb/conditions/index.jsp>) 由BMJ出版,是世界上最具有权威性的医学数据库之一,以治疗为主,涉及200多种疾病的2500多种治疗方法,每年更新一次并在不断拓展新的题目和领域(如疾病诊断)。该数据库针对每种疾病,采用严格的过程评估每种治疗方法的疗效和安全性,告诉读者哪些治疗方法有益、哪些可能有益、哪些利弊相当、哪些不可能有益、哪些可能无益或甚至有害和哪些疗效不确定。

Clinical Evidence 方便易用,但除其成员和中低收入的发展中国家外,均需付费。北京大学的循证医学中心曾将其15版翻译为中文《临床证据》于2008年出版,方便国内读者,但存在更新问题。

2. PIER( Physicians' Information and Education Resource ) 是美国内科医师学会的产品( <http://pier.acponline.org/index.html> )。主要包括疾病诊治、筛查与预防、补充/替代医学、伦理和法律问题、流程、质量测量方法和药物信息,其优点是采用多层次结构指导临床医生应用研究证据,所有问题均采用同样结构,所有推荐意见均与研究证据紧密相连。PIER 提供的推荐意见是基于严格的循证医学方法,包括精心构建问题、全面收集所有干预措施和以病人为中心的结局指标、评估单个研究的质量、采用高质量的分级系统、充分考虑患者的价值观和选择。PIER 主要涉及内科和初级保健方面的治疗问题,覆盖面欠佳。

PIER 方便易用,但需要付费。PDA 版本收费标准约100美元,ACP 成员免费。

3. UpToDate( <http://www.uptodate.com/index.asp> ) 是一电子信息资源,可从网上、台式计算机和 PDA 获取,由于其使用方便、覆盖面广和根据疾病分类收集信息,深受全科医师、专科医师和家庭医师青睐。类似于 PIER, UpToDate 为临床医生提供推荐意见,方法严谨,采用统一的结构提出问题、较全面收集相关的循证医学文献(但不够全面系统)、采用 GRADE 分级评价证据质量和提出推荐意见。UpToDate 明确承认患者价值观和选择权在临床决策中的重要性。

UpToDate 覆盖了14个医学专业的7700个临床主题,包括约80 000页正文、图片,并与 Medline 摘要、260 000条参考文献和一个药物数据库链接,每4个月更新一次。

UpToDate 使用方便,但缺乏规范检索,需要付费。个人第1年订购费为450美元,以后为350美元,图书馆订购费为10 000美元。

## (二) 证据摘要

证据摘要是经过临床专家和方法学家对原始文献和系统评价进行严格评估后,对所收集整理的信息资源作出综合、简洁的描述,以摘要形式发表。

1. ACP Journal Club 包括 ACP Journal Club( <http://www.acpjc.org/> )、Evidence-Based Medicine 杂志和系列以 ACP Journal Club 为模版的杂志,以纸质或网络版发行。ACP Journal Club 先由工作人员从130种杂志中筛选出那些方法学严格、涉及临床问题、报告了重要临床结局指标的高质量原始研究和系统评价,再让临床医生从中选择对临床有重要价值和影响的文献,以结构摘要形式进行总结,并由一名临床专家评估文献的方法和提出临床应用的建议。约150篇文献中可筛选出1篇合格文献。

ACP Journal Club 由美国内科医师学会出版,主要针对内科及其亚专业,也包括少量其他专业如儿科,月刊,需付费,有助于临床医生了解内科领域新进展。

2. InfoPOEM( <http://www.infopoems.com/> ) 类似于 ACP Journal Club,有助于了解临床新进展,重点针对家庭医学。临床人员从100多种杂志中筛选出与家庭医生常见问题相关的文献进行评估和总结。

3. Bandolie( <http://www.medicine.ox.ac.uk/bandolier/> ) 是为英国国立卫生服务中心提供的证据,全国均可使用,选择的主题涉及各临床专业,评估的证据包括评论和推荐意见。

另外,美国医学研究院( New York Academy of Medicine, [www.nyam.org/fellows/ebhc/index.html](http://www.nyam.org/fellows/ebhc/index.html) )网站列出了包括 ACP Journal Club 在内的经过评估的信息资源库,可供链接查寻。

## (三) 系统评价

1. Cochrane 系统评价数据库( Cochrane Database of Systematic Reviews, CDSR ) 发表在 Cochrane 图书馆( <http://www.interscience.wiley.com/cgi-bin/mrwhome/106568753/HOME> ), Cochrane 系统评价是 Cochrane 协作网的评价员按照统一工作手册( Cochrane Handbook for Systematic Reviews of Interventions ),在相应 Cochrane 评价小组编辑部的指导和帮助下所完成的系统评价。由于 Cochrane 协作网有严密的组织管理和质量控制系统,严格遵循 Cochrane 系统评价者手册,采用固定格式和内容,统

一的系统评价软件(RevMan)录入和分析数据、撰写系统评价计划书和报告,发表后根据新的研究定期更新,有完善的反馈和修改机制,因此Cochrane系统评价的质量比收录在Medline和其他数据库的非Cochrane系统评价质量更高。目前主要针对疾病防治、康复疗效和安全性的随机对照试验进行评价,诊断性试验的系统评价刚起步。2009年第4期的Cochrane系统评价数据库包括4027篇全文系统评价和1906篇研究方案。Cochrane系统评价可从Ovid、PubMed、光盘和Wiley网站获取。CDSR检索方便,摘要免费,全文需要付费。

2. DARE(Database of Reviews of Effects, <http://www.york.ac.uk/inst/crd/crddatabases.htm#DARE>)为评价干预措施疗效的免费系统评价数据库,是经过评价的非Cochrane系统评价摘要,涉及主题广泛。DARE既是独立信息资源,同时也包含在Cochrane图书馆中。DARE每年收录约600篇,检索简单方便。DARE对无时间评价或不能获取系统评价全文的临床医生非常实用,因此也有人认为DARE是证据摘要信息资源。

3. 临床实践指南(clinical practice guideline, CPG)是证据强度很高的信息资源,有助于指导临床决策。美国国家指南数据库(US National Guidelines Clearinghouse Database, NGC, <http://www.guideline.gov/>)收集了美国 and 全世界数千个指南并提供了2200多个指南的摘要,涉及所有主题,指南制订按照循证医学原则和方法。NGC检索简单,可同时比较多个指南。

此外,在英国国立卫生图书馆(UK National Library for Health, <http://libraries.nelh.nhs.uk/guidelinesFinder/default.asp?page=INTER>)和安大略医学会网站(Ontario Medical Association, <http://www.gacguidelines.ca/>)查寻到指南,后者指南是经过严格评估的。

#### (四) 原始研究

原始研究非常多,通常只有在上述三种数据库资源中未能检索需要的文献时才检索原始研究数据库。

1. MEDLINE是卫生研究和医疗实践的首要数据库。传统的检索方式非常复杂,主要由图书管理员进行。检索MEDLINE的途径很多,但多数通过Ovid和PubMed。

2. PubMed Clinical Queries 通过PubMed(<http://www.ncbi.nlm.nih.gov/PubMed/>)可检索整个MEDLINE数据库,其Clinical Queries(<http://www.ncbi.nlm.nih.gov/entrez/query/static/clinical.shtml>)检索可直接获得与临床应用相关的文献资料。

3. EMBASE([http://www.elsevier.com/wps/find/bibliographicdatabasesdescription.cws\\_home/523328/description#description](http://www.elsevier.com/wps/find/bibliographicdatabasesdescription.cws_home/523328/description#description))是欧洲的大型医学文献数据库,与MEDLINE类似,但重点在药物和卫生领域。由于EMBASE数据库费用高,可及性差,临床医生少有采用。EMBASE中约70%条目未包括在MEDLINE。

4. Cochrane临床对照试验中心注册库(Cochrane Central Register of Controlled Trials, CENTRAL)是随机对照试验和半随机对照试验的数据库,该数据库由Cochrane协助网组织、协调和编制,采用计算机和手工检索相结合的方法,对期刊、会议论文集、MEDLINE和EMBASE及其他文献数据库收录的刊物进行检索,确定其中的随机对照试验和半随机对照试验,为进行系统评价提供系统、全面和准确的原始资料库。

5. 中国生物医学文献数据库(CBM)是中国医学科学院医学信息研究所开发研制的综合性医学文献数据库。该数据库收录了1978年以来1600多种中国生物医学期刊及汇编、会议论文的文献题录,收录范围涉及基础医学、临床医学、预防医学、药学、中医学、中药学等生物医学的各个领域。该数据库的研制注意了与MEDLINE光盘检索系统的兼容性,其检索过程中使用的运算符及功能与MEDLINE光盘相似。检索系统具有主题词表、中英文主题词轮排表、分类表、期刊表、索引词表、作者表等多种词表辅助检索功能,检索入口多,检索功能完备。

6. 中国期刊全文数据库是中国知识基础设施工程(China National Knowledge Infrastructure, CNKI, <http://www.cnki.net/index.htm>)中最重要的数据库。CNKI是目前世界上最大的连续动态更

新的中国期刊全文数据库,收录了自1994年以来(部分回溯到1979年和创刊年)国内8200多种重要期刊,内容覆盖自然科学、工程技术、农业、哲学、医学、人文社会科学等各个领域,全文文献总量2200多万篇。本数据库集题录、文摘、全文文献信息于一体,实现一站式文献信息检索;具有知识分类导航、众多检索入口等功能。

7. 中文生物医学期刊数据库(Chinese Medical Current Contents, CMCC) 是解放军医学图书馆研制开发的中文生物医学文献书目型数据库,也是目前检索国内生物医学文献最常用的光盘数据库之一。CMCC 数据库收录了1994年以来国内正式出版发行的生物医学期刊和一些自办发行的生物医学刊物1400余种的文题录和文摘。涉及的主要学科领域有基础医学、临床医学、预防医学、药学、医学生物学、中医学、中药学、医院管理及医学信息等生物医学的各个领域。CMCC 数据库的数据与CBM 的数据从1994年后大部分是相同的,因此检索1994年后的中文医学文献选用其中之一即可。但由于CMCC 的数据更新周期为2周,因此检索最新报道的生物医学文献应用CBM 数据库。

8. 中文科技期刊数据库(VIP) 由重庆维普资讯有限公司开发,是一种综合性文献数据库,收录了1989年以来8000余种中文科技期刊,涵盖自然科学、工程技术、农业科学、医药卫生、经济管理、教育科学和图书情报等七大专辑。该数据库在高校一般有镜像站点或远程访问路径。

## 第二节 证据检索和收集的基本步骤

检索和收集证据的基本步骤如图3-1所示。

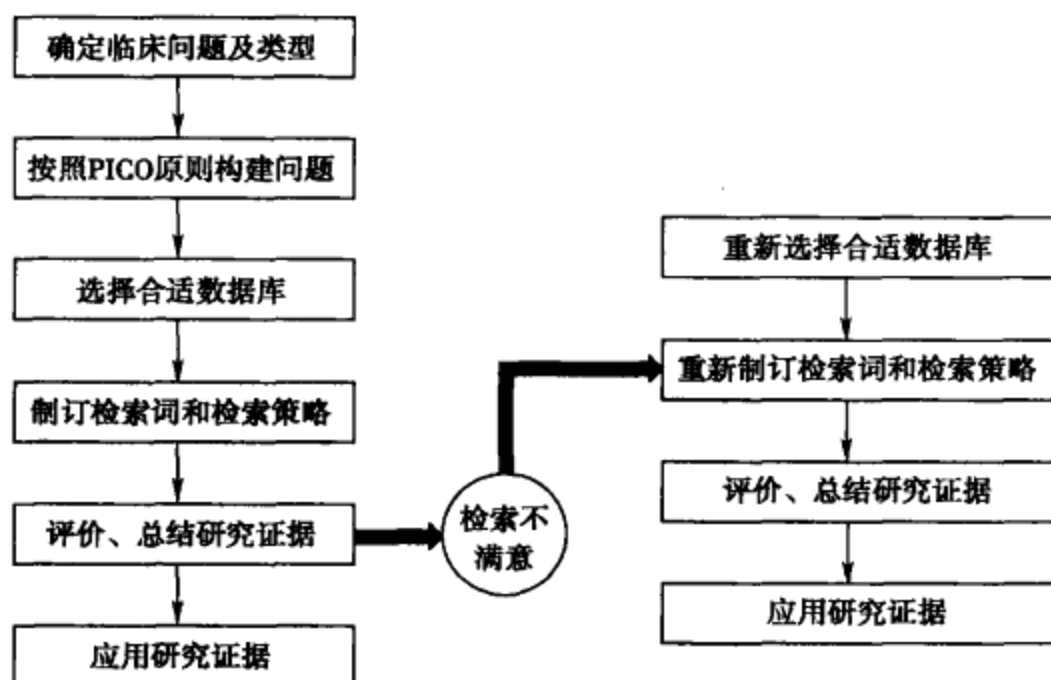


图3-1 医学文献的检索策略

### 一、确定临床问题类型和构建临床问题

常见的临床问题主要涉及疾病的病因、诊断、预防、治疗、预后及不良反应,针对每一问题,根据需要了解的信息不同又分为背景问题(“background” questions)和前景问题(“foreground” questions)。背景问题是关于某种疾病、某一诊断技术或某一干预措施的一般知识,如心房颤动是如何引起栓塞的?哪些人容易发生甲型流感?卡托普利(开博通)是如何引起干咳的?前景问题涉及临床诊断、治疗的具体知识。循证临床实践过程中,多数提出的问题均为前景问题,如某诊断试验能否准确诊断某疾病?某干预措施的疗效和安全性如何?回答前景问题,需要明确问题涉及的4个要素:研究对象(patient)、干预措施(intervention)、与干预措施比较的措施(comparison intervention)和

关心的结果(outcome),即根据 PICO 原则精心构建提出的前景问题。

明确临床问题类型及需要了解的信息、按照 PICO 原则构建临床问题,有助于正确选择数据库资源、合理选择检索词和制订检索策略。如背景问题我们可选择教材、专著和综述等,在 Cochrane 图书馆就不可能获得有关疾病病因、发病机制等方面的详细信息。而有关干预措施疗效, Cochrane 图书馆则是主要信息来源。

## 二、选择合适数据库

选择数据库,需要了解各数据库的特点、涉及专业范畴和具体临床问题类型。此外,尚需注意下列原则:

### (一) 选择计算机或网络数据库

针对临床问题,可通过四种途径获得答案:咨询同事或专家、查阅教科书或专著、手工查阅相关杂志和计算机检索相关文献数据库。咨询同事或专家是快速、有效获取信息的方法,但易受被咨询者知识更新程度的制约,常常存在一定偏倚;教科书或专著出版周期长,知识更新慢;手工查阅相关杂志能较快获得最新的信息,但非常花时间;而计算机检索克服了时空障碍,只要掌握基本的技巧,检索快速且效率高。

### (二) 尽可能选择专业数据库

综合性文献数据库如 Medline、中国生物医学文献数据库(CBM)等虽然覆盖了医学领域从基础到应用、各专业领域的资料,但有时却难以快速获得你真正需要的信息。因此,使用专业数据库更方便,易获得与专业直接相关的文献资料。但在缺乏专业数据库时,综合性文献数据仍然是最常用的信息源。

### (三) 尽可能选择最佳文献数据库

最佳证据资源(best-evidence resources)是指采用明确的方法,对研究证据的科学性和临床相关性进行严格评价后建立的数据库。由于医学文献的层出不穷,临床医生应改变传统的文献查寻方法和技巧,从针对某一问题检索所有相关文献改变为检索系统评价或 meta 分析和经专家评价过的研究证据数据库,以快速、高效获取最佳证据资源。根据“4S”模型,检索时应从证据系统(system)、证据摘要(synopses)、系统评价(syntheses)和原始研究(studies)逐级检索,原则上如果从上一级数据库检索的文献解决了提出的临床问题,则无需继续检索下一级数据库。

## 三、制订检索词和检索策略

### (一) 检索词

检索词的制订主要依据对提出的临床问题的分解,即 PICO 原则。通常检索词主要来源于 P(研究对象)和 I(干预措施),而较少采用 C(对照措施)和 O(结果指标)。当根据 P 和 I 检索结果太多时,可考虑通过 C 和 O 进行限定。

### (二) 检索策略

不同数据库,检索策略不全相同。检索策略的制订就是将检索词采用逻辑运算符“AND”、“OR”和“NOT”进行组合。

1. 扩大检索范围,提高查全率 当检索记录太少时,可以使用以下方式提高查全率:用主题词表(thesaurus)进行检索,如使用所选词的上位词进行检索、对主题词进行扩展检索、选用多个主题词检索、选用全部副主题或对副主题词进行扩展检索、选用词表提示的相关词或以前的检索词进行检索。

(1) 用自由词检索:如果一个需检索的概念由几个自由词组成一个语句,应选用最能表达该概念的最少的自由词进行检索,因为一个语句中自由词的数量与检出的文献量成反比。

(2) 用“OR”运算符:用“OR”运算符可选择新的检索词,也可把同等或同义的检索词叠加组合

起来进行检索,因此扩大了检索范围。

(3)用截词:对检索词的词根或词尾加上截词符“\*”进行扩展检索,可扩大检索词的范围,并防漏检。但用截词法耗费机时,而且容易产生假命中,应谨慎使用。

(4)用通配符检索:用通配符“?”加在检索词中进行检索,可以检索出拼法不同而意义相同或相近的词,从而扩大检索范围。

2. 缩小检索范围,提高查准率 如果检索出的文献太多,可以用以下方法来缩小检索范围:

(1)用主题词表进行检索:如选用主题词专指性不强,且该词下还有下位词,可选用下位词检索。

(2)选准副主题词进行检索。

(3)应用限定字段检索:限定字段检索时,常用的字段有TI、AU、AD、PY、CP、AB、MESH、MJME、TG、NM、PT等。如要提高通过主题词检索的准确性,可将检索结果限定在主要主题词字段进行检索。

(4)用运算符:常用缩小检索范围,提高查准率的运算符有AND、WITH、NEAR、NOT等。

完成检索策略后,针对选择的数据库进行检索。

#### 四、判断检索结果

获得检索结果后,应判断所获信息能否回答提出的临床问题。如果不能获得满意答案,应分析原因,是否数据库选择不当、是否检索词和检索策略制订不合理、还是确实该临床问题尚无相关研究证据。如果是从未经评价的数据库中检索的信息,尚需对检出的文献进行严格质量评价以确定其结果的真实性、临床重要性和适用性。

### 第三节 证据检索实例

**临床病案1:**—55岁男性患者,有2型糖尿病史18年,高血压病史12年,其血糖和血压水平一直控制良好。没有心肌梗死、心绞痛、脑血管意外和外周血管病史。最近复诊,血脂检查结果:总胆固醇5.2mmol/L, LDL-胆固醇3.0mmol/L, HDL-胆固醇1.6mmol/L,甘油三酯1.8mmol/L。你似曾记得,某次讲座中讲授者说:对于2型糖尿病患者,即使血脂水平不高,降脂治疗也可预防心血管疾病的发生。考虑到降脂治疗的长期性及他汀类降脂药可能的不良反应。因此,你认为,在向患者推荐此治疗前,需要明确有无研究证据支持这种说法。

#### (一) 提出临床问题

糖尿病患者会发生诸多并发症,例如心血管疾病、肾脏疾病、眼底病变和神经系统疾病等。糖尿病患者发生心血管疾病和因心血管疾病死亡的风险明显高于非糖尿病患者。糖尿病患者合并高血压以后,容易发生眼底、肾脏、心脏、下肢血管和脑血管疾病。因此,糖尿病合并高血压患者的降血脂治疗很重要,血脂的干预水平要严于非糖尿病患者。但是否有研究证据支持这一观点呢?

临床问题:血脂正常、无心脑血管病史的2型糖尿病患者,采用他汀类降脂药与安慰剂比较,能否预防心血管病的发生?

#### (二) 构建临床问题

按照PICO原则分解上述临床问题(表3-2),检出与上述临床问题直接相关的研究证据。

#### (三) 检索相关研究证据

不同医学文献资料在设计、实施、统计分析、结果解释和论文报告等方面存在着差异,研究质量、结果真实性和可靠性及适用性也不同。因此,检索证据时,建议首先检索经他人评估和筛选过的循证

表3-2 构建临床问题

P: 患者及问题	血脂正常、无心脑血管病史的2型糖尿病患者
I: 干预措施	采用他汀类降脂药
C: 比较措施	安慰剂
O: 结果	预防心血管病的发生

医学资源,如果未检索出需要的信息,再进一步检索未经筛选的数据库。

1. 选择数据库

(1) 首先检索经过评估或筛选的循证医学信息资源(二次文献数据库)

- Clinical evidence
- Best Evidence( Evidence-based Medicine and ACP Journal Club )
- Cochrane Library:Cochrane Database of Systematic Reviews( CDSR )
- Up To Date
- SUMSearch

(2) 再考虑检索未经评估或筛选的信息资源(原始文献数据库)

- PubMed
- EMBASE
- CBM

2. 确定检索词和检索策略

(1) 检索词: 根据构成临床问题的四要素,本病例检索可选择的检索词包括: type 2 diabetes、statin 和 cardiovascular disease。

(2) 检索策略: 上述临床问题采用检索词“type 2 diabetes”、“statin\*”和“cardiovascular disease\*”制订检索策略 [ type 2 diabetes ] AND [ statin\* ] AND [ cardiovascular disease\* ],并根据检索的数据库相应调整。

3. 检索相关数据库 我们检索了 Cochrane 图书馆,未发现1篇相关系统评价,只有1篇经济学评价的相关文献。再检索 PubMed,在“PubMed Services”下点击“Clinical queries”,输入检索词,在“Category”下点击“Therapy”,在“Scope”下点击“narrow, specific search”,获得38篇文献(1966年至2010年2月9日)。最相关的为 Colhoun 等的文章:“Primary prevention of cardiovascular disease with atorvastatin in type 2 diabetes in the Collaborative Atorvastatin Diabetes Study( CARDS ):multicentre randomised placebo-controlled trial. Lancet, 2004, 364:685-696”。

4. 判断检索结果 获得检索结果后首先应该判断该结果能否回答之前提出的临床问题,对未经评价的文献,还需要进行严格的质量评价。

针对上述问题检出的38篇文献,仔细阅读题目和摘要,结合具体情况,发现 Colhoun 等的文章“Primary prevention of cardiovascular disease with atorvastatin in type 2 diabetes in the Collaborative Atorvastatin Diabetes Study( CARDS ): multicentre randomised placebo-controlled trial. Lancet 2004;364:685-96”一文与提出的临床问题最相关。但此文献来自 PubMed,未经过质量评价,因此需要按照治疗性研究的质量评价原则分析此文献结果真实性、临床重要性、可靠性和适用性,以确定能否正确回答上述临床问题。

**临床病案2:** 一孕25周出生的早产儿,出生后30天胸骨左缘第2肋间闻及连续性杂音,脉搏增强。医生临床诊断为“动脉导管未闭(patent ductus arteriosus, PDA)”并已给予吲哚美辛(消炎痛)治疗 PDA1个疗程。婴儿非常虚弱且依赖呼吸机辅助呼吸,不便搬动进行超声心动图检查以确诊患儿是否有 PDA。那么此时,根据体格检查结果(体征)能否诊断患儿的 PDA 呢?

### (一) 提出临床问题

动脉导管未闭是早产儿最常见的先天性心脏病,若未及时诊断与处理,常可诱发或促进充血性心衰、慢性肺疾病、颅内出血和坏死性小肠结肠炎(NEC)等的发生发展。足月儿50%在生后24小时内,90%在48小时内,几乎100%在72小时内动脉导管(DA)关闭,在生后1~3个月常达到完全解剖上关闭。但若胎龄 $\leq 29$ 周,PDA发生率在0~24小时为80%,24~48小时为40%,48~72小时为7%。早产儿一旦确诊为持续性PDA,建议尽早治疗。药物(吲哚美辛等)和手术(PDA结扎)治疗PDA安全、有效。因此,应尽早确诊,减少并发症发生。

超声心动图诊断PDA最敏感准确,临床体征较迟发生,但对判断与PDA有关的远期疾病发生关系更密切。常见临床体征包括连续性杂音、心前区搏动增强、水冲脉、脉压增大或存在进行机械通气的指征。体征不同诊断意义也不同。作为标准诊断方法的超声心动图虽然准确,但该患儿病情危重、不便搬动,而床旁超声心动图机并不是任何医院均有。如果临床体征能作为筛查手段,将有助于早期诊断和治疗。尽管临床体征广泛用于PDA的初步诊断,其准确度到底如何呢?能否有助于诊断上述早产儿的PDA呢?

临床问题:对依赖呼吸机辅助呼吸的极低体重(体重 $<1000\text{g}$ )早产儿,临床体征诊断PDA的准确性如何?

### (二) 构建临床问题

按照PICO原则分解上述临床问题(表3-3),检出与上述临床问题直接相关的研究证据。

表3-3 构建临床问题

P: 患者及问题	极低体重早产儿
I: 干预措施	临床体征
C: 比较措施	超声心动图(金标准)
O: 结果	诊断动脉导管未闭

### (三) 检索相关研究证据

不同医学文献资料在设计、实施、统计分析、结果解释和论文报告等方面存在着差异,研究质量、结果真实性和可靠性及适用性也不同。因此,检索证据时,建议首先检索经他人评估和筛选过的循证医学资源,如果未检索出需要的信息,再进一步检索未经筛选的数据库。

1. 选择数据库 目前尚无专门针对诊断试验证据的数据库,只能通过综合性数据库检索诊断试验证据。

(1) 首先检索经过评估或筛选的循证医学信息资源(二次文献数据库)

- Best Evidence( Evidence-based Medicine and ACP Journal Club )
- Cochrane Library: Cochrane Database of Systematic Reviews( CDSR )
- Up To Date
- SUMSearch

(2) 再考虑检索未经评估或筛选的信息资源(原始文献数据库)

- PubMed
- EMBASE
- CBM

2. 确定检索词和检索策略

(1) 检索词: 根据构成临床问题的四要素,本病例检索可选择的检索词包括: patent arterial duct、diagnostic test、clinical examination、sensitivity、specificity、preterm。

(2) 检索策略: 上述临床问题采用检索词 patent arterial duct、sensitivity、preterm 制订检索策略 [ patent arterial duct ] AND [ sensitivity ] AND preterm, 并根据检索的数据库相应调整。



### 3. 检索相关数据库

首先检索二次文献数据库 Best Evidence 和 CDSR, 未检出相关文献。

再检索 PubMed (<http://www.ncbi.nlm.nih.gov/PubMed/>), 从页面左侧的“Clinical Queries”进入检索口“Search by Clinical Study Category”, 输入“[ patent arterial duct ] AND [ sensitivity ] AND preterm”, 在“category”下选择“diagnosis”, 在“scope”下选择“broad, sensitive search”, 检出33篇相关文献。

4. 判断检索结果 获得检索结果后首先应该判断该结果能否回答之前提出的临床问题, 对未经评价的文献, 还需要进行严格的质量评价。

针对上述问题检出的33篇文献, 仔细阅读题目和摘要, 结合具体情况, 发现“Davis P, Turner-Gomes S, Cunningham K, et al. Precision and accuracy of clinical and radiological signs in premature infants at risk of patent ductus arteriosus. Arch Pediatr Adolesc, 1995, 149( 10 ):1136-1141.”一文与提出的临床问题最相关。但此文献来自 PubMed, 未经过质量评价, 因此需要按照诊断试验的质量评价原则分析此文献结果的真实性、临床重要性、可靠性和适用性, 以确定能否正确回答上述临床问题。

(李 静)

## 第四章 证据评价的基本原则与方法

随着医学科学的飞速发展,每天均有许多医学论文发表,有许多新的研究证据产生。层出不穷而又良莠不齐的临床研究证据,只有经过严格评价,表明其具有真实性、临床重要性和适用性,才能应用于临床实践,对疾病的诊治产生积极的作用。

### 第一节 证据的分类与分级

研究证据的分类与分级是认知、评价、理解和应用证据的基础。

#### 一、证据的分类

证据分类方法众多,这里主要介绍与证据评价密切相关的两种分类方法,即以研究设计方案和以研究问题分类。

##### (一) 按研究设计方案分类

从方法学角度,可以将研究证据分为原始研究证据和二次研究证据。

原始研究证据是指直接以人群,即病人和(或)健康人为研究对象,对相关问题进行研究所获得的第一手数据,经统计学处理、分析、总结而形成的研究报告。常见的研究方法有随机对照试验、交叉试验、自身前后对照试验、同期非随机对照试验、队列研究、病例对照研究、横断面调查、病例分析和病例报告等。

二次研究证据是指在全面收集针对某一问题的所有原始研究证据的基础上,应用科学的标准,经严格评价、整合处理、分析总结而形成的研究报告。它是对原始研究证据进行二次加工后得到的更高层次的研究证据。但也必须指出其质量取决于原始研究证据的水平。常见的研究方法有系统评价、临床实践指南、临床证据手册、卫生技术评估等。

##### (二) 按研究问题分类

根据所研究问题的不同,研究证据可分为诊断、治疗、预后、病因、预防、临床经济学等研究证据。

#### 二、证据的分级

1998年, Bob Phillips、Chris Ball、David Sackett 等临床流行病学和循证医学专家共同制定了证据分级标准,2001年5月正式发表在英国牛津循证医学中心网站。该标准首次在证据分级的基础上提出了分类概念,涉及治疗、预防、病因、危害、预后、诊断、经济学分析等7个方面,更具针对性和适用性,已成为循证医学教学和循证临床实践中公认的经典标准(表4-1)。研究证据使用的推荐强度分为5级,即I级、II级、III级、IV级和V级。

此外,还有 grade 证据分级标准,参见本书其他章节。

表4-1 2001年牛津证据分级与推荐强度

推荐强度	证据级别	治疗、预防、病因研究	预后研究	诊断性研究	经济学分析
I级	I a	同质性随机对照试验的系统评价	同质性前瞻性队列研究的系统评价,或经验证的临床实践指南	同质性一流水平的诊断性研究的系统评价,或经验证的临床实践指南	同质性一流水平的经济学的系统评价
	I b	可信区间小的随机对照试验	随访率 $\geq 80\%$ 的前瞻性队列研究	纳入研究对象适当,且与金标准进行同步独立盲法比较的诊断性研究	采用适当的成本计算,对所有经过严格验证的备选医疗方案的结局进行了比较分析,包括将临床可观察到的变异结合到重要变量中进行敏感性分析
	I c	观察结果为“全或无”(某干预措施推行前某病病死率为100%,推行后低于100%,或推行前某病患者存在死亡或治疗失败现象,推行后无死亡或治疗失败)	观察结果为“全或无”的病例系列研究	绝对的特异度高即阳性者可确诊,或绝对的灵敏度高即阴性者可排除	对于干预措施分析后有明确结论:①成本低其结果佳的程度;②成本高其结果差的程度;③成本相同其结果的好坏程度
II级	II a	同质性队列研究的系统评价	同质性回顾性队列研究,或对照组未治疗的随机对照试验的同质性系统评价	同质性的但水平低于I级的诊断性研究的系统评价	同质性的但水平低于I级的经济学的系统评价
	II b	单个队列研究(包括低质量的随机对照试验,如随访率低于80%)	回顾性队列研究,或随机对照试验中未作治疗的对照组患者的追踪结果,或未经验证的临床实践指南	同步作了金标准及诊断试验,并进行了独立盲法比较,但研究对象局限且不连贯,或未经验证的临床实践指南	采用适当的成本计算,对若干备选医疗方案的结局进行了比较分析,包括将临床可观察的变异结合重要变量进行敏感性分析
	II c	结局性研究*	结局性研究*		
III级	III a	同质性病例对照研究的系统评价	—	—	—
	III b	单个病例对照研究	—	纳入研究对象适当且与金标准进行了独立盲法比较或客观比较,但部分内容对象未接受金标准试验的诊断性研究	未作准确成本计算的经济学研究,但包含在主要变量中加入临床因素作出的敏感性分析
IV级		系列病例观察(包括低质量的队列研究和病例对照研究)	系列病例观察(包括低质量的预后队列研究)	未采用盲法或未客观独立地使用金标准试验的诊断性研究,划分真阳性和真阴性的参考标准不统一的诊断性研究,纳入研究对象不适当的诊断性研究	无敏感性分析的经济学研究
V级		专家意见或基于生理、病理生理和基础研究的证据	专家意见或基于生理、病理生理和基础研究的证据	专家意见或基于生理、病理生理和基础研究的证据	专家意见或基于经济学理论的证据

注: \* 结局性研究是指描述、解释、预测某些干预或危险因素对最终结局的作用和影响的研究

## 第二节 证据评价的意义与基本要素

### 一、证据评价的意义

证据评价可以让极其繁忙的临床医生仅花费少量的宝贵时间,就可以从来源众多、良莠不齐的海量研究证据中查阅到所需要的信息,从而改进临床诊疗决策,提高医疗质量;证据评价可以为卫生行政部门决策者制定政策提供真实、可靠的依据;证据评价可以为病人选择医疗方案提供科学依据。

### 二、证据评价的基本要素

证据评价的基本要素为研究证据的内部真实性、临床重要性和适用性。

#### (一) 研究证据的内部真实性

能正确反映被研究的人群或靶人群真实状况的某一研究结果,对于它的正确(真实)程度称之为内部真实性(internal validity)。

影响内部真实性的主要因素为研究环境条件、研究对象的范围(类型的多少)以及研究设计的科学方法等。采取限制研究对象类型、严格的研究设计,消除或控制研究中有关偏倚与混杂因素的干扰,改善研究的环境条件和干预措施等手段可以改善内部真实性。

评价证据的内部真实性应重点关注该研究整体设计是否科学、研究方法是否合理、统计分析是否正确、研究结果是否支持研究结论等问题。

#### (二) 研究证据的临床重要性

研究证据的临床重要性是指其是否具有临床应用价值。循证医学强调采用客观量化指标来评价研究结果的临床意义。不同的临床研究问题其评价标准与指标不同。以评价治疗性研究证据为例,除需对每组各结局指标加以总结报告(如某结局的发生率或某观测指标的均数和标准差)外,还应报告干预措施的效果和效应值的精确度,如采用相对危险度减少率(RRR)、绝对危险度减少率(ARR)和获得一例有利结果需要防治的病例数(NNT)等客观指标,同时给出可信区间(CI)以表示估计值的精确度(详见本书治疗性研究的分析与评价专章)。

评价证据的临床重要性应重点关注证据所涉及临床问题是否明确具体、所选择的评价指标是否正确等问题。

#### (三) 研究证据的适用性

研究证据的适用性即外部真实性(external validity),是指研究结果与推论对象真实情况的符合程度,能够推广应用到研究对象以外人群的研究结果,称之为具有外部真实性。

研究证据的适用性,涉及最佳证据如何应用于循证医学实践的问题,而研究人群与其他人群的特征差异、研究对象类型以及社会环境和经济等因素将影响证据的适用性。为此,要高度重视证据的外部真实性。

评价证据的外部真实性应重点关注证据所涉及研究对象的异质性及其与拟应用对象(病人)在人口社会学特征和临床特征上差异性、拟应用对象所处环境是否具备产生证据环境所具备的人力、技术和设施设备条件等问题。

## 第三节 证据评价的基本内容与方法

### 一、证据评价的基本内容

评价研究证据时,应对研究工作的全过程进行全面的评价。

#### (一) 研究目的(假说)

是否以问题为基础确定研究目的;研究目的或假说是否明确具体,并清晰陈述;所研究的问题是

否具有临床重要性;研究假说是否具有科学性、先进性和可行性。

### (二) 研究设计

各种研究设计方案都有一定的优缺点与适用性。是否根据不同的研究问题和各种研究设计方案的科学性和可行性选择合适的设计方案;所选择的研究设计方案是否优于以往相类似或相同问题的研究设计。

### (三) 研究对象

目标人群定义是否明确;研究对象有无公认的或金诊断标准以及恰当的纳入标准与排除标准;样本的代表性如何;样本量是否足够;研究对象分组是否保证了组间均衡可比。

### (四) 观察或测量

研究变量有无明确的定义;是否采用客观指标;变量的测量方法是否恰当;结局变量是否明确、有无准确定义,是中间性指标还是结局性指标;测量指标的判断标准和临床意义要明确;测量结局变量的方法是否准确;是否采用盲法收集资料。

### (五) 结果分析

是否根据研究设计方案和资料的性质选择合适的统计分析方法;计算是否正确;研究中可能出现的误差、混杂和交互作用是否进行分析;统计推理是否恰当。

### (六) 质量控制

研究全过程可能出现的主要偏倚有哪些;是否采取了相应的控制措施;所采取的偏倚控制措施效果如何。

### (七) 结果表达

研究中观察效力有多大;研究结果的表达是否观点清晰,数据准确;是否有量效或剂量反应关系的证据;核心结果的表达是否标准化;如为阴性结果,统计学把握度是否足够。

### (八) 卫生经济学

对于干预措施是否采用成本-效果分析、成本-效益分析、成本-效用分析等评价经济效益和社会效益。

### (九) 研究结论

研究结论是否回答了研究假说;研究发现与实验室研究所得作用模式是否一致;研究所获结果在生物学上是否有可靠依据;研究发现与同类研究结果是否一致;研究结论是否可以外推;研究发现是否肯定引起现行临床实践的某种改变。

最后,评价者应全面总结以上各方面的评价结果,提出改进研究或如何使用该证据的建设性意见。

## 二、证据评价的基本方法

评价证据应先初筛临床研究证据的真实性和相关性、再确定研究证据的类型、最后根据研究证据类型按相关标准进行评价。

### (一) 初筛临床研究证据的真实性和相关性

1. 初步判定研究证据的真实性 以该研究证据是否来自经同行评审(peer-reviewed)杂志、产生证据的机构是否与自己所在的机构相似、该证据是否由某个组织所倡议且其研究设计或结果是否因此受影响等为参考指标,对研究证据的真实性进行初步的判断。

2. 初步判定研究证据的相关性 以下列3项指标为参照,对研究证据的相关性进行初步的判断。

- ①若该研究证据提供的信息是真实的,是否为自己的患者所关心的问题及对其健康有无直接影响;
- ②该研究证据是否为临床实践中常见问题,其涉及的干预措施或试验方法在自己所在机构是否可行;
- ③若该研究证据提供的信息是真实的,是否将改变现有的医疗实践。

### (二) 确定研究证据的类型

不同的临床问题,最适合的研究设计方案不同;不同的研究设计方案其技术要领和研究功效亦

不同,因此,评价研究证据前应根据其所研究的问题和所采用的研究设计方案准确判定其类型。

### (三) 根据研究证据类型进行评价

研究证据的评价应遵循临床流行病学/循证医学的原则与方法,并根据其分类属性采用相应评价标准进行科学评价。

## 三、证据评价的注意事项

为了确保对证据作出客观、全面的科学评价,评价证据时还应注意以下事项。

### (一) 方法学评价是基础

正确的研究设计方案是获得真实可靠的研究结果的根本保证,因此,方法学评价是证据评价的基础。

### (二) 证据的真实性是评价重点

证据的真实性是其生命,是作为是否采用该证据的基本依据,不具备真实性的证据是毫无价值的,因此,评价研究证据的真实性应作为重点。

### (三) 要选择恰当的评价指标

各研究设计方案分别有相应评价标准或指标。选择评价指标是否恰当,直接影响评价的结果,因此,应根据研究设计类型选择恰当的评价标准或指标。

### (四) 评价要力求全面系统

评价研究证据时应对该研究的全过程,包括选题、设计、测量、分析、结果解释等逐项逐条进行评价,并完整报告评价所获得的全部结果,包括其优点和局限性。

### (五) 评价要实事求是

研究证据来自于对病人或人群的观察或试验,由于无法严格控制各种研究条件,误差(偏倚和随机误差)、混杂只能控制而无法消除。因此,评价研究证据时务必实事求是,本着科学精神善于发现其优点、使用其有利的部分,同时客观估价其缺陷,以利于临床循证医学实践中取舍。

### (六) 正确认识阴性结果

研究者都希望获得肯定的、有效的阳性结果。同时,阳性结果文章比阴性结果文章容易发表,而且有较多机会发表在高影响因子的期刊上,会有更高的引用率。其实,否定一项无效甚至有害的干预措施,其贡献不亚于肯定一项确实有效的干预措施,只要设计科学、测量严谨、分析客观、结论正确,阴性结果同样有意义。因此,在针对某一临床问题的研究证据进行评价时,应注意不要遗漏阴性结果的证据。

(许能锋)

## 第五章 循证医学实践中常用的统计学方法

临床实践的特殊性决定了用于指导循证临床实践的“证据”，应为真实可靠、重要且又具有实用价值的最佳证据(best evidence)，而“最佳证据”的判定，则依赖于一系列原则和标准，这其中就包括统计学方法。

最佳证据来源于大量的临床研究结果，可以是单个临床研究结果，也可以是多个临床研究结果的系统评价(或 meta 分析)。尽管形式多种多样，但无论是何种形式的研究结果，都有其相应研究目的、假设、设计方案、终点效应等，而由此产生各种类型的数据资料，主要集中于 PICO 类指标的测量，如研究对象特征指标(population/patients, P)，干预或暴露测量指标(intervention/exposure, 简写 I 或 E)，对照(comparison, C)以及结果(局)指标(outcome, O)等，这些就构成了临床统计学分析与效果评价的物质基础。

根据以往临床流行病学和循证医学的教学经验，从临床研究的“创证”到循证医学的临床“用证”，均涉及统计学的正确应用问题，应用不当，会影响研究结果的真实性和可靠性。因而“最佳证据”的判定，应首先从临床研究所获得的资料、数据及其所得出的分析结论等入手，评估其真假及确定精确度的允许范围。为此，实践循证医学除了应具有临床专业知识、经验和技能之外，还应具备分析和评价临床研究证据所需的医学统计学知识和技能。

本章将从循证医学实践的角度，对涉及统计学的知识与方法应用做些阐述，以利于对证据质量的综合分析评价。有关统计学的具体公式及理论推导，可参阅医学统计学专著的相关内容。

### 第一节 临床证据的数据资料类型

临床研究及实践所用证据的数据类型，其表达形式多样，分别从不同的角度描述了各种临床现象的特征和本质，按照临床证据的数据资料基本特征，可大致归类如下：

#### 一、分类变量资料

所谓分类变量(categorical variable)，就是以个体为基本观察单位，按照研究对象的某种属性分为两类或者多个类别。若只有两个互不相容的分类，则称之为二分类变量资料，如性别分为男性与女性，疗效分为有效和无效，并发症可分为有或无等；若分类超过两个，称为多分类变量资料。按照类别间是否存在等级上的差别，又可细分为有序多分类资料(等级变量资料)以及无序多分类变量资料。如病情按照严重程度可分为轻、中、重，有等级上差异，则属于前一分类，而诸如种族、ABO 血型等指标，分类间无等级之分，则属于后者。由此可见，这类数据资料定性描述了研究对象的某种属性，本身并不是精确定量的，只是按照所隶属的类别进行计数，所以，这种资料又被称为计数资料或者离散变量资料。

分类变量资料在临床研究的数据资料中占了相当大的比例，其表达大多直接采用率和比的形式。例如，在治疗性研究中，描述试验组和对照组中事件发生率的大小，可用有效率、痊愈率、病死率、复发率等。在此基础上，又可进一步衍生出相对危险度(率比)、绝对危险度(率差)、比值比(OR)等二级指标。

又如,对于诊断试验的诊断价值大小描述,通常用敏感度、特异度以及相应的似然比等表示。

## 二、数值变量资料

所谓数值变量(numerical variable)就是临床研究中能被准确测量的各种指标,所测得数值是可以“度、量、衡”的,因此又称之为计量资料或连续性变量资料。例如临床研究中诸如体温、身高、体重、血压、血细胞计数、血脂、血糖等指标。这类资料的数值大小及其分布都有一定规律,如有的服从正态分布,有些则呈偏态分布。分布类型不同也决定了统计描述形式的不同。若资料本身服从正态或近似正态分布,则选用均数±标准差描述该组数据资料的集中趋势以及离散趋势;反之,要么通过变量变换,使其满足正态分布的条件,要么直接采用中位数(median)和四分位间距(interval of quartile range, IQR)描述。如果研究涉及不同组别间有关数值变量的相互比较,则以均数差值为效应量,应用相应的统计分析方法,进一步判断差别是否具有统计学及临床意义。

数值变量资料较之分类变量资料更能准确描述事物的变化本质,清晰反映出从量变到质变的界值及不确定的范围及程度,可为疾病早期正确诊断或防治提供较为准确可靠的信息。

## 三、等级变量资料

如前所述,在临床研究中,若多分类变量资料的类别间有程度或等级之分,称为等级变量资料。这类资料比较特殊,既可以在临床研究中预先设置,也可以结合临床实际由数值变量资料转化而来。例如:按照收缩压与舒张压水平可将高血压分为轻、中、重三类,其中1级高血压范围(SBP:140~159mmHg, DBP:90~99mmHg),2级高血压(SBP:160~179mmHg, DBP:100~109mmHg),3级高血压(SBP≥180mmHg或DBP≥110mmHg);又如高脂血症诊断,也可结合血清胆固醇和甘油三酯的不同水平,将其分为三类:合适范围(血清总胆固醇<5.20mmol/L,血清甘油三酯<1.7mmol/L),边缘高(5.23~5.69mmol/L),高血脂(血清总胆固醇>5.72mmol/L,血清甘油三酯>1.7mmol/L)。

从上例可以看出,将数值变量资料转化为等级资料,对临床病情诊断、治疗及预后的分析评价等有着重要的实用价值。如当一个个体的血糖值处于参考值范围的上限,或当某一患者的舒张期血压达到89~90mmHg以上时,则可为早期诊断提供可能。然而,对这种转换一定要慎重。这是因为当数值变量转化为等级变量资料时,毕竟会不同程度地失去量化的统计信息,对真实性自然有所影响。需在创证及循证用证时予以足够重视。

在病因/危险因素分析、治疗或预后的研究证据中,涉及资料类型往往不止一种,既有分类变量资料,也有一些相关的数值变量资料。因此,在用证与评价证据时,应具体结合变量赋值范围及其意义,全面评估研究结果的实际价值。

## 第二节 证据资料的方法学质量评价

当在循证医学实践中拟采用某一证据时,无论该证据是采用何种最佳设计方案,甚至被誉为“最佳证据”,亦不能盲从,应结合具体临床研究类型,按照病因学研究、诊断试验、临床干预性研究以及预后研究的评价原则,对其进行严格评价。其中,分析证据的基础数据资料是否可靠以及方法学质量的高低等是非常重要的一个环节。这里分别就单个研究证据与多个研究的系统评价的方法学质量评价,阐述如下。

### 一、单个研究证据的质量评价

在循证医学实践中获取的证据往往以单个研究成果居多,对其质量评价应围绕以下几个方面进行:



### (一) 资料完整性的判断

资料的完整性是指纳入分析的研究对象数量以及重要指标数应前后保持一致,不能有遗漏或进行选择分析。特别是应比较临床研究中不同阶段的研究对象数量是否前后一致,如在设计之初与统计分析阶段,研究对象数量应尽量保持相同。仍以临床试验为例,若试验前后例数一致,则说明病例资料完整;若试验前后例数不等,则应计算丢失率:丢失率=(入组例数-终末例数)/入组例数 $\times$ 100%。当丢失率 $<10\%$ 时,质量基本合格;当丢失率在 $10\%\sim 20\%$ 之间,甚至 $>20\%$ 时,则表明资料质量较差,这样结果往往会偏离真实性。

此外,还应核对设计阶段设置的重要结局指标与最终纳入分析的指标,两者是否一致,不能任意删减结局指标或有选择性加以分析,否则也会人为造成偏倚。

### (二) 组间的基线资料是否可比

在单个临床研究中,通过设置对照组增加可比性,但前提条件是除研究因素外,组间基线资料应尽量做到组间均衡。例如,试验和对照两组的重要临床基线资料应该相对一致,特别是例数、性别和年龄构成、病情严重程度、病程及其他主要影响预后因素等的组间差异应无统计学意义,这样的结果可信;否则应考虑进一步作分层分析、多因素分析等。

### (三) 重复性检验

对于临床研究中的一定观测指标,有可能在测量过程中产生偏倚,需要进行重复性检验。例如,对于临床影像学观测指标,如X线片、CT、MRI或者病理资料,除了有明确的诊断标准外,医生的观测诊断水平不同可能造成某种差异。因此,当应用此类证据于临床实践时,应检查证据的原始资料是否做过一致性检验,可信度有多大?若作了一致性检验,其Kappa值是多少?若Kappa值 $<0.4$ ,说明质差;当Kappa值 $>0.7$ 及以上,则表明质佳。

另外,对来自实验室的数据资料,还要分析实验室的操作条件、试剂是否标准化?应用试剂的批内及批间差异度有多大?是否在容许范围之内等。为此,可分别计算批内和批间差异度。其中,批内差异度=(第一次测量值-第二次测量值)/第一次测量值,批间差异度=(第一批测量值-第二批测量值)/第一批测量值。通常重复性检验中这两类差异度均应控制在 $5\%$ 以内。

### (四) 缺失值分析及其处理

在临床研究中,即使设计严格、实施规范,所收集的数据资料中,总会或多或少出现一些缺失值。缺失值(missing value)是指因种种原因不能得到观测指标的具体测量值,出现数据缺失。评判数据缺失对临床研究结果的影响程度,应视缺失其属性及比例而定。缺失主要分两种,一种称为随机性缺失,即缺失值的出现与组别、干预措施等无关,缺失一般无固定规律、是随机产生的。如临床试验中试验组与对照组均可能出现缺失值,若缺失比例相近,说明缺失与临床干预措施无关,缺失比例只要不超过 $10\%$ ,对结果影响不大。虽然随机性缺失不可避免,但可以采取一系列质量控制措施,使缺失比例控制在允许范围之内;另一种则称为非随机性缺失,这种缺失常有规律可循,要么集中出现在特定组别,要么出现于某些特别个体。假如药物的毒副作用过大,势必会造成病人的大量失访,直接表现在试验组与对照组的缺失比例出现差异,会对结果产生较大的影响。这种缺失由于与组别、干预措施等有关,在临床研究中应尽量避免,特别是在设计阶段,应考虑好相应处置措施。对于小样本临床研究,缺失值对结果的影响尤为明显,对缺失值应正确处理。

1. 分类变量资料的数据缺失及处理 一般通过敏感性分析,观察结果的稳健性,继以考察缺失值对结果的影响程度。例如,临床试验中的试验组和对照组各丢失了10例,那么可以把试验组丢失的10例,当作“无效病例”,而对照组丢失的10例,则当作“有效病例”处理,然后再作差别比较,倘若两组的率差仍有临床及统计学意义,则说明结果可靠,缺失对其影响不大;反之,若结论截然相反,说明结果对缺失值敏感,对相应结果应审慎解释。此外,还可考虑意向性分析。

2. 数值变量资料的数据缺失及处理 证据中的数值变量资料表达往往以均数 $\pm$ 标准差形式出现,有时还可得到均数差值及其 $95\%$ 可信区间(95% confidence interval, 95%CI)等方面信息,对

判断缺失值的可能影响很有帮助。假如95%可信区间的间隔小、精度高,则意味着数值资料可靠性高,结果较为可信;反之则精度差,考虑是否是由于原始数据“丢失”过多或样本数量较少所致。若数值变量资料中出现了随机性缺失值,且缺失比例在允许范围之内,如小于10%,缺失值可用该指标的平均值替代。

#### (五) 样本量的分析

如果设计科学合理,样本量大的临床研究较之小样本的临床研究,其数据资料自然更为真实可靠。然而,临床研究的样本量也不是越大越好。这里强调在被引证资料的设计中,其样本量要适当,应进行样本量估算,特别是重点考察最小效应量以及I、II型错误的水平。如在一个临床治疗性试验中,新药组假设有效率为80%,而对照组的疗效为60%,则试验组比对照组临床效果好20%,I型( $\alpha$ )错误率限制在 $<0.05$ ,II型( $\beta$ )错误率限制在 $<0.1$ ,这时可应用相应公式重新计算样本量,借以判断每组的样本例数是否合适。若遇到样本量小于40例的资料,特别是当 $P>0.05$ (无统计学意义),而效应量却又有一定的临床价值时,应计算 $\beta$ 错误率,如果其 $>0.2$ ,则表明样本量不足,扩大样本再试是必要的,而不宜简单得出结果无统计学意义的结论。

相反,若样本量很大时,即使是组间差异很小,甚至无临床价值或实际意义,也有可能出现 $P<\alpha$ 水准、差别有统计学意义的情况,这样的结果同样不可取。

#### (六) 资料的分层分析

有些大型(样本)研究或许整体上两组基线不一或效果差异不显,此时可考虑进一步作分层分析,可能会发现其中某种(些)有意义的重要差异。分层分析有助于提高统计分析质量。

## 二、多个研究证据的质量评价

循证医学实践过程中,有时会同时得到多个类似的原始研究,其结果可能一致,也可能不一致,甚至出现截然相反的结论,究竟选取或相信哪一个结果,临床医生往往陷入两难的境地。为此,应积极寻找有无针对多个研究结果的系统评价/meta分析。系统评价是指针对某一具体的临床问题全面系统地收集现有相关原始研究,用统一的科学评价标准,去粗取精,去伪存真,从中筛选出符合标准、高质量的研究,然后用定性或定量的方法进行综合,从而得出可靠的结论。当新的临床研究结果出现时,系统评价还要及时进行结果更新。而meta分析则是指系统评价中的一个定量综合分析过程,是对多个医学研究结果的数量化汇总分析。

自上个世纪末开始,基于原始研究的系统评价/meta分析数量日益增多,但存在着质量参差不齐的问题,如何对这些系统评价/meta分析的质量展开评价,一直是一个难点问题。系统评价的质量评价,可分为报告质量评价以及方法学质量评价两种。报告质量实际上反映了系统评价报告内容的完整性和全面性以及与方法学标准格式的吻合程度;而方法学质量是指系统评价及其过程中能否遵循科学标准、有效控制混杂与偏倚、使结果达到真实可靠的程度。一个系统评价的报告质量高,能为读者提供充分的信息资料,然而有很好的报告质量,并不意味着其方法学质量也高,其结果就真实可信;但倘若报告质量差、报告内容不完整,则又会增大方法学质量评价的难度,甚至无法评价。因此,系统评价的质量评价应兼顾两者,进行综合分析。鉴于质量评价带有很强的主观性,为保证评价的质量,可以借助一些现成的评价工具,特别是那些已经过信度效度反复验证的量表,如OQAQ量表(the Oxman-Guyatt Overview Quality Assessment Questionnaire)、AMSTAR、QUOROM量表(the Quality of Reporting of Meta-analysis, QUOROM)等。其中OQAQ主要用于评估方法学质量(表5-1),OQAQ量表中条目的设置主要围绕系统评价的基本步骤来进行,从选题立题、文献检索与筛选、原始文献质量的严格评价以及汇总分析等诸多方面,评估避免或减少混杂与偏倚的程度。若一篇系统评价能严格遵循科学原则,最大程度地减少偏倚和误差的出现,则该系统评价的方法学质量就高。至于汇总分析是否正确,取决于异质性检验结果及选择模型是否合适。QUOROM是测试系统评价的报告质量,并不适合作方法学质量分析。

表5-1 OQAQ表(系统评价之质量评估量表)

条目	解释	答案
1. 围绕研究问题是否采用了检索方式查找原始文献?	若标明来源类别,包括年份、具体文献源,则回答为“是”;若只提及采用手检、机检方法,回答为“部分”	1. 是;2. 部分;3. 否
2. 文献检索是否全面系统?	回答“是”者,要求检索至少三类以上文献库(源),其中至少包括计算机检索、手检,同时报告关键词与主题词	1. 是;2. 不能确定;3. 否
3. 是否报告了文献的纳入标准?	若至少报告了两条标准,回答为“是”;少于两条者为“部分”,未提及者则为“否”	1. 是;2. 部分;3. 否
4. 可能出现的文献选择偏倚是否被成功避免?	若至少两位评价者独立完成了筛选评价,且协商一致者,回答为“是”	1. 是;2. 不能确定;3. 否
5. 是否报告了纳入文献的质量评价标准?	不包括发表性偏倚的相关内容,若事先设置了真实性评价方法(如将随机、双盲、安慰剂对照或者分配方案隐匿作为文献纳入标准),则回答为“是”	1. 是;2. 部分;3. 否
6. 是否采用合理方法评价了纳入文献的真实性?	若有任何关于文献筛选或文献分析标准(内部真实性或外部真实性)的描述,则回答为“是”	1. 是;2. 不能确定;3. 否
7. 是否报告了文献结果的汇总分析方法?	具体见第三节	1. 是;2. 部分;3. 否
8. 针对研究问题,汇总分析是否合理?	未进行合并分析或合并结果不合理,则回答为“否”;若只报告了汇总结果,未提及分析过程,仍标记为“否”	1. 是;2. 不能确定;3. 否
9. 最终结论是否有数据或分析结果支持?	研究问题的主要结论若有数据支持,则回答为“是”	1. 是;2. 部分;3. 否
10. 对该系统评价的总体质量评分是多少?	基于上述9个问题的答案,对系统评价的质量进行综合评分。若一个系统评价有1个以上条目回答为“不能确定”,则说明该系统评价只存在“小的缺陷”;若条目2、4、6、8中回答为“否”者,该说明系统评价存在“大的缺陷”甚至更严重	具有明显缺陷为1分,大的缺陷为3分,小的缺陷为5分,缺陷可以忽略者分值最高,为7分

### 第三节 统计学方法的正确抉择

无论是单个研究还是多个研究的系统评价,均涉及一系列的统计学方法,不同类型、不同条件下的资料分析所选用统计学方法有所不同。若方法选用不当,会直接影响结论的真实可靠。本节主要从统计学方法基本原则与要求出发,明确证据中所选用方法是否合理,至于具体的公式和运算程序则

需参考有关医学统计学专著。

### 一、单个研究及其统计方法的正确抉择

无论何种类型的单个研究,其统计分析包括两方面内容,即统计描述与统计推断,而推断又包括假设检验与可信区间估计。判断单个研究的统计方法选用是否合理,应首先明确是否满足了相关统计方法的基本条件。例如数值变量资料的不同分布类型,正态分布,抑或偏态分布,其描述方式不尽相同:①属于正态分布的数值变量资料,直接选用均数  $\pm$  标准差;②属于对数正态分布者,则用几何均数  $\pm$  标准差;③不服从正态分布或对数正态分布者,则选用中位数及四分位间距。

#### (一) 证据统计描述的基本要求

1. 关于数值变量类证据的统计描述,列于表5-2及表5-3。

对于证据为数值变量资料,其集中趋势的描述,可选用均数、中位数、几何均数、众数等。具体应用哪一个,应考虑资料本身是否满足相应适用条件。

表5-2 数值变量资料的集中趋势指标

指标	作用	适用条件
均数	描述一组资料的平均水平或集中趋势	正态或近似正态分布
中位数	描述一组资料的平均水平或集中趋势	偏态分布或分布未知或两端无界限
几何均数	描述一组资料的平均水平或集中趋势	对数正态分布、等比资料

对数值变量类证据的离散趋势描述,可选用标准差、极差、四分位间距、变异系数等。其中,变异系数为标准差及其均数的比值。

表5-3 数值变量资料的离散趋势指标

指标	作用	适用条件
标准差	描述一组资料离散的程度	正态及近似正态分布
四分位数间距	描述一组资料离散的程度	偏态分布或分布未知或两端无界限
极差	描述一组资料离散的程度	偏态分布或分布未知或两端无界限
变异系数	描述一组资料离散的程度	正态及近似正态分布

无论是集中趋势还是离散趋势,其描述指标选择应结合数据资料的具体分布类型。然而,大量的研究表明,有相当一部分数据资料并不符合正态或近似正态分布,其统计描述仍错误选用均数  $\pm$  标准差,应用此类证据时,应慎重对待。

2. 关于分类变量资料的统计描述,列于表5-4。

表5-4 分类变量资料的常用描述指标

指标	表达方式	意义
事件发生率	事件发生例数 / 观测总例数	分析事件发生的强度和频率
构成比	单类事件发生例数 / 多类事件的例数总和	总事件数中各类事件所占比重
相对比	甲事件发生率与乙事件发生率比值	发生甲事件与乙事件相比的倍数值

对于分类变量类证据,其统计描述常用率和比。例如,循证医学实践常用事件发生率(event rate),如发病率、患病率、病死率、致残率、有效率等,用这些率表示事件发生的强度和频率;也可利用构成比表达事件发生的相对比重,如出血性脑卒中在脑卒中病人中所占的百分比;相对比的应用则

更为丰富,如相对危险度(relative risk, RR)、比值比(odds ratio, OR)等。此外,由率及构成比等可进一步衍生出一些重要临床指标,如某一事件率(死亡、有效、副效应)与另一组相应事件发生率相比而产生的绝对危险降低率(absolute risk reduction, ARR)、相对危险降低率(relative risk reduction, RRR)以及NNT等。NNT是与对照组比较,新的措施需要处理多少例数才能防止一例不良事件的发生(number needed to treat, NNT)。有关这些指标的具体计算与意义,参见本教材的相关章节,这里就不再重复阐述。

## (二) 可信区间与精确度分析

除了评价证据的统计描述是否正确外,还应该考察和分析各种证据估计的精确度(precision),这在循证医学实践中是十分重要的。

分析证据精确度主要是借助效应量及其可信区间(confidence interval, CI)或可信限(confidence limit, CL)。可信区间是按一定的概率( $1-\alpha$ )去估计总体参数所在的范围,包括准确度和精度两种属性。其中,准确度是指区间内包含总体参数的可能性,如总体均数的95%可信区间,准确度为95%,意味着在该区间范围内有95%的可能性包含总体均数,或者说从总体中做100次随机抽样,得到100个可信区间,那么理论上95个可信区间包含被估计的总体均数;精度是指可信区间的宽度,宽度越小,则精度越高。精度与样本量和准确度有关,在样本量固定的情况下,准确度越高,精度越差,99%可信区间较95%可信区间的精度差,反之亦然,因此,多数统计分析软件包常选用95%可信区间作为默认值。当然也可根据实际需要,选用90%或99%可信区间。此外,增加样本量可提高精度。

可信区间具有统计推断的功能,且与假设检验相比,可信区间能提供更多的信息。如两组样本均数比较的数值变量资料,若两组均数差值的95%可信区间不包括0,说明两总体均数差别有统计学意义,反之,无统计学意义。同时可信区间还能显示差别的程度,并由此可判断出差别程度有无实际价值或临床意义,但可信区间无法提供确切概率。

1. 数值变量类证据的精确度估计 数值变量资料的精确度估计需要样本均数、标准差、样本量等要素。如在一项临床试验研究中,为探讨传统康复治疗对脑卒中患者预后的改善效果,将39例患者纳入研究,半年后测量功能恢复评分为(35.6±28.11)分,其精确度可借助公式( $\bar{X} - t_{\alpha, v} S/\sqrt{n}$ ,  $\bar{X} + t_{\alpha, v} S/\sqrt{n}$ )估计。本例95%可信区间:35.6±(2.024×28.11/√39),下限为35.6-9.1=26.5(分),上限为35.6+9.1=44.7(分)。

2. 二分类变量类证据的精确度估计 此类证据的精确度估计需要事件发生率、样本量等信息,按照组数不同选用相应的公式。

### (1) 单组率的95%可信区间

单组率的可信区间估计公式:  $p \pm u_{\alpha} \sqrt{\frac{p \times (1-p)}{n}}$ , 其中  $p$  为样本率,  $n$  为样本含量。例如用某降压药治疗60例高血压患者,其中24例有效,有效率为40%。此例,  $p=0.4$ ,  $n=60$ , 则有效率  $p$  的95%可信区间为:  $0.4 \pm 1.96 \times \sqrt{\frac{0.4 \times (1-0.4)}{60}}$ , 其下限为0.276, 上限为0.524。

### (2) 两组率比较的95%可信区间

两组率的差值为绝对危险降低率(absolute risk reduction, ARR),  $ARR=p_2-p_1$ , 其可信区间为:  $ARR \pm 1.96 \times SE_{ARR}$ , 其中  $SE_{ARR} = \sqrt{\frac{p_1 \times (1-p_1)}{n_1} + \frac{p_2 \times (1-p_2)}{n_2}}$ 。如在某临床试验中,甲组治疗125例,病死率为12%,乙组治疗120例,病死率为25%,如表5-5。则  $ARR=p_2-p_1=0.25-0.12=0.13$ , ARR标准误为  $\sqrt{\frac{0.12 \times 0.88}{125} + \frac{0.25 \times 0.75}{120}} = 0.049$ , 因而95%CI为  $0.13 \pm 1.96 \times 0.049$ , 下限为3.4%, 上限为22.6%。

表5-5 临床试验两组的病死率分析

组别	结果		合计	病死率
	死亡	存活		
甲组	15(a)	110(b)	125( $n_1$ )	12%
乙组	30(c)	90(d)	120( $n_2$ )	25%

(3) 如果计算上例的相对危险度(relative risk, RR)及95%CI,过程如下:

$$RR=p_1/p_2, \text{服从对数正态分布,对数转换为 } \log(RR), \text{则相应的标准误 } SE_{\log RR} = \sqrt{\frac{1}{a} + \frac{1}{c} - \frac{1}{n_1} - \frac{1}{n_2}}。$$

本例  $RR=0.12/0.25=0.48$ , 则  $\log(RR)$  的95%CI  $= -0.734 \pm 1.96 \times \sqrt{\frac{1}{15} + \frac{1}{30} - \frac{1}{125} - \frac{1}{120}} = -1.31 \sim -0.167$ ,

经反对数转换,最终 RR 的95%CI 为0.272~0.846。

RR 用于病因/危险因素研究表示因果关联强度时,若暴露组发病率高于非暴露组,则  $RR>1$ ; 若暴露组与非暴露组发病率一致,则  $RR=1$ ; 如果暴露组发病率低于非暴露组,则  $RR<1$ , 这意味着暴露因素反而是有益的“保护因素”。进一步考核95%可信区间,若不包括1,则表示结果有统计学意义; 而 RR 的95%CI 包括“1”时,即表示结果无统计学意义。对于95%CI 的实际意义与统计学意义的综合判断见表5-6。

表5-6 相对危险度(RR)及其95%CI判断

RR	95% CI	结果
$RR>1$	上下限均 $>1$	危险因素且有统计学意义
$RR=1$	包括1	无实际价值及统计学意义
$0<RR<1$	上下限均 $<1$	保护因素且有统计学意义

(4) NNT 及其95%可信区间: NNT 是指与对照组相比,需要治疗多少病例才能获得1例最佳结果(number needed to treat, NNT),其95%可信区间不能直接计算,可利用 ARR 及其可信区间上、下限的倒数估算而来。本例  $NNT=1/ARR=1/0.13=7.7$ , 其95%可信区间上限为  $1/0.034=29.4$ , 下限为  $1/0.226=4.4$ 。

(5) 相对危险降低率 RRR 及其95%CI: 不能直接计算,它用1减去相对危险度的95%CI 值推算出,以上例:  $RRR=1-RR=1-p_1/p_2=1-0.12/0.25=0.52$ , RRR 的95%可信区间为0.154~0.728。

列出以上(1)~(5)的计算实例是因为在循证医学应用证据过程中,有时会缺乏95%CI 资料,这些计算方法可供借鉴。值得指出的是,有些证据特别是 RCT 研究证据,当其具有临床价值,但又无统计学意义,特别是  $P>0.05$  或更大时,需借助95%CI 进行综合判断。

### (三) 有关统计学假设检验方法的正确抉择

假设检验方法的选用与数据资料的性质、质量、分布特点及统计检验目的等有关。因此,当引用某一证据时,可通过下列简明方法评价假设检验是否合理。

1. 若证据为数值变量资料,应重点考核分析目的、设计类型、样本量以及是否满足正态性与方差齐性(表5-7),以明确统计学方法选用是否合理。

表5-7 数值变量资料比较的常用假设检验方法

分析目的	应用条件	统计方法
单个样本与已知总体均数比较	例数( $n$ )较小,样本来自正态总体	$t$ 检验
	例数( $n$ )>50或者>100以上;或例数较小,但总体标准差已知	$u$ 检验
两组成组资料比较(完全随机设计)	例数( $n$ )>50或者>100以上	$u$ 检验
	若例数( $n$ )较小,来自正态且方差齐性总体	成组设计的 $t$ 检验
	否则选用	成组设计的秩和检验
两配对资料比较(配对设计)	配对差值服从正态分布	配对设计的 $t$ 检验
	否则选用	配对设计的秩和检验
多组资料的比较(完全随机设计)	各组均来自正态总体且方差齐性	成组设计的方差分析
	否则选用	成组设计的秩和检验
配伍资料的比较(配伍设计)	各组均来自正态总体且方差齐性	配伍设计的方差分析
	否则选用	配伍设计的秩和检验

在临床研究的证据中,为消除某种(些)混杂因素的影响,常发现提供的资料是配对设计的。例如试验组和对照组的研究对象分别按照性别、年龄、病损程度进行1:1或2:1不等配对,这种形式为异体配对;另外一种形式是自体配对,如自身前后对照试验,分别测量同一个体干预前与干预后的一些观测指标。对此,应按照配对统计方法处理,取其差值进行统计分析。如治疗前SBP测量值为162mmHg,治疗后为132mmHg,其配对的前后差值为30mmHg,像这种配对资料,其价值远优于非配对的、成组设计的变量资料,真实性更好。因此,所采用的统计分析方法也有别于成组资料。

2. 若证据为二分类变量资料,同样需观察分析目的与应用条件(表5-8),判断方法的选择是否合理。

表5-8 分类变量资料比较的常见假设检验方法

分析目的	应用条件	统计方法
两组率或构成比的比较(成组设计)	$n \times p > 5$ 且 $n \times (1-p) > 5$	二项分布 $u$ 检验
	$n > 40$ 且最小 $T > 5$	四格表 $\chi^2$ 检验
	$n > 40$ 且 $1 < T < 5$	校正四格表 $\chi^2$ 检验
	$n < 40$ 或 $T < 1$	确切概率法
配对资料比较(配对设计)	$b+c > 40$	McNemar 检验
	$b+c < 40$	校正 McNemar 检验
多组率或比资料比较(成组设计)	少于1/5格子的 $1 < T < 5$	行 $\times$ 列表 $\chi^2$ 检验
	若有 $T < 1$ 或有多于1/5的格子 $1 < T < 5$	确切概率法

注:  $n$  为样本例数;  $p$  为阳性事件发生率;  $T$  为理论频数

这些假设检验方法的应用条件大多与样本含量和设计方案有关。配对设计与成组设计,所选用的方法是不同的。

(四) 多因素分析及相关要求

疾病从其发生、发展、干预到其最终的转归,是十分复杂的病理生理过程,受生物、心理、社会环境及经济等诸多因素的综合影响。因此,为了获取影响疾病发生、干预效果、预后等相关因素的证据,采用多因素分析是必要的。在将此类证据应用于临床实践之前,也应对有关多因素统计分析方法的原则和方法有所了解。

在常用的多因素分析方法中,需要事先定义应变量和自变量,构建一般表达方式为:  $Y = X_1 + X_2 + X_3 + X_4 + X_5 + \dots + X_p$ ,其中  $Y$  为应变量,常为结局指标。例如冠心病的发病与否,就可作为应

变量,用 $Y$ 表示,若发生冠心病,则定义 $Y=1$ ;若未发生,则 $Y$ 赋值为0。那么哪些因素存在与冠心病的发病有关呢?经较长期的研究发现冠心病的危险因素有高血压、糖尿病、高脂血症、低密度脂蛋白胆固醇增高、血纤维蛋白原含量增高等,这些影响因素称为自变量,分别用 $X_1, X_2, X_3 \dots$ 等表示。这些因素测量尺度与水平以及对导致冠心病发病的影响力大小不一,可在作单因素分析的基础上,进而对可能有意义的(正负效应)因素,作有关多因素分析。

多因素分析应注意样本量是否足够。为尽可能减少机遇因素的影响,通常要求样本量应为纳入分析的自变量个数的5~10倍;此外,还应注意相关自变量的赋值与标准化问题。分类变量及数值变量都可作为自变量纳入,但在分析前应按要求转换为标准数据,建立相应的数据库。这里就临床研究中最常用的几种多因素分析方法及适用范围分别描述如下:

1. 多元线性回归分析 多元线性回归分析用回归方程定量地刻画一个数值变量的应变量( $Y$ )与多个自变量( $X_1, X_2, X_3, \dots, X_p$ )间线性依存关系的大小与方向。用“ $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \dots + \beta_p X_p + e$ ”表示。其中 $\beta_0$ 为回归方程的常数项,也称截距,意义同简单线性回归,为 $Y$ 的基线水平量。 $\beta_1, \beta_2, \beta_3, \dots, \beta_p$ 为偏回归系数。如 $\beta_p$ 是指在 $X_p$ 以外的其他变量固定条件下, $X_p$ 每改变一个单位后 $Y$ 的平均变化量。 $e$ 为除去 $p$ 个自变量对 $Y$ 影响后的随机误差,也称残差。

应用多元线性回归,要考虑资料是否满足以下条件:①应变量(结果变量)要求是数值变量资料,且满足随机性与独立性,自变量可以是数值变量资料、分类变量资料、等级资料;②应变量与自变量间具有线性关系;③残差 $e$ 服从正态分布等。

2. logistic 回归模型 logistic 回归模型是研究分类变量结果与一些影响因素之间的一种多元(变量)统计方法。特点为结果变量为分类变量资料,自变量可为分类变量,也可以是连续性变量。logistic 回归模型常用于:①控制一个或多个混杂因素的情况下,探讨某个事件的发生与研究因素的关系;②探讨各种影响因素间的交互作用;③筛选危险因素;④预测事件发生。

3. COX 风险比例回归模型 生存分析是将结果事件发生及发生时间进行综合分析的一种统计分析方法。其中多因素生存分析方法首推COX风险比例回归模型。该模型应变量(结果变量)为生存时间,自变量为分类变量资料或者数值变量资料。模型一般表达为 $H_{(t)} = H_{0(t)} \exp(\beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \dots + \beta_m X_m)$ 。其中 $H_{(t)}$ 为风险函数(风险率或瞬间死亡率), $H_{0(t)}$ 为基准风险函数,是与时间有关的任意分布函数,分布与形状无明确的假定。 $X_i$ 表示与生存可能有关的影响因素, $\beta_i$ 为回归系数,由实际数据估计得到,要求在观察时间内是恒定的、不随时间变化。 $\beta_i > 0$ ,表示 $X_i$ 值越大,病人死亡的风险越大; $\beta_i < 0$ ,说明 $X_i$ 值越大,病人死亡的风险越小; $\beta_i = 0$ ,则 $X_i$ 值与病人死亡的风险无关。

这三种多元回归模型形式相似,不同之处在于应变量( $Y$ 变量)的资料类型,若 $Y$ 为数值变量资料,考虑使用多元线性回归分析;若 $Y$ 为分类变量,特别是二分类变量资料,考虑使用 Logistic 回归模型;若 $Y$ 为时间变量,则优先考虑COX风险比例模型。这三种多元分析模型中,对自变量未进行特别规定,既可为数值变量资料,也可以是分类变量资料。需要特别注意的是当自变量为无序多分类资料时,分析前要进行哑变量处理,以利于结果的正确分析与解释。

多因素分析所获得结果还应计算回归系数的95%可信区间,用以判断结果的精确程度。由于其计算过程一般较为复杂,需要借助专门的统计分析软件。

### (五) 分层分析与亚组分析

有时在临床研究中,为观察某个新治疗措施的效果,直接将试验组与对照组进行整体分析,也许不具备统计学意义,读者有可能否定其价值。然而若按病情程度等重新进行分层分析或亚组分析,则可能发现有重要临床与统计学意义的结果。

例如对颅脑血肿患者施以A、B两种不同手术方式,并比较临床疗效。A组与B组各有490例,尽管A术式从临床意义上要优于B术式,然而假设检验发现 $P=0.08$ ,按照 $\alpha=0.05$ 水准,整体比较并无统计学意义。当按颅内CT片所定量的出血灶大小(小量、中量和大量出血)分层,仍以死残率作为终点指标,再次进行分层分析后,会发现小量与中量出血灶组A式与B式手术的死残率差异



无统计学意义,而在大量出血组 A 术式则显著优于 B 术式,进一步依据出血量分层的 Logistic 分析结果发现, A 组死残率仅为 B 组的 0.7 倍; 出血量每增一个等级, 则死残危险度平均为原等级的 1.9 倍, 且具有临床与统计学双重意义 ( $P < 0.05$ )。可见, 对于一些临床研究证据, 分层分析有时对判断最佳证据质量也是有所帮助的。

## 二、多个研究及其统计方法的正确抉择

单个临床研究由于样本量往往有限, 难免会受到机遇因素的干扰和影响。如果出现多个类似的临床研究, 假设干预措施相同、设计方案又一致时, 就可将这些临床研究集中起来, 在严格评价的基础上, 进行综合量化分析, 从而获得更为精确可靠的量化结论, 这种方法就称为系统评价 (systematic review) 或 meta 分析 (meta-analysis), 在本教材中另有专章论述。

这里要强调的是任何系统评价 / meta 分析, 一定要有明确的解决某个临床问题之目的, 要有严格的设计和计划, 纳入的文献一定是高质量的研究成果, 千万不要采用低质量的证据来做这一工作, 否则, 会以讹传讹, 造成误导。此外, 进行这项工作的实施者除了需要掌握系统评价的方法外, 还应具备良好的临床专业知识、经验和临床思维方法, 这对正确认识与应用有关系统评价的证据十分关键。与单个研究的要求一样, 基于多个研究的系统评价, 其统计方法也要做到正确抉择。

### (一) 描述指标的选择

1. 二分类变量资料 结果描述可以选用比值比 (OR)、相对危险度 (RR) 等相对指标, 也可选择绝对危险度 (RD)、NNT 等绝对指标。这些结果描述指标各有优缺点, 其中 OR 使用广泛, 在样本分布及模型拟合上有一定统计优势, 但 OR 有可能被曲解, 且与其他统计量相比, 稳定性差。RR 与 OR 同为相对测量值, 当结果事件罕见时, 两者数值接近, 常被用于估计合并效应量; 而 RD、NNT 为绝对测量值, 反映基线危险度以及干预后危险度的改变量, 能提供更多的信息, 与临床关系密切, 可用于直观描述某种卫生保健服务的效果。但由于其可信区间可能随基线危险度变化而变化, 不宜用于合并效应量的可信区间估计。此外, NNT 通过 RD 计算得到, 常与时间因素有关, 只有当所纳入研究的随访时间均相同时, 才适合做合并分析。因此, 结果描述与汇总分析所用的统计量, 可以不同。当然, 理想的结果表达最好是相对指标 (如 OR、RR、RRR) 和绝对指标 (RD、NNT) 同时报告。

2. 数值变量资料 (连续性变量资料) 目前数值变量资料的结果表达仍采用均数  $\pm$  标准差形式。应用此类证据时, 要警惕可能出现的偏态数据及潜在的影响。数据资料是否呈偏态分布, 最简单的判断方法就是计算均数与标准差的比值, 若该比值小于 1.64 及以下时, 说明标准差过大, 该组数据可认定为正偏态。合并效应量可以选均数差值和标准均数差值。选均数差值 (mean difference, MD) 的最大好处是合并结果有自然单位, 易于理解和解释。当结果变量所采用的尺度或度量衡单位不一致时, 其合并效应量表达宜采用标准化均数差值 (standardized mean difference, SMD), 但应慎重解释此类结果。

3. 个体病人资料 若能直接获得个体病人资料, 可与统计师联系, 重新分析这些原始数据, 并根据具体数据类型, 选用适当方法加以描述。

### (二) 系统评价中汇总分析方法的抉择

尽管有很多汇总分析方法可供选择, 但选用哪一种模型以及如何处理异质性等方面, 仍存在一些争议。目前较为一致的看法是在决定是否进行汇总分析以及采用哪种模型, 应综合权衡以下几个方面的内容:

首先, 应结合异质性检验结果和效应量的分布假设, 选择随机效应模型或固定效应模型实施汇总分析。其中异质性分析尤为关键, 当存在较为明显的临床异质性时, 最好不要进行合并分析, 应设法弄清异质性的来源, 如干预因素、研究对象、结果测试指标以及研究质量等方面是否存在不同, 必要时考虑进行亚组分析和敏感性分析。

其次, 利用固定效应模型检验合并效应量的假设是否成立, 其结果是稳健的。合并效应量的假设

检验若有统计学意义,则表明至少其中一个原始研究的效应量是有意义的。同时应注意无论异质性是否存在,利用固定效应模型估计的结果只是所纳入研究效应量的平均值。

相对于固定效应模型,随机效应模型是假设研究的效应量不固定,但服从某种分布,一般假定为正态分布。研究间效应量的变异大小可用组间方差加以测量,并以此为权重,对效应量进行校正估计。即较小样本量的研究在合并分析中给予较大权重,较大样本量研究结果所占的权重适当减小;然而由于小样本研究的质量普遍较差,且易受发表性偏倚影响,因此,应慎重选择随机效应模型。在实际应用中,可同时采用两类模型分别计算结果,与固定效应模型结果相比,随机效应模型的估计结果更保守一些(即可信区间较宽)。若无异质性,两个模型的合并分析结果应该一致;当异质性检验有统计学意义且假设研究间效应量不固定、但服从正态分布时,应选择随机效应模型的估计结果,倘若异质性过大,应进行亚组分析或 meta 回归分析。

另外相对于不同模型,均有多种估计算法,如何操作可以向统计师寻求帮助。尽管固定效应模型的估计方法多,但估计结果间差别一般不会太大;而随机效应模型的估计方法较少,目前以基于方差倒置法和 MH 法的 DerSimonian-Laired 校正为主(表5-9)。

表5-9 常用统计方法一览表

数据资料类型	合并统计量	模型	方法
二分类变量	OR	固定效应模型	方差倒置法、Peto 法、Mantel-Haenszel 法
		随机效应模型	DerSimonian-Laired 法
	RR	固定效应模型	Mantel-Haenszel 法、方差倒置法
		随机效应模型	DerSimonian-Laired 法
	RD	固定效应模型	Mantel-Haenszel 法、方差倒置法
		随机效应模型	DerSimonian-Laired 法
数值变量资料	均数差值(MD)	固定效应模型	方差倒置法
		随机效应模型	DerSimonian-Laired 法
	标准均数差值(SMD)	固定效应模型	方差倒置法
		随机效应模型	DerSimonian-Laired 法
个体病人资料(IPD)	OR	固定效应模型	Peto 法

#### 第四节 证据的临床意义与统计学意义

无论是单个研究结果还是系统评价证据,在应用前,应综合判断其统计学意义与实际价值。假如证据有重要临床价值,则看其统计推断结果(包括假设检验及效应量的可信区间估计)。注意假设检验的结论是概率性的,有可能犯 I 型或 II 型错误。若  $\alpha$  限定为小于 0.05,则表明该试验犯 I 型错误的概率在 5% 以下,真阳性率则 >95%;若试验设计时限定 II 型错误(即假阴性错误)  $\beta < 0.1$ ,则表明所获得结论若是阴性的,其中 90% 以上是可信的,亦即检验效能(power 或称把握度) >90%。在临床实践中,往往可以遇到有些证据既有临床价值,又具统计学意义,结果自然可信;反之,无价值。有时尽管临床证据有重要意义,然而无统计学意义,这时要慎下结论;反之,如无临床价值,却有显著统计学意义,其证据的临床价值却不大,见表5-10。

表5-10 证据的统计学意义与临床意义

情况	统计学意义	临床意义	评价结论
I	有	有	可取
II	有	无	临床价值不大
III	无	有	? 或者扩大样本量继续验证
IV	无	无	不可取

因此,假设检验是否有统计学意义,并不能说明其有临床价值。绝对不能仅仅依据  $P$  值的大小就武断说某种证据的有用或者无用,否则就可能出现误判。因此,一定要将临床意义和统计学意义进行综合分析,才能得出合理的结论。

(康德英 王家良)

## 第六章 医学文献的系统评价与 meta 分析

随着医学科学日新月异的发展、研究新领域的日趋开拓、人们认识的不断深化,临床医学研究层出不穷,为决策者提供了大量科学信息。医务人员和研究者为了获得新知识、新观点和新技术以扩展视野,提高学术水平,需要阅读大量文献。但现有临床研究多数规模较小,纳入研究对象数量有限,针对同一种疾病的同一或同类干预措施的文献资料有时数量较多,质量良莠不齐,结论也不尽一致。如何从浩如烟海的医学文献信息中快速、高效率地获得所需资料,以进行科学决策,已成为我们面临的巨大挑战。这就需要有一种方法、一种标准将全面、系统收集起来的文献,去粗取精、去伪存真、综合分析,从而得出一个真实性及可信性较高的结论,以供读者参考应用,这就称为文献综述;如针对某一具体临床问题,采用一套规范、科学的方法全面收集、认真选择、严格评价和科学分析相关研究资料,得出综合可靠的结论,此乃系统评价(systematic review)。

临床医生工作繁忙,要想为解决一个临床问题,去查遍相关文献并进行综合评价,难度很大,可是,假如面对一个实际问题需要您科学回答,这就是挑战:如一个妇产科医生,面对有早产危险的孕妇,而早产婴儿往往有可能胎内发育不全,特别是因肺发育未成熟,死亡率和呼吸窘迫综合征的发生率都非常高。有专家提出,如果对可能早产的孕妇使用激素,可促进胎儿的肺发育。为了明确对可能早产孕妇使用激素后是否能减少早产儿的死亡率和呼吸窘迫综合征的发生率,你查寻了有关资料,发现有7个高质量的随机对照试验,其中5个试验结果为阴性(使用激素后未能减少早产儿的死亡率和呼吸窘迫综合征的发生率),2个试验结果为阳性,你将作何决策呢?科学回答这个问题就需要对这七篇文献进行系统评价,以助自己的科学决策。

因此,本章将重点阐述系统评价及 meta 分析的方法及相关的质量分析和评价原则,以供读者回答这一挑战的参考和应用。

### 第一节 系统评价概述

#### 一、基本概念

##### (一) 系统评价

系统评价是一种全新的文献综合方法,指针对某一具体临床问题(如疾病的病因、诊断、治疗、预后),系统、全面地收集现有已发表或未发表的临床研究,采用临床流行病学严格评价文献的原则和方法,筛选出符合质量标准的文献,进行定性或定量合成(荟萃分析, meta-analysis),得出可靠的综合结论。系统评价可以是定性的(定性系统评价, qualitative systematic review),也可以是定量的(定量系统评价, quantitative systematic review),即包含 meta 分析过程,系统评价的整个过程非常明确,使其具有独特的优点:即良好的重复性。

系统评价可为某一领域和专业提供大量的新信息和新知识,多数是可信的。但是,由于是对原始文献的二次综合分析和评价,受原始文献的质量、系统评价的方法及评价者本人的专业知识、认识水平和观点的制约,因此,读者在阅读系统评价的观点和结论时,一定要持谨慎的态度,不能盲目被动地接受。

## (二) Cochrane系统评价

Cochrane 系统评价是 Cochrane 协作网的评价人员按照统一工作手册(Cochrane reviewers' handbook),在相应 Cochrane 评价小组编辑部的指导和帮助下所完成的系统评价。由于 Cochrane 协作网有严密的组织管理和质量控制系统,严格遵循 Cochrane 系统评价者手册,采用固定的格式和内容要求,统一的系统评价软件(RevMan)录入和分析数据、撰写系统评价计划书和报告,发表后根据新的研究定期更新,有着完善的反馈和完善机制,因此 Cochrane 系统评价的质量通常比非 Cochrane 系统评价质量更高,被认为是单一的、评价干预措施疗效的最好信息资源(best single source)。

目前, Cochrane 系统评价主要针对研究疾病防治、康复疗效和安全性的随机对照试验进行评价,其方法较完善和规范,其他诸如诊断性试验的系统评价已开始进行。

## (三) Meta-分析 (meta-analysis)

Meta-分析由心理学家 Glass 1976 年首次命名,国内翻译为荟萃分析、汇总分析。就其定义目前仍然存在不同的争议。Huque 及多数专家认为“meta 分析是一种统计分析方法,它将多个独立的、可以合成的临床研究综合起来进行定量分析。”因此,如果没有明确、科学的方法去收集、选择、评价临床研究资料,而仅单纯采用统计方法将多个临床研究进行合成并不能保证结论的真实性和可靠性。

目前系统评价与 meta 分析两个名词常被混用,但系统评价不一定都包括有 meta 分析过程,而 meta 分析也不一定是系统评价。

## (四) 文献综述 (review)

文献综述又称为叙述性文献综述(narrative review)或传统文献综述(traditional review),由作者根据特定的目的和需要或兴趣,围绕某一题目收集相关的医学文献,采用定性分析的方法,对论文的研究目的、方法、结果、结论和观点等进行分析 and 评价,结合自己的观点和临床经验进行阐述和评论,总结成文,可为某一领域或专业提供大量的新知识和新信息,以便读者较短时间了解某一专题的研究概况和发展方向,解决临床实践中遇到的问题。但这种传统的文献综述,往往受综述者的主观思维以及某些选择及测量性偏倚的影响。故在接受或应用这类证据时宜持批判态度。

## 二、为什么要进行系统评价

作为一种重要的科研方法,进行系统评价(systematic review),有其一定的必然性。

### (一) 应对信息时代的挑战

每年约有200万篇生物医学文献发表在2万多种生物医学杂志上,年增长率约为6.7%。一个内科医师需要每天不间断地阅读19篇专业文献才能勉强掌握本学科的新进展、新研究结果,使得需要大量信息进行科学决策的临床医生、研究人员和卫生部门的决策者往往陷入难以驾驭的信息海洋之中。而系统评价采用严格的选择、评价方法,去粗取精、去伪存真,将真实、可靠而有临床应用价值的信息进行合成,可直接为各层次的决策者提供科学依据。

### (二) 及时转化和应用研究成果

由于疾病谱的变化,对多因素疾病如恶性肿瘤、心脑血管疾病和各种慢性疾病的治疗方法的评估,需要尽量开展大样本临床试验,特别是随机对照试验(RCT)。但实施大规模的 RCT 需要消耗大量的人力、财力和时间,往往超过一个单位的承受能力,可行性受一定的限制。而现有的临床研究虽然数量多,但多数样本量不够大,故单个试验的结果难以提供较为全面、准确和推广应用价值大的研究成果。

因此,将多个质量较高的同质临床试验结果应用系统评价方法进行合成,则可将其综合的有效措施,转化和应用于临床实践与决策。如采用累积性 meta 分析回顾性分析有关静脉链激酶治疗急性心肌梗死的临床试验,1973 年有 8 个 RCTs(2432 例患者)的 meta 分析即证明静脉链激酶能有效降低 AMI 患者的总死亡率( $P=0.01$ );1978 年达 25 个 RCTs(34 542 例患者)的 meta 分析, $P=0.001$ (包括 GISSI-1 和 ISIS-2),到 1986 年, $P=0.0001$ ,1987 年才在传统综述和教科书中推荐常规使用静脉链激酶

治疗急性心肌梗死。可见,临床应用比 meta 分析的结果整整晚了14年,这期间如早用,则可挽救多少急性心肌梗死患者的生命。所以, Murphy 等在1994年指出:1973年以后的大型临床试验,如果无医德问题也是多余的,且浪费大量经费。

### (三) 提高统计效能

针对同一临床问题的研究非常多,但因疾病诊断标准、纳入研究对象的标准、测量结果方法、治疗措施和研究设计等的差异,结果可能不一致,甚至相互矛盾。如对可能早产的孕妇使用激素的例子,纳入的7个高质量临床试验,尽管只有两个试验结果有统计学意义,但对7个临床试验进行定量系统评价,借以增加样本含量和统计效能后,总的结果却有统计学意义,即肯定糖皮质激素能有效降低早产儿的病死率。系统评价或 meta 分析在进行资料合成时,不是根据阴性或阳性研究的个数多少决定哪种治疗措施有效,而是充分考虑了各个研究的样本量大小和研究的质量。

另外,系统评价可减少有关偏倚的影响,提高研究结果的可靠性和准确性。

## 三、系统评价与叙述性文献综述的区别与联系

系统评价和叙述性文献综述均是对临床研究文献的分析和总结,目前多为回顾性、观察性的研究,也可为前瞻性系统评价。回顾性的系统评价受临床研究质量的制约,因而易受系统偏倚、随机误差的影响。因此,确定一篇综述为叙述性文献综述,还是系统评价以及其质量、价值如何,主要取决于是否采用科学的方法减少了偏倚、混杂因素的影响。

叙述性文献综述常常涉及某一问题的多个方面如糖尿病的病理、病理生理、流行病学、诊断及预防、治疗、康复的措施,也可仅涉及某一方面的问题如诊断、治疗等。系统评价或 meta 分析多是集中研究某一具体临床问题的某一方面如糖尿病的治疗或康复,具有相当的深度。因此,叙述性文献综述有助于广泛了解某一疾病的全貌,而系统评价则有助于深入了解某一具体疾病的诊治。两者的主要区别如下(表6-1)。

表6-1 叙述性文献综述与系统评价的区别

特征	叙述性文献综述	系统评价
研究的问题	涉及的范畴常较广泛	常集中于某一临床问题
原始文献来源	常未说明、不全面	明确,常为多渠道
检索方法	常未说明	有明确的检索策略
原始文献的选择	常未说明、有潜在偏倚	有明确的选择标准
原始文献的评价	评价方法不统一	有严格的评价方法
结果的合成	多采用定性方法	多采用定量方法
结论的推断	有时遵循研究依据	多遵循研究依据
结果的更新	未定期更新	定期根据新试验进行更新

## 第二节 系统评价的方法

系统评价一方面能够通过多个有争议甚至相互矛盾的小型临床研究采用严格、系统的方法进行评价、分析和合成,以解决纷争或提出建议,为临床实践、医疗决策和今后的研究起正确导向作用;另一方面,如果纳入原始研究质量不高或进行系统评价/meta分析的方法不恰当,均可影响研究结果,以至于产生不正确的信息,造成误导。因此,系统评价的方法和步骤的正确与否,对其结果和结论的真实性、可靠性起着决定性作用。

系统评价本身只不过是一种研究方法,并不仅限于随机对照试验或仅对治疗措施的疗效进行系统评价。根据研究的临床问题不同,可对病因、诊断、治疗、预后、卫生经济评价和定性研究(qualitative

research) 等方面进行系统评价; 根据系统评价纳入的原始研究类型不同(study design), 可分为临床对照试验(controlled trial) 和观察性研究(observational study) 的系统评价, 前者如随机对照试验和非随机对照试验的系统评价, 后者如队列研究和病例-对照研究的系统评价; 根据进行系统评价时纳入原始研究的方式, 又可分为前瞻性、回顾性和累积性系统评价; 根据资料分析时是否采用统计学方法(meta 分析), 还可为定性和定量的系统评价。目前, 基于随机对照试验评估治疗措施疗效的系统评价较多, 在理论和方法学上较完善且论证强度较高。

针对不同研究问题的系统评价, 其基本方法和步骤相似(图6-1), 但在检索文献的策略、评价文献质量的方法、原始文献中数据的提取以及统计分析等方面有差别, 本节将以评价治疗措施疗效的Cochrane 系统评价为例, 简述其基本方法和步骤。

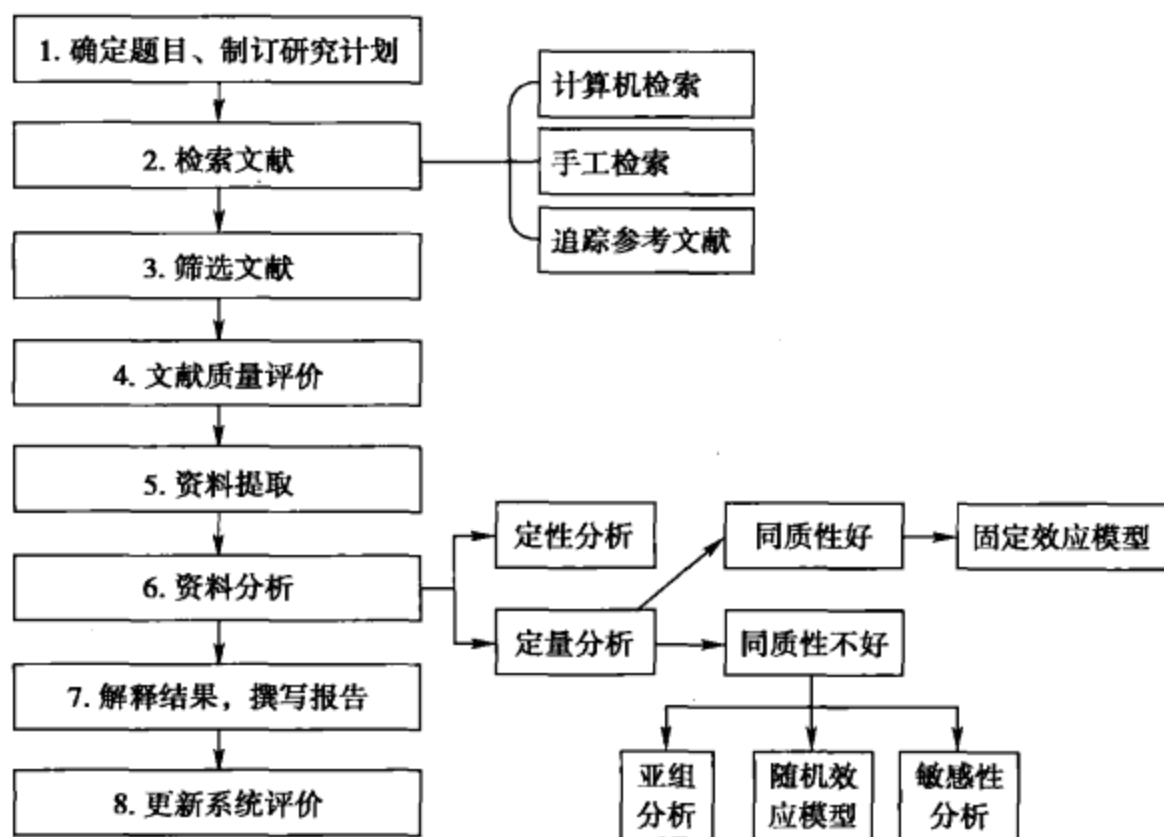


图6-1 系统评价程序图示

## 一、确立题目、制订系统评价计划书

系统评价是为医疗保健措施的管理和应用提供决策依据, 特别适用于那些干预措施的利弊, 靠单个临床研究结果难以确定, 或在临床应用过程中存在较大争议等问题的探讨。因此, 系统评价的题目主要来源于临床医疗实践中那些涉及疾病防治方面不肯定、有争论的重要临床问题。例如: 在高危人群中服用小剂量阿司匹林能否预防心脑血管病发生? 抗凝剂治疗能否预防缺血性心脏病伴心房颤动患者发生心脏事件? 急性胆囊炎患者, 早期(发病后7天内)与延缓(入院治疗后6周)行腹腔镜胆囊切除术的疗效和安全性有何差别等?

为避免重复, 首先应进行全面、系统的检索, 了解针对同一临床问题的系统评价或 meta 分析是否已经存在或正在进行。如果有, 其质量如何? 是否已经过时? 如果现有的系统评价或 meta 分析已过时或质量差, 则可考虑进行更新或重新再做一个新的系统评价。

系统评价解决的问题很专一, 涉及的研究对象、设计方案以及治疗措施需相似或相同。因此, 在确立题目时, 应围绕研究问题明确四个要素: ①研究对象的类型: 所患疾病类型及诊断标准、研究人群的特征和场所; ②研究的干预措施和进行比较的措施; ③主要研究结果的类型包括所有重要的结果(主要结果和次要结果)及严重的不良反应; ④研究的设计方案。这些要素对指导查寻、筛选和评价各个临床研究, 收集、分析数据及解释结果的应用价值均十分重要, 必须准确、清楚定义。

系统评价的题目确立后, 需要制订计划书, 内容包括系统评价的题目、背景资料、目的、检索文献

的方法及策略、选择合格文献的标准、评价文献质量的方法、收集和分析数据的方法等。

原则上,系统评价研究的问题必须在制订计划书和收集文献前就已确定,这样可避免作者根据原始文献的数据信息和结果临时改变系统评价的题目及内容,导致结论的偏倚。但由于多数系统评价是对现有文献资料的分析 and 总结,受原始文献及质量的制约,如果不了解与题目相关的资料信息和内容则难以确定一个好题目。因此,在进行系统评价的过程中如果要改变题目或评价的内容,必须明确回答原因及动机,并相应修改查寻文献和收集文献的方法。

## 二、检索文献

系统、全面地收集所有相关的文献资料是系统评价与叙述性文献综述的重要区别之一。为了避免出版偏倚和语言偏倚,应围绕要解决的问题,按照计划书中制订的检索策略(包括检索工具及每一检索工具的检索方法),采用多种渠道和系统的检索方法。除发表的论著之外,还应收集其他尚未发表的内部资料以及多语种的相关资料。

除利用文献检索的期刊工具及电子光盘检索工具(Medline、Embase、Scisearch、Registers of clinical trials)外,系统评价还强调通过与同事、专家和药厂联系以获得未发表的文献资料如学术报告、会议论文集或毕业论文等;对已发表的文章,由Cochrane协作网的工作人员采用计算机检索和手工检索联合的方法查寻所有的随机对照试验,建立了Cochrane对照试验中心注册库(Cochrane Central Register of Controlled Trials, CENTRAL)和各专业评价小组对照试验注册库,既可弥补检索工具如MEDLINE等标识RCT不完全的问题,也有助于系统评价者快速、全面获得相关的原始文献资料。

## 三、选择文献

选择文献是指根据事先拟定的纳入和排除标准,从收集到的所有文献中检出能够回答研究问题的文献资料。因此,选择标准应根据确立的研究问题及构成研究问题的四要素即研究对象、干预措施、主要研究结果和研究设计方案而制定。例如:拟探索静脉滴注硫酸镁能否降低急性心肌梗死患者的近期死亡率?围绕这一临床问题,如果确定研究对象为急性心肌梗死患者,不考虑梗死部位、患者性别、年龄,干预措施为静脉使用硫酸镁与安慰剂比较,主要研究结果为35天内的病死率,设计方案为RCT,则所选临床研究必须符合上述条件。而口服硫酸镁或静脉滴注硫酸镁与其他药物进行比较,结果为心肌梗死后35天以后的病死率或者非RCT的文献资料均不能纳入。

文献资料的选择应分三步进行(图6-2):①初筛:根据检索出的引文信息如题目、摘要筛除明显不合格的文献,对肯定或不能肯定的文献应查出全文再进行筛选;②阅读全文:对可能合格的文献资料,应逐一阅读和分析,以确定是否合格;③与作者联系:一旦被排除的文献将不再录用,因此,如果文中提供的信息不全面而不能确定,或者有疑问和有分歧的文献应先纳入,通过与作者联系获得有关信息后再决定取舍或在以后的选择过程中进一步评价。

## 四、评价纳入研究的偏倚风险

单个研究的设计和 implement 质量影响研究结果的真实性的。评价纳入研究的偏倚风险是指评估单个临床试验在设计、实施和分析过程中防止或减少系统误差(或偏倚)和随机误差的程度,以作为纳入原始文献的阈值、解释不同文献结果差异的原因、进行系统评价敏感性分析和定量分析(meta分析)时给予文献不同权重值的依据。文献评价应包括三方面内容:①内部真实性(internal validity):指单个研究结果接近真值的程度,即受各种偏倚因素如选择偏倚、实施偏倚、失访偏倚和测量偏倚的影响情况;②外部真实性(external validity或称generalizability):指研究结果是否可以应用于研究对象以外的其他人群,即结果的实用价值与推广应用的情况,主要与研究对象的特征、研究措施的实施方法和结果的选择标准密切相关;③影响结果解释的因素:如治疗性试验中药物的剂量、剂型、用药途径和疗程及依从性等因素。



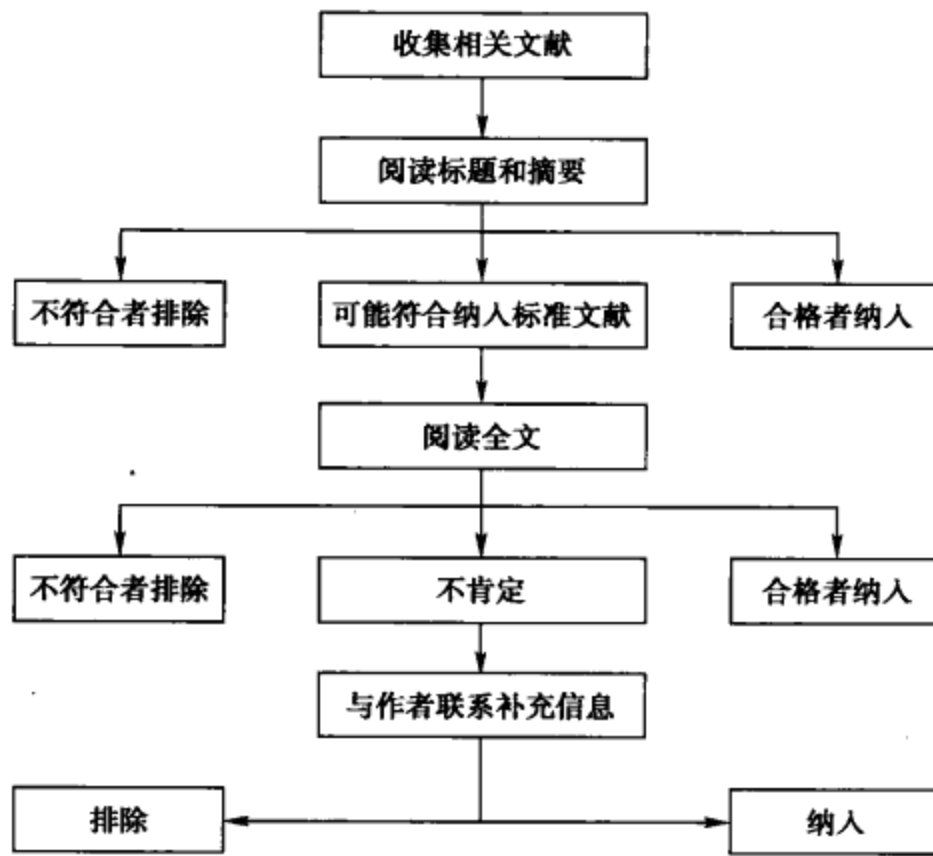


图6-2 选择文献的基本步骤

偏倚是导致研究结果偏离真值的现象,存在于临床试验中从选择和分配研究对象、实施干预措施、随访研究对象、测量和报告研究结果的每个阶段(图6-3)。因此,偏倚主要分为五种:①选择性偏倚(selection bias):发生在选择和分配研究对象时,因随机方法不完善造成组间基线不可比,可夸大或缩小干预措施的疗效。采用真正的随机方法并对随机分配方案进行完善的隐藏可避免这类偏倚的影响;②实施偏倚(performance bias):发生在干预措施的实施过程中,指除比较的措施不同外,向试验组和对照组研究对象提供的其他措施不一样。采用标准化治疗方案和盲法干预可避免实施偏倚;③随访偏倚(attrition bias):指在试验随访过程中,试验组或对照组因退出、失访、违背治疗方案等造成人数或情况不一样而产生的系统差异。对此,应尽量获得失访者的信息和对失访人员采用恰当的统计学方法处理如意向性治疗分析(intention to treat analysis)可减少其影响;④测量偏倚(measurement bias/detection bias/ascertainment bias):测量试验组和对照组结果的方法不一致所造成的系统差异,特别是主观判断研究结果时常会出现。采用统一、标准化测量方法和对研究对

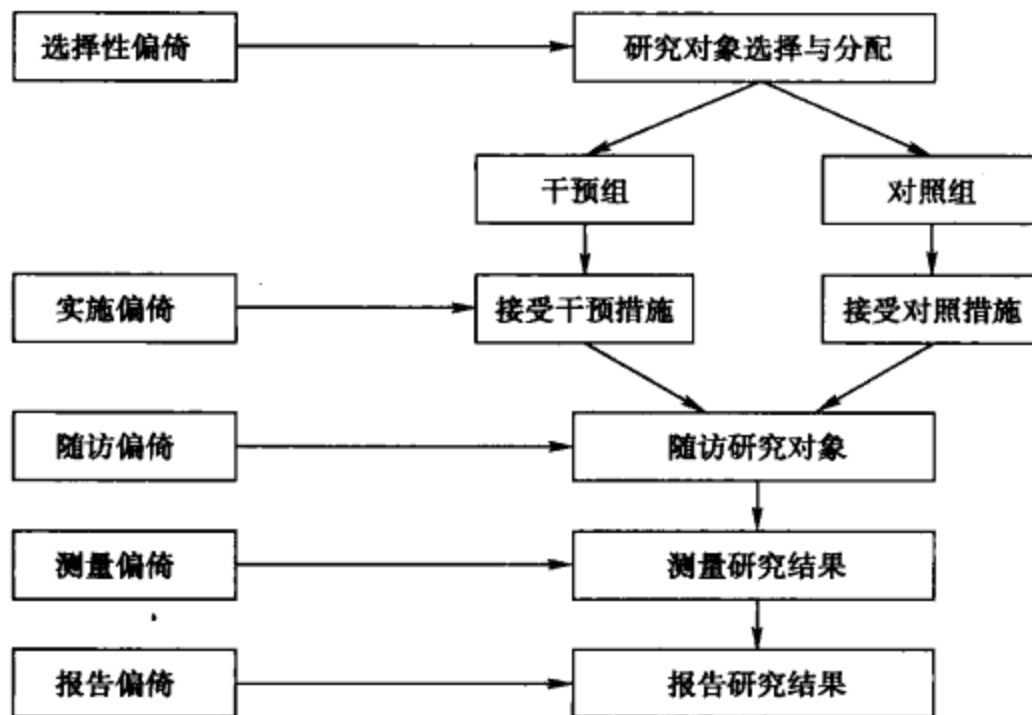


图6-3 偏倚来源

象及结果测量者实施盲法可避免其影响;⑤报告偏倚(reporting bias):指文章中报告的结果与实际分析结果间存在的系统差异。

尽管评价文献质量的方法和工具较多,但缺乏共识。评价工具可分清单或一栏表式(checklist,即有许多条目,但不给予评分)和量表评分(scale,即有许多条目,每个条目均给予评分,但可给予相同或根据重要性给予不同的权重)。迄今至少有9种以上清单(checklist)和60余种量表(scale)用于评价随机对照试验的质量,条目数从3~57个不等,一般需要花10~45分钟完成。由于这些评价方法易受文献报告质量和文献评估者的主观因素影响,Cochrane手册5.0未推荐使用任何一种清单或量表,要求采用由Cochrane协作网的方法学家、编辑和系统评价员共同制订的新的“偏倚风险评估”工具(表6-2),包括6个方面:①随机分配方法;②分配方案隐藏;③对研究对象、治疗方案实施者、研究结果测量者采用盲法;④结果数据的完整性;⑤选择性报告研究结果;⑥其他偏倚来源。针对每一项研究结果,对上述6条依次作出“是”(低度偏倚)、“否”(高度偏倚)和“不清楚”(缺乏相关信息或偏倚情况不确定)的判断。其中①②⑤用于偏倚风险评估,其余3条则需针对每一篇纳入研究中的不同研究结果加以评估,强调同一研究中不同结果受偏倚影响程度不同。偏倚风险评估结果不仅采用文字和表格描述,还要求采用图示,以更形象、直观反映偏倚情况。该评估工具对每一条的判断均有明确标准,减少了评估者主观因素影响,保证评估结果有更好的可靠性。

表6-2 Cochrane协作网的偏倚风险评估工具

评价条目	评价内容描述	作者判断
① 随机分配方法	详细描述产生随机分配序列的方法,有助于评估能否产生组间可比性	随机分配序列的产生是否正确
② 分配方案隐藏	详细描述隐藏随机分配序列的方法,有助于判断干预措施分配情况是否可预知	分配方案隐藏是否完善
③ 盲法(研究对象、治疗方案实施者、研究结果测量者,针对每一研究结果评估)	描述对受试者或试验人员实施盲法的方法,以防止他们知道受试者接受的干预措施。提供判断盲法是否成功的相关信息	盲法是否完善
④ 结果数据的完整性(针对每一研究结果评估)	报告每个主要结局指标的数据完整性,包括失访和退出的数据。明确是否报告失访/退出、每组人数(与随机入组的总人数相比)、失访/退出原因,是否采用ITT分析	结果数据是否完整
⑤ 选择性报告研究结果	描述选择性报告结果的可能性(由系统评价者判断)及情况	研究报告是否提示无选择性报告结果
⑥ 其他偏倚来源	除以上5个方面,是否存在其他引起偏倚的因素?若事先在计划书中提到某个问题或因素,应在全文中作答	研究是否存在引起高度偏倚风险的其他因素

为避免选择文献和评价文献质量人员的偏倚,可考虑一篇文章多人或盲法选择和评价,也可采用专业与非专业人员相结合的共同选择和评价办法,对选择和评价文献中存在的意见分歧可通过共同讨论或请第三方的方法解决。多人选择文献时,还可计算不同评价者间的一致性(Kappa值)。此外,应进行预试验,以摸索经验,标化和统一选择、评价方法。

## 五、收集数据

根据制订的调查表和需要收集的内容,收录有关的数据资料,其中包括:①一般资料:如评价的题目、评价者的姓名、原始文献编号和来源、评价的日期等;②研究特征:如研究的合格性、研究对象的特征和研究地点、文献的设计方案和质量、研究措施的具体内容和实施方法、有关偏倚防止措施、主要的试验结果等;③结果测量:如随访时间、失访和退出情况、分类资料应收集每组总人数及事件发

生率、连续资料应收集每组研究人数、均数和标准差或标准误等。

所有的数据资料均要输入系统评价管理软件(Review manager, RevMan),以进行文献结果的分析 and 报告。

## 六、分析资料和报告结果

对收集的资料,可采用定性或定量的方法进行分析,以获得相应的结果。

### (一) 定性分析

定性分析(non-quantitative synthesis)是采用描述的方法,将每个临床研究的特征按研究对象、干预措施、研究结果、研究质量和设计方法等进行总结并列成表格,以便浏览纳入的研究情况、研究方法的严格性和不同研究间的差异,计划定量合成和结果解释,因此,定性分析是定量分析前必不可少的步骤。

### (二) 定量分析

定量分析(quantitative synthesis)包括同质性检验(或异质性检验)、meta分析和敏感性分析。

1. 异质性检验(heterogeneity test) 系统评价或meta分析将多个研究结果合成一个效应值,不同研究间不可避免存在差异,即异质性。异质性分三类:临床异质性(clinical heterogeneity),指不同研究中研究对象、干预措施和结果测量等存在的差异;方法学异质性(methodological heterogeneity),指试验设计和质量在不同研究中存在的差异;统计学异质性(statistical heterogeneity),指不同研究中干预措施的效应值存在的差异,是临床异质性和方法学异质性导致的结果。异质性检验是指对不同原始研究之间结果的变异程度进行检验。如果检验结果有统计学意义,应解释其可能的原因并考虑进行结果合成是否恰当。确定各研究结果是否同质有两种方法:一是作图观察各研究结果的效应值和可信区间是否有重叠,如果可信区间差异太大,则放弃合成分析或采用随机效应模型。另一种方法是直接进行同质性检验(Q test, Chi-square test),在此基础上借助 $I^2$ 定量估计异质性大小,如0~40%表示异质性可能不重要,30%~60%表示有中度异质性,50%~90%表示有显著异质性,75%~100%表示有很大异质性。

2. meta分析 根据资料类型及评价目的选择效应量并对其进行定量合成分析。例如对分类变量,可选择比值比(odds ratio)、相对危险度(relative risk)、危险度差值(risk difference)和防止一例事件发生需要治疗同类患者的人数(number needed to treat, NNT)等作为效应量表示合成结果。对连续性变量,当结果测量采用同样度量衡单位时应选择均数差值(mean difference, MD),而当结果测量采用不同度量衡单位,如疼痛评分在不同研究中采用不同的量表时,则应选择标化的均数差值(standardized mean difference, SMD)。用meta分析合成结果时,可选择固定效应模型(fixed effect model)或随机效应模型(random effect model),结果采用森林图(forest plot)表示。

3. 敏感性分析(sensitivity analysis) 指改变某些影响结果的重要因素如纳入标准、研究质量的差异、失访情况、统计方法(固定效应或随机效应模型)和效应量的选择(比值比或相对危险度)等,以观察同质性和合成结果是否发生变化,从而判断结果的稳定性及程度。

## 七、解释系统评价的结果

解释系统评价必须基于研究的结果,内容应包括:

1. 系统评价的论证强度 取决于纳入研究的设计方案和每个研究的质量、是否存在重要的方法学缺陷、合成结果的效应值大小和方向、是否存在剂量-效应关系等。

2. 推广应用性 在确定系统评价结果的应用价值时,首先应考虑干预措施对患者的利弊关系,其次应考虑纳入系统评价的研究,其研究对象是否与你的患者情况相似?是否存在生物学、社会文化背景、依从性、基础危险度、病情等方面的差异。

3. 对干预措施的利弊和费用进行卫生经济分析。

4. 对医疗和研究的意义 系统评价的结果对临床医师和卫生决策者的实用价值、对今后研究的指导意义,目的在于帮助医务工作者和决策者进行正确的选择和应用,为进一步的研究指明方向。

## 八、更新系统评价

系统评价的更新是指在系统评价发表以后,定期收集新的原始研究,按前述步骤重新进行分析、评价,以及时更新和补充新的信息,使系统评价更完善。

## 第三节 meta分析

### 一、概述

系统评价可分为定性与定量评价两种。如果系统评价中纳入的原始研究缺乏有效数据或者异质性过大,那么就无法进行定量评价,只能得到定性描述结果;相反,如果符合定量分析的条件,方可考虑进行定量评价:即 meta 分析。meta 分析是将两个或多个相似研究结果进行定量综合分析的方法。广义上包括提出问题、检索相关研究文献、制定文献纳入和排除标准、描述基本信息、定量综合分析等一系列过程。狭义上,meta 分析则专指系统评价的定量分析。

因此,一个系统评价可以选用单个结局指标进行一个 meta 分析,也可选用多个结局指标实施多个 meta 分析。事实上,由于纳入研究的质量、设计类型、资料类型以及方法学等限制,只有部分系统评价可以进行定量分析。

目前,有关系统评价及 meta 分析已经在临床研究和临床实践中得以普及与推广,特别是已被广泛应用于效应量较小或存在争议的治疗性研究(主要为 RCT)、预后研究、病因学研究等,并逐步推广到剂量反应关系研究以及诊断试验的综合分析。

### 二、基本过程

meta 分析过程涉及数据提取及汇总、合并效应量估计及假设检验、异质性检验等基本内容。

#### (一) 数据提取

数据是否准确可靠,尤为关键,这是 meta 分析的基础,否则即使有先进的统计学方法,也不能弥补数据本身的缺陷。所以在收集与提取数据时,应广开渠道,通过多途径收集,确保数据全面完整;同时,采取有效的质控措施,如多人同步提取数据,防止选择性偏倚;最后对数据资料的真实性要进行严格评价,在此基础上,满足上述要求后方可进行 meta 分析。

#### (二) 数据类型及效应量的表达

可用于 meta 分析的数据主要包括以下5类:①二分类变量资料,按照某种属性分为互不相容的两类,如描述临床结局时,选用存活、死亡,复发或不复发等;②数值变量/连续性变量资料,如血压值、尿糖、CD4/CD8数等,往往有度量衡单位,且能够精确测量;③等级资料/有序多分类变量资料,按照某种属性分为多类,类与类间有程度或等级上差异。如疗效判定用痊愈、显效、有效、无效等表示。以上三类数据类型比较常见;④计数数据,即同一个体在一定观察时间内可发生多次不良事件,如心肌梗死、骨折、多次入院等;⑤生存资料,同时观察两类数据,即是否发生不良事件以及发生不良事件的时间。

数据类型不同决定了效应量的表达方式有所不同。效应量(effect size)常被定义为临床上有意义的值或改变量。当结局观察指标为二分类变量资料时,常用的效应量表达有相对危险度(relative risk, RR)、比值比(odds ratio, OR)、绝对危险度(absolute risk, AR)或 NNT 等;当结局观察指标为定量变量资料或连续性变量资料时,效应量采用均数差值(mean difference, MD)或标准化均数差值(standardized mean difference, SMD)等表达方式。对于等级资料或计数数据,可根据

实际情况转化为二分类变量资料或当作连续性变量资料进行处理,选用相应的效应量。对于生存资料,效应量表达可用风险比(hazard ratio, HR)。

在此基础上,按照统一的表格形式,将纳入研究的重要信息进行汇总,如样本量、分析方法、主要结果变量、设计方案、发表年份、具体实施时间及地点、质量控制措施等。

### (三) 异质性检验(heterogeneity test)

meta分析前尽管制定了严格的文献纳入及排除标准,最大限度地控制了异质性来源。但由于一些潜在混杂因素依然存在,如研究对象、设计方案以及统计分析模型上的差异等,均可导致异质性,若此时将结果强行合并在一起,势必会出现问题。因此,meta分析之前,应进行异质性检验,并根据异质性检验结果,来决定是否估计合并效应量;若异质性过于明显,则应探讨异质性的来源并进行相应处理。异质性检验方法主要有 $Q$ 检验法与目测图形法等。

$Q$ 检验法的无效假设为所有纳入研究的效应量均相同(即 $H_0: \theta_1 = \theta_2 = \dots = \theta_k$ ),则 $Q$ 定义为:

$$Q = \sum w_i (\theta_i - \bar{\theta})^2, \bar{\theta} = \frac{\sum w_i \theta_i}{\sum w_i} \text{ 进一步转化为: } Q = \sum_{i=1}^k w_i \theta_i^2 - \frac{(\sum w_i \theta_i)^2}{\sum w_i}$$

其中 $w_i$ 为第 $i$ 个研究的权重值。 $\theta_i$ 为第 $i$ 个研究的效应量, $\bar{\theta}$ 为合并效应量。 $k$ 为纳入的研究个数。 $Q$ 服从于自由度为 $k-1$ 的 $\chi^2$ 分布。若 $Q > \chi^2_{(1-\alpha)}$ ,则 $P < \alpha$ ,表明纳入研究间的效应量存在异质性,可进一步计算异质指数 $I^2 = \frac{Q - (k-1)}{Q} \times 100\%$ ,定量描述异质程度。

例6-1 Marinho 收集了有关含氟牙膏预防青少年龋齿的7个临床随机对照试验,试用meta分析综合评价含氟牙膏的防龋效果。资料见表6-3。

表6-3 含氟牙膏预防龋齿的疗效观察

研究	干预组		对照组		合计( $N_i$ )
	发生( $a_i$ )	未发生( $b_i$ )	发生( $c_i$ )	未发生( $d_i$ )	
Dolles (1980)	13	11	15	8	47
Forsman (1974)	174	240	56	89	559
Forsman (1974 a)	139	123	69	63	394
Hanachowioz (1984)	425	48	447	25	945
Kleber (1996)	45	32	40	39	156
Marthaler (1974)	37	13	54	5	109
Torell (1965)	113	222	169	164	668

以例6-1为例, $Q$ 检验法的计算过程见表6-4,其中 $\theta_i = \ln OR_i$ ,  $w_i = 1/Var(\ln OR_i)$ ,  $Var(\ln OR_i) = 1/a_i + 1/b_i + 1/c_i + 1/d_i$ 。

$Q = 34.76 - [(-36.61)^2/117.71] = 23.37$ ,  $\nu = 7-1=6$ ,  $P < 0.01$ ,异质性检验有统计学意义,说明研究间效应量是不同质的, $I^2 = 74.3\%$ 。

需要注意的是 $Q$ 检验法的检验效能较低,如在纳入研究数目较少的情况下,有时不能检测出异质性,出现假阴性结果,可考虑提高检验水准,如 $\alpha = 0.10$ ,以增大检验效能;相反,如果纳入研究过多,即使研究间结果是同质的,也可能出现 $P < \alpha$ 情况,即异质性检验有统计学意义。因此,对 $Q$ 检验结果的解释要慎重。

此外,还有一些图表法用于展示异质性。如标准化 $Z$ 分值图、Radial图、Forest图(森林图)、L'Abbe图等。其中通过目测森林图中的可信区间重叠程度,借以判断异质性最为常用。若可信区间大部分重叠,无明显异常值,一般可认定同质性较高。

### (四) 合并效应量估计及其假设检验

在异质性检验的基础上,选用适当的方法进行分析。若异质性不明显,同时假定理论效应量为某

表6-4 Q统计量计算过程 (以OR为例)

研究	$\ln OR_i$	$Var_{(\ln OR_i)}$	$w_i$	$w_i \cdot \ln OR_i$	$w_i \cdot \ln OR_i^2$
Dolles(1980)	-0.46	0.36	2.78	-1.28	0.59
Forsman(1974)	0.14	0.04	25.64	3.63	0.51
Forsman(1974 a)	0.03	0.05	21.89	0.69	0.02
Hanachowioz(1984)	-0.70	0.07	15.29	-10.74	7.55
Kleber(1996)	0.32	0.10	9.60	3.03	0.96
Marthaler(1974)	-1.33	0.32	3.10	-4.14	5.52
Torell(1965)	-0.71	0.03	39.42	-27.80	19.61
合计			117.71	-36.61	34.76

一固定值,纳入研究效应量间的差异是由机遇造成的,可采用固定效应模型(fixed effect model)估计合并效应量;若存在异质性,且假定理论效应量不固定、服从某种分布类型,如服从正态分布时,可用随机效应模型估计效应量及范围;若异质性过于明显,可考虑亚组分析、meta 回归分析直至放弃汇总分析,只对结果进行简单描述。

以四格表资料为例,演示合并效应量估计及假设检验过程。由于此类资料合并效应量的估计方法较多,诸如 Mantel-Haenszel 法、Peto 法、方差倒置法等。现以 Mantel-Haenszel 法为例,加以阐述(表6-5、表6-6)。

1. 估计合并效应量及95% 可信区间

表6-5 M-H法计算过程

项目	观察阳性数	理论频数	方差	$ad/T$	$bc/T$	$OR$
单个研究	$a_i$	$E_i$	$v_i$	$a_i d_i / N_i$	$b_i c_i / N_i$	$a_i d_i / b_i c_i$
汇总	$\Sigma a_i$	$\Sigma E_i$	$\Sigma v_i$	$\Sigma a_i d_i / N_i$	$\Sigma b_i c_i / N_i$	$OR_{MH}$

表6-6 Mantel-Haenszel法计算合并效应量的具体过程

研究	阳性数	理论频数	方差	$a_i d_i / N_i$	$b_i c_i / N_i$	$OR_i$
Dolles(1980)	13	14.30	2.89	2.21	3.51	0.63
Forsman(1974)	174	170.34	26.05	27.70	24.04	1.15
Forsman(1974 a)	139	138.31	21.93	22.23	21.54	1.03
Hanachowioz(1984)	425	436.46	16.86	11.24	22.70	0.50
Kleber(1996)	45	41.96	9.73	11.25	8.21	1.37
Marthaler(1974)	37	41.74	3.77	1.70	6.44	0.26
Torell(1965)	113	141.42	40.80	27.74	56.16	0.49
合计	946	984.53	122.03	104.07	142.61	0.73

$$OR_{MH} = \frac{\Sigma a_i d_i / N_i}{\Sigma b_i c_i / N_i}, OR_{MH} \text{ 对数的方差为 } Var(\ln OR_{MH}) = \frac{\Sigma P_i R_i}{2(\Sigma R_i)^2} + \frac{\Sigma(P_i S_i + Q_i R_i)}{2(\Sigma R_i)(\Sigma S_i)} + \frac{\Sigma Q_i S_i}{2(\Sigma S_i)^2}, \text{ 其中}$$

$a_i, b_i, c_i, d_i$  为四格表的实际频数,  $E_i$  为  $a_i$  的理论频数,  $R_i = \frac{a_i d_i}{N_i}; S_i = \frac{b_i c_i}{N_i}; P_i = \frac{a_i + d_i}{N_i}; Q_i = \frac{b_i + c_i}{N_i}$ ,

则  $OR_{MH}$  的95% 可信区间为:  $\exp(\ln OR_{MH} \pm 1.96\sqrt{Var(\ln OR_{MH})})$ 。

本例  $OR_{MH} = \frac{\sum a_i d_i / N_i}{\sum b_i c_i / N_i} = 104.07 / 142.61 = 0.73$ ,  $\ln OR_{MH}$  的方差为:

$$Var(\ln OR_{MH}) = \frac{\sum P_i R_i}{2(\sum R_i)^2} + \frac{\sum (P_i S_i + Q_i R_i)}{2(\sum R_i)(\sum S_i)} + \frac{\sum Q_i S_i}{2(\sum S_i)^2} = 0.0082, \text{ 则 } OR_{MH} \text{ 值为 } 0.73, \text{ 其 } 95\% \text{ 可信区间为 } 0.61 \sim 0.87.$$

2. 合并效应量的假设检验: Z 检验  $Z = \frac{\ln OR_{MH}}{\sqrt{Var(\ln OR_{MH})}}$ , 统计量 Z 服从于  $u$  分布(外文文献常表示为 Z 分布), 用于检验合并效应量是否有统计学意义。

本例  $Z = -0.3147 / 0.0906 = -3.48$ , 查 Z 表  $P < 0.001$ , 表明合并效应量有统计学意义。

实际上, 估计合并效应量以及进行异质性检验, 可以借助一些现成分析软件完成, 方便快捷。这其中首推 RevMan 软件, 图6-4则是利用该软件对例6-1资料的 meta 分析结果。

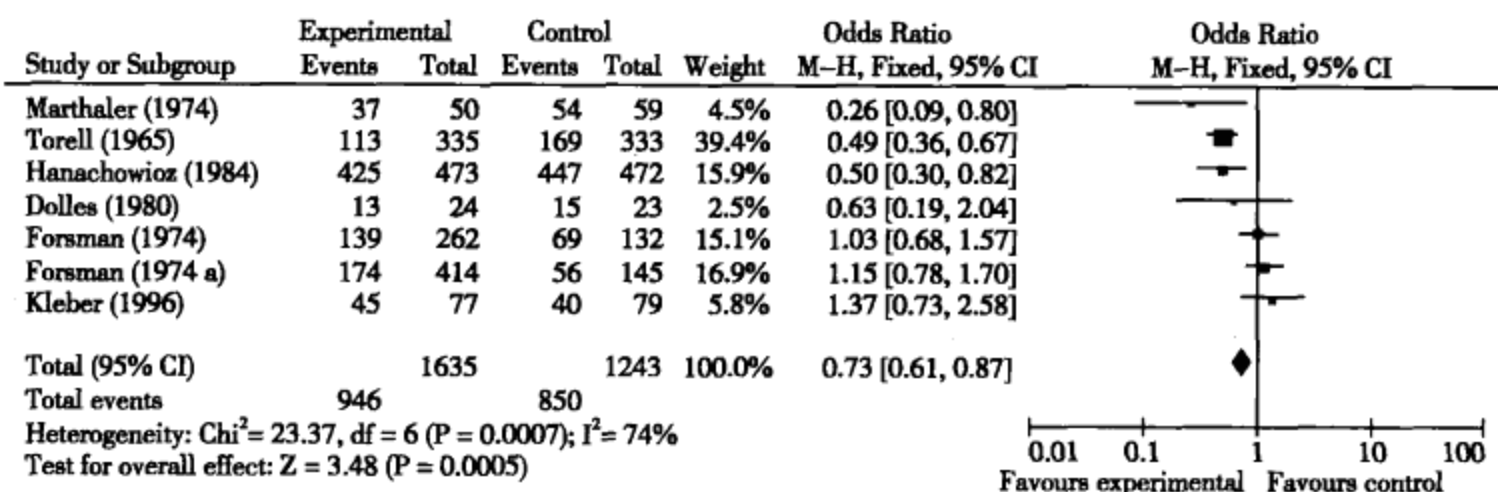


图6-4 RevMan 计算结果

图中“◆”为合并效应量图示, “0.73[0.61, 0.87]”表示合并效应量及95%可信区间; “ $Z=3.48$  ( $P=0.0005$ )”表示假设检验中的统计量及其  $P$  值。“ $\chi^2=23.37$ ,  $df=6$  ( $P=0.0007$ )”, 表示异质性检验的  $Q$  值及  $P$  值, 异质指数  $I^2=74\%$ 。

### 三、固定效应模型与随机效应模型的选择

模型选择取决于异质性检验结果以及对理论效应量的假设。如果异质性检验无统计学意义, 而且异质性小到可以忽略, 此时可认为理论效应量是固定的, 原始研究间的效应量若有差别, 也是由于抽样误差造成的, 可直接选用固定效应模型 (fixed effect model), 估计合并效应量; 反之, 如果异质性较大, 且假定理论效应量变化呈正态分布, 则应选用随机效应模型 (random effect model)。随机效应模型因将研究间的变异因子  $\tau^2$  作为校正权重, 其结果比固定效应模型结果更稳健。

#### (一) 固定效应模型

上例数据类型为二分类变量资料, 选用 MH 法固定效应模型估计得到了合并效应量。若遇到数值变量资料 (连续性变量资料), 且异质性检验无统计学意义时, 同样可选用固定效应模型进行 meta 分析, 具体过程与二分类变量资料大同小异, 只不过改用方差倒置法进行合并效应量估计。数值变量资料的效应量表达可以选择均数差值 (mean difference, MD) 和标准化均数差值 (standardized mean difference, SMD)。当纳入研究的结果变量均采用相同方式测量与表达时, 效应量可使用均数差值 (MD), 其最大的好处就是合并结果有自然单位, 易于理解。而对于那些定义相同但测量尺度不同的结果变量, 或者研究间结果变量高度不一致时, 其效应量宜采用标准化均数差值 (SMD), 但解释这类结果时应慎重。现以 SMD 作为效应量, 阐述固定效应模型的估计过程。

1. 单个研究 SMD 及 95% 可信区间估计 SMD 为标准均数差值, 可表示为  $d_i = (\bar{x}_i' - \bar{x}_i'') / s_i'$ , 其中  $\bar{x}_i'$

为第  $i$  个纳入研究其干预组的均数,  $\bar{x}_i^c$  为第  $i$  个研究对照组的均数。  $s_i^*$  为第  $i$  个研究的两组合并标准

$$\text{差, } s_i^* = \sqrt{\frac{(n_i^i - 1)(s_i^i)^2 + (n_i^c - 1)(s_i^c)^2}{n_i^i + n_i^c - 2}}。 d_i \text{ 的标准误为 } \sqrt{\frac{N_i}{n_i^i n_i^c} + \frac{d_i^2}{2(N_i - 2)}}。$$

2. 合并效应量(SMD 合并值)及95%可信区间  $d_{\text{合并}} = \frac{\sum w_i d_i}{\sum w_i}$ , 其95%可信区间为  $d_{\text{合并}} \pm 1.96 \sqrt{\frac{1}{\sum w_i}}$ ,

其中  $w_i$  为  $d_i$  标准误的平方。

3. 异质性检验  $Q = \sum w_i d_i^2 - \frac{(\sum w_i d_i)^2}{\sum w_i}$ , 进一步可计算异质指数  $I^2$ 。  $I^2 = \frac{Q - (k - 1)}{Q} \times 100\%$ 。

4. 合并效应量的假设检验  $Z = \frac{d_{\text{合并}}}{SE_{d(\text{合并})}}$ ,  $SE_{d(\text{合并})} = \sqrt{\frac{1}{\sum w_i}}$ 。

例6-2 Marinho 收集了13个有关含氟漱口剂预防青少年龋齿发生的 RCT 研究, 分别测试干预组与对照组龋失补指数增加值, 试分析含氟漱口剂的防龋效果, 具体结果如表6-7。

表6-7 龋失补指数增加值的SMD合并计算过程

研究	干预组			对照组			$s_i^*$	$d_i$	$w_i$	$d_i w_i$	$d_i^2 w_i$
	$n^i$	$\bar{x}_i$	$s_i$	$n^c$	$\bar{x}_i^c$	$s_c$					
Bastos (1989)	280	3.02	3.48	140	4.59	4.38	3.80	-0.41	91.59	-37.81	15.61
Blinkhom (1983)	190	2.65	2.31	184	3.51	2.61	2.46	-0.35	92.06	-32.16	11.23
Finn (1975)	292	3.34	3.68	161	4.27	4.21	3.88	-0.24	103.10	-24.73	5.93
Horowitz (1971)	133	0.54	1.15	123	0.72	1.00	1.08	-0.17	63.68	-10.61	1.77
Horowitz (1971a)	98	0.79	1.68	110	1.63	2.62	2.23	-0.38	50.92	-19.20	7.24
Koch (1967)	85	7.48	2.77	82	8.41	2.9	2.83	-0.33	41.18	-13.51	4.43
Koch (1967a)	117	2.58	2.7	134	2.95	2.89	2.80	-0.13	62.33	-8.23	1.09
Koch (1967b)	114	2.9	2.67	137	2.78	2.81	2.75	0.04	62.21	2.72	0.12
Mc Conchie (1977)	496	2.56	3.18	247	3.12	3.55	3.31	-0.17	164.36	-27.83	4.71
Monila (1987)	145	2.37	2.32	150	3.19	2.35	2.34	-0.35	72.60	-25.49	8.95
Radike (1973)	348	1.39	1.66	378	2.01	2.04	1.87	-0.33	178.73	-59.33	19.70
Ringelberg (1979)	341	2.78	2.94	186	3.38	3.29	3.07	-0.20	119.83	-23.44	4.58
Rugg-Gunn (1973)	222	3.74	2.49	212	5.47	3.19	2.85	-0.61	103.66	-62.85	38.10
合计	2861			2244				-0.28	1206.24	-342.47	123.47

本例合并 SMD(标准化均数差值)为 -0.28, 其95%可信区间为 -0.34~-0.23。合并 SMD 假设检验  $Z=9.86$ ,  $P<0.01$ ; 异质性检验  $Q=26.24$ ,  $P<0.01$ , 有统计学意义,  $I^2 = \frac{Q - (k - 1)}{Q} \times 100\% = 54.3\%$ 。

由于本例的异质性检验结果仍有统计学意义, 考虑选用随机效应模型。同样, 以该资料为例, 使用 RevMan 的固定效应模型分析结果见图6-5。

图中“◆”为合并效应量的图示, “-0.28[-0.34, -0.23]”表示合并标准化均数差  $SMD_{\text{合并}}$  及其95%可信区间; “ $Z=9.84$  ( $P<0.00001$ )”表示合并效应量的假设检验结果及  $P$  值。“ $Chi^2=26.13$ ,  $df=12$  ( $P=0.01$ )”, 表示异质性检验结果  $Q$  值及  $P$  值。异质指数  $I^2=54\%$ 。由于 RevMan 在估计过程中对 SMD 及标准误进行了 Hedges 校正, 结果与人工计算略有差别。

## (二) 随机效应模型

当异质性检验有统计学意义且假定真实效应量不固定但服从正态分布时, 考虑选用随机效应模型(random effect model)估计合并效应量。随机效应模型就是在固定效应模型的方差倒置法或 MH



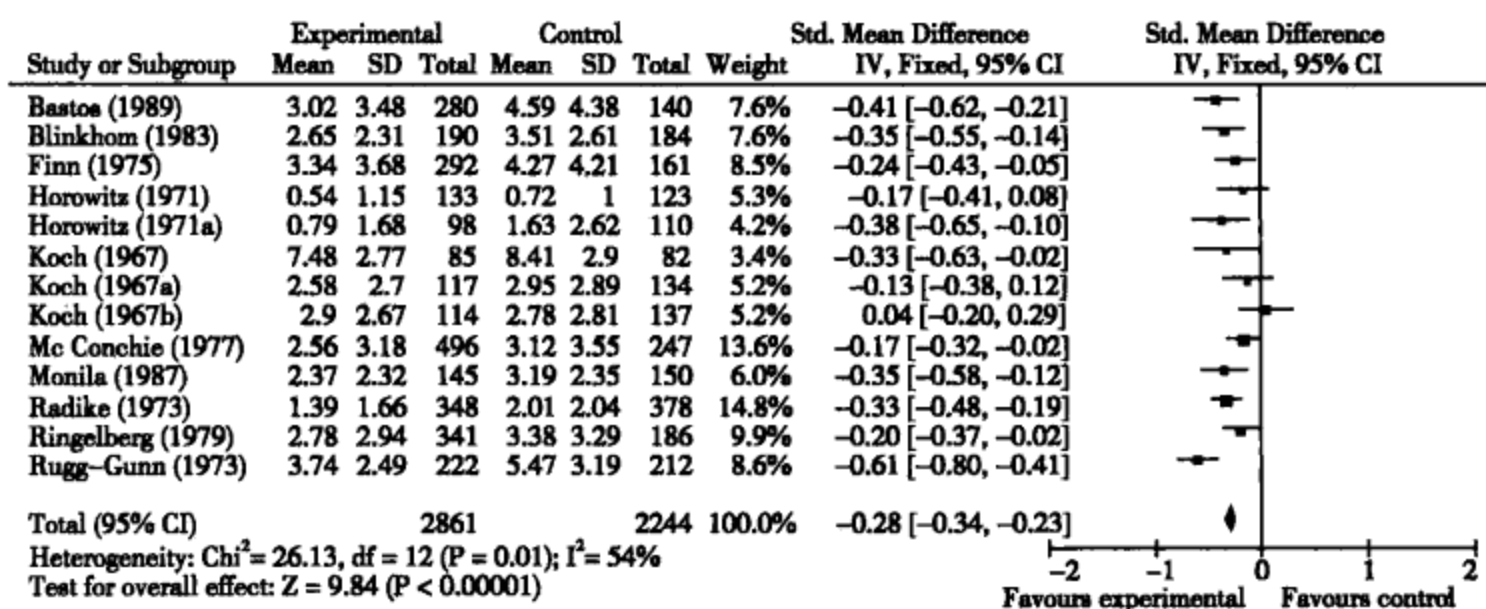


图6-5 使用 RevMan 软件的 Meta- 分析结果

法基础上采用了 DerSimonian-Laird 校正。两类模型区别在于加权方式不同,固定效应模型以每个研究内方差的倒数作为权重,而随机效应模型是以研究内方差与研究间方差之和的倒数作为权重,调整结果就是样本量较大的研究给予较小的权重,而样本量较小的研究则给予较大的权重。

1. 二分类变量资料的随机效应模型(random effect model)

该校正法是由 DerSimonian 和 Laird 于1986年提出,假设各原始研究的效应量不尽相同,以研究内方差及研究间方差之和的倒数为权重,并以此估计合并效应量。与固定效应模型相比,主要步骤相同,依次估计单个研究效应量、合并效应量及95%可信区间,最后进行假设检验。唯一不同的是需事先计算研究间方差  $\tau^2$ 。校正权重  $w_i^* = (w_i^{-1} + \tau^2)^{-1}$ 。其中  $\tau^2 = \max \left[ 0, \frac{Q - (k - 1)}{\sum w_i - (\sum w_i^2 / \sum w_i)} \right]$ ,  $Q$  为异质性

检验统计量,  $k$  为纳入分析的研究个数。若  $Q < k - 1, \tau^2 = 0$ , 若  $Q > k - 1, \tau^2 = \left[ \frac{Q - (k - 1)}{\sum w_i - (\sum w_i^2 / \sum w_i)} \right]$ 。

仍为四格表资料为例,使用随机效应模型估计合并效应量及95%可信区间:

$$OR_{\text{合并}} = \exp\left(\frac{\sum w_i^* \ln OR_i}{\sum w_i^*}\right), 95\% \text{ 可信区间为 } \exp\left(\ln OR_{\text{合并}} \pm \frac{1.96}{\sqrt{\sum w_i^*}}\right)。$$

以例6-1为例,  $Q = 23.37 > 6$ , 故  $\tau^2 = \left[ \frac{Q - (k - 1)}{\sum w_i - (\sum w_i^2 / \sum w_i)} \right] = \frac{23.37 - (7 - 1)}{117.71 - 25.77} = 0.189$ , 随机效应模型具体

估计过程如下(表6-8):

表6-8 随机效应模型的分析结果

研究	$OR_i$	$\ln OR_i$	$w_i$	$w_i^2$	$\tau^2$	$w_i^*$	$\ln OR_i * w_i^*$
Dolles (1980)	0.63	-0.46	2.78	7.74	0.189	1.82	-0.84
Forsman (1974)	1.15	0.14	25.64	657.23	0.189	4.39	0.62
Forsman (1974 a)	1.03	0.03	21.89	479.02	0.189	4.26	0.13
Hanachowioz (1984)	0.50	-0.70	15.29	233.63	0.189	3.93	-2.76
Kleber (1996)	1.37	0.32	9.60	92.25	0.189	3.41	1.08
Marthaler (1974)	0.26	-1.33	3.10	9.62	0.189	1.96	-2.61
Torell (1965)	0.49	-0.71	39.42	1553.82	0.189	4.66	-3.29
<b>合计</b>	<b>0.73</b>	<b>-0.315</b>	<b>117.71</b>	<b>3033.31</b>		<b>24.43</b>	<b>-7.67</b>

其中  $OR_{\text{合并}}$  为  $\exp(-7.67/24.43)=0.73$ , 其95%可信区间为0.49~1.09。由于固定效应模型以各研究内方差的倒数为权重, 而 D-L 校正法则以研究内和研究间方差之和的倒数为权重, 所以当异质性不明显时, 两种模型估计相同; 若存在较明显的异质性时, 结果会有差异。本例随机效应模型估计的可信区间为(0.49~1.09)明显宽于固定效应模型的估计结果(0.61~0.87), 且  $OR_{\text{合并}}$  假设检验无统计学意义。

2. 数值变量(连续性变量)资料的随机效应模型 对于数值变量(连续性变量)资料, 固定效应模型采用方差倒置法进行估计, 而随机效应模型则是在此基础上进行 DerSimonian-Laird 校正。

随机效应模型的校正权重  $w_i^*=(w_i^{-1}+\tau^2)^{-1}$ , 其中研究间方差  $\tau^2=\max\left[0, \frac{Q-(k-1)}{\sum w_i - (\sum w_i^2 / \sum w_i)}\right]$ ,  $w_i=1/v_{(di)}$ ,  $v_{(di)}=\frac{n_i^t+n_i^c}{n_i^t n_i^c} + \frac{d_i^2}{2(n_i^t+n_i^c)}$ ,  $Q$  为异质性检验统计量,  $k$  为纳入分析的研究个数。

进而 SMD 合并值  $\bar{d}_{DL}=\frac{\sum w_i^* d_i}{\sum w_i^*}$ , 其假设检验统计量  $Z=\frac{\bar{d}_{DL}}{1/\sqrt{\sum w_i^*}}$ , 95%可信区间:  $\bar{d}_{DL} \pm \frac{1.96}{\sqrt{\sum w_i^*}}$

以例6-2为例, 使用随机效应模型, 重新分析13个预防龋齿的随机对照研究, 具体结果如表6-9。

表6-9 龋失补指数增加值的SMD合并值估计(随机效应模型)

研究	干预组			对照组			$s_i^*$	$d_i$	$w_i$	$\tau^2$	$w_i^*$
	$n^t$	$\bar{x}_t$	$s_t$	$n^c$	$\bar{x}_c$	$s_c$					
Bastos (1989)	280	3.02	3.48	140	4.59	4.38	3.80	-0.41	91.59	0.013	41.81
Blinkhom (1983)	190	2.65	2.31	184	3.51	2.61	2.46	-0.35	92.06	0.013	41.91
Finn (1975)	292	3.34	3.68	161	4.27	4.21	3.88	-0.24	103.10	0.013	44.05
Horowitz (1971)	133	0.54	1.15	123	0.72	1	1.08	-0.17	63.68	0.013	34.84
Horowitz (1971a)	98	0.79	1.68	110	1.63	2.62	2.23	-0.38	50.92	0.013	30.64
Koch (1967)	85	7.48	2.77	82	8.41	2.9	2.83	-0.33	41.18	0.013	26.82
Koch (1967a)	117	2.58	2.7	134	2.95	2.89	2.80	-0.13	62.33	0.013	34.43
Koch (1967b)	114	2.9	2.67	137	2.78	2.81	2.75	0.04	62.21	0.013	34.39
Mc Conchie (1977)	496	2.56	3.18	247	3.12	3.55	3.31	-0.17	164.36	0.013	52.40
Monila (1987)	145	2.37	2.32	150	3.19	2.35	2.34	-0.35	72.60	0.013	37.35
Radike (1973)	348	1.39	1.66	378	2.01	2.04	1.87	-0.33	178.73	0.013	53.78
Ringelberg (1979)	341	2.78	2.94	186	3.38	3.29	3.07	-0.20	119.83	0.013	46.85
Rugg-Gunn (1973)	222	3.74	2.49	212	5.47	3.19	2.85	-0.61	103.66	0.013	44.16
合计	2861			2244				-0.28	1206.24		523.43

本例  $\tau^2=\left[\frac{Q-(k-1)}{\sum w_i - (\sum w_i^2 / \sum w_i)}\right]=\frac{26.24-(13-1)}{1096}=0.013$ , 合并标准化均数差值: -0.28; 合并

效应量假设检验:  $Z=\frac{\bar{d}_{DL}}{1/\sqrt{\sum w_i^*}}=\frac{-0.28}{1/\sqrt{523.43}}=-6.41$ ,  $P<0.001$ , 说明有统计学意义; 其95%可信区间

为 -0.37~-0.20。同样可以利用 RevMan 完成上述分析。由于本例的异质性检验同样有统计学意义, 相对于固定效应模型的结果(-0.34~-0.23), 显然随机效应模型结果的区间范围更宽。

#### 四、发表性偏倚的识别与分析

确保分析结果真实可靠在 meta 分析中尤为重要。在可能影响 meta 分析的偏倚中, 以发表性偏

倚的影响程度较大且较难控制,因而备受关注。发表性偏倚可使 meta 分析过分夸大治疗效应量或危险因素的关联强度,导致临床个体治疗与卫生决策的失误。

发表性偏倚是指有统计学意义的研究结果比无统计学意义的研究更容易投稿和被发表,由此而产生的偏倚。对于无统计学意义的研究,研究者可能认为意义不大,不发表或推迟发表;作为杂志编辑则更有可能对这类论文退稿。因为存在发表性偏倚,即使具备周密的检索策略和手段(如与研究个人联系),也不可能完全纳入所有相关研究。发表性偏倚的类型较多,常见的有:①当完成的临床试验得到阴性结果时,因研究者缺乏信心向国际知名的医学杂志投稿,而转投地方性杂志;②如非英语国家研究者,可能发表于本国的地方性杂志;但当得到阳性结果时,作者更愿意在国际性杂志上用英文发表,这种发表性偏倚被称为语言性偏倚;③另外还有一些论文不能发表的原因,如博士、硕士读完学位而离开原来研究单位而未能发表;④或者一些研究结果可能违背了经费提供方(如制药商)的利益,被迫搁浅不能发表;⑤出现发表性偏倚的另一种情况是一些作者为提高知名度而一稿多投,或者作为多中心研究的参研单位,同时报告各自部分结果,造成多重发表性偏倚。

现有三类比较简单的方法:即漏斗图法、剪补法以及公式法可以用来正确识别与处理发表性偏倚。其中最常用的方法就是漏斗图,其是基于样本含量(或效应量标准误的倒数)与效应量(或效应量对数)所绘制的散点图,效应量可用 RR、OR、RD 和死亡比或者其对数值等。漏斗图的前提假设是效应量估计值的精度随着样本量增加而增加,其范围也随精度增加而逐渐变窄,最后趋近于点状,其形状类似一个对称倒置的漏斗,故称为漏斗图。即样本量小的研究,数量多、精度低,分布在漏斗图的底部呈左右对称排列;样本量大的研究,精度高,分布在漏斗图的顶部,且向中间集中。利用漏斗图可以直接观察原始研究的效应量估计值是否与其样本含量有关。但当存在发表性偏倚时,则表现为漏斗图出现不对称,则呈偏态分布(图6-6)。绘制漏斗图,需要纳入较多的研究个数,原则上要求5个点以上才能进行。

图6-6所示假设为漏斗图的两情况,左图中所有研究围绕中心线对称排列,表明没有发表性偏倚,图中空心散点代表结果无效的小样本研究,小样本研究估计的效应量变异较大,出现效应量极端值机会要多于大样本研究;右图,呈不对称分布,表示存在发表性偏倚,所缺失部分恰恰为结果无统计学意义的小样本研究。基于例6-1资料,利用 RevMan5绘制漏斗图如图6-7。

图6-7显示所有研究围绕中心线对称排列,表明发表性偏倚不明显,对合并效应量的影响可以忽略。

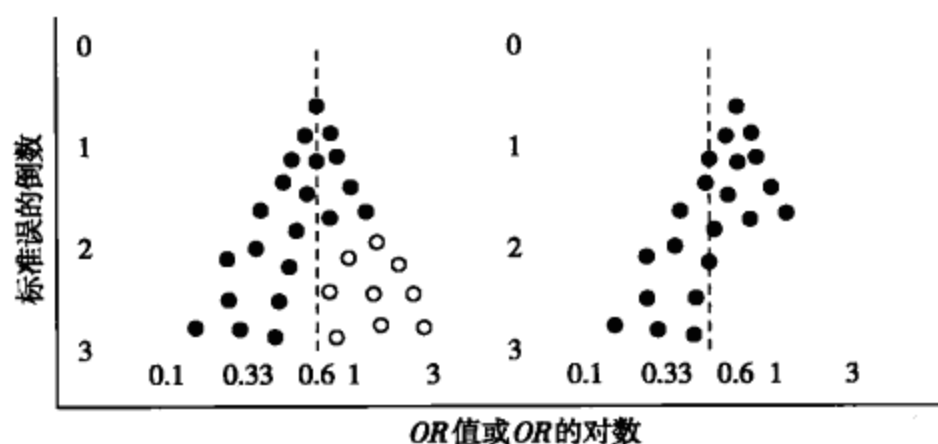


图6-6 漏斗示意图

## 五、meta回归及其他方法学进展

在临床研究中,即使研究目的完全相同,总会或多或少存在一些差别。如在药物生产厂家、剂量、研究对象年龄、病情轻重、测量时间、随访时间等方面有所不同,这些都是异质性的潜在来源。若这些因素能够被准确测量,可以选用 meta 回归模型,估计合并效应量。

$$T_i = \beta_0 + \beta_1 X_{i1} + \dots + \beta_p X_{ip} + e_i$$

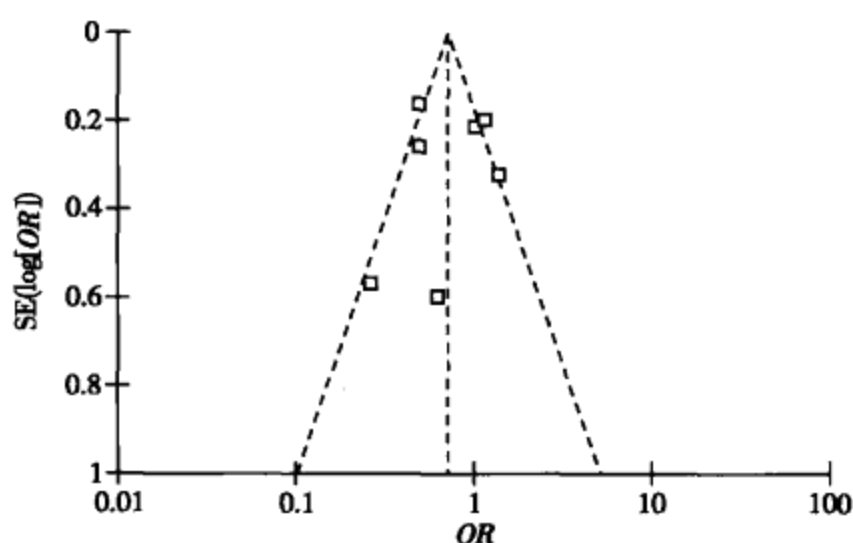


图6-7 例6-1漏斗图

其中  $X_{i1}, \dots, X_{ip}$  为混杂因素,  $\beta_0$  为固定效应,  $\beta_1, \dots, \beta_p$  为偏回归系数。

若无异质性的影响,  $\beta_1, \dots, \beta_p = 0$ , 则 meta 回归模型可简化为固定效应模型。meta 回归模型可适用于 RCT 及病例-对照研究等研究类型的 meta 分析, 也可用于敏感性分析。但 meta 回归容易产生聚集性偏倚, 特别是当资料不齐或纳入分析的研究数目较少时, 不宜进行 meta 回归分析。

尽管上述回归模型中考虑了一些混杂因素, 仍不能完全解释研究间的变异, 可进一步在模型中加入随机效应项, 那么该模型就成为混合效应模型。

$$T_i = \beta_0 + \beta_1 X_{i1} + \dots + \beta_p X_{ip} + u_i + e_i$$

其中  $u_i$  随机效应项, 其他与 meta 回归模型相同。

混合效应模型的参数估计可采用加权最小二乘法或极大似然估计法, 能够最大限度解释异质性来源。但存在两大缺点: 一是如果研究的数目较少, 不能建立混合效应模型, 二是不能进行剂量反应回归分析。

除 meta 回归模型与混合效应模型外, 其他相关方法有累积 meta 分析、迭代随机效应模型、多水平 meta 模型以及贝叶斯 meta 分析等。

## 六、慎重应用与评价 meta 分析的结果

### (一) 异质性检验与处理

若研究间有足够的同质性, 选用合适的模型(如固定效应模型或随机效应模型, 两者均可)估计合并效应量; 若存在异质性且来源已知, 采用 meta 回归模型或亚组分析, 估计合并效应量。若异质性检验有统计学意义但异质性来源未知, 可同时采用两类模型分别计算结果, 其中随机效应模型的估计结果更保守一些(即可信区间较宽)。当假设研究间效应量虽不固定、但服从正态分布时, 应选择随机效应模型的估计结果; 若异质性过大, 应放弃 meta 分析, 只对结果进行一般统计描述。

### (二) 考察 meta 分析结果的稳健性

考察 meta 分析结果的稳健性, 常常采用敏感性分析。敏感性分析(sensitivity analysis)即通过改变纳入标准(特别是那些尚有争议的研究)、排除低质量的研究或采用不同统计方法/模型分析同一组资料, 观察 meta 分析结果的变化情况, 借以考察其稳定性如何。例如在排除某个低质量研究后, 重新估计合并效应量, 并与未排除前的 meta 分析结果进行比较, 探讨该研究对合并效应量影响程度及结果稳定性。若排除后的结果未发生大的变化, 说明敏感性低, 结果较为稳健可信; 相反, 若排除后得到差别较大甚至截然相反结论, 说明敏感性高, 结果的稳健性差, 在解释结果和下结论时应非常慎重, 提示存在与干预措施效果相关的、重要的、潜在的偏倚因素, 需进一步明确争议的来源。

### (三) meta 分析结果的适用性

合并效应量实际上是多个原始研究效应量的加权平均值, 因此 meta 分析的结果在推广应用时, 应注意个体对象的特征及生物学或文化变异、干预场所、干预措施及依从性、有无辅助治疗等。不宜

推荐没有 meta 分析证据支持的建议。在无肯定性结论时,应注意区别两种情况,是证据不充分而不能定论,还是有证据表明确实无效。

#### (四) meta分析结果的时效性

meta 分析的结论同样不是一成不变的,只是对现有资料综合分析的结果,随着新的研究资料的不断纳入充实,其结论应加以更新。

## 第四节 系统评价的评价原则

近年来,系统评价或 meta 分析的数量明显增多,方法日趋复杂,对临床医务工作者和卫生决策者也产生了重要影响。但一篇系统评价或 meta 分析,并不表示其结论的绝对真实、可靠。有研究者从与 meta 分析质量有关的6个方面(研究设计、不同研究的可合成性、偏倚的控制、统计分析方法、敏感性分析、应用性)对86篇有关随机对照试验的 meta 分析进行了分析,结果发现仅28%的 meta 分析合格。因此,读者在阅读或应用系统评价或 meta 分析的结论指导临床实践前,必须对其方法和每一个步骤进行严格评价以确定系统评价的结论是否真实、可信,否则有可能被误导。评价治疗性研究系统评价的基本原则可分为以下三方面:

### 一、系统评价的结果是否真实

#### 1. 是否是根据随机对照试验进行的系统评价?

作为评价干预措施疗效“标准设计方案”的随机对照试验,如能很好地控制各种偏倚因素的影响,由此产生同质性好的系统评价被认为是论证强度最高的研究证据。而根据非同质 RCTs 及非随机对照试验进行的系统评价易受偏倚因素的影响,其系统评价的论证强度必然降低。

#### 2. 是否采用广泛和详细的检索策略检索相关文献?

从作者报告的文献检索方法中可明确收集的文献是否全面。由于标识不完整,一般的文献检索数据库如 Medline 仅能检出库中收录随机对照试验的50%,而发表偏倚即阳性结果的文章更易发表的现象可能导致系统评价出现假阳性结果。因此,全面的文献检索应包括手检相关杂志、检索会议论文集、学位论文、厂家数据库和与已发表文献作者联系。此外,如果文献检索时限制语种,也可能影响系统评价结论。收集的文献越系统、全面,则结论受发表偏倚的影响就越小,可信度就越大。

#### 3. 是否评估纳入的单个研究的真实性?

由于系统评价多为对原始文献资料的再分析和总结,因此,除了进行系统评价的方法要严格外,原始文献的质量非常重要。所以文中应详细描述评价文献质量的方法。

#### 4. 是否采用单个病例资料(或每个研究的合成结果)进行 meta 分析?

采用单个病例资料进行的 meta 分析被认为是 meta 分析的标尺(yardstick),具有根据各研究合成结果进行 meta 分析不具备的优势,如对来自不同研究的结果采用一致的定义和分界点,能从病人水平分析异质性并进行生存分析,用通常确定的亚组进行分析以检验和提出假设、通过与试验者联系可详细核查和反复校正资料,以明确随机化和随访资料的质量,通过现有病例记录系统(诸如死亡登记)更新随访信息等,将系统偏倚和机遇的影响减至最小程度。

### 二、系统评价的结果是否重要

#### 1. 不同研究的结果是否一致?

如果纳入系统评价的每个临床研究,其治疗效果相似或至少疗效的方向一致,则由此合成的结果的可信度较高。因此,作者应对各个研究结果之间的相似性,即进行同质性检验。如果同质性检验有统计学差异,则应解释差异的原因并考虑将结果进行合成是否恰当。

#### 2. 系统评价的疗效大小? 疗效的精确性如何?

在进行结果合成时,不能通过简单比较阳性研究结果和阴性研究结果的研究个数来确定系统评价的结论,而应该根据研究的质量和样本含量的大小对不同研究给予不同的权重值,并采用恰当的指标如比值比(odds ratio, OR)、相对危险度(relative risk)、均数的差值(mean difference)、防止一例事件发生需要治疗同类患者的例数(NNT)和统计方法如随机效应模型和固定效应模型等合成结果,同时计算相应的可信区间。

### 三、系统评价的结果是否能应用于我的病人

系统评价报告的结果是所有研究对象的“平均效应”,而你主管的病人不一定在研究中,因此考虑系统评价结果能否应用于你主管的具体病人时应从四个方面进行考虑。

1. 我的病人是否与系统评价中的研究对象差异较大,导致结果不能应用于你的患者?

可通过比较我的病人与系统评价中的研究对象性别、年龄、伴随疾病、疾病严重程度、病程、依从性、文化背景、社会因素、生物学及临床特征等方面的差异,并结合临床专业知识综合判断结果的推广应用性。

2. 系统评价中的干预措施在我的医院是否可行?

由于技术力量、设备条件、社会经济因素的限制,即使系统评价中的干预措施效果明显,有时在自己所在的医院却不能实施,难以应用于病人。

3. 自己的病人从治疗中获得的利弊如何?

任何临床决策必须权衡利弊和费用,只有利大于弊且费用合理时才有价值应用于病人。例如:告诉一患者其患病的真实情况有助于早期治疗和获取患者的配合,但也增加了患者的心理负担,可能降低生活质量。

4. 对于治疗的疗效和不良反应,自己的病人的价值观和选择如何?

循证医学强调,任何医疗决策的制订应结合个人的专业知识和经验、当前最佳的研究证据和患者的选择进行综合考虑,应以“病人”为中心而不是单纯治病,目前越来越强调病人参与医疗决策。

## 第五节 系统评价的应用

### 一、临床医疗的需要

随着循证医学兴起,强调任何医疗决策的制定应遵循和应用科学研究结果,即应将个人的临床专业知识与现有的最好临床科学研究结果结合起来进行综合考虑,为每个病员进行最佳诊治决策。除了高质量的原始论著外,系统评价的广泛应用正不断地改进和规范着医务工作者的医疗实践行为。例如:美国的政策研究所常应用系统评价的结果制定临床实践指南。有关为低血容量、烧伤和低蛋白血症病人常规补充白蛋白的系统评价发现,该种疗法导致英格兰和威尔士的死亡人数每年增加1000~3000人,这一结果引起了临床医师、科研人员和卫生决策者的极大关注,并呼吁禁止盲目使用白蛋白。英国伦敦 St.George 医院根据 Cochrane 系统评价结果改变了急性哮喘的治疗方案,预计一年可节约成千上万英镑。

### 二、科研工作的需要

临床科研要具有先进性、新颖性和临床价值,面对浩瀚的医学文献信息,研究人员必须查寻、阅读和评价相关领域的文献资料,掌握研究课题的历史、现状、发展趋势、存在问题、当前研究的热点与矛盾,提出选题、立题的依据,避免重复前人的工作,为研究工作提供信息资料和研究方向。目前,许多国家都非常重视高质量系统评价在临床科研中的价值。例如,英国国家医学研究会资助的临床试验,要求申请者回答是否已有相关的系统评价及结论如何,并邀请系统评价的作者参与临床试验申请

书的评审。

### 三、反映学科新动态

围绕专业发展的热点,纵览某一领域的最新文献资料,作好有关专题的系统评价,全面、深入和集中地反映该领域目前的动态和趋势、存在的问题和发展的方向,以促进学科的发展,保证不断地吸收新知识、新营养而居于学科的前沿位置。

### 四、医学教育的需要

医学教育除了向医学生传授各种疾病的共同规律和特性方面的知识外,还应该及时传授某一疾病的最新进展以及新药物、新技术的发展情况。教科书由于出版周期长,常常难以反映最新动态。因此,医学教育者需要不断地阅读有关医学文献以更新知识。而系统评价是快速获取有关知识的途径之一。有些国家的作者开始使用 Cochrane 系统评价的结果来撰写医学教科书的有关章节。

另外,广大的基层医务工作者,由于工作繁忙、文献资源有限,为了知识的不断更新,可通过阅读有实用价值的、真实可靠的系统评价,作为学习新知识的继续教育资源。

### 五、卫生决策的需要

随着人口增长、年龄老化、新技术和新药物的应用、人类健康需求层次的提高,使有限卫生资源与无限增长的卫生需求之间的矛盾日益加剧,要求各级卫生管理人员制定卫生决策时应以科学、可靠的研究结果为依据,合理分配卫生资源,提高有限卫生资源的利用率。目前许多国家在制定卫生决策时均要以医学文献资料特别是系统评价为依据。例如:1990年,魁北克的卫生技术评估委员会发表了一篇有关使用造影剂后发生副作用的 meta 分析。报告明确指出,没有证据说明使用高渗造影剂比低渗造影剂增加生命危险,仅严重副作用的发生率稍有增加。这一结果的公布使魁北克1990~1992年因使用低渗造影剂的医疗费用明显降低,净节约(除去处理严重副作用的费用)约1200万美元,即使保守估计也可节约1000万美元左右。

总之,只有采用科学、严格的方法产生的系统评价才能为临床医疗实践、医学教育、科研和卫生决策提供真实、可靠的信息,在应用系统评价结论时应该进行严格的评价。

(李 静 康德英)

## 第七章 循证医学实践的个体化原则与方法

循证医学所采取的最佳证据都是源于临床或基础医学的研究成果,这些研究的结果无论是单项研究或多个研究的系统评价,都是从有限的研究对象中所取得的综合平均效应(果),是总体效应的均值水平(共性)。其中个体间的效应,往往有不同的差异。比如,接受同一治疗措施的不同个体,有的疗效显著,而有的也许无效,这就是个体的差异(特性),造成这种差异的因素,则是多方面的,有的可知,也许有些原因不明,这在临床医疗实践中往往是很常见的现象。

应用最佳证据(共性资料)来指导具体的临床医疗实践时,除了注重它的总体效果之外,十分重要的还要注意纳入研究的患者特点,特别要注意纳入标准及排除标准(inclusion & exclusion criteria)。这是因为临床循证医学实践,是要应用证据来诊治自己的病人,这就要针对病人的个性特点,有的放矢,方可达到循证医学临床实践的目的,是执行理论联系实际的根本原则。

本章就循证医学临床实践执行个体化原则与方法进行论述。

### 第一节 最佳临床证据的特性

应用于循证医学实践的证据应该是最佳证据(best evidence),至少应具备以下几个特性。

#### 一、真实性

对于拟采用的证据首先必须分析其真实性(validity),这里最重要的是分析与评价证据的内部真实性(internal validity)。重点是:①分析是来自什么样的研究设计方案,有否恰当的对照组;②分析研究对象的诊断标准及纳入和排除标准是否明确;③分析组间的临床基线是否可比、干预措施和方法是否科学有效和安全;④分析终点指标是否确切、有何偏倚存在及采取了什么防止和处理方法、依从性如何?⑤资料收集、整理、统计分析是否合适?特别是是否进行了重要临床特点的分层分析(stratifying analysis),结果与结论是否真实和可靠。这些都是涉及证据真实性的关键因素。如果上述因素都被考虑和分析,就有助于对证据真实性的评价。不同的研究设计所产生的证据,在级别与水平上有很大差异,详见本书相关章节。

#### 二、重要性

当拟采用的证据被确认为“真实性好”之后,还要评价其有无临床应用价值。如系诊断性试验,则反映在对疾病的诊断方面究竟提高了多少敏感度、特异度及准确度;如系治疗性研究的证据,则在治疗方面究竟提高了多少疗效、安全性及成本究竟如何;在影响疾病的预后方面,有害和有利于预后的因素是哪些,各有多大的贡献等。这些都应有明确的定性和量化的指标以佐证其临床重要性的程度。

#### 三、实用性

任何最佳证据的应用和推广,都必须结合病人的实际病况、医疗条件、医务人员的知识技能水平,患者的接受程度以及社会经济状况的承受能力等。



鉴于当前临床研究的最佳证据来源于发达国家居多,由于人种、社会环境、经济水平、医疗条件乃至生物因素在不同国家差异颇大。因此,引用国外医学研究成果时,更要考虑不同的国情、种族以及病人的病情特点,切不可生搬硬套。要对具体的问题进行具体的分析,方可得出有关最佳证据是否适用的决策。

#### 四、物美价廉

经上述分析和评价的最佳或较好的证据,在采用时还应进行卫生经济学评价。原则是从多种良好证据的比较中,价格低廉、安全性及有效性可靠者应优先采用。这对广大弱势群体患者是非常重要的。对于解决“看病贵、看病难”的问题也是有一定的积极意义。

## 第二节 生物学依据

任何最佳研究成果可否应用于个体患者的诊治实际,首先应该考虑其生物学依据(biologic evidence)。例如某一药物的治病效应,从发病的生物学机制上是作用在哪个环节?为何有效?为何无效甚至发生不良反应等,均应注意有关生物学依据:

在病原生物学方面,对于感染性疾病具有十分重要的意义。例如:结核病的病原生物体是结核杆菌;病毒性肝炎的病原生物体分别是甲、乙、丙、丁、戊等型的肝炎病毒,其中乙型肝炎病毒所致慢性肝炎与肝癌有较强的相关性等,这些都为病原生物学依据。

在感染性疾病的抗生素治疗方面。例如同一种有效的抗生素,治疗同种细菌所致的感染性疾病,其敏感菌株与耐药菌株就会呈现出不同的治疗效果,也要涉及生物学的机制。

在不同人种或民族中,往往有着某些生物学方面的差异,在循证医学应用最佳证据的个体化实践中要予以注意。例如:黑人的高血压发病率较白人高且危害重,治疗高血压临床证明有效的 $\beta$ -受体阻滞剂、血管紧张素转换酶抑制剂,对黑人的高血压治疗较利尿剂的疗效差;血管紧张素转换酶抑制剂引起的血管性水肿显著,较白人的发生率高。反之,我国大凉山地区的彝族人群经调查高血压患病率却很低。又如急性或恶性肿瘤,由于某种耐药基因的影响,致使有该种基因存在的患者对化疗的治疗反应差或复发率高。因此,在疾病防治中要注意人种与民族的生物学特点。

此外,由于生物学证据不足,常常给我们在某些疾病的诊断和治疗方面带来许多困难。例如:恶性肿瘤的诊断,由于缺乏早期特异性诊断方法,使得临床早期正确诊断非常困难,进而对临床治疗与预后也有重要的影响,这就需要加强这方面的研究。

可见,在具体患者诊疗的循证医学实践中,对拟采用“最佳证据”的生物学依据的分析是很重要的。

## 第三节 病理生理学依据

任何证据的应用都需要考虑患者的病理生理特点。例如:2003年在我国流行的急性传染性非典型肺炎(SARS),部分中毒症状及肺部炎症损害明显,呼吸功能明显障碍者,适时适量地应用肾上腺皮质激素治疗取得了较好的临床效果,挽救了不少病人的生命,此为有效的证据。然而,对于那些中毒症状以及肺部损害不重、呼吸功能影响较轻的患者,应用激素则应审慎,以避免带来不良的后果。因此,应用任何有效药物的时候,一定要考虑病人的具体病理生理状况,是否与“最佳证据”的临床资料相吻合。

任何最佳证据的应用,一定要仔细考虑患者由于病理生理损害严重而导致临床病情的复杂多变性。因为临床研究的成果,就研究对象而言,相对较为单纯,因而即使是最佳证据,对临床的总体代表性(外部真实性, external validity)而言,往往有一定的局限。如果面对的对象病情较重或有某些(种)

并发症或并存症,则不会那样简单了。例如:对轻、中型高血压患者降压的随机对照试验研究,证明了 $\beta$ -受体阻滞剂及血管紧张素转换酶Ⅱ受体阻滞剂有良好的降压效果。如果患者是一位重症高血压患者( $BP>180/120\text{mmHg}$ ),或者伴有心脑靶器官损伤者,此时,宜注意被引用的资料(证据),有无轻、中、重型病例的分层比较证据,如有重型或伴有并发症且又与该患者病情相似,那么相应的证据就有被采用的价值。否则,应另作考虑。

在估价疾病预后和对有关危险因素的干预,借以改善患者的预后,提高病人的生存质量,这就要求将多种相关因素的研究证据,结合患者疾病的病理生理特点进行具体分析和评价,以估计可能发生某一事件的概率并有针对性地加以干预,从而防止或降低不良事件发生,以达到改善患者预后的目的。

#### 第四节 社会-心理及经济特点

在不同的社会和经济环境,对有关疾病的诊断及治疗性研究的最佳证据,当其用于具体患者时常有着很大的差异。例如:在发达国家或地区对冠心病的诊断可以应用冠脉造影的“金标准”进行诊断,以及采用冠脉搭桥手术或安置支架进行有效的治疗。然而其费用之昂贵加之要求高档的仪器设备和医技手段,非一般不发达地区或平民所能承担和接受。反之,在发展中国家或贫困地区对一些急性传染病、呼吸道感染、地方病、妇幼疾患、营养不良、急性腹泻等防治性研究,这些成果在发展中国家及贫困地区实践循证医学有用,而在发达国家则未必都有重要价值。因此,在应用最佳研究证据的时候,即使某些有效的诊治措施符合于患者的实际,自然还要考虑其社会经济特点。

患者是否愿意接受“最佳证据”,医生是否能认真负责地实践循证医学,这就会涉及患者以及医生的心理状态及依从性的问题。如果医生高度负责,对先进的研究成果接受的敏感性高,就会更好地应用最佳证据为自己的患者服务。在这种情况下,建立良好医患关系,患者愿意接受治疗,依从性好,自然会取得良好的效果。这里就要有医患互信与良好的心理状态,否则是难以奏效的。

因此,最佳证据用于具体患者的时候,一定要尽一切努力改善患者对接受治疗的依从性,办法是要了解患者的心理与所处的社会状态,要给患者进行接受治疗的医学知识教育;要倾听患者对治疗的意见和解除可能存在的顾虑;在治疗中要给患者提供方便的服务,尽量减少患者及其家属对接受医疗的某种不便;不断地加强交流,从而保持与发展良好的心理状态和合作关系;这是实践循证医学的重要基础之一。这方面的问题解决好了,不仅能提高医疗质量,而且可以大大地有利于缓解当前较为普遍存在的医患矛盾。

#### 第五节 证据的应用要权衡利弊

最佳证据是否可以用于对个体患者的医疗决策,自然要考虑拟采用的诊治措施能给患者带来多大的利益,同时还要考虑它们被应用后可能产生哪些不良反应及对患者造成的危害程度。因此,拟对患者采用的诊治措施必须对利、弊关系进行客观的评估,只有利大于弊方可采用。

所谓“利大”,指的是对患者拟用的新的诊治措施,其临床意义显著,对患者大有益处。显著的临床意义或称价值,一定有其量化的指标。常用者,在病因及危险因素方面,有相对危险度(relative risk, RR)、归因危险度(attribute risk, AR)、病因学分数(etiological fraction, EF)、比值比(odds ratio, OR)等;在诊断学试验方面有敏感度、特异度、准确度、患病率、预测值、似然比等指标;在治疗方面:有治愈率、有效率、病死率、绝对危险降低率(absolute risk reduction, ARR)、相对危险降低率(relative risk reduction, RRR)、与对照组比较,需要治疗多少例数才取得一例最佳效果(number needed to treat, NNT)等。关于药物不良反应(adverse drug reactions, ADR),要予以高度重视,应掌握其不良反应及程度如何?重要事件(如致残、致死)率有多大?试验组与对照组比较其不良反应的危险度增高了多

大? 治疗多少病例才发生一例重要的不良事件( number needed to harm, NNH)? 有了这些指标和数据, 就可以比较试验组和对照组的利弊比( likelihood of help vs harm, LHH), 通常用 $1/\text{NNT}:1/\text{NNH}$  比值表示, 如 $1/\text{NNT}$  越大,  $1/\text{NNH}$  越小, 则利弊比就为佳(表7-1、表7-2)。这些指标的意义及其应用将在本教材相关章节全面解释。

表7-1 干预效果测试的指标

正 效	负 效
1. 事件率 ER (event rate), 如有效率、治愈率等	事件率 ER, 如不良反应发生率
2. 绝对危险降低率 ARR (absolute risk reduction)	绝对危险增高率 ARI (absolute risk increase)
3. 相对危险降低率 RRR (relative risk reduction)	相对危险增高率 RRI (relative risk increase)
4. 需要治疗多少病例才能获得1例最佳结果 NNT(number needed to treat)	需要治疗多少病例才会导致发生1例不良反应 NNH( number needed to harm)

表7-2 干预效果的利弊比值表 (LHH)

比值	1/NNT										
	1/10	1/20	1/30	1/40	1/50	1/60	1/70	1/80	1/90	1/100	
1/10	1.0	0.5	0.3	0.3	0.2	0.2	0.1	0.1	0.1	0.1	
1/20	2.0	1.0	0.66	0.5	0.4	0.34	0.3	0.3	0.2	0.2	
1/30	3.0	1.5	1.0	0.75	0.6	0.5	0.4	0.4	0.3	0.3	
1/40	4.0	2.0	1.3	1.0	0.8	0.7	0.6	0.5	0.4	0.4	
1/50	5.0	2.5	1.7	1.3	1.0	0.9	0.7	0.7	0.6	0.5	
1/60	6.0	3.0	2.0	1.5	1.2	1.0	0.8	0.8	0.8	0.6	
1/70	7.0	3.5	2.3	1.8	1.4	1.2	1.0	0.9	0.8	0.7	
1/80	8.0	4.0	2.6	2.0	1.6	1.4	1.1	1.0	0.9	0.8	
1/90	9.0	4.5	3.0	2.3	1.8	1.5	1.3	1.2	1.0	0.9	
1/100	10	5.0	3.3	2.5	2.0	1.7	1.4	1.3	1.1	1.0	

在治疗性研究证据中, 应用 NNT 及 NNH 数据分别计算各自的倒数值, 查表7-2则可获利弊比值, 如 $1/\text{NNT}=1/50, 1/\text{NNH}=1/100$ , 利弊比 =2。

在循证医学实践中, 涉及上述有关指标的意义判断和应用, 应结合专业及临床的实际, 以评价分析利弊来决定采用或不宜采用的有关最新证据。

### 第六节 个体化干预的效果预测

应用“最佳证据”对个体患者进行干预, 究竟会产生什么效果? 可能达到什么结局? 这是患者和医生都十分关注并期望回答的实际问题。

现在让我们从一实例的探讨中去理解, 并期望读者在自己的临床实践中能举一反三。

患者,男性,70岁,患高血压20年,一直在用药物进行治疗,但血压维持在150/90mmHg左右,疗效欠满意。3年前又发现患有糖尿病(2型),接受胰岛素治疗,血糖维持在6.0~6.5mmol/L,且血脂代谢紊乱,甘油三酯2.5mmol/L,血胆固醇6.5mmol/L, HDL0.8mmol/L, LDL 3.8mmol/L。肝、肾功能正常。

患者有18年的吸烟史,现10支/日。戒烟困难。

查体:体型稍胖,身高172cm,体重74kg, BP 150/94mmHg。双肺正常,心脏无异常体征,肝脾不大,下肢不水肿。

心电图正常。

患者就医目的:询问根据现况发生冠心病或脑卒中的危险度有多大?如进行有效的治疗可以降低发生心脑血管病的危险吗?这是患者十分关心的问题,你能依据研究的最佳证据正确回答吗?

根据国内外数十年来大量的流行病学研究结果所获得的可靠证据,明确了心脑血管病发病的主要危险因素归纳如表7-3。

表7-3 心脑血管病的危险因素

高血压
吸烟
肥胖 (BMI ≥30)
缺乏体力活动
高脂血症
糖尿病
微量蛋白尿或测肾小球滤过率 GFR<60ml/min
年龄 (男性 >55岁,女性 >65岁)
早发心血管疾病家族史 (男 <55岁,女 <65岁)

[摘自 JNC 7. JAMA. 2003.289(19):2560-2572]

从该表所列的心脑血管病的危险因素中,本病例具有发生心脑血管病的危险因素为高血压、吸烟、高脂血症、糖尿病、年龄 >55岁共5项。虽然患者现在尚无心脑血管病的临床证据,但发生概率在未来是颇高的。然而这仅仅是定性结论。是否有量化的预测指标呢?如果有,那么我们进行有针对性的干预后,是否使发生心脑血管病的危险度降低呢?

现在让我们分别引用我国中国医学科学院阜外医院国家“十五”攻关项目总结的有关缺血性心脑血管疾病(ICVD)综合危险度评估工具,见图7-1中列表,可为本例患者进行评估。

根据图中列表,该患者年龄 >59岁计4分, SBP 计2分,体重指数(BMI: 74kg/1.72m<sup>2</sup>=25)计1分,胆固醇计1分,吸烟计2分,糖尿病计1分,总分为11分。从第三步表中可查出未来10年发生 ICVD 的危险度为12.8%。

国外的定量评估 ICVD 工具中,颇有重要参考价值,并可用于对尚未发生缺血性心脑血管病患者的预测,或作干预后估价危险度下降的效果预测,这里,要推荐新西兰高脂血症及高血压管理指南中所列“估计心脑血管病的危险与治疗益处( estimating cardiovascular risk and treatment benefit )图”。在该资料中,确定心血管事件为新发的心绞痛、心肌梗死、冠心病死亡、卒中或短暂性脑缺血发作、充血性心衰或外周血管综合征。

研究者根据最为主要的有关危险因素的研究结果,预测未来五年内发生上述心脑血管事件的危险度。如经过干预后,视这些可控危险因素的变化,以确定治疗将带来的益处。

为了形象化的指导干预实践,他们绘制了多维图(图7-2,彩图7-2)ICVD 预测危险度与干预效果图。

彩图7-2中主要的变量为性别、年龄、有无糖尿病、吸烟与否、血压水平、总胆固醇/高密度脂蛋白胆固醇(HDL)的比例,共6大项。图中以不同颜色的小方块图示所查病例发生 ICVD 的危险程度。

在判断危险度及治疗获益方面,列出表7-4。

缺血性心脑血管疾病(ICVD)综合危险度评估工具

男性

第一步：评分		第二步：求和		第三步：绝对危险	
年龄(岁)分值	收缩压 分值	危险因素	得分	总分	10年ICVD危险(%)
35~39 0	<120 -2	年龄	_____	≤ -1	0.3
40~44 1	120~129 0	收缩压	_____	0	0.5
45~49 2	130~139 1	体重指数	_____	1	0.6
50~54 3	140~159 2	总胆固醇	_____	2	0.8
55~59 4	160~179 5	吸烟	_____	3	1.1
	≥180 8	糖尿病	_____	4	1.5
		总计	_____	5	2.1
				6	2.9
				7	3.9
				8	5.4
				9	7.3
				10	9.7
				11	12.8
				12	16.8
				13	21.7
				14	27.7
				15	35.3
				16	44.3
				≥17	≥52.6

年龄(岁)	分值	总胆固醇	分值
<24	0	<200	0
24~27.9	1	≥200	1
≥28	2		

吸烟	分值	糖尿病	分值
否	0	否	0
是	2	是	1

10年ICVD绝对危险参考标准		
年龄(岁)	平均风险(%)	最低风险(%)
35~39	1.0	0.3
40~44	1.4	0.4
45~49	1.9	0.5
50~54	2.6	0.7
55~59	3.6	1.0

女性

第一步：评分		第二步：求和		第三步：绝对危险	
年龄(岁)分值	收缩压 分值	危险因素	得分	总分	10年ICVD危险(%)
35~39 0	<120 -2	年龄	_____	-2	0.1
40~44 1	120~129 0	收缩压	_____	-1	0.2
45~49 2	130~139 1	体重指数	_____	0	0.2
50~54 3	140~159 2	总胆固醇	_____	1	0.3
55~59 4	160~179 3	吸烟	_____	2	0.5
	≥180 4	糖尿病	_____	3	0.8
		总计	_____	4	1.2
				5	1.8
				6	2.8
				7	4.4
				8	6.8
				9	10.3
				10	15.6
				11	23.0
				12	32.7
				≥13	≥43.1

年龄(岁)	分值	总胆固醇	分值
<24	0	<200	0
24~27.9	1	≥200	1
≥28	2		

吸烟	分值	糖尿病	分值
否	0	否	0
是	1	是	2

10年ICVD绝对危险参考标准		
年龄(岁)	平均风险(%)	最低风险(%)
35~39	0.3	0.1
40~44	0.4	0.1
45~49	0.6	0.2
50~54	0.9	0.3
55~59	1.4	0.5

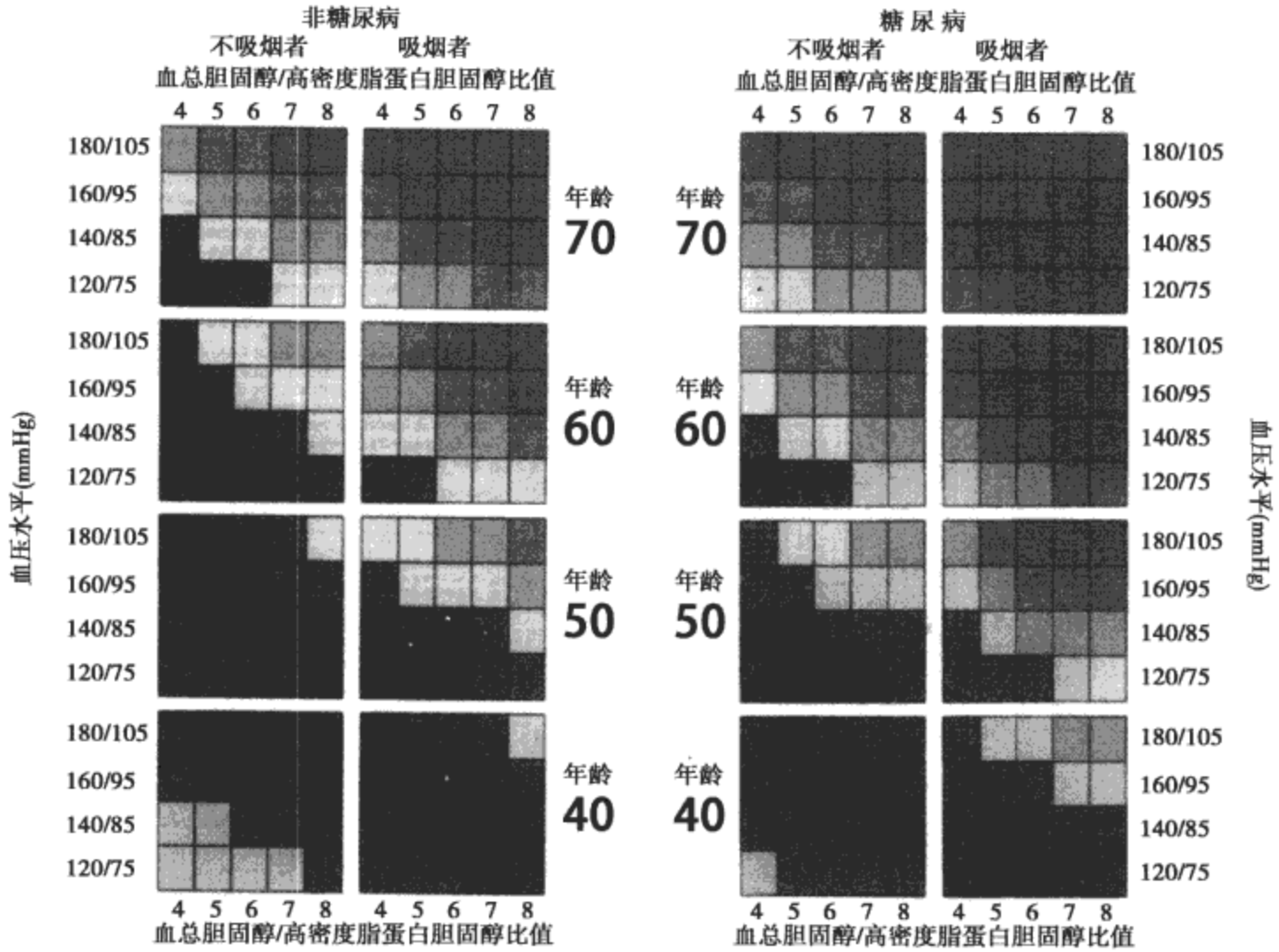
单位：收缩压：mmHg；体重指数：kg/m<sup>2</sup>；总胆固醇：mg/dl；空腹血糖：mg/dl

图7-1 缺血性心脑血管疾病(ICVD)综合危险度评估工具

表7-4 CVD的危险度和治疗获益方面估价表

危险水平 (5年发生致死与非致死CVD)	益处(1) (5年治疗100人可预防CVD)	益处(2) (治疗5年的NNT)
非常高	红色 >30%	<10
	深橙 25%~30%	11
	淡橙 20%~25%	13
高	黄色 15%~20%	16
中	绿色 10%~15%	25
低	蓝色 5%~10%	40
	紫色 2.5%~5%	80
	淡蓝色 <2.5%	120

男性的危险水平



女性危险水平

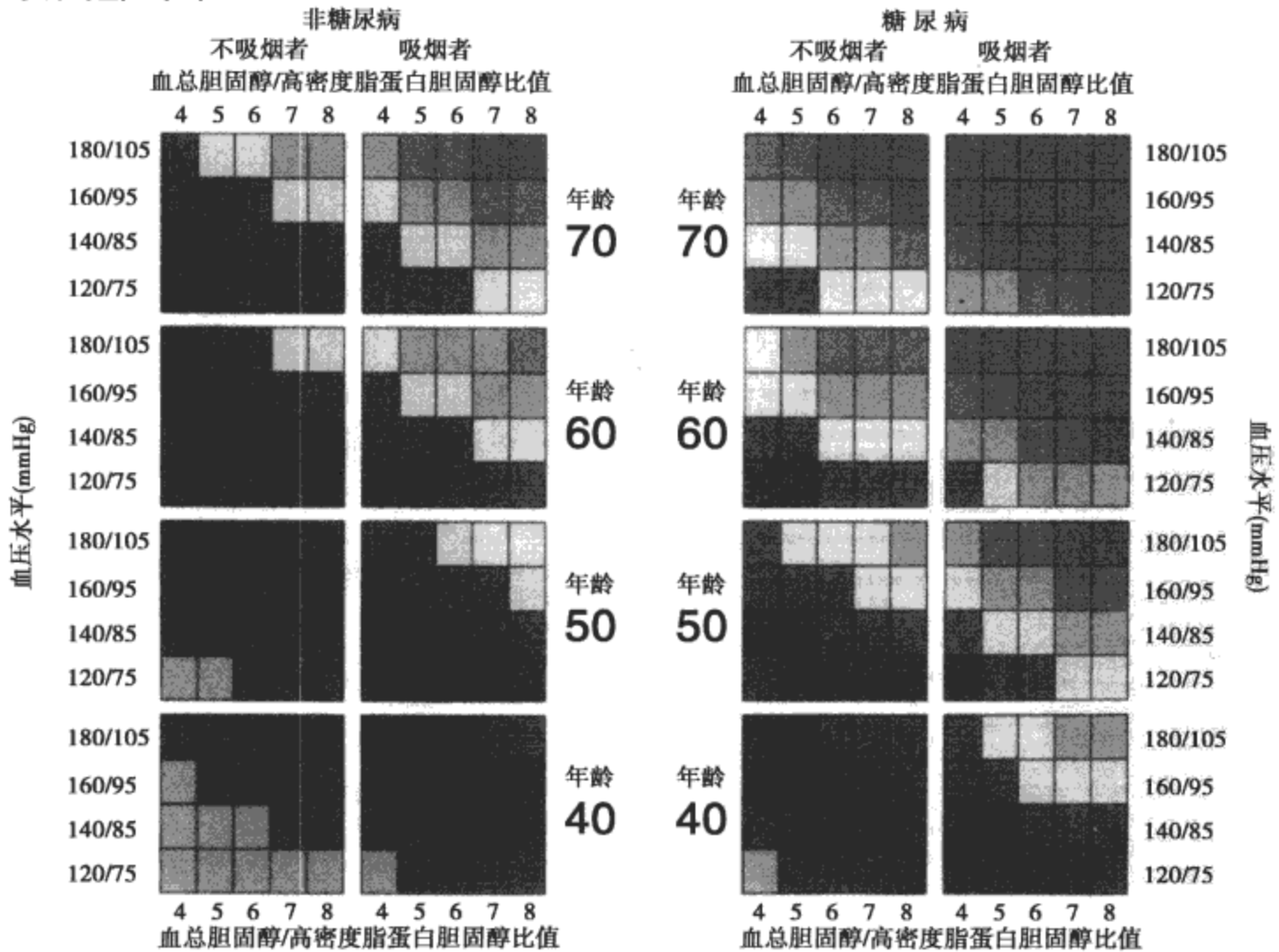


图7-2 心血管发病危险因素与治疗效益的估价

患者血清总胆固醇基线值下降20%或SBP下降10~15mmHg或DBP下降5~10mmHg,估计CVD事件危险度5年将下降1/3以上。

应用这份图表,查看该病例5年内发生CVD几率。该例为男性,有糖尿病、吸烟、血压150/94mmHg、TC/HDL=8.1。查图7-2,显示为红色小块,表示5年内发生CVD者>30%。高危性,应予积极治疗,如果能有效降低血压,控制糖尿病及高脂血症和戒烟,则可望发生CVD危险性会大大下降。其降低水平仍可查图确定。这将取决于对病人的干预及效果。

临床医学的研究工作,就是要获得对防病治病有价值的成果。然后又将其应用于防病治病的具体实践,达到有效的防病治病、保障人民健康的最终目的。但是,最佳证据的临床应用一定要联系患者的具体实际,必须执行个体化的原则与方法,并且要“对号入座”,才能达到防病治病的最佳目的。

(王家良)

## 第八章 循证医学实践的决策分析

循证医学强调任何临床诊疗决策的制订仅仅依靠临床经验是不够的,还应当基于当前最佳的科学研究成果,并充分考虑病人的意愿,只有三者有机结合制定的诊疗决策才是循证临床决策。但如何将循证决策的理念落实到具体的临床实践,保证决策的科学性和合理性,还需要科学的思维和手段,决策分析就是一种量化估计不同决策选项相对价值的方法,在循证医学实践中可以发挥重要的作用,尤其面对的临床问题复杂或者不确定时更为有用。

### 第一节 概 述

#### 一、决策及分类

决策就是指为实现一定目的而制定多个行动方案,并从中选择一个“最优的”或“最有利的”或“最满意的”或“最合理的”行动的过程,其本质是利用知识预测行动的可能后果。在实际决策中,知识有各种来源,既有经验性的也有理论性的,因此决策可以分为经验决策和科学决策,前者基于个人或群体的经验和直觉判断做出决策,相对比较主观;后者则强调在科学理论和知识的指导下,使用科学的方法或技术进行分析,从众多备选方案中选定最优方案的过程。毫无疑问,循证医学实践不排斥经验决策,但更倡导科学决策。

决策从不同的角度可以有多种分类,其中,按照决策问题具备的条件和决策结果的确定性程度可以将决策分为以下三种类型:

1. 确定型决策 是指供决策者选择的各种备选方案所需的条件都已知并能准确地知道决策的必然结果。

2. 风险型决策 是指对决策者期望达到的目标,存在着两个或两个以上的不以决策者的主观意志为转移的自然状态,但每种自然状态发生的可能性可以预先估计或可以利用文献资料得到,进行这类决策时要承担一定的风险。

3. 不确定型决策 是指决策者对各种可能出现的结果的概率无法知道,只能凭决策者的主观倾向进行决策。

#### 二、决策分析和临床决策分析

决策分析就是定量比较各种决策选择可能产生的后果和效应,从而使决策更为科学和合理的过程。决策分析包含决策主体、决策目标和一系列备选方案这三个基本要素。在临床诊疗决策制定的过程中,决策主体可以是患者,也可以是医生,但循证医学提倡医患双方共同决策的模式;决策主体对于决策问题所希望实现的目标,可以是单个目标,也可以是多个目标;备选方案要基于已有的知识做到尽可能没有遗漏。

临床决策分析(clinical decision analysis, CDA)是采用定量分析方法在充分评价不同方案的风险和利益之后,选取最佳方案以减少临床不确定性和利用有限资源取得最大效益的一种思维方式,包括诊断决策、治疗(康复)决策等。1959年, Ledley 与 Lusted 提出可以将决策分析应用于医学领域,



1967年,第一篇应用决策分析的学术论文发表,当时研究的内容是应用决策分析解决是否应对患有口腔癌而无明显颈部淋巴结转移的患者施行颈部根治术的问题。随后, Lusted 和 Kasirer 对于在医学领域中运用决策分析进行了更多的探讨。近些年来,决策分析在医学领域得到了越来越广泛的应用和发展,不仅可以帮助临床医师选择针对单个病人的治疗措施,也逐渐用于针对群体的医疗政策的制定,是一种最大限度地减少临床实践和决策失误的科学方法。

### 三、决策分析的步骤和常用方法

决策分析一般分4~5个步骤:①形成决策问题,包括提出方案和确定目标及效果量度;②用概率来定量地描述每个方案所产生的各种结局的可能性;③决策者对各种结局的价值定量化,一般用效用来表示。有了效用就能给出偏好;④综合分析和评价各方面信息,以最后决定方案的取舍;⑤有时还要对所取方案作敏感性分析。

对于不同的情况有不同的决策方法。①确定性情况:每一个方案引起一个,而且只有一个结局,这时使用确定型决策。当方案个数较少时可以用穷举法,当方案个数较多时可以用一般最优化方法。如果决策以收益为目的,那么取最大值的方案为最优方案;如果决策以减少损失为目的,那么取最小值的方案为最优方案。②风险性情况,由一个方案可能引起几个结局中的一个,但各种结局以一定的概率发生。通常在能用某种估算概率的方法时,就可使用风险型决策,例如决策树的方法。在风险型决策中,一般均按收益取最大值,损失取最小值的原则进行决策。③不确定性情况:一个方案可能引起几个结局中的某一个结局,但各种结局的发生概率未知。这时可使用不确定型决策,例如拉普拉斯准则、乐观准则、悲观准则、遗憾准则等来取舍方案。

## 第二节 决策分析的实施

决策分析的基本思想是将问题系统地分成几部分,确认各部分之中的不确定因素,然后通过回顾医学文献或者咨询专家意见得到各不确定因素的估计值,进而得到采取不同决策选项后,各种可能的结局出现的概率。具体实施时,包括以下基本步骤:问题的确认和界定范围;问题的结构化,通常采用决策树的方法;搜集相关信息,明确各种事件出现的概率;对最终结局用适宜的效用值赋值;决策树分析,计算每种备选方案的期望值;对结论进行敏感性分析。

### 一、问题的确认

决策分析的第一步是对所研究的问题给出精确的定义,并且根据问题的定义将其分解为几部分,第一部分通常是不同的决策选项,随之是不同选项之后所有可能发生的事件,而最后一部分则是决策所关心的结局。

例8-1 男性患者,50岁,体检发现左颈动脉硬化,但目前没有任何临床症状。现有证据表明,颈动脉硬化者发生脑卒中的危险性升高。因此,对于该患者是否需要治疗,就是一个决策问题,临床上可以有两种选择,一是暂时临床观察,二是行颈动脉内膜切除术。但是结合临床实际考虑,有如下可能,临床观察虽然避免了手术相关的短期危险因素(围术期死亡,手术中发生脑卒中),可以维持无症状性颈动脉硬化状态(在一段特定时间内未发生脑卒中),但是不可避免面临将来更高的脑卒中的风险。如果选择进行手术,手术虽然有益于解决问题,减少发生脑卒中的可能性,但是却有围手术期间发生脑卒中和死亡的风险。

### 二、问题的结构化

绝大多数决策分析都采用决策树(decision tree)的方法对问题进行结构化。决策树是一种能够有效地表达复杂决策问题的数学模型,按逻辑、时序把决策问题中的备选方案及结局有机组合并

用图标罗列出来,犹如一棵从左到右不断分支的树,包括一系列节点与分支。节点又可分为决策节点(decision nodes)与机会节点(chance nodes),前者以小方框“□”表示,后者以圆圈“○”表示。决策节点处分成可为决策者所控制的几个决策选项,而机会节点处则分成决策者所不能控制的一个或几个相应事件。在决策树末梢,为各方案的最后结局(outcome),各种结局必须定量描述。在每一个机会节点,其后相应事件的概率之和必须为1,也就是说,每个机会节点之后的事件必须涵盖所有可能的情况,只有这样才能保证分析的有效性。对于结局而言,结局可以是生或死,也可以是其他治疗可能带来的任何收益或风险。

基于上述原则和对临床问题的分析,例8-1可以绘制如图8-1的决策树。

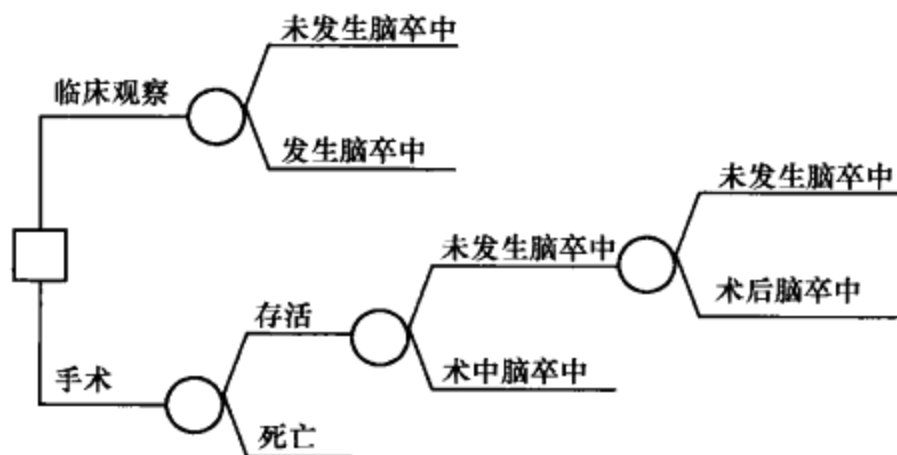


图8-1 动脉硬化处理的决策树

### 三、搜集相关信息

在决策分析中,最重要的信息是每一种可能的事件在相应条件下出现的概率。相关信息可以来自以下几方面:

#### (一) 通过文献估计概率

应用文献估计概率时,首先应当评估所用文献的有效性,来自大样本的临床试验或随访研究的概率最为可靠,但即使是大样本的研究也会存在抽样误差,因此在确定基线概率时,应根据情况确定概率的可信区间,产生一个包括最佳概率估计值(称为基线概率)、概率的上限和下限以及资料来源的表格。

根据情况不同,信息可以从一篇文献获取,也可以根据多篇文献的综合结果得到。信息单一来源大致有如下几种情况:所应用的文献是该方面唯一可以获得的数据资料;该文献是权威性的结果;研究的质量较高或者样本量很大或者该文献所代表的人群即为此次决策分析结果所要应用的人群。因此,在把单一文献估计的概率用于决策时,如果文献与我们决策问题的处理背景不同,或者病人情况有差异等,要尽量引用文献中分层分析提供的结果,然后根据经验、知识或专家的意见,参考其他研究资料,修改概率使之更符合实际要求。

如果关于某个问题有多篇文献的数据结果,可以根据情况对文献结果进行综合。推荐使用Meta分析从多篇文献中获取信息,但纳入分析的文献必须真实可靠,才可以提供比较接近真实情况的分析结果。

#### (二) 咨询专家意见以获取所需的信息

制定政策时,通过专家和专家小组可以获得所需的信息。在决策分析中也可以通过这种方式获取所需概率的估计值。但这种方法只可以在其他获取信息的方法都不可行时使用,而且任何通过这种方法获得的决策分析的结果都必须进行敏感性分析。

咨询专家小组的意见可以通过德尔菲法进行,专家们提出对所需数据的估计值,然后通过讨论和反馈对该值进行修订。

### (三) 概率估计中不确定性的估算

大多数决策分析都需要估计很多事件发生的概率,当决策模型很复杂时,某个估计中很小的偏差都有可能引起最后结果很大的改变,因此,估算概率估计中的不确定性对分析结果的影响是很重要的。

敏感性分析是估计这种不确定性最常用的方法,它可以评价概率估计的不确定性对分析结果的影响。然而通过敏感性分析同时进行多个变量的不确定性估计很困难,结果也难以理解。因此,学者们提出几种统计方法来评估同时考虑所有变量时的不确定性,例如完全贝叶斯分析以及归因分析等。

上述获取信息的方法中,通过文献估计概率最为常用。有些时候在相关的文献中无法得到所需概率,这时分析的进行常常需要依赖于专家的概率估计结果。当然也可以通过调查研究直接获得相关问题的原始资料,这样得到的结果可能很准确,但人力物力花费可能较大。另外,也有学者提出在临床应用中,可以根据个人经验“猜测”有关的概率。

对于决策分析概率的估计,要求做到以下几点:首先,用于估计概率的信息来源应该有详细的记录,估算概率的方法要有详尽的描述;其次,对于任何一个需要估计的概率,都要求尽可能采用最可靠的概率估计值;最后,在有多种途径获得所需信息时,需要在不同来源之间做出合理的选择。

获取所需要的信息后,将其填入决策树中,就可以进行下一步的分析工作。

例8-1中的决策相关信息的基线估计值如下:选择临床观察的患者有11%可能会发生脑卒中,而选择做手术会有1%的患者死亡,而术后存活的患者中有2%会在围手术期间发生脑卒中,术后“未发生脑卒中”者在以后会有5%的患者再次发生脑卒中的可能,这些估计值在决策树中各概率支下方标出(图8-2)。

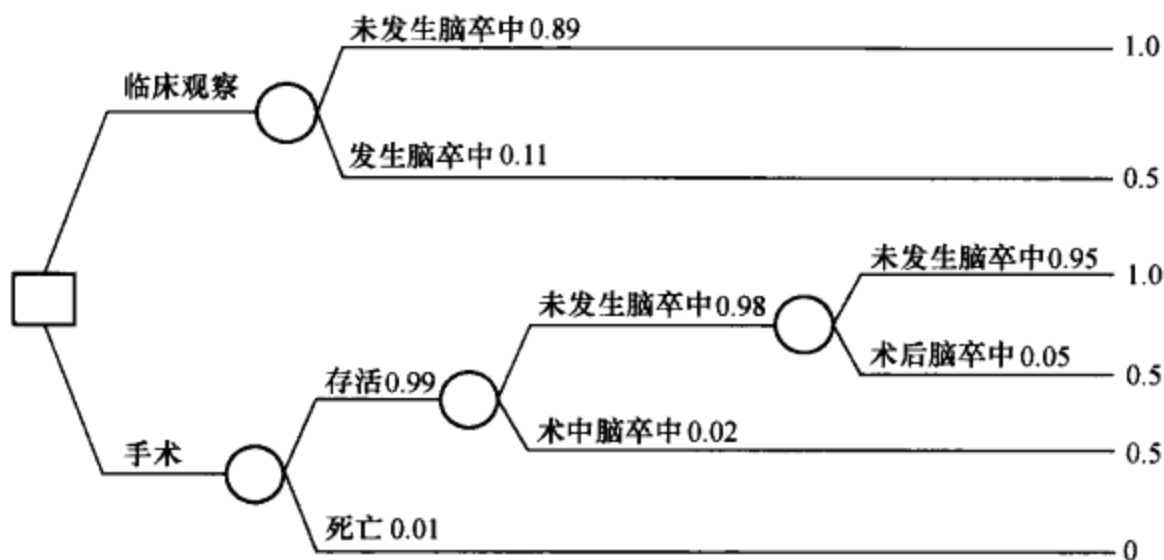


图8-2 手术与临床观察的决策树

### 四、确定最终结局的效用值

决策过程中结局的量化是决策分析的重要环节。因为决策选项有多种,每种选择的临床结局又不相同,可以是简单的二分类结局,如生存与死亡;也可以是单一属性的多项结局,如完全缓解、部分缓解、稳定、恶化;还可能是复杂的多种属性的多种结局,如一方面是生存时间的不同,另一方面生存质量也各异。因此,为了达到比较目的,需要预先将各种结局转换成同一单位的变量,这个变量就是效用(utility)。而效用值是一种表述结局相对优劣的数量化指标,通过对病人或公众进行调查获得,可以测量人们对不同健康状态的喜好程度,它是个体在不确定情况下的优先选择,表现出他们对某种健康状况的倾向和偏爱,反映了个体的主观感受,并受年龄、经济收入、教育程度等多种因素的影响。效用值通常用0~1的数值来表示,1代表完全健康,0代表死亡,也可以为负数,表示比死亡更糟糕的疾病状态,如无意识或长期卧床伴严重疼痛等。

### (一) 效用值的直接测量方法

1. 等级尺度法(rating scale) 是测量效用值最直观的方法,通常由一段直线构成,两端分别标注0和1,并有简短的描述性短语,0表示最差的健康状态比如死亡,1表示完全健康,线段上标有等间距的刻度。患者可以在线段上最能说明自己健康状态的地方做一标记,测量标记处距0端的长短,由此获得的数值即为该健康状态的效用值。例如多发性硬化是一种渐进性进展性疾病,日常生活和行为处于较差的状况,被测试者在接近0端的0.4处做标记,则该健康状况的效用值为0.4。假设死亡并不认为是最差的状态,而是在量表上的某一点,比如 $d$ ,那么其他状态的效用值可以用公式 $(x-d)/(1-d)$ 来换算( $0 < d < 1$ ,  $x$  是其他状态的值)。等级尺度法测量效用值直观,易操作,在临床上应用广泛,但该方法重测信度不高。

2. 标准博弈法(standard gamble) 是测量效用值的经典方法,已经广泛用于不同领域的决策分析,其基于效用理论的基本原理,即面对疾病被调查者有两种选择,一种是选择治疗,但治疗的结局有两种可能性:①病人完全康复,再健康生存 $t$ 年(概率为 $p$ );②当即死亡(概率为 $1-p$ );另一种选择是某种慢性疾病状态 $i$ ,生存 $x$ 年( $x < t$ )后死亡。通过对比提问法,来确定 $p$ 值。例如某种疾病手术的最好结果是手术成功再生存10年( $t$ ),最坏结果是手术失败立即死亡,如果不手术则可以维持目前的疾病状态生存5年( $x$ )。当手术成功的概率为30%( $p$ )时,患者宁愿选择不手术而生存5年,当成功的概率升至50%时,患者仍然不愿意冒险选择手术,随着手术成功概率逐步上升,假设当成功的概率为75%时,患者愿意改变主意选择手术,此时患者就在手术和不手术两种选择中保持中立,该疾病状态的效用值即为0.75。因为概率是一种比较抽象的概念,调查对象对概率的比较和动态变化常难以理解,因此在实际操作中可以使用一种叫做概率轮(chance board)的工具。这是一种由两个可以调节的部分组成的盘,采用两种不同的颜色区分,两部分的相对大小可以改变,两种颜色的比例和相应结局的概率相当。先将各种选择结果写在卡片上告诉调查对象,然后转动概率轮,两种不同颜色的比例随之改变,对应各种结局的概率也相应变化,借此可以帮助调查对象理解概率的变化并选择所偏好的结局。

3. 时间权衡法(time trade-off) 是直接对两种不同的状态作等量估计,让病人在“好的健康状态但活的时间短些”与“目前的疾病状态但活的时间长些”间作出选择,其要点在于患者愿意出让多少时间的寿命以换取更好的健康状态。比如一种是在慢性疾病状态 $i$ 下生存 $t$ 年( $t$ 为慢性疾病的期望寿命)之后死亡,另一种是接受某种治疗从而健康生存 $x$ 年( $x < t$ )之后死亡,时间 $x$ 是变化的,直至患者在两种选择中保持中立,此时 $i$ 状态的效用值即为 $x/t$ 。例如心绞痛患者如果不治疗可在疾病的状况下生存25年( $t$ ),有一种治疗方法可使心绞痛完全缓解但要减少生存年数,当生存年数缩短到14年( $x$ )时患者选择不治疗,但如果缩短到15年患者就愿意治疗,此时没有心绞痛的15年就相当于有心绞痛的25年,心绞痛的效用值计算为 $15/25(x/t)=0.6$ 。

### (二) 效用值的间接测量方法

除了用上述直接法测量效用值外,还可以用间接法进行测量。所谓间接法是因为其并不直接测得某种疾病状态的效用值,而是通过填写生存质量量表,如SF36量表等,计算量表得分,再将量表得分转换为效用值,即效用值=(量表实际得分-量表的最低得分值)/量表的得分范围。此外,还可以计算质量调整寿命年(quality-adjusted life years, QALYs),用QALYs来测量各种各样结局,使之有通约性,简化复杂的临床决策问题。

例8-1中应用质量调整寿命年(QALYs)来衡量结局,设定患者无论是保持未发生脑卒中的“无症状性颈动脉硬化状态”还是有病生存(伴脑卒中状态),他们的期望寿命都是15年,效用值(或生存质量)分别为:未发生脑卒中状态效用值=1,生存(伴脑卒中)效用值=0.5,死亡效用值=0,相应数据标在图8-2的最右侧。进一步计算他们的QALYs分别为15、7.5年和0年,获得这些资料后就可以进行下一步的工作了。

## 五、决策树的分析

所谓决策树的分析又称为决策树的折叠与平均,即指通过以上步骤所形成的决策树,计算每一种决策选项所对应的不同结局出现的情况。

例8-1在掌握了以上各种基本信息后,就可计算各机遇节点和各种方案的期望值(expected value, EV)。各机遇节点的期望值为该节点各分支概率与结局 QALYs 值的乘积之和。根据上面的决策树可以计算临床观察枝的期望值和手术枝的期望值如下:

临床观察的期望值:  $EV_c = (15 \times 1 \times 0.89) + (15 \times 0.5 \times 0.11) = 14.18 \text{ QALYs}$ 。

手术的期望值:  $EV_s = (15 \times 1 \times 0.95 \times 0.98 \times 0.99) + (15 \times 0.5 \times 0.05 \times 0.98 \times 0.99) + (15 \times 0.5 \times 0.02 \times 0.99) + (0 \times 0 \times 0.01) = 14.34 \text{ QALYs}$ 。

比较两者的期望值,手术治疗的期望值稍高,提示应该选择手术治疗,但是应该看到,两个方案的期望值相差不大,而手术治疗还有致死的可能性,虽然死亡的概率很低,但是在生与死的可能选择的情况下,医生或患者难以仅凭  $14.34 \text{ QALYs} > 14.18 \text{ QALYs}$  的结果,承担那1%的死亡风险!因此,与患者及时沟通,在知情的情况下做出决策是十分必要的。

## 六、敏感性分析

医学现象是千差万别的,一个模型建立后,可能会出现某些条件的变化,这时就会出现模型是否实用,模型的结论是否可靠的问题。例8-1中,如果认为选择做手术的病死率(1%)太高或是太低;或者1%这个概率是大概估计的,不够准确;抑或是一个区间值,不能肯定就是1%,因此,手术病死率不同,将影响到最终的决策。此时就需要进行敏感性分析(sensitivity analysis)。

敏感性分析通过观察假设变化时,决策分析的结果是否具有良好的稳定性,对其结论加以评估。当研究的结论具有较好的稳定性时,说明研究结论的可靠性较好。同时敏感性分析也可以通过观察哪些假设对结果影响最大而发现对于该分析最重要的假设,从而为问题的深入研究和进一步解决提供线索。如果一个变量在合理的范围内变动对结论无明显影响,我们就说结论对这个变量“不敏感”,反之,我们就说结论对此变量“敏感”。

这里就对例8-1进行敏感性分析,当取不同的手术病死率时,可分别计算出不同的手术治疗方案的期望值,下图显示的就是当手术病死率在一定范围内变动所对应的期望值。从图8-3中可以看出,当手术病死率为2.1%时,两种方案的期望值相当,此时的2.1%叫阈值,表明这一数值是两种方案选择的分界点。当手术病死率在小于2.1%的区间内,选择手术这一方案是合适的。

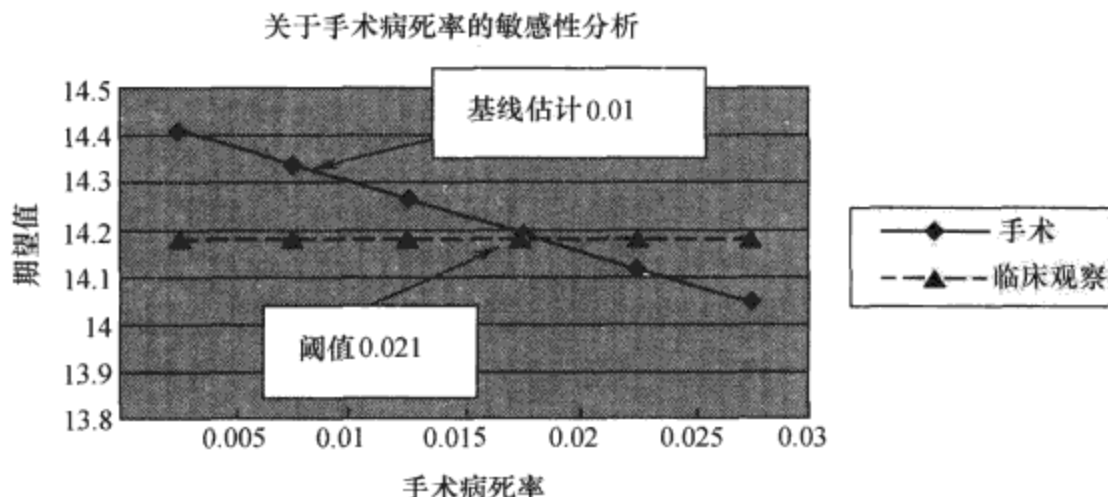


图8-3 对手术死亡率的敏感性分析

如果对伴有脑卒中生存状态的效用值也存在疑惑,同样可以对其进行阈值分析和敏感性分析(图8-4)。

关于伴脑卒中生存效用值的敏感性分析

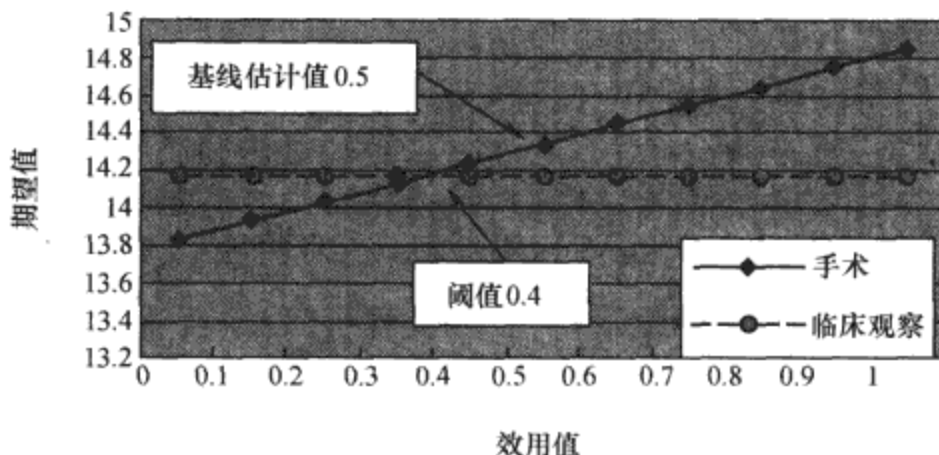


图8-4 对伴脑卒中生存的效用值的敏感性分析

敏感性分析在决策分析中具有重要的地位,它可以验证决策分析结果的可靠性,因此决策分析应该常规进行敏感性分析。

### 第三节 复杂决策问题和MARKOV模型

#### 一、复杂决策问题

例8-1只是一个简单的决策分析例子,仅有两个决策选项,而且在整个分析中需要估计事件的发生概率也只有几个。实际的问题可能会复杂得多。在医疗决策中,可供选择的决策选项可能很多,同时可能会出现多种结局,而且从干预措施到结局出现可能会经过许多过程,这些过程又有可能受很多因素的影响。在这些情况下,决策分析就变得比较复杂。在例8-1中,如果要考虑事件发生的时间因素,比如各种干预措施的远期效果,就成了复杂的决策问题。如手术存活的患者在接下来的生命过程中,不可避免的会有部分患者会再次发生脑卒中、死亡;存活的患者仍然有再次发生脑卒中、死亡的可能;对于临床观察方案同样存在类似的问题;而且相关概率也随着研究人群年龄的改变发生了改变,如50岁年龄段人群与60岁年龄段人群的死亡率就不一样,那么如何评价这两种策略的远期效果呢?本节将介绍一种有助于解决此类问题的模型: Markov 模型。

#### 二、Markov模型

疾病的预后可以描述为一系列变化的事件,病人在这一系列的事件中具有不同的危险度。例如,无症状胆结石患者可能会在任何时间急性胆囊炎发作,而如果胆囊炎发作时病人需要进行手术,那么手术的风险随病人年龄的增加而增加。如果要对无症状胆结石患者是否立即进行胆囊切除术进行决策分析,对于不进行胆囊手术的选项,按照前述建立决策树的原则,建立一系列选择节点分析患者的可能情况(正常、发生急性胆囊炎、手术死亡以及死于其他原因),即使仅仅分析患者从当前至以后30年的情况(这是一个合理的时间区间),也需要建立非常庞大的决策树,对这样的决策树几乎不可能进行分析。

可以想象,这种树形结构的很大一部分都是重复的。近十余年来,在这种问题的决策分析中,人们越来越倾向于使用 Markov 模型。Markov 模型能够更精确而且非常简洁的表示在不同健康状态间多次的或者时间不确定的转换,从而使对这种事件的决策分析能够进行。

在决策分析中,应用 Markov 模型需要四个步骤:

第一步: 定义不同的健康状态,描述在不同的健康状态之间转换的方式。

例如,对上述无症状胆结石的例子,可以认为有四种状态: 无症状胆结石、急性胆囊炎、急性胆囊炎术后和死亡。这几种状态之间的转换可以用图8-5来表示。

图中用椭圆或圆形表示不同的状态,用箭头表示不同状态之间的转换。

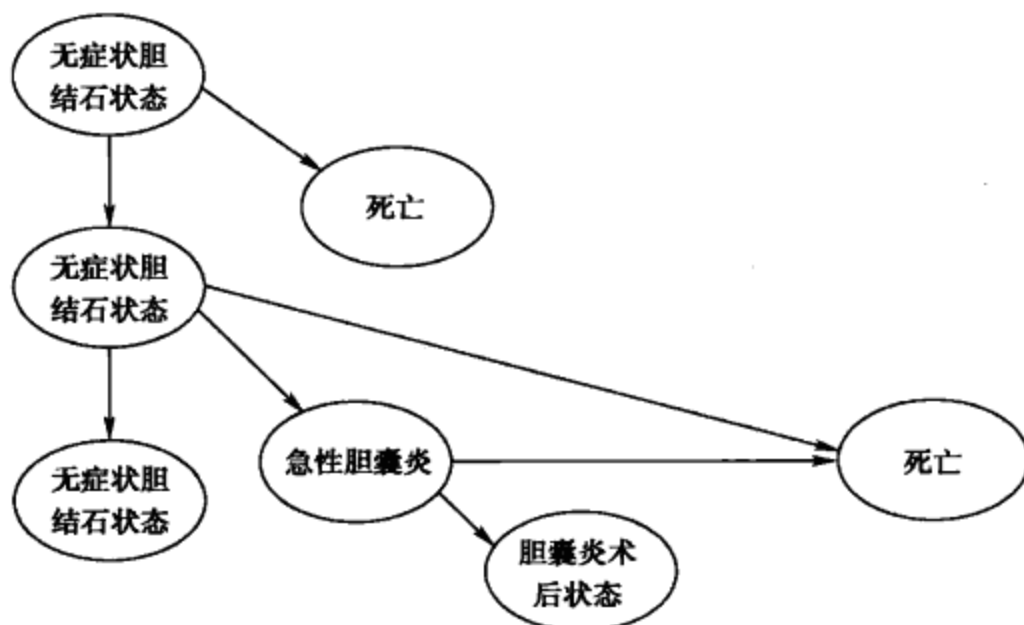


图8-5 对无症状胆结石患者不进行胆囊手术的 Markov 模型

第二步: 选择适宜的周期长度。选择计算的周期长度要根据实际情况决定。从计算上来说, 周期较长者, 计算的负荷较小。

第三步: 估计转换的概率。估计转换概率的大多数信息来源于文献。但是文献中的信息大多是用转换的率( $r$ )而不是用概率( $P$ )表示。率通常是用人时为单位的, 可以用率来计算在所确定的时间长度内的转换发生的概率, 计算公式为:

$$P=1-e^{-r}$$

此处  $e$  为自然对数。

第四步: 估算结局。Markov 模型的结局通常使用预期寿命表示。估计 Markov 模型预期寿命的方法有三种: Monte Carlo 模拟法 (Monte Carlo simulation)、Markov 队列模拟法 (Markov cohort simulation) 和矩阵代数法 (matrix algebra)。其中 Markov 队列模拟法比较简单, 它建立一定数目的虚拟队列人群, 根据转换的概率, 估算在每个周期内发生转换的人数, 直至整个队列的所有人“死亡”, 进而通过寿命表的方法估计其预期寿命。表8-1即为上例的 Markov 队列模拟法估计过程。

表8-1 用Markov队列模拟法估算预期寿命

年数	患者年龄	无症状胆结石者	患急性胆囊炎者	急性胆囊炎术后者	死亡人数
0	30	100000	0	0	0
1	31	98838	984	0	168
2	32	97676	982	992	349
3	33	96516	970	1971	544
4	34	95536	958	2934	752
5	35	94198	945	3883	974
10	40	90698	452	6600	2251
20	50	83936	206	8582	7277
30	60	71403	171	9264	19180
40	70	50509	114	7906	41472
50	80	27785	53	5061	67100
60	90	867	1	178	98954
70	100	0	0	0	100000
总计	—	3788787	13571	380022	—

整个队列的人年数为:3788787+13571+380022=4182383人年。

故平均预期寿命为41.82年。

其他两种方法比较复杂,在此不作过多介绍。感兴趣的读者可以阅读 Beck 和 Pauker 以及 Sonnenberg 与 Beck 的专著。

### 三、左颈动脉硬化是否手术治疗远期效果评价的Markov模型

例8-1如果考虑远期效果时,可以采用 Markov 模型进行决策分析。图8-6显示了观察或手术治疗后患者出现的三种结局:维持无症状性颈动脉硬化状态,发生脑卒中和死亡,这是短期结局,那么随着时间的推移,无症状性颈动脉硬化状态的患者在接下来的时间里,会有一部分人死亡,一部分人生存,生存的人中间不可避免的会有人发生脑卒中,余下的人维持无症状性颈动脉硬化状态;生存但是伴有脑卒中的人,其结局也是随着时间的进行部分人会死亡而部分人仍生存,生存的人继续这种循环,直至所有人群都到达一种结局:死亡。

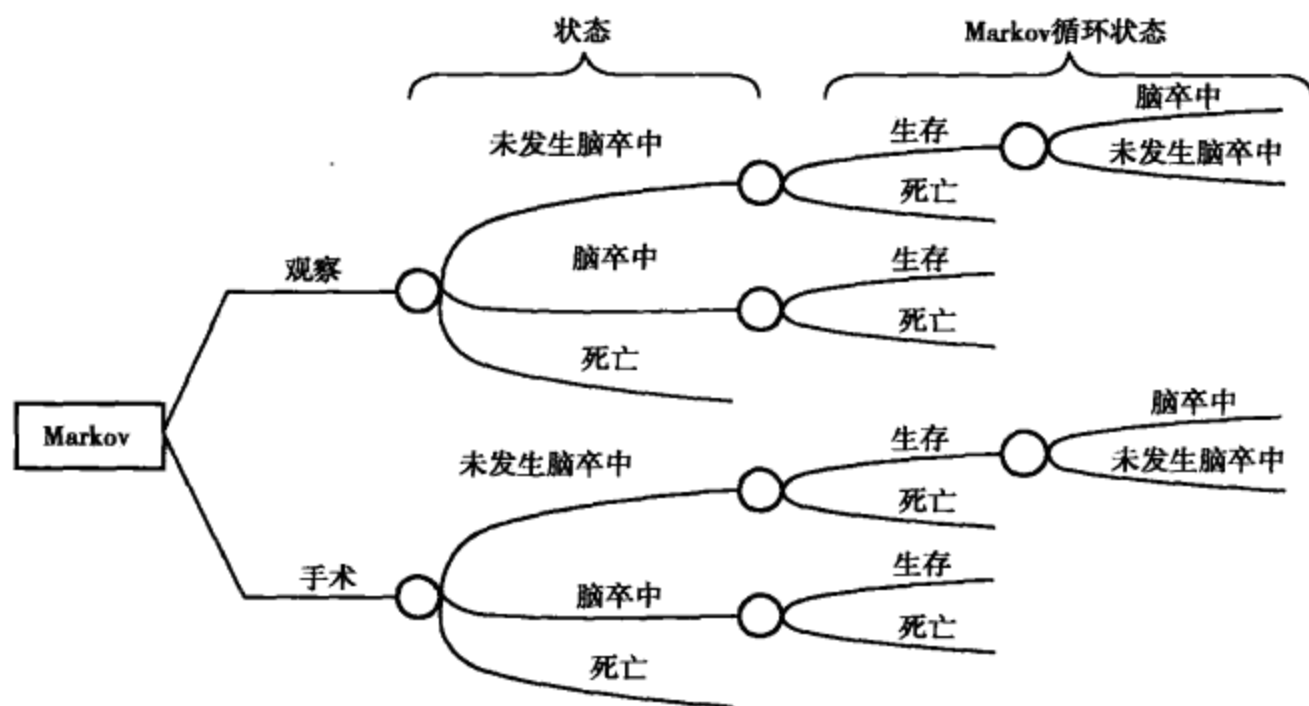


图8-6 颈动脉硬化患者的 Markov 模型示意图 (Markov 树)

用 Markov 模型来模拟他们的结局,计算结局的增益,同其他方案进行比较从而做出选择。可以看出有三个 Markov 状态:无症状性颈动脉硬化(状态1),伴有脑卒中生存(状态2),死亡(状态3)。其中状态1到状态3之病死率(5%)并不完全表示为颈动脉硬化这一疾病所致(要尽可能模拟实际情况,考虑有其他死亡原因)。它们之间的转换关系及转移概率见表8-2和图8-7。

取 Markov 循环周期为1年,那么每一个周期,“无症状性颈动脉硬化状态”为1个 QALYs,伴脑卒中生存则为0.5个 QALYs,死亡则为0。假设从手术开始当年起,根据相关文献数据估计选择手术治疗的相关数据为维持无症状性颈动脉硬化状态占97%,伴脑卒中生存的占2%,死亡为1%。此后人群在这3个状态的转移概率见表8-2。

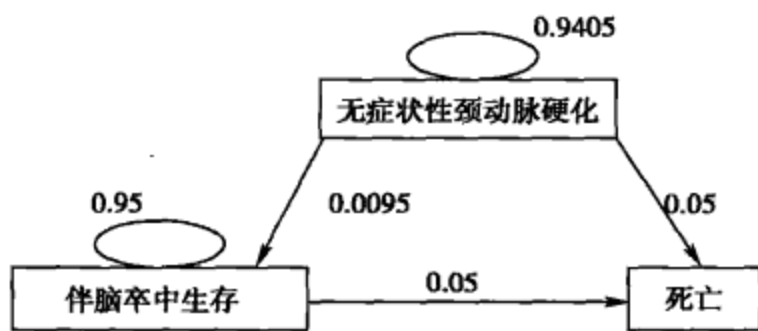


图8-7 颈动脉硬化患者各状态间的转移关系

表8-2 手术治疗后3种自然状态每年的转移概率

	无症状性颈动脉硬化	生存(伴脑卒中)	死亡
无症状性颈动脉硬化	0.9405	0.0095	0.05
生存(伴脑卒中)	0	0.95	0.05



计算原理同上,由于这一过程的计算比较麻烦,可以借助相关软件进行,如 decision maker, treeage, 比较简单的则可以运用 Excel 计算。以选择手术治疗为例,假设10万病例纳入分析,其最终结果是人均获得16个 QALYs(计算的结果见表8-3)。用同样的方法可以计算选择“临床观察”方案的人均 QALYs,然后可以比较两个方案的 QALYs。

表8-3 10万例无症状性颈动脉硬化手术治疗的Markov模型分析结果

周期数	无症状性颈动脉硬化		生存(伴脑卒中)		死亡		总 QALYs	累计 QALYs
	例数	QALYs	例数	QALYs	例数	QALYs		
1	97000		2000		1000			
2	91228	91228	2822	1411	5950	0	92639	92639
3	85800	85800	3548	1774	10652	0	87574	180213
4	80695	80695	4786	2093	15119	0	82788	263001
5	75893	75893	4744	2372	19363	0	78265	341266
6	71377	71377	5228	2614	23395	0	73991	415257
7	67130	67130	5645	2822.5	27225	0	69953	485210
8	63136	63136	6001	3000.5	30863	0	99137	551347
9	59379	59397	6301	3150.5	34320	0	62530	613887
10	55846	55846	6550	3275	37604	0	59121	672998
11	55523	52523	6754	3377	40723	0	55900	728898
50	0	0	0	0	10000	0	0	1599600
总计:								1599600
人均:								16.00

注:①无症状性颈动脉硬化状态效用值=1,生存(伴脑卒中)效用值=0.5,死亡效用值=0; QALYs: 质量调整寿命年。②本例数据引自文献: Decision analysis models: Opening the black box. Surgery, 2003, 133(1):1-4

#### 第四节 决策分析质量评估

评估决策分析的质量实质上是评估决策分析中所使用的各种原始参数的准确性和可靠性,主要标准如下:

1. 决策树各分支所用的参数是否来源于高质量的科学研究?
2. 如果使用了效用指标,效用的赋值是基于相关病人群体的调查、健康人群的调查,还是研究者个人的粗略估计?
3. 是否进行了敏感度分析,即通过改变以下四个方面的估计值,评估最终结论的稳定性?
  - (1) 各种干预的效果;
  - (2) 不良反应发生率;
  - (3) 估计的费用;
  - (4) 效用的赋值。
4. 是否考虑了所有可能涉及的花费?

总之,决策分析作为循证临床决策的辅助技术,其应用价值已经得到大多数学者肯定。但临床决策是复杂的过程,有些问题很难结构化成决策树;获取真实可用的临床信息也不容易;简便直观的效用值估计方法也有待改进;相信随着决策分析理论和方法学研究的进步,尤其是计算机辅助决策系统的开发和利用,决策分析必将在医学实践中发挥其应有的作用。

(詹恩廷 洪明晃)



## 第二篇 循证医学证据的分析与评价

### 第九章 疾病的病因与危险因素的循证分析与评价

循证医学的临床实践,对病人所患的疾病应从病因或危险因素方面寻求科学依据,以便提供科学的诊断和防治决策的证据。否则,有可能导致误诊、误治乃至误防。因此,对患者所患疾病之发病的原因(病因),或导致发病的有关危险因素的探讨和求证,有着非常重要的意义。

#### 第一节 临床实践:探求病因

当患者病后,首先要到医院诊治,临床医生接诊后,要仔细地采集患者的现病史、既往病史、家族史以及流行病史;并且要认真地体格检查以掌握重要的阳性与阴性体征;继而有针对性地选择进行相关实验室或特殊检查。根据其重要的阳性结果,结合病史和重要的临床体征,医生应用自己的医学知识、临床经验和技能,对患者的疾病作出临床诊断(包括病因学诊断)和治疗决策。

然而,人们的认识无不受科学技术水平限制,加上疾病多种多样、表现各异、病情复杂且受社会、心理等诸多方面的影响,对于某种(些)特殊的、异常的乃至突发性的卫生/疾病事件,往往一时间作出正确的临床或病因学的诊断会有很大困难。在临床方面,往往会采集不同范围和水平面的会诊,邀请一些相关学科的专家,集思广益,提出诊断假设(包括病因)和试探性相关防治措施的决策以及进一步探索或研究病因。

在临床医疗实践中,对未知病因疾病的探索和研究,一是对特殊病例提供详细的临床病理报告或系列病例报告,根据临床和病理(含尸检资料)的特点进行详细陈述与分析,在此基础上提出病因假设;二是针对所提出的假设病因,收集历史资料,作回顾性的病例对照研究(case control study)以及回顾性的队列研究以验证可能的病因;如获较为可靠病因价值的结论,则进一步进行前瞻性的队列研究或随机对照研究以证实有关的病因假设。这里仅简述从临床实践中探索不明病因的有关过程与方法。关于如何具体的研究则请参考临床流行病学的病因学研究与评价的内容。应指出的是对任何不明病因之认识,几乎总是从临床发现开始并进一步深入探索、升华认识、直至获得真实的结论。

举两例证明临床医疗实践对探索有关病因与危险因素的意义。

例9-1 在西欧20世纪60年代,怀孕妇女因妊娠呕吐,服用沙利度胺(thalidomide,反应停)其止吐效果明显,因而被广泛应用。后来有报道出生的婴儿发生双上肢畸形,类似海豹肢故称之为先天性海豹肢畸形。根据产妇产前没有病毒性感染史(如风疹病毒等)、也没有这种婴儿畸形的家族史。经医生分析:提出了与服用“反应停”有关的致畸病因学说。当该报道一发表,相继有雷同的报告出现。故引起了医学界和社会的巨大反响。于是临床医生根据这一假说,相继进行了系统的病例报道、病例对照研究及回顾性队列研究,验证了这一病因的假说。最后,由生产“反应停”的药厂召

回了该药,并停止了生产。经追踪3年后(1960~1962年)。随着该药销售量的急剧下降。发生海豹肢畸形的婴儿数量也相应下降,最后几乎等于零(图9-1)。这就证明了反应停是导致婴儿畸形的病因。

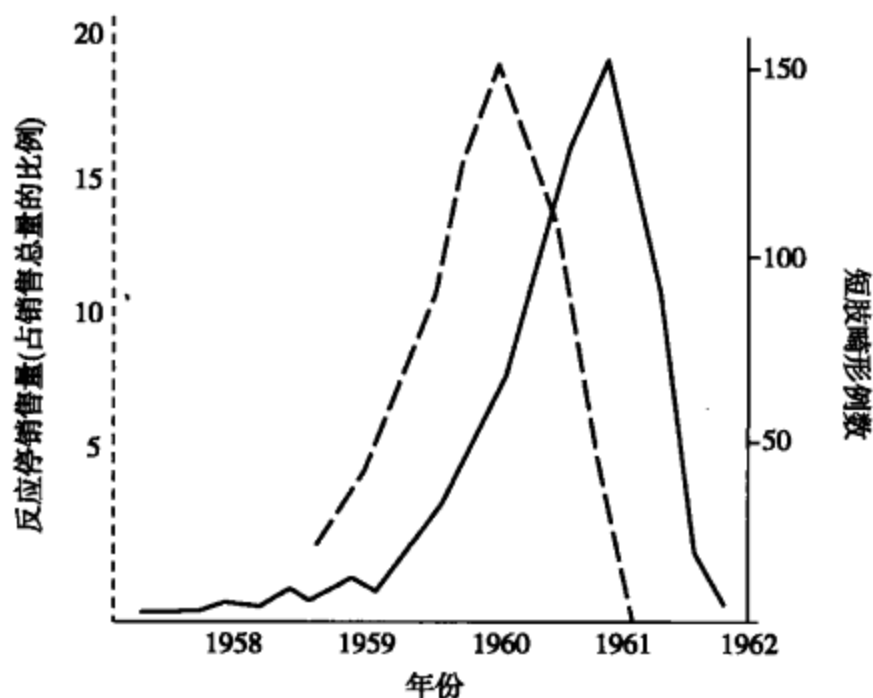


图9-1 反应停致婴儿先天性海豹肢畸形生态学研究结果

例9-2 2003年春在广东省某地发现了一些同期发病的急性呼吸道感染的病人,有的病情甚为重笃、肺部感染损伤严重、常规痰培养并未发现致病性病菌,一般的生物、毒物化验也无特殊发现,抗生素治疗缺乏效,病死率高。这种疾病显现传染力强,有着广泛的流行性。据报道,有一位患者后被称为“毒王”,跑到多个省市,感染多人且不幸病死者。我们的医护人员为抢救这种病人,有的被感染而光荣殉职! 由于对该病之病因不明、感染性强、波及多国,在国内、国际一度实行严格交通管制,对病人严格隔离,发现病人的“疫区”实行封锁、限制公共娱乐场所活动。对人居环境天天施行不同规模广泛消毒等。对人民和整个社会造成了极大的影响和损失! WHO 确定国际一流的医学检验中心对该病的病因病原体进行国际协作研究,终于通过国际广泛合作,弄清了引起该病的病因是一种冠状病毒变异株。对于这种疾病定名为急性传染性非典型肺炎或急性重度呼吸道感染综合征,简称“非典”(severe acute respiratory symptom, SARS)。这场非典在我国的暴发流行所造成的严重社会和经济损失以及对人民的生活安定的影响,确实相当于“全国没有硝烟的战争”! 这个例子可以看到尽管首先发现于临床,通过病例诊治明确了该病的临床和病例损害以及流行病学的特点。然而,它的不寻常的病因确定,是要多专业和学科的广泛合作,特别是国际大协作才能明确病因,由此制订病原学、临床诊断标准与方法和防治的科学策略。

从上例启示,无论如何,当临床发现某种用现有知识与方法难以解释的疾病和它的病因时,作为临床医生一定要有高度的敏感性,务必收集和掌握真实可靠的临床资料和证据,充分应用当代最新的医学科技知识、自己的聪明才智以及临床技能,敏感发现和提出问题,用充分的证据去验证求实,方可不断地创新、弄清和解决问题。

## 第二节 可供探讨与验证病因或危险因素的模型

疾病发病的病因或危险因素是十分复杂的,其因-果效应也受着多种因素影响,故有的人易感,也有些人因内外环境协调、免疫反应健全却十分健康。因此,在临床循证医学实践中,为求证某一疾病的病因,有以下三种在流行病学实践中总结出来的病因模型甚有价值。

### 一、流行病学三角模型

病因致病是一种因果关系。用简洁的概念关系来表达病因因果关系的模型为病因模型。对病因因果关系的不同理解或侧重,则有不同的病因模型。

疾病发生的三角模型又称流行病学三角( epidemiology triangle )。该模型认为疾病的发生要有病因(病原体)、机体(或宿主)及环境三个要素。疾病是否发生取决于这三个要素协同作用的结果。三者之间保持平衡状态,疾病就不会发生;如三者中的一个要素发生变化,失去平衡,则疾病就会发生(图9-2)。



图9-2 流行病学三角

### 二、轮状模型

又称车轮模型( wheel model )。此模型强调病因来自宿主及环境两方面。宿主处于环境包围之中,类似轮状。环境包括生物的、理化的和社会的环境,宿主是遗传作用的内核(图9-3,彩图9-3)。各部分的大小是有伸缩性的,以不同疾病而异。若环境因素作用大,则图中的外环可变大些。若是遗传病,则表示遗传作用的内核要变大些,环境“环”相应缩小些。

### 三、病因网模型

病因网模型( web of causation )将多种病因按时间顺序连接起来,形成一个病因链;多个病因链交错连接起来形成网状,因此称病因网模型。该模型可以提供完整的病因因果关系路径。例如,肝癌的病因网可以看成有三条病因链交错形成(图9-4,彩图9-4),三条链起始端分别为乙肝病毒感染、黄曲霉毒素污染食品和饮水中的藻类毒素。三条链与多个因素相互交错、相互作用最终导致肝癌的发生。该模型的优点是表达具体、清晰、系统,能很好地阐述复杂的病因因果关系。

这些模型图及其列出的若干因素可供分析与寻求有关病因时参考,当参考时,一定要带着自己的问题,应用可靠的信息联系实际,扩大视野与开拓思维,方能抓住主要矛盾,认真研讨。这样才能有助于回答与解决不明病因的问题。

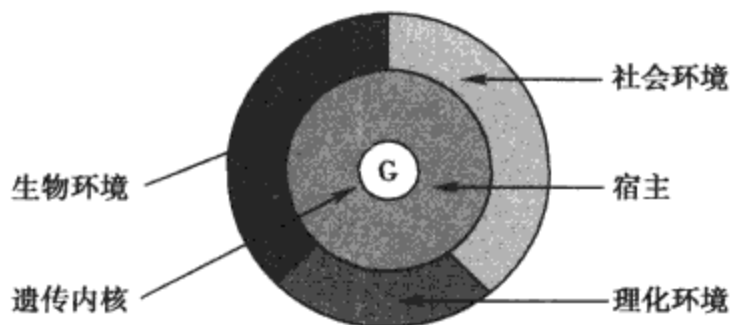


图9-3 病因的轮状模型

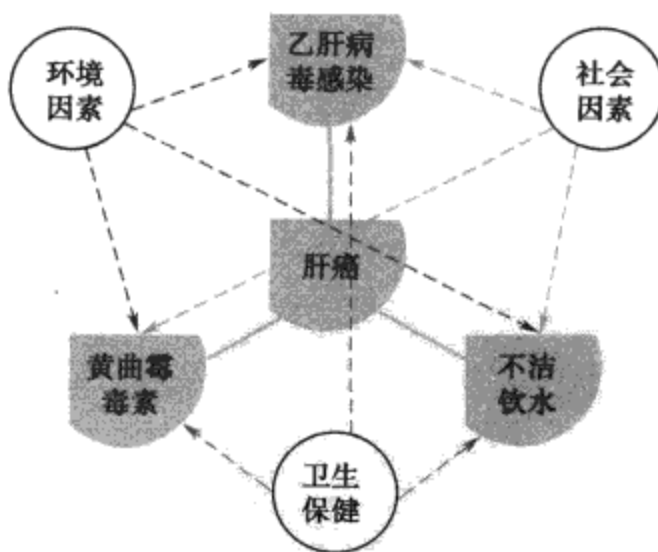


图9-4 病因网模型

## 第三节 病因与危险因素所致疾病的因果效应

不同的病因或危险因素引发相关疾病的因果效应,有着极大的差异。但是,从病因(原)或危险因素的特性以及人体发病机制的角度,将它们相互作用导致发病的因果效应,可归纳如下:

### 一、单病因(原)的因果效应

当某一病因(如病原体)侵袭人体后,引发一种特异性的疾病,这种因果效应称之为单病因的因果效应,可以传染性疾病为代表,如结核杆菌可引发人体的结核病。尽管它可定位损害肺、肠、肝、脑膜等不同脏器或部位,但都是特异性的结核病,而不是其他疾病,此即单病因的因果效应。

### 二、单病因致多病的因果效应

同一病因侵犯人体后,通过它的危害和人体的复杂发病机制与免疫反应,可以引发不同类型的多种疾病,就像俗语所说的“一石击出千层浪”。以乙型溶血性链球菌感染为例。这种病原微生物侵入人体后,可以引起急性风湿热,还可以引发急性肾小球性肾炎。这两种不同类型的疾病,经临床和流行病学研究证实,与乙型溶血性链球菌感染有着相当显著的关联。此乃一因多病的因果效应。

### 三、多病因致单病的因果效应

当多种发病的相关危险因素共存于人体时,它们对人体的危害却可引起某一种疾病。其中,危险因素在两种或以上者,其累积的危害度并不呈数学相加的结果,有时可呈类似几何级的规律。此乃多个病因致一种疾病的多因一病之因果效应。

例如:众所周知,高胆固醇血症、高血压、吸烟、糖尿病等为冠心病的发病危险因素,经人群流行病学调查证实,没有这些危险因素的人群冠心病的患病率仅为1.2%;仅有一种或两种危险因素者,其患病率则升至2%和6%;当存在三种或以上的危险因素者,其患病率竟高达31%!可见这些危险因素互相间交互为害是如何之剧,这对如何开展冠心病的有效防治提供了有力证据。

以上三种因果效应模式显示,对于疾病病因与发病的危险因素的分析 and 求证,既要注重疾病本身的临床特点和相应的病理损害定位,以及特殊(异)的实验室微观依据,还要关注宏观环境因素在发病中的效应,只有掌握了充分可靠的而不是虚假的信息,才可能对病因的因果效应作出循证的正确判断。

## 第四节 病因与危险因素的分析评价

对于病因与危险因素的分析与评价,最重要的是探求真实的因果效应。因此,作为循证医学的病因与危险因素的评价,最为重要的是病因证据的真实性(validity)。弄清了发病的病因,本身就很重( importance ),无论阳性或阴性结论都很重要!因为其有利于指导对疾病的科学诊断、治疗和预防的决策。

探求任何疾病病因,并不是凭空设想出来的,如第一节所述,其是从临床实践中遇到原因不明的疾病,从认识观一步一步地逐步临床总结、发展、发现、推导、试验、验证出来的,是从描述流行病学、分析流行病学、随机对照试验,不断反复的过程。这些特点与临床诊断、防治性的研究有所差异,因此,在病因分析与评价方面有其本身的特色。

按照循证医学分析与评价证据的共性要求,从真实性、重要性和适用性三方面予以阐述。

### 一、真实性

关于病因学研究证据的真实性评价的标准,在国际上首推由 Hill AB 制订的、并在临床流行病学实践中修订的病因学研究证据的评价标准(表9-1, Hill AB. Principles of medical statistics. 9<sup>th</sup> ed. London, Lancet, 1971. Haynes RB, Sackett DL, Guyatt GH & Tugwell P. Clinical Epidemiology. 3<sup>rd</sup> ed. LW Philadelphia, 2006:358-368)。

表9-1 关于病因学证据分析评价的条理和标准(要求)  
(organization and analysis of evidence to assess claims of causation)

1. 不同等级研究设计的证据(evidence from hierarchy of research design)
  - 真正的试验性研究(true experiments)
  - 队列研究(cohort study)
  - 病例对照研究(case control study)
  - 分析性调查(analytic surveys)
2. 因果效应的强度(strength of association)
3. 证据的一致性,特别是高质量研究的一致性(consistency-especially among studies of higher quality)
4. 从前瞻性研究所示的时间顺序(temporal sequence-from prospective study)
5. 梯度关系-通过暴露的剂量或时间来显示(gradient by dose or duration of exposure)
6. 根据流行病学、生物学以及相雷同的证据作出判断(sense-from epidemiology, biology and analogy)

我们将依照以上对病因求证的六大要求或分析与评价的内容加以探悉。

#### (一) 不同等级研究设计的证据

对于病因探讨所获得因果效应的证据,自然是源于不同的研究设计的研究结果。而证据的获得是要经历一个逐步深化研究和认识的过程。因此,对于病因证据探知不会是一开始就能选择最佳的临床试验研究设计方案-RCT进行研究的。但是,我们现在是期望回答循证医学实践中对不明病因求证,并对所获的证据(文献)进行真实性分析与评价,就应该从临床研究设计不同方案所产生证据的质量进行剖析。

1. 真正的人体试验性研究——RCT 临床真正的试验性研究即RCT(随机对照试验),是以健康人体或特定的患者为研究对象,按照试验制订纳入与排除标准,以及科学的假设要求,选择一定数量的合格研究对象,并随机暴露于病因/危险因素,然后观察一定时间、分别测量两组研究对象目标疾病的发生率,以探求该致病因素发病的因果效应,予以分析评价(详见临床流行病学有关RCT设计章节)。病因学真正临床随机对照试验的证据是最可靠的证据。

对于肯定有害的病因(病原体)对健康人体进行致病效应观察的随机对照试验是绝对违反伦理学原则的犯罪行为,因而是绝对不允许的!因此,也不可能发现它的临床随机对照试验的证据。那么,作为病因学研究是否不可以作RCT?是否在循证医学实际中查不到有关病因的真正临床试验的证据呢?否!在某些特定情况下或特定的疾病患者中,为研究疾病的因果效应,是可以进行RCT研究,也可能检索有关RCT证据的。这里大致有以下情况:

(1) 人体本身存在的某种(些)致病的危险因素(如高脂血症、高血压等):针对该种(些)危险因素,进行RCT设计,予以干预,如能预防目标病的发生,则能证明该病因/危险因素的因果效应。例如:高脂血症患者本身存在高胆固醇,其是心血管病的发病危险因素,为验证与预防心脑血管病的发生,可选择合适的无心脑血管病证据的高脂血症患者,随机分为试验组(如用他汀类药物治疗)或对照组(安慰剂组),进行RCT,最终观测降脂效应与心脑血管病的发病情况并予以评价。

(2) 不良生活嗜好或行为:当疑及为致病的危险因素(如吸烟、饮酒等):针对这类危险因素可能有致癌或致心脑血管病危险性,如能开展干预性试验(RCT),则可寻求其致病的因果效应。

(3) 治疗试验中的药物不良反应(ADR):在治疗性随机对照试验中出现的药物不良反应是在所

难免的,其性质和程度在该 RCT 中是可确定的。病因学的因果关系通过仔细分析也是可以明确的,对此的求证,将在相关章节详述。

(4) 观察性研究(observational study)获得的病因证据的验证:在医学实践中有的观察性研究获得的病因学证据,常常因受诸多的测量性偏倚的影响,真实度往往使人质疑。在不影响病人安全和符合伦理学原则的前提下,可以视具体情况开展人体试验研究以进一步明确病因是必要的。例如:系列的观察性研究证明对于绝经期妇女应用雌激素补充治疗,具有预防心脑血管病发病的效果。然而,尚有争议。因此,随后又开展了对绝经期无心脑血管病的妇女,应用雌激素预防心脑血管病的 RCT 试验,最终证明这种补充雌激素的疗法并无预防心脑血管病的作用,从而否定了这一因果效应。

上述情况说明病因/危险因素的证据,只有通过人体试验的 RCT 最为可靠,在某些特定条件下拟求证病因也是可以或可能找到此类真实证据的。

如果在循证医学实践中找不到人体的 RCT 病因证据,那么应检索与评价观察性研究的证据,用于实践中的决策参考,也是有着一定意义和价值的。

2. 队列研究(cohort study)的病因证据 探求前瞻性队列研究的病因学证据十分重要,只要是设计有足够的样本量,纳入和排除标准明确,证明入组观测的对象确实没有所观察的目标疾病、暴露病因的队列组均可被确证、非暴露病因的队列确实能证明为阴性、两个队列又能尽可能地排除相关的混杂因素(confounders)干扰、基线可比、观测终点指标一致、盲法测量与分析、追踪期足够长且失踪率<10%。如这些内容和条件均能满足,其病因及其因果效应的证据,则是可信的(cohort study 设计详见临床流行病学)。

例如:拟探求乙型肝炎病毒(HBV)是否可以诱发肝癌。20世纪80年代台湾进行了一个很好的 cohort study。其设计科学、入组样本量大。HBV(+)队列3414例,HBV(-)队列19525例,病因/危险因素 HBV 明确,终点指标为肝癌,基线状况中两队列均证明无肝癌,观察时间5年,追踪完全,研究结论:HBV 致肝癌的相对危险度为223(40/3414:1/19525)。

又如20世纪90年代,北美年轻妇女为美胸,整形医院开展了硅酮注入乳房成形术(silicone breast implant),颇受欢迎,于是接受整形手术的妇女竟达1400万人次之多,术后有数百人发现局部症状或周围纤维病变,也有发生关节炎、红斑狼疮、硬皮病患者,对此,引起了患者向法院诉讼。于是组建了国家科学组调解此件诉讼案,要求弄清是否硅酮注入乳房能发生有关结缔组织疾病(connective tissue diseases, CTD)。带着这个问题,这些专家检索了全球各种语种的有关研究报告,纳入评价分析的没有一个 RCT,但有10个队列研究(不同国家或地区的资料)。观察对象各队列达数万之众,结果表明硅酮注入乳房与所致的 CTD 没有病因致病的因果关系!这个阴性结论在循证病因学观点方面有着十分重要的意义。

3. 病例对照研究的病因证据 病因学回顾性探索最佳证据的来源,当属病例对照研究资料,因为它有拟探知的可疑病因或发病的危险因素;有采用可靠的诊断方法确诊了目标疾病的“病例组”,以及采用同样诊断标准而证明无目标疾病的“对照组”;同时它采用设计周密的方法避免研究中可影响研究质量的有关偏倚,全面地收集较为可靠的资料和相应的统计分析方法,使之获得对拟探知病因/危险因素致病的因果效应,尽可能可靠。回顾性的病因或危险因素研究资料所受偏倚的影响,往往是较多的,因而,应用这些证据时,应持慎重的批判态度,但是这种证据却可为前瞻性的病因研究提供重要信息。例如:乙型病毒致病肝癌的回顾性研究发现其较强之因果相关性的证据,就是源于它的“病例对照研究”。

病例-对照研究设计方案是回顾性研究病因/危险因素的重要方法,在分析评价这种因果关系时,一定要仔细分析源于该方案研究的证据之质量。兹引病例对照研究的科学原则作为分析评价质量的参考(表9-2)。



表9-2 病例对照研究的科学原则 (scientific principles for case control study)

(1) 在研究开始前要确定研究的假设(establish research hypothesis before the research conducted)
(2) 确定暴露是在先前的(define the exposure before hand)
(3) 对每个病人的基线状况要确定纳入和排除标准(define inclusion and exclusion criteria for each person's baseline state)
(4) 从对照组内要排除任何一个具有目标疾病早期体征或暴露病因相禁者(exclude from the control group anyone with what could be early signs of outcome state or contraindications to exposure)
(5) 检查与校正两个比较组之间有关预后因素的差异(check and adjust for prognostic factors that differ between the comparison groups)
(6) 检查选择研究对象过程中的限制选择性偏倚的措施(check a selection process that eliminates or minimize referral bias)
(7) 确立减少回忆性偏倚的方法(establish ways to eliminate or minimize recall bias)
(8) 确立减少发现(测量)偏倚的方法(establish ways to eliminate or minimize detection bias)
(9) 记录未知的暴露(因素)如未知的或未必不存在的因素(record unknown exposure as unknown, not absent)
(10) 处理参与者偏倚(manage participation bias)

(资料来源: Feinstein AR. The architecture of clinical research. Philadelphia:PA WB Saunders, 1983)

4. 现场调查的病因依据(analytic surveys) 这类分析性调查依据多来源于横断面调查资料(cross sectional study)。即当发现未知或已知病因在群体中引起某种疾病时,应对有关群体进行普查或抽样调查,以分析并初步评价有关病因及对群体造成的危害程度,但这类病因或危险因素的信息仅有参考价值。

以上有关病因与危险因素及致病因果效应的证据,其真实性是受研究设计的方案制约的。即使同种研究设计的结果,也可因设计与执行质量的不同而异。因此,为了分析与评价病因证据的真实性,除了掌握不同研究方案的科学论证强度外,还要深入分析证据来源的文献之内涵,这样才可能避免误导。现将有关病因学研究不同设计方案的论证强度与分级列表9-3。

表9-3 各种病因学研究的论证强度

研究设计类型	时间性	可行性	论证强度
随机对照试验	前瞻性	差	++++
队列研究	前瞻性	较好	+++
病例-对照研究	回顾性	好	++
横断面调查	断面	好	+
个案调查,病例报告	现在,回顾	好	+/-

(二) 因果效应的强度 (strength of association)

无论随机对照试验、队列研究或回顾性的病例对照研究,判断病因/危险因素致病因果效应自然要利用反映其强度的量化指标。如反映RCT及队列研究之因果效应的量化指标有各组的发病率(incidence)、相对危险度(relative risk, RR)、绝对危险增高度(absolute risk increase, ARI)和相对危险增高度(relative risk increase, RRI)等。而病例对照研究则用比值比(odds ratio, OR)表示(表9-4A、表9-4B)。

表9-4A RCT之病因效应分析表

组别	结果		合计
	+	-	
暴露组	a	b	a+b
非暴露组	c	d	c+d

注: 暴露组发病率(ED) = a/(a+b); 非暴露组发病率(ED) = c/(c+d)

RR = ED/ED; ARI = ED - ED; RRI = (ED - ED) / ED

表9-4B 病例对照研究之病因效应分析表

	组别	
	病例组	对照组
暴露(+)	a	b
非暴露(-)	c	d

注:  $OR=ad/bc$

按常理而言,这类指标当排除偏倚影响或经统计学校正后,自然其值越高,则病因的因果关系会更强。但在探讨有关发病的危险因素证据时,一定要联系具体的病种及有关危险因素可能造成的危害程度,做到对具体问题进行分析评价,否则就要犯教条性的错误。

例如:有文献报道吸烟与肺癌及缺血性心脏病的因果关系:①肺癌所致的死亡率,吸烟者为209/10万 vs 不吸烟者为14/10万,  $RR=15$ ,  $AR=195/10$ 万;②缺血性心脏病所致的死亡率,吸烟者为829/10万 vs 不吸烟者为572/10万,  $RR=1.6$ ,  $AR=320/10$ 万。从这些指标看,吸烟对肺癌的因果关系强度显然高于缺血性心脏病;然而,因吸烟造成的缺血性心脏病死亡率却远远高于肺癌!因此,尽管吸烟与缺血性心脏病  $RR$  偏低( $RR=1.6$ ),却仍是有意义的。

### (三) 证据的一致性

在探讨病因/危险因素因果效应的证据时,无论是前瞻性或回顾性的研究证据,应尽量收集完全,以考察证据的一致性。在不同地区或环境里探讨相同病因的因果效应,其研究的设计方案也许相同或相异,但如果结论相一致,同时又可排除混杂或其他偏倚影响的话,则证据的可信度高。例如:HBV与肝癌发病的相关性,从回顾性的病例对照研究到前瞻性的队列研究的结论较为一致,对肯定其因果效应很有价值;又如硅酮注入乳房成形术是否诱发结缔组织病,研究者从全球范围内收集与评价11个队列研究及8个病例对照研究,均获得硅酮注入乳房与结缔组织病发病没有相关性的结论,对否定该因果效应就很有意义。

### (四) 因果效应的的时间顺序

病因学因果效应的证据一定要注意时间顺序分析。作为病因致病效应,一定是因在前,果在后。因此,只有前瞻性的研究资料,且研究对象必须应用特异性高的诊断手段、有效排除早期无症状与体征的观察对象入组后,方可判断病因/危险因素之因果效应的的时间顺序。

而有关横断面调查或回顾性的病例对照研究证据,均不能反映因果效应的的时间顺序。

### (五) 因果效应的梯度关系

对于由物理因素或化学物质及药物所引起的疾病或药物不良反应,除人体的特异性过敏反应之外,反应强度通常与所接受的剂量、暴露时间或疗程有关。例如:放射性疾病患者,其接触暴露剂量越大,暴露时间越长,则受损害越厉害;出现药物不良反应的患者,凡服药量越大,疗程越长,自然反应或受害会越严重。当然,有些患者接触病因/危险因素也不一定呈现量效关系,这与人体内的发病机制和个体反应的特性有关。

此外,对某种疾病与多种危险因素共存危害的情况也应注意。例如:高血压、高脂血症、吸烟、肥胖以及高血糖(糖尿病)均为心脑血管病发病的重要危险因素。是否存在这样的危险因素及数量,对心脑血管病的发病均有显著的差异,也具有数量与效应(发病)的明显交互作用,即存在的危险因素越多则发病率越高,当然如能采取防治措施将其数量与程度降至理想水平,其发病率与病死率方会相应下降。

### (六) 根据流行病学、生物学及相同的研究证据作出判断

有些病因/危险因素的因果效应证据的判断,也许只是基于间接依据,需要进一步据实进行综合判断。例如:现在有的分子生物学研究可能发现某一基因缺陷与某一疾病有关。如遗传性溶血性贫血病,系由基因缺陷导致红细胞膜、酶及血红蛋白病变而发生贫血,也有些基础生物学研究发现某种因素与某病相关,但必须通过临床验证。

对于一些特异性高的疾病,如 SARS 或当前正在国内外流行的甲型 H1N1 流行性感冒的病原生物体研究,确定其特异性的病毒,则有特异的病因学意义,结合其发病特点与流行病的规律与证据,有助于防治的正确决策。

此外,当探讨某一疾病的病因证据尚不确定时,也许可以从雷同的因果效应获得启示。例如,非洲发生的 Birketts 淋巴瘤,被证实其病因为 EB 病毒所致,具有显著相关性。与此相仿,当初,人们从临床发现原发性肝癌患者多患有慢性肝炎,且相当比例的患者 HBV 标识物为阳性,因此,从病毒病因的雷同观点,提出肝癌发病与 HBV 有关的假说。最后,经过一系列的回顾与前瞻性的研究所证实。

以上经修正和完善的六条分析与评价病因/危险因素证据的原则,可以看做是方法论,是分析和评价病因/危险因素所应考虑的范围与内容,但并不是肯定或否定的统一标准。这里最重要的是,真正病因学或危险因素的证据是依靠科学和可行的研究设计以及执行正确合理的研究方法获取的。

## 二、重要性

经严格分析与评价的病因学/危险因素证据,除了满足真实可信外,还应该是重要的。不论是阳性结论或阴性结论,也不论它们的因果效应相关性的强度之差异如何。

### (一) 阳性的真实性证据

阳性的真实性证据,则可用于指导防治决策。如 HBV 致肝癌,则开展了 HB 疫苗的预防接种和加强 HB 的早诊早治以改善预后,预防肝癌。对于甲型 H1N1 流感,用该病毒制作疫苗,开展易感人群的广泛接种预防发病和流行等。

### (二) 阴性的真实性证据

当研究否定了可疑病因/危险因素致病的因果效应时,则有利于及时调整错误的防治决策、以“平反昭雪”。

例如:随机对照试验的证据,证实了更年期妇女应用雌激素补充治疗,对预防心脑血管病无效,因而可否定这一预防措施。又例如:经大规模的病例对照和队列研究的证据证明了硅酮注入乳房成形术与结缔组织病发病无关,从而在事实上否定了这一因果关系,有利于法律诉讼案的解决。

### (三) 因果效应相关的强度指标

因果效应相关的强度指标,如 RR、AR、RRI 等,一般而言,这些指标越高,当然因果效应的强度越大。这里还要注意其样本量大小和这些指标的 95% 可信区间。

必须再度强调的是切勿简单的依据上述指标的大小来得出肯定或否定的结论,一定要结合具体的疾病和病因/危险因素的相互效应与临床价值方能正确的理解。

## 三、适用性

对于获取的经分析与评价证明既真实又重要的证据,当然要有的放矢地指导自己循证医学实践的科学决策,并力图应用、验证和取得良好的防治疾病的效果。

### (一) 预防疾病的公共卫生决策

病因学研究获得特异性病原体病因证据,则可用于人群疾病预防。例如:甲型 H1N1 疫苗接种预防流感。

### (二) 筛选危险人群以针对性防治

当明确发病的危险因素后,则可用于筛选危险发病人群,开展有针对性的有效防治。如对高血压、高脂血症的人群予以有效治疗则可降低心脑血管病的发病与死亡率。

### (三) 制订科学的诊治决策、服务于临床实践

当明确病因与发病的危险因素后,可用于提高诊断水平和制订科学的疾病防治指南,提高临床

医疗水平。

**(四) 纠正误导措施和行为**

如获得肯定的人体试验真实的无因果关联的病因阴性证据,则有利于纠正观察研究的假阳性因果关联的结论,避免误导。

**(五) 促进临床治疗中药物不良反应的正确评价**

在临床治疗试验研究,以及日常的医疗实践中,对于患者使用药物治疗难免不发生药物不良反应(事件),一旦发生这样或那样的反应或许为治疗药物所致,或为其他因素的反应,因此,要进行药物不良反应的因果分析与评价,上述病因/危险因素的分析 and 评价的原则与方法有着重要的适用和指导价值。

文献评价实例(参见光盘辅助教材相关章节)。

(赵亚双 王家良)

## 第十章 诊断性试验循证分析与评价

临床工作中对疾病正确诊断甚为重要,任何一位临床医师在分析病例时得出的结论都可能包括错误信息,为提高临床医师的诊断水平,不仅需要研究高水平的诊断方法,而且需要对诊断性试验的临床价值进行科学的分析和评价。诊断性试验是对疾病进行诊断的试验方法,其不仅包括实验室检查,还包括病理诊断和影像诊断,如X线、CT、MRI、超声波以及放射性核素、纤维内镜、电镜等诊断方法。可以说所有获取临床信息的方法都可以被称为诊断试验,临床医生应正确认识诊断性试验的实用性与诊断价值,避免凭经验选择的盲目性或者过分相信文献资料中作者推荐检测仪器的片面性。

### 第一节 诊断性试验在临床医学中的重要性

临床医学对任何疾病进行临床观察或研究时,首先要尽可能对研究对象做出正确的诊断,一般需要收集病史、体格检查、实验室检查和(或)特殊检验,然后将所有获得的信息和数据,加以综合分析确定诊断。临床论证的过程可能极其复杂,在各项诊断性试验中,有些具有很高的诊断价值,譬如,血培养分离出特殊的病原菌、外科的手术发现、病理组织切片的诊断,这些都是具有极高肯定价值的诊断性试验,都是国际公认的诊断方法,在临床流行病学中称为“金标准”。在临床医学中,金标准是最重要的诊断性试验,循证医学最关注的是具有真实性、实用性的证据,以确定诊断。

但是,当临床医学发展到21世纪,疾病谱已经发生了变化,不少新的疾病还没有诊断的金标准。可是在临床医学中,对患者的诊断都要求提供最佳证据,因而必须对诊断性试验进行创新与深入的研究,提出新的金标准,这说明诊断性试验在临床医学发展中的重要性。临床医师在工作中,对疾病的诊断步骤和策略,要求合理地安排不同检查项目的先后次序,一定要避免遗漏关键性检查项目,尽可能简化诊断步骤,缩短确诊时间,尽量减少有创检查可能带来的并发症,降低医疗费用,从而达到提高诊断水平,降低漏诊率和(或)误诊率。因此,很好地掌握与应用循证医学的诊断策略和方法,在临床工作中定可以成为称职的临床医师。

#### 一、诊断性试验在循证医学实践中的临床意义

循证医学本身是遵循证据的临床医学,要求将最好的诊断、治疗方法提供临床应用,其是在临床流行病学基础上发展起来的,凭借临床流行病学的方法和现代信息工具的合理应用,更快、更好地将确证为最好的诊断性试验提供临床应用,使患者能及早得到正确的诊断,及时接受合理的治疗。

随着时间的推移,任何一种疾病诊断方法都会有新的发展,而且检测方法各有不同,但新的诊断性试验是否一定能超过原来的诊断方法,这要依靠对诊断性试验的评估。例如,对肺栓塞的诊断,除病史资料收集和D-二聚体(D-dimer)检测外,还有许多新开展的影像检查:除CT、MRI外,还有螺旋CT、CT肺动脉造影(CTPA)和MRI肺动脉造影(MRPA),这些方法各有千秋,虽然有的检查方法价格很贵,但都是无创性检查较为安全,由于特异度较低有时肺栓塞的诊断,还需要几种检查的结果相互验证。目前确诊肺栓塞的“金标准”仍然是以有创的肺动脉造影(PAA)为依据,这说明近年来发展的螺旋CT、CT肺动脉造影、MRI肺动脉造影,并不优于过去的肺动脉造影。

通过文献的检索,我们可以看到D-二聚体在诊断肺栓塞的敏感度可高达92%~100%,但其假阳性率很高(特异度很低),在临床应用中如该试验结果为阴性,则对排除肺栓塞有较大的价值。依据文献报道,螺旋CT诊断肺栓塞的敏感度差别较大(53%~100%),但特异度较高81%~100%,因而有较好的诊断价值。肺动脉造影的敏感度为98%,特异度95%~98%,确是很理想的诊断性试验,但发生严重并发症的可能性有0.1%~1.5%,这是一项难以逾越的缺点,以上情况说明某种疾病的诊断方法,虽有不断进展但仍需要逐步完善。

## 二、诊断性试验研究现状与发展方向

在临床医学实践过程中,诊断性试验起了很大的作用,它直接关系到每个患者的正确诊断,如果没有正确的诊断,即使使用再好的药物治疗也未必奏效。因而,在临床上需要更可靠、更正确、更实用的诊断性试验;在原有的基础上还要不断地对新的疾病研究和开发最好的诊断性试验。

循证医学要求对疾病的诊断一定要有最佳的诊断性试验为依据。但是有些疾病诊断用的“金标准”缺乏特异性,检测的阳性率不高而导致漏诊,或者出现假阳性容易误诊,这只能说明这些诊断性试验还不够理想。在这种情况下,有专家建议采用“组合性诊断标准”(constructing diagnostic criteria),例如,系统性红斑狼疮(SLE)诊断标准是1982年美国风湿学会建议的诊断标准,其中共有11项,患者若有4项符合,就可诊断为SLE,目前国际公认该项“组合性诊断标准”就是“金标准”,这种方法可提高诊断性试验的特异度减少误诊,使患者得到早期诊断和及时治疗。

我国在诊断性试验的研究方面,也在不断地发展,临床医学杂志中时有报道。在1985~1995年中华系列杂志中,共发表诊断性试验的论著50篇,对纳入病例的诊断标准,采用了“金标准”的共32篇(占64%),仍然有36%的论著没有金标准,这对诊断性试验的正确性影响极大,作者花了不少的人力、物力和财力,但却无法推广应用,缺乏实用价值。在这些论文中,应用诊断性试验评价指标的仅有10篇(20%)报道了敏感度、特异度,而没有1篇论著提到过阳性似然比,也没有盲法的报道,更没有试验重复性的报道。

2002年有作者分析了中华检验医学杂志1996~2000年,在该刊发表的诊断性试验论著111篇,其中有金标准的论著共65篇(58.5%),没有金标准的论文46篇(41.5%);即使在应用金标准的论著中,仍然有33篇(33/65,占50.8%)无法列出四格表,也无法计算敏感度和特异度。论著的样本含量,病例组有22篇(19.8%)不足30例,对照组有30篇(27.0%)不足30例。样本含量不足或严重不足的诊断性试验,论据单薄缺乏说服力。可见已在杂志上刊出的诊断性试验,还是有不少文献不可能给临床医生提供有益的证据。

我们在2001~2007年中华医学会6种核心杂志中共检出诊断性试验的论著51篇。有金标准的论著50篇(占98.0%),其中采用病理诊断作金标准的有25篇(50%)。报道了敏感度的48篇(94.1%)、特异度的46篇(90.2%),计算了准确度(或符合率)的18篇(35.3%)。采用盲法与金标准进行对照的3篇(5.9%)。报道了阳性似然比的论著有2篇(3.9%),另有2篇分别报道了ROC曲线及ROC曲线下面积(AUC)。可是,仍有10篇论著(19.6%)试验组的病例数在30例以下。

诊断性试验文献质量与10年前比较,各项指标均有明显提高。采用金标准的研究,从64.0%上升至98.0%,应用敏感度和特异度的研究从20.0%升至94.1%,采用盲法对比的研究由0增加到5.9%,然而,对阳性似然比的应用仍然不够普及。

近10年来快速诊断性试验(rapid diagnostic tests, RDTs)在热带病研究中发展迅速,例如,快速AIDS诊断试验、快速黑热病诊断试验、快速疟疾、梅毒、结核、登革热等疾病的诊断试验,都得到WHO西太地区基金的支持。各种快速的方法较多,但对其可信性(reliability)和准确性(accuracy)的评价较少。在1990年只有3种疟疾RDTs,经10年的发展,现有60个品牌200多种疟疾的RDTs,经美国CDC与WHO在2009年采用循证检验(evidence-based measures)的方法进行评估,结果只有20%(41个品种)合格。对快速结核诊断性试验19个品种进行评估,结果没有一种试验能对结核病进行正

确诊断。

因此,今后对诊断性试验的开发,要按照临床流行病学的方法学创最佳证据的要求,重视研究的质量,需事前做好设计方案、确定金标准和观测指标,而且对病例选择、样本含量、盲法试验、重复性试验等内容。要有所发现、有所创新,以期能够应用于临床而造福病人。

## 第二节 对诊断性试验研究评估的基本要求

为了对诊断性试验进行正确评价,首先要了解诊断性试验研究的基本方法是否恰当,其内容至少应包括:有确定的金标准,有新的诊断性试验方法和正确选择的研究对象,并将试验结果与金标准进行同步比较。以下所述“有病”或“无病”的病例,都是针对所研究的目标疾病而言,所谓“无病”者,是指被金标准确诊的非目标疾病的患者,而不是没有病的健康人。现将评估的基本要求简述如下:

### 一、确定金标准

诊断性试验的金标准(gold standard)是指当前临床医师公认的诊断疾病最可靠的方法,也称为标准诊断。应用金标准可以将患者,正确区分“有病”或“无病”。因此,在评价新的诊断性试验,必须要求有金标准为依据,金标准它可以包括病理活检、手术发现、微生物培养、尸检、特殊检查和影像诊断,以及长期随访的结果。

### 二、选择研究对象

诊断性试验的研究对象,应当包括两组;一组是用金标准确诊“有病”的病例组;另一组是用金标准证实为“无病”的病例组成,称对照组。

病例组应包括各型病例:如典型和不典型病例;早、中与晚期病例;轻、中与重型病例;有和无并发症者等,以便诊断性试验的结果更具有临床实用价值。

对照组可选用金标准证实没有目标疾病的其他病例,特别是与该病容易混淆的病例,以期明确其鉴别诊断价值,正常人一般不宜纳入对照组,否则可造成人为地夸大特异度。

至于样本含量,在确定定量资料参考值范围的诊断性试验,对照组的样本应满足100例(正态分布)或120例(非正态分布);在病例组也应有100例,特殊情况(少见病例)每组至少样本含量应在30例以上,尽可能地减少机遇因素的影响。此外,也可参考临床流行病学,采用样本含量的公式进行计算。

### 三、盲法对比试验结果

新诊断性试验一定要有明确的试验方法、清晰的试验程序和正确的科学依据。诊断性试验的结果,最好采用盲法与金标准进行同期对比,即要求判断试验结果的人不能预先知道,受检病例用金标准诊断为“有病”还是“无病”,以免发生疑诊偏倚,造成人为因素的影响。

判断新的诊断性试验的临床价值,应当将其结果与金标准诊断的结果进行同步对比;评估时要求每个病例必须都有两项(即金标准与新试验)检查的结果,并按要求列出四格表以便对论著中诊断性试验的实用性和临床价值进行评估,其方法如下:

1. 用金标准诊断为“有病”的病例总数为  $a+c$ 。

2. 在“有病”的病例总数( $a+c$ )中,用新的诊断性试验检测,结果阳性的病例数为  $a$ (真阳性),结果阴性者为  $c$ (假阴性)。

3. 金标准诊断“无病”的总例数( $b+d$ )中,用新的诊断性试验检测,结果阳性的病例数为  $b$ (假阳性),结果阴性者为  $d$ (真阴性)。

4. 列出四格表(表10-1),将  $a$ 、 $b$ 、 $c$ 、 $d$  的例数,准确地分别填入四格表中。如果从作者的论著资料中,找不到列出四格表所需要的数据,无法计算敏感度、特异度,则该试验就无法进行评

估。因此,要做好诊断性试验的评估,最少应当满足上述各项基本条件,否则可认为该论著的设计不合格。

表10-1 诊断试验四格表的排列

		金标准(标准诊断)		
		有病	无病	
诊断性试验	阳性(+)	a(真阳性)	b(假阳性)	a+b
	阴性(-)	c(假阴性)	d(真阴性)	c+d
合计		a+c	b+d	a+b+c+d

### 第三节 循证医学中诊断性试验常用的指标

对诊断性试验进行严格的科学评价时,通常应用的评价指标,在临床流行病学中已经详细介绍,以下将循证医学对诊断性试验评价时常用的指标进行简单的回顾。

#### 一、敏感度

诊断性试验检测为阳性的病例,在用金标准确定为“有病”的病例中所占的比例,称敏感度(sensitivity, SEN)。敏感度愈高,则假阴性的病例(漏诊率)愈少,有助于筛查相应的疾病。

$$SEN = \frac{a}{a+c}$$

#### 二、特异度

诊断性试验检测为阴性的病例,在用金标准确定为“无病”的病例中所占的比例,称特异度(specificity, SPE)。特异度愈高,则假阳性的病例(误诊率)愈低,特别有助于确定诊断。

$$SPE = \frac{d}{b+d}$$

#### 三、准确度

准确度(accuracy, ACC),诊断性试验检测为真阳性和真阴性病例总数,在全部被检病例数中所占的比例。

$$ACC = \frac{a+d}{a+b+c+d}$$

#### 四、阳性预测值

阳性预测值(positive predictive value, +PV),诊断性试验检测为阳性的全部病例中,用金标准诊断为“有病”病例所占的比例。

$$+PV = \frac{a}{a+b}$$

以上四项指标是一般诊断性试验报道中常用的指标,敏感度和特异度是两个稳定而可靠的指标;但是,准确度和阳性预测值并不具备很高的可信性,因为这两项指标与该试验患病率的高低有着相应的改变。我们用 Bayes 公式,就可以看出阳性预测值波动的原因:

$$\text{阳性预测值} = \frac{\text{患病率} \times \text{敏感度}}{\text{患病率} \times \text{敏感度} + (1 - \text{特异度}) \times (1 - \text{患病率})}$$



例10-1 某诊断性试验的敏感度为90%，特异度为60%，如被检病例为1000例，当被检病例中的患病率增高时，阳性预测值和准确度均有较明显的改变，其结果如下：

	金标准		
	有病	无病	
试验阳性 +	180	320	500
试验阴性 -	20	480	500
合计	200	800	1000

患病率 = 20% +PV=36% ACC=66% DOR=13.5

	金标准		
	有病	无病	
试验阳性 +	540	160	700
试验阴性 -	60	240	300
合计	600	400	1000

患病率 = 60% +PV=77% ACC=78% DOR=13.5

以上例证可以看出，在该试验中当患病率增高时，阳性预测值从36%上升到77%，准确度从66%上升到78%，说明这两项指标在同一项诊断性试验中并不是稳定指标，因此，该指标也不宜用于评价诊断性试验的真实性和实用性。

### 五、诊断比值比

因为诊断性试验的准确度很不稳定，所以对评价诊断性试验的价值不大，从而有作者提出诊断比值比 (diagnostic odd ratio, DOR) 的计算，该指标较准确度稳定。通过四格表中交叉乘积的比值可以计算诊断性试验的比值比。

从上例中也可看出该指标在诊断性试验中是相当稳定的指标。

$$DOR = \frac{ad}{bc}$$

### 六、阳性似然比

诊断性试验中，真阳性在“有病”患者中的比例与假阳性在“无病”例数中比例的比值，称阳性似然比 (positive likelihood ratio, +LR)。表明诊断性试验阳性时患病与不患病机会的比值，比值愈大则患病的概率愈大。关于阳性似然比如何在临床应用，后面章节还要介绍。

$$+LR = \frac{a}{a+c} \div \frac{b}{b+d}$$

$$= \frac{SEN}{1-SPE}$$

以上两项指标在诊断性试验的评价中很重要，但不少作者往往容易忽略，从例10-1可以看出诊断比值比的优越性较准确度为佳，因此，我们在今后的研究和评价中，应该予以应用和推广。

## 第四节 诊断性试验在循证医学临床实践中的应用

在循证医学中应用诊断性试验的目的，是采用某个诊断性试验，有针对性地诊断某个目标疾病。

由于使用的目的不同,因此对敏感度和特异度的要求也不一样,临床上为了筛选病例、使漏诊率降到最低,经常选用敏感度最高的诊断性试验;但是为了确诊病例、使误诊率降到最低,却要选用特异度最高的诊断性试验。循证医学为寻求最佳的临床证据,在诊断性试验中最重要的部分就是 ROC 曲线和似然比的应用以及提高敏感度或特异度的方法,现分述如下。

### 一、ROC曲线的应用

ROC 曲线(receiver operator characteristic curve)又称受试者工作特征曲线,在诊断性试验中应用的目的有两个,其一用于正常值临界点的选择,其二用于优选性质类似的诊断性试验,因此,在临床实验室工作中甚为重要。

制图方法是以该试验的敏感度(真阳性率)为纵坐标(Y轴),而以1-特异度(假阳性率)为横坐标(X轴),依照连续分组测定的数据,分别计算 SEN 及 SPE,按照平面几何方法,将给出各点联成曲线,即为 ROC 曲线。在曲线的各点中,距 Y 轴顶点直线距离最近的一点,即为正常值的最佳临界值。用该点数值区分正常与异常,其敏感度与特异度之和最大,而误诊及漏诊例数之和最小。但是,绘制 ROC 曲线不能只靠一、二次试验结果,就想找到正确的临界点,一般要求最少有5组连续分组测定数据用以制图,才能达到预期的要求。

如果同一种疾病有两种或两种以上诊断性试验方法需要进行优选时,则可用 ROC 曲线进行比较。此时,可将两种诊断性试验的 ROC 曲线制图,如图 10-1 所示,凡曲线顶点与纵坐标顶点最接近者(如 CT scan),就是两者之间最好的诊断性试验。此外,还可以比较两条曲线下的面积(area under curve, AUC),其面积大者为佳。面积的计算和显著性检验,均可使用软件包(STATA)进行运算。

临床检验用于确定诊断性试验正常值的方法较多,在连续计量资料中确定正常值的临界点,最好使用 ROC 曲线进行选择,这是国际通用的方法,其操作步骤简述如下。

例10-2 某医院采用餐后2小时血糖测定,对诊断糖尿病患者及非糖尿病患者进行连续检查,结果如表10-2,试问诊断糖尿病临界值,应确定在哪一个数值最正确?

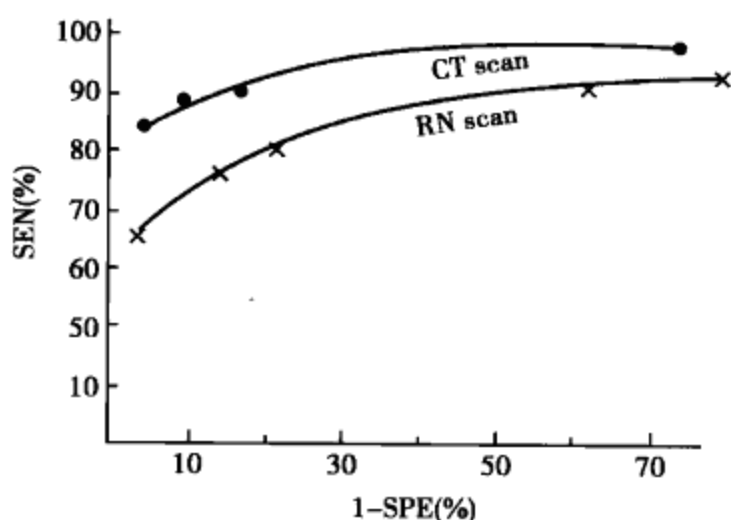


图10-1 放射性核素扫描(RN)与CT扫描诊断脑肿瘤的ROC曲线

表10-2 餐后2小时血糖测定及SEN与SPE

血糖(mg/dl)	血糖(mmol/L)	SEN(%)	SPE(%)	1-SPE
70	3.89	98.6	8.8	91.2
80	4.44	97.1	25.5	74.5
90	5.00	94.3	47.6	52.4
100	5.55	88.6	69.8	30.2
110	6.11	85.7	84.1	15.9
120	6.66	71.4	92.5	7.5
130	7.22	64.3	96.9	3.1
140	7.77	57.1	99.4	0.6

\*原表无该项数据(引自 Fletcher RH, 1996)

表中列出血糖测定值共有8组,并计算了各组的 SEN、SPE 及1-SPE,现将各组的数据,分别在纵坐标和横坐标上绘出并连成曲线,即为 ROC 曲线。如图10-2所示,在各点中距左上角直线距离最近的一点(曲线左起第4点),就是该试验的临界值。故诊断糖尿病餐后2小时血糖测定值,定为110mg/dl

(6.11mmol/L)最为恰当。

在表10-2中,可以看到随着血糖浓度的升高,敏感度下降而特异度升高。因此,敏感度和特异度中任何一个数值的提高,必然导致另一数值的降低。如选择餐后血糖100mg/dl(5.55mmol/L)为临界值,则糖尿病的漏诊率(1-SEN)为11%,而有30%的非糖尿病受试者被误诊。如果将临界值定为120mg/dl(9.66mmol/L),则可使漏诊率(1-SEN)上升到29%,而误诊率(1-SPE)则为7.5%。当餐后血糖临界值定为110mg/dl(6.11mmol/L)时,该阈值的敏感度和特异度之和为最大,这时漏诊率和误诊率之和最小。因此,餐后血糖定为110mg/dl作为临界值最适当,恰与ROC曲线确定的临界值相同。

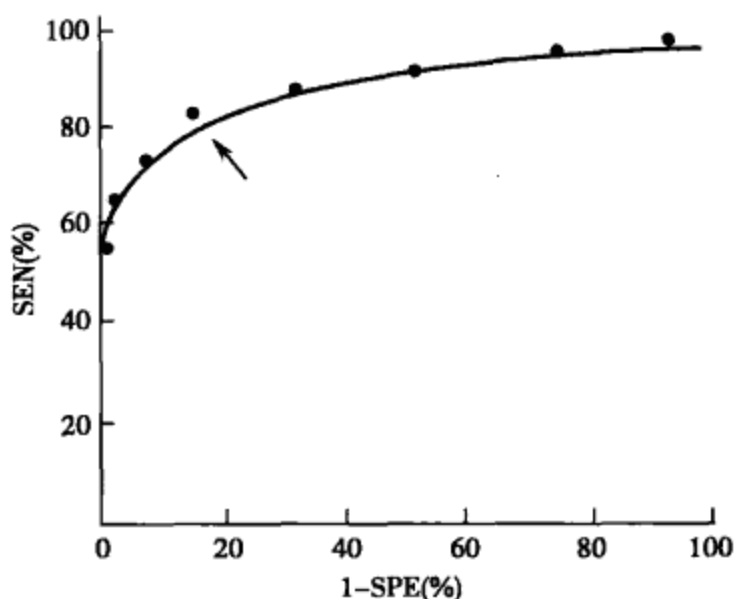


图10-2 餐后2小时血糖诊断糖尿病的 ROC 曲线

## 二、似然比的临床应用

似然比是诊断性试验综合评价的理想指标,其综合了敏感度和特异度的临床意义,而且可依据试验结果的阳性或阴性计算某病例患病的概率,以便在诊断性试验检测后,更确切地对患者进行诊断。从定义上来讲,阳性似然比(+LR)是诊断性试验的真阳性率 $[a/(a+c)]$ 与假阳性率 $[b/(b+d)]$ 之间的比值;真阳性率愈高,则阳性似然比愈大。

例10-3 研究者对贫血患者,进行血清铁蛋白(ferritin)检查,选定65 $\mu\text{g/L}$ (ng/ml)为临界值,当血清铁蛋白 $<65\mu\text{g/L}$ 则诊断为缺铁性贫血,而 $\geq 65\mu\text{g/L}$ ,则为非缺铁性贫血;共计检查2579例贫血患者,结果见表10-3。根据表10-3可计算出的一系列诊断指标,现将计算结果简述如下:

表10-3 缺血性贫血患者血清铁蛋白检查结果

		缺铁性贫血		
		是	否	
血清铁蛋白	+(< 65)	731(a)	270(b)	1001
	-( $\geq 65$ )	78(c)	1500(d)	1578
合计		809	1770	2579

(引自 Sackett DL, 2000)

$$\text{SEN} = 731/809 = 90.4\% \quad \text{SPE} = 1500/1770 = 84.7\%$$

$$+LR = \text{SEN} / (1 - \text{SPE}) = 5.9$$

如将铁蛋白测定的计量资料进行分层处理,则可计算不同检测水平的阳性似然比,如表10-4所示。

表10-4 缺铁性贫血患者血清铁蛋白分层检验结果

铁蛋白( $\mu\text{g/L}$ )	阳性似然比(+LR)
<15	51.8
15~24	8.8
25~34	2.5
35~44	1.8
45~99	0.5
$\geq 100$	0.08

(引自 Current Medical Diagnosis & Treatment, 2002)

阳性似然比的应用: 在确定某项检验或某项体征的阳性似然比后,可根据患者的病史、体征进行验前概率的估计,再根据检测结果,应用阳性似然比,便可计算该病例患病的验后概率,步骤如下:

验前概率( pretest probability )= 医师依据病史及体征做出的初步估计

$$\text{验前比值( pretest odds )} = \frac{\text{验前概率}}{(1 - \text{验前概率})}$$

验后比值( post-test odds )= 验前比值 × 似然比

$$\text{验后概率( post-test probability )} = \frac{\text{验后比值}}{(1 + \text{验后比值})}$$

例10-4 一位农村妇女36岁,面色不好,心慌乏力半年多,平时务农,每次月经量偏多,如何考虑患者的诊断?

- 根据病史: 务农、月经偏多、面色不好、乏力半年、可能有贫血及钩虫感染。
- 初步诊断: 依据病史,诊断缺铁性贫血( IDA )的可能性至少有60%( 验前概率)。
- 化验检查: Hb 65g/L, RBC  $3.40 \times 10^{12}/L$ , 血清铁蛋白  $20 \mu\text{g}/L$  ( +LR=8.8 )。
- 验前概率 =60%
- 验前比值 = $0.60 / (1 - 0.60) = 1.5$
- 验后比值 = $1.5 \times 8.8 = 13.2$
- 验后概率 = $13.2 / (1 + 13.2) = 0.93$

答: 该女性患者 IDA 的验后概率为93%。

验前概率是临床医生根据病史、体征等,估计该患者可能患病的概率;经诊断性试验(血红蛋白、铁蛋白)检查后,可利用 +LR 计算验后概率,然后对患者最后诊断。上述病例经检查后,诊断缺铁性贫血的概率,从最初估计的60% 提高到93%,确定诊断后就可及时给予相应的治疗;这说明阳性似然比在临床上确有很大的实用价值。

注意事项: 在诊断性试验的计算中,如果验前概率与该试验的患病率相同,则验后概率与阳性预测值相等。否则,不能认为验后概率就等于阳性预测值。例如,在表10-3各项数据中,可以计算患病率 = $809 / 2579 = 31.4\%$ ,阳性预测值 = $731 / 1001 = 73.0\%$ 。假定某患者的验前概率为50%,+LR 仍然是8.8,按上述方法计算验后概率为89.8%,该数值与诊断性试验的阳性预测值( 73.0% )迥然不同。

为了使读者了解阳性似然比与验前概率和验后概率之间的关系,现列表如下,我们从表10-5可以看出如果阳性似然比较高,即使验前概率在较低情况下,验后概率也会有很大的增长。例如:某患者验前概率估算只有30%,而 +LR=10,当试验结果为阳性时,查表10-5显示该患者的验后概率可高达81%。

表10-5 +LR和验前概率与验后概率的关系

阳性似然比	验前概率					
	5%	10%	20%	30%	50%	70%
10	24*	53	71	81	91	96
5	21	36	56	68	83	92
3	14	25	43	56	75	88
1	5	10	20	30	50	70
0.3	1.5	3.2	7	11	23	41
0.1	0.5	1	2.5	4	9	19

\*表中数据均为验后概率(%)

此外,我们将一些常用诊断性试验的阳性似然比(+LR)列于表10-6,供临床医师在诊断时参考。

表10-6 部分常见病、诊断性试验结果的阳性似然比

病名	金标准	诊断性试验	阳性似然比
冠心病	冠状动脉造影,狭窄≥75%	典型心绞痛发作	115
冠心病	冠状动脉狭窄(血管造影)	不典型心绞痛,有阳性病史	14
心肌梗死	心电图或尸检	肌酸激酶≥80u	7.75
深静脉血栓形成	静脉造影	深静脉血栓形成症状(疼痛、皮肤颜色改变、局部发热、压痛、周径增大3cm),全部体征伴周径增大	2.6
深静脉血栓形成	静脉造影	以上体征 <4项,且无周径改变	0.15
深静脉血栓形成	静脉彩色多普勒	血浆 D-dimer>1292ng/ml	2.0~3.1
冠心病	冠状动脉狭窄(血管造影)	心电图运动试验	
		ST 下抑≥2.5mm	39
		2~2.49mm	11
		1.5~1.9mm	4.2
		1.0~1.4mm	2.1
		0.1~0.9mm	0.92
冠心病	冠状动脉狭窄(血管造影)	放射性核素冠状动脉造影	3.6
β-溶血链球菌 咽炎	咽拭子培养	快速溶血链球菌抗原试验	15.2
腹腔积液	腹部超声波	移动性浊音	2.3
腹腔积液	腹部超声波	波动感	5.0
腹主动脉瘤	彩色超声多普勒	腹部扪诊包块 >3.0cm	2.7
胰腺癌	手术或尸检	B 超改变: 肯定阳性	5.6
		可疑阳性	2.1
		CT: 肯定阳性	26
		可疑阳性	4.8
结肠直肠癌	活检或手术	结肠镜检	5.0
结肠直肠癌	活检或手术	癌胚抗原(CEA) ≥20μg/L	3.5
		10~19μg/L	2.3
		5~9.9μg/L	1.4
贫血	Hb<110g/L 或 HCT<35%	面色苍白	3.8
肺结核	结核菌培养	痰菌: 阳性	31
		阴性	0.79

(引自 Sackett DL. Clinical Epidemiology, 1991)

为了提高患者的验后概率,如果一位患者有2项或2项以上的诊断性试验为阳性,就会有2个或2个以上的阳性似然比,然后将阳性似然比合并,就可以得到较高的验后概率,具体方法简述如下:

验前概率 = 临床医师依据病史、体征做出的估计

$$\text{验前比值} = \frac{\text{验前概率}}{(1 - \text{验前概率})}$$

验后比值 = 验前比值 × (+LR1) × (+LR2)

$$\text{验后概率} = \frac{\text{验后比值}}{(1 + \text{验后比值})}$$

例10-5 患者为40岁男性,因胸前闷胀来院就诊,经仔细询问2年前曾有不典型心绞痛,经 ECG 检查 ST 水平下移2mm,试问该患者患有 CHD 的可能性有多大?

解:40岁男性有胸前不适,估计患 CHD 的可能性有20%,根据表10-6不典型心绞痛的 +LR=14,心

心电图 ST 下移 2mm 的  $+LR=11$ , 现分别计算验前比值、验后比值及验后概率。

验前概率: 临床估计为 0.20

验前比值  $=0.20/(1-0.20)=0.25$

验后比值  $=0.25 \times 14 \times 11=38.5$

验后概率  $=38.5/(1+38.5)=97.5$

答: 该患者患有 CHD 的概率为 97%。

### 三、提高诊断性试验敏感度或特异度的方法

在临床上为了增加敏感度或特异度, 可采用多项试验的联合应用的方法, 现简述如下:

1. 平行试验(parallel test) 为提高诊断的敏感度, 同时对受检者进行几项目的相同的诊断性试验, 只要有一项试验阳性, 即可判断为患病者, 这种方法称为平行试验或并联试验。平行试验的应用可提高敏感度, 而特异度却有所下降, 在循证医学中应用平行试验的意义在于凡是诊断试验阳性的患者均可诊断目标疾病, 因而提高了敏感度, 减少漏诊病例; 可是假阳性病例有所增高, 此时应注意做好鉴别诊断。但凡诊断试验阴性者, 则可排除目标疾病, 在临床应用时甚为方便。平行试验结果的判断方法见表 10-7。

表 10-7 平行试验结果的判断方法

项目	结果		判断结果
	试验 A	试验 B	
平行 试 验	+	-	+
	-	+	+
	+	+	+
	-	-	-

平行试验计算敏感度及特异度的公式:

$$\text{平行试验 SEN} = \text{SEN A} + (1 - \text{SEN A}) \times \text{SEN B}$$

$$\text{平行试验 SPE} = \text{SPE A} \times \text{SPE B}$$

例 10-6 试验 A: SEN=65%, SPE=80%;

试验 B: SEN=75%, SPE=90%;

计算两项平行试验合并后的敏感度与特异度。

解: 按照平行试验公式进行以下运算:

$$\text{平行试验 SEN} = 0.65 + (1 - 0.65) \times 0.75 = 0.91$$

$$\text{平行试验 SPE} = 0.80 \times 0.90 = 0.72$$

答: 使用两种诊断性试验进行平行试验后, 敏感度为 91% 较前有显著提高, 这样可尽量减少目标疾病的漏诊, 适用于大样本的筛查工作; 但特异度下降明显, 应注意误诊问题。

2. 序列试验(serial test) 循证医学在临床实践中, 经常需要作出确切的诊断, 而在临床上又缺乏特异度很高的诊断性试验, 这时可在一项诊断试验阳性的前提下, 再采用另一项诊断试验, 如结果仍为阳性, 我们才作出疾病的诊断, 这样就可以提高诊断性试验的特异度。临床上可以先用安全而价格低廉的试验, 当其结果为阳性时, 再采用价格昂贵或有创性的试验进行诊断。这种采用序贯方法进行两项或两项以上的诊断性试验, 并且均取得阳性结果后, 才作出疾病诊断的多项试验方法, 称为序列试验或串联试验。但是, 用不同的诊断试验, 一项一项检查, 必然费时过多, 因而可采取一次做多项试验, 如前述 SLE 的组合诊断标准, 一次进行多项检查也就相当于序列试验。序列试验与平行试验不同之处, 在于其必须每项试验均为阳性时才能够加以合并。

序列试验计算敏感度及特异度的公式:

序列试验  $SEN(A+B) = SEN A \times SEN B$

序列试验  $SPE(A+B) = SPE A + [(1 - SPE A) \times SPE B]$

例如:急性心肌梗死患者,不同的酶学测定结果 SEN、SPE 各有高低(表10-8),单项应用易发生假阳性造成误诊,如三项同时应用,则可提高确诊的概率,从而确定疾病的诊断。

表10-8 单项酶学测定的SEN与SPE

酶学检查	SEN(%)	SPE(%)
CK	96	57
SGOT	91	74
LDH	87	91

如果要求三项结果均为阳性,才可确诊心肌梗死,则联合应用三项试验的 SPE 可高达99%,但是 SEN 则下降为76%。在序列试验中,在计算多项试验的特异度时,可先用第1、2项试验的特异度计算  $SPE(A+B)$ ,再以  $SPE(A+B)$  与第3项试验的特异度合并,计算3项试验同时为阳性的  $SPE(A+B+C)$ 。

3项试验均为阳性时,依据表10-8计算 SEN 及 SPE 的方法如下:

$$SEN(A+B+C) = 0.96 \times 0.91 \times 0.87 = 0.76$$

$$SPE(A+B) = 0.57 + [(1 - 0.57) \times 0.74] = 0.89$$

$$SPE(A+B+C) = 0.89 + [(1 - 0.89) \times 0.91] = 0.99$$

序列试验可提高 SPE 与 +PV,而降低 SEN 与 -PV;提高特异度目的在于确诊病例,使误诊率降到最低水平,从而有利于对患者及时给予正确的诊断和治疗。本例应用序列试验,当3项试验结果均为阳性时,则诊断急性心肌梗死的概率可高达99%,误诊概率仅为1%,这就达到了循证医学诊断的预期目标。

## 第五节 诊断性试验证据的评估

循证医学对诊断性试验的评价标准较为严格,是国际通用的标准,对试验的真实性、重要性及实用性,进行了认真评价,以期得到真正对临床有意义的诊断性试验,从而为循证医学提供有价值的依据,供临床医生参考与应用。1994年 Jaeschke 等应“循证医学工作组”的要求,对诊断性试验论著的评价,提出了3项9条评价原则,在2000年 Sackett 所著《循证医学》第2版中,仍然采用该评价原则。但是在被评价的文献中,对以上各项要求,未必全部具备;故在分条评定的基础上,最后对被评价的诊断性试验论著,进行一个较为全面的小结和正确的评价,以提供最佳证据,指导临床诊断与决策,见表10-9。

表10-9 诊断性试验的评价标准

评价内容
一、诊断性试验的真实性
1. 是否用盲法将诊断性试验与参考标准(金标准)作过独立的对比研究?
2. 该诊断性试验是否包括了适当的病谱(spectrum)?
3. 诊断性试验的检测结果,是否会影响到参考标准的应用?
4. 如将该试验应用于另一组病例,是否也具有同样的真实性?
二、诊断性试验的重要性
5. 是否通过该项诊断性试验,能正确诊断或鉴别该患者有无特定的目标疾病?
6. 是否进行了分层似然比的计算?
三、诊断性试验的实用性
7. 该试验是否能在本单位开展并能进行正确的检测?
8. 我们在临床上是否能够合理估算病人的验前概率?
9. 检测后得到的验后概率是否有助于我们对病人的处理?

现将各条评价标准的主要内容简述如下:

### 一、诊断性试验的真实性

循证医学对诊断性试验的要求,首先在于它的真实性,能够对患者做出正确的诊断。在众多的诊断试验中,筛选具有真实性的试验,必须要有严格的规定,目前国际上通用的评价标准如下:

1. 是否用盲法将诊断性试验与参考标准(金标准)进行独立的对比研究? 诊断性试验的研究过程中,对每个患者需进行两项试验,然后将新的诊断性试验结果与金标准诊断结果比较,才能判断该试验是否可靠、是否具有真实性。进行这项新试验的技师(或医生),事先不应知晓金标准对患者检测的结果,应在盲法下进行检测,这样可避免人为的偏倚,使该试验更具有科学性,目前使用的自动化分析检测仪,基本符合盲法要求,如果操作得当其结果应也具有真实性。最后通过分析该论著,列出四格表计算各项指标,根据 SEN、SPE 及 +LR 来确定该项诊断性试验有无临床应用价值。

2. 该诊断性试验是否包括了适当的病谱? 诊断性试验的受试患者是否包括各型病例(轻、重、治疗、未治疗)以及个别易于混淆的病例? 例如,测定血中 T<sub>3</sub>、T<sub>4</sub> 诊断甲亢,测定血糖诊断糖尿病,测定肝功能、肾功能判断肝脏和肾脏受损情况,这些都是较好的诊断性试验,当各型病例都包括在内时,这些指标可诊断疾病,又可判断病情,还可以进行鉴别诊断。

3. 诊断性试验的检测结果,是否会影响参考标准的应用? 如果我们的标准诊断(或参考标准)是确切可靠的金标准,那就不会顾及新开展的诊断性试验结果如何,这对诊断都不会有所改变。如果本来使用的诊断标准就不可靠,一旦发现新的诊断性试验结果与原来的诊断有所不同,有时就会难以取舍,必须继续观察以明确患者的诊断,然后进一步判断原来的标准诊断是否存有缺陷,以及新的诊断性试验是否真正可靠,特别是原有标准诊断的基础薄弱,多以临床症状体征为主,缺乏可信的试验指标,在这种情况下,有了新的诊断性试验,对改进原有诊断标准,提高临床诊断水平是有益的。因此,在评价过程中,一方面要考虑原有金标准是否恰当,另一方面要考虑新的诊断性试验,是否真有新的发现。

4. 如将该试验应用于另一组病例,是否也具有真实性? 一项可靠的诊断性试验,在判断其真实性时,应考虑到该试验的重复性(reproducibility),如多次测定同一标本的结果接近,说明测定数值稳定、结果可靠。因此只要疾病相同,不论在何处采用该项试验其结果都应是一致的,即使用于另一组病例,对特定的目标疾病诊断应具有同样的真实性。在新开展的病例组检测中,应该注意该组的患病率是否与以往的病例组不同,因为患病率不同的病例组,就不能使用阳性预测值和准确度作为评价的指标。

### 二、诊断性试验的重要性

诊断性试验的重要意义在于其能够正确诊断和鉴别受试者是否有特定的目标疾病,通过这项试验的结果,能够让我们作出正确诊断,这在临床上是非常重要的。这些内容可从以下两方面予以阐明:

1. 是否通过该项诊断性试验,能正确诊断或鉴别该患者有无特定的目标疾病?

我们在临床上考虑初步诊断的可能性,就可认为是验前概率,通过诊断性试验得到了阳性似然比,就可以计算验后概率。在这个过程中,诊断性试验对我们的诊断性思维能起多大的作用,这就是该试验正确诊断疾病的能力; 验后概率越高,说明该试验正确诊断目标疾病的能力就越强。因此,应依据论著中有关诊断性试验的各项数据,进行敏感度、特异度及阳性似然比的核算; 如果没有提供可运算的数据,那么该文献就没有实用的价值。

例如: 表10-5所示,当某个试验的 +LR=10时,即使验前概率只有30%,但通过试验检测结果为阳性,其验后概率可高达81%,这就使我们的初步诊断有了更大的把握,也就可以说明该试验对临床诊断的重要性。

2. 是否进行了分层似然比的计算?

要计算分层似然比,必须要有分层数据的原始资料,这样计算分层的阳性似然比,以便更确切地计算验后概率,使该试验的临床重要性有所提高。但这些资料除非在开展临床流行病学或循证医学的单位,所进行的研究工作才会有这样的报道,一般文献报道的资料都很少涉及此项内容。例如表10-4就是



分层似然比的表格,假设某女性患者,因贫血就诊初步考虑缺铁性贫血的可能性(验前概率)有50%,另外的50%是其他原因的贫血,经检查血清铁蛋白结果为 $15\mu\text{g/L}$ ,查表10-4可知 $+LR=8.8$ ;按公式计算其验后概率可达90%,大大肯定了我们最初的诊断,这说明分层的阳性似然比更具有临床诊断的重要性。

### 三、诊断性试验的实用性

在评估诊断试验的真实性与重要性之后,应明确新的试验是否能够迅速地应用于临床,使更多患者得到及时的正确诊断,并且能够取得社会效益的同时也取得经济效益,并对该试验的应用前景做出评价,以便推广应用:

#### 1. 该试验是否能在本单位开展并能进行正确的检测?

在报道的资料中,是否明确地叙述了试验的原理、试剂、操作步骤与方法、检测对象与注意事项;以便结合本单位情况,考虑可否开展该项试验,有无经济效益等。例如:设有专科门诊的医院,开展冠状动脉造影检查冠心病患者,或血液专科门诊开展血红蛋白电泳,检查长期患有小细胞贫血的患者,则阳性率较高,价值较大。如果这些检查用于基层医院,对一般的冠心病及贫血病人进行检测,不但前者会有风险存在,而且阳性率也很低,开展这样的试验实用价值不大,经济效益也会明显受到影响。

#### 2. 我们在临床上是否能够合理估算病人的验前概率?

合理的验前概率估计多与以下三个因素有关:①医师本人的临床经验;②该病在当地的发生情况(群体的患病率);③诊断性试验的应用范围是否符合我们的病例。在临床医师掌握以上基本情况后,如果新的诊断试验确实可靠,能够及时对目标疾病进行检测,这时对患者的验前概率进行恰当估计,通过检测就可以达到早期正确诊断的目的,体现诊断性试验的临床实用价值。

#### 3. 检测后得到的验后概率是否有助于我们对患者的处理?

利用阳性似然比,计算验后概率后,要了解该值是否已跨越诊断-治疗阈值(见图10-3,阈值B),如已跨越说明患者的诊断已经基本明确,就应该开始治疗,以使患者能及早获得最佳疗效。如果,验后概率没有跨越诊断-治疗阈值,仍在阈值A~B之间,则应进一步检查并给予适当治疗,以免延误病情,因此验后概率的计算,有利于患者及早得到合理的治疗,这肯定是有实用的价值。

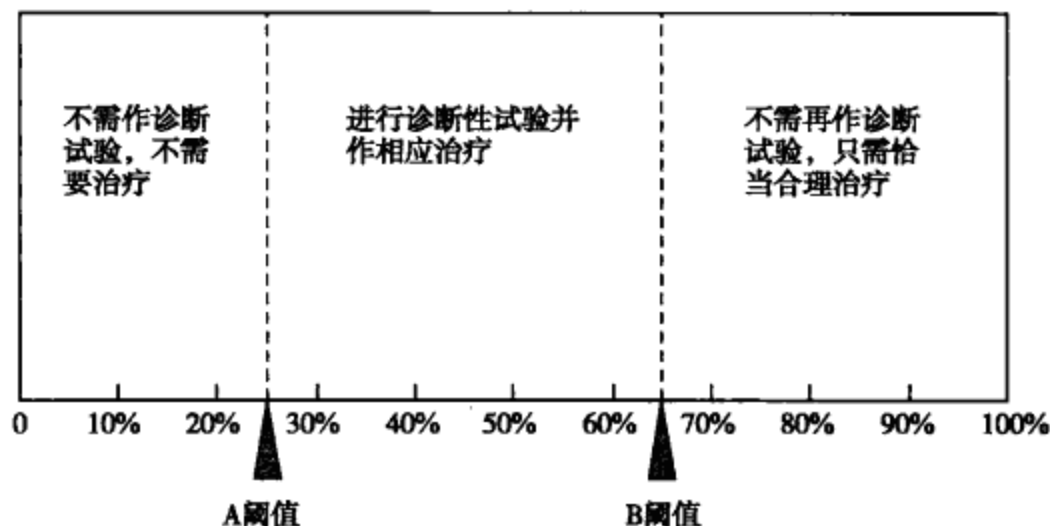


图10-3 阈值分析示意图

A 阈值, 诊断阈值 B 阈值, 诊断-治疗阈值

(引自 Sackett DL, 2000)

综合前述,循证医学对诊断性试验的3项9条评价标准,与临床流行病学应用的评价标准相比,的确要求更为严格,评估的难度也有所增加;要从真实性、重要性和实用性三方面进行评估,然后提供循证医学在诊断方面的最佳证据。我们在应用中要体会各条标准的内涵,逐渐熟悉、逐步提高,并在每篇文献评价后,进行一个恰当的小结。

文献评价实例(参见光盘辅助教材相关章节)

(王觉生)

## 第十一章 疾病治疗的循证分析评价

临床治疗性试验是一个十分丰富精彩的研究领域。在临床医学期刊发表的文献中,各种治疗性的文献一般占30%~40%,而这些研究成果(证据)的质量与水平,更是参差不齐。据BMJ所编辑的Clinical Evidence,所分析评价的全球顶级临床医学研究文献,真正有肯定价值者仅为15%!因此,在循证医学治疗实践中,如何选择与应用最新最佳证据指导自己对病人的治疗决策,是一个极大的挑战!

一旦我们的病人明确诊断后,如何治疗就成为需要解决的重大问题,应尽力获取当前最佳的有效治疗措施(证据),结合病人的实际和医疗的具体环境,进行科学的循证治疗决策,力争取得最佳效果。本章就疾病循证治疗的分析与评价进行如下阐述。

### 第一节 循证医学治疗实践应找出挑战性的问题

在日常的临床医疗实践中,一个合格的临床医生应对一般性的医疗诊治工作,不会有太大的难度。但若遇到一些具有挑战性的难题,则要费一番工夫了。这些所谓的“难题”,或许在知识更新、科技日益进步的情况下,已被他人解决,对他人并非难题,只是由于本人知识老化,而成为“难题”。因此,在循证医学治疗的实践中,要充分应用自己的医学知识和临床经验,善于从患者具体的病况中去发掘某一(些)具有挑战性的难题并积极寻找最新最好的治疗手段或方法,并付诸于实践,一旦解决了难题,则既能促进病人的康复,又能促进自己医疗水平的提高和知识的进步。

例如以下问题,在教科书或参考书中难以找到现成的答案:

- \* 静脉溶栓疗法能否降低急性心肌梗死患者的病死率?
- \* 冠状动脉内置入支架术治疗冠心病患者的远期疗效,是否优于完善的内科药物治疗?
- \* 颈内动脉内膜剥离术是否可降低缺血性脑梗死和短暂性脑缺血发作的发生率?
- \* 妊娠反应服用“反应停”后是否可引致胎儿畸形?等。

要回答和解决诸如此类的挑战性问题,需要从最新最可靠的研究成果(证据)中寻找答案,这就需要进行循证医学治疗实践的探索,以推动临床医学水平的不断进步。

### 第二节 查阅与运用最佳证据

循证医学治疗实践,首先应带着自己的、拟期望解决的临床问题去查阅有关文献。为达到“有的放矢”,这里从治疗学的角度应特别注意以下几点。

#### 一、关注治疗目的

治疗目的是通过最有效且最安全的治疗措施(或证据),获取相关疾病治疗的最大预期效果。基于患者所患疾病的不同特点,明确其相应的治疗要求或目的。如某种急性疾患,像甲型H1N1流感或急性大叶性肺炎等治疗目的是要达到治愈;而大多数慢性非传染性疾病如心脑血管疾病的治疗,往往难以达到治愈目的,因此治疗目的就变成了有效控制、缓解症状,减少若干并发症以改善生存质量、

延长寿命等;又如恶性肿瘤的早诊早治,在病灶局限而无扩散前予以手术根治,然后予以适当的化疗或放疗,其目的是防止复发等。

因此,通常应在明确某一疾病最佳治疗目的的前提下,针对患者的某一特殊问题,检索与掌握相应文献资料,方可达到有的放矢。

## 二、关注患者的具体状况

应用最佳治疗措施是为了使患者获得最佳治疗效果,但接受治疗的主体是患有该病的病人。因此,要关注病人的具体状况。例如其病情与最佳治疗措施是否能真正“对号”;即使能“对号”,那么患者的种族属性、年龄、体质、功能状况是否“合适”;即使“合适”,那么患者的经济条件、心理要求又是否能接受,特别是对带有风险的措施尤为突出,当然即使都符合,还要考虑患者的依从性问题。

作为应用某种最佳证据的具体实施者,临床医生务必要仔细耐心地关注患者的具体情况,尊重患者意愿与权利,才能达到最佳治疗目的。

## 三、关注治疗性文献的质量

由于治疗性研究文献十分丰富,临床医生若要将检索的全部相关文献予以科学评价,其难度很大。为此,许多临床科研方法学家、文献资料的分析评论家以及有关临床学家共同合作,就有关临床特殊问题,在广泛收集大量文献和严格评价的基础上,进行了可贵的二次文献研究,提供了许多治疗性文献的系统评价(systematic review),其中以 Cochrane Collaboration 项目为代表,为推动临床治疗水平的提高,作出了很好的贡献。通常只要是所收集的文献(原始研究)质量高,同时高水平评估者(专业与方法学)又能做到客观评估,这样的结论自然科学可靠(内部真实性)又具有代表性(或外部真实性),故列为一等级别的最佳证据。

英国医学杂志(BMJ)、美国内科学杂志俱乐部(ACPJC)联合编辑的 Clinical Evidence 以及英国循证医学杂志(JEBM)发表的有关原始治疗文献的详细摘要(synopses)附专家点评,为治疗性证据进行了高质量的评价,已成为临床医生实践循证医学、获取最佳证据的主要来源。

关于单个治疗性研究文献的质量分析与评价,将在下一节详述,不过就证据质量的总体而言,前瞻性研究较回顾性或断面研究的质量要好;在前瞻性研究中随机对照试验较非随机对照试验为佳;在随机对照试验中,大样本较中小样本量为好,盲法较非盲性试验质量为好。总之,单个研究中若能较好防止偏倚因素的干扰,其研究结果真实性就有保障、质量就高。

有关治疗性研究证据的质量和论证强度的等级水平、分类可参考本书相关章节。

## 第三节 单个临床治疗性证据的评价

由于提出的问题本身往往比较具体,而临床治疗涉及范围又广,因而在检索到的相关文献(证据)中很难有完全对口的系统评价或 synopses 证据,常常以单个原始研究的文献居多。因此,应用这些单个临床治疗性证据之前,要用科学的方法和质量评价标准对其进行分析与评价,首先评价研究证据的内部真实性(internal validity),如果是真实可靠的话,则进一步分析与评价该治疗措施(证据)是否有临床治疗的价值,即临床的重要性(importance);假设其不仅真实而且重要的话,则要分析这种证据(措施)有无适用性(applicability)。

现在就围绕上述“三性”阐述如何分析与评价单个临床治疗性证据。

### 一、真实性的分析与评价

#### (一) 证据是否源于真正的随机对照试验

最真实可靠的治疗性措施的效应证据,当然是来源于真正的随机对照试验(RCT)。众所周知,

无论是已知还是未知的偏倚或混杂因素,对 RCT 研究结果真实性的影响程度最低。因此,我们检索文献和获取证据应尽可能地采用 RCT 的结果。

但有些 RCT,由于设计方法不尽完善或本身存在严重缺陷,不能将其列为真正的随机对照试验。对于 RCT 应重点从以下方面加以考核:

1. 纳入 RCT 的研究对象是随机抽样的病例还是非随机抽样的病例?
2. 纳入的合格研究对象是采用什么随机方法分组的,又是通过什么程序操作的?
3. 纳入的合格研究对象是否采用分层后随机分组? 分层因素是什么?
4. 对于研究对象的随机分组是否采用隐蔽措施( *concealment* )即隐匿性随机分配( *concealed randomization* ),以使研究人员在分组病人时,不知研究对象究竟处于哪个组,观测与分析试验结果时也被“隐蔽”,于是可更好地避免选择与测量偏倚。

5. 注意试验分组后,组间基线是否平衡而无显著性差异? 如不平衡则应注意是否进行了某种校正(重要的预后因素,如年龄、病情等)。

如果以上五点都能从分析与评价的文献中找到满意的答案,那么该文献是真正的 RCT; 倘若某个随机对照试验仅仅以简单的“采用随机化分组”表达,而缺乏具体的随机化内容,则该证据的真实性值得商榷。

6. 假如收集的文献缺乏随机对照试验者,则要注意是否在收集文献时检索词限定过严或检索不全,必要时宜扩大检索范围,再检索文献以防漏检。如确实缺乏 RCT 者,则应对非 RCT 的文献进行分析与评价,以探求质佳的证据,对于非 RCT 的有关治疗性研究证据的质量水平则参考本书其他章节。但在分析与评价非随机对照试验证据时,还应注意以下几点:

(1) 凡研究所获的证据为阴性结果者,即无效或有害,或者弊大于利者,则可信度为高。因为绝大多数非随机对照试验往往是假阳性结果远较假阴率高,即报道的有效率往往水分过高。

(2) 如果是对难治的且预后很差的疾病进行的非随机对照试验,其结果(证据)显示佳,经分析而不像假阳性的证据者,则当属可信。如重症肝炎肝功能衰竭,应用中西医结合治疗降低病死率; 急性细菌感染所致败血症,应用敏感有效的抗生素治疗,能治愈者。这种情况符合“全”或“无”的规律。

(3) 有些疾病本身发病率很低,不可能搞临床随机对照试验,其证据只可能源于临床系列报道或病例报道。有些慢性非根治性疾病之患者,同时需要接受多种药物治疗,在决定弃弊保利的用药决策时,往往考虑单个病人的随机对照试验( *an individual randomized trial, n-of-1 trial* )的证据。

### (二) 所有纳入的研究对象是否随访完整,研究对象随访时间是否足够

当确认已经正确实施了随机方法之后,随之应考虑所有纳入的研究对象是否均进行了随访。因为任何观察病例的丢失,都会直接影响最后结果和证据的真实性。例如,疗效差的患者退出,在最终分析结果时,会使治疗效果偏高; 有的病人因为药物或者干预措施的副作用从治疗组中退出,可能会过低估计它的危害性。理想的情况是所有纳入的研究对象在研究过程中都没有失访,但这一点在实际临床研究中很难做到。

如何估计失访病例对试验结果真实性的影响呢? 一般原则是失访率应控制在10%以内,若失访率超过20%,研究质量会受到很大影响,结果也许会不真实。对于一般失访率影响的估计,常用的方法是将试验组全部失访的病例,在本组中当作无效病例处理,而对照组丢失的病例则全部计入有效病例内。如果仍然与原有结论一致,则可以接受原来的结果。

同时,应该确保随访期足够长,以保证获得重要的临床效应结果。例如,采用他汀类降脂药治疗高胆固醇血症患者,如果仅治疗和随访1个月,降脂效果未达到最佳,也不可能观察终点指标,如心脑血管并发症。随访的具体时间取决于目标疾病的病程特点,通常临床观察疗程至少数月,有的甚至需要1年以上方能充分显示防治措施的重要效果。

### (三) 是否对随机分组的所有研究对象进行了意向性治疗分析

被随机分配入组的病例,可因为各种原因出现失访。例如,因副作用发生出现中途停药者; 患者

依从性差而未认真按医嘱服药者以及发生沾染或者干扰者等。如果这部分研究对象不被纳入结果分析,必然会破坏随机化原则和基线的可比性,并最终影响结果的真实性。因此必须根据当初随机分组的情况对所有研究对象进行结果分析——“意向性治疗分析”(intention to treat analysis),即无论其是否接受或未接受确切的治疗药物,全部病例均按最初随机分配入组,进行最后分析。

在阅读随机对照试验文献时,要特别注意作者是否真正按照“意向性治疗分析”原则进行结果分析,而不仅仅是提及。

#### (四) 是否对研究对象、医生和研究人员采用了盲法

采用随机分组的原则可以最大限度地控制选择性偏倚。但在治疗性资料的收集过程中往往也会产生较大的测量性偏倚。如受试研究对象知道自己接受的是治疗措施、还是对照措施,研究者或结果测量者知道研究对象的具体分组情况,这些往往会夸大研究药物的效应等。实施盲法的目的是为了减少测量性偏倚以确保观察结果的真实性。盲法可以是单盲、双盲或三盲。

当无法对病人和医生实施盲法时(如外科手术),可以请其他医生评价临床记录或使用客观指标评价治疗效果。若使用临床记录,则应去除所有可能涉及破盲治疗措施的信息,以保证盲法真正的实施。

至于“双盲”中是对谁实施了盲法,不同的研究者有不同的解释。因此,在评价时不能只注意作者是否提及采用盲法,更重要的是对谁实施了盲法及实施的具体办法,以判断其正确性。

#### (五) 除试验方案不同外,各组患者接受的其他治疗方法是否相同

研究对象除了接受规定的治疗方案外,若还有意或无意采用了其他的具有类似作用的干预措施,必然也会影响结果的真实性。沾染(contamination)和干扰(co-intervention)即是常见的两种情况。前者是指对照组的患者接受了试验组的防治措施,使得试验组和对照组间的疗效差异减小;后者是指试验组或对照组接受了类似试验措施的其他处理,人为扩大或减小组间疗效的真实差异。因此,RCT要求除了研究因素之外,其他任何治疗包括支持治疗在组间都应一致,这样才可能排除各种偏倚的影响,以确保证据的真实性。

如果一项随机对照试验不符合上述五条标准的一条或多条,则需要考虑其研究结果的真实性。如果基本符合以上标准,那么,下一步就要对其重要性进行分析与评价。

## 二、治疗性证据的重要性分析与评价

一旦治疗性证据的真实性满足要求后,则对它的临床价值即其重要性应进行分析与评价,如有着重要的临床价值,方可用于具体的临床实践。

治疗性证据的重要性分析主要围绕两个方面:正面的有效性和负面的不良反应。只有疗效佳、负效小者方有临床价值。

鉴于单个研究所报导的效果证据,往往是平均效应(率或值)。因此,要注意分析其精确度的有效范围,供用证时参考。对于治疗性证据的重要性评价要注重以下两条:

#### (一) 治疗性证据的效应强度大小

1. 疗效强度 对于疗效的强度通常用率表示,即有效率、治愈率、病死率、病残率,即使有些计量的疗效指标,亦多转换为“有效率”等。

当比较分析RCT试验组与对照组的最终有效率时,可以发现两者的差异大小及检验是否有统计学意义,然而,这些“率”对量化临床重要程度略显不足,因此,在循证医学临床实践中,进一步提出了以下指标:

##### (1) 相对危险降低率(relative risk reduction, RRR)

$$RRR = \frac{CER - EER}{CER} \quad (\text{式11-1})$$

注: CER=control event rate(对照组事件率); EER=experiment event rate(试验组事件率)。

##### (2) 绝对危险降低率(absolute risk reduction, ARR):

$$ARR = CER - EER \quad (式11-2)$$

(3) 需治疗多少病例数才能获得一例最佳效果 (number needed to treat, NNT)

$$NNT = \frac{1}{ARR} \quad (式11-3)$$

例如在一个应用他汀类药物预防脑卒中的 RCT 中, 报告了治疗5年的患者观察结果, 他汀组的脑卒中发病率为4.3% (EER), 对照组的脑卒中发病率为5.7% (CER), 根据以上公式, 计算量化指标:

$$RRR = \frac{5.7\% - 4.3\%}{5.7\%} = 25\%$$

$$ARR = 5.7\% - 4.3\% = 1.4\%$$

$$NNT = \frac{1}{1.4\%} = 72$$

基于以上指标进一步结合具体病种、病情严重程度、预期最佳治疗效果、现时水平以及临床专业知识等, 客观评定证据的临床价值。

故不能简单认为危险度降到多大或NNT值多小才有价值, 特别是ARR或RRI(相对危险增加率), 必须考虑对照人群预期事件发生率大小, 方能分析出它们的实际临床价值(表11-1)。

表11-1 他汀治疗预防脑卒中5年追踪效果

	组别	脑卒中事件率	RRR	ARR	NNT
A	试验组	4.3%	25%	1.4%	72
	对照组(假设)	5.7%			
B	试验组	0.0043%	25%	0.0014%	71429
	对照组	0.0057%			

为了更好地理解或应用NNT, 特列表11-2, 从中可以看到疾病不同或同种疾病(高血压及冠心病)不同疗效指标, 其NNT差别很大。因此, 若我们判断某种疾病的治疗获益, 与可靠证据的情景不同时, 需要进行校正处理。例如: 我们对表11-2中舒张期高血压(BP 90mmHg~109mmHg)但无靶器官损害的患者拟进行严格的干预, 药物治疗1.5年后, 究竟治疗多少例这类病人才能避免一例心脑血管病或病死呢? 于是可用表中的NNT进行校正性推算。

表11-2 临床观察获得的部分有用的NNT

疾病名称	防治措施	预防的临床结局	临床结局的发生率		观察期	NNT
			CER	EER		
I型糖尿病	胰岛素治疗	糖尿病神经病变	9.6%	2.8%	6.5年	15
急性心肌梗死	链激酶+口服阿司匹林	死亡	13%	8.1%	5周	19
			22%	17%	2年	24
不稳定心绞痛	低分子肝素(与肝素比较)	死亡、心肌梗死、复发性心绞痛	23%	20%	30天	29
舒张压 (115~129)mmHg	降压药	死亡、卒中、心肌梗死	13%	1.4%	1.5年	3
舒张压 (90~109)mmHg	降压药	死亡、卒中、心梗	5.5%	4.7%	5.5年	128
良性前列腺增生	口服非那雄胺(保列治)	外科手术	10%	4.6%	4年	18

$$NNT_{\text{校正观察期}} = NNT_{\text{原观察期}} \times \left( \frac{\text{原观察期}}{\text{校正观察期}} \right) \quad (\text{式11-4})$$

那么代入公式:  $NNT_{\text{校正观察期}} = 128 \times \left( \frac{5.5}{1.5} \right) = 470$

即对该类患者严格降压治疗1.5年后,为防止一例患者心脑血管病或死亡,则需要治疗470例患者,据此,则可评价其临床价值。

2. 负效值的强度 临床治疗性试验中,特别是与安慰剂比较时,通常新药更易发生不良反应,甚至较重的药物不良反应(ADR)。因此,在分析和评价治疗性证据时,一定要注意不良反应的各组发生率、种类及强度,如同分析评价正面疗效一样,进一步测试:

(1) 相对危险增加率(relative risk increase, RRI)

$$RRI = \frac{EER - CER}{EER} \quad (\text{式11-5})$$

(2) 绝对危险增加率(absolute risk increase, ARI)

$$ARI = EER - CER \quad (\text{式11-6})$$

(3) 需治多少病例才发生一例不良反应(number needed to harm, NNH)

$$NNH = \frac{1}{ARI} \quad (\text{式11-7})$$

如上述他汀药的 RCT 中,试验组不良反应率0.05%,对照组0.03%,则该研究:

$$RRI = \frac{0.05\% - 0.03\%}{0.05\%} = 40\%$$

$$ARI = 0.05\% - 0.03\% = 0.02\%$$

$$NNH = \frac{1}{0.02\%} = 5000$$

由此可以看出,他汀类药物治疗的不良反应发生率很低,相对而言较为安全。

## (二) 疗效证据精确度的估价

上述有关疗效和负效指标的测量值,仅表示效应强度的平均水平,如疗效率80%或负效率10%等以及ARR、RRR、NNT等。但效应并不是固定于某一特定值或代表值,它有一定的变化范围,因此,就提出了效应强度的精确度(precision)问题。

有关效应强度的精确度或范围,通常用其95%可信区间(95%CI)表示,在分析和计算95%CI时,可参考本书相关章节。区间范围宽窄反映了精确度的范围,95%CI范围越窄则越精确。

## 三、治疗性证据适用性的分析与评价

上述证据的真实性与重要性经分析和评价获得肯定结论之后,那么这种有价值的证据是否可被应用于自己的临床实践中呢?这要结合自身病人的实际情况,分析与评价它的适用性。

通常以下几条规则要予以考虑:

### (一) 被评价的证据是否与病人情况不符而不能应用

1. 整体证据 任何治疗性研究所产生的有价值证据,并不是不加分析照搬照用,一定要与拟治疗的患者“对号入座”,否则就要出偏差。应重点考虑:疾病的诊断标准是否可靠,证据中的研究对象纳入标准是否与拟引证的病人相符,其生理功能与病理学的依据、病情特点、年龄、性别以及社会经济状况是否存在显著差异等,假若以上特点一致或大体一致,则该治疗性证据就可适用,否则就不可取。

2. 亚组证据 在分析与评价治疗性证据时,也应注意是否进行了亚组的疗效分析。有的证据在总体上可能缺乏适用性,然而,在一些亚组中可能有实际意义,如果患者病情与某亚组患者的病情相似,那么,这个亚组的治疗证据就有适用价值。

特别是符合下述条件者,则有着重要的适用价值:①确有生物学和临床依据者;②确有统计学和临床意义者;③亚组本身的研究假设是在研究前设计的,而不是在试验中有意而为者;④该证据在另外的研究中可被证实者。

### (二) 在我们的医疗环境里对拟采用的治疗证据是否可行?

对拟采用的有效治疗措施,需要在一定水平的医院和特定的医疗环境和条件下才能被采用,可能受医生的技术水平、医院的管理机制及设备条件、病人的意愿以及经济承受能力等限制。如冠心病的介入治疗、风湿性心脏瓣膜病换瓣手术等,即使这类治疗证明对病人有利并效果颇佳,但若不具备上述条件,也无法实施。

### (三) 基于该治疗性证据,对患者可能产生的利、弊进行估计

假若对患者拟采用某一可行性好的最佳治疗措施,那么,应对该措施可能给患者带来的利弊进行综合评价,一定要利大于弊,且要有利弊量化指标的依据。

治疗措施的利弊量化指标,最直接的是使用 NNT(益处)及 NNH(害处)。若治疗证据中无此指标,则可用两种办法帮助解决。

一是确定患者预期事件发生率(patient's expected event rate, PEER),意思是若患者不予治疗,其最终结局事件的发生率,可用治疗试验中安慰剂对照组的事件发生率(CER)表示;如无 CER 证据,也可根据临床积累的未进行治疗或缺乏特效治疗患者的观察结果作为 PEER 参考值,如急性心肌梗死患者 PEER 约15%;另外也可从亚组资料分析中获取相应的 CER 作为 PEER 等。

当获得 PEER、RRR 及 RRI 等指标值后,用下列公式则可推算 NNT 和 NNH:

$$NNT = \frac{1}{PEER \times RRR} \quad (\text{式11-8})$$

$$NNH = \frac{1}{PEER \times RRI} \quad (\text{式11-9})$$

另一个方法是应用列线图,用已知的 PEER(或称不治疗的绝对危险率)与 RRR 值固定两点,两点的连线与 NNT 线上的交叉点,即为该 NNT(图11-1)。该法非常快捷方便。

如果我们的患者确实与引用证据中的患者不同,例如发生不利结果的概率(CER)大1倍,可用校正因子 $f_i=2$ 表示;相反如不利结局发生率是原来的50%,则 $f_i=0.5$ 。 $f_i$ 值可根据我们自己的临床经验决定。用 $f_i$ 除证据中提供的 NNT 值,即作为校正后的 NNT。以干扰素处理多发性硬化症为例,NNT 为9,如我们的患者不治疗,在相同观察期内,发生致残率比试验中患者大3倍,则 NNT/ $f_i$  为9/3=3。意即用干扰素治疗危险性更高的多发性硬化症患者,每治疗3个病例,就可避免1例致残的后果发生。

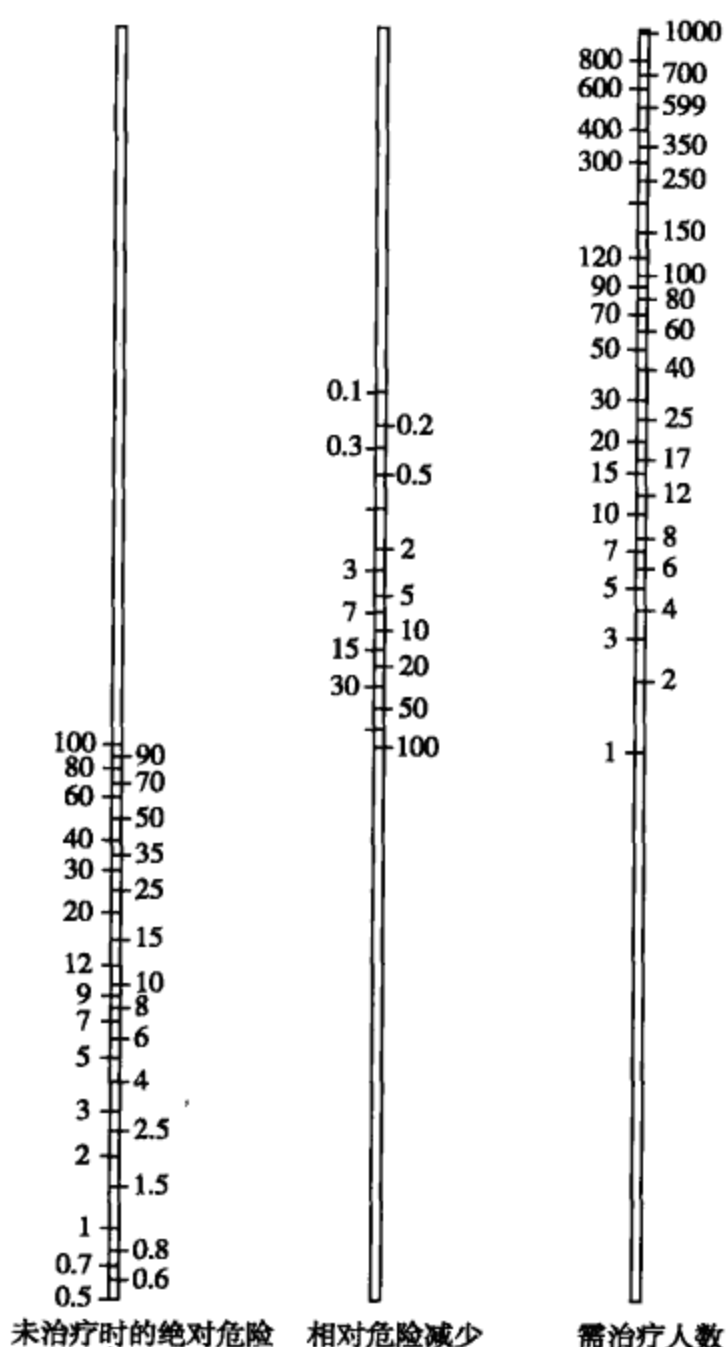


图11-1 估计 NNT 的列线图

(Chatellier et al, BMJ, 1996, 312:426-429)



同理,可利用不同的副作用发生概率,对 NNH 进行校正。即用不同假设值  $f_h$  除 NNH,  $NNH/f_h$  值愈大,则拟用的防治措施安全性愈高。反之,则对自己患者的安全性就小而不利。

#### (四) 患者对于治疗措施的价值取向与期望

在循证医学治疗实践中,对拟采用措施(证据),作为主导者的临床医生在决策时,一定要尊重患者对治疗的价值取向,即愿意接受或者不愿意接受,或愿意接受哪一种备选者(药物或有关治疗措施),而且要了解患者对治疗结局的有关期望。

1. 对任何拟采用的治疗措施或药物,一定要把安全放在第一位,即不良反应要最小化;同时还要注重疗效。因此估计与掌握治疗的利弊比,就显得十分重要。通常用治疗措施(或药物)的 NNT 与 NNH 计算其利弊比(likelihood of help vs harm, LHH)

$$LHH = \frac{\left(\frac{1}{NNT}\right)}{\frac{1}{NNH}} \quad (\text{式11-10})$$

例如他汀药物治疗预防脑卒中的 NNT 为 72, NNH 为 5000, 则:

$$LHH = \frac{\left(\frac{1}{72}\right)}{\frac{1}{5000}} = 70$$

这意味着有利面是不利面的 70 倍,是为安全、有效的治疗证据,显然 LHH 越

高越佳。

2. 如果有多种备选药物,且它们的疗效与不良反应相似或有差异,但没有显著意义。对这些备选治疗措施(或药物),则应优先选择那些成本(价格)低廉、有效安全的药物,在保证安全有效的基础上,尽可能降低医疗成本,这在一定程度上还可以帮助百姓解决“看病贵”之苦。

3. 对于任何治疗措施,一定要让患者了解清楚,有利于患者接受治疗,保持良好的依从性。

4. 在治疗过程中务必要认真观察治疗反应,关心与爱护患者,这有利于增进互信、建立和睦的医患关系,避免产生不必要的误解或纠纷。

以上系对单个治疗性研究之证据,从其真实性、重要性以及适用性三个方面进行了严格评价,学习掌握与应用这些标准和方法对循证医学治疗性实践乃是至关重要的。

## 第四节 系统评价或 meta 分析证据的评价

系统评价(systematic review, SR)及 meta 分析的定义、方法过程及指标计算在本教材的相关章节已有较为详细的介绍。现仅就其评价问题进行叙述。因为系统评价是由多个单项研究综合而成,故许多内容(尤其是评价重要性与适用性的部分)与评价单项试验是相同的,但也存在不同之处。

### 一、系统评价的真实性评价

评价系统评价的真实性,主要从以下四个方面进行:

#### (一) 是否是随机试验的系统评价

首先应该确定该系统评价是针对随机试验还是非随机化试验。如果是前者,必须明确所查找的每篇文献是否真的是随机试验,其中是否包含了非随机试验文献的混淆;必须明确文献质量的评价标准;而且对每篇合格文献都进行严格的评价;同时还要考虑文献间的同质性和异质性,同质性的结果优于异质性结果。

#### (二) 是否对文献的检索过程有详尽的描述

评价时应仔细阅读总结报告中与检索有关的方法学部分,特别是检索策略制定的合理性,达到既完整又准确。首先应该知道检索范围是否广泛,主要的医学文献数据库是否均被包括,关键词运用是否合理;除了电子检索外,是否采用包括手工检索期刊、会议记录、各种论文、制药公司的数据库

以及联系已发表文章的相关作者等多种途径;是否只局限于单一语种等。对以上内容应该进行较为详尽的描述。SR 应反映明确的方法学,如设计是否完整、文献收集的方法、纳入与排除标准、对单个 RCT 评价的质量标准、数据的收集与整理、防止偏倚的措施、统计分析方法、结果的评价等。如果对系统评价收集的随机试验,均有相互独立的真实性评价,而且彼此间有很好的一致性,则更有把握接受系统评价的证据。

### (三) 是否对单项试验做了真实性(即研究质量)的评价

在系统评价的方法学部分还应当说明调查者是如何评价每一篇入选文献的真实性,明确选用的评价标准,提出对入选文献纳入的方法,即是否由两人或多人独立进行评价。有很好的一致性时,说明系统评价结果更加可信。

### (四) 分析时采用原始病人数据还是集合数据

在进行分析时采用的数据是来自原始的病人资料还是表格资料或是已发表的报告?若利用的是原始数据,则综述结论更可信。使用原始资料还可对病人发生某一临床事件的时间进行更可靠的分析。

## 二、系统评价的重要性评价

如果系统评价的真实性不存在问题时,接下来需要研究分析系统评价结果的重要性。主要考虑:治疗效果在各个研究中是否一致?治疗效果有多大?治疗效果的精确度如何?

### (一) 不同原始研究的结果是否一致

如果每一个试验结果都显示了相似的效果或者至少是相同的方向,即有较好的同质性,那么该系统评价的结果令人相信。但在实际研究中,达到完全的同质是不可能的,表现为部分研究的结论有效,部分的无效,甚至有的表明有害。此种情况下更需要进一步分析。但最好能对异质性结果给予合理的解释。

### (二) 治疗效果有多大

有较多指标可以用于系统评价效果大小的评价,如 NNT、RRR、ARR、RRI、ARI、NNH 等。目前已有许多系统评价用 NNT 表示结果,因为 RRR 未考虑 CER 和 PEER,OR 和 RR 也具有同一缺陷且在临床实践中应用较少。故已有一些工具将 RR、OR 转换为 NNT。部分常见的转换可参见表 11-3 及表 11-4。

表11-3 OR-NNT转换表 (OR<1)

病人预期事件发生率 (PEER)	OR 值						
	0.9	0.8	0.7	0.6	0.5	0.4	0.3
0.05	209 <sup>(a)</sup>	104	69	52	41	34	29 <sup>(b)</sup>
0.10	110	54	36	27	21	18	15
0.20	61	30	20	14	11	10	8
0.30	46	22	14	10	8	7	5
0.40	40	19	12	9	7	6	4
0.50	38	18	11	8	6	5	4
0.70	44	20	13	9	6	5	4
0.90	101 <sup>(c)</sup>	46	27	18	12	9	4 <sup>(d)</sup>

(a) 此处相对风险减少 (RRR) 为 10%; (b) 此处 RRR 为 49%; (c) 此处 RRR 为 1%; (d) 此处 RRR 为 9%。(引自 John Geddes, 1999)

表11-4 OR-NNT转换表 (OR&gt;1)

病人预期事件发生率 (PEER)	OR 值						
	1.1	1.25	1.5	1.75	2	2.25	2.5
0.05	212	86	44	30	23	18	16
0.10	113	46	24	16	13	10	9
0.20	64	27	14	10	8	7	6
0.30	50	21	11	8	7	6	5
0.40	44	19	10	8	6	5	5
0.50	42	18	10	8	6	6	5
0.70	51	23	13	10	9	8	7
0.90	121	55	33	25	22	19	18

表中数字是在特定 PEER 水平时, OR 值对应的 NNT。此表适用于治疗引起的效应与副作用。(引自 John Geddes, 1999)

### (三) 证据效果的精确性如何

与对单项研究证据结果评价一样,仍需要采用95%可信区间来评价系统评价效果的精确性,以进一步估计效果的强度及临床意义。

### 三、系统评价的适用性评价

由于系统评价中的各个研究可能来自不同人群,其总体效应估计值是平均的疗效。因此,即使系统评价结果已经具有了真实性、重要性,但也不一定适用于具体的病人,而是仍需要进行适用性评价。

系统评价适用性评价原则与评价单项研究相同。即考虑:病人与研究中的病人是否存在很大的差别?治疗是否适于现实的情况?治疗对病人的潜在利益和损害有哪些?患者对治疗结局和提供的治疗方案的态度和期望如何?但与单项研究评价不同的是,系统评价会清楚地给出不同病人或不同情况下各亚组的信息结果,从而会更方便地应用于自己的病人。但在应用时必须满足亚组病人的应用条件。

在国际上,有关治疗性证据的系统评价的工作,以 Cochrane 项目做得最好,颇有价值,但对于 SR 亦应在采用的时候,采取严格评价的态度,切不可盲信。因为最佳的 SR 必以最佳的单个 RCTs 为基础,如果 RCT 质量不高、SR 研制者专业水平不高、存在有关偏倚多,那么,即使是 SR 也不一定可信。

文献评价实例(参见光盘辅助教材相关章节)。

(时景璞 王家良)

## 第十二章 药物不良反应的循证分析与评价

### 第一节 概 述

药物是治疗疾病的最重要手段之一,广泛用于临床日常工作,对控制、治疗疾病有重要意义,同时药物产生的不良反应也可能给患者带来损害,甚至威胁患者生命;随着医学科学的发展,新药物不断涌现,科学、客观评价新药物的有效性、安全性,并合理应用于临床工作,为患者服务成为临床医生所应具备的基本能力。本章对药物反应的基本概念、分类、产生原因、循证分析与评价进行介绍。

#### 一、药物不良反应的概念与分类

##### (一) 与药物反应有关的几个基本概念

药物反应是药物通过各种途径进入体内,机体对药物产生的生理、病理反应的自然过程,从严格的意义上讲,药物的治疗作用、有效性也属于药物反应范畴,而伴随药物治疗疾病的同时,药物可能引起人体的不适感,甚至机体功能或器质性的损害,此乃因药物治疗所致不良反应,是药物反应研究的重点内容。

1. 药物副作用(side effect) 药物副作用是指在正常剂量情况下出现与用药目的无关的反应。副作用与正作用(治疗所期望效应)是相对的,其差异在于用药目的不同;如阿托品有解痉与抑制腺体分泌两种作用,当作为麻醉前用药时,用药目的是抑制腺体分泌,而术后引起的肠胀气、尿潴留为它的副作用;当用于解除胆道痉挛时,口干则成为副作用。从旧有药物的副作用出发,寻找治疗疾病的新靶点,可以节省大量人力、物力、财力及时间,成为药物研究的新途径;而副作用在药物不良反应的范畴中,是指那些引起人体的不适感比较轻微,多为可逆性,停药后通常很快消失的反应。

2. 药物不良事件(adverse drug event, ADE) 药物不良事件(ADE)是指临床病人接受某种药物干预后,所发生的某种与干预相反的不良反应,而这种不良反应是否就是该药物所致,其因果关系尚待确定。因此,一旦发生 ADE 时,应力求采用科学的方法,探讨它的真正的因果关系殊为重要。

各国或各机构的“不良事件”(adverse event, AE)定义不尽相同;国际协调会议(ICH)将 AE 定义为患者在应用药物时出现的不利临床事件,该事件和治疗未必一定有因果关系。FDA 的 AE 定义为患者在应用任何剂量的药物、医疗器械、特殊营养品时,出现的可疑不良后果。该损害可由药物不良反应、用药失误、手术不当、血液输注错误等诸多原因所致。

有学者认为 ADE 的含义与 AE 相同,且易与 ADR 混淆,故建议将药物不良事件改为“可疑药物不良反应”(suspected adverse drug reaction),这是因为在 ADR 报告的实际操作中,通常是按“可疑即报”的原则进行报告的。

3. 药物不良反应(adverse drug reactions, ADR) WHO 国际药物监测合作中心对药物不良反应(ADR)的定义,指正常剂量的药物用于预防、诊断、治疗疾病或调节生理功能时,出现的有害的和与用药目的无关的反应。从定义可看出,药物不良反应的特点,第一是正常剂量,排除过量用药引起的中毒反应;第二对机体造成了伤害,伤害需要有表现形式;第三,伤害是因为所用药物引起的,即伤害与用药有因果关系。我国《药品不良反应报告和监测管理办法》第二十九条,对药品不良反应的定义:“是指合格药品在正常用法用量下出现的与用药目的无关的或意外的有害反应”。

药品不良反应是药品的一种属性,往往是不可避免的,但是通过监测分析,可以采取一些措施,减少药品不良反应的危害。

## (二) 药物不良反应分类

任何药物都可能会引起不良反应,但是由于人与人之间的个体差异,不同的人对同一种药物的不良反应表现可以有很大的差别;药物不良反应可以从不同方面进行分类:

1. 根据发生机制分类 在药物不良反应研究中,经典的分类方法是按反应是否与用药剂量有关分为:

(1) A类药物不良反应:又称为剂量相关的不良反应(dose-related adverse reactions),是因为药物的药理作用增强所致,与用药剂量呈正相关;如苯二氮草类药引起的嗜睡,抗凝药所致的出血等;该类反应多可以预测,发生率高而致死率低。

(2) B类药物不良反应:又称剂量不相关的不良反应(non-dose-related adverse reactions)。这类反应与正常药理作用及剂量无关,但与机体特异性反应有关;如青霉素过敏性休克,抗癫痫药的超敏反应等;虽然发生率很低,但难于预测,往往致死率高。

2. 根据发生频率与危害程度分类 由于药物的本身特点、制剂与生产的工艺水平,以及用药后机体反应的各异。因此,ADR发生的频率以及对人体的危害程度是十分不同的。因此,为了有效防止ADR发生与正确及时地处理。故将ADR从发生率与危害程度进行了如下分类:

(1) 国际医学科学组织委员会(Council for International Organization of Medical Sciences, CIOMS)推荐用下列术语和百分率表示药物不良反应发生频率:

- 1) 十分常见: 发生率 $\geq 10\%$ ;
- 2) 常见: 发生率 $1\% \sim 10\%$ ;
- 3) 偶见: 发生率 $0.1\% \sim 1\%$ ;
- 4) 罕见: 发生率 $0.01\% \sim 0.1\%$ ;
- 5) 十分罕见: 发生率 $< 0.01\%$ 。

(2) 根据严重程度分类: 根据药物对人体伤害的程度分为六级:

- 1级: 轻微不良反应,停药后很快好转,无需治疗;
  - 2级: 造成病人短暂损害,需要治疗或干预,但不需要住院或延长住院时间,易恢复;
  - 3级: 造成病人短暂损害,需要住院或延长住院时间(超过7天);
  - 4级: 造成病人永久性损害(系统和器官永久性损害、残疾);
  - 5级: 对生命有危险(休克、窒息、昏迷、发绀等)需要急救的症状;
  - 6级: 死亡。
- 1级为轻度,2级为中度,3级以上为重度。

## 二、药物研发与安全性研究

任何新药的研发在进入临床研究之前,首先应具备一系列的研究基础,其中通过药物化学研究以确定它的有效化学结构;通过药理与毒理学的研究以弄清其药理效应之基础以及其毒理效应,包括致畸、致基因突变和可能的致癌效应;还有它的药物代谢和动力效应等。在这一系列的试验研究中,当出现严重不良事件时,研究可能随时中断。

只有在上述一系列的试验研究中获得成功的基础上,即有充分证据证明了所研发的新药,具有对某种疾病的防治价值,并且具有安全性,方可按照国家对新药研发的规定,提供系列可靠的基础研究资料,然后向国家药品行政管理部门提出申请,经批准后,方进入临床试验阶段。

临床试验分为4期:

- I期临床试验: 初步的人体安全性评价试验;观察人体对于新药的耐受程度和药物代谢动力学。
- II期临床试验: 随机双盲对照临床试验;对新药有效性及安全性进行初步评价。

Ⅲ期临床试验:扩大的多中心临床试验;遵循随机对照原则,进一步评价有效性、安全性。

Ⅳ期临床研究:新药上市后进行的临床监测。在广泛使用条件下考察疗效和不良反应,药物上市后,使用的人群更广,表现的不良反应也会更多,有些少见的,甚至罕见的不良反应也会发现;同时长时间的监测,远期不良反应也会逐渐显现出来。

### 三、药物不良反应监测发展

虽然药物在上市前已经过动物试验和临床试验,但这些试验不足以保证药物的安全性,除了动物与人存在种属差异之外,Ⅰ期、Ⅱ期、Ⅲ期临床试验,病例数量是不够大的,试验过程也不可能为长期,且在试验中很难不受机遇因素的影响,对ADR发生率低( $<1\%$ )及在特殊人群中才能发生的不良反应往往不易被发现,所以药物上市后的安全性监测十分重要。上市后若发现严重不良反应,仍然可能采取停止使用、停产、召回等措施。历史上有过多起严重药物不良反应载入史册:

1. 氯碘喹事件 氯碘喹是1933年上市用于治疗肠道感染的药物;在50年代末期,日本出现不少人患上急性脊髓神经炎,表现为双足麻木、刺痛无力、瘫痪失明。1967年日本曾成立专门委员会对该病病因进行调查,1971年查清,该病的发生与使用氯碘喹有关,当时在日本因服用此药而引发急性脊髓神经炎病患者高达11 000多人,死亡数百人。

2. 碘胺胍剂事件 碘胺胍剂是一种消炎药,20世纪30年代在美国已广泛使用,1937~1938年,发现该药物造成肾脏功能的严重损害,发生尿毒症,导致约358人肾功能衰竭,107人死亡。

3. 反应停事件 1956~1961年,联邦德国格仑南苏药厂生产反应停(沙利度胺),用于治疗妊娠期呕吐,当时在17个国家应用,但短短的几年后,即发生了严重的畸胎事件,一些服用了该药的孕妇,所生婴儿呈四肢短小儿,形似海豹,被称之为“海豹婴儿”,当时受害婴儿达1万多人,死亡5000多人(图12-1,彩图12-1)。



图12-1 “海豹婴儿”

反应停事件(thalidomide incident)后,1968年世界卫生组织(WHO)制订了国际药物不良反应监察合作计划;最早参加国有德国、荷兰、瑞典、英国、丹麦、以色列、澳大利亚、新西兰。1970年正式成立了WHO国际药物监察合作中心(WHO Drug Monitoring Centre)。1978年迁至瑞典的东部城市乌普萨拉(Uppsala),称为世界卫生组织国际药物监测合作中心(WHO Collaborating Centre for International Drug Monitoring)。1997年WHO国际药物监测合作中心更名为乌普萨拉监测中心(Uppsala Monitoring Centre, UMC)并调整了内部组织机构。自1968年至2009年6月全世界共95个国家先后正式加入了WHO国际药物监测合作计划,另外尚有准成员国28个,中国已于1998年成为该计划的正式成员国。UMC已收到来自成员国的可疑药物不良反应(suspected adverse drug reactions)报告460余万份。这些病例报告已成为了解和评价药物安全性的重要依据之一。

## 第二节 药物不良反应诊断方法

### 一、药物不良反应诊断的意义

临床上在药物使用过程中所发生的一切不良事件,关系到医疗决策及病人的利益;严重不良事件甚至威胁患者生命;尽管发生不良事件的原因很多,可能发生问题的环节也很复杂,但明确不良事件与药物有无因果关系,是不是药物不良反应,是哪种药物的不良反应,对机体会造成什么损伤,是否会出现严重后果等问题,涉及需不需要马上停药、换药;停药后病人会出现什么问题,还有什么其他有效药物替代等医疗决策问题;还涉及是否要追究药物制造者责任等法律问题;同时还涉及其他患者能否使用,将来能不能继续使用等问题。重大药物不良反应的诊断,除了能及时挽救患者生命外,

还可能终止所有临床使用、停产、召回,挽救更多患者生命等长远问题,所以药物不良反应的诊断具有重大意义;相反不合适的停药、换药也会给临床治疗带来困难,达不到良好的治疗效果。明确不良事件与药物的因果关系,从而确诊为药物不良反应,需要科学的诊断及循证医学证据。

## 二、药物不良反应相关因素

研究药物不良反应的目的是为了预防不良反应的发生,中断不良反应的进展、改善不良反应的后果,达到以上目的必须了解与药物不良反应有关的因素。

药物进入人体后引起机体反应的表现及强度与药物及人体两方面因素有关:

### (一) 药物因素

1. 药物化学结构 化学结构相似的同类药物应用过程中常有类似的不良反应发生,如青霉素类药物能引起过敏反应,化学结构相类似的氨苄西林、羧苄西林等都能引起过敏反应。

2. 药物的剂量 药物应用过程中随着药物剂量的增加不良反应发生的概率及程度也有所增加。

3. 生产过程 药物不良反应的产生除药物有效成分外,还与生产过程的杂质、添加剂含量,氧化、分解、降解、聚合产物的微小差异等有关。如不同批次药物,不良反应发生情况有所不同。

4. 疗程 药物体内有累积效应,连续用药时间越长,发生不良反应可能性越大。

5. 药物间的相互作用 临床治疗中常有几种药物同时或先后应用,称联合用药;所联合的药物之间存在相互作用,有些药物联合会增加治疗效果,减少不良反应的发生;有些药物联合则相反,甚至出现严重不良反应,联合用药之前应查“配伍禁忌表”。

### (二) 机体因素

1. 遗传因素 药物不良反应在不同的种族的患者身上存在差异。药物在体内的代谢受代谢酶的影响,不同种族由于基因多态性,酶活性的差异引起药物在体内代谢产生差异;药物不良反应在不同种族人群中的发生率及严重程度不同。

2. 性别 女性存在月经期、妊娠期、哺乳期、绝经期等生理时期,激素水平波动较大,一般而言女性对药物不良反应较男性更为敏感。

3. 年龄 新生儿和婴幼儿各系统器官发育不成熟,肝脏对药物的解毒功能差,肾脏对药物的排泄能力差,对药物敏感性高,不良反应的发生率较高;老年人随着年龄的增加,各系统、器官功能衰退,药物不良反应发生率也有所增加。

4. 用药者的疾病状况 某些疾病能改变药代动力学、药效动力学作用,从而影响药物不良反应的发生。肝脏是药物主要代谢器官,肾脏是药物及其代谢产物的主要排泄器官,肝脏、肾脏疾病使血药浓度增加,产生药物不良反应。

## 三、药物不良反应的诊断方法

药物不良反应的诊断是判断不良事件与所使用药物之间有无因果关系,有些治疗是多种药物联合应用,还需要判断是哪个药物引起的不良反应。不良事件与药物因果关系的判断是临床工作中十分常见又非常重要的事情,有时也是相当困难的事情。ADR 因果关系评价( causality assessment )是对药物使用过程中发生的不良事件进行因果关系确认的方法,是药物安全性监测管理中一项十分重要而复杂的步骤;目前,国际上对 ADR 因果关系评价有多种方法。

### (一) Karach和Lasagna方法

Karach 和 Lasagna 方法是最常用的方法。

1. 因果关系判断的指标 可从以下几个方面判断不良事件是否与药物有关:

(1) 用药与反应出现的时间顺序是否合理;

(2) 反应是否符合该药已知的不良反应类型;

(3) 停药或降低用量,可疑不良反应能否减轻或消失;

(4) 反应症状清除后再次用药后是否再次出现同样反应;

(5) 有否其他原因或混杂因素: 相关的病理状况、合并用药、现用疗法、曾用疗法来解释。

2. 因果关系判断的标准 通过应用以上五项指标对不良事件的判断, 可初步得出以下结论:

(1) 肯定( definite): ①用药以来的时间顺序是合理的; ②该反应与已知的药物不良反应相符合; ③停药后反应停止; ④重新开始用药, 反应再现; ⑤其他原因不能解释。

(2) 很可能( probable): ①时间顺序合理; ②该反应与已知的药物不良反应相符合; ③停药后反应停止; ④无法用患者疾病来合理地解释。

(3) 可能( possible): ①时间顺序合理; ②与已知的药物不良反应符合; ③患者疾病或其他治疗也可造成这样的结果。

(4) 有条件( conditional): ①时间顺序合理; ②与已知的药物不良反应不符合; ③不能合理地以患者疾病来解释。

(5) 可疑( doubtful): 不符合上述各项标准。

## (二) 计分推算法

Naranjo 的 APS 评分法( Naranjo, 1981 Adverse drug reaction Probability Scale ) 是国际上比较常用的评价方法。评分法是对用药与反应出现的时间顺序、是否已有类似反应的资料等基本问题予以打分, 最后按所计总分评定因果关系等级, 详见表12-1。

表12-1 APS计分

问题	是	否	不详	记分
1. 该反应以前是否已有报告	+1	0	0	
2. 本例 ADR 是否在使用所疑药物后出现	+2	-1	0	
3. 当所疑药物停用后, 使用特异的对抗剂之后不良反应是否改善	+1	0	0	
4. 再次使用所疑药物, ADR 是否再次出现	+2	-1	0	
5. 是否有其他原因( 药物之外) 引起这种反应	-1	+2	0	
6. 当给安慰剂后这种反应是否能再次出现	-1	+1	0	
7. 血( 或其他体液) 的药物浓度是否为已知的中毒浓度	+1	0	0	
8. 增大药物剂量, 反应是否加重; 减少药物剂量, 反应是否减轻	+1	0	0	
9. 患者以前用过相同或类似的药物是否也有相似的反应	+1	0	0	
10. 该不良反应是否有客观检查予以确认	+1	0	0	

根据以上指标打分, 计算总分, 总分  $\geq 9$  分, 肯定有关( definite); 总分 5~8 分, 很可能有关( probable); 总分 1~4 分, 可能有关( possible); 总分  $\leq 0$  分, 可疑( doubtful)。

## 第三节 药物不良事件证据的查询与评价

### 一、带着问题查询

首先判断问题的真实性, 所要查询的不良事件是什么? 能否用客观指标检测? 可能的归类? 查询的问题是这项不良事件有人遇到过吗? 在哪里可能有记载? 假若没有怎么办? 所以需要了解所有可能的证据来源, 从最简单、最容易获取的证据入手, 制定全面的证据收集计划, 并在收集过程中, 对每项证据进行严格评价, 决定是否采信, 能否指导自己的决策。

### 二、资料检索与证据来源

有关药物不良反应或不良事件记载、研究的证据很多, 因记载目的不同, 而研究的深度不一样,



有些仅是客观地记载了有人发现过的事件,有些是对某事件进行深入研究,证实其有因果关系,或因果关系的强弱等;合理地使用证据是临床医生的基本功。就临床医生而言,可以查询获取的证据有以下几种:

1. 药物说明书 来源最方便,药物包装中均附带有药物说明书,看看是否有类似记载;当然作为医生在使用药物之前除了药物的治疗作用外,应充分了解药物的副作用、不良反应,甚至是不良事件。

2. 专著 药学专著中有药物不良反应的发生机制,理论性较强。

3. 期刊 对同一问题可能有广泛研究,信息量大,但每项研究因研究设计、研究方法等不同,其论证强度及可靠性存在差异,需认真评价。

4. 数据库 信息量浩瀚,原则为“可疑即报”;以下为国际权威数据库:

(1) <http://www.who-umc.org/> 世界卫生组织不良反应数据库( WHO Adverse Reaction Database );

(2) <http://www.fda.gov/Drugs/InformationOnDrugs/default.htm> FDA 药物批准和数据库( FDA Drug Approvals and Databases );

(3) <http://www.emea.europa.eu/htms/human/phv/communications.htm> 欧洲药物管理局药物警戒指导方针和文件( EMEA Pharmacovigilance guidelines and documents );

(4) <http://www.mhra.gov.uk/Safetyinformation/index.htm> 英国药物和健康产品管理局安全信息( Safety information:MHRA );

(5) <http://www.hc-sc.gc.ca/dhp-mps/medeff/bulletin/index-eng.php> 加拿大卫生部药物不良反应时事通讯( Canadian Adverse Reaction Newsletter )。

### 三、证据的评价

从资料检索中所获得的有关 ADR 证据是否真实可靠和可信,则必须进行严格评价( critical appraisal )。

鉴于药物临床试验最主要的目的是验证疗效,而非主要研究它的不良反应( I 期临床试验除外),又因成功的临床试验往往出现的不良反应轻,且发生率低,所以,在单个的严格的临床科研文献中,发现罕见的、特别是严重的 ADR 证据往往也是极少见的。因此,许多资料从临床科研严格设计的角度,往往证据真实性的论证强度是有限的。但是,评价 ADR 资料质量而言,仍然要抓住证据的真实性、重要性与适用性三个方面。

#### (一) 真实性

评价证据的真实性,共有4条主要的参考标准。

1. 除了治疗措施或其他相关因素之外,资料中所涉及的一组病人,其主要临床特点基线状况是否清晰界定? 而且是否组间相似或一致?

例如两组病人是否诊断清楚,组间的病例数量、性别、年龄、病情程度等是否一致。如果都是一致或相似,则有可比性。当治疗中出现了有关异常的 ADR,那么就可能与该治疗措施或其相关因素有关。

符合这一标准的往往是随机对照试验及高质量的 RCT 之系统评价( systematic review )的有关 ADR 证据,其他来自于队列研究或病例对照研究的有关 ADR 证据,其论证强度由于难免受混杂或测量等偏倚的影响,其真实性就相应的逊色了。而无对照的论述性的 ADR 报道,包括个别病例的 ADR 报告仅可作为参考依据。

2. ADR 证据是否为盲法观测的结果?

对于治疗研究主要观测的效应结果,其盲法测试与分析是至关重要的,因为它可以防止测量性偏倚,确保证据的客观真实性,这只能来源于随机盲法对照试验。其他试验设计方案由于难以执行真正的盲法,所以在 ADR 判断方面,会有不同质量的影响。

### 3. 对于 ADR 的观测追踪期是否足够长、资料是否完整?

由于常规的临床试验期限不可能持续过长,而 ADR 特别是当新药投入市场后的第Ⅳ期临床试验中,有关 ADR 证据是颇有价值的。

此外,有关 ADR 观测与追踪的资料,应注重它的完整性。通常如追踪率高于80%或高于90%者,则证据可靠,反之则 ADR 的真实性必受限制。

### 4. 对于 ADR 的证据能否满足病因/危险因素的判断标准?

- (1) ADR 发生是否能肯定因在前、果在后的因果规律;
- (2) ADR 发生是否存在量效规律;
- (3) ADR 发生是否存在消长规律:即用药则发生 ADR,撤药则消退或再用再发;
- (4) 本类 ADR 是否在其他研究报道中有着相类似或相同的结果;
- (5) 本种 ADR 是否有生物学依据。

以上涉及 ADR 的证据,如严格评价之后:确属真实、可信者,则应进一步分析它的重要性。

#### (二) 重要性

对于 ADR 证据的重要性分析与评价,目的在于确认其对病人的危害程度和它的精确程度,这必须用量化的指标表示,掌握了这类指标并确定其意义,方能提供进一步决策的依据。

#### 1. ADR 之因果效应的强度如何?

通常是依据 ADR 的总体概率及严重度分层,计算它们的相对危险度(relative risk, RR),绝对危险增高度(ARI),以及治疗多少病例方呈现一例不良反应(number needed to harm, NNH)。对于回顾性的病例对照研究的 ADR,则计算 OR;对于 RCT 及队列研究的 ADR 证据,则计算 RR;一般以  $RR > 3.0$  为有意义,而病例对照研究的  $OR > 4.0$  方可认为有意义,之所以将它们的价值列为较高范围,是由于存在若干混杂因素及偏倚因素的可能性之故。因此有人提出可视具体情况,选择可能的混杂因素来校正,如校正后的数值变化不大,则可信。

#### 2. 因果效应指标的精确性如何?

上述 ADR 因果效应指标应计算95%可信区间。凡可信区间窄,则精确度高;如果 ADR 的95%可信区间过宽,则精确度就低。

#### (三) 适用性

如经上述两步评价,其真实性好,重要性强,那么这种 ADR 证据是否在自己的医疗中能指导循证医学实践呢?这里就要进一步考虑其适用价值了。特别是以下几点:

#### 1. 被评价的 ADR 证据是否与自己的病人不一致,因而不宜采用?符合者用,不符合者弃之。

2. 权衡治疗的有益效果与 ADR 的负面效果的利弊比、指导自己的决策。计算治疗的利弊比值,用 LHH 表示(likelihood of help vs harm, LHH)。根据其值大小并结合具体的临床病人情况决策。 $LHH = (1/NNT) \div (1/NNH)$ 。

#### 3. 从治疗与 ADR 分析的决策中,宜与病人交流,尊重病人的期望、优选和要求,构建合作的医患关系。

4. 是否还有另外的备选方案,当分析与评价有关治疗的 ADR 证据后,如不十分满意,则可寻找其他备选治疗方案或方法。

## 四、临床应用

临床医生每天都面临着各种临床事件,它们或许是 ADR,或许并不是特有的治疗不良反应,然而要即刻确定正确的因果关系几乎是不大可能,但对病人必须及时处理,这里应用循证医学理论指导临床实践,正确、高效的决策十分重要。当我们疑及 ADR 时,以下临床实践决策流程图(图 12-2),可供参考。

药物不良反应是和临床日常工作息息相关的重要问题,正确掌握诊断标准,同时带着问题检索所需文献,科学判断其真实性、重要性及适用性,指导我们的医疗决策,为广大患者服务,是我们每一

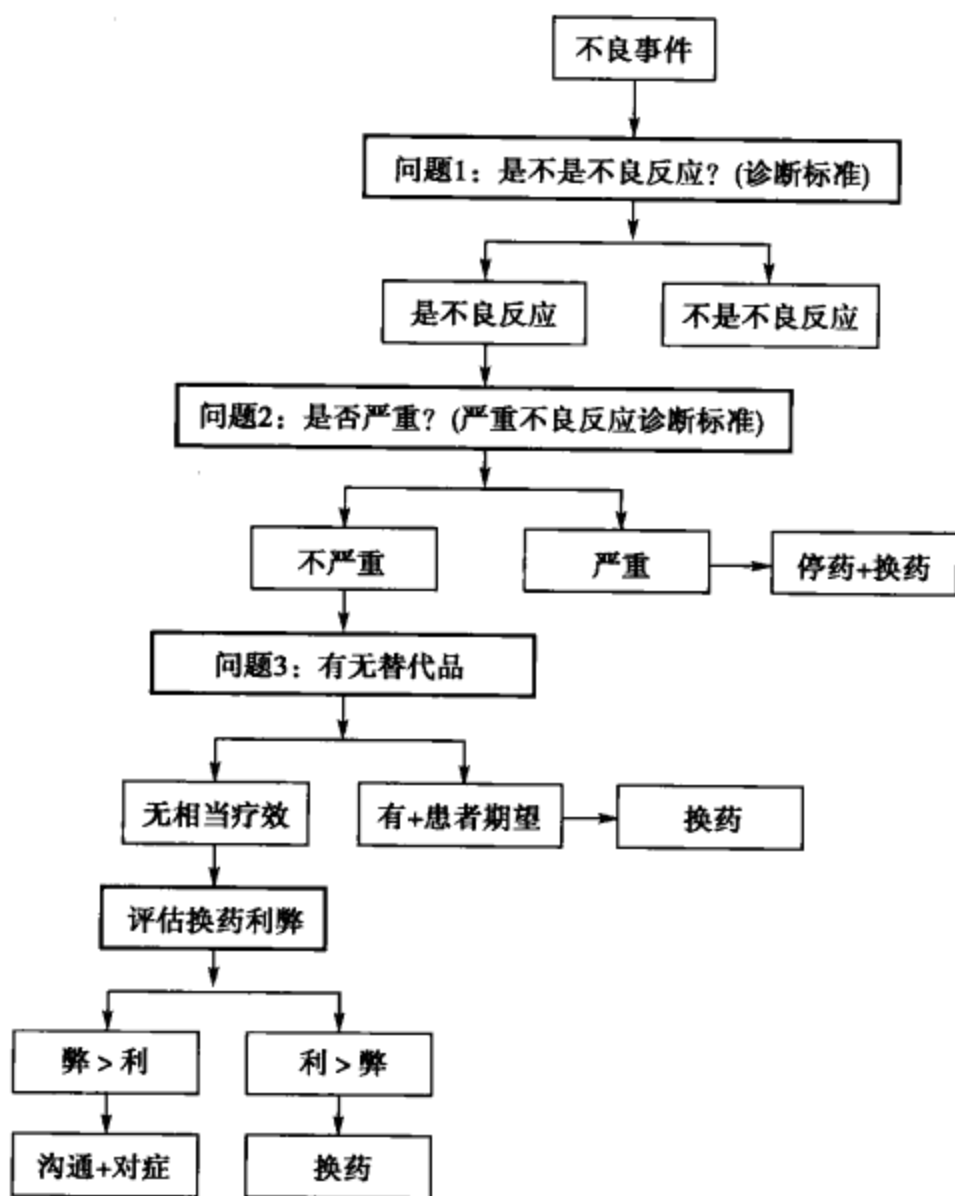


图12-2 临床治疗不良反应事件的处理流程图

个医疗工作者义不容辞的责任和义务。

(黄亚玲)

## 第十三章 疾病预后证据的循证分析与评价

在临床医疗的实践中,医生诊治病人时候,病人往往向医生提问:“我患的什么病?这种病对我的健康损害严重吗?如果严重,我是否有生命危险?危险程度有多大?有否有效的治疗措施与方法?如果没有的话,我可能活多久?等”无论该病人患的疾病为重症疾病(如冠心病急性心肌梗死)或难治性疾病(如恶性肿瘤),或者是通常可治易治的疾病,这一系列的问题,都涉及病人预后,要正确回答这些问题,首先务必掌握病人的病史、临床体征和病情以及相关病损的病理、生理依据,找出关键的临床问题,应用新近的研究成果(证据),对病人的预后进行科学的判断,并充分地应用有利的医疗环境与医疗条件,指导自己对病人的诊治决策,付诸实践并力图改善病人的预后。这些就是有关评价预后的临床循证实践。

### 第一节 预后证据分析与评价的意义

疾病的预后,指的是病人患病后,从发病到最终结局(痊愈、病残或死亡)的全过程中,根据病人的具体病情等实际情况,事前对结局事件的预测。通常用概率值表示,如痊愈率、病残率、病死率、生存率等。

对于有关疾病预后事件概率值的获得,是建立在临床科学长期观测和研究的基础上的。因此,预后证据的正确与否,必然取决于预后研究的质量。这就要涉及疾病预后研究的科研设计方案(参见临床流行病学的疾病预后研究方案)。由于影响预后的因素(有利与不利的)多伴随病人的本身,故往往只能进行观察性与分析性的研究,如队列研究(cohort study)及病例对照研究(case control study)等,通常对于疾病的自然预后是不能做人体的试验性研究(如RCT)的。因而,预后的研究证据,受有关偏倚影响的几率就要多一些。分析与评价预后证据的时候尤应注意。当然如通过干预性治疗来改善疾病预后则属例外。

预后的证据应从两方面加以分析与评价。一为有利于改善病人的预后因素,于临床循证实践中,则应充分地应用;另一为不利于病人的预后因素,则应采取相关措施予以避免或预防。根据科学证据采取“扬长避短”的措施,显然有着改善疾病预后的重大意义。

综上所述,对预后证据的分析与评价的意义可归纳为:

1. 掌握有关疾病预后结局事件的准确证据,有利于对有关疾病进行预后判断。
2. 预后证据质量取决于科学的研究设计与方法。因此,在分析和评价的时候,要用好临床流行病学的知识与方法。
3. 分析与评价预后证据的目的,在于指导医生面临病人的实际问题,应用有关影响预后的“证据”,进行改善病人预后的循证医疗决策。

### 第二节 疾病预后的时态特点

疾病的发生与发展及临床的最后结局有着各自的时态特点。通常在起病的早期病情为轻。若不及时治疗,则往往出现不同的进展,直至产生相应的结局。因此,在分析与评价预后证据的时候,要注

意疾病本身的时态特点。这里要掌握疾病的自然病史(natural history of disease)以及疾病的临床病程(clinical course)。

### 一、疾病的自然病史

疾病自然史(natural history)是指在不给任何治疗或干预措施的情况下,疾病从发生、发展到结局的整个过程。疾病的自然史包括四个时期:①生物学发病期(biologic onset)指病原体或致病因素作用于人体引起有关脏器的生物学反应,造成复杂的病理生理学改变;②亚临床期(subclinical stage)是指病变的脏器损害有所加重,但患者没有明显症状,采用灵敏度高的诊断手段,可以发现疾病已经存在;③临床期(clinical stage)指患者病变脏器的损伤更加严重,临床上出现了症状、体征和实验室检查的异常;④结局(outcome)指疾病经历了上述过程,发展到终末的结局,如痊愈、伤残或死亡等。研究疾病的自然病史对病因和预后研究、早期诊断和预防、判断治疗效果都有重要的意义。

### 二、疾病的临床病程

临床病程(clinical course)是指疾病的临床期,即首次出现症状和体征,一直到最后结局所经历的全过程,其中临床医师可采取医疗干预措施来改变其病程。

病程的概念和疾病自然史不同,病程可以因受到医疗干预(包括各种治疗措施)而发生病情改变,从而使预后发生改变。在病程早期就采取积极医疗干预措施,往往可以改善预后,在病程晚期进行医疗干预,效果就不那么明显,疾病预后就比较差。

关于自然病史与临床病程两者关系(图13-1),了解与掌握疾病在自然病史与临床病程中的不同特点,明确它们的病理损害与病情的差异,这在估价具体病人的预后情况有着非常重要的意义。

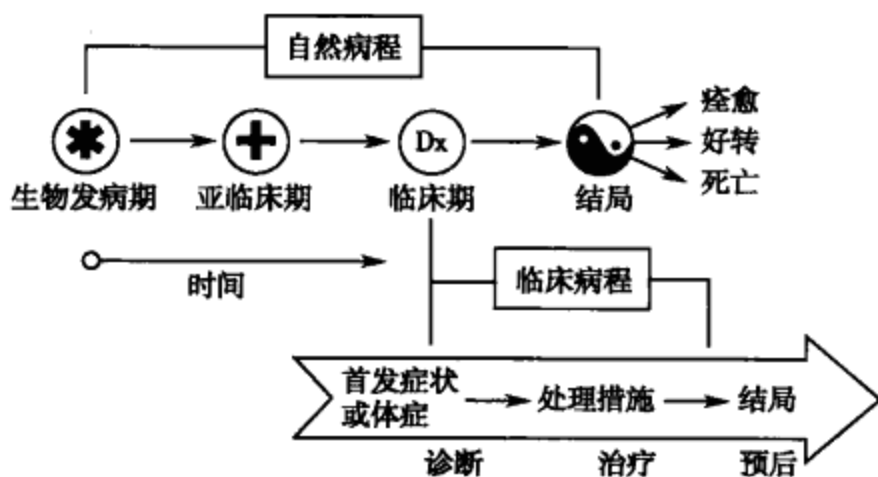


图13-1 自然病史与临床病程

## 第三节 掌握影响预后证据的有关因素

在临床医疗实践中,常常见到患着同一疾病且病情相似的病人,其预后的结局往往差异很大;此外,在分析与评价同一疾病所获得预后证据中,不同文献预后的结论也常常差别较大。这些差别的存在,有的可能是真实的,有的也许受着若干偏倚的影响导致结论缺乏真实性。因此,在分析与评价预后证据时,应掌握影响预后的因素。

### 一、疾病本身与环境的因素

1. 人口学因素 年龄、性别等。
2. 体质与心理因素 体壮与体弱、营养状态、免疫功能、个人心理状态的差异等。

3. 疾病特点 病情轻与重、病程早与晚等。

4. 有关疾病发病与预后有关的危险因素 如高血压系心脑血管发病的危险因素,同时又是影响预后的危险因素等。

5. 医疗环境 早诊与晚诊、治疗恰当与否等。

6. 社会经济因素 社会与家庭的关心照料、医疗保险、经济条件等。

上述六方面的因素均可影响病人的预后。分析证据时注意保护因素的正面效应,不利因素的负面效应等。

## 二、影响预后证据的质量因素

### (一) 集中性偏倚 (assembly bias)

所谓集中性偏倚乃是对预后证据研究报道的医疗单位,由于级别的高低不同,导致就诊的病人病情轻重各异。通常病情轻者多就医于一般水平的医院,病情重且复杂的病人往往到高级别的如“三甲”医院就诊。于是如对同一疾病的病人进行预后总体评估,一般医院病人的预后指标(如治愈率)恐怕比“三甲”医院好得多!为什么?这就是纳入观测的病人的“集中性偏倚”造成的。这种差异并不意味着前者的“高水平”或后者的“低水平”,也并不意味着前者收治病人的预后较后者好!例如:A医院为基层医院、B医院为“三甲”医院。在地震中两院各收治脑外伤伤员100名。经手术和常规治疗,A医院病死率10%,B医院病死率为35%,差值达25%!可是A医院100名病例中90%为轻度脑外伤患者,重度伤者为10%;而B医院收治重型脑外伤患者达40%,轻中度伤者占60%;A医院病死者全为重型病例(10/10),B医院的重伤患者(35/40)存活。如对脑外伤患者的预后估计,则不能得出A医院脑外伤患者的预后优于B医院的简单结论。

### (二) 存活队列偏倚 (survival cohort bias)

纳入医院观测预后的证据,往往仅是到医院救治病例的结果,无论病情的轻重,全部病例在就诊时尚存活,即使病逝于医院急诊室,仍有据可查。然而,当同一疾病未到医院即已死亡者,则无据可查,所以,在医院研究的预后证据中往往就没有院外死亡病例的信息,从而可能导致对预后证据的过好估价,这就是由存活队列偏倚所引起。

例如:对于入院就诊的急性心肌梗死患者,通常的病死率为15%左右,可是用这个概率估计整个急性心肌梗死群体的病死率可能为低,因为有些病例未到医院就已去世。这种低估的预后乃是存活队列偏倚影响所致。

### (三) 回忆性偏倚 (recall bias)

鉴于回顾性的队列研究或病例对照研究所获的有关预后证据,除了依靠病历资料外,往往还会涉及病人或亲属的回忆。这种回忆也许涉及遥远的过去(例如过去是否接触某种危害因素),因此,受着回忆是否准确或完整的若干影响,难免有丢失信息的现象。这会直接影响证据的真实性。

### (四) 失访偏倚 (lost to follow up bias)

对预后证据的分析与评价,要高度重视病例的失访偏倚,追踪率越高(即失访率越低),则预后评估的证据越可信。严格的标准至少追踪率达90%以上。但至少不应低于80%,即失访率应低于20%,否则证据的质量会受到严重的影响。

### (五) 测量偏倚 (measurement bias)

预后证据资料的来源,无论是前瞻性的观测或回顾性从病历中收集的信息,除了明确的死亡或存活的证据外,对于另外若干与预后相关的定性或定量的指标,如生化指标、影像资料,往往存在着研究者主观的测试或舍弃之可能,这类情况乃属于测量性偏倚,可影响证据的真实性。

### (六) 零时不当偏倚

判断预后的患者,彼此不在同一病程或自然病史的同一时点,故造成的差异。

上述六种偏倚,常见于对预后证据质量的影响,在分析与评价的时候应予以注意。

### 三、控制偏倚因素影响预后证据的方法

#### (一) 随机化 (randomization)

从理论上讲,两个队列进行比较,应当除研究的预后因素外,其他因素最好两组均相同,即基线状况要相同,这样才能比较该预后因素在两组中是否有差异。随机化方法进行分组是消除混杂最好的方法,真正的随机化是指每个研究对象都有同等的机会进入观察队列和对照队列,随机化分组使两组可比。

#### (二) 限制 (restriction)

在选择研究对象时,限制具有一定特征的对象进行观察,以排除其他因素的干扰。例如研究年龄是否为急性心肌梗死的预后因素,如将研究对象限定在黄种人、男性、无并发症的前壁心肌梗死病人,进行观察,这样就可以排除种族、性别、心肌梗死部位和并发症等因素的干扰和影响,就能比较清楚地反映年龄对急性心肌梗死预后的影响。但用这种方法来控制偏倚所获得的预后结论常有很大局限性。

#### (三) 配比 (matching)

配比就是为观察组的每一个研究对象匹配一个或几个具有同样因素的对照,然后比较两组的预后因素,配比的方法能消除这些因素对结果的潜在影响。许多研究者常以年龄、性别和种族作为配对条件,因为这些因素常是最常见的混杂因素,许多其他因素也可作为配对条件,如病期疾病严重程度和先前的治疗等。但千万不能把研究因素作为配对条件,否则就不能观察该研究因素在两组中的差异。

#### (四) 分层 (stratification)

分层是最常用的检出和控制偏倚的方法之一,特别是有潜在的混杂偏倚时,应用分层方法控制偏倚主要是在临床科研资料的分析阶段。分层是指将科研资料按某些影响因素分成数层(亚组)进行分析,观察研究因素是否在每层内各组间有差异,以明确该研究因素是否系独立的预后因素。

#### (五) 标准化 (standardization)

当我们比较两个率时,如果两组对象内部构成存在差别足以影响结论时,可用率的标准化加以校正,即使可能影响结果的因素受到同等的加权,则这两个率可比。这种方法称标准化(或校正)。

#### (六) 多因素分析方法

在临床预后因素研究中常比较复杂,可有多个预后因素相互作用,从而影响结局,应用单因素分析还不能将各预后因素对结局的影响分析清楚,此时应借助于多因素分析方法,多因素分析可以同时处理许多因素,可以从中筛选出与疾病结局有关的主要预后因素,及这些因素在决定预后中的相对比重。在预后因素研究中以比例风险模型-Cox 回归分析方法最为常用。

## 第四节 评价疾病预后的指标与方法

### 一、常用的指标

#### (一) 病死率

病死率(case-fatality rate)是指在患某病患者总人数中,死于该病的患者所占的比例,常用于病程短且易引起死亡的疾病。

$$\text{病死率}(\%) = \frac{\text{死于该病的患者人数}}{\text{患某病的患者总人数}} \times 100\%$$

#### (二) 治愈率

治愈率(cure rate)是指患某病治愈的患者人数占该病接受治疗患者总数的比例,常用于病程短而不易引起死亡的疾病。

$$\text{治愈率}(\%) = \frac{\text{患某病治愈的患者人数}}{\text{患该病接受治疗的总患者人数}} \times 100\%$$

### (三) 缓解率

缓解率(remission rate)是指给予某种治疗后,进入疾病临床消失期的病例数占总治疗例数的百分比,包括完全缓解率、部分缓解率和自发缓解之分,常用于长病程、低死亡的疾病。

$$\text{缓解率}(\%) = \frac{\text{治疗后进入疾病临床消失期的病倒数}}{\text{接受治疗的总病例数}} \times 100\%$$

### (四) 复发率

复发率(recurrence rate)是指疾病经过一定的缓解或痊愈后又重复发作的患者数占观察患者总数的百分比,常用于长病程低死亡的疾病。

$$\text{复发率}(\%) = \frac{\text{复发的患者数}}{\text{接受观察的患者总数}} \times 100\%$$

### (五) 致残率

致残率(disability rate)是指发生肢体或器官功能丧失者占观察患者总数的百分比,常用于长病程低死亡的疾病。

## 二、生存率分析

在生存分析(survival analysis)中可以用直接法和间接法计算生存率。生存率(survival rate)是指从疾病临床过程的一段时间内存活的病例数占总观察例数的百分比。

### (一) 直接法

本法简单、容易计算,但病例为小样本时抽样误差大,可出现后一年生存率高于前一年生存率的现象(图13-2)。

$${}_n P_0 = \frac{N - \sum_0^n (dx + wx)}{N - \sum_0^n (wx)}$$

式中  $N$  为进入研究总人数;  $dx$  为各年死于本病的人数;  $wx$  为各年失访人数。

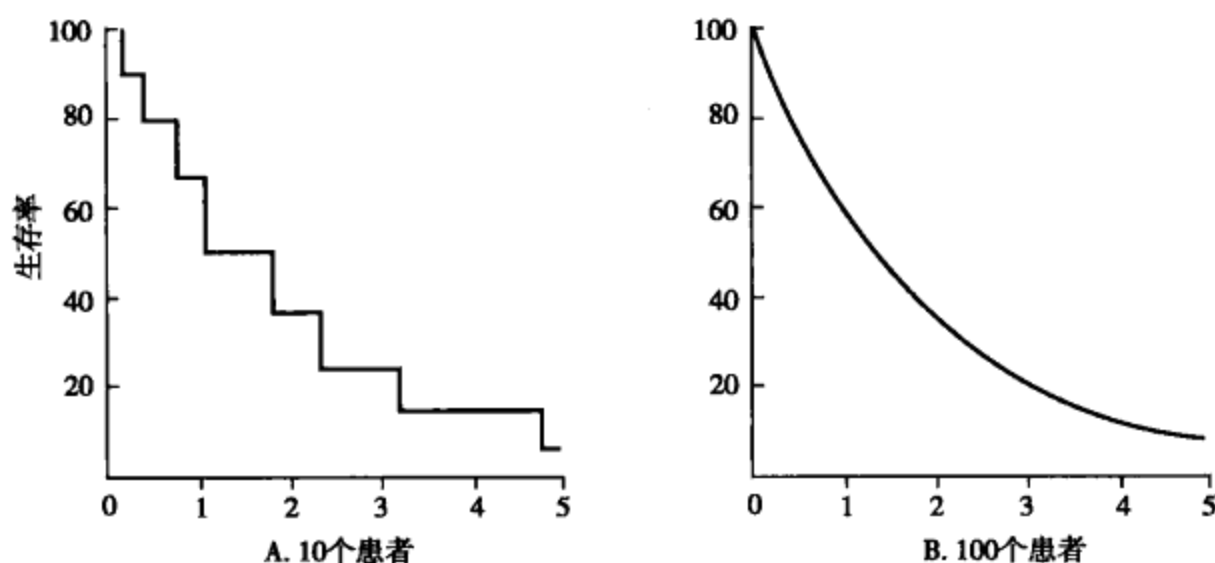


图13-2 小样本和大样本的生存曲线

### (二) 间接法

进行生存率分析,首先要对病例进行前瞻性队列研究,随访观察要以某一规定的起始点开始,每一个研究对象随访观察的起始点务必相同,终止目标和终止日期也务必有统一的规定,随访结果



可能有四种情况:①死(或复发)于本病;②死于其他疾病;③观察到规定截止时间尚存活(未复发);④失访。后三种情况的观察值都未达到预定的观察终点,都不能提供完全的信息,这种数据称终检值(censoring)。对含有终检值的资料,当例数较多时可用寿命表法分析,发病例数较少时,可用Kaplan-Meier生存率曲线分析法。寿命表法又叫条件概率法,基本步骤是先分别计算出患者进入观察后各年的生存概率,然后将各年的生存概率相乘,即得出患者进入观察后活过各年的生存率(累积生存率)。Kaplan-Meier曲线是以时间 $t$ 为横轴,生存率 $P$ 为纵轴,表示时间与生存关系的函数曲线。从曲线图可对某一病例的预期生存时间大于 $t$ 的概率进行估计。随访观察的时间单位愈小,则精确度愈高,即生存期用日计比月计为佳,计算活过各时点的生存率也是应用连乘积方法。

### (三) 生存率比较

对比不同病型、不同病情、不同治疗方法对疾病预后的影响,可以对生存率进行比较,绘制成的生存曲线图可以直观比较。时序检验(logrank test)是一种常用且较为理想的比较生存曲线的方法,其主要原理是运用 $\chi^2$ 检验,分析实际观察值与理论值之间的差值意义的大小,从而对各组之间的差异作出有无显著性的结论。

## 第五节 疾病预后证据的评价

在我们的临床实践中,针对病人的病情,为正确估价病人的预后和改善预后,对带着问题所收集的相关文献与证据,必须进行严格评价(critical appraisal),以去伪存真,再结合病人和医疗环境的具体情况,应用有用的最佳证据来指导改善预后医疗决策。

要对有关证据进行严格评价,必须采用公认的质量评价标准。现在国际上均采用由McMaster大学国际临床流行病学资源和培训中心首先制订,并由Sackett DL等主编的*Evidence Based Medicine*第1~3版(2000~2005年)逐步完善的预后证据分析与评价的原则为准。其中,首要的是评价证据的真实性(validity),共有四条原则;其次是评价证据的重要性(importance),有两条标准;最后就是评价证据的适用性(applicability),共有两条原则。这里称为原则,是因为每一条原则中,都会有具体的内容,都有着质量的差异,因此,这“三性”应予综合评价。其中真实性是基础,如证据不够真实,则无所谓重要和适用;反之,真实性好的证据,也不一定都是重要的和适用的。

### 一、预后证据的真实性评价

#### 1. 被纳入患者的代表样本,是否确定在临床病程的一个共同起点?

临床预后研究的证据,往往是从所研究的目标疾病患病群体中抽样或连续性收集的部分患者,作为代表样本来进行系统观测追踪的结果。作为纳入观察的病人,必然要符合预后研究设计的纳入和排除标准,而且要有明确的观察起始点的要求,即全部观察预后的对象应处于临床病程的同一阶段(或起点),这样其预后结局的结果或结论,才可能较真实地反映出疾病代表样本处于该病程的预后共性特点。这种情况如同100m赛跑,要求所有的运动员应在同一起跑线上,不能让部分运动员位于110m处或有的处于80m的距离,否则,其竞赛的成绩就会呈现太大的差异。因此,当评价和分析有关代表样本的预后证据时,首要的是明确其处于临床病程的什么阶段,是亚临床期、临床期或是临床结局期?显然处于不同的临床阶段的患者预后是各异的。例如恶性肿瘤的I、II期或III期患者,其预后结局的差异性肯定非常明显。互相之间是不可比的。故不能不看具体的代表样本之病情而笼统地说某病预后如何。

假设在预后证据中,能反映整个疾病群体的自然预后,则是十分理想的。例如:像冠心病患者的预后,如能包括从动脉粥样硬化之初开始直至终末的心肌梗死等结局为止全过程的预后证据,肯定是非常理想。可事实上是不大可能的,因为这个病的自然病史及临床病程太长和太复杂,人们是难以办到的。因此,也只可能从患病群体以一定方式分期分阶段抽样观察并根据分期的临床特点,进行预后

评价。如果预后研究的样本量很大,证据丰富,那么可根据实况进行临床分期的分层分析,以探讨不同病程期或病理损害的不同预后,这是十分有意义的。对此,也有利于读者根据自己病人的情况,对照地考虑其价值。

对于疾病预后研究,宜始于临床的早期阶段,即入组观测的病人都处于临床早期,此为起始队列(inception cohort),然后根据研究设计要求进行不同时期连续追踪观测或针对性地予以干预,以探讨和分析预后,如生存率分析,这种系统预后证据是很有价值的。

### 2. 研究对象的追踪观测的时间是否长、追踪是否完全?

对纳入预后观测的研究对象,应根据所探讨预后的疾病之临床病程始终的时间特点,并依预后相关指标设计相应的观测追踪时间,原则是尽可能地观测到绝大多数对象的终末指标的时间为宜,不能过短,否则可导致假阴性的误导。因此,追踪的时间宜长不宜短,这是要依具体的疾病而定,在分析与评价预后证据真实性方面,要充分应用自己的专业知识。

对预后观测患者的追踪证据,应高度重视它的追踪率,追踪率越高,则其证据的真实度会更好,因其失踪率低。然而,因种种原因(如意外死亡、迁移、交通或通讯等),必有一定的失访者,为保证证据的真实性,追踪率应 $\geq 90\%$ ,至少不低于 $80\%$ ,这里有个“5和20”的粗略规则供参考:如失访率 $\leq 5\%$ ,则对预后证据很少受偏倚的影响,结论很可信;如失踪率 $\geq 20\%$ ,则预后证据的真实性会受到严重影响。

此外,处理失踪者证据的较为严格的方法是进行敏感性分析,计算事件发生率的最低和最高值,为证据的评估者提供真实性评价之参考。例如对100例患者进行5年追踪观测,结果有4人死亡,16例失访,假设仅计算追踪完全84例(100-16)的病死率为 $4.8\%$ (4/84);如以100例计算病死率最低值为 $4\%$ ,最高病死率为 $20\%$ , $(4+16) \times 100\%/100$ 。还可进一步估计最低和最高病死率的95%可信区间,最终结果是否可信,仍应结合专业、专病的特点而定。

### 3. 结果测试的标准是否采用了盲法?

对预后结果的测试应有明确的判断标准,以防有关偏倚的影响。如最终结局是死亡或痊愈,则属无争议的硬指标,或许死因判断不一。但介于两者之间的若干患者预后如何,是好转、好转的程度,缓解、缓解的程度,病残、病残的程度等,其准确分析与评价往往困难一些。用什么方法和标准进行评价、如何测试与评价,或许受病人与预后观测者的某种主观偏倚的影响,因此,为确保预后证据评价的真实性,应在研究设计之初为不同预后结果制订客观的标准,而且宜采用盲法评价。

4. 如果亚组的预后不同,那么对于重要的预后因素以及对独立的“测试组”的患者是否进行校正?

当评价疾病预后证据真实性时,特别要注意存在哪些影响预后的因素。由于预后因素以及影响预后的偏倚因素在研究的证据中往往共存,这就有可能出现真假结果,因此,在原始资料分析中要加以鉴别。如果在原始资料的单因素分析中,被确定了的影响预后的因素(称之为“training set”或“derivation set”起源的预后因素),也许受着机遇因素或偏倚因素的影响,出现虚假的情况,因此,可以采用有关统计学方法进一步对预后因素进行分析。例如采用多因素 logistic 回归或 COX 风险比例模型分析,对预后因素判断有着重要价值。

当确定预后因素证据之后,为了探讨对不同类型病人的影响程度,往往根据亚组进行分层分析,以提供更确切的证据,为使用者提供方便。亚组分析是将所纳入的病人,依影响预后的最主要因素进行分层,归纳成几个独立的测试亚组(independent “test set” patients),然后分别与有关重要的预后因素进行校正性分析比较,从而得出合理的结论。例如:拟探讨慢性持续性心房颤动患者引起脑血管意外的预后,则可将这些病人按心脏瓣膜病有无、年龄等进行分层,分为4个独立亚组(图13-3),最后追踪各亚组脑卒中发生率以评价心房颤动的危害度。

这里应强调的是,亚组分析评价预后因素,仅适用于大样本的研究证据,且分层因素不宜过多,应选最重要的1~3个预后因素为宜,否则研究对象离散度就大。各亚组病例过少,机遇因素干扰也会

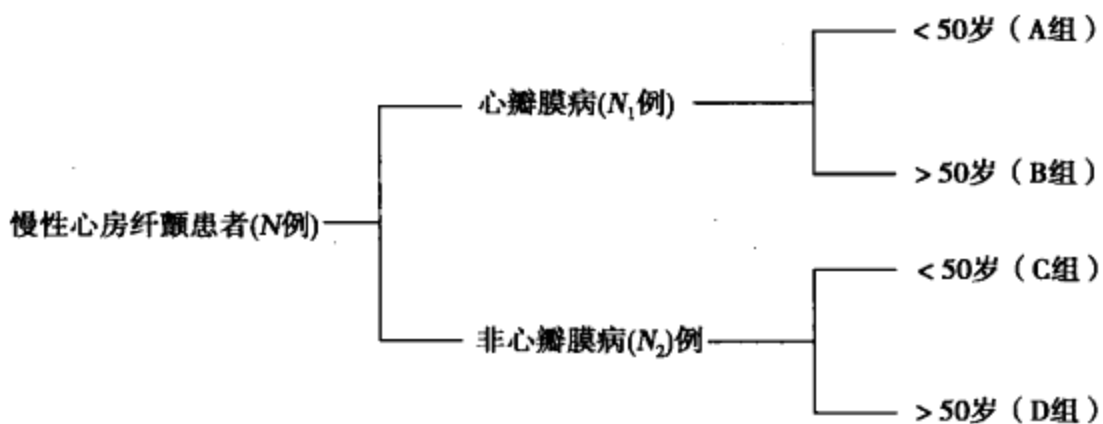


图13-3 亚组分层模式图

大,影响证据的正确判断。

此外,对有关预后因素的影响存在混杂干扰时,还可以进行统计学的校正。例如: Framingham 的一项有关风湿性心脏病(简称风心病)合并心房颤动与非风心病合并心房颤动患者发生脑卒中的预后研究中,初步结果显示,风心病合并房颤的脑栓塞率为41/1000人年,与非风心病合并房颤患者的脑栓塞率相似,但在进一步比较两组的资料时发现,有风心病的患者偏年轻,同时两组在性别、血压等方面均不平衡,经过调整年龄、性别与高血压后,显示风心病合并房颤患者并发脑栓塞的危险性是非风心病房颤患者的6倍,说明对预后的影响有着多种因素。

## 二、真实预后证据的重要性评价

经过上述真实性分析与评价,被确定为真实的预后证据之后,将对它的重要意义进行分析,以示其对疾病预后评估及促进病人改善预后决策的价值。

### (一) 预后的结果能否合理表达全时效应?

在获得真实的预后证据后,则要进一步审查效应所覆盖的时间范围有多大,是否系从始至终全时过程的观测预后结果。用这种方式表达疾病的预后,为根据不同时点所发生事件发生率计算并绘制 Kaplan Meier 生存率曲线(参见有关统计学专著),甚有价值。根据不同的曲线,可以评价有关疾病的预后状况。

兹模拟4种生存率曲线,说明不同疾病在不同时点的生存率以及中位生存率(图13-4)。

A 生存率曲线: 从始到终没有不良事件发生,意味着该病预后良好;

B 生存率曲线: 中位生存时间为3个月,12个月生存率20%;

C 生存率曲线: 中位生存时间为9个月,12个月生存率20%;

D 生存率曲线: 中位生存时间为7.5个月,12个月生存率20%。

可见 B、C、D 三种疾病在不同时点的预后: 第12个月的生存率(预后)均一样,但中位生存时间不同,表明 C 病预后要优于 B 和 D。这些结果可供预后因素重要性的相互比较。

此外,还可以应用有关事件发生率来评价预后因素效应的重要性。

### (二) 预后估价的精确度如何?

预后结局的指标,通常都用事件发生率表示,如痊愈率、病死率、致残率、生存率等。这些是率的点估计值,估计的精度是用率的95%可信区间(95% confidence interval, 95%CI)表示。95%CI 越窄,则精度越高,结果越可靠,反之,则估计精度差。

鉴于预后研究的证据,均来源于所抽样的代表样本而非患病人群总体结果,因此,预后结局的指标就难免受抽样误差(机遇因素)的影响,如果抽样的样本量足够大,那么机遇因素的影响相应较小(如小于5%),预后证据的可信度就高,反之,可信度就有很大折扣。

综上所述,在评价预后证据的精确度时,需要计算95%CI 以明确结果的精确范围,同时要注意预后证据来源样本量与机遇因素所影响的程度。

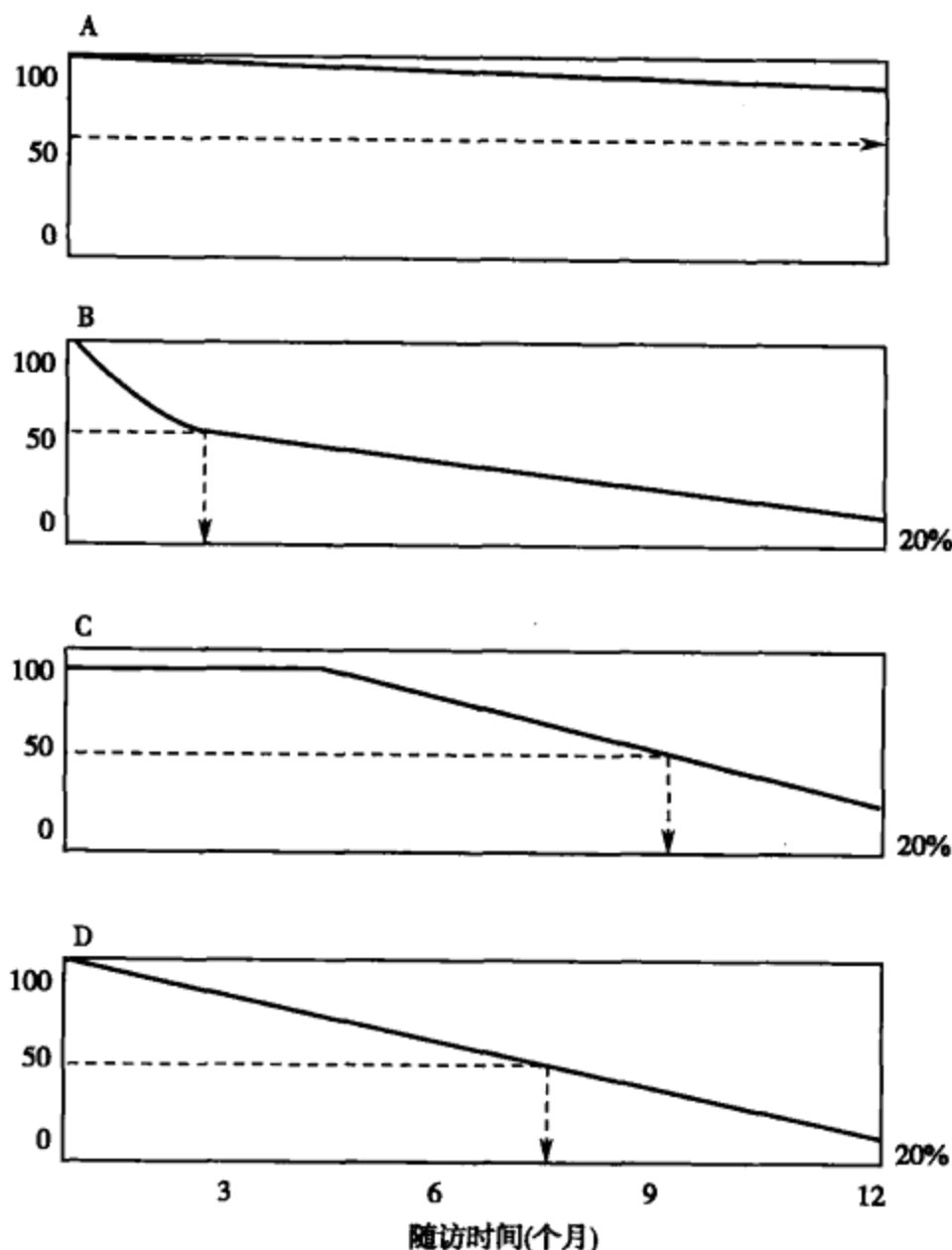


图13-4 生存曲线显示的预后

### 三、真实、重要的预后证据可否用于我们的病人

如果经过分析和评价预后证据是真实可靠且有重要价值,那么这种证据可否用于我们的临床实践、以对具体的病人作出正确的预后判断和改善病人预后的防治决策呢?这就要联系病人的实际进行个体化处理。

#### 1. 是否我们的病人情况与被评价证据中的病人情况不同,而不宜采用?

显然,为正确估价我们自己的病人预后和期望采取更有效的诊治措施以改善病人的预后,带着有关具体的临床预后问题,去检索和评价一些预后文献,是为了解决具体的临床问题,是有的放矢的工作。因此,对于真实、重要的预后证据肯定要与自己的病人“对号入座”,即人口学特征、病人病情、病程、医疗条件、医疗环境、经济状况等均为考虑范畴,能“对号”则用,不“对号”者则弃之。

#### 2. 根据预后证据拟作的临床重要决策应与病人交流

带着病人的预后问题所获的有关预测和改善病人预后的证据,无疑最后会结合病人的实际而进行相应的重要临床决策,其执行还需病人理解与合作,方能获得理想效果。因此,凡是有肯定意义的决策应与病人交流,说明改善预后决策的重要价值。其中有的可能涉及病人本身的行为或生活方式及不良的生活饮食习惯改善等。例如:急性心肌梗死患者经抢救存活,假设该患者合并高血压及高脂血症且患者又吸烟嗜酒。这些涉及预后的危险因素如不能很好控制,势必再梗死或意外事件发生率比没有这些危险因素者高。因此,为改善患者预后,就要针对影响预后的危险因素进行控制高血压及降血脂治疗的决策,并嘱病人戒烟、忌酒。这些改善预后措施的执行,很重要的是要病人理解与合作,方可达到理想的目的。

此外,即使我们所获得证据,对病人而言也许意义尚难肯定。但也可以提出相应的改善预后的建议或措施,力争取得良好效果。例如,晚期恶性肿瘤患者无论接受放化疗与否,据研究,中医或中西医联合治疗,可改善生存质量和延长生存期,如证据可靠,则可建议病人接受中医药治疗,以达到改善预后之目的。

以上评价预后证据的三条八单项原则,是在循证医学疾病预后临床实践中,如何正确识别、掌握与应用真实而重要的证据的基本原则与方法。其中,还涉及循证医学理论与方法学的基础,即临床流行病学以及医学统计的知识与方法,因此,为了更好地分析评价证据,结合实际学习有关这些相关知识是必要的。

文献评价实例(参见光盘辅助教材相关章节)。

(邱牟生 王家良)

## 第十四章 卫生经济学证据的评价与应用

### 第一节 卫生经济学及其评价的基本概念

随着人口的增长和老化、新技术和新药物的应用、人类健康需求层次的提高,医疗保健的预算远远跟不上医疗费用上涨的速度,全世界都面临着医疗资源相对不足这一问题。医院、公共卫生保健部门和卫生代理机构均面临着如何选择最佳方式,以最低的医疗资源和费用消耗获得最有效的全民保健。卫生经济学从经济学的角度研究卫生保健领域的资源如何实现最佳配置的问题,为循证医学实践提供有关经济效益方面的证据,帮助我们理性决策,真正实现基于现有最好证据,兼顾经济效益和价值取向,对可利用的资源作出最有效的安排。

#### 一、卫生经济学

卫生经济学(health economics)是经济学领域的一个分支科学,是一门研究卫生保健中的经济规律及其应用的学科,其运用经济学的基本原理和方法研究有限的资源如何向卫生行业分配,以及卫生行业内的资源如何优化配置。卫生经济学的研究目的是怎样最佳地、有效地、公平地使用稀缺的卫生资源,使之最大限度满足人们对卫生服务的需求,提高卫生服务的社会效益和经济效益。卫生经济学研究的主要内容包括:卫生服务市场,卫生服务提供体系,卫生资源的筹资、分配与使用,医疗保障制度,卫生机构的经济管理,药品经济学和卫生经济学评价。

#### 二、卫生经济学评价

卫生经济学评价(health economics evaluation)是卫生经济学研究的主要内容之一。卫生经济评价是从资源的投放(成本)和效果两个方面,对不同选择方案进行比较分析的方法。它的基本内容是确认、衡量、比较待评候选方案的成本和获益。

通常当一定量的卫生资源用于某一卫生规划时,这些资源就不可能用于其他卫生规划,究竟这些卫生资源用于哪种卫生规划可以发挥最大的效益,这就需要利用卫生经济学评价的方法对不同备选的卫生规划方案进行比较,从中选出最优方案,为决策者提供依据,以解决卫生服务的重点和优先。卫生经济学评价在确定卫生资源的投入方向、预防保健措施的选择、临床诊治方案的选择、卫生技术评估、药品经济学等方面应用广泛。

### 第二节 卫生经济学评价的基本要素

#### 一、成本

成本(cost)是指社会在实施某项卫生服务规划或方案时所投入的全部财力资源、物力资源和人力资源,在临床医疗中涉及病人的医疗成本(费用)等,通常用货币统一计量和表示。成本的计算和分析是进行经济学评价的基础。成本的种类主要包括直接成本、间接成本和无形成本。

##### (一) 直接成本

直接成本(direct costs)系直接提供卫生医疗服务过程中所花费的成本。

1. 直接医疗成本(direct medical costs) 指卫生服务过程中用于治疗、预防、保健的成本,常包括住院费、药费、诊疗费、实验室检查费、X线检查费、手术费、家庭病房费、康复费及假肢等费用。

2. 直接非医疗成本(direct nonmedical costs) 指病人因病就诊或住院所花费的非医疗服务的成本,如病人的伙食、交通、住宿、家庭看护、病人住院后家属探望的往返路费、外地病人家属的住宿费等。

### (二) 间接成本

间接成本(indirect costs),又称生产力成本(productivity costs),是指因疾病丧失生活能力甚至死亡所导致的社会资源的损失。

1. 与病残率有关的成本 由于疾病引起工作能力减退及长期失去劳动力所造成的损失,如因病损失的工资、奖金及丧失的劳动生产力造成的误工产值。

2. 与死亡率有关的成本 由于病死所造成的损失,例如,规定60岁退休,患者因病于50岁死亡,损失的10年工资、奖金都应作为间接成本计算。

### (三) 无形成本

无形成本(intangible costs)指由于疾病给病人和家属带来的精神损失和对生活质量的影响。这部分成本是难以估计的成本。

## 二、效果

效果指的是在卫生医疗的服务中,投入了多少成本,其最终结果如何呢?这就要用相关指标来表示。

### (一) 效果

广义的效果(effectiveness)指卫生服务产出的一切结果,包括用各种指标表示的结果。在成本效果分析中,效果是指因为疾病防治所带来的各种卫生方面直接结果指标的变化,如发病率或死亡率的降低、治愈率的提高以及人群期望寿命的延长。

效果可以同时或分别使用中间测量指标(intermediate measures)和健康测量指标(health measures)。前者包括症状、危险因素或有关临床测定的结果,例如溃疡的愈合率、乙型肝炎病毒E抗原的阴转率、血清胆固醇的下降程度等。后者包括病残天数、寿命年的延长、死亡数等。例如,在高血压的治疗项目中,血压下降的百分率为中间测量指标,而高血压患者远期心脑血管事件(如脑卒中和心肌梗死)的发生率则是最终健康指标。

### (二) 效益

效益(benefit)是有用效果的货币表现,即用货币表示卫生服务的有用效果。效益一般可分为直接效益、间接效益和无形效益。

1. 直接效益(direct benefit) 是指实行某项卫生计划方案之后所节省的卫生资源。如某种疾病发病率的减低,减少了诊断和治疗费用的支出,减少了人力和物力资源的消耗,这种比原来节省的支出或消耗就是该卫生计划方案的直接效益。

2. 间接效益(indirect benefit) 指某项卫生计划方案实施后所减少的间接经济损失。如由于发病率的降低,或住院人数的减少,避免患者及陪同家属的工资、奖金的损失等。

3. 无形效益(intangible benefit) 是指实行某项卫生计划方案后减轻或避免了患者肉体和精神上的痛苦,以及康复后带来的舒适和愉快等。

### (三) 效用

效用(utility)是指人们对不同健康水平和生活质量的满意程度。在成本效用分析中,表示效用的常用指标有质量调整寿命年和失能调整生命年。最常用的是质量调整寿命年。

1. 失能调整生命年(disability adjusted life year, DALY) 是指从发病到死亡所损失的全部健康寿命年,包括因早死所致的寿命损失(years of life lost, YLL)和疾病所致失能引起的健康寿命损失年

(years lived with disability, YLD)两部分。该指标是一个定量计算因各种疾病造成的早死与伤残对健康寿命年损失的综合指标,即对疾病死亡和疾病伤残而损失的健康寿命年的综合测量,是用于测量疾病负担的主要指标之一。

2. 质量调整寿命年(quality-adjusted life years, QALYs) 是用生存质量来调整期望寿命或生存年数而得到的一个指标,即将不同生存质量的生存年数换算成相当于完全健康人的生存年数,是反映人群生存质量和生存数量的一个综合指标。

计算质量调整生寿年,首先需要估计一个评价各种功能健康状况的效用值。效用值表示个体对不同健康状态的喜好程度,它是个体在不确定的情况下的优先选择,表现出他们对某种健康状况的倾向和偏爱,反映了个体的主观感受,并受年龄、经济收入、教育程度等多种因素的影响。效用值通常用0~1之间的数值来表示,1代表完全健康,0代表死亡。也可以为负数,表示比死亡更糟糕的疾病状态,如无意识或长期卧床伴严重疼痛等。表14-1是不同健康状况的效用值。效用值可以用作计算质量调整寿命年的调整权数,若个体病人在某一时间段内生存质量水平是恒定的,用该时间段内健康状况的效用值乘以生存年数就可以得到该时间段内的质量调整寿命年,将不同时间段内的质量调整寿命年数相加就可得到病人总的质量调整寿命年。

$$\text{质量调整寿命年} = \text{某种健康状况下的效用值} \times \text{生存年数} \quad (\text{式14-1})$$

表14-1 不同健康状况的效用值

健康状况	效用值
完全健康	1.00
疲劳和失眠	0.28
时常呕吐	0.55
视力受损和活动受限制	0.50
需要轮椅	0.37
死亡	0.00

### 第三节 卫生经济学评价的类型

#### 一、卫生经济学部分评价

卫生经济学部分评价(partial economic evaluation studies)包括经济研究的多种形式,这些研究也涉及成本和(或)收获的评价,但是要么没有对不同的干预方案进行比较,要么没有将成本和效果联系起来进行评价。通常有以下几种形式的卫生经济学部分评价方法:成本比较/成本分析(cost comparison/cost analysis)、成本结果描述(cost outcome description)、成本描述(cost description)、结果描述(outcome description)、疾病成本研究(cost of illness study)。

#### 二、卫生经济学全面评价

卫生经济学全面评价(full economic evaluation studies)具有两个主要特征:①要同时对两种或者两种以上方案进行比较;②评价时要考虑被评价项目的投入(成本)又要考虑项目的效果,仅比较效果,或者仅比较成本都不是完整的经济评价。不具备上述两个特征的评价,都属于经济学部分评价。

经济学全面评价是提供可靠的经济效果证据的唯一经济分析形式。经济学全面评价包括:最小成本分析、成本效果分析、成本效益分析和成本效用分析四种类型。这几种方法的差异主要体现在对结果的测量和评价方法不同。

##### (一) 最小成本分析

最小成本分析(cost-minimization analysis, CMA)也可称为成本确定分析(cost-identification



analysis)。该分析类型是指比较不同的医疗卫生措施时,备选方案的效果相同或者差异不重要,这时只需要考虑成本投入的不同,选择成本低的方案。

由于这种方法只能比较结果相同的不同措施之间的成本,故使用范围较小,其优点是直接、简单。例如,骨髓炎病人提早出院在门诊继续用抗生素治疗和常规住院相比,前者花费2271美元,比常规住院2781美元的费用低,最小成本分析结果显示早期出院方案每例病人可节约510美元。

### (二) 成本效果分析

成本效果分析(cost-effectiveness analysis, CEA)兼顾成本和效果,是通过分析成本消耗后得到的效果来确定最有效地使用资源的一种方法,也是目前在医疗保健领域中最常用的一种经济评价方法。

成本效果分析中效果的测量通常用成本效果比和增量比来表示,这就使两种不同的医疗措施,在进行比较选择时,有了相同的评价单位。

1. 成本效果比(cost/effectiveness, C/E) 是指每一效果单位所消耗的成本。例如,每个生命年、每治愈一例病人或每诊断出1例新病例所花费的成本。例如,一个高血压治疗项目使60岁男性高血压患者的舒张压从110mmHg降低到90mmHg,延长一个生命年,需要花费的成本为16 330美元,另一项用两种不同降血脂药物治疗高脂血症项目,结果显示延长一个生命年花费的成本分别59 000美元和17 800美元。可见,同样延长一个生命年,高血压治疗项目的成本效果较好。

2. 增量成本效果比(incremental cost-effectiveness ratio, ICER) 即每获得一个增加的效果单位所消耗的增量成本,它代表一个方案的成本效果与另一个方案比较而得到的结果。在研究中常以效果最低的方案作为参照,将其他方案与之相比而得,这一比值越低,则表明该方案增加一个效果单位所追加费用越低,该方案的实施意义越大。例如,对3种方案治疗上呼吸道感染进行成本效果分析,将3种方案的效果由低到高排列,并以最低效果为参照。对3组进行的增量成本效果分析如表14-2。每增加1个效果单位所花成本C组比A组多23 646.2元,B组比A组多2696.6元。

表14-2 不同方案的增量成本效果比

治疗方案	疗程成本(C)/元	总有效率%(E)	C/E	ICER (ΔC/ΔE)
A	179	81.5	219.6	-
B	580.8	96.4	602.5	2696.6
C	486.4	82.8	587.4	23646.2

### (三) 成本效益分析

成本效益分析(cost-benefit analysis, CBA)是将医疗服务的成本和效果都用货币单位来表示,用相同的单位来分析所花的成本与所得到的效益的关系。

成本效益分析常用指标为效益成本比(benefit cost ratio, BCR)和净效益(net benefit, NB),其计算公式分别为:

$$BCR = \text{某项措施的总效益} \div \text{某项措施的总成本} \quad (\text{式14-2})$$

$$NB = \text{总效益} - \text{总成本} \quad (\text{式14-3})$$

成本效益分析方法的优势在于可以进行不同医疗措施之间的比较,结果清晰、直观,常用作卫生决策者合理分配有限卫生资源的经济学依据。但这种方法的局限性在于,将健康效益转化为货币值比较困难。目前常用的转换方法主要有人力资本法和支付意愿法。

人力资本法(human capital method)是将个体当作有价值的商品资本来对待,因为他们对未来产出有潜在的贡献,也就是将人的生命看成和机器、厂房一样是一种能够带来收入的投资,如果说工厂投资10 000元,每年可以带来1000元的收入,那么,一位年收入1000元的劳动者,他的生命价值就是10 000元。

支付意愿法(willingness to pay, WTP)认为健康是可以买卖的商品,在市场交易中,人们根据自

己的意愿给健康标价,这个价格,就可以作为评价生命价值的依据。例如用飞机失事旅客生命的赔偿费等经验数据作为支付意愿的尺度。

#### (四) 成本效用分析

成本效用分析(cost-utility analysis, CUA)是通过比较项目投入成本和经质量调整的健康效益产出量来衡量卫生项目或治疗措施效果的一种经济学分析方法。

成本效用分析是成本效果分析的一种特殊形式,但是成本效果分析只能用于比较同一疾病或相同条件所采用的不同措施或方案,分析所用的效果指标应相同,如均用延长生命年或治愈率为指标。例如,比较两种假定的医疗措施,一种是治疗尿失禁的,另一种是治疗阳痿的。表14-3给出了关于这两种医疗措施的费用和效果。

表14-3 治疗失禁与阳痿的医疗措施的比较

状况	医疗费用(美元)	效果年限
尿失禁	2万	8
阳痿	1.8万	12

表14-3中的数据表明,治疗尿失禁的费用比治疗阳痿的费用要高,而且效果维持的时间要短。然而,尚无法证明尿失禁治疗比阳痿治疗的成本效果是好还是差。如果使用质量调整寿命年,进行成本效用分析就可以进行比较。表14-4中列出这两种疾病的效用值和治疗获得的质量调整寿命年,治疗尿失禁的成本效果为每质量调整寿命年需成本1万美元。治疗阳痿每获得1个质量调整寿命年需要成本1.5万美元。因此,我们得出结论:用质量调整寿命年计算,尿失禁的治疗费用比阳痿的治疗费用要低,因而其成本效果更好。

表14-4 用质量调整寿命年比较不同的医疗措施

状况	治疗费用(美元)	效果年限	效用值	QALYs	成本/QALY
尿失禁	2.0万	8	0.25	2	1.0
阳痿	1.8万	12	0.10	1.2	1.5

成本效用分析的优势是将生存质量评价引入效果评价,可以将不同的健康指标转换为可以进行比较的同一结果,因此可以比较不同医疗卫生项目的经济结果。表14-5列出不同医疗措施的成本效用分析结果。

表14-5 不同医疗措施的成本效用

治疗措施	每获得一个质量调整寿命年的成本
髋关节修复术	\$2000
肾移植	\$7500
对60岁患者心脏病突发的卡托普利治疗	\$11000
医院的血液透析	\$35000
用红细胞生成素治疗接受透析的贫血病人	\$86000
用开放式胆囊切除术加碎石术治疗胆结石	\$140000
用神经外科学治疗恶性颅内肿瘤	\$320000

成本效用分析的关键是如何确定效用值,常用的确定健康状况效用值的方法有以下3种:

1. 评价法 由相关专家根据经验进行评价,估计健康效用值或其可能的范围,然后进行敏感性分析以探究评价的可靠性,这是最简单方便的方法。

2. 文献法 直接利用现有文献中使用的效用值指标,但要注意文献研究的立场、针对的人群和研究现场等是否与自己的研究相匹配。

3. 抽样调查法 研究人员自行设计调查方案,通过对代表性样本人群进行调查而获取,先让调查对象熟悉所研究的健康状况,然后采用不同方法使调查对象对效用值进行估计。这是最精确的方法。通常采用等级衡量法、标准博弈法和时间权衡法。

(1) 等级衡量法(rating scale): 是目前最简单的一种方法。在一个标注不同健康状态的尺度上,100表述完全健康,0表示死亡,将疾病状态清楚地描述给病人后,要求病人在尺度上某一点划一条横线以表明自己目前的健康状态,划线处即为自己所得的效用值(图14-1)。

等级衡量方法描述起来非常简单,然而,这种方法对效用值的估计不太准确。为了更加精确地计算质量调整寿命年,研究者还依靠其他两种方法进行计算。

(2) 标准博弈法(standard gamble): 又称标准概率法,这是一种风险选择法(最坏和最好的结果),即在可选择范围内做出的判断。

例如,某一疾病可以手术治疗,但要冒或成功或失败的风险,手术(A)的最坏结果是死亡,最好的结果是术后可以无病生活25年(风险选择),其概率均为50%;另一方面也可以进行保守治疗(B),而不冒手术风险,但处在带病状态,效果比手术的最佳效果差。因此,可以在手术和非手术保守治疗间作一选择。

当问病人保守疗法可生存5年时,选择A还是B,病人回答选A。生存6年时?仍选A。生存7年时?病人改为选择B,也就是说病人宁愿不手术以带病状态生存7年,也不愿冒有50%可能死亡、50%可能治愈生活25年的风险,此时该病的效用值为 $7/25=0.28$ (图14-2)。

(3) 时间权衡法(time trade-off): 要求患者回答:与当前的不健康状态相比,你愿意放弃多少生存时间以换取一定健康状态的生存年数。评价时给定两种可能的结果:一种为健康地生存X年后即死亡;另一种为在某种状态(残疾状态M)下生存Y年后死亡( $Y > X$ ),不断改变X值,直到评价人认为选择两种选择方案的倾向性相等为止。此时的X/Y值即为M状态的效用值。例如,告诉心绞痛病人,如果不治疗可带病再活25(Y)年,但假设有某一种治疗可使心绞痛完全缓解,可是寿命可能要缩短些,问他无病生存时间为多少(X)年时,他才宁愿选择这一治疗,这就需要患者决策。如果患者愿意能健康地活15(X)年才选择这一治疗,否则就拒绝,于是没有心绞痛的15年就相当于有心绞痛的25年效用。心绞痛的效用值为 $X/Y=15/25=0.6$ 。



图14-1 等级衡量尺度

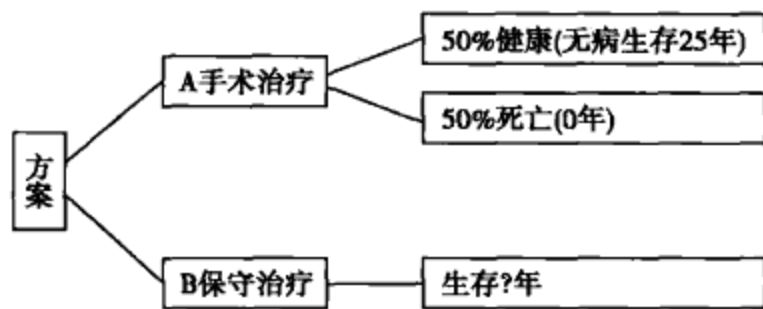


图14-2 标准博弈法

时间权衡法由于给出了在完全健康状态下生活的时间,其结果直接与效用值指标质量调整寿命年相关联,因此与各种健康状况的直接分级一致性较好。

(4) 量表测量法:除了以上3种直接测量效用值的方法,还可以用量表测量法间接测量,通过填写生命质量量表,计算量表得分,再将量表得分转换为效用值。测量生命质量的量表有很多,如:疾病影响量表(sickness impact profile, SIP), McMaster 健康指数调查表(McMaster health index

questionnaire), Nottingham 健康量表( Nottingham health profile), 医疗结果研究简化36量表( medical outcome study short-form, SF-36)、健康质量量表( quality of well-being, QWB)等。还有一些特殊量表则将着眼点放在疾病的特征方面,许多疾病都有各自特殊的量表,比如癌症(癌症病人生活功能指数 FLIC)、心血管疾病(纽约心脏协会分类量表)、糖尿病(糖尿病控制和并发症试验问卷 DCCT)和关节炎(关节炎影响量表 AIMS)等。由于不同的生命质量量表产生于不同文化背景或信仰的人群,测量的目的、内容和侧重点会有所不同,因此使用不同的量表获得的分值、得分范围可以有很大的差异,对不同人群研究结果进行生存质量比较时,尚存在难以解释或重复性差的弊端。

## 第四节 卫生经济学评价方案的设计步骤

### 一、明确所要评价的项目及要进行比较的措施

首先明确进行卫生经济学评价的目的。其次,要对比较的措施特点进行详细界定和说明。

### 二、确定经济学评价的分析角度

不同立场和观点的人对同一方案的结果可能得出不同的结论,因此经济学评价应当明确表明评价的角度(立足点),这决定了在评价中的成本和结果的定义、范围与内涵。对方案的分析决策和评价最好从整体出发,明确所要达到的目的和所要解决的问题对各方面因方案的实施所受到的影响,从整个社会的角度进行分析和评价。但实际上,由于难以获取整体评价的相关资料和时间上的限制,因此经济评价的角度通常会相对局限,从病人、医疗服务的提供者、医院的角度或某个组织的角度进行评价。

### 三、根据研究目的确定合适的经济评价方法

不同的评价目的和不同的观察角度,决定需要用何种方法进行评价。应明确选择某种评价方法的理由。

### 四、确定资料收集的方法

成本和效果信息可以用多种流行病学研究方法确定和测量。包括前瞻性研究,如随机对照试验或队列研究;回顾性研究,即根据已有的临床试验数据,回顾性补充收集方案成本和效果的数据;二次文献研究,利用已有的临床试验研究证据或系统评价证据进行研究,或利用已有的卫生经济学评价证据进行研究。运用这些研究方法时应遵循相应的流行病研究方法设计和实施原则进行。

### 五、成本和效果的确定和测量方法

所有相关的成本都应确定、收集并报告。例如,可以基于国家和当地的年度财政信息对单位成本进行估计,也通过研究机构的数据库,保险机构获取,尽可能从多种途径获取单位成本的详细信息。

效果的测量应根据所选用的评价方法确定,例如,用成本效用分析时就应报告质量调整寿命年和效用值的具体测定方法,并阐明这些方法的具体测定步骤。

### 六、处理时间对成本和效益的影响

一个方案的成本和产出在不同时间具有的经济意义是不一样的,因此在进行不同方案成本和产出比较时,为便于比较,应该消除时间的影响,将不同时间发生的投入和产出折算为同一时间的投入和产出。

贴现( discounting )就是把将来的成本和产出的价值换算成现在的价值,换算比率称贴现率( discount rate ),也叫折扣率,其是根据银行利率和物价指数综合确定。

计算公式:

$$P = \sum_{n=1}^t Fn(1+r)^{-n} \quad (\text{式14-4})$$

$P$ : 成本或效果的现在值;  $F_n$ : 成本或效果在  $n$  年时的值;  $r$ : 年贴现率;  $t$ : 项目完成的预期年限; 贴现率一般取3%~5%。

### 七、增值分析

在进行经济学评价时,不仅要比较成本/效果(效用、效益)比,还应该测定及报告增值分析(incremental analysis)的结果,即由于额外措施造成成本的增加时,其相应增加的效果(效用、效益)是多少。具体表示为一个项目比另一个项目多付的费用,与该项目比另一个项目多得到的效果(效用、效益)之比,称为增值比。

### 八、敏感性分析

在已得出上述经济评价的结果后,测定当其中几个主要的变量如价格、成本、贴现率、结果的判断标准发生变化时,或不同的经济分析类型对评价结果的影响程度,称为敏感性分析(sensitivity analysis)。如果稍微改变一下变量的数值,其经济评价的结论就会发生改变,则表明其可靠性较差。

由于对将来发生的某些情况如工资、失业率、期望寿命、治疗费、年贴现率等不能肯定,因此必须将敏感性分析作为经济评价中一项必要步骤。研究中所有不肯定的结果都应报告,对关键参数应报告其可信限及范围。

## 第五节 卫生经济学评价研究证据的评价原则

将卫生经济学评价的证据运用于卫生医疗决策之前,需要对证据的可靠性进行评价,卫生经济学证据的评价也包括三方面内容,即对研究结果的真实性、重要性和适用性进行评价。

### 一、结果是否真实

这一问题强调经济分析是否真实地反映了其中某一项措施可能提供更好的成本效果。与其他类型的研究一样,卫生经济评价的真实性取决于使用的方法是否恰当。

1. 明确该评价是否是卫生经济学全面评价。只有全面的经济评价才能提供科学、可靠的经济学证据。

2. 该评价目的是否明确,是否交代了从什么角度出发来进行经济评价的。经济分析的目的是否明确,角度是否得到广泛认同,对评价结果真实性起很大作用,也决定了你能否将这一经济评价结果应用于你的医疗实践。

3. 是否对相关的备选方案有较全面的描述,是否遗漏任何重要的备选方案。

4. 针对提出的问题,选择的经济学分析类型是否恰当。

5. 该经济学分析引用的各种备选方案效果的资料是否真实。效果的确定是通过什么研究方法获取?所用的研究方法是否符合流行病学设计和实施原则,可能影响结果的偏倚有哪些?通常可靠性最强的证据是从临床随机对照研究得出的结果。

6. 对发生在将来的成本和结果是否作了时间上的校正。贴现率是多少,如何确定,贴现后经济评价的结果是什么。

7. 成本和效果资料是否进行增量分析。

8. 是否进行了敏感性分析。

## 二、研究结果的重要性

结果的重要性首先需要考虑每种措施的成本和效果是多少,在不同亚组人群中成本和效果是多少,获得的成本和效果是否有临床意义,是否有利于资源的合理配置。其次,合理改变成本和效果的估计是否会改变经济学分析的结果,即不确定因素对结果的影响有多大,可以通过敏感性分析的结果对此进行评价。

## 三、研究结果的适用性

结果的适用性涉及其推广使用价值。有两点可以帮助判断结果的适用性:一是成本效果分析的增量比;二是你面临的医疗保健实践在多大程度上与经济分析中情况类似,是否有相似的人群,是否有相似的成本和结果。

文献评价实例(参见光盘辅助教材相关章节)。

(李亚斐)

## 第十五章 卫生技术评估与应用

科学技术成果在卫生保健领域和医疗服务系统的广泛应用,促进了医疗卫生服务水平的提高,为促进人类健康和提高生存质量作出了巨大贡献。但技术的应用在产生积极作用的同时,也会产生消极后果。例如临床医疗方案和技术程序缺乏标准化,放任了对昂贵技术的不合理使用;大量明显落后的技术不能及时淘汰,导致服务质量下降和卫生资源浪费等。因此,在享受技术进步带来益处的同时,限制技术应用已经或将要产生的负面影响和危害已成为突出的公共卫生问题。而卫生技术评估正是解决此问题的重要手段。

### 第一节 卫生技术与卫生技术评估概述

#### 一、卫生技术的定义

卫生技术(health technology)是指应用于卫生保健领域和医疗服务系统的特定知识与技术体系,包括用于预防保健、疾病控制、医疗康复的药物器械、设备设施、技术程序、医疗方案及相关的后勤支持系统和组织管理系统等。

#### 二、卫生技术评估的定义

卫生技术评估(health technology assessment, HTA)是指对卫生技术的技术特性、安全性、有效性(效能、效果和生存质量)、经济学特性(成本-效果、成本-效益、成本-效用)和社会适应性(社会、法律、伦理、政治)进行系统全面的评价,为各层次决策者提供合理选择卫生技术的科学信息和决策依据,对卫生技术开发、应用、推广与淘汰实行政策干预,从而合理配置卫生资源,提高有限卫生资源的利用质量和效率。

#### 三、卫生技术评估的历史沿革

1972年美国国会成立了全球第一个技术评估办公室(Office of Technology Assessment, OTA)。1974年7月交出第一份名为Drug Bioequivalence的评估报告:评价药品的化学成分与治疗等效关系、当前技术质量。1980年以后,丹麦、西班牙、荷兰、瑞典、加拿大等相继成立了卫生技术评估机构。如:丹麦卫生技术评估研究所(DIHTA)、瑞典国家卫生技术评估研究所(SBU)、加拿大卫生技术评估协调办公室(CCOHTA)及卫生技术评估哥伦比亚办公室(BCOHTA)、英国的卫生技术评估协调中心(National Coordinating Centre for HTA, NCCHTA)、澳大利亚的医疗服务咨询委员会(Medical Services Advisory Committee, MSAC),均为国家的医疗技术开发、应用、推广、淘汰及政府卫生政策提供科学、可靠的依据。

为加强国际间的协作,在全球范围内推广卫生技术评估,一些卫生技术评估的国际组织先后成立,如1985年成立的国际卫生技术评估协会(International Society of Technology Assessment in Health care, ISTAHC)、1993年成立的国际卫生技术评估机构网络(International Network of Agencies for HTA, INAHTA)、加泰隆卫生技术评估和研究机构(Catalan Agency for Health Technology Assessment

and Research, CAHTA)。旨在确立共同关心的课题;建立收录各成员机构评估报告的数据库、国际卫生技术评估杂志;发展和保持与其他机构的合作关系,帮助建立新的卫生技术评估网络,扩大在发展中国家影响。

2005年3月WHO借鉴循证筛选基本药物目录(essential medicines list, EML)的成功经验,启动循证制定基本卫生技术(essential health technology, EHT)的计划,将于2007年循证筛选100项最佳的卫生技术,向192个成员国推荐。

世界上卫生技术评估的发展大体分为3个阶段:1975~1985年是卫生技术评估的初级阶段,主要利用循证方法,处理卫生保健干预的效果和成本效益问题,帮助政府决策;1985年后进入第二阶段,主要探索建立与决策者间的一种良好沟通关系;在20世纪90年代以后的第三阶段则更多地致力于影响医疗机构的管理者及临床医生。评估的重点也从早期局限于大型、高技术设备扩大到微型技术、软技术、心理咨询服务等,目前已涉及更广泛的领域,如体制、社会、伦理等。卫生技术评估的重心已从纯技术学转向健康需求,在各区域内影响决策。在美国,卫生技术评估已成为进行资源配置的一个重要的信息来源。

我国于20世纪80年代引入医学技术评估概念。1994年在原上海医科大学由卫生部批准成立第一家医学技术评估中心,至1999年又相继成立3家相关机构:中国循证医学中心(原华西医科大学)、生物工程技术评估中心(浙江大学)和医学伦理研究中心(原北京医科大学)。卫生部2000年成立卫生技术管理处,领导4个中心,先后开展了多项有关医疗设备、临床技术与预防技术等方面的技术评估,如“关于伽马刀使用医学技术评估研究”、“人类辅助生殖技术”、“肝炎生物芯片检验技术”、“器官移植立法与脑死亡立法的基础研究”等。现已制定《造血干细胞库管理办法》、《人类辅助生殖技术管理办法》、《人类精子库管理办法》和《卫生技术准入管理办法》及“体外受精-胚胎移植技术标准与规程”、“人类精子库技术标准与规程”等配套措施。由科教司组织起草的《人体器官移植管理条例》、《脑死亡判定管理办法》、《脑死亡判定标准》、《脑死亡判定技术规范》、《肾脏、肝脏、心脏、角膜移植技术准入标准及眼库准入标准》等文件和技术规范已经完成。此外,还通过卫生技术评估,淘汰了35项临床检验技术,同时确立了相应的替代技术,使我国的临床检验水平迈上了一个新台阶。

#### 四、卫生技术评估的意义

卫生技术评估可为卫生行政管理部门制定公共卫生计划、配置卫生资源、研发创新与调控推广卫生技术等方面的政策提供决策依据;帮助相关组织开展卫生技术研发和市场规划;帮助卫生技术的提供者和付费者决定是否将某项卫生技术列入卫生福利计划,确定合理的报销项目和比例;帮助卫生技术的提供者和消费者合理选择卫生技术服务;总之,卫生技术评估必将有利于推进我国的新一轮医改,更有效地利用有限卫生资源解决不断增长的疾病负担和“看病贵、看病难”的问题。

##### (一) 临床医学领域

为临床医务工作者提供科学的信息和决策依据,对卫生技术的开发、应用、推广与淘汰实行政策干预,从而合理配置卫生资源。在发达国家,HTA已经被认为是解决医疗费用上涨、合理进行医疗服务及制定卫生政策的有效工具。

1990年,根据卫生技术评估的研究成果使用高渗造影剂来替代低渗造影剂,使医疗费用明显降低,净节约为1200万美元。再如:对心导管的再利用节约了医疗费用约600万美元,而取消术前常规胸片节约了700万美元。有关高新技术如器官移植、MRI等的卫生技术评估报告,对制定卫生政策和临床指南,以及合理配置资源均产生了显著的影响。

##### (二) 预防医学领域

为各层次的决策者提供合理选择卫生技术的科学信息和决策依据,合理配置卫生资源,提高有限卫生资源的利用质量和效率。可以帮助卫生保健网络和机构的管理人员获得和管理卫



生技术;帮助卫生部门的官员制定公共卫生计划;帮助卫生保健产品生产厂商进行产品开发和市场规划。

加拿大一项关于乳腺癌普查的卫生技术评估结果显示,对50~70岁的妇女进行普查其成本效果最佳,这使政府改变了过去对所有育龄妇女进行普查的政策,节约了相当的卫生保健经费,优化了卫生保健系统。美国国家卫生保健技术中心对老人保健项目覆盖政策的咨询工作,使项目开支每年可节省数亿美元。

## 五、卫生技术评估的内容

卫生技术评估的内容涉及卫生技术的技术特性、安全性、有效性、经济特性、社会和伦理适应性。

### (一) 卫生技术的技术特性

卫生技术的技术特性(technical properties)是指卫生技术的操作特性,以及是否符合该技术在设计、组成、加工、耐受性、可靠性、易使用性和维护等方面的规范。

### (二) 卫生技术的安全性

卫生技术的安全性(safety)是指卫生技术在特定的条件下,如有某种健康问题的患者、具有一定训练的医生及在特定治疗场所应用时可能出现的危险程度(不良反应的发生率和严重程度)及患者的可接受程度。例如,华法林和阿司匹林均可用于治疗心房颤动患者,虽然华法林的疗效优于后者,但导致出血的可能性却明显高于后者,且需要定期监测,因此,有的患者宁愿使用阿司匹林。

### (三) 卫生技术的有效性

卫生技术的有效性是指卫生技术在应用时改善患者健康状况的能力,包括效力(efficacy)和效果(effectiveness)。一般来说,卫生技术在严格控制的条件下或在精心挑选的患者中应用时获得的结果比在常规条件下好。效力是指在理想情况下将卫生技术应用于某一特定的健康问题,如精心设计和管理的随机对照试验,选择受试对象的标准非常严格并在条件极好的研究中心开展研究。效果是指在一般或日常条件下将卫生技术应用于某一特定的健康问题,如在社区医院由全科医生将某一卫生技术应用于各种类型的患者。例如在评估长期氧疗的作用时发现,在医院内的作用和患者出院回家后在家使用的作用差别较大。其原因是在医院内使用时有护士帮助患者操作氧发生器和监测患者是否定时应用,获得的是效力;而在家使用时由患者或其家属进行操作,是否使用有时是患者根据自身感受确定。如果今天感觉好,可能患者就不用,感觉不好时又用一下,且操作氧发生器的方法也不一定规范,获得的效果自然较医院为差。

### (四) 卫生技术的经济特性

卫生技术的经济特性(economic attributes or impacts)包括卫生技术的微观经济(microeconomic attributes or impacts)和宏观经济特性(macroeconomic attributes or impacts)。微观经济特性主要涉及某一卫生技术的成本、价格、付费情况和支付水平等,也涉及比较应用卫生技术时对资源的要求和产生的结果,如成本-效果、成本-效用和成本-效益分析。宏观经济特性包括新技术对国家健康费用的影响、对卫生资源在不同健康项目或健康领域中分配的影响以及对门诊和住院病人的影响。其次还包括对调控政策、卫生改革和技术革新的政策变化、技术竞争、技术转换和应用的影响。

### (五) 卫生技术的社会和伦理适应性

某些卫生技术如遗传试验、辅助生殖治疗技术、重要器官的移植和临终患者的生命支持系统等均涉及相关法规和社会规范,提出了一些社会和伦理问题(social and ethical concerns)。

## 六、循证医学在卫生技术评估中的作用

循证医学提高了卫生技术评估的科学性、证据强度及成本-效果。

### (一) 循证医学强调应用最佳证据

以此为指导要求卫生技术评估针对特定问题收集此前所有相关的最高级别证据,评估该项卫生

技术在临床大面积、长时间使用的安全性、有效性、适用性和成本-效果,并结合评价结论,提供决策参考,确保了评估结果的真实性和可靠性。

### (二) 循证医学强调了证据的与时俱进、不断更新

借鉴循证医学的思想和方法,卫生技术评估可建立不断更新的数据库、公共查询窗口,促进信息共享和社会监督,适时更新卫生技术评估报告。

### (三) 循证医学突出了以病人为中心

循证医学强调以结局性指标以及有重要价值的有关中间性指标判定效果,因而更能保证病人利益,取得最佳的成本-效果。

## 第二节 卫生技术评估的步骤与方法

尽管卫生技术评估所涉及技术有多种形式,在评估的范畴、方法和资料收集等方面有较大差异,但一般都遵循以下9个基本步骤:

### 一、确定评估项目

评估项目的确定主要取决于提出评估申请机构的目的、医疗实践的需要、决策者的需要。通常可以参考以下选择标准确定评估项目。

1. 造成的疾病负担重、影响的人群面广的疾病问题;
2. 昂贵的卫生技术或医疗总费用高的健康问题;
3. 临床应用中存在争议的技术;
4. 改善健康结局/降低危险度的重要干预措施;
5. 解决存在的伦理、法律、社会问题;
6. 是否有足够的资料用于评估;
7. 公众、政策的需要;
8. 是否用于制定调控、费用支付政策等。

多数组织在确定评估项目时有其自身的选择标准。卫生行政部门选择优先评估项目的标准则可能主要考虑该项技术的安全性、潜在的社会伦理和道德法律方面的影响、技术经济效果以及技术准入标准等。如卫生部注意到辅助生殖技术在现实应用中引起的重大社会影响和市场混乱,从而委托卫生技术评估研究中心进行辅助生殖技术评估;而医药企业选择优先评估项目时则往往考虑该技术(医疗设备或药品等)潜在的市场规模、能得到多大的市场份额、投资-回报率如何、安全性和功效性如何。

目前国际上卫生技术评估开展得比较好的国家如瑞典等建立了官方的技术评估优先项目选择标准,定期对根据该标准所筛选出的“重点”技术开展评估工作。

### 二、明确评估问题

明确具体要解决的问题是卫生技术评估的重要环节。评估者应清楚了解评估目的和评估结果的用户,因为不同的用户看问题的角度和所具备的专业知识差别较大,有可能影响评估的内容、报告的形式和结果的传播等。

对评估问题的明确表述至少包括以下内容。

1. 所涉及的卫生问题;
2. 所涉及的患者人群;
3. 所涉及的技术类型;
4. 技术的使用者;

5. 技术的应用场所;
6. 评估内容等。

以辅助生殖技术评估为例,一般要明确:①技术所涉及的卫生问题:人工授精针对有关精子活力问题等,试管婴儿针对受精困难等;②所涉及的患者人群:患者的数量和特征;③所涉及的干预技术:人工授精技术、试管婴儿技术、精子库等;④干预的使用者:哪类哪级医务工作者;⑤所涉及的医疗保健机构:哪些机构开展辅助生殖技术,存在何种问题;⑥评估内容:安全性、效力、效果、成本-效果、成本-效用、社会伦理问题等。

### 三、确定评估者

确定评估者取决于被评估问题的性质,已有的资金、技术、人才力量、时间和其他一些因素。评估者或评估机构一般应经过专业的培训与认证,具备相应的资质,再遵循伦理学准则,按照科学的评价标准,认真严谨、实事求是、客观公正地开展卫生技术评估工作,并承担相应的法律责任。

卫生保健决策者可以自己开展全部的卫生技术评估活动,也可以全部委托给其他评估机构,或者是把资料的收集和合成工作委托给专业评估机构,而其他工作则由自己来完成。通常小规模卫生保健服务提供者和支付者一般委托专业卫生技术评估机构开展卫生技术评估,而大规模的卫生保健服务提供者和支付者则更倾向于自己开展卫生技术评估。

### 四、收集资料

#### (一) 收集现有资料

收集充分的、可信的资料是确保卫生技术评估成功的关键。许多技术的相关资料是非常零星和散乱的,而且资料的质量也差异很大;特别是一些很新的技术,其相关资料非常少而且往往很难找到。因此,在进行文献检索时特别是在制订检索策略时应咨询信息专家,以保证合理选择数据库、主题词、自由词,获得所有相关的信息。此外,查寻资料时还要从多种数据库类型中检索以控制发表偏倚的影响。

除了传统的主流文献之外,还可以在其他一些非正规的文献中获得许多有价值的信息,例如医药行业或行政部门的专题论述、政策法规性文件、专业委员会的报告、市场研究报告、专业会议摘要、或者来自互联网上的资料和信息。互联网是信息含量非常巨大、可及性非常好的信息载体。如在进行辅助生殖技术评估时,对技术本身的特性和安全性等内容主要参考主流文献,而社会影响、技术滥用、国际热点等方面的资料则从互联网上进行了广泛收集。这种非主流文献往往时效性很好,而且提供了主流文献所没有提供的资料,但是有其固有的缺陷,通常没有经过同行专家的评阅,因此在利用时应该加以认真审查和筛选。

收集现有的相关资料是卫生技术评估的最主要任务之一,对资料收集的范围、时间、费用、质量等都应在事先进行详细计划。另外由于有些资料需要花钱才能查阅(如Cochrane数据库),而有的专业报告需要高价才能买到,因此,资料收集的费用可能会非常昂贵。

#### (二) 获取新的原始研究证据

很多卫生技术评估活动只需要利用现有的资料就可以进行,而有些评估项目需要获取新的原始研究证据。

不同类型研究设计方案的论证强度不一样。在进行原始研究时,应考虑研究类型及配套的最佳设计方案,保证结果的真实性和论证强度。一般是前瞻性研究优于回顾性研究、有对照的研究优于无对照的研究、随机化研究优于非随机化研究、大规模研究优于小样本研究、盲法研究优于非盲法研究、同期对照优于历史对照。

在研究设计上要充分考虑研究的内部真实性和外部真实性。此外,评估者在进行新的研究时必须考虑到成本和时间限制,开展经济学评估,权衡研究的边际投入和边际收益。

## 五、评价证据

卫生技术评估需要从不同类型、不同质量的研究中获得科学的证据,评估者必须对已有的资料进行全面、严格的评价。评价证据需要掌握研究的方法和统计学知识,因此评估小组中应有具备这些知识的人员。某些评估项目先由专业和证据评价方面的专家评价资料后,再由评估小组进行资料分析和合成。一般来说,证据评价包括以下4个方面:

### (一) 证据的分类

由于原始研究的设计类型不同,产生的证据强度不一样。评价证据的第一步就是按照基本的方法学类型和研究特征,采用证据表格将研究证据进行分类。证据表格一般包括研究设计特征(随机、对照、盲法),患者特点(病例数、年龄、性别),患者结局(死亡率、并发症发生率、HRQL)和统计量( $P$ 值、95%CI)。以表格方式让评价者能系统比较研究的主要特征,了解所有资料的数量和质量全貌。

### (二) 证据的分级

根据研究证据的方法学严格性,采用不同形式,按照规范的标准对每一研究进行结构式的严格评价和分级。对研究证据进行分级的方式较多,有主要针对原始研究证据的,也有同时综合针对原始研究和二次研究证据的。

进行证据分级时,不仅要考虑基础研究类型对证据质量的影响,更要考虑具体的研究设计和实施方法,对单个研究均有系统的方法评价其质量,甚至针对具体的设计方案类型,如随机对照试验,进行方法学质量评价。

许多有关卫生技术评估的研究证据均为小型、非随机的系列病例报道或单个病例报道,因此,存在方法学缺陷,难以据此确定某一技术的效力或其他特性,证据的论证强度也不高。由于被评估技术本身的特性,或者研究者难以控制某些因素,有时方法学上的缺陷是难以避免的。要确定某一新药的效力,标准的设计方法是大规模、前瞻性、双盲的同期随机对照试验。

### (三) 证据的抉择

由于方法学质量的差异,在评价单个研究证据时,每一研究证据的重要性有别。高质量研究证据对结果的影响应大于低质量研究证据。但就如何应用不同质量的研究证据,专家们在评价时意见并不统一。基本原则为:或者使用所有发表的研究证据,或者根据纳入、排除标准确定,或者根据研究质量给予不同的权重,或者校正研究结果,减少偏倚影响。

### (四) 偏倚的控制

在评价研究证据时要考虑可能存在的影响、研究实施或结果报告的利益冲突。对于研究者来说,这些利益冲突可源于与健康保健产品公司的经济利益关系,如给予工资、持有股票/专利或者为该公司的竞争者之一等。同行评审的杂志发表的证据,越来越要求公开与研究者的经济利益信息和研究资金来源,以供评价研究证据时考虑。

由于评估活动出发点的不同,不同利益集团(制药公司、学术机构、政府部门等)对同一项技术的评估可能会得出不同的结果。例如,制药公司在开发出一种新药后,往往会委托评估机构或者自己进行一些与同类药品之间的比较评估活动,而受利益驱使,它们可能希望得到对其有利的评估结果。因此,从公正客观的角度出发,各种评估报告都应该公布资助者、执行者、评估机构的隶属关系、评估方法、资料来源等信息,以供判断该评估报告的真实性和可靠性。

此外,卫生技术评估中还应采取措施确保评估人员和机构与被评估的项目和机构没有潜在利益冲突,避免评估者和评估机构被利益集团所控制,以免影响评估人员和评估机构的客观性与公正判断。

## 六、合成资料

在对研究证据的质量进行分析评价后,就要对全部资料进行综合分析。常见的综合分析方法包

括: ①定性的文献评阅: 由于缺乏严格、统一的方法学, 难以避免偏倚的影响; ②达成共识 (consensus development): 几乎所有的评估项目都或多或少地采用达成共识的方法来得出某些结果和建议, 如 Delphi 法等。由于这类方法得到的结果只能代表一组专家的意见, 因此, 在科学性上有时不能令人信服; ③ Meta- 分析: 是一种卫生技术评估中常用的对文献进行定量综合的分析方法; ④决策分析: 指通过对各替代技术间成本和效果的定量估计和比较, 决定何者是最优技术; ⑤经济学分析: 指对技术干预所引起的经济影响进行分析和评价, 主要的分析方法包括成本-效果分析、成本-效用分析、成本-效益分析等。

## 七、形成结果

卫生技术评估的主要目的就是要根据收集到的资料形成评估结果, 再根据评估结果提出政策建议。由于用来评估的研究证据的质量和说服力相差很大, 所得结果和建议的科学性和说服力也不同。因此, 结论和建议必须与研究证据的质量和强度相联系, 基于已有的证据和评估发现的结果和结论, 不能根据主观感觉进行推断。但有的评估报告所做的建议并未反映证据的相对论证强度, 读者可能错误认为所有评估报告的建议均是同样真实或具有同等权威性。

为保证评估结果能得到客观合理的应用, 评估者在评估报告中应该详细说明所用证据的来源, 并提供证据有效性和科学性的评价结果。在应用评估结论时将推荐意见级别与证据的级别相联系, 或将推荐意见级别与研究证据质量相联系。

## 八、传播结果

在得出有价值的评估结果和建议之后, 很重要的一步就是要传播结果。如果不通过各种渠道对评估结果进行广泛的、正确的、及时的宣传, 那么结果可能被曲解或误解、可能失去时效性, 也可能和那些不真实的信息相混杂, 从而就不能切实发挥卫生技术评估对决策的支持作用。

传播卫生技术评估结果和建议时, 应考虑目标人群、媒体和传播技术或策略。传播技术评估的结果有很多途径, 如及时把评估报告递交卫生行政部门并转化为相应政策法规、出版论文和专著、大会交流、在大众媒体上宣传评估结果、通过官方网站进行公告等。

## 九、监测评估结果的影响

监测评估结果的影响是一项非常有意义的工作, 因为如果评估结果未能产生积极影响, 那么花费相当的人财物进行技术评估将没有任何实际价值。

一个评估报告能够产生多大的影响不仅取决于评估报告的质量本身, 还依赖于其传播的广泛性和潜在使用者的兴趣, 并且受到不断变化的环境的影响 (卫生系统改革、经济水平变化等)。常见影响卫生技术评估结果产生作用的因素包括:

1. 提供技术服务的机构 如医院和医师所属单位情况, 综合医院或专科医院、营利性医院或非营利性医院。
2. 医务人员 医生类型如内科医师、外科医师、护士等, 专业、培训情况、所属机构和获得最新文献信息的能力。
3. 环境因素 城市或农村、经济状况、是否参加保险。
4. 评估结果 / 建议的特点 结果强度、类型和表达形式、政府的干预、费用、对提供者利益的影响。

考察评估结果的影响主要看它能否影响到政策法规的制订、能否对技术的传播和使用产生实质性影响、能否改变医生的行为以及能否改变患者的认知。前两项比较容易考察, 而后两项则相对较难评估, 而且调查医生行为改变和患者认知提升也需要相当的费用。

目前国际上也正在探索如何有效传播评估结果, 使评估结果能够更好地被潜在的使用者所利用。

### 第三节 卫生技术评估的应用

卫生技术评估是解决医院医疗费用上涨,合理使用医院医疗设备与药品以及制定卫生政策等问题的有力工具。其主要作用是为各层次的有关人员提供科学信息,进行适宜技术的选择、合理配置资源,提高卫生资源利用的质量和效率。

#### 一、应用于医疗技术管理

医院在应用医疗技术过程中出现了大量问题,一类是技术的不合理应用,另一类是该实施的技术没有实施。第三类问题是在医疗实践中对某些技术存在大量争论,以至于不知是否该应用。充足的信息是技术管理的决定因素之一,在医院管理中需了解哪些技术将要出现以及它们的效果如何,这就需要进行卫生技术评估,同样医院淘汰不适宜的技术,也需要进行卫生技术评估。

##### (一) 评估医疗新技术

为了保证服务的安全有效,并将成本控制在可接受的范围,引用任何一种技术都应有严格的卫生技术评估结果作为依据。随着新技术、新药品、新设备的不断涌现,功能和作用相同或相近的项目将越来越多。综合评价有替代效应的技术、药品和设备之间的费用与效果差异再有选择地使用,将成为卫生行政部门、医院保险机构、医务人员无法回避的问题。

##### (二) 评估淘汰医疗旧技术

包括完全废弃低效、无效、昂贵、副作用大或不符合伦理的旧技术,也包括停止或限制某些技术在某一领域的运用。例如,治疗腰背痛的方法多达120余种,卫生技术评估发现没有一种方法被证明有效,一些常用疗法如某些外科手术等甚至有害,因而应该停止使用所有昂贵的治疗,投入经费深入研究腰背痛的机制和有效干预措施。

##### (三) 评估辅助检查的适宜性

卫生技术评估发现,医院的术前常规胸片检查规定花费较大,但收效甚微,仅有30%的手术患者需要术前作胸片检查。1994年,经加拿大政府有关部门立法通过,其中52所医院废除了此项规定,既简化了检查程序,又可以每年净节约500万美元。

#### 二、应用于药品管理

评估医院治疗药品的效果与费用。根据卫生部2008年第四次国家卫生服务调查数据,次均门诊费1998年实际费用65元,2008年可比价为151元,次均住院费用1998年实际费用2515元,2008年可比价为4531元,次均门诊费用、次均住院费用分别增长了132.3%、80.2%。是否每一种药品都起到了人们在应用时所期望的疗效,就需要进行技术监测。在新药研制过程中,由于动物试验及临床试验都有一定的局限性,因此,将已应用于临床的药品再次从药理学、药效学、经济学的角度进行卫生技术评估,评价其有效性、安全性,将利用度低、价格昂贵、安全性差的药品排除在医院大门之外,使患者用上质量可靠、安全性能好、价格低廉的药品,最大限度控制药品费用比例。如瑞典一项评估治疗轻型高血压药物的报告显示:老药和新药降压效果相当,但价格仅为新药的一半,还可预防并发症和早死,从而阻止了新药的应用。英国的一个系统评价报告:使用硫酸镁在预防子痫发作、减轻呼吸系统症状、以至避免或减少围生期死亡方面,其效果比传统的冬眠合剂更好,而且硫酸镁价格便宜,临床使用更为简便。

#### 三、应用于医疗设备器械管理

评估医疗设备的配置和医疗机构大型设备的更新换代。卫生部1996年的统计资料表明,80%以上的部属医院购置了在20世纪80年代医院极少见的800M以上的X线机、CT、ECT、MRI、彩色B超、

血液净化系统。省级医院上述设备拥有率也达50%以上。据统计,到2000年全国11个省市的12个医疗单位装备了13台伽马刀。这些大型设备比较集中在我国东部、中南部和沿海地区的大中城市,卫生资源的分配呈现出明显的不平衡现象。而作为伽马刀发源地的欧洲,整个欧洲总共装备了1台。再如20世纪80年代中期,东地中海的部分产油国家未经科学评估,盲目大量引进高功率的X线机,其中2/3因为缺乏维修工程师而不能使用,造成近800万美元的浪费。

按照经济学的理论,医院在装备大型设备的前后,都要用卫生技术评估对引进设备进行可行性评价。在我国目前的经济条件下,医院管理者更应该借鉴他人的经验和方法,在不影响医疗服务质量的前提下,尽量减少固定资产的投入规模。在设备引入之后的试运行过程中,要进行包括适应证、有效性、安全性在内的综合评价,确定检查适应证,在杜绝乱检查的情况下充分提高设备的利用率。

(许能锋)

# 第十六章 临床实践指南的循证评价与应用

## 第一节 概述

### 一、临床实践指南的概念与发展

1990年,美国医学研究所提出了临床实践指南(clinical practice guideline, CPG)的定义,即系统开发的多组指导性文件,以帮助医生和病人针对具体的临床问题做出恰当处理,从而选择、决策适宜的卫生保健服务。CPG是缩小当前和最佳临床实践之间差距的临床决策工具。随着现代医学的发展,对疾病的诊治已不再由临床医师的个人经验来决定,而是需有经过正确评价的科学证据的支持。制定和推广高质量的CPG特别是循证临床实践指南,用以指导临床医生从事预防、诊断、治疗、康复、保健和管理工作的,是国际上近年来规范医疗行为、改善卫生保健质量、控制医疗费用行之有效的方法。

指南成为临床实践的一部分,最早可追溯到60余年前,但最近20多年发展特别迅速,并成为各临床专业的热点。英国卫生部近年来还特别提倡使用规范的方法来制定基于证据的循证指南(evidence-based guidelines);美国卫生部则成立了保健研究和质量局(Agency for Healthcare Research and Quality, AHRQ),每年投入数亿美元推动CPG的制定。近10年来,国际性杂志上已发表了几千种CPG。CPG的出现和发展主要与下列因素有关:

#### (一) 临床实践的极大差异

自从20世纪80年代以来,很多研究发现对同样一个临床问题,不同国家、同一国家的不同地区甚至在同一个人地区内的不同社区其处理方法各式各样,有很大的差异性。例如一项关于中国和英国对急性缺血性脑卒中处理方法的对比研究发现,19%~69%的中国医师常规使用的7种疗法(除外阿司匹林),英国医师的使用率均不超过1%(表16-1)。对于非瓣膜性心房颤动患者,用来预防脑卒中的华法林使用率在美国南部与中西部之间的差异达4倍之多。由此可见,不同地区、医院、卫生工作者之间医疗实践的巨大差异是显而易见的,而这些差异已经超过了临床、人口学及地域上的差异所能解释的范围,并令人对治疗措施的科学性受到了怀疑。基于证据的CPG则可缩小这些差异,从而规范医疗行为,使患者得到应有的合理的医疗服务。因此,CPG对于规范临床实践行为具有重要的意义。

表16-1 中英急性缺血性脑卒中治疗实践对比

治疗方法	中国医师 (%) <sup>*</sup>	英国医师 (%) <sup>†</sup>	治疗方法	中国医师 (%) <sup>*</sup>	英国医师 (%) <sup>†</sup>
甘油 / 甘露醇	69	1	低分子右旋糖酐	44	0
中药	66	0	蛇毒	30	0
阿司匹林	54	39	激素	19	<1
钙拮抗剂	53	<1			

\* 回答常规使用各种疗法的中国医师的百分比; † 回答常规使用各种疗法的英国医师的百分比

(资料来自 Chen ZM, et al. Hospital management of acute ischemic stroke in China. J Stroke Cerebrovasc Dis, 1997, 6 : 361-367)

#### (二) 医疗费用的快速上涨

有限的卫生资源不能满足对医疗保健无限增长的巨大需求是全世界面临的难题。卫生服务需



求不断增加,新的治疗方法和技术的不断增加,医疗服务手段的日益多样化、复杂化,使得医药费用呈直线攀升,各国政府和医疗保险机构对此感受到了不堪承受的压力。例如,美国的医疗费用开支巨大,年支出已超过2万亿美元,约占国民生产总值的18%,庞大的补助经费让政府不堪重负。因此,奥巴马政府提出了健康保险改革方案,但是阻力很大。所以,更加明智而不是盲目地使用有限的卫生资源已经成为共识。对于一组类似的患者,根据科学证据包括成本-效益(cost-benefit)分析、成本-效果(cost-effectiveness)分析、成本-效用(cost-utility)分析、成本最小化分析(cost minimization analysis),制订一套规范化的临床诊疗路径,对于制订医疗费用补偿政策、合理高效地使用有限的卫生资源具有重要意义。

### (三) 医疗措施的使用不当

有研究显示在临床日常诊疗实践中,大约有1/4至1/3的医疗措施存在着滥用(overuse)、误用(misuse)或使用不足(underuse)等问题。如美国的一项研究显示,约半数的普通感冒和2/3的急性支气管炎患者接受了抗生素治疗,而抗生素的滥用不但增加了患者的经济负担,而且增加了副作用和产生耐药的机会。还有调查研究则显示,对于充血性心力衰竭患者,如果没有禁忌证或不能耐受,则必须使用血管紧张素转换酶抑制剂(ACEI)和 $\beta$ -受体阻滞剂来改善患者的预后,但实际的使用率却非常低。另一项来自亚洲6个国家和地区2600例患者降脂达标调查发现,在冠心病高危患者中,只有1/3达到低密度脂蛋白胆固醇(LDL-C) $<2.59\text{mmol/L}$ (100mg/dl)的目标;12%的冠心病高危人群达到LDL-C $<1.81\text{mmol/L}$ (70mg/dl)的目标。为什么大部分患者未达标?关键是医生的认识不足。国内2004年的一项包括3000名心血管医生的问卷调查发现:12.9%不了解冠心病LDL-C目标值;43.3%不了解血脂筛查对象;32.4%不清楚调脂治疗的首要目标;26.4%仅依据化验单参考值范围判断血脂异常;20.5%认为血脂达标后即可减量或停药;42.2%认为不需要疗效监测。

综上所述,由于不同地区、不同医院及不同医务人员的临床实践存在极大的、难以解释的差异;卫生服务的需求不断增加,医药费用的直线攀升;医疗措施明显不恰当的使用;卫生资源利用不平等,以及基层医生、部分转型的全科医师业务素质偏低等等原因,制订CPG以规范临床实践引起了广泛的重视。医学界普遍认为CPG可以减少不恰当的医疗行为,改善患者的预后。因此,十多年来,制订临床实践指南成为国际热点,各国的医学团体、政府机构及其他组织纷纷制订出各种疾病诊治的CPG,通过提供必要而适当的建议来指导临床实践,从而使治疗更加充分和有效,也使医疗费用的支出更有价值。因此,循证CPG成为提高医疗质量及卫生决策质量的重要保障。然而,我国临床指南的开发刚刚起步,许多所谓的临床指南,多为一些临床经验手册。有些指南虽然在编写过程中运用了循证医学知识,但仍然不能属于真正意义上的循证临床实践指南。

## 二、临床实践指南的作用

CPG可以使临床医师提高医疗保健质量、改善临床结局,甚至对治疗指南的依从性可作为临床结局的预测指标。2005年发表的MAHLER研究就是观察心脏病科医师遵循欧洲心力衰竭诊治指南的依从性对患者临床结局(心衰发生率、心血管住院率和至心血管住院的时间)的影响。研究从欧洲6个国家随机抽取150名心脏病医师,治疗1410例心力衰竭患者(平均年龄69岁,69%为男性,NYHA II、III、IV级分别占64%、34%和2%),随访6个月。观察指标:3项依从指标(血管紧张素转换酶抑制剂、 $\beta$ -受体阻滞剂和螺内酯的应用)和5项依从指标(血管紧张素转换酶抑制剂、 $\beta$ -受体阻滞剂、螺内酯、利尿剂和强心苷的应用)。结果显示:3项指标完全依从者(100%)、中度依从者(50%~67%)和低度依从者(0~33%)心力衰竭住院率和心血管住院率分别为6.7%、9.7%、14.7%( $P<0.002$ )和11.2%、15.9%、20.6%( $P<0.001$ )。研究提示心脏病科医师对治疗指南的依从性是心血管住院的预测指标。

因此,以循证医学为基础的CPG对临床实践具有重要的意义。第一,可以提高医疗质量,给予患者最佳治疗和合理治疗。第二,可减少不同医疗机构和不同医师间医疗实践的差异。第三,可减少患

者的医疗费用。第四,可作为医疗质量检查的依据。第五,可作为医疗保险的凭证。第六,有助于医务人员的终身继续教育。一份好的 CPG 应具有真实性、可靠性、可重复性、临床实用性、临床适用性和明确的目标性,是多学科参与制作的结晶,彰显了当前的最佳临床证据。

### 三、指南与其他证据的关系

循证医学提倡在临床实践中,尽可能使用当前可得到的最佳证据、结合临床经验和患者的意愿进行诊疗方案的选择。作为循证医学资源的一部分,指南与原始研究证据和系统评价的区别在于:指南为临床医生提供具体的推荐意见指导医疗行为,是连接证据和临床实践的桥梁,更加贴近临床实践的需要。加拿大 McMaster 大学临床流行病学 Haynes 教授提出了支持循证卫生决策的循证医学证据结构的“5S”模型。这一模型充分体现了临床实践指南在循证医学证据体系中的地位及其重要的临床指导意义。

“5S”模型将原始研究(studies)作为最基层,综述(syntheses,即系统评价)作为次基层,然后是摘要(synopses,如循证期刊对原始研究和综述的简要描述),接着是循证教科书中相关的摘要、综述或研究的总结(summaries),其相对于单个摘要、综述或原始研究甚至它们的总和而言,都更具优势。比汇总性更强的资源便是系统(systems),系统通常是整合有 CPG 的计算机决策支持系统,可根据个体患者的特征(如电子病历)链接相关证据,置于模型的最顶层。系统可提醒或告知医护人员治疗的决策,因而对临床实践具有重大的指导意义。遗憾的是目前这种计算机决策支持系统寥寥无几,且未能做到及时更新。但可以预见这将是 CPG 网络资源利用研究的一个重要发展方向。

### 四、循证CPG的网络资源

目前,许多国家建立了 CPG 网站,其中比较权威和影响力的有:

1. 美国国家指南交换库(National Guideline Clearinghouse, NGC) 是 AHRQ、美国医学会(American Medical Association, AMA)和美国卫生健康计划协会(American Association for Hospital Planning, AAHP)于1998年联合制作的一个提供 CPG 和相关证据的功能完善的免费数据库。目前收集有来自全世界200多个指南制定机构提供的2 400余份指南。NGC 提供直接检索和浏览两条检索途径,并可对收集的指南进行比较。它的主要特点有:①对指南的内容进行了分类,部分指南全文可链接,可订购指南。②提供结构式摘要,可进行指南之间的比较。③提供电子论坛,交换临床实践指南方面的信息。④对指南的参考文献、指南制作方法、指南的评价和指南使用等提供有链接、说明或注释。网站: <http://www.guideline.gov>。

2. 加拿大医学会临床实践指南(Canadian Medical Association Clinical Practice Guidelines)数据库 于1995年由加拿大国家、州或地区医学卫生组织、专业协会、政府机构和专家小组共同主办并认可。指南由加拿大各地和各机构团体提供。网站提供有关键词搜索、浏览、基本检索和高级检索等多种检索途径。

3. 苏格兰学院间指南网络(Scottish Intercollegiate Guidelines Network, SIGN) 网站建于1993年,重点关注癌症、心血管疾病和心理卫生等领域。网站的栏目有指南、指南选题提示或范围、当前指南项目组指南开发的方法学等。网站: <http://www.sign.ac.uk/guidelines/index.html>。

4. 新西兰指南研究组(The New Zealand Guideline Group, NZGG) NZGG 于1996年在新西兰卫生委员会领导下建立,主要目的是为了制定和实施循证临床实践指南。该网站将指南分为四种类型:基层医疗服务管理指南、病人转诊和管理指南、第一专科评估准入标准指南和临床优先评估标准指南。网站: <http://www.nzgg.org.nz>。

其他还有德国指南库(German guideline clearinghouse)、英国临床指南网站、英国国家临床优化研究所(National Institute for Clinical Excellence, NICE)和芬兰的 EBM Guidelines 等。

## 第二节 临床实践指南制订的流程与方法

开发大量的高质量循证CPG是国际上近年来规范医疗服务、加强医疗质量管理、控制医疗费用的行之有效的方法。许多国家为开发CPG制定了符合本国国情的指南开发程序,并取得了令人瞩目的成就。

临床实践指南的制定方法一般分为两大类。

### (一) 专家共识指南制定法 (consensus guideline development)

又分为非正式和正式的专家共识制定法。前者由一组专家开会讨论,将一次或多次开会讨论后达成的共识形成推荐意见作为指南,由专业学会或政府机构进行指南发布。这种指南的推荐意见缺乏证据基础,指南易受参会人员的专业、优势、性格、组织和政治因素等影响,专家们认为有益的措施并不能保证事实上真正有益,因此这种指南的可靠性和质量较差。正式的专家共识法是就某一治疗措施给专家组提供相关研究证据的综述及可能的适应证清单,在第一次专家组会议之前,专家组成员各自对每个适应证评分以评价其适用性,量表共9分,1分为完全不适用,9分为特别适用,5分为可用或不可用。开会时专家们将小组集体评分的情况与自己的评分相比较,讨论不一致的原因,然后再次重复评分,在会议讨论的基础上修改评分。最后的评分反映了专家组成员的一致性程度。正式的专家共识指南制订法其特征仍是专家的主观意见是确定适用性的基础,虽然也考虑了研究证据,但没有将推荐意见与相关证据的质量明确地联系在一起。

### (二) 循证实践指南制定法 (evidence-based guideline development)

循证CPG的制定过程与以往撰写指南的过程有很大不同,它包括组成指南开发小组,提出相关临床问题,系统检索文献和使用正确的方法对证据进行严格评价,并结合他们的实践经验,再根据证据的级别和强度提出推荐意见。此外,还包括系统评估、推广普及、修订更新等指南推出后的工作计划,使指南能与时俱进。像任何科学程序一样,制定循证临床实践指南的方法学是基于证据的方法学,其结论或推荐意见须有可靠的证据支持。将推荐意见与相关的证据质量明确地联系在一起是循证临床指南的明显特征。本文将重点介绍。

SIGN推荐的循证指南制定较具有代表性,其开发程序为:指南开发组织——确定指南题目——组成专题指南开发小组——系统文献评价——草拟推荐建议——咨询及同行评议——发表与发行——地方应用——审计及评价。这是一个循环发展的过程,其最终目标就是改善临床结局和提高病人的健康水平。SIGN制定指南的主要步骤如下:

1. 组建指南开发小组 由来自不同地区的多学科人员(15~20人)组成。指南制定需要有四个核心技能:临床专业技能;卫生保健的实践经验;专业知识(如患者的意愿和卫生经济学);严格的评估技能。

2. 文献检索 指南开发小组确定指南拟解决的主要问题。然后,专业图书管理员和信息专家进行系统的文献检索。在Cochrane Library、Embase、Medline,重要的专业学会、协会和指南出版机构的网站,以及正在进行的试验注册资料库和其他相关的数据库反复进行检索。先检索已有的指南及系统评价,其次检索随机对照试验,最后根据所提出的问题和证据获得的数量再检索其他类型的临床试验。

3. 评价证据 指南开发小组需制定一套明确的文献纳入和排除标准,并采用一套根据临床研究设计标准清单严格评价相关文献。每份清单的结论就是一份质量量表或是证据的分级。

(1)对证据内涵和质量的评价,包括:①证据的一致性:总体一致性;入选人群特征(如年龄、性别、宗教等)的一致性;研究内容的一致性;②外部真实性:研究结果是否与实际运用时的结果一致或者相反?③针对性:证据是否直接针对指南的目标人群或者人群特征的不同将会影响最终结果?④证据容量:即患者的数量和研究的数量。

(2) 证据解释, 包括: ①病人意愿: 权衡利弊; 患者结局指标的最大改善; ②临床实践: 是否与现有的医疗实践有较大的差距; ③资源分配: 是否会导致大规模的资源重新分配, 卫生系统是否支持改进的措施?

每一篇文献至少应由两名指南制定小组成员进行评价, 如果存在分歧, 则由第三者仲裁解决。

4. 谨慎判断并提出建议 经过严格的证据评价后达成共识, 根据支持证据的强度来决定建议的等级, 并制定出指南初稿。

5. 咨询和同行评价 召开会议, 向指南小组提出疑问及对指南初稿做出评价。指南小组根据建议进一步修订指南。修订版再送同行专家进行评价。最后, SIGN 编辑组对指南进行审查并作出评价。

6. 评估 指南发布2年后再进行评估。对该领域的进展做出评价, 以决定是否出版更新指南。

7. 患者参与 病人、护理者和研究者一起工作, 确保从病人或护理者的角度参与指南制定过程, 为公众健康的新规划提供指导意见。

8. 文件存档 保存下列文件: 制定指南的原始提议; 制定指南的理由和指南涉及范围; 确定指南的关键问题; 检索策略、数据库和文献检索的时间范围; 文献评价的纳入和排除标准; 对支持建议的文献所用的方法学清单; 回答所有关键问题的证据总结表; 谨慎判断的表格; 列表说明指南小组对整体证据的质量和相关推荐分级的结论; 总结性大会和同行评议的评论及回复记录。

9. 指南的执行 指南制定与地方临床实践相孤立是一个核心问题, 应充分考虑地方的资源分配。

10. 资料来源和其他因素 制定每一份指南须耗费大量的金钱和时间。为取得预期效果, 项目必须由遵循方法学的专家管理, 在规定的时间内完成。

此方法制定的指南推荐意见有科学客观的证据依据, 令人信服; 同时又标注了推荐意见的强度, 便于使用者根据其强度是否采用其推荐意见。CPG 一般都标注推荐意见级别和证据等级, 但不同的国家和学术机构采用的标准也不同。如美国心脏学会 (American College of Cardiology, ACC)、美国心脏协会 (American Heart Association, AHA) 和欧洲心脏学会 (European Society of Cardiology, ESC) 采用 I~III 类推荐和 A~C 级证据等级分类法。I 类推荐: 益处 >> 风险, 治疗/操作应该被给予; II 类推荐: 分为 II a 和 II b 类, II a 类推荐为益处 >> 风险, 多数证据支持该治疗/操作, 而 II b 类推荐为益处  $\geq$  风险, 较少证据支持该治疗/操作; III 类推荐: 风险  $\geq$  益处, 推荐的治疗/操作无效或无益, 甚至有害。A 级: 证据来自多个大规模随机对照试验或 meta 分析; B 级: 证据来自单个随机对照试验或大型非随机对照试验; C 级: 证据来自专家共识、回顾性研究或注册研究。

为了建立评价证据和推荐意见分级的国际标准体系, 2004年由 AHRQ、NICE 及 WHO 组成了 GRADE (Grading of Recommendations Assessment, Development and Evaluation) 工作组。GRADE 涵盖所有医学专业和临床护理领域的各种临床推荐意见, 力求分级系统简单明了, 易于掌握。GRADE 系统将推荐意见分为强、弱2个级别。当明确显示干预措施利大于弊或弊大于利时, 应评为强推荐; 当利弊不确定或无论质量高低的证据均显示利弊相当时, 则视为弱推荐。证据质量分为高、中、低和极低4个等级。基于 RCT 的证据在 GRADE 4级证据质量分级系统中, 被评为高质量(表16-2)。但并非所有 RCT 的质量都一致, 下述5种情况将降低其证据质量(表16-3)。

表16-2 证据质量及其定义

质量等级	定义
高	未来研究较少可能改变现有疗效评价结果的可信度
中	未来研究可能对现有疗效评估有重要影响, 可能改变评价结果的可信度
低	未来研究很有可能对现有疗效评估有重要影响, 改变评估结果可信度的可能性较大
极低	任何疗效的评估都很不确定

表16-3 决定利益、风险、负担和费用可信度评价的因素

## 可能降低随机对照试验证据质量的因素

1. RCT 试验设计和实施质量低下, 高度提示可能存在偏倚。如未进行分配隐藏、未采用盲法、失访人数过多和过早终止试验等。
2. 研究结果的不一致性(异质性), 研究者无法对几个 RCT 结果的异质性作出合理解释时。
3. 非直接证据, 受试人群、干预措施、结局指标不同的 RCT, 研究结果为间接证据。
4. 证据不足, 如总样本量小且结局事件的数量少时, 将增加判断利弊的不确定性。
5. 报告偏倚, 包括发表偏倚。

一份好的指南应具有两个组成部分: ①对证据的总结, 以得出一种干预措施对典型患者平均效果的证据。②对如何使用该证据的推荐意见。推荐意见还应说明干预措施的利弊、局限性、最适宜的患者和人群, 以及与成本和卫生保健有关的其他因素。虽然制定临床指南要使用系统评价的证据作为基础, 但当一些常见和重要的临床问题还缺乏充分的 A 级证据时, 此时指南的建议可能基于较差的证据, 并应取得小组成员的共识。这样的指南既是现实的, 也是可以接受的。

### (三) 临床路径

临床路径 (clinical pathway, CP) 是指医院的一组人员共同针对某一病种的监测、治疗、康复和护理, 所制定的一套有严格工作顺序和准确时间要求的诊疗计划, 以减少康复的延迟及资源的浪费, 使服务对象获得最佳的医疗服务质量。其包含四个要素: ①其对象是针对一组特定诊断或操作; ②其制定是多学科知识综合的过程, 包括临床、护理、药剂、检验、麻醉、营养、康复、心理以及医院管理, 甚至有时包括法律、伦理等; ③其设计要依据住院的时间流程, 结合治疗过程中的效果, 规定检查治疗的项目、顺序和时限; ④其结果是建立一套标准化治疗模式, 最终起到规范医疗行为、减少变异、降低成本、提高医疗质量的作用。临床路径具有综合性、时效性、多专业合作性、陈列主要治疗与护理活动以及结果测量的特点。临床路径作为一种既能贯彻关键质量管理原则, 又能节约资源的标准化治疗模式, 为医院的质量建设提供了一种切实可行的管理手段, 越来越受到普遍关注。

临床路径以简单明了的流程, 将常规的治疗、检查与护理活动细化, 由各学科的专业人员将该疾病关键性的治疗、检查和护理活动标准化, 根据住院天数设计表格, 使治疗、检查和护理活动的顺序以及时间的安排尽可能达到最优化。使大多数的服务对象由入院到出院都能依此流程接受服务, 以使其获得最佳的服务, 缩短平均住院日, 减低医疗成本, 减少医疗资源的浪费。实施临床路径的目标是保证资源的有效合理使用; 降低消费和缩短住院时间; 确保护理的连续性; 促进专业化的协作配合, 为持续的质量改进提供多学科工作的基本框架。临床路径在为患者提供适宜医疗服务的同时实现资源成本最小化, 医院必须要在质量管理方面寻求较先进的质量管理方法或模式。

尽管临床路径的开发要依据临床指南, 但两者有所不同。临床指南是公认的声明, 是被系统地开发出来, 其内容要经过严格的评价、用以在特定的临床环境中帮助医务工作者对医疗活动进行决策。所以, 临床指南更具有权威性、宏观性, 适用范围更广。与之相比, 临床路径更加细化医疗过程, 关注医疗过程中的重点环节。注重对过程中无效行为的控制, 具有高度的时效性。虽然在路径开发中应该基于临床指南, 但路径所包含的主要过程通常没有经过严格的验证、可以依据医院本身的实际, 超出指南规定的范围。

## 第三节 临床实践指南的评价

循证临床实践指南现已逐渐成为制定指南的趋势。但不同的国家或学术组织针对同一种疾病可能制定了不同的指南, 这些指南质量参差不齐, 有很大的异质性。某些建议甚至互相矛盾, 给临

床决策带来极大困扰。哪些指南的质量高,建议可信?哪些指南的质量差,建议不可信?应该怎样将指南应用于临床实践?这些都是医生在临床实践过程中经常面临的问题。因此,对指南进行评价,以判断指南是否值得推荐使用或者从众多的指南中选择质量最好的应用于临床,这是应用 CPG 前的重要步骤。

## 一、对临床实践指南的评价

### (一) 临床实践指南的真实性评价

好的指南必须使用循证医学的原则和方法。强调 CPG 应建立在证据的基础上,并根据证据的可信程度对建议进行分级。评价的要点包括:

1. 指南编写者是否做了全面、可重复的文献检索?检索是在过去1年内进行的吗?
2. 是否每项建议均标明了其相关证据的等级,并提供了原始证据的链接或文献?

评价主要集中在对证据的收集、评价和合成,以及如何将推荐意见与相关的证据紧密结合方面。

### (二) 临床实践指南的重要性评价

指南是否回答了临床需要解决的重要问题?经过对 CPG 真实性评价后,还要明确指南是否回答了临床需要解决的重要问题, these 问题是临床医生必须面对的。但要注意的是临床所面临的问题相当复杂,指南不可能囊括所有的临床问题。

### (三) 临床实践指南的适用性评价

1. 在本地区,疾病负担(burden of disease)是否很低而无需参考指南?

疾病是否在本地区极少发生?我的病人是否不可能发生指南中所描述的结果?如果是,则应用指南不仅浪费时间金钱,还可能造成不必要的损害。对于个体病人,还应综合考虑:①我的病人与研究中的病人存在的差别大吗?②治疗在现有的环境条件下可以施行并使用吗?③治疗有哪些潜在的利益和损害?④对于结局和治疗,病人的观念和预期是什么?

2. 病人对治疗价值的看法(beliefs),对其利害的效度评价是否与指南中可比?
3. 执行该指南所需的成本(bargain)有多大,即同样的资源用于别的措施是否有更大的收益?
4. 对我的病人,是否有实施该指南不可克服的困难(barriers)?

实施指南的困难包括地域性的(如本地区根本无此治疗方法);传统性的(如习惯采用另一种治疗方法);权威性的(教授怎么说就得怎么做);法律性的(医生惧怕因为舍弃了常用但效果不明显的疗法会遭到起诉)或行为性的(医生无能为力或病人不能服药)。如果这些困难明显,则不值得执行指南。

因此,一个指南的成功实施依赖于与四个因素(4B),即疾病负担(burden)、价值取向(beliefs)、花费(bargain)和障碍(barriers)的吻合程度。对于这些情况患者和医生最清楚。如果没有这些情况,就可以考虑指南的应用。但应注意指南的推荐意见是原则性的,应在指南的指导下根据个体化原则诊治患者。

## 二、临床实践指南评价工具

为了科学客观地评价 CPG,不同的国家和学术团体制定了许多专门的 CPG 评价工具。美国医学研究所(Institute of Medicine, IOM)早在1990年就发表了第一个针对 CPG 的评价工具,包括效度、信度、临床实用性、临床灵活性、透明度、多学科联合开发、指南定期评价、指南开发的主要文件等8条标准。2002年美国 COGS 会议确定了18条评价指南的标准。迄今为止共有20多个评价 CPG 的工具。Vlayen J 等对 CPG 的评价工具进行了系统评价。以“practice guideline, appraisal and evaluation”为关键词检索了 Medline, Embase, and Cinah 数据库。共纳入24个评价工具。结果发现共有3个评价工具对指南的各要素进行了全面评价,其中 Cluzeau 量表是唯一经研究验证过其真实性的量表,但 Cluzeau 量表有37个条目,临床应用不方便。指南研究与评价工具(appraisal of guidelines research

and evaluation, AGREE) 在 Cluzeau 量表的基础上进行了精简。已经有研究证实其真实性, 且被广泛采用。因此, 该系统评价推荐使用 AGREE 量表来评价 CPG。目前, 美国 COGS 评价标准和欧洲 AGREE 量表是两个基本得到公认的指南评价标准的工具。以下将介绍美国 COGS 评价标准和欧洲 AGREE 量表。

(一) 美国COGS评价标准

表16-4总结了美国 COGS 会议制定的18条评价标准。

表16-4 GCP报告的COGS清单

条 目	描 述
1. 概述	提供结构性摘要包括发布日期, 指南情况(原版、修订、更新)、印刷和电子版
2. 重点	描述指南涉及的主要疾病和干预措施, 指出可替代的预防、诊断和干预措施
3. 目标	指南希望达到的目标和达到这一目标的理由
4. 使用者 / 背景	描述指南的使用者和指南应用的背景
5. 目标人群	适合指南推荐的患者人群并列排除标准
6. 制定者	区别指南制定机构与指南制定个人的潜在利益冲突
7. 资金来源或赞助人	确定资金来源或赞助人在指南制定和发布中的作用, 说明潜在的利益冲突
8. 收集证据的方法	描述文献的检索方法, 包括日期、数据库和检索标准
9. 建议分级标准	描述证据质量的分级标准和推荐强度的分级系统。推荐强度与推荐的重要性相关, 并基于证据的质量和预期获益或损害的大小
10. 综合证据的方法	描述证据如何被综合为推荐意见, 如证据表格、meta 分析、决策分析
11. 发布前评审	描述指南发布前是如何进行评审的
12. 更新计划	陈述是否有指南更新计划, 并标注本指南的有效期
13. 定义	描述不常用的术语, 并严格纠正易被指南误解之处
14. 建议与基本原则	准确陈述指南的作用和执行指南的特殊情况, 通过描述证据与推荐之间的联系来判断每一项推荐, 根据第9条来显示证据的质量和推荐的强度
15. 潜在利弊	应用指南预期的获益和潜在的风险
16. 患者意愿	当指南涉及个人选择或价值时应考虑患者的意愿
17. 法则	必要时提供图表说明
18. 指南在执行中需要考虑的事项	描述指南推荐应用时的障碍, 对任何辅助文件给予参考文献以便于应用

(二) 欧洲AGREE量表

为提高指南制定的质量, 欧洲还制定了如何撰写和评价 CPGs 的“指南”。指南研究与评价工具 (appraisal of guidelines research and evaluation, AGREE) 就是由13个国家的研究者制定的一种指南研究和评价的评估工具。并另外提供了使用该工具的培训手册。该审查工具在国际上具有较高的权威性。为目前国际指南质量评价的基础工具。

AGREE 工具的作用是为临床实践指南的质量审查提供一个框架。临床实践指南的质量, 是指在制定指南过程中可能存在的偏倚因素能够得到合理的控制, 从而确保推荐建议的内部和外部真实性及临床应用的可行性。指南审查过程中要充分考虑到指南推荐的利害和成本问题, 以及临床应用的相关问题。因此, 评估的内容包括对指南制定所采用的方法、最终推荐的内容以及应用指南相关因素的审查。指南研究与评价的评审工具不仅对指南的报告质量进行了评估, 而且对推荐结果的其他关键方面也进行了质量评估。该审查工具对指南预测的真实性, 即指南预期要达到的临床结局的可能性做出评估, 但并不涉及指南对患者临床结局影响的评估。

### 1. AGREE 量表

AGREE 量表通过6个部分(指南的范围和目的、参与人员、指南开发的严格性、指南的清晰性与可读性、指南的适用性和指南编撰的独立性)23个条目来对 GCP 进行评分(表16-5)。

表16-5 AGREE评价工具的6个部分23项具体条目

I 指南的范围和目的(涉及指南的总目的、具体的临床问题和适用的患者)
1. 明确说明指南的撰写目的
2. 明确阐述了指南所涵盖的临床问题
3. 明确阐述了指南所要应用的目标人群,应提供年龄范围、性别、临床类型及共病
II 参与人员
4. 指南制定小组的成员包括所有相关方面的专家
5. 指南考虑了目标患者的观点和选择
6. 指南的适用者已经明确规定
7. 指南发表前在适用者中已被试行
III 指南开发的严格性(涉及收集和综合证据的过程,制定和更新推荐建议的步骤方法)
8. 采用严格系统的方法检索证据,应提供检索证据的详细策略,包括使用的检索词、信息来源和文献涵盖的时间
9. 证据的纳入和排除标准有清楚的描述
10. 详细描述了形成推荐建议的方法
11. 形成推荐建议时充分考虑了干预措施的获益、风险和副作用
12. 推荐建议与支持证据之间有明确的联系
13. 指南发表前接受了制定小组以外专家的同行评议
14. 提供了更新指南的程序
IV 指南的清晰性与可读性
15. 推荐建议明确且不含糊
16. 针对不同临床状况提供了不同的治疗选择
17. 关键的推荐建议可以很容易找到
18. 提供了指南方便应用的工具(如概要、教育工具、针对患者的宣传单等)
V 指南的适用性(指南应用时可能涉及的单位、操作和费用问题)
19. 讨论了应用该指南的过程中可能遇到的困难
20. 讨论了应用该指南的过程中可能的费用
21. 提供监督和审查指南执行情况的评估指标
VI 指南编撰的独立性(推荐建议的独立性和对指南制定小组中各成员利益冲突的说明)
22. 指南的编辑工作独立于赞助单位
23. 指南制定小组成员的利益冲突在指南中有所说明

### 2. 使用说明

1) 审查者人数: 建议每个指南2~4个审查者,以增加指南评估的可靠性。

2) 评分尺度: 表中每个条目的分数为1~4分,完全符合条目要求的打4分,完全不符合的打1分,介于两者之间的根据测评人员的判断给2分或3分。另外在每项条目后都提供了补充说明的信息,仔细阅读这些信息有助于对条目所涉及问题和概念的理解,并正确合理评分。

3) 各部分得分的计算方法: 每个部分得分等于该部分中每一个条目分数的总和,并标准化为该部分可能的最高分数的百分比。标准化的每一部分得分 =  $[(\text{每一部分的实际得分} - \text{可能的最低得分}) / (\text{可能的最高得分} - \text{可能的最低得分})] \times 100\%$  (表16-6)。



表16-6 第1部分得分的计算方法

	条目1	条目2	条目3	总分
评估员1	2	3	3	8
评估员2	3	3	4	10
评估员3	2	4	3	9
评估员4	2	3	4	9
总分	9	13	14	36

最高可能分数 = 4 (完全符合) × 3 (项目) × 4 (评估员) = 48

最低可能分数 = 1 (完全不符合) × 3 (项目) × 4 (评估员) = 12

该部分的标准化总分 = (实际总分 - 最低可能分数) / (最高可能分数 - 最低可能分数)

= (36 - 12) / (48 - 12)

= 0.67 × 100%

= 67%

根据6个部分的标化百分比综合判断该指南是否值得推荐应用,分3个等级:强烈推荐(单个部分的百分比 > 50%,占6个部分的比例 ≥ 2/3)、推荐(单个部分的百分比 > 50%,占6个部分的比例 < 2/3, ≥ 1/2)、不推荐(单个部分的百分比均 < 50%)。

6个部分的得分是独立的,不能合并为一个质量评价的总分值。虽然部分得分可用来比较不同的指南,并用以确定是否使用或推荐该指南,但不能用单个部分得分作为评判指南好坏的区分标准。

AGREE 网站: <http://www.agreecollaboration.org/instrument/>。

现有评价工具普遍存在的一个缺陷是没有对建议的临床内容和制订建议的依据进行评价。AGREE 量表也缺乏具体标准来评价指南的临床内容,它只提供指南开发过程中的方法学评估,而不涉及推荐意见合理性的审查,因此还需要专家评议和临床试用来补充评估。尽管如此,AGREE 量表仍是目前评价 CPG 的最好工具之一。

### 3. 实例分析

采用 AGREE 量表对2008ESC 急性和慢性心力衰竭诊断和治疗指南进行评价。指南由来自欧洲10个国家15位专家撰写,共115页,参考文献242篇。此次是继1995、1997、2001和2005年的第5次更新,第1~6部分标准化总分得分分别为88%、58.3%、85.7%、91.6%、77.7%和100%。根据 AGREE 综合评价,该指南的临床应用为强烈推荐。

## 第四节 临床实践指南应用的原则和方法

仅仅制定出 CPG 并不能达到改善患者预后的目的。CPG 只有具体应用于个体患者时才能发挥改善患者预后、提高医疗服务质量的作用。将经过严格评价的指南应用于临床,需要考虑以下一些问题。

### 一、临床应用指南的原则

应该明确 CPG 只是为临床医生处理临床问题制定的参考性文件,不是法规。为避免不分具体情况盲目地、教条地强制性照搬使用,应注意以下原则:

1. 个体化原则 由于制定 CPG 时采用的证据绝大多数是基于人群的临床试验,推荐是对多数(典型)患者或多数情况提供的普遍性指导原则,不可能包括或解决每一个体患者所有复杂、特殊的临床问题。因此,在应用指南时,应充分考虑患者的社会人口学特征和临床特征是否与指南的目标人群一致。面对个体患者,临床医生应该在指南指导下,根据具体病情和多方面的因素个体化地选择治疗方案。而应用临床技能和经验迅速判断患者的状况和建立诊断的能力,以及判断患者对于干预措施

可能获得的效益和风险的能力是临床医生正确使用指南做出恰当临床决策的基础。

2. 适用性原则 如果患者的情况与指南的目标人群相似,可以考虑应用指南推荐的干预措施。那么就要根据本地区或医院目前的医疗条件,评估该干预措施的可行性和成本-效益比,以及患者的经济状况,对医疗费用的承受能力,医疗保健系统的覆盖支持能力等。例如各国指南均推荐急性心肌梗死早期(3~12小时内)行经皮冠状动脉介入(PCI)治疗,但我国绝大多数基层医院并无条件开展此项技术,且多数心肌梗死患者也无法承受相应的高昂费用,此时就只能采取指南建议的其他药物治疗措施。

3. 患者价值取向原则 患者或其亲属的价值取向和意愿在临床决策中具有重要的作用。指南的推荐强度越强,采取该项干预措施预期获得的效益-风险比越大,患者选择该项干预措施的可能性也越大,绝大多数患者都会选择接受该项治疗。而对于那些推荐强度较弱的干预措施而言,预期的效益-风险比则变得不确定,不同的患者可能选择截然相反的干预措施。例如下肢深静脉血栓的患者已经口服华法林1年,如果继续服用华法林可使再发下肢深静脉血栓的风险每年减少约10%,但同时需要定期检测出凝血时间且出血风险相应增加,因而一部分患者可能要放弃。

4. 时效性原则 随着医学的不断发展,每天都有大量的基础和临床试验结果问世。过去认为有效的治疗手段可能被新的证据证明无效,而过去认为无效甚至禁忌的治疗手段可能被新的证据证明有效。例如既往认为充血性心力衰竭是使用 $\beta$ -受体阻滞剂的禁忌证,但众多的大型随机对照试验却证实 $\beta$ -受体阻滞剂可以显著改善心力衰竭患者的预后。因此,新指南认为 $\beta$ -受体阻滞剂是治疗充血性心力衰竭的极重要的药物(I类推荐, A级证据)。因此,应用指南时应注意时效性,尽可能选择最新的CPG。

5. 后效评价原则 后效评价是指在患者接受根据CPG制订的方案后,对患者病情的变化进行临床随访。后效评价在整个循证临床实践中具有重要作用,也可以为指南的修订和更新提供临床资料。

## 二、指南的临床应用方法

1. 了解指南的制定方法,一项真正的循证CPG较非循证CPG的可靠性更强。

2. 对于推荐建议要注意其推荐等级与证据强度,了解其意义,以便判断推荐研究的可靠程度。

3. 根据推荐强度确定临床应用。建议表述清楚、不存在争议、采用循证医学方法制定的指南与建议表述不清、存在争议或基于专家意见的指南比较,前者的临床应用情况明显优于后者。

4. 消除指南实施中的障碍。指南在实施过程中总会面临来自社会、医疗机构和医生自身诸多的障碍。常见障碍包括:①社会因素,如某些新的治疗社保不予支付;②医生因素,盲目自信,缺乏评价证据的能力或繁忙的临床工作使其没有时间评价和实施指南;③患者因素,患者拒绝接受某些治疗;④环境因素,来源于医药公司的误导,上级医师不同意应用指南提供的证据,习惯性给予“常规治疗”等。可采用成立指南实施小组、开展循证医学教育、计算机辅助决策、多专业专家合作等有效措施来消除这些障碍。

总之,由于本地人群的基线特征、医疗卫生资源的分布都可能与指南存在差异,我们在选择指南时应尽可能选择由本地区或本国制定的指南。但遗憾的是我国制定的指南大多为传统的专家共识,指南质量较低。在选用欧美国家指南或国际性指南时,应注意考察其是否适用于自己的患者,根据患者的具体临床情况,将当前所获最佳证据与临床技能和经验相结合,考虑成本-效益比及当地卫生资源的实际情况,并充分尊重患者及其亲属的价值取向和意愿,综合以上因素做出临床决策。

(刘金来)

## 第十七章 循证医学实践的自我评价

前面的章节介绍了循证医学实践“五步曲”的前四个步骤,循证实践的第5步是后效评价,即通过临床实践中应用循证医学的理论和方法进行循证决策,必定会有成功或不成功的经验和教训,如果能够认真总结经验教训,不但自身能从中受益,不断提高自身学术和医疗水平,同时还有利于同行之间的互相交流和学习。因此,后效评价对于临床医生进行终身的继续教育和自我提高是非常关键的过程。

后效评价一个最简单的途径是通过观察一个循证证据应用于个体病例之后是否获得了预期疗效,但这又很难做得到,因为随访的难度很大,而且常常受到其他因素的影响。因此,目前的后效评价往往通过其他途径来解决,常用的方法有自我评价和效果评价。本章将重点介绍循证实践的自我评价。

### 第一节 循证实践能力的自我评价

循证医学实践中,最重要的评价方式就是对于自我设计及行动的自我评价,自我评价应该从一个循证实践者的学生时代开始,并贯穿他的整个职业生涯。这种自我认识的技能不但能够帮助临床医生意识到自身的能力和缺陷,而且也能够帮助临床医生有效地将可获得的最佳证据与临床实践整合。接下来将分步骤地对需要进行自我评价的部分进行详细介绍。

#### 一、评价“提出问题”的能力

临床实践者应问自己以下这5个问题(表17-1)。首先,我们有没有提出问题?其次,问题的描述是否符合一定的格式?在积累了一定的经验之后,可否进一步明确主要问题所在,明确自己还欠缺哪些知识,并对最初提出的问题进行修改?提出问题的过程中若遇到障碍,能不能想办法克服?是否有时间和动力随时记录临床实践中随时可能出现的新问题,以待今后解决?如果没有这个习惯,我们可能会失去一些提高临床技能的机会,应继续学习循证医学提出问题这部分内容。

表17-1 对“提出可以回答的问题”的自我评价

- |  |
|--|
| 1. 有没有提出临床问题?  |
| 2. 问题的陈述是否简洁明了,符合一定的格式?<br>A. 有关“背景”知识两部分问题<br>B. 有关诊断、管理等“核心”知识的四(或三)部分问题 |
| 3. 有没有利用绘图法来明确自己的知识缺陷,并对最初提出的问题进行修改?                                       |
| 4. 提出问题的过程中若遇见障碍,能否想法克服?   |
| 5. 有没有养成随时记录问题以待以后解决的习惯?   |

(引自 Straus SE, 2005)

在提出问题过程中,我们应该考虑提出问题的成功率是否提高了,如果我们提出问题的成功率很高,可以继续提出问题;如果成功率较低,这时可以咨询其他资深的同行或导师,也可以参加一些EBM的继续教育学习班。

## 二、评价“寻找最佳外部证据”的能力

表17-2罗列的一些问题有助于我们对寻找最佳外部证据的效果进行自我评价。有没有去寻找证据？是否了解对临床决策来说现有的最佳临床证据的来源？是否能迅速寻找到临床实践所需的硬件、软件及最佳证据？当我们开始检索时，能否从庞杂的信息来源中找到我们需要的外部证据？检索效率是否更高？例如：临床决策者可以利用掌上电脑快速的检索 CD 光盘、Cochrane Library 的系统评价、医学教科书、Medline 的原始研究文献等。

表17-2 对“寻找最佳外部证据”的自我评价

- |   |
|---|
| 1. 有没有去寻找证据？  |
| 2. 是否了解本领域内现有的最佳临床证据的来源？                                      |
| 3. 能不能迅速寻找到临床实践所需的硬件、软件及最佳证据？                                 |
| 4. 能否从庞杂的信息来源中找到有用的外部证据？                                      |
| 5. 在寻找证据的过程中有没有逐步提高检索的效率？                                     |
| 6. 在 MEDLINE 数据库中检索时有没有使用截断、布尔语言、MeSH 词头、词典、限制词及无文本智能检索等检索技巧？ |
| 7. 与文献管理人员以及同行的检索相比，我们的检索结果如何？                                |

(引自 Straus SE, 2005)

我们还应继续提问：检索 MEDLINE 数据库时是否使用了医学主题词表 (MeSH)、辞典、限制词以及无文本智能检索 (intelligent free text) 等检索技巧？是否设置了有效的检索过滤？

一个有效评价我们检索技能的方法是：请专业的文献管理员或其他专业人士重复一遍同样的检索过程，然后比较两次检索检索策略和检索结果的可用性。这样做有三方面的益处：①自我评价本身；②有机会学到更好的检索技巧；③能够获得更多的可以回答临床问题的外部证据。

如果我们仍然在检索效率和效力方面有困难，还可以与附近的医学图书馆取得联系，参加他们举办的课程或是索取有关的指南从中学习，以达到实践循证医学所需要的文献检索能力。甚至还可以邀请某位文献管理员加入我们的工作组，以提高检索水平。

实践证明：与循证医学有关的专门的文献检索培训对于提高临床实践者的文献检索能力非常关键。例如：Rachel Stark 等人曾在住院部轮转的住院医生中做了一次随机对照试验，40名参加了循证医学文献检索培训的住院医生与对照组相比，虽然在检索成功率上没有差别，但干预组在应用 Medline 和 Cochrane Library 等 EBM 资源的满意度方面要高于对照组。

## 三、评价“严格评估证据质量”的能力

表17-3列举的是评价我们在严格评价外部证据的真实性和实用性能力时涉及的问题。首先，是否确实严格评价外部证据了？如果没有，我们能不能确定原因是什么？如何克服这些障碍？需要再次强调的是：若能加入到某个小组（如各种杂志俱乐部）中，不但有助于循证医学实践，还能获得有用的反馈意见。

表17-3 对“严格评价证据质量”的自我评价

- |   |
|---|
| 1. 是否对外部证据进行了严格评估？                      |
| 2. 严格评估指南是否易于使用？                        |
| 3. 能否逐渐做到准确且熟练地使用某些严格评估的指标，如似然比、NNTs 等？ |
| 4. 有没有对严格评估进行总结？                        |

(引自 Straus SE, 2005)

如果我们确实严格评价了外部证据,则我们将更有效、更精确的应用这些证据指导临床实践。当然,在这个过程中还需进一步考虑某些效应指标(如似然比、NNTs等)的运用是否熟练和准确。这些答案可以通过将我们的评估结果与正在评价相同证据的同行的评估结果相比较后获取,或者从二次证据中所摘要的文献中提取原始数据,自行完成指标的计算,然后将计算结果与二次证据的结论进行比较。

#### 四、评价“整合外部证据与病人价值”的能力

循证实践的第四步是将经过严格评价的证据与临床经验,并考虑病人的期望相结合,做出临床决策。在自我评价这个过程中我们需要考虑以下问题(表17-4):首先,我们是否真的是在将严格评价的证据整合到临床实践中去。如果是,我们是否更有效、更精确地依据我们的个体病人调整临床决策?检验我们这方面能力的一个有效的方法是看我们能否通过临床决策以解决临床争议。

表17-4 对“外部证据与临床经验的结合”的自我评价

1 是否将严格评价的证据应用到临床实际中?
2 能否逐渐做到准确且熟练地调整严格评价的指标(例如,验前概率, NNT/f等)以适应具体的病例个案?
3 能不能解释(和解决)整合证据制定决策过程中出现的争议?

(引自 Straus SE, 2005)

## 第二节 循证实践的效果评价

经过上述四个步骤的自我评价之后,临床医生对于自身的能力和缺陷应该有了充分的认识,循证实践能力将被强化。但是另一方面,我们还应进一步评价一项有效的干预措施有没有被用于诊疗符合条件的病人,临床实践质量是否得到了提高?有多少临床实践确实有据可循?

### 一、临床实践质量是否得到了改善

从循证实践的效果角度来看,临床医生实践能力的提高对于改善临床实践质量起关键作用,表17-5列举了2个关于自我评价“改变临床实践行为”能力的问题。当我们发现有新证据表明既往的临床决策需要改变时,能不能克服旧的思维惯性而进行适当的调整?对于我们的循证实践过程,如诊断、治疗、预后等方面有没有进行过分析或审计,以评价效果?

表17-5 对“改变临床实践行为”的自我评价

1. 当有新证据表明临床实践改变时,能否克服障碍进行相应的行为调整?
2. 有没有进行检查,如对诊断、治疗或其他循证医学实践方面的审计?

(引自 Straus SE, 2005)

临床审计工作非常重要,原因是:①审计结果可以告诉我们作为临床医生,我们的表现如何;②审计通常会以多种方式进行,特别是个体反馈,这对改善我们的临床工作会产生极大的帮助,审计工作如果长期、持续、恰当地进行将会促进提供临床实践质量不断提高。审计可以有多种方式,如:随机选择病人,审计小组讨论这些病人最初的诊断、干预措施等的依据,以评价多少比例的病人被“循证”观察和处理了。国外许多医院已建立起完善的专职审计(或质量控制)委员会,具体的审计方法可参考其他专著。

### 二、多少临床实践有证可循

评价循证实践效果的另一个方面是究竟有多少临床实践有证可循? Iain Chalmers 等人于1989

年发表了一项震惊整个医学界的研究实例,即在产科使用的226种方法中,临床试验或系统综述证明:20%有效或疗效大于副作用;30%是有害或疗效可疑;50%缺乏随机试验证据。自此之后,整个医学界都在关注现有的临床实践到底在多大程度上是有证可循的。Ellis 等人于1995年4月,对牛津的某临床机构当月诊断的109名患者进行审计,对每个病例的诊断、干预进行讨论,结果显示:对其中90名(82%)患者施加的干预是有证可循的,其中53%的患者所接受的干预措施得到了一项或多项随机化临床试验或系统综述的支持,29%的患者所接受的干预措施有明显可信的非试验性研究的证据支持。另有18%的患者,他们接受的症状治疗或支持治疗虽然没能检索到有力的证据支持,但目前的干预优于其他干预或没有干预。

从此以后,世界范围内的多个临床机构都进行了类似研究,如:普外科、血液科、儿科、基础保健、麻醉科及精神科。例如:Lee JS 等人于2000年对50名胸部手术的病人进行了评价,结果50个病人中7个病人的手术治疗有随机对照试验的证据支持;32个病人有可信的非试验性研究的证据支持;11个病人的治疗没有证据支持。随着循证医学实践和教学的发展,将有越来越多的临床实践有证可循。

(李晚枫)

## 第三篇 循证医学实践的临床范例

### 第十八章 NSAIDs 预防乳腺癌的循证决策

乳腺癌是女性最常见的恶性肿瘤之一。北美和欧洲等发达国家是乳腺癌的高发区,据美国2007年统计,女性乳腺癌占有所有新发癌症病例的26%(178 480例),其死因构成比上,死于乳腺癌的比例高达15%(40 460)。亚洲是乳腺癌的低发区,但乳腺癌的发病率逐年升高,且有年轻化趋向。我国虽是乳腺癌低发地区,但乳腺癌的发病率亦逐年升高,特别是京、津、沪地区乳腺癌已成为危害妇女健康的主要恶性肿瘤。上海市乳腺癌的发病率1972年为20.1/10万,1988年为28/10万,至1993年已高达38/10万,占女性恶性肿瘤发病率的第一位。本章主要探讨如何利用循证医学的方法对药物能否预防乳腺癌等问题进行循证决策。

#### 一、疾病案例

咨询者女性,45岁,有高血压,不吸烟,少量饮酒,未绝经。外婆及母亲均死于乳腺癌,姐姐于47岁时查出乳腺癌并已进行手术治疗。因见文献报道服用阿司匹林能预防乳腺癌的发生,所以咨询服用阿司匹林是否有效?服药的剂量及频率如何?

#### 二、提出问题

这是一起关于非甾体类抗炎药(non-steroidal anti-inflammatory drugs, NSAIDs)的使用与乳腺癌患病风险性的关系的问题。

前列腺素水平的表达在乳腺癌组织中高于正常组织,而环氧化酶(cyclooxygenase, COX)是前列腺素合成的关键酶,NSAIDs的化学预防作用机制就是抑制COX,从而抑制前列腺素的合成。流行病学资料研究表明,服用阿司匹林及其他NSAIDs能够降低大肠癌的患病风险性,病例研究资料表明,其OR(odds ratio, OR)值为0.87(95%CI: 0.77~0.98);在队列研究中,相对危险度(relative risk, RR)值为0.72(95%CI: 0.61~0.85)。在食管癌及胃癌中,同样观察到阿司匹林的保护作用,RR分别为0.51(95%CI: 0.38~0.69)和0.73(0.63~0.84)。然而众多研究资料关于NSAIDs是否能降低乳腺癌患病风险性的结论并不一致。

那么,回答这名咨询者的问题,必须弄清楚以下几个问题:

1. NSAIDs或阿司匹林是否具有预防癌症发生的作用?
2. 如果具有预防作用,那么服药的剂量及频率是多少?
3. 服用NSAIDs是否具有不良反应?
4. 有无新的预防药物出现?

### 三、证据检索与评价

#### (一) 证据检索

1. 可提供的数据库检索资源 采用最简单和经济的方法在网上检索相关信息,提供证据的数据库包括:MEDLINE(1966~2008年)和PubMed(1957~2008年)题录数据库以及Springer(1950~2008年),LWW及Highwire Press等全文数据库。同时利用网络论坛(小木虫、丁香园、卓虎等)及网络引擎(学术谷歌、百度等)检索相关全文参考文献。

2. 关键词及检索策略 乳腺癌的英文关键词有:“breast tumor”、“breast cancer”、“breast carcinoma”和“breast neoplasm\*”。非甾体类抗炎药及阿司匹林的药物关键词有:“NSAIDs”、“anti-inflammatory agents”、“cyclooxygenase inhibitors”、“aspirin”和“ibuprofen”。采用AND对乳腺癌和药物的关键词两两互相组合,同时,通过“\*”来表示neoplasm或neoplasms,以增大检索范围。

3. 纳入及排除标准 检索得到的文献很多,需筛选文献,标准如下:文献是临床的研究,而非动物或细胞系的研究;文献是关于NSAIDs的使用与女性乳腺癌患病风险性的研究;所有的病例诊断均为侵袭性的或原位癌;文献属于队列研究或(巢式)病例对照研究。当遇到不同的研究文献使用同一源人群的数据时,根据以下的标准纳入文献:如果是病例对照与队列研究文献重复,纳入质量较好的队列研究的文献;其他情况,选用文献数据最多,信息最全的纳入研究。所有的文献均是以英文发表的。

4. 检索结果 共检索出45篇符合要求的文献,通过进一步的分析发现,有19篇文献因为以下原因被排除:研究良性乳腺疾病而不是乳腺癌的文献1篇;研究绝经期妇女的生存分析1篇;研究乳腺癌复发的文献1篇;有11篇文献源人群数据重复使用;有5篇文献未提供原始数据,因而无法进行综合处理。最后剩下26篇英文文献可用于综合分析NSAIDs的使用与乳腺癌患病风险性的关系研究。

(1) Rahme E, Ghosn J, Dasgupta K, et al. Association between frequent use of nonsteroidal anti-inflammatory drugs and breast cancer [J]. BMC Cancer, 2005, 5: 159.

(2) Harris RE, Beebe-Donk J, Alshafie GA. Reduction in the risk of human breast cancer by selective cyclooxygenase-2 (COX-2) inhibitors [J]. BMC Cancer, 2006, 6: 27.

(3) Slattery ML, Curtin K, et al. IL6, aspirin, nonsteroidal anti-inflammatory drugs, and breast cancer risk in women living in the southwestern United States [J]. Cancer Epidemiol Biomarkers Prev, 2007, 16(4): 747-755.

(4) Gallicchio L, Visvanathan K, et al. Nonsteroidal anti-inflammatory drugs and the risk of developing breast cancer in a population-based prospective cohort study in Washington County, MD [J]. Int J Cancer, 2007, 121(1): 211-215.

(5) Gill JK, Maskarinec G, Wilkens LR, et al. Nonsteroidal anti-inflammatory drugs and breast cancer risk: the multiethnic cohort [J]. Am J Epidemiol, 2007, 166(10): 1150-1158.

(6) Davis S, Mirick DK. Medication use and the risk of breast cancer [J]. Eur J Epidemiol, 2007, 22(5): 319-325.

(7) Ann Ready, Christine M. Velicer, Anne McTiernan, et al. NSAID use and breast cancer risk in the VITAL cohort. Breast Cancer Res Treat, 2008, 109(3): 533-543.

(8) Friis S, Thomassen L, et al. Nonsteroidal anti-inflammatory drug use and breast cancer risk: a Danish cohort study [J]. Eur J Cancer Prev, 2008, 17(2): 88-96.

(9) Marshall SF, Bernstein L, et al. Nonsteroidal anti-inflammatory drug use and breast cancer risk by stage and hormone receptor status [J]. J Natl Cancer Inst, 2005, 97(11): 805-812.

(10) Swede H, Mirand AL, et al. Association of regular aspirin use and breast cancer risk [J].



Oncology, 2005, 68(1): 40-47.

(11) Harris RE, Namboodiri K, et al. Breast cancer and NSAID use: heterogeneity of effect in a case-control study [J]. Prev Med, 1995, 24(2): 119-120.

(12) Neugut AI, Rosenberg DJ, et al. Association between coronary heart disease and cancers of the breast, prostate, and colon [J]. Cancer Epidemiol Biomarkers Prev, 1998, 7(10): 869-873.

(13) Coogan PF, Rao SR, et al. The relationship of nonsteroidal anti-inflammatory drug use to the risk of breast cancer [J]. Prev Med, 1999, 29(2): 72-76.

(14) Langman MJ, Cheng KK, et al. Effect of anti-inflammatory drugs on overall risk of common cancer: case-control study in general practice research database [J]. BMJ, 2000, 320(7250): 1642-1646.

(15) Harris RE, Namboodiri KK, et al. Nonsteroidal antiinflammatory drugs and breast cancer [J]. Epidemiology, 1996, 7(2): 203-205.

(16) Sharpe CR, Collet JP, et al. Nested case-control study of the effects of non-steroidal anti-inflammatory drugs on breast cancer risk and stage [J]. Br J Cancer, 2000, 83(1): 112-20.

(17) Cotterchio M, Kreiger N, et al. Nonsteroidal anti-inflammatory drug use and breast cancer risk [J]. Cancer Epidemiol Biomarkers Prev, 2001, 10(11): 1213-1217.

(18) Meier CR, Schmitz S, et al. Association Between Acetaminophen or Nonsteroidal Antiinflammatory Drugs and Risk of Developing Ovarian, Breast, or Colon Cancer [J]. Pharmacotherapy, 2002, 22(3): 303-309.

(19) Moorman PG, Grubber JM, et al. Association between non-steroidal anti-inflammatory drugs (NSAIDs) and invasive breast cancer and carcinoma in situ of the breast [J]. Cancer Causes & Control, 2003, 14(10): 915-922.

(20) Terry MB, Gammon MD, et al. Association of frequency and duration of aspirin use and hormone receptor status with breast cancer risk [J]. JAMA, 2004, 291(20): 2433-2440.

(21) Paganini-Hill A, Chao A, et al. Aspirin use and chronic diseases: a cohort study of the elderly [J]. BMJ, 1989, 299(6710): 1247-1250.

(22) Harris RE, Kasbari S, et al. Prospective study of nonsteroidal anti-inflammatory drugs and breast cancer [J]. Oncol Rep, 1999, 6(1): 71-73.

(23) Schreinemachers DM, Everson RB. Aspirin use and lung, colon, and breast cancer incidence in a prospective study [J]. Epidemiology, 1994, 5(2): 138-146.

(24) Johnson TW, Anderson KE, et al. Association of aspirin and nonsteroidal anti-inflammatory drug use with breast cancer. Cancer Epidemiol Biomarkers Prev, 2002, 11(12): 1586-1591.

(25) Harris RE, Chlebowski RT, et al. Breast cancer and nonsteroidal anti-inflammatory drugs: prospective results from the Women's Health Initiative [J]. Cancer Res, 2003, 63(18): 6096-6101.

(26) Jacobs EJ, Thun MJ, et al. Aspirin and other nonsteroidal anti-inflammatory drugs and breast cancer incidence in a large U.S. cohort [J]. Cancer Epidemiol Biomarkers Prev, 2005, 14(1): 261-264.

## (二) 评价证据

评价证据是循证医学实践中最为重要和关键的一步。进行荟萃分析(meta-analysis)最重要的一步是文献是否检索全面。文献检索是否全面的控制,由两名研究者分别进行文献检索,然后对检索到的文献进行核对与查漏。文献数据的提取也应由两名研究者按照统一的数据格式分别进行提取,然后进行核对,出现不一致的地方,由第三人负责核对或者两人共同核实。根据文献的不同类型,按照循证医学的评价原则,对不同的证据进行评价。评价时要充分考虑到证据的真实性、可靠性及适用性,得出准确的结论指导公共卫生决策。

1. 关于 NSAIDs 使用是否降低乳腺癌患病风险性的研究 2001年及2008年都发表了关于 NSAIDs 的使用与乳腺癌患病风险性的荟萃分析,他们纳入的文献年代分别是1980~2000年及2001~2005年,其 RR 值分别是0.82 (0.75~0.89) 及0.75 (0.64~0.88)。然而,自2005年4月以来,又有21篇关于 NSAIDs 的使用与乳腺癌患病风险性的结果不一致的观察性文献发表。再则, Friis 等关于丹麦28 695名妇女的队列研究发现,服用 NSAIDs 能增加乳腺癌患病风险性,其 RR 值为1.27 (95%CI 1.10~1.45)。因此, NSAIDs 的使用是否降低乳腺癌患病风险性依然是个未知数。

该研究荟萃分析了2008年7月以前所发表的所有病例对照及队列研究。此次分析共包括528 705名女性:其中暴露组241 050人,共发生乳腺癌23 217例;对照组287 655人,共发生乳腺癌24 539例。总的 RR 值为0.94 ( $I^2=92%$ , random model, 95%CI: 0.88~1.00,  $P=0.04$ )。由于研究的异质性较高,根据不同的研究类型,不同的研究人群及不同的参考对照(不服用药物与服用药物一组;非规则使用药物与规则使用药物一组)进行分层分析。结果见表18-1。

表18-1 NSAIDs与乳腺癌患病关系的随机效应模型的荟萃分析结果

Types of studies ( number )	$I^2$	RR/OR*	95% CI	P-value
Total ( 26 )	92%	0.94	0.88~1.00	0.04
Case-control ( 16 )	93%	0.88*	0.77~1.01	0.04
Hospital based ( 7 )	95%	1.02*	0.80~1.29	0.86
Population based ( 9 )	88%	0.84*	0.72~0.99	<0.01
Cohort study ( 10 )	91%	0.96	0.81~1.15	0.64
†Reference A ( 14 )	92%	1.02	0.95~1.10	0.59
†Reference B ( 12 )	91%	0.89	0.76~1.04	0.08

\*病例对照研究的指标用 OR; †Reference A: 不用药物 vs 用药; †Reference B: 不规则用药 vs 规则用药。(引自赵亚双,朱穗等研究)

病例对照研究中,综合的 OR 值为0.88 (95%CI: 0.77~1.01),队列研究的 RR 值为0.96 (95%CI: 0.81~1.15)。在以人群为对照的病例对照研究中发现,服用 NSAIDs 降低乳腺癌的患病风险性比以医院为对照的病例对照研究高,其 OR 值分别为0.84 (95%CI: 0.72~0.99)和1.02 (95%CI: 0.80~1.29)。虽然病例对照研究与队列研究的结果并不具有统计学意义,但是病例对照的降低作用比队列研究高,分别降低12%和4%。以人群为对照的病例对照研究也发现了同样的降低作用,然而,以医院为对照的病例对照研究却并没有发现统计学上的降低作用,其原因可能是在以医院为对照的某些与 NSAIDs 药物有关的病人(比如胃癌患者),他们并未连续服药,本应属于暴露组,却错分为非暴露组,导致高估 OR 值,从而对 NSAIDs 的保护作用造成偏移。

在以不服用任何 NSAIDs 为非暴露组及以非规则服用 NSAIDs 为非暴露组中,并没有发现有统计学差异的联系。但观察其结果发现,规则服用药物的保护作用比非规则使用作用强。

2. 阿司匹林和布洛芬与乳腺癌患病风险性的研究 阿司匹林和布洛芬是 NSAIDs 中最常用的种类,在检索的45篇文献中,研究阿司匹林的文献共20篇,研究布洛芬的文献共7篇。荟萃分析结果发现,阿司匹林和布洛芬能降低乳腺癌患病风险性,并且结果具有统计学意义,其 RR 值分别为0.91 ( $I^2=85%$ , 随机效应模型, 95%CI: 0.85~0.98,  $P=0.02$ )和0.81 ( $I^2=77%$ , 随机效应模型, 95%CI: 0.67~0.97,  $P<0.01$ )。

3. NSAIDs 的使用年限及频率与乳腺癌患病风险性的关系 此次研究中,共有14篇文献提供了 NSAIDs 服用的年限,其中有8篇文献可以按照统一标准进行年限划分(<5年和≥5年)。

12篇文献提供了 NSAIDs 使用频率的数据,但是他们都有各自不同的定义,比如:每周1片,每周几次,或者每月几片的标准。在这篇荟萃分析中,我们定义服用 NSAIDs 的频率为每天服用和每周不少于4次,见表18-2。

表18-2 NSAIDs和阿司匹林服用年限及频率与乳腺癌的关系研究

Type of drugs	Duration/frequency ( number )	$I^2$	Model	RR	95% CI	P-value
NSAIDs	<5 year ( 8 )	33%	Fixed	0.98	0.92~1.04	0.44
	$\geq 5$ years ( 8 )	79%	Random	0.94	0.84~1.06	0.32
Aspirin	<5 years ( 4 )	11%	Fixed	0.97	0.88~1.07	0.52
	$\geq 5$ years ( 4 )	85%	Random	1.03	0.82~1.31	0.75
NSAIDs	<daily ( 7 )	93%	Random	1.03	0.86~1.24	0.75
	$\geq$ daily ( 7 )	74%	Random	0.88	0.79~0.99	0.03
	<4 times/week ( 9 )	92%	Random	1.00	0.84~1.19	0.99
	$\geq 4$ times/week ( 9 )	78%	Random	0.84	0.74~0.94	<0.01

(引自赵亚双、朱穗等研究)

关于 NSAIDs 及阿司匹林的服用年限,并没有发现药物的服用时间与乳腺癌患病风险有具统计学意义的联系。然而在药物的服用频率组(每天服用和每周不少于4次),却发现了具有统计学意义的联系,其 RR 值分别为0.88 ( $I^2=74%$ , 随机效应模型, 95%CI: 0.79~0.99,  $P=0.03$ ) 及0.84 ( $I^2=78%$ , 随机效应模型, 95%CI: 0.74~0.94,  $P<0.01$ )。研究发现 NSAIDs 与乳腺癌存在剂量反应关系,但是由于缺乏 NSAIDs 的原始剂量,限制我们关于使用频率与使用年限的剂量反应关系的研究。

4. 此次荟萃分析的局限性 首先,纳入的所有文献都是观察性研究,观察性研究的内部有效性比实验性研究(RCT 实验)低。其次,关于 NSAIDs 的暴露情况的估计,都是参与者自己回忆陈述的,因此,容易导致回忆偏移和错误分类。再则,我们的文献都局限于英文文献,因此,以其他语言发表的文献就被排除了,将会导致发表偏倚。

#### 四、实践决策

##### (一) 根据文献实践决策

荟萃分析已经证明服用阿司匹林及其他 NSAIDs 能够降低大肠癌的患病风险,在食管癌及胃癌患者中,也同样观察到了阿司匹林的预防作用。Mangiapanel 的荟萃分析结果表明阿司匹林能够降低乳腺癌患病风险,但是并没有得出服用的剂量和年限剂量。

决策结论: 服用阿司匹林,无法提供剂量与频率。

##### (二) 循证医学的实践决策

虽然已有循证医学的证据证明阿司匹林能够降低乳腺癌患病风险,但是他们的文献数量少,而且文献收集仅限于2001年到2005年,具有局限性,而此次研究综合了26篇文献,结果更稳定可靠。

此次荟萃分析的结果表明, NSAIDs 的使用能够轻微降低乳腺癌的患病风险。然而此次研究存在发表偏倚,其调整的 RR 值为1.01 (95%CI: 0.96~1.07),并未发现 NSAIDs 与乳腺癌有相关性。同时, Cook 等在女性健康的随机临床对照实验中,服用低剂量阿司匹林(100mg)和维生素 E,随访长达10.1年的研究结果表明,阿司匹林并不能降低乳腺癌的患病风险,RR 值为1.01 (95%CI: 0.94~1.08),这个结果与我们的研究结果相一致。

阿司匹林和布洛芬能降低乳腺癌的患病风险性,结果具有统计学意义,其降低作用分别为9%和19%,每周服药四次以上,预防作用比较明显。但是, NSAIDs 类药物是非选择性的 COX 抑制剂,具有胃肠道和心血管方面的毒副作用性,因此要权衡利弊之间的关系,慎重服药。

研究表明 COX-2基因在乳腺癌中高表达,同时 COX-2抑制剂的胃肠道毒性作用比 NSAIDs 非选择性抑制剂的毒性小。有2篇文献是关于 COX-2抑制剂与乳腺癌关系的研究,从他们的结果可以发现, COX-2抑制剂降低乳腺癌患病风险性强于阿司匹林和布洛芬。COX-2抑制剂降低乳腺癌的患病风险似乎比其他种类的 NSAIDs 高,然而,此次研究并没有足够的去探索这种联系。但提示我们

使用特异的 COX-2 抑制剂不仅能预防癌症的发生,同时还能减少胃肠道的毒副作用。

经过前述的严格证据评价,本研究发现,随着观察性研究的深入,阿司匹林的保护作用逐渐减弱,但并没有发现服用阿司匹林能增加患乳腺癌的风险性。结合本例咨询者自身的特点,建议长期低剂量服用阿司匹林(每天100mg),不仅可以预防脑卒中,还有可能降低乳腺癌的患病风险,同时提醒患者注意体检,以期对乳腺癌做到早发现、早治疗。

## 五、小结

本章通过如何选用药物预防癌症的案例,学习如何将循证医学的知识和方法应用于卫生决策,在这个过程中,首先是要明确我们要解决什么样的问题,即阿司匹林能否预防乳腺癌,这是问题的提出环节。然后,检索证据,这是循证医学评价当中最重要的一步。重点在于如何找到全面的可信度高的证据。PubMed 等题录数据库收录了大量的文献,可重点应用。网络论坛与引擎搜索也是另一个有用的检索全文的方法。接着就进行证据的批判性评估,找到证据后,要对找到的文献进行评分,看文献的研究类型,研究对象的基本特征,随访情况,失访比例,暴露与非暴露组如何划分,结局是否明确,对照的选择标准等,剔除文献质量不高,可信度低的证据。对纳入的文献进行综合分析得出结论;根据文献的研究类型,研究人群特点等进行分层分析,探索文献内部的关联,外推药物的适用范围等。最后,就是把循证医学的证据应用到卫生决策上来,在制定卫生决策时,一定要考虑到药物的副作用及适用范围,权衡利弊。最后还要随访观察,进行后效预测。进而了解通过循证方案制定的政策效果如何,成本效益如何,不良反应有哪些等,进一步对卫生政策进行指导。

(赵亚双 朱 穗)

## 第十九章 急性心肌梗死的循证治疗

急性冠脉综合征(acute coronary syndrome, ACS)分为ST段抬高型心肌梗死和非ST段抬高型心肌梗死、不稳定型心绞痛和猝死。急性ST段抬高型心肌梗死(acute ST elevated myocardial infarction, AMI)是指在冠状动脉病变的基础上,冠状动脉血供突然中断,使相应的心肌出现严重而持久的急性心肌缺血而导致心肌坏死,心电图上表现为弓背向上的ST段抬高。绝大多数AMI是由于患者冠状动脉的不稳定斑块破裂、血栓形成导致梗死相关血管完全堵塞,心肌坏死引起心室重塑、扩大,如合并泵衰竭或心电不稳定可致死亡。在美国,每年大约有90万人发生AMI,其中22.5万人死亡。由于循证医学的发展,为临床实践提供了大量可靠的科学证据,促使AMI的诊断和治疗更加科学化、系统化、规范化,使得AMI的病死率降至5%~9%。早期再灌注,包括溶栓和直接经皮冠状动脉介入术治疗(percutaneous coronary intervention, PCI)可缩短冠状动脉阻塞时间,使心室重塑、扩大和心室功能不全限制到最小程度,减少了泵衰竭和快速恶性室性心律失常的发生率。

本章介绍1例急性心肌梗死患者入院后采取何种治疗方案进行循证实践的过程,从而增加对循证医学实践的感性认识。

### 一、疾病案例

患者男性,30岁,因“胸痛3小时”入院。患者于3小时前,休息时无明显诱因突然出现心前区闷痛,有紧缩感,疼痛难忍,呈持续性,伴有恶心、大汗淋漓及右手麻木、头晕,自服胃药(具体不详)无效,遂来医院急诊。

既往有高脂血症病史4年,高血压病史1年,血压最高为180/90mmHg,未规则服用降压药。无冠心病、糖尿病史。吸烟20余年,2包/天。

PE: T 36.5℃, P 98次/分, R 20次/分, BP 176/116mmHg。神志清楚,急性痛苦面容,自主体位,全身皮肤粘膜未见黄染。瞳孔等大等圆,双眼对光反射存在。咽部无充血,扁桃体无肿大。颈软,颈静脉未见怒张,气管居中,肝颈静脉回流征阴性。胸廓正常,肺部呼吸运动正常,呼吸规则,两肺呼吸音清,未闻及干湿性啰音。心尖搏动位置正常,心浊音界正常,心率98次/分,律齐,未闻及额外心音及杂音,周围血管征阴性,双下肢无水肿。

辅助检查: ECG结果显示(图19-1):电轴左偏, ST II、III、F、V<sub>4-6</sub>弓背向上抬高。CK(多次): 110、1601、1499、1499、1288、869u/L; CK-MB(多次): 9.0、43.0、120、94、73、50u/L。BNP<5.0 pg/ml, cTnI 3.35ng/ml。

初步诊断: 冠状动脉粥样硬化性心脏病,急性下壁、前侧壁心肌梗死,心功能I级。

### 二、提出问题

#### (一) 提出临床需要回答的问题

提出临床需要回答的问题,是循证的第一步,也是关键的一步。构建一个或数个恰当的问题十分重要。首先要有临床意义,是医生或患者所关心的问题,可以是疾病的一般知识问题,也可以是处理病人的专门知识问题;其次,问题不能太大或太小,一定要具体。一个好的问题可以帮助临床医生缩

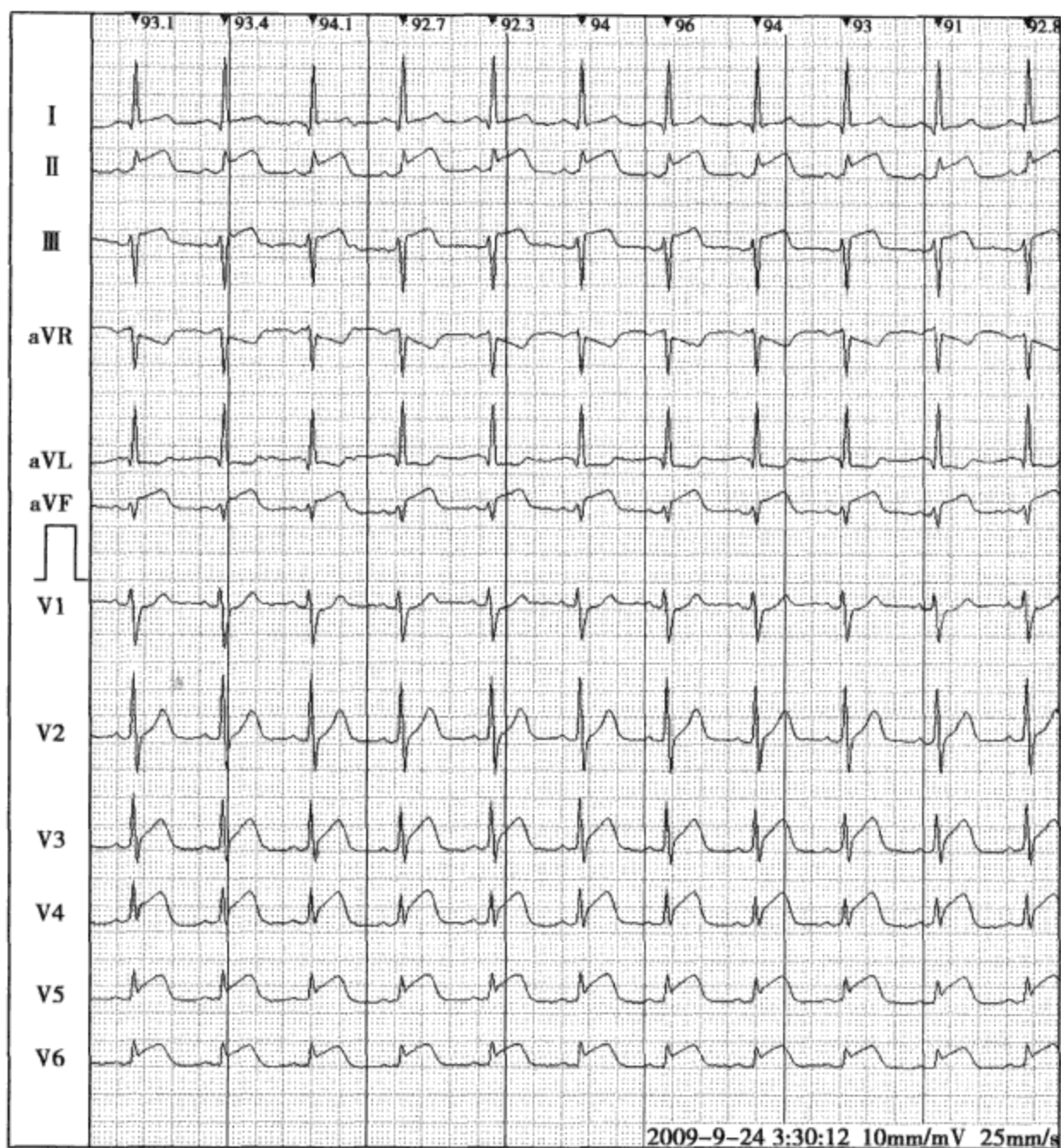


图19-1 心电图显示 ST II、III、F、V<sub>4-6</sub>弓背向上抬高(入院时)

短检索时间,迅速找到恰当的答案,且易于评价和应用。

临床上在构建问题时应根据患者的具体病情及医生和患者主要关注的临床问题,然后结合本单位的资源、条件、可行性,以及临床应用价值等综合考虑,选择范围恰当的关键问题进行处理。根据该患者的病情和临床诊断,结合医生的专业知识和临床技能,提出以下两个问题:

1. 该患者的治疗应该采用溶栓治疗还是直接 PCI 术?
2. 该患者是否需要使用利多卡因来预防室性心律失常?

(二) 将临床问题转换成可以回答的问题

采用 PICO 法,即患者(Patients)、干预(interventions)、对比(comparisons)和结局(outcomes),将临床问题转换成可以回答的问题,便于确定关键词,利于检索。表19-1是 PICO 法构建的问题范例。

表19-1 PICO法构建的问题

患者特征	干预	比较	结局	问题类型	设计类型
急性心肌梗死	溶栓	直接 PCI	再梗死率及病死率	治疗	RCT 或系统评价
急性心肌梗死	利多卡因	安慰剂	病死率	治疗	RCT 或系统评价

三、证据检索

1. 检索资源 带着上述问题,医生可以通过互联网检索各种资料库包括: Cochrane 图书馆、ACP

Journal Club、PubMed、临床实践指南库和中国生物医学文献数据库等。

2. 关键词和检索策略 本病例的关键词: 急性心肌梗死(acute myocardial infarction)、溶栓治疗(thrombolysis therapy)、利多卡因(lidocaine therapy)、PCI术(percutaneous coronary intervention, PCI)。参考要检索的数据库词典,必要时将检索词转换(转换成MeSH Terms),如acute myocardial infarction转换成acute AND myocardial infarction; thrombolysis therapy转换成thrombolytic therapy等。还可采用AND、OR或NOT布尔逻辑符,对词汇进行最佳组合后进行检索。

3. 检索结果 AMI溶栓治疗与PCI治疗对比的相关文献共检索到5篇相关指南、1篇系统评价、2篇meta分析和8篇RCT;与利多卡因预防性治疗AMI室性心律失常的相关文献共检索到相关指南3篇、meta分析3篇和RCT 2篇。

#### 四、证据评价

1. 关于该患者的治疗是采用溶栓治疗还是直接PCI术的问题 AMI的再灌注治疗近10余年来取得显著进展。大量随机对照研究已证实:早期快速、充分持久、有效开通梗死相关动脉(infarct related artery, IRA)恢复前向血流,挽救更多的濒死心肌,保护或恢复受损的心室功能,维持心电稳定,成为改善AMI患者生存率和远期预后的关键。目前再灌注治疗主要有三种方法:直接经皮冠状动脉介入术(PCI)、静脉溶栓治疗和急诊冠脉搭桥术(CABG)。后一种方法在我国开展不多,前两种治疗方法各有其优缺点,何种方法更佳,尚难定论。

(1) 指南意见: ACC/AHA急性心肌梗死治疗指南(2004、2007)、ACC/AHA/SCAI经皮冠状动脉介入治疗指南(2005)、ESC急性心肌梗死治疗指南(2008)、ESC经皮冠状动脉介入治疗指南(2005)、中国经皮冠状动脉介入治疗指南(2009)等,均提出了AMI治疗的再灌注策略,AHA/ACC急性心肌梗死治疗指南提出分两个步骤进行评估:

步骤1: 评估时间和危险性。根据4个方面进行评估: ①症状发作后的时间; ②AMI危险分层; ③溶栓风险; ④转运至熟练PCI导管室所需的时间。

步骤2: 决定首选溶栓还是PCI术。如果心肌梗死少于3小时,且介入治疗无耽搁,溶栓和PCI首选哪种都可以。

如果早期就诊(症状发作在3小时内,行介入治疗有耽搁);不适合选择介入治疗(导管室被占用或不能用;血管入路困难;缺乏熟练PCI操作相关工作人员);介入治疗时间耽搁(运输时间延长;door-to-balloon比door-to-needle时间超过60min;contract-to-balloon或door-to-balloon时间超过90分钟),则首选溶栓治疗。

如果有熟练PCI操作相关人员及有心外科支持(contract-to-balloon或door-to-balloon时间<90分钟;door-to-balloon比door-to-needle时间<60分钟);AMI高危(心源性休克;Killip分级在3级或以上);溶栓禁忌证包括增加出血危险或颅内出血;较晚就诊(症状发作后 $\geq 3$ 小时,<12小时);不能确定AMI诊断,则首选PCI治疗。

(2) 随机对照试验和系统评价: Keeley等对23个随机对照试验进行meta分析,比较了直接PCI术( $n=3\ 872$ )与溶栓( $n=3\ 867$ )治疗AMI的效果。结果显示直接PCI术在降低短期(4~6周)死亡(7% vs 9%,  $P=0.0003$ )、非致死性心肌梗死(3% vs 7%,  $P<0.0001$ )、脑卒中(1% vs 2%,  $P=0.0004$ )、联合终点(死亡,非致死性心肌梗死和脑卒中)(8% vs 14%,  $P<0.0001$ )方面均优于溶栓治疗,独立于不同的溶栓剂和是否转院行PCI术。短期复发性缺血(6% vs 21%,  $P<0.0001$ )、长期复发性缺血(22% vs 39%,  $P<0.0001$ )在PCI组显著性减少。在长期随访(6~18个月)的试验中也显示:直接PCI术治疗AMI仍然优于溶栓治疗,死亡(9.6% vs 12.8%)、非致死性心肌梗死(4.8% vs 10.0%)和联合终点(死亡,非致死性心肌梗死和脑卒中)(12% vs 19%)均显著性减少。结论是直接PCI术治疗AMI优于溶栓治疗。支持直接PCI治疗AMI作为I类推荐(A级证据)的临床研究包括PAMI、GUSTO-Ⅱb、C-PORT、PRAGUE-1、PRAGUE-2和DANAMI-2试验。

Cochrane 图书馆有一个系统评价。该研究是评价直接 PCI 术治疗 AMI 是否优于溶栓治疗。研究终点为总病死率、再梗死率、各种卒中、死亡和再梗死的复合终点、复发性缺血、大出血和冠脉搭桥术。结果显示10个试验共入选2573例患者,与溶栓治疗比较,直接 PCI 显著减少试验终点的短期病死率( $RRR=2\%$ ,  $95\%CI\ 5\%\sim 50\%$ )、再梗死率( $RRR=54\%$ ,  $95\%CI\ 39\%\sim 66\%$ )、复发性缺血( $RRR=32\%$ ,  $95\%CI\ 5\%\sim 50\%$ )、各种卒中( $RRR=66\%$ ,  $95\%CI\ 28\%\sim 84\%$ ),以及死亡和再梗死复合终点( $RRR=46\%$ ,  $95\%CI\ 30\%\sim 58\%$ )。大出血发生率无显著性差别( $RR=1.18$ ,  $95\%CI\ 0.73\sim 1.90$ )。在复合终点的降低方面,直接 PCI 术与强化 t-PA 溶栓治疗比较获益较少( $RR=0.70$ ,  $95\%CI\ 0.51\sim 0.97$ ),而与链激酶溶栓治疗比较获益较多( $RR=0.30$ ,  $95\%CI\ 0.17\sim 0.53$ )。作者的结论是在有直接 PCI 术经验的医学中心,直接 PCI 术是再灌注的优选方案。然而,在大多数情况下,适当的溶栓治疗仍被认为是极好的再灌注方案。

2. 关于预防性使用利多卡因能否降低急性心肌梗死患者病死率的问题 20世纪80年代以前,虽然无随机对照试验证据支持利多卡因能减少室性心律失常引起的死亡,但多数专家将预防性使用利多卡因作为 AMI 的标准治疗方案,直到多项随机对照试验和 meta 分析结果发表,才改变了人们的观念。

(1) 指南意见: ACC/AHA 急性心肌梗死治疗指南、ESC 急性心肌梗死治疗指南和中国急性心肌梗死诊断和治疗指南等都将 AMI 患者预防性使用利多卡因作为 III 类推荐(B 级证据)。

(2) 系统评价和随机对照试验: 1988年, MacMahon 等对14项随机对照试验进行荟萃分析,以评价预防性使用利多卡因对 AMI 患者早期室颤和病死率的影响。6961例入组肌内注射利多卡因试验,2194例入组静脉注射利多卡因试验,分别随访1~4小时和24~48小时。结果显示: 室颤发生103例,死亡137例。利多卡因组减少室颤的风险约1/3( $95\%CI\ 3\%\sim 56\%$ ),而死亡风险却增加1/3( $95\%CI\ 2\%\sim 95\%$ ),但统计学无显著性差异。作者的结论是因为事件发生率较少、随访期限短、某些特殊死亡原因的资料不全,即便是荟萃分析也难以得出预防性使用利多卡因有益或有害的结论,未来需要更大规模或更长随访期的临床试验。随后, Hine LK 等(1989)也对 AMI 患者预防性使用利多卡因对病死率的影响进行了 meta 分析。共纳入14个随机对照试验,其中6个院前试验( $n=7656$ )和8个住院试验( $n=1407$ )。按照治疗方式、报告时间间隔、患者分类评估死亡数据。院前 RCTs 表明病死率无显著性差异( $risk\ difference, RD=0.0184$ ,  $95\%CI\ 0.0481\sim 0.012$ )。住院期间 RCTs 表明使用利多卡因治疗病死率显著性增加( $RD=0.029$ ,  $95\%CI\ 0.004\sim 0.055$ )。结果证实 AMI 住院前监护的患者给予利多卡因并不能减少病死率,住院期间监护的无并发症 AMI 给予利多卡因可增加病死率。在1999年, Sadowski ZP 等系统评估了21个随机对照试验共12 385例患者,其中利多卡因治疗组6223例,非利多卡因治疗组6164例。结果显示,利多卡因治疗组室颤的减少无显著性差异( $OR\ 0.71$ ,  $95\%CI\ 0.47\sim 1.09$ ),而病死率却显著性增加( $OR\ 1.12$ ,  $95\%CI\ 0.91\sim 1.36$ )。结论是 AMI 患者不推荐常规预防性使用利多卡因。

在大型研究 GUSTO-I 和 GUSTO-II b 试验中,共研究了43 704例经溶栓治疗的 AMI 患者,分别有16%和3.5%的患者接受了预防性利多卡因治疗。结果显示: 治疗组24小时病死率较低( $OR\ 0.81$ ,  $95\%CI\ 0.67\sim 0.97$ )、住院病死率( $OR\ 0.90$ ,  $95\%CI\ 0.81\sim 1.01$ )和30天病死率有降低的趋势( $OR\ 0.92$ ,  $95\%CI\ 0.82\sim 1.02$ )。在调整基线特征后,用或不用利多卡因,病死率相似( $OR$  分别为0.90和0.97)。

## 五、实施决策

经过资料检索、整理、分析,结合该患者的具体病情(患者胸痛发作超过3小时),以及医院的技术水平和直接 PCI 术的数量,并考虑患者的经济条件以及家属的意见后,决定采用直接 PCI 术。本例患者无任何室性心律失常,不应采用利多卡因进行预防性治疗。

## 六、后效评估

入院后立即给予阿司匹林300mg 嚼服,氯吡格雷600mg 口服,以及低分子肝素皮下注射,进行抗



凝、抗血小板治疗,拟行急诊冠脉造影术和直接 PCI 术。冠状动脉造影所见:右冠优势型,左主干、前降支、回旋支未见明显狭窄。右冠状动脉血流部分再通,右冠中段见一狭窄病变,狭窄约80%,病变远端见血栓影(图19-2)。术中采用抽吸导管抽出血栓,并行直接支架术,置入药物支架1枚(图19-3)。术后患者胸痛消失,心率75次/分,律齐,未闻及杂音,双肺呼吸音清。术后 ECG 检查结果显示,ST 段回落正常(图19-4)。出院前行心脏超声检查,结果显示室间隔增厚,左室壁运动节段性异常,下壁运动稍减弱,EF 62%,FS 34%,CO 5.0L/m<sup>2</sup>,CI 2.3。患者住院一周出院,出院后继续行双抗血小板(阿司匹林100mg 每天1次和氯吡格雷75mg 每天1次)治疗、降脂和降压治疗,随访病情稳定。

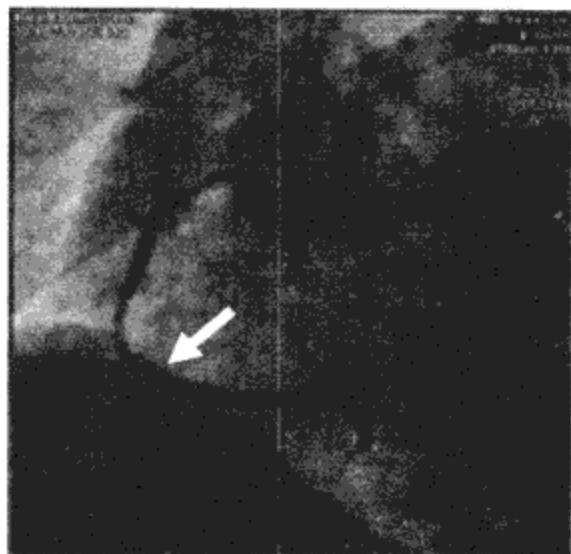


图19-2 冠脉造影显示右冠中段见一80%狭窄病变,病变远端见血栓影(箭头)



图19-3 采用抽吸导管抽出血栓,并行直接支架术,置入支架一枚(箭头)



图19-4 心电图显示抬高的 ST II、III、F 段回落(PCI 术后)

## 七、小结

本章通过1例急性心肌梗死患者的循证治疗实践,学习如何将循证医学的理论和方法用于解决临床的治疗问题。在这个过程中,通过五个步骤来实践循证医学。第一步,采用 PICO 法提出一个具体而又有重要意义的临床问题,这是第一个关键步骤。第二步是寻找证据,设计恰当的检索策略则是这一步成功的关键。学习检索技巧并多实践,才能掌握较好的检索策略,提高文献的查全率和查准率。第三步是评价证据,采用循证医学的评价方法对证据进行严格评价,从证据的真实性、可靠性、临床价值及其适用性作出评价,得出确切的结论并指导临床决策。这也是循证实践中最难的一个步骤。第四步是实施决策,采用当前最佳的证据,并结合医生的临床经验和技能,同时考虑病人的意愿,给予病人最好的临床治疗。最后一步是后效评价,就是了解循证治疗后的疗效,病人的不良反应以及病人和家属对治疗的评价。

(刘金来)

## 第二十章 2型糖尿病合并严重冠心病时 罗格列酮的使用

### 第一节 临床情景

#### 一、病历摘要

患者穆某某,女性,67岁,主因“发现血糖升高10余年,心前区疼痛5年加重0.5年”入院。既往有高血压、脂肪肝、高脂血症病史。入院前10年因外阴瘙痒同时发现血糖升高达15mmol/L,伴消瘦,诊为“2型糖尿病”,曾用“二甲双胍0.5g 每天3次、伏格列波糖0.2mg 每天3次”降糖治疗,因血糖控制不理想,最终改用“罗格列酮4mg 每天1次、诺和龙2mg 每天3次及二甲双胍0.5g 每天3次”治疗。5年前患者因严重心前区疼痛住院,冠状动脉造影显示:3支冠状动脉(左冠状动脉主干、左前降支及右冠状动脉)均明显动脉硬化造成狭窄,由于阻塞严重,已不适宜进行冠脉介入及冠脉搭桥术治疗,无奈之下,应用硝酸酯类等扩冠药物治疗,取得了良好的治疗效果;平时应用上述降糖药物长达5年,空腹血糖控制在5~7mmol/L,餐后血糖控制在7~8mmol/L, HbA<sub>1c</sub> 6.2%;近2周加用他汀类调脂药(阿托伐他汀20mg 每天1次)及阿司匹林100mg 每天1次;无夜间阵发性呼吸困难发生,偶有双下肢水肿发生。查体: P 70次/分, BP 155/80mmHg, BMI 72.72kg/m<sup>2</sup>,腰围 115cm;发育正常,营养中等,肥胖体型。全身皮肤黏膜未见黄染。无颈静脉怒张,双肺呼吸音清,心界向左下扩大,心音低,心律齐,各瓣膜听诊区未闻及杂音;腹软,无压痛,肝脾肋下未及,双下肢无水肿,四肢肌力V级,双侧足背动脉搏动减弱。化验检查: UA 243 μmol/L, TC 6.67mmol/L, TG 2.48mmol/L, LDL 3.87mmol/L, HDL 1.66mmol/L;尿微量白蛋白检测: 15.6 μg/min;眼科检查: 白内障,无眼底出血;心电图: 慢性冠状动脉供血不足;胸部 X 线示主动脉型心脏;双下肢动脉彩色多普勒超声示: 双下肢动脉粥样硬化伴多发斑块,双侧胫后动脉闭塞;腹部 B 超显示: 脂肪肝及胆囊多发性结石;心脏彩色多普勒超声显示: 严重节段性室壁运动异常,左室舒张功能减低;神经电图显示: 右侧正中神经、尺神经损害(感觉纤维受累)。初步诊断: ①2型糖尿病,糖尿病周围血管闭塞,糖尿病周围神经病变;②冠状动脉性心脏病心绞痛;③高血压2级;④高脂血症。

#### 二、查房讨论

甲医生: 本患者临床特点是: ①老年女性伴肥胖;②较长的2型糖尿病史,长期应用罗格列酮4mg 每天1次、诺和龙2mg 每天3次及二甲双胍0.5g 每天3次等药物控制血糖,血糖达标;③并发较严重的糖尿病大血管病变,冠心病尤为严重,冠状动脉造影提示3支冠脉病变,不能进行“介入”及“搭桥”治疗,但万幸的是以硝酸酯类等扩冠药为主的内科药物治疗冠心病效果满意。④本患者合并高血压、高胆固醇血症、高甘油三酯血症。根据患者上述情况,我认为可以继续维持上述治疗。

乙医生: 我有一点不同的意见,本患者虽然血糖控制达标,通过药物治疗冠心病效果满意;但是,最近有研究证据显示: 在应用罗格列酮治疗的2型糖尿病患者中,心肌梗死及因心血管原因死亡的风险存在着统计学意义上的增加;这个患者冠心病这么严重,我认为不应该继续应用罗格列酮,应立即停用该药。

甲医生: 我不同意乙医生的意见,本患者虽然2型糖尿病病史较长以及合并严重的冠心病,血糖控制达标,但那是在“罗格列酮4mg 每天1次、诺和龙2mg 每天3次及二甲双胍0.5g 每天3次”三种药物的作用下才控制达标的,若停用罗格列酮血糖势必再度升高;况且,本患者应用罗格列酮多年,并没

有心脏病加重的迹象;所以,我不同意停用该药。

丙医生:我同意两位医生的分析,正如甲医生所说:“本患者应用罗格列酮多年,并没有心脏病加重的迹象”;但是,也正如乙医生所说:“最近确实有研究证据显示:应用罗格列酮治疗2型糖尿病时心血管原因死亡的风险增加”。情况是这样的,2007年著名的《新英格兰医学杂志》(NEJM)刊出了一篇对42项涉及罗格列酮临床试验的荟萃分析,结果显示,在采用罗格列酮治疗2型糖尿病的15 500患者中,心肌梗死及因心血管原因死亡的风险存在着统计学意义的增加,与对照组相比,罗格列酮组心肌梗死(MI)的风险比值为1.43(95%可信区间:1.03~1.98);心血管性死亡风险比值为1.64(95%可信区间:0.98~2.74)。但是,就在《新英格兰医学杂志》发布研究结果不久,另一著名医学杂志《柳叶刀》发表评论阐述,NEJM公布的有关罗格列酮大幅增加心脏病风险及导致患者死亡率增加的研究结果与之前的几项大规模研究结论(ADOPT、DREAM等)并不一致。建议等待专门评价罗格列酮心血管预后及糖尿病血糖调节的研究,即RECORD研究结果公布以后再下定论。2009年6月,在美国ADA第69届年会上,“RECORD研究”和“2型糖尿病患者冠脉搭桥血运重建(BARI-2D)研究”结果公布,并发表在《柳叶刀》医学杂志上。我们在这样的事实面前需要经历一次循证临床实践的过程,我们期望通过自己的循证实践来回答这样的问题。于是,在丙医生的带领下,他们开始并最终完成了一次完美的循证临床实践过程。

### 三、实践循证医学基本步骤

第一步:把临床实践中需要解决的问题转化为一个可回答的问题。(提出问题)

第二步:查找并获得可以回答这个问题的最佳临床证据。(获取证据)

第三步:评价证据的真实性(研究方法及其研究过程的可靠程度)、重要性(研究结果的临床意义)及临床适用性。(评价证据)

第四步:将评价结果与医生的临床经验、疾病的发病机制、病人的个体情况及价值观结合并综合分析。(应用证据)

第五步:总结上述步骤在实施过程中的经验及教训。(总结经验)

## 第二节 循证临床实践过程

### 一、提出问题

#### (一) 如何构建理想的临床问题

构建出理想的临床问题是循证医学实践的第一步,也是最重要的一步,可能也是整个循证医学实践过程中最难的一步,如何构建出好的临床问题,我们提出如下建议:

1. 临床问题来源于深入的临床实践,应随时发现病人急需解决的问题;同时,要将“病人需要解决的问题”进一步“归类”,使之纳入病因问题、临床表现及鉴别诊断问题、诊断试验问题、预后问题及预防治疗问题的范畴,归类后需再进一步了解问题的实质及内涵。

2. 构建一个好的临床问题必须包含5个必备的组成部分(PICOT原则):①病人和(或)问题;②主要的干预措施;③对比措施;④重要的临床结局;⑤观察时间。

3. 重视临床问题的“筛选”及“保管”。“筛选”即选择重要的问题加以回答;“保管”即要将你认为重要的问题保存在笔记本、纸质卡片、电脑文件夹、录音笔、手机等存储媒介上。

#### (二) 本例临床防治性问题的提出及分析

首先,明确初始临床问题是什么。对于这位老年女性,有高血压、2型糖尿病、冠心病史且体重超重者;并且,冠心病已经失去介入及搭桥机会;在这种情况下,医生建议详细的饮食控制方案,同时医生又建议长期应用罗格列酮4mg 每天1次、诺和龙2mg 每天3次及二甲双胍0.5g 每天3次等药物控制血糖,以硝酸酯类等扩冠药为主的内科药物治疗冠心病;应用上述药物效果很好,血糖控制达标,心绞痛发作明显减少,并且没有充血性心力衰竭发生;高胆固醇血症及高甘油三酯血症应用他汀类药物可以解决。

问题是,有的医生在查房讨论时提出“严重冠心病时最好不使用罗格列酮类药物”这样的问题,但又有医生持反对意见;在“使用”与“不使用”的争论焦点上就形成了一个临床初始问题,要解决好这样的问题,就要运用循证临床实践的技能和方法。下一步,需要将问题转换,转换问题目的是为了便于使得初始临床问题“标准化”,经过“标准化”的临床问题可以很方便的从中获得文献检索关键词,以组成检索策略及获取相关的证据。治疗性研究所涉及的问题一般是由医生根据患者具体的临床治疗情况提出,该病例提出的问题是希望了解“糖尿病合并严重冠心病时能否使用罗格列酮”。

根据PICOT原则将“糖尿病合并严重冠心病时能否使用罗格列酮”转换成“标准化”的临床问题,即“2型糖尿病合并严重冠心病的患者,应用罗格列酮控制血糖是否增加心肌梗死及心血管原因的死亡风险?”

## 二、获取证据

### (一) 防治性问题的最佳证据

本例的临床问题属于临床实践范畴,治疗性研究证据设计方案不尽相同,论证强度分级也不一样。RCT的系统评价或RCT的结果是证明某种疗法的有效性和安全性最可靠的依据(金标准)。在没有这些金标准的情况下,可依次使用其他证据,但其可靠性逐级降低。单个防治性研究的设计按照提供证据的强度水平,依次有RCT、队列研究、病例-对照研究、系列病例观察、专家意见等。目前国际上公认的是源于RCT的系统评价论证强度最高。

### (二) 数据库选择

1. 首先使用已经过整理的医学文献数据库,例如: Cochrane library; Up to date; Clinical evidence; Best evidence ( ACP journal club and evidence based medicine ); Sumsearch; Tripdatabase。

2. 其次使用未经过整理的医学原始文献数据库,例如:PubMed(最好从 Clinical Queries 频道进入)。

### (三) 确定关键词和制定检索策略

检索证据时,可以根据自己的检索能力和时间进行初级检索和高级检索,下面推荐的主题词和副主题词是便于高级检索时使用。

#### 1. 推荐检索治疗性证据的主题词

clinical protocols	patient selection
feasibility studies	random allocation
pilot Projects	sample size
reproducibility of results	comparative study
research design	placebos

double-blind method

#### 2. 推荐检索防治性证据的副主题词

drug therapy	therapeutic use
--------------	-----------------

3. 检索防治性证据的关键词可以根据 PICOT 原则获得

4. 推荐的关于防治证据的检索策略(高级检索)

(1) clinical or control\$ or compare\$ adj ( trial\$ or study or studies ). ti, ab, sh。

(2) ( multicenter\$ or multicenter\$ or multi-centre\$ or multi-center\$ ). ti, ab, sh。

(3) ( blind\$ or mask\$ or placebo\$ ). ti, ab, sh。

(4) 1 or 2 or 3。

### (四) 本病例循证检索过程

1. 我们推荐首先检索已经过整理的医学文献数据库,例如: Cochrane library、Up to date、Clinical evidence、Best evidence ( ACP journal club and evidence based medicine )、Sumsearch、Tripdatabase,或者从 OVID 循证医学数据库中选择相应的内容,等;但是,由于种种原因不是每一位工作在第一线的临床医生都能很方便、很及时的获取信息;所以,对于各级一线的临床医生来讲,恐怕最经济实用、最便捷的证据查寻首选

的仍是: 未经过整理的医学信原始文献数据库, 例如: PubMed(最好从 Clinical Queries 频道进入), 检索好条目后通过一站式服务获取原文, 这也是一个比较有效的方法。

2. 根据 PICOT 原则, 以其所包含的5个基本要素作为关键词: “diabetes mellitus rosiglitazone、coronary heart disease、myocardial infarction、heart failure、mortality risk、treatment”。还可以加入一些防治研究方法学的特征性关键词, 如“RCT、systematic review”等。最后通过 Pubmed 中的 Clinical Queries 频道检索出了如下的证据文献:

3. 从2009年 N Engl J Med.360卷第24期中筛选出与临床问题密切相关的一篇随机对照试验, 题目为“BARI 2D Study Group. A randomized trial of therapies for type 2 diabetes and coronary artery disease.。从2009年 Lancet 373卷第20期中筛选出题目为 Rosiglitazone evaluated for cardiovascular outcomes in oral agent combination therapy for type 2 diabetes ( RECORD ): a multicentre, randomised, open-label trial 的证据。最后, 根据摘要的信息找到全文出处在“2007年 N Engl J Med 上题目为 Effect of rosiglitazone on the risk of myocardial infarction and death from cardiovascular causes.”的文章。

### 三、评价证据

在获得的防治性研究证据中, 可以采用高质量的系统评价或循证防治指南及单个随机对照试验的研究证据。高质量的系统评价是最高级别的循证医学证据; 单个随机对照研究证据较易获得, 如果它是质量佳且样本量较大的随机对照试验, 也有着重要的临床价值。因此, 对它们研究质量真实性和研究结果重要性的评价至关重要, 评价的主要原则如下:

#### (一) 系统评价及荟萃分析证据的严格评价

1. 系统评价的概念 系统评价是针对某一具体问题的临床专题, 通过全面收集全世界所有已发表或未发表的临床研究, 按照减少偏倚和随机误差的原则, 选出合格的文献, 通过严格评价进行定性分析和定量分析, 获得较为客观可靠的结论, 该结论随着新研究结果的出现及时更新。meta 分析 / 荟萃分析是一种客观的、定量的统计分析方法, 用来整理归纳及综合分析原始证据。

当系统评价包含 meta 分析时, meta 分析可以获得准确可靠的证据, 但是, meta 分析有时会得出与大规模临床试验相互矛盾的结果, 《新英格兰医学杂志》1997年的一篇研究文章发现35%的 meta 分析结果与后来的大规模随机对照临床试验结果不完全一致; 例如: 镁制剂治疗心肌梗死, 小样本的 Meta 分析结果有效, 但大规模随机对照试验得出的结论却相反。

2. 系统评价(综述)的评价方法, 详见本教材相关章节。

源于高质量 RCT 所做的系统评价, 其质量和证据强度是可信的, 然而不可盲信, 因此, 评估系统评价 / meta 分析时应特别引以为戒的是不能只检索单一的数据库; 不能选用研究结果差别很大的研究(研究是否同质); 纳入的原始研究样本不能很小。

3. 本案例防治性问题所涉及的系统评价 本患者患有糖尿病, 合并较严重的冠心病, 罗格列酮的安全性问题自然是最优先考虑的问题, 2007年一项荟萃分析报告与其他对照组相比, 罗格列酮治疗组心肌梗死和心血管原因死亡的发生率较高。回顾性荟萃分析了临床试验数据库的42个试验, 报告了与安慰剂或其他药物相比, 罗格列酮显著增加心肌梗死的发生危险有统计学意义, 心肌梗死: odds ratio=1.43, 95% CI 为1.03~1.98,  $P=0.03$ ; 同时也报告了与安慰剂或其他药物相比, 罗格列酮增加心血管原因死亡的危险无统计学意义, 心血管原因死亡: odds ratio=1.64, 95%CI 为0.98~2.74,  $P=0.06$ 。

对于这样的结论每一个医生在使用罗格列酮时都不免有一些担心, 系统评价 / 荟萃分析是循证医学中高级别的证据, 医生肯定会引起足够的重视; 特别是本篇荟萃分析的结论与先前的一些涉及罗格列酮与心血管并发症的大型循证医学研究结果不一致, 同时又因为涉及患者的治疗和处理原则, 关系重大, 我们需慎之又慎, 所以在使用循证医学证据时, 荟萃分析的结果同样要认真的给予严格评价。首先我们看到该荟萃分析确实是选用单个随机对照试验做的系统评价, 这一点在文章中作了比较详细的阐述。对于是否针对临床问题进行了系统全面的文献检索、是否对纳入的原始研究文献进行了严格的

真实性评价以及是否采用原始研究资料进行系统评价分析,文章均未作详细的阐述。另外,本荟萃分析整合的多个研究的数据中,对于每一个单项研究的严格评价结果未作详细阐述;另外,荟萃中涉及多项研究不是专门为研究罗格列酮心血管副作用而设计的,所以入选的病人条件尚需评价。再者,这个荟萃分析的数据综合是“结果综合”,而非“原始数据综合”。另外,不良事件确定性的结论用荟萃分析并不是最严格的方法,荟萃分析应结合针对该疾病患者的大规模长期的临床试验才是最科学和严格的调查药物安全和有效性的方法,我们认为荟萃分析是用来提出假设而不能得出确定的回答。

系统评价/荟萃分析虽然是论证强度很高的循证医学证据,但是,由于原始研究质量不同,而且在撰写荟萃分析时方法学上的差异,其结论可能不尽相同,所以将此类研究证据用于病人身上我们既要非常重视,同时又要非常谨慎。

## (二) 单个随机对照试验证据的严格评价, 参见本教材相关章节。

1. 本案例所涉及单个随机对照试验的评价 本案例的证据涉及两个大型临床试验: RECORD 研究及 BARI-2D 研究,我们分别对其进行评价:

### (1) RECORD 研究的评价:

1) 真实性评价: RECORD 研究是2001年启动的多中心研究,由欧洲、新西兰和澳大利亚共25国家的364个研究中心参加,是一个大型、前瞻性、随机、开放、对照的研究;该研究目的是评估: 两组2型糖尿病患者均应用了二甲双胍或磺脲类药物,在此基础上其中一组加用罗格列酮,另一组不加用罗格列酮,评价上述两组患者的心血管终点事件及血糖的控制情况。研究共入选了4447例2型糖尿病患者,其中一组使用二甲双胍或磺脲类药物加用罗格列酮2220例,另外一组二甲双胍与磺脲类药物联合用药而未加用罗格列酮2227例。入选标准: 2型糖尿病、年龄40~75岁、BMI > 25.0kg/m<sup>2</sup>、用二甲双胍或磺脲类药物(允许使用的磺脲类药物包括格列本脲、格列奇特、格列美脲)单药治疗、HbA<sub>1c</sub> > 7.9%。排除标准: ①前3个月曾因严重心血管事件而住院; ②计划进行心血管介入治疗; ③有心力衰竭史或曾因心衰接受治疗。该研究设计终点目标是比较在二甲双胍或磺脲类药物治疗的基础上加用罗格列酮的患者与二甲双胍和磺脲类药物联合治疗的患者首次发生心血管住院或死亡(包括心脏病发作、充血性心力衰竭和脑卒中)的时间。

真实性评价分析: 本研究对象在研究之初通过中心随机进行了随机分配; 试验对随机方法进行了隐藏,由中心人员设计随机方案,并按此方案将受试者进行分配; RECORD 研究受试者特征: 年龄40~75周岁, BMI > 25.0kg/m<sup>2</sup>,糖化血红蛋白水平介于7.0% 到9.0% 之间,治疗组已经服用最大耐受剂量的二甲双胍或磺脲类药物(格列本脲,格列奇特或格列美脲),对照组依据糖化血红蛋白7.0% 为治疗达标值而调整药物剂量。当糖化血红蛋白 > 8.5% 时,罗格列酮组加用第三种口服药,对照组改用胰岛素治疗。两组患者有相似的人口学特征、种族、糖尿病家族史和糖代谢特征以及相似的人组治疗,故两组基线具有较好的可比性。研究采用了意向性治疗分析原则。只有7% 的患者失访,无法最后确认这些患者是否发生了严重的心血管并发症。

总结: 试验采用了随机的方法对受试者进行分配,采用中心随机方法进行了隐藏,失访率较低,结果中进行了意向性治疗分析,试验组和对照组的初始基线资料具有较好的可比性,5.5年的观察过程中除干预因素外,其他治疗具有可比性,对于设定的目标来讲,观察时间足够长,遗憾的是本研究对受试者、数据收集者未采用盲法。总的来说,该随机对照试验质量好,真实可靠。

2) 重要性评价: ①主要终点结果显示: 罗格列酮治疗组与二甲双胍和磺脲类药物联合应用组相比,心血管病住院治疗和心血管病死亡的总体发生率没有统计学差异,它们的风险比为0.99,95% 可信区间是0.85~1.16,心血管事件包括冠心病发作、脑卒中等,在被统计的所有数据中,发现了罗格列酮使心力衰竭的风险增加了一倍,但罗格列酮不会增加总体心血管发病率或死亡率的风险。②次要终点结果显示: 罗格列酮对心血管死亡复合终点(包括心血管死亡、心肌梗死或脑卒中)的影响较小,罗格列酮小于二甲双胍和磺脲类,罗格列酮组154例事件或6.9%,二甲双胍和磺脲类165例事件或7.4%,其风险比值为0.93,95% 可信区间为0.74~1.15,差异没有统计学意义。心肌梗死方面没有统计学差异(罗格列酮组2220名患者中64例出现心肌梗死(占2.9%),对照组2227名患者中56例出现心肌梗死(占2.5%),风险比

为1.14,95%可信区间为0.80~1.63。脑卒中出现的更少(罗格列酮组发生46例事件,占2.1%,对照组发生63例事件,占2.8%),风险比为0.72,95%可信区间为0.49~1.06。通过分析可以发现,与最常用的糖尿病药物二甲双胍和磺脲类药物相比,罗格列酮不会增加心血管死亡或住院的总体风险。③平均糖化血红蛋白结果显示,RECORD研究再次显示罗格列酮在长期血糖控制方面优于二甲双胍和磺脲类药物。在为期5.5年的研究结束时,罗格列酮组的糖化血红蛋白水平较磺脲类药物组低0.29%,较二甲双胍组低0.26%,差异均具有统计学意义。④总结:本研究最终证实罗格列酮对2型糖尿病患者,在心血管病和心血管死亡的发生中,与其他降糖药物组对照,并无显著性差异。罗格列酮对这两个主要研究终点均无不良影响。但是,所有的数据分析中,使用罗格列酮的不利证据就是心衰的风险增加了一倍,尽管如此,心血管死亡和中风终点事件并没有增加,罗格列酮的治疗仍达到了非劣效性的标准。

## (2) BARI-2D 研究的评价

1) 真实性评价: BARI-2D 研究目的是评价2型糖尿病合并稳定性缺血性心脏病的患者,单纯使用药物治疗或药物治疗同时联合即时血运重建术,比较两组方法的优劣。BARI-2D 研究共纳入了2368名2型糖尿病合并稳定型冠心病的患者,将他们分为两组,一组接受即时血运重建和强化药物治疗。另一组接受单纯的强化药物治疗,然后再进一步将上述两组随机分组接受胰岛素增敏治疗或胰岛素供给治疗,该研究直接比较了胰岛素与胰岛素增敏剂对糖尿病患者心血管死亡或生存率的影响。入选标准:所有入选患者均患糖尿病、年龄>25岁、经冠状动脉造影显示一支或多支血管存在50%狭窄(可以确诊的冠心病)、所有患者均不需要立即行手术控制严重的不稳定型心绞痛(不需要马上进行血管成形术)。研究主要终点:全因死亡。研究次要终点:死亡、心肌梗死、脑卒中。

真实性评价分析: BARI-2D 研究采用针对缺血控制措施及血糖控制措施的2×2析因设计,研究对象在研究之初进行了随机分配。BARI-2D 研究基线资料比较,研究对象分为胰岛素增敏组( $n=1183$ )、胰岛素补给组( $n=1185$ )、手术组(CABG+OMT or PCI+OMT)( $n=1176$ )、口服药治疗组( $n=1192$ )四组,它们在比较了年龄(岁)、男性比例(%),HbA1c(%),糖尿病病程(年)、有MI病史的患者比例(%),有CHF病史的患者比例(%),脑血管意外TIA患者比例(%),外周动脉疾病(%),以前接受过血运重建术(%)等项目,比较无显著差异。随访时间足够,长达5年。5年的治疗过程中,手术治疗及药物治疗组均接受 $\beta$ -受体阻滞剂(手术组84%,药物组88%)、ACEI/ARB(手术组91,药物组92%)、他汀类调脂药(手术组95%,药物组95%)、阿司匹林等药物(手术组94%,药物组94%);5年治疗过程中,降糖药物分配比例合适,增敏剂组80%~91%研究对象使用罗格列酮,胰岛素组90%~92%研究对象使用胰岛素。研究报告没有交代该研究是否进行了意向性治疗分析。

总结: 试验采用了随机方法对受试者进行分配,失访率较低,试验组和对照组的初始基线资料具有较好的可比性,5年的观察过程中除干预因素外,其他治疗具有可比性,对于设定的目标来讲,观察时间够长。遗憾的是研究报告没有交代该研究是否进行了意向性治疗分析。总的来说,该随机对照试验质量好,真实可靠。

2) 重要性评价: 对于2型糖尿病合并稳定性缺血性心脏病的患者,在强化药物治疗基础上进行积极血运重建治疗,它包括经皮冠状动脉介入术(PCI)及冠状动脉旁路移植术(CABG),与强化药物治疗相比发现血运重建治疗在减少死亡或主要心血管事件的终点方面并无优势,立行血运重建术组生存率为88.3%,强化药物治疗组生存率87.8%,95%可信区间为-2.0~3.1,  $P=0.97$ ; 胰岛素增敏组与胰岛素补给组临床预后无差异,使用胰岛素增敏治疗方案与用胰岛素治疗方案相比,增敏剂组生存率是88.2%,胰岛素治疗组生存率是87.9%;同时,增敏剂降低HbA1c作用更强,增敏剂使55.5%的患者HbA1c<7.0%,胰岛素仅使40.0%的患者HbA1c<7.0%。

## 四、应用证据

### (一) 循证医学证据适用性评价的基本内容

我们认为这个环节在循证医学实践中是最重要的环节。当获得的循证治疗证据被确定为真实及



有重要临床价值后,下一步就是要评价其是否能够用于自己的患者,这时就要根据患者的具体临床情况,将所获得的最佳证据与医生的临床经验及患者家属意愿三者有机结合在一起,结合临床患者的具体情况来应用最佳证据。

## (二) 本患者的临床决策

### 1. 判断是否自己的病人与研究证据中病人的情况有差异?

RECORD 研究目的是评估在二甲双胍或磺脲类药物治疗的基础上加用罗格列酮,对比二甲双胍和磺脲类药物联合治疗,评价两组患者的心血管终点事件及血糖的控制情况。BARI-2D 研究目的是评价合并冠心病的2型糖尿病患者使用胰岛素增敏剂治疗与提供胰岛素治疗对死亡率和心血管事件的影响。我们的病人患有2型糖尿病,体型肥胖,并合并严重三支冠状动脉病变,不伴心力衰竭,与上述 RECORD 研究及 BARI-2D 研究中入选的患者情况相似。

### 2. 判断是否研究证据提供的治疗方法适合现实环境?

目前我国市场上有多种罗格列酮,包括进口和国产品种,在我国很多医院均可获得。目前,我们有较好医疗保健体系,病人医疗费用承受能力较强;医生本人所在的医院有条件开展此项治疗;同时,病人随诊方便,实施该治疗措施以后可随时在监察和随访下慎重用药。

### 3. 判断是否证据提供的治疗措施利益大于弊端?

上述 RECORD 研究及 BARI-2D 研究告诉我们罗格列酮控制血糖效果优于经典的口服降糖药及胰岛素;上述两个试验结果提示罗格列酮不引起额外的心血管死亡增加。

### 4. 判断是否了解了患者及家属对你的治疗措施的态度和意愿?

通过与患者及家属沟通,充分讲解应用罗格列酮的必要性,并向家属讲明此药的副作用以及如何防止副作用的发生等问题,家属及患者表示了充分的理解,愿意配合治疗。

### 5. 是否选对了运用证据的时机及有效的切入路径?

我们知道应用罗格列酮时,如果患者年龄大于70岁、糖尿病病程大于10年、曾有心力衰竭病史、目前存在水肿、肝肾功能减退等情况,应慎用此药物,我们的这例患者存在上述情况,使用罗格列酮时应特别小心谨慎,密切观察,同时应选择好用药时机。

### 6. 判断是否具有足够的生物学及病理生理学解释?

罗格列酮的作用机制与特异性激活一种核受体:过氧化物酶体增殖因子激活的  $\gamma$ -型受体(PPAR  $\gamma$ )有关。在人类,PPAR  $\gamma$  受体分布在一些胰岛素作用的关键靶组织,如脂肪组织、骨骼肌和肝脏等。PPAR  $\gamma$  核受体的作用是调节胰岛素反应基因转录,而胰岛素反应基因参与控制葡萄糖产生、转运和利用。另外,PPAR  $\gamma$  反应基因也调节脂肪酸代谢,在脂肪组织上胰岛素介导的葡萄糖转运子 GLUT-4 的表达增加。罗格列酮与其他口服降糖药联合治疗不仅能显著降低血糖、血脂,同时通过调节炎症因子,罗格列酮可以进一步改善胰岛素抵抗,稳定血管内皮功能,防止  $\beta$  细胞功能衰退及血管并发症的发生。

结论:通过上述分析,针对该患者可以慎重的应用罗格列酮,但须密切观察水肿等不良反应的发生。

## 五、经验总结

1. 在血糖控制方面罗格列酮的研究证据表明,患者经过平均治疗5.5年后,血糖控制水平优于二甲双胍和磺脲类药物。说明罗格列酮对于合适的患者仍然是一种重要的糖尿病治疗用药。

2. 罗格列酮是否增加了因心血管事件入院和死亡的风险,RECORD 研究结果和 BARI-2D 研究结果给出了很完整的答案,也就是说2型糖尿病合并冠心病的患者应用罗格列酮并不是禁忌证。

3. 所有的数据中,对罗格列酮唯一的不利证据就是心衰的风险增加了一倍,心衰是一种严重的疾病,临床医生开处方时要早期密切关注,如果出现不良反应如水肿应立即停药。对有心衰病史的糖尿病患者而言,罗格列酮不推荐使用。这一点在《新英格兰医学杂志》的系统评价中已有明确阐述,应引起我们足够地重视。

(邸阜生)

## 第二十一章 慢性乙型肝炎抗病毒的循证治疗

乙型病毒性肝炎(hepatitis B, HB)是由乙型肝炎病毒感染(hepatitis B virus infection, HBV infection)引起的传染性肝病之一。美国肝病研究协会(AASLD)2007年制定的慢性乙型肝炎临床指南中指出:急性HBV暴露后发生慢性乙肝病毒感染(chronic hepatitis B virus infection, chronic HBV infection)的几率,在HBeAg(+)母亲所生新生儿约为90%,在5岁以内的婴幼儿及儿童约为25%~30%,在成人则不足5%。免疫抑制的患者在急性HBV感染后较易发展为慢性HBV感染。目前,全球约有20亿人感染过乙肝病毒,造成了约3.5亿的慢性HBV感染者。我国属HBV感染高流行区,世界范围内的HBV感染者半数出现在中国,一般人群的乙型肝炎表面抗原(hepatitis B surface antigen, HBsAg)检出率约9.09%,接种与未接种乙型肝炎疫苗人群的HBsAg阳性率分别为4.51%和9.51%。慢性乙型肝炎(chronic hepatitis B, CHB)患者约为3000万,慢性乙型肝炎患者可因其肝炎的反复发作,增加其发生肝硬化、肝细胞癌(HCC)的危险性。对一项684例慢性乙型肝炎患者的前瞻性研究表明,慢性乙型肝炎患者发展为肝硬化的估计年发生率为2.1%。在6岁以前受感染的人群中,约25%成年时发展为肝硬化和HCC。

CHB的自发缓解率极低,在欧美等西方国家自发性乙型肝炎e抗原(hepatitis B e antigen, HBeAg)血清转换率为每年8%~15%,而在亚洲国家尚不足1%。慢性乙肝可能发展的严重后果,迫使患者急切地希望获得有效治疗,并激发了医生治疗慢性乙肝的积极性。目前美国肝病研究协会(AASLD)、欧洲肝病学会(EASL)以及亚太肝病学会(ARASLD)的专家学者均认为,控制CHB,阻止或延缓发生肝硬化、HCC以及CHB相关的失代偿性肝病的最有效治疗就是抗病毒治疗。

虽然目前对慢性HBV感染的治疗取得了长足的进展,现行可供选择治疗CHB的抗病毒药物主要有干扰素(interferon, IFN)和核苷(酸)类似物[nucleos(t)ide analogs, NAs]两大类,IFN包括传统 $\alpha$ 干扰素(interferon alpha, IFN $\alpha$ )和聚乙二醇IFN $\alpha$ (PEG-IFN $\alpha$ ),NAs有拉米夫定(LMA)、阿德福韦酯(ADV)和恩替卡韦(ETV)等。前者的优点是疗程相对固定,HBeAg血清学转换率较高,疗效相对持久,耐药变异较少,其缺点是需要注射给药,不良反应较明显,不适于肝功能失代偿者。后者的优点是口服给药,抑制病毒作用强,不良反应少而轻微,可用于肝功能失代偿者,其缺点是疗程相对不固定,HBeAg血清学转换率低,疗效不够持久,长期使用可产生耐药变异,停药后可出现病情恶化等。因此,如何针对不同特征的患者采取不同的治疗方案,尤其是抗病毒药物的选择是临床亟待解决的问题。采用循证医学原则指导慢性肝炎的治疗,对帮助患者正确对待自己的疾病,规范医生合理的用药,具有重要的现实意义和科学价值。

因目前用于慢性乙型病毒性肝炎抗病毒治疗的药物众多,本章仅以 $\alpha$ -干扰素为例介绍慢性乙型病毒性肝炎抗病毒治疗的循证过程。

### 一、疾病案例

患者男性,36岁,工人。3年前体检发现:乙肝五项HBsAg(+),HBeAg(+),抗-HBc(+);HBV-DNA(+);肝功能明显异常。曾接受过中、西医治疗,给予护肝片、葡醛内酯(肝泰乐)、维生素B<sub>1</sub>、维生素C等药物口服,同时给予强力宁每日1次静脉滴注,但病情始终无明显好转。近1个月来症状反而加重,再次就诊。既往史:家庭中父亲、两个姐姐均患有慢性乙肝,其父亲已患肝硬化,呈明显的“家庭聚集现象”。

查体: 面黄、体瘦, 精神萎靡, 巩膜轻度黄染。心、肺检查无异常。腹部平坦, 肝、脾未触及, 肝区叩痛阳性, Morphy 征阴性。脐周围及左半结肠叩诊呈鼓音, 右半结肠肠鸣音活跃。移动性浊音阴性。辅助检查: 乙肝五项HBsAg(+)、HBeAg(+)、抗-HBc(+); HBV-DNA  $4.47 \times 10^8$ ; 肝功能: ALT 170u/L、AST 120u/L; 超声波检查提示慢性肝损害。

临床诊断: 慢性肝炎、乙型(活动性)。

## 二、提出问题

### (一) 提出临床需要回答的问题

根据该病例的疾病史及临床检查结果, 确诊该患者属于慢性乙肝患者。其病情发展的过程为乙肝病毒的活跃复制, 如果治疗不及时, 效果不理想将导致其肝脏反复的炎症坏死, 以及其后的纤维形成和累积。进一步可能会发展到肝硬化, 最终进展为晚期肝衰竭及原发性肝癌的严重结局。

患者及其家属关心的问题是: 患者是否应该开始抗病毒治疗? 如果做抗病毒治疗, 哪种药物效果好些?

慢性乙型肝炎治疗的总体目标为长期最大限度抑制或消除HBV, 减轻肝细胞炎症坏死和肝纤维化, 延缓和阻止疾病进展, 减少和防止肝脏失代偿、肝硬化、HCC 及并发症发生, 从而改善患者的生存质量和延长存活时间。对于该患者而言, 如何终止慢性乙肝的病情进展, 减少其最终严重结局的发生是治疗的关键问题。

结合患者的病情及其期望, 医生在治疗过程中可以提出以下问题:

对这例活动性慢性乙肝病例是否应采取抗病毒药物治疗?  $\alpha$  干扰素的真正疗效如何? 即患者通过接受治疗能获得的最大效益是什么? 治疗措施又会给他带来什么样的不良反应和危害?

### (二) 将临床问题转换成可以回答的问题

根据前面章节提到的PICO法, 将上述临床问题包含的四个要素: 患者(patients)、干预(interventions)、对比(comparisons)和结局(outcomes), 转化为可以回答的问题, 问题的构建见表21-1。

表21-1 PICO法构建的问题

患者特征	干预	比较	结局	问题类型	设计类型
HBeAg(+), 慢性乙型病毒性肝炎患者	$\alpha$ -干扰素	对照措施	血清ALT复常, HBV DNA水平下降, HBeAg转阴或不伴抗-HBe转阳, 以及肝组织学的改善等	治疗	RCT 或系统评价

## 三、证据的检索

### (一) 经验的证据

在抗病毒药物得以广泛应用之前, 临床医生给予慢性乙肝患者抑制肝脏炎症的药物, 以减轻肝脏的炎症, 恢复肝脏功能, 称为保肝护肝治疗, 其观察指标是以临床症状改善和肝功生化指标复常为目的。但这种治疗方法无法解决肝脏炎症坏死过程的反复发作, 特别是在某些诱因的影响下, 极易诱发肝炎病变的再次活跃。肝脏炎症的反复活跃, 造成炎症坏死纤维组织增生, 并不断累积形成肝纤维化及肝硬化。因此, 临床医生不得不给这类患者抗纤维化的药物治疗, 试图阻止或延缓纤维化的进展。但这种基于经验的治疗措施, 既缺乏可靠的检测指标, 又很难对其效果作出评价。在意识到慢性乙肝病毒感染是慢性肝炎病情进展的始动原因后, 开始给予慢性乙肝患者抗病毒治疗。作为这类患者重要的病因治疗, 一旦获得成功, 即直接对病毒复制的抑制, 将会带来其后的一系列相关的结果。临床可观察到患者肝功能逐步改善, 反映其肝脏炎症的缓解和平息, 患者达到相对的病情稳定。

目前, 临床上用于CHB 抗病毒治疗的药物越来越多, 1992年, 最早用于慢性乙型肝炎治疗的药

物—— $\alpha$  干扰素 (interferon alpha, IFN $\alpha$ ) 在美国获准上市, 近年来用于治疗 CHB 的核苷(酸)类似物 (Nas) 发展很快, 品种也越来越多。但临床经验亦证实, 抗病毒药物治疗慢性乙肝的疗效是有限的, 仅对小部分能达到治疗目标, 特别是对众多的慢性乙肝病毒携带者几乎没有作用。究竟哪些类型的慢性乙肝患者更适合抗病毒治疗, 如何选择合适的患者给予抗病毒治疗, 以期从治疗获得最大受益等问题? 都是经验性证据所无法解决的, 而需要完善的临床研究来总结相关的证据。

## (二) 循证医学的证据

1. 检索资源 可采用多种检索手段检索相关生物学方面的资源, 具体检索方法与证据来源详见相关章节。

### 2. 原始研究证据

(1) 中文文献: 以“慢性乙型肝炎”、“ $\alpha$ -干扰素”和“随机对照试验”为关键词, 检索中国生物医学文献数据库, 初步检索2005年至今的中文文献, 检出117篇文献, 初步阅读获得与本例相关的10篇文献用于评价。

(2) 英文文献: 以“慢性乙型肝炎”(chronic hepatitis B)、“抗病毒治疗”(antiviral therapy)和“ $\alpha$ -干扰素”(alpha-interferon)为关键词, 计算机检索PubMed, 利用PubMed的Clinical Queries制定检索策略。

在Clinical Queries检索界面, 键入“chronic hepatitis B”, “alpha-interferon”, 用运算符“and”连接, 然后点“therapy”一项, 再点broad sensitive search选择项, 得到检索策略为:(Chronic Hepatitis B)AND(alpha-interferon)AND(clinical [Title/Abstract]AND trial [Title/Abstract])OR clinical trials [MeSH Terms]OR clinical trial [Publication Type]OR random\* [Title/Abstract]OR random allocation [MeSH Terms]OR therapeutic use [MeSH Subheading]), 检出1381篇文献, 文献量较大, 需要花费大量时间去剔除不适用的文献。此时, 在Clinical Queries检索界面可以取narrow, specific search选择项, 得到检索策略为:(Chronic Hepatitis B)AND(alpha-interferon)AND(randomized controlled trial [Publication Type]OR(randomized [Title/Abstract]AND controlled [Title/Abstract]AND trial [Title/Abstract])), 将检索文献类型限定在“randomized controlled trial”, 特异度提高, 检出235篇文献。利用PubMed的limits功能, 进一步将检索范围限制在: 发表年限: 近5年; 研究对象: 人; 发表语种: 英语, 检索得到参考文献72篇, 数量较少, 准确度提高。初步阅读这72篇文献的标题和摘要, 筛选出8篇文献与我们的临床问题相关, 以进一步详细阅读。

其中一篇于2009年发表的多中心随机对照试验与本例密切相关, 可用来评价:

Zheng Y, Zhao L, Wu T, et al. Efficacy of consensus interferon in treatment of HBeAg-positive chronic hepatitis B: a multicentre, randomized controlled trial. *Virology*, 2009, 6:99-105.

3. 系统评价的证据 从循证医学角度出发, 在决定是否给予慢性乙肝患者抗病毒治疗, 以及应给予何种药物治疗时, 基于随机对照试验基础上的系统评价结果是最有力和最有用的证据。Cochrane图书馆生产的系统评价, 对纳入的RCT有严格的标准, 并经过严格的质量评价, 其得出的结论更可靠。在Cochrane图书馆, 以“chronic hepatitis B”和“antiviral therapy”检索, 未检出与本例相关的系统评价。利用PubMed的Clinical Queries提供的“systematic reviews”检索功能, 以“chronic hepatitis B”和“antiviral therapy”检索, 检出17条记录, 其中1条记录与本例密切相关:

Shamliyan TA, MacDonald R, Shaukat A, et al. Antiviral therapy for adults with chronic hepatitis B: a systematic review for a National Institutes of Health Consensus Development Conference. *Ann Intern Med*, 2009, 150(2):111-124.

4. 临床实践指南证据 临床实践指南 (clinical practice guideline, CPG) 指人们针对特定的临床情况, 制定出一套系统的能帮助临床医生和患者做出恰当处理的指导意见。在指南的指导下结合患者的具体病情做出诊断和治疗的决策, 有助于循证医学的原则在临床医疗实践中得到更好的贯彻和实施, 规范临床医生的医疗行为, 提高医疗服务质量。若是能检索出慢性乙型肝炎抗病毒治疗的指南,

可以节省临床医师查找证据的时间,如指南中标明了证据的级别,还可以省去文献评价的时间,因此,基于循证医学基础上建立起来的临床实践指南是指导临床实践的最佳依据。

(1) 国外的临床实践指南:从美国国立指南数据库(National Guideline Clearinghouse, NCG)检索到35个乙型病毒性肝炎的临床实践指南,其中一个是关于慢性乙型病毒性肝炎治疗方面的指南:

Adefovir dipivoxil and peginterferon alpha-2a for the treatment of chronic hepatitis B. National Institute for Health and Clinical Excellence (NICE) - National Government Agency [Non-U. S.]. 2006 Feb. 37 pages. NGC:004902

指南给出的有关药物治疗的意见如下:① Peginterferon alpha-2a(聚乙二醇化干扰素 $\alpha$ -2a)是治疗成人慢性乙型肝炎(HBeAg阳性或阴性)的首选。②阿德福韦酯可用于治疗成人慢性乙型肝炎(HBeAg阳性或阴性),如果:用IFN $\alpha$ 和PEG-IFN $\alpha$  2a治疗无效;初次治疗后复发;用IFN $\alpha$ 和PEG-IFN $\alpha$  2a治疗耐药或有禁忌证。③阿德福韦酯不应在拉米夫定治疗前使用,可以单用或与拉米夫定联合使用,当用拉米夫定治疗导致病毒耐受,或拉米夫定耐受很快会发生(例如,病毒复制迅速),而且拉米夫定耐药可能会导致不良反应发生(如代偿性肝病)。

另外,美国肝病研究协会(AASLD)2007年制定了新版慢性乙型肝炎临床指南,参考全球最新循证医学证据,以帮助临床医师和其他卫生工作者对慢性乙型肝炎病毒(HBV)感染预防、诊断和处理。主要内容包括高危人群中HBV感染者筛选、慢性乙型肝炎的健康教育和预防、慢性HBV感染的自然史及专业术语、慢性HBV感染者及患者的评估和处理、慢性乙型肝炎的抗病毒治疗等5个方面,并包括推荐处理意见41条。

其中关于成人HBeAg阳性慢性乙型肝炎患者抗病毒治疗方面的推荐意见如下(括号内罗马字母为证据等级):

1) ALT升高 $>2 \times$  ULN(正常上限值)或肝活检显示中/重度肝炎,伴有HBV DNA $>20\,000$  IU/ml的患者应当考虑抗病毒治疗(I)。①代偿性肝病患者在实施治疗前应当先观察3~6个月,以判断是否发生自发性HBeAg血清转换(II-2)。②若患者出现黄疸及ALT明显升高(肝炎发作),应当立即实施治疗(III)。③可选用已获批准的六种抗病毒药物中的任何一种药物治疗,包括普通干扰素 $\alpha$ 、聚乙二醇化干扰素 $\alpha$ 、拉米夫定、阿德福韦酯、恩替卡韦或替比夫定,但优先选用聚乙二醇化干扰素 $\alpha$ 、阿德福韦酯或恩替卡韦(I)。

2) ALT持续正常或轻微升高( $<2 \times$  ULN)的患者,一般不应当实施抗病毒治疗(I)。①对于一过性或轻微ALT升高,特别是40岁以上患者,应考虑肝活检(II-3)。②若肝活检证实有中或重度炎症坏死,或显著肝纤维化,可以实施抗病毒治疗(I)。

(2) 我国的临床实践指南:2005年12月10日,中华医学会肝病学会分会和中华医学会感染病学分会组织国内有关专家,在参考国内外最新研究成果的基础上,按照循证医学的原则,制订了我国的《慢性乙型肝炎防治指南》,其中推荐意见所依据的证据共分为3个级别5个等次,文中以括号内斜体罗马数字表示。本章只摘录与本例有关的抗病毒治疗参考意见部分。

指南中HBeAg阳性慢性乙型肝炎患者抗病毒治疗的推荐意见:

对于HBV DNA定量 $\geq 1 \times 10^5$ 拷贝/ml, ALT水平 $\geq 2 \times$  ULN者,或ALT $<2 \times$  ULN,但肝组织学显示Knodell HAI $\geq 4$ ,或 $\geq G2$ 炎症坏死者,应进行抗病毒治疗。可根据具体情况和患者意愿,选用IFN $\alpha$ (ALT水平应 $<10 \times$  ULN)或核苷(酸)类似物治疗。对HBV DNA阳性但低于 $1 \times 10^5$ 拷贝/ml者,经监测病情3个月,HBV DNA仍未转阴,且ALT异常,则应抗病毒治疗(III)。

1) 普通IFN $\alpha$  5Mu:可根据患者的耐受情况适当调整剂量,每周3次或隔日1次,皮下或肌内注射,一般疗程为6个月(I)。如有应答,为提高疗效亦可延长疗程至1年或更长时间(II)。应注意剂量及疗程的个体化。如治疗6个月无应答者,可改用其他抗病毒药物。

2) PegIFN $\alpha$ -2a: 180mg,每周1次,皮下注射,疗程1年(I)。剂量应根据患者耐受性等因素决定。

3) 拉米夫定: 100mg,每日1次口服。治疗1年时,如HBV DNA检测不到(PCR法)或低于检测下

限, ALT 复常, HBeAg 转阴但未出现抗-HBe 者, 建议继续用药, 直至HBeAg 血清学转换, 经监测2次(每次至少间隔6个月), 仍保持不变者可以停药(II), 但停药后需密切监测肝脏生化学和病毒学指标。

4) 阿德福韦酯: 10mg, 每日1次口服。疗程可参照拉米夫定(II)。

5) 恩替卡韦: 0.5mg(对拉米夫定耐药患者为1mg), 每日1次口服。疗程可参照拉米夫定。

临床医生应根据自己的专业知识和临床经验, 在综合考虑患者具体病情及个人意愿的基础上, 在指南的原则框架下确定个体化治疗方案(图21-1)。

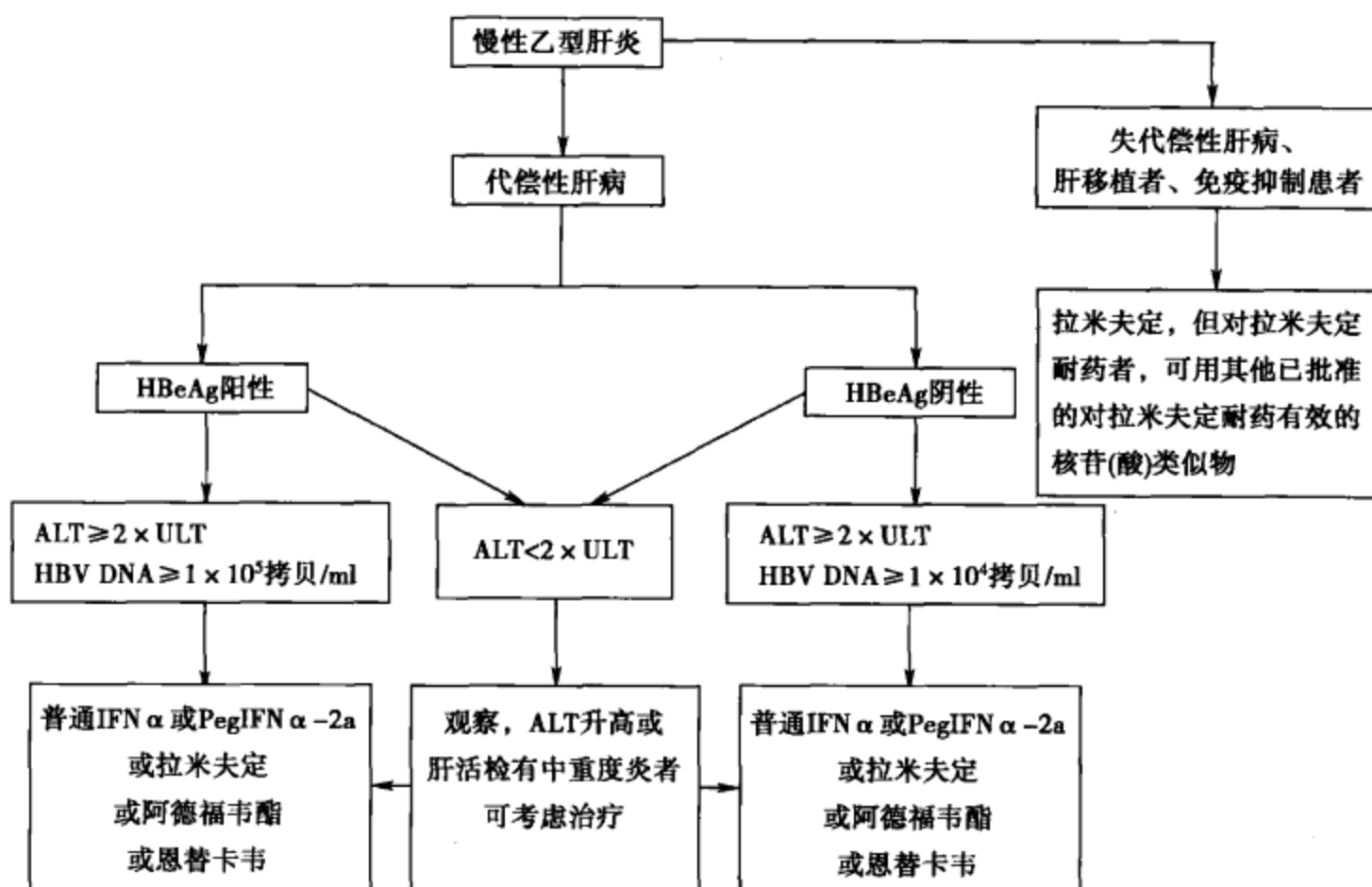


图21-1 慢性乙型肝炎抗病毒治疗流程图

(引自中国《慢性乙型肝炎防治指南》, 2005)

#### 四、证据评价

应用循证医学评价防治性研究文献的质量标准, 对单个研究文献的真实性、临床意义及实用性进行评价, 主要评价要点如下:

##### 1. 原始研究证据

(1) 中文文献: 仔细阅读检索的10篇文献, 纳入的样本量有差异, 由48~230例不等, 样本量大多较小。文献中都提到随机, 但都未交代随机的方法。没有一篇文献提到盲法设计。对于脱落的研究对象都未提及, 文章中也未用到ITT 等分析处理。研究真实性较差, 因此, 文献报道的结果不予评价与应用。

(2) 英文文献: 郑永利、赵连三等人做的多中心随机对照试验将144例HBeAg 阳性慢性乙型肝炎患者随机分为两组, 一组给予9 μg CIFN, 一组给予50 μg INF-α 1b, 每周3次, 连续给药24周, 治疗结束后随访24周, 比较两组患者的血清ALT 复常率、HBV DNA 阴转率和HBeAg 阴转率等。结果显示: 24周治疗结束时, 两组的ALT 复常率分别为52.8% 和52.9%, HBV DNA 阴转率分别为23.6% 和20.0%, HBeAg 阴转率在两组内只达到了15.3% 和14.3%。

该研究采用多中心随机对照试验的设计方案, 文章中给出了病例的诊断、纳入和排除标准, 并交代了随机分配方法及随机分配方案的隐匿。但结果中提到5例脱落病例(CIFN 组1例, INF-α 1b 组4例), 却未说明脱落病例的结局, 统计分析也未采取特殊的分析方法处理。另外, 文章中还涉及药物不良反应的结果。

总而言之,单个研究证据显示,无论是与安慰剂比较还是与其他抗病毒药物比较,如果以近期病毒复制指标阴转为终点评价指标,即HBeAg和(或)HBVDNA转阴,不同的试验其疗效差异很大,如接受 $\alpha$ 干扰素3Mu,每周3次,皮下注射,治疗疗程不少于4个月的慢性乙肝患者,其HBeAg阴转率大多报道在20%~40%之间。

文献证据还显示药物疗效的差异随报道的地区、种族、纳入患者的特点、患者感染HBV的年龄,特别是患者的e抗原状态、血中HBVDNA水平、治疗前丙氨酸氨基转移酶(ALT)水平等,而有很大不同。且不同种类药物产生的副作用发生的可能性也不同。如 $\alpha$ 干扰素治疗患者会出现发烧、全身关节肌肉疼痛、流感样症候群、骨髓抑制、精神抑郁及脱发等副作用。因而来自文献的证据仍不足以帮助医生和患者作出接受何种药物治疗的决策。

2. 系统评价的证据 该系统评价的目的在于评价成人慢性乙型肝炎抗病毒治疗的疗效,文献来源于1990年至2008年发表的干扰素(IFN $\alpha$ 和PEG-IFN $\alpha$ )、拉米夫定、阿德福韦酯、恩替卡韦和替比夫定治疗慢性乙型肝炎的随机对照试验。有严格的文献纳入、排除标准及质量评价标准,最终纳入了93篇文献代表了60个独立的随机对照试验,样本量为20~1367名患者,大部分受试对象来自亚洲(64%),白种人占30%。

系统评价结果显示:只有少量文献报道了远期疗效。如死亡,只有13篇低质量文献提示这些药物不会降低死亡率;只有2篇小样本短期疗效试验提示IFN $\alpha$ 可降低肝硬化的危险性,但没有统计学意义;其他远期疗效指标如肝衰竭、肝细胞癌等,都没有足够的证据表明这些抗病毒药物能降低风险。

对于短期疗效,60个随机对照试验结果提示没有一个治疗方案能够改善所有的近期疗效指标。其中有关 $\alpha$ 干扰素治疗慢性乙型病毒性肝炎疗效的结果为:只有低质量的证据提示IFN $\alpha$ 能清除HBsAg(2RCTs,211名患者);IFN $\alpha$ 使HBeAg转阴(3RCTs,351名患者),血清转化(2RCTs,304名患者)和ALT正常化(2RCTs,131名患者)结果还提示:约50%的患者用抗病毒治疗会出现副反应。但是,大部分研究排除了肝、肾功能不全和合并其他并发症的患者。

该系统评价检索了Medline、PubMed、the Cochrane Library和其他数据库,尽管语言限定为英语,但研究者同时也检索了Medline中其他语言发表的随机对照试验,检出10篇中文文献摘要用于系统评价,目的在于尽量控制语言偏倚。而且该系统评价大部分受试对象来自亚洲,因此评价结果对于我们的临床实践应该有一定的指导意义,但由于研究质量较低,且研究样本、干预措施、结局测量指标判定标准之间有显著的异质性,因此该系统评价未作定量分析,即meta分析。结论:无充分证据表明抗病毒治疗有远期疗效,而且由于一些不一致的结论,使我们也无法确定近期疗效指标是否为评价疗效的可靠指标。因此这是一个虽然较为真实,但临床价值相对较小的证据,依然无法解决本例的临床问题。

3. 临床实践指南证据 由于目前不同国家、不同的学术组织针对同一种疾病制定了不同的临床实践指南,指南质量也参差不齐。因此,对于临床实践指南也应进行评价,以判断指南是否值得推荐使用。具体的临床实践指南的评价标准详见相关章节。

综合评价国际、国内的慢性乙型肝炎防治指南,虽然都给出了抗病毒治疗的推荐,但是只有美国肝病研究协会2007版慢性乙型肝炎临床指南推荐意见是基于I级证据基础之上,美国国立指南数据库未列出证据的级别,我国指南中抗病毒治疗的推荐意见是建立在III级证据的基础之上。实质上,正是由于原始研究证据质量低下,才导致临床实践指南无法依据高质量的证据制定更可靠的推荐意见。

## 五、实施决策

在使用证据为患者作出临床决策时,循证医学强调的是将最佳证据、临床经验及患者的价值观综合考虑。因此,在决定是否为本例患者进行抗病毒治疗之前,首先应考虑以下四个方面的问题:

### 1. 我们所面对的患者是否与研究证据纳入的患者相似?

本例患者为成年HBeAg阳性慢性乙型肝炎患者,前述系统评价纳入的患者64%来自亚洲,且都为成年人;我国《慢性乙型肝炎防治指南》针对成年HBeAg阳性慢性乙型肝炎患者给出了推荐治疗意见,

因此, 这些证据可用于指导临床实践。

#### 2. 采用治疗性证据的可行性如何?

干扰素类药物虽然疗程相对固定, HBeAg 血清学转换率较高, 疗效相对持久, 耐药变异较少, 但需要注射给药, 不良反应较明显, 不适于肝功能失代偿者。

关于经济可行性方面, 查阅相关文献显示抗病毒治疗比一般的综合保肝治疗更有优化的成本效果比, 其中以拉米夫定组的成本效果比最优, 干扰素组次之。

#### 3. 治疗措施对患者是否利大于弊?

虽然经过循证医学评价后, 从所获证据显示, 抗病毒治疗对于慢性乙肝患者有一定的短期疗效。但正如前述证据评价过程中显示: 大部分原始研究证据真实性不高, 系统评价证据虽然较为真实, 但临床价值相对较小。而且前述系统评价证据显示: 约50% 的病人用抗病毒治疗会出现副反应。所以对于是否应采用抗病毒治疗, 还应考虑患者的意愿以及治疗可能给患者带来的不利影响, 以综合分析评价。可以采用防治措施导致患者受益和有害可能性比值比(likelihood of help vs harm, LHH)的方法来帮助决策。

#### 4. 考虑患者的价值观及对疗效的期望

医生可将以上所查到的证据及干预措施的利弊分析告诉患者及其亲属, 结合患者及亲属的意见综合考虑干预措施。

本例患者为男性成年HBeAg 阳性慢性乙肝患者, HBV DNA 定量 $\geq 1 \times 10^5$ 拷贝/ml, ALT 水平远远 $\geq 2 \times \text{ULN}$ , 且患者经过多年保肝治疗效果不理想, 因此患者及家属强烈要求抗病毒治疗, 依据我国的《慢性乙型肝炎防治指南》, 首选普通IFN $\alpha$  5Mu, 隔日1次, 皮下注射, 注意观察病情变化, 适当调整药物种类和剂量, 以求达到抗病毒治疗的最佳效果。

### 六、后效预测

由于慢性乙型肝炎发病机制复杂, 自然进程多变, 个体差异很大, 治疗影响因素众多, 治疗结果难于预测, 确切效果需要进行长期的随访观察。

该病例经给予IFN $\alpha$  5Mu 隔日注射后, 于3个月疗程时肝功能完全恢复正常, 同时HBV DNA 水平降至 $1.51 \times 10^4$  (PCR法), 但未现HBeAg 阴转。后效预测基于疗程结束后如实现HBeAg 阴转, 则可能获得较稳定的持续病毒复制停止和肝组织学的病变改善, 对远期不良结局的发生亦有良好的影响。如未能实现HBeAg 阴转, 则病毒复制可能重新出现, 肝脏炎症活跃可能复发。因此IFN $\alpha$  抗病毒治疗趋于失败, 此时可考虑调整治疗方案, 给予拉米夫定或阿德福韦酯治疗。

### 七、小结

慢性乙型病毒性肝炎是我国常见的慢性传染病之一, 严重危害人民健康。由于临床可供选择的抗病毒药物种类较多, 临床医生在选用抗病毒药物时有一定难度。本章通过介绍一个成年HBeAg 阳性慢性乙肝案例的循证治疗, 利用循证实践的5个步骤去解决临床常见的治疗问题。从第一步提出临床问题到最终的后效评价, 应用循证医学评价抗病毒药物治疗慢性乙型肝炎的证据, 能较为真实地提供医生和患者抗病毒治疗的确切疗效和可能的副作用, 为临床决策提供依据。

通过本例分析发现, 慢性病毒性肝炎抗病毒治疗目前所面临的关键问题在于原始研究文献质量较差, 疗效评价试验中采用中间疗效指标的文献较多, 而采用远期疗效评价指标的文献很少, 这也为制定临床实践指南和临床医生选择治疗方案增加了难度。因此, 慢性病毒性肝炎抗病毒治疗究竟是否利大于弊尚需进一步研究, 一方面应提高临床研究的方法学质量, 使研究的真实性更强, 另一方面还应逐步积累更多的远期疗效追踪观察, 方能最后对慢性乙肝抗病毒治疗作出最终评价。

(李晚枫)



## 第二十三章 结核化疗中预防应用保肝药的循证实践

我国是结核病高负担国家,WHO在2009年结核控制报告中表明,我国每年新发现的初治及复发结核病患者即达98万左右,加上原患结核尚未治愈的患者,每年接受抗结核治疗的病人超过100万。最常用的治疗方案是由WHO推荐的直接面视下短程化疗方案,要求患者接受利福平、异烟肼、吡嗪酰胺、乙胺丁醇、链霉素等药物的联合使用6~9个月。这些药物在杀灭结核菌的同时,也可以导致不同频率、不同严重程度不良反应,其中影响最大且发病率较高的为肝损害,据估计,其发生率约为11.9%。如果病人在接受抗结核治疗过程中出现了肝损害,则不仅可引起严重不良后果如肝功衰竭甚至死亡等,还可能被迫暂停或结束抗结核治疗,从而导致病人的结核病治疗失败、迁延不愈等。为了避免这种情况出现,目前部分医生在临床诊治结核病时,常在抗结核治疗开始时,即给予保护肝脏的药物,以期达到预防肝损害发生的目的。但由于保肝药物的研发初衷多为治疗已存在的各种原因造成的肝损害,对无肝损害的人群预防性使用是否确实有效,是否安全目前尚存在一定争议。本章将探讨面对一个即将开始抗结核治疗的病人,如何应用循证医学的思想和方法来确定是否应给予其预防性保肝药物。

### 一、疾病案例

患者男性,60岁,汉族,因咳嗽咳痰1个月至综合医院就诊,后被转诊至当地结核病预防控制中心,X线片显示右肺上中叶阴影,左肺无,痰涂片阳性,诊断为Ⅲ型肺结核。由于病人为初次治疗结核病,因此计划给予其6个月的初治病人标准化疗方案,前2个月为强化期,需隔日一次服用利福平、异烟肼、吡嗪酰胺、乙胺丁醇分别为0.6、0.6、2.0、1.25g,共计30次,后4个月为巩固期,需隔日一次服用利福平、异烟肼各0.6g,共计60次。

患者无高血压、糖尿病等慢性疾病,目前未服用其他药物,无肝炎病史和酗酒史,无肝肾疾病家族史,开始抗结核治疗前检查肝肾功能主要结果为:谷丙转氨酶(ALT)18mmol/L,谷草转氨酶(AST)12mmol/L,总胆红素(TB)5.2mmol/L,尿素氮(BUN)2.5mmol/L,乙肝表面抗原检查阴性。

### 二、提出问题

面对上述病人,临床医生的问题是该不该给他处方预防性保肝药物,这是一个简易的表达,而当需要把它转化成具体的临床问题时,我们就会发现,这一问题包括了几个方面:

1. 预防性保肝药物具体有哪些药?
2. 这些药物的预防效果如何?
3. 这些药物的安全性如何?
4. 哪种或哪类药物的疗效最佳? 安全性最好?

那么,是否要将每个方面均视作一个临床问题呢?答案是否定的。一个好的临床问题不应该过于宽泛或狭窄,应当恰如其分的反映临床疑问,并有利于确定关键词,以指导循证实践的后续步骤。在这个例子当中,由于上述四方面均是需要解决的疑问,而其中疗效又是最关心的问题,所以将问

表22-1 PICO格式的临床问题

患者特征 (P)	干预 (I)	比较 (C)	结局 (O)	问题类型	设计类型
接受抗结核治疗者	预防性使用保肝药物	1. 干预组vs 安慰剂或空白对照组 2. 干预组vs 其他药物干预组	肝损害发生率	治疗	随机或非随机的对照试验及系统评价

题综合起来,可以列成PICO格式(表22-1),最终表述为“对接受抗结核治疗的病人,预防性使用保肝药物能降低肝损害的发生率吗?哪种保肝药物降低肝损害发生率的效果最好?”

### 三、证据检索和评价

#### (一) 二次研究证据检索

二次研究证据包括系统评价、指南等,常用的资源如Cochrane图书馆、《Clinical Evidence》杂志、美国国立指南数据库等均具有高度的可信性和实用性。我们确定临床问题后,首先搜索二次研究证据,可节约大量的时间精力,得到较高质量证据。

当使用“tuberculosis+hepa\*+prevent\*”为主要关键词检索了上述三个资源和PubMed中的系统评价,其中《Clinical Evidence》和美国国立指南数据库中均未涉及抗结核治疗中预防性应用保肝药问题,而Cochrane图书馆和PubMed中则检到了同一篇系统评价,具体如下:

- Liu Q, Garner P, Wang Y, et al. Drugs and herbs given to prevent hepatotoxicity of tuberculosis therapy: systematic review of ingredients and evaluation studies. [J]. BMC Public Health, 2008, 8:365-373.

这篇系统评价检索了MEDLINE、EMBASE、LILACS、CINAHL、Cochrane 临床试验登记库等外文资源和CNKI、维普网络、万方网等中文资源,纳入标准为2007年4月份前发表的所有在接受抗结核治疗人群中为预防肝损害而使用药物的试验类文章,共计查到85项研究,其中中国研究77项、印度2项、俄国4项、乌克兰2项。这些研究评价了30种不同的保肝制剂或草药,但均为小样本、设计不良、使用中间指标结局的研究,且安全性报告少。系统评价的结论认为目前尚无可信的证据能证明保肝药物的预防效果。

通过阅读这篇系统评价,可以发现预防性应用保肝药物存在药物种类繁多、效果及安全性不明的情况,仍然不能解决前面提出的临床问题,且二次证据均存在发表时证据时间滞后的现象,因此,下一步计划进行原始研究检索。

#### (二) 原始研究证据检索

1. 检索资源 鉴于大部分研究均在中国进行,并用中文发表,我们将检索资源定为中国期刊网全文数据库(CNKI)医药卫生部分和中华医学会期刊数据库,查找1998年~2007年以中文发表的关于抗结核治疗中预防性应用保肝药物疗效及安全性的临床试验研究,检索时间为2008年6月。

##### 2. 关键词和检索策略

第一步,对不同检索关键词进行组合,以寻找敏感性和特异性均较高的检索策略。测试关键词包括四个方面的词汇,一类是结核相关词汇,如“结核”、“TB”、“癆”,一类是保肝药相关词汇,如“保肝”、“护肝”,一类是肝损害相关词汇,如“肝损伤”、“肝损害”、“药物性肝炎”、“肝功能异常”,一类是预防相关词汇,如“预防”、“防治”。经过不同组合后发现,由于不同杂志和不同作者的关键词用词不一,很多文章均缺乏上述四类词汇的某一类或某几类,如全部包括则虽特异性高,但将丢失较多信息。因此,决定采用较宽泛的词汇进行初检,而后以查到的具体药物名称再进行特异性较高的检索。

第二步,以关键词“结核+药物”,出版年限“1998~2007年”限定进行模糊匹配检索,通过阅读

查到文献的标题及部分综述文章确定主要保肝药物的药名和商品名,包括:硫普罗宁、凯西莱、诺宁、水飞蓟宾、西利宾胺、当飞利肝灵、可予宁、水飞蓟葡甲胺片、益肝灵、甘草酸单铵、复方甘草酸苷、甘草甜素、美能、强力宁、甘草酸二铵、知肝保、甘利欣、甘平、谷胱甘肽、还原型谷胱甘肽、古拉定、泰特、葡醛内酯、护肝片。

第三步,以上述药物名称+结核作为关键字再次检索。人工过滤确定与研究主题相关的文章并下载全文。

### 3. 纳入及排除标准

(1)纳入标准为:①研究对象为应用抗结核药治疗的结核病患者,年龄、性别不限,用药前肝功能正常;②预防用药定义为保肝药与抗结核药同时应用,疗程、剂量不限,给药途径为口服或静脉滴注;③试验药物为一种保肝药或多种药物联用,对照组为安慰剂、空白或其他阳性对照;④研究设计类型为随机或非随机的临床对照试验。

(2)排除标准为:①无对照试验;②历史性对照研究;③病例对照研究;④按病人特点进行分配的试验(如性别、年龄等);⑤综述;⑥动物实验、细胞组织学研究;⑦因非预防性目的接受保肝药物治疗的研究。

4. 检索结果 文献检索及纳入排除流程如图22-1,最终纳入文献94篇。

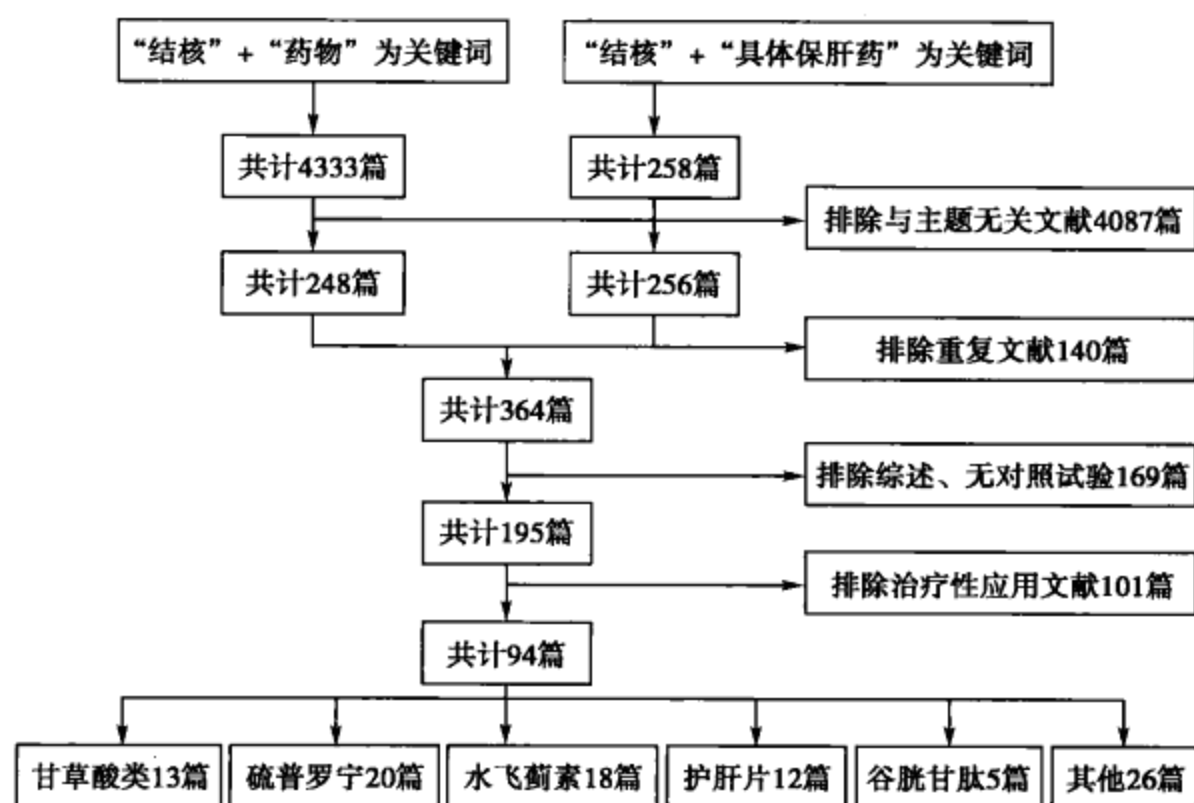


图22-1 文献纳入排除流程图

对纳入文献的基本情况观察发现,保肝药物主要是硫普罗宁、水飞蓟宾类、甘草酸类、护肝片、谷胱甘肽,少数涉及齐墩果酸、螺旋藻及中药丸剂等,药物种类共计42种。42篇文章使用空白照,另外52篇采用其他阳性对照,以肌苷片、葡醛内酯为主。各研究的样本量在40~1216人之间,中位数149人,多为16~80岁的成年人,结核病多为肺结核,但有16篇文章的研究对象也包括肺外结核病人。各研究采用的肝损害诊断指标及指标切点值各有不同,74项研究(占总研究数的78.7%)使用生化指标而非临床症状或停药来做判断,其中56篇文章仅以生化指标超过正常值上限即列为肝损害。

### (三) 评价证据

首先对所有的纳入试验,使用评价研究质量学质量的JADAD 三条目量表进行评分,满分为5分,总分1~2分者为低质量研究,总分3~5分为高质量研究。评分标准和各条目得分文献数如表22-2,没有总得分超过4分的文献,总分为1、2、3分的文献分别为25、63、5篇,可见多数文章质量不佳。

表22-2 JADAD三条目量表定义及文献数

条目	得分	文献数		
		2分	1分	0分
随机化	恰当: 如计算机产生的随机数字或类似的方法(2分)			
方法	不清楚: 描述为随机试验, 但未告知随机分配产生的方法(1分)	5	63	26
	不恰当: 如采用交替分配或类似方法的半随机化(0分)			
盲法	恰当: 使用完全一致的安慰剂或类似的方法(2分)			
	不详: 试验称为双盲法, 但未交代具体的方法(1分)	0	0	94
	非盲法: 未采用双盲法或盲的方法不恰当(0分)			
失访与退出	具体描述了撤除与退出的数量和理由(1分)	1	93	1
	未报告撤除或退出的数目或理由(0分)			

JADAD 评分标准有着简便易行、可量化比较的优点, 其所采用的3条目分别表明了临床试验研究设计的3个最重要方面。我们纳入的研究中仅有5篇文章使用了正确的随机化分配, 有14篇文章使用了伪随机分配而将自己的研究视为RCT, 另外53篇研究则没有意识到应在文献中报告随机化分配的具体方法。纳入研究没有一篇采用安慰剂和盲法, 因此其研究结果不可避免地要被质疑受到了安慰剂效应、研究对象和研究者的主观判断等带来的影响。绝大多数文献都报告无失访和退出, 体现了国内研究的研究对象依从性均非常好, 这是国内临床试验的优势之一, 但100%的随访率, 也有可能是因为存在选择偏倚, 或在研究后期有意删除了失访资料等现象。

其次, 考虑到JADAD 量表得分不高的原因除研究设计不良外, 也可能是由于研究报告不规范, 欲评估不规范报告的情况, 则需使用临床试验报告统一标准规范(consolidated standards of reporting trials, CONSORT)。CONSORT 声明旨在提供一个撰写RCT 报告的规范指南, 以保证所有研究相关信息均被涵盖且清楚描述, 来帮助编辑、读者进行发表审核、临床决策等, 自1997年发表以来受到了全球研究者及主要医学杂志的重视。我们将CONSORT 清单拆分成28个具体条目, 进行逐条评价。评价结果见表22-3, 可见大多数文献都有一些条目未报告清楚, 尤其是在标题中描述随机化、样本量的计算、分配隐藏、随机的具体实施过程、盲法、研究对象流程图、ITT 分析、研究效应大小及精确度报告、研究的局限性探讨等重要方面, 所有文章均未报告。对随机化方法的统计发现, 在19篇说明了随机产生方法的文章中, 仅有5篇文献是使用随机数字表法进行的正确随机分组, 其余14篇均采用登记号、入院顺序等伪随机方法分组。对于干预方法和内容的描述部分, 仅有29篇文献指出了试验药物的生产厂家, 均未提及药物批号。没有研究对研究对象进行分年龄、肝病史、治疗措施的亚组分析。所有文献的条目总完成率仅为31.7%。另外, 多篇文献有文章结构不清, 报告顺序颠倒的现象, 如本应出现在结果部分的研究对象基本情况描述, 常被放在方法中, 本应在方法中说明的观察结局定义则常在结果中被说明, 有些文章在结果部分没有报告不良反应, 却在讨论部分报告等。这些报告内容缺失固然可能与期刊杂志篇幅要求等有关, 但也反映了研究者对RCT 设计实施及结果解释的知识不足, 以及国内部分期刊杂志对CONSORT 声明的了解及应用尚不够充分。

表22-3 CONSORT质量评价结果

条目	描述	研究数 (百分比)	
		方法恰当 报告详尽	方法不当 或未报告
标题与摘要	标题中描述随机化	0	82 (100%)
	摘要中描述随机化	30 (36.6%)	52 (63.4%)
介绍			
背景	科学背景与原理解释	55 (67.1%)	27 (32.9%)
方法			
研究对象	研究对象的人选标准	11 (13.4%)	71 (86.6%)
	研究对象的排除标准	4 (4.9%)	78 (95.1%)
	资料收集的地点与背景	82 (100%)	0
干预	干预措施的实施方法	82 (100%)	0
	干预措施的详细内容	29 (35.4%)	53 (64.6%)
目标	设定的目标及假说	47 (57.3%)	35 (42.7%)
结局	主要、次要结局的定义及测量方法	46 (56.1%)	36 (43.9%)
样本大小	样本量的确定方法	0	82 (100%)
随机化	随机化序列的产生方法	5 (6.1%)	77* (93.9%)
分配隐藏	分配是否隐藏	0	82 (100%)
实施	随机化的实施人员及方法	0	82 (100%)
盲法	盲法对象	0	82 (100%)
统计学方法	各主要结局的分析方法及其他分析方法	41 (50%)	41 (50%)
结果			
研究对象的流动	研究对象的流动情况 (流程图)	0	82 (100%)
研究对象的征集	研究对象的征集过程及随访日期范围	20 (24.4%)	62 (75.6%)
基线数据	基线人口学和临床特征	31 (37.8%)	52 (63.4%)
分析的数量	各不同分析中研究对象的总人数 (分母)	82 (100%)	0
	进行ITT分析	0	82 (100%)
	结果用绝对数表示	81 (98.8%)	1 (1.22%)
结局和估计	报告每个组的主要、次要结局的效应大小及精确度	0	82 (100%)
辅助分析	报告亚组分析及调整分析及哪些是预先设定的分析哪些是探索性分析	2 (2.4%)	80 (97.6%)
不良反应事件	各干预组的不良反应事件	28 (34.1%)	54 (65.9%)
讨论			
解释	结合研究假设、潜在偏倚及不精确的来源等综合解释研究的局限性及为了避免偏倚所做的努力	0	81 (98.8%)
可推广性	试验结果的外部真实性	31 (37.8%)	51 (62.2%)
证据总体	结合现有证据的全面解释	20 (24.4%)	62 (75.6%)

\*：含14篇使用伪随机分配方法的文章

通过上述质量评价可见,目前国内该领域的研究尚局限于小样本、无安慰剂对照、非盲法的临床试验,无论是研究的设计还是研究结果的报告规范性均需进一步改善、提高。

另外,对纳入研究的疗效研究结果进行统计发现,无论是与空白对照还是与阳性对照药物比较,当结局指标是肝损害是否发生时,研究均报告阳性结果,呈现“一边倒”现象,图22-2为与空白比较和与阳性对照比较的研究的RR值分布图,可见除一项差异无统计学意义的研究报告RR值为1.33外,其他所有研究RR值均小于1,怀疑可能存在一定的发表偏倚。

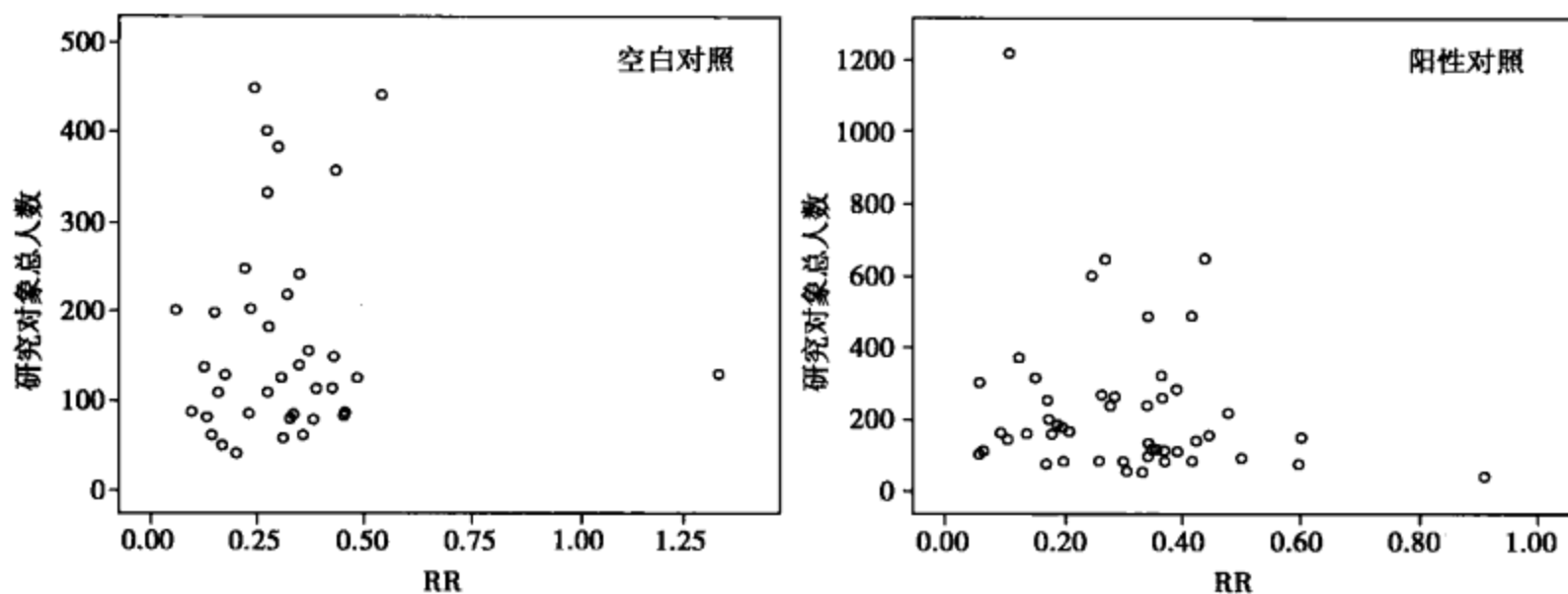


图22-2 纳入研究的RR值分布图

除肝损害发生率指标外,有6篇文献报告使用保肝药物可降低肝损害的严重程度。有6篇文章比较了试验组与对照组肝损害出现时间的差异,结论并不统一。另有28篇文献报告试验组因肝损害中断结核化疗的中断率低于对照组。

纳入文献中有66篇文献未提及保肝药是否引起不良反应,22篇文献报告保肝药未引发不良反应,6篇文献报告应用保肝药引起皮疹、皮疹伴瘙痒、胃肠道反应或发热等不良反应,但均未进行不良反应与药物的关联性评价。

综上,对检索文献的综合评价表明,现有研究均为无安慰剂对照、非盲法的试验,采用的肝损害判断标准不尽合理,缺乏亚组分析,且报告信息不清,总体质量不佳,可能会受到各种偏倚的影响,因此无法提供可靠的疗效和安全性证据,即我们无法从现有研究中得到足够信息判断预防性使用保肝药对抗结核药物致肝损害的疗效及其应用的安全性,也无法判断何种药物是最佳保肝药物。由于证据的真实性受到挑战,也无法在此基础上对证据的临床重要性进行评价。

#### 四、实践决策

通常,我们收集到证据,评价完证据的真实性和临床重要性之后,即开始考虑证据的适用性,包括病人是否与研究证据中的病人特征相似,研究中的干预是否在当地可行,治疗对病人的潜在益处和害处如何,以及对于这些益处和害处病人的观念和预期如何。但经过上述证据评价过程,可知目前尚无针对这一临床问题的优质证据,因此无法使用研究中的数据来判断是否应予病人保肝治疗。在这种情况下,只能从定性的治疗益害以及病人的观念方面着手进行决策(表22-4)。

综合上述益害分析及与病人沟通后,我们认为由于保肝药物的预防疗效不能确认,患者年纪较大,两类药物联用发生其他不良反应的风险可能相对较高,同时患者肝功能正常,无肝病史酗酒史,无肝病家族史,发生肝损害的风险有限,且患者因保肝药物不可报销而对其费用有一定抵触,决定暂不予病人保肝药物,但要求在开始服用抗结核药物2周和1月时各来复查肝功能一次,在服药过程

表22-4 治疗的益害及病人观念分析表

益害及病人观念		进行预防性保肝治疗	不进行预防性保肝治疗
肝损害	是否发生	可能发生	可能发生
	概率大小	0.64%~38.64% (中位数7.29%)	11.9%
	后果	治愈率或好转率98.1%	治愈率或好转率98.1%
	费用	保肝治疗: 30~500元/月 肝损害: 200~20 000元	200~20 000元
皮疹等不良反应	是否发生	可能发生	无
	概率大小	皮疹等: 1% (不确定) 其他风险*: 不确定	/
	后果	皮疹等: 100% 痊愈 其他风险*: 不确定	/
	费用	皮疹等: 约10~200元 其他风险*: 不确定	0元
	病人观念	肝损害和不良反应的风险都不大, 都可以接受	

\* 部分保肝药物如硫普罗宁曾被报告可引起过敏性休克甚至死亡等, 此种风险无法估计

中出现不适也要及时检查肝功。

## 五、后效评估

依照上述循证决策, 给予患者抗结核治疗药物并在患者开始服药后2周复查肝功能, 其ALT、AST、TB 的检查结果分别为24、23、15.1mmol/L, 服药1个月时再次复查上述指标分别为17、20、8.3mmol/L, 其他检测如血常规、肾功能检测等均未见任何异常。患者未报告有不适现象, 服药2个月后痰菌转为阴性, 6个月短程化疗结束时痰菌检查阴性, X线片显示右肺阴影已消散, 结核治愈。

## 六、小结

本章介绍了一个是否给予预防性药物的循证决策过程, 作为药物疗效评价和安全性评价的循证实践过程, 这个例子遵循了一般的步骤和方法, 但它仍有自己的特殊性, 即在某个临床问题没有最佳证据时, 应当如何决策。众所周知, 随机对照试验和随机对照试验的

系统综述是疗效评价的最佳证据，但现实是，目前仍有众多的问题没能被科学的阐明，仍有众多的空白领域需要进一步的研究。作为临床医生，遇到这种情况，一方面应当尽量关注相关研究的最新发展，及时获得优质信息；另一方面，也应在现有证据的条件下，结合病人自身的情况及治疗措施的益害两方面做出判断。在条件具备时，可进行定量的利弊似然比（likelihood of help vs harm, LHH）计算，即比较采用治疗可降低的疾病风险与带来不良后果风险，同时结合病人对疾病风险和不良后果风险的打分，综合计算该治疗对病人的利弊可能比。在没有良好证据提供上述风险信息数据时，仍应遵循相同的原则进行近似的定性估计。本章所举的例子中，对病人个体来说，发生肝损害的风险相对较小，而预防性治疗的疗效不能确定，且可能带来其他风险，因此权衡的结果是不给予保肝治疗。但如果面对的临床问题是对一个严重疾病患者，如肾癌患者，考虑是否给予其某种药物治疗，而该种药物尚无最佳证据支持其疗效，但已知其安全性较好，则决策结果可能会倾向于给予治疗。另外，除上述利弊比外，治疗的成本效果、效益比对经济条件有限的患者来说也非常重要，决策时应予一定的考虑。

（夏愔愔 詹思延）



## 主要参考文献

1. BARI 2D Study Group. A randomized trial of therapies for type 2 diabetes and coronary artery disease. *N Engl J Med.* ,2009,360:2503-2515.
2. Ceri J Phillips. *Health economics: An introduction for health professional.* Massachusetts, USA: Blackwell Publishing Ltd,2005.
3. Keeley EC, Boura JA, Grines CL. Primary angioplasty versus intravenous thrombolytic therapy for acute myocardial infarction: a quantitative review of 23 randomised trials. *Lancet*,2003,361:13-20.
4. Sackett DL, et al. *Clinical Epidemiology. A basis science for clinical medicine.*2nd ed. Boston: Little Brown, 1992.
5. Sackett DL, Richardson WS, Rosenberg W, et al. *Evidence-based Medicine. How to practice & Teaching EBM.*2nd ed. London: Churchill Livingstone, 1997.
6. 王家良. 临床流行病学. 第3版. 上海: 上海科技出版社,2009.
7. 王家良. 循证医学. 北京: 人民卫生出版社,2005.
8. Beth Dawson, Robert G Trapp. *Basic & clinical biostatistics.*3rd ed. Lange Medical Books/McGraw-Hill, Medical Publishing Division,2001.
9. Komajda M, Lapuerta P, Hermans N, et al. Adherence to guidelines is a predictor of outcome in chronic heart failure: the MAHLER survey. *European Heart Journal*,2005,26:1653-1659.
10. Liz Trinder, Shirley Reynolds. *Evidence-Based Practice—A CRITICAL APPRAISAL.* Oxford: Blackwell Publishing company,2000.
11. Perrillo R. Benefits and risks of interferon therapy for hepatitis B. *Hepatology*,2009,49(5 Suppl):S103-111.
12. 祁国明. 卫生技术评估. *中华医学科研管理杂志*,2000,13:3-4.
13. The Task Force on the management of ST-segment elevation acute myocardial infarction of the European Society of Cardiology. Management of acute myocardial infarction in patients presenting with persistent ST-segment elevation. *European Heart Journal*,2008,29: 2909-2945.
14. Guyatt G. An emerging consensus on grading recommendations? *Chin J Evid-based Med*,2007,7:1-8.
15. 陈杰, 贾继东. 2007年美国肝病学会慢性乙型肝炎防治指南推荐意见介绍. *临床肝胆病杂志*,2007,23: 83-87.
16. Antman EM, Anbe DT, Armstrong PW, et al. ACC/AHA guidelines for the management of patients with ST-elevation myocardial infarction—a executive summary: A report of the American College of Cardiology/American Heart Association Task Force on practice guidelines (writing committee to revise the 1999 guidelines for the management of patients with acute myocardial infarction): developed in collaboration with the Canadian Cardiovascular Society. *Circulation*,2004,110:588-636.
17. 中华医学会肝病学分会/中华医学会感染病学分会. 慢性乙型肝炎防治指南.2005.
18. Field MJ, Lohor KN. *Clinical practice guidelines: directions for a new program committee to advise the public health service on clinical practice guidelines.* Institute of Medicine. National Academy Press Washington DC, 1990.
19. Antman EM, Hand M, Armstrong PW, et al.2007 focused update of the ACC/AHA 2004 guidelines for the

- management of patients with ST-elevation myocardial infarction: A report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines: Developed in collaboration with the Canadian Cardiovascular Society endorsed by the American Academy of Family Physicians: 2007 Writing group to review new evidence and update the ACC/AHA 2004 guidelines for the management of patients with ST-elevation myocardial infarction, writing on behalf of the 2004 writing committee. *Circulation*, 2008, 117 : 296-329.
20. SIGN ( Scottish Intercollegiate Guidelines Network ). Report on a recommended referral document. SIGN Publication No.31, 1998.
  21. De Luca G, Biondi-Zoccai G, Marino P. Transferring patients with ST-segment elevation myocardial infarction for mechanical reperfusion: A Meta-regression analysis of randomized trials. *Ann Emerg Med*, 2008, 52:665-676.
  22. Dickstein K, Cohen-Solal A, Filippatos G, et al. ESC guidelines for the diagnosis and treatment of acute and chronic heart failure. *European Heart Journal*, 2008, 29: 2388-2442.
  23. Bleser LD, Depreitere R, Waele KD, et al. Defining pathways. *Journal of Nursing Management*, 2006, 14: 553-563.
  24. Magid DJ, Calonge BN, Rumsfeld JS, et al. For the National Registry of Myocardial Infarction 2 and 3 Investigators. Relation between hospital primary angioplasty volume and mortality for patients with acute MI treated with primary angioplasty vs thrombolytic therapy. *JAMA*, 2000, 284:3131-3138.
  25. MacMahon S, Collins R, Peto R, et al. Effects of prophylactic lidocaine in suspected acute myocardial infarction: an overview of results from the randomized, controlled trials. *JAMA*, 1988, 260:1910-1916.
  26. 赵亚利, 崔树起, 彭晓霞. 国内临床指南发展现状及国内外指南比较分析. *中国全科医学*, 2005, 8:593-596.
  27. Hine LK, Laird N, Hewitt P, et al. Meta-analytic evidence against prophylactic use of lidocaine in acute myocardial infarction. *Arch Intern Med*, 1989, 149 : 2694-2698.
  28. 杨茗, 董碧蓉. 临床医生如何评价和应用临床实践指南. *中国循证医学杂志*, 2008, 8:235-238.
  29. Sadowski ZP, Alexander JH, Skrabucha B, et al. Multicenter randomized trial and a systematic overview of lidocaine in acute myocardial infarction. *Am Heart J*, 1999, 137: 792-798.
  30. Boersma E, Maas AC, Deckers JW, et al. Early thrombolytic treatment in acute myocardial infarction: reappraisal of the golden hour. *Lancet*, 1996, 348:771-775.
  31. Alexander JH, Granger CB, Sadowski Z, et al. Prophylactic lidocaine use in acute myocardial infarction: incidence and outcomes from two international trials. *Am Heart J*, 1999, 137:799-805.
  32. Evidence-based Medicine Working Group. Evidence-Based Medicine - A new Approach to Teaching the Practice of Medicine. *JAMA*, 1992, 268:2420-2425.
  33. Drummond MF, Richardson WS, O' Brien BJ, et al. How to use an article on economic analysis of clinical practice A. Are the results of the study valid? *JAMA*, 1997, 277:1552-1557.
  34. Greenbalgh T. How to read a paper: The basics of evidence based medicine .2nd edition. London: BMJ Books, 2003.
  35. Home PD, Pocock SJ, Beck-Nielsen H, et al. Rosiglitazone evaluated for cardiovascular outcomes in oral agent combination therapy for type 2 diabetes ( RECORD ): a multicentre, randomised, open-label trial. *Lancet*, 2009, 373 ( 9681 ):2125-2135.
  36. Zijlstra F, Hoorntje JC, deBoer MJ, et al. Long-term benefit of primary angioplasty as compared with thrombolytic therapy for acute myocardial infarction. *N Engl J Med*, 1999, 341:1413-1419.
  37. 王家良. 循证医学. 北京: 人民卫生出版社, 2001.
  38. AGREE. Appraisal of Guidelines for Research & Evaluation: AGREE Instrument Training Manual [EB/OL]. The AGREE Collaboration. January, 2003.
  39. Di Marco V, Craxi A. Chronic hepatitis B: who to treat and which choice of treatment? *Expert Rev Anti Infect*

- Ther, 2009, 7:281-291.
40. Paul Glasziou, Chris Del Mar, Janet Salisbury. Evidence-based Medicine Workbook. London: BMJ Publishing Group, 2003.
  41. 李幼平, 敬媛媛, 艾昌林. 循证医学与卫生技术评估. 医学与哲学, 2007, 28:75-76.
  42. 李幼平. 循证医学. 北京: 高等教育出版社, 2003.
  43. 王家良. 临床流行病学- 临床科研设计、测量与评价. 第2版. 上海: 上海科技出版社, 2001.
  44. Sharon E. Straus, W. Scott Richardson, Paul Glasziou, et al. 循证医学实践和教学. 詹思延译. 北京: 北京大学医学出版社, 2006.
  45. 中华医学会心血管病学分会. 急性心肌梗死诊断和治疗指南. 中华心血管病杂志, 2001, 29: 710-725.
  46. 中华医学会心血管病学分会, 中华心血管病杂志编辑委员会. 经皮冠状动脉介入治疗指南(2009). 中华心血管病杂志, 2009, 37(1): 4-25.
  47. Milos Jenicek. Foundations of evidence-based medicine. New York: The Parthenon Publishing Group Inc, 2005.
  48. Nissen SE, Wolski K. Effect of rosiglitazone on the risk of myocardial infarction and death from cardiovascular causes. N Engl J Med, 2007, 356(24):2457-2471.
  49. Nunn CM, O' Neill WW, Rothbaum D, et al. Long-term outcome after primary angioplasty: report from the primary angioplasty in myocardial infarction (PAMI-I) trial. J Am Coll Cardiol, 1999, 33:640-646.
  50. O' Brien BJ, Heyand D, Richardson WS, et al. How to use an article on economic analysis of clinical practice B. What are the results and will they help me in caring for my patients? JAMA, 1997, 277(22):1802-1806.
  51. Rosenberg W, Donald A. Evidence-based medicine: an approach to clinical problem-solving, BMJ, 1995, 310:1122-1125.
  52. 李立明. 流行病学. 第6版. 北京: 人民卫生出版社, 2007.
  53. AGREE Collaboration. Development and validation of an international appraisal instrument for assessing the quality of clinical practice guidelines: The AGREE project. Qual Saf Health Care, 2003, 12:18-23.
  54. Morgan M, Keeffe EB. Diagnosis and treatment of chronic hepatitis B: 2009 update. Minerva Gastroenterol Dietol. 2009, 55(1):5-22.
  55. Sackett DL, Straus SE, Richardson WS, et al. Evidence-based medicine: How to practice and teach EBM. London: Churchill-Livingstone, 2000.
  56. 樊宏, 刘越. 国内外卫生技术评估研究现状及应用. 国外医学: 卫生经济学分册, 2007, 24:72-76.
  57. 王家良. 临床流行病学. 第2版. 北京: 人民卫生出版社, 2004.
  58. 徐德忠. 循证医学入门: 临床科研方法与实例评价. 第2版. 西安: 第四军医大学出版社, 2006.
  59. Andersen HR, Nielsen TT, Rasmussen K, et al. A comparison of coronary angioplasty with fibrinolytic therapy in acute myocardial infarction. N Engl J Med, 2003, 349(8):733-742.
  60. David Machin, Michael J. Campbell, et al. Design of studies for medical research. New York: John Wiley & Sons Ltd, 2005.
  61. Edwcod M. Critical appraisal of epidemiological studies and clinical trials. 2nd edition, London: Oxford Univ. press, 1998.
  62. BMJ publishing Group. Clinical Evidence, A Compendium of the best available evidence for effective health care. London:BMJ, 1999.
  63. Manchikanti L, Singh V, Helm S, et al. A critical appraisal of 2007 American College of Occupational and Environmental Medicine (ACOEM) practice guidelines for interventional pain management: an independent review utilizing AGREE, AMA, IOM, and other criteria. Pain Physician, 2008, 11(3): 291-310.
  64. Sackett DL, Straus SE, Richardson WS, et al. Evidence-based Medicine. How to practice and teaching EBM. 2nd edition. Edinburgh:Churchill Livingston, 2000.

65. Sharon E. Straus, W. Scott Richardson, Paul Glasziou, et al. Evidence Based Medicine: How to practice and teach EBM. 3rd Edition. New York: Elsevier/Churchill Livingstone, 2005.
66. Zheng Y, Zhao L, Wu T, et al. Efficacy of consensus interferon in treatment of HBeAg-positive chronic hepatitis B: a multicentre, randomized controlled trial. *Virology*, 2009, 6:99-105.
67. 胡善联. 卫生经济学. 上海: 复旦大学出版社, 2004.
68. 王吉耀. 循证医学与临床实践. 第2版. 北京: 科学出版社, 2006.
69. 应向华, 曹建文, 陈洁, 等. 卫生技术评估的基本步骤. *中国卫生资源*, 2005, 8:37-39.
70. Tierney Jr LM, McPhee SJ, Papadakis MA. *Current Medical Diagnosis & Treatment*. 41st ed. New York: McGraw-Hill Co., 2002.
71. 杜丰, 苏雪松, 李德天, 等. 注射用硫酸头孢噻利治疗急性细菌性感染随机对照临床研究. *中国临床药理学杂志*, 2009, 25(3):195-200.
72. Antman EM, Anbe DT, Armstrong PW, et al. ACC/AHA guidelines for the management of patients with ST-elevation myocardial infarction—a executive summary: A report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines (writing committee to revise the 1999 guidelines for the management of patients with acute myocardial infarction): developed in collaboration with the Canadian Cardiovascular Society. *Circulation*, 2004, 110:588-636.
73. 王家良. 循证医学. 北京: 人民卫生出版社, 2005.
74. Vlayen J, Aertgeerts B, Hannes K, et al. A systematic review of appraisal tools for clinical practice guidelines: multiple similarities and one common deficit. *International Journal for Quality in Health Care*, 2005, 17(3): 235-242.
75. Wang H, Wang QH, Wang HF, et al. A clinical study of the efficacy and safety of secretory human interferon alpha-2a treatment for chronic hepatitis B. *Zhonghua Gan Zang Bing Za Zhi*, 2004, 12(10):589-592.
76. 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. *中国循证医学杂志*, 2008, 8(2):127-133.
77. 李静. 卫生技术评估的基本方法. *中国循证医学杂志*, 2003, 3(4):315-320.
78. 舍曼·富兰德, 艾伦·C·古德曼, 迈伦·斯坦诺. 卫生经济学. 第3版. 北京: 中国人民大学出版社, 2004.
79. Caroline M Mitchell, Jane E Hitti, Kathy J Agnew, et al. Comparison of oral and vaginal metronidazole for treatment of bacterial vaginosis in pregnancy: impact on fastidious bacteria *BMC Infectious Diseases*, 2009, 9:89-95.
80. Silber S, Albertsson P, Avilés FF, et al. Task Force for Percutaneous Coronary Interventions of the European Society of Cardiology. Guidelines for percutaneous coronary interventions. The Task Force for Percutaneous Coronary Interventions of the European Society of Cardiology. *Eur Heart J*, 2005, 26(8): 804-847.
81. 程晓明. 卫生经济学. 北京: 人民卫生出版社, 2003.
82. 王滨有. 临床实用流行病学. 哈尔滨: 黑龙江科学技术出版社, 2002.
83. Arase Y, Tsubota A, Saitoh S, et al. Randomized, controlled trial of natural interferon-alpha therapy for e-antigen-positive chronic hepatitis B patients. *Hepatology*, 2002, 23(2):98-104.
84. 顾杰, 姜林娣. 健康状况的效用值和测量方法. *中华行为医学与脑科学杂志*, 2009, 18(1): 84-85.
85. Lohr K, Field M. A provisional instrument for assessing clinical practice guidelines. In *Guidelines for clinical practice: from development to use*. Washington DC: National Academy Press, 1992.
86. 王家良. 临床流行病学. 第3版. 北京: 人民卫生出版社, 2008.
87. Smith SC, Feldman TE, Hirshfeld JW, et al. ACC/AHA/SCAI Writing Committee to Update 2001 Guidelines for Percutaneous Coronary Intervention. ACC/AHA/SCAI 2005 guideline update for percutaneous coronary intervention: a report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines (ACC/AHA/SCAI Writing Committee to Update 2001 Guidelines for Percutaneous Coronary

- Intervention). *Circulation*, 2006, 113(7): e166-286.
88. 吴明. 卫生经济学. 北京: 北京医科大学出版社, 2002.
  89. Higgins JPT, Green S. *Cochrane Handbook for Systematic Reviews of Interventions*. Version 5.0.2 [updated September 2009]. The Cochrane Collaboration, 2008. Available from [www.cochrane-handbook.org](http://www.cochrane-handbook.org).
  90. Lohr K, Field M. *Clinical practice guidelines: directions for a new program*. Institute of Medicine, National Academy Press, Washington DC, 1990.
  91. Shamliyan TA, MacDonald R, Shaukat A, et al. Antiviral therapy for adults with chronic hepatitis B: a systematic review for a National Institutes of Health Consensus Development Conference. *Ann Intern Med*, 2009, 150:111-124.
  92. Spencer B, King, Sidney C, Smith, John W, Hirshfeld, et al. 2007 focused update of the ACC/AHA/SCAI 2005 guideline update for percutaneous coronary intervention: a report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines: 2007 Writing Group to review new evidence and update the ACC/AHA/SCAI 2005 guideline update for percutaneous coronary intervention, writing on behalf of the 2005 Writing Committee. *Circulation*, 2008, 117(2):261-295.
  93. Ayoub WS, Keeffe EB. Review article: current antiviral therapy of chronic hepatitis B. *Aliment Pharmacol Ther*, 2008, 28(2):167-177.
  94. Cluzeau FA, Littlejohns P, Grimshaw JM, et al. Development and application of a generic methodology to assess the quality of clinical guidelines. *Int J Qual Health Care*, 1999, 11: 21-28.
  95. Cucherat M, Bonnefoy E, Tremeau G. WITHDRAWN: Primary angioplasty versus intravenous thrombolysis for acute myocardial infarction. *Cochrane Database Syst Rev*, 2007, 18: CD001560.
  96. Férrir G, Kaptein S, Neyts J, et al. Antiviral treatment of chronic hepatitis B virus infections: the past, the present and the future. *Rev Med Virol*, 2008, 18(1):19-34.
  97. Haynes RB. Of studies, syntheses, synopses, summaries, and systems: the "5S" evolution of information services for evidence-based healthcare decisions. *Evid Based Med*, 2006, 11: 162-164.
  98. Cook NR, Lee IM, Gaziano JM, et al. Low-dose aspirin in the primary prevention of cancer: the Women's Health Study: a randomized controlled trial. *JAMA*, 2005, 294(1):47-55.
  99. Davis P, Turner-Gomes S, Cunningham K, et al. Precision and accuracy of clinical and radiological signs in premature infants at risk of patent ductus arteriosus. *Arch Pediatr Adolesc*, 1995, 149(10):1136-1141.
  100. Dube C, Rostom A, Lewin G, et al. The use of aspirin for primary prevention of colorectal cancer: a systematic review prepared for the U. S. Preventive Services Task Force. *Ann Intern Med*, 2007, 146(5):365-375.
  101. Friis S, Thomassen L, Sorensen HT, et al. Nonsteroidal anti-inflammatory drug use and breast cancer risk: a Danish cohort study. *Eur J Cancer Prev*, 2008, 17(2):88-96.
  102. Gonzalez-Perez A, Garcia Rodriguez LA, Lopez-Ridaura R. Effects of non-steroidal anti-inflammatory drugs on cancer sites other than the colon and rectum: a meta-analysis. *BMC Cancer*, 2003, 3:28.
  103. Guyatt GH, Rennie D, Meade MO, et al. *Users' Guides To The Medical Literature: A Manual for Evidence-Based Clinical Practice*. Chicago:AMA Press, 2008.
  104. Haynes RB. Of studies, syntheses, synopses, and systems: the "4S" evolution of services for finding current best evidence. *ACP J Club*, 2001, 134(2):A11-A13.
  105. Haynes RB. Of studies, syntheses, synopses, summaries, and systems: the "5S" evolution of information services for evidence-based health care decisions. *ACP J Club*, 2006, 145(3):A8.
  106. Jackson, R. Guidelines on preventing cardiovascular disease in clinical practice. *BMJ*, 2000, 320: 659-661.
  107. Jaeschke R, Guyatt GH, Sackett DL, and the Evidence-Based Medicine Working Group. Users' guides to Evidence-based Medicine. How to use an article about a diagnostic test. *JAMA*, 1994, 271(5):389-391, 1994,

- 271(9):703-707.
108. Khuder SA, Mutgi AB. Breast cancer and NSAID use: a meta-analysis. *Br J Cancer*, 2001, 84(9):1188-1192.
109. Mangiapane S, Blettner M, Schlattmann P. Aspirin use and breast cancer risk: a meta-analysis and meta-regression of observational studies from 2001 to 2005. *Pharmacoepidemiology Drug Safety*, 2008, 17(2):115-124.
110. MILTON C. 临床决策分析. 曹建文, 译. 上海: 复旦大学出版社, 2005.
111. Moher D, Schulz KF, Altman DG. The CONSORT statement: revised recommendations for improving the quality of report of parallel group randomized trials. *BMC Medical Research Methodology*, 2001, 1-2.
112. Paul Glasziou, Chris Del Mar, Janet Salisbury. *Evidence-based Practice Workbook*. 2nd ed. Canberra: Blackwell Publishing, 2007.
113. Petitti DB. *Meta-analysis, Decision analysis and Cost-effect analysis: Methods for Quantitative Synthesis in Medicine*. 2nd ed. London: Oxford University Press, 2000.
114. Rathbun SW, Raskob GE, Whitsett TL. Sensitivity and specificity of helical computed tomography in the diagnosis of pulmonary embolism: A systematic review. *Arch Intern Med*, 2000, 132: 227-232.
115. Stein PD, Hull RD, Ghali WA, et al. Two decades of hospital practice trends for diagnosing deep vein thrombosis and pulmonary embolism. *Arch Intern Med*, 2003, 163:1213 -1219.
116. Stephen G, Jerome. *Decision analysis*. *N Eng J M*, 1987, 316:250-259.
117. Straus SE, Richardson WS, Glasziou PP, et al. *Evidence-based medicine: How to practice and teach EBM*. 3rd edition. Edinburgh: Churchill Livingstone, 2005.
118. WHO's Certified. WHO Stop TB Department: Global tuberculosis control: epidemiology, strategy, financing, WHO report 2009[DB/OL]. [http://www.who.int/entity/tb/publications/global\\_report/2009/pdf/full\\_report.pdf](http://www.who.int/entity/tb/publications/global_report/2009/pdf/full_report.pdf)
119. 郭美华, 马满玲, 尹海燕. 114例硫普罗宁不良反应文献分析. *中国新药杂志*, 2006, 15(16):1403-1405.
120. 洪明晃. 第十六章临床决策分析// 王家良. *循证医学*. 北京: 人民卫生出版社, 2005: 145-157.
121. 李大林. 第十三章决策分析// 詹思延. *循证医学与循证保健*. 北京: 北京医科大学出版社, 2002.
122. 吕筠. 提出研究的问题// 詹思延. *循证医学和循证保健*. 北京: 北京医科大学出版社, 2002: 5-14.
123. 唐金陵, Muir Gray. *循证医学: 循证医疗卫生决策*. 北京: 北京医科大学出版社, 2004:148-154.
124. 夏愔愔, 詹思延. 国内抗结核药物不良反应发生率的综合分析. *中华结核和呼吸杂志*, 2007(6):419-423.
125. 杨克虎. *循证医学*. 北京: 人民卫生出版社, 2007.
126. 臧焰, 秦晓光. 诊断性试验文献中的诊断标准问题. *中华检验医学杂志*, 2002, 25: 127-128.
127. 翟振国, 伍燕兵, 王辰. 肺血栓栓塞症的诊断策略. *中华医学杂志*, 2003, 83: 703-704.
128. 詹思延. *循证实践与教学*. 第3版. 北京: 北京大学医学部出版社, 2006.
129. 张俊华, 商洪才, 张伯礼. 系统评价和Meta分析质量的评价方法. *中西医结合学报*, 2008, 6:337-339.
130. Noteboom JT, Allison SC, Cleland JA, et al. A primer on selected aspects of evidence-based practice to questions of treatment. Part 2: interpreting results, application to clinical practice, and self-evaluation. *J Orthop Sports Phys Ther*, 2008, 38(8):485-501.
131. Anderson T. Evaluating Diagnostics. *TDR news*, 2009, (83):12-17.
132. Stark R, Helenius IM, Schimming LM, et al. Real-time EBM: from bed board to keyboard and back. *J Gen Intern Med*, 2007, 22:1656-1660.

# 英汉名词索引

95% confidence interval, 95%CI

95% 可信区间

35

## A

absolute risk reduction, ARR

绝对危险降低率

39, 68

accuracy, ACC

准确度

98

ACP Journal Club

美国内科医师学会杂志俱乐部

19

acute coronary syndrome, ACS

急性冠脉综合征

175

acute myocardial infarction

急性心肌梗死

177

adverse drug event, ADE

药物不良事件

118

adverse drug reactions, ADR

药物不良反应

68, 118

an individual randomized trial, n-of-1 trial

单个病人的随机对照试验

110

applicability

适用性

109, 131

appraisal of guidelines research and evaluation, AGREE

指南研究与评价工具

160, 161

assembly bias

集中性偏倚

128

attribute risk, AR

归因危险度

68

attrition bias

随访偏倚

51

## B

“background” questions

背景问题

22

benefit cost ratio, BCR

效益成本比

139

benefit

效益

137

best-evidence resources

最佳证据资源

23

biologic evidence

生物学依据

67

biologic onset

生物学发病期

127

burden of disease

疾病负担

160

## C

case control study

病例对照研究

85, 126

case-fatality rate

病死率

129

categorical variable

分类变量

33

chance nodes

机会节点

76

China National Knowledge Infrastructure, CNKI

中国知识基础设施工程

21

Chinese Medical Current Contents, CMCC

中文生物医学期刊数据库

22

clinical course

临床病程

127

clinical decision analysis, CDA

临床决策分析

74

clinical heterogeneity

临床异质性

53

clinical pathway, CP

临床路径

159

clinical practice guideline, CPG

临床实践指南

21, 154, 190

clinical question

临床问题

10

- |   |                      |            |
|---|----------------------|------------|
| clinical stage  | 临床期                  | 127        |
| Cochrane Central Register of Controlled Trials, CENTRAL           | Cochrane 临床对照试验中心注册库 | 21         |
| cohort study  | 队列研究                 | 90, 126    |
| co-intervention   | 干扰                   | 111        |
| computerized decision support system, CDSS                        | 计算机决策支持系统            | 18         |
| concealed randomization   | 隐匿性随机分配              | 110        |
| confidence interval, CI   | 可信区间                 | 39         |
| confounders   | 混杂因素                 | 90         |
| consolidated standards of reporting trials, CONSORT               | 临床试验报告统一标准规范         | 198        |
| contamination   | 沾染                   | 111        |
| cost-benefit analysis, CBA  | 成本效益分析               | 139        |
| cost-effectiveness analysis, CEA                                  | 成本效果分析               | 139        |
| cost-minimization analysis, CMA                                   | 最小成本分析               | 138        |
| cost-utility analysis, CUA  | 成本效用分析               | 140        |
| cost  | 成本                   | 136        |
| Council for International Organization of Medical Sciences, CIOMS | 国际医学科学组织委员会          | 119        |
| critical appraisal  | 严格评价                 | 123, 131   |
| cure rate   | 治愈率                  | 129        |
| <b>D</b>  |                      |            |
| decision nodes  | 决策节点                 | 76         |
| decision tree   | 决策树                  | 75         |
| diagnostic odd ratio, DOR   | 诊断比值比                | 99         |
| direct costs  | 直接成本                 | 136        |
| direct medical costs  | 直接医疗成本               | 137        |
| direct nonmedical costs   | 直接非医疗成本              | 137        |
| disability adjusted life year, DALY                               | 失能调整生命年              | 137        |
| disability rate   | 致残率                  | 130        |
| discount rate   | 贴现率                  | 142        |
| doer  | 证据提供者                | 4          |
| <b>E</b>  |                      |            |
| educational prescription  | 教育处方                 | 13         |
| effectiveness   | 效果                   | 137, 147   |
| efficacy  | 效力                   | 147        |
| epidemiology triangle   | 流行病学三角               | 87         |
| essential health technology, EHT                                  | 基本卫生技术               | 146        |
| essential medicines list, EML                                     | 基本药物目录               | 146        |
| etiologic fraction, EF  | 病因学分数                | 68         |
| evidence-based guidelines   | 循证指南                 | 154        |
| evidence-based journal abstracts                                  | 循证杂志摘要               | 18         |
| evidence based medicine, EBM                                      | 循证医学                 | 1          |
| external validity   | 外部真实性                | 30, 50, 67 |
| <b>F</b>  |                      |            |
| fixed effect model  | 固定效应模型               | 53, 56     |
| "foreground" questions  | 前景问题                 | 22         |
| forest plot   | 森林图                  | 53         |



gold standard	<b>G</b>	金标准	97
health economics evaluation	<b>H</b>	卫生经济学评价	136
health technology assessment, HTA		卫生技术评估	145
health technology		卫生技术	145
hepatitis B, HB		乙型病毒性肝炎	188
heterogeneity test		异质性检验	53
importance	<b>I</b>	重要性	109, 131
inception cohort		起始队列	132
incremental cost-effectiveness ratio, ICER		增量成本效果比	139
indirect costs		间接成本	137
intention to treat analysis		意向性治疗分析	51, 111
internal validity		内部真实性	30, 50, 66, 109
International Society of Technology Assessment in Health care, ISTAHC		国际卫生技术评估协会	145
interval of quartile range, IQR		四分位间距	34
likelihood of help vs harm, LHH	<b>L</b>	利弊比	69, 115
lost to follow up bias		失访偏倚	128
mean difference, MD	<b>M</b>	均数差值	43, 53, 54
measurement bias		测量偏倚	128
measurement		测量	3
median		中位数	34
methodological heterogeneity		方法学异质性	53
missing value		缺失值	35
narrative review	<b>N</b>	叙述性文献综述	47
National Guideline Clearinghouse, NGC		美国国家指南交换库	156
non-steroidal anti-inflammatory drugs, NSAIDs		非甾体类抗炎药	169
number needed to harm, NNH		需治多少病例才发生一例不良反应	113
number needed to treat, NNT		需治疗多少病例数才能获得一例 最佳效果	68, 112
numerical variable		数值变量	34
observational study	<b>O</b>	观察性研究	49, 90
odds ratio, OR		比值比	39, 64, 68
outcome		结局	127
overuse		滥用	155
parallel test	<b>P</b>	平行试验	104

patient's expected event rate, PEER	患者预期事件发生率	114
performance bias	实施偏倚	51
positive likelihood ratio, +LR	阳性似然比	99
positive predictive value, +PV	阳性预测值	98
post-test probability	验后概率	102
precision	精确度	39
pretest probability	验前概率	102
<b>Q</b>		
qualitative research	定性研究	48
qualitative systematic review	定性系统评价	46
quality-adjusted life years, QALYs	质量调整寿命年	78, 138
quantitative systematic review	定量系统评价	46
<b>R</b>		
random effect model	随机效应模型	53
randomization	随机化	129
rapid diagnostic tests, RDTs	快速诊断性试验	96
rating scale	等级尺度法	78
recall bias	回忆性偏倚	128
receiver operator characteristic curve	ROC 曲线	100
recurrence rate	复发率	130
relative risk increase, RRI	相对危险增加率	113
relative risk reduction, RRR	相对危险降低率	39, 68, 111
relative risk, RR	相对危险度	39, 53, 54, 64, 124
remission rate	缓解率	130
reporting bias	报告偏倚	52
restriction	限制	129
Review manager, RevMan	系统评价管理软件	53
<b>S</b>		
safety	安全性	147
Scottish Intercollegiate Guidelines Network, SIGN	苏格兰学院间指南网络	156
selection bias	选择性偏倚	51
sensitivity analysis	敏感性分析	53, 62, 79, 143
sensitivity, SEN	敏感度	98
serial test	序列试验	104
side effect	药物副作用	118
specificity, SPE	特异度	98
standard gamble	标准博弈法	78, 141
standardized mean difference, SMD	标准化均数差值	43, 54
statistical heterogeneity	统计学异质性	53
subclinical stage	亚临床期	127
survival analysis	生存分析	130
survival cohort bias	存活队列偏倚	128
survival rate	生存率	130
suspected adverse drug reactions	可疑药物不良反应	120
synopses	证据摘要	18

systematic review,SR  
system

系统评价 5, 43, 46, 109, 115, 123  
证据系统 18

technical properties  
thalidomide incident  
The New Zealand Guideline Group, NZGG  
thesaurus  
time trade-off

## T

技术特性 147  
反应停事件 120  
新西兰指南研究组 156  
主题词表 23  
时间权衡法 78,141

user  
utility

## U

证据应用者 4  
效用 77, 137

validity

## V

真实性 66, 88, 131

web of causation  
wheel model  
willingness to pay, WTP

## W

病因网模型 87  
车轮模型 87  
支付意愿法 139

years lived with disability, YLD

## Y

健康寿命损失年 137

## 汉英名词索引

95% 可信区间	95% confidence interval, 95%CI	35
Cochrane 临床对照试验中心注册库	Cochrane Central Register of Controlled Trials, CENTRAL	21
ROC 曲线	receiver operator characteristic curve	100
<b>A</b>		
安全性	safety	147
<b>B</b>		
报告偏倚	reporting bias	52
背景问题	"background" questions	22
比值比	odds ratio, OR	39, 64, 68
标准博弈法	standard gamble	78, 141
标准化均数差值	standardized mean difference, SMD	43, 54
病例对照研究	case control study	85, 126
病死率	case-fatality rate	129
病因网模型	web of causation	87
病因学分数	etiologic fraction, EF	68
<b>C</b>		
测量	measurement	3
测量偏倚	measurement bias	128
车轮模型	wheel model	87
成本	cost	136
成本效果分析	cost-effectiveness analysis, CEA	139
成本效益分析	cost-benefit analysis, CBA	139
成本效用分析	cost-utility analysis, CUA	140
存活队列偏倚	survival cohort bias	128
<b>D</b>		
单个病人的随机对照试验	an individual randomized trial, n-of-1 trial	110
等级尺度法	rating scale	78
定量系统评价	quantitative systematic review	46
定性系统评价	qualitative systematic review	46
定性研究	qualitative research	48
队列研究	cohort study	90, 126
<b>F</b>		
反应停事件	thalidomide incident	120

- |             |   |            |
|-------------|---|------------|
| 方法学异质性      | methodological heterogeneity  | 53         |
| 非甾体类抗炎药     | non-steroidal anti-inflammatory drugs, NSAIDs                         | 169        |
| 分类变量        | categorical variable  | 33         |
| 复发率         | recurrence rate   | 130        |
| <b>G</b>    |   |            |
| 干扰          | co-intervention   | 111        |
| 固定效应模型      | fixed effect model  | 53, 56     |
| 观察性研究       | observational study   | 49, 90     |
| 归因危险度       | attribute risk, AR  | 68         |
| 国际卫生技术评估协会  | international society of technology assessment in health care, ISTAHC | 145        |
| 国际医学科学组织委员会 | Council for International Organization of Medical Sciences, CIOMS     | 119        |
| <b>H</b>    |   |            |
| 缓解率         | remission rate  | 130        |
| 患者预期事件发生率   | patient's expected event rate, PEER                                   | 114        |
| 回忆性偏倚       | recall bias   | 128        |
| 混杂因素        | confounders   | 90         |
| <b>J</b>    |   |            |
| 机会节点        | chance nodes  | 76         |
| 基本卫生技术      | essential health technology, EHT                                      | 146        |
| 基本药物目录      | essential medicines list, EML   | 146        |
| 急性冠脉综合征     | acute coronary syndrome, ACS  | 175        |
| 急性心肌梗死      | acute myocardial infarction   | 177        |
| 疾病负担        | burden of disease   | 160        |
| 集中性偏倚       | assembly bias   | 128        |
| 计算机决策支持系统   | computerized decision support system, CDSS                            | 18         |
| 技术特性        | technical properties  | 147        |
| 间接成本        | indirect costs  | 137        |
| 健康寿命损失年     | years lived with disability, YLD                                      | 137        |
| 教育处方        | educational prescription  | 13         |
| 结局          | outcome   | 127        |
| 金标准         | gold standard   | 97         |
| 精确度         | precision   | 39         |
| 决策节点        | decision nodes  | 76         |
| 决策树         | decision tree   | 75         |
| 绝对危险降低率     | absolute risk reduction, ARR  | 39, 68     |
| 均数差值        | mean difference, MD   | 43, 53, 54 |
| <b>K</b>    |   |            |
| 可信区间        | confidence interval, CI   | 39         |
| 可疑药物不良反应    | suspected adverse drug reactions                                      | 120        |
| 快速诊断性试验     | rapid diagnostic tests, RDTs  | 96         |
| <b>L</b>    |   |            |
| 滥用          | overuse   | 155        |
| 利弊比         | likelihood of help vs harm, LHH                                       | 69, 115    |
| 临床病程        | clinical course   | 127        |

- |               |   |                 |
|---------------|---|-----------------|
| 临床决策分析        | clinical decision analysis, CDA                     | 74              |
| 临床路径          | clinical pathway, CP                                | 159             |
| 临床期           | clinical stage                                      | 127             |
| 临床实践指南        | clinical practice guideline, CPG                    | 21, 154, 190    |
| 临床试验报告统一标准规范  | consolidated standards of reporting trials, CONSORT | 198             |
| 临床问题          | clinical question                                   | 10              |
| 临床异质性         | clinical heterogeneity                              | 53              |
| 流行病学三角        | epidemiology triangle                               | 87              |
| <b>M</b>      |   |                 |
| 美国国家指南交换库     | National Guideline Clearinghouse, NGC               | 156             |
| 美国内科医师学会杂志俱乐部 | ACP Journal Club                                    | 19              |
| 敏感度           | sensitivity, SEN                                    | 98              |
| 敏感性分析         | sensitivity analysis                                | 53, 62, 79, 143 |
| <b>N</b>      |   |                 |
| 内部真实性         | internal validity                                   | 30, 50, 66, 109 |
| <b>P</b>      |   |                 |
| 平行试验          | parallel test                                       | 104             |
| <b>Q</b>      |   |                 |
| 起始队列          | inception cohort                                    | 132             |
| 前景问题          | "foreground" questions                              | 22              |
| 缺失值           | missing value                                       | 35              |
| <b>S</b>      |   |                 |
| 森林图           | forest plot   | 53              |
| 生存分析          | survival analysis                                   | 130             |
| 生存率           | survival rate                                       | 130             |
| 生物学发病期        | biologic onset                                      | 127             |
| 生物学依据         | biologic evidence                                   | 67              |
| 失访偏倚          | lost to follow up bias                              | 128             |
| 失能调整生命年       | disability adjusted life year, DALY                 | 137             |
| 时间权衡法         | time trade-off                                      | 78, 141         |
| 实施偏倚          | performance bias                                    | 51              |
| 适用性           | applicability                                       | 109, 131        |
| 数值变量          | numerical variable                                  | 34              |
| 四分位间距         | interval of quartile range, IQR                     | 34              |
| 苏格兰学院间指南网络    | Scottish Intercollegiate Guidelines Network, SIGN   | 156             |
| 随访偏倚          | attrition bias                                      | 51              |
| 随机化           | randomization                                       | 129             |
| 随机效应模型        | random effect model                                 | 53              |
| <b>T</b>      |   |                 |
| 特异度           | specificity, SPE                                    | 98              |
| 贴现率           | discount rate                                       | 142             |
| 统计学异质性        | statistical heterogeneity                           | 53              |

## W

外部真实性	external validity	30, 50, 67
卫生技术	health technology	145
卫生技术评估	health technology assessment, HTA	145
卫生经济学评价	health economics evaluation	136

## X

系统评价	systematic review, SR	5, 43, 46, 109, 115, 123
系统评价管理软件	Review manager, RevMan	53
限制	restriction	129
相对危险度	relative risk, RR	39, 53, 54, 64, 124
相对危险降低率	relative risk reduction, RRR	39, 68, 111
相对危险增加率	relative risk increase, RRI	113
效果	effectiveness	137, 147
效力	efficacy	147
效益	benefit	137
效益成本比	benefit cost ratio, BCR	139
效用	utility	77, 137
新西兰指南研究组	The New Zealand Guideline Group, NZGG	156
需治多少病例才发生一例不良反应	number needed to harm, NNH	113
需治疗多少病例数才能获得一例最佳效果	number needed to treat, NNT	68, 112
序列试验	serial test	104
叙述性文献综述	narrative review	47
选择性偏倚	selection bias	51
循证医学	evidence based medicine, EBM	1
循证杂志摘要	evidence-based journal abstracts	18
循证指南	evidence-based guidelines	154

## Y

亚临床期	subclinical stage	127
严格评价	critical appraisal	123, 131
验后概率	post-test probability	102
验前概率	pretest probability	102
阳性似然比	positive likelihood ratio, +LR	99
阳性预测值	positive predictive value, +PV	98
药物不良反应	adverse drug reaction, ADR	68, 118
药物不良事件	adverse drug event, ADE	118
药物副作用	side effect	118
乙型病毒性肝炎	hepatitis B, HB	188
异质性检验	heterogeneity	53
意向性治疗分析	intention to treat analysis	51, 111
隐匿性随机分配	concealed randomization	110

## Z

增量成本效果比	incremental cost-effectiveness ratio, ICER	139
沾染	contamination	111
诊断比值比	diagnostic odd ratio, DOR	99
真实性	validity	66, 88, 131

证据提供者	doer	4
证据系统	system	18
证据应用者	user	4
证据摘要	synopses	18
支付意愿法	willingness to pay, WTP	139
直接成本	direct costs	136
直接非医疗成本	direct nonmedical costs	137
直接医疗成本	direct medical costs	137
指南研究与评价工具	appraisal of guidelines research and evaluation, AGREE	160, 161
质量调整寿命年	quality-adjusted life years, QALYs	78, 138
治愈率	cure rate	129
致残率	disability rate	130
中国知识基础设施工程	China National Knowledge Infrastructure, CNKI	21
中位数	median	34
中文生物医学期刊数据库	Chinese Medical Current Contents, CMCC	22
重要性	importance	109, 131
主题词表	thesaurus	23
准确度	accuracy, ACC	98
最佳证据资源	best-evidence resources	23
最小成本分析	cost-minimization analysis, CMA	138