

根据国家卫生计生委、中央编办、国家发展改革委、教育部、
财政部、人力资源社会保障部、国家中医药管理局联合发布的
《关于建立住院医师规范化培训制度的指导意见》编写

国家卫生和计划生育委员会住院医师规范化培训规划教材

循证医学

Evidence Based Medicine

主编 | 王吉耀 何 耀



人民卫生出版社
PEOPLE'S MEDICAL PUBLISHING HOUSE

根据国家卫生计生委、中央编办、国家发展改革委、教育部、
财政部、人力资源社会保障部、国家中医药管理局联合发布的
《关于建立住院医师规范化培训制度的指导意见》编写

国家卫生和计划生育委员会住院医师规范化培训规划教材

循证医学

- | | | |
|--------------------|----------------|--------------------|
| 1. 内科学 心血管内科分册 | 17. 外科学 神经外科分册 | 33. 医学遗传学 |
| 2. 内科学 呼吸与危重症医学科分册 | 18. 外科学 胸心外科分册 | 34. 预防医学 |
| 3. 内科学 消化内科分册 | 19. 外科学 泌尿外科分册 | 35. 口腔医学 口腔全科分册 |
| 4. 内科学 血液内科分册 | 20. 外科学 整形外科分册 | 36. 口腔医学 口腔内科分册 |
| 5. 内科学 肾脏内科分册 | 21. 骨科学 | 37. 口腔医学 口腔颌面外科分册 |
| 6. 内科学 内分泌科分册 | 22. 小儿外科学 | 38. 口腔医学 口腔修复科分册 |
| 7. 内科学 风湿免疫科分册 | 23. 妇产科学 | 39. 口腔医学 口腔正畸科分册 |
| 8. 内科学 感染科分册 | 24. 眼科学 | 40. 口腔医学 口腔病理科分册 |
| 9. 儿科学 | 25. 耳鼻咽喉头颈外科学 | 41. 口腔医学 口腔颌面影像科分册 |
| 10. 急诊医学 | 26. 麻醉学 | 42. 重症医学 |
| 11. 皮肤性病学 | 27. 临床病理学 | 43. 循证医学 |
| 12. 精神病学 | 28. 临床检验医学 | 44. 医学科研方法 |
| 13. 神经病学 | 29. 放射影像学 | 45. 医学伦理学实践 |
| 14. 全科医学 | 30. 超声医学 | 46. 医患沟通技能训练 |
| 15. 康复医学 | 31. 核医学 | 47. 住院医师英语手册 |
| 16. 外科学 普通外科分册 | 32. 肿瘤放射治疗学 | |

策划编辑 蔡向阳 刘彬 海珊
责任编辑 蔡向阳 樊京娜

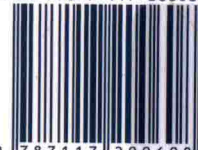
封面设计 李蹊
版式设计 单斯



人卫社官网 www.pmph.com 出版物查询, 在线购书

人卫医学网 www.ipmph.com 医学考试辅导, 医学数据库服务, 医学教育资源, 大众健康资讯

ISBN 978-7-117-20060-8



9 787117 200608 >

定价: 30.00 元

国家卫生和计划生育委员会住院医师规范化培训规划教材

循证医学

Evidence Based Medicine

主 编 王吉耀 何 耀

副主编 徐佩茹 祁艳波 王聪霞 王小钦

编 委 (按姓氏笔画排序)

王小钦 (复旦大学附属华山医院)

王吉耀 (复旦大学附属中山医院)

王聪霞 (西安交通大学医学院第二附属医院)

祁艳波 (齐齐哈尔医学院)

吴尚洁 (中南大学附属湘雅二医院)

何 耀 (中国人民解放军总医院)

罗 坤 (新疆医科大学第一附属医院)

姜林娣 (复旦大学附属中山医院)

徐佩茹 (新疆医科大学第一附属医院)

黄悦勤 (北京大学第六医院)

学术秘书 金雪娟 (复旦大学附属中山医院)

刘 森 (中国人民解放军总医院)

人民卫生出版社

图书在版编目 (CIP) 数据

循证医学 / 王吉耀, 何耀主编. —北京: 人民卫生出版社, 2014

国家卫生和计划生育委员会住院医师规范化培训规划教材

ISBN 978-7-117-20060-8

I. ①循… II. ①王…②何… III. ①临床医学—医师—职业培训—教材 IV. ①R4

中国版本图书馆 CIP 数据核字 (2014) 第 279696 号

人卫社官网	www.pmph.com	出版物查询, 在线购书
人卫医学网	www.ipmph.com	医学考试辅导, 医学数据库服务, 医学教育资源, 大众健康资讯

版权所有, 侵权必究!

循 证 医 学

主 编: 王吉耀 何 耀

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

购书热线: 010-59787592 010-59787584 010-65264830

印 刷: 北京铭成印刷有限公司

经 销: 新华书店

开 本: 850 × 1168 1/16 印张: 12

字 数: 330 千字

版 次: 2015 年 2 月第 1 版 2015 年 2 月第 1 版第 1 次印刷

标准书号: ISBN 978-7-117-20060-8/R · 20061

定 价: 30.00 元

打击盗版举报电话: 010-59787491 E-mail: WQ@pmph.com

(凡属印装质量问题请与本社市场营销中心联系退换)

出版说明

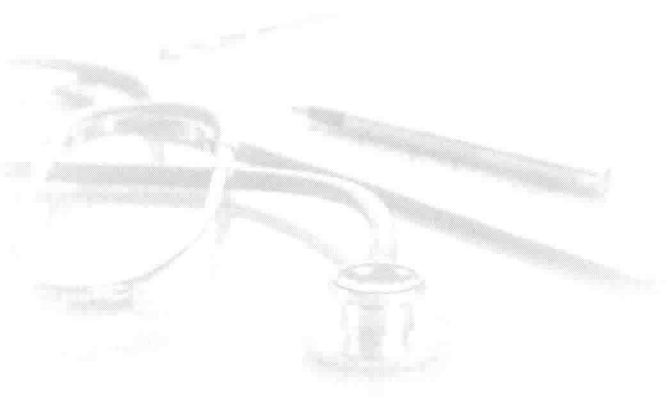
为深入贯彻国家卫生计生委、中央编办、国家发展改革委、教育部、财政部、人力资源社会保障部、国家中医药管理局联合发布的《关于建立住院医师规范化培训制度的指导意见》文件精神,满足全国各地住院医师规范化培训的要求,在国家卫生和计划生育委员会科教司领导和支持下,全国高等医药教材建设研究会、全国住院医师规范化培养教材评审委员会组织编写了《住院医师规范化培训规划教材》,人民卫生出版社正式出版。

本套教材的编写原则是:①坚持“三个对接”:与5年制的院校教育对接,与执业医师考试对接,与专科医师的准入和培训对接;②强调“三个转化”:在院校教育强调“三基”的基础上,本阶段强调把基本理论转化为临床实践、基本知识转化为临床思维、基本技能转化为临床能力;③强化“三个临床”:早临床、多临床、反复临床;④提高“四种能力”:职业道德、专业能力、人际沟通与团队合作能力、教学与科研的能力;⑤培养“三种素质”:职业素质、人文素质、综合素质;⑥实现“三医目标”:医病、医身、医心。不仅要诊治单个疾病,而且要关注患者整体,更要关爱患者心理。

本套教材强调“规范化”和“普适性”,实现培训过程与内容的统一标准和规范化。其中临床流程、思维与诊治均按照各学科临床诊疗指南、临床路径、专家共识及编写专家组一致认可的诊疗规范进行编写。在编写过程中不断地征集带教老师和学员意见并不断完善,实现“从临床中来,到临床中去”。本套教材的编写模式不同于本科院校教材的传统模式,注重体现PBL和CBL的教学方法,符合毕业后教育特点,并为下一阶段专科医师培训打下坚实的基础。

本套教材共47种。根据新近印发的《住院医师规范化培训内容与标准(试行)》的文件要求,分为临床学科(42种)、医学人文(5种)两类。本套教材充分考虑各学科内亚专科的培训特点,能够满足不同地区、不同层次的培训要求。

本套教材是在全面实施以“5+3”为主体的临床医学人才培养体系,深化医学教育改革,建设和建设一支适应人民群众健康保障需要的临床医师队伍的背景下组织编写的,希望全国广大住院医师培训基地在使用过程中提供宝贵意见。



国家卫生和计划生育委员会住院医师规范化培训规划教材

教材目录

序号	教材名称	主编	副主编
1	内科学 心血管内科分册	张 澍 霍 勇	陈 红 高海青 何 奔 周玉杰
2	内科学 呼吸与危重症医学科分册	王 辰 高占成	康 健 王 虹 李海潮 代华平
3	内科学 消化内科分册	唐承薇 张澍田	陈旻湖 房静远 陈卫昌 王蔚虹
4	内科学 血液内科分册	黄晓军 吴德沛	王健民 邵宗鸿 侯 明 卢振霞
5	内科学 肾脏内科分册	梅长林 余学清	陈江华 陈 楠 付 平 倪兆慧
6	内科学 内分泌科分册	童南伟 邢小平	郭晓蕙 肖海鹏 余学锋 陈 兵
7	内科学 风湿免疫科分册	张奉春 栗占国	鲍春德 刘 毅 毕黎琦 杨念生
8	内科学 感染科分册	魏 来 李太生	范学工 张文宏 党双锁 赵龙凤
9	儿科学	申昆玲 黄国英	母得志 薛辛东 罗小平 黄松明
10	急诊医学	于学忠 黄子通	陆一鸣 陈玉国 陈旭岩 张连阳
11	皮肤性病学	张学军 涂 平	徐金华 高兴华 陆前进 晋红中
12	精神病学	唐宏宇 方贻儒	李占江 刘铁桥 胡 建 贾福军
13	神经病学	贾建平 陈生弟	黄一宁 洪 震 周 东 唐北沙
14	全科医学	于晓松 季国忠	霍洪军 赵 钢 李双庆 王 敏
15	康复医学	励建安 黄晓琳	燕铁斌 何成奇 岳寿伟 吴 毅
16	外科学 普通外科分册	刘玉村 朱正纲	王 杉 胡三元 刘青光 程南生
17	外科学 神经外科分册	李新钢 王任直	赵世光 游 潮 刘建民 康德智
18	外科学 胸心外科分册	胡盛寿 王 俊	孙立忠 高长青 庄 建 肖颖彬
19	外科学 泌尿外科分册	叶章群 周利群	黄翼然 张小东 吴 斌 黄 翔



序号	教材名称	主编	副主编
20	外科学 整形外科分册	祁佐良 李青峰	郭树忠 王晓军 郭 澍 江 华
21	骨科学	裴福兴 陈安民	翁习生 阎作勤 林建华 贺西京
22	小儿外科学	孙 宁 郑 珊	冯杰雄 刘文英 高 亚 董 蓓
23	妇产科学	杨慧霞 狄 文	王建六 赵 霞 薛凤霞 漆洪波
24	眼科学	黎晓新 王宁利	许 迅 刘奕志 刘 平 沈 晔
25	耳鼻咽喉头颈外科学	韩东一 肖水芳	许 庚 唐安洲 张 榕 潘新良
26	麻醉学	刘 进 于布为	王国林 李文志 赵国庆 任家顺
27	临床病理学	陈 杰 步 宏	王连唐 李 挺 吴 强 戚基萍
28	临床检验医学	王 前 王建中	府伟灵 李 莉 续 薇 欧启水
29	放射影像学	郭启勇 王振常	胡道予 龚启勇 滕皋军 刘士远
30	超声医学	姜玉新 张 运	王金锐 田家玮 唐 杰 李建初
31	核医学	黄 钢 李亚明	李 方 王全师 石洪成 王 铁
32	肿瘤放射治疗学	王绿化 朱广迎	郎锦义 郭小毛 马 骏 刘晓冬
33	医学遗传学	邬玲仟 张 学	赵彦艳 张咸宁 余细勇 刘睿智
34	预防医学	朱启星 傅 华	张正东 王 彤 宿 庄
35	口腔医学 口腔全科分册	周学东 白玉兴	宋宇锋 刘洪臣 章锦才 徐 欣
36	口腔医学 口腔内科分册	凌均策 陈 智	孙 正 牛玉梅 俞立英 潘亚萍
37	口腔医学 口腔颌面外科分册	俞光岩 王慧明	王佐林 周 诺 胡勤刚 董福生
38	口腔医学 口腔修复科分册	周延民 陈吉华	高 平 陈 江 余占海 麻健丰
39	口腔医学 口腔正畸科分册	王 林 沈 刚	周 洪 邓 锋 毛 靖 王建国
40	口腔医学 口腔病理科分册	钟 鸣 王 洁	李铁军 陈 宇 周 峻 肖 晶
41	口腔医学 口腔颌面影像科分册	王铁梅 余 强	郑广宁 傅开元 程 勇 曾东林
42	重症医学	于凯江 杜 斌	管向东 王祥瑞 马晓春 康 焰
43	循证医学	王吉耀 何 耀	徐佩茹 祁艳波 王聪霞 王小钦
44	医学科研方法	陈世耀 刘晓清	张宏家 吕 明 肖志波
45	医学伦理学实践	邹和建 陈晓阳	纪宗正 张 欣 杨 薇 王兆良
46	医患沟通技能训练	李惠君 郭 媛	王 颖 刘惠军 韩新生 曹素艳
47	住院医师英语手册	唐熠达 冉志华	蔡世荣 潘 慧 金泽宁 李 刚

全国住院医师规范化培养教材

评审委员会名单

总 顾 问 赵玉沛 沈晓明 王德炳 刘德培 吴孟超
刘允怡 陈灏珠 钟南山 杨宝峰

顾 问 (按姓氏笔画排序)

冯友梅	武汉大学	赵 群	中国医科大学
吕兆丰	首都医科大学	柯 杨	北京大学
闫剑群	西安交通大学	桂永浩	复旦大学
步 宏	四川大学	陶立坚	中南大学
张英泽	河北医科大学	黄 钢	上海交通大学医学院
张学军	安徽医科大学	雷 寒	重庆医科大学
汪建平	中山大学	腾皋军	东南大学医学院
郑树森	浙江大学医学院	瞿 佳	温州医科大学

主 任 委 员 王 辰 陈贤义

副主任委员 (按姓氏笔画排序)

于晓松	中国医科大学第一附属医院	陈 红	北京大学人民医院
万学红	四川大学华西临床医学院 / 华西医院	陈旻湖	中山大学附属第一医院
王 虹	南京医科大学第一附属医院	林建华	福建医科大学附属第一医院
刘继红	华中科技大学同济医学院附属同济医院	唐安洲	广西医科大学附属一院
李 青	国家卫生和计划生育委员会医药卫生科技发展研究中心	郭启勇	中国医科大学附属盛京医院
杜 贤	人民卫生出版社	高海青	山东大学齐鲁医院
邵 洁	上海交通大学医学院附属瑞金医院	阎作勤	复旦大学附属中山医院



委 员 (按姓氏笔画排序)

于凯江	哈尔滨医科大学附属第二医院	陈 椿	福建医科大学附属协和医院
毛 颖	复旦大学附属华山医院	陈卫昌	苏州大学附属第一医院
王 前	南方医科大学南方医院	陈昕煜	国家卫生和计划生育委员会科技教育司
王以朋	北京协和医院	周玉杰	首都医科大学附属北京安贞医院
王共先	南昌大学第一附属医院	罗天友	重庆医科大学附属第一医院
占伊扬	江苏省人民医院	胡娅莉	南京大学医学院附属鼓楼医院
申昆玲	首都医科大学附属北京儿童医院	费广鹤	安徽医科大学第一附属医院
伍伟锋	广西医科大学第一附属医院	赵龙凤	山西医科大学第一临床医院
刘 彬	吉林大学第一医院	赵增仁	河北医科大学第一医院
刘建国	天津医科大学总医院	唐北沙	中南大学湘雅医院
刘青光	西安交通大学医学院第一附属医院	徐剑铖	第三军医大学第二附属医院 (新桥医院)
朱晒红	中南大学湘雅三医院	贾建国	首都医科大学宣武医院
汤宝鹏	新疆医科大学第一附属医院	贾明艳	北京医学教育协会
许 迅	上海市第一人民医院	高 亚	西安交通大学医学院第二附属医院 (西北医院)
吴一龙	广东省人民医院	高 炜	北京大学第三医院
张东华	哈尔滨医科大学附属第一医院	高长青	中国人民解放军总医院
张成普	中国医科大学附属盛京医院	诸葛启钊	温州医科大学附属第一医院
张学文	吉林大学中日联谊医院	龚启勇	四川大学华西临床医学院 / 华西医院
李占江	首都医科大学附属北京安定医院	董 蓓	青岛大学医学院附属医院
李海潮	北京大学第一医院	谢苗荣	首都医科大学附属北京友谊医院
沈 晔	浙江大学医学院附属第一医院		
狄 文	上海交通大学医学院附属仁济医院		
邱海波	东南大学附属中大医院		

主编简介



王吉耀

二级教授,复旦大学附属中山医院消化科主任医生,博士生导师,上海医学院内科学系主任、复旦大学临床流行病学培训中心和循证医学中心主任、复旦大学学术委员会委员。国际临床流行病学网常务理事、中国临床流行病学网主席、美国消化学会资深委员(AGAF)、美国肝病学会和国际流行病学协会会员。中华医学会临床流行病学分会前任主任委员。*Frontier of Medicine*、《复旦大学学报医学版》、《胃肠病学和肝病学》三本杂志的副主编。

从事教学工作45年,主编全国高等学校教材七年制《内科学》和国家十五、十一五规划教材7/8年制《内科学》,分别获全国和上海市优秀教材一等奖。主编《循证医学与临床实践》。负责的《内科学》列入国家级首批精品课程、《循证医学》为上海市精品课程。任《实用内科学》第14版主编。长期从事胃肠病及肝病的临床和研究,发表论文200多篇,SCI收录50多篇,包括发表于*Lancet*、*Hepatology*、*J Hepatology*。研究成果分别获国家科技进步二等奖(第二完成人)、上海市医学科技一等奖和三等奖、教育部科技进步二等奖、中华医学科技三等奖、上海市科技进步二等奖和三等奖(均为第一完成人)、上海市科技进步一等奖(第7完成人)等10余项。曾先后获得上海市三八红旗手、上海市优秀教育工作者、上海市高校教学名师、上海市高尚医德奖等荣誉。



何耀

现任解放军总医院老年医学研究所所长、北京市衰老与相关疾病研究重点实验室主任、全军保健医学重点实验室主任、香港大学公共卫生学院荣誉教授、中华预防医学会流行病学分会副主委、中国老年学会老年医学委员会副主委、全军流行病学专业技术委员会副主委、《中华流行病学杂志》和《中华保健医学杂志》副主编、国家疾病预防控制专家委员会成员兼老年卫生组副组长。

从事临床流行病学和循证医学教学工作已28年。在老年常见病、生活方式相关的代谢性疾病及烟草危害的流行病学研究、健康风险评估与健康风险管理、临床流行病学教学等领域取得多项研究成果,发表中英文论著200余篇,累计引用800余次。相关成果以其重要的公共卫生意义被收录WHO和美国卫生署的技术报告、英国医学年鉴、中国控烟报告和中国心血管病报告。曾获省部级科技进步一等奖一项、二等奖三项,国际心脏病联盟4年评选1名的“青年研究者奖”、陕西省首届“青年科技奖”、第九届“吴阶平医学研究奖”等学术奖励。

副主编简介



徐佩茹

教授,博士生导师,现任新疆医科大学第一附属医院副院长,新疆维吾尔自治区循证医学研究所所长,中华医学会儿科分会委员,中华医学会儿科分会呼吸学组睡眠障碍协作组副组长,中国医师协会循证医学专委会委员,循证医学教育部网上合作研究中心管理委员会委员,新疆医学会儿科分会主任委员。《中国循证医学杂志》编委。先后主持国家级等 10 余项课题。先后发表论文 86 篇。获宋庆龄儿科医学奖、自治区科技进步三等奖。



祁艳波

现任齐齐哈尔医学院公共卫生学院院长,黑龙江省医学会临床流行病学分会委员,齐齐哈尔市营养学会常务副理事长,齐齐哈尔市养生学会副理事长。从事教学工作 28 年。为省级精品课程《预防医学》、《医学统计学》主讲教师。曾获黑龙江省高等教育教学成果二等奖、齐齐哈尔医学院优秀教学成果奖、优秀教学质量奖、A 级教师等奖项。



王聪霞

医学博士,教授,博士生导师。西安交通大学医学院二附院心内科副主任。陕西省临床流行病学分会副主任委员。长期从事起搏、冠脉、射频消融等介入性诊疗工作。研究方向为冠心病、心律失常和心血管介入。国家科技奖励和自然科学基金评审专家。发表论文 92 篇,SCI 论文 16 篇;主持国家和省级课题 5 项;主持省厅级成果 4 项;主编专著 6 部,参编教材 2 部。



王小钦

医学博士,复旦大学附属华山医院血液科教授,博士生导师。现任复旦大学循证医学中心副主任,上海医学会临床流行病学分会候任主任委员,中华医学会临床流行病学分会委员和工作秘书,中华医学会血液学分会红细胞学组委员。研究方向为血液学、临床流行病学和循证医学。负责多项大型国际合作课题、国家级和省部级课题。发表论文 120 余篇,其中 SCI 论文 20 篇。主编和参编医学专著 10 部。获上海市科技进步三等奖和上海市医学奖三等奖。

前 言

循证医学是指导 21 世纪临床医生在作出临床诊疗决策时必须具备的理念和方法。为了培养住院医师的循证临床思维能力、临床循证实践能力和发现问题、分析问题与解决问题的能力,在国家卫生计生委住院医师规范化培训目标的指导下,由人民卫生出版社组织全国从事循证医学领域的专家,在广泛综合国际上权威的、最新的循证医学专著与教科书,并结合我国学者的有关论著的基础上,通过拟定编写提纲、共同编写、同行评议、交叉审阅、主编终审等环节,历时一年多,终于完成了本教材的编写工作。编者不仅是临床一线的专家,也是临床流行病学专家,其中主编、副主编更是我国循证医学领域的奠基人与开拓者。

本教材分三篇。内容包括循证医学的概论、方法与实践应用。通过案例引导学习,住院医师将掌握有效的搜索医学文献的方法;学习如何成为一个细致而又有批判性的读者;学习如何把新的研究证据运用到自己的临床实践中。“知识点”模块重点突出了每章的核心知识,以助记忆。通过本课程的学习,住院医师将掌握病因探索、诊断试验的准确性、治疗效果、疾病预后、药物的不良反应等临床实践常见问题的相关证据评估和解读;学习系统综述、临床指南评估的基本技能;了解临床决策和卫生政策决策的相关内容。最重要的是通过上述学习,住院医师将掌握目前所能获得的最佳证据结合临床知识和经验以及患者的需求,尽责、明白和明智地应用到临床实践中,使病人得到更好的照顾。

本书可以作为临床住院医师规范化培训的教科书,也可以作为各级临床医疗卫生人员学习循证医学的推荐读物。

作为主编,我们衷心感谢各位编委及学术秘书在本书的编写和出版过程中所作的努力。

由于受篇幅及编者水平的限制,书中难免存在不足之处,恳请读者不吝赐教,我们将十分感谢。

王吉耀 何 耀

2015 年 2 月

目 录

第一篇 总 论

第一章 循证医学概论.....	2
第一节 循证医学的概念.....	2
第二节 实施循证医学的步骤.....	4
第三节 循证证据的分级.....	7
第二章 根据临床问题获取最佳证据.....	11
第一节 提出需要解答的临床问题.....	11
第二节 常用的循证医学证据资源.....	13
第三节 获取证据的策略.....	16

第二篇 各 论

第一章 病因及危险因素.....	28
第一节 基本概念.....	28
第二节 评价原则.....	36
第二章 诊断试验.....	46
第一节 基本概念.....	46
第二节 评价原则.....	50
第三章 治疗效果.....	57
第一节 基本概念.....	57
第二节 评价原则.....	62
第四章 疾病预后.....	69
第一节 基本概念.....	69
第二节 评价原则.....	73
第五章 生命质量评价.....	79
第一节 基本概念.....	79
第二节 评价原则.....	81
第六章 临床经济学.....	88
第一节 基本概念.....	88



第二节 临床经济学评价方法	90
第三节 临床经济学文献评价原则	95
第七章 系统综述与 Meta 分析	98
第一节 基本概念	98
第二节 评价原则	104
第八章 循证实践指南	114
第一节 基本概念	114
第二节 评价原则	115

第三篇 循证实践

第一章 循证决策和决策分析	122
第二章 循证实践指南的临床应用	128
第三章 病因研究的循证实践	133
第一节 前瞻性队列研究原文剖析	133
第二节 前瞻性队列研究文献评价	137
第三节 前瞻性队列研究临床思维与决策	139
第四节 病例对照研究原文剖析	139
第五节 病例对照研究文献评价	142
第六节 病例对照研究临床思维与决策	144
第四章 诊断试验的循证实践	146
第一节 诊断试验原文剖析	146
第二节 诊断试验文献评价	149
第三节 诊断试验临床思维与决策	152
第五章 临床治疗的循证实践	153
第一节 Meta 分析的原文剖析	153
第二节 Meta 分析的文献评价	155
第三节 Meta 分析的临床思维与决策	156
第四节 生活质量的原文剖析	156
第五节 生活质量的文献评价	158
第六节 生活质量的临床思维与决策	158
第七节 临床卫生经济学原文剖析	158
第八节 临床卫生经济学文献评价	159
第九节 临床卫生经济学临床思维与决策	160
第十节 不良反应的循证实践	160
第六章 预后研究的循证实践	163
第一节 前瞻性队列研究原文剖析	163



第二节 前瞻性队列研究文献评价.....	166
第三节 前瞻性队列研究临床思维与决策.....	167
第四节 随机对照研究原文剖析.....	168
第五节 随机对照研究文献评价.....	171
第六节 随机对照研究临床思维与决策.....	172
中英文名词对照索引.....	174

第一篇

总 论



第一章 循证医学概论

临床问题

有一位81岁的老年男性患者,近一年来,快步行走或上楼时,有时会出现胸骨后压迫感,每次持续约数分钟,休息后自行缓解。检查发现静息心电图正常,心电图运动试验运动中未出现明显症状,运动后Ⅱ、Ⅲ、aVF ST段水平压低 $\geq 1\text{mm}$,诊断为慢性稳定性心绞痛。根据既往临床经验,医生建议在最佳药物治疗(药物治疗)的基础上,进行经皮冠脉介入治疗(PCI),以改善预后和缓解症状。患者不希望进行手术,但他同时关心如果仅仅接受药物治疗,近几年内是否会发生急性心肌梗死或死亡?在这种临床情境下,如果实践循证医学?

第一节 循证医学的概念

循证医学(evidence based medicine, EBM)是临床流行病学理论和方法学在临床医疗实践中的具体应用。循证医学的核心思想是:在充分考虑患者意愿的条件下,医务人员应该认真地、明智地、深思熟虑地把从科学研究中获得的最佳证据运用到临床决策。其内容可以概括为三个基本要素:①临床研究的最佳证据;②临床医生的经验与技能;③患者的基本价值观与愿望(图 1-1-1)。

数千年来,传统临床思维方式一直以“临床经验为基础”,临床决策往往根据医生自己的临床经验与直觉,以及听从专家的意见和权威参考书中的观点。由于经验推理与专家意见缺乏严谨的科学研究设计方法学的保证,其所得到的结论有时候会带有偏倚,这可能会导致错误的临床决策。传统的权威性参考书也存在弊端,如时间的滞后性,一些真正有效的

新疗法未被编录,而一些实际无效甚至有害的疗法,因从经验和理论上推论可能有效,被长期广泛地使用。循证医学是以“证据为基础”的临床医学思维方式,它更强调对发表的文献证据进行严格分级的前提下,利用最新、最可靠的证据解决具体的临床问题,并充分考虑患者的需求和意愿进行取舍。因此,循证医学不同于传统经验医学(表 1-1-1)。

每位临床医生都非常重视自己的临床经验,这些经过日积月累的辛勤劳动所积累的临床经验非常宝贵,忽视临床经验是一种对待循证医学不正确的态度。权威和专家在临床工作中积累的宝贵经验,使他们在病史、查体、诊断方面,具有敏锐的洞察力。对患者无偏倚的观察、准确的判断是循证医学的前提,没有来自患者正确的第一手信息就可能导致错误的诊断,在错误诊断引导下的最佳治疗证据没有任何价值。另外,缺乏临床实践经验的医生即使得到了最好的证据,也

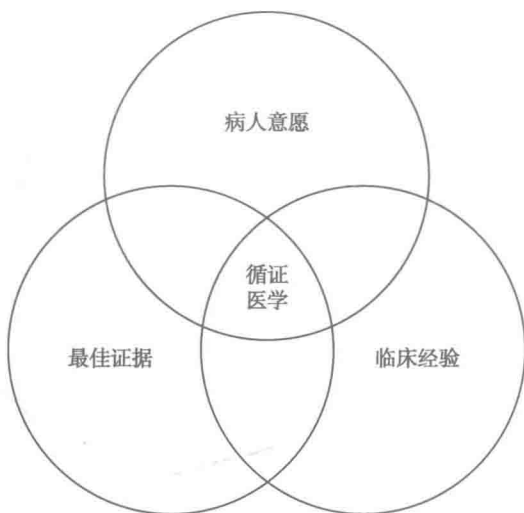


图 1-1-1 循证医学的概念



表 1-1-1 循证医学与传统经验医学的区别

项目	循证医学	传统经验医学
证据来源	强调系统全面的文献检索,并对相关研究进行科学的评价;强调证据分级;证据来源以多中心、大样本、随机双盲对照试验、前瞻性研究、科学的 Meta 分析为主	以临床经验为基础,高年资医生的指导、权威专家的意见、药品说明书、个人主观推理以及传统教科书与医学杂志上的零星研究报告
评价疗效的指标	患者的最终结局,如生存率、重要临床事件的发生率、致残率、生命质量、临床经济学指标等	通常以患者临床症状与体征的改善、实验室检查结果的变化等中间指标来评价疗效
临床决策的依据	对患者的诊治决策是结合当前可得到的最佳临床研究证据、医生的临床经验及患者的意愿进行取舍	对患者的诊治决策往往根据医生自己的临床经验与直觉,或者听从专家的意见和权威参考书中的观点
临床决策的规范性	有序,规范化	无序,不规范

可能不会准确地使用,因为最好的临床证据在用于每一个具体患者时,必须因人而异,需要结合临床资料进行取舍。循证医学在重视经验的同时,更强调临床医生必须掌握寻找、评价和利用医学证据的技能,如果缺乏最好、最新的外部证据指导,临床医生可能采用已经过时的旧方法,给患者造成损害。充分考虑患者的价值观与意愿,必须重视向患者提供医疗服务的证据和这些证据的说服力,这高度地体现了医生对生命的尊重。一位优秀的临床医生,应该既具备丰富的临床经验,又能依据现有最好的科学依据来指导临床实践,同时具有人文关怀的情怀,三者缺一不可。

目前,EBM 的理念和方法已经涵盖到所有的临床医学学科以及护理学、卫生事业管理、公共卫生、卫生决策等在内的各种学科,形成了以循证思维为主体的多个分支学科群。临床医师通过循证医学的学习,将学会如何批判性地评价不断涌现的新的医学研究信息,并把科学的研究证据运用到自己的临床实践中,以及将学会如何处理临床实践中的不确定性。临床医生实践循证医学,分为 3 个层次水平:①接受者(replicators):简单地接受专家关于已有的最佳证据的意见;②使用者(users):能够用高效率的方式检索到已有的最佳证据;③实践者(doers):能够进行文献评阅,系统综述和 meta- 分析。

循证医学推动临床医学思维方式发生了深刻变化,掌握信息(information mastery, IM)、批判性评价(critical appraisal, CA)和知识的转化(knowledge translation, KT)已经成为临床医生必须具备的 3 项技能。

本例临床情境中,根据临床经验,治疗策略是“药物治疗+PCI”。这位 81 岁的老年男性患者的意愿是“不做 PCI,仅采用药物治疗,再多活几年”。在循证医学实践中,医生首先需要寻求证据,以了解目前药物治疗与 PCI 手术治疗的预后(主要不良心脏事件,包括死亡、非致命性再梗死、靶血管再血管化治疗)是否存在差异。如果存在,有多大?然后应用证据,判断目前的研究证据是否可以应用到这位老年患者,在患者知情同意的基础上,进行治疗决策。

知识点

循证医学基本概念:循证医学是指在充分考虑患者意愿的条件下,医务人员应该认真地、明智地、深思熟虑地把从科学方法中获得的最佳证据运用到临床决策。

循证医学的三个基本要素:①临床研究的最佳证据;②临床医生的经验与技能;③患者的基本价值观与愿望。



第二节 实施循证医学的步骤

临床医生可以分 5 个步骤来实践循证医学：①提出临床问题；②寻找证据；③评价证据；④应用证据；⑤后效评价，即对实施结果进行追踪和再评估，修正错误，发现更好的方法。

一、确定一个需要回答的问题

提出一个恰当的临床问题十分重要，是循证的第一步，也是关键所在。一个构架完好的问题，可以帮助临床医生缩短寻找证据的时间，快速检索找到适合的证据，并且易于评价和应用。常见的临床问题包括：①病因：如何确定疾病的原因；②预防：如何通过确定和改变危险因素来降低疾病发生的机会，如何通过筛检早期诊断疾病；③诊断与鉴别诊断：为了肯定或排除某一诊断，在考虑待用诊断试验精确性、正确性、可接受性、费用、安全性的基础上，如何恰当选择诊断试验并解释其结果；④治疗：如何选择对患者有好处而无害处的治疗手段，从效果及花费来决定是否值得采用；⑤预后：如何估计患者可能产生的临床过程以及可能产生的并发症。

将在病因、预防、诊断、治疗、预后各方面的临床情况转换为一个来自具体患者的明确问题时，可采用国际上常用的 PICO 格式（表 1-1-2）。

表 1-1-2 临床问题的 PICO 格式

PICO	含义	内容
P population 或 participants	何种疾病或患病人群	何种疾病患者 何种年龄、性别、合并症人群等
I Intervention	干预措施	诊断 治疗 暴露(病因)等
C comparator 或 control	对比因素	无 安慰剂 其他治疗组 金标准等
O outcome	与患者相关联的结局	生存率 重要临床事件发生率 生命质量 临床经济学指标等

本例临床情境中，P= 慢性稳定性心绞痛老年男性患者；I= 单纯药物治疗；C= 药物治疗 + PCI；O= 主要不良心脏事件发生率。

二、寻找回答临床问题的最佳证据

（一）确定检索资源

互联网的发展使查找和检索原始文献很容易，大多数人习惯于使用一个或多个不同的搜索引擎，例如 Google、百度等。网络的主要优势是可以迅速更新信息，然而，互联网上的医学证据质量可能存在广泛的问题，认识到这一点，至关重要。对中国医生来说，目前最简单和经济的方法是使用循证医学网上信息资源，循证信息资源帮助临床医生整合所有的最好证据以解决一个特定的临床问题，例如提供治疗证据的 Cochrane 图书馆的系统评价、ACP Journal Club（其从 100



余种生物医学刊物中,按循证医学文献要求选择论著,对其进行摘要,并对其文献的临床应用价值进行评论)和 TRIP Database (其同时显示来自默克诊疗手册、治疗指南、Cochrane 图书馆的 Meta 分析、PubMed 中原始研究的相关文献,并直接与这些资料库进行链接,得到摘要或全文)等均可提供大量的循证信息。另外,临床实践指南(National Guideline Clearinghouse)集中了全世界不同国家所制定的治疗指南,直接将证据应用于指南中,并对证据进行了评价定级。我国中华医学会各学科分会颁布的国人特定疾病的临床实践指南也是提供循证医学证据的一类重要资源,可作为临床医生处理临床问题的参考性文件。如果查到新近的、标有证据级别、高质量的相关临床指南,可以省去文献评价的环节。

(二) 制订检索策略

循证医学资源可划分为 6 个等级:①原始研究(例如 MEDLINE);②循证研究摘要(例如循证医学杂志的摘要);③系统评价(例如 Cochrane 系统综述);④循证证据提要(例如 ACP Journal Club);⑤证据综合(例如循证医学教科书);⑥计算机决策支持系统。而获取最佳证据的快捷途径则应从等级资源的最高层开始,避免阅读不必要的文献。由于等级资源的最高层证据资源的不完善性,而且同一类资源库又有多种,逐个检索将费时而烦琐,可以使用跨文献库的检索引擎,例如 TRIP Database 可以同时检索多种文献库。事实上,并没有单一的最佳证据来源,重要的是要了解各种可用的高质量、“最佳”的证据来源以及他们的相对优点和缺点。临床医生必须具备证据检索的基本技能(见第二章)。

本例临床情境中,该医生首先查询了 Cochrane 图书馆的系统评价,没有查询到新近的相关系统综述。通过 TRIP Database 找到 2007 年发表在新英格兰医学杂志(N Engl J Med)上的一项对比 PCI 与强化药物治疗改善稳定性心绞痛患者预后的 RCT——COURAGE 研究。另外,通过中国生物医学文献数据库(CBM),找到中华医学会心血管病分会 2007 年 3 月颁布的中国《慢性稳定性心绞痛诊断与治疗指南》。

三、评价证据

对得到的证据进行真实性(validity)评估,理解和掌握这些研究的证据,才能应用到临床实践。真实性指一项研究产生的结论的正确性及可靠度,即所得的结果是否反映了欲进行研究结果的真实情况。真实性包括内部真实性和外部真实性。内部真实性是临床研究能正确反映研究人群中实际应该产生的研究结果,与研究设计的科学性相关。如果具有内部的真实性的结果推广到研究人群以外的同类人群仍有效,则该项研究结果具有外部真实性,其常与样本的选择有关。

通过循证医学信息资源寻找得到的证据,通常都已经进行了评价定级,可以根据不同证据水平与推荐级别(见本章第三节)直接应用到临床实践。现今,每天有超过 2000 篇新的医学研究论文出版,新的研究信息涌入意味着现代医疗实践在不断地飞速改变和发展。这就要求临床医生能够熟练地浏览医学的新进展,批判性地看待研究报告,客观地、快捷地从浩瀚的医学信息海洋中提炼出有科学性的最佳证据。根据临床问题完成证据检索后,可针对研究要素对找到的文献进行快速评估,当评阅一项研究时,有必须回答的几个关键问题,如果答案是否定的,那么它很少值得进一步阅读。表 1-1-3 列出了快速评阅不同临床问题证据真实性评估需要回答的关键问题。



表 1-1-3 快速评阅文献需要回答的关键问题

临床问题类型	关键问题
疗效	1. 研究类型为系统综述或 RCT 吗? 2. 是否实行分配隐藏和随机分组? 3. 对患者的分组,医师和患者是否双盲? 4. 除了需要评估的治疗措施外,两组是否得到相同的干预? 5. 被研究患者的随访是否完整? 失访率有多高? 6. 资料总结分析是否采用 ITT 分析?
不良反应	1. 研究类型是系统综述、RCT、队列研究、病例-对照研究,或与临床问题患者特征相同的单个患者随机对照试验(n-of-1)吗? (在不良反应研究中,RCT 研究比较少用,因为常常没有足够的样本量。) 2. 患者是否除了特定的治疗不同以外,其他所有重要方面的特征都相近? 3. 是否用同样的方式测量了两组的治疗和不良反应状况? 4. 是否客观地评价了不良反应? 评价不良反应是否采用了盲法? 5. 随访是否足够长并且完整? 6. 关于不良反应的研究结果是否满足了能确定因果关系的原则?
预后	1. 研究类型为系统综述或队列研究吗? 2. 患者是否具有代表性? 3. 随访是否足够长并且完整? 4. 失访是否严重? 失访的原因是什么? 5. 预后指标的定义是否明确? 其测量有无偏倚? 6. 比较时是否控制了混杂因素?
诊断与筛查	1. 研究类型是系统综述或横断面研究吗? 2. 患者是否具有代表性? 3. 试验是否与“金标准”试验进行了独立、“盲法”的比较? “金标准”的使用是否合理? 4. 是否每个被测者都做参照试验进行评价? 5. 所研究患者样本是否包括临床试验中将使用该诊断试验的各种患者? 6. 是否提供了准确性指标估计值的精度?

本例临床情境中,COURAGE 研究从证据来源看,新英格兰医学杂志是影响因子最高的医学期刊。按以下问题快速评阅该研究的真实性:

1. 研究类型为 RCT 吗? 是
2. 是否实行隐藏分组和随机分组? 是
3. 对患者的分组,医师和患者是否双盲? 是
4. 除了需要评估的治疗措施外,两组是否得到相同的干预? 是
5. 被研究患者的随访是否完整,失访率有多高?

PCI 组 107 例失访(107/1149,9.3%),药物组 97 例失访(97/1138,8.5%),两组没有差异。在随访期中,有药物治疗组 1/3 的患者转入 PCI 组,可能给结果的判断带来偏倚。

6. 资料总结分析是否采用 ITT 分析? 是,研究结果报告有精确的 HR(95%CI)。在平均 4.6 年的随访中,全因死亡与非致命 MI 这两项主要终点事件,PCI 与仅接受医药治疗的分组之间并无差异(20.0% vs 19.5%;HR=1.05;95% CI=0.87~1.27;P=0.62)。

快速评阅初筛证据的真实性结果,可以认为该研究报告值得精读。



四、把得到的证据应用到自己的患者

通过证据找寻与证据评价,您是否找到了答案?您找到的答案是否是高质量的证据?结果的作用大小和精确性(95%可信区间)如何?是否有临床意义?你的患者是否与研究证据中纳入的患者特征相似(考虑诊断标准、纳入和排除标准、临床特征等)?研究中的干预措施在当地可行性如何?是否受到条件限制(技术、经费、观念)?您能否对此作定量的评价?您的患者和家属同意您的意见吗等等,将这些信息与整合临床专业知识与患者的期望进行整合,作出最佳的临床决策。

本例临床情境中,患者符合 COURAGE 研究其入组标准,但 COURAGE 研究中,患者来自北美,从我国的临床证据看,2007 年 3 月颁布的中国《慢性稳定性心绞痛诊断与治疗指南》中,指出对于低危的稳定性心绞痛患者,包括强化降脂治疗在内的药物治疗在减少缺血性事件方面与 PCI 一样有效。对于相对高危患者及多支血管病变的稳定性心绞痛患者,PCI 缓解症状更为显著,生存率获益尚不明确。因此,综合临床经验,患者的意愿和 COURAGE 研究的证据,帮助患者根据他的价值观和意愿谨慎权衡利弊,推荐了 COURAGE 研究中采用的方案:指导患者改变生活方式(戒烟、运动)为最重要的基本治疗,充分使用循证药物,并密切关注症状是否转为强烈,以确保在仅使用最佳药物治疗未能充分控制症状的前提下,为其提供早期血管重建。

五、最后一步就是随访效果评价

不确定性是临床实践中不可避免的难题之一。在将证据与患者的个体因素结合作出医疗决策并予以实施后,更要对整个医疗过程进行总体评价,以提高下一次循证医疗的水平,在这个过程中要注意结果评定的客观性。随着循证医学的发展对于证据强度的认识有一个渐进的过程。随着医疗水平的发展,新干预措施的出现,最佳证据也会随着时代更新。

本例临床情境中,患者接受最佳医药治疗,通过专科门诊随访。在最佳药物治疗法单独使用一年后,并未出现明显负面影响,例如出现不稳定心绞痛或心肌梗死。药物治疗作为第一治疗策略仍然合理。

知识点

实施循证医学的步骤:循证医学可以分 5 个步骤来实施:①提出合适的临床问题,采用国际上常用的 PICO 格式;②寻找证据:获取最佳证据的快捷途径应从循证医学证据等级资源的最高层开始,无法解决则转向下一层;③评价证据:根据临床问题完成检索后,可针对研究要素对寻找到的文献进行快速评估;④应用证据:需要考虑患者是否与研究证据中纳入的患者特征相似,诊断试验或治疗是否受到技术、经费、患者的观念条件限制等对临床问题做出诊疗决策;⑤后效评价:随访患者,并对效果进行评价。

第三节 循证证据的分级

明确区分对待不同来源的证据是循证医学的重要内容之一。循证证据的分级包括证据水平(level of evidence, LOE)和推荐级别(class of recommendation, COR)两个方面。证据水平是对

某项干预措施证据强度的精确估计。推荐级别除需要考虑是否获益的证据强度外,还需要综合考虑风险和可能造成的损害大小,以及分别从临床医生、患者、政策制定者角度谨慎权衡利弊,作明确实用的诠释。高质量证据不一定意味着强推荐。

不同的临床问题,具有不同的特性,需要特定的标准来区分证据;而且循证医学对于证据强度的认识有一个渐进的过程,因此在循证医学文献中,常出现不同的证据质量和推荐强度分级标准。尽管不同的证据分级系统之间有差异,但它们的目相同:按质量从高到低对证据进行分级,帮助临床医生认识证据是否为高质量,干预措施是否利大于弊或利弊相当,还是利弊关系不确定。

目前,被国际上广泛接受和使用的证据等级划分标准主要来自牛津大学循证医学中心(Oxford Centre for Evidence-based Medicine)在2001年制定的证据等级标准,以及在2004年“推荐分级的评估、制定与评价”工作组推出的将各个分级标准综合而形成的GRADE标准(Grading of Recommendation Assessment, Development and Evaluation, GRADE)。

一、牛津循证医学中心的证据水平与推荐级别

牛津循证医学中心的证据水平分为5级,推荐建议分为A、B、C、D四级(表1-1-4)。

表 1-1-4 牛津循证医学中心证据水平分级及推荐依据

推荐级别	证据水平	临床问题(病因、治疗、预防)	临床问题(预后)
A	1a	同质性良好的 RCTs 系统综述	同质性良好的队列研究系统综述
	1b	95% 可信区间较窄的单项 RCT	单项起点一致的队列研究,随访率 >80%
	1c	全或无(传统治疗全部无效)	系列病例报告全部死亡或全部生存
B	2a	同质性良好的队列研究的系统综述	对回顾性队列或 RCT 空白对照组的系统综述
	2b	单项队列研究及质量差的 RCT	单项回顾性队列或 RCT 空白对照组的随访
	2c	结局研究	结局研究
	3a	同质性良好的病例对照研究的系统综述	
	3b	单项病例对照研究	
C	4	系列病例分析或质量差的病例对照研究、系列病例报告、质量差的队列研究(例如随访率 <80%)	
D	5	没有分析评价的专家意见或在病理生理基础上的意见	

A 级:证据极有效,推荐;B 级:证据有效,可推荐,也可能在将来更高质量的新证据出现而改变;C 级:证据在一定条件下有效,因谨慎应用研究结果。D 级:证据的有效性局限,只在较窄的范围内有效

二、GRADE 证据水平与推荐级别

GRADE 系统将证据质量分为高、中、低、极低四级,见表 1-1-5。GRADE 系统中,就干预措施效果而言,最可靠的证据来自多个 RCT 的系统综述,其次是单个 RCT。虽然基于 RCT 得出的证据一开始被定为高质量,但 RCT 的证据可能会因为下面 5 个因素而降低:研究的局限性、研究结果不一致、间接证据、结果不精确以及报告有偏倚。观察性研究(如队列研究和病例对照研究)一开始被归为低质量,但若某干预措施疗效显著;证据显示存在剂量效应关系;存在各种可能导致疗效显著性降低的偏倚时,观察性研究证据的等级将可能提高。

GRADE 系统推荐级别只分为“强”、“弱”两级。当明确显示干预措施利大于弊或弊大于利时,将其列为强推荐。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。在推荐强度分级时,GRADE 除考虑证据质量外,其他一些因素也会影响推荐意见的强弱,包括患者价值观和意愿的不确定性或可变性,尽管该证据质量高,但不同价值观和意愿的患者选择



表 1-1-5 GRADE 系统证据质量及其定义

证据级别	定义
高质量 (A)	1. 进一步研究也 <u>不可能改变</u> 该疗效评估结果的可信度 2. 我们 <u>非常确信</u> 真实的效应值接近效应估计值
中等质量 (B)	1. 进一步研究很可能影响该疗效评估结果的可信度,且可能改变该评估结果 2. 对效应估计值我们有 <u>中等程度</u> 的信心:真实值有可能接近估计值,但仍存在二者很不相同的可能性
低质量 (C)	1. 进一步研究极 <u>有可能</u> 影响该疗效评估结果的可信度,且该评估结果很可能改变 2. 我们对效应估计值的 <u>确信程度有限</u> :真实值可能与估计值大不相同
极低质量 (D)	1. 任何疗效评估结果都很 <u>不确定</u> 2. 我们对效应估计值 <u>几乎没有信心</u> :真实值很可能与估计值大不相同

1. 早前定义;2. 当前定义

会有不同,就不能确定为强推荐。另外,还考虑到资源的合理利用,推荐措施的获益程度与风险以及成本比。例如,预防短暂缺血性脑卒中患者中风复发,阿司匹林成本低,为强推荐;氯吡格雷联合阿司匹林成本高,为弱推荐。

本例临床情境中,按照牛津循证医学中心标准,证据水平为 1b,推荐级别为 A 级。按照 GRADE 系统标准,证据水平为中等高质量。从疗效看,COURAGE 研究证明对于稳定性心绞痛患者,在标准的药物治疗基础上,再加上介入治疗并不降低死亡率与心肌梗死发生率;从利弊关系看,标准的药物治疗使用方便、成本低;从患者的价值观和意愿看,稳定性心绞痛老年患者可能更重视 PCI 手术的风险而非其延寿作用;因此考虑推荐级别为“强”。

另外,参考美国心脏病学会基金会 (ACCF) 和美国心脏协会心血管病诊断与治疗指南的证据分级,COURAGE 研究证据水平为 B 级和推荐级别为 I 类。

知识点

证据水平:证据水平是对某项干预措施证据强度的精确估计。
推荐级别:推荐级别除需要考虑是否获益的证据强度外,还需要综合考虑风险和可能造成的损害大小,以及分别从临床医生、患者、政策制定者角度谨慎权衡利弊,作明确实用的诠释。高质量证据不一定意味着强推荐。
常用的证据等级划分标准有牛津大学循证医学中心的证据等级标准以及 GRADE 系统标准。

思考题

- 简述实施循证医学的步骤。
- 下列哪项不属于循证医学的要素
A. 最佳证据
B. 临床情况
C. 患者的需求和价值观
D. 专家意见
E. 临床经验
- 有关实施循证医学的步骤的描述,哪项是不正确的
A. 构建临床问题,应该采用 PICO 格式

学习
笔记



- B. 最佳证据的来源是临床实践指南
- C. 不能通过 Google 等搜索引擎寻找最佳证据
- D. 证据质量分级高的证据不一定是强推荐
- E. RCT 得出的证据也可能存在局限性

参考答案

2. D 3. B

(王吉耀 金雪娟)

参考文献

1. 王吉耀. 循证医学与临床实践. 第3版. 北京:科学出版社,2012.
2. Evidence-Based Medicine Working Group. Evidence-based medicine. A new approach to teaching the practice of medicine. JAMA, 1992, 268 (17):2420-2425.
3. Straus SE, Glasziou P, Richardson WS, Haynes RB. Evidence-based-medicine. How to practice and teach EBM. 4th ed. Edinburgh: Churchill Livingstone, Elsevier, London, 2011.
4. Purcell GP, Wilson P, Delamothé T. The quality of health information on the Internet. BMJ, 2002, 324: 557-558.
5. Haynes B. Of studies, syntheses, synopses, summaries, and systems: the “5S” evolution of information services for evidence-based healthcare decisions. Evid Based Nurs, 2007, 10 (1): 6-7.
6. Dicenso A, Bayley L, Haynes RB. Accessing pre-appraised evidence: fine-tuning the 5S model into a 6S model. Evid Based Nurs, 2009, 12 (4): 99-101.
7. Guyatt GH, Oxman AD, Vist GE, et al. GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. BMJ, 2008, 336 (7650): 924-926.
8. Mayer D. Essential Evidence-based Medicine. 2nd ed. UK: Cambridge University press, 2010.
9. Rothman SM. Health advocacy organizations and evidence-based medicine. JAMA, 2011, 305 (24): 2569-2570.
10. Boden WE, O'Rourke RA, Teo KK, et al. Optimal medical therapy with or without PCI for stable coronary disease. N Engl J Med, 2007, 356: 1503-1516.
11. 中华医学会心血管病分会, 中华心血管病杂志编辑委员会. 中国慢性稳定性心绞痛诊断与治疗指南. 中华心血管病杂志, 2007, 35 (3): 195-202.
12. Yancy CW, Jessup M, Bozkurt B, et al. 2013 ACCF/AHA guideline for the management of heart failure: a report of the American College of Cardiology Foundation/American Heart Association Task Force on Practice Guidelines. J Am Coll Cardiol, 2013, 62 (16): e147-239.
13. Balshem H, Helfand M, Holger J, et al. GRADE guidelines: 3. Rating the quality of evidence. Journal of Clinical Epidemiology, 2011, 64: 401-406.
14. <http://www.gradeworkinggroup.org/society/index.htm>

第二章 根据临床问题获取最佳证据

临床问题

医疗过程中,临床医师、患者本人以及其家属等决策者经常会因为知识的局限性提出关于疾病的各种各样的临床问题。例如:抑郁会增加绝经后妇女骨折的风险吗?阿司匹林对急性心肌梗死患者有害吗?在互联网高速发展的当代,患者也会关注网上新闻和医学的报道,例如一名正在服用孟鲁司特的运动型哮喘患儿父母可能会疑惑“孟鲁司特是否是治疗运动型哮喘儿童的最好药物?”本章重点阐述如何根据临床问题获取最佳证据。

第一节 提出需要解答的临床问题

在临床诊疗中,病人、住院医师和专家都会提出各种临床问题。临床问题根据问题提出不同,分为病因、诊断、治疗和预后四种类型。运用 PICO 原则构建临床问题是将初始临床问题转变为医学科研可以回答的临床问题,有利于住院医师就某一临床问题进行文献检索。运用 PICO 原则构建临床问题是循证临床实践的第一步,为循证临床实践提供方向。住院医师对病人诊治过程就是一个不断提出问题、寻找证据、解决问题的过程,所以,住院医师恰当运用 PICO 原则构建临床问题,在循证临床实践中至关重要。

临床问题的构建运用 PICO 原则,包括 4 个基本成分,见图 1-2-1。

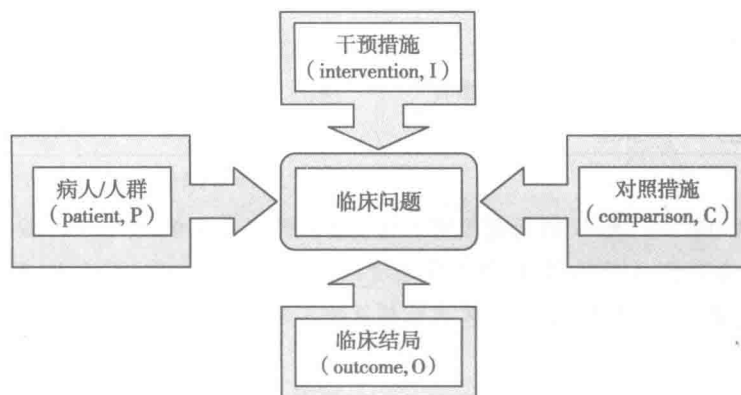


图 1-2-1 临床问题构建的 PICO 原则

病人/人群 (patient/population, P): 患者或者人群及其临床特征。包括与需要解决的临床问题密切相关的患者或者与患者相似的人群的临床特点。

干预措施 (intervention or exposure, I): 关注的处理措施,如暴露因素、药物、外科手术、诊断试验、预后因素等。

对照措施 (comparison, C): 与干预措施或者暴露因素相比较的措施,在治疗问题上,多为常规治疗、安慰剂作为对照措施,在诊断性问题上,对照措施常是诊断某种疾病的金标准。

临床结局 (outcome, O): 住院医师关注的暴露或处理措施导致的患者相关结局,多为远期指



标,如死亡率、功能改善、不良反应发生率等。

病因问题的构建。患者,女性,60岁,股骨干骨折1天。近两年来,常有腰酸背痛、乏力等症状。一年前,曾发生肱骨骨折,已治愈。半年前,曾在社区医院因情绪低落诊断为抑郁症。无神志障碍,无继发性骨质疏松症和代谢性骨病,已自然绝经。

原始问题:骨折和抑郁症有关系吗?

剖析问题:根据询问病史了解到,病人多次骨折,且有抑郁症史,相关文献报道绝经后妇女的抑郁和其骨密度有一定关系。且病人60岁,已自然绝经,符合人群特征。

构建问题:抑都会增加绝经后妇女骨折的风险吗?(表1-2-1)

表 1-2-1 构建后的病因问题

PICO	信息	PICO	信息
患者(P)	绝经后妇女	对照因素(C)	没有抑郁症
暴露因素(I)	抑郁症	结局(O)	骨折发生率

诊断问题的构建。患儿13岁,因听力障碍来医院耳鼻喉科检查,现确诊为感音神经性耳聋。需进一步检查确诊是否是内耳畸形。

原始问题:需要进行什么检查才能确诊内耳畸形?

剖析问题:据临床经验,先天内耳畸形是导致感音神经性耳聋的主要原因,目前为止,只通过影像学检查无法确诊,但是MRI可以判定内耳骨迷路正常的病人是否伴有听神经发育异常。故考虑选用MRI检查来初步诊断内耳畸形。

构建问题:MRI诊断感音神经性耳聋患者为内耳畸形的把握度有多大?(表1-2-2)

表 1-2-2 构建后的诊断问题

PICO	信息	PICO	信息
患者(P)	感音神经性耳聋患者	对照措施(C)	金标准
干预措施(I)	MRI	结局(O)	内耳畸形

治疗问题的构建。患儿20天前无明显诱因出现阵发性咳嗽,有痰不易咳出,清晨及下午多见,无气喘气憋,无呼吸困难,患儿神志清,精神欠佳,饮食及睡眠欠佳,否认异物吸入史,无发热,无腹泻,无呕吐,无声嘶,无犬吠样及痉挛性咳嗽。

原始问题:用什么药治疗?

剖析问题:根据案例,了解到治疗的对象是运动型哮喘患儿,通过查证得知孟鲁司特在治疗运动型哮喘患儿的疗效是有效的,而患者和住院医师更多关注的是孟鲁司特和其他哮喘用药(常规治疗)相比,孰优孰劣?预后方面,患儿家属十分关心通过治疗后患儿是否会复发。

构建问题:孟鲁司特与常规治疗相比,能降低运动型哮喘患儿喘息的复发率吗?(表1-2-3)

表 1-2-3 构建后的治疗问题

PICO	信息	PICO	信息
患者(P)	运动型哮喘患儿	对照措施(C)	常规治疗
干预措施(I)	孟鲁司特	结局(O)	喘息的复发率



预后问题的构建。患者 60 岁,男性,一天前出现头昏、右侧肢体麻木、无力,症状持续加重,3 小时前出现口角歪斜,右侧肢体无力加重行走困难,左侧视力下降,伴有意识不清。

原始问题:抗血小板药对脑卒中患者有效吗?

剖析问题:根据临床诊疗经验,抗血小板药对改善急性缺血性脑卒中患者有一定的帮助,但是对于患者而言,更多的是关注急性缺血性脑卒中后,未来的生存率和致残率等预后结局。

构建问题:抗血小板药与不用抗血小板药物相比能改善急性缺血性脑卒中患者的生存率和致残率吗?(表 1-2-4)

表 1-2-4 构建后的预后问题

PICO	信息	PICO	信息
患者(P)	急性缺血性脑卒中患者	对照措施(C)	不用抗血小板药
干预措施(I)	抗血小板药	结局(O)	患者的生存率和致残率

知识点

临床问题构建的 PICO 原则:临床问题的构建的 PICO 原则,包括以下 4 个基本成分:①病人/人群(patient/population,P);②干预措施(intervention or exposure,I);③对照措施(comparison,C);④临床结局(outcome,O)。

第二节 常用的循证医学证据资源

目前,大量循证医学证据资源涌现,高质量的临床研究、系统综述、文献摘要、循证指南、信息系统等日趋增加和完善,住院医师可以就自己感兴趣的问题,检索并利用最新的信息,为临床诊疗服务提供证据。但是,住院医师繁忙的工作与浩瀚的医学文献的海洋之间形成了矛盾,使住院医师难以跟上知识信息时代的发展步伐。通常表现为住院医师没有足够的精力花费大量的时间查寻和筛选所需信息。所以,目前如何才能从医学文献中高效和准确地获取自己需要的最佳的科学证据,是摆在临床医师面前的一个非常重要的问题。

循证医学证据资源证据划分为 6 个等级,简称为“6S”等级结构(图 1-2-2)。住院医师可以根据需要,通过获得的最高级别的信息资源来寻找临床实践证据,开展循证临床实践。“6S”更多地考虑了临床医生的时间、工作条件和环境等,把循证临床实践理念和循证临床实践操作者的实际状况相结合。“6S”是循证临床实践的现实模式,也是对循证医学未来发展的展望。

一、计算机决策支持系统

计算机决策支持系统(systems)是循证医学证据资源的最高等级,它勾勒了一个

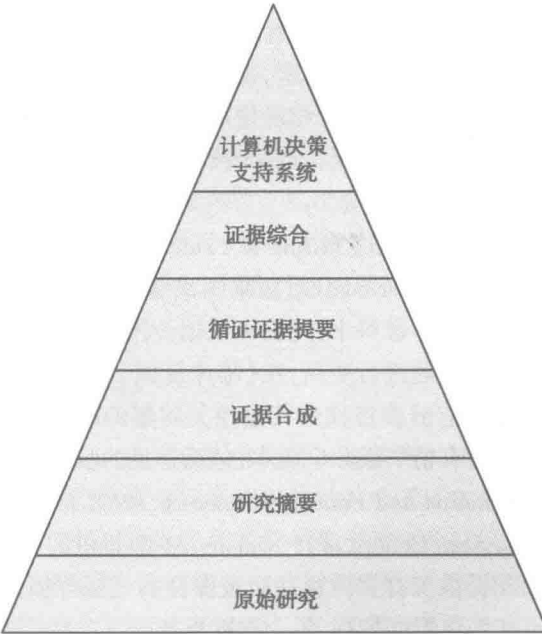


图 1-2-2 循证医学证据资源“6S”等级结构图

学习
笔记



超速、便捷的计算机医学信息的人机对话系统。一个完美的计算机决策支持系统包含完整且精准的医学信息。通过住院医师和计算机的简答互动过程,能够解答住院医师在临床中遇到的各种重要临床问题。如住院医师在循证临床实践时,通过给予计算机决策支持系统某种链接或者命令,把临床问题输入计算机决策支持系统,计算机决策支持系统获得指令后,通过搜索医学信息,如电子病历和临床研究,最终给出最佳的临床诊断决策,为住院医师循证临床实践提供证据。简单来说,我们可以认为使用计算机决策支持系统的过程是一个标准的医学信息查询过程。

同样,计算机决策支持系统也可以在特定的时间向住院医师发出某些提示性信息,如病人病历记录书写错误及更改建议。住院医师可以通过查询计算机决策支持系统,以便能够迅速获取相关问题的医学知识的更新内容。住院医师在撰写系统综述时,也可以通过计算机决策支持系统查找医学信息,并进行严格的质量评价。患者也可以通过计算机决策支持系统的使用,及时了解自己的疾病的诊断、治疗和预后。通过计算机决策支持系统,可以缓解医疗信息不对称带来的医患矛盾,使医生和患者实现双赢。

目前,医学电子信息系统还不能达到上述完美的状态。虽然全国各地的医院已经建立了医学电子信息系统,由于多数医院或科研机构采用自制的医学电子信息系统,给实现医学电子信息的共享带来一定的困难,但是,部分地区已经实现了医院间医学电子信息系统的共享。医学电子信息共享的实现为医学科研和临床实践提供了重要的帮助。

值得注意的是,完美的计算机决策支持系统能够为查询者提供最快最好的当前最佳的证据,但是住院医师不能单纯地依靠计算机决策支持系统证据做出的诊疗决策,而需要运用个体化原则,把计算机决策支持系统中的证据和病人的实际情况以及病人的意愿结合起来才能作出循证决策。研究者应尽量能够追溯到信息源头,通过计算机决策支持系统链接到系统综述和原始研究,评价信息的真实性、重要性和实用性,以确保为病人提供最佳的循证诊疗方案。

二、证据综合

证据综合(summaries)的主要形式是循证医学教科书与循证临床实践指南。目前,尽管有很多出版商和作者都采用“循证医学”一词,但却不了解循证医学的真正含义,或不能审慎地提供真正的循证医学内容,所以不是每一本教科书都符合循证医学教科书的要求。循证医学教科书要求至少一年修改一次,每一个章节或者部分应该有最近修改日期;采用明确的方法用于选择和评价证据,详尽介绍所使用的方法和程序,使其具有可重复性;在推进某种临床实践的时候,应给出相应的证据,以便于读者能够查找原始出处和研究细节,从而了解、支持、论述证据的详尽内容。

《临床证据》(*Clinical Evidence*)由《英国医学杂志》(*British Medical Journal, BMJ*, www.clinicalevidence.com)出版集团出版,是首推的循证医学教科书,《临床证据》是住院医师必读的循证医学教科书。《临床证据》中的医学信息以高质量的医学证据作为依据,并定期对每一个医学主题进行更新,在《临床证据》中,不仅可以查阅到需要查找问题的医学信息,而且可以通过相应链接查找到其他相关问题的医学信息。2008年,国内出版了中文翻译版本的《临床证据》,有精华版和全版本,包括常见的临床治疗证据。《医师信息与教育资源》(*the Physician's Information and Education Resource, PIER*, <http://pier.acponline.org/index.html>)由美国医师学会(American College of Physicians, ACP)提供的一本供住院医师阅读的在线循证医学教科书。该书明确提供了有关内科和初级保健的证据等级。《当代》(*UpToDate*, <http://www.uptodate.com>)以光盘和网络形式发行,每个季度更新,有广泛的参考信息,《当代》没有对其纳入的文献进行质量评价,但是主要证据有 Medline 摘要。这至少为住院医师提供了一个公平的机会评价和判断支



持的证据。以上推荐的循证医学教科书均来自国外,不管是在线版还是中文翻译版,在运用循证医学教科书进行循证医学临床实践的时候,均要符合中国当前的国情、医院的条件和考虑病人的实际情况。

三、循证证据提要

循证证据提要(syntheses of synopses)是一种把原始研究和综述按固定格式提炼的摘要。在没有现成的计算机决策支持系统和循证医学教科书的情况下,住院医师可以通过查询相关问题的研究和综述,获得很好的信息资源。但是,由于住院医师工作的繁忙,很难有时间查找和使用原始的循证医学资源,原始研究和综述的详细的文章内容又是住院医师需要阅读的部分。如何才能解决这一矛盾呢?循证证据提要能帮助住院医师在最短的时间内获取相关问题的医学信息。不管是循证证据提要,还是循证题目均尝试着为查询者提供简单且足够的信息,例如循证题目“血管紧张素受体阻滞剂与血管紧张素转化酶抑制剂治疗冠心病患者有效性和安全性的meta分析”。如果在题目中不能获得信息,可以通过阅读循证摘要获取。我们可以通过检索《循证医学》等查阅到循证题目和循证摘要。住院医师通过循证题目和循证摘要的检索和阅读其所提供的信息,即便在不清楚全文信息的情况下也能够采取相应的措施,为病人提供最佳的临床诊疗。

四、证据合成

证据合成(syntheses)即系统综述,例如Cochrane系统评价。如果没有计算机决策支持系统,查找到的摘要又不能提供更多的详细的信息,可以通过《Cochrane图书馆》(www.cochrane.org)查找系统综述。系统综述主要针对当前的预防和治疗问题为卫生保健的干预措施效果提供简明扼要的信息。系统综述是通过广泛收集全球已发表和未发表的证据、严格评价证据,定量合成,定性分析得到的科学综述。如果Cochrane图书馆中没有您感兴趣的题目,Medline是您的另一个选择。还有一个尚佳的资源是Ovid的《循证医学综述》(EBMR),它能够提供对Cochrane和Cochrane系统综述的一站式检索。例如“环孢素对肾病综合征的作用”,对Cochrane系统综述数据库(Cochrane Database of Systematic Reviews, CDSR)、循证综述摘要数据库(Database of Abstracts of Reviews of Evidence, DARE)进行综述检索,用“环孢素”(cyclosporine)和肾病综合征(nephritic syndrome)作为检索词,可以得到5条系统综述的记录。

五、研究摘要

研究摘要(synopses of studies)主要形式是循证摘要的杂志。如果没有更好的途径找到您所需要的信息,我们可以采用循证研究摘要。在Embase可以找到独立的试验研究的摘要。这里的研究摘要和上面提及的循证证据提要类似,不同在于研究摘要还包括独立的试验研究的摘要。这些信息是以一些高质量研究的结构摘要的形式描述,不仅符合严格的质量评级标准,并且以临床相关问题和兴趣的形式收集成册。

六、原始研究

研究(studies)是指发表在各种杂志上的原始研究。建立计算机决策支持系统、提炼摘要、合成证据均需要时间和精力。建立一个完美的计算机决策支持系统仍然是一个循证临床实践的理想状态。提炼摘要和合成证据,根据质量要求不同,通常需要花费半年至数年的时间。如果找不到计算机决策支持系统、证据综合、循证证据提要、系统评价以及循证研究摘要,那我们必须开始寻找原始研究。我们可以在检索网络数据库获取相关医学信息,如原始文献数据库Medline、Ovid等。

知识要点

循证医学证据资源“6S”等级结构:从最高级别至最低级别依次为:①计算机决策支持系统(systems);②证据综合(summaries);③循证证据提要(syntheses of synopses);④证据合成(syntheses)⑤研究摘要(synopses of studies);⑥原始研究(studies)。

第三节 获取证据的策略

由于循证医学中的核心部分是运用当前最佳证据,因此循证临床实践获取证据的获取方式与一般临床研究不同,只要针对问题找到最佳的证据即可,因此获取证据的策略是依据证据等级“6S”金字塔从高到低依次检索。问题不同,证据等级由高到低排列不同。

一、病因学问题证据获取的策略

病因学问题证据等级由高到低依次为循证临床实践指南、基于病例对照研究的系统评价、单个病例对照研究、病例系列报告、个案报告、体外与体内试验、没有证据基础的专家经验。

在临床工作中发现多数帕金森病(Parkinson's disease, PD)患者同型半胱氨酸水平偏高。同型半胱氨酸是蛋氨酸的代谢产物,是一种含硫氨基酸。正常生理情况下同型半胱氨酸在体内维持较低的水平。有研究提示高同型半胱氨酸血症与冠心病、动脉粥样硬化、脑梗死、痴呆、帕金森病等有关,但研究结果尚不一致。

1. 提出问题 根据 PICO 原则提出问题,高同型半胱氨酸血症是否与 PD 相关? P: 帕金森病(Parkinson's disease); I: 高同型半胱氨酸血症(hyperhomocysteinemia); O: 二者相关性(the correlation)。

2. 确定本研究为病因学问题。

3. 获取证据文献检索策略 计算机检索 Cochrane Systematic Review Database、PubMed(1966—2012.8)、Embase(1980—2012.8)、ISI、Spring、CBM(1980—2012.8)、VIP(1989—2012.8)、CNKI(1978—2012.8)、万方数据库(1980—2012.8)和中国优秀硕士、博士学位毕业论文数据库(1979—2012)。以“同型半胱氨酸”、“帕金森病”为中文检索词,“homocystein”、“Parkinson's disease”为英文检索词,同时结合各个数据库的主题词和自由词,运用逻辑符、通配符和范围运算符等制定检索模式。分别以不同方式进行检索。此外通过追查纳入文献的参考文献,并手工检索相关会议文献及未公开发表的文献,以 Cochrane Systematic Review Database 为例(图 1-2-3、图 1-2-4)。

初检出 1418 篇文献,经初筛选出文献 1029 篇,进一步阅读全文及剔除重复、数据不全文献

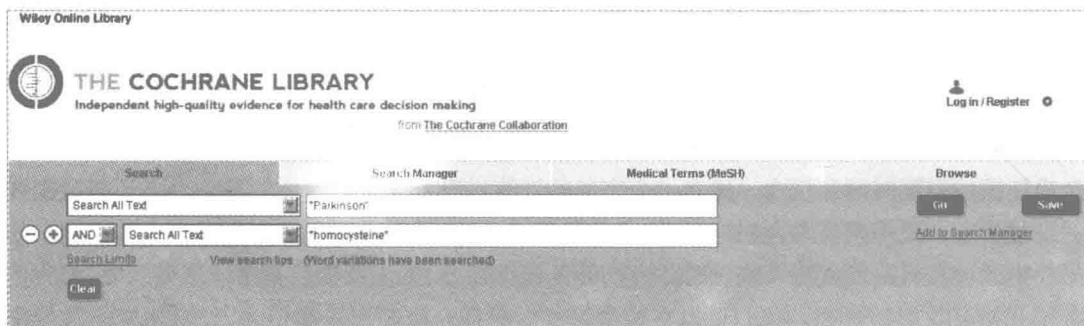


图 1-2-3 Cochrane 系统综述数据库检索式

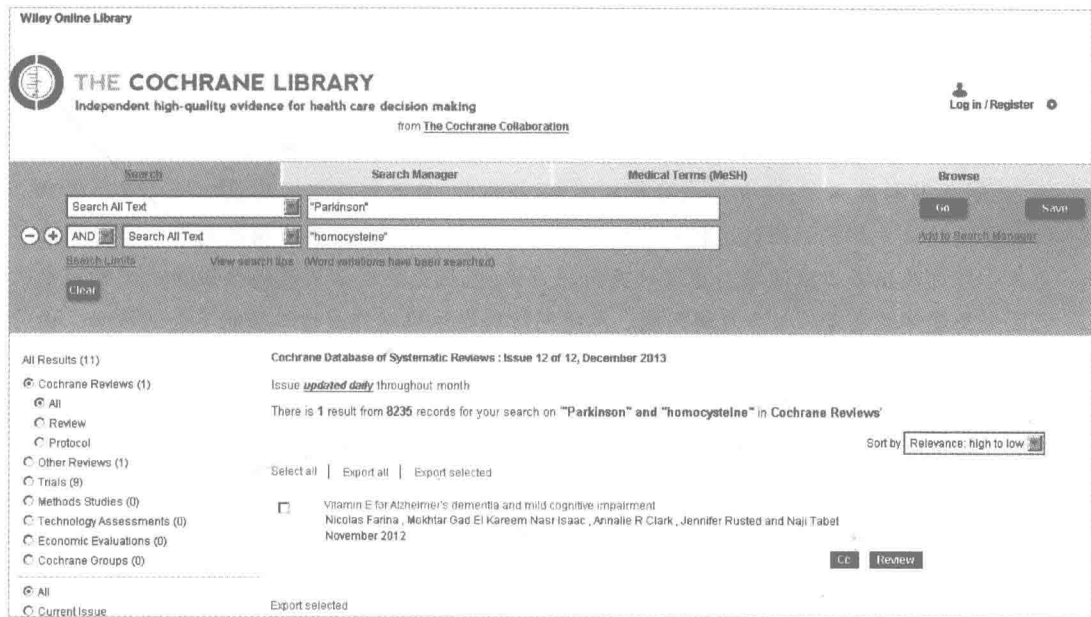


图 1-2-4 Cochrane 系统综述数据库检索式

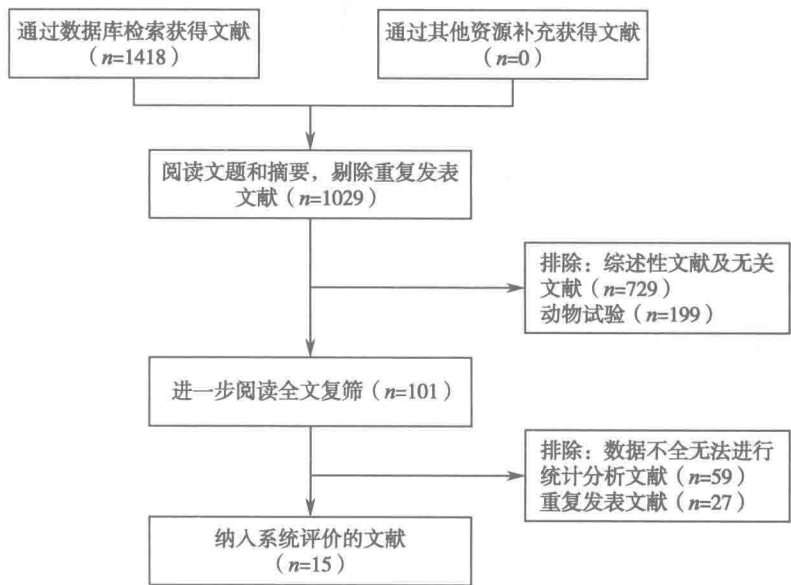


图 1-2-5 文献筛查流程

后最终纳入 15 个研究。文献筛查流程图见 1-2-5。

二、诊断问题证据获取的策略

诊断问题证据等级由高到低依次为循证临床实践指南、基于诊断试验研究的系统评价、单个诊断试验研究、病例系列报告、个案报告、体外与体内试验、没有证据基础的专家经验。

目前尚没有单纯收集诊断证据的专门数据库。但是，许多综合性循证临床证据库包含了诊断试验的研究证据。首先，使用已经过滤的医学信息资源（二次文献数据库）。使用二次文献数据库可以节省时间。如 Best Evidence (ACP journal club and evidence based medicine)、Cochrane Library、UptoDate、Sumsearch 和 Tripdatabase。其次，使用未过滤的医学信息资源（原始文献数据



库),如PubMed(最好从Clinical Queries路径进入)和Embase。再次,确定关键词和制定检索策略。

患者,70岁,男性,因“气憋2周,加重伴呼吸困难2小时”就诊于医院急诊科。有高血压病史2年,父亲死于心肌梗死。有吸烟史15年。查体:血压160/90mmHg,心率87次/分,无明显颈静脉充盈;肺部(-);心界不大,心音正常,各瓣膜区未闻及病理性杂音。胸片和心电图均未见异常。面对这样一个呼吸急促的老年患者,需要尽快确定患者的呼吸困难是源于肺部疾病或心脏问题。患者的高血压病史、吸烟史和直系亲属的心血管疾病家族史使医生高度怀疑患者的心力衰竭问题。单凭问诊和查体难以作出左心功能不全的诊断。由于是急诊,只能利用一些简便快捷的检查方法即检测血浆脑钠尿肽(brain natriuretic peptide, BNP),即B型钠尿肽辅助诊断。

1. 提出问题 根据PICO原则提出问题,血浆B型脑钠尿肽对诊断充血性心力衰竭的意义。P:充血性心力衰竭患者(patients with congestive heart failure);I:血浆脑钠尿肽含量(brain natriuretic peptide, BNP);C:超声心动图(金标准)(echocardiography);O:诊断充血性心力衰竭(congestive heart failure)。

2. 确定本问题属于循证诊断问题。

3. 获取证据文献检索 本病例中,以PubMed为检索资源,对构建问题时的四个要点进行组合,构成检索策略,对所期望的高级别证据进行查找(图1-2-6、图1-2-7、图1-2-8)。

三、预防治疗问题或危害研究证据获取的策略

预防治疗问题或危害研究证据等级为证据等级由高到低依次为循证临床实践指南、基于随机对照试验的系统评价、单个随机对照试验、基于队列研究的系统评价、单个队列研究、基于病例对照研究的系统评价、单个病例对照研究、病例系列报告、个案报告、体外与体内试验、没有证据基础的专家经验。

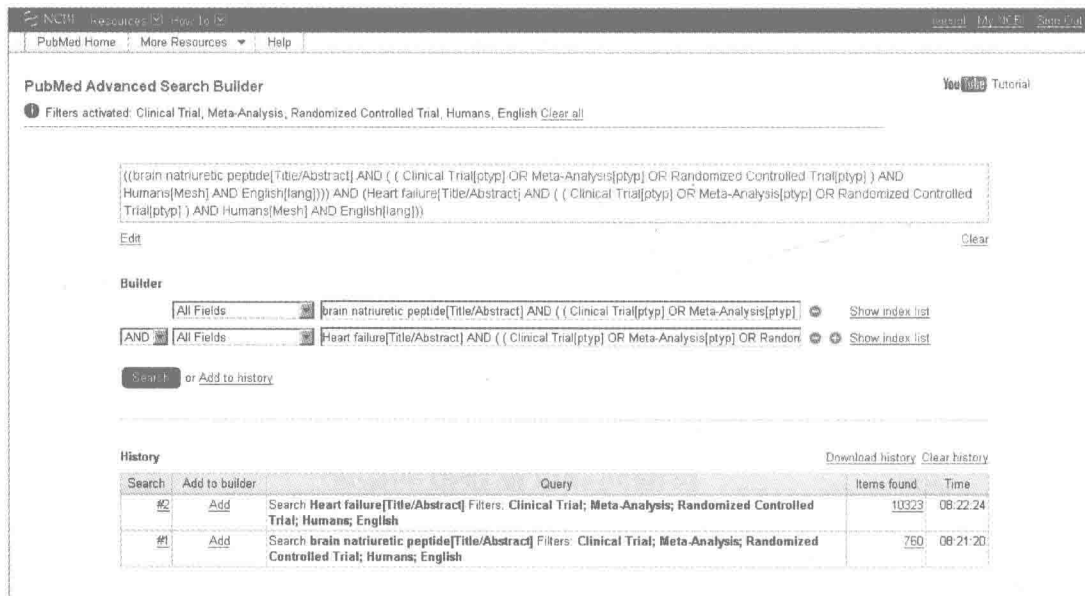


图 1-2-6 PubMed 检索示例

以“B-type natriuretic peptide”作为检索词,在“limited”中限定:Field:Title/Abstract,Limits:English,Clinical Trial,Meta-Analysis,Randomized Controlled Trial,Humans,检索得到106篇文献。以同样的方式检索“Heart failure”得6002篇文献

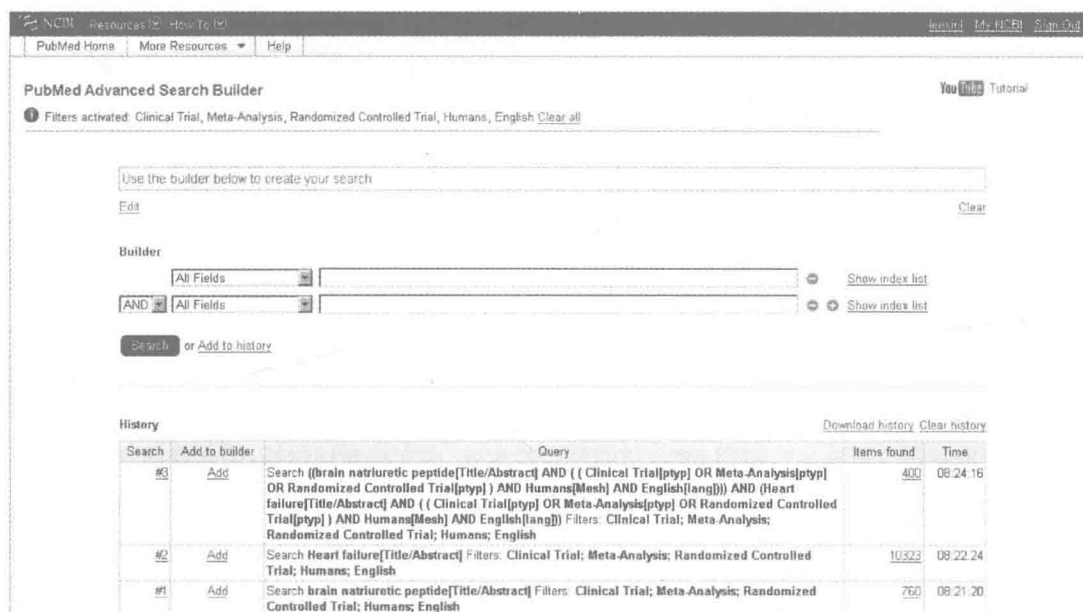


图 1-2-7 PubMed 检索示例
将两个检索结果以“AND”相加,得 75 篇文献

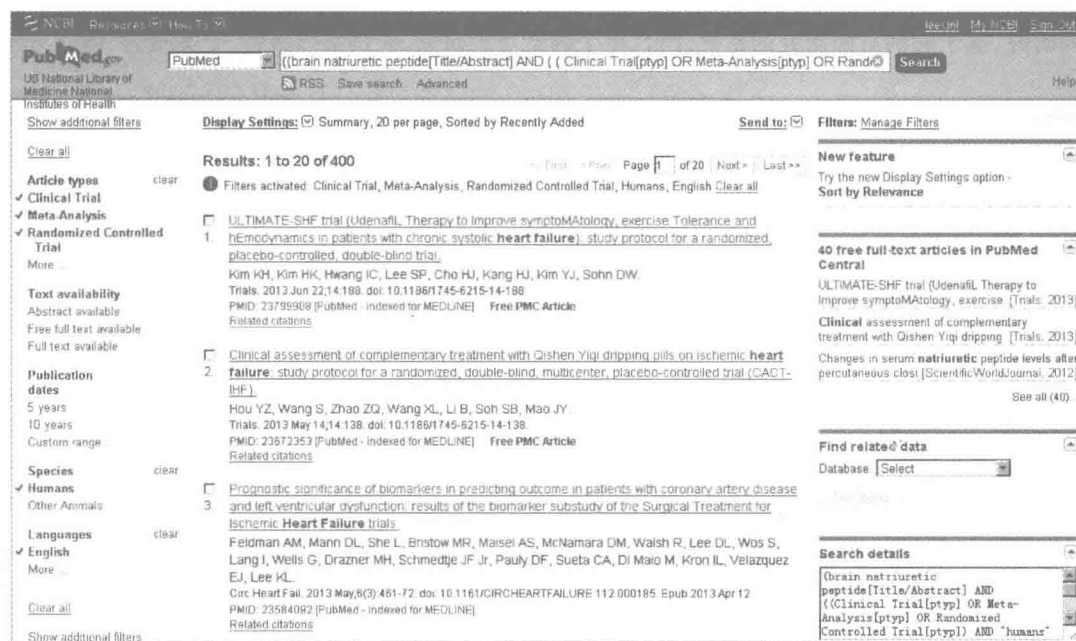


图 1-2-8 PubMed 检索示例

浏览文献题目和摘要,逐一筛查,发现一篇相关文献:“对急诊室诊断为心力衰竭的患者进行 BNP 的快速测定”(Maisel AS, Krishnaswamy P, Nowak RM, et al. Rapid measurement of B-type natriuretic peptide in the emergency diagnosis of heart failure. N Engl J Med, 2002, 347: 161-167)



患者,女性,63岁。主诉为肢体不自主抖动5年。完善相关检查后诊断为帕金森病,开始规律口服复方左旋多巴(美多巴)片,半片/次,2次/天,随着病情进展,药物逐渐加量。一年前患者出现右下肢不自主抖动,美多巴量增至半片/次,4次/天,效果仍不太理想。查体:表情呆滞,呈现面具脸,神志清,精神可,肢体可见不自主震颤,呈静止性,四肢肌力5级,双下肢腱反射亢进,双侧病理征阴性。雷沙吉兰(rasagiline)是第2代不可逆的高选择性单胺氧化酶-B(monoamine oxidase-B, MAO-B)抑制剂,能通过神经保护作用延缓PD的疾病进程。

1. 提出问题 根据 PICO 原则提出问题雷沙吉兰(rasagiline)与美多巴联合使用是否可以增加PD的疗效。P:帕金森病患者(Parkinson's disease);I:雷沙吉兰与美多巴联合使用(rasagiline and madopar);C:单用美多巴(madopar);O:治疗效果(therapeutic effect)。

2. 确定本例为循证治疗问题。

3. 获取证据文献检索 按照 PICO 策略制定检索式,中文检索词包括:帕金森病,雷沙吉兰,单胺氧化酶 B 抑制剂;英文检索词包括:parkinson*, rasagiline, monoamine oxidase type B inhibitor, randomized controlled trial。我们已经计算机检索 CNKI、VIP、Wanfang Data、Cochrane library (2013 年第 3 期)、WEB OF KNOWLEDGE、Ovid (EBM-REVIEWS*、Medline)、PubMed。RCT 限定检索遵循 Cochrane 系统评价手册 5.1.0 RCT 的检索策略,其中检索采用自由词与主题词相结合的方式,所有检索策略通过多次预检索后确定,并根据具体数据库调整策略。检索时间均从建库到 2013 年 10 月。示例:PubMed 数据库检索策略见图 1-2-9~图 1-2-11。

四、预后研究证据获取的策略

预后研究证据等级由高到低依次为循证临床实践指南、基于队列研究的系统评价、单个队列研究、基于病例对照研究的系统评价、单个病例对照研究、病例系列报告、个案报告、体外与体

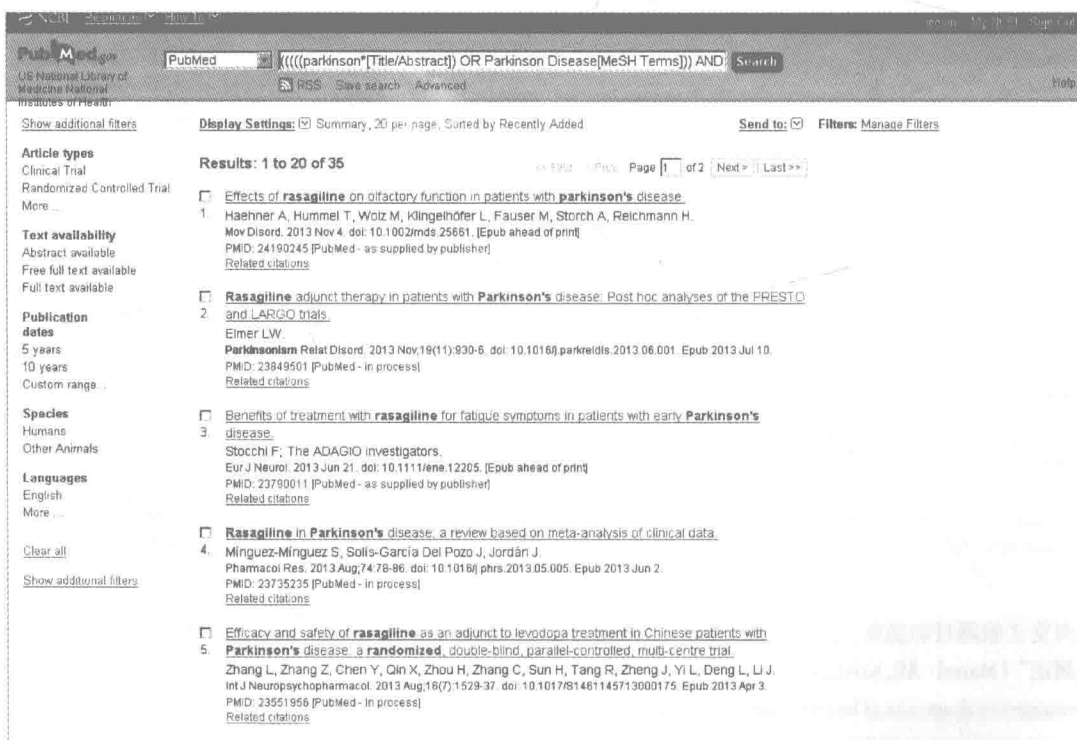


图 1-2-9 PubMed 数据库检索策略示例

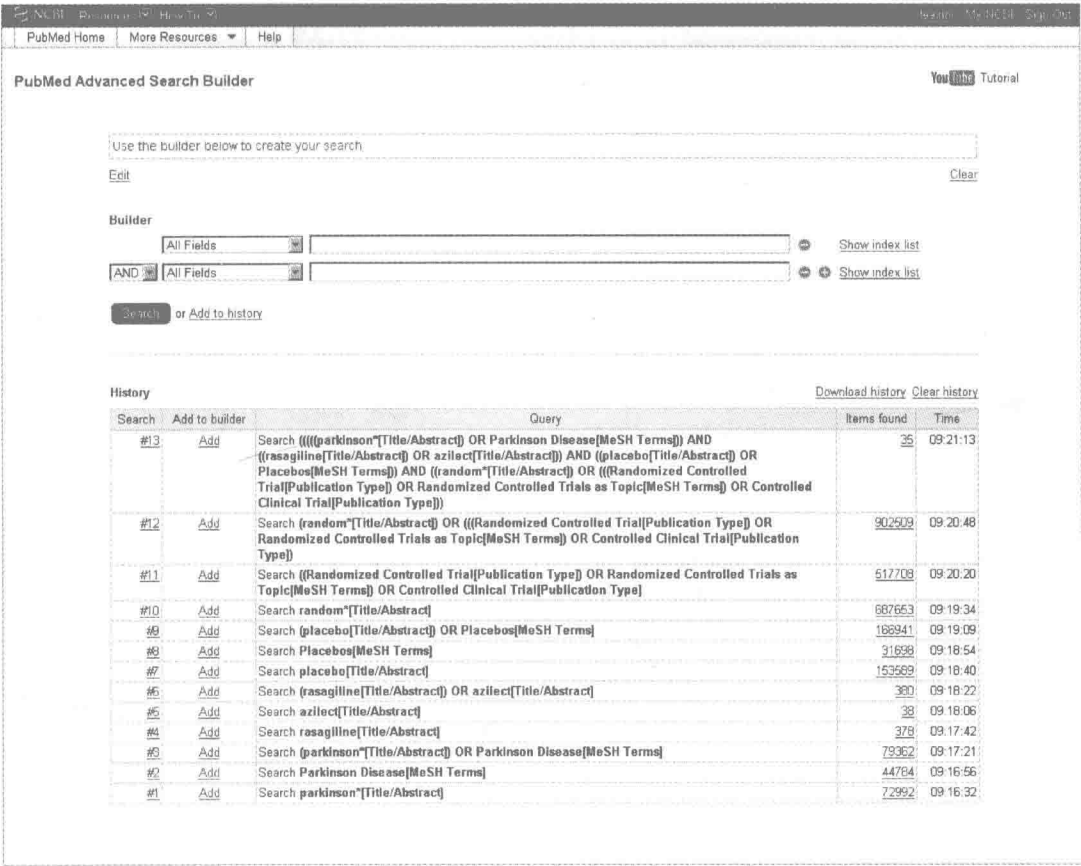


图 1-2-10 PubMed 数据库检索策略示例

学习
笔记

PubMed 的检索策略

- #1 parkinson* [Title/Abstract]
- #2 "Parkinson Disease" [Mesh]
- #3 "(#1) OR #2
- #4 rasagiline [Title/Abstract]
- #5 azilect [Title/Abstract]
- #6 (#4) OR #5
- #7 placebo [Title/Abstract]
- #8 "Placebos" [Mesh]
- #9 (#7) OR #8
- #10 random* [Title/Abstract]
- #11 "Randomized Controlled Trial" [Publication Type] OR "Randomized Controlled Trials as Topic" [Mesh] OR "Controlled Clinical Trial" [Publication Type]
- #12 (#10) OR #11
- #13 (((#3) AND #6) AND #9) AND #12

内试验、没有证据基础的专家经验。

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(basic search)和高级检索(advanced search)。下面推荐的主题词和副主题词是便于高级检索时使用。

推荐检索预后证据的主题词有:Cohort Studies、Prognosis、Disease Progression、Treatment Outcome、Longitudinal Studies、Medical Futility、Prospective Studies、Treatment Failure、Follow-up Studies、Morbidity、Incidence、Mortality、Prevalence、Survival Rate、Death Infant Mortality、Survival Analysis 和 Maternal Mortality。

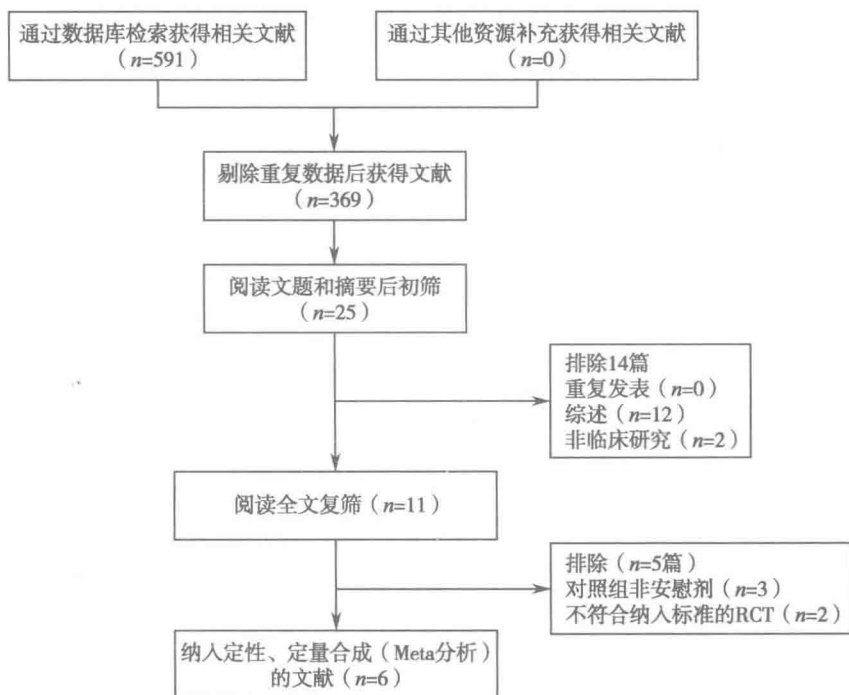


图 1-2-11 文献筛查流程

推荐检索预后证据附加于主题词后的副主题词有:mortality (mo) 和 epidemiology (ep)。

检索预后证据的关键词(自由词)是根据 PICO 原则获得,有利于进行初级检索时使用。

推荐的关于预后证据高级检索的检索策略:①(prognos \$ or outcome \$ or follow-up or predict \$). ti, ab, sh.; ②exp disease progression; ③《natural \$ or disease \$) adj (progress \$ or course \$ or histor \$). ti, ab, sh.; ④4.1 or 2 or 3; ⑤exp cohort studies; ⑥(cohort \$ or compar \$ or longitudinal \$ or prospective \$ or multivariate or reproducib \$) ti, ab, sh.; ⑦5 or 6; ⑧4 and 7。

患者,男性,71岁。2012年10月3日以“突发言语不清,口角歪斜1天”为主诉入院。在医院完善头颅CT及头颅MRI检查后诊断为右侧缺血性脑卒中,经积极治疗后好转。但造成左上肢功能丧失,能依靠拐杖行走。2013年10月初来医院复查。有15年高血压病史,血压一直控制不佳。超声心动图显示左心室肥厚及轻度左心功能不全。心电图提示窦性心律。血管超声显示双侧颈动脉狭窄<50%。目前服用的药物有阿司匹林 300mg/d、血管紧张素转化酶抑制剂和噻嗪类利尿剂。

1. 提出问题 根据 PICO 原则提出问题,首次脑卒中后的复发率、致残率和病死率? P: 首次脑卒中患者(a first-event stroke); I: 任何治疗措施(any treatment); C: 未行干预或治疗(no intervention or treatment); O: 复发率、致残率和病死率(The recurrence rate, morbidity and mortality)。

2. 该问题为循证预后问题。

3. 获取证据文献检索 本案例中,首先确定在医学信息量最大的 PubMed 进行检索,希望能找到前瞻性队列研究证据或次一级的病例一对照研究证据。首先确定主题词,在 Medical Subject Heading (MeSH) 中输入自由词“stroke”后,发现其主题词是“cerebral infarction”。我们可以看一下选择不同检索式进行初级检索和高级检索的结果:如果单用“cerebral infarction”,可检索到相关文献 955 篇;联合“epidemiology”、“recurrence”和“prognosis”,产生检索式:“cerebral infarction” AND (“epidemiolog*” OR “recurren*” OR “prognosis”),可检索到相关文献 379 篇文章;



再联合“cohorts”,产生检索式“cerebral infarction”AND(“ep-idemiolog*”OR“recurren*”OR“prognosis”)AND“cohort\$”,可检出相关文献 157 篇。然后通过摘要逐一筛查,发现一篇相关文献:“首次脑卒中后复发的长期风险”(Long-term risk of recurrent stroke after a first-ever stroke)。该文献以下简称“脑卒中再发危险”(图 1-2-12~图 1-2-14)。

在 ACP journal club 中检索,“cerebral infarction”AND“recurrence”AND“prognosis”检索到 2 篇;在 Sumsearch 中检索,“cerebral infarction”AND“recurrence”AND“prognosis”找到 38 篇;在 Tripdatabase 检索“cerebral infarction”AND“recurrence”AND“prognosis”AND“cohort study”找到 50 篇;在 PubMed 中检索,“stroke”AND“cohort\$”AND(“epidemiolog*”OR“recurren*”OR“prognosis”),共找到 157 篇。

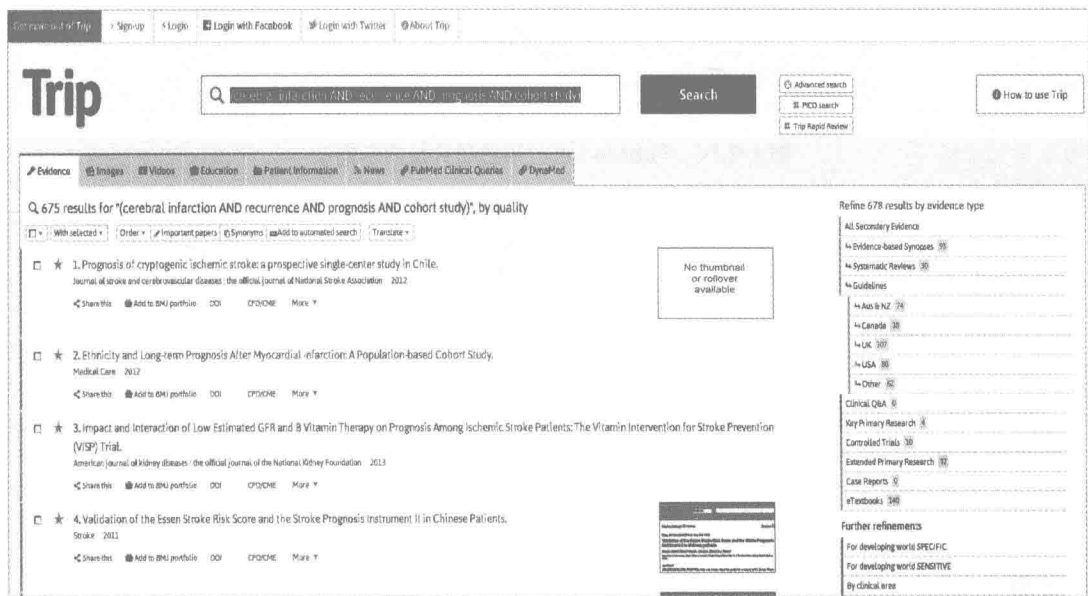


图 1-2-12 PubMed 进行预后研究检索示例一

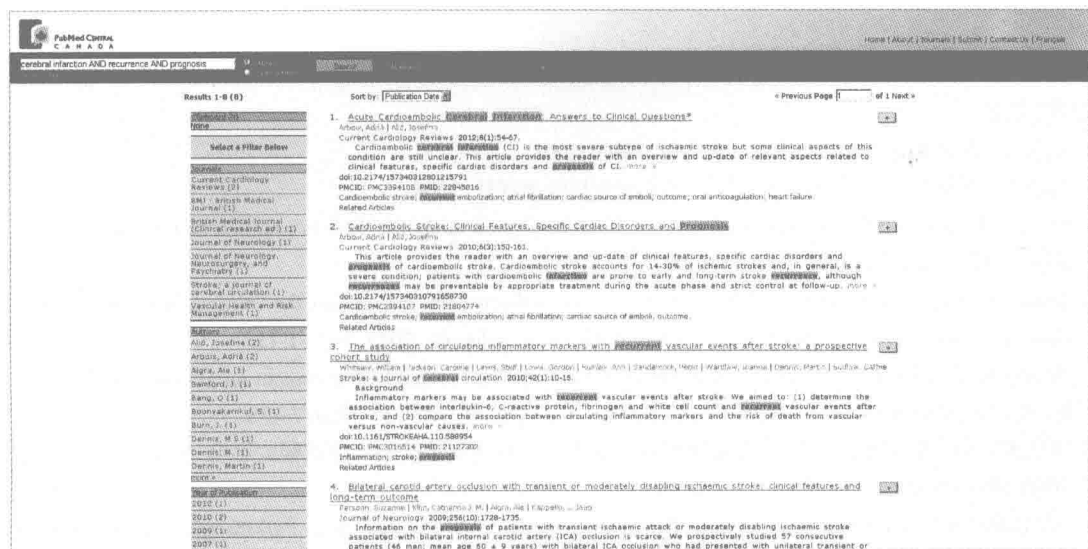


图 1-2-13 PubMed 进行预后研究检索示例二



图 1-2-14 PubMed 进行预后研究检索示例三

知识点

获取证据的策略:①依据证据等级“6S”金字塔从高到低依次检索;②临床问题不同,获取证据的检索策略也存在差异,应根据临床问题灵活运用;③检索式没有统一标准,同一疾病,同一网站,没有统一的检索式;不同网站,不同检索库,逻辑符不同,逻辑符布尔逻辑符和非布尔逻辑符之分。

思考题

1. 何为 PICO 原则?
2. 何为“6S”?
3. 查找循证医学证据常用的网站有哪些?

(徐佩茹)

参考文献

1. 董碧蓉. 循证临床实践. 北京:人民卫生出版社,2008.
2. 詹思延. 循证医学实践与教学. 北京:北京大学医学出版社,2006.
3. Straus SE, Glasziou P, Richardson WS, et al. Evidence-based medicine, how to practice and teach it. Fourth Edition. Churchill: Livingstone elsevier, 2011.
4. Jenner P, Morris HR, Robbins TW, et al. Parkinson's disease-the debate on the clinical phenomenology, aetiology, pathology and pathogenesis. Journal of Parkinson's disease, 2013, 3(1): 1-11.
5. Nevrlly M, Kanovsky P, Vranova H, et al. Effect of levodopa and entacapone treatment on plasma homocysteine levels in Parkinson's disease patients. Parkinsonism Relat Disord, 2009, 15(6): 477-478.
6. 王红梅, 葛许华, 周联生, 等. 帕金森病患者左旋多巴治疗对血浆同型半胱氨酸的影响. 中华神经科杂志, 2006, 39(9): 609-611.
7. ADAGIO Study Investigators. A double-blind, delayed-start trial of rasagiline in Parkinson's disease. N Engl J Med, 2009, 361(13): 1268-1278.
8. Naoi M, Maruyama W. Monoamine oxidase inhibitors as neuroprotective agents in age-dependent



- neurodegenerative disorders. *Curr Pharm Des*, 2010, 16 (25): 2799-2817.
9. Robottom BJ. Efficacy, safety, and patient preference of monoamine oxidase B inhibitors in the treatment of Parkinson's disease. *Patient Prefer Adherence*, 2011, 5: 57-64.
 10. Amit T, Weinreb O, Bar-amor, et al. The neuroprotective mechanism of 1-(R)-aminoindan, the major metabolite of the anti-parkinsonian drug rasagiline. *J Neurochem*, 2010, 112 (5): 1131-1137.
 11. Keating GM, Lyseng-Williamson KA, Hoy SM. Rasagiline: a guide to its use in Parkinson's disease. *CNS Drugs*, 2012, 26 (9): 781-785.

第二篇

各 论



第一章 病因及危险因素

临床问题

医疗过程中临床医师经常需要考虑是否某种危险因素或治疗措施对患者有害,例如:吸烟是否增加患肺癌的危险?口服避孕药是否增加患乳腺癌的危险?而在互联网高速发展的当代,患者也会关注网上新闻和医学的报道,例如:一名正在服用钙通道阻滞剂的高血压患者就曾携带文献询问医师“钙通道阻滞剂是否增加患癌症的危险”,要回答这些问题,就需要在有关病因学研究(或治疗措施副作用研究)的信息中寻找证据,那么这些证据的真实性如何?重要性怎么判断?患者的情况是否适用?结论又是否可信呢?

第一节 基本概念

一、病因与危险因素

病因(causation of disease)那些能使人群发病概率升高的因素,就可以认为有病因关系存在;当它们之中一个或多个因素不存在时,人群疾病频率就会下降。

以结核病患病为例:结核杆菌作用于易感机体,导致机体罹患结核病,就是结核病的病因。

危险因素(risk factor)目前,慢性非传染性疾病已经是危害人类健康的主要疾病,慢性疾病由于发病比较隐匿,病程缓慢,病因复杂,从单一的患病个体去研究疾病病因会十分困难,因此,需要以相应群体作为研究对象对有关的发病因素进行宏观分析探讨,因而提出了危险因素这一概念。依据流行病学调查资料认为与健康及其相关状态有关联,并有重要预防意义的一些个人行为或生活方式特征、环境暴露、一些先天性或遗传性特征等,都可称之为危险因素。危险因素一词的使用范围相当广,具有下列任何一种含义者即可称为危险因素:①某个因素或暴露状态与特定结果(如疾病的发生)的概率增大有关联,但不一定是因果联系,该因素为一危险标记;②一种能增加某种疾病或其他特定结局发生的概率,该因素为一决定因素;③一种决定因素能通过干预措施改变其影响和作用,并通过这种干预降低疾病或其他特定结局发生的概率。为了避免混淆,这种决定因素可以称为可预防危险因素。值得注意的是,危险因素虽然与疾病的发生、消长有一定的因果关系,但缺乏充分依据能阐明其明确的致病效应。

有研究提示:缺乏体育锻炼的人群当中肥胖率较体育锻炼较多的人群为高,且当这些人群增加运动时间,改善运动习惯后其人群的肥胖率会有所下降,故缺乏体育锻炼是肥胖的危险因素。且属于可预防的危险因素。

另,有研究发现,在A型性格人群中冠心病的发病率较普通人群为高,A型性格是冠心病的危险因素之一。



二、病因学研究的测量指标

(一) 相对危险度

相对危险度(relative risk,RR)是在随机对照研究和队列研究中,表示因果相关性强度的指标,是指暴露组发病或死亡的危险性与非暴露组发病或死亡的危险性之比,即病因暴露组的发病率与未暴露组发病率的比值,或治疗组副作用的发生率与非治疗组副作用发生率

以“孕妇服用反应停是否会增加其子女患海豹肢畸形的危险”为例,将随机对照研究或队列研究的结果总结于下列四格表(表 2-1-1)。

表 2-1-1 反应停致海豹肢畸形效应结果

		海豹肢畸形		合计
		+	-	
反应停	服用	a(10)	b(14)	a+b(24)
	未服用	c(51)	d(21 434)	c+d(21 485)

以该表为例,服用反应停的海豹肢畸形发生率是 $a/(a+b)$ (即 I_e),未服用组海豹肢畸形发生率是 $c/(c+d)$ (即 I_0),相对危险度即为:

$$RR=[a/(a+b)]/[c/(c+d)]=\frac{I_e}{I_0}=\frac{10/24}{51/21\,485}=175.5$$

RR 越远离 1,则越有价值。在上述服用反应停是否增加后代患海豹肢畸形危险的队列研究中 $RR=175.5$,其意思是服用反应停的孕妇后代患海豹肢畸形的机会是未服用反应停组的 175.5 倍。

(二) OR 值

OR(odds ratio)即比值比,其意义表示病例组中暴露于该因素者与未暴露者的比值为对照组中该项比值的倍数。

仍以“服用反应停是否增加后代患海豹肢畸形的危险”为例,将病例对照研究的结果总结于下列四格表(表 2-1-2)。

表 2-1-2 反应停致海豹肢畸形效应病例对照结果

		病例组(海豹儿的母亲)	对照组(正常儿的母亲)
反应停	既往曾服用	a(12)	b(2)
	既往未曾服用	c(188)	d(298)

$$OR=(a/c)/(b/d)=ad/bc=12\times 298/2\times 188=9.5$$

表示海豹儿母亲的暴露率(服用反应停)是正常儿母亲的 9.5 倍,表示既往服用反应停的孕期妇女产出海豹肢畸形患儿的危险是未服用者的 9.5 倍。

(三) 归因危险度

归因危险度(attributable risk,AR)又叫特异危险度、危险度差(risk difference,RD)和超额危险度(excess risk),是暴露组发病率与对照组发病率相差的绝对值,它表示危险特异地归因于暴露因素的程度。

学习
笔记



$$AR=I_e-I_0=\frac{a}{n_1}-\frac{c}{n_0}$$

由于 $RR=\frac{I_e}{I_0}$, $I_e=RR \times I_0$
所以 $AR=RR \times I_0 - I_0 = I_0(RR-1)$

表 2-1-3,从 RR 看,吸烟者患肺癌的危险是不吸烟者的 10.7 倍,患心血管疾病的危险是不吸烟者的 1.7 倍,吸烟对肺癌的作用较大,病因联系较强;但从 AR 看,与不吸烟者比,吸烟增加 45.43/10 万人年的肺癌死亡率,126.43/10 万人年的心血管疾病死亡率,吸烟对心血管疾病的作用较大,控烟对预防心血管疾病死亡所取得的社会效果将更大。

表 2-1-3 吸烟与肺癌和心血管疾病的 RR 与 AR 比较

疾病死亡率	吸烟者	非吸烟者	RR	AR
	(1/10 万人年)	(1/10 万人年)		(1/10 万人年)
肺癌	50.12	4.69	10.7	45.43
心血管疾病	296.75	170.32	1.7	126.43

(四) 归因危险度百分比

归因危险度百分比(attributable risk proportion or percent,ARP,AR%)又称为病因分值(etiological fraction,EF)或归因分值(attributable fraction,AF),是指暴露人群中的发病或死亡归因于暴露的部分占全部发病或死亡的百分比。

$$AR\%=\frac{I_e-I_0}{I_e} \times 100\%$$

或

$$AR\%=\frac{RR-1}{RR} \times 100\%$$

以表 1-3 为例计算肺癌的归因危险度百分比

$$AR\%=\frac{50.12-4.69}{50.12} \times 100\%=90.6\%$$

说明吸烟者中发生的肺癌有 90.6% 可归因于吸烟。

(五) 人群归因危险度

人群归因危险度(population attributable risk,PAR)与人群归因危险度百分比(PAR%),人群归因危险度百分比也叫人群病因分值(population etiologic fraction,PEF)或人群归因分值。PAR 是指总人群发病率中归因于暴露的部分,而 PAR% 是指 PAR 占总人群全部发病(或死亡)的百分比。

PAR 和 PAR% 的计算公式如下:

$$PAR=I_t-I_0$$

I_t 代表全人群的率, I_0 为非暴露组的率

$$PAR\%=\frac{I_t-I_0}{I_t} \times 100\%$$

另外,PAR% 亦可由下式计算:

$$PAR\%=\frac{P_e(RR-1)}{P_e(RR-1)+1} \times 100\%$$

式中 P_e 表示人群中有某种暴露者的比例,从该式可看出 PAR% 与相对危险度及人群中暴露者的比例的关系。

学
习
笔
记



继续以表 2-1-3 的数据资料为例,已知非吸烟者的肺癌年死亡率为 0.0469‰ (I_0),全人群的肺癌年死亡率为 0.2836‰ (I_t),则

$$(1) PAR = I_t - I_0 = 0.2836\text{‰} - 0.0469\text{‰} = 0.2367\text{‰}$$

$$(2) PAR\% = \frac{I_t - I_0}{I_t} \times 100\% = \frac{0.2367}{0.2836} \times 100\% = 83.5\%$$

从计算结果可知,虽然吸烟导致肺癌的 AR% 达 90.6%,但因人群中只有部分人吸烟,故其 PAR% 仅为 83.5%。

(六) NNH/NNT

还有一种对临床医师和患者更直观、更易理解的指标,就是 NNH (number need to harm),其意思是导致额外一例疾病的发生需要暴露在可疑危险因素中易感个体的人数(或是导致一例副作用的发生需要接受治疗措施的病人数,而这种治疗措施被怀疑与副作用的发生有关)。另一个与之计算方法相同的指标是 NNT (number need to treat),其意思是防治一例不良事件的发生需治疗的病例数。对 NNH 计算,一般是进行副作用的研究,则:

$AR\% = [\text{试验组事件发生率} - \text{对照组事件发生率}] / \text{试验组事件发生率}$
意思是事件的发生归因于暴露于试验因素的百分比,又称相对危险增加率(RRR)。

$$AR = [\text{试验组事件发生率} - \text{对照组事件发生率}]$$

意思是与对照组相比试验组发生事件的绝对危险增加的水平。又称绝对危险增加率(ARR)。而 NNH 即为 ARR 的倒数,即

$$NNH = 1/ARR$$

钙通道阻滞剂是否增加患癌症危险的队列研究(Lancet, 1996, 348:493-497)中,用药组癌症发生率为 3.03%,未用药组癌症发生率为 2.17%。即:

$$RRR = (3.03\% - 2.17\%) / 3.03\% = 28.03\%$$

$$ARR = 3.03\% - 2.17\% = 0.86\%$$

$$NNH = 1/0.86\% = 116$$

意思是需要 116 人服用钙通道阻滞剂才导致额外一例癌症的发生。

该案例是研究治疗药物副作用的,故其指标为 NNH。

如果研究治疗措施的保护作用,则一般选择 NNT 作为计算指标,此时:

$$ARR = [\text{对照组事件发生率} - \text{试验组事件发生率}]$$

意思是与对照组相比试验组发生事件的绝对危险减少的水平。

而 NNT 即为 ARR 的倒数,即:

$$NNT = 1/ARR$$

血管紧张素转化酶抑制剂治疗急性心梗(AMI)可使病死率(治疗 4 周)与安慰剂比较有所降低,其试验组病死率为 9.1%,安慰剂对照组病死率为 9.7%。

$$ARR = 9.7\% - 9.1\% = 0.6\%$$

$$NNT = 1/0.6\% = 167$$

表示采用血管紧张素转化酶抑制剂治疗 167 例 AMI 患者可挽救 1 例死亡。

三、病因学研究的主要方法与设计方案

病因学研究探索的是疾病发生的原因及相关因素间的相互作用和各因素对疾病发生、发展的影响。在病因学研究中,必须强调的一个概念是任何与结局(疾病)有联系的原因性暴露都必



须发生在结局之前。但是发生在结局(疾病)之前的暴露未必就是真正的原因(病因)。病因学的研究可以为临床诊断、治疗和预防提供依据,同时可以为制定相应的医学决策提供依据,从而获得疾病防治的最大效益。本章就临床常用的病因学研究类型概述见表 2-1-4。

表 2-1-4 常用病因学研究设计类型简表

			特点	用途
观察性研究	描述性研究	病例报告	快,无对照,无设计	用于提供病因线索
		横断面研究	有设计,无对照	描述分布,寻找病因线索
	分析性研究	病例对照研究	由果及因,按有无疾病分组	初步验证因果关系
		队列研究	由因及果,按暴露状况分组	验证因果关系
实验性研究		随机对照试验	随机化分组,人为干预	验证病因,研究副作用

我们以一种原因不明的晶体后纤维增生症的病因研究过程为例介绍常用的病因学研究方法。

(一) 病例报告

病例报告(case reports)是指详细地介绍某种罕见病的单个病例或少数病例。常是发现某药物临床应用中发生严重或罕见不良反应的第一线索。

1942 年,美国眼科医师 Terry 首先报道了发生在波士顿早产儿中的一种原因不明的晶体后纤维增生症(retrolental fibroplasia,RLF)5 例,均见于出生后 3~6 个月的早产儿,他对该 5 例病人的临床特征和组织病理特征都进行了详细描述和报告,发现致盲的原因是晶状体后的纤维组织块,它的出现与以前已知眼病的特征不符合。该病例报告提出了一种原因不明性的致盲性疾病,提出的研究问题引起了研究者的关注,随后,欧美一些国家也报告了类似病例。

(二) 横断面研究

横断面研究(cross-sectional study)又称现况研究、患病率研究(prevalence study)、疾病频率调查(disease frequency study)。通过对特定时点(或期间)和特定范围内人群中的有关因素与疾病或健康关系的描述,即调查这个特定群体中的个体是否患病和是否具有某些特征的情况,从而描述所研究的疾病(或某种健康状况)以及有关因素在目标人群中的分布,为研究的深入提供线索和病因学假设。

接续上例,研究者进一步查阅地理分布的相关资料。发现继波士顿之后,美国其他城市、其他国家也有报告,但发病率差别大:英、法、瑞典和澳大利亚病例较多;设备较好的医院、教学医院病例较多。(美国 1951 年儿科学会议材料,表 2-1-5)

表 2-1-5 不同城市、医院和人种 RLF 发生情况

城市	医院类型	白人		有色人种	
		早产儿数	病因不明失明	早产儿数	病因不明失明
新奥尔良	慈善医院	90	0(0)	210	0(0)
	私人医院	100	9(9%)	31	0(0)
纽约	部分免费医院	1541	188(12%)	759	2(0.3%)
芝加哥	公费医院	600	2(0.3%)	600	2(0.3%)
	教学医院	100	35(35%)	30	3(10%)
普洛维顿斯	部分免费医院	115	10(8.7%)	67	0(0)
合计		2446	244(10%)	1697	7(0.4%)



儿科医师 Rothmund 等调查了不同出生体重早产儿中的 RLF 失明发生率(表 2-1-6)。Engle 等报道了不同孕周早产儿的各种 RLF 表现的发生率(表 2-1-7)。发现 RLF 绝大多数发生在早产或低体重儿中,且体重愈低,发病率愈高。

表 2-1-6 不同体重婴儿的失明发生率(马里兰州)

出生时体重(g)	婴儿数	原因不明性盲(%)
<1000	12	33.3
1001-1500	129	17.8
1501-2000	548	0

表 2-1-7 不同孕周早产儿中 RLF 不同表现的发生率

妊娠周数	婴儿数	RLF		
		急性	后遗症	失明
≥31 周	51	29(57%)	6(12%)	1(2%)
≤30 周	42	28(67%)	15(36%)	6(14%)

根据以上横断面调查的结果,我们可以看出这种新发的 RLF 多出现在条件较好,医疗条件较先进的医院和城市地区,且在低体重儿中多发,提示其病因可能与新兴的医疗措施有关。

(三) 病例对照研究

病例对照研究(case-control study),又称回顾性研究,基本原理是以一组已经确诊的患有某种特定疾病的病人作为病例组,以一组不患有该病但具有可比性的个体作为对照组,通过询问、实验室检查或复查病史,搜集既往各种可能的危险因素的暴露史,测量并比较病例组与对照组中各因素的暴露比例,经统计学检验,若两组某因素的差别有统计学意义,则可认为该因素与疾病之间存在着统计学上的关联。在评估了各种偏倚对研究结果的影响之后,再借助病因推断技术,推断出某个或某些暴露因素是否为疾病的危险因素,从而达到探索和检验病因假说的目的。这是一种回顾性的、由结果探索病因的研究方法,是在疾病发生之后去追溯假定的病因和检验病因假说的一种研究方法。

美国眼科医师 Kinsey 和 Zachanias 回顾性地调查了 1938~1947 年在波士顿医院出生的低重儿 372 例,发现其中 RLF53 例,并从患儿亲友、经治医师及病历中收集相关可疑危险因素的资料。

结果:当时尚未采用规范的病例对照研究。

一胎或多胎、性别、有无先天畸形、白细胞计数、照射 X 线、输血、眼部或非眼部感染等因素无意义。护理因素与 RLF 的关系如表 2-1-8。

表 2-1-8 早产儿护理情况与 RLF 的关系

护理情况	正常婴儿	RLF 婴儿
平均护理天数	46	56
在暖箱平均天数	12	19
平均输氧天数	12	21

作者进一步将病例按两个时期(1938~1942,1943~1947)分别统计研究,结果后一时期 RLF 病例显著增多,护理上变化有增加应用铁剂、维生素 A、维生素 D 改为水溶剂和更多给早产儿输氧。



根据病例-对照研究的定义,可以看出,上述案例并不是一个典型的病例对照研究,但是采用了病例对照研究的基本研究思路,那如果一个典型的病例对照研究应该如何设计呢,就本案例而言,选择一组患 RLF 的早产儿作为病例组,在同一医院或地区选择同期生产的眼睛正常的患或不患其他疾病的早产儿作为对照组,回顾性收集两组患儿的人口学特征,医疗措施接触史,尤其是根据上文案例提示的应用铁剂,维生素 A、维生素 D 以及输氧等的具体接触剂量、时间等以及他们母亲的相关病史,测量并比较病例组与对照组中各因素的暴露比例,计算相关指标(如 OR)并进行统计学检验,从而判断和验证哪些或者哪种新近引入的医疗措施是可能导致 RLF 的病因。

(四) 队列研究

队列研究(cohort study)是分析性流行病学研究中的重要方法之一,类似的名称还有前瞻性研究(prospective study)、随访研究(follow-up study)、纵向研究(longitudinal study)。其验证病因假设的效能优于病例对照研究。其基本原理是将特定人群分为暴露于某因素与未暴露于某因素的两组人群或不同暴露水平的几组亚人群,追踪观察期内各自的发病结局并比较两组或各组间的发病率或死亡率,从而判定暴露因子与发病有无关联及关联程度大小的一种分析性研究方法。依据研究对象进入队列时间及终止观察的时间不同,队列研究分为前瞻性队列研究、历史性队列研究和双向性队列研究。

不同地区学者均注意到 RLF 发病与婴儿输氧方式的关系,分别进行了历史性队列研究,例如 Campbell(1951)在澳大利亚墨尔本,Chase 和 Evans(1952)在美国伯明翰某医院的调查,结果见表 2-1-9、表 2-1-10。

表 2-1-9 墨尔本三个单位 RLF 发病率

机构	供氧情况	婴儿数	RLF 数	发病率(%)
A	40%-60% 氧气常规	123	23	18.7
B	紫绀时应用	44	3	6.8
C	紫绀时应用	14	1	7.1

表 2-1-10 某医院不同时期 RLF 发病率

时期	婴儿数	存活婴儿		RLF	
		例数	%	例数	%
1931~1948(很少用氧)	412	132	32.0	1	0.8
1949~1950(广泛用氧)	69	26	37.7	5	19.2
1950.6~1951(按需供氧)	54	24	44.5	0	0.0

可以看到,当时的研究者也是采用了队列研究的设计思路,将入院的早产儿按照高浓度持续给氧和按需给氧分组,观察不同组间 RLF 发生率的差别,从而判断给氧和 RLF 发病之间是否存在关联以及关联程度的大小。

(五) 随机对照试验

随机对照试验(randomized control trials, RCTs)是在人群中进行的、前瞻性的、用于评估医学干预措施效果的实验性对照研究。它把研究对象随机分配到不同的比较组,每组施加不同的干预措施,然后通过适当时间的随访观察,比较组间重要结局发生频率的差别,以定量估计不同措



施的作用或效果的差别。除对照和随机分组外,随机对照试验通常还会采用分组隐匿、安慰剂、盲法、提高依从性和随访率、使用维持原随机分组分析等降低偏倚的措施。随机对照试验是目前评估医学干预措施效果最严谨、最可靠的科学方法。

随机对照实验的特点:①属于前瞻性研究:干预在前,效应在后,因果论证强度高。②随机分组:采用随机的方法把研究对象分配到实验组或对照组,以控制研究中的偏倚和混杂。③具有均衡可比的对照组:实验流行病学研究中的对象均来自统一总体的样本人群,其基本特征、自然暴露因素等应相似。④有人为干预措施:是与观察性研究(自然状态)一个根本的不同点。

为了进一步证明早产儿给氧与 RLF 发病的因果关系,一些学者进行了实验流行病学研究(干预性研究、临床试验)。如 Patz 等(1952)在华盛顿某医院将出生体重在 3.5 磅(1575g)以下的婴儿按入院顺序交替分为两组:第一组给 65%~90% 氧 4~7 周;第二组只给 40% 氧 1~2 周,其他治疗与护理完全相同,追踪观察其 RLF 发生情况,结果如表 2-1-11。

Kinsey 和 Hamphi(1954)报告了美国 18 个医院合作研究的结果。他们将体重 1500g 或 1500g 以下的早产儿按入院顺序随机分为两组:一组为高氧组给予 50% 的氧 28 天,另一组按需要供氧,结果如表 2-1-12。

表 2-1-11 不同浓度氧气治疗与 RLF 的关系

组别	婴儿数	正常婴儿数	RLF 例数		
			轻度	中度	重度
1	23	11	3	7	2
2	37	31	4	2	0

表 2-1-12 不同剂量的氧气对 RLF 发生的影响

治疗	正常(%)	异常(%)	1 级	2 级	3 级	4 级	5 级
急性 RLF (%)							
高氧组	28.0	72.0	25.0	23.0	13.0	9.0	2.0
按需供氧组	70.0	30.0	15.0	8.0	5.5	1.0	0.5
瘢痕 (%)							
高氧组	75.4	24.6	5.7	3.8	3.8	3.8	7.5
按需供氧组	94.0	6.0	1.6	1.6	0.8	1.6	0.4

注:分级愈高,病情愈重

此外还发现,给氧的天数愈长,则 RLF 的发生率愈高。

综上,病因学研究可有多种设计类型,每种设计类型与它的论证强度存在密切关系,一个好的研究设计类型除了满足上述的时间顺序和可重复性要求,还要能较好地控制各类偏倚的干扰,所获结论不易被后来的研究所否定。一般而言,在因果论证强度上,实验性研究大于观察性研究,有对照的研究大于无对照的研究,以个体为分析单位的研究大于以群组为分析单位(生态学)的研究。防治效应的因果研究最好采用随机对照试验。病因研究最好采用前瞻性队列研究,如果有去除病因的干预试验则更好。当然,研究设计类型的选择同研究所处的进展阶段、研究者的资源条件和医学伦理有关。实验性研究尤其需要考虑伦理问题,它有时因为伦理问题难以进行。实验性研究控制偏倚的能力大于观察性研究,研究结论本身更可靠,但实验的条件可能脱离真实生活环境,使它推论到现实情况时受限。而观察性研究因为更接近真实生活环境,推论到现实情况时更可信,但其研究结论本身较容易受到混杂因素的干扰。研究者可根据实际条件和需



要,选择适宜的设计类型进行病因学研究探索。各种病因学研究的论证强度比较见表 2-1-13。

表 2-1-13 各种病因学研究的论证强度

设计类型	性质	可行性	论证强度
随机对照试验	前瞻性	较差	++++
队列研究	前瞻性	较好	+++
病例对照研究	回顾性	好	+
横断面研究	断面	好	+
叙述性研究	前瞻/回顾	好	+/-

知识点

病因和危险因素的区别:病因需要经过因果推导的严格验证,而危险因素是流行病学上发现的增加疾病发生概率的因素。可以说,寻找危险因素是发现病因的前期阶段。

RR 与 OR 的区别:RR 和 OR 都是表示研究因素与疾病结局间关联强度的指标,其区别在于,RR 一般用于前瞻性队列研究或临床试验研究中,而 OR 值是在病例对照研究和横断面研究中运用的 RR 的替代性指标。

RR 与 AR 的区别:RR 与 AR 都是表示关联强度的重要指标,彼此密切相关,但其流行病学意义却不同。RR 说明暴露者与非暴露者比较相应疾病危险增加的倍数;AR 则是指暴露者与非暴露者比较,所增加的疾病发生的数量,如果暴露因素消除,就可减少这个数量的疾病发生。前者具有病因学的意义,后者更具有疾病预防和公共卫生学上的意义。

RR 与 PAR、AR 与 PAR% 的区别:RR 和 AR 是通过比较暴露组与对照组,说明暴露的生物学效应,即暴露的致病作用有多大;而 PAR 和 PAR% 则是通过比较暴露组与全人群,说明暴露对一个具体人群的危害程度,以及消除这个因素后该人群中的发病率或死亡率可能降低的程度,它们既与 RR 和 AR 有关,又与人群中暴露的比例有关。

常用病因学研究设计类型见表 2-1-4。

各种病因学研究的论证强度比较见表 2-1-13。

第二节 评价原则

为增加病因学研究的科学性,提高研究结论的可信度,前人从研究实践中总结了具有指导性意义的评阅标准,为后来者检验自身及他人病因学研究提供借鉴,以提高研究的效率和准确性。

一、真实性

(一) 研究对象是否有明确的定义,各组之间除了暴露因素外,其他重要的方面是否相似、可比

在病因学研究中,对于研究对象应有明确的定义,纳入的病例组和对照组分别是如何选择的,可以代表哪类群体,这关系到研究结果的外推性和外推范围的判定,同时也关系到研究对象内部真实性的判定。另外,需要注意的是,除暴露因素外,病例组或者试验组与对照组之间在人口学特征如年龄、性别、职业、教育程度等重要因素上应相似可比,其目的是避免由于这些重要的非研究因素的组间不同而造成研究偏倚。



病例组选自 1989 年 12 月~1992 年 11 月间在西安市三所医科大学附属的教学医院中的女性冠心病患者,其诊断标准为 WHO 诊断标准确诊的心肌梗死患者或经冠状动脉造影确诊的冠心病患者。

对照组有三个来源:入院初可疑为冠心病,后经冠脉造影证实的冠脉无狭窄者;排除冠心病症状、更年期综合征、心律失常和非心源性胸痛等疾患的内科门诊病人;和一个按国际上推荐方法所设的普查人群中随机抽样的非冠心病对照组。

对三种来源的对照进行人口统计学特征的组间比较,其差异均无显著性意义。因此,三种对照合并为一组与病例组进行下面的分析。经统计检验显示病例组和对照组在婚姻状况、职业和教育程度方面类似,但病例组年龄偏大。病例组的平均年龄为 58.0 ± 5.39 ,大于对照组的平均年龄 55.0 ± 5.05 ($t=3.69, P=0.002$),所有对象均在西安居住二十年以上。

该文献对于病例和对照的选择均有明确的描述,如何选择,入组时间为何,判断标准是什么都有明确的叙述。而后在结果中对除暴露因素(被动吸烟)外的其他重要因素如婚姻状况、职业和教育程度以及年龄都进行了比较,发现病例组和对照组在年龄上存在差异,故而在之后的结果分析中应注意由年龄差异造成的偏倚,将年龄引入多元统计方程进行控制再看暴露因素对结局的影响。

(二) 试验组和对照组有关暴露和结局的测量方法是否相同——是否采用盲法或客观的方法测量暴露和结局

在病因学研究中,对于所致疾病的诊断标准以及测量结果指标的方法在试验组和对照组间应保持一致,而且观测方法宜为盲法,这样才能保证结果的真实性。这是因为,观察者可能不自觉地更关心试验组或者病例组的对象,更主动、更详尽地随访追查是否发生了疾病或者既往有无暴露。同样,发生了疾病的受试对象也更关注自己的病况,更容易回忆过去曾经接触的暴露因素。如果观察者和受试者都不知道研究假设、不知道研究分组的情况,盲法如实记录观测结果,就能避免上述偏倚造成的影响,增加结果的真实性。

文献中有关病例组和对照组的相关描述:病例组选自 1989 年 12 月~1992 年 11 月间在西安市三所医科大学附属教学医院中的冠心病患者,其诊断标准为 WHO 诊断标准确诊的心肌梗死患者或经冠状动脉造影确诊的冠心病患者。冠状动脉造影采用 Judkins 技术并由两名经验丰富的放射科医师分别诊断。确诊冠脉病变的标准是经冠状动脉造影显示至少一支主要冠脉狭窄 $\geq 50\%$ 。两名放射科医师诊断的一致率在 90% 以上。冠脉狭窄 $\leq 50\%$ 者未纳于本研究。

对照组有三个来源:入院初可疑为冠心病,后经冠脉造影证实的冠脉无狭窄者;排除冠心病症状、更年期综合征、心律失常和非心源性胸痛等疾患的内科门诊病人;和一个按国际上推荐方法所设的普查人群中随机抽样的非冠心病对照组。后两组经 WHO 诊断标准排除冠心病并且经心电图平板运动试验检查为阴性。

家庭被动吸烟定义为与吸烟的丈夫共同生活长于 5 年。单身女性的暴露状态与丈夫不吸烟的妻子同视。工作场所被动吸烟定义为不吸烟的女性与某个(些)吸烟的同事在同一办公室或车间共同工作长于 5 年。长期远离工作场所烟草暴露环境者视为无工作场所暴露。所有研究对象的无暴露或暴露年限均大于 5 年。

对 10% 研究对象的调查过程进行同期录音;对 1/3 住院研究对象的丈夫进行再次调查,核对其对妻子调查数据的可靠性;单盲双人重复调查约 30% 医院对照。此外,26 名误诊患者(入院时按初诊冠心病者调查,后经冠脉造影排除的非冠心病者)作为误诊组,与一般对



照组和病例组分别进行被动吸烟暴露史的比较,以检验调查者在询问调查中的主观偏倚。

可以看到在该案例文献中,作者对病例组 and 对照组的结局变量:冠心病,暴露因素:被动吸烟,均有客观统一的评价标准,采用了单盲的方法进行了信息调查,并采用双人重复抽查的方式控制调查质量,对丈夫和妻子的问答进行录音核对,有效地避免了调查偏倚,较好地保证了研究信息的真实性。

(三) 随访时间是否足够长,随访是否完整

研究某些疾病特别是慢性非传染性疾病发病危险因素的致病效应时,由于其自然病程长,往往需要足够时间才能观察到结果的发生,观察期过短则易得假阴性的结果。以“吸烟增加肺癌危险”为例,如果受试者仅被随访了几周,我们则无法区分阴性结果的真实性,是吸烟的确没有增加肺癌罹患的风险?还是随访期短、癌症还没有表现出来?

有一项研究钙通道阻滞剂是否增加患癌症危险的队列研究(Lancet,1996,348:493-497)平均随访了3.7年,在服用钙通道阻滞剂的人群中共随访了1549人年,但仅发生47例癌症。

可见类似研究如果仅随访几周是不会有结果的。

同样,随访中途丢失的病例最好不超过总观察人数的10%,一旦超过20%,则结果很可能失去真实性,因为中途退出的病例可能在某些重要特征上与仍然留在研究中的病例有很大的差别。

(四) 结果是否满足因果推论的要求

1. 是否因在先,果在后,满足因果时间顺序 在评价某一病因学研究时,如果能明确危险因素(或治疗措施)的出现早于疾病(或副作用)的发生,则研究结果的真实性高。如果结局和研究因素同时出现在一个 person 身上,谁是因、谁是果,必须持慎重的态度。例如,糖尿病患者往往有心血管疾病,对谁先谁后不能草率下结论。因果效应时相顺序的确定有赖于前瞻性研究(如队列研究或干预性研究),而回顾性、横断面调查等则不能确定因果效应的时相顺序。也因此,干预性研究和队列研究的因果推断强度强于病例对照研究以及横断面研究。

有研究高血压与高血脂关系的文献描述如下:正常胆固醇(TC)与高TC人群高血压(HBP)患病率比较调查结果显示,高TC血症患病率为18.98%;高TC血症人群HBP患病率为40.80%,明显高于正常TC人群HBP患病率(31.40%),各年龄组患病率随年龄升高而上升,差别均有统计学意义($P<0.05$),见表2-1-14。

表 2-1-14 各年龄组正常 TC 和高 TC 血症人群 HBP 患病率

年龄组(岁)	人数	高 TC 患病率(%)	HBP 患病率(%)	
			正常 TC	高 TC
15~	55	7.27	1.96	0.00
25~	122	13.11	7.55	6.25
35~	204	14.22	12.57	13.79
45~	257	23.35	24.87	38.33
55~	184	22.83	41.55	52.38
65~	179	21.23	64.54	68.42
75~	63	19.04	80.39	50.00
合计	1064	18.89	31.40	40.80



该文献是一个横断面调查,采用了随机整群的抽样方法,对比了两类人群中高血压的发病率差别,发现高血脂的人群中高血压的发病率也高,但是并不能说明高血脂与高血压何者发生在前,何者发生在后,即不能明确区分两调查因素的时间顺序,故无从判断二者之间的因果关联,也不能下何者是何者病因的结论。

2. 有无剂量效应关系 危险因素(或治疗效应)存在剂量效应关系是指其致病效应与有关危险因素的剂量或暴露时间具有显著的相关性。也就是说当病因可以分级处理时(根据量化或特征),随着级别的变化,可以影响疾病在人群的发病率变化。这种关系可以制成相关图,得一形如阶梯的曲线,称剂量效应反应曲线。

在医疗实践中,治疗措施的疗效和毒副作用往往存在剂量效应关系,在治疗剂量时,可显现疗效,但过量时,则出现中毒反应。当病因学研究(或治疗措施副作用研究)呈现剂量效应关系时则结果的真实性较高。

不同暴露水平组间冠心病、缺血性脑卒中、肺癌、慢性阻塞性肺病(COPD)、所有与烟草有关的疾病以及全死因死亡的 Kaplan-Meier 生存曲线比较见图 2-1-1,除缺血性脑卒中($P=0.082$)和肺癌($P=0.107$)外大多数死因的时序检验(Log-rank 检验)结果均具有显著性差异(P 值为 $0.047\sim 0.001$)。

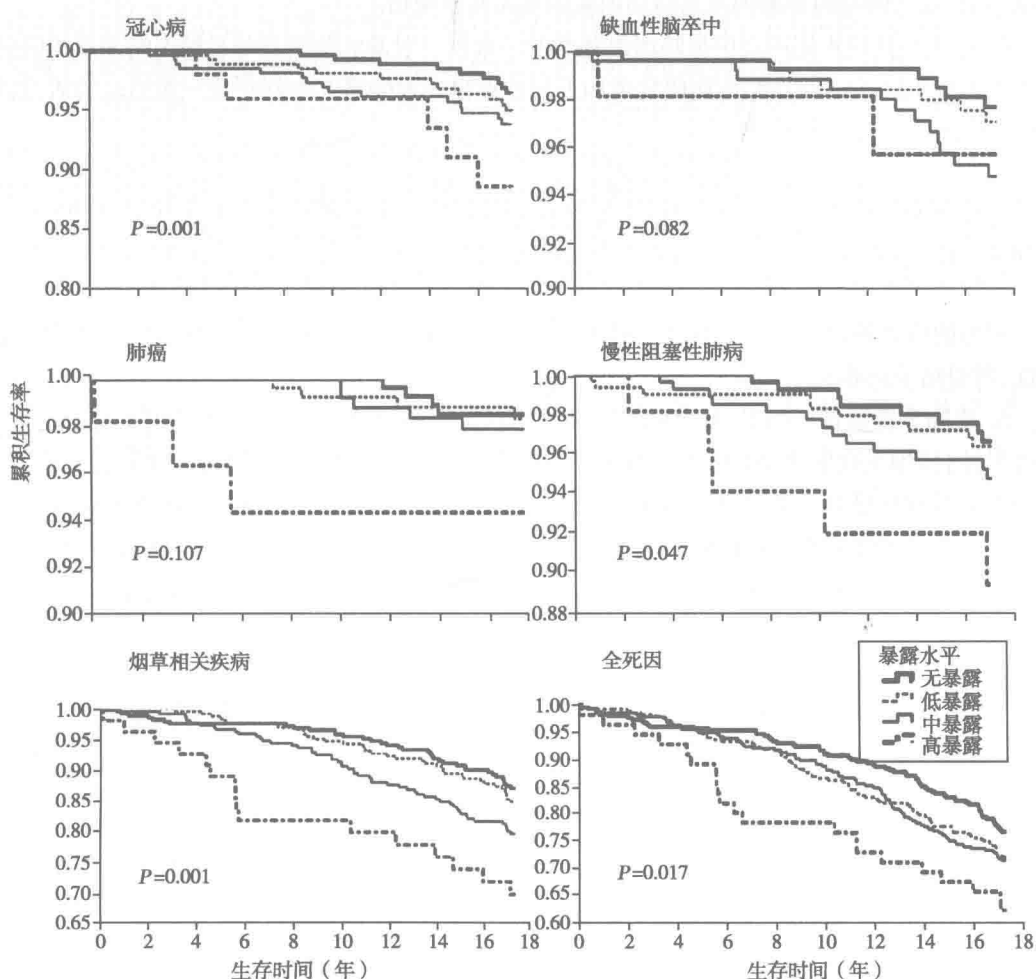


图 2-1-1 二手烟不同水平暴露与冠心病、缺血性脑卒中、肺癌、慢性阻塞性肺病、烟草相关疾病死亡及总死亡的生存曲线比较

该文献为一研究二手烟暴露与慢性阻塞性肺病及烟草相关死亡关系的随访时间长达17年的队列研究,研究者按照二手烟暴露量将人群分组,进行队列研究,绘制冠心病、缺血性脑卒中、肺癌、慢性阻塞性肺病、烟草相关疾病死亡及总死亡与二手烟暴露量关系的生存曲线,发现随着二手烟暴露量的增加,人群中上述各类疾病的死亡也增加。增加了二手烟暴露是上述疾病死亡病因之一的真实性。

3. 是否有停止暴露或减少暴露后发病率下降的研究 疾病在人群中的分布特点和消长的变化,往往与相关的危险因素消长的变化相吻合,当危险因素存在时,该病的发病率和患病率往往较高,反之,当其减弱或消除时,该病的发病率及患病率也随之下降;在治疗措施的副作用研究中,符合流行病学规律表现为终止治疗措施伴随副作用的减弱或消失,重新开始治疗措施时,副作用再次出现。

反应停致胎儿海豹肢畸形的例子,反应停销售高峰时,海豹肢畸形的出生率随后也达高峰,当采取干预措施停止生产和销售反应停后,该畸形的发生率也明显下降,这就是流行病学研究病因致病的规律。

所以,当病因学研究(或治疗措施副作用研究)结果符合流行病学规律,即暴露减少后发病也减少时,则其研究的暴露因素与结局因素因果关系较密切。

4. 有无在其他不同研究中反映因果联系的一致性 对某危险因素与某种疾病关系的研究,如果在不同地区、不同时间、不同研究者和不同的设计方案的研究中都获得一致结论的话,这种病因学的因果效应就较可信。

吸烟引起肺癌的病因学研究,在世界上不同国家和不同时间至少有10次以上的队列研究、50余次的病例对照研究得出相似的结论,说明吸烟与肺癌的因果关系较为真实。

尽可能收集相同性质高质量研究结果,并进行系统评价和Meta分析,可以定量评价结果真实性,所得结果的真实性较高。

5. 生物学合理性 如果病因学研究(或治疗措施副作用研究)揭示的因果关系有生物学的可解释性(病理生理的改变),则可增加因果联系的证据强度和结果的真实性。在上述钙通道阻滞剂是否增加患癌症危险的队列研究(Lancet, 1996, 348:493-497)中,研究者提出这样的假设:钙通道阻滞剂可能干扰细胞的凋亡因而导致癌症的发生。但如果缺乏生物学上的合理解释,否定因果关系时要慎重,因为受科学水平的限制,可能现在无法合理解释的因素,若干年后可能会在其他研究中得到解释。

1747年Lind发现海员的坏血病与食用水果蔬菜有关,百年后才分离出维生素C,最终确定坏血病是维生素C缺乏所致。

同样源于上述RLF病因学研究:为了进一步验证给氧与RLF发病的关系,研究者还开展了动物实验。选择出生后数周的小鸡,其视网膜血管的发育类似临产前胎儿。将小鸡暴露于60%~80%的氧气中数天,发现这样浓度的氧对小鸡视网膜产生严重损害。当小鸡移置于一般空气中,部分血管又得以重新开通。

这个实验从生物学上证明了给氧的确会导致婴儿眼睛的损害,使之罹患RLF,给予了高浓度给氧是RLF病因以生物学证据的支持。



随着当代生命科学的飞速发展,把临床流行病学的宏观研究结果与分子生物学、基因工程等微观研究结果相结合,将促进病因学研究的深化。

二、重 要 性

(一) 因果联系强度的大小

评价病因学研究结果的重要性常借助于一些定性或定量指标。如在临床试验中,定性指标有事件发生率(如病死率、生存率、治愈率等)、相对危险度(RR)、 OR 值、归因危险度百分比(ARP , $AR\%$)、人群归因危险度(PAR)与人群归因危险度百分比($PAR\%$)以及需治疗多少例患者才能发现一例不良反应(NNH)等。对于定量指标,则较为单一,主要是计算组间均数差值。有关主要的定性指标的定义和计算方法我们在第一节中已经进行了探讨,下面以文献中的实例描述来看如何通过定性和定量指标来评价一篇病因学研究。

1954 年至 2009 年间,共有 69 995 例住院及门诊的癫痫患者被纳入此项研究。通过年龄及性别配对,与 660 869 例正常人群和 81 396 例癫痫患者的健康亲属进行对照分析。在性别、年龄、癫痫类型、是否合并精神症状及癫痫后的诊断时间等方面进行了敏感性分析,从而评价各因素的区别。

随访共发现 6155 例(8.8%)癫痫患者死亡,平均年龄为 34.5 岁;死亡危险比大幅度增加(与正常人群的死亡危险比 OR 值为 11.1;与健康亲属相比为 11.4),死亡时间明显提前。过早死亡的患者(972 例)中 15.8% 为外因引起,包括非车祸因素(5.5)及自杀风险(3.7)明显增加。在这些外因死亡的患者中,75.2% 的患者合并有精神疾病、抑郁以及药物滥用情况。

这项研究提示,对于癫痫患者的治疗与管理,应将降低因外因导致的过早死亡作为首要任务。精神合并症状在癫痫患者的过早死亡中扮演了重要角色。

该文献采用的是定性指标 OR 值来表明研究因素在病人过早死亡中的重要性,文中描述:与正常人群相比的 OR 值为 11.1;与健康亲属相比为 11.4,其意义表明合并精神症状的癫痫患者其早死的危险是正常人群的 11.1 倍,是其健康亲属的 11.4 倍,该文献还可报告 ARP 以及 $PAR\%$ 来进一步说明预防该研究因素的人群保护意义。

在讲述 RR 和 OR 值的概念时,我们曾说: RR 或 OR 值越远离 1,则越有价值。一般来说,根据他们的值判断因果关联强度的大小(表 2-1-15)。

表 2-1-15 RR/OR 值及其所表示关联强度

RR/OR 值		关联强度
0.9~1.0	1.0~1.1	无
0.7~0.8	1.2~1.4	弱
0.4~0.6	1.5~2.9	中
0.1~0.3	3.0~9.9	强
<0.1	>10	很强

据此我们可以看出,文献中的研究因素与早死之间的关联是很强的。值得注意的是,现在研究的慢性病大多是社会环境生物等多因素共同作用导致的,故而其单个病因的 RR/OR 值并不会很大,如其值达到 2~3 则可认为关联强度较强。

(二) 研究结果的精确性即可信区间是否较窄

除评价因果关系的联系强度即临床重要性外,还需评价其精确性,即判断研究结果的统计学意义。统计学意义的判定可通过假设检验和区间估计加以实现。若假设检验的 P 值小于预

学
习
笔
记



先设置的检验水准(一般为 0.05),则可认为组间差异有统计学意义。区间估计就是计算 RR 或 OR 值的 95% 可信区间(95%CI, CI: confidence interval),如果 95%CI 范围较狭小、下限和上限值不包括 1.0,则其精度高,有统计学意义。这里要强调的是组间差异的大小与 P 值和 95%CI 无必然联系,若组间差异无临床意义, P 值再小或者 95%CI 再窄也无临床应用价值。

表 2-1-16 列出了孕妇产前服用非甾体抗炎药(NSAIDs)与流产危险的关系;表 2-1-17 列出了孕妇产前服用阿司匹林与流产危险的关系。可以看出,与未服用药物的孕产妇相比,产前服用 NSAIDs 的孕妇流产的危险比是 1.8,服用阿司匹林的危险比是 1.6,数值上相差不大,但是服用 NSAIDs 的孕妇流产危险比的 95%CI 范围不包含 1.0,服用阿司匹林的 95%CI 范围包含 1.0,说明前者具有统计学意义,精确度较好,而后者不具有统计学意义。

表 2-1-16 孕妇产前服用非甾体抗炎药(NSAIDs)与流产危险

NSAIDs	流产		RR (95%CI)
	是($n=162$)	否($n=871$)	
未服用($n=980$)	149	831	1.0
服用($n=53$)	13	40	1.8(1.1~3.2)

表 2-1-17 孕妇产前服用阿司匹林与流产危险

阿司匹林	流产		RR (95%CI)
	是($n=162$)	否($n=871$)	
未服用($n=980$)	149	831	1.0
服用($n=22$)	5	17	1.6(0.6~3.2)

三、实用性

(一) 我们自己的病人是否与文献报道的病人非常不同

这一部分主要是考虑你所要解决的临床案例中的病人是否确实不同于文献中的研究对象,是否存在本质的差别以致文献结果对你的临床决策毫无帮助。

文献报道,骨质疏松在中年人群中男性高于女性,提示中年男性应注意骨质疏松的预防,如果我们面对的是老年女性,她们已进入绝经期失去了雌激素的保护作用,因而对于骨质疏松是较老年男性更高危的群体;如果我们面对的主要就诊群体是老年人,文献中的结论就不适用,对老年女性更应注意提示她们预防骨质疏松。

(二) 病人可以到的益处和危害是什么

参照文献资料估计所要解决的临床案例发生不良反应的可能性大小以及由此使病情可得到改善的可能性。对于不良反应的可能性大小估计,需要首先确定你的病人发生不良反应的可能性与文献报告病人发生不良反应可能性的比值(F),用文献报告的 NNH 除以 F ,即得到你的病人发生不良反应的 NNH 。

如一队列研究显示,使用非甾体抗炎药,发生胃肠道出血副反应的 NNH 为 2000,而你了解你的病人发生该副反应的可能性是该队列研究中研究对象的 4 倍,即 $F=4$,则你病人的 $NNH=2000/4=500$,即每采用非甾体抗炎药治疗 500 例此类病人可能有 1 例发生胃肠道出血。



(三) 病人的意愿、期望

由于不同的病人在权衡治疗措施与不良反应的效应时看法不同,针对具体病人进行治疗决策时,可根据病人意见改变 F 值,然后运用文献数据将 NNH 与 NNT 进行比较,进而进行治疗决策。

使用抗凝剂治疗深部静脉栓塞,如果病人认为发生颅内出血的严重性是发生静脉血栓的 2 倍,则 F 值变为 2,可将文献报告的 NNH 除以 2 得到该患者期望的 NNH 值;反之,使用血管紧张素转化酶抑制剂治疗 AMI,如果病人认为发生不良反应咳嗽的严重性是因 AMI 死亡严重性的 1%,则 F 值为 0.01,将文献报告的 NNH 除以 0.01 得到该患者期望的 NNH 值。

一般来说,如果 $NNT > NNH$,则应考虑改变治疗措施。

(四) 措施是否可行

在实际应用文献研究成果时,还应注意文献介绍的治疗措施应用到我们的医疗实践中是否可行,包括在经济上、技术上以及患者接受程度上是否达到可以运用的程度。

现在有多篇研究证实 BRCA1 基因和 BRCA2 基因会大大增加罹患乳腺癌的风险,而在 2006 年 11 月号的《自然》杂志遗传学分册上,美国研究者表示,携带有 BRIP1 突变基因的妇女患乳腺癌的风险性是不携带该突变基因者的二倍。

据美国癌症学会统计,在美国妇女中,携带 BRCA1 或 BRCA2 突变基因的人数不到 1%。BRCA1 和 BRCA2 的正常功能是肿瘤抑制基因,因此它们突变后将失去抑制恶性肿瘤产生的功能。尽管这种新发现的基因在人群中的突变率很低,但这项新发现能帮助乳腺癌研究者判断在乳腺癌患者中到底有多少归因于该基因的突变。

而中国妇女中,该基因的携带率未见文献报道,检测该基因在技术上可行,但花费不菲,对于绝大多数妇女来说也没有必要,故而即使研究文献证实该类基因为乳腺癌的病因,在实际工作中并不适合推荐来咨询者检测以预测排除乳腺癌。

知识点

病因研究文献的评价原则:

真实性:

1. 研究对象是否有明确的定义? 各组之间除了暴露因素外,其他重要的方面是否相似可比?

2. 试验组和对照组有关暴露和结局的测量方法是否相同? 即是否采用盲法或客观的方法测量暴露和结局?

3. 随访时间是否足够长,随访是否完整?

4. 结果是否满足因果推论的要求? 是否因在先,果在后,满足因果时间顺序? 有无剂量效应关系? 是否有停止暴露或减少暴露后发病率下降的研究? 有无在其他不同研究中反映因果联系的一致性? 生物学合理性如何。

重要性:

1. 因果联系的强度的大小?

2. 研究结果的精确性如何? 即可信区间是否较窄?

实用性:

1. 自己的病人是否与文献报道的病人非常不同?



2. 病人可以得到的益处和危害是什么?
3. 病人的意愿、期望是什么?
4. 措施是否可行?

思考题

一、选择题

1. 病因的定义是
 - A. 引起病理变化的因素
 - B. 只要病因存在,必然会引起疾病
 - C. 引起疾病发生概率升高的因素
 - D. 引起疾病发生的有统计学关联的各种因素
 - E. 引起疾病发生概率升高且符合因果联系标准的因素
2. 某广告声称“用金霉素治疗 1000 名上呼吸道感染的儿童有 900 名在 72 小时内症状消失”。可以得出结论“应选择金霉素治疗儿童的上呼吸道感染”
 - A. 正确
 - B. 不正确,因为没有对照组
 - C. 不正确,因为没有进行统计学检验
 - D. 不正确,因为样本量太小
 - E. 不正确,因为未说明研究对象是如何选择的
3. 一位研究者为了探讨新生儿黄疸的病因,他选择了 100 名患此病的婴儿,并同时在同一所医院选择 100 名未患此病的婴儿,然后查阅了婴儿母亲的孕期保健和产科分娩记录,以确定产前和分娩中各种暴露因素,该研究属于
 - A. 队列研究
 - B. 病例对照研究
 - C. 临床试验研究
 - D. 横断面研究
 - E. 随机对照试验

二、问答题

1. 病因学研究文献的评价标准是什么?
2. 在吸烟与冠心病的病例对照研究中,选择病例与对照各 300 例,病例组中有 120 例吸烟,对照组中有 75 例吸烟。问题:①将上述结果以四格表的形式列出。②写出 OR 值的计算公式,求出 OR 值,并说明意义。

参考答案

一、选择题

1. E 2. B 3. B

二、问答题

2. 答:①吸烟与冠心病的关系四格表。
② $OR = ad/bc = 120 \times 225 / (75 \times 180) = 2$ 。表示:吸烟者发生冠心病的危险是不吸烟者发生冠心病危险的 2 倍(或增加了 1 倍)。



吸烟情况	有无冠心病		合计
	患病	未患病	
吸烟	120	75	195
不吸烟	180	225	405
合计	300	300	600

(何耀 孙业桓)

参考文献

1. 王家良. 循证医学. 第2版. 北京:人民卫生出版社,2001.
2. 李立明. 流行病学. 第6版. 北京:人民卫生出版社,2007.
3. Dan Mayer. Essential Evidence-Based Medicine. New York:Cambridge University Press,2004.
4. Sharon E. Straus. Evidence-Based Medicine. Edinburgh London New York Oxford Philadelphia St Louis Sydney Toronto:Churchill Livingstone Elsevier,2011.
5. Seal S,Thompson D,Renwick A,et al. Truncating mutations in the Fanconi anemia J gene BRIP1 are low-penetrance breast cancer susceptibility alleles. Nature genetics,2006,38(11):1239-1241.
6. 施侣元,李立明. 现代流行病学词典. 北京:人民卫生出版社,2010.
7. He Y,Lam TH,Li LS,et al. Passive smoking at work as risk factor for coronary heart disease in Chinese women who have never smoked. Br Med J,1994,308:380-384.
8. Fazel S,Wolf A,Långström N,et al. Premature mortality in epilepsy and the role of psychiatric comorbidity:a total population study. The Lancet,2013,382(9905):1646-1654.
9. Li D-K,Liu L,Odouli R. Exposure to non-steroidal anti-inflammatory drugs during pregnancy and risk of miscarriage:population based cohort study. Br Med J,2003,327(7411):368-373.
10. 方顺源,朱晓霞,金达丰,等. 高血压与肥胖、高血脂、高血糖及高尿酸血症的关系. 中国慢性病预防与控制,2006,14(1):33-35.

学习
笔记

第二章 诊断试验

临床问题

临床医师每天都需要对患者进行诊断,除了收集患者病史特点和临床体格检查是医师进行诊断的关键步骤之外,怎样为患者选择诊断检查?所选择的诊断试验准确性如何?怎样来评价所做检查的临床价值?这都是临床诊断的关键。为提高临床诊断水平和临床科研质量,临床医师有必要系统地学习和了解临床诊断试验以及诊断研究评价方法,了解和熟悉诊断试验准确性判断的标准,本章的目的正在于此。

第一节 基本概念

诊断试验(diagnostic test)是指应用临床各种试验、医疗仪器等检查手段对就诊的患者进行检查,从就诊者实验室检查结果来诊断和鉴别诊断疾病的试验。

诊断性研究(diagnostic study)是研究对疾病进行诊断的试验方法,包括对各种实验室检查、各种影像学检查以及放射性核素、内镜等诊断方法的研究。诊断性研究可以涉及临床医学各领域及环节。在临床工作中,疾病准确且快速的诊断甚为重要。为了提高临床诊断水平和效率,不仅需要发现高水平的诊断试验应用于临床,同时也需要对现有的诊断试验进行科学的评估和判定,以指导临床医师正确的认识和应用。

一、诊断试验准确性评价的常用指标

(一) 灵敏度与漏诊率

灵敏度(sensitivity, Sen)也称敏感度或真阳性率,指一项诊断试验能将真正有病的人正确诊断为患者的能力,或采用金标准(gold standard)诊断为“有病”的病例中,此项诊断试验检测为阳性例数的比例。

漏诊率是指一项诊断试验将真正有病的人错误地诊断为非患者的比率,也称假阴性率。

灵敏度越大,试验发现阳性的可能性越大,而若试验结果为阴性,则更容易排除被检查者患病的可能,即漏诊率越低。

(二) 特异度和误诊率

特异度(specificity, Spe)也称真阴性率,指一项诊断试验能将真正无病的人正确诊断为非患者的能力。或采用金指标诊断“无病”的例数中,诊断试验结果为阴性的比例。

误诊率指一项诊断试验将实际无病的人错误诊断为患者的比率,也称假阳性率。

真阴性例数越多,特异度越高,误诊率则越低。

(三) 准确性

准确性(accuracy, Acc)指诊断试验中真阳性和真阴性在总检例数中的比例。灵敏度和特异度是诊断性试验的准确性方面的两个基本特征。一个理想的诊断试验,灵敏度和特异度都应达到100%,即假阳性和假阴性均为零。



有研究 A,研究者以 823 例胸部 X 线提示肺部病变的患者其肺 CT 检查结果进行诊断试验准确性研究,以病理作为判定肺癌的金标准,结果见表 2-2-1。

表 2-2-1 肺癌螺旋 CT 诊断研究评价

CT 诊断	真正“肺癌”病人	非“肺癌”病人	合计
阳性	a(385 例)	b(17 例)	a+b(402)
阴性	c(42 例)	d(379 例)	c+d(421)
合计	a+c(427)	b+d(396)	a+b+c+d(823)

$$\text{灵敏度 (Sen)} = a/a+c \times 100\% = 90.16\%$$

$$\text{漏诊率} = 1 - \text{灵敏度} = 9.8\%$$

$$\text{特异度 (Spe)} = d/b+d \times 100\% = 95.7\%$$

$$\text{误诊率} = 1 - \text{特异度} = 4.2\%$$

$$\text{准确性} = 92.8\%。$$

准确性反映诊断试验的灵敏度和特异度,准确性高的试验其灵敏度和特异度也高,但准确性不能反映灵敏度和特异度单方面的特性。在临床实践中,就具体某疾病选择怎样的诊断试验?对于灵敏度和特异度有怎样的要求?

对于周围型肺癌的诊断,由于恶性肿瘤是一类预后极差的疾病,一旦明确诊断,对患者和患者家属无疑带来的是“噩耗”。因此,对于各类恶性肿瘤的明确诊断,势必要求选择特异度高的诊断试验,以降低误诊率。

而如果在 SARS(严重急性呼吸综合征)疫情期间,由于 SARS 的严重传染性和病情发展迅猛性,如果出现临床漏诊,将会出现疫情扩散、病人出现生命危险的严重后果。因此,对 SARS 病人的诊断,就势必要求选择灵敏度高的诊断,以降低漏诊率。

二、诊断试验临床应用评估指标及意义

(一) 阳性预测值与阴性预测值

预测值是评估诊断试验价值的指标,表示诊断试验的临床意义。是指在已知某诊断试验结果为阳性或阴性前提下,正确评判有无疾病的概率,分为阳性预测值(positive predictive value, +PV)和阴性预测值(negative predictive value, -PV)。

阳性预测值是指诊断试验阳性结果中真正有疾病的概率;阴性预测值是指诊断试验阴性结果中真正无病的概率。

(二) 阳性似然比与阴性似然比

似然比是指在诊断试验中患者出现某种检测结果(如阳性或阴性)的概率与非患者出现此种结果的概率之比,说明患者出现该结果的机会是非患者的多少倍,可分为阳性似然比(positive likelihood ratio, +LR)和阴性似然比(negative likelihood ratio, -LR)。

阳性似然比是诊断试验中,真阳性率与假阳性率的比值。表明该诊断试验阳性时,患病与不患病机会的比值,比值愈大,则患病的机会愈大。一般认为 $+LR \geq 10$ 预示该诊断试验具有较高的临床价值。

$$\text{阳性似然比 (+LR)} = \text{灵敏度} / (1 - \text{特异度}) = \text{灵敏度} / \text{误诊率}$$

阴性似然比是诊断试验中,假阴性率与真阴性率的比值。表明判断阴性的可能性是正确判断阴性可能性的倍数,即该诊断试验为阴性时,患病与不患病机会的比值,比值越小,试验的价值越大,一般认为 $-LR \leq 0.10$ 预示该诊断试验有较高的诊断价值。



阴性似然比 $(-LR) = (1 - \text{灵敏度}) / \text{特异度} = \text{漏诊率} / \text{特异度}$

患者 A, 男, 45 岁, 因“右侧胸、背疼痛 6 个月”门诊就诊。患者 6 个月前无明显诱因出现右侧胸痛, 呈隐痛, 伴有右侧对应部位的背部隐痛, 夜间明显, 伴有咳嗽, 多为干咳, 不伴咯血, 气促等。起病来体重减轻约 4kg。吸烟近 30 年, 15 支/日。既往常有咳嗽。门诊医师给予患者肺部 CT 检查, 发现“右下肺团块状结节影, 周围型肺癌可能性大”。患者 A 完成 CT 检查并确定为阳性, 该患者真正患肺癌的概率有多大? 以表 2-1-1 诊断研究的结果为例。

$$\text{阳性预测值 (+PV)} = [a / (a+b)] \times 100\% = 95.77\%$$

$$\text{阴性预测值 (-PV)} = [d / (c+d)] \times 100\% = 90\%$$

$$\text{阳性似然比 (+LR)} = \text{灵敏度} / (1 - \text{特异度}) = 20.96$$

$$\text{阴性似然比 (-LR)} = (1 - \text{灵敏度}) / \text{特异度} = 0.10$$

从此研究结果得出结论: 从阳性预测值的结果解释, 患者 A 可能患肺癌的概率是 95.77%。肺部 CT 检查有较高的阳性似然比, 患者 A 为患肺癌的几率是不患肺癌的 20 倍。

(三) 患病率

患病率 (prevalence, Prev) 是指诊断试验的全部例数中, 真正“有病”例数所占的比例。灵敏度和特异度提示有病患者和无病对象出现诊断试验阳性和阴性结果的机会分别有多大, 但临床医师需要了解的是, 诊断试验结果阳性或阴性时, 患病和不患病的机会有多大, 即诊断试验结果对疾病判断的可靠性怎样。诊断试验的阳性结果并不意味着肯定“有病”, 阳性结果提示有病的机会取决于诊断试验的灵敏度、特异度以及检测人群中该病的患病率 (阴性结果提示“无病”的机会也同样受到这些因素的影响)。但患病率对预测值的影响要比灵敏度和特异度的影响更为重要。受试对象的患病率愈低, 阴性结果提示无病的把握愈大, 而阳性结果提示有病的把握愈小。

当诊断试验用于患病率很低的人群时, 即使灵敏度很高, 阳性预测值也不会很高, 即在阳性结果中可能存在较多的假阳性。同样, 当用于患病率很高的人群时, 即使特异度很高的诊断试验, 阴性结果中仍然会有不少假阴性结果的出现。

因此, 临床医师必须注意到, 在解释诊断试验阳性或阴性结果时应考虑不同地区、不同人群, 估计患病率的高低, 因受试对象的差异, 其试验结果的意义将存在区别。

有研究 B, 研究者对肺癌高危人群 126 例体检对象肺部 CT 的结果进行分析, 见表 2-2-2。

表 2-2-2 肺部 CT 诊断肺癌研究的评价

CT 诊断	真正“肺癌”病人	非“肺癌”病人	合计
阳性	a (5 例)	b (8 例)	a+b (13)
阴性	c (2 例)	d (111 例)	c+d (113)
合计	a+c (7)	b+d (119)	a+b+c+d (126)

表 2-2-2 结果与表 2-1-1 研究结果对比, 见表 2-2-3。

表 2-1-3 结果可见, 在不同患病率的群体中, 肺部 CT 检查作为一种诊断试验准确性几乎相同的情况下, 诊断试验阳性或阴性结果对临床病人诊断的判断, 即阳 (阴) 性似然比、阳 (阴) 性预测值有很大的差别。



表 2-2-3 两个患病率不同的群体肺部 CT 诊断试验评价比较

研究	患病率	灵敏度	特异度	准确性	阳性似然比	阴性似然比	阳性预测值	阴性预测值
研究 A	97.2%	90.2%	95.7%	92.8%	20.96	0.10	95.7%	90%
研究 B	5.5%	71.4%	93.2%	92.0%	10.14	0.31	38.8%	98%

(四) 验前概率

验前概率 (pre-test probability) 是指临床医师在应用某一诊断性试验前估计该疾病的患病概率,即该疾病的患病率,通常可以从文献中查找得到。

(五) 验后概率

验后概率 (post-test probability) 是指在进行完某一诊断性试验后,根据需诊断的疾病的验前概率和该诊断试验的阳性似然比计算而得出,即在不同患病率情况下该试验的预测值。

单纯的灵敏度和特异度不能反映诊断性试验的全貌,而似然比通过比较有病人群和无病人群出现该诊断试验的阳性和阴性结果的概率来反映该诊断试验的特征,不受患病率的影响,较单纯的灵敏度和特异度更全面和更稳定地诠释诊断试验。因此,似然比是诊断性试验综合评价的理想指标,能够依据灵敏度、特异度、试验的阳性或阴性结果,计算出患病的概率,即验后概率,便于在诊断试验之后,更确切地对患者作出诊断。

患者 A 为表 2-1-1 研究中的一位受试者,那么在肺部 CT 阳性结果的前提下,该患者可能患肺癌的概率计算如下:

验前概率(患病率) 97.2%

验前比数 $= 0.972 / (1 - 0.972) = 34.7$

验后比数 = 验前比数 $\times$ 阳性似然比 $= 727.3$

验后概率 = 验后比数 $/ (1 + \text{验后比数}) \times 100\% = 99.8\%$

结论:患者 A 在此肺部 CT 阳性的结果下,患肺癌的概率从验前的 97.2% 上升至验后的 99.8%。

患者 A 为表 2-1-2 中的一位受试者,那么在肺部 CT 阳性结果的前提下,该患者可能为真正的病人的概率计算如下:

验前概率(患病率) 5.5%

验前比数 $= 0.055 / (1 - 0.055) = 0.058$

验后比数 = 验前比数 $\times$ 阳性似然比 $= 0.59$

验后概率 = 验后比数 $/ (1 + \text{验后比数}) \times 100\% = 37.1\%$

结论:患者 A 在此肺部 CT 阳性的结果下,患肺癌的概率从验前的 5.5% 上升至验后的 37.1%。

(六) ROC 曲线

ROC 曲线又称受试者工作曲线(receiver operator characteristic curve)在诊断试验中,通过多次连续分组测定的数据进行制图。制图时以该试验灵敏度(真阳性率)为纵坐标,以 1- 特异度(假阳性率)为横坐标,依据连续分组测定所得数据,分别计算出的灵敏度和特异度标入图中,连成曲线,即为 ROC 曲线(图 2-2-1)。曲线上的任意一点代表某项诊断试验的特定阳性标准值所相对应的灵敏度和特异度对子。

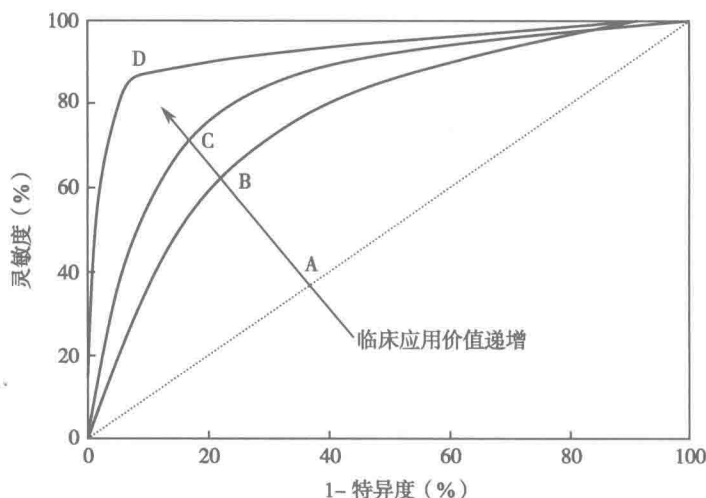


图 2-2-1 ROC 曲线示意图

以图 2-2-1 ROC 曲线为例, A、B、C、D 四条曲线分别代表四个诊断试验: 由 A-D, 诊断试验的临床价值逐渐递增。好的诊断试验其曲线远离中间的对角线, 并十分接近左上角(如图中所示 D 曲线)。差的试验则非常接近中间的对角线(如图中所示 A 曲线)。通常可以通过计算曲线下的面积(AUC)来评价试验的准确性, 面积越大, 表示试验越好。同时, 一般选择以曲线距左上角最近的一点为正常值的最佳临界点, 这点下的面积最大。用该点区分正常与异常, 其灵敏度和特异度均为最高, 即误诊和漏诊病例最少。

知识点

诊断试验性能指标: 灵敏度、特异度、阳(阴)性预测值、阳(阴)性似然比、验前概率、验后概率、ROC 曲线与曲线下面积。

灵敏度与特异度临床应用区别: 灵敏度高的试验主要用于排除无病的对象, 此时阴性结果最有意义。特异度高的试验主要用于确诊有病的对象, 此时阳性结果最有意义。

预测值与似然比的区别: 阳(阴)性预测值是试验阳性(阴性)时患病(不患病)的概率, 其受患病率影响。而阳(阴)性似然比是该诊断试验阳性(阴性)时, 患病与不患病机会的比值。不受患病率影响。

ROC 曲线的临床运用: 绘制 ROC 曲线可用来决定正常值, 还可以通过曲线下面积比较不同诊断试验的优劣。

第二节 评价原则

为增加诊断试验研究的科学性, 提高诊断试验研究结论的可信度, 前人从研究实践中总结了具有指导性意义的评阅标准, 为后来者检验自身及他人诊断试验研究提供借鉴, 以提高诊断试验研究的效率和准确性。评阅标准有 8 个部分(表 2-2-4)。

表 2-2-4 诊断试验的文献评阅标准

1. 诊断试验与金标准盲法比较和评价	5. 诊断试验所确定的正常值是否合理、可靠
2. 纳入研究病例的分析与评价	6. 系列试验的应用是否正确
3. 病例来源和研究工作的安排是否叙述	7. 诊断试验方法叙述及其重复性分析
4. 诊断试验的重复性及其临床意义是否明确	8. 诊断试验的实用性如何



(一) 是否将研究的诊断试验与金标准进行了盲法比较

诊断性研究中诊断试验的金标准(gold standard)是指当前临床医师公认的诊断疾病最可靠的方法,也称为诊断标准,它能正确区分“有病”与“无病”。临床诊断常用的金标准,包括病原学诊断、病理学诊断、外科手术发现、特殊的影像学诊断、临床综合性诊断标准、长期临床随访所得出的肯定结果等均可作为金标准,但应注意,有些疾病诊断的金标准是相对的,又被称为相对金标准。随着医学的发展以及对疾病进一步深入的认识,相对金标准在不断地进行完善中。

金标准的选择及其可靠性直接影响到诊断性研究结论的可靠性,对同一组对象,采用不同的诊断金标准可能对某一诊断试验得出不同的结论,因此必须严格掌握。

诊断试验的准确性评定,即诊断性试验是否准确反映检查对象有病或无病,是研究的关键。诊断性试验准确性评定主要是与金标准比较而言。为了消除人为偏倚,应用盲法对比则更为科学。因此,被研究的诊断性试验在临床应用之前,应与金标准进行盲法比较,根据四格表计算出灵敏度、特异度、阳性和阴性预测值、准确性和似然比等。

某诊断性研究采用痰培养出结核菌结果作为诊断肺结核的金标准,对血液结核 PCR 诊断试验的结果进行评价(表 2-2-5)。

表 2-2-5 患者结核 PCR 诊断试验结果

TB-PCR	检出结核菌	未检出结核菌	合计
阳性	45	5	50
阴性	8	62	70
合计	53	67	120

此研究以病原学诊断为金标准,对结核的血液学诊断试验进行科学的评价。

(二) 研究中纳入病例的选择是否有代表性

诊断试验的研究对象,应当包括用金标准确定“有病”的研究对象,即病例组;和用金标准证实“无病”的研究对象,其中所谓“无病”的研究对象,是指没有金标准诊断的目标疾病,而不是完全无病的正常人。

不同应用目的的诊断性研究应选择不同的研究对象,如评价筛选诊断试验价值的,研究对象可选择普通人群;评价临床诊断试验价值的,则应选择病例作为研究对象。同时,病例还应包括不同病期、不同病情程度(轻、中、重)、有无并发症以及典型和非典型病例等,以使研究的结果具有代表性。对照组可选择金标准确认“无病”的其他病例,最好包含由金标准确认无该病,但有易于与该病混淆的其他病例,这样的对照更具临床诊断,特别是鉴别诊断价值。正常人一般不宜纳入“无病”对照组,如果在试验的初级阶段纳入了正常人进入对照,则应对其结果谨慎对待。

在上面对血液 TB-PCR 诊断试验的评估研究中,纳入研究并接受血液检测的受检人群包括轻、中、重度肺结核患者,也包括原发型、继发型肺结核各型患者,这样得出对该诊断试验的评价结果才科学可信。

(三) 研究对象的来源是否正确叙述

患病率对诊断试验的预测值等最后分析结果有很大的影响。研究对象不同的来源,即各级医院就诊患者不同的结构组成,势必出现研究对象某疾病的患病率的不同,最后对研究结果产



生较大的影响。对于某一诊断试验其灵敏度和特异度是相对固定的,而在临床工作中,我们更注重的是诊断试验的阳性预测值,同时又要减少假阳性例数,希望通过该诊断试验能提高对疾病诊断的预测,即充分发挥诊断试验的效率。由于患病率对疾病的预测值将产生较大的影响,因此,选择不同患病率的人群来应用适当的诊断性试验,是临床医师提高诊断性试验效率的关键之一。

肺部螺旋 CT 诊断试验在不同患病人群的应用结果就有几种不同的预测值:①在发现肺部结节的人群中,验前估计患病概率为 97.2%,验后阳性预测值为 99.8%,应用肺部 CT 试验后,预测仅增加 2.6%;②在高危的吸烟体检人群,验前估计患病概率为 5.5%,验后阳性预测值为 37.1%,应用该试验后,预测增加 32.1%;因此,研究中应正确叙述研究对象的来源,以增加研究结果的科学性和论证强度。

(四) 诊断性试验是否具有很好的重复性

重复性(repeatability)又称精密性(precision)或可靠性(reliability),即诊断试验需重复操作,其重复性是诊断试验临床应用的先决条件。诊断试验要求重复测定值应处于相对稳定状态,即多次测定同一标本,应结果接近,方法可靠。

(五) 诊断性试验的正常值的确定是否合理、可靠

正常值的含义应准确定义及清楚叙述,不同含义的正常值可直接影响正常值的数据。当有病人群和无病人群测定值的频数分布曲线有重叠时,应采取科学的方法,(如正态分布数据采用均数 $\pm 2s$,非正态分布用中位数或百分位数以及应用 ROC 曲线等),正确地划分正常和异常的临界点,即找到最佳临界点。是否为最佳临界点直接影响到诊断性试验的灵敏度和特异度以及其他重要指标。

(六) 联合试验的选择是否合理、科学

诊断试验的联合方式包括:平行试验(parallel tests)和序列试验(serial tests)。

1. 平行试验 为提高诊断的灵敏度,同时做几种目的相同的诊断试验,只要其中一种试验阳性,即可判断为患者。平行试验的应用可提高灵敏度和阴性预测值,但却降低了特异度和阳性预测值,也就是减少了漏诊率,却增加了误诊率。这种方法在临床应用时,需谨慎考虑鉴别诊断,尽量减少误诊。

平行试验(parallel tests)的评价方法见表 2-2-6。

表 2-2-6 平行试验结果的评价方法

项目	结果		评价结果
	试验 A	试验 B	
平行试验	+	-	+
	-	+	+
	+	+	+
	-	-	-

假设:试验 A:Sen = 80% Spe = 60%

试验 B:Sen = 90% Spe = 90%

平行试验计算:

$$\text{Sen} = \text{SenA} + (1 - \text{SenA}) \times \text{SenB} = 0.80 + (1 - 0.80) \times 0.90 = 0.98$$

$$\text{Spe} = \text{SpeA} \times \text{SpeB} = 0.60 \times 0.90 = 0.54$$



某诊断性研究对胸腔积液患者进行胸水的 PPD、LDH 和 ADA 三项指标进行检测,以此诊断胸腔积液患者结核性胸腔积液的可能见表 2-2-7。

表 2-2-7 胸腔积液患者胸水诊断试验结果

ADA 试验 (A)	LDH 试验 (B)	结核性胸水患者	非结核性胸水患者
+	-	8	4
-	+	16	6
+	+	60	2
-	-	16	578
合计		100	590

ADA 试验 (A):

$$\text{Sen} = (8+60)/100 \times 100\% = 68\%$$

$$\text{Spe} = (6+578)/590 \times 100\% = 98.9\%$$

LDH 试验 (B):

$$\text{Sen} = (16+60)/100 \times 100\% = 76\%$$

$$\text{Spe} = (4+578)/590 \times 100\% = 98.6\%$$

平行试验计算:

$$\text{Sen} = \text{SenA} + (1-\text{SenA}) \times \text{SenB} = 0.68 + (1-0.68) \times 0.76 = 0.92$$

$$\text{Spe} = \text{SpeA} \times \text{SpeB} = 0.98 \times 0.98 = 0.96$$

此结果表明,平行试验增加了试验的灵敏度,但降低了试验的特异度。

2. 序列试验 当临床缺乏单一的特异度高的诊断试验时,为提高临床诊断的特异度,设计一系列的诊断试验,决定前后顺序进行,结果必须一系列试验均为阳性,才能作出患病的诊断。序列试验 (serial tests) 提高了特异度和阳性预测值,但同时降低了敏感度和阴性预测值,即临床上序列试验降低了误诊率,却也可能增加了漏诊机会。当临床所用的多项诊断试验特异度都不高时,可采用序列试验比较适当。

序列试验 (serial tests) 的评价方法:

(1) 必须设计的一系列诊断试验均为阳性。

(2) 计算方法:

A, B, C, D, …… 系列试验

$$\text{Sen}_{(\text{序列})} = \text{SenA} \times \text{SenB} \times \text{SenC} \times \text{SenD} \cdots \cdots$$

计算出序列试验的敏感度;

$$\text{Spe}_{(\text{序列})1} = \text{SpeA} + [(1-\text{SpeA}) \times \text{SpeB}]$$

$$\text{Spe}_{(\text{序列})2} = \text{Spe}_{(\text{序列})1} + \{ [1-\text{Spe}_{(\text{序列})1}] \times \text{SpeC} \}$$

$$\text{Spe}_{(\text{序列})3} = \text{Spe}_{(\text{序列})2} + \{ [1-\text{Spe}_{(\text{序列})2}] \times \text{SpeD} \} \cdots \cdots$$

按此规律,最后计算出序列试验的特异度。

准确评价联合试验的敏感度、特异度和准确性,同时,应和每个单项试验的敏感度、特异度和准确性进行比较。在一系列同类试验中评比选择最佳联合试验方法。

以表 2-2-7 中检测结果作为联合诊断试验的计算基础,序列试验后的灵敏度和特异度的计算为:

$$\text{Sen}_{(\text{序列})} = 0.68 \times 0.76 = 0.516$$



$$\text{Spe(序列)}1 = 0.989 + [(1 - 0.989) \times 0.986] = 0.999$$

此结果表明,序列试验增加了试验的特异度,但降低了试验的灵敏度。

(七) 诊断性试验的操作方法是否仔细叙述

为使被研究的诊断性试验得以在临床应用,应对其具体操作方法进行仔细地叙述,包括操作步骤、使用仪器及试剂规格、试验前、后被检者有何安排或限制(如饮食、药物等限制)以及结果评判方法等。

(八) 诊断性试验的临床实用性如何

综合评定被研究的诊断试验的临床实用性,包括临床应用是否方便、对病人有无伤害、准确度是否高、结果是否容易判断以及成本效益分析是否好。

知识点

金标准的选择:金标准的选择及其可靠性直接影响到诊断性研究结论的可靠性,必须慎重选择。

研究对象的选择:研究对象的选择直接影响诊断试验的灵敏度和特异度。病例组应包括不同严重程度的病例,对照组应包括由金标准确认无该病,但有易于与该病混淆的其他病例,而不是健康正常人。了解研究对象的来源,才能决定是否可以用于纳入研究的病人。

平行试验与序列试验:平行试验的应用可提高灵敏度和阴性预测值,但却降低了特异度和阳性预测值;序列试验提高了特异度和阳性预测值,但同时降低了敏感度和阴性预测值。

思考题

1. 临床医师在应用诊断试验诊断某患者时,下列哪项指标是临床应用价值意义最大的

- A. 敏感度
- B. 阳性预测值
- C. 阳性似然比
- D. 患病率
- E. 验后概率

2. 某镇人口 10 000 人,估计糖尿病的患病率为 2%,用检查血糖含量的方法诊断糖尿病患者。并规定以血糖含量以 10mmol/L 及以上为阳性,试验的灵敏度和特异度分别为 30% 和 90%。①请绘制四格表,并填入近似数据。并计算下列数值:假阳性率、假阴性率、+PV 和 -PV。②若诊断试验的灵敏度增加,血糖诊断水平为 7.2mmol/L,灵敏度为 80%,特异度为 80%,请绘制四格表,并填入相应数据,计算假阳性率、假阴性率、+PV 和 -PV。③从问题①与②所得数值你看出什么问题? 如果让你制订糖尿病普查计划,你采用 7.2mmol/L 还是 10mmol/L 为诊断标准? 请说明理由。④若诊断试验的标准仍为 7.2mmol/L(试验的灵敏度、特异度均为 80%)但人群患病率较原来为高,约 4%。调查 10 000 人口,请绘制四格表,计算假阳性率、假阴性率、+PV 和 -PV。⑤从问题②和③所得数据,可以看出什么问题?

参考答案

1. E

2. 答:① 绘制四格表及假阳性率、假阴性率、+PV 和 -PV 计算如下:



血糖检测	真正“糖尿病”	非“糖尿病”	合计
≥10mmol/L	60	980	1040
<10mmol/L	140	8820	8960
合计	200	9800	10 000

假阳性率(误诊率)=1- 特异度 =1-90%=10%

假阴性率(漏诊率)=1- 灵敏度 =1-30%=70%

+PV=60/1040=0.0576

-PV=8820/8960=0.984

② 绘制四格表及假阳性率、假阴性率、+PV 和 -PV 计算如下：

血糖检测	真正“糖尿病”	非“糖尿病”	合计
≥7.2mmol/L	160	1960	2120
<7.2mmol/L	40	7840	7880
合计	200	9800	10 000

假阳性率(误诊率)=1- 特异度 =1-80%=20%

假阴性率(漏诊率)=1- 灵敏度 =1-80%=20%

+PV=160/2120=0.0754

-PV=7840/7880=0.9949

③ 从问题①与②所得数值可以看出,建议采用血糖检测≥7.2mmol/L 为诊断糖尿病指标。因为采用≥7.2mmol/L 标准,漏诊率与误诊率均较低。

④ 若诊断试验的标准仍为 7.2mmol/L (试验的灵敏度、特异度均为 80%)但人群患病率较原来为高,约 4%。调查 10 000 人口,绘制四格表,计算假阳性率、假阴性率、+PV 和 -PV 如下：

血糖检测	真正“糖尿病”	非“糖尿病”	合计
≥7.2mmol/L	320	1920	2240
<7.2mmol/L	80	7680	7760
合计	400	9600	10 000

假阳性率(误诊率)=1- 特异度 =1-80%=20%

假阴性率(漏诊率)=1- 灵敏度 =1-80%=20%

+PV=320/2240=0.1428

-PV=7680/7760=0.9896

⑤ 从问题②和③所得数据,可以看出患病率影响诊断试验的阳性、阴性预测值。

(王吉耀 吴尚洁)

参考文献

1. Straus SE, Glasziou P, Richardson WS, et al. Evidence-based medicine. 4th ed. Elsevier; Churchill Livingstone, 2011.

2. 王吉耀. 循证医学与临床实践. 第 3 版. 北京: 科学出版社, 2012.

3. Dan Mayer. Essential Evidence-based Medicine. UK: Cambridge University Press, 2004.

4. 王家良. 临床流行病学. 第 2 版. 上海: 上海科学技术出版社, 2001.

5. 刘爱忠, 黄民生. 临床流行病学. 第 2 版. 长沙: 中南大学出版社, 2010.

6. 陈立章, 吴尚洁. 循证医学与实践. 长沙: 中南大学出版社, 2012.

7. 李建, 阿不都艾尼·吐尔洪, 居来提·艾尼瓦尔, 等. 不同检测方法对周围型肺癌诊断的意义. 新疆医学,

学
习
笔
记



2011,41 (5):12-14.

8. 蒋晶,低剂量螺旋 CT(LDCT)在肺癌筛查中的临床价值. 中国医药指南,2013,11 (7):227-228.

9. Department of Clinical Epidemiology and Biostatistic,McMaster University Health Sciences Centre . How to read clinical journals; IV. To determine etiology or causation. Can Med Assoc J, 1981, 124:985-990.

第三章 治疗效果

临床问题

有一位中年男性,活动后感胸痛,冠脉造影显示冠脉前降支狭窄最严重处达 75%,诊断为冠心病、稳定性心绞痛。临床医师明确诊断后面临最重要的问题是采用什么样的治疗方法?该治疗方法的有效性如何?经该方法治疗后能否降低不良事件的发生?此类问题均为循证医学治疗性研究的相关问题,应用 PICO 模式可将此临床问题转化为:稳定性心绞痛患者(P),应用介入+药物治疗或药物治疗(I、C),治疗效果(疼痛症状缓解、死亡、心脑血管事件发生率)(O)如何?

第一节 基本概念

临床医师在对疾病作出正确的诊断后,面临的下一个重要问题就是如何对患者进行最安全、有效的治疗。鉴于目前新药、新疗法的不断出现,使疾病的治疗药物和治疗手段趋于多样化,如何判断此治疗方法是否有效以及疗效如何,需要从治疗性研究中寻找答案。

治疗性研究与评价是在临床实践中以患者为研究对象,应用医学科研的理论和方法,通过科学、严谨的设计和精确的测量对所研究或选择治疗的效果进行客观的评价,以达到提高治愈率,降低病残率及病死率,提高生存质量,改善人类健康的目的。

治疗性研究的特点主要有 3 个方面:①研究对象具有特殊性。治疗性研究的对象是患者,首先患者具有社会属性,受个体的生理特点、心理状态、文化水平、经济水平以及所处的自然和社会环境的影响,同一种疾病其临床表现相差可能十分显著,对同一治疗方案,治疗效果也可能不同,且因其心理状态、经济水平和文化水平,还可能影响患者对研究的依从性。其次,必须保证患者安全的前提下进行临床试验,且务必符合伦理学要求,《赫尔辛基宣言》中指出:“凡涉及人的生物医学实验,必须遵循科学的原则。应建立在足够的实验室和动物实验及科学文献认识的基础之上。”因此,治疗性研究要求更高、实施难度更大。②必须设立对照组。治疗性研究存在很多干扰或混杂因素,如果不去掉这些混杂和干扰因素,很难判断结果的真实性。因此,通过设置对照,不但可以排除一些混杂和干扰因素对研究的影响,而且能够增加实验结果的可信度和说服力。③必须有人为的干预。干预措施是根据研究目的确定的欲观察或欲施加的,并能够作用于患者引起直接或间接效应的措施,治疗性研究干预措施的目的是人为干预疾病的自然病程,治疗性研究的评价即为评价干预效应。

1986 年 V-HeFT I 的随机双盲研究,入选 642 例慢性心力衰竭患者,分为肼屈嗪+二硝酸异山梨醇酯组、呋塞米组及安慰剂组,分别给予肼屈嗪+二硝酸异山梨醇酯、呋塞米和安慰剂后随访 3 年,观察终点——死亡率。结果发现肼屈嗪+二硝酸异山梨醇酯与安慰剂或呋塞米相对能够降低心力衰竭的死亡率。

临床治疗性试验是一种前瞻性和实验性研究的方法,在人体通过比较实验组和对照组的结



果,从而确定某项干预措施的效果,按对照设立的不同,分为以下7种类型,其中以随机对照试验最常用,其效率、可信度均较高。

一、随机对照试验

随机对照试验(randomized control trials, RCTs)是应用随机分配的方法,将合格的研究对象随机地分配到实验组和对照组,同时接受相应的试验措施,在一致的条件下或环境中同步的进行研究和观测试验的效应,并通过客观的效应指标对试验结果进行科学的测量和评价。RCTs方案采用了随机分组、设立对照、实施盲法等措施确保了研究对象的代表性,防止了若干混杂因素或偏倚因素的干扰,保证了两组之间的均衡可比性,目前被公认为治疗性研究的金标准方案。RCTs的设计流程如图2-3-1所示。

(一) 研究对象的来源

治疗性研究不可能把所有希望研究的患者均纳入研究,只能从总体人群中选择一定的研究人群,通常为确保总体中每一位患者均具有相同的被选择的机会,同时使本次科研结果具有总体的代表性,研究可以采用随机抽样(random sampling)方法,目前随机抽样方法主要采用的有简单随机抽样、系统随机抽样和分层随机抽样3种。

(二) 诊断、纳入、排除标准

治疗性研究的试验对象均为某种疾病的患者,因此必须有公认、明确的诊断标准及比较客观的诊断指标,尽可能按照国际疾病分类标准或全国性学术会议规定的标准来选择研究对象,但依照诊断标准明确的病例不一定均选为研究对象,还应该根据研究目的和具体条件通过纳入、排除标准界定和限制研究对象的范围,同时纳入标准也界定了研究者希望该干预措施或该研究的结果对未来所适用患者的范围。

(三) 随机分组

随机分组(random allocation)与前述随机抽样目的不同,前者是通过随机的方法,每个研究对象都有相同的机会进入治疗组和对照组,使治疗组和对照组已知和未知的影响因素均匀地分布于两组,即确保两组基线可比;同时不受研究者或患者的主观意愿的干扰,最终达到治疗组和对照组均衡可比的目的。

在临床治疗性研究中成功实施随机分配不但需要科学的随机化方法,而且需要在实施过程中隐藏随机分组(concealed random allocation),目前隐藏随机分组的方法有按顺序编码、不透光、密封的信封、中心随机系统、编号或编码的瓶子或容器、中心药房准备的药物等,以前3种方法最为常用。隐藏随机分组可以控制选择偏倚,使试验结果更具有真实性。有研究表明,一般情况下,与充分使用随机隐匿的临床试验结果相比,不充分随机隐匿的或者随机隐匿方案不清楚的临床试验结果往往会夸大效应估计值。

(四) 设立对照

设立对照、随机分组与盲法(blindness)是随机对照试验的三大基本原则。设立对照即设立

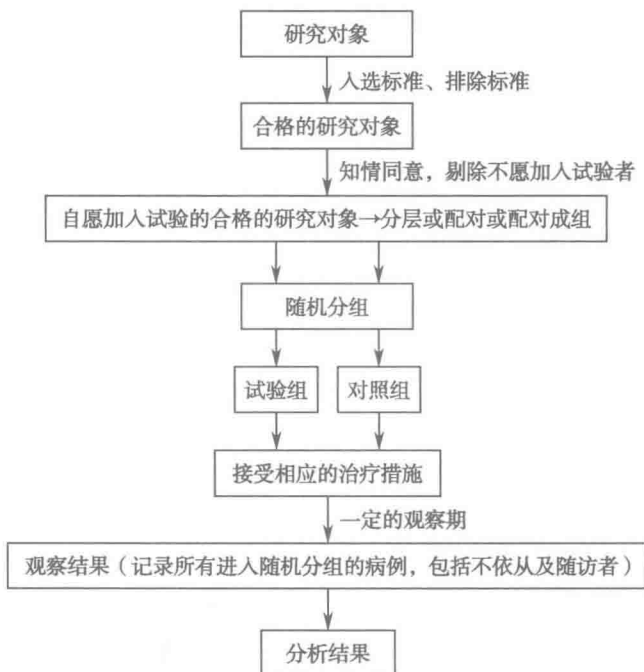


图 2-3-1 RCTs 设计流程图



条件类似、诊断一致的一组对象,接受某种与治疗组不相同的措施,目的是和治疗组的结果进行对照比较,以证明两组或多组间结果的差异及程度。该组研究对象除不接受干预措施外,其他方面如试验条件、观察指标和效应标准等均应与治疗组相同。这种以对照比较的方法可以消除一些影响疾病转归的因素,如疾病的自然转归作用、回归中位作用、霍桑效应、安慰剂效应等,使两组间临床结局之差真实地反映治疗特异作用的大小。

(五) 盲法

盲法 (blindness) 治疗性研究一般以人为研究对象,而人(研究者或医师、患者、资料收集和分析者)的主观意识不可避免地会对研究结果产生某些干扰作用,为了控制这种干扰作用,在实施过程中最好采用盲法。盲法对于判断的结果是某些软指标(如疼痛、乏力)时更为重要,如评价某一药物对心绞痛的疗效,医师可能为了得到阳性结果而暗示患者,而患者为了迎合医师而导致该指标的可信度下降,产生偏倚。盲法一般分为单盲(single blind)、双盲(double blind)和三盲(triple blind)3种。

(六) 选择效应指标

治疗性研究选择研究指标时需要注意以下几点:①选用的指标需与回答的主要问题有密切的关系,即指标的关联性。②尽量多用客观指标,少用主观指标。③尽量选用灵敏度和特异度较高的指标。④在选择指标时,也应该注意指标的类型,一般而言,计量资料均为客观测量的定量指标,产生的观察性偏倚较小;计数指标为定性指标,若依靠主观判断来定性计数,则易产生偏倚;半定量指标用于等级资料。⑤选择指标时还需考虑经济成本、损伤性和可接受性。⑥此外,循证医学提倡的主要观察指标是满意的终点指标,即重要事件的临床发生率、病死率、致死率、生存率等。这也是医生和患者最关心的治疗结果,更具客观性和可靠性。

(七) 干预措施

治疗性研究中对要研究的治疗药物或疗法称为干预措施。干预措施的不同等级称为干预措施的水平,在整个治疗性研究中,要求干预措施及其水平在整个试验过程中始终如一,按统一标准进行。

稳定性心绞痛患者可行冠脉造影明确诊断,若狭窄程度严重,需行经皮冠状动脉腔内血管成形术(percutaneous transluminal coronary angioplasty, PTCA)以解除狭窄。在PTCA过程中可能造成心肌损伤,应用何种药物可减少这种损伤是临床医生关注的问题。此处以PTCA中应用阿托伐他汀减少心肌损伤的研究(Atorvastatin for Reduction of MYocardial Damage during Angioplasty, ARMYDA)为例说明RCTs的设计要点。

研究对象的来源:研究阿托伐他汀能否减少心肌损伤,研究者不可能将所有劳力性心绞痛欲行冠状动脉造影患者均纳入研究范围,故只选取2002年9月1日至2004年3月1日入院欲行冠状动脉造影的患者;纳入和排除标准:纳入标准为:①典型的劳力性心绞痛;②负荷试验阳性;③择期行冠状动脉造影者。排除标准为:①3个月内患有急性心肌梗死者;②不稳定性心绞痛;③在研究开始磷酸肌酸激酶同工酶、肌钙蛋白I、肌红蛋白高于正常上限;④肝酶(谷草转氨酶/谷丙转氨酶)升高;⑤左室射血分数<30%;⑥肾衰竭,血肌酐>3mg/dl;⑦既往有肝脏或肌肉疾病史;⑧目前服用他汀类治疗者;随机分组:通过随机化方法将研究人群分为阿托伐他汀预处理组和安慰剂组,而后者是为了确保本次研究结果可应用到研究总体人群;设立对照:将纳入合格的研究对象随机化为阿托伐他汀预处理组(治疗组)和安慰剂预处理组(对照组),除干预措施不同外,两组基线水平均衡可比,两组间效应结局之差可真实反映在冠状动脉造影中阿托伐他汀可以减少心肌损伤;盲法的选择:研究采用双盲;干预措施:在治疗组每天给予阿托伐他汀40mg预处理,共7天。效应指标的选择:选用磷酸肌酸激酶同工酶、肌钙蛋白I、肌红蛋白来评价阿托伐他汀在PTCA中心肌损伤情况。



二、非随机对照试验

非随机对照试验(non-randomized control trials, NRCTs)与 RCTs 相同,均属于同期对照试验,所不同的是 RCTs 的研究对象在分组时是按照随机化的方法进行的,而 NRCTs 研究对象的分配是由主管研究的医师决定,或者由患者或家属的意愿决定。由此可知,其优点是方便、简单,容易被医师和患者接受,依从性较高,缺点是难以保证各组间结果的均衡可比性和比较的合理性。

在研究马来酸左旋氨氯地平降压治疗中,纳入 100 名 1 级或 2 级高血压患者,按就诊顺序,前 50 例为治疗组,后 50 例为安慰剂组。并未按照随机化分组的原则。共观察 8 周,比较两组血压降低的水平,此处应用的就是 NRCTs。

三、非同期随机对照试验

非同期随机对照试验也称为历史性对照试验(historical control trials, HCTs),即是将当前的患者全部安排到实验组接受新疗法,而以过去某一时期患同种疾病且接受过传统疗法的患者作为对照组,也可以将相关文献资料或本单位的历史资料作为对照组,比较两组患者的治疗效果和价值。其优点是实施方便,但其缺点较多,首先研究者安排患者时可能会有意无意地对患者进行选择,易出现选择性偏倚;其次,在不同的历史条件下,患者情况、医疗条件、医生水平等均有变化,使所得资料无可比性;最后历史性资料可能存在缺乏设计、记录不全、前后条件不一致等问题,影响了试验结果的真实性。故治疗性研究中一般不主张进行 HCTs。

四、单病例随机对照试验

RCTs 是目前治疗性研究的金标准方案,但它反映某个干预措施对研究对象的平均效应,对于某个特定的患者可能无效,因此,对于研究个体差异性较大的慢性疾病且需长期治疗者,或心理精神性疾病的治疗性研究,或同时服用多种有效或无效药物需要进行筛选决策的可以进行单病例随机对照试验(N of 1 RCTs)。N of 1 RCTs 是对单个病例进行双盲、随机、多次交叉的试验,是一种以单个病例自身作为对照,对某种药物与安慰剂或另一种药物的疗效进行评价。研究思路见图 2-3-2。优点是设计和执行简单易行,患者依从性较好,但缺点是基线可比性可能较差,且样本量较小、数据较少,易出现 II 型错误。

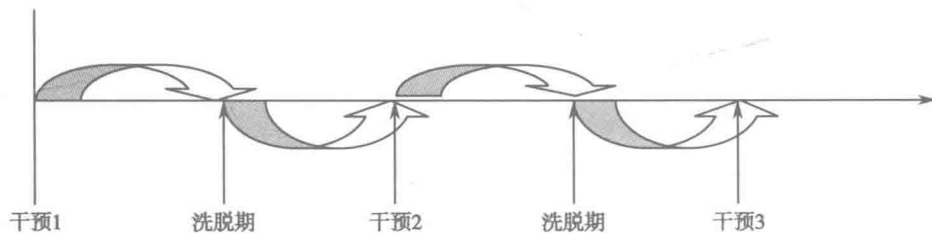


图 2-3-2 单病例随机对照试验设计思路

Estrda CA 等学者对 1 例 65 岁男性中度高血压患者进行 N of 1 RCTs。试验分为 3 轮,每轮分为两期,每期均为 3 周,每轮随机确定治疗期(大蒜素 500mg,口服,每日 2 次)和安慰剂期。考虑到药物的后续作用,每期的第一个星期作为洗脱期,不纳入结果分析。大蒜素及安慰剂被制成有相同颜色、气味和包装的胶囊。患者于每天中午取坐位测量血压,两次测量后取平均值记录,纳入资料分析,最后研究结果显示大蒜素有一定的疗效。



五、交叉对照研究

交叉对照研究(crossover study design)是指将纳入的研究对象随机分配到实验组和对照组,分别给予不同的干预措施,经过一个治疗效应期及洗脱期后,再将实验组和对照组接受的干预措施互换,评价两个阶段的干预效果。这种研究方案采用随机化、盲法、内对照、组间对照的方法,既节约了样本量、减少了伦理学问题、消除了个体差异,又提高了两组的均衡可比性。不足之处是首先交叉对照研究只能适用于一些临床上反复发作、病情稳定的疾病的研究,如冠心病、高血压、哮喘等非根治性治疗,对于那些急重症疾患(如心肌梗死)或不能回复到初始阶段治疗前状况的疾病及那些不允许停止治疗让病情回复到初始阶段的疾病(心力衰竭),均不能采用交叉对照研究方案。其次,交叉对照的设计方案也决定了在进行第二阶段的干预措施时,可能不会完全回复到初始状态。再次,由于实验周期长,可能会存在依从性下降的问题。

地尔硫革治疗肥厚型心肌病的交叉对照试验,研究对象随机分配为两组,一组接受地尔硫革治疗,另一组接受安慰剂治疗,经过一段治疗期后,两组进行交叉互换,最后综合比较地尔硫革和安慰剂对肥厚性心肌病的治疗效果。

六、自身前后对照研究

自身前后对照研究(before after study in the same patients)与交叉对照研究相比,自身前后对照研究没有将研究对象随机分配,而是将干预措施随机分配,先后安排给所有的研究对象,最后将两次观测的结果进行比较的一种设计方案。与交叉对照研究相同,两阶段期间也要经历一个洗脱期。适应范围与应用指征同交叉对照研究,所不足的是研究的科学性不及交叉试验。

一项临床试验采用自身前后对照研究的方法研究复方丹参滴丸治疗稳定性心绞痛的疗效。62例患者分别接受两种方法(复方丹参滴丸和安慰剂)的治疗,每一种方法治疗时间为1个月,经过2个治疗阶段后(2个阶段之间有2周的洗脱期),比较两组治疗方法后西雅图心绞痛量表中文版评定的心绞痛发作情况结果。发现口服复方丹参滴丸可以降低心绞痛的发作。

七、序贯试验

序贯试验(sequential trial)又称序贯分析,与上述6种研究不同的是,序贯试验设计可事先不规定样本量,而是随着试验进展情况而定。其试验设计思路是对每一个或每一对研究对象进行试验后,立即分析试验结果,再决定下一步试验,直至出现规定的试验结果后终止试验,所以称之为序贯试验。优点是减少样本量、避免浪费研究资金,缺点是只适用于仅以单一指标作结论依据的试验。

其他研究方法有病例报告、病例分析,但其样本量少、缺乏对照,偏倚较大,只适用于治疗性研究的初级阶段。

此外,在临床实践中,疾病的预防占有重要的地位。循证医学指出,临床医师应该“认真地、明智地、深思熟虑地运用在临床研究中得到的最新、最有力的科学信息来治疗患者”。预防性研究包括筛查试验(screening test)、健康指导(counseling interventions)、免疫接种(immunization)和药物预防(chemoprophylactic regimens)研究。通过采用科学的设计、衡量和评价的方法开展预防性研究,可使医务工作者们采用有效的预防措施,从而预防疾病的发生。

知识点

治疗性研究的用途:①对于新药物、新疗法的临床疗效、不良反应进行评价;②澄清治疗过程中有争议的问题;③指导临床治疗方案的决策;④通过治疗验证病因。

常用的随机分组方法:①简单随机分组;②区组随机分组;③分层随机分组;④中心随机化分组。

样本量的估计:样本量估计是 RCTs 设计中的重要问题之一,样本量的大小与以下 3 方面因素有关:①干预措施预期作用的大小。干预措施预期作用越大,需要的样本量越小;②本次研究所需要达到的精确度、灵敏度及 I、II 类统计错误的概率。精确度与研究样本量成正比,灵敏度及 I、II 类统计错误的概率与样本量成反比;③研究分组越多,所需样本量愈大。

第二节 评价原则

全世界每年约有 200 万篇有关生物医学的文章发表在 2 万余种生物医学杂志上,然而针对某一专题的医学文献中真正有用的不足 15%。即使“金标准”方案的 RCTs 也存在诸多问题,以 1999 年至 2000 年 35 种中华类医学杂志 164 篇 RCTs 为例,其中只有 15% 真正实现了随机分配,29% 基线描述清楚且具有可比性。因此,临床医师应该对检索到的原始文献进行严格的分析和评价。首先对此项研究的内部真实性进行评价,如果文献具有真实性,则进一步分析与评价该治疗措施是否具有临床治疗价值,即临床的重要性;假设其不仅具有真实性,而且具有重要性,则需要分析这种措施有无适用性(表 2-3-1)。

表 2-3-1 治疗性研究的文献评价原则

1. 评价治疗性文章的真实性

- (1) 研究对象是否真正的进行了随机分配及随机分配隐藏
- (2) 治疗组和对照组除干预措施外,各组接受的其他方法是否一致
- (3) 所有纳入的研究对象是否均完成了随访,研究对象随访时间是否足够长
- (4) 是否根据随机分组的情况对所有患者进行意向分析
- (5) 是否对研究对象、医师和研究人員采用盲法

2. 评价治疗性文章结果的重要性

- (1) 治疗性研究效果的大小
- (2) 治疗作用的精确性估计

3. 评价文献结果的适用性

- (1) 能否将文献的研究结果用于自己的患者
- (2) 我们目前的医疗环境能否采用这种治疗方法
- (3) 评估治疗措施对患者的利与弊

一、真实性的分析与评价

(一) 研究对象是否真正地进行了随机分配及随机分配隐藏

临床治疗性研究结果受多种已知和未知因素的影响,而治疗措施只不过是其中一个因素。在非随机的治疗性研究中,因治疗组和对照组之间的变异性较大,易产生各种偏倚,有时甚至得出错误的结论。而真正进行随机分组的 RCTs 研究方案则可以避免人为的选择性偏倚;同时采用随机分配隐藏的方法,则更能有效地防止试验观测时的偏倚;在研究人数足够大的情况下,除了有效地防止已知的偏倚外,还可使未知的偏倚在随机分组的条件下,达到相互之间的相对平



衡而消除其影响,增强两组的可比性。

对于 RCT 应重点从以下方面进行评价:①随机化分组及随机分组隐藏的方法应该有以下具体的交代:对于随机分组应交代清楚分配序列产生的方法,如简单随机分组、区组随机分组还是分层随机分组,对于随机分组隐藏应交代清楚是按顺序编码、不透光、密封的信封、中心随机系统、编号或编码的瓶子或容器、还是中心药房准备的药物等。只有这样,读者才能理解是否为真正的随机化和随机分配隐藏。②试验组与对照组的研究例数是否相等或相近,彼此间差异有没有统计学意义。一般来说,两组患者人数相差的比例不宜太大,否则会影响把握度的大小。③治疗组和对照组的基线有无明确的描述,两组间是否可比?如果两组基线状态不一致,则应注意在结果分析时是否作了基线的分层比较和校正,如果进行了分层校正与比较,则会增强研究结果的论证强度。④如果采用分层随机分组,则需注意分层因素的数目,若数目过多,可能会使每组研究人数过少,最终各组的样本量不符合实际,影响实验结果的真实性。

(二) 治疗组和对照组除干预措施外,各组接受的其他方法是否一致

临床治疗性研究除了接受规定的干预措施外,患者或医师为了使治疗更有效、更安全,常常会增加一些有关的治疗性药物,此时应该注意避免干扰和沾染两种影响试验结果真实性的情况,以确保证据的真实性。干扰是指治疗组或对照组接受了类似干预措施的其他处理,人为地扩大或缩小了治疗组和对照组间疗效的真实差异,而沾染是指对照组接受了实验组的防治措施,使治疗组和实验组之间的疗效差异减小,如在观察应用小剂量阿司匹林预防短暂性脑缺血发作及脑卒中的研究中,可能因对照组成员,多次应用阿司匹林治疗感冒而出现沾染。

(三) 所有纳入的研究对象是否均完成了随访,研究对象随访时间是否足够长

患者的随访是否完整,对于决定结果的可靠性十分重要。由于不能耐受不良反应、症状已缓解、患者搬离开原地址或拒绝接受某项检查,在试验的某一点上,需要测定患者的结果时,找不到该患者,使结果评定的可靠性大大降低,一般原则是失访率应控制在 10% 以内,若失访超过 20%,研究质量会受到很大的影响,结果可能就会不真实。

同时,应该保证随访期足够长,以确保能够获得重要的临床效应结果。如采用美托洛尔缓释片观察心力衰竭患者的临床获益情况,若仅观察 1 周,就不可能观察到所期待的结果。随访的时间一般取决于所研究疾病的病程长度,通常临床观察疗程至少数月,有的甚至 1 年以上才能充分显示治疗措施的重要效果。

(四) 是否根据随机分组的情况对所有患者进行意向分析

随访过程中不可避免会存在失访问题,如果这部分失访者没有被纳入结果分析,必然会破坏随机化原则和基线的可比性,最终影响结果的真实性。因此,必须依据最初的随机分组情况进行意向性分析(intention to treat, ITT),即对所有纳入随机分配的研究对象只要坚持完成随访,不管是否坚持分配的治疗,是否完成治疗,均按入组时的分组对结果进行分析。意向性分析可保持两组的可比性,较分析时将失访者排除所得的结果更接近实际情况,但其可靠性和真实性与真实值是有差异的,这是在研究中不得已才采用的统计分析方法。例如比较手术治疗和服药治疗冠心病的疗效,随机分配至手术治疗组的患者,有些因为病情太重或在等待手术期间发生了其他情况(如脑卒中或心肌梗死)以至于根本未进行手术,此时研究者若将这些患者(其预后必然不良)的结果归入服药组而不是手术组进行分析,即使是一项无效的手术也会似乎有效。而实际上这种手术“疗效”并不是来自接受手术者的实际得益,而是由于从手术组中系统地排除了预后最差的患者。

(五) 是否对研究对象、医师和研究人员采用盲法

在确定和判断最终结果时,是否采用盲法具有重要的意义。盲法可以有效地排除测量性的偏倚,确保结果判断的真实与可靠。凡治疗组与对照组可施行盲法治疗却没有施行者,其结果的论证强度必然会受影响;在治疗过程中,不能用盲法的,如内科治疗与外科治疗的比较,若采用盲法进行统计分析,可提高可信度。但施行盲法时需注意伦理学问题。



有一项多国、随机、双盲研究,研究替米沙坦与缬沙坦的降压效果是否相当。研究纳入1000例轻中度高血压患者,研究采用随机化分组,随机接受替米沙坦,40~80mg,每日1次;或缬沙坦,80~160mg,每日2次。除干预措施外,各组接受的其他方法一致;所有纳入的研究对象进行8周的随访,失访率为6%。同时对各组研究对象在入组时行意向性分析。确定和判断最终结果时,采用双盲。

二、重要性的分析与评价

(一) 治疗性研究效果的大小

评价治疗性研究结果的大小时常用率表示,如有效率、治愈率、病死率、病残率,但考虑到这些率对量化的临床重要程度略显不足,因此,在比较分析RCTs治疗组与对照组的最终有效率时,我们用危险度—表示一个给定结果的频数分布(危险度是概率,波动于0.0~1.0之间。概率为0.0表示事件不会发生,概率为1.0表示事件必然会发生)来估计临床疗效大小。

1. 相对危险度 相对危险度(relative risk,RR)指治疗组某结局事件的发生概率与对照组该结局事件的发生概率之比。如果结局事件是不良事件、恶性事件(病死、致残等)且 $RR < 1.0$,则说明治疗能使不良事件的发生概率降低。如果 $RR > 1.0$,则说明治疗不但没有降低不良反应的发生,反而增加了不良反应的发生。反之,如果结局事件是良性事件(如治愈),且 $RR > 1.0$,则说明治疗使良性结局发生的概率增加;如果 $RR < 1.0$,则说明该治疗不但无效,反而对该结果有害,使良性结局发生率降低,影响治疗效果。

学
习
笔
记

用 β 阻滞剂预防心肌梗死后死亡的随机对照试验,治疗组和对照组分别用 β 受体阻滞剂和安慰剂进行治疗,1年后病死率分别为0.75%和1.00%,那么此研究中心心肌梗死发生的相对危险度为0.75%/1.00%,即0.75, $RR < 1$,说明经过治疗可以使心肌梗死后死亡事件发生概率降低。

2. 相对危险度减少 相对危险度减少(relative risk reduction,RRR)是指对照组与治疗组有关临床事件发生率之差占对照组事件发生率的百分比,表示经过治疗后,治疗组比对照组有关临床事件发生的相对危险度下降的水平,通常RRR在25%~50%或以上,方有临床意义。

一项大型的多中心研究目的是研究卡托普利是否能降低急性心肌梗死早期病死率和并发症。结果发现经卡托普利治疗组病死率为9.1%,安慰剂组病死率为9.7%,那么 $RRR = (9.7\% - 9.1\%) / 9.7\% = 6.2\%$,表明经卡托普利治疗的危险性的减少是安慰剂组的6.2%。

3. 绝对危险度减少 绝对危险度减少(absolute risk reduction,ARR)是指治疗组和对照组不良结果事件的绝对差别。即对照组不良事件发生率减去治疗组不良事件发生率,此值越大越好。

心肌梗死存活全球研究-2(ISIS-2)评价阿司匹林在心肌梗死中的作用,最终结果阿司匹林组心血管病死亡率为9.4%,对照组心血管病死亡率为11.8%,则 $ARR = 11.8\% - 9.4\% = 2.4\%$,由于RRR仅是相对数,有时ARR下降很小,而RRR看上去却很大,以ISIS-2为例, $RRR = (11.8\% - 9.4\%) / 11.8\% = 20.3\%$,可见虽然相对危险度降低较大,但绝对危险度减少较小,有时相对危险度是同一数字,但绝对危险度差别较大。

4. 需要治疗的患者人数 需要治疗的患者人数(number needed to treat,NNT)是指挽救一



个患者免于发生严重的临床事件,需治疗具有发生这种事件危险性的患者人数,这一指标对于评价某一治疗措施的临床价值包括经济价值有重要意义,该值越小,其治疗方法的临床价值越大。虽然常用 RRR 来评估临床疗效大小,但 RRR 是一个相对数,不能仅根据 RRR 的结果来断定治疗效果,而要结合未治疗患者发生不良结果事件的危险性大小来判断。如果 RRR 不变,那么未治疗患者发生不良结果事件的危险性就越大,则患者从该治疗中的获益越大,为防止 1 次事件所需治疗的患者人数越少。不同的事件发生率也许 RRR 相似,但意义却不相同,而同时应用 ARR 及 NNT ,则评价就更加完善、可靠。

如 A 药治疗甲病,治疗组的残疾发生率为 39%,对照组的残疾发生率为 50%, $RRR=(50\%-39\%)/50\%=22\%$, $ARR=50\%-39\%=11\%$;而采用 B 药治疗甲病,治疗组的残疾发生率为 0.00039%,对照组的残疾发生率为 0.00050%, $RRR=(0.00050\%-0.00039\%)/0.00050\%=22\%$, $ARR=0.00050\%-0.00039\%=0.00011\%$,由此可见,这两种药应用于临床,其实际的意义是不一样的,很明显 A 药用于临床,其实际意义更大些。再如上例 A 药和 B 药治疗甲病的疗效,两药疗效的 RRR 相同,但对于 A 药: $ARR=11\%$, $NNT=9$;对于 B 药: $ARR=0.00011\%$, $NNT=909\ 090$,可见应用 A 药为防止 1 次不良事件所需治疗患者人数是 9 人,而 B 药则需 909 090 人。故其临床意义显而易见。

(二) 治疗性研究效果的精确性

RR 、 RRR 、 ARR 及 NNT 是有关效应大小的指标,仅表示效应强度的平均水平,但效应并不是固定于某一特定值或代表值,也有一定的变化范围,因此,我们需要用 95% 可信区间 (confidence intervals, CI) 表示精确度或范围。95%CI 表示真正的治疗作用 95% 的机会落在这段可信区间中,或者说,如果进行 100 次类似的临床试验,其中有 95 次试验的 95% CI 中会包括总体真正的疗效值。且 95%CI 范围越窄,对于真实的 RRR 估计越准确。对于 RRR 的 95%CI 有以下几种情况: RRR 的 95%CI 下限 >0 说明治疗组的效果明显大于对照组;若 RRR 的 95%CI 上限 <0 说明治疗措施对患者是有害的;若 RRR 的 95%CI 上限 >0 ,其下限 <0 ,说明有可能治疗组疗效优于对照组,也有可能治疗组疗效等于对照组,还有可能治疗组疗效比对照组差。

学
习
笔
记

ASCOT 研究为国际多中心参与的随机对照试验,主要比较氨氯地平 / 培哚普利与阿替洛尔 / 苧氟噻嗪两种抗高血压治疗方案对预防心血管事件的影响。主要终点结局见表 2-3-2。可知心血管死亡率的相对危险度为 0.76,95% CI 为 0.65~0.90,可信区间下限大于 0,说明治疗组氨氯地平 / 培哚普利的疗效明显优于阿替洛尔 / 苧氟噻嗪组。

表 2-3-2 氨氯地平 / 培哚普利组的疗效与阿替洛尔 / 苧氟噻嗪组相比终点事件情况

终点事件	RRR 95%CI
首要终点事件	
非致死性心梗 + 死亡	0.9 (0.79~1.02)
次级终点事件	
非致死性心梗 + 死亡	0.87 (0.76~1.00)
终末期冠心病	0.87 (0.79~0.96)
冠心病事件	0.84 (0.78~0.90)
全因死亡	0.89 (0.81~0.99)
心血管死亡率	0.76 (0.65~0.90)
致死和非致死性卒中	0.77 (0.66~0.89)
致死和非致死性心衰	0.84 (0.66~1.05)



三、适用性的分析与评价

(一) 能否将文献的研究结果用于自己的患者

被评价的文献的真实性和重要性获得肯定后,不能不加分析的搬来使用,而应该注意文献中所研究的患者与我们目前需要治疗的患者大体上是否一致,如诊断标准是否可靠,我们的患者是否在文献的纳入标准之内、排除标准之外,还应该注意我们目前的患者的生理功能、病情阶段、年龄、性别、社会经济情况是否存在显著差异等。

在一些情况下,我们发现患者的情况与总体相比缺乏适用性,然而,在某些亚组中可能存在实际意义即患者病情与亚组中的患者病情相似,那么,这个亚组的治疗措施就有实用价值,但应注意,如果亚组分类过多,这种有效性可能是由于机会引起的,不一定真的具有实用价值,因此在区分亚组患者的结果能否适用时需考虑以下几点:①确有生物学上和临床上的意义;②确有统计学意义;③确保研究前已经分出这个亚组,且有其他研究证实;④确保研究中划分出的是有限的亚组。

(二) 我们目前的医疗环境能否采用这种治疗方法

有的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实施,而且此项治疗措施还可能受医生的技术水平、医院的管理机制及设备条件等限制。如开篇提到的问题,我们已从所研究的文献中获知,李某应行 PTCA,此时需要考虑该患者所就医的医院是否具有开展介入治疗的条件。这些条件包括:①能在患者住院 90 分钟内施行经皮冠状动脉介入术(percutaneous coronary intervention, PCI);②心导管室每年施行 PCI>100 例并有心外科待命的条件;③施术者每年施行 PCI>30 例;④AMI 直接 PTCA 成功率在 90% 以上;⑤在所有送到心导管室的患者中,能完成 PCI 者达 85% 以上。若该医院不具备这些条件,就不宜施行急诊 PTCA。此外,我们还应该考虑患者及医疗保险系统能否支付该项治疗措施的费用,如不可能时,就不能将这项治疗措施应用于自己的患者。

(三) 评估治疗措施对患者的利与弊

在对患者进行一项新的治疗前必须考虑此项治疗措施对患者的利与弊,然后综合评价,一定要在利大于弊的基础上实施治疗措施。治疗措施利弊量化的指标最直接的是使用 NNT 和发生 1 例不良反应需要的病例数(number needed to harm, NNH)。如一种治疗费用低廉,使用方便、依从性高且安全,则我们会毫不犹豫地治疗,即使是用几百例去挽救 1 例患者的生命,但事实上这种安全、廉价的治疗措施可能并不存在,反而会有一定的不良作用出现,我们报告这些不良作用时,可以用 NNH 来估计治疗的利弊。

如心律失常抑制试验,与安慰剂组相比氟卡尼虽能减少心肌梗死后室性心律失常的发生,但却明显的增加了病死率,其弊大于利,因此不宜用于患者的治疗。再如对低危患者的 β 受体阻滞剂预防心肌梗死后死亡的 RCTs 研究中 $NNT=400$, $NNH=10$, 对高危患者的 β 受体阻滞剂预防心肌梗死后死亡的 RCTs 研究中 $NNT=40$, $NNH=10$, 即说明对于低危组每减少 1 例死亡可造成 40 例患者产生疲劳感,而对于高危组每减少 1 例死亡只造成 4 例患者产生疲劳感。故对于高危组其不良作用较小,可以适用于高危患者。

此外,在治疗前,作为主管医师,应充分地告知患者此项治疗的效果、不良后果、不良反应及费用,尊重患者的决定,若患者不愿意接受此种治疗方案,不可强行实施。

**知识点**

治疗性研究的文献评价原则见表 2-3-1。

治疗性研究效果大小的评价指标:①相对危险度(relative risk, RR);②相对危险度减少(relative risk reduction, RRR);③绝对危险度减少(absolute risk reduction, ARR);④需要治疗的患者人数(number needed to treat, NNT)。

治疗性研究效果精确性的评价指标:可信区间(confidence intervals, CI)表示精确度或范围,通常用 95% CI 表示。95% CI 范围越窄,估计越准确。

应用 NNT 的注意点:① NNT 是表示疾病与干预结果的关系,只有在条件和结果相同时,才可直接比较不同干预措施的 NNT ;②是一种点估计,临床研究中得到的 NNT 值可能与真实的 NNT 有差别。可通过计算 NNT 95% CI 来解决这一问题;③对于具体的某个患者, NNT 没有固定值;④当不同处理之间的比较建立在相似的基础上时,这些比较才有意义。

思考题

1. 下列哪项属于 RCT 的设计
 - A. 专家的会诊意见
 - B. 队列研究
 - C. 随机对照试验
 - D. 病例对照试验
 - E. 注册研究
2. 下列哪项属于随机对照试验控制偏倚的措施
 - A. 配比
 - B. 随机分组
 - C. 多元回归分析
 - D. 意向性分析
 - E. 明确入选标准
3. 下列哪项不属于随机对照试验的基本原则
 - A. 随机分组
 - B. 对照原则
 - C. 盲法
 - D. 干预
 - E. 意向性分析
4. 下列哪项情况不属于研究结果意向性分析的范畴
 - A. 能耐受不良反应
 - B. 患者搬离开原地址
 - C. 拒绝接受某项检查
 - D. 患者患有另一种疾病
 - E. 患者转组

5. SAVE 试验研究卡托普利对 AMI 后无症状的左室功能不全患者的作用,该研究观察了 2231 名 AMI 后无症状的左室功能不全的患者,将其随机化分为卡托普利组和安慰剂组,于 AMI 后 3~16 天开始给予相应治疗,平均随访 42 个月,结果显示卡托普利组总死亡率为 20.4%,安慰剂组总死亡率为 24.6%,其 ARR 和 NNT 分别为



- A. 4.2%,24
- B. 42%,2
- C. 42%,24
- D. 4.2%,2
- E. 24%,42

参考答案

1. C 2. B 3. D 4. D 5. A

(王聪霞)

参考文献

1. Dan Mayer. Essential Evidence-based Medicine. UK: Cambridge University Press, 2004.
2. Pasceri V, Patti G, Nusca A, et al. Randomized trial of atorvastatin for reduction of myocardial damage during coronary intervention results from the ARMYDA (Atorvastatin for Reduction of Myocardial Damage during Angioplasty) study. Circulation, 2004, 110(6): 674-678.

第四章 疾病预后

临床问题

预后问题是临床医师每天要面临的问题,当患者确诊某疾病后,非常关注的问题是如何治疗,治疗效果如何,是否容易复发,平均可以存活多少年,最长能够活多少年,注意哪些方面可以延长寿命,这些都是有关预后的问题。

第一节 基本概念

预后研究(prognostic study)也称为结局分析(outcome analysis)。预后(prognosis)是指疾病发生后,对将来发展为各种不同后果(痊愈、复发、恶化、伤残、并发症和死亡等)的预测或事前估计,通常以概率表示,如治愈率、复发率、5年生存率等。预后研究就是关于疾病各种结局发生的概率及其影响因素的研究。

凡影响疾病预后的因素都可称预后因素(prognostic factors),若患者具有这些影响因素,其病程发展过程中出现某种结局的概率就可能发生改变。预后因素的研究有助于临床医师进行医学干预,包括筛检、及时诊断、积极治疗和改变患者影响健康的不良行为等,从而为改善患者疾病预后而作出努力。患者的预后往往是多个因素决定的,所以预后因素的分析需要应用多因素统计方法,Cox 回归模型是最常用的统计方法。

对于急性髓系白血病(AML)患者,染色体是重要的预后因素,也对治疗方案的选择起到了指导性作用,例如对存在不良染色体的患者建议缓解后早期进行异基因造血干细胞移植,而对存在良好染色体核型的患者不建议早期异基因移植。年龄也是重要的预后因素,>60岁的老年患者预后较差,生存期短于年轻人。患者的预后往往有多个因素,例如对于AML患者而言,年龄、白细胞计数、染色体核型、血小板计数、治疗方案等都是独立的预后因素。

一、用于预后研究各种率的定义

(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。常用于病程短且容易死亡的疾病,如各种传染病、急性中毒、心脑血管疾病的急性期和迅速致死的癌症。

$$\text{病死率}(\%) = \frac{\text{死于该病的患者人数}}{\text{患该病的患者总人数}} \times 100\%$$

(二) 疾病死亡率

疾病死亡率(disease-specific mortality)是指一定的时期内(通常指一年),某一人群中因为某病死亡的人数所占的比例,一般以1/10万或1/万为单位。



$$\text{死亡率} = \frac{\text{一定时期内死于某病的人数}}{\text{同期平均人口数}} \times 10 \text{ 万} / 10 \text{ 万 (或 } 1 \text{ 万} / 1 \text{ 万)}$$

病死率和死亡率是两个不同的概念,例如 2012 年某区 40 例 AML 患者死亡,该区共有人口 10 万人,白血病患者 200 人,白血病死亡率是 40/10 万,病死率是 20%(40/200)。

(三) 治愈率

治愈率(cure rate)系指患病治愈的患者人数占该病接受治疗患者总数的比例。

$$\text{治愈率}(\%) = \frac{\text{患某病治愈的患者人数}}{\text{患该病接受治疗的总患者人数}} \times 100\%$$

(四) 缓解率

缓解率(remission rate)系指进行某种治疗后,进入疾病临床消失期的病例数占总治疗例数的百分比。有完全缓解率、部分缓解率和自发缓解率之分。

$$\text{缓解率}(\%) = \frac{\text{治疗后进入疾病临床消失期的病例数}}{\text{接受该种治疗的总病例数}} \times 100\%$$

(五) 复发率

复发率(recurrence rate)系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。例如,100 例急性白血病患者诱导治疗后,70 例达到缓解,1 年后有 30 例患者复发,复发率为 42.8%。

$$\text{复发率}(\%) = \frac{\text{复发的患者例数}}{\text{接受观察的患者总数}} \times 100\%$$

(六) 总体生存率

总体生存率(overall survival rate, OS)系指从疾病临床过程的某一点开始(一般为确诊时间),一段时间后存活的病例数占总观察例数的百分比。生存率常用于长病程致死性疾病,如各种癌症,病程较短的癌症可用 1 年生存率,一般癌症用 5 年生存率表示预后。例如 AML 的 5 年生存率为 20%,表明从诊断 AML 开始有 20% 的患者可以生存 5 年以上。

$$n \text{ 年生存率}(\%) = \frac{\text{活满 } n \text{ 年的病例数}}{n \text{ 年内观察的总例数}} \times 100\%$$

预后研究仅仅报道生存率是不够的,例如图 2-4-1 中央层主动脉瘤、肺癌、HIV 感染、100 岁以上老人 4 种情况 5 年生存率均为 10%,但不能反映 5 年间生存率的变化情况。夹层主动脉瘤早期死亡率很高,但如果可以度过早期的危险期,以后死亡风险趋于稳定。

(七) 无病生存率

无病生存率(disease-free survival rate, DFS)常用于癌症的结局判断,指疾病经过治疗达到临床缓解后,没有临床疾病复发或死亡的患者占有所有临床患者的比例。

100 例 AML 患者经过某一化疗方案治疗,有 70 例达到完全缓解,30 例未达到完全缓解,70 例完全缓解的患者 3 年内有 20 例复发,各种原因死亡 5 例,则应用该化疗方案治疗 AML 的 3 年无病生存率为 45%(45/100)。



(八) 无进展生存率

无进展生存率 (progression-free survival rate, PFS) 常用于癌症的结局判断, 指疾病诊断或进入临床试验随机化分组后, 没有进展或死亡的患者占有所有临床患者的比例。例如 100 例肺癌患者随机分配生物治疗组, 3 年内 75 例病情无进展, 20 例病情有进展, 各种原因死亡 5 例, 则 3 年无进展生存率为 75% (75/100)。

二、中位时间

(一) 中位生存时间

中位生存时间 (median survival time) 又称为半数生存期, 即当累积生存率为 0.5 时所对应的生存时间, 表示只有 50% 的患者可以活过这个时间。例如 AML 的中位生存期为 20 个月, 说明诊断为 AML 后只有 50% 的患者可以活过 20 个月。

从生存曲线可以知道中位生存时间, 图 2-4-1 我们可以看到夹层主动脉瘤的中位生存时间不足半年, 肺癌不足 2 年, HIV 感染约为 2.5 年, 大于 100 岁的老年人 1.5 年。

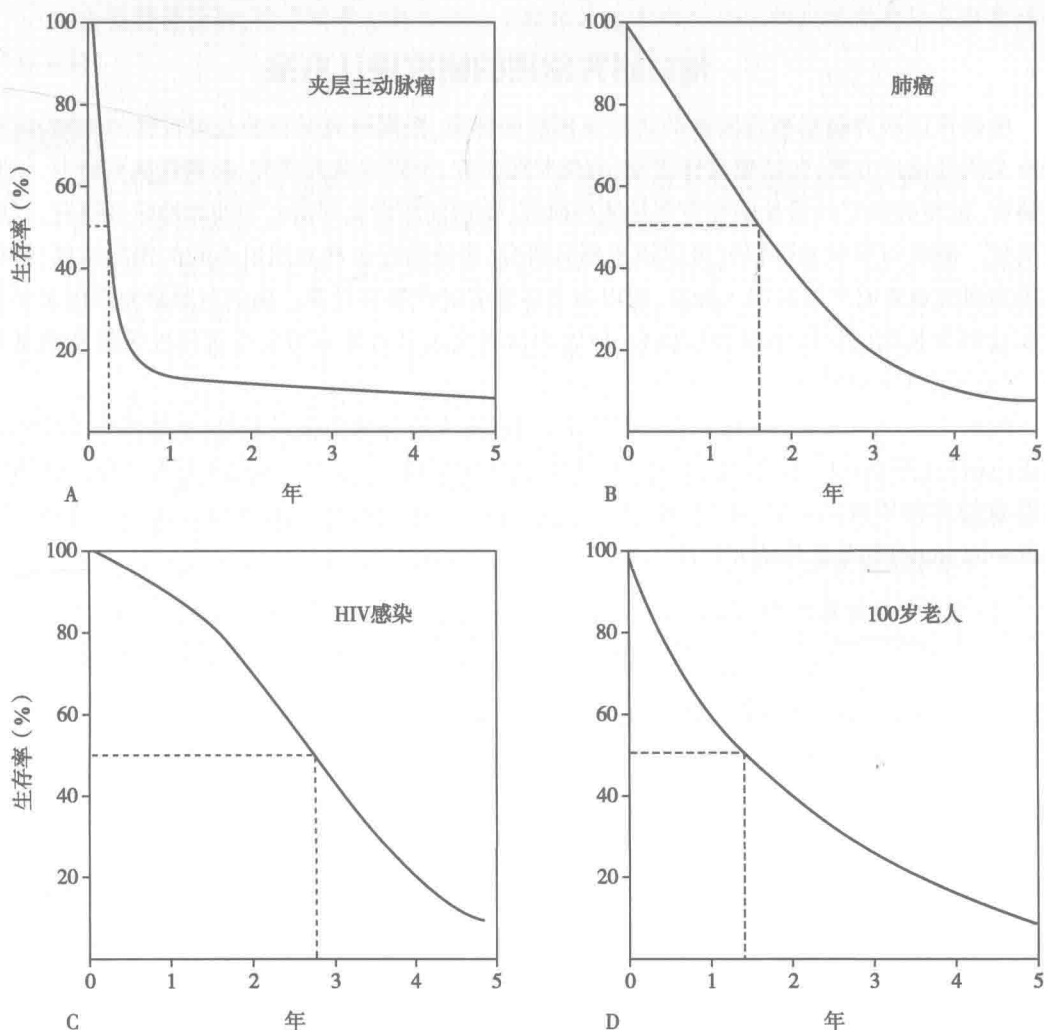


图 2-4-1 4 种不同人群的生存曲线

5 年生存率均为 10%, 但生存曲线明显不同, 说明仅报道 5 年生存率有很大的局限性, 虚线为中位生存时间



(二) 中位无病生存时间

中位无病生存时间(median disease-free survival time)的基本概念同中位生存时间,如果 AML 的中位无病生存时间为 18 个月,表明有 50% 的患者可以在无临床疾病的状态下存活 18 个月。无病生存时间一般从疾病缓解之日、手术切除之日开始算起,到疾病复发或死亡为止。

(三) 中位无进展生存时间

中位无进展生存时间(median progression-free survival time)指从规定的随访起始点开始,直至疾病进展或死亡的中位时间。在药物疗效评价中,常用随机化之日为起始点。

比较分子靶向药物吉非替尼和常规化疗作为一线治疗肺癌的长期疗效,新诊断的肺癌患者随机分两组,一组为吉非替尼治疗组,一组为常规化疗组,随访 5 年,观察 PFS 和 OS 等预后指标。随机化分组之日为观察起始点,OS 是指随机化之日起到任何原因死亡的时间,PFS 是指随机化之日起到疾病进展的时间。比如有 1 名患者,2013 年 1 月 1 日诊断为肺癌,随机分组入化疗组,化疗后疾病稳定,直至 2013 年 10 月 31 日出现疾病进展,2013 年 12 月 31 日死亡,该患者的 OS 为 12 个月,PFS 为 10 个月。

三、预后研究常用的研究设计方案

疾病预后研究包括预后因素的研究及预后的评定,根据研究的目的及可行性的原则,可选择有关研究设计方案,包括描述性研究、病例对照研究、回顾性队列研究、前瞻性队列研究、试验性研究,但预后研究的最佳研究方案是队列研究,包括回顾性队列研究和前瞻性队列研究,以后者为佳。随机对照试验性研究可以用于预后研究,明确治疗组 and 对照组不同的预后结局,但往往因为研究对象有严格的纳入标准,所以对全体患者的代表性较差。病例对照研究常用于罕见疾病或需要长期随访疾病的预后研究,但因为该研究设计方案容易发生选择性偏倚和测量偏倚,结论的证据不够强。

队列研究(cohort study)由一组人群组成,可以将人群分为两组或多组,从暴露到结局的方向进行研究(图 2-4-2),是明确疾病发病率和自然史的最佳方法,也可以用于研究由单一暴露因素导致的多种研究结局,发病率研究(incidence study)、纵向研究(longitudinal study)、随访研究(follow-up study)都是队列研究的同义词。

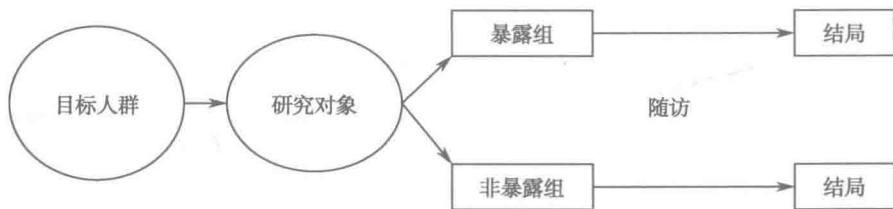


图 2-4-2 队列研究设计图

中美联合上海市白血病协作组完成的 623 例 AML 的预后研究为例,说明前瞻性队列预后研究的基本步骤(本章末参考文献 2)。

1. 确立研究的疾病、研究目的 研究原发性、初治 AML 的预后。
2. 确定观察指标、研究因素 结局观察指标为完全缓解率、3 年复发率、3 年生存率、中位生存时间、分析影响生存的预后因素。
3. 确定研究方案 本研究采用前瞻性队列研究的设计方案进行预后研究。确定 AML



的诊断标准,定义 OS 为从诊断之日到死亡或最后的随访之日,确定随访终止时间为 2008 年 5 月 1 日。

4. 收集队列,登记资料 前瞻性收集 2003~2007 年的上海市白血病协作组 24 家医院的原发性、未治疗的初发 AML 患者 623 例,所有骨髓标本统一送到中美实验室进行统一诊断,以防误诊。登记病例信息和所有相关的实验室资料,包括身份证号码、联系电话、地址、主管医师的联系信息等。

5. 随访 每 6 个月随访一次,同时随访患者本人(或家属)和主管医师,随访内容包括症状、体征、治疗方案、并发症、血象和骨髓检查结果。多种随访途径对减少失访有很大的作用。

6. 统计分析

(1) 病例一般资料分析:患者年龄、性别、WHO 亚型、染色体等资料的分析。

(2) 完全缓解率、复发率的计算。

(3) 生存率、中位生存时间、生存曲线:应用 Kaplan-Meier 方法计算生存率,绘制生存曲线。

(4) 单因素分析:应用 Log-rank 方法统计与生存相关的预后因素,得出与预后相关的因素如 WHO 亚型、染色体分组等。

(5) 多因素分析:把单因素分析中 $P < 0.1$ 的预后因素纳入 Cox 回归模型进行多因素预后分析。

知识点

预后的概念:预后是指疾病发生后,对将来发展为各种不同后果的预测或事前估计。

预后因素的概念:预后因素是指影响疾病预后的因素。

死亡率与病死率的区别:死亡率与病死率是两个不同的概念,计算方法也不同。病死率是指在患某种疾病患者总人数中,死于该病的患者所占的比例,一般以百分号(%)为单位。死亡率中的病死率是指一定的时期内(通常指 1 年),某一人群中因为某病死亡的人数所占的比例,一般以 1/10 万或 1/ 万为单位。

OS、DFS 和 PFS 的含义:总体生存率(overall survival rate, OS)、无病生存率(disease-free survival rate, DFS)和无进展生存率(progression-free survival rate, PFS)是临床常用的描述患者生存情况的指标,特别常用于肿瘤等慢性疾病。①OS 是指确诊或随机化之日起到任何原因死亡的时间;②PFS 是指规定的随访起点(如随机化之日)起到疾病进展的时间。

预后研究的最佳设计方案与步骤:①前瞻性队列研究是预后研究的最佳设计方案;②队列研究的基本步骤是收集队列、随访、确定结局、统计分析。

第二节 评价原则

检索到预后的文献后,根据病人特征、研究问题、研究设计方案等选择合适的文献,例如有多篇 AML 预后的文献,选择较新的、大样本、多中心、前瞻性研究的文献,对其质量及其研究结论是否真实可靠,应进行评价。文献评价包括 3 个方面:真实性、重要性和实用性。评价的原则和标准可归纳为 8 条(表 2-4-1)。

(一) 观察预后的研究对象是否都处于同一起始队列、研究的对象是否能代表被研究疾病的目标人群

1. 尽量采用起始队列 预后研究要求各队列的研究对象观察疾病预后的起始点一定要统一,可以是症状首发时间,疾病确诊时间或治疗开始时间,务必明确,不应存在杂乱的零点时间,



表 2-4-1 预后研究文献的评价原则

1. 真实性	2. 重要性
(1) 队列的起始点是否相同? 队列是否有代表性?	(1) 报告预后研究的结果是否完整?
(2) 随访是否足够长, 是否完整?	(2) 研究结果的精确性如何? 即可信区间是否较窄?
(3) 判断结局时是否有客观的结局标准, 是否采用盲法?	3. 实用性
(4) 是否对影响预后研究的重要因素进行了统计学的校正?	(1) 我们自己的患者是否与文献报道的患者非常不同?
	(2) 研究结果是否有助于治疗方案的制定和是否有助于对患者及其亲属作出解释?

如研究脑卒中的预后因素, 纳入的研究对象应是首次发作的脑卒中患者, 排除第 2 或第 3 次发作者。对入选的研究对象处于病程的哪一个阶段必须有清楚地叙述。所选择的零点时间最好是处于病程的早期, 即起始队列 (inception cohort)。例如以心脏重症监护室内的心肌梗死患者为研究队列来研究心肌梗死的生存率, 就会错误地提高了生存率, 因为一些在急诊室就死亡的患者没有被纳入研究。

有一队列研究, 共有 AML 患者 500 例, 如果以诊断白血病之日为研究起始点, 3 年生存率为 35%, 中位生存时间为 18 个月, 如果以化疗第一天作为研究起点, 3 年生存率为 45%, 中位生存时间为 36 个月。因为有少数患者诊断后还没有来得及治疗就因出血或感染而死亡了, 生存时间只有几天, 把这些患者统计在内时, 生存率就会偏低, 中位生存时间也将相应缩短。

2. 研究对象应该具有代表性 只有当研究对象有代表性时, 该文献获得的结论才有普遍适用性, 即外推性, 结论可以应用于相似的人群。例如来自社区医院的患者和来自三级医院的患者在疾病的严重程度上会存在显著的差异, 三级医院的患者比较严重, 社区医院的患者病程较早或较轻, 得出的预后分析结论必然不同, 结论的代表性和外推性也不同。

(二) 随访时间是否足够、随访是否完整

1. 随访时间和截尾数据 由于预后因素常常存在于不良结局发生之前一段较长的时间, 因此随访时间必须足够长, 以便发现关注的研究结果。如果随访时间很短, 只有一小部分患者达到了我们感兴趣的结果, 如肿瘤发生、康复、复发或是不良事件的发生, 这样就不能反映该疾病预后的真实情况。并且随访必须完整, 在理想情况下, 应当将所有纳入研究的对象从疾病早期一直随访到完全康复、复发或死亡, 但事实上难以做到, 因此存在一定的失访率。到观察截止点为止, 没有发生想要观察的结局事件, 称为截尾数据 (censored data), 包括失访、存活、死于其他疾病 (图 2-4-3)。

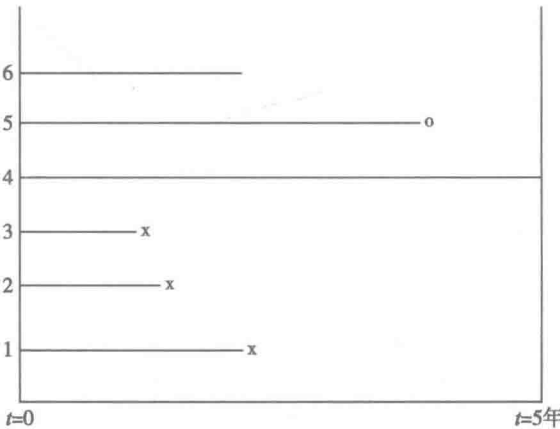


图 2-4-3 截尾数据

注: 6 例患者, 3 例死亡 (x 表示), 第 4 例患者在随访 5 年结束时仍存活, 第 5 例患者失访 (o 表示), 第 6 例患者死于其他疾病, 所以第 4、5、6 例患者均为截尾数据



随访 AML 100 例患者,随访时间到 2013 年 7 月 1 日,有 74 例患者死亡,5 名患者失访,1 名患者死于心肌梗死,20 名患者存活,属于截尾的患者有 26 名。

2. 判断失访对结论的影响 一般遵从“5 和 20”原则,失访率 <5%,其研究结果偏倚少,结果可靠,如失访率 >20% 则严重影响结果真实性,5 和 20 之间结果比较可靠。亦可通过敏感性分析来估计对结论的影响,比较“最高”和“最低”发生率,如两者相差不大,则结果可信,如两者相差很大,则研究结果不可信。

100 例患者,5 例失访,未失访的患者中 40 例死亡,则粗死亡率为 42.1%(40/95),如果把 5 例失访均算入死亡,则死亡率为 45%(45/100),如果把 5 例失访均算为存活,则死亡率为 40%(40/100),最低死亡率为 40%,最高为 45%,相差不大,说明失访对结论的影响较小。

(三) 判断结局有无客观标准,是否采用了盲法

观察疾病预后的终点,即结局应有客观的标准。在研究开始前,研究者必须对结局提供明确的定义,要有客观的测量标准。有些结局容易确定如死亡,但大多数结局,如痊愈、残疾、复发、生存质量改变等,都需要有客观的标准,以避免临床医生在判断预后结局时产生分歧,从而影响预后研究的结论。判断预后结局属“硬”指标,如“死亡”、“残疾”等可以不用盲法判断,如结局属“软”指标容易产生偏倚,应该采用盲法判断,例如疼痛程度的判断。

(四) 是否对影响预后研究的重要因素进行了统计学的校正

预后研究中可能存在各种混杂因素,从而影响预后研究的结论。因此在下结论时应应对这些因素应用统计学方法进行校正。Framingham 的研究者报道风湿性心脏病心房颤动患者的脑卒中发生率为 41/1000 人年,与非风湿性心脏病心房颤动患者的脑卒中发生率十分接近。但风湿性心脏病患者比非风湿性心脏病患者更年轻。对患者的年龄、性别和高血压状态进行校正后,风湿性心脏病心房颤动患者脑卒中的发生率是非风湿性心脏病心房颤动患者的 6 倍。校正的方法最简单的是分层分析,较为复杂的校正方法是多因素分析法如 Logistic 回归及 Cox 模型分析,适用于有多个混杂因素的校正。

有研究骨髓增生异常综合征的一个亚型难治性贫血伴多系发育异常的预后因素,表 2-4-2 为该研究的 Cox 回归分析结果,从表可知年龄、中性粒细胞绝对值、血红蛋白数量为独立的预后因素,而血小板计数、国际预后积分系统(IPSS)不是独立的预后因素。在校正了其他预后因素后,年龄 ≥60 岁的患者死亡的风险是 <60 岁患者的 1.77 倍($HR=1.77$)。风险比(hazard ratio, HR)的含义与相对危险度(RR)的含义相同。

表 2-4-2 难治性贫血伴多系发育异常 Cox 模型多因素预后分析

因素	回归系数 β	HR (95%CI)	Z	P
年龄 (≥60 岁)	0.57	1.77 (1.17~2.68)	2.69	0.007
中性粒细胞绝对值 (<1.0 × 10 ⁹ /L)	0.70	2.01 (1.31~3.10)	3.18	0.001
血红蛋白 (<90g/L)	1.38	3.97 (1.98~7.93)	3.90	0.000
血小板 <30 × 10 ⁹ /L	0.31	1.36 (0.90~2.07)	1.46	0.144
IPSS 积分分组	0.09	1.10 (0.56~2.14)	0.28	0.783



(五) 报告预后研究的结果是否完整

预后研究的定量结果是在一段时间内发生结局的事件数。比如报告生存情况有三种方法：①某一时间点的生存率，如1年生存率、5年生存率等；②中位生存时间(median survival time)，即观察到50%的研究对象死亡的随访时间；③生存曲线(survival curves)能够了解预后，可随时间而变化。例如所报告的1年生存率均为20%，但两条生存曲线形态不同，一条显示中位生存时间为3个月，提示疾病早期预后就很差，另一条显示中位生存时间为9个月，提示疾病早期预后好，随着时间推移而逐渐恶化。因此生存曲线可以了解预后的全貌。完整地报告预后研究结果应当同时报告某一时点的生存率、中位生存时间以及生存曲线。

Kaplan-Meier生存曲线是常用的生存分析方法，又称为乘积极限法，横坐标为生存时间，纵坐标为累积生存率，从图中可以大概估计中位生存时间(天、月、年)，1年、2年、3年等的生存率(%),也可以比较不同分组的生存曲线是否不同。

骨髓增生异常综合征患者根据预后积分系统可以分为低危组、中危组、高危组，从图2-4-4可以看到不同的预后分组——低危组、中危组和高危组之间的生存时间和生存率有显著差异，高危组中位生存时间约21个月，中危组约32个月，低危组约49个月。每一个向下的折线代表有1例患者死亡。曲线的左侧数据结论较精确，右侧因为不断有死亡患者，研究人数逐渐减少，结论精确性降低。生存曲线不能任意延长来估计将来的生存趋势。

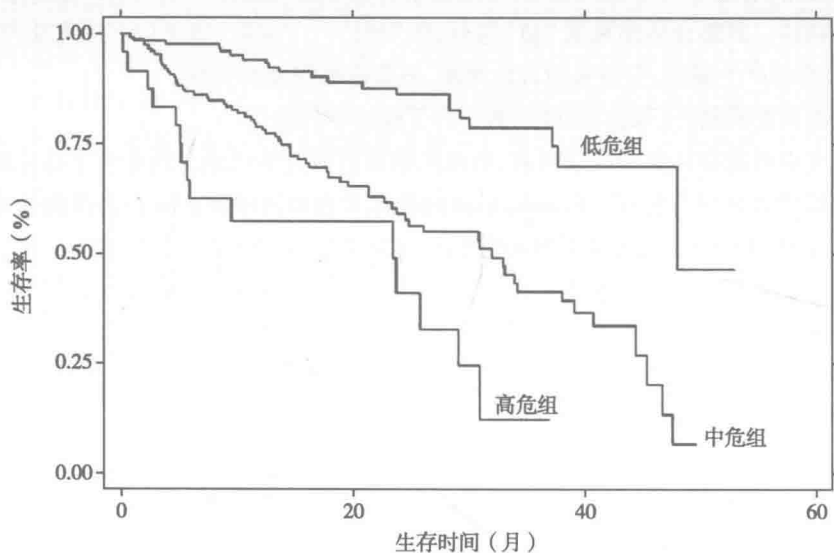


图 2-4-4 Kaplan-Meier 生存曲线

(六) 研究结果的精确性如何

除了报道生存率、生存时间、生存曲线，还应当报告预后估计的精确度，即预后结局概率的95%可信区间(95%CI)。对预后因素的研究可用相对危险度和绝对危险度等来表示，同时也要报告95%CI。95%CI较窄，说明样本量足够大，结果精确性高，结论外推性好，对总体预后的估计更精确。如果一项研究中某预后因素的HR为2.1，95%CI为0.7~2.9，说明该因素无统计学差异，也不能说明是独立的预后因素，未必有临床价值。

(七) 我们自己的患者是否与文献报道的患者非常不同

我们自己的患者是否可以采纳文献报道的结果，还要关注我们的患者与文献报道的研究对象是否在年龄、性别、疾病特征等方面相似或存在极大的不同？如果没有极大的差异，就可以应用研究结论。



(八) 研究结果是否有助于治疗方案的制定和是否有助于对患者及其亲属作出解释

研究结果是否直接有助于治疗方案的取舍? 研究结果是否有助于对患者及其亲属作出解释? 例如一项可信、精确度高的研究结果显示疾病具有良好的预后, 则十分有助于向焦虑的患者及其家属做出解释而使其放心。另外, 一项质量高的研究结果显示疾病预后不良, 就可以与患者和其家属进行有关不良预后结局的讨论。这同样也有实用价值。

知识点

预后性研究的文献评价原则见表 2-4-1。

预后性研究设计注意点: ①研究对象应该有统一的随访起点; ②研究对象应该具有代表性, 注意避免样本的偏倚; ③判断患者的结局应该采用客观标准, 尽量用盲法判断; ④多因素分析是常用的校正混杂因素的统计方法; ⑤报告预后结局的指标主要包括生存率、中位生存时间、生存曲线; Kaplan-Meier 生存曲线是常用的生存分析方法。

截尾: 截尾数据包括失访、研究结束时尚存活、死于其他疾病三种情况。

失访对结果的影响: 判断失访对结论的影响一般遵从“5 和 20”原则。失访率 $<5\%$, 其研究结果偏倚少, 结果可靠, 如失访率 $>20\%$ 则严重影响结果真实性, 5 和 20 之间结果比较可靠。

风险比 (hazard ratio, HR) 的含义与相对危险度 (relative risk, RR) 的异同点: HR 与 RR 都是队列研究中分析暴露因素与发病的关联程度的指标。RR = 暴露组的发病或死亡率 / 非暴露组的发病或死亡率; HR = 暴露组的风险函数 $h_1(t)$ / 非暴露组的风险函数 $h_2(t)$, t 指在相同的时间点上。一般可以认为 HR 与 RR 的含义一样, 但 HR 有时间因素在内, 换句话说, 包含了时间效应的 RR 就是 HR; 在预后性研究资料中, RR 考虑了终点事件的差异, 而 HR 不仅考虑了终点事件的有无, 还考虑了到达终点所用的时间及截尾数据。

思考题

1. 进行 100 例心力衰竭患者的生存分析, 在 3 年中有 60 例患者属于截尾, 下列哪项不属于截尾的原因

- A. 在 3 年内死于其他疾病
- B. 患者退出研究
- C. 患者罹患了其他更致命的疾病
- D. 患者随访时间不足 3 年
- E. 患者死于心力衰竭

2. 有关生存曲线的描述, 下列哪项是正确的

- A. 如果有患者失访时, 是对生存事件的无偏估计
- B. 从零点开始对生存率的估计
- C. 失访结束后仍然存活的研究对象的比例
- D. 离开队列的人数的率
- E. 估计研究队列将来的累积生存率

3. 为了研究多发性硬化症的临床过程, 研究者随访了已经完成的随机对照临床试验的患者, 患者来自三级医院, 在确诊时入组, 符合统一的诊断标准。10 年期间, 每年随访患者。10 年后 40% 患者仍然能够行走。下列哪项影响研究的可信性

- A. 零点不一致
- B. 迁移偏倚



- C. 结论的普遍性
D. 测量偏倚
E. 没有采用生存分析
4. 研究吸烟与周围动脉疾病的关系,继续吸烟患者与戒烟患者相比,HR 值为 3,下列哪项说明了 HR 的意义
- A. 继续吸烟患者比例除以戒烟患者的比例
B. 可以从病例对照研究中获得
C. 不能判断与预后的关系
D. HR 的含义与相对危险度一样
E. 根据继续吸烟患者和戒烟患者的发生率进行计算
5. 有两个研究均是关于急性髓系白血病的预后研究,一个研究从诊断白血病开始随访,计算中位生存时间为 18 个月,另一个研究从首次化疗之日开始随访,得出的中位生存时间为 30 个月,两个研究的结论明显不同的最可能的原因是什么
- A. 不同的研究对象
B. 不同的研究医院
C. 诊断标准不同
D. 不同的起始队列
E. 不同的种族
6. 有一项研究随访 200 例研究对象,其中失访 20 例,5 例死亡,请问该失访率对研究结论有无重要影响,为什么
- A. 没有影响,因为研究对象 200 例,样本量比较大
B. 有影响,因为失访率 10%,>5%
C. 没有影响,因为失访率 10%,<20%
D. 没有影响,因为最高死亡率和最低死亡率相差较小
E. 有影响,因为最高死亡率和最低死亡率相差较大

参考答案

1. E 2. B 3. C 4. D 5. D 6. E

(王小钦)

参考文献

1. Straus SE, Glasziou P, Richardson WS, et al. Evidence-based medicine. 4th ed. Elsevier:Churchill Livingstone, 2011.
2. 中美联合上海市白血病协作组. 上海市 623 例成人急性髓系白血病非选择性病例的 WHO 亚型分布、初治疗效及预后. 中华血液学杂志, 2010, 31 (2): 102-107.
3. Wang XQ, Ryder J, Gross SA, et al. Prospective analysis of clinical and cytogenetic features of 435 cases of MDS diagnosed using the WHO (2001) classification: a prognostic scoring system for predicting survival in RCMD. Int J Hematol, 2009, 90: 361-369.

第五章 生命质量评价

临床问题

生命质量评价是临床实践中一种重要的医学评价技术,日益受到重视和运用。目前生命质量测定广泛用于心脑血管疾病、糖尿病、风湿病、老年病、肿瘤性疾病、外科术后患者等慢性疾病诊断、处理后的疗效评价;新药临床疗效评价;人群和患者的健康状况评价;预防保健措施的效果评价;资源分配、计划和决策的制定等方面。例如,评价类风湿关节炎疾病活动度和严重程度时,仅检查关节生理功能如关节的活动度、压痛、肿胀等,存在的缺陷在于这些临床资料均是针对单个关节逐一进行,不能有效反映需多个关节同时参与的复杂运动能力,更不能反映疾病对患者心理、社会功能、角色功能等影响程度,这时就需要进行生命质量研究,那么怎样进行生命质量评估呢?结果的真实性如何?在不同患者、不同疾病阶段的可重复性又如何呢?

第一节 基本概念

一、生命质量的定义

1993年世界卫生组织(World Health Organization, WHO)生命质量评估组定义生命质量(quality of life, QOL)为“个体在不同的文化背景和价值体系下,与个体目标、期望、标准以及所关心的事物有关的生存状况的体验”。该定义是建立在WHO“健康”定义的基础上,WHO对“健康”的定义为:“健康,不仅意味着没有疾病,而且是生理、心理及社会功能等方面都要处于一种完全的良好状态”;是顺从于生物-心理-社会医学模式的转变。生命质量的内涵强调依从于每一个个体独特的生活体验,具有明显的主观特性,在不同的种族和文化群体中,生命质量的内涵可以是不同的。

一名类风湿关节炎女性患者表现为双手近端指间关节、腕关节肿胀和压痛,关节活动包括双手指、腕屈伸明显受限,即患者存在生理功能损伤。穿衣、梳头、开水龙头等困难,无法进行整理房间等家务,患者生活自理能力明显受阻。由于疼痛、关节功能受限和疲劳,患者不能步行超过1km,显著影响睡眠,体重下降,出现整体功能状况受到影响。患者自述无法上班,所喜爱的看电影和打网球活动均停止,其失业状态和娱乐活动不能完成是社会功能受限的表现。由于受疾病和外部环境变化的影响,患者情绪低落,心理状态恶化。因此,仅观察患者关节炎的症状与体征等与疾病诊断相关的临床指标(关节疼痛与肿胀、血沉、C反应蛋白等)可能会低估疾病对患者影响的严重程度,而进行生命质量的测定能更全面地反映疾病不良影响的真实情况,为制定综合干预措施提供依据和评价方法。

二、健康相关生命质量的定义

从事临床的医务工作者们为了全面评价健康服务对患者的影响,把生命质量的理论和医学



实践结合起来,提出了“健康相关生命质量”(health-related quality of life, HRQL),即“在疾病、医疗干预、个体经济收入、老龄化、社会环境变化等影响下的健康状况,是与经济、文化背景和价值取向相联系的主观满意度。”

三、健康相关生命质量的评估

由于生命质量的大多数内容为受试者的主观体验,因此常用量表来评价和动态随访目标人群生命质量的水平和变化。

(一) 量表的构成

量表(scale)通常构架为生理功能、心理功能、社会功能、总体状况、角色功能等不同的维度(dimension 或称为领域 domain,或亚量表 subscale);每一维度又可以包含不同方面(facet),如生理功能可以分为日常生理功能、认知水平、角色功能等三个方面,躯体功能分为疾病相关症状、治疗不良反应等方面;每一个方面又可以由多个不等的条目(item)构成。

(二) 量表的分类

量表主要分为两种:一类是测量患者的一般健康状态的普适量表(generic QOL instrument);另一类是疾病特异性的专用量表(specific QOL instrument)。这两种类型量表评估的是患者健康相关生命质量的不同方面,因而具有互补性。

普适量表包含常规生活活动方面和心理状况,其条目内容可以适用于不同种类的疾病,适用于不同种族、不同语言人群。普适量表的优势在于不受疾病诊断影响,能够直接比较不同疾病的严重性、采取不同干预措施的利弊,比如类风湿关节炎与糖尿病、高血压的治疗比较,有利于决策者权衡不同疾病的损害或治疗对病人的影响,从而对有限的资源作出有效的决策。然而,这类量表常常未包含与研究疾病特征有关的条目,忽视了研究疾病所受影响的重要功能方面,因而该类量表用于特殊疾病评估时的信度往往不高,可能存在对某些条目应答率低。

疾病专用量表是为某一类疾病或该类疾病某一亚型制定的量表,能有效地反映该类疾病对患者生命质量的影响,适用于该类疾病不同干预措施的比较。但是,专用量表也有它的局限性,因为不同地区文化传统、信仰、风俗、生活方式等不同,有时需要建立不同的量表,也无法对不同的疾病进行同等的评价。

类风湿关节炎是一种常见的慢性疾病,健康相关生命质量是治疗类风湿关节炎疗效评估的主要指标之一,评价类风湿关节炎 HRQL 的量表需能识别出疾病的严重度。常用的普适量表包括健康效用指数(the Health Utilities index, HUI)、欧洲多维健康量表(EuroQOL, EQ-5D)和六维健康量表(Short Form 6D, SF-6D)等;类风湿关节炎的专用量表有类风湿关节炎生命质量调查表(RAQOL)和躯体功能评估量表(Health Assessment Questionnaire, HAQ)等。一项纳入 313 例类风湿关节炎患者的研究中比较了这些量表对疾病严重度的评价效力,如预期的一样,疾病专用量表 RAQOL、HAQ 在住院、误工、需要专职医疗和家庭服务的患者中得分较不住院、没有误工、不需要专职医疗和家庭服务者高,而应用普适量表 HUI2、EQ-5D、SF-6D 则不能更好地识别出类风湿关节炎的严重病例(表 2-5-1)。

表 2-5-1 不同量表评价类风湿关节炎疾病严重度

参数	平均值(标准差)		总效用值(平均数,标准差)		
	RAQOL	HAQ	HUI2	EQ-5D	SF-6D
因类风湿关节炎住院					
是	14.9(7.9)	1.38(0.73)	0.63(0.21)	0.60(0.11)	0.61(0.26)
否	12.3(8.2)	1.05(0.77)	0.72(0.19)	0.63(0.13)	0.67(0.23)



续表

参数	平均值(标准差)		总效用值(平均数,标准差)		
	RAQOL	HAQ	HUI2	EQ-5D	SF-6D
效力	0.32	0.44	0.45	0.25	0.24
因类风湿关节炎误工误学					
是	16.4(7.7)	1.27(0.71)	0.67(0.19)	0.59(0.11)	0.59(0.28)
否	10.1(7.8)	0.83(0.75)	0.76(0.19)	0.67(0.13)	0.72(0.22)
效力	0.81	0.60	0.47	0.66	0.52
因类风湿关节炎需专职医疗或家庭服务					
是	15.1(7.9)	1.34(0.72)	0.66(0.20)	0.59(0.12)	0.60(0.26)
否	11.0(8.1)	0.80(0.74)	0.75(0.18)	0.65(0.13)	0.71(0.22)
效力	0.51	0.74	0.47	0.48	0.46

(三) 患者报告结果

HRQL 可以由观察者(如医护人员)判定,如步行、穿衣或打开瓶盖。近年来国外多采用患者报告结果(patient reported outcomes, PROs)的形式,即患者报告他们自我感受、正在从事的活动以及他们对自身健康及生活满意程度的信息。PROs 包括了来自于患者的一切报告。通过在医院诊所、病人记录日记或其他可能的方式获得病人对疾病或治疗的反应,可以是单项指标的测量、事件记录、症状主诉或有关健康相关生命质量量表填写等。PROs 从患者角度了解其疾病状况和身体的总体感觉,有利于患者参与治疗决策,有助于了解疾病负担,为健康资源分配决策提供指南。PROs 已成为重要的临床研究评价指标如新药的疗效等。

学
习
笔
记

一项新型口服药物治疗类风湿关节炎的随机双盲对照研究中,采用了 PROs 评估药物对患者疼痛、躯体功能和健康状态的作用,PROs 测定项目包含了患者对关节炎疼痛的自我评估、患者对疾病活动性的自我评估、HAQ 和 SF-36 量表的填写。研究结果显示与安慰剂相比,应用新型药物治疗后更多的患者关节疼痛缓解,获得了有临床意义的 HAQ 分值下降以及包括躯体、心理评估在内的 SF-36 各亚量表的改善,并随着药物剂量的增加,改善程度更显著。PROs 为治疗措施有效性和安全性的评估提供直接来源于患者的依据,尤其是外部观察者或客观检测工具无法获得的主观感受和主动活动能力,随着对 PROs 方法学的不断深入,PROs 能有效而敏感地反映治疗措施的综合疗效,可实现与其他评价指标的比较和整合,在临床决策中扮演重要的角色。

知识点

- 生命质量的定义:生命质量包含了生理健康、心理健康、社会功能以及个体机体功能等内容;生命质量的定义和内涵体现尊重个体的价值观,尊重个体的社会价值。
- 健康相关生命质量的评估形式:主要以量表的形式实现。
- 普适量表与疾病专用量表的区别:两者最主要的区别为是否针对特定疾病设计。
- PROs 的特征:将患者的主观体验转化成客观数据,使患者成为医疗保健行为的主动者。

第二节 评价原则

卫生保健系统不仅以减轻疾病负担为己任,而且越来越意识到患者优先的重要性,包括独



立自主的意愿,维持重要活动的能力,达到日常生活各方面的自我保健意识以及其他有关患者生命质量的因素,而不仅仅是疾病控制以及死亡。实际上,美国国家食品药品监督管理局(FDA)要求临床试验必须包含 HRQL 评价以证明其有效性。在此背景下,有关 HRQL 的研究文献数量近十年逐年递增,并且仍在快速增长中。

虽然目前尚缺乏统一的 HRQL 研究及相关文献报道质量的评价系统,许多 HRQL 研究的国际组织和专家致力于制定相关推荐或指南,以保证高质量 HRQL 研究开展和文献报道,其中具有一定影响力的包括特别针对 PROs 的临床试验 CONSORT 声明扩展版(2013 年)、国际生命质量研究协会(ISOQOL)关于在以病人为中心的结局和比较效果研究中应用 PROs 的最低标准推荐(2012 年)、基于专家共识的健康状态测定工具选择标准(COnsensus-based Standards for the selection of health status Measurement INstruments, COSMIN)(2010 年)以及 FDA 发布的 PROs 测定工具指南(2009 年)等。

一、HRQL 量表分析的文献评价

开发新的 HRQL 测定工具如量表、对已有的测定工具进行修改或将其应用于新的疾病之前均需对该测定工具的可靠性、有效性及反应性进行分析研究。FDA 指南旨在指导 HRQL 测定工具的建立, COSMIN 为寻找高质量的 HRQL 测定工具提供了标准,因此,现将根据 FDA 指南和 COSMIN 标准的具体内容介绍 HRQL 测定工具分析研究文献评价的主要方面(表 2-5-2)。

表 2-5-2 HRQL 分析型研究文献评价的主要方面

主要方面		检验参数或考察内容举例
可靠性	重测信度或评定者内信度	组内相关系数
	评定者间信度	组间相关系数
	内部一致性	Cronbach's α 系数
有效性	内容效度	所有条目对量表构成方面的适用性 所有条目与研究人群特征相关性 所有条目与量表测定目的相关性 患者对量表内容的理解能力
	结构效度	区分和聚合效度 效标效度
	效标效度	ROC 曲线(连续性变量) 敏感度和特异度(二分类变量)
	文化调适效度	目标语言翻译本 参与翻译专家本 条目的翻译和回译 译本调整的策略 预实验
反应性		最小显著差异 患者随时间的变化

英国专家研发的狼疮生命质量量表(LupusQoL)是应用于系统性红斑狼疮(systematic lupus erythematosus, SLE)的专用量表,目前已翻译成 77 种语言,在 51 个国家的患者中使用。LupusQoL 包含了 34 个条目组成的 8 个亚量表,分别为躯体健康、疼痛、医疗计划、亲密关系、他人负担、情绪健康、体型和乏力。现以 LupusQoL 简体中文版应用于中国患者的研究为例,介绍 HRQL 量表分析研究文献评价中涉及的具体内容。

(一) 可靠性

可靠性(reliability)即信度,是对测定工具所得结果的稳定性和对变化的反应性的评价。影响量表信度的三大因素是测定内容、测定时间和评定者。对这三大因素常用的检验方法有重测



信度、评定者间信度、内在一致性等。

1. 重测信度(test-retest reliability) 如果测量和重新测量间的时间间隔足够长,可以使同一患者记忆效应最小化,而且重复测量所包含的信息量最大。对于缓解和复发交替或周期性疾病来说,重测效度可能难以完成或不可能完成。常应用 Kappa 值来检验两次测量结果一致性程度,若是连续变量可采用组内相关系数(intraclass correlation coefficients, ICC)进行统计。一般来说, Kappa 在 0.4~0.75 或 ICC>0.6,说明重测信度较好;若 Kappa 或 ICC>0.75,则极好。

共 220 例 SLE 患者纳入该项研究,随机抽取了 20 例患者相隔两周时间两次完成量表,结果所有亚量表 ICC 均大于 0.80,提示该量表有极高的重测信度(表 2-5-3)。

表 2-5-3 LupusQoL 应用于中国狼疮患者的信度

亚量表	条目数	重测信度	内部一致性信度
躯体健康	8	0.892	0.890
疼痛	3	0.956	0.913
医疗计划	3	0.927	0.918
亲密关系	2	0.880	0.965
他人负担	3	0.974	0.931
情绪健康	6	0.970	0.961
体型	5	0.877	0.811
乏力	4	0.836	0.824

2. 评定者间信度(inter-reviewer reliability) 能反映不同测定者间的一致性,测定工具管理的标准化和对测定者进行该标准的培训是提高评定者间信度的前提。多采用 Kappa 值或组间相关系数进行检验。

3. 内部一致性信度(internal consistency reliability) 如 α 系数(Cronbach's alpha)它是检验亚量表内部条目之间一致性水平的指标。理想情况是每一个亚量表 α 系数均 ≥ 0.70 ,达到这一水平说明量表各条目所测内容具有同源性。需注意的是,在临床研究中,在缺乏重测信度的前提下,内部一致性信度将受到质疑。

表 2-5-3 列出 LupusQoL 8 个亚量表 α 系数,可见各亚量表 α 系数均大于 0.80,表明该量表内部一致性好,结合重测信度结果,可以判断 LupusQoL 应用于中国患者有较好的可靠性。

(二) 有效性

有效性(validity)即效度,是指量表包含内容是否全面反映被检测者生命质量内涵,是否实现生命质量测定目的。效度是检测量表测定内容与真实情况的吻合程度,也就是检验量表的真实性和准确性。效度评价包括内容效度、结构效度、效标效度、文化调适效度等。

1. 内容效度(content validity) 表明测定的内容能否真实反映或真正代表所要测定现象,是首要关注的量表特性之一,但内容效度涉及的方面非常广泛,且无法应用定量方法进行检验,故对内容效度优劣的评价存在困难,仅能观察文献中有关内容效度的描述是否存在、准确以及是否全面和详细。内容效度包括条目生成(如条目生成方法,条目来源、选择、编辑和删减,条目评估的定量工具,研究对象等)、数据收集方式和工具监管模式、回忆周期、反应选择项、测定工具标准化和培训、患者理解能力、条目和亚量表的评分系统、应答者和管理者责任等。

LupusQoL 中国研究的文献报道在方法学部分详细介绍了 SLE 患者特征以及数据收集、计算和统计的方法,为确保内容效度,LupusQoL 简体中文版是在已使用的繁体中文版基础上进

行修订,且研究者进行了预试验,评价所有条目是否被患者理解,并根据反馈情况进行了语言表达上的调整。这些报道内容虽不够全面,但尽可能保证了LupusQoL 简体中文版的内容效度。

2. 结构效度 (construct validity) 是指测定工具测到所要测量理论结构和特质的程度,表明研究结果与假设之间的一致性,即测定工具是否真正测量到假设检验。结构效度又可分为区分效度 (discriminant validity) 和聚合效度 (convergent validity)。区分效度是应用测定工具能识别出具有或不具有某种特征的检测对象。

聚合效度是指不同测量方法测定同一特征时测量结果的相似程度。

LupusQoL 中国研究应用 SLE 疾病活动指数 (SLEDAI) 和国际狼疮组织损伤指数将患者分为稳定期和活动期,观察 LupusQoL 量表鉴别出活动患者的能力,即区分效度检验(见表 2-5-4),结果显示除体型外,稳定期和活动期患者在其余各亚量表上的得分具有显著性差异,表明 LupusQoL 能有效识别出疾病活动的患者。此外,Mann-Whitney 非参数检验的结果显示 LupusQoL 与 EQ-5D 在相似亚量表的得分上呈现出较强的相关性,即 LupusQoL 与 EQ-5D 具有较好的聚合效度(表 2-5-5)。

表 2-5-4 基于疾病活动度的 LupusQoL 区分效度

亚量表	SLEDAI		P
	0~4	>4	
躯体健康	87.43 (13.17)	57.77 (25.70)	0.000
疼痛	87.24 (16.18)	57.58 (27.59)	0.000
医疗计划	79.86 (21.29)	53.03 (28.01)	0.000
亲密关系	62.34 (31.06)	37.93 (30.16)	0.000
他人负担	68.19 (26.65)	49.49 (31.73)	0.002
情绪健康	78.05 (20.31)	56.19 (31.27)	0.000
体型	75.34 (18.08)	72.12 (15.61)	0.113
乏力	83.00 (16.53)	68.18 (21.78)	0.000

表 2-5-5 LupusQoL 的聚合效度

LupusQoL 亚量表	EQ-5D 亚量表	Spearman's r
躯体健康	普通活动	-0.630
疼痛	疼痛 / 不适	-0.778
情绪健康	焦虑 / 抑郁	-0.761
医疗计划	普通活动	-0.560

3. 校标效度 (criteria-related validity) 是指所建量表和已有的标准测定结果(“金标准”)一致程度。标准测定结果可以是成熟的量表、临床指标或长期临床随访结果。效标效度分为两种:一种为平行效度 (concurrent validity),即用所评定量表和标准测定工具同时测定一组研究对象,评价两者测定结果的一致性。另一种是预测量表 (predictive validity),量表能否预测未来事件(如疾病复发、治疗反应等)。然而,大多数生命质量研究很难找到“金标准”,如在研究中使用了标准测定工具,则需提供它的基本原理、敏感度、特异度和预测值等信息。

4. 文化调适效度 (cross-cultural validity) 由于社会、经济、信仰、文化背景等诸多方面的差异,同一量表应用于不同国家、不同民族、不同语言人群过程中,需要对量表进行翻译和文化调适。此类文献报道应注明源语言版本的发展情况和目标语言版本的翻译过程,进行翻译和回译的专家信息,翻译者间是否独立,文化差异的调整内容,检验患者理解难易程度的预试验等。



LupusQoL 简体中文版研究的文献报道在背景资料中介绍了英文版的开发和应用情况,并特别指出简体中文版源于已在中国台湾应用的繁体中文版。两名专家(一位风湿病学专家和一位中文语言学家)分别独立完成简体中文版的翻译,并由多学科共识委员会讨论制定统一版本。随后,随机选择 6 例 SLE 门诊患者进行预试验,根据反馈信息,调整“整理花园、画画”为“搬运气罐、米袋”,“品质”改为“质量”等。另外 6 例 SLE 门诊患者完成修订版本后没有提供新的建议。该文献较详细地描述了量表的文化调适过程,预试验结果显示 LupusQoL 简体中文版可以被中国患者理解和使用。

(三) 反应性

反应性(reaction)随时间变化或干预措施后,病人的情况发生改变,相应生命质量也发生变化,量表反映这种变化的能力称之为反应性。临床研究常以“统计学显著”来判定疗效差异,但仅仅数值上有显著差异不代表一定具有临床意义。患者需改善多少,患者或医护人员才认为此改变具有重要性或有意义,该阈值即为最小显著差异(minimal important difference, MID)。MID 作为定量的反应性指标经常被用于治疗疗效评估和比较的研究。

应用药物成瘾者生存质量评定量表(QOL-DA)和 SF-36 评定吸毒者戒毒前和戒毒 12 周后生命质量变化,结果显示 QOL-DA 能敏感地反映出吸毒者生命质量的改变,而 SF-36 未能反映出这种改变(表 2-5-6)。

表 2-5-6 戒毒前后 QOL-DA 和 SF-36 量表得分的变化

项目	戒毒前	戒毒 12 天后	<i>t</i>	<i>P</i>
QOL-DA	132.85 ± 40.33	148.00 ± 32.79	2.62	0.01
SF-36	96.14 ± 16.77	96.34 ± 17.33	0.33	0.74

二、HRQL 量表应用的文献评价

越来越多的临床研究将生命质量评价结果尤其是 PROs 作为研究的主要终点或次要终点,所获得的数据被用于临床决策和健康政策制定。高质量 HRQL 应用型研究的开展和文献报道是保证所得数据准确性、可利用性的前提。CONSORT 声明(表 2-5-7)和 ISOQOL 推荐均对提高 PROs 应用型研究报道质量制定了规范,可作为 PROs 应用型研究文献评价的参考标准。

表 2-5-7 2013 年 CONSORT 扩展版有关 PROs 的内容

框架	条目	针对 PROs 的扩展内容
标题和摘要	摘要	标识出 PROs 作为主要终点或次要终点
介绍	背景和目的	PROs 评价的背景和基本原理 阐明 PROs 假设及列出相关亚量表
方法学	研究对象	无特殊性,除非 PROs 被用于入选或分层标准
	结果	提供或引用 PROs 测定工具的效度和信度包括完成 PROs 人员和数据收集信息(纸质版、电子版、电话询问等)
	样本量	不要求,除非 PROs 作为主要终点
随机化	统计学方法	明确阐述处理缺失数据的统计学方法
结果	研究对象流程图	清楚注明基线和各随访时间点 PROs 局数据的数量
	基线数据	收集的 PROs 基线数据



框架	条目	针对 PROs 的扩展内容
讨论	病例量分析	有关 PROs 结果的
	结果和预测	来自于各亚量表和不同时间节点的多维 PROs 结果
	补充分析	与 PROs 相关的分析
	局限性	PROs 特异的局限性 普及和临床实践的意义
	解释	PROs 数据与其他临床结局如生存数据的相关性

托珠单抗是一种白介素(IL)-6受体的拮抗剂,被应用于治疗类风湿关节炎,能显著改善患者关节肿痛。一项托珠单抗治疗类风湿关节炎的随机双盲安慰剂对照研究中,将PROs作为次要研究终点,现根据CONSORT声明的主要内容对该篇文献进行评价。

方法学中对结果的定义和收集:在该项研究中运用的PROs测定工具包括视觉模拟刻度法(visual analogue scale, VAS)、HAQ、SF-36、SF-6D、慢性疾病治疗功能评估表(Functional Assessment of Chronic Illness Therapy, FACIT)。介绍了应用各项测定工具的主要研究目的和各生命质量量表的亚量表构成,但没有提供信度和效度数据。明确注明了PROs由患者完成及进行测定的时间节点。

统计学方法:明确定义了PROs各项测定工具的MID如VAS为10毫米,HAQ大于0.22个点等。详细介绍了缺失数据产生原因及处理方法,缺失数据源于早期退出、接受副作用治疗等,主要分析中对缺失数据不进行插补,但应用最后观察结转法插补缺失数据进行敏感性分析。

结果:该文献没有提供研究对象流程图。但详细提供了基线数据包括患者年龄、性别、病程等一般资料以及PROs得分,提供了实现最小显著差异以上的患者缓解率。逐一列出了各项PROs测定工具的检测结果,托珠单抗较安慰剂显著改善患者疼痛、躯体功能、精神情绪、社会功能等生命质量评测内容。进一步分析了各测定工具的检测结果随时间的变化趋势,随着时间的延伸,托珠单抗治疗组的改善情况更明显。

讨论:该文献指出托珠单抗治疗后HRQL的改善在治疗早期即可发生,成为能改善类风湿关节炎临床表现和生命质量的有效治疗措施。但作者未提及应用PROs评估托珠单抗治疗类风湿关节炎疗效的局限性,也没有分析PROs与其他临床指标的相关性。

该文献运用多维PROs测定工具对托珠单抗治疗类风湿关节炎疗效进行评估,总体上,还是较好地符合CONSORT声明。但对PROs测定工具的原理和特性介绍较少,是否适用于类风湿关节炎患者存在疑问,也没有分析PROs与常用的疾病活动性或严重性指标的相关关系,无法获知PROs对类风湿关节炎预后的预测能力。

总之,生命质量的研究开拓了临床医学研究的一个新领域,而在该领域应用科学方法建立和评价生命质量可以帮助临床医师进一步复习和评价有关文献,并指导临床医师正确和合理地选用药物,同时还有助于医疗管理部门对资源的分配和医疗的决策。

知识点

重测信度的定义:重测信度是在相同条件下,不同时间点,同一患者完成量表所能达到的一致性程度,是体现量表稳定性的重要指标。

内容效度的评定:内容效度缺乏统一的评定指标,但在文献报道中应尽可能地详细描述内容效度的相关信息。

区分效度和聚合效度:区分效度和聚合效度均是重要的效度评价指标,其难点在于参照工具的选择上,应选择公认的有效的测定工具与需评价的量表进行比较。



量表的适用范围:反应性与量表用途和目标人群特征密切相关,决定了量表的适用范围。

健康相关生命质量(health-related quality of life,HRQL):除了疗效评价外,HRQL也被用于疾病预防、疾病诊断和预测预后等方面,亟须 HRQL 文献质量评价系统的建立。

思考题

1. 生命质量评价主要的应用领域不包括
 - A. 慢性病干预的疗效评价
 - B. 新药临床试验
 - C. 人群和病人的健康状况评价
 - D. 基础病理研究
 - E. 资源分配、计划和决策的制定
2. 若您拟设计一份类风湿关节炎生命质量评价量表,一般不纳入以下哪个方面
 - A. 药物不良反应
 - B. 日常生活能力
 - C. 关节疼痛程度
 - D. 关节活动度
 - E. 社会角色能力
3. 健康相关生命质量量表分析的主要内容包括
 - A. 重测信度
 - B. 内容效度
 - C. 结构效度
 - D. 反应性
 - E. 以上均是

参考答案

1. D 2. A 3. E

(姜林娣)

参考文献

1. Marra CA, Woolcott JC, Kopec JA, et al. A comparison of generic, indirect utility measures (the HUI2, HUI3, SF-6D, and the EQ-5D) and disease-specific instruments (the RAQoL and the HAQ) in rheumatoid arthritis. Soc Sci Med, 2005, 60 (7): 1571-1582.
2. Coombs JH, Bloom BJ, Breedveld FC, et al. Improved pain, physical functioning and health status in patients with rheumatoid arthritis treated with CP-690,550, an orally active Janus kinase (JAK) inhibitor: results from a randomised, double-blind, placebo-controlled trial. Ann Rheum Dis, 2010, 69 (2): 413-416.
3. Wang SL, Wu B, Leng L, et al. Validity of LupusQoL-China for the assessment of health related quality of life in Chinese patients with systemic lupus erythematosus. PLoS One, 2013, 8 (5): e63795.
4. Strand V, Burmester GR, Ogale S, et al. Improvements in health-related quality of life after treatment with tocilizumab in patients with rheumatoid arthritis refractory to tumour necrosis factor inhibitors: results from the 24-week randomized controlled RADIATE study. Rheumatology (Oxford), 2012, 51 (10): 1860-1869.
5. Calvert M, Blazeby J, Altman DG, et al. Reporting of patient-reported outcomes in randomized trials: the CONSORT PRO extension. JAMA, 2013, 309 (8): 814-822.

第六章 临床经济学

临床问题

临床实践中医生经常需要考虑是否昂贵的药物一定有更好的疗效,或者手术治疗是否一定比非手术治疗的预后更好,如5-羟色胺再摄取抑制剂抗抑郁药一定比三环类抗抑郁药疗效好吗?对于大多数患者来说,一般都会认为昂贵的药物一定更有效。常常挂在患者嘴上的话是“只要能治好病,多贵的药都舍得买。”这样做值得吗?要计算不同临床治疗方案的成本,要回答价格不同的治疗方案是否效果不同的问题,用什么作为临床治疗收益的评价指标?如何测量治疗方案的效果和效益?什么是效用及其测量的指标?

第一节 基本概念

一、定 义

卫生经济学是经济学在卫生领域中的应用,在发展过程中产生若干分支,包括临床经济学、保健经济学、卫生计划经济学、卫生技术经济学、医院经济管理学、医学经济学等。临床经济学(clinical economics)是临床医生应用经济学的原理和方法评价临床诊断、预防和治疗技术与措施的经济学效果,探讨影响合理利用有限资源的因素,指导临床医生在临床实践中作出科学的决策。

二、意义和目的

(一) 意义

我国是一个发展中国家,卫生投入不足,卫生总费用只占国内生产总值(GDP)的5%左右,而社会各界又明显感受到卫生费用的迅速上涨。卫生费用上涨的原因很多,既有人口老龄化、疾病谱改变、服务可及性增加、技术进步等客观原因,也有医疗补偿机制、供方诱导服务、需方浪费等因素。从供方看,我国卫生服务体系中存在着一些资源浪费的现象。目前,按服务收费的医院补偿模式和不合理的价格体系,刺激高新技术、高价药物的利用,重治轻防,“诱需争盈”现象表现明显。卫生费用的压力或许是经济学发展的外在动力。因此,将经济学的理论和方法应用于医药卫生领域的研究有着十分重要的理论和实践意义。

(二) 目的

1. 论证某卫生规划或卫生活动实施方案的可行性 通过临床经济学评价论证该方案是否具有经济可行性,即通常所说的该方案是否合算。
2. 比较改善同一健康问题的各个方案 改善同一健康问题可能会有多种方案,通过临床经济学评价对这些方案进行比较,从中选择出解决该健康问题的最佳方案。
3. 比较改善不同健康问题的各个方案 各个卫生规划或卫生活动方案所解决的问题不尽相同,通过临床经济学评价比较各个方案,可以从经济学的角度确定哪个方案最有意义,最有价值优先实施。



三、方 法

(一) 成本测量

一般而言,在临床经济学评价中,将成本分成直接医疗成本(direct medical cost)、直接非医疗成本(direct non-medical cost)、间接成本(indirect cost)和无形成本(intangible cost)四类。相应地,效益(benefit)是用货币表示卫生服务的有用效果,也可按此分类。

(二) 常用评价方法

1. 成本最小化分析 成本最小化分析(cost minimization analysis, CMA)是在效果、效用和效益没有差别的条件下,选择成本低的方案,即成本最小化分析,这是一种特例。比如,采用抗抑郁药物和重复经颅磁刺激两种方法治疗抑郁症,经过成本计算,达到相同的疗效时,药物治疗的成本小。

2. 成本效果分析 成本效果分析(cost-effectiveness analysis, CEA)是将某卫生规划或卫生活动每个方案的成本与效果相联系进行分析与评价。比如,治疗某种癌症采用化疗和手术治疗两种方法,成本相近而手术治疗提高生存率的效果好。

3. 成本效益分析 成本效益分析(cost-benefit analysis, CBA)是将某卫生规划或卫生活动每个方案的成本与效益相联系进行分析与评价。比如,采用电子病历诊治患者,虽然购买软件和调试费用较纸版病历记录高,但工作效率大大提高,诊治患者数量增加,获得更高的效益。

4. 成本效用分析 成本效用分析(cost-utility analysis, CUA)是将各个卫生规划或卫生活动实施方案的成本与效用相联系起来考虑,从而比较评价选择各种不同的方案。成本效用分析在进行产出测量时,把各个方案的不同结果都转化为效用指标,比如质量调整寿命年、失能调整寿命年、质量调整预期寿命等,使得各个方案的结果都使用一致的指标来表示。

四、应用的领域

(一) 应用于预防保健领域

对于某种疾病可以有不同的预防措施或者不同的干预人群,通过临床经济学评价可以选择最为经济的预防保健措施,或者选择最需要实施预防保健措施的人群,从而使相同的资源使用获得最大的收益。

(二) 应用于技术评估领域

当今世界高科技日新月异,现代化的诊疗技术层出不穷,通过临床经济学评价可以使人们了解各项新技术的花费以及对个体健康状况的改善,从而选择适宜的新技术。

(三) 评价并比较疾病的各种治疗方案以选择最佳方案

对于同一种疾病可以有不同的治疗方案。

(四) 应用于药品研究领域的药物经济学

从临床经济学的角度,将治疗疾病药品的花费与治疗疗效相联系,比较可以治疗相同疾病的不同药品,或者比较治疗不同疾病的不同药品。由此得出相关结论,为决策部门分配资源,为病人选择治疗方案提供依据。

(五) 评价并比较各项投资方案并作出决策

面对各种健康问题,人们有各种各样的解决方案有待投资并予以实施。众所周知,改善同一健康问题,可以加强预防保健领域的投资,也可以增加医疗领域的投资。在医疗领域,可以加强专科医院的建设,也可以加强社区初级卫生保健站的建设。但是卫生资源是有限的,卫生事业管理者和决策者可以通过临床卫生经济学评价的方法决定投资领域和投资方案,从而使有限的资金取得最大的收益。

**知识点**

卫生经济学与临床经济学的基本概念:①卫生经济学是经济学在卫生领域中应用的新兴交叉学科;②临床经济学的定义是应用经济学的原理和方法评价临床诊断、预防和治疗技术与措施的经济学效果;③临床经济学评价的意义是探讨影响合理利用有限卫生资源的因素,指导临床实践。

卫生经济学有四种常用的评价方法:①成本最小化分析;②成本效果分析;③成本效益分析;④成本效果分析。

第二节 临床经济学评价方法

一、成本的测量和分析

(一) 成本的定义

成本(cost)是商品经济的价值范畴,是商品价值的组成部分。人们要进行生产经营活动或达到一定的目的,就必须耗费一定的人力、物力和财力等资源,其所耗费资源的货币表现及其对象化称之为成本。并且随着商品经济的不断发展,成本概念的内涵和外延都处于不断的变化发展之中。

中国成本协会 2005 年出版的《成本管理体系术语》标准中的第 2.1.2 条中对成本术语的定义是:为过程增值和结果有效已付出或应付出的资源代价。应付出的资源代价指应该付出,但目前还未付出,而且迟早要付出的资源代价。资源代价是总合的概念;资源指凡是能被人所利用的物质;在一个组织中资源一般包括人力资源、物力资源、财力资源和信息资源等。该定义的成本是广义的概念,不是狭义的概念。

(二) 成本的含义

成本有三方面的含义。第一,成本属于商品经济的价值范畴,即成本是构成商品价值的重要组成部分,是商品生产中生产要素耗费的货币表现;第二,成本具有补偿的性质,它是为了保证企业再生产而应从销售收入中得到补偿的价值;第三,成本本质上是一种价值牺牲,它作为实现一定的目的而付出资源的价值牺牲,可以是多种资源的价值牺牲,也可以是某些方面的资源价值牺牲;甚至从更广的含义看,成本是为达到一种目的而放弃另一种目的所牺牲的经济价值,在经营决策中所用的机会成本就有这种含义。我国处在多种所有制的生产主体同时并存的体制下,成本的含义应与目前的经济体制相适应。

(三) 成本的分类

临床经济学常用的成本分类有 7 类。

1. 直接成本和间接成本 根据生产费用计入产品成本的方式所划分的直接成本(direct cost)和间接成本(indirect cost)。

直接成本是指在成本核算中可以直接计入某一成本核算对象的费用,表示卫生服务成本,指的是卫生项目实施和卫生活动直接消耗的资源或所花的代价,即将资源用于直接提供疾病诊断、治疗、服务、预防等花费的成本;临床经济学评价中常用直接医疗成本和直接非医疗成本,前者包括疾病诊治所需的药品花费、检查花费等的成本,后者包括为就医所需的交通费、餐费、住宿费等成本。

间接成本指在成本核算中不能直接计入成本核算对象,而必须按一定的标准分摊于不同成本核算对象的费用,表示社会成本,指的是由于疾病而丧失的资源。间接成本是指在卫生活动



的实施与开展过程中,所引起的间接的代价,如医院的行政管理成本、辅助科室成本、固定资产折旧等。

2. 固定成本和可变成本 按成本与医疗服务量的关系分为固定成本(fixed cost)和可变成本(variable cost)。固定成本指成本总额在一定时期和一定业务范围内,不受业务量增减变化的影响而固定不变的那部分成本。可变成本指成本总额随业务量增减而变化的那部分成本。

3. 可控成本和不可控成本 按成本的可控性划分可分为可控性成本(controllable cost)和不可控性成本(uncontrollable cost)。可控成本指通过管理活动能够直接控制其变化的那部分成本;不可控成本指那些通过管理活动不能改变其数额的成本,是实施规划方案所必须负担的最低成本。

4. 有形成本和无形成本 有形成本(tangible cost)是在实施或接受医疗服务中所消耗的产品或服务的成本,其特点是伴随着资源的耗费而发生的。无形成本(intangible cost)是因疾病引起的或因实施医疗服务引起的患者及其家属在躯体或精神上的不便、痛苦、忧虑或紧张等负性情绪波动,以及引起的意愿、声誉受损或社会不安定等。无形成本也叫隐性成本,其特点是发生并不伴随资源的耗费。

5. 平均成本与边际成本 平均成本(average cost)是单位产出或服务的资源消耗,即总成本、总服务量或总产出。边际成本(marginal cost)是多提供一单位产品或医疗服务所需增加的成本量。

6. 疾病成本和治疗成本 疾病成本(cost of illness)是疾病的经济负担,因疾病造成的资源耗费和代价,包括因病导致的生产能力丧失或失能以及死亡的损失、患者及其家属的误工损失,以及疼痛或痛苦等无形成本等。治疗成本(cost of therapy)是为诊断、治愈、缓解、控制疾病所消耗的资源或代价。

7. 机会成本和沉没成本 机会成本(opportunity cost)是将某种具有多种用途的有限资源用于某种特定用途时所放弃的,但置于其他用途时可能带来的最大收益。当面临多个选择机会时,因选择了某个机会而不得不放弃其他机会,因而放弃了利用其他机会所可能获得的收益,在所放弃的所有机会中可能获得的最大收益就是所做选择的机会成本。沉没成本(sunk cost)是以往发生的、与当前决策无关的成本。

(四) 成本分析

成本分析(cost analysis)指利用成本核算资料及其他有关资料,全面分析成本水平及其构成的变动情况,研究影响成本升降的各个因素及其变动的原因,寻找降低成本的规律和潜力。通过成本分析可以正确认识和掌握成本变动的规律性,不断挖掘内部潜力,降低产品成本,提高经济效益;通过成本分析还可以对成本计划的执行情况进行有效的控制,对执行结果进行评价,肯定成绩,指出存在的问题,以便采取措施,为编制下一步成本计划和作出新的决策提供依据,给未来的成本管理指出努力的方向。医疗服务成本的分析是对服务项目的实际成本进行分析,其目的在于了解医疗卫生机构的成本现状和趋势,认识成本变动的规律,寻求挖掘潜力的措施、途径和方法,努力降低医疗成本,又称为成本最小化分析(cost minimization analysis, CMA)。

二、成本效果分析

(一) 效果测量

效果测量就是产出测量。广义地讲,效果指相关卫生规划或卫生活动的方案实施后所取得的结果,可能是好的结果,也可能是不好的结果。比如,实施预防接种可以提高人群对传染性疾病的免疫率,从而降低传染病的发病率,这是该卫生规划的方案实施后取得的好结果。相反,如果在预防接种的过程中,由于某些原因造成冷链中断或疫苗污染,由此引发不必要的传染病流行,该卫生规划就取得了不好的结果。

狭义地讲,效果指的是好的、有用的结果,也就是能够满足人群需要、给人们带来好处或满



足感的结果。除了前面所举的提高免疫率的例子,通过卫生规划或卫生活动的方案实施,各种健康指标、卫生问题改善指标和服务利用指标等的改善都属于这个范围。效果指标既可以是绝对指标,比如就诊人次数的增加,早期诊断例数的增加,治愈病人数的增加等;也可以是相对指标,比如床位利用率的提高,发病率的下降,孕产妇死亡率的下降,婴儿死亡率的下降等。一般而言,在分析中提到的效果如果没有特殊说明,都是指狭义的效果。

(二) 成本效果分析的定义

成本效果分析(cost-effectiveness analysis, CEA)是指将某卫生规划或卫生活动每个方案的成本与效果相联系进行分析与评价。可以根据下面的原则对每个方案进行评价,从而确定每个方案是否可行,并且比较各个方案以确定其中的最佳方案。将投入产出相联系进行评价即为成本效果分析。

(三) 成本效果分析的基本原则

成本效果分析的基本原则:①相关卫生规划或卫生活动方案的成本尽量低,同时取得的效果尽量好。②明确卫生规划或卫生活动方案的实施是否存在成本上限,也就是预算约束。③明确卫生规划或卫生活动方案的实施是否存在期望效果下限。④成本效果分析中成本采用的是货币形式,而效果却采用的是健康指标、卫生问题改善指标或卫生服务利用指标等。因此,在成本效果分析过程中,不同方案之间的效果应该具有可比性。

一位30岁的妇女泌尿道感染反复发作已有数年,平均每年发作3次,可以使用抗菌药物预防疗法,也可以发作时再治疗。两种方法的效果都是使泌尿道感染不再发作,可以通过成本效果分析选择最佳治疗方案。调查分析表明,如果发作时治疗,一次费用至少需要126元,则每年的总治疗费用为 $126 \times 3 = 378$ 元;如果采用长期口服复方磺胺甲基异噁唑预防,可将每年平均发作的次数降至0.15次,则每年的总治疗费用为 $126 \times 0.15 = 18.9$ 元,加上预防药物的费用为85元,总计每年需要花费 $85 + 18.9 = 103.9$ 元;可以看出两种方案虽然取得效果相同,但是采用抗菌药物预防疗法花费要少得多,从临床经济学的角度可以将其选择为最佳的治疗方案。

三、成本效益分析

(一) 效益测量

效益测量即产出测量,效益是将相关卫生规划或卫生活动方案实施所获得的有用结果以货币的形式表达。人力资本法是对疾病治疗方案的健康产出根据患者避免损失的实际工资进行计算。陈述偏好法是采用调查的方式,让受访者表达疾病治疗方案所带来的效益,以此对偏好进行估算。显示偏好法是估算个人对于某种产品或服务的偏好,在实际购买决策行为中观察个人的偏好强度。

(二) 成本效益分析的定义

成本效益分析(cost-benefit analysis, CBA)是通过比较某卫生规划或卫生活动项目的全部成本和效益来评估项目价值的一种方法,其基本原理是针对某项支出目标,提出若干实现该目标的方案,运用一定的技术方法,计算出每种方案的成本和收益,通过比较方法,并依据一定的原则,选择出最优的方案。成本效益分析作为一种经济决策方法,将成本费用分析法运用于政府部门的计划决策之中,以寻求在投资决策上如何以最小的成本获得最大的效益。常用于评估需要量化社会效益的公共事业项目的价值。比如,现有一个卫生规划拟治疗冠心病患者,经过该规划方案的实施,治愈冠心病患者,会减少就医费用,包括诊疗费,住院费,检查费,药品费等各种费用;减少由于就医所造成的额外费用,包括市内或远程交通费,额外营养费,外地住宿费等;



由于许多患者原来患病时需要专人看护,治愈后患者本人和负责看护的家庭成员都可以重返工作岗位获得收入,而且心情愉快,家庭幸福。所有这些减少的费用和增加的收入都是该冠心病患者治疗规划实施所获得的效益。

(三) 成本效益分析的基本原则

成本效益分析的基本原则:①相关卫生规划或卫生活动实施方案的成本尽量低,同时取得的效益尽量好。②成本效益分析中卫生规划或卫生活动实施方案的产出是用效益描述,也就是货币的形式表达,因此可以直接比较各个方案本身的成本与效益。③对于某一具体的方案,应该考虑在方案周期内所有的资金发生情况,包括所有的成本投入和效益产出。只有当所获得的总效益不低于同期所投入的总成本时该方案才是可行的。④在进行不同方案之间的比较时,可以首先计算各个方案的效益与成本的比较值,然后再将各个方案的“比较值”进行比较,该“比较值”越大,该方案从成本效益分析的角度就越有意义,越有价值,可以确定为优选方案。⑤由于各个方案的成本和效益可以发生在不同的年份,因此需要考虑资金的时间价值。通过年利率将各年的成本和效益都折算为同一基准年的现值,从而便于进行不同方案之间的比较。⑥只要卫生规划或卫生活动的效果可以转化为货币的形式,就可以利用成本效益分析方法比较不同卫生规划或卫生活动的方案,并选择最佳方案,并不要求方案所取得的效果具有可比性。

比利时的一项研究表明,用疫苗接种预防肺炎感染,在成年人中可延长存活大约2年,但每年的成本是11 250欧元;而老年人则可延长存活9年以上,每年的成本为125欧元。因此,结论是支持接种老年人的方案效益更好。

四、成本效用分析

(一) 效用测量

效用(utility)测量就是产出测量,效用是指人们所获得的满足感。众所周知,各种卫生规划和卫生活动,甚至整个卫生行业的最终目标都是改善人群的健康水平,提高人群的生活质量,使人们获得更大满足感。只有达到这个目标,卫生规划和卫生活动才是最终有意义的。效用多采用质量调整寿命年(quality adjusted life year, QALY)、伤残调整寿命年(disability adjusted life year, DALY)来衡量。QALY可以直接或间接进行测量,直接测量采用标准博弈法(standard gamble technique, SG)、时间权衡法(time trade-off, TTO)和人群权衡法(population trade-off, PTO);间接测量采用健康质量调查表(quality of well-being, QWB)、健康效用指数(health utility index, HUI)和欧洲五维度健康质量表(EuroQol 5-Dimensions, EQ-5D);也可将生存质量简表36(short form 36, SF-36)转化成效用。

(二) 成本效用分析的定义

成本效用分析(cost-utility analysis)是将各个卫生规划或卫生活动实施方案的成本与效用联系起来考虑,从而评价选择各种不同的方案。在某种意义上,成本效用分析是成本效果分析的一种发展。成本效用分析在进行产出测量时,把各个不同方案的不同结果都转化为效用指标,比如质量调整寿命年、失能调整生命年、伤残调整寿命年等。由于各个方案的结果都使用一致的指标来表示,与成本相联系对方案进行评价,比较解决不同健康问题的方案更为方便,不像成本效果分析严格要求方案效果的可比性。成本效用分析用同一个效用测量单位来比较不同卫生项目和政策,重视个人的生活质量,因此比简单的成本效益分析更全面地分析效益。

(三) 成本效用分析方法

1. 质量调整寿命年为效用指标 生命年作为效用指标没有考虑由于生存质量差异所造成人们满足感的差异,可以采用质量调整寿命年(quality adjusted life year, QALY)来弥补这个缺点。



计算不同生命质量(健康状况)的存活年数相当于多少生命质量(健康状况)为完全健康的存活年数,再与生命数相乘,计算所得的生命年数为质量调整寿命年,用于表示各个卫生规划或卫生活动的方案实施后所获得的效用。

可以赋予不同的生活质量(健康状况)以不同的质量权重,完全健康的质量权重为 1,死亡的质量权重为 0,其他生活质量的权重介于 0 与 1 之间。通过使用不同健康状况的质量权重,可以将不同健康状况的生命年数转化为统一的质量调整寿命年。可以看出,质量权重的确定是一个非常重要的环节。在实际工作中,可以对某个卫生规划的目标人群进行调查,了解他们对各种健康状况的评价,从而获得质量权重;也可以使用专家判断法来确定质量权重。获得各个卫生规划或卫生活动的方案实施所能够获得的质量调整寿命年之后,便可以进行相关的成本效用分析。

以质量调整寿命年作为效用指标时的评价原则:①对于某个具体的卫生规划或卫生活动的实施方案,如果该方案的实施可以获得的质量调整寿命年大于 0,那么这个方案是有意义的,可以采纳;②比较不同卫生规划或卫生活动的实施方案,计算各个方案获得单位质量调整寿命年所需要花费的平均成本,平均成本最低的方案是最优的方案,可以优先选择。

慢性肾衰竭的病人既可以采用血液透析疗法,也可以采用肾移植的疗法。利用临床卫生经济学评价方法,测量病人在整个治疗过程中与治疗疾病相关的所有花费、治疗后带来的寿命延长以及生存质量的改善,并且进行成本效用分析,可以从临床经济学的角度评估,因人而异地选择最佳方案。

2. 伤残调整寿命年为效用指标 伤残调整寿命年(disability adjusted life year, DALY)指从发病到死亡所损失的全部健康年,包括因早死所致的寿命损失年和疾病所致伤残引起的健康寿命损失年两部分,是综合评价各种非致死性健康结果(包括各种伤残状态)与早死的效用指标,可以用来衡量人们生命与健康状况的改善情况。

伤残调整寿命年的目标与质量调整寿命年是一致的,都是为了不仅仅考虑各种卫生规划或卫生活动对减少早死所做的贡献,而且还考虑这些规划与活动对于改善非致死性健康状况所做的贡献。与质量调整寿命年不同的是,伤残调整寿命年的重点是确定与选择各种状态的伤残权重,而质量调整寿命年的重点是确定与选择各种状态的质量权重。伤残调整寿命年计算的是健康的损失,而质量调整寿命年计算的是健康的获得。

在确定伤残权重时,主要考虑不同伤残状态对人们生活的影响。不同的疾病,不同的伤残状态可能影响人们不同的具体功能,比如偏瘫的病人终日卧床,日常活动比如吃饭、个人卫生及大小便均需要别人的帮助;类风湿性关节炎的病人由于关节活动受限,日常生活也需要别人的帮助,虽然病人的具体状态不一样,但是疾病和伤残状态对他们的生活所造成的影响却基本相同,因此在确定伤残权重时,可以考虑赋予两个状态以相同的权重。

以伤残调整寿命年作为效用指标时的评价原则:①对于某个具体的卫生规划或卫生活动的实施方案,如果该方案的实施可以挽救的伤残调整寿命年大于 0,那么这个方案是有意义的,可以接受;②比较不同卫生规划或卫生活动的实施方案,计算挽救每一伤残调整寿命年损失所需要花费的平均成本,平均成本最低的方案为最优方案,从经济学的角度应该优先选择。

自 1985 年到 1991 年荷兰全国卫生保险基金会对心脏移植、肝移植和肺移植三种技术进行评估,在达到相同健康指标(获得 1 DALY)的情况下,三种技术的花费分别为 44 000 美元,30 000 美元和 63 000 美元。由此得出结论,从临床经济学评价的角度评估肝移植技术是最佳的。



知识点

临床经济学常用的成本分类:成本分析是临床经济学评价的首要步骤。常用的成本分类包括直接成本和间接成本、固定成本和可变成本、可控成本和不可控成本、有形成本和无形成本、平均成本和边际成本、疾病成本和治疗成本、机会成本和沉没成本等。

成本效果分析中效果的含义与意义:成本效果分析中的效果是各种健康指标的改善和服务利用指标等的改善。成本低、效果好的治疗方案为最佳方案。

成本效益分析中效益的含义与意义:成本效益分析中的效益是将卫生活动方案实施所获得的有用结果以货币的形式表达。在进行不同方案之间的比较时计算各个方案的效益与成本的比值,然后再将各个方案的“效益成本比”进行比较。

成本效用分析中效用的含义与意义:成本效用分析中的评价指标是质量调整寿命年和伤残调整寿命年。对不同的生活质量(健康状况)赋予不同的质量权重。在确定伤残权重时主要考虑不同伤残状态对人们生活的影响。

第三节 临床经济学文献评价原则

一、临床经济学评价的定量和定性分析的关系

(一) 与国家的关系

临床经济学评价与国家的整体规划有关,比如国家目前最需要解决的问题是教育问题,则在国家宏观规划中,即使从临床经济学评价的角度评价某卫生行业的规划及活动是相当合理的,非常值得实施的,但是也需要与教育问题相协调,甚至作出让步。

(二) 与卫生行业的关系

临床经济学评价与卫生行业的整体规划有关,比如卫生行业制订的目标是全球消灭天花,但是鉴于天花的例数已经非常少,大规模监测需要花费很高的费用,这些费用如果投入到其他领域可能获得的产出会更明显,因此从经济学的角度考察似乎不太合理。但是即便如此,仍然要花费大量的费用用于监测天花,遵循传染病防治的全球策略。

(三) 与人群的关系

临床经济学评价与人群以及国家的价值取向有关,如前所述,国家和卫生行业规划重点本身就是与人群以及国家的价值观相联系的。有的价值观是目前已经被所有人群或国家达成共识的,比如充分利用资源,避免浪费,使有限的资源发挥最大的作用;有的价值观在不同的人群或地区还存在分歧,比如如何对待弱势群体,如何对待精英人群,精英人群是否应该给予优先的照顾或保护;比如同一人群不同年龄,即成年人、儿童和老年人的健康寿命年是否相同,如果不同,差异程度有多大;比如现在的一个健康年是否等于将来的一个健康年,如果不同,孰高孰低;所有这些问题都有待讨论,以便使临床经济学评价可以在一定的价值标准下进行。

上述问题的确定有学者称为临床经济学定性评价,可以通过定性的方法,比如专题小组访谈,个人深入访谈,特尔非法等予以确定。

二、临床经济学评价的分析角度

卫生经济学评价可以从多种角度进行,比如从医院的角度,从政府的角度,从社会的角度,从个人的角度。从什么角度分析对理解一项研究的结果非常重要,因此在进行分析评价之前,应该首先确定评价分析的角度。



三、临床经济学评价受时间的影响

有时临床经济学评价中涉及的方案可能会持续几年甚至几十年,在不同的时间发生投入和产出。不同时间发生的投入和产出所具有的经济意义是不一样的。因此在进行投入和产出的比较时,应该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,也就是消除时间对投入和产出的影响,从而便于进行比较。

四、敏感性分析

敏感性分析是临床经济学评价中一个不可缺少的部分。由于测量和计算过程中存在着一定程度的不确定性,通过敏感性分析改变假设条件或改变在一定范围内的估计值,观察其结果或结论的稳定性,从而评估重要参数对评价结果的影响程度,尤其可以确定能够影响分析结论的因素,从而便于对分析结果进行修正,并且在今后的研究中重点考虑这些因素。

五、临床经济学文献评价的内容

临床经济学文献评价的内容包括:①明确研究问题,包括明确研究目的、报告对象和目标人群,应明确评价的用途和药物的适用人群,纳入标准及排除标准。②选择合适的研究角度。研究角度包括社会、卫生保健系统、第三方付费者(医疗保险)、卫生服务提供者和病人等多个方面。不同的研究角度纳入的成本不同,研究者应根据研究目标确定其研究角度,推荐采用社会角度,评价报告中必须清楚阐明研究角度。③确定对照,理想情况下新药应该与目前最适用于成本效果分析的方法进行比较,一般是与常规治疗或最低成本治疗方法比较。④选择研究设计与分析方法。⑤应根据研究角度确定评价的成本范围,成本可分为直接成本和间接成本。两者又分别包括卫生服务内和卫生服务外的成本。卫生服务系统内的直接成本是指直接与治疗干预有关的固定及可变成本。如预防、诊断和治疗成本等;卫生服务系统外的直接成本是指与治疗干预有关的非医疗成本,如病人的交通费、营养费等;卫生服务系统内的间接成本是指由于治疗干预而节约或增加的其他医疗成本,包括健康生命年延长时期内的与干预有明确直接关系的医疗成本;卫生服务系统外的间接成本主要是生产力损失的成本,也包括其他成本(如教育)。从社会角度出发,至少应包括直接成本和卫生系统外的间接成本。有条件时可做有无此种成本的敏感度分析。⑥结果测量,临床效果首选临床终点效果指标如发病率、死亡率和健康相关生命质量如疾病通用量表、疾病专用量表来估计。效益测量方法有人力资本法、陈述偏好法(包括意愿支付法和意愿接受法)和显示偏好法来估计疾病治疗的各种效益。效用多采用质量调整寿命年、伤残调整寿命年等指标进行测量。⑦进行贴现,如果研究的时间超过1年,就应该对成本进行贴现(discount),推荐采用3%的贴现率和做贴现率为0~10%变化时的敏感度分析。⑧计算平均和增量成本效果比。⑨敏感度分析,计算成本和效果的各种参数、贴现率都可进行敏感度分析。在参数较少时采用单纯法和极端值分析法,在参数较多和模型设计时采用 Monte Carlo 模拟进行概率敏感度分析。⑩评价结论,根据投入和产出分析的结果及其判别原则,确定待评价的方案是否可行,或者从多个备选方案中选择一个最佳方案。

知识点

临床经济学评价的定量和定性分析的选择:取决于分析的角度,并受时间的影响。

临床经济学文献评价的十项内容:①明确研究问题;②选择合适的研究角度;③确定对照;④选择研究设计与分析方法;⑤应根据研究角度确定评价的成本范围;⑥结果测量;⑦进行贴现;⑧计算平均和增量成本效果比;⑨敏感度;⑩评价结论。



思考题

1. 直接医疗成本不包括
 - A. 药费
 - B. 护工费
 - C. 挂号费
 - D. 就医交通费
 - E. 实验室检查费
2. 临床经济学评价的角度是
 - A. 医院
 - B. 患者
 - C. 社会
 - D. 医疗保险
 - E. 以上都是
3. 成本效果分析中的效果指标不包括
 - A. 误差率
 - B. 有效率
 - C. 痊愈率
 - D. 发病率
 - E. 生存率
4. 某方案可以接受是由于
 - A. 效益成本比 >0
 - B. 效益成本比 >1
 - C. 效益成本比 <1
 - D. 效益成本比 <0
 - E. 效益成本比 $=1$
5. 从发病到死亡所损失的全部健康年是
 - A. GDP
 - B. YEAR
 - C. DALY
 - D. ONLY
 - E. QALY

参考答案

1. D 2. E 3. A 4. B 5. C

(黄悦勤)

参考文献

1. 黄悦勤. 临床流行病学. 第3版. 北京: 人民卫生出版社, 2010.
2. 萨克斯. 贫穷的终结: 我们时代的经济可能. 邹光, 译. 上海: 上海人民出版社, 2007.
3. Eisenberg, JM. Clinical economics: A guide to the economic analysis of clinical practice. JAMA, 1989, 262: 2879-2886.
4. 胡善联. 药物经济学. 北京: 高等教育出版社, 1995.

第七章 系统综述与 Meta 分析

临床问题

患儿 10 天前无明显诱因出现双小腿疼痛,压痛明显,不能站立,伴发热,最高达 39.0℃,无畏寒寒战,血常规 WBC $2.47 \times 10^9/L$,ANC $0.4 \times 10^9/L$,Hb $93.6 \times 10^9/L$,PLT $95 \times 10^9/L$,予头孢孟多抗感染 1 天后,热稍退,疼痛未见明显改善,遂住院检查:双踝关节软组织肿胀并关节积液;1/4 日行骨穿,骨髓检查示 ALL-L1。诊断为急性淋巴细胞白血病。家长对白血病相关知识不了解,表示对疾病治疗的忧虑和担心。

从这个临床情景中可以看出,这是一个有关治疗的问题,医疗卫生干预的效果评价主要采用随机对照试验(RCT)的研究设计。在面临此问题时,我们首先可能做的是检索、阅读和评价有关的高质量的临床试验,从中找到有关问题的证据,据此作出合理的决策。但是,我们经常会发现就某一具体临床问题可能存在几篇甚至几十篇文献。可能需要花费大量的时间来筛选和评价这些文献,这在临床实践中往往难度很大。因此,为了解决临床医生难以花费大量时间检索和阅读原始研究的困难,系统地总结原始研究证据的 Meta 分析(meta analysis)应运而生,这一类研究一般被称作系统综述(systematic review)报告形式之一。高质量的系统综述是临床医生、研究者、决策者和患者的最佳证据。1989 年英国产科医生 Iain Chalmers 对短疗程、低价格类固醇药物治疗有早产倾向孕妇的随机对照试验(RCT)进行了总结归纳。这一系统综述结果表明低价格类固醇药物可以显著降低婴儿死于早产并发症的风险,这一药物的推广使得欧洲新生儿死亡率降低了 30%~50%。1994 年美国国立卫生研究院(NIH)制定官方政策,鼓励使用此法,美国每个婴儿可节省 3000 多美元,全国每年可节省 15.7 亿美元。由此可见,系统综述 /Meta 分析在指导临床实践和卫生决策中的重要指导意义。同时其科学、快速处理海量信息的方法在促进知识转化、缩短研究与实践的差距中亦发挥着重要作用。

第一节 基本概念

系统综述(systematic review,SR)、Meta 分析(meta analysis)是循证医学中经常提到的术语,与循证医学的发展密切相关。1992 年国际上正式提出循证医学的概念。同年在英国牛津正式建立了英国 Cochrane 中心,次年建立了世界 Cochrane 协作网,为循证医学发展提供系统评价证据。但是系统综述不等于循证医学,它是循证医学中经常使用的一种重要证据。证据是循证医学的基础,质量是证据的关键。实践循证医学首先要学会如何从大量的临床文献中找出当前最佳的证据并加以应用,有条件者还应进行原始临床研究为循证医学提供高质量证据。而系统综述和 Meta 分析正是循证医学重要的研究方法和最佳证据的重要来源之一,是当前临床医学各专业使用最频繁的工具之一。

一、系统综述

系统综述是一种全新的文献综合方法,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



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得出综合结论的方法。同时,随着新的临床研究结果的出现进行及时更新。系统综述可以是定性系统综述(qualitative systematic review),也可以是定量系统综述(quantitative systematic review)。系统综述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。高质量的系统综述是目前级别最高的证据之一,被临床指南广泛引用。系统综述方法发展很快,其主题已由最初的干预措施疗效领域拓展到病因、诊断、预后、不良反应和动物实验等多个方面。

系统综述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、系统综述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,读者在阅读系统综述的观点和结论时,一定要持谨慎的态度,不能盲目被动地接受(见本章第二节)。

二、Cochrane 系统综述

Cochrane 系统综述是 Cochrane 协作网的评价人员按照统一的工作手册(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综述。Cochrane 协作网定义 Cochrane 系统综述:全面收集符合纳入与排除标准的经验性证据来回答某个研究问题,用清楚、明确的方法减少偏倚,提供可靠的研究结果以便得出结论,做出决定。

Cochrane 系统综述是一个设计周密科学、运作高效协同的系统工程,包括题目、研究方案、全文过程都实行注册,从入口把关设计质量,实行过程监督把关过程质量,指导规范化发表把关出口质量。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 reviewers' handbook,采用固定的格式和内容要求,统一的系统综述软件(RevMan)录入和分析数据、撰写系统综述计划书和报告,发表后根据新的临床研究的出现定期更新,有着完善的反馈和完善机制,因此,Cochrane 系统综述的质量通常比非 Cochrane 系统综述质量更高,成为卫生干预措施效果最有价值的信息来源。Cochrane 系统综述是循证决策与实践的重要证据,是循证医学研究与实践的重要纽带。Cochrane 系统综述有标准格式,便于电子出版和随后的更新,并且有专用的软件 RevMan 用于制作系统综述。Cochrane 系统综述制作流程如下:①确定题目与注册:首先根据系统综述制定原则,确定题目,并与 Cochrane 相关评价小组联系,表达研究兴趣。Cochrane 系统评价小组经过 2~3 周的评审,若同意注册,将会通过 E-mail 发送题目注册表,否则提出建议修改题目,或告知另换题目或参加系统评价相关知识和技能培训。②撰写计划书:计划书(protocol)又叫研究方案,Cochrane 系统评价小组将会在计划书形成期间,免费提供相关培训资源、软件、制作流程和模板等信息,帮助作者完善计划书;同时提供包括检索策略、方法学实施计划和内容等方面建议。计划书需经过 Cochrane 系统评价小组内外 3~5 位评审专家审评及多次修改完善,同意后方能发表在 Cochrane 图书馆(Cochrane library)。③制作系统综述:系统综述(systematic review)强调计划在先,制作时应根据已发表的计划书所描述的方法撰写,经过小组同行评审,作者多次修改,合格后才在 Cochrane 图书馆全文发表。一篇系统综述从计划书到全文草稿提交的间隔时间原则上不超过 12 个月,到全文完成和发表最多 2 年时间。若计划书超过 2 年还未完成则可能被要求退出。不同的小组也会根据制作进展、文献查找等具体情况作相应调整(注意:完成的系统评价若想在国际英文杂志上发表,必须事先经过相关小组的同意。为确保权威性和质量,维护其版权,Cochrane 协作网规定,未经相关小组授权,任何作者不得擅自将已注册的系统综述以任何形式提前或重复发表)。④更新系统综述:发表后的系统综述需要随时接受反馈意见和新发表的原始研究,并根据反馈意见和新的原始研究不断更新(updating)。Cochrane 协作网要求其作者每 2 年更新一次系统综述。主要是对系统综述中相关方法学的内容和检索部分进行更新,实际中经常是对新查找到的研究和检索时间进行更新。如果可能,Cochrane 小组将根据已制定的检索策略,在需要更新之前 3~6 个月检索相关数据库,提供可能相关的研究,帮助作



者更新其系统综述。

三、Meta 分析

Meta 分析(meta analysis)由 Beecher 于 1955 年最先提出。英国心理学家 G.V.Glass 1976 年首先将医学文献中对多个同类研究统计量的合并方法称为“meta analysis”,并且 Glass 最早在教育学研究中使用了 Meta 分析,现已广泛应用于医学健康领域,针对关于诊断、治疗、预防和病因方面的问题而进行综合评价。从 20 世纪 80 年代中期开始被引入到临床随机对照试验以及观察性的流行病学研究中。20 世纪 80 年代末,该方法被引入我国,中文译名有荟萃分析、二次分析、汇总分析、集成分析等。Meta 分析的定义目前尚有不同意见,Cochrane 图书馆将 Meta 分析定义为:Meta analysis is a statistical technique for assembling the results of several studies in a review into a single numerical estimate,即 Meta 分析是将系统综述中的多个研究结果合并为单个量化指标的一种统计学方法。而 David Sackett 等人在 *Evidence Based Medicine* 一书中,将 Meta 分析定义为:A systematic review that uses quantitative methods to summarize the results,即运用定量方法汇总多个研究结果的一种系统综述方法。“系统综述”常与“Meta 分析”交叉使用,意义相同。现多认为 Meta 分析是系统综述的一种类型,但系统综述不一定是 Meta 分析,Meta 分析作为一种统计分析方法也应用于其他研究。

医学研究中,传统的文献综述的文献收集和复习不够全面与严格,在处理同一问题的多个研究结果时,一般不进行文献评价,也不考虑文献的质量,通常只是汇总同类研究中某类结论的多少,即平等的(等权重方法)对待多个同类研究结果而得出结论,这种等权重的文献综述方法至少存在两个问题:一是将质量不相同的多个研究人为地判为相同;二是将样本含量大小(权重)不相等的多个研究平等地对待。因此,这种文献综述的方法很难保证研究结果的真实性和可靠性,尤其当多个正反结果的研究数量相当时,很容易让人产生困惑或误解。Meta 分析是对多个同类研究结果进行合并汇总的分析方法,能从统计学角度达到增大样本含量,提高检验效能的目的。尤其当多个研究结果不一致或都没有统计学意义时,采用 Meta 分析可得到更加接近真实情况的综合分析结果。当系统综述的数据资料适合使用 Meta 分析时,用 Meta 分析可以克服传统文献综述的上述两大问题,提高分析结果的可靠性;当数据资料不适合做 Meta 分析时,系统综述只能解决文献评价的问题,不能解决样本含量的问题,因此,对其分析结论应慎重。应特别注意:不按系统综述标准操作规范实施,或未经严格文献评价的研究,即使用 Meta 分析也不一定是系统综述的研究,更难说是高质量研究。

四、系统综述的报告质量规范

系统综述和 Meta 分析在医疗保健领域已变得越来越重要。临床医生通过阅读它们来更新自己在该学科中所掌握的信息,同时它们还常常成为制定临床实践指南的证据基础。与所有的研究一样,系统综述价值的大小取决于已经存在的研究、研究的发现以及报告的明晰与否。与其他形式发表的文献一样,系统综述的报告质量千差万别,这会限制读者对综述文章优缺点的评估。先前已经有过一些文章评价综述报告的质量。1987 年,Mulrow 对 4 本当时处于领先水平的医学期刊在 1985 年与 1986 年发表的 50 篇综述文章进行了评估,发现没有一篇符合所有 8 项明确的科学规范,例如对纳入研究的质量进行评估。同年,Sacks 及其同事对 83 个 Meta 分析进行了报告充分程度的评价,评价涉及 6 个领域 23 个特征,总体来说报告质量较低。

1996 年,为了使 Meta 分析报告中的这些不尽如人意的情况得到改善,由临床流行病学家、临床学家、统计学家、从事 Meta 分析的研究人员以及来自英国和北美的对 Meta 分析感兴趣的编辑共 30 人组成专家小组,对 RCT 的 Meta 分析报告质量进行了方法学的评价,即 QUOROM 声明(Quality of Reporting of Meta Analyses, QUOROM),并提出了统一的报告格式。



2009 年为了进一步提高系统综述和 Meta 分析文章报告的质量,由国际著名专家组成的系统综述和 Meta 分析优先报告的条目(Preferred Reporting Items for Systematic Reviews and Meta Analysis, PRISMA)小组在国际重要医学期刊包括《英国医学杂志》、《临床流行病学杂志》、《内科学年鉴》和美国《公共科学图书馆医学杂志》等同步发表了《系统综述与 Meta 分析优先报告条目:PRISMA 声明》。该标准的制定对于改进和提高系统综述和 Meta 分析报告质量起到重要作用。该声明较以往制定的《随机对照试验 Meta 分析报告质量》(Quality of Reporting of Meta Analyses, QUOROM),即《QUOROM 声明》更加全面、完善。由于近年来国内外对系统综述的高度重视,发表的文章数量也越来越多,因此有必要规范报告的标准,提高报告的质量。这份声明对系统综述和 Meta 分析类文章报告的 27 个条目及流程图进行了定义和介绍,对相关条目进行了详细的解释和说明(表 2-7-1,图 2-7-1)。

表 2-7-1 PRISMA 2009 检查表

项目	编号	条目清单
标题		
标题	1	明确本研究报告是系统综述、Meta 分析,还是两者兼有
摘要		
结构式摘要	2	提供结构式摘要,包括背景、目的、资料来源、纳入研究的标准、研究对象和干预措施、研究质量评价,数据综合的方法、结果、局限性、结论和主要发现,系统综述的注册号
前言		
理论基础	3	介绍系统综述研究的理由和依据
目的	4	通过对研究对象、干预措施、对照措施、结局指标和研究类型(participants、interventions、comparisons、outcomes、study design, PICOS)5 个方面为导向的问题提出所需要解决的清晰明确的研究问题
方法		
方案和注册	5	如果已有研究方案,则说明方案内容并给出可获得该方案的途径(如网址),并且提供现有的已注册的研究信息,包括注册号
纳入标准	6	将指定的研究特征(如 PICOS 和随访的期限)和报告的特征(如检索年限、语种和发表情况)作为纳入研究的标准,并给出合理的说明
信息来源	7	针对每次检索及最终检索的结果,描述所有文献信息的来源(如资料库文献,与研究作者联系获取相应的文献)
检索	8	至少说明一个资料库的检索方法,包含所有的检索策略的使用,使得检索结果可以重现
研究选择	9	说明纳入研究被选择的过程(包括初筛、合格性鉴定及纳入系统综述等步骤,还可包括纳入 Meta 分析的过程)
资料提取	10	描述资料提取的方法(例如预提取表格、独立提取、重复提取)以及任何向报告作者获取或确认资料的过程
资料条目	11	列出并说明所有资料相关的条目(如 PICOS 和资金来源),以及作出的任何推断和简化形式
单个研究存在的偏倚	12	描述用于评价单个研究偏倚的方法(包括该方法是否用于研究层面或结局层面),以及在资料综合中该信息如何被利用
概括效应指标	13	说明主要的综合结局指标,如危险度比值(risk ratio)、均值差(difference in mean)
结果综合	14	描述结果综合的方法,如果进行了 Meta 分析,则说明异质性检验的方法

学
习
笔
记



续表

项目	编号	条目清单
研究偏倚	15	详细评估可能影响数据综合结果的可能存在的偏倚(如发表偏倚和研究中的选择性报告偏倚)
其他分析	16	对研究中其他的分析方法进行描述(如敏感性分析或亚组分析、Meta 回归分析),并说明哪些分析是预先制定的
结果		
研究选择	17	报告初筛的文献数,评价符合纳入标准的文献数以及最终纳入研究的文献数,同时给出每一步排除文献的原因,最好提供流程图
研究特征	18	说明每一个被提取资料的文献的特征(如样本含量、PICOS 和随访时间)并提供引文出处
研究内部偏倚风险	19	说明每个研究中可能存在偏倚的相关数据,如果条件允许,还需要说明结局层面的评估(见条目 12)
单个研究的结果	20	针对所有结局指标(有效性或有害性),说明每个研究:(a)各干预组结果的简单合并数据,以及(b)综合效应值及其可信区间,最好以森林图形式报告
结果的综合	21	说明每个 Meta 分析的结果,包括可信区间和异质性检验的结果
研究间偏倚	22	说明研究间可能存在偏倚的评价结果(见条目 15)
其他分析	23	如果有,给出其他分析的结果(如敏感性分析或亚组分析、Meta- 回归分析,见条目 16)
讨论		
证据总结	24	总结研究的主要发现,包括每一个主要结局的证据强度,分析它们与主要利益集团的关联性(如医疗保健的提供者、使用者及政策决策者)
局限性	25	探讨研究层面和结局层面的局限性(如偏倚的风险)以及系统综述的局限性(如检索不全面,报告偏倚等)
结论	26	给出对结果的概要性的解析,并提出对未来研究的提示
资金支持		
资金	27	描述本系统综述的资金来源和其他支持(如提供资料)以及资助者在完成系统综述中所起的作用

来源:Moher D, Liberati A, The PRISMA Group (2009). Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. PLoS Med 6(6):e1000097. doi:10.1371/journal.pmed1000097 For more information, visit: www.prisma-statement.org.

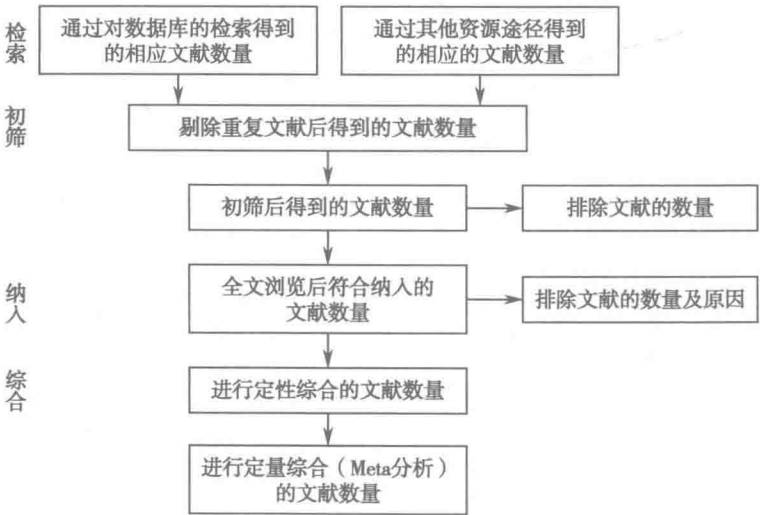


图 2-7-1 PRISMA 声明文献筛选四阶段流程图



五、系统综述的报告质量规范

除了随机对照试验,观察性研究也是临床流行病学研究的重要组成部分,包括多种研究设计,如队列研究、病例对照研究和横断面调查等,主要用于描述疾病的分布、探索和检验暴露与疾病之间的因果关联。但由于观察性研究很难证明排除了一些偏倚,也不可能完全去除混杂效应,如果观察性研究过程中确实存在相同的系统误差,Meta 分析只会加大这些偏倚,产生统计学上的假象。换言之,如果原材料就有缺陷,那么系统综述结论的真实性也会受到损害。因此,观察性研究的 Meta 分析结果应采取科学的态度进行解释,重点应放在检查研究结果异质性的可能来源上。1997 年 4 月,由美国疾病预防控制中心资助,召集临床实践、现场干预、统计学、流行病学、社会科学以及生物医学编辑等方面的 27 名专家组成专题研究小组,讨论并制定了流行病学中观察性研究的 Meta 分析(Meta-analysis of observational studies in epidemiology, MOOSE)的报告规范(表 2-7-2)。

表 2-7-2 MOOSE 报告规范详情表

报告要求
研究背景
定义研究问题
陈述研究问题假设
确定研究结局
暴露 / 干预措施
研究设计类型
研究人群
文献检索策略
文献检索的资格(如图书管理员和调查员)
文献检索策略,包括文献检索的时间范围和使用的关键词
尽可能获取所有文献,包括研究文献作者的个人通信
检索的数据库和档案库
采用的检索软件及其版本号,包括使用的特殊功能(如进行主题词及其下位词的扩展检索)
手工检索(如已有文献的参考文献清单)
列出纳入和排除的文献,以及判断标准
处理非英语文献的方法
处理只有摘要和未发表文献的方法
介绍个人通信的情况
研究方法
描述检索文献是否符合研究问题
数据整理和编码的基本原则(如有完善的临床编码规则或便于编码)
数据分类和编码的记录(如多个文献评价者,盲法,以及文献评价者之间的一致性)
混杂的评估(如入选研究中病例和对照的可比性)
评价研究质量,包括对质量评价者采用盲法,对研究结果的可能预测值进行分层分析或者回归分析
评价研究异质性
详细介绍统计分析模型,以便能重复该研究(如详细描述采用的固定效应模型或者随机效应模型,采用该研究模型分析研究结果的理由,剂量反应关系模型,或者累积 Meta 分析)
提供合适的统计图、表

学
习
笔
记

**研究结果**

绘图总结入选各研究和汇总研究结果

列表描述入选各研究结果

研究结果的敏感度分析(如亚组分析)

研究结果统计学稳健性的指标

讨论

定量地评价偏倚(如发表偏倚)

解释排除标准的合理性(如排除非英语文献)

评价入选研究的质量

研究结论

导致观察到结果的其他可能原因

概括性的结论(如根据研究所得的数据,在评价文献涉及的领域,对研究结论进行适当地外推)

为以后该问题的研究提供指导意见

公布研究资助来源

知识点

Cochrane 系统综述的特征:Cochrane 系统综述有 6 个特征:①研究目的和纳入与排除标准明确;②方法学清楚并可重复;③检索策略系统,能全面收集符合纳入与排除标准的研究;④评价纳入研究的真实性,如评价偏倚风险;⑤纳入研究特征与结果的表达及合成系统;⑥定期更新。

Cochrane 系统综述的制作步骤:Cochrane 系统综述制作步骤:①确立题目、撰写系统综述计划书(formulating question and developing protocol);②检索文献(locating studies);③筛选文献(selecting studies);④对纳入的文献质量进行评价(assessment of study quality);⑤资料提取(extracting data);⑥资料分析并形成结果(analyzing and presenting results);⑦结果的解释(interpreting results);⑧系统综述的更新与改进。

Meta 分析的统计学过程:Meta 分析的基本内容包括:①合并统计量的选择;②异质性检验;③合并统计量的检验;④森林图/漏斗图;⑤敏感性分析与亚组分析。

Meta 分析的分类:①根据纳入的原始研究设计类型不同(study design),可分为临床对照试验(controlled trial)和观察性研究(observational study)两种类型的 Meta 分析。前者如随机对照试验和非随机对照试验,后者如队列研究和病例对照研究的 Meta 分析;②根据纳入的原始研究数据类型(data type)不同,可分为二分类资料(dichotomous)的 Meta 分析、连续性资料(continuous)的 Meta 分析和生存资料的 Meta 分析。

PRISMA 声明:《PRISMA 声明》由 27 个条目的清单(见表 2-7-1)以及一个四阶段的流程图(图 2-7-1)组成。PRISMA 声明的目的在于帮助作者改进系统综述和 Meta 分析的撰写和报告。主要针对的是随机对照试验的系统综述。PRISMA 也可以用于已发表系统综述的严格评价。PRISMA 并非测量系统综述质量的工具。像其他循证医学出版物一样,PRISMA 也将是随时更新的。

第二节 评价原则

制作系统综述/Meta 分析的最终目的是为了使用,但因制作者的水平良莠不齐,故质量也参



差不齐。因此,在使用前一般需要对系统综述 /Meta 分析的质量进行评价。如同原始研究一样,学者们研制了一些用于系统综述 /Meta 分析的质量评价工具。

OQAQ 量表 (Oxman-Guyatt Overview Quality Assessment Questionnaire) 是用于评价系统综述真实性的最常用工具。OQAQ 量表共有 9 个方面 10 个条目,前 9 个条目可以评为充分 (报告并正确使用) 和不充分 (没有报告或不正确),第 10 个条目是对整个文献质量进行打分。评价者根据前面 9 个问题的情况给 1~7 分,1 分代表该系统综述不符合标准,7 分则代表该系统综述完全符合。若前 9 项有 1 个以上条目回答为“不能确定”,则表明该系统综述存在“小的缺陷”;若条目 2、4、6、8 中有回答为“否”者,则说明存在“大的缺陷”。一般有明显缺陷者,其综合评分为 1 分;大的缺陷评为 3 分;有小的缺陷者评为 5 分;缺陷可被忽略者为最高分 7 分 (表 2-7-3)。

表 2-7-3 OQAQ 量表

条目	描述	条目	描述
1	是否报告了文献检索方法 Were the search methods reported	6	对纳入研究的质量评价是否全面、恰当 Was validity assessed appropriately
2	检索是否全面 Was the search comprehensive	7	是否报告了纳入研究的数据合并方法 Were the methods used to combine study reported
3	是否报告了研究的纳入标准 Were the inclusion criteria reported	8	纳入研究的结局是否适合合并 Were the findings combined appropriately
4	是否避免了纳入研究的选择偏倚 Was selection bias avoided	9	系统综述的结论是否得到了报告数据的支持 Were the conclusions supported by the reported data
5	是否报告了对纳入研究真实性评价的标准 Were the validity criteria reported	10	此系统综述的总体科学性如何 What was the overall scientific quality of the overview

学
习
笔
记

针对本章病例中患儿家属所关心的问题,据你了解,粒细胞集落刺激因子 (G-CSF)、粒细胞巨噬细胞集落刺激因子 (GM-CSF) 等造血生长因子能调节造血细胞的增殖和分化,已用于临床治疗或预防化疗引起的中性粒细胞减少、发热和感染。为了回答“粒细胞集落刺激因子治疗急性淋巴细胞白血病是否安全和有效”这一临床问题,你在 Cochrane Library 及 MEDLINE 中进行检索,检索策略为 (G-CSF OR “granulocyte colony-stimulating factor”) AND (ALL OR “Acute Lymphoblastic Leukemia”),未发现相关系统综述文献。进而你在中国知网 (CNKI) 中进行检索,检索策略为“粒细胞集落刺激因子”并含“急性淋巴细胞白血病”,发现一篇能回答该问题的系统综述文章:李培、席亚明、徐建旺,等. 粒细胞集落刺激因子治疗急性淋巴细胞白血病疗效和安全性的系统评价. 该系统综述评价了 2000~2009 年发表的关于粒细胞集落刺激因子治疗急性淋巴细胞白血病疗效和安全性的随机对照试验。该文章有可能回答了你的问题,因此我们按照 OQAQ 量表评价以下这篇文献的质量。

(一) 是否报告了文献检索方法以及检索是否全面

系统综述的制作,一旦确定好研究问题,在文献收集之前,要撰写计划书。系统综述本质上属于回顾性研究。如果作者熟悉这些研究结果,那很有可能会改变其确定问题的方法,包括设定选择标准,纳入何种干预和比较措施,选择何种观察结局等。因此计划书的撰写可以避免系统综述制作者仅根据检索不全的原始文献结果报告进行分析而导致的结论偏倚。系统综述的文章必须清楚的定义研究问题,应按照 PICOS 原则构建临床问题:P 表示 patient or population (患



者或人群),I 表示 intervention(干预措施),C 表示 comparison(对照措施),O 表示 outcome(结果,即干预措施的影响),S 表示 study(研究类型)。

系统综述的检索策略(search strategy)制定应尽可能系统和全面,必须明确研究相关的检索来源名称、检索时间(检索起止日期)和文献语种。为全面查找所有相关临床研究,凡是可能收录了与研究问题相关的原始研究数据库均应考虑在内,不限定语种和时间。系统综述检索来源主要包括:①综合性文献数据库资源(如 MEDLINE、EMBASE)、Cochrane 对照试验中心数据库(Cochrane Central Register of Controlled Trials,CENTRAL)。②专业数据库,如专业小组资料库、中医药库等。③查找其他相关资源,包括在研临床试验库;人工检索相关杂志、灰色文献(如内部报告、会议论文)和已发表研究参考文献;检索美国科学引文索引数据库(science citation index,SCI)或与研究通讯作者联系等。同时,在计划书中应注明是否进行人工检索。

本例中,通过阅读该篇系统综述的题目、研究目的及摘要,发现这篇系统综述的临床问题具体、明确,与我们的问题非常相关。其检索策略为:计算机检索 Cochrane Library、PubMed、EMBASE、中国期刊全文数据库和维普中文科技期刊数据库。中英文检索词:g-csf、Granulocyte Colony-Stimulating Factor、gm-csf、Granulocyte-Macrophage Colony-Stimulating Factor、acute lymphoblastic leukemia、细胞集落刺激因子、急性淋巴细胞白血病等。检索时间限定在 2000 年 1 月至 2009 年 1 月。

通过阅读该系统综述可以看出,作者对原始文献的检索是较为全面的,对常用的中外文数据库都进行了检索。对于检索的结果,作者也做了明确的描述:初检出 853 篇文献,通过阅读题目及摘要初筛纳入 39 个研究,进一步阅读全文后最终纳入 6 个研究。该系统综述没有绘制“漏斗图”,可能是由于纳入研究数量较少所致。

可以看出,该系统综述对于原始研究的检索是较为全面的。

(二) 原始研究的纳入标准是否合适以及是否避免了纳入研究的选择偏倚

为了确定一个系统综述是否包含了适合的原始研究,读者需要知道系统综述作者挑选研究的标准。如果一个系统综述没有明确的纳入标准/排除标准,或者预先制定的标准不明确,或者纳入的研究和预先制定的标准不一致,就可能存在研究者根据研究结果随意纳入研究的现象,从而产生偏倚。因此为减少选择偏倚,应尽量制定明确的、严格统一的文献纳入和排除标准;在选择文献时,应该至少由两个人采用盲法独立进行。

1. 研究设计类型 系统评价的研究设计类型选择原则就是根据研究目的,选择能回答所关注问题的研究设计类型,优选当前方法学质量较高的研究设计类型。如涉及治疗的问题,随机对照试验(RCT)是最佳研究设计。而病因或危险因素的问题则选择队列研究(cohort study)或病例对照研究(case control study)。

2. 研究对象的类型 选择研究对象要使用明确的标准来界定疾病或所关注状况,疾病或健康问题的种类或亚型,疾病的诊断标准;研究对象的重要特征,如年龄、性别、种族、疾病类型及病程、研究实施地、诊断主体以及需要排除的研究对象等。

3. 干预措施与对照措施 具体说明所关注的干预措施以及作为对照的处理措施。干预措施内容是:相关措施使用的具体强度、频率、实施方式、实施主体以及是否需要实施者进行培训或实施者资质等。此外,还应考虑到干预措施的变化在多大程度上对研究对象和关注的结局指标产生不同的影响。若干预措施是药物,还应考虑药物的制剂、给药途径、剂量、持续时间和频率等因素。若是复杂干预(如教育或行为干预),应界定其共同的核心特点。

4. 结局指标 Cochrane 系统综述应当包括所有对用户有意义的结局指标。主要涉及三方面内容:确立结局指标、确立结局指标测量方式和确立结局测量的时间。



本例中,该系统综述根据研究目的,设定了原始研究的纳入与排除标准。

纳入标准:

- 1. 研究设计 随机对照试验(RCT)。
- 2. 研究对象 经骨髓细胞学、化学染色和免疫分型等确诊为 ALL 的患者,符合 FAB 诊断标准或张之南主编的《血液病诊断及疗效标准》,年龄、性别不限。
- 3. 干预措施 对照组:化疗药物及支持治疗,加用安慰剂或空白对照;试验组:在对照组基础上加用 G-CSF 或 GM-CSF。
- 4. 结局指标 主要结局指标:随访 1 年以上的总体生存率(overall survival, OS)、诱导化疗结束后的完全缓解率(complete remission, CR)及生存质量(quality of life, QOL)等。次要结局指标:感染情况(包括感染率、血培养阳性率、发热人数)、复发率及不良反应等。

排除标准:排除骨髓移植后给予细胞集落刺激因子治疗的研究。

可以看出,该系统综述中纳入了对于疗效评价问题来说最佳的研究类型。研究对象定义明确,有明确的诊断标准,比较组设定合理,结局指标纳入了所有决策相关的重要结局。由此可见,原始研究的纳入标准是合适的。

(三) 是否报告了对纳入研究真实性评价的标准以及对纳入研究的质量评价是否全面、恰当

研究结果的可信度由研究的质量决定,评估研究的质量就是对研究结果的可信度的评价。研究的质量由研究设计和具体研究采取的各种偏倚控制措施所决定。原始研究的质量是系统综述质量的基础。因此,评价原始研究的质量以确定系统综述结果的真实性非常重要。

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,可给予相同分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 手册未推荐使用任何一种清单或量表,而要求采用由 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 2-7-4),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、直观地反映偏倚情况。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。

为避免选择文献和评价文献质量人员的偏倚,可考虑一篇文章由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(表 2-7-4)。

表 2-7-4 Cochrane 协作网的偏倚风险评价工具

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,以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,以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

学
习
笔
记



续表

评价条目	评价内容描述	作者判断
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	盲法是否完整
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(由系统综述者判断)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

本例中,该系统综述对纳入的每项研究均进行了质量评价。按照 Cochrane 协作网的偏倚风险评价工具,对纳入文献的随机方法、分配隐藏、盲法、意向性分析及随访、选择性报告研究结果和其他偏倚来源等进行评价。经过评价显示:纳入的 6 项研究中,对是否随机分配、是否采用盲法、是否隐藏了分配方案三项均为不清楚,仅有 1 篇未提及随访情况,其余 5 篇随访时间均在 1 年以上,纳入研究失访率均 <20%。

综上所述,原始研究质量一般,方法学不足导致的偏倚可能会很大。

(四) 是否报告了纳入研究的数据合并方法,数据合并是否合适

系统综述/Meta 分析要对纳入研究的结果进行合并,最后得到一个综合所有研究的“平均结果”。如果纳入系统综述的每个研究的结果相似或方向一致,则由此合成的结果其可信度较高。但是,一般来说,纳入到系统综述中的研究总是不同的,系统综述/Meta 分析中所纳入原始研究间的差异或变异叫作异质性(heterogeneity),分为方法学异质性、临床异质性和统计学异质性三种,在临床方面(如病人特征)和研究设计方面的差别都可能引起原始研究间的异质性。因此,在进行系统综述过程中,对纳入的研究进行异质性检验是非常重要的。读者应当关注系统综述是否报告了异质性检验的结果。

异质性检验(tests for heterogeneity),又称同质性检验(tests for homogeneity),Meta 分析前需要对多个研究结果进行异质性检验,以判断多个研究是否具有同质性。按统计学原理,只有同质的资料才能进行多个研究统计量的合并,反之则不能。异质性检验的方法,目前多用 χ^2 检验(Chi-square test)。若异质性检验结果为 $P>0.10$ 时,可认为多个同类研究具有同质性,可选择固定效应模型(fixed effect model)计算其合并效应统计量。当多个研究结果经异质性检验后若 $P\leq 0.10$,可认为多个研究结果有异质性。纳入研究的异质性大小可用 I^2 来衡量, I^2 的计算公式如下:

$$I^2 = \frac{Q - (k-1)}{Q} \times 100\%$$

式中的 Q 为异质性检验的 χ^2 值, k 为纳入 Meta 分析的研究个数。在 RevMan 中, I^2 是可用于衡量多个研究结果间异质程度大小的指标,用于描述由各个研究所致,而非抽样误差所引起的变异(异质性)占总变异的百分比。在 Cochrane 系统综述中,只要 I^2 不大于 50%,其异质性可以接受。

当异质性检验出现 $P\leq 0.10$ 时,则显示异质性的存在,这时,寻找异质性的原因将成为该系



统综述的主要分析内容。首先应分析导致异质性的原因,如设计方案、测量方法、用药剂量、用药方法、疗程长短、病情轻重、对照选择等因素是否相同,必要时应与原文作者取得联系。由这些原因引起的异质性可用亚组分析(subgroup analysis)Meta 回归进行分析和处理。若经这些方法分析和处理后,多个同类研究的结果仍然不能解决异质性时,而且认为合并的结果仍具有一定的实践意义,可选择随机效应模型(random effect model)合并研究结果。需特别注意的是,随机效应模型是针对异质性资料的统计处理方法,不能代替导致异质性的原因分析。如果异质性过于明显,而且研究在临床和方法学方面存在明显的不同,可彻底放弃对结果的合并,只对不同研究的结果做分别的报告和讨论。

在系统综述中,研究人员从个体研究中收集数据,利用定量的方法(Meta 分析)对这些数据进行整合。Meta 分析需要将多个同类研究的结果合并(或汇总)成某个单一效应量(effect size)或效应尺度(effect magnitude),即用某个合并统计量反映多个同类研究的综合效应。系统综述中常见的资料类型有三类:一类是二分类资料,即每一个体只能处于两种状态中的一种,如阴性与阳性,有或者无等。这样的资料可选择比值比(*OR*)、相对危险度(*RR*)或危险差(*RD*)为合并统计量,用于描述多个研究的合并结果。*RR* 或 *OR* 均是相对测量指标,其结果解释与单个研究指标相同,而 *RD* 是两个率的绝对差值。第二类资料为连续性变量资料,如身高、体重、血压、血糖水平等,可选择均数差(mean difference, *MD*)或加权均数差(weighted mean difference, *WMD*)或标准化均数差(standardized mean difference, *SMD*)为合并统计量。*WMD* 即为两均数的差值,消除了多个研究间的绝对值大小的影响,以原有的单位真实地反映了试验效应;*SMD* 可简单地理解为两均数的差值再除以合并标准差的商,它不仅消除了多个研究间的绝对值大小的影响,还消除了多个研究测量单位不同的影响,尤其适用于单位不同或均数相差较大资料汇总分析,但标准化均数差是一个没有单位的值,因而对 *SMD* 分析的结果解释要慎重。第三类资料为生存时间,常常用于发生频率比较高的事件如癌症的治疗研究,处理这类资料时通常会用风险比(hazard ratio, *HR*)来测量治疗的效果。

无论采用何种方法计算得到的合并统计量,都需要进行统计推断——假设检验(hypothesis test)的方法检验多个同类研究的合并统计量是否具有统计学意义,常用 $z(u)$ 检验,根据 $z(u)$ 值得到该统计量的概率(P)值。若 $P \leq 0.05$,多个研究的合并统计量有统计学意义;若 $P > 0.05$,多个研究的合并统计量没有统计学意义。

可信区间是按一定概率估计总体参数(总体均数、总体率)所在范围(区间),如 95% 的 *CI*,是指总体参数在该范围(区间)的可能性为 95%。可信区间有估计总体参数和假设检验两个用途。若要利用样本资料得到总体值(参数)时,常用可信区间进行估计,如均数的可信区间、率和 *OR* 值的可信区间等。可信区间的范围愈窄,用样本指标估计总体参数的可靠性就愈好;反之其可靠性就愈差。可信区间可用于假设检验,95% 的可信区间与 $\alpha=0.05$ 的假设检验等价,99% 的可信区间与 $\alpha=0.01$ 的假设检验等价。

当试验效应指标为 *OR* 或 *RR* 时,其值等于 1 时试验效应无效,此时其 95% 的可信区间若包含了 1,等价于 $P > 0.05$,即无统计学意义;若其上、下限不包含 1(均大于 1 或均小于 1),等价于 $P < 0.05$,即有统计学意义。当试验效应指标为 *RD*、*WMD* 或 *SMD* 时,其值等于 0 时试验效应无效,此时其 95% 的可信区间若包含了 0,等价于 $P > 0.05$,即无统计学意义;若其上、下限不包含 0(均大于 0 或均小于 0),等价于 $P < 0.05$,即有统计学意义。

Meta 分析的结果通常以森林图(forrest plot)的形式表达(图 2-7-2)。森林图的平面直角坐标系以一条垂直的无效应线(横坐标刻度为 1 或 0,即无统计学意义的值)为中心,*RR* 和 *OR* 无效竖线的横轴尺度为 1,而 *RD*、*MD* 和 *SMD* 无效竖线的横轴尺度为 0。用平行于横轴的多条线段表达每个被纳入研究的效应量及其可信区间(confidence interval, *CI*),可信区间中间的正方形图形(或其他图形)的大小代表研究的权重,正方形的中心是该研究效果的点估计值。综合效应一



般用菱形表示,以区别具体研究的结果,菱形的中心点是综合结果的点估计值,其宽度表示可信区间。若某个研究 95% 可信区间的横线与无效应线相交或相切,即该研究无统计学意义,反之,若该横线落在无效应线的左侧或右侧不与无效应线相交,该研究有统计学意义,详见下述实例分析。

为了解二甲双胍对多囊卵巢综合征的治疗作用,研究者收集了以患者排卵数为评价指标的 7 个随机对照试验的结果(表 2-7-5)。

表 2-7-5 二甲双胍对多囊卵巢综合征的治疗性研究

<i>k</i> 个研究	二甲双胍组		对照组		OR	OR 的 95%CI	
	排卵数 (<i>n</i>)	治疗 总数(<i>N</i>)	排卵数 (<i>n</i>)	治疗 总数(<i>N</i>)		下限	上限
Fleming2002	37	45	30	47	2.62	0.99	6.90
Jakubowicz2001	8	28	0	28	23.6	1.29	433.02
Nestler1996	5	11	1	13	10.00	0.94	105.92
Nestler1998	12	35	1	26	13.04	1.57	108.36
Ng2001	3	9	3	9	1.00	0.14	7.10
Vandermolen2001	1	12	1	15	1.27	0.07	22.72
Yarail2002	6	16	1	16	9.00	0.94	86.52
合计	72	156	37	154			

来源:Jonathan M Lord,Ingrid HK Flight,Robert J Norman.Metformin in polycystic ovary ayndrome:systematic review and meta analysis.BMJ,2003,327:951

该数据资料在 RevMan 5.0 软件中的计算结果见图 2-7-2。

图 2-7-2 左侧所示为 7 个独立研究的数据;中间为个独立研究的固定效应模型(fixed effect model) OR 值及 95% 可信区间(95%CI)的计算结果,如 Fleming 2002(第 1 个)的研究,其 OR 为 2.62,95%CI 为 0.99~6.90;右侧所示为 7 个独立研究的森林图(forest plots),该图的竖线为无效线,即 OR=1,每条横线为该研究的 95% 可信区间上下限的连线,其长短直观地表示了可信区间范围的大小,线条中央的小方块为 OR 值的位置,其方块大小为该研究权重大小。若某个研究 95%

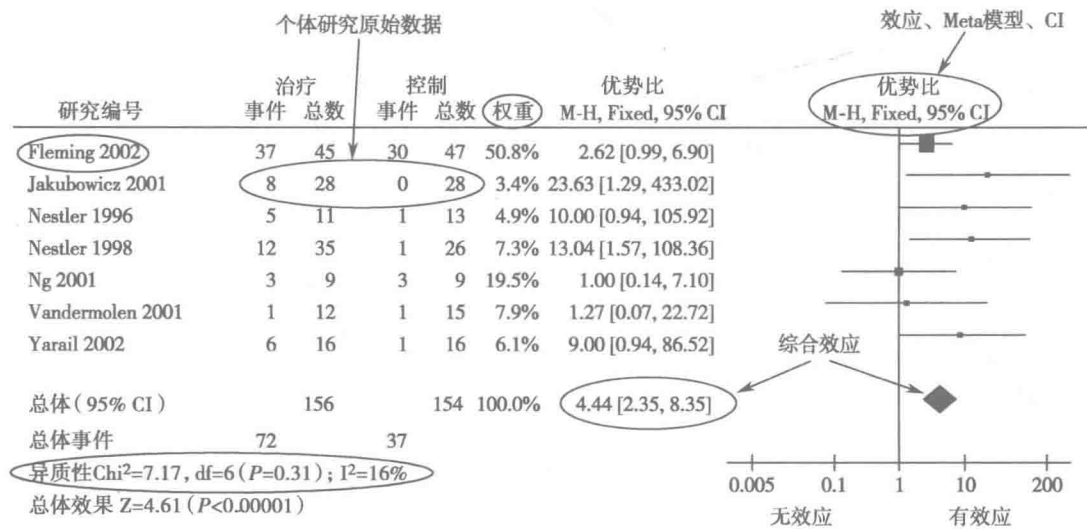


图 2-7-2 7 个二甲双胍治疗多囊卵巢综合征的 Meta 分析结果(森林图)



可信区间的线条相交或相切于无效应线,该研究无统计学意义,反之,若该横线落在无效应线的左侧或右侧,该研究有统计学意义。从图 2-7-2 中可见,在 7 个研究中第 1、3、5、6 和第 7 个研究 OR 的 95%CI 都包含了 1(下限小于 1,上限大于 1),即无统计学意义,认为二甲双胍无效,而其余 2 个研究的 95%CI 的上下限都大于 1,认为二甲双胍有效;中间底部所示为该 7 个研究的 Meta 分析结果:异质性检验(test for heterogeneity) χ^2 值和 P 值,该实例 $\chi^2=7.17$, $P=0.31$, $I^2=16\%$,合并效应量 $OR_{\text{合并}}=4.44$,其 95%CI(2.35~8.35)

根据上述分析结果,可认为这 7 个二甲双胍治疗多囊卵巢综合征研究资料具有同质性(异质性检验 $\chi^2=7.17$, $P=0.31$, $I^2=16\%$),因此,合并效应量 OR 采用固定效应模型, $OR_{\text{合并}}=4.44$,其 95%CI(2.35~8.35),可认为二甲双胍治疗多囊卵巢综合征有效。

本例中,异质性检验结果显示所纳入的研究异质性可以接受。Meta 分析结果显示 G-CSF 组在成人 ALL 总体生存率方面优于对照组。G-CSF 可提高成人 ALL 患者的总体生存率,尚不能证明其有提高完全缓解率、改善生存质量、减少感染情况、降低复发率的作用,还有待进一步研究证实。

(五) 此系统综述总体科学性

通过对检索出的这篇系统综述使用 OQAQ 量表进行质量评价,可以看出该篇系统综述总体科学性一般,主要原因在于该研究存在局限性:①所有纳入文献均未描述具体的随机方法和分配隐藏方法,存在选择性偏倚的可能性;②6 篇研究均未明确指出使用盲法,存在测量和实施偏倚的可能性;③受试者种族、年龄、疾病危险分层、具体实施的化疗方案、G-CSF 的使用方法、剂量及随访时间等不完全一致,可能影响该研究结论的可靠性。

学习
笔记

知识点

OQAQ 量表质量评价的重点:OQAQ 量表不涉及发表质量和研究的重要性,主要针对系统综述中容易产生偏倚的几个关键环节(是否进行了全面的文献检索;如何减少在文献选择、数据提取和质量评价过程中偏倚的产生;对原始研究的质量评价是否采取恰当的评价工具和方法;研究数据合并是否恰当;研究结论是否客观),是目前较常用的工具之一。

在文献检索方面,读者应特别关注:①研究者是否尽可能地使用了各种主要的文献检索方法;②电子检索方法是否全面、系统并清楚地陈述了检索策略;③检索使用的关键词是否全面、得当,是否遗漏了重要的关键词;④是否对各种检索结果有明确的描述;⑤当研究数目较大时,是否用漏斗图等方法评估了发表偏倚的可能性,并对潜在的偏倚进行了测量和控制。

漏斗图(funnel plots)是用每个研究的处理效应估计值为 x 轴,样本含量大小为 y 轴的简单散点图(scatter plot)。对处理效应的估计,其精确性是随样本含量的增加而增加,小样本研究的效应估计值分布于图的底部,其分布范围较宽;大样本研究的效应估计值分布范围较窄。当无偏倚时,其图形呈对称的倒漏斗状,故称为“漏斗图”。实际使用时应注意:作 Meta 分析的研究个数较少时不宜做漏斗图,一般推荐作 Meta 分析的研究个数在 5 个及以上时才需做漏斗图。

纳入/排除标准的制定依据:①研究对象的类型(types of participants);②研究的干预措施(types of interventions)和作为对照的措施(types of comparisons);③研究结局指标(types of outcome measures);④研究的设计方案(types of studies)。

纳入研究时常见的偏倚及控制方法:①纳入标准偏倚(inclusion criteria bias):如果研究者根据某种需要制定了“有偏向的”纳入标准,可能会引入人为的偏倚;②筛选者偏倚

(selector bias):在筛选过程中受研究者主观意愿的影响而引入的偏倚;③制定客观、严密的纳入标准,是控制纳入偏倚和筛选者偏倚最重要的方法。另外,盲法筛选是控制筛选者偏倚常用的方法,即筛选者对一项研究纳入与否的判断独立于该项研究的结果。

评价原始研究的质量对系统综述的作用:①可作为纳入标准之一;②用于解释研究结果间的异质性;③用于敏感性分析;④作为综合结果时赋予权重的依据,即质量越高赋予的权重越大。

亚组分析与敏感性分析:①亚组分析(subgroup analyses)是依据研究的临床或方法学特征进行分组,分别估计各组内研究的结果,并对各组的综合结果进行比较,如果它们之间的差异存在统计学意义,则提示分组的因素可能是异质性的原因之一;②敏感性分析(sensitivity analyses)是通过剔除具有某种特征的少数研究,来观察剩余的多数研究合并的结果是否会发生变化,如果结果发生了明显的改变,则说明剔除的研究可能是引起异质性的原因。

系统综述小结:系统综述可以为临床医生提供一个快速的、可靠的、综合的证据。阅读系统综述时需要应用一定的原则对其质量进行评价。OQAQ量表就是用于评价系统综述真实性的最常用工具。一篇系统综述的质量常常取决于原始研究收集的全面程度和质量,以及提取分析数据的方法。因此,评价系统综述时,首先要明确该系统综述的研究问题与自己面临的问题是否相关,其次是评价该系统综述的质量,最后依据综述的证据,结合病人的需要作出适合具体病人的决策。

思考题

1. Meta 分析在合并各个独立研究结果前应进行
 - A. 相关性检验
 - B. 异质性检验
 - C. 回归分析
 - D. 关联性分析
 - E. 标准化
2. 如果漏斗图呈明显的不对称,且统计学检验提示研究分布不对称,此结果说明
 - A. 异质性统计学检验效能不够
 - B. Meta 分析中各独立研究的结果存在异质性
 - C. Meta 分析中各独立研究的结果具有一致性
 - D. Meta 分析可能存在发表偏倚
 - E. Meta 分析的合并效应值有统计学显著意义
3. 有关 Meta 分析的描述不正确的是
 - A. Meta 分析是一种观察性研究
 - B. Meta 分析能排除原始研究中的偏倚
 - C. Meta 分析可以增加检验效能
 - D. Meta 分析的目的是比较和综合多个同类研究的结果
 - E. 针对随机对照试验所做的 Meta 分析结论更为可靠
4. 系统综述中异质性检验的目的是
 - A. 评价研究结果的统计学显著性
 - B. 增加统计学检验效能



- C. 评价一定假设条件下所获效应合并值的稳定性
 - D. 检查各个独立研究的结果是否具有-致性(可合并性)
 - E. 估计偏倚的大小
5. 发表偏倚是指
- A. 由于我们更偏向于发表样本量大的有“统计学意义”的研究结果而造成的偏倚
 - B. 由于世界上几个主要的医学文献库绝大部分来自发达国家而造成的选择偏倚
 - C. 由于研究者根据需要自定一个不合适的纳入标准所引起的选择偏倚
 - D. 由于研究结果的筛选过程中筛选者主观偏好的影响而引入的偏倚
 - E. 由于数据库数据标引不准确致使所需文献未被检出而产生的偏倚

参考答案

1. B 2. D 3. B 4. D 5. A

(祁艳波)

参考文献

1. 刘鸣. 系统评价、Meta 分析设计与实施方法. 北京:人民卫生出版社,2011.
2. 王家良. 循证医学. 第2版. 北京:人民卫生出版社,2010.
3. 李幼平. 循证医学. 第2版. 北京:高等教育出版社,2009.

第八章 循证实践指南

临床问题

在医疗实践过程中,临床医师经常需要考虑如何提高医疗质量和病人安全,而患者经常考虑如何选择最佳诊疗模式,决策者经常需要考虑如何使卫生保健更加有效。例如,如何选择破裂的颅内动脉瘤手术时机;破裂颅内动脉瘤最佳治疗方式是什么;是否需要在健康体检时进行颅内动脉瘤的筛查;是否需要在遥远偏僻 5A 风景区附近医院建血库;要进行这些决策前,需要循证临床实践指南来帮助临床医师、患者、决策者进行合适的决策。

第一节 基本概念

20 世纪 80 年代,为了提高以及保证医疗质量、确保医疗保健的连续性、降低医疗成本,在全球范围内开展了临床实践指南(clinical practice guideline, CPG)指导医疗实践的医学运动。它是以系统评价为依据,经过专家讨论后由专家学会经过严格评价和筛选后制定,具有权威性和实践意义的临床指导意见。1990 年,美国医学研究所提出了临床实践指南的定义:针对特定的临床问题,经系统研究制定后发布,帮助医生和患者作出恰当的判断的指导性意见,从而选择和决策适宜的卫生保健服务。目的在于:①改善临床结局,提高医疗质量;②以患者为中心、尊重患者价值观的临床决策;③减少临床实践中的不恰当差异,确保病人安全;④促进医疗资源的合理配置;⑤科学合理研究结果,使用临床决策清晰、透明。指导临床实践的指南与原始的临床研究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。

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。这些指南存在很大的异质性,质量参差不齐,某些建议甚至互相矛盾,给临床决策带来极大的困扰。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。随着循证医学的发展,循证临床实践指南(evidence-based clinical practice guidelines, E-CPGS)成为解决上述问题的方法。E-CPGS 针对某一特定的问题,特定的人群,由特定的组织和人员按照规范化的流程,集合当前最佳的证据,根据证据等级和推荐制定,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。它增加了指南的科学性、针对性、实用性。循证临床实践指南的特点:①清晰地界定的临床问题以及人群;②规范的证据质量评价和分级:科学规范地获取评价证据、根据规范统一标准进行证据质量评价并给予分级;③根据证据等级结合病人的价值观进行推荐,从而达到证据的临床转化;④实时更新的特点:如果有新的证据出现,指南应做相应的调整。同样循证临床实践指南也有缺点:费时、费力,医院、相关组织、临床医生应用指南时需要经过专门培训。

知识点

循证临床实践指南与临床实践指南的异同点:两者均为了医疗质量和病人安全。但循证临床实践指南的制定更加清晰、规范、科学、实用。



第二节 评价原则

一、评价原则

(一) 循证临床实践指南真实性评价

循证临床实践指南的核心是指导临床实践决策,因此它的真实性尤为重要。循证临床实践指南真实性评价包括:①指南针对的问题是否清晰(包括人群、不同干预方案、患者所有的重要结局):一个高质量的临床指南会考虑对患者重要的结局如治愈、生活质量、死亡、致残等。②指南制定流程是否规范,一般指南开发包括如下流程:明确临床问题、指南开发团队组织运营、提取以及评价证据、从证据到推荐、修改和更新。③是否由多学科人参与(相关不同学科临床专家、临床工作人员、方法学家等)。④证据提取是否全面、方法是否科学。⑤是否严格评价证据质量并分级:如 GRADE (Grading of Recommendation, Assessment, Development and Evaluation) 证据质量评价与分级。⑥是否根据证据质量分级并结合指南应用背景作出谨慎推荐。确定推荐意见是循证临床实践的难点和核心。指南的推荐意见首先要来自研究的证据,对证据进行质量评价后形成证据级别,然后结合患者的社会、经济、人文、政治、需求、利弊分析等来决定推荐强度即是否推荐。因此,不是所有的高级别证据都是强推荐,即高级别证据也有弱推荐;反之亦然。

(二) 循证临床实践指南重要性评价

循证临床实践指南重要性评价就是评价研究合并后的新的研究结果有多重要。通过系统评价和 Meta 分析、决策分析、成本效益分析后,临床医生通常要知晓不同干预措施对患者的每一个重要和(或)关键临床结局影响的利弊差异具体是多少,备选措施的把握度到底有多大,能够节约多少钱,如在治疗成人急性腰背痛的循证临床实践指南中认为神经根减压手术能加速患者恢复,也必须清晰地告知恢复的速度到底是几天、几星期或是若干月。

(三) 循证临床实践指南实用性评价

1. 指南是否回答了临床需要解决的问题 应用 AGREE 量表评估 CPG 的质量。选择出质量好、可信度高、实用性强的指南后还需要明确指南是否回答了临床需要解决的问题。有时候筛选出的指南可能并未包含临床医生关心的问题,此时可能需要重新筛选指南。值得注意的是临床面临的问题相当复杂。现有指南不可能囊括所有的临床问题,如果针对某个临床问题不能从指南中获取证据,则可能需要查找其他的证据,如系统评价、随机对照研究等。

2. 指南的时效性 随着生物学和医学技术不断发展,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。因此,应尽可能选择最新的指南。例如,既往认为心力衰竭是使用 β 受体阻滞剂的禁忌证,而现在的大型随机对照试验却证实 β 受体阻滞剂可以改善心力衰竭患者的预后,新的指南均认为 β 受体阻滞剂是治疗心力衰竭的重要药物。

3. 患者的临床情况是否与指南目标人群相似 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,而临床工作中却需要将指南建议应用于个体患者。因此,一些医生总是担心自己面临的患者与指南的目标人群存在不同而拒绝应用指南。实际上,应用指南的重要步骤之一是分析临床面临的个体患者与指南的目标人群具有哪些相似性。而不是存在哪些不同之处。合格的 CPG 应该明确而详尽地描述指南的目标人群。尽可能减少临床医生重读原始文献的麻烦,便于临床应用。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。如果某些重要特征与指南的目标人群存在显著差异时,医生就应该根据自己病例的实际情况谨慎使用指南建议。值得指出的是,个体患者的临床特征极少与指南目标人群的特征完全相似。总会存在或多或少的差异,应用指南只是尽可能减少临床决策的不确定



性,而不可能消除这种不确定性。

4. 本地区(或医院)的医疗条件及患者的经济状况如何 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。下一步就需要根据本地区(或医院)目前的医疗条件,评估该干预措施的可行性和费用效益比,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,世界各国指南均推荐急性心肌梗死发作早期数小时内行经皮冠状动脉介入疗法,但我国绝大多数基层医院并无条件开展这项治疗,而且大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。

5. 认识并克服指南在实施过程中可能遇到的障碍 指南在实施过程中会面临来自于医疗机构、社会和医生等各方面的障碍。常见的障碍来源包括:①临床医生,如临床医生盲目自信,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;②环境因素,如来源于制药公司的误导,上级医师不同意采纳指南提供的证据,习惯性给予“常规治疗”;③患者因素,如患者拒绝接受某些治疗;④社会因素,如某些新的治疗措施社保不予支付;⑤其他因素,指南本身也可能成为指南实施的障碍,主要体现在指南表述含糊不清或其建议与临床“常规治疗”存在矛盾。为避免这些障碍需要采取的措施是: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。

6. 患者(或其亲属)的价值取向和意愿 患者(或其亲属)的价值取向和意愿在临床决策中的重要性已经受到临床医生的日益重视。指南的推荐强度越强,采取该项干预措施预期获得效益/风险比就越大,患者(或其亲属)选择该项干预措施的可能性也就越大。而对于那些推荐强度较弱的干预措施而言,预期的效益/风险比变得不再确定,不同的患者可能选择截然相反的干预措施。在这种情况下考虑患者(或其亲属)的价值取向和意愿就显得尤为要。例如,口服阿司匹林可使心肌梗死后死亡率降低约 25%,而且阿司匹林引起消化道出血的风险小,价格便宜,几乎所有的患者都会选择接受阿司匹林治疗;另一方面,如已经口服华法林 1 年的下肢深静脉血栓的患者,如果继续服用华法林,则再次发生下肢深静脉血栓的几率会减少 10%,但是出血的风险会相应增加,而且需要定期检测出凝血时间。在这种情况下,一部分患者可能会选择继续服用华法林,而另一部分患者则可能选择停用,此时患者的价值取向和意愿就显得尤为重要。

指南的质量评价和报告规范每个机构不太一样,目前由 11 个欧洲国家和加拿大共同研究制订的 AGREE 量表已被世界各国广泛用于评估指南的质量和报告规范。包括 WHO 在内的许多国际性学术机构也采用 AGREE 量表来评估他们所制订的指南,具有较高的权威性。AGREE 评价表通过 6 个方面(指南的范围与目的、利益相关者的参与度、指南开发的严格性、表述明确和清晰程度、指南的可应用性、编辑工作的独立性)共 23 个条目对 CPG 进行评分(表 2-8-1)。

表 2-8-1 AGREE 评价工具的具体条目

指南的范围和目的

1. 指南的撰写目的有明确说明
2. 指南包括的临床问题有详细的描述
3. 指南针对的目标人群有明确说明

利益相关者的参与度

4. 指南制定的小组成员包括所有相关方面的专家
5. 指南考虑了目标患者的意愿和喜好
6. 对指南使用人群有清楚的说明
7. 指南发表前已经在目标使用者中进行了预先测试

指南开发的严格性

8. 采用严格、系统的方法检索证据



续表

9. 证据的纳入标准有清楚的描述
10. 形成推荐建议的方法有清楚的描述
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. 推荐建议和证据之间存在客观联系
13. 指南发表前接受了制定小组之外专家的同行评议
14. 提供了更新指南的程序
表述明确和清晰程度
15. 推荐建议专业而不含糊
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. 提供了方便的应用工具(如概要、教育工具、针对患者的宣传单等)
指南的可应用性
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
编辑工作的独立性
22. 指南编辑工作独立于其他经济实体
23. 指南制定小组成员的利益冲突在指南中有表述

二、如何使用 AGREE II

(一) 获取相关信息

评估人员在评估前应努力寻找有关指南制订过程的所有信息。这些信息可能与推荐建议存在于同一个文件中,也可能被总结在一篇独立的技术报告、已出版的论文或政策报告中(例如指南纲要)。因此,在评估前应认真阅读指南及其相关文件。

(二) 评估人员的数量

建议每个指南至少需要 2 个评估人员,最好由 4 个评估人员进行评价,这样可以增加指南评估的可靠性。

(三) 确定评分等级

AGREE 的每个条目和两个全面评价条目均以 7 分表评价(表 2-8-2)。这个评分等级衡量了该条目标准执行的程度。如果没有与 AGREE 条目相关的信息或报道的概念非常差,则为 1 分;如果报道的质量很高,有用户手册里完整的标准和清楚的理由,则为 7 分;当 GREE II 条目报道不能满足全部标准或理由,则根据不同情况给予 2~6 分。分值分配取决于报道的完整性和质量。当更多的标准被满足和理由更充分时,则分值增加。

表 2-8-2 AGREE 针对每个条目的评分等级和注释

1	2	3	4	5	6	7
很不同意						很同意
注释						

(四) 填写注释

在每一个条目后面都会有一个注释框,可以在框中填写之所以这样回答的原因。例如,可能因为没有信息、该条目不适用,或者是该信息描述方法不尽如人意而回答“很不同意”。

学习
笔记



(五) 计算各领域得分

分别计算 AGREE 每个质量领域的得分,6 个领域的评分是独立的,不能合并为一个质量评分。每个领域得分等于该领域中每一个条目分数的总和,并标准化为该领域可能的最高分数的百分比(表 2-8-3)。如果条目没有包括,需要适当修改最大可能分值和最小可能分值的计算方法。对指南的评价必须有较高的水准,这十分重要,这个标准和理由是指导而不是取代评价者的评价。因此,AGREE II 条目没有为 7 分表中每一分值提供清楚的解释。4 个评价员给领域 1(范围和目的)的评估分数见表 2-8-3,最高可能分数=7(很同意)×3(条目)×4(评价员)=84;最低可能分数=1(很不同意)×3(条目)×4(评价员)=12;该领域的标准化总分应为:(实际总分-最低可能分数)÷(最高可能分数-最低可能分数)=(53-12)÷(84-12)=0.5694×100%=57%。

表 2-8-3 AGREE 各领域得分的计算方法举例

评价员	条目 1	条目 2	条目 3	总分
评价员 1	5	6	6	17
评价员 2	6	6	7	19
评价员 3	2	4	3	9
评价员 4	3	3	2	8
总分	16	19	18	53

(六) 得分说明

这 6 个领域的得分各自独立,不能将其合计为一个总的质量得分,尽管这些领域的分数可以用来比较指南,帮助决定是否应用或推荐一个指南。但是不能对得分设立一个阈值去评价指南的优劣。

(七) 综合评估

AGREE II 最后一部分为综合评估(表 2-8-4),要求用户对指南的质量作一个评判,并要考虑在评价过程中应用的标准。用户也会被问及是否将推荐这个指南的使用。两个总体评价包括:给指南总的质量评分和“我将推荐使用这个指南”。用户也可以提出具体的修正意见,仅仅制定出循证临床实践指南并不能达到改善患者预后的目的。指南只有具体应用于个体患者时才能发挥提高诊断,改善患者预后,提高医疗服务水平的作用。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。

表 2-8-4 AGREE 对指南总体质量的评分表

1	2	3	4	5	6	7
可能最低质量						可能最高质量

知识点

指南真实性评价:主要评价指南制定过程是否科学规范。

循证临床实践指南实用性评价:证据质量与推荐之间的关系。要结合病人具体情况以及所在医疗机构的实际情况而定。指南中高级别的证据有可能会是弱推荐,低级别的证据也有可能会是强推荐。

思考题

1. 循证临床实践指南的定义是什么?



2. 如何评价循证临床实践指南?

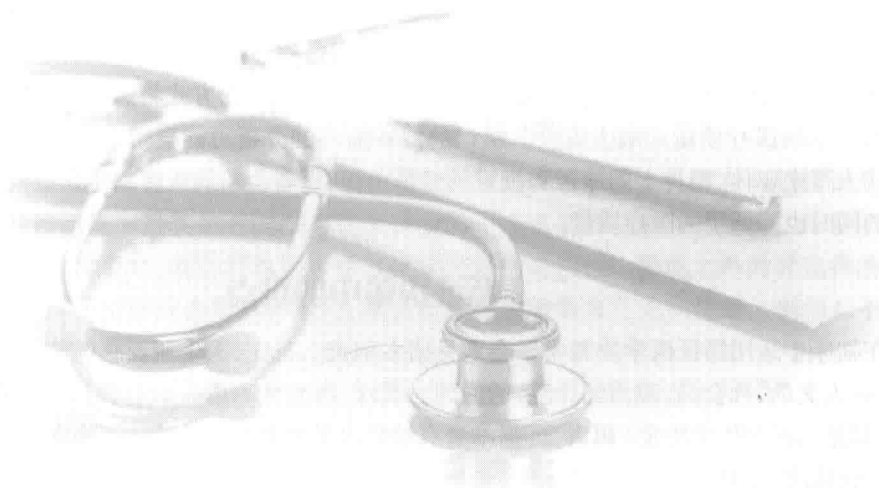
(徐佩茹)

参考文献

1. Straus SE, Glasziou P, Richardson WS, et al. Evidence-based medicine. 4th. edition. London: Elsevier Churchill Livingstone, 2011.
2. Dan Mayer. Essential Evidence-based Medicine. 4th edition. UK: Cambridge University Press, 2004.

第三篇

循证实践



第一章 循证决策和决策分析

一、循证决策的概念

医生每天都要面对不同临床症状和体征的病人,针对病人的具体情况应用不同的诊断方法,提出适当的治疗措施,采用相应的护理措施,分析疾病发生可能的危险因素,为其预后进行判断。所有上述过程称之为临床决策。随着医学对疾病认识的加深和病人对医疗质量的要求增高,越来越多的经典医疗决策受到挑战。究其原因主要是由于这些决策不恰当或不能达成全球一致,经不起时间或空间的检验。最近一项研究表明:仅有一半的医生会用已经被证据证实的药物来挽救糖尿病和心脏骤停患者的生命。另一项研究表明:美国每年约有 98 000 人死于不恰当的或错误的医疗决策。从此可以判断,许多临床医生所作的临床决策是武断的或是多变的。人们往往认为这是由于地域差别造成的,可是就在相同的地域内和相同的医院级别,同一种手术率差别也非常大。如在美国缅因州,一个地区子宫切除手术率是 20%,而另一个地区则为 70%。在英国不同的地区,前列腺手术、冠脉搭桥术、甲状腺手术的手术率差别可高达 300%。世界 13 个大城市中膝关节置换手术的差别为 700%、颈内动脉剥脱术为 2000%。就同一个病人,让同一位医生间隔一段时间先后判断两次,有 40% 的可能性两次决策不一致。

如何进行个体化的诊疗?如何科学有效地进行临床决策?循证临床决策是解决上述问题的新模式,它不强调直觉、非系统的临床经验和基于病理生理推理作为临床决策的依据,而是运用循证医学的方法,将患者的具体情况(一般情况、症状、体征、辅助检查)、价值观和当前最佳的证据进行整合,为患者所患疾病的病因、诊断、治疗、预防、危害、预后乃至花费作出适合个体决策的一个过程。具体来说就是证据的临床评价、应用过程。诸多的因素影响这个过程,因此循证临床决策比任何时候都要困难和复杂。医生整合临床信息的方式将影响临床决策的准确性,从而大大影响医疗质量和病人安全。错误的或不恰当的分析可能会带来灾难性的后果。轻者会给病人带来躯体、精神上的痛苦和沉重的经济负担,重者会给病人带来生命危险,在影响患者安全的同时也降低了的医疗质量。

二、循证决策中的难点

在临床中运用循证医学决策常常会遇见诸多困难。首先,医生所面临的情况很多而且非常复杂,有人文的、社会的、政治的、经济的原因。其次,医生并不事先知道自己所作决策的后果是什么,但是又必须作出决定。再者,证据本身有缺陷或是根本没有好的证据来支持决定。表 3-1-1 列出了临床决策困难常见的原因。

表 3-1-1 临床决策困难常见的原因

因素	解释
临床问题的复杂性	临床决策受到的多方面影响
决策后果的不确定性	临床研究结果的多样性
必须作出决定	自认为必须作出决策
巨大的安慰效应	疾病的自愈(很多情况下仅仅宣教就可以解决问题)
病人的高期望	患者和社会对医疗的过高期望



续表

因素	解释
政府的高期望	政府要求医疗保健服务物美价廉
专家的压力	专家要求做事的方式和他一样
追逐热门技术	医疗机构必须有新技术而且鼓励开展新技术

临床医生通过评价、应用和推荐最佳证据给病人,为病人决策服务。要求临床医生在三个方面进行训练:证据的评价、深入了解病人情况和通过与病人良好的沟通深知患者的价值取向。要么通过对最佳证据的评价,然后和病人的实际情况和价值取向进行结合,为病人作出决策;要么通过临床指南进行决策分析。下面就用实例来进行临床决策。

三、循证决策流程和注意点

①提出并构建问题(asking question);②针对问题进行证据查询(acquiring evidence);③评价证据(appraising evidence);④临床应用证据(applying evidence);⑤后效评价(after assessment)。如何提出并构建问题、如何进行证据查询、如何评价证据在前面的章节都有描述,在此不一一赘述。本章重点讲述在临床决策过程中如何科学、正确、合理地应用证据,并进行后效评价。其取决于以下几个关键步骤:

(一) 病人的人群特点 / 干预措施 / 对照 / 结局(PICO)是否与证据中的 PICO 相似或一致

我的病人是否适合使用证据,病人的依从性如何,医生的依从性如何。广为人知的 PICO 方法即病人 / 干预 / 对照 / 结果 / 环境,在进行循证临床决策时的价值逐渐被广大医生和病人所接受。在进行临床决策时,首先将证据中的 PICO 与现实中的病人的 PICO 进行比较,只有证据中 PICO 的每一项与现实病人 PICO 相似或一致,证据才有可能应用于病人,才能进行决策。首先是证据中研究的人群特征与将要进行决策的人的基本特征相符合(包括所有影响诊断、治疗、预后的所有已知的生物学特性(如年龄、性别、族别、病情的轻重缓急、疾病的不同类别等)和社会学特征(如经济、政治文化、教育等)。如在证据中研究的病人主要是重病人而你面对的是一个轻病人,证据等级再高仍然不适用于你的病人,也不能为你的病人的诊疗作出决策。同理,证据中的干预措施与你将要实施的干预措施也要相似或一致,包括确保治疗安全性和疗效所需的一系列诊断性试验、监测设备、干预能力、技能及其他技术要求。尤其在有创性治疗试验中,临床医生的技能是衡量其是否能参与临床试验的一个重要条件。当普通医生的技能达不到参与试验的医生标准时,就必须慎重考虑该试验结果的适用性。除此之外两者选择的对照是一致的。如一种干预措施选择多种对照,就会给临床决策带来巨大的挑战。例如对于接受保守(无创性)治疗的非 ST 段抬高的急性冠脉综合征,欧洲心脏病学协会推荐使用抗凝药物,由于对照不同,就有磺达甘肝素、肝素和依诺肝素三种药物选择,如果对照不一样,所选药物就不一样。有时证据实施的环境同样也能影响决策。例如,为富有的经济发达地区制定的中耳炎的抗生素治疗证据是不适用于一个来自贫困地区的患者,即使证据再好也不能为你的病人的治疗进行决策。

(二) 充分权衡不同诊疗方法利弊基础上的利弊平衡

循证临床决策时患者和医生都必须权衡各种治疗方案的利弊。对预期利弊的最佳估计及决策者对这些估计的信心均会影响其最终决定。利弊平衡差距越大,越容易作出临床决策。差别越小,越不容易作出临床决策,要取决于病人的价值取向甚至花费。如某项治疗的利远远大于弊或者弊远远大于利,就很容易作出决策使用这项治疗或不使用这项治疗。如给早产倾向的孕妇使用类固醇激素可以降低胎儿发生呼吸窘迫综合征的风险,此证据,利远大于弊,对于此类患者很容易作出决策,即使用类固醇激素。

学习
笔记



(三) 证据质量

不同的组织和研究设计对证据的质量定义不同。临床流行病学文献中使用的“质量”通常是指单个研究的内在效度(如偏倚风险)的判断,就对照试验而言一般考虑的是随机、随机隐藏、盲法以及意向性分析的运用或有无;对观察性研究而言一般考虑的是恰当的测量暴露与结局以及适当的控制混杂。但是对照研究和观察性研究均要考虑失访问题,以及设计、实施和分析过程中影响偏倚风险的其他因素。在系统评价中证据质量是指对效应估计值正确的把握程度,但是在循证临床决策中证据的质量是我们对效应估计值足以支持某种决策的把握程度。运用最佳证据为临床实践决策时不仅要考虑偏倚风险,而且还要考虑影响偏倚风险的其他因素,如不精确、不一致、间接性、发表偏倚以及效应量的大小以及量效关系。

(四) 病人及其家人的价值观和价值取向以及医患沟通

循证临床决策非常强调并确保决策符合病人及其家人的价值观和意愿。原因之一是患者和临床医生的价值观可能存在本质的区别。不同个体患者之间价值观的差异也很大。例如,Devereaux 及其同事调查了 63 名医生和 61 名患者,问题是“您能接受 100 名患者中出现多少例严重的消化道出血事件,您仍愿意开具处方或服用华法林以预防 100 名患者中的 8 次卒中事件(4 次小卒中,4 次大卒中)吗?”虽然医生的回答多种多样,但多数患者都很重视卒中风险的预防,愿意接受 22% 的出血风险以降低 8% 的卒中风险。即便患者的价值观和意愿存在明显差异,但只愿意接受较小的出血风险以降低 8% 的卒中风险的患者仅占少数。这些数据表明,为了确保临床决策符合患者的价值观和意愿,有必要让患者参与决策。患者有效参与决策要求其理解其他可选治疗方法的利与弊,包括这些方法对生存质量的影响。如何将这些通常都很复杂的信息有效地传达给患者,这对于临床医生无疑是一个巨大的挑战。为此,近年来研究者制作了决策辅助工具,以通俗易懂的方式向患者传达各种可选的治疗方法及其相关结局,但选择的复杂性依然存在。

(五) 花费

在临床实践中并非所有高质量的、利远远大于弊的证据就能顺利地应用于临床。例如一项治疗措施很好,质量也很好,但是花费却非常昂贵,它应用于临床的障碍就很大。

四、循证决策的方法

临床决策中疾病的诊断决策非常关键和重要,而且非常特殊,常常需要清楚地了解诊断试验的规律。另外,在临床中常常会遇见证据级别不高或是病人以及家人有意愿参与决策的情况,常常采用决策树分析法进行决策。下面就这两种特殊的临床决策作一个简要的叙述。

(一) 疾病的诊断决策

医生每天都要采集和分析大量不同病人的临床信息,对疾病进行诊断决策,比如选用何种检查可以比较准确地诊断疾病,也就是说要选用更加准确的诊断试验。进行正确科学有效的诊断决策的前提是正确认识和使用诊断试验。另外,对个体病人进行诊断与治疗决策时,会有这样的局面:病人是需要新的诊断试验还是不需要新的诊断试验而直接治疗?所有这样的诊断决策需要正确认识诊断试验的规律,具体的内容见本书第二篇第二章以及第三篇第四章。

(二) 决策树分析法

临床上许多事件都是随机的,比如病人的疗效和预后,为了形成以后可以多次使用的临床策略时,可以使用计量的决策分析,比较常用的是决策树分析法。下面结合一个例子简要说明如何用决策树分析法进行临床决策,如颅内动脉瘤有两种治疗模式:开颅手术夹闭和介入手术栓塞治疗,而且颅内动脉瘤有出血的和未出血的状态,如果采用开颅手术,手术复发率低,但是有可能造成神经功能损害,如果采用介入手术复发率高,但是病人恢复快。这样的情况下该如何决策呢?



决策树分析由三部分组成,即决策点、机遇点和结局,分别用小方框、小圆圈和小三角表示。决策点在决策树上用小方框表示,在这一点决策者可以选择可供选择方案中的一种,用线条表示,连接决策点和机遇点。机遇点在决策树中用小圆圈表示,代表不同的病情、获得的不同的检查结果以及不受医生控制的事件,如出血的颅内动脉瘤和未出血的颅内动脉瘤。针对不同的治疗方式和不同的病情就会有不同的结局,在决策树中用小三角表示。外加一些线条、文字说明以及数字描述构成了决策树(图 3-1-1)。下面举例说明如何绘制决策树以及如何分析。

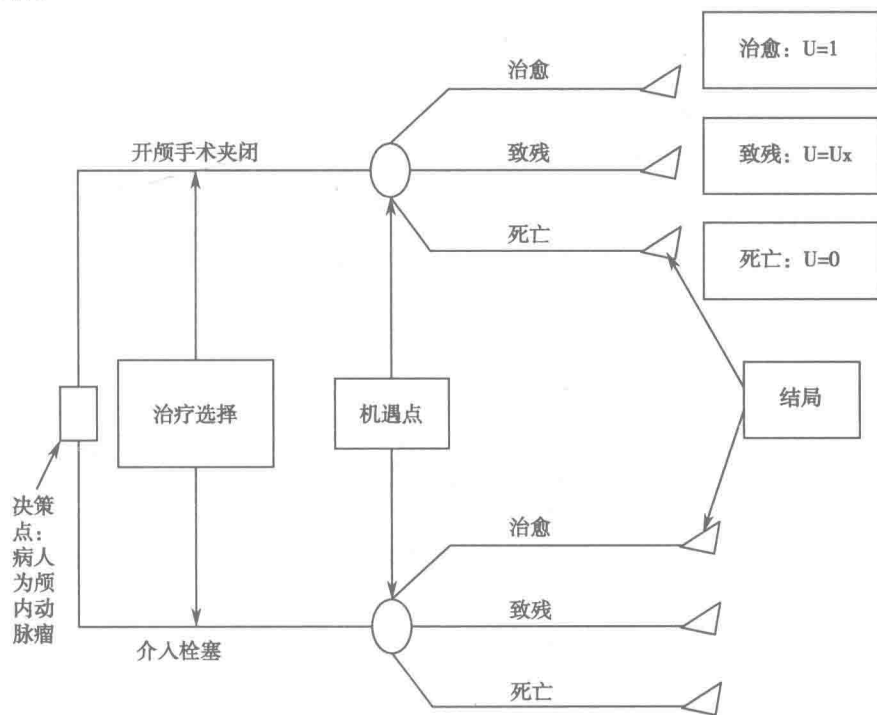


图 3-1-1 决策树构成

第 1 步:明确所有的备选方案。本案例中该种疾病在决策点上有两种可供选择的方案——开颅手术夹闭和介入手术栓塞。

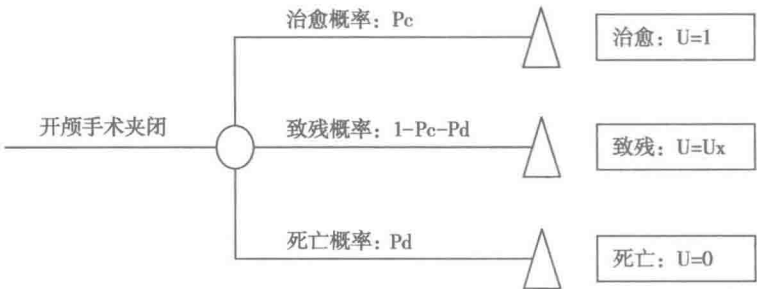
第 2 步:列出不同方案以及不同病情组合的结局。不管选用哪种治疗方案,病人的结局取决于一系列机会事件。也就是说无论选择开放手术还是介入手术,病人有可能会出现死亡、功能损害和治愈这几种结局。在决策树中都要列出。

第 3 步:明确各种结局可能出现的概率。开放手术或介入手术的治愈率、致残率、致死率可以从文献中来,也可以根据经验进行推测。

第 4 步:将最终的临床结局用适宜的效用值赋值。效用值是对患者健康状态较好的测量,通常在 0~1 之间,如治愈为 1,死亡为 0。有时用寿命年或质量调整寿命年表示。

第 5 步:计算每一个备选方案的期望值,选择期望值最好的备选方案为决策方案。计算期望值是从“树尖”向“树根”方向计算。开始向“树根”的方向进行计算,将每一个机会结所有的不同状态效用值与其发生概率分别相乘,其总和为该机会结的期望效用值。在每一个决策臂中,各机会结的期望效用值分别与其发生概率相乘,其总和为该决策方案的期望效用值。在决策树中的次级决策结,应选择可提供最大期望效用值的决策臂计算,而忽略其他臂(图 3-1-2)。

第 6 步:应用敏感分析对决策分析的结论进行测试。这是决策分析的最后一步,是检测决策分析的真实性。敏感性分析所要回答的问题是:当概率和结局效应值在一定合理的范围内变化时,决策分析的结论会变化吗?(图 3-1-3、图 3-1-4 所示)



开颅手术夹闭期望值 (E)= $P_c(1)+(1-P_c-P_d)U_x+P_d \times 0$

图 3-1-2 决策树中开颅手术夹闭的期望值

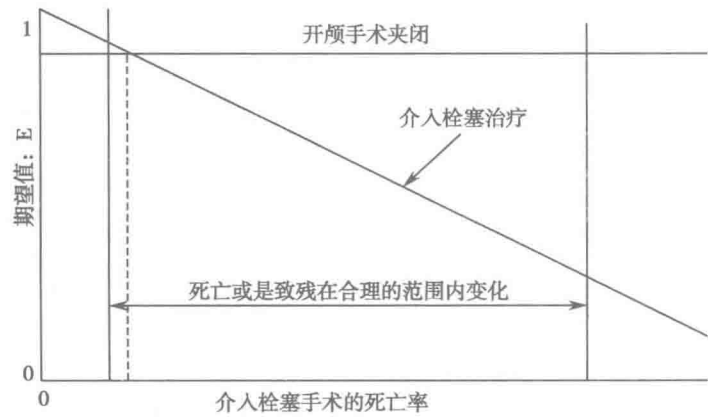


图 3-1-3 颅内动脉瘤治疗单一决策敏感性分析简易模式图

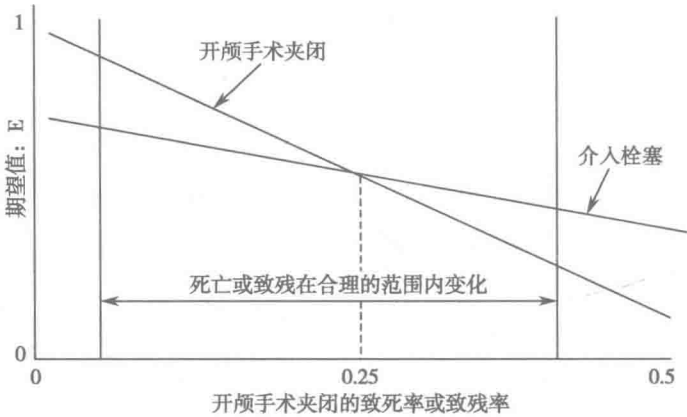


图 3-1-4 颅内动脉瘤治疗复杂决策敏感性分析简易模式图

经过上述 6 个步骤就形成了颅内动脉瘤手术的决策分析,该决策分析包括了所有的治疗模式以及可能的结果,可能的选择方案的概率比较符合临床基本情况,对决策树也进行了敏感性分析,可以适用于实际的病人。

最后,合理地应用于临床后,对其疗效以及安全性进行实际评价,看在应用过程中有哪些确实方便有用,同时也找出证据使用中的病人障碍、医生障碍,以及在医院层面证据使用的障碍并提出解决问题的方法,更加有益于证据的应用和评价。

知识点

循证决策的概念:运用循证医学的方法,将患者的具体情况(一般情况、症状、体征、辅助检查)、价值观和当前最佳的证据进行整合,为患者所患疾病的病因、诊断、治疗、预防、危害、预后乃至花费作出适合个体决策的一个过程。

证据质量:一般指单个研究的内在效度(如偏倚风险)的判断。不同的研究类型证据质量的判断不同。

循证决策流程:①提出并构建问题(asking question);②针对问题进行证据查询(acquiring evidence);③评价证据(appraising evidence);④临床应用证据(applying evidence);⑤后效评价(after assessment)。

循证决策中难易度与以下有关:①证据中的 PICO 与实际的病人的 PICO 的一致,决策容易实施;②利弊差异越大,决策越容易;③证据质量越高,决策越容易;④病人价值取向满足度越高,决策越容易实施;⑤决策物美价廉就容易实施。

思考题

想一想,就所学习的内容对临床实践过程中遇到的具体案例(诊断、治疗、预后、病因)该如何进行循证决策分析?

(罗 坤)

参考文献

1. Straus SE, Glasziou P, Richardson WS, et al. Evidence-based-medicine. How to practice and teach EBM. 4th ed. Churchill Livingstone, Elsevier, 2011.

第二章 循证实践指南的临床应用

临床问题

患者男性,45岁,以“突发头痛3小时”为主诉入院。头颅CT平扫提示“蛛网膜下腔出血”。临床考虑:“颅内动脉瘤破裂自发性蛛网膜下腔出血”。如何依据循证临床实践指南进行循证临床实践决策呢?循证临床实践指南临床应用步骤见表3-2-1。

表 3-2-1 循证临床实践指南临床应用步骤

基本步骤	主要项目	具体内容
1. 明确具体的临床问题	病人:新发颅内动脉瘤致蛛网膜下腔出血的成人患者 证据:循证临床实践指南	用循证临床实践指南对颅内动脉瘤性蛛网膜下腔出血成人患者进行临床决策
2. 查询循证临床实践指南	相应的循证临床实践指南网站去检索	基于循证实践指南的循证决策分析
3. 评价循证指南	指南的真实性评价 结果的重要性评价 指南的临床应用性评价	指南的制定过程和方法 指南的评价量表选择
4. 应用循证指南解决临床问题(PICO原则)	应用证据的病人条件;应用后可能带来的利与弊;如何处理可能出现的问题	指南是否回答了临床需要解决问题 是否能指导颅内动脉瘤性蛛网膜下腔出血成人患者的临床诊疗 指南的时效性 是否是最近制定的 临床人群是否与指南目标人群相似 拟应用指南地区的医疗条件和经济状况如何 认识并克服指南实施过程中的障碍 病人或家属的价值取向如何 后效评价 目前医生所在医院的有关颅内动脉瘤性蛛网膜下腔出血成人患者的临床效果如何

一、提出问题

如何利用指南对颅内动脉瘤破裂致自发性蛛网膜下腔出血的成人患者进行诊疗决策。问题的目标人群是“颅内动脉瘤破裂致自发性蛛网膜下腔出血的成人患者”,目的是利用循证临床实践指南来为病人进行决策分析。那么颅内任何部位的动脉瘤破裂致蛛网膜下腔出血的成人患者就是指南的目标人群,而未破裂的颅内任何部位的动脉瘤的成人患者则不是指南的目标人群。其他的专家共识、系统评价等则不是所需的证据。



二、针对临床问题查询循证临床实践指南

在 National Guideline Clearinghouse 网站上查询,输入检索词“aneurysmal subarachnoid hemorrhage”后有 11 条相关指南。其中第一条题目为“Guidelines for the Management of Aneurysmal Subarachnoid Hemorrhage:A Statement for Healthcare Professionals From a Special Writing Group of the Stroke Council,American Heart Association”的指南符合我们的要求。

三、从真实性、重要性和临床应用性来评价指南

首先此指南由美国心脏病协会(AHA)下属的脑卒中专业组为此种特殊疾病所写的循证临床实践指南的再次更新。在背景中描述了疾病的重要性(高发病率)和危险性(高致死和致残率),接着描述了指南制定的小组是多个专业小组,证据的检索是全面的,完全遵照了循证医学的理念,由高级别向低级别过度。首先查找随机对照试验或是非随机的同期队列研究,假如没有如此研究就查询病例分析或是非随机的历史对照的队列研究,这些研究仅仅用于一些特殊的问题。证据的推荐也是依据标准的 AHA 证据等级标准。AHA 将治疗和诊断预后证据级别都分为三级,即 A 级、B 级、C 级。治疗性 A 级证据是指证据来源于多个随机对照试验、B 级为证据来源于单个随机对照试验或多个非随机试验)、C 级为证据来源于专家共识。诊断性或预后 A 级证据是指证据来源于多个前瞻性队列研究且由互不知情的评价者参照统一标准判断、B 级为证据来源于单个 A 级研究或多个病例对照研究或多个由互不知情的评价者参照统一标准判断的研究)、C 级为证据来源于专家共识。AHA 将治疗和诊断预后证据的推荐级别分为三级即 I 级、II 级、III 级。I 级推荐是指在这种情况下,绝大多数患者和医生以及医疗机构认为该方案是有效的 / 有用的,会采纳该方案。II 级推荐指治疗和措施的有效性和实用性上有冲突或是分歧。在这种情况下,多数患者和医生以及医疗机构认为该方案是有效的 / 有用的,仍然有相当的患者和医生以及医疗机构会根据实际情况采用另外的方案。II a 证据或是共识认为治疗和措施有效性和实用性不够充分,II b 证据或是共识认为治疗和措施有效性和实用性很不充分。III 级推荐是指证据或是大多数意见认为措施或治疗是无效的和没有用的,在这种情况下,绝大多数患者和医生以及医疗机构认为该方案是无效的 / 无用的,甚至是有害的。会拒绝该方案。无论是 I 级还是 III 级推荐我们都认为是强推荐,因为它要么推荐要么强烈反对即不推荐(表 3-2-2)。

表 3-2-2 证据级别与推荐强度

I 级推荐	在这种情况下,绝大多数患者和医生以及医疗机构认为该方案是有效的 / 有用的,会采纳该方案
II 级推荐	在这种情况下,多数患者和医生以及医疗机构认为该方案是有效的 / 有用的,仍然有相当的患者和医生以及医疗机构会根据实际情况采用另外的方案 II a:多数证据认为该方案有效 II b:部分证据认为该方案有效
III 级推荐	在这种情况下,绝大多数患者和医生以及医疗机构认为该方案是无效的 / 无用的,甚至是有害的。会拒绝该方案
治疗证据级别:	
A	资料来自多个随机对照试验
B	资料来自单个随机对照试验或多个非随机试验
C	专家共识
诊断 / 预后证据级别	
A	资料来自于多个前瞻性队列研究且由多个专家使用统一标准的评价指标盲审
B	资料来自单个 A 级证据或多个病例对照研究或多个使用开放、非标准的评价指标评价的研究
C	专家共识

学
习
笔
记



如何应用指南来指导诊疗

1. 是否回答了临床需要解决的问题? 该指南就颅内动脉瘤破裂致自发性蛛网膜下腔出血的诊疗用循证医学的方法给予了全面准确而清晰的回答。囊括了该疾病的一般情况如高发病率、高致死率、高致残率,分别清晰地阐述了疾病的危险因素以及方式办法,自然病程以及预后,如何就临床症状体征进行诊断决策,如何急救,防止再出血的措施,治疗模式的选择(手术还是介入),详细的交代了围术期的管理,并发症的处理以及就该疾病的预后也给予了充分的交代。因此不仅是医生,而且也是病人及其家属针对此问题进行临床决策的工具。

2. 指南的时效性是什么? 该指南是对 1994 年指南的更新,是最新的指南。

3. 患者的临床情况是否与指南目标人群相似? 案例中的患者与指南目标人群基本一致,无论是基本情况(如年龄、出血方式、我国该病的流行情况),还是有关疾病的危险因素以及方式办法,自然病程以及预后,如何诊断,如何急救,防止再出血的措施,治疗模式的选择(手术还是介入),围术期的管理,并发症的处理等方面都是该患者和医生将要面临的问题。

4. 本地区(或医院)的医疗条件及患者的经济状况如何? 如果该患者就诊医院的医疗条件如诊断水平、治疗水平以及患者的经济情况基本与指南中相似,则可以使用指南。譬如,如果上述病人因发病就诊于某家医院,该医院有诊疗颅内动脉瘤破裂致蛛网膜下腔出血的经验,每年有 120 例手术的经验,且医院有相应的螺旋 CT(computed tomography)可以急诊实施 CTA(computed tomography angiography)的经验,患者意愿可以接受手术治疗。

5. 认识并克服指南实施过程中可能遇到的障碍? 此指南实施过程中的障碍可能如下:医生的障碍:在颅内动脉瘤的认识不够,或者手术技术(夹闭或介入)水平欠缺。影像诊断误差大。譬如该患者就诊了一家没有相应专业人员的医院,该指南无法应用,患者只有转院。

医院的障碍:颅内动脉瘤诊疗相应的绿色通道没有建立或是不通畅。没有相应的检测手段,如没有 CTA 检查手段,没有相应的专业人员或是监测手段以及监护人员和措施不到位。

病人的障碍:经费不足以负担诊疗费用,拒绝相应的治疗,或是病人不能授权且家人又不在病人跟前,无法签署相关同意书以及对医院和医生的不够信任等。

社会的障碍:社保可能不予支付相应的费用,比如介入手术昂贵的材料费用等。

6. 患者(或其亲属)的价值取向和意愿如何? 患者(或其亲属)的价值取向和意愿在临床中是非常重要的。尽管指南的推荐强度很强,也就是采取该项干预措施预期获得效益的风险比大,但是它并不符合患者(或其亲属)的价值观,仍然不会选择该措施,反之亦然。在指南中有关如何选择诊断颅内动脉瘤性蛛网膜下腔出血的措施中建议所有疑似的病人应行头颅 CT 平扫(B 级证据,强推荐),但是 CT 没有发现蛛网膜下腔出的病人推荐腰穿(B 级证据,强推荐)。家人在亲人疾病突如其来的危急心理状态下以及害怕有创带来的风险,尽管指南是 B 级证据强推荐,大多数 CT 没有发现出血的病人及其家人仍然不会去选择腰穿,而会直接选做 CTA 来进一步检查。

四、应用指南

对此类病人的诊断、治疗、预后以循证临床实践指南为依据进行临床策,也就是如何推荐。将指南推荐给患者取决于 5 个方面的因素:①充分权衡不同治疗方案利弊基础上的利弊平衡。利弊越不平衡,越是强推荐。利弊相当的方案是弱推荐。②证据质量。当一项干预措施的利弊大小不确定时,作为推荐去支持或反对某个方案就会有问题。③患者的价值观和意愿的不确定性或是多变性。④费用。⑤医生的专业能力,比如上述患者就诊的医院仅有手术夹闭的医生或是仅有手术介入栓塞的医生或者两者都没有,这时医生的专业特长就决定了病人最终的治疗模式。就颅内动脉瘤的手术治疗而言,手术夹闭和介入栓塞两种不同的治疗措施,都是 B 级证据强推荐,也就是说利益平衡和证据质量都相似,假如手术夹闭和介入栓塞的医生都很有经验,患



者的价值观取向或是费用问题决定患者选择何种手术方式,如果患者觉得介入手术更为微创愿意承担介入栓塞导致的出血风险,他/她就会选择介入栓塞。假如上述患者就诊的地区,介入栓塞的费用远远大于手术费用,或者相反,而这费用差距病人承担不了,就会选择物美价廉的治疗模式(表 3-2-3)。

表 3-2-3 证据级别与推荐强度之间的关系

	I 级推荐 利 >>> 弊	II a 级推荐 利 >> 弊	II b 级推荐 利 ≥ 弊	III 级推荐 弊 ≥ 利
A 级证据 来自多个人群研究 资料来自于多个 随机对照试验研 究或 Meta 分析	对于患者:绝大多 数患者会采纳该 方案 对于医生:绝大多 数患者应该接受 该方案 对于政策制定者: 该推荐方案绝大 多数情况下会被 采纳作为政策	在这种情况下,多 数患者和医生以 及医疗机构认为 该方案是有效的/ 有用的,仍然有相 当的患者和医生 以及医疗机构会 根据实际情况采 用另外的方案	在这种情况下,部分 患者和医生以及医 疗机构认为该方案 是有效的/有用的, 是可以考虑的一种 方案	在这种情况下,绝 大多数患者和医 生以及医疗机构 认为该方案是无 效的/无用的,甚 至是有害的。会 拒绝该方案
B 级证据 来自有限的人群 研究 资料来自于单个 随机对照试验研 究或多个非随机 试验研究	对于患者:绝大多 数患者会采纳该 方案 对于医生:绝大多 数患者应该接受 该方案 对于政策制定者: 该推荐方案绝大 多数情况下会被 采纳作为政策	在这种情况下,多 数患者和医生以 及医疗机构认为 该方案是有效的/ 有用的,仍然有相 当的患者和医生 以及医疗机构会 根据实际情况采 用另外的方案	在这种情况下,部分 患者和医生以及医 疗机构认为该方案 是有效的/有用的, 是可以考虑的一种 方案	在这种情况下,绝 大多数患者和医 生以及医疗机构 认为该方案是无 效的/无用的,甚 至是有害的。会 拒绝该方案
C 级证据 来自非常有限的人 群研究 来自专家共识、 标准治疗或系列 病例报告	对于病人:绝大多 数患者会采纳该 方案 对于医生:绝大多 数患者应该接受 该方案 对于政策制定者: 该推荐方案绝大 多数情况下会被 采纳作为政策	在这种情况下,多 数患者和医生以 及医疗机构认为 该方案是有效的/ 有用的,仍然有相 当的患者和医生 以及医疗机构会 根据实际情况采 用另外的方案	在这种情况下,部分 患者和医生以及医 疗机构认为该方案 是有效的/有用的, 是可以考虑的一种 方案	在这种情况下,绝 大多数患者和医 生以及医疗机构 认为该方案是无 效的/无用的,甚 至是有害的。会 拒绝该方案

学
习
笔
记

五、对应用指南后的效果进行后效评价

根据指南对上述的病人进行了诊疗,会发现由于地理、环境、人文、社会、经济、政治等的不同,而有相应的指南,最好是当地的指南最适用于当地的病人,因此会对该指南进行修正。或者在使用指南的过程中发现问题而进行重新论证,重新提升某个问题的证据等级或是推荐强度。

思考题

- 1. 循证临床实践指南临床实施的障碍有哪些?



2. 如何构建循证临床实践指南实施的监测反馈网络?

(罗坤 徐佩茹)

参考文献

1. Straus SE, Glasziou P, Richardson WS, et al. Evidence-based medicine How to practice and teach it. 4th ed. Elsevier Churchill Livingstone, 2011.
2. Joshua B, Bederson E, Sander Connolly, et al. Guidelines for the Management of Aneurysmal Subarachnoid Hemorrhage: A Statement for Healthcare Professionals: From a Special Writing Group of the Stroke Council American Heart Association. Stroke, 2009, 40: 994-1025.

第三章 病因研究的循证实践

临床问题

医疗过程中临床医师经常需要考虑是否某种危险因素或治疗措施对患者有害。例如,一名长期服用钙通道阻滞剂的 70 岁老年女性高血压患者就曾询问医生:“服用钙通道阻滞剂是否增加患癌症的危险”。要回答这些问题,就需要在有关病因学研究(或治疗措施副作用研究)中寻找证据,那么这些证据的真实性如何?结论是否可信?患者的情况又是否适用呢?

PICO 问题:

P:70 岁老年女性,患有高血压

I:服用钙通道阻滞剂(治疗高血压的一线药物)

C:相对于未服用钙通道阻滞剂

O:副作用是发生癌症

寻找证据:

我们以“钙通道阻滞剂”和“癌症”为检索词进行检索,检索到病因或治疗副作用相关的文献后,首先阅读题目和摘要,根据调查对象特征、研究问题、研究设计方案等选择合适的文献,如有多篇关于钙通道阻滞剂与癌症风险关联研究的文献,选择大样本、多中心、前瞻性的研究。经过检索、查阅和评价文献,最终发现目前尚没有关于“钙通道阻滞剂”与“癌症”风险相关的指南或系统综述,单个研究报告中有 2 个研究样本量较大、设计较严谨。分别对这两篇文献进行评价。

在阅读本章前请先阅读第二篇第一章“病因及危险因素”,有助于深入理解。

第一节 前瞻性队列研究原文剖析

一、文献题目和来源

题目:老年人群服用钙通道阻滞剂和癌症发病率的研究

来源:The Lancet, 1996, 348:493-497

二、研究目的与设计方案

研究利用 EPESE 老年人群队列数据(the Established Populations for Epidemiologic Studies of the Elderly),通过平均 3.7 年的队列追踪,分析长期服用钙通道阻滞剂是否增加癌症发生的风险。

设计方案(design):见图 3-3-1。

三、研究场所

美国马萨诸塞州波士顿地区、爱荷华州部分农村地区和华盛顿地区、康涅狄格州纽黑文市。

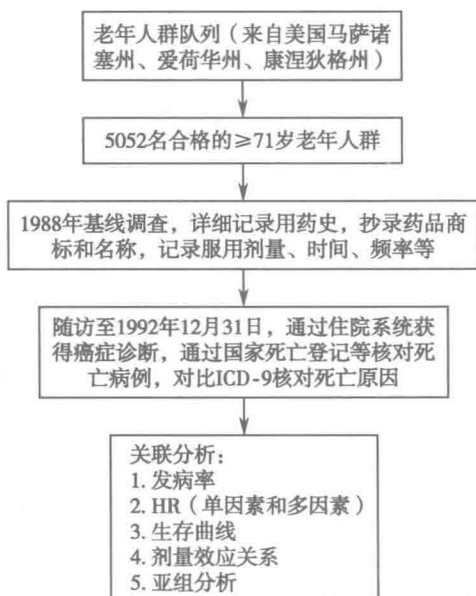


图 3-3-1 设计流程图

四、调查对象

1. 诊断标准 1988 年开始调查至 1992 年 12 月 31 日期间，所研究队列人群中，医院病例诊断为癌症或死亡原因按照 ICD-9 编码为 140~208 的，判断为阳性结局。其中，病例诊断依据医院住院病例的信息，个体均有来自 Medicare 管理系统的出院诊断。死亡病例诊断依据来自于亲戚随访、死亡讣告和国家死亡登记等。

2. 纳入标准 EPESE 队列建立于 1982 年，在美国马萨诸塞州波士顿地区、爱荷华州农村和华盛顿地区通过整群抽样，对 65 岁及以上老年人群进行调查。同期在康涅狄格州纽黑文市通过分层随机抽样，获得社区老年样本人群。对其中基线资料完整的人群进行随访。

3. 排除标准 资料不完整、1988 年前已经死亡、既往患有癌症、现服用抗癌药物者。

4. 样本量 未报告样本量的计算方法及是否满足统计学要求，仅报告纳入 5052 例老年人。

五、暴露(或干预措施)和结局

详细询问调查对象药物服用情况，药品种类通过查看药瓶记录并将商品名转化为药物化学名。记录服用钙通道阻滞剂的类别、剂量和服用时间。

主要结果的测量指标：癌症的发病率，风险比(hazard ratio, *HR*)，生存曲线，剂量效应关系，不同钙通道阻滞剂与不同癌症类别风险的 *HR*。

六、主要统计方法

绘制 Kaplan-Meier 生存曲线，Cox 回归模型进行多因素分析。

七、主要研究结果

1. 调查对象基本情况 共纳入 5052 名调查对象，其中服用钙通道阻滞剂的 451 人，未服用者 4601 人。对两组调查对象的基本情况进行分析，可见两组人群年龄、性别、人种、血压、体重指数、吸烟、饮酒等无差异，但在患有冠心病、脑卒中、高血压、肢体残疾、入院次数等身体健康状况方面有差异。结果见表 3-3-1。



表 3-3-1 调查对象一般情况

特征	未服用钙通道阻滞剂 (n=4601)	服用钙通道阻滞剂 (n=451)	P
人口学特征			
平均年龄	79.3 (0.1)	79.0 (0.3)	>0.1
女性	64.5%	64.1%	>0.1
人种			>0.1
白人	94.2%	93.6%	
黑人	5.1%	6.4%	
其他	0.7%	0	
共病情况			
冠心病	21.9%	71.2%	<0.0001
心衰	9.7%	29.3%	<0.0001
高血压	81.7%	90.0%	<0.0001
脑卒中	9.2%	12.6%	0.016
糖尿病	15.7%	26.2%	<0.0001
服药情况			
β 受体阻滞剂	11.6%	23.3%	<0.0001
ACEI	4.4%	8.6%	<0.0001
利尿剂	33.5%	50.6%	<0.0001
地高辛	11.%	25.1%	<0.0001
硝酸盐类	6.2%	38.8%	<0.0001
非甾体抗炎药	12.5%	14.6%	>0.1
阿司匹林	29.7%	37.3%	<0.0001
雌激素类	0.8%	0.2%	>0.1
糖皮质激素	3.2%	3.8%	>0.1
香豆素	2.6%	6.0%	<0.0001
收缩压	136 (0.4)	138 (1.3)	>0.1
舒张压	73 (0.2)	72 (0.6)	0.033
体重指数	24.4 (0.1)	24.8 (0.3)	>0.1
吸烟状况			>0.1
现在不吸烟	90.1%	91.8%	
1~19 支 / 天	4.9%	5.3%	
≥ 20 支 / 天	3.8%	2.7%	
饮酒 ≥ 28.35g/d	5.0%	4.4%	>0.1
肢体残疾	45.4%	62.7%	<0.0001
入院次数	1.2 (0)	2.2 (0.1)	<0.0001

ACEI: 血管紧张素转换酶抑制剂

2. 癌症发病率和死亡率 自 1988 年纳入队列开始随访,至病例诊断发生癌症、因癌症造成的死亡或随访至 1992 年 12 月 31 日为生存时间,平均随访 3.7 年,共计 18 774 人年。发生癌症 420 例,癌病发病率为 22.4/1000 人年。其中 169 名调查对象死于癌症,死亡率为 9/1000 人年。

3. Cox 模型分析 如果把不良事件编码为 1,则 $HR>1$ 是危险因素, $HR<1$ 是保护因素。结

学
习
笔
记



果如表 3-3-2 所示, HR 为 1.72, 即服用钙通道阻滞剂者发生癌症的风险是未服用者的 1.72 倍 (95%CI: 1.27~2.34, $P=0.0005$)。对于不同的钙通道阻滞剂, 服用维拉帕米和硝苯地平两类药物发生癌症的风险比未服用钙通道阻滞剂者高, 差异有统计学意义。而服用地尔硫草类药物与癌症风险无统计学关联。

表 3-3-2 服用钙通道阻滞剂与发生癌症的关系

	阳性 结局	人年数	人 / 每 1000 人年	未调整		调整多因素后 *	
				HR (95%CI)	P	HR (95%CI)	P
全部人群							
未服用钙通道阻滞剂	373	17 225	21.7	1.00(参考值)	—	1.00(参考值)	—
服用钙通道阻滞剂	47	1 549	30.3	1.42 (1.05~1.92)	0.032	1.72 (1.27~2.34)	0.0005
按钙通道阻滞剂分类							
维拉帕米	18	420	42.9	2.03 (1.26~3.25)	0.004	2.49 (1.54~4.07)	0.0002
硝苯地平	16	480	33.3	1.56 (0.94~2.51)	0.083	1.74 (1.05~2.88)	0.031
地尔硫草	13	641	20.3	0.94 (0.54~1.63)	>0.1	1.22 (0.70~2.12)	>0.1

* 调整性别、年龄、民族、吸烟、饮酒、是否心衰、入院次数等因素

4. 不同癌症类别 作者进一步分析了服用钙通道阻滞剂与发生不同癌症类别的关系, 结果见表 3-3-3。可见, 服用钙通道阻滞剂与增加子宫 / 卵巢癌、淋巴 / 血液系统癌症风险有统计学意义, 分别为 3.69 (95%CI: 1.22~11.14) 和 2.57 (95%CI: 1.13~5.83)。对于胃癌、结肠癌、直肠癌、乳腺癌、前列腺癌、泌尿系统癌症等, 服用钙通道阻滞剂增加其发生风险, 但结果没有统计学差异。对于肝 / 胆 / 胰腺癌、皮肤癌, 服用钙通道阻滞剂与其发生风险的 HR 接近 1。对于肺癌, 服用钙通道阻滞剂与其发生风险 $HR < 1$, 但没有统计学差异。不同癌症类别结果不一致, 有可能是服用钙通道阻滞剂对不同癌症类别的真实作用效果不同, 但也可能由于追踪时间短, 样本量小, 结果不稳定, 仍需要长时间的追踪和大样本的数据来证实。

表 3-3-3 服用钙通道阻滞剂与发生不同癌症类别的关系

癌症类别 (ICD-9 编码)	阳性结局	HR (95%CI)
胃癌 (151)	13	3.64 (0.96~13.76)
结肠癌 (153)	65	1.98 (0.90~4.38)
直肠癌 (154)	23	1.32 (0.31~5.74)
肝 / 胆 / 胰腺癌 (155~157)	24	1.15 (0.26~4.96)
肺癌 (162)	56	0.21 (0.03~1.52)
皮肤癌 (172~173)	14	1.11 (0.14~8.62)
乳腺癌 (174)	31	1.65 (0.49~5.55)
子宫 / 卵巢癌 (182~183)	23	3.69 (1.22~11.14)
前列腺癌 (185)	58	1.99 (0.93~4.27)
膀胱 / 尿道 / 肾癌 (188~189)	38	1.57 (0.55~4.47)
淋巴 / 血液系统癌 (200~208)	46	2.57 (1.13~5.83)

5. 剂量效应关系 危险因素的剂量效应关系是指其致病效应与有关危险因素的剂量或暴露时间具有显著的相关性, 即当病因可以分级处理时 (根据量化或特征), 随着级别的变化, 可以影响疾病在人群中的发病率或死亡率的变化。本研究也发现, 将服用钙通道阻滞剂的剂量分为低剂量组、中剂量组、高剂量组, 相对于未服用钙通道阻滞剂者, 其发生癌症的风险呈现明显的



剂量效应关系(图 3-3-2)。

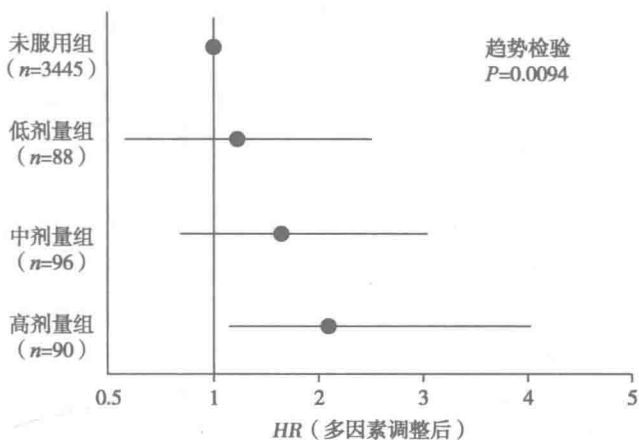


图 3-3-2 服用不同剂量钙通道阻滞剂与发生癌症风险的关系

6. 亚组分析 作者根据有可能影响癌症发生的因素进行分组,分析服用钙通道阻滞剂与发生癌症风险的关系。结果显示,相对于未服用钙通道阻滞剂者,71~79 岁年龄组和 ≥ 80 岁年龄组的 HR 分别为 1.77(95%CI: 1.20~2.63)和 1.55(95%CI: 0.93~2.59);男性和女性两组 HR 分别为 1.64 (95%CI: 1.09~2.46)和 1.81(95%CI: 1.13~2.90);白种人和黑种人 HR 分别为 1.60(95%CI: 1.15~2.22)和 3.42 (95%CI: 1.25~9.39);现在吸烟和现在不吸烟两组 HR 分别为 1.44 (95%CI: 0.55~3.73)和 1.78 (95%CI: 1.29~2.47);饮酒量 $<28.35\text{g/d}$ 和饮酒量 $\geq 28.35\text{g/d}$ 两组的 HR 分别为 1.76 (95%CI: 1.27~2.42)和 1.26(95%CI: 0.42~3.73);高血压组和未患高血压组的 HR 分别 1.71(95%CI: 1.23~2.37)和 1.99(95%CI: 0.78~5.07)。亚组分析结果进一步印证了服用钙通道阻滞剂增加癌症发生的风险。

八、结 论

老年人群中,长期服用钙通道阻滞剂者发生癌症的风险是没有服用者的 1.72 倍,且存在明显的剂量效应关系。

第二节 前瞻性队列研究文献评价

(一) 调查对象是否有明确的定义,各组之间除了暴露因素外,其他重要的方面是否相似、可比

是的。研究基于既往构建的老年队列人群,有明确的纳入排除标准。新发癌症病例的诊断和因癌症死亡病例的诊断有严格的依据,所研究病例均为队列期间新发病例,有统一的起始时间。

根据是否服用钙通道阻滞剂将研究人群分为两组。分析显示,两组人群年龄、性别、人种、血压、体重指数、吸烟、饮酒等无差异,但在患有冠心病、脑卒中、高血压、肢体残疾、入院次数等身体健康状况方面有差异。同时作者进行了亚组分析,对于不同身体健康状况亚组人群进行分析,结果显示,服用钙通道阻滞剂能显著增加癌症发生风险。

(二) 试验组和对照组有关暴露和结局的测量方法是否相同——是否采用盲法或客观的方法测量暴露和结局

本文中两组暴露的测量,均为调查员详细询问调查对象药物服用情况,药品种类通过查看药瓶记录并将商品名转化为药物化学名。详细记录钙通道阻滞剂类别、服用剂量和服用时间。根据是否服用钙通道阻滞剂将调查对象分为两组。



结局为癌症的发生或因癌症导致的死亡,是客观测量的指标。

未采用盲法。因此,在本研究中,观察者可能不知不觉地更关心服用钙通道阻滞剂的患者,更主动、更详尽地随访追查是否发生了癌症。同样,发生了副反应的患者更关注自己的病况,更容易回忆过去曾经接触的暴露因素。如果观察者和患者都不知道研究假设、不知道研究分组的情况,盲法如实记录观测结果,就能避免上述偏倚造成的影响,增加结果的真实性。

(三) 随访时间是否足够长、随访是否完整

该研究是前瞻性队列研究,随访时间从 1988 年开始调查至诊断癌症至因癌症死亡或随访至 1992 年 12 月 31 日,平均随访了 3.7 年,共发生 420 例癌症。在服用钙通道阻滞剂的人群中,共追踪了 1549 人年,发生 47 例癌症。若本研究仅随访几周,得到阴性结果,则无法区分阴性结果的真实性,是钙通道阻滞剂的确没有增加癌症的风险?还是随访期短、癌症还没有表现出来?

队列研究中,随访丢失的病例最好不应超过总观察人数的 10%。一旦超过 20%,则结果很可能失去真实性,因为中途退出的病例可能在某些重要特征上与仍然留在研究中的病例有很大的差别。本研究采用的队列人群是既往有良好基础的队列人群,且通过住院信息和死亡登记系统等获得调查对象诊断癌症或因癌症死亡的信息。但需要注意的是,尽管研究者尽可能采取各类措施获得阳性结局的资料,但仍有可能存在缺失,如部分癌症患者未去医院就诊,或死亡登记不准确等。

(四) 是否因在先,果在后,满足因果时间顺序

本文采用了队列研究的设计类型,即按照暴露情况进行分组,前瞻性追踪观察结局的发生,病因发生在前,疾病结局发生在后,因果现象发生的时间顺序是合理的。且本文在调查对象的排除标准中,去掉了既往患有癌症的对象。即研究中观察到的癌症病例,均为新发病例,尽可能排除了“因”和“果”同时存在而无法判断时序性的情况。

(五) 有无剂量效应关系

有。随着每日服用钙通道阻滞剂剂量的增加,发生癌症的风险逐渐增加。

(六) 有无在其他不同研究中反映因果联系的一致性

钙通道阻滞剂增加患癌症危险的病因学研究,已有多队列研究和病例对照研究,但结论不尽一致,说明钙通道阻滞剂与癌症的关系尚不明确。而本文为两者间具有相关性的假说提供了支持性证据。

(七) 生物学合理性

本文作者提出这样的假设:跨膜的钙信号可以激活细胞凋亡,细胞内钙的增加可以刺激细胞分化,钙通道阻滞剂可以抑制钙的跨膜转运并进而抑制细胞的凋亡和分化,因而促进癌症的发生。并引用多篇既往研究文献证据予以支持。

(八) 因果联系强度的大小

评价因果联系强度的指标越远离 1,则因果关联强度越大。本研究中,服用钙通道阻滞剂与增加癌症风险的 $HR=1.72$,其意义为服用钙通道阻滞剂者发生癌症的风险是未服用者的 1.72 倍。

比较调整前后的 RR 或 OR 值,可在一定程度上辅助判断:如果在调整混杂因素后, RR 、 OR 或 HR 值大幅下降,那么调整前后的 RR 或 OR 值都不可信。本研究在调整多项可能的混杂因素后, HR 值增至 1.72,说明这些可能的混杂因素并没有对结果造成明显的影响。

(九) 研究结果的精确性

RR 、 OR 或 HR 值的 95%CI 即反映了研究结果的精确性。如果 95%CI 范围较狭小、下限和上限值不包括 1.0,则其精度高,有统计学意义。本文中报告了 HR 的精确性,调整前 HR 为 1.42 (95%CI: 1.05~1.92),多因素调整后 HR 为 1.72 (95%CI: 1.27~2.34)。

(十) 我们自己的患者是否与文献报道的患者非常不同

本文篇首提出临床问题的患者为老年女性,年龄 70 岁,诊断为高血压,该文献为基于美国



70 岁以上老年人群的数据,患者与文献报道的病例临床背景相似,可以采用文献的结果。但也要考虑到人种的差异,是否美国老年人群与中国老年人群在基因、环境(包括膳食文化等)有所不同,仍需谨慎地下结论。

(十一) 患者有可能的收益或风险

本研究中,治疗组癌症发生率为 3.03%,对照组癌症发生率为 2.17%。

$$\text{即: } AR=3.03\%-2.17\%=0.86\%$$

$$NNH=1/0.86\%=116$$

表示需要 116 人服用钙通道阻滞剂才导致额外 1 例癌症的发生。说明服用钙通道阻滞剂导致癌症发生的这一副作用较小。

本文未提供服用钙通道阻滞剂对于降低总死亡和心脑血管疾病死亡的数据,因此查阅文献,2005 年 *Journal of Hypertension* 发表的随机对照试验研究显示,在我国 50~79 岁的中老年人群中,服用维拉帕米能使 2 年死亡率从 3.1% 下降至 2.3%,心脑血管疾病死亡从 2.1% 下降至 1.5%。

$$\text{即: } AR(\text{总死亡})=0.8\%$$

$$NNT(\text{总死亡})=0.6\%$$

$$AR(\text{心脑血管疾病死亡})=1/0.8\%=125$$

$$NNT(\text{心脑血管疾病死亡})=1/0.6\%=167$$

表示治疗 125 例患者可挽救 1 例死亡;治疗 167 例患者可挽救 1 例心脑血管疾病死亡。

AR、NNH、NNT 的定义详见本书第二篇第一章。

(十二) 研究结果是否有助于治疗方案的制订,是否有助于对患者作出解释

通过本文的剖析,对于该患者提出的“服用钙通道阻滞剂是否增加患癌症的危险”这一问题进行解答。目前尚没有定论,但根据这篇文献的分析结果,服用钙通道阻滞剂会增加患癌症的风险。且目前所面临的患者为老年女性,70 岁,与本研究人群较为相似。但仍需注意人种的差异,且需要考虑这名老年女性患者,若不服用钙通道阻滞剂降压,有无其他可替代的降压治疗方案?若有,替代的治疗方案是否会增加其他风险?仍需结合患者的实际情况综合考虑。

第三节 前瞻性队列研究临床思维与决策

本文为一篇设计规范的病因研究报告,采用队列研究的设计类型,利用来自美国三个地区的数据,5052 例调查对象的大样本研究,平均 3.7 年的追踪时间,且采用了规范的生存分析统计方法。但研究存在一些不足,如服用钙通道阻滞剂这一暴露的数据为基线调查的资料,未对随访期间服药情况进行追踪;结局诊断仅依据医院或死亡登记,会造成漏报;老年人群代表性不足,发生癌症的风险高于一般人群;是否有其他可能的混在因素,会影响服用钙通道阻滞剂与发生癌症的风险,如血压水平等。

第四节 病例对照研究原文剖析

一、文献题目和来源

1. 题目:钙通道阻滞剂和癌症风险
2. 来源:JAMA, 1998, 279: 1000-1004

二、研究目的及设计方案

采用病例对照的研究设计,利用包含 9513 名病例和 6492 名对照的美国中老年人数据,分

析服用钙通道阻滞剂是否增加癌症发生的风险。

自 1976~1996 年,在美国马萨诸塞州、纽约、费城、巴尔的摩四个城市,由经过专门培训的护士访问员,对因癌症第一次入院或非肿瘤入院的病例进行详细的问卷调查。考虑到 1980 年及以前钙通道阻滞剂的使用较少,且从服用药物至引发癌症发生需要一段时间,本文将调查对象限制在 1983~1996 年。此外,40 岁以下人群较少服用钙通道阻滞剂,作者又进一步将调查对象限制为 ≥ 40 岁人群。

三、调查对象

1. 病例的选择标准 1983~1996 年间,40~69 岁,患有原发性癌症,且首次诊断时间在问卷调查之前,既往无癌症史(除外非黑色素皮肤癌)。共纳入 9513 名病例。

2. 对照的选择标准 1983~1996 年间,40~69 岁,因为非恶性情况入院,且入院原因与使用降压药物无关的对象。排除既往有癌症史(除外非黑色素皮肤癌)、首次诊断与降压药物有关如因心血管疾病入院、或二次诊断疾病与降压药物有关等。共纳入 6492 名对照。

3. 诊断标准 病例和对照的诊断均来自于医院病例信息或病理报告,且诊断信息的获取采用了盲法,即调查人员不知道调查对象是否暴露于服用钙通道阻滞剂的情况。

4. 样本量研究 未报告样本量的计算方法及是否满足统计学要求,仅报告所纳入研究的病例组为 9513 名,对照组为 6492 名。

四、干预措施

研究者详细询问了调查对象在入院前规律服用药物的情况。对于每一种药物,记录药名、服用起始时间、服用年数等。研究者同时记录了钙通道阻滞剂、 β 受体阻滞剂和 ACEI 类药物的服用情况。考虑到药物诱发肿瘤发生需要一段时间,研究者按照药物服用时间(<1 年和 ≥ 1 年)分别进行分析。

五、主要统计方法

采用非条件多元 logistic 回归调整可能的混在因素后,计算服用钙通道阻滞剂与癌症风险的 OR 值。对于直肠癌等男性女性均可能患有的癌症,则采用全部人群进行分析。对于前列腺癌等仅有男性患有的癌症,病例和对照组仅选择男性人群;而对于乳腺癌等仅有女性患有的癌症,病例和对照组仅选择女性人群进行分析。

六、主要研究结果

1. 调查对象基本情况 9513 名病例,平均年龄为 56.0 岁,41.0% 为男性,84.0% 为白种人,29.0% 现在吸烟,27.0% 肥胖(体重指数 $\geq 28.0\text{kg/m}^2$),23% 自我报告既往有服用降压药物史。6492 名对照,平均年龄为 52.0 岁,42.0% 为男性,68.0% 为白种人,32.0% 现在吸烟,35.0% 肥胖,23.0% 自我报告既往有服用降压药物史。其入院原因包括:2448 例消化道或尿道问题(胆石症、尿石症、胃溃疡、十二指肠溃疡等),2075 例急性病症(急性感染、阑尾炎、创伤等),1969 例其他问题(骨关节病、良性肿瘤、盆腔炎、疝修补等)。

2. 病例组和对照组服用药物情况及 OR 值 对于服用钙通道阻滞剂时间 ≥ 1 年的人群,其平均服药时间在病例组和对照组分别为 3.8 年和 3.7 年,其对发生癌症的 OR 值为 1.1,而服用 β 受体阻滞剂和 ACEI 对发生癌症的 OR 值也为 1.1。同样,对于服用药物时间 <1 年的人群,三种药物与癌症的 OR 也无统计学意义(表 3-3-4)。

3. 不同影响因素下服用钙通道阻滞剂与癌症的风险 在不同性别、年龄、服药时间等情况下,分析服用钙通道阻滞剂与癌症的风险,OR 值均在 1.0 左右,无统计学意义(表 3-3-5)。



表 3-3-4 病例组和对照组服用钙通道阻滞剂、β 受体阻滞剂、ACEI 类药物情况

服用药物情况	钙通道阻滞剂		β 受体阻滞剂		ACEI	
	病例	对照	病例	对照	病例	对照
从不服用	8 855	6 011	8 238	5 769	9 047	6 163
服用时间 <1 年	481	352	1 059	591	348	247
OR (95%CI)	1.1 (0.9~1.3)		1.1 (1.0~1.3)		1.1 (0.9~1.3)	
服用时间 ≥ 1 年	162	121	151	102	102	74
OR (95%CI)	1.0 (0.8~1.3)		1.0 (0.8~1.4)		1.2 (0.9~1.7)	
服用时间不清楚	15	8	65	30	16	8

表 3-3-5 服用钙通道阻滞剂与癌症的风险

特征	病例	对照	OR (95%CI)
男性	266	162	1.2 (0.9~1.5)
女性	215	190	1.0 (0.8~1.2)
年龄			
40~64	328	281	1.1 (0.9~1.3)
65~69	153	71	1.0 (0.7~1.3)
服药时间			
<1 年	22	21	0.8 (0.4~1.5)
1~4 年	309	232	1.1 (0.9~1.3)
≥ 5 年	147	97	1.2 (0.9~1.5)
地尔硫草			
曾经服用	142	105	1.0 (0.8~1.4)
服用时间 ≥ 5 年	44	35	1.0 (0.6~1.6)
硝苯地平			
曾经服用	163	132	1.0 (0.8~1.3)
服用时间 ≥ 5 年	53	31	1.5 (0.9~2.3)
维拉帕米			
曾经服用	172	111	1.2 (0.9~1.5)
服用时间 ≥ 5 年	51	31	1.1 (0.7~1.8)

4. 不同癌症类别 按照癌症发生部位和类别,分析服用钙通道阻滞剂、β 受体阻滞剂和 ACEI 类药物与癌症的风险。结果显示,在 48 个 OR 值中,仅有 5 个有统计学意义,分别是肾癌,与钙通道阻滞剂 OR 为 1.8 (95%CI:1.1~2.7),与 β 受体阻滞剂 OR 为 1.8 (95%CI:1.3~2.5),与 ACEI 的 OR 为 1.9 (95%CI:1.2~3.0);呼吸道癌(非肺部),与 β 受体阻滞剂 OR 为 2.4 (95%CI:1.4~4.3);子宫癌,与 ACEI 的 OR 为 2.1 (95%CI:1.2~3.7)。而其余癌症与服用三种药物的 OR 值无统计学意义。而既往有多个研究结果显示,肾癌与高血压或降压药物相关。此外,作者对服药时间 ≥ 5 年的调查对象进行分析,结果显示钙通道阻滞剂与癌症风险的 OR 值均为 1.0 左右,即服用钙通道阻滞剂与增加癌症的风险无关(表 3-3-6)。

七、结 论

本研究结果显示,除了肾癌,服用钙通道阻滞剂与癌症总体发生风险或单类别癌症发生风险无关。

学
习
笔
记



表 3-3-6 服用药物情况与不同癌症类别的 OR 值

癌症类别	人数	钙通道阻滞剂		β 受体阻滞剂		ACEI	
		服 药 人数	OR (95%CI)	服 药 人数	OR (95%CI)	服 药 人数	OR (95%CI)
乳腺	2893	92	1.1 (0.8~1.4)	264	1.1(0.9~1.3)	75	1.0(0.8~1.4)
结肠	1004	46	0.9 (0.7~1.3)	124	1.1(0.9~1.4)	36	1.0 (0.7~1.5)
肺	994	35	0.9 (0.6~1.4)	91	1.0(0.8~1.4)	17	0.8 (0.5~1.5)
前列腺	823	110	1.3 (0.9~1.7)	140	1.2(0.9~1.6)	89	1.2 (0.9~1.7)
恶性黑色素瘤	597	16	1.6 (0.8~3.0)	67	1.2(0.9~1.7)	7	1.3 (0.5~3.1)
直肠	490	19	1.0 (0.6~1.7)	63	1.4(1.0~1.9)	14	1.0 (0.6~1.8)
膀胱	364	27	1.3 (0.8~2.0)	46	1.2(0.8~1.7)	20	1.4 (0.8~2.4)
胰腺	349	20	1.1 (0.7~1.8)	43	1.3(0.9~1.8)	14	1.1 (0.6~1.9)
卵巢	282	12	1.1 (0.6~2.0)	27	1.0(0.7~1.6)	10	1.2 (0.6~2.4)
肾脏	279	31	1.8 (1.1~2.7)	49	1.8(1.3~2.5)	22	1.9 (1.2~3.0)
尿道	264	19	1.0 (0.6~2.7)	36	1.1(0.7~1.7)	23	2.1 (1.2~3.7)
淋巴系统	217	5	0.4 (0.2~1.1)	16	0.7(0.4~1.2)	4	0.5
白血病	205	13	1.2 (0.6~2.2)	24	1.3(0.8~2.1)	2	0.2
食管	144	12	1.8 (0.9~3.5)	7	0.5(0.2~1.0)	1	0.2
胃	141	5	0.8 (0.3~2.1)	13	1.0(0.5~1.8)	2	0.5
骨关节	114	2	0.4	10	0.9(0.4~1.7)	3	0.7
呼吸道(非肺)	98	5	1.7 (0.6~4.7)	18	2.4(1.4~4.3)	2	1.3
胆囊	90	5	1.0 (0.4~2.7)	9	1.0(0.5~2.0)	5	1.6 (0.6~4.1)
睾丸	43	0	0	2	1.1	1	1.2
子宫	40	0	0	2	0.5	0	0
甲状腺	33	2	1.5	1	0.4	0	0
肝	29	2	1.0	5	1.5(0.6~4.3)	1	0.7
小肠	20	3	2.9	2	0.9	0	0

第五节 病例对照研究文献评价

(一) 调查对象是否有明确的定义,各组之间除了暴露因素外,其他重要的方面是否相似、可比

是。本研究中病例和对照的选择有明确的纳入排除标准,病例的诊断均来自于医院病例信息或病理报告。

研究者对于病例组和对照组的基本特征进行了报告,包括平均年龄、性别、人种、现在吸烟状况、肥胖状况、既往服用降压药物史等,但研究者未对两组人群进行统计学检验。读者也可以自行进行统计学 t 检验(针对数值型变量)或 χ^2 检验(针对分类变量)。若两组人群在重要指标上有差异,则有可能影响服用钙通道阻滞剂与癌症的风险。需要在结果分析时加以注意,可采用分层或多因素调整的方法,以去掉这些指标的影响。例如,本研究中报告了服用钙通道阻滞



剂与肾癌发生相关,但既往也有研究指出患高血压的人更容易得肾癌,那么在本研究中,若病例组和对照组人群中患高血压的比例不同,则无法判断服用钙通道阻滞剂人群增加的癌症患病率是钙通道阻滞剂引起的还是高血压引起的。

(二) 试验组和对照组有关暴露和结局的测量方法是否相同——是否采用盲法或客观的方法测量暴露和结局

本文中暴露的测量为调查员询问调查对象在入院前规律服用药物的情况。对于每一种药物,记录药名、服用起始时间、服用年数等。病例和对照按照明确的标准选择,且诊断信息均来自于医院病例和病理报告。

诊断信息的获取采用了盲法,即调查人员不知道调查对象是否暴露于服用钙通道阻滞剂的情况。因此避免了调查者偏倚,结果真实性更高。而上一篇文献中,未采用盲法,则有可能使观察者更关心服用钙通道阻滞剂的患者,更主动、更详尽地随访追查是否发生了癌症,导致结果不真实。

(三) 随访时间是否足够长、随访是否完整

本研究是病例对照研究,也叫回顾性研究,因此不涉及随访时间这一问题。

(四) 是否因在先、果在后,满足因果时间顺序

本文采用了病例对照的设计类型,即按照是否患有癌症进行分组,回顾性追踪疾病发生前有可能的暴露情况,不符合因果现象发生的时间顺序。因此,病例对照研究设计的病因论证强度比前瞻性队列研究的论证强度低。

(五) 有无剂量效应关系

研究者将服用钙通道阻滞剂的服用时间分为<1年、1~4年和≥5年三类,分别计算其与癌症发生风险的关系,结果显示,*OR*值均为1.0左右,无统计学意义,即服用钙通道阻滞剂与癌症发生风险无剂量效应关系。

但研究者未报告服用药物的剂量、服用频率等。若有,可补充分析这些信息与癌症风险有无剂量效应关系。

(六) 有无在其他不同研究中反映因果联系的一致性

钙通道阻滞剂增加患癌症危险这一问题,已有多个队列研究和病例对照研究,但结论不尽一致,说明二者关系尚不明确。上一篇文献的结论为支持二者的相关性提供了证据,而本文的结论指出,服用钙通道阻滞剂与癌症(除外肾癌)风险无关。同时,作者指出,上一篇文献中所得到的升高的*HR*值,有可能是由于发生的病例数较少,研究结果不稳定造成的,所发生癌症结局数目小的变动都可能会对*HR*造成较大的影响。而本文采用病例对照研究,纳入的癌症病例有9513名,因此研究结果更稳定。

(七) 生物学合理性

目前关于钙通道阻滞剂与癌症有关的理论依据多认为:钙通道阻滞剂可能干扰细胞的凋亡因而导致癌症的发生。但研究者也指出,目前这个理论仍缺乏确凿的证据支持。且有些研究指出,钙通道阻滞剂能降低致癌作用,减少癌症发生。因而两者间的确凿关系仍需要进一步的研究。

(八) 因果联系强度的大小

在病例对照研究中,评价因果联系强度的指标为*OR*。越远离1,则越有价值。本研究中,服用钙通道阻滞剂与增加癌症风险的*OR*为1.1(95%CI:0.9~1.3),即服用钙通道阻滞剂者发生癌症的风险是未服用者的1.1倍,但95%CI包含1,即该指标没有统计学意义,也就是说,本研究认为服用钙通道阻滞剂与癌症发生风险没有关联。

(九) 研究结果的精确性

本研究报告了*OR*值的95%CI。且相对于前一篇研究,本文样本量较大,可信区间较窄,即



精确性较高。

(十) 我们自己的患者是否与文献报道的患者不同

研究为基于美国 40~69 岁老年人群的数据,患者与文献报道的病例临床背景相似,可以采用文献的结果。但也要考虑到人种的差异,仍需谨慎地下结论。

(十一) 研究结果是否有助于治疗方案的制订,是否有助于对患者作出解释

通过本文的剖析,对于该患者提出的“服用钙通道阻滞剂是否增加患癌症的危险”这一问题进行解答,目前尚不能认为服用钙通道阻滞剂能够增加癌症风险。但同时需要指出,目前有多篇这方面的文献,结论不一致,且该研究为美国老年人群,目前尚未看到基于我国老年人群的质量较高的大规模数据分析。这名老年女性患者是否继续服用钙通道阻滞剂应结合其高血压患病和控制情况、自身健康状况、有无其他副作用或其他替代药物等综合情况作出判断。

第六节 病例对照研究临床思维与决策

本文为一篇设计规范的病因研究报告,采用病例对照研究的设计类型,利用来自美国四个地区的数据,纳入了包括 9513 名病例和 6492 名对照的人群,是较为少见的针对癌症的大样本研究,且采用了规范的统计分析方法。但研究存在一些不足,如未统计服用钙通道阻滞剂的频率和剂量这部分信息,不能完整地分析其与癌症的剂量效应关系;是否有其他可能的混在因素;对照组均为医院病人,可能会存在一些影响癌症发生的混杂因素,存在选择偏倚。若能在医院对照之外,再设置一组社区对照人群,即健康人群对照,增加对照的可比性。若两组对照结果分析均显示无关,则更加印证了服用钙通道阻滞剂与癌症发生风险无关。

病因学研究的目的是明确疾病发生的原因,掌握其发病机制,为正确的诊断、评估危害程度、有效地预防和治疗、控制和消灭疾病作出合理的医疗决策。当明确了疾病的病因(或治疗的副作用)后,就可采取有针对性的、有效的医疗决策,使患者获得最好的治疗效果或果断采取干预措施,降低发病率,减少社会负担。病例对照研究和队列研究都是病因学研究中最常用的研究设计类型。前者相对更省力、省钱、省时间,易于组织实施,特别适用于罕见病的研究。但存在一些缺点,如信息获取时难以避免回忆偏倚、暴露与疾病的时间先后难以判断,不适于研究人群中暴露比例较低的可能病因。而队列研究由于暴露资料的收集在结局发生之前,因此较好地避免了回忆偏倚,且时间顺序合理,验证病因假说的能力强。但缺点是耗费的人力、物力、财力和时间较多,容易发生失访偏倚,不适合于发病率低的疾病。

就本研究而言,鉴于两篇文献结论相反,目前的证据尚不能充分支持“服用钙通道阻滞剂增加患癌症的危险”这一假说。第一篇在 Lancet 发表的队列研究,采用前瞻性设计,符合因果论证的时间顺序。但研究中对暴露和结局的测量未采用盲法,有可能存在检出偏倚和错分偏倚;且研究对象年龄较大,发生癌症的风险高于一般人群。另外,由于癌症作为罕见事件结局,其 HR 增高有可能由于阳性病例数的结果不稳定造成。而第二篇发表在 JAMA 的病例对照研究,病例数较大,并获得了有可能的影响因素等信息。但也存在一些不足,包括对照组设置为住院病人而非健康对照人群,因此有可能造成病例组 and 对照组的差异被低估。两篇文献设计均存在不足,且由于研究人群特征不同,研究结果不一致。对于临床医生和研究者,一方面可以在这个领域组织开展或关注大规模多中心前瞻性的科学研究,进一步寻找确凿的证据;另一方面在临床实际工作中,注意观察服用钙通道阻滞剂病人的远期预后,尤其是癌症的发病和死亡是否增加,进一步积累病例资料。在临床决策上,通过计算得到 NNH 和 HHT,其利弊得失从目前的证据难以确定,故临床医生应谨慎决定患者是否应采用钙通道阻滞剂进行治疗。总之,在面临病因或治疗副作用的问题时,临床医生在医疗决策过程中,要坚持利大于弊的原则,结合循证医学的证据和临床工作经验,依据目前最新的临床实践指南,同时充分考虑患者的自身情况和意愿,再作出



最合适的临床决策。

(何耀 刘森)

参考文献

1. 李立明. 流行病学. 第6版. 北京:人民卫生出版社,2007.
2. 黄悦勤. 临床流行病学. 第3版. 北京:人民卫生出版社,2002.
3. 李立明. 临床流行病学(八年制). 第2版. 北京:人民卫生出版社,2011.
4. Sharon E. Straus. Evidence-Based Medicine:How to Practice and Teach It. 4th ed. New York:Churchill Livingstone,2010.
5. Rosenberg L,Rao RS,Palmer JR,et al. Calcium channel blockers and the risk of cancer. JAMA,1998,279(13):1000-1004.
6. Pahor M,Guralnik JM,Ferrucci L,et al. Calcium-channel blockade and incidence of cancer in aged populations. Lancet,1996,348(9026):493-497.
7. Liu L,Zhang Y,Liu G,et al. The Felodipine Event Reduction (FEVER) Study:a randomized long-term placebo-controlled trial in Chinese hypertensive patients. J Hypertens,2005,23(12):2157-2172.

学
习
笔
记

第四章 诊断试验的循证实践

临床问题

患者,男性,42岁,HBsAg、HBeAg、HBcAb均阳性,HBV-DNA 5×10^7 ,ALT、AST一直正常。来门诊询问是否需要抗病毒治疗。根据慢性乙型肝炎防治指南,尽管转氨酶正常,但如果肝活检检查结果有2级以上炎症或者肝纤维化,就应该抗病毒治疗。因此,临床医师建议患者先做肝穿刺活检检查。然而,患者不愿意做有创的肝活检,又因地处边远山区,希望能通过做无创性检查诊断肝纤维化。

PICO 问题:

P:慢性乙型肝炎患者

I:无创检查诊断

C:金标准(肝穿刺活检)

O:诊断肝纤维化

寻找证据:

可以参考第一篇第二章,根据临床问题获取最佳证据中的检索策略,寻找最恰当的文章。本例临床情境中,该医生首先查询了Cochrane图书馆的系统评价,没有查询到新近的相关系统综述。通过TRIP Database找到一篇采用FibroScan和生化指标组合的诊断方法代替肝穿刺活检,然而,因当地没有FibroScan仪器。因此,继续限定检索范围,关注用生化指标组合来诊断肝纤维化的文章。找到一篇2012年发表在国际肝脏病杂志(Hepatol Int)上的应用临床指标组合预测肝纤维化的论文。该诊断试验报告是否能用于该患者,首先要对该文章进行剖析和评价,其次参照国际诊断准确研究报告的标准(Standards for Reporting Diagnostic Accuracy, STARD)要求,对检索到的文献进行剖析,在阅读本章前请先阅读第二篇第二章“诊断试验”,以便于深入理解。

第一节 诊断试验原文剖析

一、文献题目和来源

题目:Establishment and validation of a simple noninvasive model to predict significant liver fibrosis in patients with chronic hepatitis B

来源:Hepatol Int, 2012, 6: 360-368

二、研究目的

建立包括临床常用生化指标的无创模型诊断乙肝患者肝纤维化诊断模型。

三、研究设计方案

与金标准(肝穿刺活检)进行对照,建立无创诊断模型,并进行内部和外部病例验证,研究设



计流程见图 3-4-1。

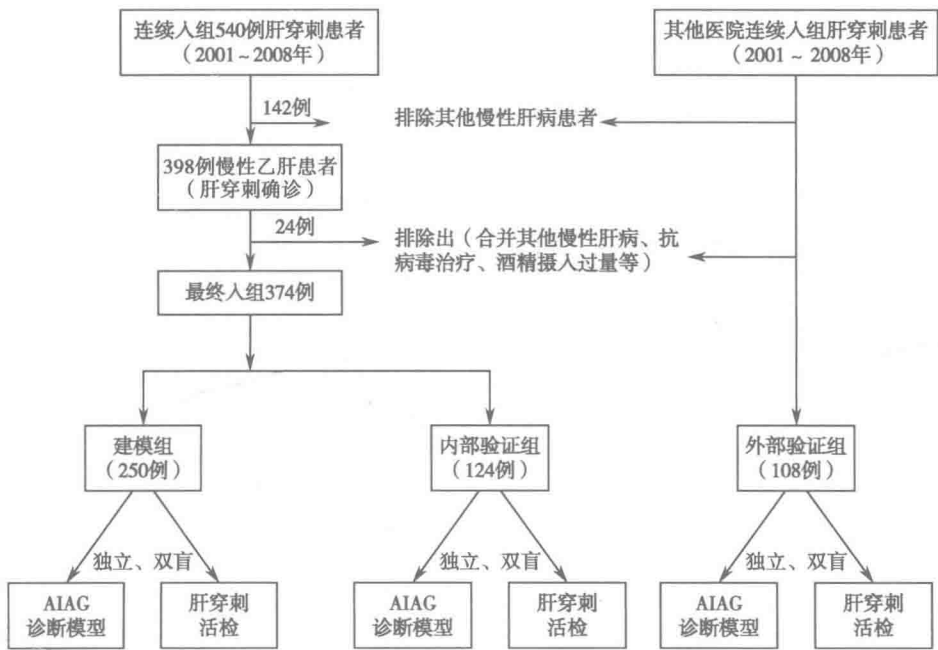


图 3-4-1 研究流程图

四、研究场所

我国三家三级甲等医院。

五、研究对象

1. 慢性乙型肝炎患者,HBsAg 阳性、ALT 异常 >6 个月。
2. 排除合并有其他慢性肝病、抗病毒治疗、酒精摄入过量等患者。

六、金标准的选择

肝穿刺活检。由病理医师对肝组织病理片评判实行盲法,METAVIR 纤维化评分系统,F2、F3、F4 均为显著肝纤维化。

七、主要研究结果

1. 在建模组 250 例研究对象中,通过单因素分析,比较“无显著肝纤维化组”与“有显著肝纤维化组”中,差异具有显著性的各项临床特征指标。纳入分析的指标包括性别、年龄、HBsAg、HBcAb、慢性乙型肝炎家族史、外周血白细胞计数、血红蛋白、血小板计数、血清丙氨酸氨基转移酶(ALT)、天门冬氨酸氨基转移酶(AST)、总胆红素(TB)、结合胆红素(CB)、 γ -谷氨酰转肽酶(GGT)、白蛋白、凝血酶原时间(PT)、凝血酶原时间国际标准化比值(INR)、HBV-DNA 和总胆固醇水平。将单因素分析中,两组比较 $P<0.5$ 的变量进入多元 Logistic 回归分析,构建诊断显著肝纤维化的回归模型。最终进入模型的变量为:年龄(Age)、凝血酶原时间国际标准化比值(INR)、白蛋白(Albumin)和 γ -谷氨酰转肽酶(GGT),取首字母,命名为 AIAG 乙肝肝纤维化诊断模型(AIAG 模型)。得到诊断慢性乙型肝炎患者显著肝纤维化公式 AIAG 如下:

$$P=-7+0.03 \times \text{年龄(岁)}+9 \times \text{INR}-0.08 \times \text{白蛋白(g/L)}+0.004 \times \text{GGT(U/L)}$$

$$\text{AIAG}=e^P/(1+e^P)$$



2. 绘制 AIAG 乙肝肝纤维化诊断模型的 ROC 曲线,计算 ROC 曲线下面积(AUC),选取最佳界值点。构建诊断试验四格表,计算 AIAG 模型在不同界值点时的敏感度、特异度、阳性预测值、阴性预测值以及似然比。确定 $AIAG < 0.32$ 和 > 0.72 为无或者有显著肝纤维化的截点,AIAG 在 $0.32 \sim 0.72$ 之间的患者,需要做肝穿活检进一步确定有无显著肝纤维化。结果见表 3-4-1。

表 3-4-1 AIAG 乙肝肝纤维化诊断模型在不同截点时的诊断性能比较

诊断试验(显著性肝纤维化)	总例数	肝穿刺金标准 (显著性肝纤维化)		Sen	Spe	PPV	NPV	+LR	-LR
		无(F0~1)	有(F2~4)						
		n	n						
低界值 0.32				96%	39%	66%	90%	1.59	0.09
无(≤ 0.32)	49	44	5						
有(> 0.32)	201	68	133						
		112	138						
高界值 0.72				50%	96%	93%	61%	11.2	0.52
无(≤ 0.72)	176	107	69						
有(> 0.72)	74	5	69						
		112	138						
最佳界值 0.60				67%	83%	83%	67%	3.97	0.39
无(≤ 0.60)	138	93	45						
有(> 0.60)	112	19	93						
		112	138						

Sen: 敏感度; Spe: 特异度; PPV: 阳性预测值; NPV: 阴性预测值; +LR: 阳性似然比; -LR: 阴性似然比

3. 对 AIAG 乙肝肝纤维化诊断模型进行内部验证与外部验证,一致性较好。

4. 对 3 组 482 例对象进行汇总,比较 AIAG 诊断模型与 APRI、Forns 指数以及 FIB-4 诊断肝纤维化性能的优劣。通过绘制不同乙肝肝纤维化诊断模型的 ROC 曲线,比较 ROC 曲线下面积(AUC),选取最佳诊断模型,见图 3-4-2。

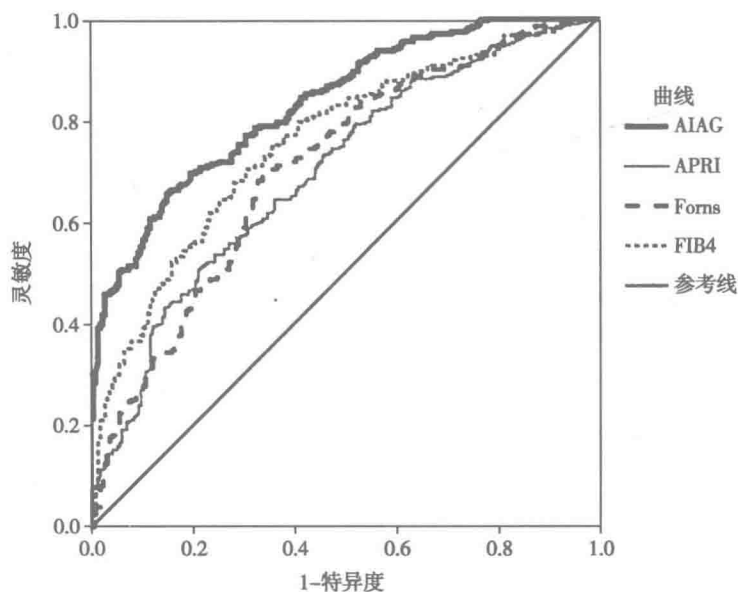


图 3-4-2 AIAG 诊断模型与 APRI、Forns 指数以及 FIB-4 诊断肝纤维化的 ROC 曲线比较



第二节 诊断试验文献评价

一、评价文章的科学性

对查询到相关文章进行科学性进行评价,只有设计合理,具有科学性,其结果才是可信的。对照诊断试验科学性的评判标准,该临床医生作出了如下评价:

(一) AIAG 诊断试验是否与合适的“金标准”进行了比较

本文献中,采用了合适的“金标准”——肝穿刺活检,由病理医师对肝组织病理片采用 METAVIR 评分系统,F2、3、4 均为显著肝纤维化。

(二) AIAG 试验与金标准试验的比较,是否采用了独立、“盲法”的方法

是的,在接受了金标准后,下一个问题是,同一患者诊断试验的结果和金标准试验结果的测定是否是互相独立进行评价的。本文献中,对肝组织病理片进行评判的病理医生是不知道患者的情况和化验结果的,采用了独立、“盲法”的评判。值得注意的是,虽然病理学诊断结果通常被认为是最佳的金标准,不同的病理学专家对 METAVIR 评分系统病理诊断结果可能也会存在测量变异(variation)。本案例中,组内所有的病理结果由同一位病理学家诊断,因此,避免了观察者之间的变异,因为即使是经验相同的病理科医师,在对一患者肝活检的病理诊断,诊断出 METAVIR 评分可能也不一致;也避免了由于研究病理科医师经验不同,影响到 METAVIR 评分,金标准结果可信。诊断试验所用的生化指标,有专职实验人员测定。AIAG 公式,则由医生代入公式计算。因此,操作者的技能与熟练程度可靠。

(三) 是否每个被测者都做肝活检进行评价

本文献中,试验将有或无显著肝纤维化的患者都与金标准肝活检结果进行了比较。

(四) 研究中纳入病例的选择是否有代表性

是,但可能存在一定的局限性。最理想的是研究样本与被测人群完全相同,这样的样本是有代表性的。但是,临床研究中常常很少有可能做到。本文献中,纳入的患者是连续入院的患者,研究者从 3 家三级甲等医院中接受肝穿刺患者的连续求诊者中得到样本。连续入选病例,避免了入组对象的在选择性偏倚(selection bias)。然而本文献中,患者样本的疾病谱可能存在局限性。因为诊断试验的目标人群为所有可能有肝纤维化的乙型肝炎患者,考虑到肝穿刺活检的创伤性和患者的依从性,接受肝穿刺患者的病情可能更重,患者的样本中许多病情较轻的患者被排除在外,这样的结果可能会夸大了试验的灵敏度。在研究流程图中,最好能够提供可能有肝纤维化的乙型肝炎患者的筛查人数,即有多少例符合肝穿刺入选标准而没有接受肝脏穿刺的患者例数。这部分患者的特征信息如何,没有阐述,存在局限性。另外,临床特征与生化指标为是否存在缺损值情况? 研究分析中,对缺损值是如何处理的? 生化指标测定时间点如何定义? 文章没有给出说明。

(五) 诊断试验是否有可重复性与可再现性

是。诊断试验的重复性(repeatability)是指对同一个患者,在同等条件下,试验结果是否可以重复,其差异值由最纯净的随机误差引起。再现性(reproducibility)是指试验在不同实验室,不同的观察者等任何可以预想的临床情况,对同一个患者,在同等条件下,试验结果是否可以重现。重复性与可再现性反映了诊断试验的稳定程度。本文献中,技术上可靠与精确。作者详细介绍诊断试验的具体实施操作方法,使其他研究有可能达到相同的实验条件,验证试验的可再现性。另外,诊断试验是用年龄(age)、凝血酶原时间国际标准化比值(INR)、白蛋白(albumin)和 γ -谷氨酰转肽酶(GGT)构建了 AIAG 乙型肝炎纤维化诊断模型。这些临床常用生化指标的测定都是可靠与精确的,具有可重复性。尤其值得肯定的是,研究者还提供了观察者



间(外部验证)以及观察者内(内部验证)其结果一致,证明了 AIAG 模型在内部和外部验证时均具有较高的稳定性。

总之,本案例中,AIAG 诊断试验基本满足上述 1~5 条科学性评价标准,其中第 4 条疾病谱可能存在的局限性,在循证诊断实践者使用该诊断试验,应予以注意。

二、估计临床应用的重要性

(一) 估计疾病的验前概率

验前概率(pre-test probability)是指患者在做诊断试验或检查前,患这种病的概率,可理解为诊断前对某病发生可能性的一个判断,临床上为该疾病的构成比,普查或筛检应以患病率为验前概率。因此,在临床实践中,验前概率随着就诊对象来源不同及医疗环境不同而有较大差别。验前概率通常从文献资料、本地的调查数据以及根据临床实践中遇到此类患者的概率估计得到。根据 2006 年全国乙型肝炎流行病学调查表明,我国 1~59 岁一般人群 HBsAg 携带率为 7.18%,现有的慢性 HBV 感染者约 9300 万人,其中慢性乙型肝炎患者约 2000 万例。慢性乙型肝炎患者中,约 50% 会发生肝纤维化。在本案例中,3 组 482 例患者,肝纤维化的验前概率 = $272/482 \times 100\% = 56\%$ 。因此,开发无创性诊断试验有社会价值与意义。

(二) 诊断试验的性能:灵敏度、特异度、似然比

本文献中,诊断试验性能参数汇总结果见表 3-4-2。可见,AIAG 模型用于预测显著性肝纤维化,具有较高的诊断价值。需要注意的是,灵敏度与特异度是评价诊断试验真实性的基本指标,是诊断性试验的固有特性,不受患病率影响。但是,敏感度、特异度会受研究对象的疾病谱中病情轻重的影响,如果有病组均为晚期患者,则敏感度会增高;如果无病组纳入的是正常人,则特异度增高。因此,本文献中,纳入病例的病情可能偏重,可能会存在 AIAG 灵敏度被高估,而特异度被低估的情况。另外,患病率对诊断试验的预测值等最后分析结果有很大的影响,本研究对象代表了三级医院的患者情况。

表 3-4-2 AIAG 诊断显著性肝纤维化的诊断试验结果

肝活检	AIAG	n(%)	灵敏度	特异度	PPV	NPV	+LR	-LR
F0~F1	<0.32	112(44.8%)	96%	39%	66%	90%	1.519	0.09
F2~F4	>0.72	138(55.2%)	50%	96%	93%	61%	11.2	0.52

(三) 诊断性试验正常值的确定是否合理、可靠

本研究用 ROC 曲线确定正常值的截点,有较高的正确性。试验组、内部验证组、外部验证组 ROC 曲线下面积值(95% 可信区间)分别为:0.842(0.795~0.888)、0.806(0.730~0.882)和 0.866(0.800~0.933),均超过 0.80,可以认为正常值的确定是合理、可靠的。

(四) 诊断性试验的临床实用性如何

在大样本的慢性乙肝患者中进行探索,以肝穿刺结果为金标准,从临床常用的常规指标中筛选出年龄(age)、凝血酶原时间国际标准化比值(INR)、白蛋白(albumin)和 γ -谷氨酰转肽酶(GGT)这 4 个最具诊断价值的常用指标,建立了 AIAG 这样一个指数模型,用于预测显著性肝纤维化,发现其具有较高的诊断价值:当 AIAG<0.32 时,阴性预测值为 90%;当 AIAG>0.72 时,阳性预测值为 93%。并且 AIAG 模型在内部和外部验证时均具有较高的稳定性,与其他的肝纤维化诊断模型相比具有更高的诊断准确性。

AIAG 诊断模型所用的指标均为临床常规的检测指标,更为简便易行,几乎所有的医院均能常态化检测,且诊断的准确性高,易于推广。AIAG 模型可使约 50% 慢性乙型肝炎的患者(AIAG<0.32 或 >0.72)避免进行肝穿刺活检,从而节约医疗资源。



三、将临床研究结果用于自己的患者

(一) 结果是否适用于并可提供给我自己的患者

是。回到临床情景中,这位 42 岁的男性患者的特征与查询到文献中的患者相吻合。在应用到患者时,需要在患者的知情同意下,考虑试验的可行性(feasibility)与可接受性(acceptability)。这些临床常用生化指标的测定在最基层的医院都能进行,简单易行,与肝穿刺活检相比,价格低廉,并不会给患者带来不适,患者也愿意接受检查。试验结果可以改变对患者的治疗策略,改善患者的预后。因此,可以给这位患者开出医嘱做 AIAG 诊断试验。

目前在复旦大学附属中山医院的官方网站上已经向患者免费提供该计算公式,只要输入以上四项指标就能得到其肝纤维化评分,十分方便。

(二) 诊断试验结果是否改变了对患病率的估计

本临床情景中,根据既往的临床经验,估计该患者肝纤维化的比例(验前概率)为 55%,根据贝叶斯定理(Bayes theorem),可以估计患者的验后概率。阴性试验结果的验后概率(阴性预测值)公式为:

$$\text{阴性预测值} = \frac{(1 - \text{患病率}) \times \text{特异度}}{(1 - \text{患病率}) \times \text{特异度} + \text{患病率} \times (1 - \text{灵敏度})}$$

阳性试验结果的验后概率(阳性预测值)公式为:

$$\text{阳性预测值} = \frac{\text{患病率} \times \text{灵敏度}}{\text{患病率} \times \text{灵敏度} + (1 - \text{患病率}) \times (1 - \text{特异度})} \times 100\% \quad (\text{公式 3-4-1})$$

根据公式 3-4-1,如果 $\text{AIAG} < 0.32$,患者没有肝纤维化的概率:

$$P > \frac{0.45 \times 0.39}{0.45 \times 0.39 + 0.55 \times (1 - 0.96)} \times 100\% > 88.9\% \quad (\text{公式 3-4-2})$$

根据公式 3-4-2,如果该 $\text{AIAG} > 0.72$,患者有肝纤维化的概率:

$$P > \frac{0.55 \times 0.5}{0.55 \times 0.5 + (1 - 0.55) \times (1 - 0.96)} \times 100\% > 93.9\%$$

患者的常规肝功能试验结果凝血酶原时间国际标准化比值(INR)为 1.15、白蛋白为 35g/L、 γ -谷氨酰转肽酶(GGT)为 50U/L,年龄为 42,代入复旦大学中山医院网站的 AIAG 公式计算后得到 $\text{AIAG} = 0.88 (> 0.72)$,因此该患者在没有计算 AIAG 前,他有 50% 的可能性有显著纤维化,而将他的肝功能检测指标代入公式后,其有显著纤维化的概率上升到 93%。

因此,是的,AIAG 诊断试验结果改变了对患者是否有肝纤维化概率的估计。

(三) 诊断试验结果是否改变了对患者的处理

是。AIAG 检测乙型肝炎患者肝纤维化的真实性与可靠性均较高。研究提供的诊断阈值和治疗阈值(图 3-4-3)。用 AIAG 诊断有无显著性肝纤维化。当 $\text{AIAG} < 0.32$ 时,有 90.7% 的概率患者没有显著性肝纤维化,患者有肝纤维化的可能性为 $\text{AIAG} > 0.32$ 人的 1/10;而如果 $\text{AIAG} > 0.72$ 时,有 93.6% 的可能性是显著性肝纤维化,即显著性肝纤维化的可能性为 $\text{AIAG} < 0.72$ 的 11 倍。



图 3-4-3 AIAG 诊断模型的诊断阈值与治疗阈值



因此,临床上认为,对 $AIAG < 0.32$ 的患者可以暂时不做肝活检,也不抗病毒治疗,可以随访观察; $AIAG > 0.72$ 的患者显著性肝纤维化可能性很高,因此也可以不做肝活检,但应抗病毒治疗。 $AIAG$ 在 0.32 和 0.72 之间,不能肯定有无显著性肝纤维化,故仅需要对这部分患者做肝活检。因此,大约有 50% 病人可以避免肝活检。这个患者的 $AIAG > 0.72$,其有显著纤维化的概率为 93%,因此,不再做肝活检,而开始抗病毒治疗。

第三节 诊断试验临床思维与决策

我们在开出医嘱做某项试验前应考虑以下几个方面:

1. 验前概率是多少,对患者诊断还有多大疑问,是否需要做这项检查,如果验前概率非常高或非常低,临床医生则可以考虑不做诊断试验。因为如果验前概率高于治疗阈值,即可立即开始治疗;而如果验前概率低于诊断阈值,则基本排除该病。因为试验即使阳性,也不可能是该病。

2. 该项检查如果漏诊或误诊会对患者带来多大危害?

3. 这项检查的似然比能否改变进一步临床决策。当然还应考虑到做该项检查的危险性、费用以及做该项检查的迫切性。对于是否需要做进一步试验,可根据以下情况决定:如果验前概率很高,试验结果阴性时,则假阴性的可能性很大,必须再做确诊试验。如果验前概率很小,试验结果阳性时,则假阳性的可能性很大,也必须再做确诊试验。如果验前概率很高,试验结果阳性时,即可肯定诊断,不必做进一步试验,除非该试验特异性特别低。如果验前概率很小,试验结果阴性时,即可排除该诊断,除非该试验敏感性特别低。

4. 有多项试验可供选择时,应该先选择无创的检查,还应选择性价比高的。在开每一张检查单时,应考虑做这项检查对于病人诊断的利与弊。如果检查有不良反应和风险,应根据患者情况考虑患者能否耐受,检查的风险是否大于可能得到的诊断带来的益处。做一些有风险的检查,必须要让患者知情同意。让每位患者在做检查前了解为什么必须做这项检查,做这项检查的利与弊,检查结果对进一步诊断和治疗的意义。如果是自费项目,患者经济状况能否负担均应考虑进去。而与患者谈话内容应该有循证医学的证据。

循证医学通过实施上述步骤使医生对患者的处理更具科学性。

(王吉耀)

参考文献

1. Sheng-di Wu, Yan-jun Ni, Li-li Liu, et al. Establishment and validation of a simple noninvasive model to predict significant liver fibrosis in patients with chronic hepatitis B. *Hepatol Int*, 2012, 6(1): 360-368.

第五章 临床治疗的循证实践

循证医学在近 20 年得到了迅速的发展,其核心思想是任何医疗决策的确定均应以客观的临床科研证据为依据,其理论与研究方法是医务人员提高医疗服务质量和效率的重要和有效途径。在临床医师的日常诊疗过程中,一方面应以疾病的病理生理学机制为理论基础,以既往对疾病治疗的成功经验为参考,以现有最佳临床研究结果为依据,并且与患者的选择相结合,最终制定出最佳的临床医疗方案与决策;另一方面要求临床医师不仅仅需要考虑治疗效果,而且更应该关注治疗方案的不良反应、治疗后患者的生活质量及因疾病和治疗所产生的卫生经济学问题。

第一节 Meta 分析的原文剖析

临床问题

老年男性患者,以“活动后胸骨后疼痛 1 年,加重 1 月”之主诉入院。既往有高血压病史 3 年,且有糖尿病病史 10 年,口服“瑞格列奈,2mg,每日 1 次”,血糖控制较差。入院行心电图示:V₁~V₄ 导联、II、III、aVF 导联 ST 段压低;超声心动图未见明显异常,左室射血分数:0.48。行冠状动脉造影提示:左主干+三支血管病变。目前患者属于多支血管病变,对于所发现的冠状动脉病变情况,有两种积极的治疗方案,一是进行经皮冠状动脉介入治疗(PCI),一是进行冠状动脉旁路移植术(CABG)。那么,哪一种治疗方案治疗效果较好?哪一项治疗术后患者有更佳的生活质量?哪一项治疗方案卫生资源消耗最低?最终我们究竟为患者推荐哪一种治疗呢?

PICO 问题:

P:冠心病左主干+三支血管病变患者

I:CABG

C:PCI

O:治疗 1 年后心脑血管事件发生的概率

寻找证据:

该问题是有关治疗的典型的且常遇到的问题,由于循证临床决策中的核心部分是运用当前最佳证据,因此它的证据获取方式与一般临床研究不同,只要针对问题找到最佳证据即可,由此认为获取证据的策略是根据证据等级金字塔从高到低依次检索。该问题是预防性治疗问题,故证据等级由高到低顺序为:循证临床实践指南/基于随机对照试验的系统评价/单个随机对照试验/基于队列研究的系统评价/单个队列研究/基于病例对照研究的系统评价/单个病例对照研究/病例系列报告/个案报告/体外、体内试验/没有证据基础的专家经验。依据检索策略,我们未查到相关循证临床实践指南,故进行随机对照试验的系统评价文献检索。我们以“CABG、PCI”作为检索词,在 Medline 数据库中进行检索,设定文章类型为“review”,出版时间为“近 5 年(2009~2013 年)”,查到 106 篇系统综述,筛选出符合要求的文献如下。



一、文献题目和来源

题目: A Meta-analysis of 3773 patients treated with percutaneous coronary intervention or surgery for unprotected left main coronary artery stenosis

来源: JACC, Cardio-vascular Interventions, 2009, 2(8): 739-747

二、研究目的

通过对冠心病存在左主干 + 三支血管病变的患者行 PCI 或 CABG 术后主要不良心血管及脑血管事件 (Major Adverse Cardiac and Cerebrovascular Events, MACCE) 的发生率进行分析, 从而对两者的治疗效果进行系统评价。

三、研究背景

近年来, 冠心病发病率逐年增加, 多起病突然且变化快, 为了寻找新的治疗措施及提高其治愈率, 有必要调查选用 CABG 或 PCI 治疗冠心病心绞痛随机对照试验 (RCT) 或系统评价的现状, 并评价其能否为临床应用提供高质量证据。

四、研究方案

该文献为 Meta 分析, 纳入 10 篇文献, 共 3773 人, 平均随访 3 年。文献检索主要通过电子检索, 电子数据库包括 BioMed Central 数据库、ClinicalTrials.gov、Google Scholar 及 PubMed 数据库及其他美国心脏病学会、美国心脏协会的主要会议和导管射频治疗杂志上的随机数据集。

五、研究对象

1. 诊断标准 2008 年 WHO 诊断标准。
2. 纳入标准 (including criteria) ①合并糖尿病的无保护的左主干狭窄患者对于行 CABG 或 PCI 均无特殊限制; ②随访至少 1 年; ③有 MACCE 记录和存活情况记录; ④有危险评分相关报道的患者至少超过 30 人; ⑤发表在同行评议的杂志上, 或在主要心血管病会议上有随机数据集的展示。
3. 排除标准 ①相同研究重复的文章; ②文章中仅提及 CABG 或是 PCI, 而非两项均提及者; ③研究未包括糖尿病或为仅包含 1 型糖尿病患者; ④研究中没有比较 CABG 和 PCI 的结局者; ⑤研究中使用其他外科手术联合 CABG 者; ⑥非随机对照试验; ⑦综述、书信及观点; ⑧不能萃取结果的研究。
4. 样本量 3773 例。

六、主要结果的测量指标

1. CABG 和 PCI 术后直至术后 3 年死亡率的 OR 值作为主要终点。
2. MACCE (死亡、心肌梗死及卒中) 及术后靶血管重建的 OR 值作为次要终点。

七、主要统计方法

该文献为 Meta 分析, 统计软件采用 R (R Development Core Team 2007) package rmeta 2.14 版, 并使用 Mantel-Haenszel 模型来合并固定效应模型的 OR 值, 应用 Woolf's test 来检验异质性, 对于每次进行敏感性分析均省略一项研究。此外, 该文献结局事件 (死亡率和 MACCE 事件) 通过 Kaplan-Meier 生存曲线来评估; 样本大小通过已报道的 CABG 或 PCI 的随访时间的对数正态分布的均值和方差模型来评估。



八、主要研究结果

通过 Meta 分析,该研究得出 CABG 和 PCI 术后死亡、心肌梗死和卒中(MACCE)在 1 年($OR:0.84$; 95% $CI:0.57\sim1.22$)、2 年($OR:1.25$; 95% $CI:0.81\sim1.94$)及 3 年($OR:1.16$; 95% $CI:0.68\sim1.98$)是相似的;靶血管重建在 PCI 组明显高于 CABG 组于 1 年($OR:4.36$; 95% $CI:2.60\sim7.32$)、2 年($OR:4.20$; 95% $CI:2.21\sim7.97$)和 3 年($OR:3.30$; 95% $CI:0.96\sim11.33$)所观察的期间;死亡率方面,PCI 和 CABG 在 1 年($OR:1.00$; 95% $CI:0.70\sim1.41$)、2 年($OR:1.27$; 95% $CI:0.83\sim1.94$)和 3 年($OR:1.11$; 95% $CI:0.66\sim1.86$)无明显差异。

九、结 论

文献纳入的患者与我们临床遇到的患者相似,可以将文献的结果应用于我们的患者。基于以上分析,文献中 PCI 组患者在 1 年、2 年、3 年中总 MACCE 事件及血运重建事件发生率均较 CABG 组高。由此可见,如果只考虑疗效,与 PCI 术相比,CABG 至少可以降低 20 个月的血运重建发生率,降低死亡率和心脑血管事件的发生率。

第二节 Meta 分析的文献评价

该文献是 Meta 分析,属二次研究证据,以下将从研究结果的真实性、重要性和适用性对该证据进行评价。

(一) 研究的对象是否能代表被研究疾病的目标人群

是。纳入病例有明确的诊断标准,避免了诊断水平不一致造成的误诊。该文章共纳入了 10 篇文献,为多中心荟萃分析,样本量较大,基本能够代表冠心病左主干+三支血管病变的人群。

(二) 随访时间是否足够,随访是否完整

基本是。该研究是 Meta 分析,随访时间为手术后 3 年,中位随访时间为 34 个月,最长随访时间为 40 个月,最短随访时间为 20 个月。失访率平均约 8%,按照 5% 和 20% 原则判断,<20% 失访率属于可以接受的范围。

(三) 判断结局有无客观标准,是否采用了盲法

观察指标为 MACCE,属于硬指标。即使不采用盲法也能客观判断结局。

(四) 是否对患者进行意向性分析

该研究纳入的病例均为随机对照实验,在结果的分析中因为试验组存在部分患者由于各种原因未能实施手术治疗,故在结果的分析中采用了意向性分析。因此该证据真实性好。

(五) 研究结果的精确性

该研究对结果报道了 95%CI。例如,该研究得出 CABG 和 PCI 术后死亡、心肌梗死和卒中(MACCE)在 1 年($OR:0.84$; 95% $CI:0.57\sim1.22$)、2 年($OR:1.25$; 95% $CI:0.81\sim1.94$)及 3 年($OR:1.16$; 95% $CI:0.68\sim1.98$)是相似的;靶血管血运重建方面,在 1 年($OR:4.36$; 95% $CI:2.60\sim7.32$)、2 年($OR:4.20$; 95% $CI:2.21\sim7.97$)和 3 年($OR:3.30$; 95% $CI:0.96\sim11.33$)其 PCI 组明显高于 CABG 组;死亡率方面,PCI 和 CABG 在 1 年($OR:1.00$; 95% $CI:0.70\sim1.41$)、2 年($OR:1.27$; 95% $CI:0.83\sim1.94$)和 3 年($OR:1.11$; 95% $CI:0.66\sim1.86$)无明显差异。

(六) 文献的研究结果能否用于自己的患者,目前的医疗环境能否采用这种方法

可以。本文篇首提出的临床问题的患者属于老年男性人群,年龄 62 岁,诊断冠心病左主干+三支血管病变,患者与文献报道的病例临床背景相似,可以采用文献的结果。只要不存在明显不同的临床背景,如年龄、病情轻重、特殊临床表现等则可以应用文献的结论。目前国内已拥有具备这两种手术技术的多家医院,目前的医疗环境可选用这种方法。

第三节 Meta 分析的临床思维与决策

本文为一篇较好的 Meta 分析,样本量大,保证了样本的代表性,减少了样本偏倚,而且纳入的均为随机对照研究。随访时间较长,结果的精确性好。可适用于日常遇到的患者。

冠心病为慢性疾病,特别是面对左主干+三支血管病变的患者,其病情重,故仔细选择合适的治疗方式可减少患者的痛苦,提高患者的生存率。CABG 和 PCI 作为两种手术方式,如何选择疗效更佳的治疗方法是医务工作者关心的问题。临床医师应用这些证据处理患者后,还应该对这些结果应用的效果进行评价,即后效评价。比如此患者实际疗效是否如文献所述?如果这位患者的疗效特别好或特别差,需要及时总结经验教训,积累到一定的病例后也可以自己进行疗效研究。

第四节 生活质量的原文剖析

由以上探讨可知,单纯考虑疗效,我们可以认为对于多支血管病变的冠心病患者建议其考虑 CABG,但是,随着社会的进步,经济的发展,临床医师和患者对治疗的要求不仅仅局限于一些确定的终点生物指标如死亡、心脑血管事件等,而且开始关注患者发病时的心理状态和患者所承受的痛苦。面对这些问题,临床治疗目的由最初的症状控制转向对治疗转归和生活质量等远期疗效的重视。而生活质量评价不同于实验室指标,它主要是个体主观感受的表达和体现,因此我们需要通过生活质量量表对其进行评估。对于相关生活质量文献的真实性的评价,实质也是对量表的评价。若量表不经过信度、效度和反应度的检验,我们就有理由怀疑这篇文章的真实性。

临床问题

老年男性患者,以“活动后胸骨后疼痛 1 年,加重 1 月”之主诉入院。既往有高血压病史 7 年,血压基本维持在正常范围;有糖尿病病史 15 年,口服“瑞格列奈,2mg,每日 1 次”,血糖控制较差。入院行心电图示:V₁~V₄ 导联、II、III、aVF 导联 ST 段压低;超声心动图大致正常,左室射血分数:0.48。行冠状动脉造影提示:左主干+三支血管病变。目前患者存在多支血管病变,基于冠状动脉病变情况,哪一项治疗术后患者会具有更佳的生活质量?带着这一问题,我们进行临床治疗决策的循证实践。

PICO 问题:

P:冠心病左主干+三支血管病变患者

I:CABG

C:PCI

O:治疗 1 年后的生活质量

寻找证据:

首先依据检索策略,进行文献检索,以“quality of life”and “PCI”and “coronary-artery bypass surgery”为检索词在 Cochrane 图书馆查阅相关系统综述,未检索到相关文献;其次,可选择 Dare 数据库进行检索,寻找相关系统综述及随机对照试验,仍未找到;最后,在 PubMed 数据库中搜寻相关随机对照研究。

一、文献题目和来源

题目:Quality of life after PCI with drug-eluting stents or coronary-artery bypass surgery

来源:N Engl J Med, 2011, 364 (11): 1016-1026



二、研究目的

通过对冠心病存在左主干+三支血管病变的患者行 PCI 或 CABG 术评价术后生活质量的优劣。

三、研究设计方案

通过西雅图心绞痛量表(Seattle Angina Questionnaire,SAQ)、健康调查表(The Medical Outcome Study 36-Item Short-Form,SF-36)来评价经皮药物洗脱支架植入术(DES-PCI)和 CABG 后的生活质量,本篇文献主要应用两个量表进行生活质量的评价。文献中对 SAQ、SF-36 量表进行信度、效度及反应度的检验。

四、主要结果的测量指标

应用 SAQ 和 SF-36 量表对行 DES-PCI 或 CABG 的冠状动脉左主干病变或患有 3 支血管病变患者基线状况、术后第 6 个月及术后第 12 个月生活质量情况进行评价。

五、主要研究结果

通过对接受 DES-PCI 或 CABG 的冠状动脉左主干病变或 3 支血管病变患者基线状况、术后第 6 个月及术后第 12 个月进行生活质量评价,发现第 6 个月和第 12 个月的 SAQ 和 SF-36 评分均明显高于基线状况,说明无论经过 PCI 或是 CABG 治疗,患者生活质量均有明显提高;进一步比较发现 SAQ 心绞痛发作状况得分在第 6 个月和第 12 个月 CABG 组(6 个月, 92.8 ± 14.5 ; 12 个月, 93.8 ± 14.2) 高于 DES-PCI 组(6 个月, 91.1 ± 17.3 ; 12 个月, 92.4 ± 17.5), 均有统计学意义,但在同一时间两组之间的差别较小,均为 1.7;两组在第 1 个月和第 6 个月无心绞痛发作的情况是相似的,但在第 12 个月 CABG 组略高于 DES-PCI 组(76.3% vs 71.6%, $P=0.05$);而 SAQ 其他亚组量表和 SF-36 量表对两组的评分或是相似,或是在第 1 个月 DES-PCI 组略高。由此,我们可以认为,CABG 组在第 6 个月和第 12 个月心绞痛发作情况较优于 DES-PCI 组,但获益程度较小(表 3-5-1)。

表 3-5-1 应用 SAQ 表评估心绞痛发作状况

变量	PCI	CABG	两组间差异 (95%CI)	P
SAQ 评分				
基线	69.6 ± 25.8	69.9 ± 26.7	$-0.1 (-2.7 \sim 2.4)$	0.91
1 个月	90.2 ± 17.6	88.7 ± 88.9	$-1.3 (-3.1 \sim 0.5)$	0.17
6 个月	91.1 ± 17.3	92.8 ± 14.5	$1.7 (0.1 \sim 3.3)$	0.04
12 个月	92.4 ± 17.5	93.8 ± 14.2	$1.7 (0.2 \sim 3.2)$	0.03
症状改善(%)				
1 个月	54.7	52.4	$-2.4 (-7.5 \sim 2.8)$	0.37
6 个月	56.5	57.4	$1.0 (-4.2 \sim 6.2)$	0.71
12 个月	57.6	58.3	$0.6 (-4.6 \sim 5.9)$	0.81
未发生心绞痛(%)				
基线	22.2	22.1	$-0.2 (-4.2 \sim 3.9)$	0.94
1 个月	64.4	61.6	$-2.8 (-7.7 \sim 2.1)$	0.27
6 个月	68.5	72.0	$3.5 (-1.2 \sim 8.2)$	0.14
12 个月	71.6	76.3	$4.1 (0.1 \sim 9.2)$	0.05

注:正值代表 CABG 效果较 PCI 好,负值代表 CABG 效果较 PCI 差



六、结 论

我们可以将研究结果应用于临床患者。我们的患者与文献纳入标准相似,患者对生活质量的要求可以通过该文献来回答,经过 CABG 术后,患者的生活质量较 DES-PCI 略好。此时,我们仅综合分析疗效与生活质量评估,可以建议患者行 CABG 术。

第五节 生活质量的文献评价

一、真实性评价

该文献只是说明了对患者随机分为观察组和对照组,但并没有指出采用何种方法进行的分组,因此不能判断是否进行了真正的随机化分组。该证据的观察组和对照组患者在年龄、性别等方面具有基线可比性。该证据表明研究中失访率低。该证据均按原始分组进入分析过程,采用的是意向性分析。

二、重要性评价

该研究精确度高。采用 SAQ 和 SF-36 量表对生活质量进行评估。

三、可行性评价

可用于临床遇见的患者,CABG 与 PCI 技术成熟可行。对于案例中的患者来说,并没有明确的表述其对治疗措施的价值观与意愿,有待于同患者进一步沟通交流,从而得出患者对于治疗措施的意见及所希望达到的生活质量。

第六节 生活质量的临床思维与决策

本文为随机对照试验,样本量大,保证了样本的代表性,减少了样本的偏倚。通过 SAQ 和 SF-36 量表对生活质量进行评估,结果有一定的精确性。通过对文献的原文剖析及讨论评价,可以认为 CABG 治疗冠心病多支血管病变是有效的,且较为安全。此外,术后生活质量高于接受 PCI 治疗者。可以将该证据结果运用于案例中的患者进行相应的治疗,对于可能出现的与证据不符的情况应通过干预措施进一步的解决处理,并不断地进行循证实践。

第七节 临床卫生经济学原文剖析

面对我国短缺的卫生资源实情,对于冠心病(多支血管病变)伴有糖尿病的患者,我们不得不考虑 CABG 和 PCI 哪一种更经济。临床医生在为患者作出决定时,不仅要考虑到临床结果,即诊断或治疗措施的有效性、危险性(安全性)及社会学效果(即患者的生命质量),而且还应该考虑到经济学效果(即卫生资源的消耗)。因为医生面对的不仅仅是眼前的一位患者,而是大量类似的患者。经济分析是一种定量的分析方法,同时从资源的使用(成本)和临床结果(效果,包括有效性、危险性和生命质量)两方面比较不同措施。近年来,临床经济分析越来越受到重视,在国家制定政策、医院确定重大决策、医生处理临床问题等方面均得到了广泛应用。临床经济分析可以采用临床试验有关的结果和同时收集的成本资料,也可以综合文献报告的有关治疗、诊断、预后等研究结果估计效果。

参见本章开头提出的临床病例。当我们面临临床问题时,我们平时实践中所采用的方法并



不一定适合目前的患者,常常需要咨询上级医生或者查阅文献,后者实际上就是采用循证医学的原理去帮助我们解决一系列的临床问题。大型的临床试验可以提供卫生经济学评价的可靠数据。

寻找证据:

结合上述临床病例,在 Medline 上通过输入 cost-effectiveness, PCI, CABG, diabetes mellitus, multivessel coronary artery disease, 可以查找到来自大型临床试验 FREEDOM 的文献,题目为 Cost-effectiveness of percutaneous coronary intervention with drug eluting stents versus bypass surgery for patients with diabetes mellitus and multivessel coronary artery disease results from the FREEDOM trial。

FREEDOM 研究是一项大型前瞻性的随机对照试验,通过比较糖尿病多支冠脉病变的患者行 DES-PCI 与 CABG 两种治疗方法的临床结果,进一步随访 5 年后两种治疗方法的终点事件发生率。从美国卫生保健系统的角度出发,明确地提出研究目的——比较糖尿病患者行 DES-PCI 与 CABG 两种治疗方法的成本效果分析。比较了治疗冠心病多支血管病变的两种临床措施。随访时间为 5 年,同时应用模型对终生进行临床经济学分析。该模型与文献纳入标准冠心病多支血管病变且合并糖尿病可以行 DES-PCI 或 CABG 者一致。

第八节 临床卫生经济学文献评价

(一) 研究的设计

FREEDOM 研究是一项大型前瞻性的随机对照试验,研究者从美国卫生保健系统的角度出发,对 DES-PCI 与 CABG 两种治疗方法进行成本分析,故我们认为该文献提供了完整的经济评价。文献明确地提出研究目的——比较糖尿病患者行 DES-PCI 与 CABG 两种治疗方法的成本效果分析。对于 CABG 患者,虽然初始手术成本较 PCI 低,但总住院成本仍较高,为 8622\$。在随访的 5 年,PCI 者由于较高频率的血运重建和门诊药物治疗使成本升高。在试验期间,虽然 CABG 患者有少许的获益,尽管如此,累计 5 年每个患者成本即为 3641\$。当试验结果被延伸到终生范围时,相对于 DES-PCI, CABG 仍然是较为经济有效的。

(二) 成本效果分析应该在随机对照试验基础上进行

FREEDOM 研究属于随机对照实验,是在 FREEDOM 研究得出临床结果的基础上进行的。并明确指出随访时间为 5 年,同时应用模型对终生进行临床经济学分析。该模型与文献纳入标准——冠心病多支血管病变且合并糖尿病可以行 DES-PCI 或 CABG 者一致。此外,该文献也考虑到生活质量问题,应用欧洲五维度三水平量表对两种方法进行了生活质量评估。但该文献没有考虑依从性问题,对于已行 DES-PCI 者,有可能再次进行 DES-PCI,实际上临床工作中可能只做一次 DES-PCI 手术,术后由于各种原因终止了治疗,降低了依从性。该文献对两种治疗方法的各成本(如医疗保险费用、手术成本、术后住院费用及其他成本)的测量均进行了明确的规定。遗憾的是该文献并没有对间接医疗成本、间接成本及无形成本进行统计。

(三) 该文献所选择的经济学分析类型恰当

文献引用的各种备选方案的资料真实可信,经济学分析确定了所有的成本和效果,并选择了可靠的估计方法。增量分析是在增加效果或效用(生命质量属于成本效用分析)时需要增加的成本,是在做临床决策时需要考虑的重要方面。该文献,中对成本和效果资料分析进行了增量分析,增量分析结果显示:行 CABG 每增加一个质量调整生命年,需要 8132\$,且这个增量成本效果比小于 20 000\$。此外,该文献在进行成本效果分析时均考虑到了贴现,为 3%。

(四) 是否进行了敏感性分析

是。



第九节 临床卫生经济学临床思维与决策

在得到了两种措施经济分析的精确性评价以及经济分析结果之后,我们考虑结果是否可以应用于我们的患者。主要从两方面考虑:一是成本效果分析的增量比,二是患者在多大程度上与经济分析中的病例人群相似。对于糖尿病合并3支血管病变的患者,CABG与PCI对照相比时,如果PCI成本增加而疗效减少,当然不会选择PCI;而如果成本减少效果增加,则肯定会接受PCI。文献中纳入标准与我们的患者条件相似。但是,该文献纳入的是国外的患者,产生的费用均以国外的收费标准且以美元计算,可能与国内的情况存在差距。

综合上述分析,我们通过CABG与PCI术的疗效、生活质量和卫生经济学循证实践,可以建议患者行CABG术,它不但可以降低MACCE,减少血运重建,提高生活质量,而且可以具有较低的卫生资源消耗。

第十节 不良反应的循证实践

在临床工作中,我们面对的是患者,对于3支血管病变的患者无论行CABG还是PCI,医师均会要求患者口服阿司匹林肠溶片,尤其对于行PCI手术者,以防止支架内再次形成血栓。但我们也发现,在服用阿司匹林获益的同时,还会出现了一些胃肠道反应、阿司匹林哮喘、过敏反应等。此外,还会出现少见的大疱性表皮松解型药疹、中毒性低血糖、听力损害、肝脏毒性、肾损害、心脏毒性、诱发癫痫习惯性和成瘾性等,这些即为阿司匹林的不良反应。世界卫生组织对不良反应(adverse drug reaction, ADR)的定义为“一般是指在正常用量和用法情况下,药物在预防、诊断、治疗疾病或调理生理功能时所发生意外的、与防治目的无关的不利或有害的反应”,它包括副作用、毒性反应、特异质反应、后遗效应、继发反应、过敏反应、致畸、致癌、致突变反应和依赖性等。需要注意区别的是药物的不良事件(adverse drug event, ADE),它是指某种药品在治疗期间所发生的任何不利的医疗事件(包括偶发事件),但该事件不一定与该药品有因果关系。

不良反应在临床治疗过程中很常见,据报道,美国每年约大于200万名患者因ADR导致病情恶化,其中10.6万名患者可因ADR而致死;据WHO评估,中国每年约有5000万人住院,其中至少有250万人是因ADR而住院,50万人是严重的ADR,每年约19万人死亡,从而增加医药费40亿元。针对如此严重的损失,目前各国均建立了药物不良反应的监测机构,我国也建立了国家食品药品监督管理局(China Food and Drug Administration, CFDA),主管全国药品不良反应监测工作(图3-5-1),形成了国家药品不良反应监测中心。



图 3-5-1 CFDA 网络平台



一、原文剖析

文献题目: Association of aspirin use with major bleeding in patients with and without diabetes

来源: JAMA, 2012, 307 (21): 2286-2294。

研究方案: 属于大型多中心的队列研究。该文献纳入从 2003 年 1 月 1 日起至 2008 年 12 月 31 日止意大利 12 个卫生局在某一指定时期的患者 4 069 869 人, 使用低剂量阿司匹林治疗者 598 418 人, 排除服用阿司匹林少于 75 天后因出血而住院的患者, 纳入研究以前仅服用降糖药者, 及年龄小于 30 岁或大于 95 岁者共 356 574 人后, 剩余 241 844 人。同期 3 471 451 人未服用低剂量阿司匹林者, 排除 1 671 998 纳入前仅服用降糖药和年龄小于 30 岁、大于 95 岁者, 剩余 1 799 453 人。随机抽取 845 707 人与治疗组 241 844 名患者通过 1:1 倾向性评分匹配方法, 最终形成 372 850 人组成的队列研究。其中服用低剂量阿司匹林(治疗组)186 425 人, 对照组(未服用低剂量阿司匹林)186 425 人, 两组人数相等, 两组的基线状况也有明确的对比分析。

二、文献评价

评价不良反应研究结果真实性的原则: 患者是否有明确的分组; 是否以同样的方式和手段测量了两组的治疗和不良反应状况; 随访时间是否足够长; 是否能确定因果关系。该文献属于大型多中心的队列研究, 研究服用阿司匹林的不良反应。经过 5.7 年的平均随访期, 共发现 6907 例出血; 但由于是队列研究, 没有使用盲法。ADR 研究的实质是病因研究, 因此对于评价 ADR 研究的真实性应该考虑是否有因果效应时序、致病因素的剂量效应、脱离接触 - 再接触效应、有无理论基础及在其他研究中是否出现过等因果关系的原则。该研究符合理论基础及前因后果的时序原则。

该研究在经过平均随访期 5.7 年后, 服用小剂量阿司匹林者发生全部出血事件率是 5.58 每 1000 人年 (95%CI: 5.39~5.77), 对照组为 3.60 每 1000 人年 (95%CI: 3.48~3.72), 其 RR 值为 1.55 (95%CI: 1.48~1.63)。可以看出无论服用小剂量的阿司匹林还是不服用小剂量的阿司匹林, 均有一定的出血率。由于该研究是队列研究, 其 RR 值 >3.0 认为有意义。由此看来, 小剂量阿司匹林引起全部出血事件意义不大。在亚组分析中, 糖尿病患者出血事件增加与服用小剂量的阿司匹林无关, 其 RR 值为 1.09 (95%CI: 0.97~1.22), 其 CI 包括 1, 无统计学意义。若不考虑小剂量阿司匹林的服用, 糖尿病与出血事件独立相关 (RR, 1.36; 95%CI: 1.28~1.44)。基于上述结果, 我们认为小剂量阿司匹林可以引起出血反应, 糖尿病与出血反应相关, 但意义均不大。

该文献中的患者与我们的患者相似。PCI 术后的患者均开始服用阿司匹林 100mg, 符合文献中小剂量阿司匹林的标准 (<300mg, 且从研究开始服用阿司匹林到发生出血至少 75 天), 年龄在 30~95 岁之间, 则与文献患者相似, 文献中研究结果即可以应用。

三、临床思维与决策

研究表明, 平均 1000 例接受小剂量阿司匹林治疗 5.7 年将会出现 1 例出血事件。利大于弊, 我们也可以考虑使用阿司匹林肠溶片; 最后, 在治疗过程中, 我们应该选择更少发生不良反应的方法。如果由于治疗仅 10 例患者即发生 1 例严重的 ADR 如死亡, 毋庸置疑, 医师在得到这样的结论时, 会停用此种治疗方法。若治疗措施的 ADR 并不是非常严重, 但如果有 ADR 更小的备选措施, 也应考虑更换以尽可能减少或避免 ADR。

综合分析, 我们通过比较 CABG 与 PCI 术的疗效、生活质量、卫生经济学的循证实践和术后药物的不良反应发生情况, 可建议患者行 CABG 术, 以降低 MACCE、减少血运重建、提高生活质量、降低卫生资源消耗和减少术后服药期间不良反应的发生。

(王聪霞)



参考文献

1. Giorgia DB, Giuseppe L, Antonio D, et al. Association of aspirin use with major bleeding in patients with and without diabetes. *JAMA*, 2012, 307 (21): 2286-2294.
2. Elizabeth AM, Michael EF, Valentin F, et al. Cost-effectiveness of percutaneous coronary intervention with drug eluting stents versus bypass surgery for patients with diabetes mellitus and multivessel coronary artery disease results from the FREEDOM trial. *Circulation*, 2013, 127 (2): 820-831.
3. Hursh Naik, Anthony J, Tarun Chakravarty, et al. A Meta-Analysis of 3773 patients treated with percutaneous coronary intervention or surgery for unprotected left main coronary artery stenosis. *JACC*, 2009, 2 (8): 739-747.
4. Liu JP, Lin Hui, McIntosh H. Chinese medicinal herbs for chronic hepatitis B. *Cochrane Database Syst Rev*, 2001, (8): CD001940.
5. David JC, Ben VH, Patrick WS, et al. Quality of life after PCI with drug-eluting stents or coronary-artery bypass surgery. *N Engl J Med*, 2011, 364 (11): 1016-1026.

第六章 预后研究的循证实践

临床问题

有一名 40 岁中年男性,发热、白细胞轻度升高,诊断为急性髓系白血病(AML),M4 亚型,染色体正常核型。确诊后他和临床医师关心的问题是:如何治疗?如果获得诱导缓解后如何进行巩固强化治疗?如果进行化疗,平均可以存活多少年?最长能够活多少年?哪些因素可以影响预后?

PICO 问题:

P:急性髓系白血病

I:化疗

C:比较缓解后传统化疗方案或大剂量阿糖胞苷方案的长期疗效

O:中位生存时间、3 年和 5 年长期生存率、预后因素

寻找证据:

以“急性髓系白血病”和“预后”或“生存率”或“中位生存”为检索词进行检索,检索到预后的文献后,首先阅读题目和摘要,根据病人特征、研究问题、研究设计方案等选择合适的文献,例如有多篇 AML 预后的文献,选择较新的、大样本、多中心、前瞻性研究的文献,对其质量及其研究结论是否真实可靠进行评价。

下面分别以两篇不同设计方案的预后研究文献为例,一篇为前瞻性队列研究,另一篇为随机对照研究,举例说明预后评价方法在临床实践中的应用。在阅读本章节前请先阅读第二篇中“疾病预后”章节,以便于深入理解。

第一节 前瞻性队列研究原文剖析

一、文献题目和来源

题目:上海市 623 例成人急性髓系白血病非选择性病例的 WHO 亚型分布、初治疗效及预后
来源:中华血液学杂志,2010,31(2):102-107

二、研究目的及背景

通过对上海市 AML 的化疗效果(包括完全缓解率、生存率和生存时间)的分析,并与国外的疗效进行同期比较,了解我国 AML 治疗的现状,并与 20 世纪 90 年代上海地区的疗效进行历史对照,了解近 20 年来 AML 治疗效果的变化。主要研究重点在 AML 的预后和预后因素。

三、研究设计方案

前瞻性队列研究和生存分析方法见图 3-6-1。

四、研究场所

中美联合上海市白血病协作组 24 家医院,均为二、三级医院。多中心研究。

五、研究对象

1. 诊断标准 2001 年 WHO 诊断标准。
2. 纳入标准 2003 年 6 月 1 日至 2007 年 4 月 30 日期间上述 24 家医院凡发现有可疑白血病的成人患者(年龄 ≥ 18 岁),均收集临床资料并采集血标本和骨髓标本送中美联合实验室进行形态学、免疫学、细胞遗传学、分子生物学(MIC)分型诊断,其中符合 AML 诊断标准的共 623 例。为连续样本。
3. 排除标准 <18 岁的未成年人,复发的病人。
4. 样本量 623 例。

六、干预措施

化疗方案:确诊后大多数病例采用 DA(柔红霉素、阿糖胞苷)、IA(去甲氧柔红霉素、阿糖胞苷)或 HA(高三尖酯碱、阿糖胞苷)标准剂量诱导缓解方案,部分患者用其他蒽环类药物(米托蒽醌、比柔比星等)加阿糖胞苷治疗或用其他化疗方案如依托泊苷(VP16)加阿糖胞苷、FLAG(氟达拉滨、阿糖胞苷、G-CSF)等诱导缓解,部分高龄患者、低增生性白血病或有并发症病例(9.8%)采用 CAG(阿克拉霉素、阿糖胞苷、G-CSF)或小剂量的姑息化疗。

七、主要结果的测量指标

缓解率,复发率,1 年、2 年、3 年生存率,中位生存时间,生存曲线。

确诊之日至死亡或末次随访日期为总体生存时间(OS),随访至死亡或 2008 年 5 月 1 日,中位随访时间为 21 个月,最长随访时间为 56 个月。失访率约 12%。

八、主要统计方法

生存率和中位生存时间用 Kaplan-Meier 法计算,绘制 Kaplan-Meier 生存曲线,Log-rank 时序检验进行单因素分析,Cox 回归模型进行多因素分析。

九、主要研究结果

1. 中位生存时间和生存率 所有 623 例患者中位诊断年龄为 51 岁(18~88),中位生存时间为 18 个月(95%CI:14.9~22.4),达到 CR 的患者中位生存时间为 34 个月(95%CI:27~)。1 年、2 年和 3 年生存率为 59.0%、43.0% 和 30.8%。

本研究的结果与其他研究结果的比较,与上海市 1984~1994 年的研究结果比较见表 3-6-1。因为本研究随访时间较短,尚不能得出 5 年生存率。

2. 生存曲线 伴有重现性遗传学异常 AML、AML- 伴多系发育异常(AML-MD)和不属于上述分类的 AML 的中位生存时间分别为 >50 个月(根据目前随访,尚不能得出准确的时间)、4.2 个月和 12 个月(图 3-6-2)。说明伴有重现性遗传学异常 AML 的预后较好,AML- 伴多系发育异常的预后最差。

生存曲线还可以提供每年的生存率的数据,如伴有重现性遗传学异常 AML 的 1 年、2 年和

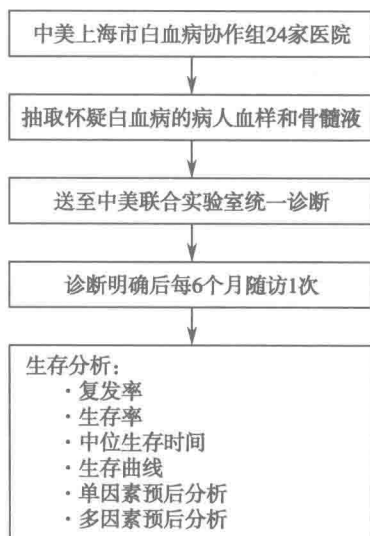


图 3-6-1 设计流程图



表 3-6-1 本组资料治疗效果与其他研究的比较

指标	本研究 <i>n</i> =623	日本 <i>n</i> =109	美国 <i>n</i> =1213	中国天津 <i>n</i> =196	上海 1984~1994 年 <i>n</i> =1526
完全缓解率(%)	66.5	70	66	78.1	51
中位生存时间(月)	18	14.4	18	19.3	12
3 年生存率(%)	30.8	—	32	31.5	18.5
5 年生存率(%)	—	31	24	29.2	13.3
3 年复发率(%)	55.1	34.8	—	—	63

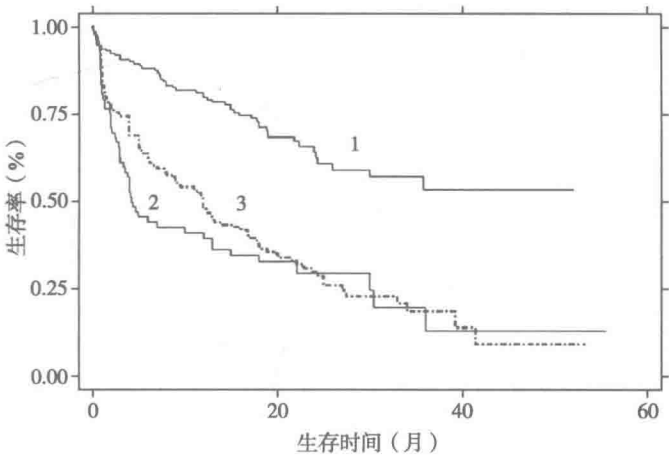


图 3-6-2 按照 WHO 分型的急性髓系白血病 Kaplan-Meier 生存曲线
注:1 为伴有重现性遗传学异常的 AML;2 为 AML- 伴多系发育异常;3 为不属于上述分类的 AML

3 年生存率为 79.6%、64.0% 和 53.1%;AML-MD 的 1 年、2 年和 3 年生存率为 40.9%、29.3% 和 13.0%;不属于上述分类的 AML 的 1 年、2 年和 3 年生存率为 47.8%、30.0% 和 18.5%。

3. Cox 模型多因素分析 AML- 伴多系发育异常的 CR 率和中位生存时间均较差,用年龄、发病时白细胞计数、血红蛋白、血小板和染色体预后分组校正后,发现有无多系发育异常对 AML 生存并无影响,对 AML 生存影响较大的因素为染色体不良核型、高龄和发病时高白细胞计数(表 3-6-2)。而血红蛋白和血小板计数对生存的影响较小。

表 3-6-2 急性髓系白血病的 Cox 模型生存预后因素分析

预后因素	HR (95%CI)	P
WHO 分型		
伴有重现性遗传学异常 AML	1	
AML-MD	1.05 (0.61~1.82)	0.85
不属于上述分类 AML	0.93 (0.56~1.55)	0.80
年龄(岁)		
<60	1	
≥60	1.99 (1.50~2.67)	<0.01
白细胞 (× 10 ⁹ /L)		
<4	1	
4~30	1.35 (.99~1.85)	0.07
≥30	1.94 (1.35~2.78)	<0.01

学
习
笔
记



续表

预后因素	HR (95%CI)	P
血红蛋白 (g/L)		
<60	1	
60~100	0.75 (0.54~1.04)	0.08
≥100	0.66 (0.42~1.01)	0.06
血小板 ($\times 10^9/L$)		
<20	1	
20~50	0.95 (0.65~1.39)	0.80
≥50	0.70 (0.50~1.02)	0.06
染色体预后分组		
良好	1	
中等	3.33 (1.83~5.92)	<0.01
不良	4.77 (2.66~8.53)	<0.01

如果把不良事件编码为 1, 则风险比 (HR) >1 是危险因素, <1 是保护因素。该研究把死亡编码为 1, 存活编码为 0, 结果如表 3-6-2 所示, 年龄 ≥ 60 岁的患者死亡风险是年龄 <60 岁患者的 1.99 倍 ($HR=1.99$)。染色体预后分组是重要的预后因素, 以染色体预后良好组为基准 ($HR=1$), 预后中等组相对于良好组死亡的风险增加 3.33 倍, 预后不良组相对于良好组死亡风险增加 4.77 倍。而血红蛋白和血小板较高者, 死亡风险下降 ($HR<1$), 例如血红蛋白 $\geq 100g/L$ 的患者与 $<60g/L$ 的患者比较, 死亡风险相对下降了 34%。

十、结 论

上海市 AML 治疗水平较 20 年前有明显提高, 短期治疗效果如 CR 率、3 年生存率等已基本达到国际先进水平, 但长期疗效尚有待提高, 复发率较高。

第二节 前瞻性队列研究文献评价

(一) 观察预后的研究对象是否都处于同一起始队列, 研究的对象是否能代表被研究疾病的目标人群

是。以急性白血病确诊之日为起点计算生存时间, 都是新发的住院病例, 有统一的起始时间, 属于同一起始队列。纳入病例有明确的诊断标准, 并且统一在中美联合实验室诊断, 避免了诊断水平不一致造成的误诊。研究对象来自上海市 24 家医院, 为多中心研究, 而且是 2003~2007 年诊断的所有 AML 患者, 属于连续样本, 不是选择性样本, 样本量较大, 基本能够代表上海市急性白血病的人群。

(二) 随访时间是否足够, 随访是否完整

基本是。该研究是前瞻性队列研究, 随访时间从诊断至死亡或 2008 年 5 月 1 日, 中位随访时间为 21 个月, 最长随访时间为 56 个月, 最短随访时间为 12 个月。对于 2006 年和 2007 年诊断的白血病患者而已, 随访时间只有 1~2 年, 随访时间不够长, 所以无法得出 5 年生存率。失访率约 12%, 按照 5% 和 20% 原则判断, $<20\%$ 失访率属于可以接受的范围。

(三) 判断结局有无客观标准, 是否采用了盲法

观察指标为死亡或存活, 属于硬指标, 即使不采用盲法也能客观判断结局。

(四) 是否对影响预后研究的重要因素进行了统计学的校正

是。该研究先采用单因素分析, 再采用 Cox 回归多因素分析, 校正了多个混杂因素。发现



对 AML 生存影响较大的因素为染色体不良核型、高龄和发病时高白细胞计数。

(五) 报告预后研究的结果是否完整

是。报道了缓解(CR)率,复发率,中位生存时间,3年生存率,有 Kaplan-Meier 生存曲线和 Cox 模型多因素分析结果。AML 患者 CR 率约为 67%,中位生存时间 18 个月,3 年生存率为 30.8%,3 年内的复发率为 55%。报道结果比较完整。图 3-6-2 提示染色体核型和 WHO 分型是重要的预后因素。

年龄对白血病的预后有影响,但该研究只介绍了总体患者的生存率等指标,没有对年龄进行分层统计,所以无法得知 40 岁的患者生存率是多少。所以统计分析方面比较粗犷。

(六) 研究结果的精确性

该研究对部分结果报道了 95%CI,有的结果没有报道。例如,所有 623 例患者中位生存时间为 18 个月(95%CI:14.9~22.4),达到 CR 的患者中位生存时间为 34 个月(95%CI:27~),因为随访时间不够长,达到 CR 患者的 95%CI 无法计算出来。年龄 ≥ 60 岁的老年 AML 相对于 <60 岁的患者,死亡的风险比(HR)为 1.99,95%CI 是 1.50~2.67,说明老年人死亡的危险性是年轻人的 1.99 倍。

(七) 我们自己的患者是否与文献报道的患者不同

没有不同。本文篇首提出临床问题的患者为上海市患者,年龄 40 岁,诊断 AML,该文献为上海市白血病协作组的数据,患者与文献报道的病例临床背景相似,可以采用文献的结果。只要没有非常不同的临床背景,如年龄、病情轻重、特殊临床表现等就可以应用文献的结论。

(八) 研究结果是否有助于治疗方案的制订,是否有助于对患者及其亲属作出解释

是。该文研究有重要的临床意义和实用性。对 3 年生存率及影响生存的预后因素作了比较客观、可信的评价,对临床医师及病人家属了解急性白血病预后及预后因素都有重要价值。结合文献和中国 AML 治疗指南,对本文的患者提出治疗建议,并告诉家属该患者预后属于中等,平均生存时间约 2 年,活过 3 年的概率超过 30%。积极、定期治疗是生存的关键。

第三节 前瞻性队列研究临床思维与决策

本文为一篇较好的预后研究报告,上海市 24 家医院多中心协作,大样本研究,并且为连续研究,保证了样本的代表性,减少了样本偏倚。而且采用规范的生存分析统计方法。所不足的是有一定的失访率,部分病例随访时间不够长,对年龄等预后因素没有进行分层统计其生存率、复发率、中位生存时间,所以无法准确知道 40 岁的患者的这些数据,只能知道成人 AML 患者的总体生存情况。

临床医师接诊类似的患者后,一定而且必须与家属深入沟通,告知患者的短期疗效(如 CR 率),也要告知长期疗效(如 3 年、5 年生存率,中位生存时间等),家属和患者知道这些信息后可以主动参与循证医学的临床决策,患者的意见是循证决策的重要组成部分,这样一方面有利于家属和患者积极配合治疗,另一方面也可以减少医患矛盾。绝大多数的医患矛盾均是信息不平衡、沟通不够造成的。

临床医师应用这些证据处理患者后,还应该对这些结果应用的效果进行评价,即后效评价。比如这个患者实际预后是否如文献所述?如果这个患者的预后特别差或特别好,要及时总结经验教训,积累到一定的病例也可以自己进行预后研究,只有这样医学才会进步,证据才会越来越多。任何一个临床医师是证据的应用者,同时也是证据的生产者。



第四节 随机对照研究原文剖析

一、文献题目和来源

题目: A randomized comparison of 4 courses of standard-dose multiagent chemotherapy versus 3 courses of high-dose cytarabine alone in postremission therapy for acute myeloid leukemia in adults; the JALSGAML201 Study

来源: Blood, 2011, 117: 2366-2372

二、研究目的及背景

在年龄 <65 岁的 AML 患者获得第一次完全缓解 (CR) 后, 比较两种治疗方案对患者预后的影响。一种方案为 4 个疗程的标准剂量、多种化疗药物组合的传统方案 (CT), 另一种方案为 3 个疗程的大剂量 Ara-C (HiDAC)。

三、研究设计方案

为随机对照临床试验, CONSORT 流程图见图 3-6-3。

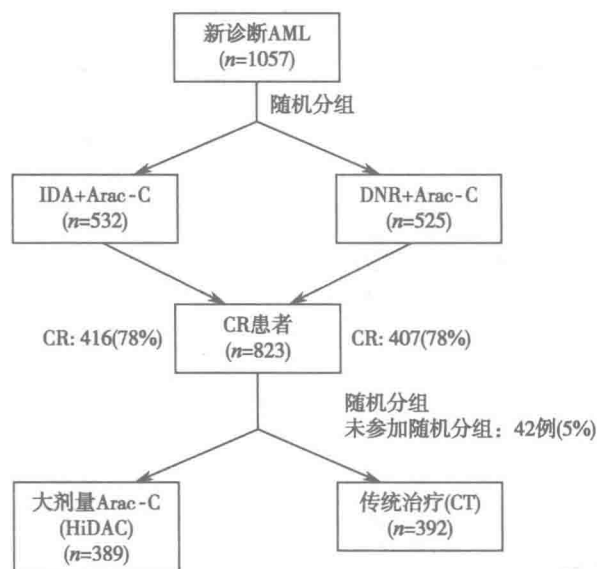


图 3-6-3 CONSORT 流程图

四、研究场所

日本 129 个研究中心。多中心研究。

五、研究对象

1. 诊断标准 (diagnostic criteria) FAB (French-American-British) 诊断标准。
2. 纳入标准 (including criteria) 2001.12~2005.12 期间新诊断的 AML 患者, 年龄 15~64 岁, 为连续样本。同时无明显心肺疾病, 肝肾功能正常, ECOG 状态评分在 0~3 分。
3. 排除标准 (excluding criteria) 排除 M₃ 亚型, 有骨髓增生异常综合征和化疗史的患者。
4. 样本量 (sample size) 1057 例。



六、干预措施

化疗方案:第一次随机分组,分别用 IDA+Ara-C 和 DNR+Ara-C 诱导缓解方案进行诱导治疗,共有 823 例患者取得 CR,CR 后再次随机分组,分别用 3 个疗程的大剂量 Ara-C 和 4 个疗程的标准传统方案,42 例患者因为各种原因未参加第二次随访分组。

七、主要结果的测量指标

第一次随机化后的观察指标是 CR,第二次随机化后的主要观察指标是无病生存(disease-free survival, DFS),次要观察指标是总体生存(OS)和 3 级以上的不良反应。DFS 定义为从 CR 的第一天开始到白血病复发各种原因死亡,OS 定义为从白血病诊断的第一天到各种原因死亡。中位随访时间为 48 个月(5~78 个月)。

八、主要统计方法

Kaplan-Meier 法估计 DFS 和 OS,单因素预后分析采用 Log-rank 方法,多因素预后分析采用 Cox 回归模型。

九、主要研究结果

1. 5 年 DFS 和 OS:HiDAC 组和 CT 组的 5 年 DFS 分别为 43% 和 39% ($P=0.724$),5 年 OS 分别为 58% 和 56% ($P=0.954$) (图 3-6-4)。在染色体预后良好组,HiDAC 组和 CT 组的 5 年 DFS 分别为 57% 和 39% ($P=0.050$),5 年 OS 分别为 75% 和 66% ($P=0.174$) (图 3-6-5)。在染色体预后中等和不良组中,两组的 DFS 和 OS 无统计学差异。

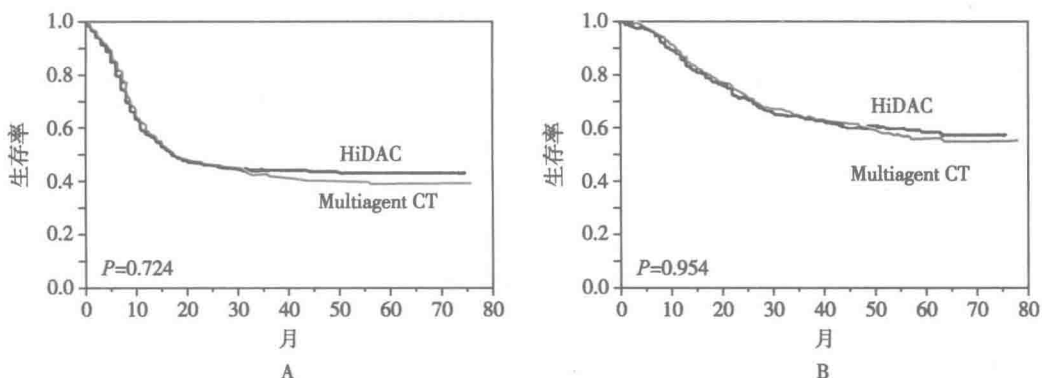


图 3-6-4 两组患者 DFS 和 OS 的生存曲线

A: 大剂量 Ara-C (HiDAC) 组的 DFS 是 43%, 传统化疗 (CT) 组为 39% ($P=0.724$)

B: HiDAC 组和 CT 组的 5 年 OS 分别是 58% 和 56% ($P=0.954$)

2. 预后因素 应用 Cox 回归模型进行多因素预后分析,对于 DFS 而言,发病时白细胞 $\geq 20 \times 10^9/L$ 、诱导疗程数 2 次、年龄 >50 岁为独立的预后因素,校正了这些因素后缓解后的强化方案(应用传统化疗)并非独立的预后因素。对于 OS 而言,年龄 >50 岁、诱导疗程数 2 次、发病白细胞 $\geq 20 \times 10^9/L$ 、原始细胞过氧化酶(MPO)阳性率 $<50\%$ 为独立的预后因素,而校正了这些因素后缓解后的强化方案不是独立的预后因素(表 3-6-3)。

3. 不良反应 在 HiDAC 组,白细胞减少更明显,持续时间更长,感染率更高(表 3-6-4)。

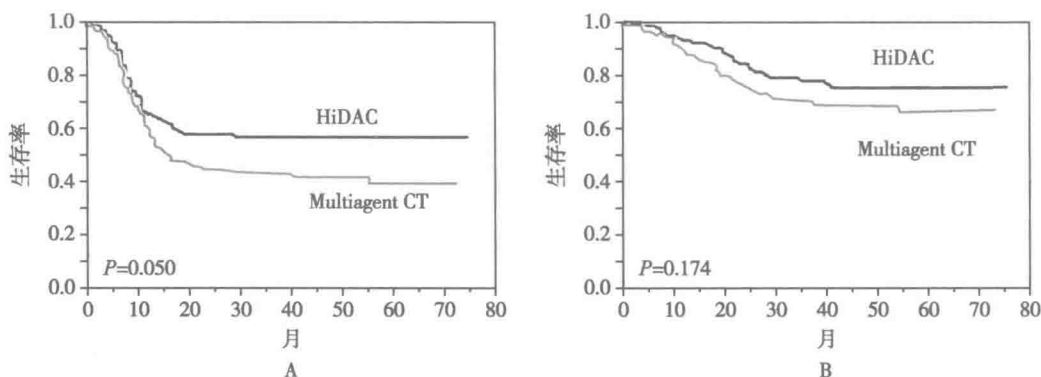


图 3-6-5 染色体预后良好组的 DFS 和 OS

A:HiDAC 和 CT 组的 5 年 DFS 分别为 57% 和 39% ($P=0.050$)B:HiDAC 和 CT 组的 5 年 OS 分别为 75% 和 66% ($P=0.174$)

表 3-6-3 应用 Cox 回归进行多因素预后分析

变量	分类	HR	P
DFS			
发病时 WBC 计数	$\geq 20 \times 10^9/L$	1.49	<0.0001
诱导缓解次数	2 个疗程	1.50	0.0006
年龄	>50 岁	1.33	0.0028
强化方案	传统化疗	1.04	0.7128
OS			
年龄	>50 岁	2.00	<0.0001
诱导缓解次数	2 个疗程	1.58	0.0033
发病时 WBC 计数	$\geq 20 \times 10^9/L$	1.41	0.0070
MPO 阳性的原始细胞比例	<50%	1.42	0.0149
强化方案	传统化疗	0.96	0.7768

DFS:无病生存;WBC:白细胞;OS:总体生存;MPO:过氧化物酶

表 3-6-4 两组不良反应发生情况比较

	大剂量 Ara-C 组	传统化疗组	P
第一次化疗后最低白细胞($\times 10^9/L$)	0.17	0.40	<0.0001
第一次化疗后白细胞 $<1.0 \times 10^9/L$	13 (0~40)	12 (0~36)	0.0005
持续时间(天)			
感染发生率(%)	20.9	14.5	<0.001
出血发生率(%)	0.8	0.7	0.601
30 天病死率(%)	0.9	0.6	0.389

十、结 论

完全缓解后,传统化疗与大剂量 Ara-C 治疗的疗效相当。在染色体预后良好的患者中 HiDAC 的 DFS 优于传统化疗。



第五节 随机对照研究文献评价

该研究属于 RCT 设计,进行两种化疗方案对长期预后的影响研究。在应用预后研究文献评价原则进行评价之前,应该先用治疗研究的文献评价原则评价其真实性,包括:

(一) 患者是否采用真正的随机化分组

在该文中只是简单写到采用随机化分组,但没有介绍具体随机化分组的方法。

(二) 随机化方案是否隐藏

未介绍随机化隐藏的方法。

(三) 在试验开始时两组是否均衡可比

进行了两组临床特征基线情况的比较,如年龄、白细胞、染色体等因素,说明两组可比性好。

(四) 除了试验方案外,两组其他治疗是否相同

基本是。介绍了支持治疗,如血小板输注和细胞因子应用的指征。但因为有 129 个研究中心,各个不同研究中心应用的指征可能存在差异。

(五) 样本量是否足够

研究设计时计算了样本量,使把握度达到 80% 以上,样本量 >1000 例。

上述评价提示 RCT 研究基本是真实可靠的,再进一步用预后研究文献评价原则进行评价。

(一) 观察预后的研究对象是否都处于同一起始队列,研究的对象是否能代表被研究疾病的目标人群

是。急性白血病均为原发性、新诊断的病例,排除了继发性白血病和复发难治的病例。该文对 DFS 和 OS 有明确的定义,DFS 定义为从 CR 的第一天开始到白血病复发各种原因死亡,OS 定义为从白血病诊断的第一天到各种原因死亡。

该研究为多中心研究,研究对象来自 129 个研究单位,1000 多例病例,样本量大,代表性好。诊断标准明确,在各个研究中心诊断后,有一个专门的专家小组复核每一例纳入病例的血片和骨髓片,确保诊断的准确性。

(二) 随访时间是否足够,随访是否完整

基本是。该研究中位随访时间为 48 个月(5~78 个月),从图 3-6-4 可以估计中位 DFS 约为 20 个月,但尚不能得出中位 OS 时间,对于主要观察指标 DFS 而言随访时间足够,但对于次要观察指标 OS 而言,随访时间尚不够。

该文没有直接报道失访率,根据文章描述,所有患者均有完整的随访结果。

(三) 判断结局有无客观标准,是否采用了盲法

观察结局为 CR、DFS、OS 和不良反应,均有明确的定义。因为观察指标缓解、死亡等均为客观的指标,不容易产生测量性偏倚,可以不采用盲法。

(四) 是否对影响预后研究的重要因素进行了统计学的校正

应用了 Cox 回归模型进行了多因素的分析,对影响预后的主要因素进行了校正。

(五) 报告预后研究的结果是否完整

是。报道了 CR 率、复发率、DFS、OS,采用 Kaplan-Meier 生存曲线形式展示了结果,也报道了 Cox 回归分析后的 HR 值和 P 值。没有报道具体的中位生存时间,但读者可以根据生存曲线进行估计。

(六) 研究结果的精确性

没有报道 95%CI,特别是表 3-6-3 中 HR 应该同时写出 HR 的 95%CI。

(七) 我们自己的患者是否与文献报道的患者不同

没有。该研究为多中心、大样本量研究,亚洲日本患者,与我们中国患者应该相似,不会有



很大的人种差异。

(八) 研究结果是否有助于治疗方案的制订,是否有助于对患者及其亲属作出解释

该结果对临床医生提供了治疗决策,有助于对患者进行解释。本章节开篇提到的临床问题中 AML 患者染色体正常,属于染色体预后中等组,应用 HiDAC 和传统化疗可以取得相类似的长期预后,所以两种治疗方案均可以选择。

日本该研究报道 5 年生存率约在 56%,上一篇前瞻性队列研究报道上海市 AML 患者的 3 年生存率仅有 30%,为什么都是亚洲人群,都是研究 AML 患者的预后,不同的设计方案得到的生存率却存在很大的差异?除了化疗的规范程度不同(我国某些医院化疗剂量、疗程、间歇时间等还存在不规范的地方,没有严格按照临床实践指南进行治疗),最主要的原因可能是日本研究为 RCT 研究,对病例的纳入标准有比较严格的要求,例如年龄为 15~64 岁,平均年龄为 46 岁,而且排除了肝肾功能障碍、有心肺疾病的患者,要求 ECOG 评分为 0~3 分,排除了体能状态很差的患者。而上海研究纳入了所有成年 AML 患者,病例没有选择性,属于连续病例,年龄 18~88 岁,平均年龄 51 岁,不排除有并发症和基础疾病的患者,也包括了体能状态很差的患者。所以 2 个研究总体生存率有较大的差异。上海研究更接近真实的研究,RCT 研究的主要缺点就是病例有严格的纳入标准,导致结论的外推性较差。

本文篇首提到的患者较年轻,只有 40 岁,没有其他并发症,根据日本的报道,如果能够按照指南进行规范化治疗,5 年生存率可能达到 50% 以上。

第六节 随机对照研究临床思维与决策

本文为一篇较好的预后研究文献,应用 RCT 设计,研究不同的化疗方案的预后,日本 129 个研究中心的多中心协作,大样本研究,采用正确的预后研究生存分析方法。所不足的是随机分组时没有具体介绍随机化方法和随机化隐藏方案,不符合 RCT 规范写作报道的 CONSORT 声明。部分病例随访时间不够长,没有报道 HR 的 95%CI。

RCT 研究是疗效评价的金标准,其主要缺点是病例有严格的纳入标准,导致结论的外推性较差,操作难度大,对伦理学要求高,所以常用于几种治疗方法的疗效评价,一般不用于全部患者的预后研究。但如果放宽纳入标准,接近真实人群的 RCT 研究,也是预后研究的一种较好的方法。

当研究同一问题,但不同的文献有不同的结论时,临床医师在看文献的时候要考虑其设计方案是否合理、研究对象的纳入和排除标准是否不同(是连续病例还是选择性病例)、治疗方案是否有差异、随访时间长短、结局判断标准等各个方面,综合判断文献的科学性和实用性,进行临床决策。

该研究提出大剂量 Ara-C 和传统化疗都可以取得较好的、相当的短期和长期疗效,对我们的患者治疗方案的选择有指导意义。因为大剂量 Ara-C 组应用后白细胞减少更明显,持续时间长,感染率高,如果某患者一般情况差,或有基础疾病,就不建议大剂量 Ara-C,可以选择用传统化疗。年轻的、一般情况好的患者可以选择大剂量 Ara-C。

循证医学的实践包括提出问题、寻找证据、评价证据、应用证据和后效评价。所以临床医师除了日常的临床工作外,还需要进行方案应用后的后效评价,积累经验,丰富证据。

(王小钦)

参考文献

1. 中美联合上海市白血病协作组. 上海市 623 例成人急性髓系白血病非选择性病例的 WHO 亚型分布、初治疗效及预后. 中华血液学杂志, 2010, 31(2): 102-107



2. 中华医学会血液学分会. 成人急性髓系白血病(非急性早幼粒细胞白血病)中国诊疗指南(2011年版). 中华血液学杂志, 2011, 32(11): 804-807.
3. Miyawaki S, Ohtake S, Fujisawa S, et al. A randomized comparison of 4 courses of standard-dose multiagent chemotherapy versus 3 courses of high-dose cytarabine alone in postremission therapy for acute myeloid leukemia in adults: the JALSGAML201 Study. Blood, 2011, 117(8): 2366-2372.

中英文名词对照索引

B

边际成本 (marginal cost).....	91
标准博弈法 (standard gamble technique, SG).....	93
病例报告 (case reports).....	32
病例对照研究 (case-control study).....	33
病人 / 人群 (patient/population, P).....	11
病死率 (case-fatality rate).....	69
病因 (causation of disease).....	28
病因分值 (etiologic fraction, EF).....	30
不可控性成本 (uncontrollable cost).....	91
不良反应 (adverse drug reaction, ADR).....	160

C

超额危险度 (excess risk).....	29
沉没成本 (sunk cost).....	91
成本 (cost).....	90
成本分析 (cost analysis).....	91
成本效果分析 (cost-effectiveness analysis, CEA).....	89, 92
成本效益分析 (cost-benefit analysis, CBA).....	89, 92
成本效用分析 (cost-utility analysis).....	89
成本效用分析 (cost-utility analysis).....	93
成本最小化分析 (cost minimization analysis, CMA).....	89, 91
重测信度 (test-retest reliability).....	83
重复性 (repeatability).....	52

D

单盲 (single blind).....	59
队列研究 (cohort study).....	34, 72
对照措施 (comparison, C).....	11

F

反应性 (reaction).....	85
非随机对照试验 (non-randomized control trials, NRCTs).....	60
复发率 (recurrence rate).....	70

G

GRADE 标准 (Grading of Recommendation Assessment, Development and Evaluation, GRADE).....	8
干预措施 (intervention or exposure, I).....	11
构建问题 (asking question).....	123
固定成本 (fixed cost).....	91
归因分值 (attributable fraction, AF).....	30
归因危险度 (attributable risk, AR).....	29
归因危险度百分比 (attributable risk proportion or percent, ARP, AR%).....	30

H

横断面研究 (cross-sectional study).....	32
后效评价 (after assessment).....	123
缓解率 (remission rate).....	70
患病率研究 (prevalence study).....	32
患者报告结果 (patient reported outcomes, PROs).....	81

J

机会成本 (opportunity cost).....	91
疾病成本 (cost of illness).....	91
疾病频率调查 (disease frequency study).....	32
疾病死亡率 (disease-specific mortality).....	69
间接成本 (indirect cost).....	89, 90
健康调查表 (The Medical Outcome Study 36-Item Short-Form, SF-36).....	157
健康相关生命质量 (health-related quality of life, HRQL).....	80
健康质量调查表 (quality of well-being, QWB).....	93
交叉对照研究 (crossover study design).....	61
结构效度 (construct validity).....	84
结局分析 (outcome analysis).....	69
截尾数据 (censored data).....	74
金标准 (gold standard).....	51
精密性 (precision).....	52
聚合效度 (convergent validity).....	84
绝对危险度减少 (absolute risk reduction, ARR).....	64



K

- 可变成本 (variable cost).....91
 可靠性 (reliability).....52, 82
 可控性成本 (controllable cost).....91
 可信区间 (confidence intervals, CI).....65

L

- 历史性对照试验 (historical control trials, HCTs).....60
 量表 (scale).....80
 临床结局 (outcome, O).....11
 临床经济学 (clinical economics).....88
 临床实践指南 (clinical practice guideline, CPG).....114
 临床应用证据 (applying evidence).....123
 灵敏度 (sensitivity, Sen).....46

M

- 盲法 (blindness).....59

N

- 内部一致性信度 (internal consistency reliability).....83
 内容效度 (content validity).....83

O

- 欧洲五维度健康质量表 (EuroQol 5-Dimensions, EQ-5D).....93

P

- 平均成本 (average cost).....91
 平行试验 (parallel tests).....52
 评定者间信度 (inter-reviewer reliability).....83
 评价证据 (appraising evidence).....123

Q

- 起始队列 (inception cohort).....74
 区分效度 (discriminant validity).....84

R

- ROC 曲线 (receiver operator characteristic curve).....49
 人群病因分值 (population etiologic fraction, PEF).....30
 人群归因危险度 (population attributable risk, PAR).....30
 人群归因危险度百分比 (PAR%).....30
 人群权衡法 (population trade-off, PTO).....93

S

- 三盲 (triple blind).....59
 伤残调整寿命年 (disability adjusted life year, DALY).....
93, 94
 生存质量简表 36 (short form 36, SF-36).....93
 生命质量 (quality of life, QOL).....79
 时间权衡法 (time trade-off, TTO).....93
 双盲 (double blind).....59
 随机抽样 (random sampling).....58
 随机对照试验 (randomized control trials, RCTs).....34, 58
 随机分组 (random allocation).....58

T

- 特异度 (specificity, Spe).....46
 贴现 (discount).....96
 推荐级别 (class of recommendation, COR).....7

W

- 危险度差 (risk difference, RD).....29
 危险因素 (risk factor).....28
 文化调适效度 (cross-cultural validity).....84
 无病生存率 (disease-free survival rate, DFS).....70
 无进展生存率 (progression-free survival rate, PFS).....71
 无形成本 (intangible cost).....89, 91

X

- 西雅图心绞痛量表 (Seattle Angina Questionnaire, SAQ).....157
 相对危险度 (relative risk, RR).....29, 64
 相对危险度减少 (relative risk reduction, RRR).....64
 效益 (benefit).....89
 效用 (utility).....93
 校标效度 (criteria-related validity).....84
 需要治疗的患者人数 (number needed to treat, NNT).....64
 序贯试验 (sequential trial).....61
 序列试验 (serial tests).....53
 循证临床实践指南 (evidence-based clinical practice guidelines, E-CPGS).....114
 循证医学 (evidence based medicine, EBM).....2

Y

- 验后概率 (post-test probability).....49
 验前概率 (pre-test probability).....49, 150
 阳似然比 (positive likelihood ratio, +LR).....47



阳性预测值 (positive predictive value, +PV).....	47
药物的不良事件 (adverse drug event, ADE).....	160
意向性分析 (intention to treat, ITT).....	63
阴性似然比 (negative likelihood ratio, -LR).....	47
阴性预测值 (negative predictive value, -PV).....	47
隐藏随机分组 (concealed random allocation).....	58
有效性 (validity).....	83
有形成本 (tangible cost).....	91
预后 (prognosis).....	69
预后研究 (prognostic study).....	69
预后因素 (prognostic factors).....	69

Z

真实性 (validity).....	5
诊断试验 (diagnostic test).....	46
诊断性研究 (diagnostic study).....	46
证据查询 (acquiring evidence).....	123
证据水平 (level of evidence, LOE).....	7

直接成本 (direct cost).....	90
直接非医疗成本 (direct non-medical cost).....	89
直接医疗成本 (direct medical cost).....	89
治疗成本 (cost of therapy).....	91
治愈率 (cure rate).....	70
质量调整寿命年 (quality adjusted life year, QALY).....	93
中位生存时间 (median survival time).....	71
中位无病生存时间 (median disease-free survival time).....	72
中位无进展生存时间 (median progression-free survival time).....	72
准确性 (accuracy, Acc).....	46
自身前后对照研究 (before after study in the same patients).....	61
总体生存率 (overall survival rate, OS).....	70
组内相关系数 (intraclass correlation coefficients, ICC).....	83
最小显著差异 (minimal important difference, MID).....	85

[General Information]

书名=循证医学

作者=王吉耀，何耀主编

页数=176

SS号=13919675

DX号=

出版日期=2015.02

出版社=北京人民卫生出版社