



普通高等教育“十一五”国家级规划教材



卫生部“十一五”规划教材

全国高等医药教材建设研究会规划教材

全国高等学校教材

供基础、临床、预防、口腔医学类专业用

# 临床流行病学

第 3 版

主 编 王家良 王滨有

副主编 刘续宝



人民卫生出版社

## 全国高等学校教材

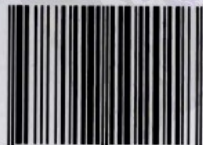
### 供基础、临床、预防、口腔医学类专业用

- |                  |                      |
|------------------|----------------------|
| 1. 医用高等数学 / 第5版  | 27. 传染病学 / 第7版       |
| 2. 医学物理学 / 第7版   | 28. 眼科学 / 第7版        |
| 3. 基础化学 / 第7版    | 29. 耳鼻咽喉-头颈外科学 / 第7版 |
| 4. 有机化学 / 第7版    | 30. 口腔科学 / 第7版       |
| 5. 医学生物学 / 第7版   | 31. 皮肤性病学 / 第7版      |
| 6. 系统解剖学 / 第7版   | 32. 核医学 / 第7版        |
| 7. 局部解剖学 / 第7版   | 33. 流行病学 / 第7版       |
| 8. 组织学与胚胎学 / 第7版 | 34. 卫生学 / 第7版        |
| 9. 生物化学 / 第7版    | 35. 预防医学 / 第5版       |
| 10. 生理学 / 第7版    | 36. 中医学 / 第7版        |
| 11. 医学微生物学 / 第7版 | 37. 计算机应用基础 / 第4版    |
| 12. 人体寄生虫学 / 第7版 | 38. 体育 / 第4版         |
| 13. 医学免疫学 / 第5版  | 39. 医学细胞生物学 / 第4版    |
| 14. 病理学 / 第7版    | 40. 医学分子生物学 / 第3版    |
| 15. 病理生理学 / 第7版  | 41. 医学遗传学 / 第5版      |
| 16. 药理学 / 第7版    | 42. 临床药理学 / 第4版      |
| 17. 医学心理学 / 第5版  | 43. 医学统计学 / 第5版      |
| 18. 法医学 / 第5版    | 44. 医学伦理学 / 第3版      |
| 19. 诊断学 / 第7版    | 45. 临床流行病学 / 第3版     |
| 20. 医学影像学 / 第6版  | 46. 康复医学 / 第4版       |
| 21. 内科学 / 第7版    | 47. 医学文献检索 / 第3版     |
| 22. 外科学 / 第7版    | 48. 卫生法 / 第3版        |
| 23. 妇产科学 / 第7版   | 49. 医学导论 / 第3版       |
| 24. 儿科学 / 第7版    | 50. 全科医学概论 / 第3版     |
| 25. 神经病学 / 第6版   | 51. 麻醉学 / 第2版        |
| 26. 精神病学 / 第6版   | 52. 急诊医学             |

策划编辑... 窦天舒 李向东  
责任编辑... 李向东 樊京娜  
封面设计... 郭 森  
版式设计... 郭 森 魏红波



ISBN 978-7-117-10201-8



9 787117 102018 >

定价(含光盘): 32.00 元



普通高等教育“十一五”国家级规划教材  
卫生部“十一五”规划教材  
全国高等医药教材建设研究会规划教材

全国高等学校教材

供基础、临床、预防、口腔医学类专业用

# 临床流行病学

第 3 版

主 编 王家良 王滨有

副主编 刘续宝

编 者 (以姓氏笔画为序)

方 芳 (四川大学华西临床医学院)

王吉耀 (复旦大学上海医学院)

王家良 (四川大学华西临床医学院)

王素萍 (山西医科大学)

王滨有 (哈尔滨医科大学)

刘续宝 (四川大学华西临床医学院)

孙业桓 (安徽医科大学)

时景璞 (中国医科大学)

李 革 (重庆医科大学)

李 静 (四川大学华西临床医学院)

李绍忱 (山东大学公共卫生学院)

林果为 (复旦大学上海医学院)

洪明晃 (中山大学附属肿瘤医院)

徐 苓 (北京协和医学院)

徐德忠 (第四军医大学)

康德英 (四川大学华西临床医学院)

黄悦勤 (北京大学附属第六医院)

廖晓阳 (四川大学华西临床医学院)

学术秘书 康德英 洪 旗

人民卫生出版社

**图书在版编目 (CIP) 数据**

临床流行病学/王家良等主编. —3 版. —北京:  
人民卫生出版社, 2008. 6  
ISBN 978-7-117-10201-8

I. 临… II. 王… III. 临床流行病学-高等学校-  
教材 IV. R181.3

中国版本图书馆 CIP 数据核字 (2008) 第 062851 号

本书本印次封底贴有防伪标, 请注意识别。

**临床流行病学  
第 3 版**

**主 编:** 王家良 王滨有

**出版发行:** 人民卫生出版社 (中继线 010-67616688)

**地 址:** 北京市丰台区方庄芳群园 3 区 3 号楼

**邮 编:** 100078

**网 址:** <http://www.pmph.com>

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

**购书热线:** 010-67605754 010-65264830

**印 刷:** 三河市富华印刷包装有限公司

**经 销:** 新华书店

**开 本:** 787×1092 1/16 印张: 19

**字 数:** 510 千字

**版 次:** 2002 年 8 月第 1 版 2008 年 6 月第 3 版第 8 次印刷

**标准书号:** ISBN 978-7-117-10201-8/R·10202

**定价 (含光盘):** 32.00 元

版权所有, 侵权必究, 打击盗版举报电话: 010-87613394

(凡属印装质量问题请与本社销售部联系退换)



# 全国高等学校五年制临床医学专业 第七轮 规划教材修订说明

全国高等学校五年制临床医学专业卫生部规划教材从第一轮编写出版至今已有30年的历史。几十年来,在卫生部的领导和支持下,以裘法祖院士为代表的一大批有丰富临床和教学经验、有高度责任感的老教授和医学教育家参与了本套教材的创建和每一轮的修订工作,使我国的五年制临床医学教材不断丰富、完善与更新,形成了一套课程门类齐全、学科系统优化、内容衔接合理的规划教材。本套教材为推动我国医学教育事业的改革和发展做出了历史性巨大贡献。正如老一辈医学教育家亲切地称这套教材是中国医学教育的“干细胞”教材,由她衍生出了八年制和研究生两套规划教材。今天,全国一大批在临床教学、科研、医疗第一线的中青年教授、学者继承和发扬了老一辈的优良传统,积极参与了本套第七轮教材的修订和建设工作,并借鉴国内外医学教育教学的经验和成果,不断完善和提升编写的水平和质量,已逐渐将每一部教材打造成了精品,使第七轮教材更加成熟、完善和新颖。

## 第七轮教材的修订从2006年5月开始,其修订和编写特点如下:

●在全国广泛、深入调研基础上,总结和汲取了前六轮教材的编写经验和成果,尤其是对一些不足之处进行了大量的修改和完善,并在充分体现科学性、权威性的基础上,更考虑其全国范围的代表性和适用性。

●依然坚持教材编写“三基、五性、三特定”的原则。

●内容的深度和广度严格控制在五年制教学要求的范畴,精练文字压缩字数,以更适应广大五年制院校的要求,减轻学生的负担。

●在尽可能不增加学生负担的前提下,提高印刷装帧质量,根据学科需要,部分教材改为双色印刷、彩色印刷,以提升教材的质量和可读性。

●适应教学改革的需求,实现教材的系列化、立体化建设,本轮大部分教材配有《学习指导与习题集》、《实验指导》、《教师用书》以及配套光盘等,且与教材同期出版。

第七轮教材共52种,新增1种,即《急诊医学》。全套教材均为卫生部“十一五”规划教材,绝大部分为普通高等教育“十一五”国家级规划教材,分两批于2008年出版发行。

# 第七轮 教材目录

1. 医用高等数学 / 第5版 主编 张选群
2. 医学物理学 / 第7版 主编 胡新珉
3. 基础化学 / 第7版 主编 魏祖期
4. 有机化学 / 第7版 主编 吕以仙
5. 医学生物学 / 第7版 主编 傅松滨
6. 系统解剖学 / 第7版 主编 柏树令
7. 局部解剖学 / 第7版 主编 彭裕文
8. 组织学与胚胎学 / 第7版 主编 邹仲之 李继承
9. 生物化学 / 第7版 主编 查锡良
10. 生理学 / 第7版 主编 朱大年
11. 医学微生物学 / 第7版 主编 李 凡 刘晶星
12. 人体寄生虫学 / 第7版 主编 李雍龙
13. 医学免疫学 / 第5版 主编 金伯泉
14. 病理学 / 第7版 主编 李玉林
15. 病理生理学 / 第7版 主编 金惠铭 王建枝
16. 药理学 / 第7版 主编 杨宝峰
17. 医学心理学 / 第5版 主编 姚树桥 孙学礼
18. 法医学 / 第5版 主编 王保捷
19. 诊断学 / 第7版 主编 陈文彬 潘祥林
20. 医学影像学 / 第6版 主编 吴恩惠 冯敢生
21. 内科学 / 第7版 主编 陆再英 钟南山
22. 外科学 / 第7版 主编 吴在德 吴肇汉
23. 妇产科学 / 第7版 主编 乐 杰
24. 儿科学 / 第7版 主编 沈晓明 王卫平
25. 神经病学 / 第6版 主编 贾建平
26. 精神病学 / 第6版 主编 郝 伟
27. 传染病学 / 第7版 主编 杨绍基 任 红
28. 眼科学 / 第7版 主编 赵堪兴 杨培增
29. 耳鼻咽喉-头颈外科学 / 第7版 主编 田勇泉
30. 口腔科学 / 第7版 主编 张志愿
31. 皮肤性病学 / 第7版 主编 张学军
32. 核医学 / 第7版 主编 李少林 王荣福
33. 流行病学 / 第7版 主编 王建华
34. 卫生学 / 第7版 主编 仲来福
35. 预防医学 / 第5版 主编 傅 华
36. 中医学 / 第7版 主编 李家邦
37. 计算机应用基础 / 第4版 主编 邹赛德
38. 体育 / 第4版 主编 裴海泓
39. 医学细胞生物学 / 第4版 主编 陈誉华
40. 医学分子生物学 / 第3版 主编 药立波
41. 医学遗传学 / 第5版 主编 左 伋
42. 临床药理学 / 第4版 主编 李 俊
43. 医学统计学 / 第5版 主编 马斌荣
44. 医学伦理学 / 第3版 主编 丘祥兴 孙福川
45. 临床流行病学 / 第3版 主编 王家良 王滨有
46. 康复医学 / 第4版 主编 南登崑
47. 医学文献检索 / 第3版 主编 郭继军
48. 卫生法 / 第3版 主编 赵同刚
49. 医学导论 / 第3版 主编 文历阳
50. 全科医学概论 / 第3版 主编 杨秉辉
51. 麻醉学 / 第2版 主编 曾因明
52. 急诊医学 主编 沈 洪

## 全国高等学校临床医学专业第五届教材评审委员会

名誉主任委员 裘法祖

主任委员 陈灏珠

副主任委员 龚非力

委员 (以姓氏笔画为序)

于修平 王卫平 王鸿利 文继舫 朱明德 刘国良 李焕章 杨世杰

张肇达 沈 悌 吴一龙 郑树森 原 林 曾因明 樊小力

秘 书 孙利军

## 第3版前言

2004年世界卫生组织(WHO)对临床流行病学给予了极高评价。指出这门学科从群体层面和定量研究的方法出发,在推动全球卫生研究、创造最佳的研究成果、促进人类健康事业方面作出了突出贡献。英国医学杂志组织了一些专家,对国际顶级临床医学杂志发表的临床研究文献,以临床流行病学对研究质量评价的标准为准绳,系统地评价出最佳的研究成果(证据),编印并出版了《Clinical Evidence》,推广到临床医疗实践,促进了医疗质量的提高。这些国际性的重大举措,意味着临床流行病学在医学领域中举足轻重的地位。因此,学习、掌握和应用本教材的基本理论、知识和方法,对培养医学生的科学素质是十分重要的。同时,也是医学教育要“面向世界,面向未来”之需。

本教材第1版和第2版面世后,接受了全国高等医学教育和医学继续教育社会实践的考验,相继荣获“教育部优秀教材二等奖(2002)”和“全国高等医药优秀教材一等奖(2005)”,表明本教材是一本深受欢迎的优秀教材。

根据全国高等医药教材建设研究会和卫生部教材办公室的指示精神,在这次第3版的修订中,扩增了有关高等医学院校并诚邀学术造诣高的教授参与了编写,我们结合本学科的国际进展和我国高等医学教育的实际,充分发挥了新编委会的集体智慧,对第3版的内容作了一些创意性修订:

1. 在丰富与发展原版核心内容的基础上,进一步强调掌握医学文献的评价要素和分析的思维能力(新增第5章);
2. 强化医学生如何在科学探索与创新知识(科学研究)的实践中,搜集、管理与应用新知识(新增第4章);
3. 培养医学生面对“卫生突发事件”,如何应用临床流行病学和流行病学的正确方法应对和研究(新增第20章);
4. 鉴于循证医学已列为专门教材,故新版中不再保留该章。关于临床流行病学与循证医学的关系在绪论中阐述。
5. 为了更利于本学科的教学实践,特增版了辅助教材和光盘资料。

我们共同的愿望是在全国一流教材的基础上,紧跟现代科学发展的步伐,不断地吸收最新最佳证据(current best evidence),与时俱进地丰富本教材的科学内涵,使之永葆一流的教材水平。从而更好地服务于我国高等医学教育和人才培养。在本版的修订中,尽管全体编委尽心尽力,毫无疑问地会有某些不足或谬误,敬希应用本教材的师生和同道,给予批评和指正。

在本教材的修订中,得到了四川大学华西临床医学院和华西医院领导的大力支持,特致以衷心的感谢!本教材资料录入,图表绘制,编排和辅助教材的编辑以及部分内容的编写,光盘资料的设计制作等,均为学术秘书康德英和洪旗两位老师负责完成,对他们的辛勤劳动和奉献,致以诚挚的感谢!

主编 王家良 王滨有

2008年2月



<b>第一章</b>	<b>绪论</b>	1
第一节	历史与现状 / 1	
第二节	临床流行病学的学科特点 / 2	
	一、临床流行病学的概念 / 2	
	二、临床流行病学必须是以临床医学为主体的多学科合作 / 2	
	三、临床流行病学的研究对象是病人及其群体 / 3	
	四、临床流行病学力求研究结果的真实性与可靠性 / 3	
第三节	流行病学与临床流行病学的关系 / 3	
第四节	临床流行病学与循证医学 / 4	
第五节	临床流行病学研究的方法学 / 5	
	一、设计 / 6	
	二、测量 / 7	
	三、评价 / 8	
第六节	临床流行病学对临床医学的作用与价值 / 9	
	一、为临床科研提供科学的方法学 / 9	
	二、促进临床循证医学实践, 提高医疗水平 / 10	
	三、服务于医学教育, 培养高质量的人才 / 10	
<b>第二章</b>	<b>疾病负担与临床研究重点的确定</b>	12
第一节	疾病负担的概念 / 12	
	一、定义 / 12	
	二、疾病负担的特点 / 12	
	三、研究疾病负担的意义 / 13	
第二节	疾病负担测量的指标与意义 / 13	
	一、发病指标 / 13	
	二、死亡指标 / 14	
	三、残疾失能指标 / 15	
第三节	疾病负担测定的资料来源与方法 / 21	
	一、疾病监测 / 21	
	二、疾病统计 / 22	
	三、人口学调查 / 23	
	四、残疾失能的估计 / 24	
	五、经济负担的估计 / 24	
第四节	疾病负担指标在医学研究和决策中的应用 / 26	
	一、确定国家或地区性疾病研究的重点 / 26	
	二、确定重点研究与干预的对象 / 28	

**第三章 临床医学科研的选题与立题** ..... 31

## 第一节 临床医学研究的特点 / 31

- 一、研究的对象是病人 / 31
- 二、临床科研的干预措施要安全有效 / 31
- 三、研究的场所 / 31
- 四、医德 / 32

## 第二节 选题与立题的原则 / 32

- 一、选择疾病负担重大的疾病进行研究 / 32
- 二、研究问题要明确具体 / 33
- 三、要有创新性 / 33
- 四、医学研究的公正性 / 33
- 五、选择足够的研究对象 / 33
- 六、抉择合理的研究设计方案 / 34
- 七、干预措施要安全有效 / 34
- 八、研究措施执行的可行性 / 34
- 九、经费支持 / 34
- 十、伦理学原则 / 34
- 十一、预测研究成果的价值 / 35

## 第三节 临床科研课题的类别 / 35

- 一、国家级课题 / 35
- 二、部、省级研究课题 / 35
- 三、国际合作研究课题 / 35
- 四、新药及新技术的临床试验 / 35
- 五、其他项目 / 35

## 第四节 临床选题和立题的程序 / 36

- 一、研究问题的提出与选题 / 36
- 二、充分掌握现有的科技信息 / 36
- 三、确定研究的关键问题 / 37
- 四、撰写研究计划书 / 37

## 第五节 临床科研立题研究的评价标准 / 38

- 一、是否为国家或地区性危害人民健康的重大疾病 / 38
- 二、研究的重点是否明确 / 38
- 三、是否掌握了本研究涉及的最新科技信息 / 38
- 四、是否具有创新性 / 38
- 五、是否可行 / 38
- 六、预期的成本-效果 / 38
- 七、医德 / 39

**第四章 现代医学研究的信息资源与管理** ..... 40

## 第一节 制定检索策略与文献信息查询 / 40

- 一、制定检索策略 / 40



二、文献信息检索 / 40	
三、检索过程与结果评价 / 41	
四、文献信息检索流程实例演示 / 41	
第二节 常见的医学信息资源 / 42	
一、常见的医学信息源种类与分布 / 42	
二、电子文献检索数据库 / 43	
三、灰色文献资源库 / 43	
第三节 对检索文献的严格评价 / 45	
一、常见文献的严格评价工具 / 45	
二、WWW 来源文献的评价工具 / 45	
第四节 电子文献及非常规文献信息的引用格式 / 46	
一、电子出版物文献的基本引用格式与要素 / 46	
二、WWW 网页 / 47	
三、CD-ROM (光盘) / 48	
四、私人邮件信息 / 48	
五、网上公告栏信息 / 48	

## 第五章 如何阅读与评价医学研究文献..... 49

第一节 阅读与评价医学文献的重要性 / 49	
一、医学文献的种类 / 49	
二、阅读和评价医学文献的重要性 / 50	
第二节 高效率阅读医学文献的基本步骤 / 51	
一、明确阅读文献的目的 / 51	
二、熟悉文献的基本结构 / 51	
三、选择性阅读文献 / 51	
第三节 如何评价医学文献 / 52	
一、初筛临床研究证据的真实性和相关性 / 52	
二、确定研究证据的类型 / 54	
三、根据研究类型评价医学文献 / 54	
四、评价医学文献的一般原则 / 54	

## 第六章 医学文献综述与系统评价..... 56

第一节 文献综述的目的与特点 / 56	
一、科学研究工作的需要 / 56	
二、反映学科新动态 / 56	
三、继续教育的需要 / 57	
四、临床医疗和管理决策的需要 / 57	
第二节 文献综述的类型 / 57	
一、叙述性文献综述 / 57	
二、系统评价 / 57	
三、叙述性文献综述与系统评价的区别与联系 / 58	





第三节	叙述性文献综述的撰写方法 / 59
一、	确定文献综述的目的和选题 / 59
二、	广泛收集文献 / 59
三、	阅读和评价 / 59
四、	撰写文献综述 / 60
第四节	系统评价的撰写方法 / 60
一、	确立题目、制订系统评价计划书 / 61
二、	检索文献 / 61
三、	选择文献 / 61
四、	评价文献质量 / 62
五、	提取数据 / 62
六、	分析资料和报告结果 / 63
七、	解释系统评价的结果（讨论和结论） / 63
八、	更新系统评价 / 64

## 第七章 临床医学研究设计的基本原则 ..... 65

第一节	随机化原则 / 65
一、	概述 / 65
二、	随机化方法 / 65
三、	随机化分配的优缺点 / 70
四、	随机化的应用 / 71
第二节	对照的原则 / 71
一、	概述 / 71
二、	设置对照的意义 / 71
三、	对照的种类 / 72
第三节	盲法试验 / 74
一、	概述 / 74
二、	盲法的分类 / 74
三、	非盲法评定 / 75
第四节	均衡性原则 / 76
一、	概述 / 76
二、	资料的均衡性分析 / 76
三、	保证均衡性的方法 / 77

## 第八章 临床医学研究对象的来源与选择 ..... 79

第一节	样本的来源 / 79
一、	社区人群 / 79
二、	医院和专门防治机构 / 80
第二节	样本的抽样方法 / 80
一、	个体随机抽样 / 81
二、	整群随机抽样 / 81



三、系列随机抽样 / 81	
四、非随机抽样 / 81	
第三节 诊断标准 / 82	
一、诊断与疾病诊断 / 82	
二、疾病诊断标准的制定 / 82	
第四节 纳入与排除标准 / 83	
一、纳入标准的制定 / 83	
二、排除标准的制定 / 83	
第五节 影响样本代表性的因素 / 84	
一、样本含量 / 84	
二、随机原则 / 84	
三、无应答和失访 / 84	
四、依从性 / 85	
五、选择性偏倚 / 85	
<b>第九章 临床研究设计的主要方案与运用</b> .....	87
第一节 随机对照试验 / 88	
一、概况 / 88	
二、应用范围 / 89	
三、设计原则和特点 / 90	
四、设计模式 / 91	
五、随机对照试验的主要优缺点 / 91	
六、结果的统计分析方法 / 92	
七、其他类型的随机对照试验 / 92	
第二节 交叉试验 / 93	
一、概述 / 93	
二、应用范围 / 93	
三、设计模式及特点 / 94	
四、结果分析 / 94	
五、优缺点 / 94	
第三节 自身前后对照试验 / 95	
一、概述 / 95	
二、应用范围 / 95	
三、设计模式及特点 / 95	
四、结果分析 / 95	
五、优缺点 / 96	
第四节 队列研究 / 96	
一、概述 / 96	
二、研究设计模式 / 97	
三、研究对象的选择和资料收集 / 97	
四、样本量的估算 / 98	
五、资料的分析 / 98	



	六、应用范围 / 99	
	七、应用实例：新生儿黄疸与母亲使用催产素的关系 / 100	
	八、优缺点 / 100	
第五节	非随机同期对照试验 / 101	
	一、概述 / 101	
	二、设计模式 / 102	
	三、优缺点 / 102	
第六节	病例对照研究 / 102	
	一、概述 / 103	
	二、研究模式 / 103	
	三、研究对象的选择和资料收集 / 103	
	四、研究样本的估算 / 104	
	五、资料的整理与分析 / 104	
	六、应用范围 / 105	
	七、应用实例：雌激素与阴道腺癌关系的病例对照研究 / 105	
	八、优缺点 / 106	
	九、巢式病例对照研究 / 106	
第七节	横断面研究 / 106	
	一、概述 / 107	
	二、研究设计模式 / 107	
	三、应用范围 / 107	
	四、优缺点 / 108	
第八节	叙述性研究 / 108	
	一、概述 / 108	
	二、应用范围 / 108	
	三、设计模式 / 108	
	四、优缺点 / 109	
<b>第十章</b>	<b>影响研究质量的误差与防止措施</b> .....	110
第一节	概述 / 110	
	一、系统误差和随机误差 / 110	
	二、真实性和可靠性 / 112	
第二节	选择偏倚 / 113	
	一、类型 / 113	
	二、控制措施 / 118	
第三节	信息偏倚 / 119	
	一、类型 / 119	
	二、控制措施 / 121	
第四节	混杂偏倚 / 122	
	一、概念 / 122	
	二、正混杂和负混杂 / 123	
	三、控制措施 / 124	





## 第五节 机遇 / 125

一、机遇的定义与概念 / 125

二、机遇与偏倚的关系 / 125

三、机遇的控制方法 / 126

**第十一章 临床研究资料统计分析的基本方法** ..... 127

## 第一节 临床研究的数据收集与整理 / 127

一、临床研究的数据来源 / 127

二、临床研究数据的收集内容与范围 / 127

三、临床研究数据的收集方法 / 127

四、临床研究数据的整理与管理 / 128

五、质量控制 / 129

## 第二节 统计分析的准备与筹划 / 129

一、评价数据质量 / 129

二、设计统计分析框架图 / 130

## 第三节 常用的统计分析方法 / 131

一、统计描述方法 / 131

二、假设检验方法 / 131

三、可信区间 / 140

## 第四节 统计分析方法的选择 / 142

一、与研究目的及分析目的有关 / 142

二、与数据类型以及变量个数有关 / 142

三、与设计方案有关 / 143

四、与应用条件有关 / 143

## 第五节 统计分析结果的正确解释与评价 / 143

一、统计分析结果的正确表达 / 143

二、正确解释统计结果 / 143

三、统计学意义与临床意义的综合评价 / 144

四、统计分析结果的真实性评价原则 / 145

**第十二章 临床研究中的伦理学** ..... 146

## 第一节 概述 / 146

## 第二节 生物医学研究中的伦理问题 / 146

一、研究对象的特殊性 / 146

二、研究结果的不确定性 / 147

## 第三节 涉及人体生物医学研究的伦理基本原则 / 147

一、涉及人体医学研究的国内外伦理法规 / 147

二、涉及人体研究的基本伦理原则 / 148

## 第四节 知情同意 / 151

一、知情同意的目的 / 151

二、知情同意的过程 / 151

三、知情同意书的基本内容 / 152



## 第五节 伦理委员会 / 153

一、伦理委员会的目的与作用 / 153

二、伦理委员会的审查内容 / 153

三、伦理委员会的组建与运作 / 153

**第十三章 病因与危险因素的研究与评价** ..... 157

## 第一节 病因与危险因素研究的意义 / 157

一、有助于疾病的临床诊断与治疗 / 157

二、有助于疾病的预防 / 157

## 第二节 病因与危险因素研究的基本概念 / 158

一、病因与危险因素 / 158

二、寻找病因的指南清单 / 162

三、疾病自然史对病因学研究的意义 / 162

## 第三节 疾病病因与危险因素研究的基本过程与方法 / 163

一、提出病因假设 / 163

二、验证病因假设的要素 / 165

三、验证病因/危险因素假设的主要研究方法 / 166

四、病因学因果关联的统计学指标 / 168

五、防止偏倚干扰, 慎重分析病因效应的关联 / 168

## 第四节 疾病病因与危险因素研究的评价原则 / 170

一、病因学研究结果的真实性的评价 / 170

二、病因学研究结果的重要性评价 / 172

**第十四章 诊断试验的研究与评价** ..... 173

## 第一节 研究和评价诊断性试验的意义 / 173

## 第二节 诊断性试验研究的基本方法 / 173

一、确定诊断试验的金标准诊断 / 174

二、试验对象的选择和样本量的确定 / 174

三、诊断性试验的数据简化 / 175

四、列四格表 / 176

五、盲法比较诊断性试验的结果 / 176

## 第三节 诊断性试验的评价指标及其临床应用 / 176

一、敏感度与特异度 / 176

二、准确度 / 177

三、预测值与患病率 / 178

四、似然比 / 179

## 第四节 受试者工作特性曲线 / 181

一、确定诊断试验的参考值 / 182

二、不同试验诊断同一疾病优劣的比较 / 182

## 第五节 参考值的确定 / 183

一、参考值的概念 / 183

二、确定参考值的基本方法 / 183



第六节	多项试验的联合诊断 / 184
一、平行试验 / 184	
二、序列试验 / 185	
三、试验的独立性 / 186	
第七节	诊断试验结果的一致性评价 / 186
一、Kappa 值及其意义 / 186	
二、观察或诊断结果的一致性评价方法 / 187	
第八节	诊断试验的评价标准 / 187
一、诊断试验与金标准进行独立的盲法比较 / 187	
二、纳入受试者的合理性 / 187	
三、充分告知研究的地点、环境、受试者来源 / 188	
四、诊断性试验的重复性 / 188	
五、诊断性试验的参考值是否合理、可靠 / 188	
六、如果是联合诊断试验, 是否每一个试验的敏感性与 特异度等重要指标都进行了测量 / 188	
七、对诊断试验操作步骤、注意事项等是否有详细描述 / 188	
八、诊断试验的实用价值是否予以描述 / 188	

## 第十五章 治疗性研究设计与评价 ..... 189

第一节	治疗性研究与评价的重要性及基本条件 / 189
一、治疗性研究评价的重要性 / 189	
二、治疗性研究评价的基本条件 / 189	
第二节	临床治疗性研究的设计 / 191
一、立题的依据和研究的目的 / 191	
二、研究方案的选择 / 191	
三、研究对象的选择 / 195	
四、样本量的估计 / 195	
五、试验药物或措施的选择 / 197	
六、试验效果测试指标的选择 / 197	
七、干预实施方法及随访观察期的确定 / 197	
八、试验结果的整理分析 / 198	
第三节	影响临床治疗性研究质量的常见因素及处理方法 / 199
一、机遇 / 199	
二、偏倚 / 199	
第四节	临床治疗性试验研究的评价原则 / 200
一、真实性评价 / 200	
二、重要性评价 / 201	
三、实用性评价 / 203	

## 第十六章 疾病预后的研究与评价 ..... 204

第一节	疾病预后的概念 / 204
一、疾病预后及其研究的意义 / 204	



二、疾病的自然病史 / 204	
三、临床病程 / 205	
四、预后因素 / 205	
第二节 疾病预后评定方法及其指标 / 206	
一、疾病预后的评定指标 / 206	
二、生存分析 / 207	
第三节 疾病预后研究方法 / 207	
一、疾病预后研究常用设计方案 / 207	
二、疾病预后研究设计中的若干注意事项 / 208	
三、疾病预后因素的研究方法 / 208	
第四节 预后研究中常见的偏倚及其处理方法 / 209	
一、预后研究中常见的偏倚 / 209	
二、偏倚的处理方法 / 210	
第五节 疾病预后研究的评价原则 / 211	
一、观察预后的研究对象是否都处于同一起始队列 / 211	
二、研究的对象是否能代表被研究疾病的目标人群 / 211	
三、随访时间是否足够? 随访是否完整 / 212	
四、判断结局有无客观标准, 是否采用了盲法 / 212	
五、是否对影响预后研究的重要因素进行了统计学的校正 / 212	
六、报告预后研究的结果是否完整 / 212	
七、研究结果的实用性和临床意义怎样 / 213	

## **第十七章 临床决策分析与评价** ..... 214

第一节 概述 / 214	
第二节 临床决策的类型 / 215	
第三节 决策的步骤与原则 / 215	
第四节 决策树与期望值 / 216	
第五节 概率估计 / 218	
第六节 效用分析 / 219	
第七节 不确定型决策 / 222	

## **第十八章 临床经济学分析与评价** ..... 224

第一节 概述 / 224	
一、临床经济学的概念 / 224	
二、临床经济评价的必要性 / 224	
第二节 临床经济评价的基本要求及分析评价的步骤 / 225	
一、临床经济学评价的要求 / 225	
二、临床经济学分析与评价的基本步骤 / 226	
第三节 成本的种类及其测量 / 227	
一、直接成本 / 227	
二、间接成本 / 227	
三、无形成本 / 228	



#### 第四节 临床经济学分析与评价的类型 / 228

- 一、最小成本分析 / 228
- 二、成本-效果分析 / 228
- 三、成本-效用分析 / 230
- 四、成本-效益分析 / 232
- 五、临床经济分析类型的比较 / 232

#### 第五节 临床经济学分析结果的真实性评价 / 232

- 一、立场是否明确 / 232
- 二、是否比较了所有相关的临床措施 / 233
- 三、是否提供了完整的经济评价 / 233
- 四、成本和临床效果测量是否准确 / 233
- 五、结果成本是什么, 是否做了增量分析 / 233
- 六、结论是否随着成本和结果的敏感性改变而变得不可靠 / 233
- 七、研究中评估得出的成本和结果是否和人群中的基线接近 / 234
- 八、是否讨论了读者关心的问题, 包括伦理问题 / 234

### 第十九章 健康相关生存质量的测量与评价 ..... 235

#### 第一节 健康相关生存质量的基本概念 / 235

- 一、产生背景 / 235
- 二、健康相关生存质量的定义 / 235

#### 第二节 健康相关生存质量的测试工具(量表) / 236

- 一、通用量表 / 237
- 二、专用量表 / 237
- 三、量表选用的注意事项 / 238
- 四、HRQL 量表测试的一般要求 / 238

#### 第三节 HRQL 量表的基本功能与质量测评 / 239

- 一、HRQL 量表的基本功能 / 239
- 二、HRQL 量表的信度与效度评价 / 240

#### 第四节 HRQL 在临床研究及卫生政策决策中的应用 / 241

- 一、HRQL 用于慢性非传染性疾病的疗效测量与临床决策 / 241
- 二、HRQL 用于药物临床试验以及药物流行病学研究 / 241
- 三、HRQL 用作评估社区人群和特殊人群的总体健康状况 / 241
- 四、HRQL 还可用于效用评价与卫生经济学评价 / 242

#### 第五节 健康相关生存质量研究的评价原则 / 242

- 一、研究结果是否真实可靠 / 242
- 二、HRQL 研究结果有多大的临床意义 / 242
- 三、HRQL 的测试结果是否有助于自己的病人处治 / 243
- 四、HRQL 的研究能否促进临床实践, 有利临床决策分析 / 243

### 第二十章 突发性公共卫生事件研究设计与应急处治 ..... 250

#### 第一节 突发性公共卫生事件概述 / 250

- 一、概念 / 250



二、分类 / 251	
三、特征 / 251	
第二节 突发性公共卫生事件调查研究 / 252	
一、调查研究的意义 / 252	
二、暴发调查 / 252	
三、研究设计的要点 / 256	
第三节 突发性公共卫生事件的预防和处理 / 257	
一、预防措施 / 258	
二、应对措施 / 259	

## **第二十一章 临床科研设计报告书的撰写 ..... 261**

第一节 基本概念 / 261	
第二节 研究设计报告书的基本内容与撰写方法 / 261	
一、立题依据 / 261	
二、研究目的 / 262	
三、科研假设 / 262	
四、设计方案 / 262	
五、研究地点 / 262	
六、研究对象 / 262	
七、样本大小的确定 / 263	
八、干预方法 / 263	
九、研究因素 / 263	
十、资料收集和分析 / 264	
十一、预期结果 / 264	
十二、质量控制 / 264	
十三、预试验 / 264	
十四、医学伦理道德 / 265	
十五、管理和时间安排 / 265	
十六、经费预算 / 265	

## **第二十二章 临床科研论文的撰写原则和方法 ..... 266**

第一节 医学科研论文的质量基础 / 266	
一、学术水平 / 266	
二、写作水平 / 266	
第二节 临床医学科研论文的撰写格式与内容 / 267	
一、题目 / 267	
二、作者 / 267	
三、摘要 / 268	
四、关键词 / 268	
五、导言或引言 / 268	
六、材料与方法 / 268	
七、结果 / 269	



八、讨论 / 271	
九、致谢 / 271	
十、参考文献 / 271	
第三节 如何写好医学科研论文 / 271	
一、酝酿准备 / 272	
二、拟定提纲 / 272	
三、完成草稿 / 272	
四、最后定稿 / 272	
主要参考文献 .....	273
中英文名词索引 .....	276

# 第一章 绪 论

临床流行病学 (clinical epidemiology) 是一门新兴的临床医学基础科学 (a basic science for clinical medicine), 是临床医生从事临床医学科学研究和指导临床医疗实践十分有用的理论与方法学, 对促进高等医学教育和临床医学的发展有着重要的价值。本学科的学术地位已被国际确认。

## 第一节 历史与现状

自 20 世纪 30 年代 John R. Pual 提出临床流行病学的概念以来, 经历了 30 多年, 通过了 Alvan R Feinstein、David L Sackett 等辛勤地耕耘, 在临床研究和医疗实践中, 创造性地将流行病学及卫生统计学的原理和方法, 有机地与临床医学相结合。继后, 又将卫生经济学及社会医学与临床医学相关的内容, 融合于临床医学, 从而更加发展和丰富了临床研究的方法学, 通过科学实践, 逐步深化了对疾病的发生、发展和转归整体规律的宏观量化的认识, 提高了临床医学的研究和对疾病诊断和治疗的水平, 于是创建了现代的临床流行病学。

在 20 世纪 80 年代初期, 在美国洛氏基金会卫生部主任 Kirr White 和 Scott Halstead 等的发起和支持下, 于 1982 年建立了国际临床流行病学网 (International Clinical Epidemiology Network, INCLEN) (图 1-1), 在美国、加拿大和澳大利亚建立了七个一级国际临床流行病学资源和培训中心 (International Clinical Epidemiology Resource and Training Center, CERTC), 已为全球有关国家的著名医科大学培训了大量临床流行病学的高级专业人才, 此后, 在亚、非及拉丁美洲相继建立了地区性的临床流行病学资源和培训中心 (Regional Clinical Epidemiology Resource and Training Center), 包括我国的华西医科大学及上海医科大学在内。这些中心担负着全球与地区性或本国的人才培训和研究工作。

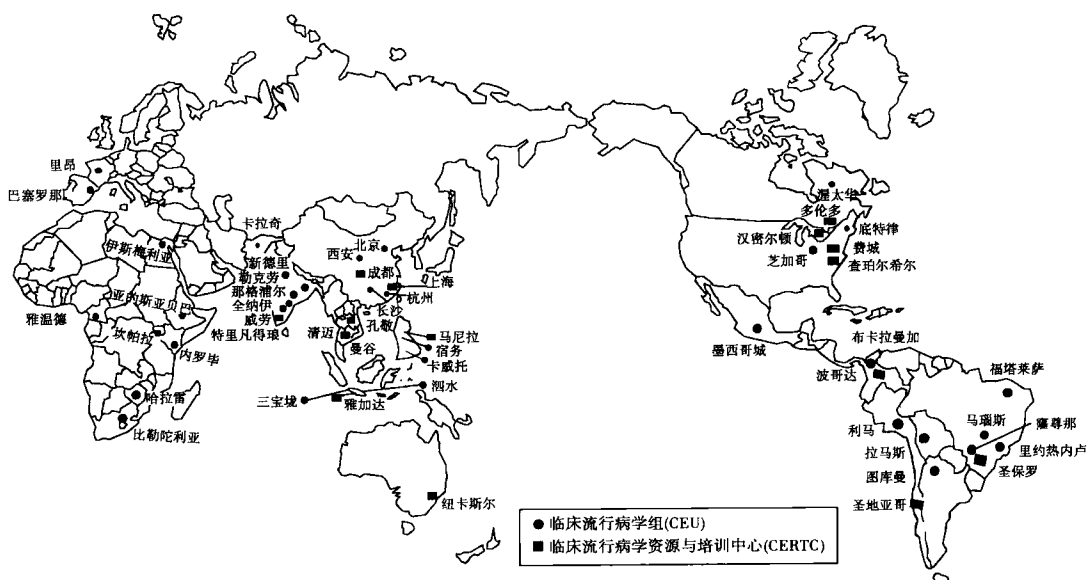


图 1-1 国际临床流行病学网及其培训中心、临床流行病学组在世界各国的分布图





INCLLEN 的宗旨是：“在最可靠的临床依据和最有效使用卫生资源的基础上，促进临床医学实践，从而致力于改善人民健康。为达此目的，本工作网内的各国临床医生、统计师及社会学家要共同奋斗，以建立和维持最佳的医学研究和医学教育的能力和水平，这些是致力于改善人民健康的最重要的条件”。(INCLLEN mission statement: We are dedicated to improving the health of disadvantaged populations, particularly in low-and middle-income countries, by promoting equitable health care based on the best evidence of effectiveness and the efficient use of resources.)

在我国卫生部的领导和支持下，1983 年在我国原华西医科大学、上海医科大学、广州中医学院建立了三个“设计、衡量、评价 (DME)”的国家培训中心 (the centre of design, measurement and evaluation on the medical research, D. M. E)；对临床本科生和研究生开设了临床流行病学课程，培养了高质量的医学人才，并启动了更多的医学院校建立 DME 教研室 (组) 或开展教学、培训工作。在 1989 年建立了中国临床流行病学网 (China Clinical Epidemiology Network, CHINACLEN)，与 INCLLEN 对口接轨；在 1993 年建立了中华医学会临床流行病学分会，为本学科的发展打下了良好的组织基础。实践证明，国际临床流行病学的蓬勃发展，特别是在加强国际卫生研究能力、对重大国际健康问题进行合作研究、促进卫生医药的研究事业、尤其是立足于重点为发展中国家人民服务、致力于卫生保健的公正性作出了很好贡献，具有十分重要的影响。因而获得了世界卫生组织及国际医学界的重视和支持。在 2004 年 WHO 正式发表了一个专题报告，其中对临床流行病学的贡献作了高度评价：“…… by establishing the discipline of Clinical Epidemiology made tremendous contributions towards studying disease & clinical intervention at the population level. Their advances ultimately promoted quantitative methods of measuring disease which enabled the reliable assessment of treatment outcomes in various population”。(临床流行病学这一学科的建立，对群体层面的疾病研究和临床干预作出了巨大贡献。其进展从根本上升华了测量疾病的定量方法，使之在各种群体层面上能够可信地评价干预治疗的结果。)

毫无疑问，21 世纪我们将会在国际上看到，本学科的发展一定会对临床医学革新，特别是对危害人民健康的若干重大疾病的研究和有效的防治发挥更为巨大的作用和贡献。

## 第二节 临床流行病学的学科特点

### 一、临床流行病学的概念

临床流行病学是在临床医学的领域内，引入了现代流行病学及统计学等有关理论，创新了临床科研的严格设计、测量和评价的临床科研方法学，用宏观的群体观点及相关的量化指标，从患者的个体诊治并扩大到相应特定患病群体的研究，探讨疾病的病因、诊断、治疗和预后的整体性规律，力求排除或防止偏倚因素的干扰，确保研究结果的真实性，使获得的研究结论有充分的科学依据，并对防病治病的循证实践有重要实用价值。

临床流行病学的理论原则和方法学均可广泛应用于临床各个学科和专业，目的在于促进临床医学的进步，合理地利用有限的医药资源，提高医疗质量，因而具有临床医学的普遍意义。

### 二、临床流行病学必须是以临床医学为主体的多学科合作

临床流行病学的定位在临床医学，临床流行病学家必须在自己的临床研究和诊治实践



中,应用流行病学的群体观和量化的观点,不断创新临床研究方法和促进本学科的发展,从而不断地创新研究并提供新成果供临床应用。因此,从事于临床医学的临床流行病学工作者应是不脱离临床实践的具有良好临床基础的临床医生。为了创新临床科研方法学,临床流行病学家必须与流行病学家、卫生统计学家、卫生经济学家和基础医学家紧密结合,互相学习,互相促进,在临床研究和医疗实践中,实现多学科交叉,发挥各自优势,方能发展临床流行病学并促进临床医学有效地为人民服务的学科使命。因此,临床流行病学是以临床医学为主体的多学科相合作的“边缘性学科”。

### 三、临床流行病学的研究对象是病人及其群体

临床流行病学的研究对象是以源于医院的个体病例为基础,并扩大到相应的患病群体;或者是从某一疾病患者的整群中去随机抽样以获得具有群体代表性的研究对象。所组成的“群体性患者”或这种研究对象特定疾病的群体性(specified population of disease)乃为本学科的“流行病学”特征。但其理论体系、研究范畴、应用实践等方面却有别于流行病学。这就突破了传统的临床医学局限于医院内的框框,使医院内与人群中对疾病的诊治研究相互结合,无疑对疾病的早期发现与防治,以及对疾病发生、发展和转归规律的认识,会更加全面与深入,因此,本学科对临床医学的发展有重要意义和价值。

### 四、临床流行病学力求研究结果的真实性与可靠性

临床流行病学的精华还在于强调在临床医学研究中,应用科学的方法学,强化科研设计,最重要和最根本的是排除各种偏倚、混杂因素的影响,确保研究结果的真实性(validity)和研究结论的可靠性(reliability),使得科学研究获得的成果能够用于指导临床诊治的决策。促进循证医学(evidence-based medicine)的临床实践,有利于推动临床医学从经验医学向临床医学科学化的方向发展。

## 第三节 流行病学与临床流行病学的关系

按照国际流行病学学会对流行病学的定义:即流行病学是研究特定人群中的健康相关状况或疾病事件的分布及其决定因素,同时要应用这种研究的结果去解决健康问题(Epidemiology: The study of the distribution and determinants of health-related states or events in specified populations, and application of this study to control of health problems. IEA, 4th edition. A dictionary of Epidemiology, P62, 2001)。

可见,该学科是定位于人类社会中的特定人群,目的是研究人群的健康相关状况和疾病事件的分布和它们发生的相关因素,并应用研究的结果去有效地改善人民健康并控制相关的疾病。例如:某些急性传染病一天花、霍乱、伤寒、病毒性肝炎、SARS 等等在人群中的发生与流行以及应用研究出的特异性疫苗和相应措施在人群中进行有效防治,以解决人群中的健康问题。

临床流行病学则是临床学家学习与应用了流行病学的原理与方法,结合了临床医学的实际,研究特定疾病的“群体”分布及其临床特征,包括病因、危险因素、诊断、治疗以及预后等等,以期获得最佳研究结果,并且要应用这些具有群体特征性的研究结果,面对个体患者的具体情况,作出科学预测性的防治决策,以解决具体病人的健康问题,可见,它所涉及的是具体的疾病及其在临床方面的防治。这与流行病学是有所差异的。因而,解决的人类健康或疾病防治等问题具有各自的侧重点,可是在学科的理论和方法学上却有着密切的关联。正如临床流行病学的创始人 Paul J. R 早年提出的“临床流行病学”是将流



行病学家在人群中用于研究疾病的定量化的概念和临床医学（生）日常用于对个体患者诊治决策之间的紧密融合。“A marriage between quantitative concept used by epidemiologists to study disease in populations and decision-making in the individual case which is the daily fare of clinical medicine”（IEA: 4th edition. A Dictionary of Epidemiology, P30, 2001）

Fletcher RH 在论证临床流行病学与流行病学两者关系时，颇为精辟地认为：临床流行病学本身是源于两大学科，说它是“临床”是因为要回答的是临床问题并用最佳证据指导临床决策；说它是“流行病学”是因为用以回答临床问题的方法，很多是由流行病学家所创用的，同时对个体患者医疗决策的最佳证据往往源于患病群体的较大样本的研究成果。

当前，我国临床流行病学的发展，在专业队伍中，一部分为临床医生经过正规的临床流行病学专业培训而成为临床流行病学“家”；另一部分为流行病学家面向临床医学的临床研究与教学实践而“过渡”成为临床流行病学家。正如前述这种学科彼此交叉和结合，必将有利于科学的发展。

## 第四节 临床流行病学与循证医学

近十几年来，由于在国际上出现了“循证医学”（evidence-based medicine, EBM）的概念，又由于它具有显著的“临床性”，目的是要应用当代最佳的研究结果（current best evidence）来提高对患者的诊治水平，所以十分令人关注和欢迎。

什么是“循证医学”呢？循证医学是临床医生在对病人的诊治决策中要谨慎、确切和明智地应用最近最佳的证据，并且必须与自己的临床专业知识相结合，而不能生搬硬套。同时要充分权衡这种决策对病人的利弊关系，要以病人的利益第一；当然这种决策的付诸执行，还必备五个前提：

第一，要有高素质的临床医生：既有良好的医学基础知识和临床专业技能，又有高尚的医德和全心全意为病人服务及专业进取的精神，这是实践循证医学的最重要条件。

第二，要有最佳的研究证据（成果）：源于医学研究中不断产生的最新成果（current best evidence）、这些成果经过严谨评价而确认是真实的、有重要临床意义的而且临床又有实用价值的，方为最佳证据（best evidence）。

第三，要懂些临床流行病学的基本知识与方法学：由于分析和评价最佳证据的方法与标准，都源于临床流行病学的基本理论与临床研究质量的评价原则（标准），又由于某种原因造成某些“证据”可能存在着以劣充优的现象，因此，在实践循证医学中引用“最佳证据”的时候，一定要结合临床实际持批判性的态度。当对某种（些）证据有所疑虑的时候，应查其原始来源并用临床流行病学有关理论与方法进行分析和评价，这样方可避免被误导或误用，并可避免对病人的伤害。

第四，要把病人的利益放在医疗决策的首位：在循证医学实践中应用任何医疗决策并施于病人的时候，务必要以病人利益为第一，要求对病人的疾病控制和康复有利，尽可能地防止或减轻不良反应。所引用的最佳证据，应掌握量化利弊比的依据。同时，还应平等地与病人交流，尊重病人的人格和选择的权利，营造和谐的医患关系，取得病人对医疗决策的良好依从性（compliance）。

第五，要有良好的医疗环境：任何最佳证据或诊疗措施付诸实施，一定要有相应的医疗环境和条件，否则就不能落实。例如：确诊心脏瓣膜病的病人，需要做瓣膜置换治疗方能取得最佳疗效。这就要求医院心脏外科有着良好的医疗硬件设备和相关技术条件。否



则，最佳效果的瓣膜置换就不能实施。

至于循证医学实践于临床医疗，则首先必须准确地发掘病人存在并需要解决（回答）的临床问题；根据临床问题去检索相关的资料及文献；对收集到的文献进行严格的评价，以找出“最新最佳证据”；针对病人的具体情况，权衡利弊，将最佳证据用于循证医疗决策并付诸执行；最后分析与评价最终结果。这就是实践循证医学的“五步曲”。

上述诸点均表明循证医学乃是临床医生对具体病人诊治的一种医疗模式，与传统临床模式的最大差异在于前者强调应用最新最佳的证据来指导医疗决策，强调任何决策要有量化指标科学依据，并充分尊重病人的价值取向和意愿选择。

那么，循证医学与临床流行病学是一种什么关系呢？从上一节中了解到临床流行病学的本身是应用科学的临床研究方法学，研究并产出最佳的研究结果，而且应用这些研究证据来指导临床医疗决策，从而提高医疗水平，达到保障人民健康的目的。20世纪90年代，循证医学这个概念的产生、应用与发展，实际上是在现代临床流行病学的发源地之一的加拿大 McMaster University，工作在临床医疗和教学第一线的临床流行病学家们，应用了临床流行病学的理论原则与方法指导临床医疗教学实践的一种创新和发展。正如 Fletcher RH 所述，“循证医学是临床流行病学用于对病人（诊治决策）的一个现代名词”（Evidence-based medicine is a modern term for the application of clinical epidemiology to the patients）。

## 第五节 临床流行病学研究的方法学

由于患者临床特点、病程、心理状态和社会经济状况等各异，构成了不同的疾病或同一疾病的不同患者临床状况的复杂性。而临床研究及诊断的资料来源是患者的病史、体检发现的阳性及重要的阴性体征、实验室及某些特殊检查的结果，这些资料的获得又来源于多种专业部门和人员，因此，对病史的收集整理是否符合实际，体征是否准确，实验等检查结果是否真实可靠，这里可能受着多种偏倚（bias）和（或）混杂（confounder）的影响，这些都直接关系到临床研究和诊治的质量。

面对着病人的复杂临床状况以及多源性的临床资料，如果没有科学的方法来识别及防止各种偏倚因素的干扰，那么就很难保证临床研究结果和诊治疾病的科学性和可靠性。

传统的临床科研或经验总结，往往利用的是现存资料，即使是前瞻性的研究，由于缺乏严格设计，资料质量的真实可靠性往往有所缺陷，制约了临床医学的研究水平，这就是问题的症结所在，不解决这类问题，就很难使临床科研产生质的飞跃。因此，从20世纪70年代起临床流行病学得到广泛的认同和快速发展。这个学科就是面对临床医学实践中的复杂情况，应用流行病学和统计学的规律和方法学，并与临床实践相结合，创造性地建立了临床科研设计（design）、测量（measurement）与评价（evaluation）的临床科研方法学（DME）并应用于复杂的临床医学研究实践。

在具体地应用临床流行病学研究方法时，应抓住以下几大要素：

（1）疾病/病人（patients/population）：即研究的是什么疾病、是什么样的病人、他们（她）们选自何处、诊断及纳入/排除标准是否明确、样本量需求多少、是否愿意参与试验、依从性如何等等。

（2）研究设计（design）的方案是什么、科学性与实用性如何。

（3）干预措施（intervention）是否有科学依据，有效性和安全性的前期试验依据，具体的应用方法与指征、疗效与不良反应的指标、有否并用辅助疗法等等。

（4）对照组的设置（comparison）：对照组对象的入选标准和方法、来源何处、同期



或非同期、对照的措施是安慰剂或有效对照等。

(5) 研究的结果 (outcomes): 有益 (有效) 及不良反应 (有害) 的结果, 反映这些结果的指标依据, 判断的方法以及校正与否等等。

(6) 时间因素 (time): 干预疗程或追踪的时间有多久, 是否能满足药效或不良反应显效的时间要求, 即通过试验研究, 能在人为设计的水平上可以回答试验组与对照组效应显著差异最低限度水平 (minimally important difference, MID) 时所需的时间。对于某些疾病的追踪观察时间长短, 要注意疾病的自然病史的规律。

上述要素可用各自英文名词的第一字母缩写: “PDICOT”, 然后分别在科研设计方法学 (DME) 中落实。现将本学科研究的方法学分述如下。

## 一、设计 (design)

### (一) 抉择合理的设计方案

要根据不同性质的临床研究课题, 对下列 (表 1-1) 方案在注重科学性和可行性的前提下, 进行抉择。表内的 “+” 号数目表示论证强度及可行性。(+) 的数目越多则越佳。见本教材第七章。

表 1-1 依不同性质的研究课题抉择不同的研究方案

研究的性质	备选方案	论证强度	可行性
病因/危险因素 (aetiology/risk factor)	—— 随机对照试验 (randomized controlled trial, RCT)	++++	—
	—— 队列研究 (cohort study)	+++	+++
	—— 病例对照研究 (case-control study)	+	+++
	—— 描述性研究 (descriptive study)*	±	++++
诊断性试验 (diagnostic study)	—— 金标准方法对照, 系列诊断指标评价	++++	+++
防治性研究 (prevention/treatment)	—— 随机对照试验 (RCT)	++++	++
	—— 交叉试验 (cross-over study)	++	++
	—— 前-后对照试验 (before-after study)	+	++
	—— 病例对照研究 (case-control study)	±	+++
	—— 描述性研究 (descriptive study)	±	++++
预后研究 (prognosis)	—— 队列研究 (cohort study)	+++	++
	—— 病例对照研究 (case-control study)	+	+++
	—— 描述性研究 (descriptive study)	±	++++
	—— 疾病的现况调查	±	++++
	—— 叙述性研究	±	++++

\* 含横断面调查 (cross-sectional study)

### (二) 研究对象的正确诊断

研究的对象, 一定要符合临床公认的诊断标准, 同时按课题设计的具体要求, 确定具体对象的纳入标准及排除标准, 这样, 就会确保研究对象的可靠性 (见本教材第八章)。

### (三) 样本的来源与样本含量的估算

临床研究所涉及的研究对象, 往往都源于医院的就诊病人, 可以是系列病人, 也可以从患病群体中做随机抽样, 视具体情况而定。当需要从人群中去收集研究对象时, 则应按设计要求从人群中随机抽样, 以获得足够量的研究样本 (病人), 至于需要多少量的样本, 要根据有关研究设计的试验与对照组间的假设差异性水平, 以及容许的 I 型错误及 II 型错



误的水平,应用相关的样本含量的计算公式,计算合适的样本需要量。有时为方便计算,也可以采用查表法确定样本含量。应用这类方法以尽可能地减少机遇(chance)因素的影响,把可能产生的假阳性或假阴性的错误控制在可容许的程度(见本教材第11章)。此外,还可以防止样本量不足而造成假阴性的错误结论,以及避免样本量过多导致不必要的资源及时间的消耗。

#### (四) 试验措施要有创新性

试验的干预措施要有创新性,具备其有效性和安全性的科学依据,而且对接受试验措施的对象要有安全性的保障。

#### (五) 确定试验观察的期间要合适

临床试验观察的期间,要根据试验终点的设计指标而定,如终点指标是痊愈、死亡、有效、无效等,那么大多数试验对象达到终点需要的时间即定为试验观察期间,对于通常药物治疗性临床试验的观察期间确定,宜采用MID原则,它的确定要有生物学及临床的试验依据。观察过短易致假阴性结论,过长也可致资源的浪费。

#### (六) 要有识别和防止偏倚因素干扰的对策

由于临床研究的复杂性,在研究的全过程各个环节,不可避免地存在着各种偏倚因素的干扰,可以说整个研究的过程都是排除干扰力争求真的过程,因此在设计中要估计研究中可能存在的偏倚影响因素,要有具体的防止和排除干扰的对策,确保研究结果真实可靠(见本教材第十章)。

#### (七) 要正确地应用统计学分析方法

在研究设计中,根据所采用的有关测试的指标及试验干预措施可能产生的预期结果及其相关资料,考虑正确地采用有关统计学的分析方法。临床研究的资料是丰富多彩的,有定量的、定性的,也有等级的;有的是配对的,有的是非配对的;有的研究结果需要多组比较,也有的仅两组间的结果相比较;有的是需要做单因素分析,也有的需要做多因素的分析;总之,不同质的研究资料,应采用不同的统计学方法,予以分析比较和评价,这对于提高研究质量是很重要的(见本教材第十一章)。因此,所涉及被采用的统计分析方法是在研究设计的阶段就应基本确定,而不能是在研究结束后所获的资料和相关数据,想应用什么统计学的方法就用什么方法,这样做,则是欠科学的。

由于计算机统计软件的广泛应用,为临床医学工作者提供了十分有用的统计学工具。对临床医生而言,统计学的分析再也不是望而生畏的了。

#### (八) 注重科研道德

临床科研的对象是病人,因此,任何研究应以病人的利益为第一,在研究的设计和执行各个环节都要尊重病人的权利,维护病人的利益(见本教材第十二章)。

上面的八点,在科研设计方案中应加以充分考虑。

## 二、测量(measurement)

当某种致病因素接触人体后,往往可以引起人体致病;当某种药物或治疗措施用于病人治疗后,会发生治疗效应。这些致病因素的致病或药物的治疗效应,需要采用一些方法和指标来发现和测量,以作为诊断和治疗效果的依据。为了准确地做到对效应的测量,宜注意下列几点。

#### (一) 试验的措施一定要有反应性和可度量性

试验性的致病因素或治疗的药物,其本身要有致病的或治疗的作用,而且这种作用要能客观地反映出来并被临床及实验室等检查方法及指标度量。



## (二) 测量的方法要求有良好的敏感性和特异性

当致病或治疗效应发生后, 采用的测量方法要有良好的敏感性加以发现, 并有良好的特异性予以确定, 否则, 有发生漏诊(测)或误诊错误的危险。

## (三) 测量指标的判断标准和临床意义要明确

测量临床效应的指标, 归纳起来有计量的标准(如像实验室的生化指标、血压测量等等), 也有计数的指标(如病人痊愈、病残、病死数及其率等, 还有等级指标, 如病情的轻、中、重等; 疼痛的轻、中、重等的分级)。还可分为生理功能性的以及病理性的指标等等。这些指标所获得的数据及程度, 都要注意临床意义。对于任何测量指标的结果判断也应有公认的临床判断标准, 如有效、无效以及恶化等, 至于某些计量指标所反映的治疗前后均数的差值及其统计学的差异性, 虽然有重要的价值, 但更应用临床意义来判断而不宜简单地凭  $P$  值大小作结论。

## 三、评价 (evaluation)

临床研究的结果是否真实可靠、临床意义和价值有多大、研究结果能否能用于临床实践及其适用程度等等, 都是研究者对自己研究成果的评价需要回答的问题, 同时也是临床医生拟采用某一研究成果(证据)于循证医学临床决策时应予严格评价的问题。

第一, 研究的真实性 (validity)。

不论研究的结果是阳性的或阴性的, 一定要根据研究的客观指标作出科学的解释和结论, 确保研究的真实性及其可信程度。

真实性评价的关键, 应注重研究的证据是来自何种设计方案及其论证的强度、对照组的有无以及设置是否恰当、研究对象的诊断标准是否可靠、纳入/排除标准如何、样本量是否足够、组间重要的基线状况是否可比、有无相关偏倚因素存在以及是否采取了相应的防止或处理的措施、依从性如何、对相应的试验观测指标及资料所采用的整理、统计分析方法是否恰当等等。

第二, 临床的重要性 (clinical importance)。

如果研究的结果真实性好的话, 那么临床的意义和价值又如何呢? 按照临床流行病学评价的原则, 一定要有量化的指标加以论证。

评价临床研究结果重要性的指标, 归纳为: 事件发生率 (event rate) (如病死率、生存率、治愈率……)、绝对危险降低率 (absolute risk reduction, ARR)、绝对受益增高率 (absolute benefit increase, ABI)、相对危险降低率 (relative risk reduction, RRR)、相对受益增高率 (relative benefit increase, RBI)、需治疗多少例患者才能获得一例最佳效果 (number needed to treat, NNT) 以及需治疗多少例患者才能发现一例不良反应 (number needed to harm, NNH), 根据它们的具体数据计算可各自的 95% 可信区间。此外, 用于诊断性试验评价患者亦有一系列量化指标 (见本教材第十四章)。这些量化指标都能清楚地表明: 试验组与对照组相比所示的具体有效以及不良反应程度, 便于临床评价。

传统的临床研究证据, 特别是定量化的资料, 常常计算组间均值以互相比, 如果  $P$  值  $< 0.05$ , 则认为组间差异性显著, 往往下结论认为某组的效果优于另一组的效果。这里要强调的是组间差值的大小究竟有无临床意义, 如果没有临床意义的话,  $P$  值再小, 也无临床应用价值。因此, 统计学分析组间差异的显著性并不能完全代表临床的意义。所以, 对于任何临床研究的结果, 务必要注重临床的价值, 如果研究的结果具有临床意义的话, 那么必须应用正确的统计学方法对结果进行显著性检验, 以评价临床差异的真实程度, 即肯定结果的真阳性、真阴性的概率以及检验效能的水平。当某种研究结果既有临床意义, 又有统计学的差异显著性时, 即能作出肯定性的结论; 如仅有临床的意义而统计学



差异并不显著时，不能因此而否定临床的价值，此时应计算Ⅱ型错误和检验效能的水平；如果一个研究的结果既无临床意义，又无统计学显著差异，则这种研究的结果应予否定（表 1-2）。

表 1-2 临床与统计学意义评价

	临床意义	统计学意义	结论
A	+	+	+
B	+	-	+~±
C	-	-	-

临床研究的结果，还应做卫生经济学的评价，计算其成本-效果（cost-effectiveness），成本-效益（cost-benefit）以及成本-效用（cost-utility）并进行比较和评价，分析它们可产出的社会效益及经济效益，以肯定那些成本既低，其效果又佳的研究成果，使之能推广应用（见本教材第十八章）。

第三，研究成果的实用性（applicability）。

临床研究往往是解决某种（些）重大疾病的早期正确诊断以及有效防治或改善疾病预后等为其目的的，因此，它们的针对性是颇为明确的。因此，对于研究的成果之实用性，要分析它们有无实用价值、有多大的实用价值，利弊比有多大，在什么样的医疗环境和条件下可以采用或推广，宜作实事求是的估价。切不可脱离自己的环境技术条件和病人的实际情况，盲目地接受或推行最佳证据。

## 第六节 临床流行病学对临床医学的作用与价值

### 一、为临床科研提供科学的方法学

临床流行病学对临床医学最重要的贡献之一是提供科学的研究方法学。针对研究课题的性质提供严格的科学设计，防止或排除一切可能出现的偏倚因素的干扰，应用严谨的测量指标及评价质量的标准以及合理的统计分析方法，力求研究的结果真实可靠，产出最佳的研究证据，为提高临床医疗质量服务。

临床研究的成果，特别是高质量的最佳证据（best evidence）是防病治病的重要武器，国际上据世界卫生组织统计，每年投入的卫生研究费用逾 700 多亿美元；我国政府在“十一五”国家重大疾病防治研究领域，极大地加强了经费投入，期望能通过加强卫生研究，产出高质量的研究成果以促进防病治病的水平。根据“十五”国家重大疾病防治攻关课题的质量分析，近 30%~40%的课题因为研究设计方法学的缺陷严重地影响了研究的质量，是为教训。所以一旦研究课题确立后，应用临床流行病学的理论、原则和方法来规范和指引研究是成功的关键之一。

临床流行病学指导临床研究，设计了一种效应环的原理，首先通过调查了解有关疾病对国家和人民造成的疾病负担程度（burden of disease），以掌握危害人类健康的主要疾病；从特定群体的角度针对它们发病的病因和危险因素，采取针对性的有效防治措施，并不断地总结其效率和效果；通过研究实践，收集有关研究成果予以综合评价和进一步地改进，再付诸实践，最后验证危害人类健康之疾病负担减轻的程度（图 1-2）。在这样一个循环效应下，要不断地创造新方法、新经验、新措施、新成果，服务于临床医学。



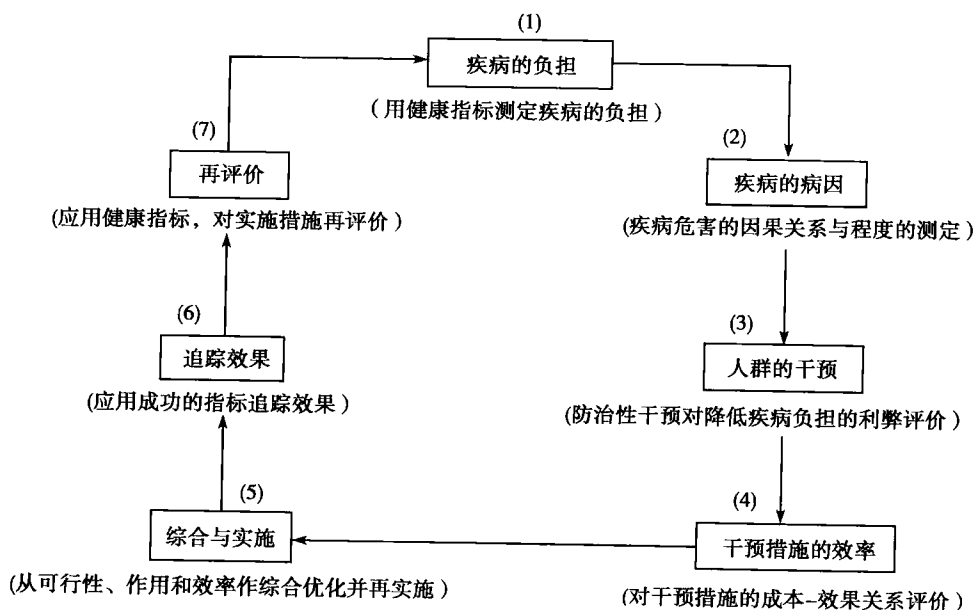


图 1-2 临床流行病学效应测量环

## 二、促进临床循证医学实践，提高医疗水平

由于当代社会经济的快速发展，科技的日益进步，特别是生物科学与信息技术的突飞猛进，推动着医学及人类健康水平的提高。为了使“人人享受健康”（health for all），必须要大量真实、重要并且能较为普遍适用的医学研究的成果，用以提供为人民的健康服务，这是 21 世纪医学研究人员面临的巨大挑战！因此，国际及国内医学界为加强医学研究正付出着巨大的努力，促进了丰富多彩的新药、新疗法、新诊断不断地问世，并涌向临床医学领域。有的成果确实可以提高临床诊治水平，促进临床医学水平的进步；有的非但无效，反而可能给患者带来严重的不良反应；或者既往认为有临床价值且被普遍接受的治疗措施或药物，但经严格研究证明为无效或无益者。因此，临床医生迫切需要从新的成果中去辨别真伪，去伪存真以提高自己临床诊治水平，对此临床流行病学为循证医学提供了对疾病病因、诊断、治疗及预后的一系列评价的原则、标准与方法。

临床流行病学的理论和方法既用于临床医学研究，促进了优秀研究成果的产出——即“创证”或者称之为“创造新知识”；同时又用于循证医学指导临床实践，即通过应用临床流行病学有关严格分析和评价的标准，去识别最新最佳研究成果（证据），联系临床实际予以应用，促进临床诊治水平的提高，这个过程可谓“用证”或者称之为“将真知转化到医疗实践”。“创证”为的是“用证”，而“用证”更需要不断创造更好的“证”！两者是互为促进的整体，而临床流行病学则是它们的理论和方法学的基础。

## 三、服务于医学教育，培养高质量的人才

临床流行病学的精髓是培养医学生的科学思维，从而为今后从事研究工作打下基础。因此，对医学生开设临床流行病学的科研设计课程，着重学习临床科研的基本理论、基本知识和基本方法，通过理论联系实际，提高学生鉴别真伪的科学批判能力，这对培养高质量的医学生，使之在毕业后的临床医疗实践中，更好地搞好临床医疗和科研工作，对于发展临床医学具有重要的战略意义。



临床流行病学是临床医师在医学实践中从事科研、医疗及教学十分有用的工具和科学的方法。学习、掌握和创造性的应用它，将会更加开阔自己的视野，活化自己的思维，改进临床工作的质量，提高自己的学术水平。这就是临床流行病学作为临床基础学科的实质。

(王家良 李绍忱)

# 第二章 疾病负担与临床研究重点的确定

## 第一节 疾病负担的概念

### 一、定义

疾病负担 (burden of disease) 是指疾病 (或伤害)、早死对患者、家庭、社会和国家所造成对健康和 (或) 经济、资源方面的损失程度。

它是一种衡量不同的疾病对人群健康伤害程度的国际通用健康指标。在一个国家或地区可以应用这一指标值的大小, 进行依次排序, 对指导全局的有关疾病的研究和防治, 有着重要的导向性价值 (图 2-1)。

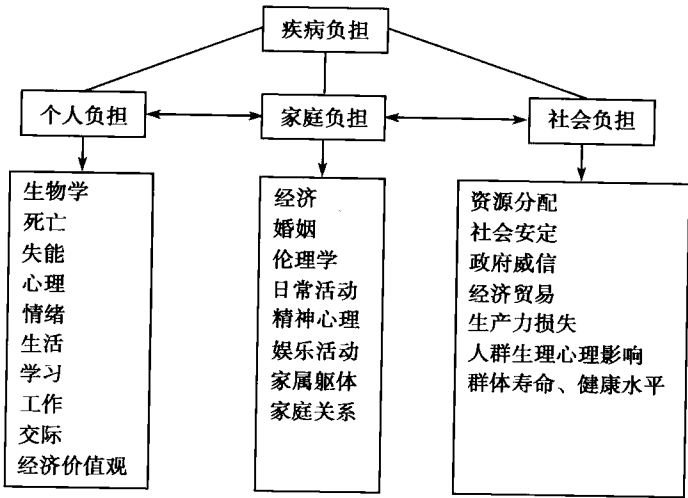


图 2-1 疾病负担模式图

周尚成《国外医学》卫生经济分册 2004, 21 (4): 147

由于不同国家或地区的经济发展水平、人文习俗、生态环境、种族等, 情况各异。因而, 疾病的负担显然不一, 同时上述等方面的情况动态变化, 也会促进疾病谱及其疾病负担发生相应变化, 因而 1993 年世界银行在发展报告中, 提出了“全球疾病负担”(global disease burden, GDB) 的概念, 用于研究国家控制疾病的优先重点及确定基本的一揽子卫生服务。

### 二、疾病负担的特点

1. 健康是一个复杂而又抽象的生物学现象, 其结局是多维的。所以疾病负担的度量应该包括对生命数量和生命质量以时间为单位的综合度量。即对健康的多维测量。
2. 不同时期研究疾病负担的思路和使用的方法、指标有所不同。疾病负担研究从原先单纯考虑死亡损失, 已发展到整合死亡和失能两方面来评价。疾病负担综合性评价, 使其研究得到了一定的发展。
3. 不同年龄组的人群疾病谱不同, 所以疾病负担也有较大的差异: 人从出生到老,



当处于不同的年龄阶段,所患疾病往往有其特点,因而疾病负担也有特色,由此造成的影响也就各有意义。如1岁以内的婴儿主要死亡原因是:①出生时低体重(小于2500g);②围产期与分娩有关的疾病;③先天性出生缺陷;④呼吸系统疾病;⑤传染病。婴儿早死虽然对年轻组人群的预期寿命可造成严重影响,但对劳动生产率没有影响,因此疾病的经济负担并不大。相反,25~64岁的青壮年,他们大多死于心脏病、肿瘤、脑血管病(脑卒中)、肝硬化、意外伤害等,因此造成预期寿命的损害比婴儿或儿童小,但他们是主要劳动力,因而造成的社会经济损失较大。65岁以上的老年人多死于心脏病、肿瘤、脑卒中、糖尿病等老年病和慢性病,对预期寿命和对社会经济的损失相对较轻。在发展中国家,加强医疗卫生服务的力度,针对人群中不同年龄阶段所患的多发常见疾病的特点采取医疗保健措施,是减少早死和伤残、降低疾病负担的重要举措。

### 三、研究疾病负担的意义

1. 从社会宏观角度研究疾病负担,可帮助我们了解不同人群和不同疾病给社会带来的负担,是确定卫生工作重点的基础。同时可以帮助我们针对受损害的目标人群,进行社区诊断(community diagnosis),确定包括医疗和预防在内的卫生服务重点。

2. 疾病负担这一研究方法的应用,在医疗实践和卫生管理工作中发挥着巨大的积极作用。近年来,已扩至全球范围,即应用“全球疾病负担”(GDB)的概念,研究世界各国及不同地区的疾病负担,并进行比较性的研究,故为卫生事业全球化的标志。

3. 对疾病负担的研究,可帮助我们了解疾病对人群的危害程度和规律,借此用于评价其危害性。

4. 可为卫生干预、卫生规划、卫生部门评价及卫生计划,提供一个可比较的科学测量标准。从而对确定疾病的医疗和预防决策提供可靠的依据。

## 第二节 疾病负担测量的指标与意义

### 一、发病指标

常用测量疾病负担的主要指标包括:发病指标、死亡指标和残疾失能指标等三个方面。

#### (一) 发病率 (incidence)

1. 定义 表示在一定期间内,一定人群中某病新病例出现的频率。

$$\text{发病率} = \frac{\text{一定期间内某人群中某病新发病例数}}{\text{同时期暴露人口数}} \times k$$

$k=100\%$ , 1000/千, 或 10000/万……

观察时间单位可根据所研究的疾病病种及研究问题的特点决定。通常多以年表示。

2. 分子与分母的确定 分子是一定期间内某一人群中的新发病人数。但对发病时间难确定的一些疾病可将初次诊断的时间作为发病时间,如恶性肿瘤、精神病等。分母中所规定的暴露人口是指可能会发生该病的同一人群的总人口数,对那些不可能患该病的人(如传染病的非易感者、已接种疫苗有效者)不应计入分母内,但实际工作中不易实现,当描述某些地区某集团的某病发病率时,分母多用该集团该时间内的平均人口。如观察时间以年为单位时,可为年初与年终人口之和的平均人口数或以当年7月1日的人口数表示。

发病率可按不同特征(如年龄、性别、职业、民族、种族、婚姻状况、病因等)分别



计算,此即发病专率。由于发病率的准确度可受很多因素的影响,所以在对比不同资料时,应考虑年龄、性别等的构成,进行发病率的标化。

3. 应用 在流行病学中,可用作描述疾病的分布,它能反映疾病发生的比率,它的变化意味着病因因素的变化。通过比较不同人群的某病发病率来帮助确定可能的病因,探讨发病因素,提出病因假说,评价防治措施的效果。发病率的准确性取决于疾病报告,登记制度以及诊断的正确。

### (二) 罹患率 (attack rate)

该指标和发病率一样,也是人群新病例数的指标。通常多指在某一局限范围,短时间内的发病率。观察时间可以日、周、旬、月为单位。适用于局部地区疾病的暴发,食物中毒、传染病及职业中毒等暴发流行情况。其优点是可以根据暴露程度精确地测量发病几率。

### (三) 患病率 (prevalence rate)

1. 定义 患病率也称现患率。是指某特定时间内总人口中某病新旧病例所占比例。患病率可按观察时间的不同,分为期间患病率和时点患病率两种。时点患病率较常用。通常患病率时点在理论上是无长度的,一般不超过一个月。而期间患病率所指的是指特定的一段时间,多超过一个月。

$$\text{时点患病率} = \frac{\text{某一时点一定人群中现患某病新旧病例数}}{\text{该时点人口数 (被观察人数)}} \times k$$

$$\text{期间患病率} = \frac{\text{某观察期间一定人群中现患某病的新旧病例数}}{\text{同期的平均人口数 (被观察人数)}} \times k$$

$k=100\%$ , 1000/千, 或 10000/万……

期间患病率实际上等于某一特定期间开始时患病率加上该期间的发病率。

2. 应用 患病率通常用来表示慢性病的发生或流行情况。影响患病率升高的因素包括:①病程延长;②未治愈者的寿命延长;③新病例增加(即发病率增高);④病例迁入;⑤健康者迁出;⑥易感者迁入;⑦诊断水平提高;⑧报告率提高。

影响患病率降低的因素包括:①病程缩短;②病死率高;③新病例减少(发病率下降);④健康者迁入;⑤病例迁出;⑥治愈率提高。

当某地某病的发病率和该病的病程在相当长时间内保持稳定时,患病率、发病率和病程三者的关系是:

患病率=发病率×病程,即:  $P=ID$

式中:  $P$ ——患病率;

$I$ ——发病率;

$D$ ——病程。

这可用于推算某些疾病的病程。患病率常用于表示病程长的慢性病发病的情况,可为医疗设施规划,医疗质量的评估和医疗费用的投入等提供科学的依据。

## 二、死亡指标

### (一) 死亡率 (mortality rate)

1. 定义 表示在一定期间内,在一定人群中,死于某病(或死于所有原因)的频率。是测量人群死亡危险最常用的指标。其分子为死亡人数,分母为发生死亡事件的总人口数(通常为年中人口数)。常以年为单位。多用千分率、十万分率表示。

$$\text{死亡率} = \frac{\text{某期间内 (因某病) 死亡总数}}{\text{同期平均人口数}} \times k$$

$k=100\%$ , 1000/千, 或 10000/万……



死于所有原因的死亡率是一种未经调整的率，也称死亡粗率（crude death rate）。死亡率也可按不同特征如年龄、性别、职业、民族、种族、婚姻状况及病因等分别计算，此即死亡专率。计算时应注意分母必须是与分子相对应的人口。对不同地区死亡率进行比较时，须注意不同地区人口构成不同，而存在差异，故需将死亡率进行标化后，才可进行比较。

2. 应用 死亡率是用于衡量某一时期，一个地区人群死亡危险性大小的一个指标。例如发展中国家死亡率水平仍然高得令人难以接受，儿童死亡率较发达国家高出 10 倍，如果贫穷国家的儿童死亡率能降低到富裕国家的水平，则每年就会少死 1100 万儿童。在这些可防止的死亡中，几乎有一半是因腹泻和呼吸系统疾病造成的，营养不良则又雪上加霜，使其加重。此外每年有 700 万成人死于可防止的或用较低费用就可以使其治愈的疾病，如仅结核就可致死 200 万人，约 40 万妇女死于妊娠及其妊娠并发症。发展中国家母婴死亡率平均是发达国家的 30 倍。

某些病死率高的恶性肿瘤，其死亡率与发病率十分接近，其死亡率基本上可以代表其发病率，而且其死亡率准确性高于发病率，因此常用作病因探讨的指标。

死亡专率可用于提供某病死亡在人群、时间、地区上的变化的信息，以用于探讨病因和评价防治措施。

## （二）病死率（fatality rate）

1. 定义 是表示一定时期内（通常为 1 年），患某病的全部病人中因该病死亡者的比例。

$$\text{病死率} = \frac{\text{某时期内因某病死亡人数}}{\text{同期患某病的病人数}} \times k$$

$$k = 100\%$$

2. 应用 病死率表示确诊疾病的死亡概率，它可表明疾病的严重程度，也可反映医疗水平和诊断能力，通常多用于急性传染病，较少用于慢性病。但是在比较不同医院的病死率时，须格外小心。因为医疗设备好，规模较大的医院接受危重型病人比较小的医院要多，因而大医院有些疾病的病死率可能高于小医院。所以用病死率作为评价不同医院的医疗水平时，要注意可比性。

# 三、残疾失能指标

## （一）病残率（mobility rate）

某一人群中，在一定期间内每百（或千、万、十万）人中实际存在的病残人数。即是指通过询问调查或健康检查，确诊的病残人数与调查人数之比。可说明病残在人群中发生的频率。也可对人群中严重危害健康的任何具体病残进行单项统计。它是作为人群健康状况的评价指标之一。

$$\text{病残率} = (\text{病残人数} / \text{调查人数}) \times k$$

$$k = 100\%, \text{或 } 1000/\text{千} \dots\dots$$

## （二）潜在减寿年数（potential years of life lost, PYLL）

为了避免死亡年龄不同所造成的影响，以便能更加合理地反映和评价疾病造成的死亡负担，1982 年美国疾病控制中心（CDC）提出应用潜在减寿年数这一指标。

1. 定义 潜在减寿年数是指某病在某年龄组人群死亡者的期望寿命与实际死亡年龄之差的总和。它是以期望寿命为基础，计算不同年龄死亡所造成的潜在寿命损失年，强调了早死对健康的影响，故能定量地估计了疾病造成早死的程度。PYLL 计算是对每例死亡者计算其死亡时的实际年龄与其相应的期望寿命年之差，再取总和。



## 2. 计算公式

$$PYLL = \sum_{i=1}^e a_i d_i$$

式中： $e$ ——预期寿命（岁）。

$i$ ——年龄组（通常计算其年龄组中值）。

$a_i$ ——剩余年龄， $a_i = e - (i + 0.5)$ ，意义为：当死亡发生于某年龄（组）时，至活满  $e$  岁还剩余的年龄。由于死亡年龄通常以上一个生日计算，所以尚应加上一个平均值 0.5 岁。

$d_i$ ——某年龄组的死亡人数。

潜在减寿年数是根据死亡年龄对期望寿命有明显影响，平均死亡年龄大时，对期望寿命影响较小；反之，当平均死亡年龄小时，对期望寿命的影响则较大这一原理提出的。在考虑死亡数量的基础上，以期望寿命为基准，进一步衡量死亡造成的寿命损失，强调了早亡对健康的损害。用潜在减寿年数来评价疾病对人群健康影响的程度，能消除死亡者年龄构成的不同对预期寿命损失的影响。

该指标不仅考虑到死亡率水平的高低，且考虑到死亡发生时的年龄估计对预期寿命的影响。该项指标可用来计算不同疾病或不同年龄组死者总的减寿年数。PYLL 计算是对每例死亡计算死亡年龄与潜在生命上限之差，再取总和。

### 3. 用途

(1) PYLL 是人群疾病负担测量的一个直接指标，也是评价人群健康水平的一个重要指标。

(2) 通过计算和比较各种原因所致的寿命减少年数，可反映出各种死因对人群的危害程度。

(3) 该指标可用于将某一地区（县）和另一标准地区（或省）相比较。

(4) 在卫生事业管理中，筛选确定重点卫生问题或重点疾病时，潜在减寿年数是一个很有用的指标，同时也适用于防治措施效果的评价和卫生政策的分析。

4. 优点 计算简便、易于理解，结果直观。潜在减寿年数的分析不仅可用于了解居民过去和现在的卫生健康水平，对卫生防疫工作的设计、计量和估计也是必需的。

5. 缺点 该指标只能反映死亡这样一种形式或结局的疾病负担情况。它对超出期望寿命的死亡的疾病负担却难以评价。如对老年人死亡的计算时，超过生命上限的老年人死亡对指标没有贡献，这与通常情况相悖，而且与社会对老年健康的重视和卫生资源对老年人的分配情况不相符合。

### （三）伤残调整寿命年（disability adjusted life year, DALY）

1. 定义 伤残调整寿命年是定量化的计算因各种疾病从发病到死亡造成的早死与残疾，对健康寿命年损失的一个综合指标。是将由于早死造成的损失和因伤残造成的健康损失二者结合起来进行测算的。

疾病可给人类健康带来包括早死与残疾（暂时失能与永久残疾）两方面的危害，这些危害的结果均可减少人类的健康寿命。DALY 的测算可以科学地对发病、失能、残疾和死亡进行综合分析，是生命数量和生命质量以时间为单位的综合度量。可克服通常方法中存在的片面性、主观性及局限性。

在研究人类疾病负担的过程中，为了全面地、准确地、客观地评价不同地区的卫生状况，有效地防治疾病，降低疾病负担以合理分配卫生资源，在世界银行和 WHO 有关部门的支持下，Murray 及 Lopez 提出了 DALY 这一指标，并开始应用于全球疾病负担的分析。一个 DALY 表示一个健康寿命年的损失，它反映的是疾病状态下损失寿命年或通过干预挽回的寿命年损失。



DALY 特别强调其指标对不同社区、不同国家和不同人种除了人的年龄和性别方面的差异之外,不考虑人的其他生物学及社会学等方面的差异,即有公正性和可比性,是其优点。

## 2. DALY 计算公式

一个人损失 DALY 数量的一般公式如下:

$$DALY = \int_{x=a}^{x=a+L} DC_x^{-\beta x} e^{-r(x-a)} dx$$

式中:  $x$ ——年龄;

$a$ ——发病年龄;

$L$ ——残疾(失能)持续时间或早死损失的时间;

$D$ ——残疾(失能)权重(0~1);

$DC_x^{-\beta x}$ ——该指数函数可用于计算不同年龄的生存时间;

$r$ ——贴现率;

$e^{-r(x-a)}$ ——连续贴现函数;

$\beta$ ——年龄权重函数的参数。

根据短暂性失能或永久性残疾的不同严重程度,可将残疾及失能分成 6 类并赋予不同的权重值(表 2-1)。0 代表完全健康,1 代表死亡,权重值在 0~1 之间。当发生疾病遗留短暂性或永久性残疾时,其剩余的期望寿命年应乘以残疾权重进行折算。

表 2-1 残疾分类及权重值

	残疾水平	残疾权重值
1.	在下列方面至少有一项活动受限:娱乐、教育、生育、就业	0.096
2.	在下列方面有一项大部分活动受限:娱乐、教育、生育、就业	0.22
3.	在下列方面有两项或两项以上活动受限:娱乐、教育、生育、就业	0.40
4.	在下列所有方面大部分活动受限:娱乐、教育、生育、就业	0.60
5.	日常生活如做饭、购物、从事家务,均需靠器具的帮助	0.81
6.	日常生活如吃饭、个人卫生、大小便均需人帮助	0.92

(CL Murray, 1994)

## 3. 应用

(1) 导向宏观的研究和防制决策:根据不同地区、人群的特点(年龄、性别等),以及有关疾病的 DALY 信息,可对疾病研究的重点选择以及重点疾病的防治决策提供依据,有导向价值,如通过干预后的 DALY 动态变化,又可用于效果评价以及疾病谱动态指标。

目前,国际上常用伤残调整寿命年(DALY)来估计疾病负担,图 2-2 及表 2-2 均显示全球慢性非传染性疾病的负担超过了传染病的疾病负担。

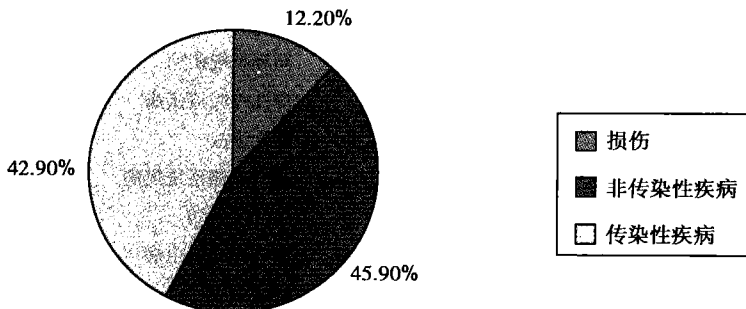


图 2-2 2001 年全球疾病 DALY 统计

资料来源: WHR2002





表 2-2 世界不同年代 DALY 的主要原因

次序	1990 年	2000 年	2020 年 (预测)
1	肺炎及其他呼吸道疾病	缺血性心脏病	心脏病
2	腹泻病	下呼吸道感染	抑郁症
3	新生儿出生儿童疾病	脑血管病	车祸
4	抑郁症	慢性阻塞性肺病	脑血管意外
5	心脏病	艾滋病	支气管炎及肺气肿
6	脑血管意外	围产期疾病	肺炎及其他呼吸道疾病
7	结核病	腹泻病	结核病
8	麻疹	结核病	战争
9	车祸	车祸	腹泻病
10	先天缺陷	支气管肺癌	艾滋病

又如 1992 年上海市的疾病负担构成, 一年共损失 162.52 万个 DALY, 占整个 DALY 损失的 51.32%, 伤残占 48.68%。其中, 传染性疾病占 6.31% (10.25/162.52), 非传染性疾病占 84.14% (136.74/162.52), 外伤性疾病占 9.55% (15.53/162.52)。说明上海市疾病模式已经由传染性疾病转型为以非传染性疾病为主的疾病过程。因此这些 DALY 信息当然对宏观的卫生研究和疾病重点的防治有重要的导向意义。

(2) 临床经济评价。DALY 用于成本效果分析, 研究不同病种, 不同干预措施挽回一个 DALY 所需的成本, 以求采用最佳干预措施来防治重点疾病, 使有限的资源产生更大效益。

(3) 计算疾病负担的顺位。DALY 注重伤残对寿命质量的影响, 因此 DALY 的顺位较之以前常用的死因顺位方法有很大区别。表 2-3 显示 1992 年上海市前 10 位疾病负担原因与前 10 位死因的比较。由于 DALY 损失综合考虑了伤残和死亡两方面的因素, 因而对一些死亡率较低, 但严重影响人类生存质量的疾病, 如精神疾患给予了较高的伤残权数。同理, 先天异常和一些由非致死性意外事故引起的中老年常见的肌肉骨髓系统疾病, 虽然对生存质量有很大的影响, 在前 10 位疾病负担中分别排列第 10 位和第 6 位, 却未出现在前 10 位的死因中。这是符合现代“生物-心理-社会医学”模式的特点的。

表 2-3 1992 年上海市前 10 位疾病负担原因与前 10 位死因比较

顺位	前 10 位死亡原因		前 10 位疾病负担原因	
	病因	率 (1/10 万)	病因	DALY/千
1	循环系统疾病	221.57	精神疾患	30.52
2	恶性肿瘤	183.94	恶性肿瘤	22.11
3	慢性呼吸道疾病	152.40	循环系统疾病	17.79
4	意外事故	37.90	慢性呼吸道疾病	12.27
5	消化道疾病	27.05	意外事故	9.38
6	精神疾患	20.51	肌肉骨骼系统疾病	5.36
7	传染病寄生虫病	16.83	消化道疾病	4.93
8	糖尿病	12.89	传染病寄生虫病	3.90
9	蓄意事故	10.42	糖尿病	3.12
10	生殖泌尿系统疾病	9.57	先天异常	2.69

(夏毅等, 2000)



不同年龄引起的死亡和残疾的相对生命价值是不同的, 一个健康生命现在的损失价值与将来的损失价值是不等值的。众所周知, 青壮年工作 and 生产劳动对社会的贡献大于老人和儿童。所以其生命价值也高于老人和儿童, 假设正常寿命年的相对价值等于或大于 1.0。从出生时的 0 开始, 急剧上升, 在 25 岁时达最高值, 之后随年龄增加逐渐下降。小于 10 岁的儿童或大于 50 岁的老人其寿命相对价值均小于 1.0, 见图 2-3。每个特定年龄因死亡而损失的伤残调整寿命年也不同。如一个新生女婴死亡时损失 32.5 个 DALY, 30 岁的成年妇女死亡损失 29 个 DALY, 而在 60 岁死亡时则损失 12 个 DALY。通常男性较女性稍低 (图 2-4)。

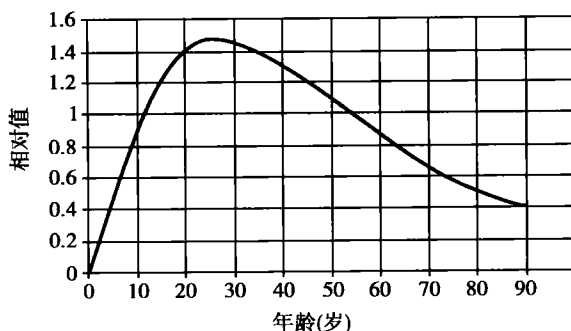


图 2-3 不同年龄每一寿命年的相对价值  
(据 World bank, 1993)

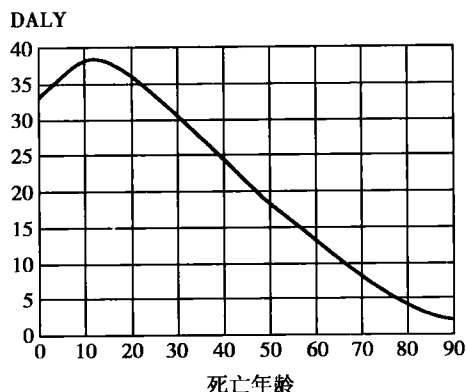


图 2-4 不同年龄死亡时减少的 DALY (女性)  
(据 World bank, 1993)

1990 年全世界共损失 136 亿个 DALY。这相当于 4200 万个新生婴幼儿死亡, 或等于 8000 万个 50 岁的人死去。因早死导致的损失占整个 DALY 损失的 66%, 而由残疾造成的损失占 34%。在发展中国家约 67% 的 DALY 损失是由早死所致。表 2-4 示全球疾病负担按性别、疾病种类及损失种类的分布情况。

表 2-4 1990 年不同性别、原因和损失类型的疾病负担 (百万 DALY)

性别和结果	疾病类型		
	传染性*	非传染性	外伤
男性			
死于未成年	259	152	70
残废	47	146	39
女性			
死于未成年	244	135	33
残废	74	142	20

\* 包括由母亲传染的疾病和分娩前后传染的疾病 (据 World Bank, 1992)

表 2-4 结果提示人们对传染性疾病给人类带来的危害决不能掉以轻心, 这些儿童疾病占全球疾病负担的 1/4, 而这些疾病有些是很容易预防或很容易治愈的。对成人来说传染病的威胁也不容忽视, 如性传播疾病和结核病加起来约占全球疾病负担的 7%。

#### 4. DALY 的缺陷

(1) 虽然 DALY 可将疾病造成的早死和失能合并考虑来反映疾病对人群造成的负担, 但由于该指标存在有主观性。



(2) DALY 未能对疾病给人群造成的心理社会负担、家庭负担和社会负担予以充分评价。

#### 5. DALY 应用时的注意点

(1) 健康寿命年损失包括早逝和残疾（暂时性失能与永久残疾）两个方面。在计算非致死性疾病的健康寿命年损失时，宜根据不同的疾病严重程度给予相应的权重。

(2) 不同性别、年龄的寿命价值是不等价的，发病或死亡造成的健康寿命年损失是不同的。而其他影响健康结局的因素诸如种族、社会阶层、职业、教育等，因涉及复杂的问题而不再予以考虑。

(3) 健康寿命的现在损失与将来损失的社会价值也是不等价的，因而，根据适时的具体情况，应考虑适当调整贴现率，以减少低年龄组死亡或残疾对总的健康寿命年损失的影响。

### (四) 健康寿命年

1998 年 Hyder 等人提出了一个试图将疾病的致死及致失能效果结合在一起的新的测量疾病负担的指标——健康寿命年 (health life years, HeaLY)。

$$HeaLY = L_1 + L_2$$

式中： $L_1$ ——该人群中因患某种疾病死亡而损失的健康寿命年  
 $= P \times I \times CFR \times [E(A_0) - (A_f - A_0)]$ ;

$P$ ——人群的总人数；

$I$ ——该人群中某种疾病每年每千人口的发病率；

$CFR$ ——该病的病死率；

$A_f$ ——因该病死亡时的平均年龄；

$A_0$ ——因该病发病时的平均年龄；

$E(A_0)$ ——年龄为 ( $A_0$ ) 时的期望寿命，采用标准期望寿命；

$L_2$ ——该人群中因患某种疾病失能而损失的健康寿命年  
 $= P \times I \times CDR \times D_e \times D_i$

$CDR$ ——患该病人因该病失能的比例；

$D_e$ ——失能权重；

$D_i$ ——此病平均病程。

计算 HeaLY 时应注意：①任何年龄的生命是等价的，如在 25 岁生存的一年和在 65 岁生存的一年是等价的。② HeaLY 用时间贴现对早逝和失能期间的健康寿命损失进行调整。

HeaLY 与 DALY 一样，均以发病为起点，以一种疾病发病后其自然史作为基本框架，来评价患病和死亡的综合效应。其计算公式更简化及易于理解。HeaLY 从疾病的发病开始，根据疾病的自然史考虑疾病引起死亡的情况及不同年龄组死亡的影响。同时更充分地考虑到发病期间失能对健康的影响。这对宏观地认识疾病和控制疾病有十分重要的意义。

应用 DALY 和 HeaLY 时，其不足之处在于所需资料在发展中国家常常不易得到，此外，由于采用标准期望寿命，也常使估计结果与实际略有差异。

### (五) 质量调整寿命年

质量调整寿命年 (quality adjusted life years, QALYs) 是一种健康状态和生命质量的正向综合测量指标。其基本思想是把生存时间按生存质量高低分为不同阶段，每阶段给予不同的权重 (0~1 间取值)，从而得到质量调整寿命年。一个 QALYs 反映一个健康生存年，即它可反映在疾病状态下或干预后剩余 (经过调整) 的健康寿命年数。例如，急性心肌梗死的患者，经治后，虽然活了 10 年，但是他的生存质量仅及健康人的 50%，因



此, 他的 QALY 为 5 年。QALY 是 20 世纪 80 年代后期才发展并逐步完善起来的。通常认为它是一个正向的指标。

### 第三节 疾病负担测定的资料来源与方法

分析研究疾病负担的关键, 是需要有足够的、有价值的关于疾病与健康及其相关的信息资料。获取的主要途径是通过描述流行病学方法, 调查疾病在不同人群、不同地区、不同时间的分布情况, 获取所需的数据、信息资料, 并计算疾病负担有关指标。此外, 通过常规、系统的疾病监测 (surveillance), 亦可获得有关疾病负担的资料。

#### 一、疾病监测

疾病监测也叫公共卫生监测, 是了解疾病在人群中发生、发展、分布及消长规律和长期动态趋势的重要方法, 可对制定、改进防治对策和措施提供重要的科学依据, 有利于卫生事业的科学决策。通过在人群中长期、系统、持续地收集有关资料, 进行科学的分析与评价。

随着历史的发展, 疾病监测的范围与作用也在发生变化, 已从最初对个人监测发展到广泛应用到公共卫生领域, 包括危险因素, 行为等有关现象进行检测, 目前则更紧密的将监测与社区内的疾病控制与预防活动联系起来, 这样来源于基层人群的资料, 将会使疾病监测日趋完整。

##### (一) 监测目的

监测目的是为了对卫生决策、卫生干预提供信息。具体包括以下几点: ①估计人群中疾病发生的频率及其在人群、时间、地区的分布, 动态地监测疾病发展的趋势。②通过对人群中现患率, 发病率变化的长期监测的分析, 评价干预措施的效果。③确定某病的高危人群和低危人群, 为制定合理的干预措施提供依据。④确定影响疾病传播、蔓延和发生发展的危险因素。⑤为制订防治疾病的策略确定重点。

##### (二) 监测方式

监测方式主要包括 4 种:

1. 一般人群监测 是对人群中, 尤其是亚人群中疾病分布, 长期变化趋势进行监测, 了解人群中疾病变化规律。我国现行的疫情报告制度即属于这种监测方式。

2. 哨点监测 是在总体 (如全国或某一地区) 的人群中, 通常以 1% 的人口比例, 按照设计要求, 在不同地点选定若干疾病监测哨点, 依据疾病监测标准化的工作程序和指标, 系统地收集有关资料, 以反映人群的基本卫生状况。哨点选择应与该监测系统的监测目的相吻合。

选择哨点原则 首先是分层原则, 根据一定指标将城市、农村再细分为不同类型, 保证样本在不同卫生状况的地区其人口比例与全国相似, 分层设岗保证地理分布的均衡性。其次是有可行性, 由于监测工作是对该地区长期的观察, 而不是进行一次性的地区调查, 所以所选的地区应该必须具备一些基本条件: 即该地区须具备领导重视、组织健全、有保证正常工作的条件。如交通条件, 以便检查、培训。还要有经过培训素质较高、工作主动的工作人员, 否则收集不到所需要的信息资料, 不能反映真实情况。在同类地区可允许有不超过 15% 的少量不能胜任工作的监测点进行调换。

3. 被动监测 (passive surveillance) 是下级单位常规地将监测数据资料向上级机构报告。上级单位属被动接受。如常规法定传染病报告。这种监测方式虽然易于进行, 且比较普遍, 但其结果存在明显的质量的问题, 漏报情况较为严重。



4. 主动监测 (active surveillance) 是根据特殊需要, 上级单位亲自调查收集或要求下级单位尽力去收集有关数据资料。如 2003 年我国 SARS 流行状况的主动监测。

### (三) 疾病监测系统可提供的基础资料

1. 社会人口学资料 包括人口总数、年龄构成与性别比例; 劳动力总数, 就业人数及职业分类; 人口密度; 人口出生率及死亡率; 育龄人口生育率、计划生育率、人口增长率。

2. 社会经济学资料 国民生产总值和人均国民生产总值 (GNP); 人均收入, 人均消费水平及家庭卫生保健费用开支; 文化教育情况, 住房条件等。

3. 基本健康水平资料 粗死亡率, 新生儿死亡率, 婴儿死亡率, 儿童死亡率, 期望寿命; 法定传染病发病率和死亡率; 主要职业患病率; 主要女性所患疾病的患病率; 儿童青少年主要疾病发病率、患病率; 死因构成; 低体重新生儿发生比例; 幼儿及少年生长发育情况。

4. 卫生资源及卫生服务资源资料 卫生保健机构数, 设施数, 床位数, 医药卫生保健人员数及结构比例; 医疗保险制度, 各类医疗保健卫生费用支付的比例; 平均人均医疗卫生服务人口数及服务半径; 卫生经费来源及投入总数, 人均金额; 人均门诊费用, 住院费用; 医疗卫生保健机构基本设备水平及利用率; 计划免疫覆盖率; 特殊疾病的防治状况; 新法接生率, 新生儿保健率, 儿童保健率, 孕妇保健率等。

### (四) 疾病监测步骤

1. 收集资料 按照所监测疾病的项目, 尽可能地完整、准确地收集有关流行病学、临床以及社会人口学相关资料。

2. 资料分析 包括研究疾病自然史, 变化趋势, 流行过程的影响因素, 薄弱环节及防治效果等。要准确, 防偏倚求真实。

3. 印刷和反馈资料 资料分析结果上报及通知有关单位和个人以便采取相应防治措施。

### (五) 疾病监测的种类

1. 传染病监测 每个国家都有其各自规定的不同监测病种。2004 年我国颁布的“传染病防治法”中规定报告的病种分为甲、乙、丙三类共 37 种。甲类传染病包括鼠疫、霍乱, 因其传染性强, 病死率高, 易引起大流行。所以应对其采取强制管理措施。乙类传染病的危害较甲类稍小, 但对艾滋病和炭疽病人也应采取某些强制性防治措施。对丙类传染病主要是在监测区中进行监测管理。WHO 把疟疾、流行性感冒、脊髓灰质炎、流行性斑疹伤寒和回归热列为国际监测的传染病。

传染病监测的主要内容包括: 监测人群基本情况, 人口资料, 出生、死亡、生活习惯、风俗习惯、生产情况、文化教育水平, 经济状况, 居住条件, 人口流动状况; 传染病的人群、地区、时间的动态分布情况; 人群的易感性; 传染源, 动物宿主, 媒介昆虫; 病原体的毒力、耐药性与型别; 防疫措施效果评价; 预测流行, 开展病因学与流行病规律研究。

2. 非传染病监测 非传染病监测日益受到人们重视, 不同国家不同地区其监测内容不同, 我国已开始对恶性肿瘤、心血管疾病、高血压、出生缺陷等进行检测, 根据不同监测目的, 其监测内容也不同, 如营养监测、围生期监测、婴儿死亡率监测、水质监测及食品监测等。

## 二、疾病统计

疾病的统计资料是制定卫生政策, 防治规划及评价卫生工作质量和效果的重要依据,



也可为病因学研究提供基础资料。

做好疾病统计工作必须具备下列三项条件：

1. 要有一个统一的、较完整的、得到人们公认的国际疾病分类 (international classification of diseases, ICD)，这样才能使疾病统计资料得以正确地整理，并使其资料更具可比性。

疾病分类是按照既定的统一标准将疾病、损伤和死因纳入相应的一定类目的一种系统分类。ICD 可用于对记载在多种类型的健康和生命记录上的疾病和其他健康问题进行分类。ICD 的目的是允许对不同国家或地区及在不同时间收集到的死亡和疾病数据进行系统的记录、分析、解释和比较。ICD 可将疾病诊断和其他健康问题用字母数字进行编码，从而易于对数据进行储存、检索和分析，并可提供相应的其他健康状况信息。对于流行病学和许多健康管理问题来说，ICD 已成为国际标准的诊断分类。它适用于对各人群组一般健康状况的分析、对疾病发病和患病的监测以及与其他变量有关的健康问题。

虽然这种统一的疾病分类对不同地区间的疾病和死因资料的比较提供了便利和可能，但是随着科学的进步和发展，人们对疾病和死因的认识也在不断发生变化，新的认识、新的病种也在不断出现，这就增加了重新制定或修改分类标准的可能性与必要性。ICD 自创建以来，约每隔 10 年修订一次，现已有 ICD<sub>10</sub> 问世。我国现行的仍以 ICD<sub>9</sub> 为主。

ICD<sub>10</sub> 的创新之处在于其采用了一种字母数字编码方案，即在四位数水平上的第一位使用一个英文字母，后面跟着三个数字。第四位数在一个小数点之后。因此可能的编码数字范围从 A<sub>00.0</sub> 到 Z<sub>99.9</sub>。其效果是使编码框架的容量比 ICD<sub>9</sub> 扩大一倍多，并使绝大多数能使用唯一的一个或一组字母，每个字母可提供 100 个三位数类目。在可使用的 26 个字母中，已使用了 25 个，字母 U 被留下来为增加和更改使用。编码 U<sub>00</sub>-U<sub>49</sub> 用来暂时分配给某些病因不明的新疾病。编码 U<sub>50</sub>-U<sub>99</sub> 可用于研究，如为了一个特殊项目而检验一种替代的亚分类时。ICD<sub>10</sub> 共分为 21 章。

为了量化失去健康寿命的全部损失，世界银行和 WHO 共同研究指出，可以 ICD<sub>9</sub> 为基础，找出 190 种主要疾病作为 DALY 的研究病种，这些疾病几乎可以包括全部死亡和 95% 的伤残。

2. 正确地规定疾病统计指标，以便从几个必要的方面反映疾病统计本身的一些特征。
3. 正确地安排收集疾病统计资料的程序，以便保证取得完整可靠的原始资料。

### 三、人口学调查

主要调查研究人口数量、组成及其变动，了解居民健康状况及社会生活条件对居民健康的影响。人口学调查研究可分两类。

1. 人口静态研究 指对某一时间断面上相对静止的人口状态。是研究人口数量的主要方法，其资料主要靠调查方法取得，人口普查可获得某一时间断面上人口数。实际工作中，不仅需要人口总数，也需要分组的人口数。人口分组采用社会标志及年龄、性别标志。社会标志以不同社会阶层、经济收入、职业、文化水平、民族等项为主。人口年龄、性别组成是将人口按年龄、性别分组。

2. 人口动态研究 是看人口的变化。所谓人口动态是指在一定时间内由于出生、死亡和迁移所形成的变动。包括机械变动和自然变动。在医学中应用较多的是“自然变动”。表示人口自然变动的指标，可分为表示出生，死亡，平均寿命及再生育水平等四类。如我国近 50 年来死亡率急剧下降。20 世纪 60~70 年代出现出生高峰，之后出生率开始明显下降，这将会产生在未来的几十年内我国人口年龄构成将会出现中老年人口



比重激增的明显变化。疾病负担有效性的估计,取决于需要有流行病学调查资料 and 人口资料的准确性。由于评价疾病负担所需的资料很多,甚至在发达国家流行病学所需的资料仍有许多空白,在发展中国家也有许多是不可靠的。所以,保证资料的质量是进行各种调查的前提。

#### 四、残疾失能的估计

如前所述,疾病负担常采用潜在减寿年数(PYLL)及伤残调整寿命年(DALY)等指标进行估计(详见第二节)。

#### 五、经济负担的估计

疾病的后果可造成死亡和伤残。显而易见,死亡是个十分明确的指标,易于被人认识,而伤残却是个模糊的指标,所以过去的一些研究分析常常忽略伤残带来的损失,但应用DALY可以计算因伤残带来的寿命损失和社会经济损失。虽然老年人和儿童的健康寿命损失DALY相对较高,但从经济损失来看,15~44岁年龄组的寿命损失对社会经济的影响更大,从创造社会财富的角度来说,儿童由于疾病和死亡造成的生命损失并未带来社会经济的损失。

由于不同疾病所造成的经济负担不同,所以常常可通过对不同疾病所造成的经济负担来确定主要的卫生问题。例如非传染病造成的经济损失和DALY损失均远远超过传染病和意外伤害。如表2-5所示:1990年全国非传染病造成的经济损失为8998.6亿元、累积寿命损失为1162.9百万人年,而传染病造成的经济损失则为3864.3亿元、累积寿命损失为499.4百万人年。

同样,也可通过成本-效果分析(cost-effectiveness analysis, CEA)来分析、评价或比较成本消耗后获得的有用或有效果。就是要以最少的资源投入获取最大的效益,最大的效益可以是最小的疾病负担。因此可用减少一个DALY所需的费用多少来评价卫生方案或干预措施的效果。例如有一个研究高血压治疗的项目,曾对60岁男性高血压病患者舒张压从110mmHg降低到90mmHg,延长一个寿命年需花费的成本为16330美元,而另有一个应用两种不同的降血脂药物治疗高胆固醇血症的项目,结果其延长一个寿命年花费的成本则分别为59000美元和17800美元,由此可见,前一个项目的经济效益较高。

还可应用类似的分析进行临床决策。如有一项治疗措施,花费3000美元即可挽回2个寿命年,所以每挽救一个寿命年需1500美元。而另一项新措施花费成本6000美元就可挽救3个寿命年,即每挽救一个寿命年需2000美元,所以该结果显示新措施的成本比老措施高,但前者挽救的寿命少。

研究不同国家、不同地区的疾病经济负担,对确定疾病防治的优先重点,合理分配有限的卫生资源,提供卫生计划和决策有着十分重要的意义。

衡量疾病造成的经济负担可通过计算直接费用和间接费用来表示。疾病的直接费用是由于患病而接受医疗服务所消耗的医疗服务人力资源和物质资源。包括年治疗费用或其他卫生服务费用、药品及医疗仪器检查、检验的费用等。通常可采用发病率和患病率的方法来计算直接费用。慢性病的经济损失多采用患病率的计算方法。

间接费用表示因病所造成的如饮食、营养、交通、陪护等其他方面的损失支出。当确定国家的主要卫生问题、优化卫生资源的配置,以便能更好地提高人群健康水平,发展卫生保健事业时,需注意不同地区经济发展的水平不一。所以应该依据本地区的卫生资源、卫生事业发展目标、政策及成本效用分析的结果来考虑。



表 2-5 1990 年全国按病因统计的疾病负担

疾病名称	DALY 损失 (百万)			经济损失 (亿元)	
	男性	女性	合计	男性	女性
传染病、妇科及围产期疾病	277.4	222.0	499.4	2146.5	1717.8
传染病和寄生虫病					
肺结核	24.4	34.7	59.1	188.8	268.5
性传播疾病	33.3	0.8	34.1	257.7	6.2
腹泻	21.7	20.7	42.4	167.9	160.2
儿童疾病	8.1	9.3	17.4	62.7	72.0
肺炎	2.8	4.0	6.8	21.7	31.0
肝炎	2.2	4.5	6.7	17.0	34.8
疟疾	0.0	0.1	0.1	0.0	0.8
热带病	2.3	3.8	6.1	17.8	29.4
沙眼	2.6	1.1	4.7	27.9	8.5
肠道蛔虫	30.6	32.6	63.2	236.8	252.3
呼吸道感染	69.0	60.0	129.0	533.9	464.3
妇科	25.0	0.0	25.0	193.5	0.0
围生期疾病	54.4	50.4	104.8	420.9	390.0
非传染性疾病	555.9	607.0	1162.9	4301.6	4697.0
肿瘤	72.0	113.1	185.1	557.1	875.2
糖尿病	4.1	3.6	7.7	31.7	27.9
营养及内分泌	38.6	27.7	66.3	298.7	214.3
神经及精神疾病	78.5	81.7	160.2	607.4	632.2
感觉器官	9.0	7.5	16.5	69.6	58.6
心血管	15.0	133.4	148.4	281.8	1032.2
呼吸道	85.4	96.0	181.4	660.8	742.8
消化道	37.8	49.9	87.7	292.5	386.1
生殖泌尿系统	12.8	21.8	34.6	99.0	168.7
肌肉骨骼系统	40.0	13.3	53.3	309.5	102.9
先天性异常	35.4	34.8	70.2	273.9	269.3
口腔	8.9	9.2	18.1	68.9	71.2
创伤	138.5	197.6	336.1	1071.7	1629.0
意外事故	84.7	148.3	233.0	655.4	1147.5
蓄意伤害	53.8	49.3	103.1	416.3	381.5
总计	971.8	1026.6	1998.4	7519.8	7943.8

(梁万年, 2002)





## 第四节 疾病负担指标在医学研究和决策中的应用

### 一、确定国家或地区性疾病研究的重点

科学准确的疾病负担指标，有利于国家或地区确定疾病研究和干预的重点。确定重点一般包括确定重点疾病、确定高危人群和主要的危险因素，只有这样才能有的放矢地发挥防治疾病，增进人群健康的目的。

确定研究重点时应注意：时间范围要明确；应估计到存在地区性差异，所以不同地区的重点可能不同；所确定的重点必须要明确，不能含糊不清；问题的顺位和复杂程度应该明确；所确定的重点疾病必须要有行之有效的防治办法，有经费的保障；选择的重点病种成本效益必须提高。

#### 确定重点疾病

1. 原则 重点疾病通常根据下列两方面情况确定：一是该病的发病频率高，其中包括发病率、患病率高。二是严重致残及高死亡率的疾病，即疾病负担严重的疾病。如世界部分国家全死因中，归因于高血压的疾病负担（DALY）从 1.4%~16.3% 不等，我国为 4.8%，是除了俄罗斯、日本和美国之外最高的国家（表 2-6）。

表 2-6 2002 年世界部分国家全死因中归因于高血压的 DALY 比例（%）

国家	男	女	合计
尼日利亚	1.3	1.5	1.4
美国	5.2	5.4	5.3
巴西	4.0	3.7	3.9
墨西哥	2.0	2.1	2.0
巴基斯坦	2.3	2.2	2.2
俄罗斯	14.1	19.6	16.3
印度尼西亚	4.0	4.5	4.2
印度	1.7	1.8	1.8
孟加拉国	3.6	2.9	3.2
日本	8.6	5.4	7.1
中国	4.8	4.8	4.8

确定重点疾病的有效干预研究应具备以下条件：①所确定的重点疾病，其流行病学问题已经或基本清楚。如其病原体、传染源、传播途径、易感人群及其分布特征等。②有行之有效的防治办法。③有保证这些办法实现的自然条件和社会条件。对具备这些有利条件而尚未有效控制者，则应列为有效干预的重点研究。

对那些有较高发病率和危害严重的疾病，但又可能被预防和被控制的疾病应列为防治工作的重点予以优先考虑，以期对疾病流行进行控制，减少危害程度，最终达到消灭的目的。

对有些疾病负担严重但至今尚无有效预防及控制措施的疾病，应列为临床流行病学以及基础医学研究工作的重点，以期弄清疾病的本质，研究有效的防治方法，以降低疾病负担。

2. 确定重点疾病的依据及方法 可根据疾病对人群健康不良影响的大小来确定重点疾病。主要表现在疾病引起的发病、死亡、减少寿命、致残、影响正常生产、学习和工作等方面。可根据疾病负担诸指标，如 DALY、PYLL、发病率、死亡率及其构成等，确定



重点疾病。应用这些指标列出疾病谱和死亡谱，可反映出某时期、某地区的重点疾病或不同时期的变化情况。例如 1999 年澳门地区前 10 位死因顺位见表 2-7，可作借鉴。

表 2-7 1999 年澳门地区死亡率（1/10 万）、标化死亡率、PYLL、DALY

死亡原因（ICD-9）	死亡人数	死亡率 (1/10 万)	标化死亡率 (1/10 万)	PYLL (人年)	DALY (人年)
肿瘤	366	84.33	65.24	3892.64	5254.69
心血管疾病	263	60.60	41.75	1132.80	2523.36
脑血管疾病	178	41.01	31.44	990.68	2310.24
呼吸系统疾病	168	38.71	27.66	990.93	2148.08
损伤和中毒	115	26.50	24.88	4005.69	2861.05
症状、体征和某些不明确情况	59	13.59	略	495.66	758.40
消化系统疾病	53	12.21	略	470.58	723.09
泌尿生殖系统疾病	41	9.45	略	205.89	502.89
传染病和寄生虫病	33	7.60	略	383.78	480.66
内分泌、营养代谢及免疫疾患	28	6.45	略	100.12	335.73

（陈穗芬，中国预防医学杂志，2001，12（2）：253-256）

此外，表 2-8 显示 2001 年我国部分城市前 10 位死亡原因。可看出男女前 10 位死因中，不论男性或女性，恶性肿瘤、脑血管病、心脏病、呼吸系统疾病、损伤和中毒均为最主要死因。过去传染病占主导地位逐渐转变为以非传染病为主。在制定疾病的重点防治和科研计划时必须考虑到这些疾病负担特点。

表 2-8 2001 年部分主要疾病死亡专率及死亡原因构成

顺序	死亡原因	合计死亡专率 (1/10 万)	占死因总人数 %	死亡原因	男性死亡专率 (1/10 万)	占死因总人数 %	死亡原因	女性死亡专率 (1/10 万)	占死因总人数 %
1	恶性肿瘤	135.59	24.93	恶性肿瘤	163.77	27.64	恶性肿瘤	106.27	21.55
2	脑血管疾病	111.01	20.41	脑血管疾病	118.33	19.97	脑血管疾病	103.39	20.97
3	心脏病	95.77	17.61	心脏病	97.50	16.46	心脏病	93.98	19.06
4	呼吸系统病	74.64	13.36	呼吸系统病	76.42	12.90	呼吸系统病	68.70	13.93
5	损伤和中毒	31.92	5.87	损伤和中毒	39.31	6.63	损伤和中毒	24.24	4.92
6	内分泌、营养代谢及免疫疾病	17.18	3.16	消化系统疾病	20.47	3.45	内分泌、营养代谢及免疫疾病	20.50	4.16
7	消化系统病	17.04	3.14	内分泌、营养代谢及免疫疾病	13.52	2.36	消化系统病	14.76	2.74
8	泌尿、生殖系病	8.55	1.57	泌尿、生殖系病	8.90	1.50	泌尿、生殖系病	8.18	1.66
9	精神病	5.37	0.99	神经病	5.94	1.00	精神病	5.60	1.13
10	神经病	5.20	0.96	精神病	5.15	0.87	神经病	4.43	0.90
	十种死因合计		92.00	十种死因合计		92.78	十种死因合计		91.02

注：统计范围包括北京等 36 个市全市或部分市区资料（中国卫生统计年鉴 2002）



其次可依据 DALY 及潜在减寿年数 (PYLL) 确定重点疾病, 这种方法不仅可揭示各种原因对该人群健康的危害程度, 也能反映出对各年龄组人群的危害大小。为确定不同年龄组重点疾病提供了科学依据。

例如, 从表 2-9 对我国 1990 年至 1994 年几种主要疾病的分析可以看出, 感染性疾病的重要性相对下降, 慢性病、意外伤害的重要性上升, 用 5 年平均死亡率, 5 年平均 PYLL 和 DALY 指标来描述感染性疾病 (包括产科、围生期疾病), 非感染性疾病及意外伤亡等三大类疾病的分布比例, 结果显示既往严重危害人类健康的传染病虽有所下降, 但依然占据相当比例, 慢性病、意外伤害也严重威胁着人群健康。

表 2-9 1990~1994 年我国几种主要疾病的平均死亡率、期望寿命减损年限、疾病负担占总死亡和本类疾病的比例

	平均死亡率 (10/万)			期望寿命减损年限 (PYLL)			疾病负担 (DALY)		
	死亡率	构成比 (%)	本类疾病构成比 (%)	例数	构成比 (%)	本类疾病构成比 (%)	数	构成比 (%)	本类疾病构成比 (%)
感染性、围产期及妇科疾病	51.98	10	—	199286	26	—	509.4	25	—
传染病	19.2	4	37	46454	6	23	250.6	12	49
肺结核	7.20	1	14	7983	1	4	59.1	3	12
肺炎	19.69	4	38	82128	11	41	113.1	6	22
围生期疾病	9.85	2	19	64932	8	33	198.2	10	39
窒息	5.29	1	10	35061	5	18			
早产	1.92		4	12601	2	6			
慢性病	412.18	79		379762	50		1167.7		58
肿瘤	84.13	16	20	87429	11	23	185.1	9	16
心血管	148.57	28	36	87854	11	23	281.8	14	24
疾病									
呼吸系统疾病	107.67	21	26	35511	5	9	181.4	9	16
COPD	120.83	20	25	31051	4	8	111.7	6	10
意外伤害	57.19	11		186633	24		336.1	17	
自杀	15.11	3	26	37614	5	20	69.8	3	21
交通事故	9.96	2	17	31017	4	17	45.3	2	13
溺水	6.78	1	12	33215	4	18	24.1	1	7
合计	524.56			765682			2013.2		

(杨功焕, 1996)

## 二、确定重点研究与干预的对象

### (一) 具有危险因素的人群

优先防治疾病的重点确定之后, 还必须要考虑该地区范围内哪些人群应被列入重点保护对象及重点高危人群。所谓高危人群通常指虽然每个人都处在患病和死亡的危险之中, 但有些人群却具有更高的危险性, 发生某些疾病或有产生严重的健康问题的危险。我们把这种具有发病危险和处于 (或暴露于) 该危险因素之下而尚未发病者的人群称之为高危



人群。

WHO 的《2002 年世界卫生报告》中指出,危害人类健康的最大原因为:高血压、吸烟、酗酒、胆固醇过高、肥胖症、营养不良、缺铁、危险性生活、无洁净水、固体性燃料释放的烟害等。表 2-10 显示世界 10 个主要疾病负担危险因素中的 5 个,即高血压、高胆固醇、肥胖、体力活动少和摄入水果、蔬菜少,是和饮食与体力活动有强联系的。大约有 75% 的 CVD 死亡可归因于大多数危险因素,而少数几个危险因素,如高胆固醇、高血压、飞盘、吸烟和饮酒,导致了大半的慢性病负担。

例如,有高血压、高血脂、超体重、高盐饮食和吸烟习惯的人群是心、脑血管疾病的高危人群。有癌前病变和受致癌因素影响的人群是好发恶性肿瘤的高危人群;近亲婚配是发生某些出生缺陷的高危人群。可依据流行病学信息和方法,确定高危人群,明确重点保护对象是制定相应的防治措施。

表 2-10 用 DALY 为指标衡量十大主要危险因素对疾病负担的百分位数

发展中国家				发达国家	
高死亡率国家	(%)	低死亡率国家	(%)		(%)
营养不良	14.9	饮酒	6.2	吸烟	12.2
不安全性行为	10.2	高血压	5.0	高血压	10.9
不卫生水	5.5	吸烟	4.0	饮酒	9.2
固体燃料室内污染	3.7	营养不良	3.1	高胆固醇	7.6
缺锌	3.2	超重	2.7	超重	7.4
缺铁	3.1	高胆固醇	2.1	蔬菜、水果低摄入	3.9
维生素 A 缺乏	3.0	固体燃料室内污染	1.9	缺乏体育活动	3.3
高血压	2.5	蔬菜、水果低摄入	1.9	药瘾	1.8
吸烟	2.0	缺铁	1.8	不安全性行为	0.8
高胆固醇	1.9	不卫生水	1.7	缺铁	0.7

## (二) 年龄特点

疾病负担的大小都与年龄有着明显的相关性。如婴儿死亡率相当高,随着年龄增长逐渐下降,5~14 岁时最低。14~40 岁逐渐上升,40 岁以后死亡率迅速增加,几乎每 10 年增加 1 倍。糖尿病、心血管疾病、关节炎、癌等可在高年龄组中多发,其发病率均随年龄的增加而升高。某些严重的出生缺陷发生亦与母亲的年龄有关,小于 15 岁或大于 35 岁孕妇所生畸形儿的几率较高,由此观之,确定重点疾病的研究,其被干预的对象选择,显然有着年龄的特点。

不同年龄组的疾病负担情况可为卫生事业管理部门提供制定防治策略与计划的依据。因腹泻引起的 80% 以上的 DALY 损失是因 5 岁以下儿童受感染所致。寄生虫感染多集中于 5~14 岁儿童中间。传染病的疾病负担约一半以上是由 15~44 岁年龄组人群所承担,缺血性心脏病 60% 以上发生在 60 岁以上的人群(表 2-11)。

## (三) 性别

性别可作为判断高危人群的一个依据。从整个世界看每 1000 个女性的疾病负担要比男性低 10%。女性因早死而损失的 DALY 比男性少,但因残疾而损失的 DALY 却差不多与男性相等。传染病在性别分布上的差异主要反映了感染机会的多少。例如女性所承受的与性传播疾病有关的负担要比男性大得多。所以应采取相应有效的措施来控制许多这种过度的女性负担。胆囊炎、胆结石及地方性甲状腺肿均以女性多见,可能与其解剖、生理和内分泌的差异有关。除乳腺癌、宫颈癌外我国恶性肿瘤的死亡率均表现出男性高于女性。



表 2-11 1990 年不同年龄组和性别的四种主要疾病的负担 (百万 DALY)

疾病和性别	年龄 (岁)					总计
	0~4	5~14	15~44	45~59	60 以上	
腹泻						
男性	42.1	4.6	2.8	0.4	0.2	50.2
女性	40.7	4.8	2.8	0.4	0.3	48.9
寄生虫传染病						
男性	0.2	10.6	1.6	0.5	0.1	13.1
女性	0.1	9.2	0.9	0.5	0.1	10.1
结核						
男性	1.2	3.1	13.4	6.2	2.6	26.5
女性	1.3	3.8	10.9	2.8	1.2	20.0
缺血性心脏病						
男性	0.1	0.1	3.6	8.1	13.1	25.0
女性	***	***	1.2	3.2	13.0	17.5

\*\*\* 不足 5 万

(世界银行数据)

#### (四) 职业

有些疾病的发生与职业有关,如粉尘作业致尘肺,有害气体致呼吸系统疾病等,因此,从事与职业致病因素接触者,视为高危人群,予以重点干预。又如从事其职业的人以不同方式威胁其他人群的健康,也应列为防护的高危人群之列。像患有肠道传染病或者成为病原携带者,从事饮食服务行业或水源工作者,则对其周围的人群危害甚大,亦应作为防治和监测的重点人群。

因此,有关重点研究课题的研究,凡与职业因素相关者,要充分注意有关直接或间接影响及易感的高危人群作为研究对象。

#### (五) 民族

不同民族因遗传、饮食习惯、地理条件、卫生条件、文化水平、宗教信仰、风俗习惯等因素不同,其患病与死亡水平也不同。凡涉及种族有关重点疾病的研究,纳入的研究对象务必注重民族特点。

#### (六) 行为及风俗习惯

已证实某些行为如吸烟、饮酒等是某些疾病的主要病因及致死因素,吸烟者特别是学龄儿童、青少年及妇女受害较深,是主要的高危人群。吸烟与疾病的死亡、失能、生命质量有关,除可引起某些肿瘤、心血管疾病、呼吸系统疾病外,对人类生殖也有很大危害,易发生自发性流产、胎儿及新生儿死亡、早产、低体重、围生期死亡或发病。所以吸烟的孕妇是一个特殊的高危人群。被动吸烟对人群健康同样也有很大危害。

男性同性恋、吸毒者、性乱者是艾滋病及某些性传播疾病感染机会较高的高危人群。某些风俗习惯如喜食生鱼、未熟肉类者与某些疾病的发生密切相关。

(王滨有 时景璞)

## 第三章 临床医学科研的选题与立题

临床医学研究的目的在于探索人类疾病发生、发展和转归的规律，提高对疾病的诊断和防治水平，消除或减轻疾病对人体的危害，改善预后，提高人类的健康水平。

由于危害人类健康的疾病众多，造成的疾病负担程度各异，对疾病的病因和发病机制的认识程度的差异，诊断与防治措施的效果不一等等，造成了临床医学的复杂性。此外，由于当代经济的全球化、科学技术以及信息技术的快速发展，人们对自身健康质量的关注以及对防病治病的要求越来越高。因此，如何选择与确立具有重要意义的课题进行研究，是具有挑战性的重大问题。

临床医学的研究，应遵循国家的卫生工作方针与政策，以预防为主、防治结合，改善人民的健康，促进我国临床医学事业的兴旺发达。从而也有利于社会安定和国家经济建设的发展。

### 第一节 临床医学研究的特点

临床医学的研究不同于基础医学的实验性研究，也有异于预防医学的群体健康性研究，其特点表现为以下几个方面。

#### 一、研究的对象是病人

临床医学研究的对象是病人，病人是作为有关疾病的“载体”。因此，患了疾病的病人，他们都是受着疾病的折磨，经受着不同程度的痛苦，承受着肉体、精神和经济等方面的压力，又鉴于各自的病理生理特点、心理情感状态、文化水平及所处的自然与社会环境的差异，疾病的临床病程和病情复杂，因此，即使同一种疾病，病理损害相似，而临床表现的个体差异往往十分显著，这就构成了临床科研对象的复杂性。所以，作为临床研究的对象，务必符合有关疾病的诊断标准，同时还要制订合适的纳入与排除标准，以使被研究的对象具有相对的匀质性而有利于研究的可比性。

#### 二、临床科研的干预措施要安全有效

任何投入临床试验治疗的药物或治疗措施包括诊断性措施，一定要有临床试验前的科学依据，证明对病人具有安全性和对疾病的诊治具有有效性，方可进行立题研究。例如，新药的临床治疗试验，一定要有如药物化学成分、药效学依据、病理和毒理、药物代谢动力学等可靠的基础医学实验研究及一期临床试验证据，证明安全有效者，且要符合有关法规方可立题进行正式的临床试验。因此，对于某种（些）缺乏科学依据且无疗效证据的药物或“措施”是不允许用病人做试验研究的。

为了对比新的试验措施或药物的疗效，需要设立对照组，而对照组的研究措施亦要保证病人的安全性。安慰剂的使用应根据治疗性试验的具体情况和要求使用（参见第九章）。

#### 三、研究的场所

临床研究基本上是在医院范围内进行，但是当涉及病因或危险因素的致病效应、疾病的早期诊断，以及疾病的早期治疗或一、二级预防时，则要面向社区，筛选具有临床前期或仅有早期轻微临床表现而尚能照常工作的患病群体进行研究。这种包括院内与院外患者的综合性群体研究，是临床流行病学的研究特点。这也是临床医学研究在今后的必经



之途。

#### 四、医德

按照世界医学协会关于人体临床试验的赫尔辛基宣言要求，凡是以人体为研究对象的临床研究，所使用的试验药品或措施，都必须具有充分的科学依据，要安全、有效，保证无损于病人的利益。对于接受试验的患者，要明确地解释接受治疗或试验措施的目的、意义、可能发生的副作用。坚持自愿的原则，要尊重病人的人格，不能欺骗研究对象。如果接受试验，要签署知情同意书，以防日后发生争议。即使在试验进程中，患者亦有退出的权利。因此，任何涉及以病人为对象的临床试验，在正式立题研究的时候，要高度注意医德问题，必须首先向有关伦理委员会申请，接受伦理委员会的审查，通过后方可进行研究（详见本书第十二章）。

### 第二节 选题与立题的原则

尽管国家各级机构确定的科研课题投标指南能够为研究者提供研究方向和课题的选择，然而，研究人员仍需要掌握临床科研选题立题的原则，方能充分发挥科研人员的积极性和创造性。

#### 一、选择疾病负担重大的疾病进行研究

从疾病负担的调查与分析中，选择高发病率、高致残率、潜在减寿年数（potential years of life lost, PYLL）或伤残调整寿命年（disability adjusted life year, DALY）损失大的、造成疾病负担重、对社会安定影响大、波及范围地域广的疾病为重点课题进行研究。如恶性肿瘤、脑血管疾病、呼吸系统以及心血管疾病、急性传染病、人畜共患的有关疾病、新生儿疾病等疾病负担是十分突出的，为国家确定的重点防治研究疾病（详见第二章）。

人体暴露于致病因素后，具有不同的易感性和抵抗力，因此，应将高危人群列为重点防治研究的对象。例如：吸烟者多为肺癌发病的高危人群；乙型肝炎病毒感染致慢性肝炎者为肝癌发病的高危人群；高血压、糖尿病、高脂血症等患者又为心脑血管病的高危人群，而环境、水源与食物污染严重的地区，往往是肿瘤高发区和高危人群所在等等。

随着社会的进步和医学的发展，疾病谱、病原生物体及其疾病负担在不断地发生变化，而且人们的认识总是落后于不断变化的客观现实，即要求不停顿的去研究与探索。因而研究重点的选择也必然会相应地变动。例如，我国在建国初期，鼠疫、霍乱和天花等急性、烈性传染病疾病负担非常严重，为防治研究重点。当被控制或消灭后，则以监控为主；但随着经济发展，人民生活水平的提高，生活节奏的加快，生态环境的变化等，肿瘤、心脑血管病等非传染性疾病上升为主要健康问题。因此，研究的选题就应随着国家的人民疾病负担状况的改变与之相适应。此外，我国幅员辽阔、人口众多，各地区的自然、社会和经济环境有着显著的差异，疾病谱及疾病的负担也存在地区性的差异。因此，防治病的选题方面也就有着地方特点，各地区都有相应的研究重点。例如，南方有关省区的血吸虫病，广东地区的鼻咽癌，河南林县及四川盐亭地区的食管癌等为其地区性的研究重点。

国家重大疾病防治研究课题的选择与决策，除了依靠专家的积极性外，根据一些先进的国际经验，还应吸收有关卫生医疗人员、行政管理人員和社区的人民代表参与，共同讨论，以听取各方面的意见和建议，从科学性和防病治病的现实需要性等多个维度进行综合



分析,方能作出符合国情的防病研究的科学决策。

## 二、研究问题要明确具体

选择研究课题,首要的是确定研究的焦点,即具体和明确地提出期望要解决的问题。例如,艾滋病的防治研究被选定为国家的重点课题,它所涉及的问题十分广泛,如病因及发病的危险因素、诊断、治疗、预后及预防的研究等。从临床研究的角度,研究的重点应当是探索有效的防治措施和改善患者的预后,这是目前迫切需要解决的实际问题。任何具体的研究要想一揽子地去研究和解决全部问题,往往是不可能的。因此,一定要根据自己的基础、专业特长、兴趣爱好、工作条件、技术力量、经济支持等,实事求是地选择某一个关键问题去重点研究,绝不能在一个研究中企图去解决多个问题。否则易造成困难的局面,达不到预期的目的。

如果属于综合性的重大课题,由于涉及的内容和研究方法的不同,可根据具体内容,以主题为中心,设若干分题(子题)。分题是主题的一部分,与主题密切相连。诸分题最终结果的有机综合,则构成总课题的研究整体。

## 三、要有创新性

研究课题的选择,要有自主创新性,有新的见解和新的特色,而不能跟着别人去盲目地重复或赶“时髦”。为此,课题涉及的内容,应力求居于该领域的研究前沿和具有领先水平,或者在过去研究的基础上有所突破和改进,如采用现代科学技术,在分子生物学水平上研究候选基因、基因多态性与疾病发生的联系。应用蛋白质组学技术探讨靶向药物等有可能取得比以往在组织、细胞水平更为精确的结果。即使对于重大疾病防治措施的研究,也应剖析既往或现存有效干预的最佳水平,有所新的突破,否则,就会造成人力、物力和财力的损失或浪费,而研究本身往往也无什么意义。

## 四、医学研究的公正性

医学研究应面向广大人民的健康问题。我国是世界上最大的发展中国家,在国内不同的地区经济发展的水平也存在着巨大差异,社会上的贫富差别也颇显著,城乡医疗卫生和保健事业的水平企求相应的亦很不同。因此,临床医学研究课题的选择,应该结合我国的国情、地方特点以及广大人民(不分贫富)对健康和有效防治疾病的要求予以优先考虑,以充分体现医学研究的社会公正性。

此外,国家或地方的卫生研究经费投入也是有限的,人力的资源与水平也会受着某些限制。因此,在选择卫生研究课题的时候,也应该考虑将有限的资源投入到最大多数需要解决的健康和有效防治疾病的研究项目方面,方能体现卫生研究资源利用的社会公正性。特别是人民当家做主的社会主义国家,更应树立为大多数人民健康研究的社会公正性的典范。

## 五、选择足够的研究对象

对疾病病因及危险因素的研究,要有足够的接触及不接触暴露因素的观察对象;诊断及防治试验要计算所需求的最低样本数量。因此,在选题的时候,一定要根据实际情况充分考虑研究对象的来源和数量。如果在一个地区或单位,在规定期间内有足够对象的来源者则可行。否则要考虑组织具有相同或相似水平的地区、单位进行多中心协作,以保证样本量的需求。这是因为临床试验所采用的新的试验措施或药物,都有发展的相对时间限制,期望尽可能快地验证其应用的临床效果和价值。如果因为研究对象来源的不足而导致





试验时间延长,于是有可能被更新研究的药物或措施所代替,从而造成原试验研究的逊色甚至失去意义。如果研究对象充足,视具体情况可从特定人群中或医院的患者群体中,按设计要求去随机抽样。

## 六、抉择合理的研究设计方案

根据立题研究的课题性质,抉择科学性及其可行性良好的设计方案,这是保障所研究课题获得成功的关键之一,详见本教材第九章。

## 七、干预措施要安全有效

选择任何干预措施和诊断方法进行研究,都要确保安全、有效。即使对尚无有效治疗的疾病进行治疗性试验,其试验的药物或措施,亦要有科学依据,特别是药效学的实验依据(参考本章第一节)。要有先进性、创新性,否则就没有研究的必要。

## 八、研究措施执行的可行性

选题的时候,对拟采用的研究措施,一定要考虑执行的可行性。

1. 仪器设备及其实验室条件 试验研究的措施及试验结果,都需要相应的仪器设备进行测试。例如:涉及抗生素或化学治疗药物的耐药性基因或肿瘤病因学的研究,有关抑癌或致癌基因的研究,需要具备分子生物学实验的仪器设备和实验条件,否则,即使研究的问题选得好,客观上却不可行。

2. 要有执行研究课题的配套人才 要有一支能胜任负责科研设计、实施和观察,资料收集,统计分析,以及科研的总体指导和管理等技术力量较强的专业队伍,这是保证选题立题后能够执行的重要条件。

3. 干预措施和测试的指标一定不能过繁 采取的试验干预措施或药物,在剂型、用法、用药的途径等方面,应力求简单易行。测试效应的指标不应过多,在保证准确可靠的前提下,选用敏感性和特异性好的指标,做到少而精,既可以保证执行的可行性,也可以减少假阳性或假阴性偏倚的影响,避免误导研究的结果和结论。

4. 注意社会、文化和宗教等的接受可行性 临床试验主要是以病人作为研究对象。因此,要注意社会习俗、文化素养以及宗教信仰等因素及其临床试验措施的可接受性。否则,就可能产生负面影响。

5. 注意研究对象的依从性,即使一种很好的防病治病措施,如果价格过贵或应用方法过繁,或缺乏对患者足够解释及关爱,则研究对象的依从性往往不能保证,对此,在研究课题立题的时候,要充分注意。

## 九、经费支持

在选题和立题时,务必考虑经费的来源和保证,要作实事求是的预算,以保证仪器设备、实验试剂、药品开支、劳务报酬、学术交流等必需的费用。不切实际地过高或过低的经费预算是不利于所立课题的正常运作的。

## 十、伦理学原则

任何临床研究课题的立题,务必要符合伦理学原则,遵守国际公认的“赫尔辛基宣言”的精神。因此,按照国际惯例,临床试验的研究课题,必须经过有关科研伦理委员会审查通过后,方有资格立题研究,详见本教材第十二章。



## 十一、预测研究成果的价值

研究者要根据科学的依据,对所选择和立项研究的课题可能获得科学成果及其防病治病的效果和价值,要进行实事求是的预测,估计它的成本—效果,可否被推广应用,以及可能产生的社会效益及经济效益等。这些因素及指标,对于选题申报具有很重要的参考价值。

## 第三节 临床科研课题的类别

临床医学研究,一定要为国家的防病、治病及保障人群健康的根本任务服务,这也是临床科研选题立项的基础。为了保障人力和财力的投入能够取得预期的效果,在科研的选题与立项方面,国家的卫生和科技主管部门,针对有关疾病对国家和人群健康造成的疾病负担情况,在广泛征求医务人员和研究人员建议的基础上,组织有关专家专门讨论,并根据国家卫生工作的方针政策,选择重要的有关疾病立项研究,制订详细的研究课题招标指南,向全国或地区公开招标。立项的种类如下:

### 一、国家级课题

涉及对国家重大疾病的防治范畴,如每五个年度的重大疾病防治研究的攻关项目,国家自然科学基金招标项目等。研究者按照国家的招标立项的项目要求,有针对性地提出研究课题的申请书,经专家评审通过,并报经政府主管部门批准后,则正式立项进行研究。

### 二、部、省级研究课题

凡未能纳入国家级,而确需进行研究以提高疾病防治水平,改善人群健康的研究项目,特别涉及有关地区性需要的重大研究项目,国家卫生部以及省级卫生、科技主管部门,亦分别制订课题的招标指南,指导研究人员按要求提出科研课题的申请,评审与批准程序与国家级相同。部省级重点支持的范围,主要是对防病治病具有实效的应用性研究课题。

### 三、国际合作研究课题

随着经济发展的全球化,日益扩展了我国在国际舞台上的地位,卫生和疾病亦具有全球化的性质。因此,自改革开放以来,我国参与国际性医学卫生的学术合作研究有很大进展,有许多重大疾病与国际医学界世界卫生组织(WHO)和有关国际基金会共同进行合作研究,如像艾滋病、心血管病、肿瘤以及新近出现的严重急性呼吸综合征(SARS)、禽流感、结核病等研究课题。

### 四、新药及新技术的临床试验

药厂生产的新药经卫生行政部门审批合格后,进行临床治疗试验,或新仪器(器械)的临床应用试验,这些并不是临床研究者自己主动进行选题与立项的研究项目,而是对新药或新的医疗仪器及技术的临床效果的验证。

### 五、其他项目

凡在国家和各级政府部门招标项目之外,临床医学工作者根据自己所处的环境,为提高疾病诊治等各方面的需要,按照选题立项的原则和方法,自拟研究课题所谓自选课题付



诸实践者是相当普遍的。许多都可获得本单位的大力支持,当获得成功后可以产生很好的效益。此后也可与国家各级课题对号再行申请。

临床医生在自己的临床医疗实践中,发现临床的新问题、新经验或教训,往往收集系列相关病例,进行回顾性的经验总结分析和探讨,可能的因果关系,也是十分重要的研究类型,往往为前瞻性的研究选题提供重要信息和依据。

## 第四节 临床选题和立题的程序

### 一、研究问题的提出与选题

临床研究问题的提出,离不开防病治病的具体实践。当在临床医学实践中发现了某种不易解释的临床现象或某种疾病严重地威胁人类的健康时,必然促使人们要去探索其病因及发病的危险因素,弄清楚诊断,为有效地防治提供科学的依据;当某些疾病对患者造成巨大的负担而缺乏有效治疗对策的时候,就会迫使临床学家和基础医学学科及药学专家们去研究发掘有效的防治药物与措施,以挽救患者;对某些难治的、预后不良的急性或慢性疾病,往往会激励临床学家去研究一些改善预后的措施以减轻病残,降低病死率和改善患者的生存质量。人们的认识会在这个过程中逐步地深化和完善,从而找准研究的课题和解决的重点。

新问题的出现,成为人们去研究和解决的动机。从多种认识中去比较分析,去粗存精,弃伪存真,以选出最重要、最能影响全局的问题去研究,这就是从提出的问题进行选题研究的主要过程。

为了有助于对选题的理解,这里以艾滋病的研究为例。众所周知,艾滋病是20世纪末威胁全人类健康的重大问题之一,据2003年底WHO报告艾滋病(HIV)感染者,全球达4000余万人群,发病者达500万人,死亡者达300余万。当今世界投入研究的资源是十分惊人的。从临床病例的发现,对患者的病例-对照研究和流行病学调查,弄清了该病的传播途径和发病的危险因素。在此基础上,病毒学和免疫病学家研究,发现了致病的人类免疫缺陷病毒(HIV),导致患者的免疫功能严重损害,成为继发感染和致瘤的原因,相继为艾滋病的特异性诊断提供了有力的武器;发病机制的阐明对临床病征的解释和治疗提供了科学的依据。尽管如此,临床却缺乏有效的治疗手段。生物医学家们正全力地研究治疗的药物,以期挽救患者生命并改善预后。在预防研究方面,疫苗的研制与应用、针对发病危险因素的干预难度很大,涉及社会经济、精神文明和道德行为的方方面面。诸多未解决的问题,需要医学家及社会学家从中选择研究,一一突破和解决。这就是对艾滋病的研究及其需要解决的重点问题。

又如2002年底至2003年初从我国发生的严重急性呼吸综合征(SARS),导致国内及国际上部分地区的传播,经WHO组织国际合作研究确定病原为变异的冠状病毒。尽管全球的病例仅8000多例,但其病死率高(5%~10%),影响甚大。当前重点是研究该病的特异性的早期诊断方法以及有效的防治措施,改善患者的预后,这些是亟待解决的重大课题。因此,国际卫生研究发展委员会(COHRED)提出卫生研究的选题原则,强调两点:一为国家重大的疾病;二为这种疾病对全球人民的健康具有重要的影响。对具有这两方面特点的疾病应重点选题进行研究。因此,这些原则与事例可作为研究者选题之参考。

### 二、充分掌握现有的科技信息

人们的认识水平和临床经验是十分有限的。当出现陌生或不易解释的临床现象,或者



发生了某种未知疾病，临床医学家必须回答这些挑战性的问题。这些问题，可能是医学领域里的未知现象，或者是不太清楚的事件，也可能是自己不知而他人已知的事件。因此，除了应用自己的智慧之外，要带着问题去认真地检索医学文献，充分掌握科技信息资料，认真进行分析，以了解被研究问题的历史、现状以及尚待探讨的关键问题。抓住了矛盾与问题，则应进一步地追寻采用什么样的研究方法去研究解决这类具有挑战性的问题。

在查阅文献的时候，一定要采用临床流行病学/循证医学的原则和方法，进行严格的科学分析和评价，吸收有充分科学依据、真实性和水平高的文献的精华；抛弃有偏倚因素影响的、缺乏科学依据的文献。这样，对研究课题的选择才不致发生误导。这里要十分强调：好的文献并不一定都是阳性的结果，阴性的研究结果并不意味着是质量不好的文献，标准取决于文献的真实性。

为研究选题而查阅文献，要充分应用现代科学的检索手段和方法，针对研究的课题，尽可能收集敏感度及特异度高的文献加以借鉴，切忌视野狭窄、主观片面、草率从事、结论武断（详见本书第四章）。

### 三、确定研究的关键问题

由于人的精力和时间是有限的，应根据学术力量、知识的水平、研究的基础和研究的具体条件从诸多需要解决的问题中，参照有关选题的原则，选择研究的关键和切实可行的重大研究问题立项研究。对此，要实事求是，要估计研究成功的把握度，以充满信心地完成研究任务，达到预期的目的。这里切记的是不要贪多求全，企图通过一个研究课题来解决许多问题，实践证明这样的想法和做法，获得成功的概率是十分罕见的。

### 四、撰写研究计划书

立项研究的课题，应按照科研计划书的格式和内容，撰写科研计划书（详见本书第二十一章）。

最后，将临床科研的选题和立项的主要步骤，归纳总结见图 3-1。

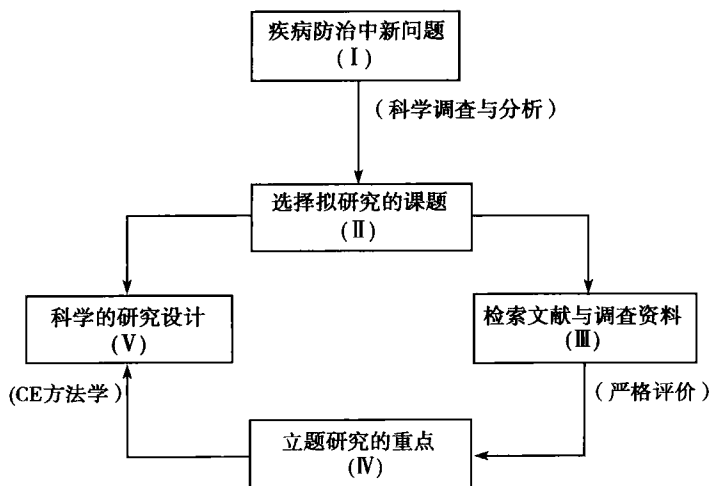


图 3-1 临床科研选题与立项步骤



## 第五节 临床科研立项研究的评价标准

对于临床研究立项的质量评价，有以下参考标准供研究者考虑：

### 一、是否为国家或地区性危害人民健康的重大疾病

这里要注重国家或地区性疾病负担的现状，要重视对危害人民健康重大的疾病研究，要考虑临床研究能为大多数人服务。当然，还要结合研究者和研究机构的实际水平与研究的能力，实事求是地为国家或地区有效防治重大疾病、研究创新方面作贡献。此外还要考虑社会的公正性和需求性。

### 二、研究的重点是否明确

一个科研课题必须有明确的研究重点，即重点要解决的关键问题是什么？例如：对 SARS 的特异性早期诊断研究，重点究竟是什么？是分离病毒确诊或是应用特异性抗体检查早期患者的病毒抗原？还是检验 SARS 患者血中的抗体？方法是什么？是否准确可信？预期价值等等问题必须明确回答，方能论证研究的重点。

一个研究课题，如涉及的问题面宽，想通过一个研究解决多方面的问题，必然缺乏重点，其结果也不会达到预期的目的。

### 三、是否掌握了本研究涉及的最新科技信息

立项研究与确立研究的重点，务必要通过检索文献资料，查阅与掌握相关的最新科技信息，这是避免无效重复及使得自己掌握最新知识，做到创新的最为重要的环节。应力争完善并对新知识做到严格评价，取其精华、弃其糟粕，掌握可靠的最新最佳证据方能在探索中继往开来，使立项研究更具科学性。

### 四、是否具有创新性

一个好的研究课题，务必要有自主性的创新性，重复他人的“研究”是缺乏生命力的，如像有的研究人员从国外引进一点技术或试剂，在国内照人家作的方法，找一些病人采集一些标本，作一下人家做过的试验，获得某种结果，包括分子生物学的“研究结果”进行报道，尽管在国内算首次，但算不上“自主创新”，因为毕竟是重复人家的，创新应该是自己真正的创造。此外，亦应避免盲目地与国际“接轨”和脱离国情的“追新”！所以在创新上应认真鉴别，尽管创新难，但在研究工作方面又确属重要。

### 五、是否可行

对课题承担与执行者的学术技术力量、工作基础、设备条件、研究设计的方案和路线、干预措施、研究对象的来源和数量上的保障以及经费的预算，是否可行，要仔细评价。

### 六、预期的成本-效果

这里要考核研究课题预期可能产生的效果。例如提高诊断的水平、提高疗效水平、降低病死率、病残率、伤残调整寿命年等的指标以及为此而投入的成本，计算其成本-效果和作相应的卫生经济分析，这样才可能看出一个研究课题所产生的社会和经济的效益。这对于研究成果今后转化为防病治病实践的推广应用是有预测的参考价值。



## 七、医德

临床科研务必要保障病人的安全及权益，尊重病人的人权和人格为最高医德标准，而课题的立题应通过伦理学组织的审查批准方可实施。对此，任何临床试验都应遵照世界医学学会所制订的赫尔辛基宣言的精神。

(王家良 李绍忱)

## 第四章 现代医学研究的信息资源与管理

在生物医学及信息技术高度发达的今天，卫生研究的全球化，已成为必然。随着全球范围内卫生研究经费的投入不断加大，在不同水平、不同层次进行的研究数量剧增，随之而来的是医学文献信息呈几何倍数的增长，每年新增的文献数量达数百万篇，且有加速增长之势。此外还有数量庞大的未公开发表文献、灰色文献等，由此形成了海量的医学信息资源。

如何从中追踪研究前沿，掌握新近进展，保持“与世界顶级文献同步”，临床医生和研究者面临着巨大挑战。因此，学会查询、挖掘、评价现有医学信息资源，从中提炼出有价值的、真实可靠的证据和信息，也是知识更新和自我发展的必由之路。

### 第一节 制定检索策略与文献信息查询

选题立题之初，应围绕研究目的，检索所有相关文献信息，全面剖析研究背景、现状以及存在问题。但应该查询哪些医学信息源？可及的信息资源或文献数据库又有哪些？能否做到查新查全？要回答这些问题，需要事先做好一系列筹划设计，包括构建可查询的检索问题，确定检索范围和文献信息资源库，制定检索策略，评估检索结果等。

#### 一、制定检索策略

文献信息查询之前，要先确立可用于检索的研究问题，并将研究问题具体化。一个能用来检索的临床研究问题，常由 PECOT 内容组成：即人群或病人（population/patients）、暴露或干预因素（exposure/intervention）、比较（comparison）、结果（outcomes）和时间（time）等要素。然后将上述研究问题解析，分解成多个独立词汇，作为初始检索用词。再参考检索数据库词典，从中选择与初始检索用词最接近的标准检索词。目前最常用的标准检索词为美国国立医学图书馆编制的医学主题词表（MeSH）中的主题词。鉴于文献检索本身费时费力，甚至可能得到相互矛盾的检索结果，事先制定好检索策略很有必要。

根据检索目的和范围，合理选用检索用词（包括主题词和副主题词），并与逻辑运算符（如逻辑和“AND”，逻辑或“OR”，逻辑非“NOT”）、位置运算符（如 WITH, NEAR, IN）、范围运算符（“=”，“>”，“<”等）、截词符（“\*”）以及通配符（“?”）等搭配组合后，形成检索策略，有助于在短时间内实现查新查全、最大程度检索出与研究问题高度相关的文献或信息等。

制定检索策略时应同时兼顾检索的敏感性（sensitivity）和特异性（specificity）。若提高查全率，要制定高敏感性的检索策略，扩大检索范围，如选用“逻辑或”、截词符等；若要提高查准率，需制定高特异性的检索策略，缩小检索范围，提高相关文献的检出率，如选用“逻辑和”、位置运算符等。

#### 二、文献信息检索

选择信息资源库与检索工具，全面检索文献信息。目前应用较多的文献数据库和检索工具有：MEDLINE（美国“医学索引在线，或 PubMed”）、EMBASE、中国生物医学文



献数据库 (Chinese Biomedical Literature Database, CBM) 等。这些生物医学文献数据库, 收录文献范围广、包罗万象、条目数量庞大, 但彼此间收录互有交叉、质量不一。若只单独选择某个文献数据库, 难免挂一漏万。应在专业人员的指导下, 选择多个资源数据库, 同步进行检索。具体检索流程见图 4-1。

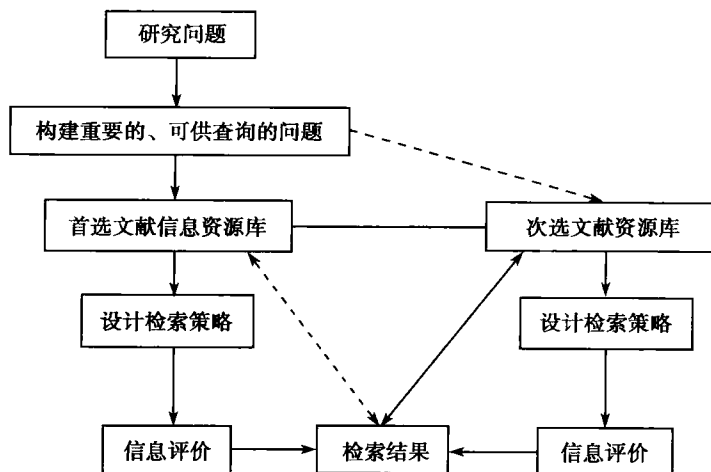


图 4-1 文献信息检索流程

### 三、检索过程与结果评价

全面评价检索质量, 避免漏检或误检。应对检索策略、检索过程以及文献信息检索结果的完整性、真实性和重要性等进行综合评价。

### 四、文献信息检索流程实例演示

1. 提出一个可用于检索的研究问题。如: “降压治疗能否降低 II 型糖尿病伴有高血压病人的病残率和病死率? ”。在该研究问题中, 病人 (patients) 为 “II 型糖尿病伴有高血压病人”, 干预 (intervention) 为 “降压治疗”, 比较 (comparison) 为 “与未降压组相比”, 结果 (outcomes) 为 “病残率和病死率”。

2. 制定检索策略, 实施检索。

(1) 选择一个文献检索数据库: 如 PubMed。

(2) 确定检索主题词, 设计检索策略。

#1 “diabetes type 2/hypertension” [MeSH] Limits: Clinical Trial, Human

#2 hypertensive Type 2 diabetic [MeSH] Limits: Clinical Trial, Human

#3 “management” OR “therapeutic”

#4 “mortality” OR “mobility”

#5 (#1 OR #2) AND #3 AND #4

(3) 检索结果: 共检索出 52 篇相关文献 (截至 2007 年 7 月)。

1) Erdmann E, Dormandy JA, Charbonnel B, Massi-Benedetti M, Moules IK, Skene AM; PROactive Investigators. The effect of pioglitazone on recurrent myocardial infarction in 2445 patients with type 2 diabetes and previous myocardial infarction: results from the PROactive (PROactive 05) Study. J Am Coll Cardiol. 2007 May 1; 49 (17): 1772-1780. Epub 2007 Apr 16. PMID: 17466227 [PubMed-indexed for MEDLINE];





2) Weber MA, Bakris GL, Dahlof B, Pitt B, Velazquez E, Gupte J, Lefkowitz M, Hester A, Shi V, Weir M, Kjeldsen S, Massie B, Nesbitt S, Ofili E, Jamerson K. Baseline characteristics in the Avoiding Cardiovascular events through Combination therapy in Patients Living with Systolic Hypertension (ACCOMPLISH) trial: a hypertensive population at high cardiovascular risk. Blood Press. 2007; 16 (1): 13-19. PMID: 17453747 [PubMed-indexed for MEDLINE];

其余 50 篇文献题目略。

3. 检索结果与文献信息评价 需对获取文献的质量、时效性等展开严格评价。注意所获得文献信息一般有 3~6 个月的滞后期。

## 第二节 常见的医学信息资源

发展中国家的临床研究者，在查询文献信息时，经常会遇到截然相反的两重困境。一方面是有用或高质量信息不足或缺乏。由于经费缺乏或其他条件限制，大量文献原文无法获取，形成“信息孤岛”；另一方面则是信息泛滥问题，大量杂乱的灰色信息无处不在。研究者又面临着如何从海量的信息资源中提取有价值信息的困境。

为此，了解现有医学信息资源的类别与分布，掌握一些基本的查询方法，学会利用互联网或其他电子信息资源，将有助于及时追踪研究领域的最新进展，全面了解专业学科的最新发展动态。

### 一、常见的医学信息源种类与分布

医学信息来源很广，包括电子文献检索数据库（如 MEDLINE），课题研究进展报告，学术会议资料，引文索引，论著和教科书以及个人经验交流等。在图 4-2 中，主要医学信息源若按照可及数量和关联度大小排列，就形成了金字塔状的分布结构。其中与研究问题高度相关的信息源主要是电子文献检索数据库，这些数据库以固定的标准词汇建立索引，利用索引可以进行精确的检索。在图 4-2 金字塔中，越向下的部分，与研究问题相关的文献分布密度越低，塔底为 WWW 来源信息，密度最低，大部分为噪声信息或灰色文献。但若只查询高密度医学信息源，可能会漏掉部分有用文献信息，但要查全这部分信息，则又可能花费大量的时间和人力，得不偿失。因此需要结合检索目的，在两者之间进行权衡。

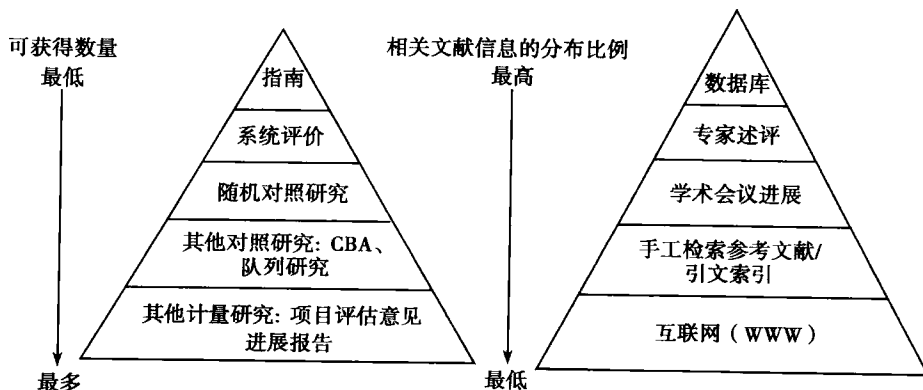


图 4-2 医学文献信息金字塔



同样医学信息级别分布,也呈金字塔状,高级别文献信息数量少,分布在塔尖,低级别文献信息数量庞大,分布在塔底。一般来讲高级别文献信息是由低级别信息综合提炼加工而成。例如,临床实践指南是最高级别的证据,综合萃取了系统评价及 RCT 等证据信息,因而临床实践指南的发表数量要少于 RCT 原始研究文献。因此,评价查询结果时,应判断获取的文献信息类型、级别分布是否呈金字塔状。

## 二、电子文献检索数据库

电子文献索引与全文数据库收录了绝大部分原始研究文献,收录的文献类型包括指南、系统评价、RCT 试验研究及其他计量研究等。全文数据库则收录了原始文献的全部内容,如 Cochrane 系统评价数据库。这些电子文献数据库允许综合应用逻辑、位置运算符等制定检索策略,提高了文献的查准率和查全率。开始接触电子文献数据库时,最好寻求图书管理员的帮助,利用熟悉的检索方法,实现快速有效的检索。表 4-1 中罗列了部分电子文献数据库。

表 4-1 电子文献数据库一览

电子文献数据库类别	收录内容与范围
MEDLINE (美国国立医学图书馆) <a href="http://www.pubmed.gov">http://www.pubmed.gov</a>	收录了自 1966 年以来医疗卫生方面的文献,如医疗卫生、口腔、护理、生物、生理学及与健康相关的人文与信息科学
EMBASE 数据库 (Elsevier 公司) <a href="http://www.embase.com">www.embase.com</a>	收录了 1974 年以来医疗卫生、药学方面的文献以及摘录了护理、口腔、兽医与心理卫生方面的文献
医学图书馆网 <a href="http://www.bireme.br/bvs/i/ibd.htm">http://www.bireme.br/bvs/i/ibd.htm</a>	收录多个生物医学文献全文数据库。如 LILACS、PAHO、REPIDISCA、Desartres、ADOLEC、AdSaude
Cochrane 图书馆 <a href="http://www.thecochranelibrary.com">www.thecochranelibrary.com</a>	收录了循证医学相关的全文文献,并随时更新。包括 Cochrane 系统评价数据库, Cochrane 注册临床试验库, NHS 经济评价数据库,卫生技术评估数据库等
CINAHL-医学、卫生与护理文献类索引 <a href="http://www.cinahl.com/">http://www.cinahl.com/</a>	收录了主要英文文献中涉及护理、职业卫生、生理治疗及其他相关领域的文献
PsycINFO (由美国心理学会建立) <a href="http://www.apa.org/psycinfo/products/psycinfo.html">http://www.apa.org/psycinfo/products/psycinfo.html</a>	收录了心理学研究及相关学科的文献,如:管理、教育、社会工作、精神病学、药物和护理

## 三、灰色文献资源库

尽管电子文献数据库是临床研究的主要医学信息源,但仍存在一些不足。如大多只能得到摘要,无法获取文献全文。

其次是存在发表性偏倚。有大约 25% 临床研究在 10 年内并无文献发表,因无法被电子文献数据库收录而出现缺失。同时若一项研究结果有统计学意义,论文发表年限平均为 5 年;倘若得到阴性结果,则很少被发表,如公开发表文献中有 85% 报告结果有统计学意义,一些主要医学杂志(如 BMJ、JAMA 和柳叶刀杂志)只有 10% 阴性结果的论文被发表。但即便被发表,发表年限一般会延迟至 8 年。因而电子文献数据库中收录的阴性结果文献可能被严重低估。

另外文献的收录也存在地域性偏倚。如在一个收录了 3000~4000 种杂志的文献数据库中,只有 2% 的文献来自中低收入的发展中国家,而发表在这些国家国内杂志上的文献超过 90% 并没有被主要电子检索数据库收录。



因此,为弥补电子文献数据库的上述缺陷,可查阅一些灰色文献数据库作为重要补充。灰色文献对临床研究也很重要。如有研究表明,系统评价和 Meta 分析中若不考虑灰色文献,将可能导致效应量被高估 10% 以上。

灰色文献是指由各级政府部门、学术机构、商业与其他机构生产的,不以赢利为目的的印刷品和电子产品。形式多种多样,可以是各种报告(进展报告、技术报告、统计报告、市场研究报告),也可以是论文,讨论进展,技术说明与标准,非营利性的译文,著作,技术与商业文件,非营利性的官方文件等等。但这些文献信息未经过严格评价,缺乏可信度,所以称为灰色文献。

较为有用的灰色文献数据库有国家图书数据库与目录库、相关学术组织或机构网页及链接数据库等。除此之外,网上灰色文献库中还包括数量庞大的、有潜在价值的网页,这些网页是由一些小的或不知名的机构、非政府组织(NGO)、学术研究中心、私人公司及咨询公司等创建。

### (一) 国家图书数据库与目录

建立与发展国家图书数据库与目录,对国内外研究者均很重要。例如,我国出版了科学论文文摘与索引,收录的文献,大部分以中文发表。表 4-2 列出了一些国家级图书数据库。

表 4-2 国家文献数据库一览

发达国家图书数据库	
AMICUS-加拿大国家图书馆	<a href="http://amicus.nlc-bnc.ca/aaweb/amilogine.htm">http://amicus.nlc-bnc.ca/aaweb/amilogine.htm</a>
澳大利亚国家图书馆目录	<a href="http://www.nla.gov.au/catalogue">http://www.nla.gov.au/catalogue</a>
大英图书馆	<a href="http://www.bl.uk/">http://www.bl.uk/</a>
美国国立医学图书馆 LOCATOR+	<a href="http://locatorplus.gov">http://locatorplus.gov</a>
国家技术信息服务目录	<a href="http://www.ntis.gov">http://www.ntis.gov</a>
加拿大科学与技术信息所	<a href="http://cat.cisti.nrc.ca/">http://cat.cisti.nrc.ca/</a>
中低收入的发展中国家图书文献数据库	
克罗地亚中央医学图书馆	<a href="http://smk.mef.hr/cml.htm">http://smk.mef.hr/cml.htm</a>
古巴医学杂志全文文摘 (INFORMED)	<a href="http://www.informed.cu">http://www.informed.cu</a>
中国学术期刊出版物中心服务网	<a href="http://www.library.pitt.edu/gateway/">http://www.library.pitt.edu/gateway/</a>
非洲杂志在线	<a href="http://www.inasp.ino/ajol/index.htm">http://www.inasp.ino/ajol/index.htm</a>
俄美医学图书馆联盟 (MEDLIBET)	<a href="http://www.pinpen.com/preview/medlibnet5.htm">http://www.pinpen.com/preview/medlibnet5.htm</a>

国家图书数据库与目录可能无网络版和光盘版,检索需要付费,检索系统也需要进一步完善,如增加一些检索辅助功能等。

### (二) 相关组织机构网页

一些组织机构,如国际发展研究中心(IDRC)、世界银行、卫生研究发展委员会(COHRD)或卫生研究世界论坛(GFHR)等正逐步将其内部文件、报告发布到网上。对此类组织机构网页查询,可分类进行检索。

1. 发展组织机构类,如世界银行、USAID、CIDA、DFID、WHO、UN、AusAID 和欧盟等。这些机构因是大量卫生研究的资助方,定期会将研究进展、技术报告等发布在网页。

2. 专业组织机构类,如国际 HIV/AIDS 联盟,疟疾国际基金会,国际抗结核和肺病联合会等。这些组织和机构随时发布与疾病有关的最新信息以及研究进展报告等。

3. 国家级研究机构类,如非洲医学研究基金会,印度医学研究会等。这些研究机构



也会定期发布资助项目的研究报告。

### (三) WWW 网页链接查询

互联网目前已有超过 500 亿个人网页, 可用 Google 和 Yahoo 等搜索引擎检索。最近在对 7 个 Cochrane 系统评价进行再评价的一项研究中, 通过检索 WWW 网页查询未公开发表文献, 结果共耗时 21 小时评价了 429 个相关网页, 从中新发现了 9 个临床试验。

由于缺乏索引, 同时搜索范围广, 涉及整个 Internet 网, 查询很费时, 即使有检索策略, WWW 检索结果的精度仍很差, 会得到大量无关文献。

### (四) 其他专门灰色文献数据库

一些特殊文献数据库专门收集不同类型的灰色文献。例如, 学位论文摘要 (dissertations abstracts) 收录了自 1861 年以来超过 1000 所著名大学的硕士与博士论文。欧洲灰色文献信息系统 (SIGLE) 则收录了在欧洲范围内的未公开发表的论文、学位论文、进展报告以及其他非正式文献等。

## 第三节 对检索文献的严格评价

对检索到的医学文献信息, 还须对其质量进行严格评价。不同研究类型的文献, 均有相应的严格评价工具或标准。如 JAMA 用户指导手册、CASP 严格评价技巧项目网等提供了一系列严格评价标准, 用来评估系统评价、随机对照试验、病例对照研究、队列研究、描述性研究、诊断试验和经济学评价研究等不同类型文献。

### 一、常见文献的严格评价工具

由于随机对照试验采用了随机、盲法以及设置对照组, 最大限度地控制了混杂和偏倚对结果的影响, 确保了结果的真实性, 被认为是一种论证强度较高的设计方案, 因而在临床研究中备受推崇, 相关文献发表很多, 成为临床证据的重要来源。对随机对照试验的质量评价方法发展很快。Moher 早在 1995 年就列举了 34 种随机对照试验质量评价工具, Juni 等通过手检又额外发现了另外 14 种工具。但现有的评价工具还不是很严格, 工具不同, 质量评价结果也不尽相同。其中 Jadad 工具已得到初步认可, 该工具设置的评价内容包括: 是否随机、盲法质量如何、是否有对失访或中间退出的描述、方案分配是否隐藏等 (详见第五章、第十五章)。

对系统评价文献的质量评价方法就是与 QUOROM 中设置的项目逐一进行比对。QUOROM 是用来规范系统评价的标准报告格式, 包括 18 个条目, 其中 9 个循证条目覆盖了系统评价报告中的摘要、简介、方法与结果等方面内容 (详见第六章)。

JAMA 用户指导手册针对定量研究文献设置了 5 项指导性评价原则: ①研究对象是否与研究问题有关? ②研究对象的选择是否有充足的理由? ③数据资料的收集方法是否与研究目的与场所匹配? ④数据资料的收集是否完整充分, 足以描述观察事件? ⑤资料的分析是否合适以及发现的结果是否被充分证实? 这 5 项原则对一般定量研究文献的评价均有一定的参考价值, 已被广泛的应用到临床研究文献初步评价中。

### 二、WWW 来源文献的评价工具

与上述常规文献相比, WWW 来源文献与报告大多为灰色文献, 因此, 对其严格评价显得尤为重要。尽管目前有一些专用工具评价 WWW 文献信息质量, 但这些文献信息大部分还不能算做真正的研究文献。另外, 尽管工具很多, 如仅在 1997~2002 年 6 年间就发现了 98 种评价工具, 但没有一个工具提供了信度与效度的检验结果。这些评价工具的



实际测试结果与预期设想的测量结果是否一致也并不清楚。因此,在这里并没有具体列出这些评价工具,最好还是使用上述被广泛认可的严格评价工具与标准。

## 第四节 电子文献及非常规文献信息的引用格式

近年来,越来越多的研究报告及相关资料以电子文献格式发表,在撰写论文与计划书,引用这些电子出版物时,应正确使用引用格式。电子出版物的引用格式不同常规文献,主要是基于以下特点:

1. 资料的多样性与媒介载体的多样性 不同类型的电子文献均有各自的引用及表达格式,资料和媒介的多样性导致出现多种排列格式。电子文献信息主要来源有:①基于印刷出版物的在线文档;②电子杂志文献;③电子文献数据库;④WWW 网页;⑤CD-ROMS;⑥公告栏信息;⑦私人电子邮件等。引用格式不尽相同。

2. 电子文献信息的动态性 非固定、动态性是电子文献信息的特色,也许昨天发现的文献,第二天就移到了其他地方或者内容发生改变,甚至完全消失。

3. 引文信息有可能不完整 例如一些网页信息残缺,没有作者姓名或者发表年限,甚至没有题目等。

4. 引用格式多样化 现有温哥华格式、美国心理学会格式(APA)、哈佛格式等三种通用引用格式。其中温哥华格式已被国际医学杂志编辑委员会所接受。另两种格式也有所应用。

### 一、电子出版物文献的基本引用格式与要素

电子信息文献的主要引文要素包括:①作者或编者;②发表年限;③题目(论文或书籍);④(杂志名);⑤(卷号);⑥(期号);⑦媒介类型;⑧(页码);⑨获得途径;⑩进入时间等。

对于某个特定的参考文献,引文要素并不要求完全具备,如上述带括号要素,属于可选项。但下面三种要素则是电子参考文献所必备的。

#### (一) 媒介类型

一些文献信息可通过多种媒介途径获取,如 CD-ROM(光盘)、印刷品或在线方式等。因此电子参考文献必须标明获取途径。

范例 1: COHRED Working Group on Research to Action and Policy. 2000, *Lessons in Research to Action and Policy. Case Studies from Seven Countries*, [Online] COHRED, Geneva, Switzerland, COHRED Document 2000.10, Available: <http://www.cohred.ch/documents-COHREDweb/Issues-papers/researchtoaction.pdf>. [3 February 2003] (哈佛格式)

该文献为 COHRED 工作组的一份报告,可以纸质版或 PDF 的文件格式获得。若仅是通过在线阅读浏览(下载打印或在显示器上阅读),应在题目加上标识(online)。

范例 2: Hodnett, E. D. 2000. 'Support during pregnancy for women at increased risk' (Cochrane Review), in *The Cochrane Library*, [CD-ROM], Issue 2, Update Software, Oxford, UK. (哈佛格式)

范例 2 为一篇系统评价,来自 Cochrane 系统评价数据库,采用哈佛引用格式。Cochrane 评价文献可以通过 Cochrane 图书馆光盘或者借助 OVID 在线获得。在此例中,需要在题目中加注(CD-ROM)。



## (二) 获得途径

出版商与出版地对于电子文献意义不大,获取途径才是引文中最重要的内容。互联网上的文献或者网页,需提供 URL 地址。若同时有多个在线数据库与电子杂志汇编,需提供单个数据库或电子杂志汇编的名称。具体的说明格式包括:“从××处拷贝”或“可从××处获得”等。

### 范例 3: 电子杂志文献

Notess GR. (1997). On the Net: Internet Search Techniques and Strategies. *ONLINE Magazine* [Online], 21 (4) Available: <http://www.onlineinc.com/onlineinc/mag/JulOL97/net7.html> [1999, November 18] (温哥华格式)

注意随时保持检查更新,确保所提供的 URL 地址能正常使用。为此,Web 地址(URL)要记录准确,地址中的上标、下标、符号(如#, @, ~)等均应如实写出,最好是直接将 URL 地址复制粘贴。

美国心理协会(ASA)建议在撰写初稿、投稿、论文最后定稿以及评价证据等几个时点,检查与更新 URL。若在此过程中,引用文献不能获取,应改用其他文献替代,或者干脆删掉不用。

## (三) 获取日期

互联网信息具有动态性,这就意味着位置、内容会经常变换,甚至完全消失。因此,获得文献的具体日期应包括在引文中。温哥华与哈佛格式在引文的末尾处均用方括号标出。而 ASA 格式则是将获取的日期与途径合并在一起(见范例 4)。同时应注意与前两种格式不同,ASA 格式并不要求标明媒介类型。

范例 4: Vanden Bos, G. Knapp (2001). Role of reference elements in the selection of resources by psychology undergraduates. *Journal of Bibliographic Research*, 5, 117-123. Retrieved October 13, 2001 from <http://jbr.org/articles.html>

## (四) 引文序列号和页码的编排格式

温哥华格式是按照引用的顺序,每个参考文献给予一个序列号。哈佛与 ASA 格式则在正文中插入作者与日期,并用圆括号标出。不论何种格式,引用文献的页码必须标明。

一些电子文献,特别是 WWW 网页,页码的意义不大。例如一篇电子文献只有一页网页,但可打印 50 页。对此,不用页码标明文献的具体位置,但可以利用符号¶或者用缩写“para”注明段落序号。如范例 5: (Myers, 2000, para. 5) (APA 格式)

## 二、WWW 网页

在引用 WWW 网页时,若不能确定网页的具体作者,可在作者姓名位置用网页名称代替,不能用推测的作者姓名。温哥华格式允许直接采用网站组织或负责人名称替代。这在作者不明以及文献转移或丢失时很有帮助。也可以从网站组织或负责人处,获取找到该引用文献的途径与方法。

在引用一般网页时,浏览器中显示的题目就作为引文题目,否则标为“无题文献”。在参考文献中,应标明提供的日期是版权日期,还是更新日期。若没有具体日期,可用缩写 n. d 表示。

范例 6: Kleiber PB, Holt ME, Swenson JD. (n. d.). *The Electronic Forum Handbook: Study Circles in Cyberspace*, [Website of the Civic Practices Network], [Online] Available: <http://www.cpn.org/cpn/sections/tools/manuals/electronic-handbook1.htm> [2002, October 8] (温哥华格式)



### 三、CD-ROM (光盘)

光盘出版物与纸质出版物引用格式大致相同，但必须标明媒介类型。

范例 7: Rosselli D., Otero A. & Dennis R. 2003. 'Demand Creation for Health Research in Developing Countries' in *INCLIN Leadership and Management Program*, [CD-ROM], INCLIN Trust, Manila, Philippines. (哈佛格式)

### 四、私人邮件信息

引用私人邮件信息应该包括如下要素：①作者姓名；② [作者电子信箱地址]；③信息中的一行添加为题目；④通讯日期作为发表日期；⑤通讯种类（如主送信件，转发或抄送信件）；⑥收件人姓名与电子邮件地址用来标明获取途径；⑦访问日期。

范例 8: 刘×××。2003 年 2 月 3 日，第三次医学信息学学术会议，主送：张×××。获取途径：kang×××@126.com (哈佛格式)

注意在引用他人电子信件信息，一定事先征求发件人本人的同意，遵守知识产权的相关规定。

### 五、网上公告栏信息

网上公告栏信息的引用与私人邮件信息一样，应该包括如下要素：①作者姓名；② (作者电子信箱地址)；③信息中的一行添加作为题目；④公告发布日期作为发表日期；⑤网上公告信息栏地址用作说明获取途径；⑥访问日期等。

(康德英 廖晓阳)

# 第五章 如何阅读与评价医学研究文献

临床医生每天面临许多不同的病人，常常出现某些问题是我们现有的知识不能回答的。而随着时间的飞逝，我们现有的知识和临床技能也在逐渐过时。如何为患者提供最新的信息，帮助他们进行医疗决策？如何更新我们自己的知识，紧跟医学发展的前沿？这是每一位医务工作者面临的巨大挑战。

1956 年哈佛大学医学院院长 Burwell 教授曾说过：医学生在校期间所接受的知识中，有一半在 10 年内将证明是错误的，而糟糕的是，没有一位教师知道，哪一半是错误的。而最新研究发现，每 3 年，70% 的知识可能已过时。

因此，如何阅读与评价医学研究文献，在知识爆炸的今天，就显得十分重要。

## 第一节 阅读与评价医学文献的重要性

### 一、医学文献的种类

医学文献中的临床研究分为两类（图 5-1）：①原始研究证据（primary studies）：即原始论著，分为试验性研究（experimental studies）和观察性研究（observational studies）；②二次研究证据（secondary studies）：即根据原始论著进行综合分析、加工提炼而成，包括 Meta-分析、系统评价、综述、评论、述评、实践指南、决策分析和经济学分析等。

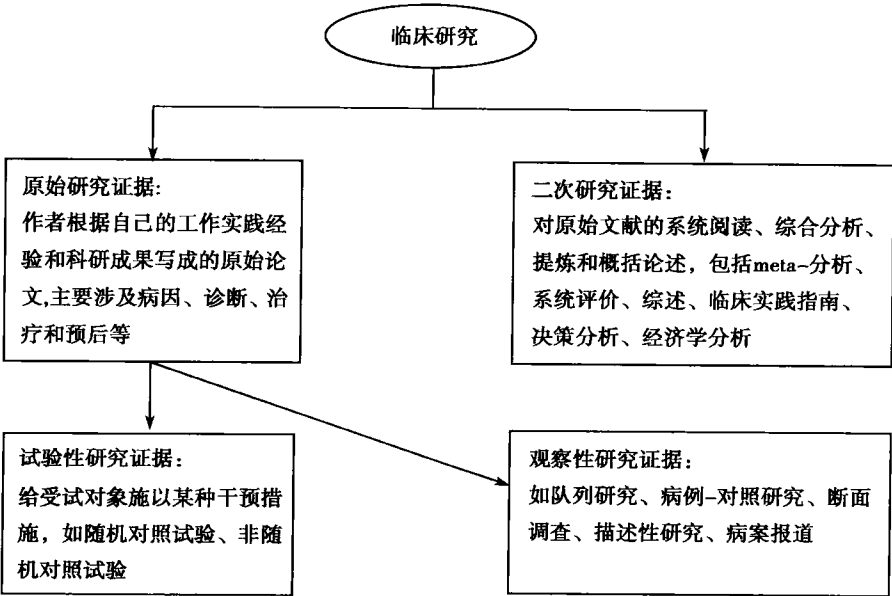


图 5-1 临床研究分类





## 二、阅读和评价医学文献的重要性

### (一) 阅读医学文献的重要性

在日常工作中，为了解决临床医疗实践中有关疾病病因、诊断、治疗、预后的问题，掌握新近的诊断技术和治疗措施，跟上医学发展前沿；为了在教学工作中向学生传授最新的知识；为了在科研工作中了解某一研究领域的历史、现状、发展趋势和存在问题，提出今后的研究方向；为了保持知识的不断更新，我们必须不断阅读医学文献资料。

然而，全世界每年有 200 多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，一个内科医师需要每天不间断地阅读 19 篇本专业文献才能基本掌握本学科的新进展、新研究结果。因此，医务人员应该改变阅读文献的方法和技巧，把自己从一个医学信息的记忆库变成医学信息管理者 (From memory repositories to information managers)，从考虑如何跟上医学领域的新进展转变为明确医学领域的哪些新进展是需要跟上的 (From “How do I keep up with new developments in medicine?” to “What developments in medicine do I need to keep up with?”)，充分利用有限的时间和精力阅读高质量、有临床价值的文献，因此，如何提高阅读效率，就显得非常重要了。

### (二) 评价医学文献的重要性

医学文献作为临床实践、医学教育、医学研究的证据，其质量至关重要。不同来源、不同类型研究在设计、实施、统计分析和论文报告方面，其质量并不一致，因此，在应用时必须进行严格评价 (critical appraisal)，将真实、可靠、有实用价值的文献资料用于工作中。

1. 证据来源复杂 随着计算机信息技术和医学信息的迅猛发展，患者越来越容易获得各种医学知识并寻求医务人员的解释。而各种媒体如报纸、收音机、电视、非专业的流行杂志和国际互联网等提供的医学信息和对疾病的建议，有时相互矛盾或缺乏严格的科学依据，患者因医学知识的限制缺乏鉴别真伪的能力，使医务人员每天都面临许多类似问题而不堪重负。

2. 证据质量良莠不齐 虽然全世界每年有 200 多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，但针对某一专题的医学文献中真正有用的不足 15%，多数文献未经同行严格评价。即使发表在最著名的医学杂志上的文章也不一定完美无缺。对发表在许多医学杂志上的临床试验进行分析发现，从设计、实施、结果分析和文章撰写等方面均存在较大缺陷。结果是，许多诊断试验和治疗方法未经严格评估就进入临床常规应用，给患者造成严重危害。临床医务人员面临的挑战是如何应用真实、最新的医学信息为患者治病，尽可能保存功能、减少痛苦和症状以延长患者的寿命，提高患者的满意度。要做到这点，不仅要求医务人员有高度的热情，还要掌握基本技能，包括掌握严格评价医学文献的技巧，在信息的海洋中系统、全面而又快速、有效地获取所需要的临床医学研究文献，掌握快速阅读和正确评价临床医学文献的基本原则和方法，筛选出真实、有临床意义的研究证据应用于临床实践，为患者作出最佳的医疗决策。

3. 临床研究证据必须结合病人具体情况 我们主管的病人与临床研究证据中的研究对象存在性别、年龄、并存症、疾病严重程度、病程、依从性、社会因素、文化背景、生物学及临床特征的差别，即使是真实、可靠且具有临床价值的研究证据，也不一定能直接应用于每一个病人，医务人员必须根据自己的临床专业知识、分析评价相关临床研究证据，综合病人的具体情况和选择，作相应调整。切忌生搬硬套，以防误导。



## 第二节 高效率阅读医学文献的基本步骤

### 一、明确阅读文献的目的

提高阅读文献的效率,必须首先明确阅读文献的目的,明确希望从文献中获得什么样的信息,以指导选择目标杂志、数据库和文献类型。例如,要了解 $\beta$ -受体阻滞剂在心力衰竭患者中的应用价值,应先查寻有无相关的系统评价或高质量文献综述,因为这类文献浓缩了大量原始文献的信息,特别是系统评价,其严格的方法学使文献的结论具有很高的真实性和可信度,可节省读者逐篇阅读和评价原始文献的时间和精力,快速和有效地获取有价值的信息资源。如果没有相关的系统评价或综述,再查寻和阅读原始文献。

### 二、熟悉文献的基本结构

临床实践中,要正确选择和应用某种药物,医师必须熟悉该种药物的成分、药代动力学、适应证、禁忌证和有效剂量范围等,否则就可能选药不当,造成病情延误和增加患者的经济负担。快速阅读医学文献正如临床选药一样,必须了解文献的基本结构和组成,及每一部分重点阐述的问题,否则浪费大量阅读时间却一无所获。例如,大多数原始论著均包括摘要、前言、材料和方法(或对象和方法)、结果、讨论(包括结论)和参考文献六部分,如果了解某篇文献的结论是否适合于自己的病人,可直接阅读方法学部分了解其设计方案、病例的选择标准等以判断其结论的应用范围,无需从头读到尾。

### 三、选择性阅读文献

既然繁忙的医务工作者不可能博览所有相关文献,掌握选择性阅读目标杂志中的医学文献以获得丰富信息的技巧就十分重要。

#### (一) 阅读感兴趣和有临床应用价值的文献

如果时间和精力有限,应根据文献的题目、著者和摘要等选择自己感兴趣或对临床实践有价值的文献精读。如果不感兴趣,则无需浪费宝贵的时间。

#### (二) 快速浏览文献

浏览使读者能快速了解文章的基本结构和内容,有助于放弃阅读一些貌似有临床实用价值,实为有关高度专业化的实验室工作或复杂的数字合成的文献,这样做就能节省阅读时间。

#### (三) 集中阅读文献的方法学部分

通过前面两步骤后,读者可以决定是否有必要精读这篇文献。如何精读呢?多数读者喜欢仔细阅读摘要、结果和讨论部分,有的读者甚至只读讨论部分,认为节省时间而不丢失信息。而用小字体印刷的方法学部分常常被忽略。实践证明,上述阅读方法并不正确。阅读文献应首先阅读方法学部分。如果一篇文献的设计、实施和统计分析都是错误或不当的,无论其结果多么诱人或作者的讨论多么深刻,其结论均不真实、不可靠。对方法学有严重缺陷的文献,不值得花时间精读全文。

#### (四) 可保留对文献的最后裁决权

各级临床医师常常结合临床问题阅读文献,但多数临床医师在应用文献结果处理主管病人时并未对其研究的真实性、可靠性和临床价值进行严格评价(critical appraisal),有



可能被低质量的文献所误导。因此,读者阅读医学文献时应该采用临床流行病学/循证医学评价文献的原则和方法对文献进行严格评价,并结合病人的具体情况综合考虑,而不能盲目遵从文献得出的结果和结论。选择性阅读文献的基本步骤见图 5-2。

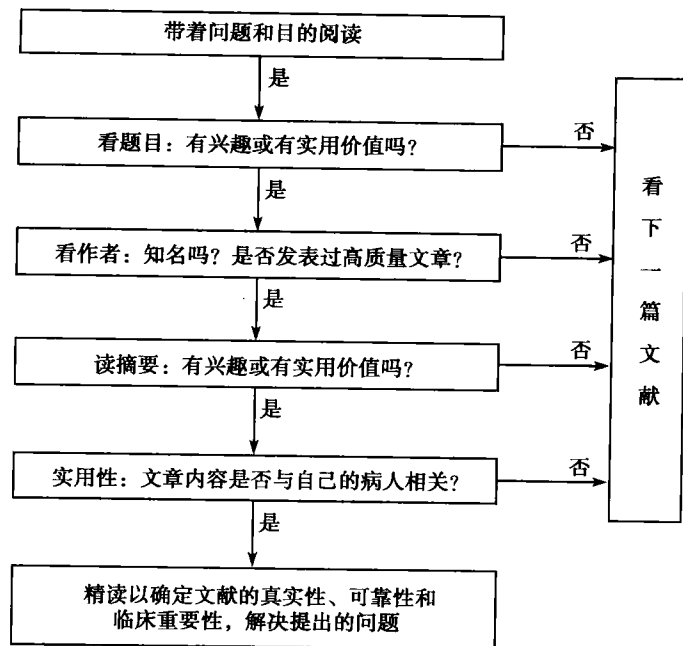


图 5-2 选择性阅读文献的基本步骤

### 第三节 如何评价医学文献

评价临床医学文献是阅读文献的重要步骤,目的在于避免被动地接受文献的观点和结论,而且对文献的价值给予公正、科学的评定,以便将科学、可靠、有临床价值的医学文献应用于临床、教学、科研和卫生政策制订中,提高我国有限卫生资源的利用率,改善人民健康。

但繁忙的临床医生往往难以花数小时时间阅读和严格评价一篇文章,因此应有一套简洁、有效的筛选、评价方法,让临床医生能快速、有效地判断文章的真实性和临床实用价值,为患者作出科学的诊治决策。为此,加拿大 McMaster 大学的临床流行病学和统计学教研室的专家们从 1981 年开始即发表了系列有关疾病病因、诊断、治疗和预后的评价标准,帮助临床医生严格阅读和评价临床研究证据。JAMA 从 1992 年至 2002 年已发表了 26 种研究证据的评价原则,美国医学会 (JAMA & Archives Journals) 已将其以手册和专著的形式于 2002 年发表。

评价临床研究证据的步骤可分为三步:①初筛临床研究证据的真实性和相关性;②确定研究证据的类型;③根据研究证据的类型评价其真实性和适用性。

#### 一、初筛临床研究证据的真实性和相关性

阅读和评价临床研究证据的第一步,阅读者应问自己:“这篇文章是否值得花时间精读?”,要回答这个问题,可参考表 5-1 中的 6 个简单问题。



表 5-1 初筛临床研究证据的真实性和相关性

这篇文章是否值得花时间精读?	是	否
1. 这篇文章是否来自经同行评审 (peer-reviewed) 的杂志?	继续	停
2. 这篇文章的研究场所是否与你的医院相似, 以便结果真实时可应用于你的患者?	继续	停
3. 该研究是否由某个组织所倡议, 其研究设计或结果是否可能因此受影响?	暂停	继续
阅读这篇文章摘要的结论部分, 确定相关性		
4. 如果文章提供的信息是真实的, 对我的患者的健康有无直接影响, 是否为患者所关心的问题?	继续	停止
5. 是否为临床实践中常见的问题, 文章中涉及的干预措施或试验方法在我的医院是否可行?	继续	停止
6. 如果文章提供的信息是真实的, 是否会改变现有的医疗实践?	继续	停止

### 1. 这篇文章是否来自经同行评审 (peer-reviewed) 的杂志?

多数国内外医学杂志的文章均经过同行评审后才发表, 但也有少数杂志未经同行评审或其目的是通过广告获得经济效益。

有同行评审的杂志上发表的文章均经过了严格的评审过程, 尽可能筛除有严重缺陷的文章, 提高了发表文章的质量。一般来说, 杂志收到研究人员投稿后, 先由编辑部确定该文章是否适合在本杂志上发表, 如果可以接受, 再送给至少两名评审人员。同行评审人员都不是编辑部成员, 常常为某一领域和研究方法学上的专家。同行评审的目的是为了筛除设计不合理、不重要或难以解释、有缺陷的文章, 以提高文章设计和统计分析方法的质量。如内科学年鉴 (Annals of Internal Medicine) 每年可收到约 1200 篇论著, 编辑部会筛除一半, 剩下的一半每篇由至少两名评审员评审, 最终只发表 15%。因此, 尽管称不上完美, 同行评审仍被公认为提高医学文献报告质量的重要方法。不可否认, 即使经过很好的同行评审, 发表偏倚 (publication bias) 也不可能避免, 阳性结果和大样本研究常常比阴性结果的研究容易被发表。

### 2. 这篇文章的研究场所是否与你的医院相似, 以便结果真实时可应用于你的患者?

这个问题可以通过阅读作者的单位或进行研究的场所确定。如果你在乡村医院工作, 阅读的文章是在某个大学的专科病房所进行的研究, 你就要考虑其结果应用到你的病人时可能存在环境条件以及患者本身的差异, 当然这不是拒绝这篇文章的重要理由, 但如果差异太大, 应谨慎考虑。

### 3. 该研究是否由某个组织所倡议, 导致其研究设计或结果可能受到影响?

这个问题主要考虑研究资金的来源可能导致的偏倚。大多数杂志要求研究人员说明研究资金的来源。临床医师应注意: 药厂在同行评审的杂志上主办的专刊往往带有促销性质, 其题目容易误导医师和患者, 且多采用商品名, 或者不像正刊一样经过同行评审。另外, 发表在增刊上的随机对照试验的质量往往不如正刊。这并不是说具有商业目的的研究都存在偏倚。相反, 许多设计完善的大型研究均有药厂资助。但如果一个研究有药厂或其他商业组织资助, 应要求研究人员保证其研究的设计和结果并未因此而受到影响。

下面的三个问题涉及临床相关性, 可通过阅读摘要的结论和相关部分获得。临床相关性对医生和病人均很重要。一般来说, 如果一篇文章涉及的问题临床很少见, 或涉及的治疗或诊断方法在你的医院根本没有开展, 这样的文章不值得阅读。即使阅读也只是满足了你的学术好奇心, 不会影响你的临床实践。因此, 以病人为中心的临床研究证据远比以疾病为中心的研究证据实用。如一篇文章描述了某种方法用于筛选某种癌症的敏感度和特异度 (以疾病为中心的证据), 而另一篇文章描述经过这种筛选试验可以提高患者的生存质量或寿命 (以病人为中心的证据), 你应该首先阅读后一篇文章。

### 4. 如果文章提供的信息是真实的, 对我的患者的健康有无直接的影响, 是否为患者



所关心的问题?

可以通过阅读文章摘要的结论部分初步解决这个问题。例如,如果某篇文章的结论为通过某种治疗方法,脑卒中患者偏瘫肢体的肌电图有明显改善,并未涉及肌力和活动能力;但对于患者、医生来说,可能更关心的是经过治疗后偏瘫肢体的肌力是否改善、能否活动,因此这篇文章提供的信息并不是你的患者所关心的问题。

5. 是否为常见临床问题,文章中涉及的干预措施或试验方法在你的医院是否可行?

如果文章涉及的问题在临床实践中经常遇到,且研究的干预措施或试验方法在你的医院也有条件实行,这样的文章值得深入阅读。

6. 如果文章提供的信息是真实的,是否会改变现有的医疗实践?

如果文章涉及的干预措施或试验方法,你过去未在类似的病人中使用过,也许新的尝试可能获得意外的收获,因此,有必要继续阅读这篇文章。

综上,数分钟就可以回答上述6个问题,从而决定你是否值得花时间精读某一篇文章。当然,这些条款均着重于“带着病人的临床问题”去阅读文献,至于为专题科研而收集相关文献或为更新知识去阅读文献则通常可不受这些条款的“制约”。

## 二、确定研究证据的类型

如果你决定继续阅读某一篇文章,下一步就要明确该研究的目的和要解决的临床问题是什么?这可通过阅读文章的摘要,必要时文章正文的前言以确定研究的目的。一般来说,原始论著回答的主要问题有四类:病因、诊断、治疗和预后(表5-2)。而二次研究证据尚有meta-分析或系统评价、临床指南、决策分析或经济学分析等。

表5-2 原始研究涉及的主要临床问题及其常用的设计方案

临床问题	常用设计方案
病因:评价某种因素是否与疾病的发生有关?	队列研究或病例-对照研究
诊断:评价某一诊断试验的真实性和可靠性?或评价某一试验在应用于人群时检测临床前期病例的准确性?	诊断性试验的设计(将新的试验与金标准进行比较)
治疗:评价某种治疗方法如药物、外科手术,或其他干预措施的效果?	随机、双盲、对照试验
预后:确定疾病的结局	队列研究、病例-对照研究

## 三、根据研究类型评价医学文献

临床医学文献的评价应采用临床流行病学/循证医学的原则和方法。根据研究的侧重点如疾病病因、诊断方法、治疗措施、预防、预后、系统评价、卫生经济学和决策分析等的不同,选择不同的具体评价的原则和方法,详细的评价内容详见本教材相关章节。

## 四、评价医学文献的一般原则

不论评价哪一种临床医学文献,都应从文献的真实性、临床重要性和实用性三方面进行综合考虑。

### (一) 文献的内在真实性

内在真实性是评价研究证据的核心。文献的内在真实性(internal validity)是指就该文章本身而言,其研究方法是否合理、统计分析是否正确、结论是否可靠、研究结果是否支持作者的结论等。例如:评价治疗性研究,应考虑合格病例是否随机分配到不同的治疗



组？随机化方法是否完善隐匿？统计分析时是否按随机分配的组别将全部研究对象纳入分析？是否采用盲法等？如果一篇文献内在真实性有缺陷，自然就没有多大的价值。

### （二）文献的临床重要性

文献的临床重要性是指研究结果本身是否具有临床价值。评价研究结果的临床价值主要采用一些客观指标，而不同的研究类型其指标不同（详见相关章节）。例如：治疗性研究可采用相对危险度降低率（relative risk reduction, RRR）、绝对危险度降低率（absolute risk reduction, ARR）和与对照组比较，多减少一例某种事件的发生需要治疗的病例数（number needed to treat, NNT）等判断某种治疗措施的净效应及其临床价值；而诊断性试验则采用敏感度、特异度、阳性和阴性预测值、似然比及 ROC 曲线等指标判断某种诊断试验的价值。

### （三）文献的适用性

一篇真实性强且具有重要临床价值的文献（研究成果），是不是能被采用，亦即适用性（applicability）是应予考虑的。如能被广泛推广应用，则其外部真实性（external validity）或推广性（generalizability）佳。这里重要的是要考虑应用这类成果（证据）时，自己医院的硬件设施与医疗环境以及自己病人的具体病情及其自愿性，乃至经济承受能力等。例如，大型临床试验和系统评价均证实使用  $\beta$ -受体阻滞剂对心力衰竭病人有益，但当你主管的心衰病人有糖尿病、目前正在使用胰岛素治疗、有明显血脂增高时，是否立即使用  $\beta$ -受体阻滞剂就需要仔细权衡其利弊，而不能盲目遵从文章结论。

（李 静 廖晓阳）

## 第六章 医学文献综述与系统评价

假设1: 作为一名呼吸科医师, 你了解有关各种戒烟方法的疗效和安全性, 为吸烟者或病人戒烟提供最新的研究证据。为此, 你检索了MEDLINE, 发现有上千篇临床试验, 涉及的戒烟方法也非常多, 仅研究尼古丁替代品的临床试验就有150篇, 且结论也不一致。由于临床工作太繁忙, 你既不清楚该相信哪些文章的结果, 也没有足够的时间、精力去阅读众多的文章。也许你希望能有一篇文章将所有相关信息综合在一起, 以助你的医疗决策。

假设2: 早产儿因肺发育未成熟, 死亡率和呼吸窘迫综合征的发生率都非常高。有专家提出, 如果给可能早产的孕妇使用激素, 可促进胎儿的肺发育。为了明确对可能早产孕妇使用激素后, 是否能减少早产儿的死亡率和呼吸窘迫综合征的发生率, 你搜寻了有关资料, 发现有7个高质量的随机对照试验, 其中5个试验结果为阴性(使用激素后未能减少早产儿的死亡率和呼吸窘迫综合征的发生率), 2个试验结果为阳性, 你将作何结论?

在当今知识爆炸和医学研究层出不穷的时代, 医务人员和研究者要想在有限时间内获得更多新知识、新观点和新技术以扩展视野, 提高学术水平, 需要阅读大量的文献。但是, 全世界每年有200多万篇生物医学文献发表在2万多种生物医学杂志上, 一个内科医师需要每天不间断地阅读至少19篇本专业的文献才能基本掌握本学科的新进展。Sackett等调查了一周之内内科临床医师阅读医学文献的时间, 发现阅读时间的中位数不超过90分钟。而高年资住院医师以上的各级医师中, 竟然有15%~40%在过去一周内未阅读过任何医学文献。可见, 如果要求人人都博览群书, 似乎是不太现实的。为了能节省阅读时间和尽快获得本专业的最新进展和信息, 多数临床医师喜欢阅读医学文献综述。如何科学地利用已有的医学信息, 撰写高质量医学文献综述, 为繁忙的临床医务工作者提供大量真实、可靠的最新证据, 对于临床医疗实践和医学科研工作, 至关重要。

医学文献综述(review)是对原始文献的第二次加工, 受原始文献的质量、进行综述的方法及综述者本人的认识水平和观点的制约, 读者在阅读或接受文献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。因此, 了解、熟悉并掌握医学文献综述方法和质量评价的标准是十分重要的。

### 第一节 文献综述的目的与特点

高质量的医学文献综述有重要的实用价值, 表现在四个方面:

#### 一、科学研究工作的需要

面对浩瀚的文献资料, 需要分析、评价和综合, 才能了解和掌握与科研工作相关学科和领域里的有关知识的历史和现状、存在的问题、当前研究的热点与矛盾、发展的趋势, 敏感地剖析矛盾, 抓住关键, 找准方向, 突破难点。因此, 文献综述是医学研究选题和立题的基础, 借助综述可为研究工作提供信息资料和研究方向。

#### 二、反映学科新动态

围绕专业发展的热点, 纵览某一领域的最新文献资料, 做好有关专题的文献综述, 全



面、深入和集中地反映该领域目前的新动态、新理论、存在问题和发展方向，为促进学科的发展，保证不断地吸收新知识、新营养而居于学科的前沿位置服务。

### 三、继续教育的需要

Ransey 等发现医学生掌握新知识的情况与从医学院毕业的年限呈显著负相关关系。而最新研究发现，每 3 年，70% 的知识可能已过时或被新的知识取代。这就形象地说明，随着时间的飞逝，我们现有的知识和临床技能需要不断更新。

因此，对于广大的基层医务工作者，由于工作繁忙，文献资源有限，为了知识的不断更新，需要有实用价值的、真实可靠的文献综述，作为广大医务人员的继续教育资源，使临床医师在最短的时间内吸取尽可能多的新知识、新观点和新技术，帮助他们及时扩展视野。

### 四、临床医疗和管理决策的需要

医疗和卫生管理决策不能单凭个人的主观意志和经验，必须基于科学的研究证据。以某一关键问题为中心去筛选医学研究文献，通过综合和评价，写出科学性强、实用价值大的文献综述，指导临床医疗实践和卫生决策，促进医疗和管理决策科学化。

## 第二节 文献综述的类型

根据综述方法的不同，分为两大类：叙述性文献综述和系统评价。

### 一、叙述性文献综述

叙述性文献综述 (narrative review) 为传统的文献综述，是由作者根据特定的目的和需要或兴趣，收集有关的文献资料，采用定性分析的方法，对论文中阐述的研究目的、方法、结果、结论和观点等进行分析和评价，用自己的观点和判断，将一系列相关文献，经过综合归纳、条理化，综合成文。如果作者具有一定的水平，对专业有较深入的了解，收集的文献充分，经过认真阅读和评价，撰写时有明确、严格、科学、系统的方法，就会写出高质量的文献综述，否则可能产生误导。

Oxman 等曾对撰写传统文献综述的专家及他们撰写的文献综述采用下述四条原则进行评价：①是否收集了所有的相关研究文献；②是否无偏倚地列出了所有的参考文献；③是否对引用的文献进行了科学的评价；④是否对文献资料进行了恰当分析和总结。结果发现：①专家们互相之间对撰写文献综述时是否遵循了上述四条原则，均不能达成共识；②按上述原则对专家们自己撰写的文献综述进行评价，发现对上述原则的依从性与专家的专业知识水平呈负相关关系，即专业知识水平越高，对此原则的依从性越低。这些原则无论对作者撰写文献综述或者读者阅读与分析评价叙述性文献综述的质量和水平，都有重要的参考价值。

### 二、系统评价

系统评价 (systematic review) 是一种全新的文献综合方法，指针对某一具体临床问题 (如疾病的病因、诊断、治疗、预后)，系统、全面地收集全世界所有已发表或未发表的临床研究，采用临床流行病学的原则和方法严格评价文献，筛选出符合质量标准的文献，进行定性或定量合成 (meta-analysis, 荟萃分析或汇总分析)，得出综合可靠的结论。





系统评价可以是定性的（定性系统评价，qualitative systematic review），也可以是定量的（定量系统评价，quantitative systematic review），即包含 Meta-分析过程。系统评价的整个过程非常明确，使其具有独特的优点：良好的重复性，可为某一领域和专业提供大量的新信息和新知识，最有效地利用有限的卫生资源为人类健康服务。

系统评价是将经过严格评价（critical appraisal）的高质量研究论著，加以综合分析和评价，使得多个研究成果综合为一体，从而获得更为科学可靠的结论，推广应用于临床实践的高级系统工程。进行系统评价，需要具备较深厚的学术理论基础和丰富的临床经验，经过专门训练，掌握了系统评价的方法学之后，方能承担这一任务。

现有的系统评价，无论在数量上或质量上都不能满足临床医生和医学决策者的需要。因此，有关国家的临床医学专家、医学研究的方法学家、系统评价的专业人员以及临床用户联合起来共同成立了一个国际性的组织—Cochrane 协作网（Cochrane Collaboration），并相继在有关国家成立了 Cochrane 中心，目的在于协调本国收集原始的临床研究资料，或进行系统评价，或将国际上发表的系统评价翻译为本国语言进行发表传播，为临床医学实践，特别是为疾病的治疗、预后和康复提供了大量高质量的科学依据。Cochrane 协作网通过杂志、电子刊物即 Cochrane 图书馆（Cochrane Library）和国际互联网（Internet）的形式将系统评价传播到世界各国，使全球医务工作者均能受益。

Cochrane 系统评价是 Cochrane 协作网的评价人员按照统一工作手册（Cochrane reviewers' handbook），在相应 Cochrane 评价小组的指导和帮助下所完成的系统评价。由于 Cochrane 协作网有严密的组织管理和质量控制系统，严格遵循 Cochrane 系统评价者手册，采用固定的格式和内容，统一的系统评价软件（RevMan）录入和分析数据、撰写系统评价计划书和报告，发表后根据新的研究定期更新，有完善的反馈和修改机制，因此 Cochrane 系统评价的质量比非 Cochrane 系统评价更高，被一些学者认为是单一性评价干预措施疗效的最好信息资源（best single source）。

Cochrane 协作网产生的系统评价，在临床医学界的重大影响在于①肯定一些有效的疗法并推广应用；②否定一些无效或有害的疗法并予抛弃；③发现某些有希望、但缺乏足够依据的疗法，建议开展进一步的研究，促进了某些重大课题的实施。因此，这种形式的文献综述具有对临床科学研究及疾病防治实践的导向性。目前，Cochrane 系统评价的结果被作为许多发达国家卫生决策的依据，影响着这些国家的医疗实践、卫生决策、医疗保险、医学教育、医疗科研和新药开发，促进 21 世纪的医学从经验医学向循证医学转化。

### 三、叙述性文献综述与系统评价的区别与联系

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

叙述性文献综述常常涉及某一问题的多个方面如糖尿病的病理、病理生理、流行病学、诊断及预防、治疗、康复的措施，也可仅涉及某一方面的问题如诊断、治疗等。系统评价或 meta-分析均为集中研究某一具体临床问题的某一方面，如糖尿病的治疗或康复，具有相当的深度。因此，叙述性文献综述有助于了解某一疾病的全貌，而系统评价则有助于某一具体疾病的诊治。二者的主要区别见表 6-1。



表 6-1 叙述性文献综述与系统评价的区别

特征	叙述性文献综述	系统评价
研究的问题	涉及的范畴常较广泛	常集中于某一临床问题
原始文献来源	常未说明、不全面	明确, 常为多渠道
检索方法	常未说明	有明确的检索策略
原始文献的选择	常未说明、有潜在偏倚	有明确的选择标准
原始文献的评价	评价方法不统一或未评价	有严格的评价方法
结果的综合	多采用定性方法	多采用定量方法
结论的推断	有时遵循研究依据, 较主观	多遵循研究依据, 较客观
结果的更新	未定期更新	定期根据新试验进行更新

### 第三节 叙述性文献综述的撰写方法

叙述性文献综述是针对某一问题选择性评价部分医学文献并展开广泛讨论。有专家倡导, 针对临床问题的传统文献综述 (clinical review article), 如果能基于已有的系统评价和 Meta 分析展开讨论, 提供的信息将更具说服力和指导意义。并且, 此类文献综述应按照循证的格式进行撰写, 包括选择主题、检索文献、评价文献、证据分级等, 称为循证临床综述 (evidence-based clinical review)。

#### 一、确定文献综述的目的和选题

既然文献综述主要是反映学科动态, 以服务于科研选题、立题, 继续教育, 临床和卫生决策为目的, 故所选题目的内容必须能反映该学科领域的新理论、新知识和新技术, 反映学科前沿、最新的进展以及该领域某些有争议的问题。

临床综述的选题, 应选择常见的、在诊断和治疗方面有新进展的临床问题, 强调新信息的真实性和对变革临床实践的价值。例如, 眼科和家庭医生对单纯、无并发症的角膜擦伤病人的标准治疗模式包括用眼罩、散瞳剂、抗菌眼液。很少有人怀疑戴眼罩是否有益, 因为这是“常识”, 无须质疑。但随机对照试验和系统评价均显示, 让角膜擦伤的患者戴眼罩, 与不戴眼罩比较, 可引起更多症状, 并延缓擦伤的愈合。

#### 二、广泛收集文献

文献综述的目的和题目确定后, 应选定相关的主题词, 制订正确的检索策略, 充分利用各种检索工具和文献资料库, 广泛收集相关文献。此外, 还可以利用期刊杂志每年最末一期附录的文题索引、专著或教科书及其有关参考文献等作为文献综述的信息资源。应注意的是, 专著或教科书的参考文献不一定是最新的文献。如果综述的目的是反映学科新知识和新进展, 应收集近年来最新的文献。

针对临床问题的循证综述, 在检索文献时, 应重点检索循证医学资料库, 检出与题目相关的、高质量的系统评价、meta 分析、单个随机对照试验和临床实践指南等, 汇总他们的推荐意见。如果未发现相关资料, 应引用具有权威性专家的共识意见, 并明确指出尚缺乏高质量研究证据支持目前的临床应用。

#### 三、阅读和评价

收集的文献常常较多且质量差别较大, 因此应根据文献综述的目的和内容, 制订文献入选的标准以选择符合要求的文献, 并认真阅读和分析, 采用临床流行病学/循证医学的



原则和标准进行评价,以真实可靠的文献作为支持文献综述结论的基础。

对循证临床综述,同样应对支持结论的文献资料的论证强度和真实性进行评价,并对研究证据进行分级,而这种分级不只是强调研究的设计类型,如随机对照试验、队列研究、病例-对照研究的重要性,更重要的是强调文献资料的内在真实性和总体研究质量的重要性。

#### 四、撰写文献综述

将经过严格评价的高质量、结果真实、可靠的文献进行整理归纳、综合成文。文献综述的撰写格式,可分为题目、正文、结语、引用的参考文献等四个部分。在正式撰写前,务必按照目的和要求以及上述四个部分的内容,先设计好撰写的大纲和设置好撰写的条目及内容,然后根据该大纲的要求,逐项地将收集的文献资料条理化地归纳综合,进行撰写。

对循证临床综述,专家建议分为四部分:①前言:定义综述的题目并阐明综述的目的;②方法:阐明文献检索的方法及检索的资料库,最好能制订选择文献的标准,包括证据的质量评价和分级方法,以避免有意识选择支持你结论的文献。如果文献资料的结果存在不一致性,应分析和阐明;③讨论:可采用经典的综述格式,包括临床问题涉及的病因、病理生理、临床表现、诊断、鉴别诊断、治疗、预后及未来研究方向等。若存在争议、未解决的问题、新进展、不同观点、明显的利益冲突或可能影响证据强度的偏倚,综述应全面分析。突出有证据支持的观点,若证据不足,则强调共识性观点。若缺乏共识,应阐述目前较公认的实践行为或讨论一种或多种合理的方法,同时指出这些建议缺乏确凿证据支持。④参考文献:包括当前最新的和最重要的支持主要观点的资料。

总之,对于一篇好的叙述性文献综述,要求题目要简明,内容要精练,重点要突出,条理要清楚,观点要明确,结论要科学。

### 第四节 系统评价的撰写方法

系统评价多数是对已有的研究资料进行评价、分析,属回顾性研究,因此其质量必然受原始研究质量的影响;同时,系统评价的方法、步骤的正确与否,对其结果和结论的真实性、可靠性起着决定性的作用。

系统评价本身只不过是一种研究的方法学,并不仅限于随机对照试验或仅对治疗措施疗效进行系统评价。根据研究的临床问题不同,系统评价可分为病因、诊断、治疗、预后、卫生经济评价和定性研究(qualitative research)等方面的系统评价;根据系统评价纳入的原始研究类型不同(study design),可分为临床对照试验(controlled trial)和观察性研究(observational studies)的系统评价,前者如随机对照试验和非随机对照试验的系统评价,后者如队列研究和病例-对照研究的系统评价;根据进行系统评价时纳入原始研究的方式,可分为前瞻性、回顾性和累积性系统评价;根据资料分析时是否采用统计学方法(meta分析),可分为定性和定量的系统评价。目前,由于根据随机对照试验所进行的系统评价在理论和方法学上较完善及其论证强度较高,所以有关随机对照试验或评估治疗措施疗效的系统评价较多,而其他类型的系统评价如诊断试验、病因学研究、非随机对照试验等正在研究之中。

针对不同研究问题的系统评价,其基本方法和步骤相似,但在检索文献的策略、评价文献质量的方法、从原始文献收集的数据和统计分析方法等方面有差别。由于Cochrane的系统评价具有严格、系统的研究方法,且定期更新,而有关治疗性研究的系统评价最多、方法也较完善,本节将以评价治疗措施疗效的Cochrane系统评价为例,简述其基本方法和步骤。



## 一、确立题目、制订系统评价计划书

系统评价是为医疗保健措施的管理和应用提供决策的依据,特别适用于某些干预措施的利弊根据单个临床研究结果难以确定,或在临床应用过程中存在较大差异时。因此,系统评价的题目主要来源于临床医疗实践,涉及疾病防治方面不肯定、有争论的重要临床问题,以帮助临床医师进行医疗决策。例如,在高危人群中服用小剂量的阿司匹林能否预防心脑血管病的发生?抗凝剂治疗能否预防缺血性心脏病伴心房纤颤患者发生心脏事件等。

系统评价解决的问题很专一,涉及的研究对象、设计方案以及治疗措施需相似或相同。因此,在确立题目时,应围绕研究问题明确四个要素:①研究对象的类型:所患疾病类型及其诊断标准、研究人群的特征和场所;②研究的干预措施或进行比较的措施;③主要研究结果的类型,包括所有重要的结果及严重的不良反应;④研究的设计方案:如治疗性研究主要选择随机对照试验,病因或危险因素研究选择病例-对照研究和队列研究等。这些要素对指导查寻、筛选和评价各个临床研究,收集、分析数据及解释结果的应用价值十分重要,必须准确、清楚定义。

系统评价的题目确立后,需要制订计划书,内容包括系统评价的题目、背景资料、目的、检索文献的方法及策略、选择合格文献的标准、评价文献质量的方法、收集和分析数据的方法等。

## 二、检索文献

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

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

## 三、选择文献

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

文献资料的选择应分三步进行:①初筛:根据检索出的引文信息如题目、摘要筛除明显不合格的文献,对肯定或不能肯定的文献应查出全文再进行筛选;②阅读全文:对可能



合格的文献资料，应逐一阅读和分析，以确定是否合格；③与作者联系：一旦被排除的文献将不再录用，因此，如果文中提供的信息不全面而不能确定，或者有疑问和有分歧的文献应先纳入，通过与作者联系获得有关信息后再决定取舍或在以后的选择过程中进一步评价。

## 四、评价文献质量

评价文献的质量是指评估单个临床试验在设计、实施和分析过程中防止或减少系统误差（或偏倚）和随机误差的程度，以作为纳入原始文献的阈值、解释不同文献结果差异的原因、进行系统评价敏感性分析和定量分析（meta 分析）时给予文献不同权重值的依据。为此，对于入选的文献，需要应用临床流行病学/循证医学评价文献质量的原则和方法，进一步分析评价。文献的评价应包括三方面内容：①内在真实性（internal validity）：指单个研究结果接近真值的程度，即受各种偏倚因素如选择偏倚、实施偏倚、失访偏倚和测量偏倚的影响情况；②外在真实性（external validity 或称 generalizability）：指研究结果是否可以应用于研究对象以外的其他人群，即结果的实用价值与推广应用的条件，主要与研究对象的特征、研究措施的实施方法和结果的选择标准密切相关；③影响结果解释的因素：如治疗性试验中药物的剂量、剂型、用药途径和疗程以及依从性等因素。

在进行系统评价过程中，文献质量的评价强调对内在真实性的评估，即是否存在各种偏倚因素及其影响程度。偏倚主要来源于四个方面：①选择性偏倚（selection bias/allocation bias）：发生在选择和分配研究对象时，因随机方法的不完善造成组间基线不可比，可夸大或缩小干预措施的疗效。采用真正的随机方法并对随机分配方案进行完善的隐匿可避免这类偏倚的影响；②实施偏倚（performance bias）：发生在干预措施的实施过程中，指除比较的措施外，向试验组和对照组研究对象提供的其他措施不一样。避免的措施是标化治疗方案和对研究对象及实施研究措施者采用盲法；③随访偏倚（attrition bias）：指在试验的随访过程中，试验组或对对照组因退出、失访、违背治疗方案的人数或情况不一样造成的系统差异。尽量获得失访者的信息和对失访的人员采用恰当的统计学方法处理，如意向分析法（intention to treat analysis）可减少其影响；④测量偏倚（measurement bias/detection bias/ascertainment bias）：测量试验组和对对照组结果的方法不一致所造成的系统差异，特别是主观判断研究结果时。因此，采用统一、标化测量方法和对研究对象及结果测量者实施盲法可避免其影响。

Cochrane 协作网要求，根据影响研究质量的重要因素，推荐在评价时至少应包括以下几方面：是否采用真正的随机方法？随机分配方案是否完善隐匿？影响研究结果的重要因素的基线在组间是否可比？是否对研究对象、治疗方案实施者、研究结果测量者采用盲法？是否有研究对象失访、退出、违背治疗方案并在分析时作恰当处理（意向分析法）？当然，这些重要因素也仅仅限于治疗性研究质量评价。

为了避免选择文献和评价文献质量人员的偏倚，可以考虑一篇文章多人或盲法选择和评价，也可采用专业与非专业人员相结合的共同选择和评价的办法，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人的方法进行解决。此外，应进行预试验，以摸索经验，标化和统一选择、评价方法。

## 五、提取数据

根据制订的调查表和需要收集的内容，收录有关的数据资料，其中包括：①一般资料：如评价的题目、评价者的姓名、原始文献编号和来源、评价的日期等；②研究特征：如研究的合格性、研究对象的特征和研究地点、文献的设计方案和质量、研究措施的具体



内容和实施方法、有关偏倚防止措施、主要的试验结果等；③结果测量：如随访时间、失访和退出情况、分类资料应收集每组总人数及事件发生率、连续性资料应收集每组研究人数、均数和标准差或标准误等。

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

## 六、分析资料和报告结果

对收集的资料，可采用定性或定量的方法进行分析，以获得相应的结果。

### （一）定性分析（non-quantitative synthesis）

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

### （二）定量分析（quantitative synthesis）

定量分析包括三个方面：同质性检验（或异质性检验）、Meta-分析和敏感性分析。

1. 同质性检验（homogeneity test）指对不同原始研究之间结果的变异程度进行检验。如果检验显示各研究间结果差异有统计学意义，应解释其可能的原因并考虑进行结果合成是否恰当。确定各研究结果是否同质有两种方法：一是作图观察各研究结果的效应值和可信区间的重叠程度，如果可信区间差异太大、重叠程度小，则表示可能存在统计学异质性。另一种方法是进行同质性检验，如果同质性检验显示  $P \leq 0.1$ ，则表示有统计学异质性。

2. meta 分析 应根据资料的类型及评价目的选择效应量和统计分析方法。例如：对于分类变量，可选择比值比（odds ratio）、相对危险度（relative risk）、危险度差值（risk difference）和比对照组多减少一例事件发生需要治疗同类患者的人数（number needed to treat, NNT）等作为效应量表示合成结果。对于连续性变量，当结果测量采用同样度量衡单位时应选择加权均数差值（weighted mean difference），而当结果测量采用不同的度量衡单位，如疼痛评分在不同研究中采用不同的量表时，则应选择标化的均数差值（standardized mean difference）。进行 meta 分析合成结果时，可选择固定效应模型（fixed effect model）或随机效应模型（random effect model）。meta 分析的结果采用森林图（forest plot）表示。

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

## 七、解释系统评价的结果（讨论和结论）

解释系统评价必须基于研究的结果，内容应包括：

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

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

3. 对干预措施的利弊和费用进行卫生经济分析。



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

## 八、更新系统评价

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

总之，无论哪一类性质的文献综述，只要选题有意义、收集的文献真实可靠，评阅的方法无偏倚，所提供的信息能够如实地反映学术发展的新进展和新观念，具有良好的实用性，对临床科研、医疗及医学教育和决策无疑有着指导价值和参考意义。

(李 静 李 革)

## 第七章 临床医学研究设计的基本原则

临床医学研究是以病人及相应群体为研究对象，研究和论证某个或某些研究因素对疾病病因、诊断、治疗、预后等方面产生的效应或影响。与理化实验及动物实验不同，人体作为最高形式的生命体，除具有生物界普遍存在的个体差异外，还具有思维、心理及社会活动带来的差异；另外，就疾病本身而言，还存在病程、病情、诊断与治疗等因素带来的影响，增加了临床医学研究的复杂性。因此，临床医学研究应该遵循随机、对照、盲法等基本原则，尽量减少已知和未知影响因素的干扰，以保证研究结果的真实性、可靠性。

### 第一节 随机化原则

#### 一、概述

随机化 (randomization) 是采用特殊手段，使总体或样本中每个个体发生某事件的概率均等。在临床医学研究中，采用随机化的目的主要是为了在选取样本和将研究对象分组时，防止来自研究者与被研究者两个方面的主观因素的干扰，避免结果偏离真实值。随机化主要应用在抽样和分组两个环节，因此，又分为随机抽样与随机分组两种形式。

##### (一) 随机抽样 (random sampling)

由于临床医学研究的目标总体常常是无限总体，如某种疾病的所有患者或产生该疾病的人群，在研究中以全部总体为研究对象的可行性很小；另一方面，由于人力、财力、时间等因素的限制，也不可能把所有对象都纳入研究，因此，只能从总体中抽取一定的样本进行研究。随机抽样是在抽样过程中，采用随机化方法，使总体中所有对象都有同等的机会被抽取进入研究样本。随机抽样的目的是保证样本的代表性，避免发生选择偏倚。但在很多临床研究中，研究对象仅来源于医院某疾病的门诊及住院患者，数量不大，不需要通过随机抽样获取样本，而往往采用某段时间内的连续病例作为研究对象。

##### (二) 随机分组 (random allocation)

随机分组又叫随机分配，它是指在研究样本确定后，进一步采用随机的方法，使研究对象以同等的机会被分配进入试验组 (experimental group) 或对照组 (control group) 中。随机分组的主要目的是提高组间的均衡性，减少非研究因素的干扰。

#### 二、随机化方法

随机化原则是针对每个独立观察单位 (sampling unit) 进行随机抽样或随机分配。因此，在实施随机化方法之前，应根据研究目的及相应的设计方案确定研究的独立观察单位，即研究因素针对的具体对象。临床医学研究通常是以单个个体 (如某病患者) 为独立观察单位。但是，有时因研究目的或研究因素的特殊性，不能以单个个体为独立观察单位，而需要采用某一特定组群 (cluster)，如一对夫妻、一个家庭、一个班组，甚至一个社区作为独立观察单位。例如，研究生殖健康教育与育龄夫妻避孕效果的关系，就常以一对夫妻作为一个独立观察单位。

##### (一) 简单随机法 (simple randomization)

简单随机法主要指采用抛硬币、抽签、掷骰子、随机数字表，甚至利用计算器、计算





机产生的随机数字进行抽样或分组的方法。

1. 抛硬币、抽签等方法主要应用在研究涉及的例数较少时。例如欲在 10 个观察单位中抽取两个作为研究对象,可制作 10 个签,其中 8 个为“无”,2 个为“有”,抽中“有”签者作为研究对象。在随机分组时,可事先规定硬币“正”、“反”两面中的一面为试验组、另一面则为对照组,然后取一硬币随意抛掷并让研究对象依次与硬币落下所得的“正”、“反”结果相对应,从而实现简单随机分组。但是,在样本不大时,硬币抛掷出的“正”、“反”面次数可能出现较大差异,使得不同试验组间观察对象数也存在较大差异,需进一步调整。但临床医学研究的观察对象往往较多,使用以上几种方法进行抽样或分组,操作过程较为复杂,实施起来很不方便。

2. 随机数字表是实现随机化的一种较为简单、实用的方法。数理统计学家根据概率论原理编制了随机数字表(表 7-1),表中排列的数字无论以行、列或斜向顺序都呈随机状态。利用随机数字表,既可以进行随机抽样,也可以进行随机分组。

表 7-1 随机数字表

编号	1~10	11~20	21~30	31~40	41~50
1	22 17 68 65 84	68 95 23 92 35	87 02 22 57 51	61 09 43 95 06	58 34 82 03 47
2	19 36 27 59 46	13 79 93 37 55	39 77 32 77 09	85 52 05 30 62	47 83 51 62 74
3	16 77 23 02 77	09 61 87 25 21	28 06 24 25 93	16 71 13 59 78	23 05 47 47 25
4	78 43 76 71 61	20 44 90 32 64	97 67 63 99 61	46 38 03 93 22	69 81 21 99 21
5	03 28 28 26 08	73 37 32 04 05	69 30 16 09 05	88 69 58 28 99	35 07 44 75 47
6	93 22 53 64 39	07 10 63 76 35	87 03 04 79 88	08 13 13 85 51	55 34 57 72 69
7	78 76 58 54 74	92 38 70 96 92	52 06 79 79 45	82 63 18 27 44	69 66 92 19 09
8	23 68 35 26 00	99 53 93 61 28	52 70 05 48 34	56 65 05 61 86	90 92 10 70 80
9	15 39 25 70 99	93 86 52 77 65	15 33 59 05 28	22 87 26 07 47	86 96 98 29 06
10	58 71 96 30 24	18 46 23 34 27	85 13 99 24 44	49 18 09 79 49	74 16 32 23 02
11	57 35 27 33 72	24 53 63 94 09	41 10 76 47 91	44 04 95 49 66	39 60 04 59 81
12	48 50 86 54 48	22 06 34 72 52	82 21 15 65 20	33 29 94 71 11	15 91 29 12 03
13	61 96 48 95 03	07 16 39 33 66	98 56 10 56 79	77 21 30 27 12	90 49 22 23 62
14	36 93 89 41 26	29 70 83 63 51	99 74 20 52 36	87 09 41 15 09	98 60 16 03 03
15	18 87 00 42 31	57 90 12 02 07	23 47 37 17 31	54 08 01 88 63	39 41 88 92 10
16	88 56 53 27 59	33 35 72 67 47	77 34 55 45 70	08 18 27 38 90	16 95 86 70 75
17	09 72 95 84 29	49 41 31 06 70	42 38 06 45 18	64 84 73 31 65	52 53 37 97 15
18	12 96 88 17 31	65 19 69 02 83	60 75 86 90 68	24 64 19 35 51	56 61 87 39 12
19	85 94 57 24 16	92 09 84 38 76	22 00 27 69 85	29 81 94 78 70	21 94 47 90 12
20	38 64 43 59 98	98 77 87 68 07	91 51 67 62 44	40 98 05 93 78	23 32 65 41 18
21	53 44 09 42 72	00 41 86 79 79	68 47 22 00 20	35 55 31 51 51	00 83 63 22 55
22	40 76 66 26 84	57 99 99 90 37	36 63 32 08 58	37 40 13 68 97	87 64 81 07 83
23	02 17 79 18 05	12 59 52 57 02	22 07 09 47 03	28 14 11 30 79	20 69 22 40 98



续表

编号	1~10					11~20					21~30					31~40					41~50				
24	95	17	82	06	53	31	51	10	96	46	92	06	88	07	77	56	11	50	81	69	40	23	72	51	39
25	35	76	22	42	92	96	11	83	44	80	34	68	35	48	77	33	42	40	90	60	73	96	53	97	86
26	26	29	13	56	41	85	47	04	66	08	34	72	57	59	13	82	43	80	46	15	38	26	61	70	04
27	77	80	20	75	82	72	82	32	99	90	63	95	73	76	63	89	73	44	99	05	48	67	26	43	18
28	46	40	66	44	52	91	36	74	43	53	30	82	13	54	00	78	45	63	98	35	55	03	36	67	68
29	37	56	08	18	09	77	53	84	46	47	31	91	18	95	58	24	16	74	11	53	44	10	13	85	57
30	61	65	61	68	66	37	27	47	39	19	84	83	70	07	48	53	21	40	06	71	95	06	79	88	54
31	93	43	69	64	07	34	18	04	52	35	56	27	09	24	86	61	85	53	83	45	19	90	70	99	00
32	21	96	60	12	99	11	20	99	45	18	48	13	93	55	34	18	37	79	49	90	65	97	38	20	46
33	95	20	47	97	97	27	37	83	28	71	00	06	41	41	74	45	89	09	39	84	51	67	11	52	49
34	97	86	21	78	73	10	65	81	92	59	58	76	17	14	97	04	76	62	16	17	17	95	70	45	80
35	69	92	06	34	13	59	71	74	17	32	27	55	10	24	19	23	71	82	13	74	63	52	52	01	41
36	04	31	17	21	56	38	73	99	19	87	26	72	39	27	67	53	77	57	68	93	60	61	97	22	61
37	61	06	98	03	91	87	14	77	43	96	43	00	65	98	50	45	60	33	01	07	98	99	46	50	47
38	85	93	85	86	88	72	87	08	62	40	16	06	10	89	20	23	21	34	74	97	76	38	03	29	63
39	21	74	32	47	45	73	96	07	64	52	09	65	90	77	47	25	76	16	19	33	53	05	70	53	30
40	15	69	53	82	80	79	96	23	53	10	65	39	07	16	29	45	33	02	43	70	02	87	40	41	45
41	02	89	08	04	49	20	21	14	68	86	87	63	93	95	17	11	29	01	95	80	35	14	97	35	33
42	87	18	15	89	79	85	43	01	72	73	08	61	74	51	69	89	74	39	82	15	94	51	33	41	67
43	98	83	71	94	22	59	97	50	99	52	08	52	85	08	40	87	80	61	65	31	91	51	80	32	44
44	10	08	58	21	66	72	68	49	29	31	89	85	84	46	06	59	73	19	85	23	65	09	29	75	63
45	47	90	56	10	08	88	02	84	27	83	42	29	72	23	19	66	56	45	65	79	20	71	53	20	25
46	22	85	61	68	90	49	64	92	85	44	16	40	12	89	88	50	14	49	81	06	01	82	77	45	12
47	67	80	43	79	33	12	83	11	41	16	25	58	19	68	70	77	02	54	00	52	53	43	37	15	26
48	27	62	50	96	72	79	44	61	40	15	14	53	40	65	39	27	31	58	50	28	11	39	03	34	25
49	33	78	80	87	15	38	30	06	38	21	14	47	47	07	26	54	96	87	53	32	40	36	40	96	76
50	13	13	12	66	99	47	24	49	57	74	32	25	43	62	17	10	97	11	69	84	99	63	22	32	98

摘自 Fisher R A and Yates F. Statistical Tables for Biological, Agricultural and Medical Research, P128. Oliver and Boyd, 1957 [1]

(1) 随机抽样：例如，欲在某医院门诊随访的 250 名高血压患者中随机抽取 20 名作为一种降压新药的试验对象。先将 250 名高血压患者从 1~250 顺序编号，然后从随机数字表任何一行的任何一列开始，顺序产生 20 个 00~99 之间的随机数，每个数字乘以 250 取前 3 位数，超过 250 号的，取前 2 位数，重复的数弃去重抽，最后，以对应编号的患者组成参加本次试验的样本。但应注意，一旦确定行、列或斜向顺序后，中途不能任意更改方向。



(2) 随机分组：例如，欲将合乎试验要求的 20 名患者，随机分配为试验组 (T)，对照组 (C) 2 组。首先将患者从 1 至 20 编号，然后从随机数字表的任何一行、任何一列开始，顺序产生 20 个 00~99 之间的随机数。按事先规定，对应的随机数为“奇数”者进入试验组，为“偶数”者进入对照组。如从表 7-1 第 9 行，8 列开始，按从左到右，从上到下顺序获取 20 个随机数字并依次列入表 7-2。

表 7-2 20 例患者查随机数字表分配情况

患者编号	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
随机表数	52	77	65	15	33	59	05	28	22	87	26	07	47	86	96	98	29	06	58	71
归组	C	T	T	T	T	T	T	C	C	T	C	T	T	C	C	C	T	C	C	T

结果显示两组例数不等，则应该进一步遵循一定的随机方法，对两组的例数进行调整。

3. 利用计算机或计算器产生的随机数字实现随机化，是大样本研究中最常用的一种方法。大多数具有编程功能或数字计算功能的计算机软件都可以通过随机数发生器产生的随机数。计算机产生的随机数是取值在 0~1 之间（不包括 0 和 1 本身）均匀分布的随机变量。

例如，欲产生 5 个 0~999 的随机数，一般需事先指定一个种子数 (seed)，相当于在随机数字表上指定行和列。可在 SAS (6.12 版本) 中指定种子数为 8888，产生 5 个随机数：0.94732、0.14850、0.63843、0.53516、0.20371。因计算机产生的随机数是在 0~1 之间，要得到 0~999 的随机数，只需将每个数乘以 1000 并取整，据此，以上 5 个随机数转换为：947、148、638、535、959。同一软件用相同种子数所产生的随机数是一样的（统计学上称为重现性），而不同软件所得结果不一样。用函数计算器的“RAN”键，也可以产生取值在 0~1 之间的随机变量。每按一次“RAN”键，可产生一组随机数字。

得到随机数字序列后，可按照与应用随机数字表相同的方法进行随机抽样或随机分组。

必须指出，交替将受试对象分到试验组和对照组的分配方法 (TCTCTC……) 不属于简单随机分配法。因为，当研究者预先知道下一位患者将分配到哪一组时，可能在主观上对患者的病程早晚、病情轻重等进行一定的取舍，导致研究对象选择上的偏倚。

## (二) 分层随机法 (stratified randomization)

分层随机法是将研究对象按某一特征进行分组 (层)，然后在各层中采用简单随机的方法抽取研究对象组成样本，实现分层随机抽样 (stratified sampling)；或在各层中按简单随机分配的方法，分出试验对象与对照组对象，最后将各层试验对象合在一起作为试验组、将各层对照对象合在一起作为对照组，实现分层随机分组 (stratified allocating)。分层随机抽样可以保证各“层”都有一定研究对象进入样本，提高了样本的代表性；分层随机分组可以保证各“层”都有对象进入试验组或对照组，提高了试验组间的均衡性。

在分层随机抽样中，往往是以对观察值变异影响较大的因素作为分层因素 (stratifying factor)；在分层随机分组中，主要以研究对象中某些可能产生混杂作用的特征作为分层因素，如研究对象的重要临床特征或预后因素 (包括年龄、性别、病情、有无并发症等)。在临床医学研究中，可根据以下原则考虑分层因素：①选择所研究疾病或其并发症的危险因素；②选择对所研究疾病预后有明显影响的因素；③遵循最小化原则，将分层因素控制到最低限度，否则将致分层过多，造成组内研究对象过度分散的不利局面。

例如，进行慢性心房纤颤复律后抗心律失常药物维持治疗的随机对照研究。考虑到慢性



心房纤颤患者的预后与病因、心脏大小及心房纤颤病程长短有密切关系，因此，宜针对上述三个因素（a. 病因：瓣膜性心脏病与非瓣膜性心脏病；b. 心脏大小：心胸比例 $\geq 0.50$ 和 $< 0.50$ ；c. 病程：心房纤颤病程 $\geq 6$ 个月及 $< 6$ 个月）进行分层随机分组（图 7-1）。

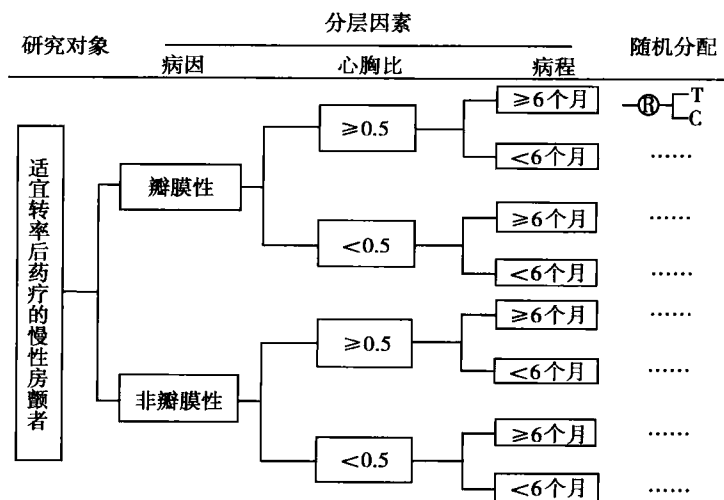


图 7-1 分层随机分组

经图 7-1 所示的分层随机分配过程，可以使病因、心脏大小、病程这三个对治疗效果和预后有较大影响的因素在试验组（T）与对照组（C）间保持平衡。

在临床实验性研究中，当样本量达数百甚至上千时，不需要进行分层随机分组。一方面，大样本经过简单随机分配后，不同试验组间的基线状况往往具有可比性；另一方面，需要有完整的研究对象名单后才能进行分层随机分组，对临床大样本而言，可操作性不大。如果在资料分析时发现试验组之间均衡性不好，还可以通过分层分析进行统计学处理。分层随机与分层分析是两个不同的概念和方法。分层随机是在设计阶段，为避免一些可能对研究结果产生较大影响的非研究因素在试验组之间分配不均而采取的一种控制措施，其目的是增强组间的可比性。而分层分析是在试验结束后，针对某些可能对研究结果产生影响的因素而采用的一种统计分析方法。如果研究的样本量较小，对资料分层后有些层组可能出现对象数为“零”的情况，则不适宜进行分层分析。

### （三）区组随机法（block randomization）

区组随机法主要用于随机分组。区组随机分组是先将研究对象分为不同区组，然后再对每一区组内的研究对象用简单随机法进行分配。在临床医学研究中，主要是按固定的研究对象数来划分区组。每一区组的研究对象数一般按分组数的倍数来确定。如研究分为试验组、对照组两组，则区组例数可选 2、4、6、8 等，但区组例数越大，研究对象在分配时的排列组合越复杂。例如某研究分为试验组（T）、对照组（C）2 组，确定区组例数为 4。首先，研究对象按进入试验的先后顺序，每 4 个人一组，然后再对每一区组的 4 个研究对象分别根据随机数字表进行随机分组。区组随机分组的特点之一就是分组后各组人数相等。而对一个区组的 4 个对象按随机数字表进行随机分配时很可能出现 2 组例数不等的情况，必须进行适当调整。

区组随机分组有两方面优点：①保证各组人数相等；②便于逐渐累积临床病例。如上例，每积累 4 例研究对象即可进行分组并开始试验，不需要把所有样本全部收集齐后再来分组、展开试验。



#### (四) 系统随机抽样法 (systematic sampling)

系统随机抽样简称系统抽样或机械抽样 (mechanical sampling), 是随机抽样方法之一。系统抽样即先将总体的观察单位按某一特征的顺序 (如按入院先后顺序) 编号, 再根据抽样比例将其分为若干部分, 先从第一部分随机抽取第一个观察单位, 然后按一固定间隔在第二、第三……等各部分抽取观察单位组成样本。例如, 欲从 2000 个观察单位中抽取 100 个组成样本, 即抽样比例为 5% (抽样间隔为 1/20), 可先从第 1~20 (第一部分) 之间随机抽出一个观察单位, 如为 12 号, 此后按每隔 20 抽取一个单位, 即 32、52、72……至 1992 号组成样本。

系统抽样的优点是简单易行, 适用于大样本的流行病学调查, 样本的观察单位在总体中分布均匀, 抽样代表性较好。缺点是如果总体各单元的排列顺序存在一定周期性, 以这种方法进行抽样则可能出现较大的偏倚。例如, 某工厂工作证编号中末位数代表车间, 用该法按工作证顺序抽样则可能抽到的全部是同一车间的对象, 使样本不能代表全厂的情况。

#### (五) 整群随机法 (cluster randomization)

整群随机法是以现成的群体 (社区、街道、乡、村、医院、病房等) 而不是个体为单位, 进行抽样或分组。在整群随机抽样 (简称整群抽样) 中, 抽到的群体中的所有观察单位, 都将作为研究样本。例如, 采用整群抽样方法, 从某地区的 20 所小学中随机抽取 2 个学校, 并对这 2 个小学的全部学生进行视力检查, 以了解该地区小学生近视率。同样, 在整群随机分组中, 被分到试验组的群体中的每个观察单位, 都作为试验对象, 被分到对照组的群体中的每个观察单位, 都作为对照对象。

采用整群随机法要求群间的变异越小越好, 否则将影响样本的代表性或组间的可比性。一般情况下, 用相同的样本含量, 整群抽样的抽样误差最大, 整群随机分组组间的可比性最小。但是, 整群随机法具有节约人力、物力、方便、容易实施等优点, 在实际工作中可行性较好, 适用于大规模研究。

#### (六) 多级抽样法 (multi-stage sampling)

多级抽样是一种从大到小多个级别进行的抽样方法。首先从总体中随机抽取范围较大的单元, 称为一级抽样单元 (例如省、市), 再从抽中的一级单元随机抽取范围较小的二级单元 (如区、街道), 若抽样到此为止称为二级抽样, 若再继续往小范围抽样, 则称为多级抽样。

在一些大规模研究 (全国高血压抽样调查、全国糖尿病患病率调查等) 项目中, 多级抽样成为唯一实用的抽样方法。在具体实施的时候, 多级抽样常常与上述各种基本抽样方法结合使用。

#### (七) 半随机化法 (quasi-randomization)

不采用随机数字产生序列号, 而是根据被纳入研究对象的出生日期、住院号或门诊号等分为单、双数, 分别将单数者分为 1 组, 双数者分为另一组, 这种方法称半随机化法。

当进行大样本研究时, 如社区人群的试验, 利用半随机化分配受试者既简便、实用, 同时也能较好地实现组间均衡。但在临床试验中, 研究对象往往较少, 半随机化很难达到组间均衡, 因而该方法的使用受到限制。

### 三、随机化分配的优缺点

#### (一) 优点

1. 随机分配保证了研究组之间的均衡、可比。特别是采用分层随机分配, 可使一些



影响疾病进程的重要临床特征, 以及一些已知或未知的混杂因素在研究组间保持均衡, 有效避免了这些因素对研究结果的干扰。

2. 随机分配结合盲法, 可以使研究结果更加客观, 提高了研究结果的准确性。

## (二) 缺点

1. 在随机分配过程中, 为了保证组间的均衡、可比, 往往使研究对象局限于符合条件的患者, 因而可能导致研究对象代表性的降低, 使研究结果的推广范围受到局限。

2. 随机化分配往往需要累积一定数量的研究对象后才能实施, 增加了临床研究的难度。

3. 随机分配可能违背研究对象的个人意愿, 可能存在医德问题。

## 四、随机化的应用

在利用随机化方法时, 应注意以下问题:

1. 在利用随机化方法进行研究对象分组时, 必须对随机化方案进行保密, 该过程称为随机化分配的隐匿 (concealment)。从某种意义上讲, 随机化分配的隐匿比采用随机的方法产生不可预测的随机分配序列更加重要。如果随机分配序列产生者与试验执行者 (分配受试对象入组者) 为同一人, 试验执行者就可以预先知道下一位受试对象将被分入哪一组。在这种情况下, 很难避免研究人员因主观因素造成的对受试对象的刻意选择, 甚至改变分配方案, 导致研究结果产生偏倚。因此, 随机化分配方案隐匿的原则就是让分配受试对象入组的试验执行者事先不知道随机分配方案的具体内容, 不知道下一位入选的对象将进入哪一组。实现随机化分配隐匿的方法有中心电话随机系统、药房控制随机分配方案、利用编号或编码的药品容器、按顺序编码、密封入不透光的信封等, 研究者可根据具体情况选择应用。

2. 为了保证试验的可靠性, 研究中所用随机化方法、随机数等均应详细记录。如果是用计算机产生的随机数, 需要报告产生随机数的程序、种子数等; 如果是从随机数字表获得随机数, 必须说明所用随机数表、起始页、起始行、起始列及获取方法。特别是在新药的临床试验中, 随机数必须具有重现性, 产生随机数的参数及程序应与盲底一起封存。

## 第二节 对照的原则

### 一、概述

所谓“对照 (control)”, 指的是在临床研究过程中, 设立条件类似、诊断一致的可供相互比较的组别。

“有比较才有鉴别”, 这是辩证唯物主义认识事物的基本法则。临床医学研究主要解决某个或某些研究因素对疾病病因、诊断、治疗、预后等方面产生的效应的有无、大小或安全性的高低等问题, 但临床事件的结局有着很大的不确定性, 因此, 需要通过对照比较才能得出可靠结论。

### 二、设置对照的意义

在临床医学研究中, 除了干预措施 (处理因素) 的作用以外, 还有很多因素可能影响研究对象的临床结局:

(1) 不能预知的结局 (unpredictable outcome): 由于个体的生物学变异和社会、心理



因素的影响,导致临床上患相同疾病的个体,其临床特征(临床表现、病情轻重、预后等)存在较大差异。例如感冒、细菌性痢疾、病毒性肝炎等许多传染性疾病都有明显的自愈倾向,而在高血压、糖尿病、胃溃疡等慢性病的病程中,也存在病情的自行缓解与加重现象,即使是恶性肿瘤,其生存时间长短、并发症的发生等情况也各不相同。

(2) 霍桑效应(Hawthorne effect):指人们因成为研究中特别感兴趣和受关注的对象而产生的一种正向心理——生理效应,这种效应与他们接受的干预措施的特异性作用无关。霍桑效应一般是一种正向效应,即与干预措施预期效果同向的效应。但有时也可能因为厌恶某医生或不信任某医疗单位而产生负向效应。

(3) 安慰剂效应(placebo effect):某些研究对象,即使仅使用了安慰剂,也可能表现出病情好转等不该有的效果,其实质是某些研究对象由于依赖医药而产生的一种正向心理效应。

(4) 向均数回归(regression to the mean)现象:指一些极端的临床症状和体征有向均数接近的倾向。如血压水平处于高限5%的人,即使不给任何处理,经过一段时间后,其血压值也可能有所降低(向均数靠近)。

(5) 潜在未知因素的影响:由于知识的局限性,很可能还存在一些影响干预效应的因素,但目前尚未被人们所认识。

因此,在临床研究中设立合理的对照非常必要,其主要目的是消除非处理因素的干扰,鉴别试验性与非试验性效应,减少或消除实验误差。

### 三、对照的种类

#### (一) 按照研究的设计方案分类

1. 同期随机对照(concurrent randomized control) 按严格规定的随机化方法将研究对象同期分配到试验组和对照组,以此方案设置的对照类型称为同期随机对照。

同期随机对照的优点主要有三方面:①由于采用了随机化分组方法,可以较好地保证试验各组之间的均衡可比,有效避免了潜在未知因素对试验结果的影响;②设置同期对照,可以同时对各试验组进行观察,有效避免了因试验先后顺序对结果的影响,使研究结果更有说服力;③由于多数统计方法都是建立在随机样本的基础之上,采用本设计类型更有利于资料的统计分析。

同期随机对照的缺点是:①需要有一半对象充当对照,因此所需样本量较大;②在有些情况下可能涉及医德方面的问题。

2. 自身对照(self control) 将一组受试对象分为前后两个阶段,分别施加不同的干预措施,然后比较两个阶段的两种处理效应的差异。一般在前一阶段结束后应有一段时间间隔,称为洗脱期。设置洗脱期的目的是为了避开前一阶段的处理效应对后一阶段的处理效应产生影响。

自身对照主要适用于慢性反复发作性疾病的治疗性研究,这一类疾病病程长,病情变化不大,允许进行分阶段处理与观察。采用自身对照可以消除个体差异,减少一半样本量,并且保证每个受试对象都接受有效治疗的处理;缺点是难以保证两个阶段的病情完全一致,可能存在处理先后对结果的影响。

3. 交叉对照(cross-over control) 将两组受试对象分两个阶段进行试验。如第一组第一阶段试验A措施,间隔一段洗脱期后再试验B措施;第二组第一阶段试验B措施,间隔一段洗脱期后再试验A措施,然后对比A、B两种措施的效果。

交叉对照同样可以保证每个受试对象都接受有效治疗的处理,消除个体差异,提高统计效率,并且可以避免因试验先后顺序对结果的影响;但是,该方法的应用也主要限于慢



性复发性疾病。

4. 配对对照 (matching control) 以可能对研究结果产生影响的混杂因素 (如年龄、性别、病情等) 为配比条件, 为每一个试验对象选配一个以上的对照。通常采用 1:1 或 1:2 配对。

配对对照的优点是可以保证比较组之间在这些主要影响因素上的均衡性, 避免已知混杂因素对结果的影响。

5. 非随机同期对照 (non-randomized concurrent control) 有同期对照, 但试验组与对照组未严格按随机化原则进行分组。例如在多中心临床试验中, 以一个医院的对象作为试验组, 另一个医院的对象作为对照组进行比较。

非随机同期对照的设置简便易行, 易为医生和患者接受; 缺点是由于非随机分配, 可能因选择偏倚导致两组基线情况不一致, 可比性较差。

6. 历史对照 (historical control) 试验仅设一组接受新干预措施的受试对象, 并将该组的处理效果与过去采用的旧干预措施的效果进行比较。因此, 历史对照属于非随机、非同期对照。

历史对照的资料可以通过查阅医学文献和医院病历记录获得。优点是节省时间、经费, 不存在医德问题; 但通常既往的资料中的研究对象、条件、环境等都很难与本次试验保持一致, 可比性差, 一般不宜采用。在特殊情况下, 如对一些预后极差的疾病, 采用历史对照还是有一定说服力。如恶性滋养层细胞瘤在 20 世纪 50 年代单用手术治疗, 预后极差。随着化疗和放疗的进展, 联合化疗的有效率达 70%~90%, 对这两个历史阶段中, 同一疾病的不同治疗效果进行历史对照, 证明现在的治疗效果明显提高, 预后显著改善, 为人们所公认。

## (二) 按照对照组的处理措施分类

在以上各种对照设置方案中, 试验组的处理一般是施加某种需要评价的干预措施 (处理因素), 而对照组也需要采用相应的对照措施 (对照因素)。目前使用较多的对照措施有有效对照、空白对照、安慰剂对照等。

1. 有效对照 即以目前临床公认的有效处理方法 (如治疗某病常规、有效的治疗方法) 施加给对照, 然后与试验组处理措施 (新治疗方法) 的效果相比较。这类试验通常采用随机双盲设计, 受试者随机地分配至试验组和有效对照组。是临床治疗性研究中最常用的对照方法。

在有效对照中, 施加给对照组的处理措施效果稳定, 试验期间能保证对照组成员受到合理治疗, 较少引起伦理学及医德方面的问题。

2. 空白对照 (blank control) 对照组在试验期间不给任何处理, 仅对他们进行观察、记录结果, 并将其与试验组的效应进行比较。

在空白对照中, 对照组成员在试验阶段未得到任何治疗, 可能造成不良后果, 应用时要特别谨慎。空白对照不能用于急、重或有较重器质性病变的病人, 仅用于病情轻、稳定, 即使不给任何治疗也不会导致病情恶化的疾病 (如近视), 否则将产生伦理和医德方面的问题。

3. 安慰剂对照 (placebo control) 安慰剂是不具有真正治疗或致病效应的制剂。将安慰剂的效应与试验措施的效应进行比较称为安慰剂对照。常用的安慰剂为没有药理活性, 但其色、形、味均与试验的新药完全相同的制剂, 如淀粉、乳糖、维生素等。若从注射途径给药, 多采用生理盐水作为安慰剂。

虽然安慰剂对照本质上也是一种空白对照, 但其效应不同。因为, 使用安慰剂后可产生安慰剂效应。采用安慰剂对照的主要目的是为消除主观因素的影响, 便于盲法的实施。





安慰剂对照与空白对照一样，主要用于病情轻、稳定，或目前没有有效治疗方法的疾病。同时，也应注意伦理学和医德方面的问题。

### 第三节 盲法试验

#### 一、概述

盲法试验 (blind trial) 是指临床研究过程中，指标的观测、数据的收集和结论的判断均在不知道研究对象分配的组别、不知接受的是试验措施还是对照措施的前提下进行。此外，盲法还用于对研究资料的分析与报告。

临床医学研究要求对每一个研究对象的诊断、疗效或预后等作出准确、可靠的测量和评价，否则将出现测量偏倚 (measurement bias)，导致研究结果偏离真实值。在临床试验中，测量偏倚既可以产生于研究人员方面，也产生于研究对象方面。如果研究对象知道自己分在对照组，则可能因主观因素的影响，反映出较差的疗效或预后，甚至会对治疗失去信心而中途退出；而分入治疗组的受试对象可能因感觉受益而较多反映有利的结果。另一方面，如果研究人员知道受试对象的分组情况，则可能因希望研究得出阳性结论而有意或无意识地暗示受试对象，诱导治疗组对象多报告有利结果，对照组对象多报告不利结果。进行盲法试验的主要目的就是为了解决可能来自研究者或受试者的主观因素所导致的偏倚。导致研究结果的失真性。

#### 二、盲法的分类

盲法试验的基本原则就是让受试对象和/或研究人员不知道哪些人接受的是新试验措施，哪些人接受的是对照措施。在一项临床研究中，往往涉及研究对象、干预措施执行者、结果测量者、统计分析者等，根据“盲”的对象不同，一般可将盲法分为单盲、双盲、三盲 3 种类型。

##### (一) 单盲

1. 定义 对于研究对象的分组及所施加的处理因素 (如选用药物) 情况，只有研究者知道，而受试对象不知道，叫做“单盲” (single-blind)。所以，“单盲”主要是盲研究对象。另外，对检验人员、病理学检查人员等隐瞒患者的诊断与治疗情况，便于他们客观报告实验结果，也属于单盲。

##### 2. 优点

(1) 方法简单，容易进行。

(2) 干预措施执行者和结果测量者知道受试对象的分组情况，便于更好地观察、掌握病情。一旦发生病情变化，可以及时、恰当调整处理方案或采取其他干预措施，使受试对象在临床试验过程中的安全更加有保证。

(3) 减少了因受试对象主观因素对研究结果的影响。

3. 缺点 单盲不能避免研究方主观因素造成的影响。主管医生可能通过许多方法去影响患者的疗效，比如，医生对接受新疗法的患者观察特别仔细，护士对新疗法组患者更加关心和热情，这些都可能影响或暗示受试对象产生不同的反应。

##### (二) 双盲

1. 定义 受试对象和试验执行者 (干预措施执行者及结果测量者) 双方均不知道分组情况，不知道受试者接受的是哪一种干预措施，称为“双盲” (double-blind)。

2. 实施方法及注意事项



(1) 双盲往往用于临床随机对照试验之中,如配合隐匿措施,则更能增强研究的科学性。

(2) 双盲试验要有严格的管理制度和办法,组织严密,操作规范。在双盲试验过程中,需要有“局外”的管理、监督者,他们不直接参加临床研究的观测和数据收集,仅参与研究设计、药物编码的控制与保密、资料的保管、分析等,而受试对象和研究的执行者只知道每个试验对象的药物编号,如 A 和 B。待试验结束并完成数据分析后,方能“揭盲”。

(3) 双盲试验要求各组的药物在外观的形状、大小、颜色,给药的途径、办法、次数上保持一致,若有疏漏,则可能导致泄密,失去“盲法”的作用。药物临床试验时,若试验药与对照药的剂型不一样,可采取“双盲双模拟”办法:试验药+与对照药剂型一样的安慰剂;对照药+与试验药剂型一样的安慰剂,并分别编上密码。若制剂的效应维持时间不同,也可以采用“双盲双模拟”方案。

(4) 在盲法试验过程中,一旦某一受试对象的病情突然发生变化,如出现严重的副作用、治疗无效或病情加重等,不应该为追求资料的完整性而继续试验,必须立即停止试验,予以“破盲”,并同时使用必要的治疗措施。因此,在盲法试验进行之前,需事先制定出让主管医师明确的停药或换指针,以免给受试者带来不良后果。同时,“破盲”应局限在较小范围内进行,尽量减少由于“破盲”对双盲实施的影响。

### 3. 优缺点

(1) 优点:是临床试验最常采用的一种盲法形式,可以有效避免受试对象和试验执行者主观的偏倚因素对试验结果的影响。

(2) 缺点:①在管理上缺乏灵活性;②有特殊副作用的药物容易被破盲;③双盲试验不适用于危重患者。

## (三) 三盲

1. 定义 三盲(triple-blind)是指受试对象、试验执行者和资料分析与报告者三方均不知道受试者接受的是哪一种干预措施,全部采用编号密封。它可以避免双盲法在资料分析阶段的测量偏倚。

### 2. 优缺点

(1) 可以使偏倚减到最小的程度,使受试者的效应反应和评价结果更符合客观情况。

(2) 设计复杂,执行难度较大,常因医德、沾染、补偿、失密等问题而难以坚持。

(3) 当受试对象的病情突然发生变化时,难以及时掌握治疗情况,造成处理延误。

所以,尽管三盲试验是减少偏倚最有效的方法,但在实际工作中使用并不普遍。在许多临床研究中,医师既是试验设计者与观察者,也是资料分析和结果评价者,很难真正做到三盲。

## 三、非盲法评定

### (一) 定义

非盲法评定,又称为开放试验(open trial),即受试对象和研究者均知道试验组和对照组的分组情况,以及所给予的干预措施。

并非所有临床医学研究都适用盲法,例如,比较手术疗法和保守疗法治疗某病的效果,探讨功能训练或针灸疗法的疗效,评定生活习惯(吸烟、运动、饮食)对疾病的影响等。在这种情况下,盲法试验难以实施,只能采用非盲法评定;另外,如果临床试验的效应指标是明确的硬指标(如存活或死亡),很少受主观因素影响,采用非盲法评价也可以获得比较满意的结果。



## (二) 适用范围

1. 在一次研究中,需要对多种因素进行分析。如不仅希望观察药物的疗效,同时也要评价影响疗效或发病的有关因素。

2. 研究者希望更安全、更周到地作出决策(例如患者是否需要继续治疗,药物是否需要增减,是否需要配合其他药物等),使医疗决策更灵活。

## (三) 优缺点

1. 易于设计和实施,研究者了解分组,便于对受试对象及时作出处理。

2. 最主要的缺点是容易产生测量偏倚。

# 第四节 均衡性原则

## 一、概述

通过比较具有不同特征的人群在疾病发生、发展过程中的差异,来研究这些特征与疾病间的关系,是流行病学研究问题、解决问题并作出推论的逻辑学基础。但是,为了保证研究结果的准确性、可靠性,要求比较组之间要均衡可比,即必须符合均衡性原则。

所谓均衡性是指,比较组之间除了处理措施(研究因素)不同以外,其他可能影响研究结局的因素(非研究因素)处于相同或相近水平。也就是说,试验组与对照组之间比较的基础是具有相同的背景(background),只有在这种情况下,才能排除其他伴随因素的混杂作用,保证对比结果的准确与可靠。如果试验组与对照组之间均衡性良好,称为“同质”(homogeneity),或者叫“齐同可比”(control at the same background)。在临床研究中,影响研究结果的背景情况又叫做基线(baseline),它主要是指研究措施执行之前,被研究对象的临床特征,包括病史、临床特征及实验室检查等资料。根据研究具体的情况,基线资料可以是测定的自然数值,也可以是按不同因素分类以后的等级数据。如病情的轻、中、重,病程的早、中、晚,按月均收入划分的不同经济等级等。

## 二、资料的均衡性分析

在临床研究中,为了保证比较组之间的均衡性,必须在研究措施执行之前对各组间可能影响研究结果的重要基线资料进行均衡性分析。

例如:一项关于十二指肠溃疡治疗的研究(中华内科杂志,1991,30:226),将研究对象随机分为治疗组(94例)和对照组(106例)。研究者在分组后对两组对象的主要临床特征及溃疡的状况均进行了基线比较(表7-3、表7-4)。

表 7-3 两组患者一般情况比较

	治疗组 (n=94)		对照组 (n=106)	
	n	(%)	n	(%)
性别				
男	73	77.7	79	74.5
女	21	22.3	27	25.5
年龄(岁)				
18~30	18	19.1	27	25.5
31~40	35	37.2	38	35.8
41~50	19	20.2	20	18.9



续表

	治疗组 (n=94)		对照组 (n=106)	
	n	(%)	n	(%)
51~60	15	16.0	13	12.3
61~70	6	6.4	6	5.7
>70	1	1.1	2	1.9
嗜好				
吸烟	49	52.1	69	65.1
饮酒	9	9.6	4	3.8
饮茶	42	44.7	52	49.1

表 7-4 两组患者溃疡状况比较

	治疗组 (n=94)		对照组 (n=106)	
	n	(%)	n	(%)
溃疡灶数				
单个	88	93.6	92	86.8
两个	6	6.4	14	13.2
第一个溃疡长径 (nm)				
3~5	42	44.7	51	48.1
6~10	44	46.8	48	45.3
>10	8	8.5	7	6.6
第二个溃疡长径 (nm)				
3~5	2	2.1	8	7.5
6~10	4	4.3	5	4.7
>10	0	0	1	0.9

经统计学检验, 两组各项指标差异均为  $P>0.05$ , 说明治疗前两组的基本情况没有差别, 组间均衡性良好, 为研究结果的准确与可靠奠定了基础。采用统计学方法, 对资料均衡性进行分析的过程又叫做均衡性检验。

### 三、保证均衡性的方法

在临床研究中, 可以从研究设计、研究实施和资料处理等方面采取措施, 以尽量提高研究组间基线资料的均衡性。

#### (一) 研究设计

1. 限制研究对象的人选条件 为了提高研究对象整体的齐性, 可以在研究设计中, 通过设定纳入、排除标准, 对受试对象的基线条件进行限制, 避免研究对象基线差异过大。例如, 为评价某降糖药治疗 2 型糖尿病的效果, 可以限定患者的年龄、体重、并发症等基本情况, 甚至将研究对象的空腹血糖限定在一定范围, 排除血糖过高的患者。在此基础上再进行分组, 组间基线情况的差异可以有效降低。

2. 匹配设计 匹配 (matching) 也称配比, 即在设计时要求对照组在某些因素和特征上与试验组保持一致。匹配条件包括性别、年龄等一般情况, 也可以是病型、病程、病情等临床特征。根据设计的不同, 匹配可以分为群体匹配 (频数匹配) 和个体匹配两种类型。在临床试验中, 通过匹配设计, 使对照组与试验组在可能影响研究结果的非研究因素



(背景因素)上保持一致,将有效提高组间均衡性。

3. 分层随机分配 因机遇的影响,一般的随机化分配,难以保证一些可能影响研究结果的重要因素在试验组与对照组之间均衡分布。而采用分层随机分组,研究者可将重要的危险因素或预后因素设为分层因素,将患者分层后再通过严格的分层随机分组,可以在一定程度上提高比较组间基线的可比性。例如,研究冠心病治疗效果时,按患者的年龄、性别、血清胆固醇水平等进行分层随机分配,可以有效保证研究组间以上因素的均衡可比。

4. 制定统一的检测标准 基线资料的准确性将直接影响均衡性检验的结果。为使基线资料能真实反映研究对象的实际情况,必须制定统一的检测标准。例如测定体重指数,必须事先规定量体重的时间是在餐前还是餐后;测定血压值,需规定要连续测多少次,是以最低值还是平均值为准,同时,还需要统一测量时间,是早晨还是晚上。如果没有按标准进行测量,甚至不同的组采用不同的测量方法,会使结果产生较大差异。

## (二) 实施过程

在研究实施阶段,应在测定基线数据后,尽快实施干预。如果间隔时间过久,患者的基线情况可能发生改变,进而破坏组间的均衡性。例如,一般认为一项有效的治疗在3个月之内可减少10%~15%的并发症发生。如果在采用历史对照设计的慢性病治疗效果的研究中,取得基线资料1个月后才开始实施处理,则可能已经有5%的患者病情自行好转,经治疗3个月后将会有15%~20%的患者减少了并发症的发生,这样就夸大了对疗效的评价。

## (三) 资料处理

关于研究基线状况的分析,重点是比较试验组和对照组主要临床特点以及性别、年龄和样本量是否均衡、可比。组间应不具差异性,如果具有差异的显著性,如 $P < 0.05$ ,则基线不可比或称为基线不均衡。此时,应据主要影响疗效或预后的因素,将两组的试验对象进行分层分析,如各层均衡可比,则真实性佳,否则,会导致研究质量的下降。

上述四大临床科研设计的基本原则及其实施,对保证科研的高质量十分重要!但是,临床科研的对象是病人,鉴于病后的社会-心理-生物学效应的十分复杂性,加之科研设计与执行者的学术水平以及研究环境和工作条件的复杂性,因此,在研究实践中执行以上原则,就应对具体情况、具体分析,科学地实事求是创造性地应用和发挥同样也是十分重要的,切忌教条主义。

(李 革 李 静)

## 第八章 临床医学研究对象的来源与选择

研究对象是科学研究的要素之一。临床观察性和实验性研究的研究对象无不是选择一组样本作调查研究，从理论上讲研究某种疾病的全部病例最为理想，可以取得完整、无一遗漏的结果，其实不然，似乎避免了抽样误差却因调查者和被调查者的众多介入反而增大了产生系统误差的可能性。如果在一个小的空间，例如在医院范围内，收集一定时间内所有该病的病例也不可避免地存在一定偏性，这种缺少代表性的研究限制了它的科学性。所以临床研究在选择研究对象的时候，必须讨论样本的来源、抽样的基本方法、选择研究对象的条件或标准以及研究所需要的样本数量，最后还要审视它对总体的代表性。

### 第一节 样本的来源

临床科研的样本可来源于社区人群中的病例，也可来源于医院就诊的病人；依据设计的要求还需要来源于暴露或未暴露某些可疑致病危险因素的健康人群等等。现就通常的样本来源分述如下。

#### 一、社区人群

##### （一）传染病报告

新中国成立初期就依法规定了传染病的报告制度，责任报告人发现传染病病人或疑似者，依法填写传染病报告卡，按规定的时限逐级上报疾病预防控制（卫生防疫）部门，于是可以从卫生防疫或隔离治疗单位取得样本，有利于疑难传染病的防治研究。例如我国政府已将新型冠状病毒所致非典型肺炎（SARS）列为法定传染病，有利于该传染病的监测以及防止复燃和再次流行。这种样本来源的质量和 research 价值取决于传染病报告的漏报率和订正报告的实施情况。

##### （二）疾病报告登记

主要指一些慢性病的发病与死亡报告。20 世纪 70 年代末期，我国上海、北京等城市开始肿瘤登记；80 年代初期，天津市开展了高血压、冠心病、脑卒中、肿瘤四病的登记报告；浙江省在全省范围内建立心血管病防治体系，逐渐形成了以社区为基础的慢性病防治模式。近年来，国家卫生部门正在逐步推行慢性病社区综合防治规划，社区诊断所取得的基线资料为常见病的研究提供了样本来源和足够数量的新病例，同时可得到个体危险因素和环境致病因素的信息，有利于对一些起病缓慢且隐匿的疾病进行临床医学研究。但由于疾病报告的网络与体系正在初建，有待于进一步完善，以充分满足调查研究的需要。

##### （三）疾病监测

为制定预防和保健措施，考核其防治效果而长期连续地调查、收集和分析人群中的疾病动态、分布和影响因素称为疾病监测（surveillance of diseases）。国外疾病监测始于 20 世纪 40 年代的美国疾病控制中心（CDC），我国的疾病监测可追溯到 20 世纪 50 年代初期的传染病监测，1980 年以后形成体系并扩大了疾病的监测范围，先是原中国医学科学院流行病学研究所牵头在全国 13 个省、市、自治区建立了疾病监测点，至今已有 71 个，覆盖 1000 万人口。其后是京、津、冀地区实施的 120 万人心血管病监测的 MONICA 计划。疾



病监测要求进行发病与死亡报告, 计算发病率、死亡率等人群疾病负担指标, 同时要收集与发病有关的个体危险因素和实验室的各项检测结果, 为专题研究提供了条件。

#### (四) 普查和筛查

在发病率或患病率较高的地区和人群中, 采用简便易行、真实性较好的检测手段做疾病普查 (census) 或筛查 (screening), 所发现的病例可作为研究的样本来源。例如在流行地区进行血吸虫病的普查, 以及在高危人群中进行食管癌、原发性肝癌、鼻咽癌等肿瘤和 HIV/艾滋病的筛查等等。普查和筛查可以在较短时间内获得基线特征较齐全的足够数量新病例, 从中又可得到可比性较好的对照, 是一种比较理想的样本来源。然而一般普查和筛查要耗费相当大的人力和财力, 是否切实可行, 要做成本与效益等卫生经济学评价。

#### (五) 现场

这里所指的现场是疾病发生、流行所在的空间范围, 可以是城乡社区或学校、工矿、企业、宾馆、医院不等。传染、中毒常以突发事件出现, 时间短暂, 调查研究者需亲临现场, 就地取得样本病例。例如 2005 年 5~6 月间, 四川省 12 个地市所发生的人感染高致病性猪链球菌的暴发流行, 采用现场研究查明了病原、血清型别, 并以现场的病例进行了病例对照研究, 阐明了传播因素及其联系的强度。现场研究的特点是以新发病例或新感染者为主, 同时也为调查研究提供了即时的环境条件, 但是由于疾病过程短暂, 边调查边进行医疗、卫生等处理, 有些前瞻性的研究受到一定的限制。

## 二、医院和专门防治机构

### (一) 医院

医院是临床研究样本来源的主要场所, 可涉及门诊及住院病人。医院的技术力量强、设备齐全, 短时间内可以得到足够数量、依从性好的病例和对照, 又便于质量控制, 历来不少病因学研究的突破和药物疗效的评价均来自医院研究。医院及其所积累的病例成为临床医生从事科学研究不可多得的有利条件, 但它对总体的代表性可能不足, 容易发生选择偏倚。

### (二) 专门防治机构

我国针对某些严重危害人群健康的疾病, 如结核病、麻风病、性病、地方病、寄生虫病等设立了专门的防治机构。此外还有对危害工矿企业职工健康的职业病设立专门了防治系统, 以及针对妇幼健康与疾病有专门的妇幼保健系统等。通过日常性的防治, 可基本掌握本部门或本系统的发病资料, 成为这类疾病调查研究理想的样本来源。但由于收治病种单一, 应用范围会受到一定的限制。

### (三) 健康体检中心

为了适应人们日益增高的健康意识和需求, 近年来全国各省市普遍以医院为依托成立了健康查体 (体检) 中心, 此举不仅可以发现新的或早期病例, 也能检出临床前期者。由于大多为集体体检, 因此, 筛查出的病例在职业、经济和文化背景相近, 成为一些专题研究和干预试验理想的样本来源, 但这一资源还没有得到充分利用。

## 第二节 样本的抽样方法

临床研究不可能将研究疾病的所有病人集中在一起后才开始进行, 只能是从中抽取一定数量的样本进行调查研究。为了尽量降低因抽样产生的抽样误差对研究结果的影响, 应按设计的要求计算最适的样本含量。如此, 使样本的研究结果做到在质与量的两个方面较为真实和可靠地反映出总体的特征, 为疾病的认识、防治效果的评价提供可信的依据。因



此,在临床科研设计中,对样本抽样方法的选定和应用就显得十分重要了。

无论采用何种方法抽取一定数量的样本,均应按照科学的设计方案限制机遇(chance)因素对研究结果的影响。但是机遇在科学研究中,尤其是以人体为主要研究对象的临床研究中是不可避免的,通常需要加以限定,如把出现假阳性(第一类错误)限定在小于5%( $\alpha < 0.05$ )的水平上,将可能出现的假阴性(第二类错误)限定在不超过10%( $\beta < 0.1$ )或20%( $\beta < 0.2$ )的水平上,后者意味着得出或检出具有阳性结果的把握度( $1-\beta$ )在90%或80%以上。

## 一、个体随机抽样

### (一) 入选个体的随机抽样

当合格对象多,而需要的样本量又有限时,可以从合格的研究对象中做个体随机抽样。通常用随机数字表等做简单随机抽样。

### (二) 入选病例的随机抽样

在回顾性研究中,常利用医院病历或病例登记册选择病例,如所研究病种的病例数量较多,而需求的样本量又有限时,也可采用简单随机或系统随机抽样的方法,从中抽取所需的样本数。

### (三) 分层随机抽样

如在样本量不大的研究课题中,组间疾病分型、病情不同可影响研究结果,为了保证组间的可比性,可按病情轻、中、重进行分层后,再随机抽样。另外,在多个医院协作研究中,由于医院级别不同,收治病人的病情与病程可能参差不齐,为保证所抽取的样本能反映该种疾病的临床特征和全貌,也可按上述特征分层后再随机抽取样本。

## 二、整群随机抽样

在涉及社区人群健康的大型研究中,往往需要很大的样本量,方能反映总体的真实情况。但事实上又不可能对整个社区人群进行普查。对此,可按照群组单位对社区人群进行整群随机抽样,然后对抽取群组中的全部个体进行调查研究,借以反映总体的健康状况。这种群组单位可以是行政建制,如县、乡、村或区、街道办事处、居委会,也可以是工厂、车间、班组等。依据样本量的大小还可进行多级(阶段)整群抽样。

## 三、系列随机抽样

若临床研究所需样本不大或者病例本身就很稀少时,可按照估算的样本量,以病人就诊的先后顺序纳入研究,直至满足样本量为止。为保证抽取样本的随机性,不得随意更改进入样本的顺序。

## 四、非随机抽样

非随机抽样带有极大的随意性,易产生选择偏倚等,很难保证研究质量和所得结果的真实性(validity),所以一般不宜采用。在制定详尽合理的纳入标准与排除标准的情况下,借用“限制”方法控制非研究因素的影响,可以有条件地应用非随机抽样。如突发病因或发病不明的疾病、发病率低的疾病的应急调查研究中可以采用非随机抽样。艾滋病(AIDS)发生与流行的早期,在美国加州曾采用类似的方法研究当时病因不明、传播途径不清又引起社会强烈反响的AIDS。SARS发生与流行的早期,对病原学、传播途径、治疗原则、病死率所见各异,与病例样本量小且仅来源局限的非随机样本不无关系,所以该种抽样,可能出现选择性偏倚,研究结果的外推性(generalizability)也受到一定的限制。





### 第三节 诊断标准

选定研究对象的首要条件是必须符合疾病的诊断标准，它是研究对象纳入的基本前提，同时又是保证研究质量与真实性的基础。

#### 一、诊断与疾病诊断

诊断 (diagnosis) 的词义是把一个事物与另一个事物区分开来。疾病诊断是从认识疾病到确定疾病的一个过程，包括病因学诊断、病理组织学和病理生理学诊断等，为临床治疗和预后判断以及决策提供可靠的依据。

#### 二、疾病诊断标准的制定

##### (一) 制定的依据

临床研究中，所选定的研究对象是患者时，一定要依据诊断标准 (diagnostic standard) 予以明确的疾病诊断。凡属国际疾病分类中的疾病都有相应的科学诊断标准，而诊断标准的制定又受着科学水平和认识水平的限制，所以任何疾病诊断标准都是随着科学发展和人们对疾病认识水平的提高而逐渐合理完善的。从 1979 年至今 30 多年来，我国对高血压诊断标准曾有多次修订与更改，现执行的是 1999 年 2 月由世界卫生组织/国际高血压协会 (WHO/ISH) 所确定的标准。糖尿病也是如此，1999 年 10 月我国采纳了同年由 WHO 咨询专家与国际糖尿病联盟 (IDF) 所制定的推荐标准。

诊断标准的制定可分为两类：

1. 金标准诊断法 凡属于特异性强且被解剖、病理以及生物学研究肯定的临床诊断称为金标准 (gold standard) 诊断。像肿瘤的病理学诊断、遗传性疾病 (染色体或基因异常) 的分子生物学诊断、传染病的病原学诊断、免疫性疾病的免疫学诊断、冠心病的影像学冠脉造影诊断等，已广泛用于诊断/筛查以及疾病预后研究。

2. 临床诊断标准 凡缺乏金标准诊断法者，可按临床发病特点、临床症状、体征和实验室检查，由有关专家商讨和提议制定出的标准。像全国性专业学会讨论制定出的风湿热、类风湿病、肺心病、脂肪肝的标准，都是临床诊断标准。此外，还可进一步根据临床病情特点以及治疗与预后的不同，制定疾病的分型或分级诊断标准。

在制定国内疾病诊断标准时，最好能参照国际通用标准，取得诊断标准的一致，便于国际间的比较和交流。如心肌梗死和上述的高血压和糖尿病标准等。

##### (二) 国际疾病分类

疾病的诊断是表明一种疾病在解剖、病理形态和生理上都有它相对独立的特征，能够实现疾病的分类。国际疾病分类 (international classification of diseases, ICD) 是国际通用的疾病分类标准，我国已于 20 世纪 80 年代开始采用 ICD，并由北京协和医院负责组织编译、培训。

国际疾病分类始于 1893 年，最初只是一个简单疾病死亡分类，以后逐步修改完善，至 1976 年，经第九次修订后，就形成了一个科学合理的分类系统 (ICD-9)，该系统同时兼顾了临床、医疗管理和健康保险等多方需求。1991 年又完成了第十次修订，形成 ICD-10 英文修订版，将原来的 17 章扩充到 21 章，建议各成员国于 1993 年采用，希望能够成为较长时间使用的版本。

##### (三) 临床前期的判定

达到临床诊断只是表明已处于疾病自然史的临床期。随着科技进步以及生物医学技术



的日益成熟,临床诊断时机大幅前移。临床上将处于临界状态的时间段,称为临床前期,在该阶段出现了一些轻微的临床体征及(或)实验室检测异常,虽未达到诊断标准,却极有可能进展到临床期。例如出现空腹血糖损害(IFG)、糖耐量降减(IGT)划归为糖尿病前期(prediabetes)。美国将正常血压的高临界值(120~139/80~89mmHg)称为高血压前期(prehypertension),此时血清胆固醇、低密度脂蛋白胆固醇已出现边缘升高。

目前疾病临床前期的预防与控制已受到广泛关注。疾病只有早诊早治,才能降低发病率、病死率、改善生存质量,最终促进人类健康水平的整体提升。

#### (四) 诊断标准是确定研究对象的首要条件

诊断标准对研究对象的确定至关重要。如果不把研究对象限定在统一诊断标准的基线上,研究结果难以置信,必将会过高或过低地估计某种研究因素与某种结局间的联系,失去参考价值,所以对一项研究项目或成果的评价首先是审核研究对象是否按统一诊断标准确定和选择。

疾病诊断时,所采用的检验方法和仪器型别都应符合诊断标准的统一规定。选择的对象原则上是被金标准确诊了的病例,如在诊断试验研究中,冠心病、溃疡病应分别以动脉造影、胃镜检查结果为准。

### 第四节 纳入与排除标准

符合统一诊断标准是选择研究对象的首要条件,然而,符合诊断标准的研究对象,却不一定都符合研究设计的要求。因为临床研究对象的病情轻重不一、病程可能不同,合并症和并发症存在差异,同时心理因素、文化和社会背景等也不尽相同,使得探讨某一种因素效应时,受到诸多非处理因素或非研究因素的影响。因此在选择研究对象时还应制定纳入和排除的标准,最终确定符合研究设计要求的合格对象(eligible subject),从而使研究因素相对单一,排除某些非研究因素的混杂,确保研究的质量,并为重复试验或进一步研究提供基础。

#### 一、纳入标准的制定

按照研究课题的科学构思,以及暴露或干预因素研究拟达到的目的,制定符合研究课题要求的纳入标准(inclusion criteria)。

纳入标准的要点是从复杂的群体中,选择临床特点和人口学相对同质的对象进行研究。例如在西欧和以色列等9个国家、703个医疗中心所进行的高血压干预研究中(INSIGHT, 1994年9月到1999年7月),鉴于钙拮抗剂对伴有糖尿病的高血压病人的疗效结论不一,与应用血管紧张素转换酶抑制剂(ACEI)相比,药物作用机制可能不同,为此研究了硝苯地平控释片(nifedipine)对伴有糖尿病的高血压病人的干预效果。(INSIGHT研究对象的纳入标准为:①白人、年龄55~80岁;②血压>150mmHg(20.0kPa)/95mmHg(12.7kPa)或收缩压>160mmHg(21.3kPa);③同时至少有一种心血管病危险因素(家族史、高胆固醇血症、吸烟、左室肥厚、心梗史、糖尿病等)。

在多中心试验研究中,各研究单位应以统一的纳入标准选择研究对象。纳入标准的制定应简明扼要,不宜设置过多的限制条件,否则会降低研究结果的推广应用价值。

此外,在纳入研究对象时还应尽可能选择新患病例,使其可以从中获益。凡是纳入的研究对象必须签署知情同意书(informed consent form)。

#### 二、排除标准的制定

如上所述,研究对象的来源、病情、社会经济地位、心理状态以及所接受的治疗措施



等都会影响研究结果。为了防止这些因素的干扰,提高试验研究结果的可靠性 (reliability), 只有纳入标准还不能更好地控制临床上千变万化的各种非研究因素, 因此, 应根据研究目的以及干预措施, 制定相应的排除标准 (exclusion criteria), 使研究对象处在同一个基线上, 以便能真实反映出研究效应。例如 ACEI 用于治疗慢性心力衰竭是临床治疗的一大进展, 但由于对其降低血压和肾功能损害抱有疑虑, 存在用量不足的问题。为此, 北京西拉普利 (cilazapril) 治疗心力衰竭协作组进行了大剂量治疗的安全性和疗效的临床试验, 确定 NYHA (纽约心功能分级) II ~ IV 级、收缩压  $>90\text{mmHg}$  ( $12.0\text{kPa}$ )、左室射血分数  $<45\%$  等作为纳入标准, 同时制定了排除标准: ①瓣膜疾病; ②不稳定性心绞痛; ③急性肺水肿; ④血肌酐  $>177\text{mmol/L}$ ; ⑤对 ACEI 药物过敏; ⑥双侧肾动脉狭窄; ⑦血清转氨酶超过正常上限 3 倍。如此, 可以对药物的适用范围、适应证和禁忌证及其疗效等进行全面评价。

另外, 有试验药物过敏或不良反应者、病情危重随时有可能发生意外事件者亦应列入排除范围。除非有专业研究需要, 一般孕妇不宜列为新药评价试验的研究对象。

## 第五节 影响样本代表性的因素

为确保研究结果的真实性和可靠性, 选择的研究对象一定要有对总体的代表性, 并对影响样本代表性的因素予以控制。毋庸置疑, 所抽取的样本必须能够充分代表总体的特征, 样本含量则是影响其代表性的重要因素之一。

### 一、样本含量

临床研究的样本含量应是最适的样本大小, 即按照总体客观存在的性质与特征以及研究者能容忍的误差风险而决定的样本含量。样本过小, 往往使检验效能过低, 结论缺乏依据; 反之, 样本量过大, 会造成人力、时间和经济的过度耗费。虽然当前有人提倡大样本的前瞻性随机对照试验, 以获得可靠性良好的研究成果, 但毕竟会受到多种因素的制约, 难以推广应用。所以具有最适样本含量的临床研究, 既可以准确估计总体效应的大小及范围, 又能保证临床研究的可行性。有关样本含量的计算方法详见相关章节与参考书。

### 二、随机原则

为使样本反映总体的性质和特征, 对研究对象的选择应实施随机的原则, 这里包括随机抽样 (random sampling) 和随机分组 (random allocation)。随机抽样是按随机的原则从总体内抽取若干个病例、个体或群组作为样本, 总体内每个单元有同等机会或概率被抽作样本, 使样本的框架 (sampling frame) 符合总体构成。随机分组是将研究对象随机地分配到试验组或对照组, 每个研究对象都有同等的机会进入试验组或对照组, 使两组的临床特征和影响预后的一些未知的因素能够均衡分布, 具有良好的可比性, 必要时还需做均衡性检验。

### 三、无应答和失访

无应答 (non-response) 是指由于研究对象或研究者的种种原因而造成研究对象对调查问题的不予回答, 因而丢失一部分研究对象的信息, 直接影响到研究结果对总体的代表性。产生无应答的常见原因有调查对象不了解调查的意义或存有戒备、恐惧心理而拒绝回答, 也可因调查者所采用的调查方法、技巧、态度不当而引起的回避等等。

由此引起的无应答偏倚 (non-respondent bias) 可使研究结果失真, 所以应尽力设法



控制不应答人数,保持对总体的代表性,使样本能够推论到总体,并对它的可信度作出估计。一般认为应答率(response rate)应在80%以上。

失访(withdrawal)与无应答相似,也会影响样本代表性。失访常见于前瞻性研究,由于这类研究随访时间长、观察人数又多,或因人口的迁徙、流动、死亡等等原因导致一部分研究对象的失访。通常失访率应控制在10%以内。

总之,为确保研究的真实可靠性,应针对无应答和失访的种种原因,采取相应的措施设法补救,力争将其控制在允许范围之内。

## 四、依从性

依从性(compliance)是指受试者对治疗措施及其他需要控制因素的依从和执行程度。依从性的好坏直接影响研究或试验的真实性,可以通过询问、服用药物量的计算、临床反应和药物代谢产物的测定等方法做依从性评价。这些评价方法的可行性和可靠性有着不同程度的限制,一旦造成不依从的事实,将难以排除它对结果带来的影响。

依从性如何,虽然不像无应答或失访那样直接减少人数、降低检验效能,但由于对治疗、干预措施的不依从而潜在地影响研究结果的真实性,使所调查的样本无形中降低了对总体的代表性。

提高依从性的关键在于研究设计时应确定好研究对象的来源,提出防止不依从出现的措施。一般来源于医院的样本依从性好于社区人群,而医院的住院患者又优于门诊患者。临床试验在选择受试对象时,事先应讲明试验的意义和利害关系,征得本人的同意,并签署知情同意书,这既有利于提高依从性,也符合医学伦理原则。

此外,在临床研究的其他环节,如调查表的设计、调查方式方法、药物剂型、给药途径等都会影响受试者的依从性。

## 五、选择性偏倚

选择研究对象时所发生的系统误差称为选择偏倚(selection bias),如果发生这种偏倚,也将影响样本对总体的代表性。常见的选择偏倚分述如下。

### (一) 入院率偏倚

因入院的机会不同而产生的偏倚称入院率偏倚(admission rate bias),是由 Berkson (1946) 提出并加以论证的,所以又称 Berkson 偏倚。

临床研究的对象主要局限于同一个医院,患者是否入院治疗以及进入哪一级医院并不是完全随机的,受到患者本人和医院双方诸多因素的制约,如患者的经济、病情、交通条件、福利和医院的技术、设备、规模、信誉等。所以,来自同一个医院的样本并不是一个无偏样本,对研究结论的解释宜取慎重的态度。为了提高样本的代表性,研究设计时,就要提出防止选择偏倚的相关措施。如病例对照研究可从不同等级的多个医院选取病例,从社区人群中的普查或筛查中选择部分病例,设多个对照组,最好设一组社区人群随机抽样的对照组,可以最大限度地防止入院率偏倚。

### (二) 现患-新病例偏倚

Neyman 在 1955 年利用现患病例进行病例对照研究时提出了现患-新病例偏倚(prevalence-incidence bias),所以又称 Neyman 偏倚。研究对象选自现患病例,实际上是该种疾病的幸存者,而未包括该病的死亡病例以及病程短、不典型或处于潜伏期的病例。一些慢性病的现患病例大多病程长,生活习惯或行为可因干预而改变或得到改善,长期的治疗也会降低某些危险因素的水平,如高胆固醇血症的控制、戒烟等。已有实例证明,与队列研究相比较,病例对照研究往往过低地估计危险因素与疾病的联系,所以只以现患和典型病



例作为研究对象，限制了样本对总体的代表性，解释研究结果时应留有余地。为了减少这种偏倚的影响，力求从普查或筛查时发现的新病例中随机地选择研究对象，或者可以在不同医院或地区选取病例作为研究对象，可以认为是防止现患-新病例偏倚（prevalence-incidence bias）的一种可行措施。

### （三）检出征候偏倚

临床上的患者往往是出现了不适症状后才去就医，但同一征候（如发热、头痛、出血、白细胞增多、肝脾肿大）可为不同的疾病所致；不同疾病也会有相同的征候重叠出现（如各个系统的感染都会有发热、白细胞增多，胃溃疡和肝硬化都会发生呕血等）。此外，征候轻微或处于潜伏期的患者，也不一定到医院就诊。因此，凡从医院内选择的研究对象就有可能受到检出征候偏倚的影响，以致研究的结论不能完全反映出总体的特征，甚至于可以造成虚假的联系。例如有回顾性研究报道，女性激素避孕药物是引起女性子宫内膜癌的危险因素，后经前瞻性研究予以否定了。为什么会造成上述虚假的因果联系的研究结论呢？这是因为服用女性激素药物易发生阴道出血的征候，促使患者到医院就诊，行妇科检查，增加了检出子宫内膜癌的机会，于是得出两者存在因果联系的虚假结论。如果能以普查、筛查或疾病监测所发现的病例作为研究对象，可以防止检出征候偏倚（detection sign bias）的发生。

（李绍忱 刘续宝）

## 第九章 临床研究设计的主要方案与运用

医学研究分为基础医学、临床医学和预防医学研究三部分，其共同目标是研究人类疾病的发生和发展规律，探索正确、早期诊断的方法，并进行有效地防治，借以提高人民健康水平。

基础医学研究与临床研究有密切关系，但也有很大的不同。基础医学研究主要采用实验室研究，研究对象为动物模型或离体的组织、细胞，研究因素和实验条件可以进行严格控制，如两组动物的年龄、性别和体重。临床研究是以病人为研究对象的医学科学研究，由于病人的生理特点、文化水平、经济地位、民族及宗教信仰各不相同，患病以后的病情、病程变化受社会-心理-生理综合因素的影响，即使患同一种疾病，病理损害相似，临床表现却有明显差异。同时，由于病人对临床研究的依从性不同，以及许多不可控制因素的影响，使临床研究比基础研究更容易出现系统误差（偏倚）和随机误差（机遇），影响临床研究的真实性和可靠性。临床研究要获得正确的结果和结论，最有效的方法就是要有严格的研究设计，尽可能控制和减少偏倚和机遇的影响。因此，如何选择与研究课题相应的论证强度高、可行性好的设计方案，是保证临床研究质量的关键因素。

传统的流行病学将研究的设计方案分为两大类：观察性研究（observational studies）和试验性研究（experimental studies）。试验性研究设计者可以人为控制试验条件，随机分组、根据研究目的设置合理的对照、盲法观测试验结果，以探讨某因素与疾病的联系。因此试验性研究的论证强度较高，结论较可靠。常用的试验性研究有随机平行对照试验、交叉试验、前-后对照试验等。观察性研究与试验性研究最主要的区别在于研究者不能人为地控制试验条件，分组系自然形成，只能尽可能控制非研究因素的影响，以求得结论的真实、可靠，因此观察性研究的论证强度常不及试验性研究。常用的观察性研究有描述性研究、病例-对照研究和队列研究。

但考虑到临床研究的特点、实用价值以及各类设计方案获得研究结果和结论的论证强度，参考加拿大 McMaster 大学的分类方法，将常用于疾病病因、诊断、预防与治疗、预后等方面的设计方案，依照它们的论证强度和研究者是否能主动控制试验条件，分为四个级别（表 9-1）。

表 9-1 临床研究设计方案分级及其特点

级 别	特 点	常用设计方案
一级设计方案	<ul style="list-style-type: none"><li>● 前瞻性研究设计方案：试验开始时尚无研究的结果</li><li>● 同期对照</li><li>● 偏倚因素及研究措施可以被主动控制</li></ul>	<ul style="list-style-type: none"><li>● 随机对照试验</li><li>● 半随机对照试验</li><li>● 交叉试验</li></ul>
二级设计方案	<ul style="list-style-type: none"><li>● 前瞻性研究设计方案</li><li>● 有对照</li><li>● 偏倚因素及研究措施不能被主动控制</li></ul>	<ul style="list-style-type: none"><li>● 队列研究</li><li>● 前-后对照试验</li><li>● 非随机同期对照试验</li></ul>
三级设计方案	<ul style="list-style-type: none"><li>● 多数结果在研究开始时已存在</li><li>● 可有对照</li><li>● 偏倚因素及研究措施不能被主动控制</li></ul>	<ul style="list-style-type: none"><li>● 横断面研究</li><li>● 病例-对照研究</li></ul>
四级设计方案	<ul style="list-style-type: none"><li>● 无对照</li><li>● 偏倚因素较多</li></ul>	<ul style="list-style-type: none"><li>● 病例分析</li><li>● 病例报告</li><li>● 专家评述</li></ul>

根据表 9-1 设计方案的类别，本章将分别对各种设计方案的设计模式、适用的研究范围、方案的优缺点及相应结果的分析方法作较为系统的论述，以便读者在临床研究的实践中选用。



## 第一节 随机对照试验

### 一、概况

随机对照试验 (randomized controlled trial, RCT) 是采用随机分配的方法, 将合格的研究对象分别分配到试验组和对照组, 然后接受相应的试验措施, 在一致的条件或环境中, 同步地进行研究和观测试验的效应, 并用客观的效应指标对试验结果进行科学的测量和评价。

1926 年, 由 Fisher 首次将随机对照试验应用于农业实验, 而随机对照试验被应用于临床研究以评价治疗措施的疗效是 1946 年。1946 年先后有两个随机对照试验, 分别评价链霉素治疗肺结核和免疫措施治疗百日咳的效果, 并发表于 1948 年和 1951 年。近 50 多年来, 随着理论和方法的日趋成熟, 随机对照试验被公认为评价干预措施疗效的金标准或标准方案而广泛应用于临床研究中, 为疾病治疗、预防和康复提供了大量真实、可靠的依据。同时, 随着 RCT 的研究的新发现, 还不断地在更新着人们的传统医疗意识, 例如:

#### (一) 既往认为有效的治疗措施经随机对照试验证实无效

过去, 某种治疗措施或药物是否应用于临床实践, 主要取决于专家或顾问的意见。而临床医师选择治疗措施也主要根据个人既往治疗患者的成功经验或对疾病的病理生理学机制的理解, 至于某种治疗措施的真正疗效或副作用有多大, 受哪些因素的影响, 难以进行科学的评价。在临床实践中, 某些治疗措施的效果非常明显, 医生根据临床经验即可进行判断, 如青霉素治疗大叶性肺炎, 外科手术治疗阑尾炎等。但慢性非传染性疾病, 因其为多因素致病, 其治疗措施的疗效并不十分显著并受多种因素的影响, 如患者个体的疾病特点、经济和社会因素、医疗水平等, 要明确某种治疗措施对该病的确切疗效, 有必要进行严格评估。例如, cyproheptadine HCL 和 trimeprazine tartrate 被广泛应用于治疗慢性严重瘙痒, 但随机对照试验证明, 两种药物的疗效与安慰剂相似。

#### (二) 病理生理机制推论有效的治疗措施, 经随机对照试验证实无效或有害

由于疾病发病机制的复杂性和认识水平的局限, 单纯根据疾病的病理生理机制、实验室研究结果推断某种干预措施在人体的疗效, 有时可能误导。例如, 从疾病的病理生理进程来看, 心肌梗死患者发生室性心律失常是猝死的重要危险因素, 因此我们有充足理由对此类患者常规使用抗心律失常药物, 但随后的随机对照试验证明, I 类抗心律失常药虽然能抑制室性心律失常, 却增加病人的病死率; 短效钙拮抗剂虽然能有效降低高血压患者的血压水平, 却增加心肌梗死和死亡风险。因此, 要明确某种治疗措施的短期和长期疗效或副作用, 必须进行以人体为研究对象的临床试验。

#### (三) 其他类型研究设计方案可能夸大或缩小治疗措施的真实效果

由于临床研究的复杂性, 研究的质量也受多种因素的影响, 如设计方案的强度、研究对象的选择和分配方法、是同期对照还是历史对照、是否采用盲法测量结果、是否控制各种偏倚因素的干扰等。1977 年, Chalmers 对 32 篇抗凝剂治疗急性心肌梗死的临床试验进行了分析和评价, 其中 6 篇 RCT, 8 篇非随机临床对照试验 (controlled clinical trial, CCT), 18 篇历史对照试验 (historical control trial, HCT), 与 RCT 比较, 死亡的相对危险度降低率在 HCT 和 CCT 中分别夸大了 35% 和 6%; 1982 年, Sacks 对 50 篇 RCT 和 56 篇 HCT 进行了比较分析, 结果发现, 79% 的 HCT 证明新的治疗方法优于传统治疗方法, 仅 20% 的 RCT 证明新的治疗方法优于传统治疗方法。可见, 由于各种因素的影响,



可能导致临床研究结果偏离真实的情况。

#### (四) 单个随机对照试验的局限

随机对照试验能真实、客观地评价干预性措施的疗效，被公认为评价预防、治疗和康复措施疗效的“金标准或金方案”。但是，单个随机对照试验由于研究对象、研究地点、样本含量、干预措施等的差异，研究结果的代表性局限，且针对同一临床问题的众多随机对照试验可能结果不一致，甚至结论相反，造成临床医生和患者难以进行医疗决策。理想的解决方法是开展大规模、多中心的临床试验，但因此类研究需要有效的组织管理、大量的经费、高素质的研究人员，这在许多单位难以实施，为此，1979年，英国著名流行病学学家 Archie Cochrane (已故) 首先提出进行系统评价 (systematic review)，以某一具体临床问题为基础，系统、全面地收集全世界所有已发表或未发表的随机对照试验，采用临床流行病学严格评价文献的原则和方法，筛选出符合质量标准的文献，进行定性或定量合成 (即 meta 分析)，去粗取精，去伪存真，得出综合可靠的结论，并随着新的随机对照试验的出现及时更新，随时提供最新的知识和信息作为决策依据，以改进临床医疗实践和指导临床研究的方向。因此，这有利于克服样本量不大的单个 RCT 的局限性。

## 二、应用范围

随机对照试验虽然被公认为“最佳的治疗性研究设计方案”，但不能用于研究和解决所有的临床问题。在某些情况下，使用随机对照试验是不可行或不恰当的，如诊断性研究、病因学研究、疾病预后的自然病史等。

#### (一) 临床治疗或预防性研究

随机对照试验最常用于治疗性或预防性研究，借以探讨某一干预或预防措施 (药物、治疗方案、筛查方法等) 的确切疗效，为正确的医疗决策提供科学依据。

#### (二) 在特定的条件下，可用于病因学研究

多数情况下，病因学研究不适于采用随机对照试验，将某种致病因素和危险因素施加于人体，进行致病效应的研究是不符合医德的。例如，要了解吸烟在肺癌发病中的作用，人为设计一随机对照试验，将原本不吸烟的研究对象随机分配入吸烟组或不吸烟组，随访数年，比较两组肺癌的发生率，显然既不可行，也有违医德。

但在特定的条件下，随机对照试验也可以用于病因学因果效应研究。应用的前提是：尚无充分证据证明某种可能致病因素对人体有危害，但又不能排除它与疾病的发生有关。如果已有研究证明某一因素对人体有害，就不允许将该因素用于人体进行随机对照试验。

例如，妇产科为预防早产儿因缺氧带来的大脑损害和对今后智力发育不全的影响，曾对早产婴儿均施以高浓度的氧气疗法，几乎被常规应用。后来发现经此治疗的婴儿，出现了眼晶体后纤维组织增生，导致不同程度的视力障碍，严重者失明。经分析推论，认为可能与高浓度氧疗有关，为证实这种因果效应，于是采用了随机对照试验，一组早产儿继续用高浓度氧疗，另一组则用低浓度氧疗。经追踪观察分析，上述视力障碍确与高浓度氧疗有关，于是，临床上就淘汰了这一疗法。后来，这一病征，被命名为 Terry's Syndrome。

#### (三) 非临床试验的系统工程

随机对照试验还可应用于非临床试验的系统工程如教育学和农业。例如，要评价循证医学教育模式与传统医学教育模式的教学效果，可将条件相似的学生随机分配进入任何一组，课程结束后进行短期或长期效果的评估。





### 三、设计原则和特点

#### (一) 研究对象随机分配入组，避免选择性偏倚

随机对照试验中，采用随机化的方法制订分配方案，并对分配方案进行隐藏，使合格的研究对象均有同等机会进入试验组或对照组，不以研究人员或研究对象的主观意愿为转移，可避免选择性偏倚的干扰。

#### (二) 增强组间的可比性

在随机对照试验中，采用随机化方法分配研究对象，在足够样本量的情况下，可使若干已知和未知、或可被测量和不能够被测量的可影响疗效或预后的因素，在组间分布中维持相对均衡，从而有利于基线的可比性。如果样本量不太大，不能保证影响预后或疗效的主要因素在组间均匀分布，导致基线不可比的话，这时可采用分层随机分组的方法，保证组间的可比性（参考本书第七章）。

#### (三) 试验对象的特点

用于治疗性随机对照试验的对象，一定是需要进行治疗的，不治疗通常对患者的健康是不利的。对于某种自限性疾病，不需特殊治疗且较短期间就可以痊愈者，显然就不适宜选入为研究对象，因为如将其纳入研究，或许会出现与治疗无关的假阳性反应。

用于病因或危险因素致病效应的随机对照试验的观察对象，在试验开始前，肯定不应患有被该病因或危险因素所致的相关疾病。否则，又可以引出错误的阳性结论。

所有参与随机对照试验的研究对象，根据医德原则，病人应知情并自愿，不应强迫参加。

#### (四) 试验的同步性，条件的一致性

随机对照试验的两组（或多组）对象，均应同步性地开展研究，不能先做试验组，后作对照组，或者相反；而且试验的条件和环境，应保持一致，不能将试验组患者作住院治疗，对照组作门诊治疗，或者相反。因为两组对象的试验观察，在时相上的不同步，环境条件的不一致，显然会影响研究的结果，从而有可能导致错误的结论。因此，随机对照试验，一定强调同步性和环境的一致性。

#### (五) 试验期间的一致性

对试验组和对照组的对象，试验期间应保持一致，这是随机对照试验的又一特点。不能使试验组观察期长于对照组，或者相反，因为两组观察期间不一致，本身就可以造成试验结果的差异，导致研究结论偏离真实性。

#### (六) 研究结果于试验结束时方可获得

随机对照试验系前瞻性研究，试验的结果一定是试验对象接受相应研究措施之后，并经历了一段效应期，方可获得阳性或阴性的结果。

因此，与回顾性研究相反，试验开始时并没有研究的结果。但试验初如出现了试验终点观察的结果者，要注意偏倚因素的影响。例如，应用青霉素和安慰剂对乙型溶血性链球菌感染者预防风湿热发作的随机对照研究，在青霉素组，有的用药后第二天，出现了急性风湿热发作的反应，显然就不能下青霉素预防风湿热无效的结论。因为发病者于试验初就处于风湿热的亚临床期而未被发现，这种试验初出现了试验“结果”系为试验对象选择性偏倚造成的。

#### (七) 保证统计分析结果的真实性

由于采用随机对照原则，获得的资料结果与其他设计方案相比，往往真实可靠，受偏倚因素影响小，使得统计分析有可靠的科学基础。



## 四、设计模式

### (一) 设计模式

随机对照试验的设计模式如图 9-1。试验的研究对象必须采用公认的诊断标准确定，可从患病群体（目标人群）中随机抽样，也可来自住院或门诊的连续性非随机抽样的样本，再根据试验设计中确定的纳入和排除标准，选择符合标准且自愿参加试验的患者，采用明确的随机化方法将合格的研究对象随机分配入试验组或对照组，接受相应的干预措施，经过一段恰当的观察期后，测量治疗后的效果。根据结果的资料类型，采用相应的统计学方法进行分析、处理以评价干预措施的真实疗效及其组间差异。

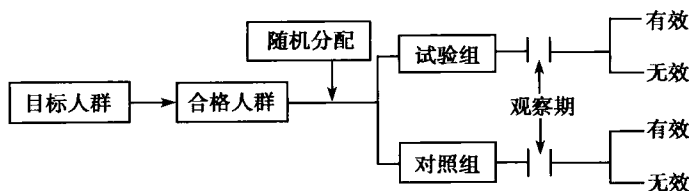


图 9-1 随机对照试验的设计模式

### (二) 结果分析模式

根据图 9-1 的设计模式，可列出四格表（表 9-2），将试验组和对照组的结果分别填入相应的表格内，对两种干预措施的疗效进行分析和比较评价。

表 9-2 随机对照试验结果分析四格表

	结果		合计
	有效	无效	
试验组	a	b	a+b
对照组	c	d	c+d
合 计	a+c	b+d	N

两组疗效比较采用卡方检验。

## 五、随机对照试验的主要优缺点

随机对照研究作为临床医学研究中论证强度最高、设计最佳的设计方案，具有某些优点，也存在某些缺点。

### 1. 优点

(1) 组间可比性好：随机分配研究对象，特别是在某些情况下，将研究对象按影响结果的某些重要因素进行先分层后再随机分配进入试验组和对照组，使组间的基线状况的保持相对一致，增加可比性。

(2) 防止选择性偏倚好：采用随机分配和分配方案的完善隐藏，在选择和分配研究对象时可以较好地防止人为因素的影响，即使存在不为人知的偏倚或混杂因素，也可能维持组间的相对平衡。

(3) 研究对象的诊断确切：对被研究的对象，采用严格、一致的诊断、纳入和排除标准，有利于读者验证研究结果和确定研究结果的推广应用价值。

(4) 盲法衡量和分析结果，结果更真实、可靠。随机对照试验中，如果能够采用盲法衡量研究结果，则可避免研究人员和患者所导致的测量性偏倚对结果的影响，增加结果的



真实性和可靠性。

(5) 高质量的单个 RCT, 可成为系统评价的可靠资源。

## 2. 缺点

(1) 随机对照试验比较费时, 人力与财力支出较大。

(2) 随机对照试验常常有严格的纳入、排除标准, 使入选的研究对象具有良好的同质性, 但也导致其研究结果的代表性和外在的真实性受到一定的局限。

(3) 安慰剂不恰当的应用、对照组措施选择不当, 或让受试对象暴露于某种有害致病危险因素, 则会违背医德的原则。

## 六、结果的统计分析方法

随机对照试验中, 应根据研究的类型(病因或防治性研究)、资料的种类(计数或计量资料)、研究的分组数、资料的分布(正态或非正态分布)、影响研究结果的相关因素等, 抉择相应的统计分析方法, 如卡方检验、秩和检验、t 检验、方差分析及其两两比较、多因素分析、时效分析等。对于病因学研究的随机对照试验, 尚可采用危险度指标如相对危险度、归因危险度、病因分值等评价致病因素或危险因素与疾病发生之间的强度关系; 对于治疗性研究, 尚可采用相对危险度、相对危险度降低率、绝对危险度降低率、预防一例不良事件的发生需要治疗同类疾病患者的例数(number needed to treat, NNT)等表示。

## 七、其他类型的随机对照试验

### (一) 半随机对照试验(quasi-randomized controlled trial)

与随机对照试验设计相似。唯一区别是试验对象分配方式不同。它是按半随机分配方式, 即按试验对象的生日、住院日或住院号等的末尾数字的奇数或偶数, 将试验对象分配到试验组或对照组, 接受各自的试验措施。

半随机对照试验由于分配方式的关系, 容易受选择性偏倚的影响, 造成基线情况的不平衡, 因此, 虽然花费的时间、精力、财力并不亚于随机对照试验, 其结果的真实性和可靠性却不及随机对照试验。

### (二) 非等量随机对照试验(unequal randomization control trial)

指试验对象按一定比例(通常为 2:1 或 3:2)随机分配入试验组或对照组。主要应用于新药疗效验证时, 由于病人来源和研究经费有限而研究者希望尽快获得结果。

但是随着试验组病例数的增多, 检验效能会随之降低, 特别是当试验组的病例比例超过 75% 时, 检验效能明显降低。

### (三) 整群随机对照试验(cluster randomized controlled trial)

多数随机对照试验的随机分配单位为单个的患者, 但在某些特殊情况下, 以单个个体为分配单位是不合适的。

例如, 设计一种预防心血管病的特殊膳食食谱, 拟与普通饮食比较, 观察预防心血管疾病发生的效果。假若一家 4 口人中, 有 3 个被选为合格的试验对象, 其中有可能 1 人分配到特殊膳食组, 2 人分配到普通饮食组。在日常生活中, 一家人不可能长期制做两种不同膳食供用, 即使做到了, 两组对象互换不同饮食的现象, 也不可避免, 于是就会发生沾染和干扰, 影响研究结果。

在上述情况或相类似的有关试验中, 显然单个体不宜作为试验的分配单位。于是, 就可以以一个家庭, 一对夫妇, 一个小组甚至一个乡镇等作为随机分配单位, 将其随机地分配入试验组或对照组, 分别接受相应的措施, 进行研究。此类试验称为整群随机对照



试验。

整群随机对照试验在设计上与一般随机对照试验一样，不同之处在于因随机分配的单位不同，导致样本含量的计算和结果的分析方法有所差异，所需样本含量较大。

#### (四) 单个患者的随机对照试验 (number of one randomized control trial, n-of-1 trial)

对于患慢性疾病的单个患者，采用单个患者的随机对照试验，以确定多种治疗措施中哪一种对其有效，以避免服用多种药物，浪费经费和受某些无效甚至有害药物的影响。

1. 定义 对单个患者，采用多种药物作随机对照试验，以筛选出对该患者有确切疗效的药物。

2. 方法 将所有“有效”的药物与其安慰剂配对，以每对药物为一个单位，采用随机分配的方式决定每对药物的使用顺序；对每对药物，同样以随机分配的方式决定试验药物和安慰剂的使用顺序。依药物疗效发生和达到稳定所需的时间来决定药物的观察期，所有的试验药物观察期应保持一致，以便比较。通常采用双盲法，以利于试验结果的评价。

3. 应用 适用于慢性复发性疾病，如：冠心病的心绞痛，偏头痛，支气管哮喘等，或者患者服用多种药物，但不明确哪一种有效、哪一种无效。

(李 静)

## 第二节 交叉试验

### 一、概述

交叉试验 (cross-over design) 是指试验中的试验组和对照组，在整个试验过程中通过前后两个阶段互相交叉的方式，分别先后接受两种不同试验措施的处理，最后评价试验结果的一种临床试验性研究的设计方案。主要用于临床干预措施的研究和评价，是随机对照试验的一种特殊类型。它兼有随机对照试验和自身前后对照试验的优点，属于一级设计方案。

在试验开始前，符合纳入标准的试验对象先进入观察期，了解研究对象是否处于一个相对稳定的状态，例如在比较和评价新的治疗高血压的方案中，观察期的研究对象的血压已达正常水平，则不宜列入研究对象，以免影响研究结果的判断。经过一定时间观察后，将符合纳入标准的研究对象通过随机的方法分为试验组和对照组，分别接受第一阶段试验，如试验组首先接受方案 A，而对照组接受方案 B，经过一定时间试验观测，并获得相应的结果之后，按设计要求经过“清洗期”(washout period) 然后进入第二阶段的交叉试验。即两组进行交换，试验组换为方案 B，而对照组换为方案 A。按照第一阶段相同的指标，观测第二阶段的治疗反应。当试验结束时，将其结果与第一阶段之结果进行综合分析和评价。

在交叉试验设计时，一定要注意前后两个试验阶段药物清洗期的时间设置，通常以药物的 5 个半衰期为宜，不宜过短或过长，否则易受偏倚干扰的影响。

### 二、应用范围

一般而言，交叉试验仅适用于慢性病、且不易根治并需要药物维持治疗的某些疾病的研究。如高血压病、冠心病心绞痛、支气管哮喘等等。因为这些疾病通常不用药物治疗就会发作。因此对于某些一治即愈者是不能采用交叉试验设计方案的。此外，在新药的开发和研究中，I 期临床试验也常常采用交叉试验来观察药物的毒副反应，以便减少或消除个



体间的偏倚的影响，减少样本含量的需要。

### 三、设计模式及特点

设计模式（图 9-2）：交叉试验通过随机的方法将受试者分成两组，甲组先接受方案 A，乙组先接受方案 B，两组同时进行观察。经过一定清洗期后，甲乙两组再交换接受方案 B 和方案 A。将两个不同时期观察到的结果综合分析并进行比较。

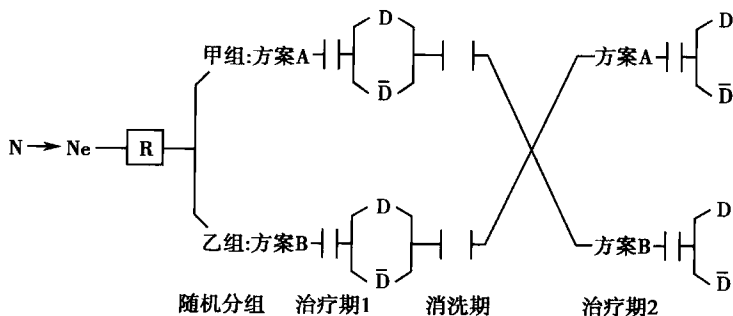


图 9-2 随机分组交叉设计模式

### 四、结果分析

交叉试验中每个受试对象都要接受两种措施的处理，必然得出两种方案的观察结果。由于系自身前后对照，结果分析对定性资料应采用配对卡方检验，而定量资料则采用交叉试验的配对 t-检验和秩和检验（详情参见相关章节及统计学）。根据前后两方案的处理，将观察的结果分别列入配对四格表（表 9-3），即方案 A 和 B 均有效的为 a，均无效的为 d，方案 A 有效而 B 无效的为 b，而方案 A 无效而 B 有效则为 c。

表 9-3 交叉试验结果分析四格表

		方案 B		合计
		有效	无效	
方案 A	有效	a	b	a+b
	无效	c	d	c+d
		a+c	b+d	N

### 五、优缺点

#### （一）优点

1. 交叉试验采用随机分组、盲法测量和同期对照的方法，具有随机对照试验的优点，消除或减少了选择性偏倚、测量性偏倚等误差对试验结果的影响。
2. 每个受试对象先后接受两种方案处理，具有自身前后比较，消除了个体差异，同时也获得组间比较的结果，可减少样本量。

#### （二）缺点

1. 应用范围受到限制，只适用于慢性复发性疾病。
2. 试验过程包括一定洗脱期，如过短难以避免两种措施的重叠作用，过长影响试验周期，甚至使患者长时间得不到应有的治疗，影响病情。
3. 如果试验周期长，容易发生失访、退出、依从性下降等问题。



4. 每个病例在接受第二阶段治疗时, 很难保证病情处于试验第一阶段开始时的相似状态。

(方 芳)

### 第三节 自身前后对照试验

#### 一、概述

自身前后对照试验 (before-after study in the same patient) 系指每一个受试对象, 先后接受试验和对照两种不同措施进行试验研究, 最后将两次先后观测的结果进行比较的一种设计方案。自身前后对照试验是以个体自身为对照, 它可以避免个体差异对结果的影响。在研究过程中, 试验和对照两种措施的先后安排可以是随机的, 也可以是非随机的, 但最佳决策是采用随机方法选择试验措施或对照措施作为第一阶段的试验。如方案 A 随机地选入第一阶段研究试验, 那么, 受试者先接受方案 A 的干预试验, 当完成试验观测任务后, 则停止用药并总结前阶段的试验结果。然后进入清洗期, 清洗期结束后, 更换为方案 B 开始第二阶段的试验研究。同样按照第一阶段方案 A 的测试指标观测相应的结果, 完成后则将前-后两阶段的结果进行分析和比较。

#### 二、应用范围

自身前后对照设计, 每例受试者均要在前-后不同阶段接受试验和对照两种措施, 因此和交叉试验一样仅适用于慢性反复发作疾病的防治性研究。

#### 三、设计模式及特点

自身前后对照试验属于前瞻性研究设计, 符合研究的纳入对象随机或非随机的在第一阶段接受一种措施的试验, 然后经过一定的洗脱期后, 受试者开始接受第二阶段的第二种措施。当完成试验后, 将前-后的试验结果进行分析比较 (图 9-3)。

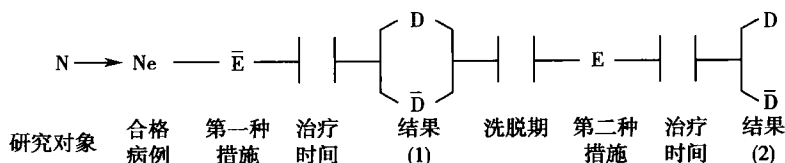


图 9-3 自身前后对照研究设计示意图

#### 四、结果分析

参与自身前后对照研究的患者必须完成两个阶段的研究才能纳入结果分析, 因此每例均有前后两种措施处理后获得的结果。受试对象可能有四种情况 (表 9-4), a 为两种方案均有效, b 为方案 1 无效而方案 2 有效, c 为方案 1 有效而方案 2 无效, d 为两种方案均无效。自身前后对照研究的结果属于配对资料, 故定量资料采用配对  $t$  检验; 而定性资料采用配对  $\chi^2$  检验。



表 9-4 自身前后对照研究结果分析四格表

第二种措施	第一种措施	
	有效	无效
有效	a	b
无效	c	d

## 五、优缺点

### (一) 优点

1. 每例受试者以自身为对照，可以消除个体差异，减少样本含量，节约时间和成本。
2. 研究过程中，每例受试者均有接受试验和对照两种措施的机会，具有公平性。
3. 减少了自愿者偏倚和研究人员意愿偏倚。
4. 可以实现试验措施的标准化。
5. 试验中可以采用盲法，并且可以用随机的方法安排前后的实验措施，提高结果的可信度。

### (二) 缺点

1. 自身前后对照试验分前后两个阶段，所以很难保证两阶段的起始点完全一致，可能影响两个阶段的可比性。
2. 试验的应用范围有限，只适用于慢性复发性疾病。
3. 清洗期过长可能影响患者的及时治疗。

(方 芳)

## 第四节 队列研究

队列研究 (cohort study) 在疾病病因/危险因素、疾病预后乃至非随机前瞻性治疗性的研究中，论证强度较高，所获结果可靠性较强，能颇好地揭示两事件间客观存在的因果关系。故掌握队列研究理论，亦有助于对临床试验的理解。

### 一、概述

#### (一) 概念

队列研究，又名群组研究、定群研究，是将一群研究对象（队列），按是否暴露 (exposure) 于某因素分为暴露组与非暴露组（对照组），并随访适当长时间，比较两组之间所研究疾病（或事件）的发病率（或发生率）、治愈率或死亡率差异，以研究这种（些）疾病（或事件）与暴露因素之间的关系。

“暴露”此术语在流行病学中应用广泛，其词意为“接触”、“受到”、“具有”等。在临床流行病学中，常指接受某种诊疗措施或接触某些致病因子。

#### (二) 特点

队列研究的特点为：①研究开始时，研究对象均无所研究的疾病或事件，需经一段时间随访才能发现病例（或事件结局）；②研究对象按暴露与否分组，其暴露状况已客观存在，研究者不能将其随机化分配；③人群的暴露及其变化由研究者调查与记录；④从因果关系视之，因在前，果在后，由因寻果；⑤可计算两组的发病率、发生率、治愈率、死亡率和相对危险度 (relative risk, RR)。



### (三) 分类

按研究的时间可分:

1. 前瞻性队列研究 (prospective cohort study) 指从“现在”开始的前瞻性的队列研究。本节主要讨论此队列研究。

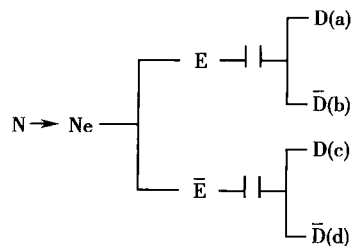
2. 回顾性队列研究 (retrospective cohort study) 是回顾过去的某个时期 (阶段), 调查某人群是否暴露于某种 (些) 因素, 连续追溯至“现在”, 以研究此因素与某疾病 (事件) 之间的因果关系。

3. 双向性队列研究 (ambispective cohort study) 回顾性队列研究 (在时间上) 追溯至“现在”后, 再延续进行下去的研究。

上述三类研究中, 第一、三类队列研究方案科学性较强, 但所需时间长、人力物力多; 第二类研究方案虽较省事, 却偏倚颇多, 故真实性较差。临床研究接触人群面不广, 研究的问题要求明确, 研究对象诊断清楚, 结局指标与目的清晰, 故常采用前瞻性队列研究方案; 但由于目前国内的条件有限, 若进行严密设计、细致分析、谨慎结论, 也可应用回顾性队列研究获得新的发现。如西方的两个前瞻性队列研究发现保留乳房 (保乳) 治疗的早期乳腺癌与接受根治性的患者长期存活率相同, 从而肯定了保乳治疗的临床意义。然而, 我国复旦大学附属肿瘤医院在 1995~2002 年间的 4 038 例乳腺癌患者中, 应用 1:2 配比的回顾性队列研究方法, 规定了纳入、排除标准、配比条件、两组各自的手术方法和结局指标, 选择 127 例为保乳 (暴露) 组, 254 例为根治 (非暴露) 组。随访至 2004 年 11 月, 发现: 保乳组 5 年局部区域复发率为 3.39%, 根治组为 1.84%, 两者无显著差异 ( $P=0.55$ ); 5 年总存活率则分别为 96.73% 和 97.57% ( $P=0.37$ ), 5 年无病存活率分别为 86.04% 和 91.57% ( $P=0.66$ )。由此, 作者推测: 中国人早期乳腺癌保乳手术的长期存活率和根治术的无明显差异, 和西方学者的前瞻性队列研究结果相似。可见, 规范的回顾性队列研究也可取得比较可靠的结果。

## 二、研究设计模式

目标人群确定之后, 设计的主要步骤: 按研究目的与条件, 选择符合设计要求的合格对象全部人群或部分抽样人群, 根据其暴露与否分组, 并随访观察疾病或事件发生状况 (图 9-4)。应注意二点: ①分组为非随机化分配, 因为研究对象为“暴露”与否是事实上已经存在的个体, 及其相应的队列 (定群); ②图中的结局 (outcome), 仅标“发病”与否, 尚可代表“治愈”、“死亡”或“事件发生”等。



N: 研究对象  
Ne: 合格的研究对象  
E: 暴露组  
 $\bar{E}$ : 非暴露组  
D: 发病(事件)  
 $\bar{D}$ : 无病(事件)

## 三、研究对象的选择和资料收集

研究对象的选择和有关资料的收集, 均为复杂而细致的工作, 若有偏倚, 将使研究结果不可靠甚至导致研究失败。注意事项:

图 9-4 队列研究的设计模式

1. 若在社区中进行研究, 则应选择人口相对稳定、便于随访的地区; 所在地区或单位, 医疗卫生条件较好, 能获得可靠的资料, 确保研究质量; 抽样的目标人群应为被研究因素暴露率高的人群, 如研究冠心病和高血脂、高血压、糖尿病的关系, 应选择 40 岁以上人员进行抽样。

2. 应有明确的诊断标准以及纳入和排除标准; 纳入的研究对象确无被研究的疾病





(事件)。

3. 被研究的因素应有明确标准和灵敏度、特异度均较高的检测方法。例如对高胆固醇血症与正常胆固醇水平的定义,标准应该明确;又如 HBV 致肝癌的研究,HBsAg 的实验诊断必须使用酶联免疫试验(ELISA)或放射免疫试验(RIA)。

4. 收集的资料应全面而准确。设计时应制订各类相关的调查表及其调查手册,在正式实施调查前应进行预调查并根据情况对调查表做相应修改;同时还应对调查员进行培训,考核合格后才可上岗。

设计时尚须注意:和被研究疾病或事件以及暴露因素相关的其他主要因素均应全面收集,以排除混杂。如研究脑卒中预后时,调查表不仅应包括一般情况:性别、年龄、婚姻、文化程度,还应有个人行为及疾病既往史:吸烟、饮酒、高血压、冠心病、陈旧性梗和心房纤颤、脑卒中(脑梗死、脑出血、自发性蛛网膜下腔出血)及糖尿病史;不仅应包括临床特征:入院时心率、血压、Glasgow 量表分值、NIHSS 量表分值、OCSP 分型及并发症等,尚应有治疗状况:溶栓、抗凝、神经保护、降纤,还应有实验室和影像学结果(血常规、肝功、肾功、凝血系列、同型半胱氨酸、血糖、血脂、心电图、颈部血管 B 超、影像学),住院天数、本次入院结局等。

5. 对两个组即暴露与非暴露组的设置,临床研究中通常多在同一群体中进行,谓“内对照”;若研究某种环境因素的致病效应,则对照组(非暴露组)应在无该因素的地区或人群选择,谓“外对照”。

6. 结局事件的确诊,应有公认的标准诊断为依据。例如:HBV 导致肝癌,HBV 导致脑血管意外等。

## 四、样本量的估算

估计样本含量之前,必须确定下述参数:

1. 非暴露人群或全人群中被研究疾病的发病率( $P_0$ ) 可通过查阅文献或预调查获得。

2. 暴露人群中的发病率( $P_1$ ) 也可通过查阅文献或预调查获得;或者已知相对危险度,可以用公式  $RR = P_1/P_0$ , 求  $P_1$ 。

3. 相对危险度  $RR$  为暴露人群与非暴露人群中发病率或死亡率之比。 $RR$  亦可通过查阅文献或预调查获得。

4.  $\alpha$  值 第一类错误概率,即假阳性率,也为统计学上的显著性水平。若将  $\alpha$  值定为 0.05,则按此进行研究所确认的某病与病因之间的联系,估计错的概率 $\leq 5\%$ 。

5.  $\beta$  值及把握度(power)  $\beta$  为第二类错误概率,即错误拒绝无效假设的概率,也称假阴性率。把握度 $=1-\beta$ 。若将把握度定为 0.90 ( $\beta$  为 0.10),则有 90%的把握能发现某种病因与疾病之间确实存在的联系。把握度愈高,样本量愈大。为保证研究的可靠性,把握度至少应在 0.80。

估计样本量,常用两种方法:查表法与公式法。具体公式和步骤,请查阅有关书籍。应注意,由于失访等原因,实际观察的样本数应比估计数多 10%~15%。

## 五、资料的分析

对暴露组与非暴露组,在按照设计要求收集资料后,应认真核对,保证资料的真实可靠。然后对资料进行分析。分析格式见表 9-5。



表 9-5 队列研究结果分析用四格表

	有病 (事件)	无病 (事件)	合计
暴露组	a	b	a+b
非暴露组	c	d	c+d
合 计	a+c	b+d	N

根据表中数据可计算有关评价指标:

1. 发病率 (治愈率、死亡率、病死率、并发症发生率等)

暴露组的发病率  $I_1 = a / (a+b) (\%)$

非暴露组的发病率  $I_0 = c / (c+d) (\%)$

对两组 (率) 差异显著性的比较, 可作  $\chi^2$  检验。

2. 相对危险度: 为暴露组和非暴露组两个发病率的比值。

$$RR = [a / (a+b)] \div [c / (c+d)] = I_1 / I_0$$

还应计算 RR 的 95% 可信区间。当 RR 的 95% 可信区间不包括 1.0 时, 则表明被研究因素和被研究疾病 (事件) 相关。RR 大小的意义, 应视具体疾病并参考其临床价值而定。

3. 特异危险度或归因危险度 (attributable risk, AR) 指暴露组的发病率 ( $I_1$ ) 与非暴露组的发病率 ( $I_0$ ) 之差。AR 表示暴露者中完全由某暴露因素所致疾病的发病率。公式如下:

$$AR = I_1 - I_0$$

$$\text{或 } AR = I_0 (RR - 1)$$

4. 特异危险度百分比 (attributable risk percent, AR%) 暴露人群中完全由暴露因素所致发病的比例。即暴露组中排除其他未知因素的作用之后, 完全由暴露所致的发病率 ( $I_1 - I_0$ ) 占暴露组发病率 ( $I_1$ ) 的百分比。以此可评价暴露对疾病发生作用的大小。所以, AR% 又称病因分值 (etiological fraction, EF) 或相对危险增高率 (relative risk increase, RRI)。AR% 主要与 RR 的高低有关。公式如下:

$$AR\% = \frac{I_1 - I_0}{I_1} \times 100\% \text{ 或 } AR\% = \frac{RR - 1}{RR} \times 100\%$$

5. 人群特异危险度百分比 (population attributable risk percent, PAR%) 表示人群中由于暴露 (某因素) 所致的发病率占人群发病率的百分比。PAR% 不仅考虑了暴露的 RR, 而且与某因素在人群中的暴露率 ( $P_e$ ) 有关。PAR% 在领导部门制订预防策略时需加以考虑, 也可用于疾病预防的宣传教育, 它可提示在完全控制该暴露因素后人群中某病发病率可下降的程度。因此, 又称之为人 群病因分值 (population etiological fraction, PEF)。公式如下:

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

## 六、应用范围

队列研究在病因与疾病预后研究中有很重要的应用价值。此外, 亦可应用于疾病治疗性研究。“初治急性白血病 P-糖蛋白及其预后的前瞻性队列研究” 即是一例。多药耐药性 (multidrug resistance, MDR) 是导致白血病化疗最终失败的主要原因, 而 P-糖蛋白 (Pgp) 的过度表达可能是前者发生的主要机制。许多研究已证实 Pgp 在难治性复发性白血病和初发患者中的表达。作者尝试应用队列研究, 进一步了解 Pgp 与 MDR 之间的关系, 并评价 Pgp 预测白血病预后的价值。该研究在上海市三级和二级医院, 收集了 48 例



初治急性白血病 (AL) 的连续性病例, 并按美、英、法 (FAB) 诊断标准和分类法进行了分类: 急性髓细胞性白血病 (AML) 35 例, 急性淋巴细胞性白血病 (ALL) 12 例, 髓、淋混合细胞性白血病 1 例。发现其中 Pgp 阳性者 23 例, 阴性者 25 例。应用标准化疗方案, 结局指标为: 完全缓解 (CR) 和未缓解即耐药 (NR)。经两疗程的治疗, Pgp 阴性者完全缓解 (CR) 率为 68%, 而阳性者仅为 22%; 经单因素 logistic 回归分析, Pgp 阴性者取得 CR 的可能性是阳性者的 7.65 倍 ( $RR=7.65$ , 95%  $CI$  为 2.1~28.0,  $P=0.002$ )。经年龄、性别、WBC、FAB 亚型和骨髓原 (早) 幼细胞比例等变量校正后显示, Pgp 与 CR 仍显著相关 ( $RR=9.3$ , 2.2~39.0,  $P=0.002$ )。此研究虽然所用时间不长, 却在我国验证了 Pgp 阳性与初治急性白血病多重耐药的相关性, 并提示 Pgp 的状态可能可作为后者的一种预测指标, 由此将提高初治急性白血病的治疗水平。

## 七、应用实例: 新生儿黄疸与母亲使用催产素的关系

关于新生儿黄疸是否与母亲临产时使用催产素有关, 从 1972 年起, 即在妇产科领域内引起了不少争论。至 1986 年, 曾有 6 篇论文的研究结果持赞成态度。而另 5 个研究却与之相反。台湾国泰综合医院妇产科陈福荣医师认为, 必须进行各种条件控制严格的前瞻性研究, 以得出可信的结论, 澄清两者之间的关系。陈医生的主要研究设计与结果如下:

### (一) 研究对象与研究方法

#### 1. 纳入标准与排除标准

(1) 纳入标准: 孕妇健康、胎儿足月经阴道生产, 胎儿为单胎头位, 新生儿出生后至少能在婴儿室观察 5 天。

(2) 排除标准: 有溶血性疾病与低体重婴儿。

作者将 1986 年 5 月至 12 月符合上述标准的 206 例孕妇作为研究对象。

#### 2. 分组: 根据产妇的生产情况, 分成三组:

I 组: 60 例, 完全由催产素导产;

II 组: 95 例, 自然生产加催产素催生;

III 组: 51 例, 完全为自然产程, 未用催产素。

3. 给药方式: 静脉点滴给药, 用含电解质的溶液溶解催产素, 催产素浓度为 5U/500ml。

### (二) 结果

1. 三组研究对象的均衡性 作者将三组之间可能影响研究结果的主要变量作了比较 (表 9-6), 结果表明这些影响因素在三组之间均无显著差异。

2. 三组新生儿总血清胆红素水平 比较三组新生儿出生后第一、第三、第五天的总血清胆红素, 用  $t$  检验处理, 未见显著差异 ( $P>0.05$ ) (表 9-7)。

3. 三组新生儿高胆红素血症比例的比较 若将 1 次或多次的血胆红素测值  $\geq 10\text{mg/dl}$  作为高胆红素血症, 则三组新生儿间的高胆红素血症的比例,  $\chi^2$  检验无显著差异 (表 9-8)。若将此“界线”定为  $\geq 12\text{mg/dl}$  或  $\geq 15\text{mg/dl}$ , 三组间仍无统计学差异。

由此, 作者认为: 当生产过程中, 使用的催产素量不多, 且输入的溶液含有电解质时, 即如上述研究的给药方式, 则新生儿高胆红素血症与应用催产素无关。

因为陈医生的设计比较严密、规范, 既符合临床流行病学的要求, 又不脱离临床实际状况, 故虽仅进行了近一年的研究, 却比较圆满地回答了妇产科学界长期争论的一个难题, 值得仿效。

## 八、优缺点

1. 优点 可设立前瞻性的同期对照, 除暴露因素接触与否之外, 各队列中纳入的观



察对象、诊断标准和纳入/排除标准以及观测指标与项目等均可做到标准化；并随时对整个研究过程加以记录，收集的资料可靠性强；能计算发病率与  $RR$ ，因而论证强度较高，易确立两事物间的因果关系。

2. 缺点 费时长，样本量往往较大，需较大投入，易发生失访偏倚等。

表 9-6 三组间主要变量的比较

变 量	I 组	II 组	III 组
性别 (男/女)	28/32	44/51	25/26
婴儿体重 (均数, g)	3 228	3 270	3 286
孕期 (周)	40.1	40.9	39.9
母乳喂养 (%)	60	55	66
Agar 记分 (1min/5min)	8.9/9.6	8.7/9.3	8.8/9.5
器械助产率 (%)	28	35	33
硬膜外麻醉 (例数)	2	2	1
地美露 (例数)	2	3	2
AFP (ng/ml)	126 347	112 478	109 288

表 9-7 三组新生儿的血清总胆红素 (mg/dl, 均数 $\pm$ SD)

时间	I 组	II 组	III 组
第一天	4.9 $\pm$ 1.5	5.1 $\pm$ 1.7	4.9 $\pm$ 1.4
第三天	9.2 $\pm$ 2.2	9.1 $\pm$ 3.0	9.6 $\pm$ 2.5
第五天	9.9 $\pm$ 2.8	9.1 $\pm$ 3.4	10.0 $\pm$ 3.3

表 9-8 三组新生儿高胆红素血症的发生率\*

血清胆红素 (mg/dl)	I 组 (N=60)		II 组 (N=95)		III 组 (N=50)	
	人数	%	人数	%	人数	%
$\geq 10$	39	65.0	56	58.9	26	52.0
$\geq 12$	17	28.3	34	35.8	17	34.0
$\geq 15$	3	5.0	3	3.2	3	6.0

\*  $P>0.05$

(徐德忠)

## 第五节 非随机同期对照试验

### 一、概述

非随机同期对照试验是临床传统采用的一种研究设计。系指试验组和对照组的受试对象不是采用随机的方法分组，而是由病人或医生根据病情及有关因素人为的纳入试验组或对照组，并进行同期的对照试验。

非随机对照试验是前瞻性的研究。常用于比较临床不同干预措施的效果。该试验在研究对象的分组分配上，由于人为的因素，往往会造成试验和对照两组之间在试验前即处于不同的基线状态，缺乏可比性。在研究过程中也难以盲法评价试验结果，造成许多已知和

未知的偏倚影响观测结果的真实性。但在临床实际工作中，有些情况下不适宜做随机对照试验，例如外科手术治疗、急重症病人抢救或贵重药物的选用等。因此，只能根据具体情况将患者分入试验组或对照组。其研究结果的论证强度虽远不及随机对照试验，但是，在尚无随机对照试验结果或不能获得随机对照试验结果的时候，还是应该予以重视的，尤其是对于病例量大的非随机同期对照试验研究，仍有重要价值。不过在分析和评价研究结果的价值及意义的时候，应持审慎的科学态度。

## 二、设计模式

设计模式与结果分析基本同 RCT，仅在纳入对象分组时不采用随机方法。

### (一) 设计模式（图 9-5）

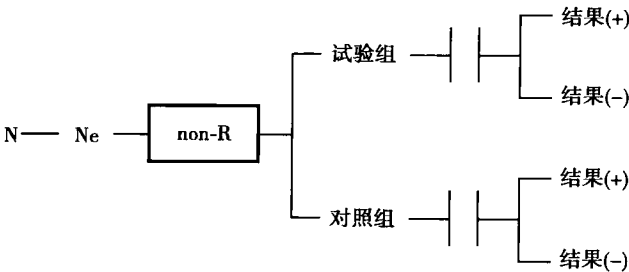


图 9-5 非随机同期对照试验设计模式

### (二) 结果分析（表 9-9）

表 9-9 非随机同期对照试验四格表

	结果		合计
	+	-	
试验组	a	b	a+b
对照组	c	d	c+d
合 计	a+c	b+d	N

## 三、优缺点

研究方案的可行性好，易为临床医生和患者接受，依从性较好；但是由于选择性偏倚和测量性偏倚的影响，使结果的真实性下降，结论的论证强度减弱。

(方 芳)

## 第六节 病例对照研究

病例对照研究（case-control study）是临床回顾性研究最有实用价值的研究设计方案，它具较严格的配对对照，可以在一定程度上防止混杂因素的干扰，对假设的病因及危险因素因果关系，乃至治疗、预后等方面的研究均有重要意义。可为前瞻性研究提供重要信息，从而可进一步验证假设，确定因果关系。随着病例对照研究方法的不断完善和临床科研的需求，其应用范围已日益扩大。



## 一、概述

1. 概念 病例对照研究是一种回顾性研究。其选择所研究疾病或事件的病人作为病例组，无此病（或事件）的病人作为对照组，调查两组对某个（些）因素或防治措施的暴露情况，比较两组间暴露率或暴露水平的差异，以研究该疾病或事件与这个（些）因素或防治措施的关系。若病例组某因素或措施的暴露率或暴露水平明显高于对照组，且研究过程又无明显的偏倚，则该因素或措施与所研究的疾病或事件有联系（图 9-6、表 9-10）。

2. 特点 ①在疾病（事件）发生后进行，已有一批可供选择的病例；②研究对象按疾病（事件）发生与否分成病例组与对照组，分组的方法有两种：成组法（group comparison）和配比法（matching）；③调查的被研究因素或措施的暴露情况，由研究对象从现在对过去的回顾；④从因果关系视之，结果已发生，由果推因；⑤仅能了解两组的暴露率或暴露水平，不能计算发病率。

3. 改良方法 病例对照研究简单易行、十分常用，但又具有一定的局限性，故建立了许多改良方法。其中最具价值的为巢式病例对照研究，将在下面专节论述。

## 二、研究模式

病例对照研究对象的分组不是随机化分配。因为发生疾病（事件）与否已经客观存在，研究者不能主观加以控制（图 9-6）。

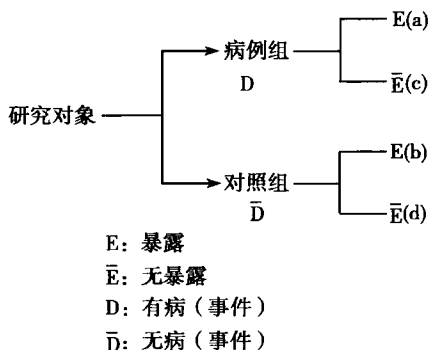


图 9-6 病例对照研究的设计模式

## 三、研究对象的选择和资料收集

研究对象选择的原则有二：第一，具有代表性：病例组应能代表靶人群中患该病的总体；对照组能代表靶人群中未患该病的总体；第二，具有可比性：病例组与对照组在年龄、性别、居住地、社会经济文化等主要特征方面应可比。

1. 病例组的选择 被选择的病例，诊断必须正确可靠，不能将诊断不明或误诊的病例作为“病例组”的研究对象，否则会产生错误分类偏倚而造成低估疾病与暴露因素的关系；其次，被选择的病例，应具有暴露于调查（研究）因素的可能性，否则应予排除。例如，探讨口服避孕药与某些疾病的关系时，对做过绝育术或因其他原因而忌用口服避孕药者则不能选；此外，应纳入新病例作为研究对象。

关于病例组的来源，宜在同一地区不同水平的医院选用一个时期内符合要求的连续性病例，以在一定的程度上防止选择性偏倚的影响。若条件许可，可选择社区总体人群中的全部病例（适于患病率低的疾病）或者从中随机抽样，其代表性更好。

2. 对照组的选择 被选择的“对照”，必须确实排除所研究疾病，否则，也可出现错误分类；其次，“对照组”的研究对象亦应具有暴露于被研究因素之可能性。对照组对象的选择，应与“病例组”同源。

### 3. 病例组与对照组的比较方式

（1）成组法：按和病例可比的原则，选择一定数量的对照。对照与病例的数量不需成严格的比例关系。此法较配比法易于实施，但不易控制混杂因素；

（2）配比法：每一个病例选择一个或几个对照，使病例与对照配成对（pair），而对照在某些重要特征（如年龄、性别等）方面应与其相配的病例相同或基本相同。这些特征称



之为配比因素 (matching factor)。配比使病例组与对照组有可比性, 且可较好地控制混杂因素。

病例与对照的比例, 一般为 1:1, 也可 1:2, 但不超过 1:4。应注意: 被研究的因素不能作为配比因素。配比的因素不应过多, 否则容易发生“配比过度 (overmatching)”, 不仅影响结果的可靠性, 而且浪费精力。

在研究中, 可设多组对照, 如既选医院的病人, 又选亲属或邻居作为对照。这不仅扩大了对照的来源, 而且减少偏倚, 增强代表性; 同时还可研究疾病与被研究因素在不同水平之间的关系或发现另外一些线索。

病例对照研究的资料收集, 也应注意防止信息偏倚, 自调查表的设计、调查员的培训、调查询问的方式方法至资料的核对、复查、录入等均应避免偏倚。

#### 四、研究样本的估算

在病例对照研究的设计中, 对两组样本量的估算, 也需先掌握或估计 5 种参数, 即病例组 and 对照组各自对被研究因素的暴露率; 相对危险度 (RR) 或优势比 (OR); 容许的  $\alpha$  值和  $\beta$  值。根据上述参数, 查阅有关统计学专著, 查表或应用相应的计算公式估算样本量。样本量无法估计时, 则可用经验的累积法处理, 即先调查病例和对照各 100 例, 然后作统计学处理, 若无显著性差异, 则可再增加适当的样本量。

#### 五、资料的整理与分析

资料收集后首先进行核对, 防止误差; 然后按研究对象的各种特征以及暴露因素逐项整理统计, 最后进行卡方检验, 计算 OR 值和其他指标; 若存在混杂因素则应做分层分析, 涉及多因素者则需进行多因素分析。在此基础上, 对被研究因素和疾病的关系作结论。

资料整理后, 按成组法或配比法分别列出下述两种四格表 (表 9-10、表 9-11), 供分析用。成组法四格表的数字比较简单; 配比法则不同。如 1:1 配比病例对照研究, 四格表 (表 9-11) 中的数字 (a, b, c, d) 不是单个病例或对照数, 而是病例偕对照的对子数: 病例和对照均有暴露史的对子数, 填入 a 格, 而均无暴露史的对子数则填入 d 格; 病例无暴露史而对照有暴露史的对子数填入 b 格; 病例组有暴露史而对照组无暴露史者, 则填入 c 格。

表 9-10 成组病例-对照研究结果分析用四格表

暴露史	病例组	对照组	合计
有	a	b	a+b
无	c	d	c+d
合计	a+c	b+d	N

优势比:  $OR=ad/bc$

表 9-11 1:1 配比病例对照研究结果分析用四格表

对照组	病例组	
	有暴露史	无暴露史
有暴露史	a	b
无暴露史	c	d

配比  $\chi^2 = (|b-c| - 1)^2 \div (b+c)$   $OR=c/b$



指标的计算:

1. OR 值 病例对照研究通常不能获得发病(发生)率,故无法计算 RR; OR 为优势比(odds ratio),是相对危险度(RR)的估计值。

成组病例对照研究时,OR 为四格表中 4 个数字的对角乘积比:  $ad/bc$ 。

配比病例对照研究的 OR 值计算较复杂。1:1 配比时的公式: 配比  $OR = c/b$ ; 配比  $\chi^2 = (|b-c| - 1)^2 \div (b+c)$ 。

2. 特异危险度百分比(AR%) 由于队列研究中 AR% 可以 RR 值计算,故病例对照研究中也可用 OR 估计 AR%。公式如下:

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

3. 人群特异危险度百分比(PAR%) 病例对照研究中也可用 OR 估计 PAR%。公式如下:

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

## 六、应用范围

病例对照研究主要用于发病危险因素的研究,但也可用于临床回顾性治疗与探索预后因素的研究等。

1. 病因研究 从 20 世纪中叶开始已有大量有关疾病病因的病例对照研究。著名的,如 Doll 的吸烟与肺癌关系研究;Herbst 的雌激素与阴道腺癌关系研究。

2. 筛检试验效果评价的研究 如国外 20 世纪 80 年代已开始应用病例对照研究,对宫颈涂片检查宫颈癌、乳房照片检查乳腺癌、X 线胸透筛检肺癌等进行评价。

3. 治疗效果的评价 病例对照研究对于发生率很低的某些疾病或事件很适用,因为此时进行 RCT 很难。如 Horwitz 用改良的病例对照研究方法评价了利多卡因控制心肌梗死后心室颤动的作用,解决了研究 30 多年仍无定论的问题。

## 七、应用实例:雌激素与阴道腺癌关系的病例对照研究

美国妇产科医生 Herbst 1966~1969 年在波士顿发现 7 例女青年阴道腺癌,而其分布与以往有很大差异:①阴道癌占女性生殖系统癌的 2%,阴道腺癌仅占阴道癌的 5%~10%,非常罕见;而此 7 例全是腺癌(类型异常);②1930~1965 年间 Massachusetts 总医院和 Pondville 州立医院中仅 2 例阴道腺癌;而 Vincent 医院 4 年间就有 7 例(时间、地区分布异常集中);③既往 2 例年龄均大于 25 岁;而此 7 例全在 15~22 岁之间(年龄分布异常)。这些分布的差异(属于描述性流行病学范畴)提示作者:当时在该地区有某种(些)因素与阴道腺癌异常发病有关。按此假设,Herbst 以病例对照研究方法进行了病因的探索。

Herbst 将此 7 例病人,加上 1969 年另一医院的 1 例 20 岁阴道腺癌患者共 8 例病人作为病例组,病例均经病理确诊;每例病人配 4 个对照,共 32 个对照。对照均选自同医院(病房)出生、出生日期与病例不超过 5 天的女青年。对病例、对照及其母亲的几十种有关因素,由经过训练的家庭访问员用标准调查表进行了调查,结果发现:母亲怀孕期用雌激素(己烯雌酚)治疗与以后她们的女儿发生阴道腺癌的联系最显著( $P < 0.00001$ );母亲怀孕时出血( $P < 0.05$ )和母亲以往流产史( $P < 0.01$ )与疾病发生也有一定联系。然而,以往流产史与怀孕时阴道流血却是应用己烯雌酚的临床指征。因此,作者认为母亲妊





娠早期开始持续服用己烯雌酚显著地增加了其女儿青春期发生阴道腺癌的危险性。又指出：这些病人出生的时间恰符合临床上开始将雌激素广泛用于怀孕的时间。据此研究结果，又鉴于阴道腺癌的严重性，美国食品与药物管理局（FDA）撤消了怀孕妇女使用己烯雌酚的批准书。

Herbst 在描述性流行病学结果的启示下，进行的这个研究，设计合理、计算明确、结论客观，揭示了阴道腺癌的病因，是临床医生用病例对照研究方法进行少见病病因探索的一个范例。

## 八、优缺点

1. 优点 病例对照研究所需样本量小，病例易找，因此工作量小，所需人力、物力较少，易于进行，出结果快，对患者无损害；可以对一种疾病的多种病因或诊断、治疗与预后的相关因素进行研究。

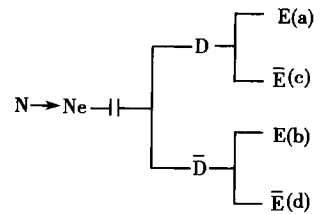
2. 缺点 主要受到回忆偏倚的影响，合理的对照选择又较困难，偏倚可能较大；此外，不能计算发病率，只能推算出优势比，论证强度不如队列研究高。

## 九、巢式病例对照研究

病例对照研究与队列研究作为疾病病因和治疗等观察性研究的主要方法，已为国内外广泛使用。由于此两种方法各有其优势与不足，且两法的优势与不足正好相互补充：病例对照研究的优势是队列研究的不足，前者的不足又是后者的优势，因此，在实践过程中产生了一种扬长避短的新研究类型——巢式病例对照研究。

巢式病例对照研究（nested case control study），又称套式、嵌入式病例对照研究。运作开始时，按队列研究的方式进行：选择一队列，收集基线资料，采集所研究生物学标志的组织或体液标本储存备用，继之随访至能满足病例对照研究样本量的病例数为止。将这些病例作为病例组，按病例进入队列的时间与性别、年龄等配比条件，从同一队列中选择 1 个或数个非病例作对照，抽取病例与对照的基线资料并检测收集的标本，资料按病例对照研究配比法处理，其他的非病例及其资料、标本均可不用。其研究设计模式见图 9-7，结果分析可参照表 9-11 或有关公式。

1992 年 Ross 报道了上海地区肝癌与尿中黄曲霉素生物学标志关系的巢式病例对照研究结果，在 18 244 名中年男性队列中发现 22 例肝癌，每例配 5 或 10 个对照，检测研究开始时的尿样发现，黄曲霉素 B<sub>1</sub> 及其代谢产物和 DNA 加成物的 OR 值为 2.3~6.2，经调整混杂因素后，OR 值为 3.8（1.2~12.2）。作者认为，这是当时关于黄曲霉素作为肝癌致病因素最直接的证据。



N: 研究对象  
Ne: 合格的研究对象  
E: 暴露  
 $\bar{E}$ : 无暴露  
D: 全部发病（事件）病例  
 $\bar{D}$ : 抽取的非病例（事件）

图 9-7 巢式病例对照研究的设计模式

（徐德忠）

## 第七节 横断面研究

横断面研究（cross-sectional study）通常归于描述性研究范畴。但其研究设计，研究对象的选择，研究因素的调查及其结果分析较其他描述性研究严密、规范；对患者（或事



件发生者)与非患者的特征及其影响因素又可以进行某些比较性研究分析,因此,科学性较强。为进一步进行分析性研究(如病例对照研究与队列研究)与实验性研究(如RCT)打下基础,因而在预防医学与临床上应用较广。有学者认为,横断面研究也具有分析性流行病学的成分。

## 一、概述

1. 概念 横断面研究是指某一时点或相当短的时间内对某一人群疾病(或事件)的患病(或发生)状况及其影响(暴露)因素进行的调查分析。由于在短的时间内,如一天、一周或一个月,且调查的是患病现状,因此又称现况研究,或现患率研究(prevalence study)。

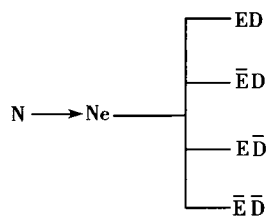
2. 观测指标 需根据不同的目的,选用不同的指标。除暴露因素外,对疾病或事件的频率,如患病率、抗体阳性率、抗原或病毒携带率、其他生物学标志阳性率、疾病诊治率、疫苗接种率、伤残率等,可根据不同的要求,选用不同的指标。

患病率(prevalence),可按调查时间的长短分成时点患病率(point prevalence)与期间患病率(period prevalence)。时点患病率的时间常指某一天(如12月31日)或更短的时间;期间患病率则可以比较长:周、月甚至更长些。调查的病例数(即计算患病率时的分子),应包括调查时点或期间内,新发生的、正在患病的,以及在此期间内发病后病故的全部病例。

3. 分类 现况研究按对象选择的不同方式分成两类:普查与抽样调查。

## 二、研究设计模式

现况研究主要是通过抽样或普查方式,研究目标人群的疾病或事件的发生率及其暴露(包括诊治预防措施)状况,模式如图9-8所示。



N: 研究对象

Ne: 合格的研究对象

E: 暴露

Ē: 无暴露

D: 发病(事件)

D̄: 无疾病(事件)

图9-8 横断面研究的设计模式

## 三、应用范围

现况研究是分析性研究的基础,应用颇为广泛,主要有以下几种。

1. 描述群体中疾病的患病率与感染率、事件发生率;
2. 初步了解与疾病或事件发生的有关因素;
3. 初步评价筛检与诊断试验、治疗措施的效果和疾病预后的影响因素,以及预防措施在人群中的作用;
4. 研究人群中医疗卫生服务的需求及其质量的评价。

如成都市第一人民医院2003年报告的一项现况研究很有价值。为了解成都地区糖代谢障碍发生状况及其相关因素,作者以整群抽样方式调查了市区居民7288例的人口统计学指标、社会经济因素、心血管病危险因素和空腹血糖。结果发现,在成都市区,糖尿病知晓率为22.55%,治疗率17.39%,控制率5.59%;推算血糖升高者约10.6万,糖尿病患者约6.82万;血糖随年龄增高,血糖水平与血压、甘油三酯、胆固醇、体质指数、心率、尿酸明显相关,其中以收缩压及甘油三酯联系更为密切。作者认为,成都人群血糖水平、糖尿病患病率均有增高,而治疗与控制率低;血糖与多种危险因素有伴随关系,与高血压常共存。显而易见,在此研究的基础上,研究者可对糖尿病的发生和流行规律与防治策略以及其对心脑血管病的作用,进行深入的分析性或干预性研究。



## 四、优缺点

1. 优点 容易实施，科学性较强；如设计规范，研究对象代表性好，一次研究可观察多种疾病（事件）的发生状况及多种相关的因素。
2. 缺点 不能得出确切的因果关系；如大规模调查，需投入很多人力物力。

（徐德忠）

## 第八节 叙述性研究

### 一、概述

叙述性研究（descriptive study）是临床医生熟悉和广泛应用的一种研究方法。它是对现成的临床资料进行归纳、分析并得出结论，或对某些临床新出现的疾病病因或表现特征进行描述、总结或报告的一类研究。包括了个病例报告、系列病例分析、专家评述、杂志评论、专家经验和编者的话等等。

与其他设计方案相比，叙述性研究带有更多的主观性，其结论的科学性当然不及专门的研究，但往往反映出作者临床经验积累的水平、对当前学科发展和相关领域研究结果的关注程度。由于缺乏严格的设计，又多为回顾性的研究，并且缺乏严格的对照，在结果的测量与评价中都难以避免各种偏倚因素，故往往影响结果的真实性和重复性。

### 二、应用范围

叙述性研究可用于疾病病因的初步探讨、分析疾病诊断的手段、评价疾病防治措施的效果、判断疾病预后的相关因素等。它常常是分析性研究或试验性研究的前驱，如在SARS发现和研究过程中，从个案的报道、传染病流行病学分析，从而导致病原学研究发现和分离出变异冠状病毒，最终对疫苗的研制以及今后可能投入人群的预防。这里，可以清楚地看到，任何一种未知原因的疾病的发生和发现，常常都是从个案及系列病例报道为始动的，因此，叙述性研究与报道仍然是临床研究不可或缺设计方案。

### 三、设计模式

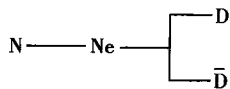
叙述性研究包括的种类很多，难以规范其设计模式。但常见的两类研究是从因到果和从果到因的方案。

#### （一）从因到果的设计模式

从疾病的可能病因观察其致病效应，或通过干预措施观察对疾病的治疗和预防效果。例如，接受化疗的恶性肿瘤病人，观察其化疗的不良反应或并发症的发生率、发生的时间、严重程度和恢复的时间等。

##### 1. 设计模式（图 9-9）

2. 结果分析（表 9-12） 计数资料常采用率表示，并以 95% 的可信区间来表示结果的可信度。计量资料则用均数 ± 标准差表示。



N: 目标人群

Ne: 纳入研究对象

D: 结果阳性者

$\bar{D}$ : 结果阴性者

图 9-9 从因到果的设计模式



表 9-12 从因到果的结果分析

		结果		
		有	无	合计
病因或干预措施	有	a	b	a+b
	无	c	d	c+d

观察结果的指向————→

(二) 从果到因的设计模式

从已获得结果中分析和总结可能的病因或干预措施的效果。例如在肺癌患者中，调查吸烟的历史。从高血压患者并发症分析和评价降压药的作用和效果等。

1. 设计模式（图 9-10）

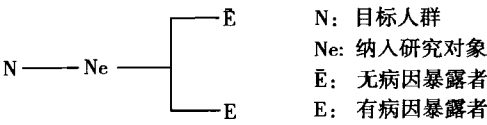


图 9-10 从果到因的设计模式

2. 结果分析（表 9-13） 同前。

表 9-13 从果到因的结果分析

		结果	
		有	无
病因或干预措施	有	a	b
	无	c	d
	合计	a+c	b+d

观察结果的指向←————

四、优缺点

(一) 优点

- 1. 研究容易实施，节省人力财力，短期易出结果；
- 2. 是很多临床分析性研究和试验性研究的基础。

(二) 缺点

- 1. 没有对照组，结果缺乏对比研究的说服力；
- 2. 不能控制偏倚和混杂因素对结果的影响，使研究的结果论证强度弱，重复性差。

(方 芳)

## 第十章 影响研究质量的误差与防止措施

临床医学研究的目的是通过严谨的设计、测量和评价,探讨疾病的病因和危险因素、疾病的发生和发展、疾病的正确诊断、疾病的预后和防治,为临床决策提供科学的证据。临床科研的所有结论都应该以真实可靠的证据为基础。然而,由于各种生物、心理、社会因素的复杂作用,研究结果的真实性和可靠性不可避免地受到影响。为了获得真实可靠的临床研究结果,就应该在研究的设计、实施和分析的每一环节中认识、控制和消除导致错误结论的各种误差。

### 第一节 概 述

#### 一、系统误差和随机误差

##### (一) 误差

在任何研究中,尤其是以人为研究对象的临床流行病学研究中,无论设计如何严谨和具有科学性,总会有误差存在,这是因发生在不同水平的变异(variability)所致。在个体水平上有个体变异(individual variability)和测量变异(measurement variability),在群体水平上有个体间遗传变异(genetic variability between individuals)、环境变异(environmental variability)和测量变异,而在抽样水平上可以在抽样方法(manner of sampling)、样本量(sample size)和测量方法等环节中发生变异。在测量变异中有随机误差(random error)和系统误差(systematic error),前者又称为机遇(chance),后者又称为偏倚(bias)。

##### (二) 随机误差

随机误差是由某事件发生的概率造成的测量值与真实值之间的差异,是随机性的,通常造成的差异在真实值的两侧波动,并且是较小的误差;可以通过统计学中计算标准误的方法估计其大小,并可以通过扩大样本量使其减小,但不可能完全消除。

##### (三) 系统误差

系统误差是人为造成的错误,是由于研究设计、实施和分析不正确造成的测量值与真实值之间的差异,是非随机性的,通常偏离真实值的一侧,且一旦出现均是较大的误差。若系统误差,即研究中的偏倚能够被识别和纠正,其影响基本可以消除。

以图 10-1 可以说明随机误差和系统误差的特性。A 靶相当于一项研究或测量所得的结果既无系统误差的影响,又无随机误差的影响;C 靶相当于研究结果有系统误差(偏倚),但随机误差较小;B 靶相当于研究结果的系统误差小,但随机误差较大;D 靶相当于研究结果既有较大的系统误差,又有较大的随机误差。

在研究的设计、实施、分析和推断过程中存在的各种对暴露因素与疾病关系的系统误差和随机误差可以血压测量的结果加以说明(图 10-2)。采用动脉内插管所测量的血压的确是血压的真实值,但仅有在少数情况下可以采用该方法测量血压,而在大多数情况下均采用血压计测量血压,因此不可避免地产生了系统误差,使测量值均偏离在真实值的一侧;当采用血压计多次测量体表血压时,尽管血压计和测量方法没有错误,但难以获得完全一致的结果,这就是随机误差的影响。

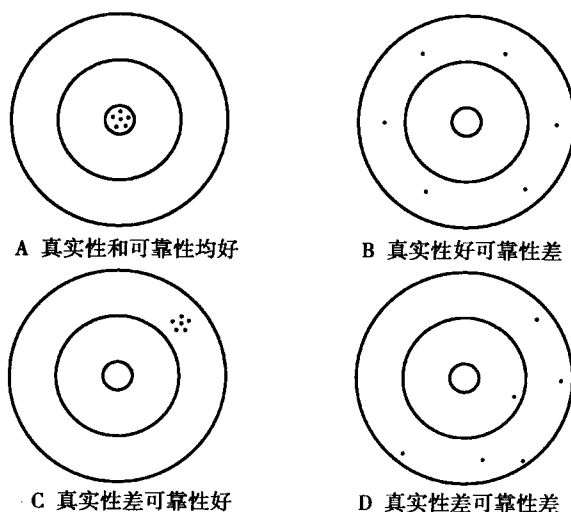


图 10-1 真实性和可靠性的特点

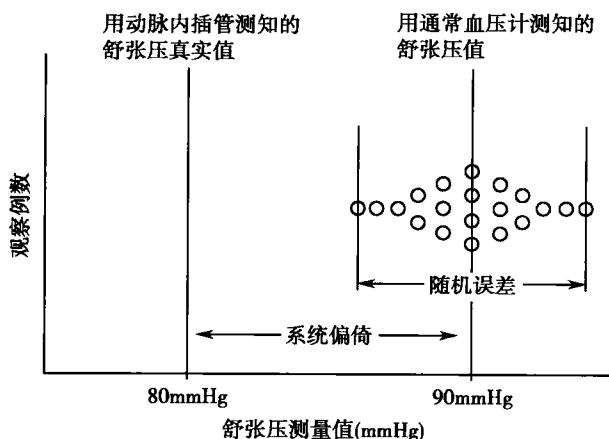


图 10-2 系统误差和随机误差间的关系

#### (四) 偏倚

由系统误差所致的错误称为偏倚，它是人为造成的误差，可以通过一定的手段加以防止、控制或消除。为了保证研究结果的真实性，首先必须对偏倚因素加以识别。

例如，用某种新进口的西药与某种活血化瘀中药治疗冠心病，结果显示西药组疗效优于中药组。评价这项研究的结论，认为有两种可能性：一种可能该结论是真实的结果；另一种可能是偏倚所致。如果西药组比中药组病情轻，那么分组时带来的选择偏倚会导致两组疗效的差异。如果西药服用剂量小，且无特殊味道，而使该组对象比“中药组”能按时服用，结果研究对象的依从性偏倚使两组的疗效产生差异。如果西药是风行一时的新药，而中药是人人皆知的老药，医生和患者都倾向认为新的西药疗效好，结果在观察和测量疗效时，就会因主观因素造成信息偏倚。此外，如果西药组的研究对象多为青壮年，而中药组的研究对象多为老年人，而年龄与冠心病及疗效均有关，就会带来混杂偏倚。由此可见，偏倚可以产生于研究过程的任何一个环节，从研究设计到实施，以及最后的资料分析和结论推导，都可能出现偏倚。在描述性和分析性流行病学研究中都不可避免地存在偏倚的影响，无论是探索病因的研究，还是临床疗效观察，以及诊断试验，都普遍存在偏倚的



影响。

偏倚是造成歪曲研究结果真实性的主要原因，其作用主要是降低真实性。偏倚使研究结果偏离真实值的大小和方向取决于偏倚的特点和严重程度，有些偏倚因素会夸大结果的真实性，有些偏倚因素则会减小结果的真实性，不同方向的偏倚因素都会使研究结果失去临床价值。目前，国内外对偏倚尚无统一的分类，在各种临床流行病学书籍和文献中，已总结出几十余种偏倚，其分类方法主要是描述性的。本章采用 Meittinen (1976) 提出的建议，将临床实践和科研中所出现的偏倚按其性质和产生的阶段归纳为三大类，即选择性偏倚 (selection bias)、信息偏倚 (information bias) 和混杂偏倚 (confounding bias)。选择偏倚主要在研究设计阶段产生，信息偏倚主要在研究实施阶段出现，而混杂偏倚主要是由于设计和资料分析阶段对混杂因素未加以控制和消除而影响研究结果的真实性。

## 二、真实性和可靠性

### (一) 真实性 (validity)

真实性又称为效度或准确性，是反映一项研究或测量能够获得正确结论的程度。要使一项研究或测量产生真实的结果，就要控制和防止误差。消除系统误差的影响，就可以提高临床医学研究的真实性，

从科研结论的效果来看，真实性可分为内部真实性 (internal validity) 和外部真实性 (external validity)。内部真实性是一项研究的结果可以正确反映研究对象所代表的人群真实状况的外延性，即科研结论可以推论到目标人群或称靶人群的范围。如果在研究人群中结果是不真实的，那就有理由怀疑该结论是否可以应用到其他人群。内部真实性是由研究本身的范围所决定的。因此，如果一项研究能够提供在限定的研究人群中作用的真实估计，它就具有内部真实性。提高内部真实性的方法常包括限定研究对象的类型和进行研究的环境条件，以减少外部因素的影响。外部真实性是一项分析结论能够应用到其他人群的外延性。在严格控制的环境中所获得的结果可能不适用于一些状况，例如老年人、病情严重者，或是与研究对象有不同人口学特征的病人。临床医生对外部真实性常常有特殊的兴趣，因为他们必须确定一项研究发现是否能够适用于临床实践。确定一项研究结论是否能够进行一般化的推论，需要判断调查中所包括的研究对象类型、临床医生所诊治的病人类型，以及研究人群与其他人群临床特点的差异。

科研结论的内部真实性和外部真实性均需要认真考察，以确定该结论是否可以推论到靶人群以外的其他人群中，例如，美国进行了一项声誉很高的退伍军人高血压病治疗效果的临床试验研究。研究对象是 143 例 30~73 岁退伍军人，平均舒张压 115~129mmHg，均无高血压并发症。研究设计严谨，采取随机分组，双盲观察分析，研究对象自始至终配合很好。研究结果显示研究组对病人采用双氢克尿塞和利血平联合用药能明显降低心、脑、肾等并发症。经统计学处理研究组与对照组差异有显著性。研究过程采取了各种措施防止了偏倚发生，因而公认该研究结论的内部真实性是较好的。但是这一药物对女性和非退伍军人是否均有效，对舒张压高于 130mmHg 和已有高血压并发症的病人是否也有效，该研究是无法解决这些问题的。因此，评价临床科研的质量，需要考查其内部真实性和外部真实性。

### (二) 可靠性 (reliability)

可靠性又称为信度或精确性，是一项研究或测量能够多次获得同一结论的稳定程度，与样本量的大小或测量的次数有关。样本量越大，或测量次数越多，其平均值越趋向于真实值，可靠性也就越好。可靠性受随机误差的制约，减小随机误差，就可以提高结果的可靠性。



图 10-1 表示真实性和可靠性的特性。一项研究或测量的结果可比作射靶, 箭中靶心的结果相当于一项研究或测量所得的结果既有很好的真实性, 也有很好的可靠性, 如 A 靶。如果不能瞄准靶心而偏向一方, 即系统地偏离靶心射击, 虽然偏向一方, 但多次射击的靶点几乎都集中于一点附近, 即相当于研究结果虽然无满意的真实性, 但有很好的可靠性, 如 C 靶。反之, 射击时靶心瞄得很准, 但各次射击的靶点十分分散, 远离靶心或左或右, 或上或下, 这相当于研究结果真实性尚可, 但可靠性很差, 如 B 靶。最后一种情况是既未瞄准靶心, 各次射击又十分分散, 即相当于研究结果既无满意的真实性, 也无满意的可靠性, 如 D 靶。

## 第二节 选择偏倚

选择偏倚是由于不正确地选择了研究对象组成试验组和对照组等不同的比较组, 使从研究开始的时候, 各比较组研究对象就存在除研究因素以外的其他因素分布的不均衡性, 即选择出的研究对象或样本人群与其代表的总体间的某些特征具有系统的差别, 因而导致研究结果与真实情况之间产生差异。在各种临床流行病学研究设计中都可能产生选择偏倚, 尤其是在病例对照研究和临床试验中较为常见。根据选择偏倚产生的原因, 国内外临床流行病学家归纳和描述了多种情形, 以利于对其作用和影响充分认识。

### 一、类型

#### (一) 入院率偏倚 (admission rate bias)

又称伯克森偏倚 (Berkson bias)。Berkson 等于 1946 年首先对此类偏倚进行了描述。在利用医院资料进行病例对照研究探索病因时, 由于疾病症状的严重程度、患者就医的条件、人群对某一疾病的认识程度, 以及医疗保健制度和社会文化经济等诸多因素的差异, 使病人出现了不同的住院率, 因而可能会夸大或掩盖某因素与某疾病的真实联系, 即由于具有某因素和不具有某因素的患者入院率不同, 导致该因素与研究疾病形成虚假联系。下面用 Lilienfeld 的模拟例子加以说明。

假定某人群中患 A 病和 B 病者各为 1000 人, 患 A 病和 B 病的人群中各有 20% 的人具有某因素 X, 而 A 病和 B 病与 X 因素并无真正的因果联系, 如表 10-1。

表 10-1 人群 A 病、B 病患者数及 X 因素频率

疾病	有 X 因素者	无 X 因素者	总人数	X 因素频率%
B (病例)	200	800	1000	20
A (对照)	200	800	1000	20

设患有 A 病和 B 病的入院率分别为 50% 和 20%, 因有 X 因素者的入院率为 40%, 则实际入院人数应为:

(1) 患 A 病又具有 X 因素的 200 人中, 因 A 病入院率为 50%, 则入院人数为 100 人, 余下的 100 人中 40% 的患者因具有 X 因素而入院, 即入院人数为 40 人, 二者合计入院人数为 140 人。

(2) 患 A 病而不具有 X 因素的 800 人中, 因 A 病而入院人数为 400 人。

(3) 按照同样原理计算, 因 B 病具有 X 因素的 200 人中入院人数为 104 人。

(4) 患 B 病而不具有 X 因素的 800 人中入院人数 160 人。

将上述各类入院人数列入表 10-2。





表 10-2 不同入院率入院患者与 X 因素的关系

疾病	有 X 因素者	无 X 因素者	总人数	X 因素频率%
B (病例)	104	160	264	39.4
A (对照)	140	400	540	25.9

$$\chi^2 = 15.215 \quad df = 1 \quad P < 0.01$$

计算 B 病与 X 因素的关联强度比值比 OR 为:

$$OR = (104 \times 400) / (140 \times 160) = 1.86$$

由此可见, 由于不同的入院率而使疾病与某特征因素之间造成虚假关联, 即 B 疾病与 X 因素并不存在真实联系, 而是 A 病、B 病及 X 因素引起的入院率差异, 使以医院患者作为病例对照研究的对象来源所获的研究结果产生了偏倚, 即入院率偏倚。

下面以一组实例证实入院偏倚的实际作用。1977 年加拿大的临床流行病学家将某地区的几种疾病与可疑药物的关系进行了分析, 分别计算以社区人群为基础的和以医院病人为基础的病例对照研究的比值比, 两者之间差异有显著性, 以此说明入院率偏倚给结果带来的误差。从理论上讲, 以社区人群为基础的病例对照研究的结果优于以医院病人为基础的病例对照研究的结果 (表 10-3)。

表 10-3 两类调查对象估计药物与疾病关联强度的比较

可疑药物	临床疾病	比值比 (OR)		P
		以社区为基础	以医院为基础	
水杨酸类	过敏	1.15	0.18	0.02
水杨酸类	疲乏	2.09	0.72	0.02
缓泻药	运动骨骼系疾病	1.53	5.07	0.06
缓泻药	关节炎类风湿病	1.48	5.00	0.01
安眠药	循环系疾病	6.38	3.72	0.32
维生素类	过敏	1.76	0.00	0.01
维生素类	外伤	0.61	1.92	0.11
心脏病类	循环系疾病	30.65	19.17	0.47
心脏病类	关节炎类风湿病	3.46	49.92	<0.01

(David Sackett, 1977)

## (二) 错误分类偏倚 (misclassification bias)

由于疾病的诊断标准不明确, 或者诊断方法不完善, 结果是错误地将病例判断为非病例而归类于对照组, 将非病例判断为病例而归类于病例组, 产生错误分类偏倚, 影响结果的真实性。

例如, 进行人格障碍的病例对照研究, 确定病例采用了筛查量表和诊断量表两种工具, 而确定对照由于条件限制仅采用了筛查量表一种工具, 结果对照组的对象中就可能存在漏诊的假阴性病例, 从而发生误将病例判断为非病例的错误分类偏倚, 影响结果的真实性。

## (三) 检出征候偏倚 (detection sign bias)

亦称为揭露伪装偏倚 (unmasking bias)。当某一因素与某种疾病无因果联系, 但因该因素能促使类似该病的症状或体征出现, 因而具有该症状或体征的患者急于求医, 结果接受检查的机会增加, 使其中患该病的病人检出率被人为地提高了, 导致该因素与疾病的虚假因果联系的错误结论。



例如，1975 年 Ziel 等人进行病例对照研究来探讨服用雌激素与发生子宫内膜癌之间的关联。研究者从美国某州的 110 万妇女的调查结果得出两者之间存在着高度关联的结论，即服用雌激素是子宫内膜癌的危险因素。随后又有几位研究者得出了类似的结果。然而到 1979 年，Horwitz 和 Feintein 提出质疑，认为这是一种由检出征候偏倚导致的虚假联系。他们指出，服用雌激素会刺激子宫内膜增生，易发生子宫出血，因而无症状子宫内膜癌患者因服用雌激素出现子宫出血，以致其求医和接受检查的机会增加，从而早期子宫内膜癌患者就可以在检查子宫出血的过程中被发现。而那些虽然患子宫内膜癌，但从未服用过雌激素的患者，因无子宫出血的症状而未能及时就诊接受检查，其被诊断的机会就很少。由此，子宫内膜癌病例中服用雌激素的比重被人为夸大了（表 10-4）。

表 10-4 子宫内膜癌与雌激素关联的病例对照研究

服用雌激素史	子宫内膜癌病例	对照
有	54	29
无	40	159
合计	94	188

$OR = (54 \times 159) / (40 \times 29) = 7.40$

为此，Horwitz 等人进行了深入的研究，调查同一所医院肿瘤科和妇科中子宫内膜癌的病例，发现服用雌激素的患者中有 79% 为第一期患者，而在未服用雌激素的患者中只有 58% 为第一期患者。由此证明，由于服用雌激素引起子宫出血，促使患者尽快就诊检查，则早期患者被诊断的机会就大于无子宫出血症状的患者。这就是检出征候偏倚带来的疾病与因素的虚假联系。

（四）无应答偏倚（non-respondent bias）

在临床流行病学调查中，那些因各种原因不回答所提出问题的人称为无应答者，无应答者可能在某些重要的特征或暴露方面与应答者有区别。若无应答者超过一定的比例，将会影响研究结果的真实性和可靠性，这就是无应答偏倚。

Taylor 于 1966 年报告了一项研究以说明这种偏倚。研究者调查和检查了美国西北部铁路系统工作的男职工冠心病的患病分布情况。虽然采取了许多措施要求和鼓励全体职工参加该项研究，但仍然只有占规定年龄组的 73.6% 的职员和 58.2% 的扳道工参加该项目。其他都不愿意参加。初步调查结果是职员冠心病的现患率为 43%，扳道工为 24%，二者的差异有统计学意义，六年后研究人员重新检查了上述员工的健康记录，包括死亡者的死因，并同时得到了六年前参加和未参加冠心病患病率调查研究项目的员工健康资料。经过分析表明，上次调查时部分患冠心病的扳道工因害怕由于疾病被解雇而未参加，因此证明职员与扳道工冠心病现患率的差异是由于无应答偏倚所致的，是人为的系统误差对结果的影响。

造成无应答的原因是多种多样的，有研究表明年龄大，文化水平低的对象应答率比较低。国外有学者报道在一次医师的信访调查中，年龄大于 65 岁的医师无应答率为 20%。一般说来，城市居民应答率高于农村居民；关心个人健康的人乐于参加调查研究并愿意接受检查，而对自己健康不重视、或躯体及精神状况不佳的人则不愿意或不主动参加这种研究工作。此外，在不同的疾病各类患者中会出现复杂的无应答偏倚，如有学者报道尘肺的现患率在最愿意接受检查的人中较高，结核病现患率则在不愿意接受检查的人中较高。这种现象常见于癌症、慢性呼吸道疾病、冠心病等慢性病，即那些最易患病的人不十分愿意接受调查和检查，其患病水平则有别于真实情况。

为了满足真实性的标准，Alderson 认为应答率要超过 90%，Sackett 提出应答率不应



低于 80%，因此，公认的应答率最低限应为 80%。

### (五) 志愿者偏倚 (volunteer bias)

与无应答偏倚相对应的是，来自特殊群体的志愿者，其心理因素和躯体状况与非志愿者有差别，且对研究的依从性可能优于一般人群，以该类人群的样本作为研究对象所获得的暴露结局会明显不同于非志愿者，由此影响结果的真实性，就称为志愿者偏倚。

例如，在一次预防冠心病的研究中，志愿者非常关注自我健康，主动进行体育锻炼，并采用低脂饮食，而非志愿者多是社会经济和健康状况较差的人群，其中许多人患慢性病或酒瘾，显而易见，志愿者与非志愿者二组人群存在明显差别，由他们组成不同的观察组，其研究结果无疑会产生偏倚。

### (六) 失访偏倚 (withdraw bias/loss of follow-up)

是无应答的另一种表现形式。在队列研究和干预性研究中，由于观察时间较长，研究对象中有人不能坚持而退出队列，有人迁居，有人死亡，有人因药物副作用而停止治疗，当观察终止时能够分析结果的人数远少于进入观察时的人数，这种现象对研究结果的影响，称为失访偏倚。

现在以假设的例子说明失访偏倚，如表 10-5A、表 10-5B 所示。

表 10-5A 行为型别与冠心病的关联研究\*

对象	患冠心病	未患冠心病	合计
A 型行为组	40	160	200
B 型行为组	20	180	200

$RR=2.0$  \* 为假设的真值

表 10-5B 失访对行为型别与冠心病关联的影响

对象	患冠心病	未患冠心病	合计
A 型行为组	32 (40-8)	144 (160-16)	176 (200-24)
B 型行为组	18 (20-2)	162 (180-18)	180 (200-20)

$RR=1.8$

由上表所见，若观察全部队列，A 型行为者患冠心病的相对危险度为 2 (表 10-5A)。而当 A 型行为冠心病患者的失访率为 20%，A 型行为者总失访率为 12%，同时 B 型行为冠心病患者的失访率为 10%，B 型行为者总失访率为 10% 的情况下，则相对危险度下降到 1.8 (表 10-5B)，这说明不同的失访率歪曲了行为型与冠心病的关联，即研究结论的真实性由于失访偏倚而降低。

### (七) 现患-新病例偏倚

又称奈曼偏倚 (Neyman bias) 或患病率及发病率偏倚 (prevalence-incidence bias)，在病例对照和现况研究中，研究对象一般是某病的幸存者，而不包括死亡病例、轻型和亚临床型病例，以及痊愈者，尤其是较严重的致死性疾病因病程较短，会同时存在轻型和无症状的亚临床型的患者。而在队列研究中，可以观察到各种临床类型的病例。因此，前类研究选择的是现患病例，后类研究选择的是新发病例，两类病例疾病状况肯定会有差别，所得到的某因素与某病的关联就会出现偏倚，即为患病率及发病率偏倚。

现患病例与新发病例的差别还体现在另一种情形，即某病的幸存者因疾病或其他因素改变了生活习惯，从而降低了某危险因素暴露水平或在调查时夸大或缩小了病前生活习惯上的某些特征，若进行现况研究或病例对照研究，选择现患病例为研究对象，所获得的某危险因素的暴露就可能不代表过去的情况。而如果进行队列研究，由于是客观地观察暴露



于危险因素的频率和程度，故能代表观察期新发病例的暴露情况，这两类病例所得到的某因素与某疾病的关联必然有差异。

例如，Friedman 等人以高胆固醇水平与冠心病关联的实例来说明患病率与发病率偏倚。在队列研究中，高胆固醇患冠心病的相对危险度为 2.4；而在病例对照研究中，高胆固醇者患冠心病的比值比为 1.16。进一步分析发现，在队列研究中，高胆固醇者占冠心病新发病例的 42.3%。而在病例对照研究中，高胆固醇者只占冠心病现患病例的 25.1%。这说明冠心病的存活病例改变了病前生活习惯，故根据他们的观察结果所获得的胆固醇水平与冠心病无明显关联的结论就是不真实的。

### (八) 易感性偏倚 (susceptibility bias)

在队列研究和临床试验中，观察对象可能因为各种主客观因素不同，暴露于危险因素的概率不同，使得各比较组对所研究疾病的易感性有差异，这就可能夸大或缩小了暴露因素与疾病的关联强度，产生易感性偏倚。此外，若两个观察组处于同一疾病的不同阶段或不同的临床类型，这两组治疗效果的差异就不一定是药物疗效的差异，而是易感性偏倚造成的系统误差。

例如，1971 年 Stern 等报告，在口服避孕药的妇女中，子宫颈发育异常的频率比使用避孕工具的妇女高很多，该结果提示口服避孕药是子宫颈异常发育的危险因素。而深入调查以后发现，子宫颈发育异常的妇女在选择避孕方法时愿意选择对宫颈影响小的口服避孕药。由此可见若选择不同特征的对象进行比较，易感性偏倚就会导致因果关系倒置的错误结论。

另一例证是，若观察麻疹在成人和婴儿中的发病率，由于婴儿普遍对麻疹易感，故无论观察何种暴露因素，都会出现婴儿组发病率高的现象，这是典型的易感性偏倚造成的误差，有时会被忽视。

### (九) 成员偏倚 (membership bias)

与易感性偏倚有类似之处的成员偏倚是由于组成观察组的成员与一般人群在各方面，尤其是健康状况有差异所导致的系统误差。

例如，加拿大有学者进行了体育锻炼与心肌梗死关系的队列研究，得到参加规律的体育锻炼能减少心肌梗死复发的结果，如表 10-6 所示。

表 10-6 体育锻炼与心肌梗死复发的队列研究

参加规律体育锻炼	心肌梗死复发	无心梗死复发	合计
是	7	59	66
否	18	46	64
	25	105	130

$RR=0.38$

之后，美国有学者进行了体育锻炼与心肌梗死的随机对照试验，否定了二者之间的关联，见表 10-7。

表 10-7 体育锻炼与心肌梗死复发的随机对照研究

随机参加耐力训练	心肌梗死复发	无心梗死复发	合计
是	28	359	387
否	21	345	366
	49	704	753

$RR=1.26$



经过对两种研究的分析,发现在队列研究中因不能控制参加体育锻炼的条件,故健康状况较好的对象更愿意参加体育锻炼。而在随机对照试验中,用随机的方法决定对象是否参加耐力训练,由于成员偏倚所致的体育锻炼对心肌梗死复发有保护作用的错误结果就不存在了。

#### (十) 排除偏倚 (exclusive bias)

未按照同样的原则或比例从研究组和对照组中排除某些研究对象所引起的偏倚称为排除偏倚。

例如,在一项抗高血压药利血平与乳腺癌关联的研究显示两者存在因果关系。在考察对象选择方法时发现,对照组排除了心血管疾病的患者,而病例组并未排除,结果产生了排除偏倚,导致了利血平与乳腺癌的虚假联系。在之后的一项研究中避免了排除对象的不一致性,结果证明利血平与乳腺癌并无因果关联。

#### (十一) 非同期对照偏倚 (non contemporary bias)

随着医学的发展,疾病的定义、诊断标准、临床表现、治疗方法,以及疾病的危险因素均会随着时间的推移而发生变化。若研究中使用了不同期的病例作为对照,它们之间的不可比性就会产生系统误差,称为非同期对照偏倚。

例如,当前用口服红霉素治疗猩红热,其治愈率优于四十年前青霉素等的综合治疗,并发症也很少。而事实上这种现象并非由于红霉素的疗效真正优于青霉素,而是因为猩红热的病原体乙型溶血性链球菌的毒力减弱,患者病情也随之减轻,并发症亦减少。若不考虑不同时期疾病的变化而直接进行比较,就会因非同期对照偏倚而影响结果的真实性。

#### (十二) 迁移性偏倚 (migration bias)

在队列研究或干预研究中,当病人以原来所在的队列或观察组迁移到另一队列或观察组,称为迁移偏倚。造成这种偏倚的原因是在队列研究和干预试验中,不可避免地发生一些成员退出原队列,进入另一队列。两队列或观察组迁移出成员的数目是非随机不均衡的,这样就会破坏原来设计精良的队列或观察组的结构均衡性,降低二者的可比性,从而影响最终的观察结果。

例如,一项劳动强度与心血管疾病关系的队列研究,近 4000 名码头工人按劳动强度分为高低劳动强度两组进行观察,追踪观察 22 年比较两组的冠心病发病率。但是,在观察期间,高劳动强度的码头工人自动向低劳动强度的职业迁移,若直接将两组进行比较,必然会受到迁移偏倚的影响。为避免迁移偏倚的作用,研究者对研究对象重新分组,调整其他因素,并每年定期检查一次,最后得出了较为真实的结论:低劳动强度工人的心力衰竭发病率为高劳动强度工人的 2 倍。

#### (十三) 诊断机会偏倚 (diagnostic enter bias)

该偏倚与入院率偏倚相类似,即罹患某种疾病的病人,由于各方面的原因,如疾病的严重程度、经济状况、医疗条件、地理条件等的差异,使得获得诊断的机会不同。例如,边远山区的农民,由于医疗条件差,交通不便及经济水平低下,他们获得诊断的机会与城市居民有明显差别,其疾病的早期诊断与治疗的机会无疑是不同的,若二者进行比较,则诊断机会偏倚会影响研究结果。

以上各类偏倚都是国内外流行病学家总结和加以描述的,其共同特点是在对象选择时出现了系统误差,人为地夸大或减小了研究因素及研究疾病的关联,是研究设计阶段经常出现的错误。深刻地了解这些偏倚的特点和产生的原因,有利于加以控制和防止。

## 二、控制措施

选择性偏倚主要在设计阶段出现,因此,为了避免和减少产生这类偏倚的可能性,应



慎重进行研究设计。具体采用以下几种方法：

### （一）随机化

为尽量使两个比较组除研究因素以外其他各种条件都保持均衡，应采取随机分配的方法，对研究对象进行随机分组。可用单纯随机化（simple randomization）、区组随机化（block randomization）和分层随机化（stratified randomization）的方法使每个研究对象进入各比较组的机会均等。将不同病情、不同特征的研究对象均衡地分配在各比较组中，就可防止选择偏倚。

### （二）设立对照

在临床试验中，可设立两个或多个对照组，其中之一应来自一般人群，其他对照组可以来自医院，这样既可以代表社区一般人群，又可以代表医院内不同类型的病人。然后对试验组 and 不同对照组的主要基线状况进行比较，以判断是否有选择偏倚存在。如果采用各种对照组所获得的结果无明显差别，即可表明选择偏倚存在的可能性比较小。同时还要注意考察不同对照组获得相似的结果是否会由于各对照组的选 择偏倚程度相同所致，以免影响结果的真实性。

### （三）严格诊断标准

在设计阶段应明确研究对象的人组标准和排除标准，尽量选用国内外一致公认的诊断标准，并根据纳入/排除标准选择研究对象。在研究实施阶段，要严格遵守，不能轻易改动。否则，要影响入选对象，造成对研究真实性的影响。

### （四）提高应答率

在临床研究中应该采取各种措施鼓励应答，尽量提高应答率，防止或减少失访，减少选择性偏倚。如果出现了无应答或失访，要针对产生的原因采取补救措施。如果无应答率或失访率超过 10%，研究结果的推论就应慎重。如有可能，应在无应答者或失访者中进行随机抽样调查以获得应答，并将抽样结果与应答者的结果相比较，若结论一致，则表明无应答或失访对结果影响不大；若差异明显，则出现选择偏倚的可能性很大。此外，也可在资料分析时加以处理，即对试验组无应答或失访对象作为无效或阳性事件发生者；对照组中无应答或失访对象作为有效或阴性事件发生者，再经统计学分析，假设两者的结果相近而无显著差异，则无应答或失访对研究无明显影响。否则，要慎重地作出结论。

## 第三节 信息偏倚

在资料收集阶段，由于观察和测量方法上有缺陷，使各比较组所获得的信息产生系统误差，即为信息偏倚，又称为观察偏倚（observation bias）和测量偏倚（measurement bias）。产生信息偏倚的原因主要是诊断或结果判断的标准不明确、既往资料不准确或遗漏、对各比较组采用了不一致的观察或测量方法，以致获得错误信息影响了结果的真实性。

### 一、类型

#### （一）诊断怀疑偏倚（diagnostic suspicion bias）

当研究者事先已经知道研究对象的暴露史，怀疑他们已患有某种疾病，因此在诊断或判定治疗效果时，主观上作出对预期结果有利的判断，故而对暴露者使用多种诊断手段，进行详细的检查，并提高方法的灵敏度，使暴露组的诊断率和检出率提高，而对于非暴露组则因不怀疑他们患有某种疾病而对诊断和检查不够认真，这样各比较组的资料就会出现系统误差，影响结论的真实性，即为诊断怀疑偏倚。

诊断怀疑偏倚多见于队列研究和临床试验，在病例对照研究也可见到，在诊断亚临床



病例，鉴别是否存在药物的毒副作用和主要靠临床印象作出诊断的疾病，尤其容易发生这种偏倚。

例如美国弗明汉心脏研究（Framingham Heart Study）有一项规定，即病人若在发病后几分钟内死亡，医师未发现其他病因或既往病史不能提供其他死因，则可确定为冠心病猝死。而事实上蛛网膜下腔出血、某些代谢病和呼吸系统疾病也可引起突然死亡。同时，由于公认心血管危险因素与猝死有关，尤其是患者亦应具备这些危险因素时，医生在填写死亡证明书时就会将死因归入冠心病。进一步考察发现，只有 38% 的死亡病例生前有冠心病症状。这说明心血管危险因素与猝死的关联由于诊断怀疑偏倚而被人为地夸大了。

### （二）暴露怀疑偏倚（exposure suspicion bias）

当研究者认为某病与某因素有关联时，主观地采用不同深度和广度的调查或观测方法，对可能有因素暴露的病例组及未暴露的对照组探索可疑的致病因素，由此带来的偏倚称为暴露怀疑偏倚。该偏倚多见于病例对照研究。若调查表设计不完整、一问多答、调查的深度和广度不一致、暗示性启发式询问、记录不完整等，均可出现暴露怀疑偏倚，夸大可疑因子与疾病的联系。

例如，1962 年 Nishiyama 和 Kaventos 等人报告了使用不可比的方法调查儿童甲状腺癌病例与其放射性物质的照射史，若按照常规的调查方法或根据病人的医疗记录，病人的放射性物质照射史分别为 28% 和 0%。若往深入询问，核查同一研究对象用两种方法所获得的照射史分别为 47% 和 50%。由此可见用不可比的调查方法获得的照射率有明显的差异，暴露怀疑偏倚对结果的真实性有不容忽视的影响。

### （三）回忆性偏倚（recall bias）

指各比较组研究对象回忆以往发生的事件或经历，在准确性和完整性方面存在的系统误差。产生回忆性偏倚的原因常常是由于调查询问的因素或事件发生的频率很低，未给研究对象留下深刻的印象；或者因调查发生久远的事件而研究对象对此已记忆模糊或遗忘；还可能病例组的病人因患病而对过去的暴露史反复思索，甚至家属也帮助提供线索，以至于夸大了暴露情况，而对照组的非病人对调查不够重视，未认真回忆暴露史，因此使病例组和对照组提供的既往史的准确性和完整性差异较大；此外，可能由于某种原因，有的研究对象故意夸大或降低致病因素的暴露水平。以上种种原因都会导致回忆偏倚，影响结果的真实性。

例如，1956 年 Stewart 发表了幼儿白血病与母孕期接受 X 线照射的病例对照研究，结果显示多数白血病患儿的母亲于本次孕期接受 X 线照射的比例大于正常儿母亲的对照组，其中以 X 线检查腹部和骨盆的差别最明显，如表 10-8 所示，作者认为幼儿白血病与母孕期接受 X 线照射有关联。

表 10-8 白血病患者母孕期 X 线照射史比较

	X 线照射腹部			X 线照射其他部位		
	病例	对照	比例	病例	对照	比例
婚后至本次怀孕前	109	121	0.90 : 1	213	184	1.16 : 1
本次怀孕	178	93	1.91 : 1	117	100	1.17 : 1

上述推论引起了一场争论，问题的焦点之一是回忆偏倚。有学者认为上述两组妇女孕期 X 线照射史可能不同，但不能排除两组妇女回忆上的差异。因为幼儿患白血病以至死亡会给母亲心理带来创伤，调查时她们会比较认真地回忆孕期各方面的情况，常带有主观性。对照组母亲则无上述经历，可能漫不经心地回忆，于是回忆偏倚就使幼儿白血病与母

孕期 X 线照射的关联被人为地夸大了。临床实践表明，当询问病人既往病史时，回答的真实性和完整性随发病时间与调查时间的时间间隔加大而降低。

**(四) 临床资料遗漏偏倚 (missing clinical data bias)**

由于临床资料正常、阴性、未测量或测量未作记录所造成的临床资料遗漏，与完整的临床资料之间存在系统的差异，称为临床资料遗漏偏倚。例如，在统计分析病历资料时，未发现某项内容，因而导致了此类误差。

**(五) 家庭信息偏倚 (family information bias)**

向家庭成员调查某成员的既往病史或暴露史，若该成员是新发患者或久病不愈的患者，则倾向于提供更多的阳性信息；而被调查成员是健康者，则可能提供更多的阴性信息，其中一部分可能为假阴性，这就产生家庭信息偏倚。

例如，Schull 研究风湿性关节炎的家庭聚集性，以健康人为对照组。首先询问风湿性关节炎患者其父母患风湿性关节炎的家族史，结果呈现该病有家庭聚集性。然后以风湿性关节炎患者的健康同胞兄弟姐妹为对照询问其父母风湿性关节的患病情况，对比结果是，患病的家庭成员比健康的家庭成员提供了更多的阳性信息。见表 10-9。

表 10-9 风湿性关节炎患者与其同胞提供家庭遗传信息的对比

家族患病情况	风湿性关节炎病例组		患者同胞对照组	
	人数	%	人数	%
父母均未患病	11	27	20	50
父母一方患该病	23	58	17	42
父母双方患该病	6	15	3	8
合计	40	100	40	100

**(六) 顺序偏倚 (sequence bias)**

当研究按一定时间顺序进行，季节、气候、温度，以及研究对象或研究者机体状况发生系统的变化，即由于顺序规律而发生的系统误差，称为顺序偏倚。例如，用药物治疗慢性支气管炎，有时治疗试验要持续几个月，则季节的变化可影响治疗效果，如从冬季开始治疗到夏季，其结果就会优于从夏季开始治疗到冬季的疗效。

**(七) 不接受测量偏倚 (unreceptive measure bias)**

由于测量方法会造成损伤、羞辱、侵犯个人权力和隐私，则研究对象逃避和拒绝检查，若这种情况在病例组和对照组发生不均衡，则会产生不接受测量偏倚，影响结果的真实性。

**(八) 不敏感测量偏倚 (non-sensitive bias)**

当某试验的检测结果不足以测出临床有意义的变化或差异，即假阴性带来的误差，称为不敏感测量偏倚。

例如，用免疫电泳法测定乙型肝炎表面抗原的灵敏度较低，以此来检测就会出现假阴性，即产生不敏感测量偏倚。

**二、控制措施**

信息偏倚主要是在资料收集阶段由于不正确的测量方法而获得不真实的信息。为防止信息偏倚产生，通常应采取以下方法：

**(一) 采用盲法收集资料**

为消除研究者和研究对象主观因素的影响，可采用盲法，即在病例对照研究中让调查





者不了解研究对象的疾病诊断去询问和测量暴露史，在队列研究和临床试验中让观察者不了解研究对象的暴露情况和分组状态，以此收集和测量到的阳性结果就能保证其真实性。此外，尽可能在作出诊断前收集资料，亦可有效地减少信息偏倚。

## （二）收集客观指标的资料

在实际工作中使用盲法收集资料有时不可行，因此尽可能收集客观的定量指标，如利用实验室方法，避免开放式问题，以及查阅病历等医学记录，以此减少收集资料中的系统误差。

## （三）广泛收集各种资料

在可能的情况下，收集资料的范围尽可能有意识地扩大，不但收集详细的疾病资料，还可收集一些虚变量，即与疾病和暴露因素关系不密切的资料，借以分散调查者和研究对象对某因素的注意力，减少主观因素造成的误差。

## （四）保证研究人员的科学态度

研究开始之前要对调查人员进行统一的培训，并保证其掌握调查方法的一致性。研究开始以后要严格按照调查工作手册收集资料，不能轻易更改标准，操作技术应熟练，记录要准确，最大限度地保证资料的真实性。

# 第四节 混杂偏倚

## 一、概念

在慢性疾病病因研究中，当研究暴露于某因素与某疾病之间的关系时，由于一个或多个既与疾病有制约关系，又受暴露因素密切相关的外部因素（extraneous factor）的影响，从而掩盖或夸大了所研究的暴露因素与该疾病的联系。这种影响所带来的误差称为混杂偏倚或混杂。那些外部因素称为混杂因素（confounding factor）。混杂因素是一个与暴露因素和疾病都有关系的因素，它是所研究疾病的独立的危险因素，也是非暴露组中一个危险因素，并在人群中的分布与暴露因素的分布相关。而要注意区别的是混杂因素并非因果关系中的中间变量。

例如，一项高血清胆固醇水平与心肌梗死的病例对照研究，从以往的研究得知，心肌梗死的危险性与肥胖症有关，总胆固醇水平亦与肥胖症相关。此项病例对照研究发现，60名心肌梗死的病人中36人有高胆固醇，占60%。而60名对照中仅24人有高胆固醇，占40%。该结果提示高总胆固醇水平与心肌梗死危险性增加相关。当将肥胖症与非肥胖症者分开后考察二者的关联，结果就随之改变了。在肥胖症者中，40名心肌梗死的患者中34人发现有增高的胆固醇水平，占85%；而20名对照中发现有18人胆固醇水平增高，占90%。在非肥胖症者中，20名心肌梗死的患者中2人胆固醇水平增高，占10%；而40名对照中有6人胆固醇水平增高，占15%。由此可见，无论是肥胖症者还是在非肥胖症者中，胆固醇水平增高的现象在对照中比心肌梗死病人中更普遍存（表10-10）。

表10-10 血清胆固醇水平与心肌梗死危险性的关系及肥胖症因子的混杂作用

分组	高胆固醇	低胆固醇	合计	肥胖症者			非肥胖症者		
				高胆固醇	低胆固醇	合计	高胆固醇	低胆固醇	合计
心肌梗死患者	36	24	60	34	6	40	2	18	20
对照	24	36	60	18	2	20	6	34	40
合计	60	60	120	52	8	60	8	52	60



在此项研究中，60 名肥胖症者中 52 人胆固醇水平增高，占 87%，而 60 名非肥胖症者中仅 8 人胆固醇水平增高，占 13%。显而易见，外部因素肥胖症因子对结果起了混杂作用，即满足两个条件，一是在非暴露情况下与研究疾病相关，二是与暴露相关但不是暴露的结果。与选择偏倚和信息偏倚不同，混杂偏倚可以在结果分析时进行评价，通过分析暴露与疾病的关联发生改变而说明混杂作用的存在。混杂作用并不是“全或无”的，它可在不同研究中产生不同的作用。

二、正混杂和负混杂

(一) 正混杂偏倚

指由于混杂因素的作用，使暴露因素与疾病之间的关联被夸大。现以性格类型与冠心病关系的病例对照研究加以说明。在 257 名冠心病患者 A 型性格者有 178 人，占 69%，而在 514 名对照中 A 型性格者有 243 人，占 47%。粗比值比 (*cOR*) 为 1.50。结果表明 A 型性格者较 B 型性格者易患冠心病。然而吸烟的作用不容忽视，于是将上述资料按是否吸烟分为两层，结果见表 10-11。

表 10-11 性格类型与冠心病关系的分层分析

分组	吸烟			不吸烟		
	A 型性格	B 型性格	合计	A 型性格	B 型性格	合计
冠心病	168	34	202	10	45	55
对 照	189	38	227	54	233	287
合 计	357	72	429	64	278	342
aOR	0.99			0.96		

按吸烟与否分层分析后，校正比值比 (*aOR*) 小，均不到 1。说明性格类型与冠心病的关联由于吸烟的混杂作用而夸大。

(二) 负混杂偏倚

指由于混杂因素的作用使暴露因素与疾病的关联被人为地减弱。现以接触粉尘与慢性呼吸道疾病关联的病例对照研究为例加以说明。在 600 名慢性呼吸道患者中接触粉尘者为 200 名，占 33%，而对照组的 5400 名对象接触粉尘者有 1800 人，占 33%，*cOR* 为 1.00。该结果提示接触粉尘与慢性呼吸道疾病无关联。进一步考察年龄的作用，将研究对象分为小于 50 岁和大于 50 岁两组进行分层分析，结果见表 10-12。

表 10-12 接触粉尘与慢性呼吸道疾病关系的分层分析

分组	小于 50 岁			大于 50 岁		
	接触粉尘	未接触粉尘	合计	接触粉尘	未接触粉尘	合计
慢性呼吸道疾病	110	380	490	90	20	110
对 照	390	2620	3010	1410	980	2390
合 计	500	3000	3500	1500	1000	2500
aOR	1.94			3.13		

由此可见，按年龄组分层后，校正比值比 (*aOR*) 均大于 1，说明接触粉尘与慢性呼吸道疾病的关联由于年龄的混杂作用而减弱。



### 三、控制措施

#### (一) 限制 (restriction)

在研究设计阶段对研究对象的选择条件加以限制, 将已知存在混杂因素的对象不纳入研究, 规定各比较组在人口学特征上近似或在疾病特征上相同。例如, 研究年龄对急性心肌梗死预后的影响, 研究对象限制为 40~69 岁男性病人, 且无并发症。这样限制就控制了年龄、性别、并发症的混杂作用。但如果严格限制研究对象, 则会影响研究对象的代表性, 使研究结果在一般人群推论受到限制。

#### (二) 配比 (matching)

将可疑混杂因素作为配对因素, 使各比较组同等分配具有同等混杂因素的对象, 以此来消除混杂作用。各种流行病学研究都可用配比的方法消除混杂因素的影响。配比可分为个体配比 (pair matching) 和频数配比 (frequency matching)。前者是每一个暴露组或病例组的对象选择一个或几个非暴露或非病人的对照人群的对象作为对照, 组成对子, 每个对子具有某些相同的特征, 在资料分析时亦不拆开对子。后者亦称为成组配比, 即在获得暴露组或病例组后, 根据可疑的混杂因素在暴露组或病例组中的分布情况, 选择与其相同或相似的对照组, 使混杂因素在两组中均衡分布以消除其影响。

通常将年龄和性别作为配比因素, 因为这两个因素与许多疾病的发生、发展和预后关系密切。在临床流行病学研究中可根据研究目的和专业选择可疑的混杂因素进行配比。但注意在一项研究中配比因素不能过多, 其原因一是会增加获得符合条件对象的难度, 二是配比因素不能在研究加以分析, 可能会损失信息。

#### (三) 随机化 (randomization)

在设计阶段, 采取随机抽样的方法选择研究对象, 并采取随机分组的方法将研究对象分配到各比较组中, 使除研究因素以外的各种因素, 包括未知的混杂因素均衡地分布在各组中, 以消除混杂作用。

#### (四) 分层 (stratification)

在研究资料分析阶段, 将已知的或可疑的混杂因素按其不同水平分层, 然后再分别加以分析。这种方法适合于设计和实施阶段出现误差, 已无法更改资料, 经过分层分析, 可以控制混杂因素的影响。分层有单纯分层分析法和 Mantel-Haenszel 分层分析法。

#### (五) 标准化 (standardization)

按照统计学标准化的方法, 将需比较的两个率进行调整, 使可疑的混杂因素在两比较组中得到同等加权从而获得有可比性的标准化率, 以此避免混杂因素的影响。

#### (六) 多因素分析 (multivariate analysis)

当欲控制的混杂因子较多, 由于样本含量的限制使分层分析不适用时, 可采用各种流行病学多因素分析方法。20 世纪 60 年代起, Cornfield 提出了 Logistic 回归模型。经过几十年的发展, Logistic 回归模型目前已成为现代流行病学危险因素研究的首选方法。不论在病因学研究或预后的研究中, 危险因素或预后因素与疾病的关系都是非常复杂的, 各种危险因素或预后因素之间可以相互影响, 它们对结果的影响大小也不相同。采用 Logistic 回归模型进行多变量分析, 能在复杂关系中平衡多种混杂因素的作用, 进一步筛选出主要的危险因素或预后因素, 并反映其在决定发病以及预后中的相对比重。随着电子计算机统计软件的发展, 多因素分析方法应用得更为广泛, 可以有效地消除混杂因素的影响。



## 第五节 机 遇

### 一、机遇的定义与概念

对于一个假定完全避免了偏倚的样本作观察时,所得的结果仍会与真实情况有一定差异。这是由于来自于测量过程本身或生物学变异所产生的随机变异(random variation)造成的,这种单纯由于机会引起的差异称为机遇。换言之,机遇是某种结局可能出现的概率,而非观察本身的原因引起的,与系统误差的性质不同。

在临床研究中,无论何种设计,都不可能在整个人群和全部病例中进行,而只能从中抽取一部样本进行研究。因此就会不可避免地产生抽样误差,其特点是每次测量结果总是在真实值上下波动而始终不能消除与真实值的差异,这就是机遇。同偏倚一样,机遇在临床研究的各个阶段中广泛存在,是影响研究结果重复性,即可靠性的重要原因。研究和应用机遇的规律,了解其在研究设计、测量和评价中的影响,是临床流行病学的重要内容之一。

通常以两种方法检查机遇在临床研究中的作用,第一种是假设检验,第二是“估计”。假设检验是20世纪初就开始应用的传统的统计学检验,而“估计”的方法是用统计学的技术估计可能包含真实值的范围。由于机遇的影响,使得检验结果除了有真阳性和真阴性两种情况外,还有可能出现假阳性和假阴性。假阳性称为Ⅰ类错误或 $\alpha$ 错误,假阴性称为Ⅱ类错误或 $\beta$ 错误。对于 $\alpha$ 错误和 $\beta$ 错误的估计,就是检验单纯由于机遇引起错误结果的可能性大小。传统的假设检验是用 $P$ 值来估计发生假阳性错误的可能性,而用“估计”的方法是用可信区间来显示机遇的影响,常用95%可信限表示代表真实值的可能性为95%,也就是错误的可能性小于5%。用“估计”的方法优点是可估计测量值与真实值之间的差距范围。

影响机遇发生的重要因素是样本量,若在研究中反复抽样,假定没有偏倚影响,随着样本量增加,测量值会越来越接近真实值。尽管从理论上讲,为尽可能减少机遇的影响,样本量应该越大越好,但实际上样本不能无限制地扩大。因此要在机遇的影响减少到容许接受范围内决定样本量。样本量的决定因素主要为观察对象个体间的差异、研究结果的差异及 $\alpha$ 和 $\beta$ 值。如果研究对象个体差异小,研究结果各比较组间相差较大,则需要较少的样本量。如果要求尽可能地限制 $\alpha$ 和 $\beta$ 错误的可能性,就需要较大的样本量。

### 二、机遇与偏倚的关系

机遇和偏倚都是影响结果可靠性和真实性的重要因素,但由于它们产生的原因不同,应该加以区别。以血压测定为例,偏倚和机遇的关系和区别见图10-2。以动脉导管直接测定血压可以得到血压真实值,但通常只能使用袖带式血压计测定以获得体外血压的间接测量值,它与真实值之间的差异即为偏倚。当用袖带式血压计测量血压时,尽管认真控制条件,反复测定,但所获得的测量值总有差异,都围绕均数上下波动,在一个范围内变化,这就是机遇的影响。

从理论上讲,偏倚是研究设计不正确以及实施中的不规范等因素所造成的人为误差,因此可以通过完善的研究设计、正确客观的测量方法和适当的分析方法加以控制和消除,使其影响减少到最低程度。这也是认识分析偏倚的原因并进行处理的目的。对于已知的偏倚可以通过一系列方法加以控制,而对于未知的偏倚,只有通过严格的随机化加以控制,并在资料分析阶段采取有效的方法加以处理,以使偏倚的影响减弱到最小程度。



机遇是所有抽样研究固有的误差，只要不是以全部人群为研究对象，不可能纳入全部被研究的患者，因而机遇总是不可避免地存在。因此，只能通过扩大样本量，用统计学的方法将抽样误差限制在能够接受的范围之内，并估计其的大小，评价其对研究结果重复性，即可靠性的影响程度。

### 三、机遇的控制方法

首先，在研究设计阶段，要根据研究目的与假设，确定  $\alpha$  和  $\beta$  错误，即假阳性和假阴性的允许接受范围，然后来确定样本量。若要限制出现假阳性和假阴性的可能性，就要采用较大的样本量来进行研究。

例如，一项临床研究比较两种药物对抑制肿瘤生长的效果。预期新药疗效为 40%，高于老药 10%，要求有 80% 的检验效能，即把握度  $(1-\beta)$ ，故  $\beta$  错误为 20%，且单侧检验  $\alpha$  为 5%，以此条件计算出每组最少需为 280 例，共 560 例病人。因为两比较组间差异较小才需要这样大的样本数。反之，如果两比较组差异较大，或提高  $\alpha$  和  $\beta$  错误的水平，就可适当减少样本量。关于样本量的计算，应参阅有关统计学书籍。

其次，为控制机遇的影响，应在研究实施阶段严格控制测量条件，尽量使每次测量时的各种因素保持齐同，提高研究结果的可靠性。

最后，在资料分析阶段，对研究结果进行统计学处理分析，可获得不同的假阳性和假阴性值，并计算 95% 可信限，准确地估计机遇影响的范围。

(黄悦勤)

# 第十一章 临床研究资料统计分析的基本方法

临床研究的整个过程均要涉及统计学相关知识。在设计之初，要根据研究分析目的、数据资料类型等选择合适的统计分析方法，估算样本含量；在组织与实施阶段，可利用统计描述方法进行监测与质量控制，有针对性地收集数据资料，确保临床研究资料的收集质量与数据完整，防患于未然，避免数据资料的大量缺失；在研究完成后，又要利用选定的统计方法进行分析和评价研究结果。

借助统计学方法可以帮助研究者对繁杂的临床科研数据资料进行系统地归纳和总结，以便从中去粗存精，去伪存真，得到有价值的结果，是临床科研必不可少的工具。

## 第一节 临床研究的数据收集与整理

在临床研究中，数据资料的收集与整理均非常重要。原始数据的真实可靠，是临床研究成功的重要保证。

### 一、临床研究的数据来源

临床研究的数据来源很多，例如病历、研究记录表、调查表或问卷、专题调查与实验记录、实验室检验数据、统计报表、年鉴等均可为研究所用。

病历作为重要的临床工作记录，同时也是临床研究的重要数据来源。收集病历信息时应特别注意收集的质量，做到如实、全面、真实可靠。

如果研究对象是社区病人，则需要根据研究方案与目的，设计专门的调查表或问卷，供资料收集之用。

### 二、临床研究数据的收集内容与范围

临床研究的资料收集常集中于 PIO 类指标，如研究对象特征指标（population/patients, P），干预或暴露测量指标（intervention/exposure，简写 I 或 E）以及结（局）测量指标（outcome，简写 O）等。这些指标可大致归为以下四种类型：

#### （一）单纯生物学指标

即临床常用的一些硬指标，如病死率、不良事件发生率、痊愈率、复发率以及其他一系列有关人体生化、生理学指标等临床传统观察指标。

#### （二）健康相关生存质量及其衍生指标

随着疾病谱及医学模式的转变，一些全面反映病人健康与生存状态的指标，如 HRQL、QALYs、DALYs 等，在临床研究中已得到广泛应用。

#### （三）临床经济学指标

如直接医疗成本、间接医疗成本等一系列费用指标，可用于成本效果分析（CEA）、成本效益（CBA）分析、成本效用（CUA）分析等。

#### （四）人口特征指标

包括性别、年龄、种族、职业、教育程度及其他一些社会经济学指标等。

### 三、临床研究数据的收集方法

在确定了临床资料的收集范围与内容后，通过何种途径、方法收集则成为关键，这也



直接决定了数据收集的质量好坏。

### (一) 设计专门的资料收集工具

根据研究内容, 设置基本条目和备查条目, 形成专门的资料收集工具, 如调查/研究记录表等。基本条目是指与研究目的密切相关、必不可少的内容。备查条目是用于质量控制的一些项目。在临床研究中, 还可适当增加一些备选条目, 以便尽可能多地收集信息。

若填写项目或回答问题较多时, 可以采用一人一表的格式, 每份表只填写一个研究对象的所有相关内容; 若项目较少时, 可采用一览表的形式, 即在一份表内同时填写多个研究对象的相关信息。

### (二) 确定采集方式

采集方式主要有直接观测与访问等。直接观测是指研究人员直接到现场对观察对象进行观察或测量, 得到相关数据信息, 如临床研究中有关体检及实验室资料的收集就采用这种方式, 直接观测得到的数据较为客观真实。

访问法需要研究对象的配合, 通过研究对象自己回答问题来完成资料的收集。常见的访问方式有: ①面对面 (face to face) 访问法: 研究人员在现场, 通过研究对象自己填写或口头问答完成数据的收集; ②电话访问法: 研究人员通过电话问答的方式收集信息, 但该法有时会因电话变更、调查内容过多, 依从性较差, 可能出现较高的失访率; ③信访法: 以普通邮件或电子邮件的方式, 将表格直接寄给研究对象, 填好后寄回。本法虽然节省人力、物力与财力, 但收集质量和依从性难以保证, 应答率最低。

## 四、临床研究数据的整理与管理

当然原始数据还需进一步整理与归纳, 使其系统化与条理化后, 方能用于统计分析。传统的数据整理方法有打孔法、卡片法等, 伴随着计算机的普及与信息技术的发展, 现在临床研究大多采用建立数据库, 进行无纸化数据管理。

### (一) 赋值与定量化

资料整理的重要环节就是赋值与定量化。对于数值变量资料, 像血脂、血糖水平本身就被准确测量, 不存在赋值和定量化问题, 只是当有缺失值时, 才需做相应处理。对于分类变量资料, 则需要重新赋值。如对有序多分类资料, 可根据实际测量尺度采用等间距或非等间距赋值 (例如, 临床疗效分类中, 无效为 0, 有效为 1, 显效为 2, 痊愈为 3); 而对无序多分类资料, 就要复杂一些, 需采用哑变量方法赋值。例如研究中涉及黄、白、黑三个种族, 不能直接将黄种人、白种人、黑种人依次赋值为 1、2、3。这是因为三个种族并没有等级之分, 但在赋值后反而人为出现不同级别。对此, 可通过设置两个哑变量加以解决, 如规定凡是黄种人, 哑变量 1 赋值为 1, 余为 0; 凡为黑种人, 哑变量 2 赋值为 1, 余为 0。转换结果如下:

原来分类	哑变量赋值	
	哑变量 1	哑变量 2
黄种人	1	0
黑种人	0	1
白种人	0	0

哑变量的设置个数为分类个数减 1。如 ABO 血型包括 A、B、AB、O 四种类型, 需要设置 3 个哑变量。

### (二) 数据录入与建库

对收集到的临床研究数据, 为方便管理, 常借助现成的数据库管理软件, 如 AC-



CESS、EXCEL 以及 Visual Foxpro 等, 建立数据库。

建库应首先构建数据库结构, 统一定义字段及属性。字段又称变量, 反映了研究对象的某种共同属性, 例如病人的年龄、性别、身高、体重、血压等。定义字段包括确定字段名、字段属性(如字符型或数字型等)与字段大小。

结构建好后, 就可开始录入数据。通常将一个研究对象的所有字段信息称为一个记录,  $n$  个记录就构成了一个数据库。

## 五、质量控制

质量控制措施实际上应贯穿于临床数据资料的收集、整理与分析的全过程。数据质量是统计分析的根本, 否则再好的统计学方法也不能弥补数据上的缺陷。

在资料收集过程中, 一些混杂因素会影响数据的收集质量。例如病人自身的症状与体征在资料收集时不稳定, 处于剧烈变化之中, 像第一次测定的血压值为 148mmHg/88mmHg, 而一周后为 158mmHg/80mmHg; 又如对自诉症状重复测量时, 回忆性偏倚可导致两次测量结果不一致; 又比如资料收集者(包括测量者、调查者)本身也可能带来偏倚, 特别是在测量临床软指标(如生存质量等)时, 其态度的好坏与提问方式等都将影响到病人是否能如实回答问题; 此外测量工具(或仪器)本身也可能因系统误差导致结果失真, 特别是在大规模多中心临床试验中, 若各中心实验室条件不一, 使用的仪器、试剂与度量衡单位不统一, 会造成数据混乱, 直接影响结果的真实性。

因此, 需要采取一系列质控措施, 如对资料收集人员进行严格的培训, 制定统一标准与操作规程, 采用盲法测量, 选择信度与效度俱佳的测量工具等, 以保证原始资料的收集质量。

录入数据时, 同样应采取质量控制措施, 例如制定严格的录入规范和说明、双输入法等, 以确保数据录入的准确可靠。

## 第二节 统计分析的准备与筹划

在统计分析之前, 还需要进行一些准备工作, 如评估欲分析的数据质量, 发现数据有无缺失和异常; 同时应结合研究目的, 设计统计分析方案, 选择恰当的分析方法, 以减少统计分析的盲目性。

### 一、评价数据质量

统计分析前, 需要从整体上把握数据的基本特征以及质量, 发现有无极端值与异常值等。

#### (一) 极端值、异常值与缺失值

极端值(extreme value)又称离群值, 是指那些远离大多数测量值的极端数值, 要么极大, 要么极小。这些值会直接造成结果不稳定, 甚至夸大或歪曲结果, 得到错误结论。特别是在小样本的临床研究中, 极端值的作用尤为明显。判断极端值是否为异常值, 需结合临床或专业知识, 异常值常为临床专业知识无法解释的测量值。

缺失值(missing value)是指因种种原因不能得到观测指标的具体测量值, 出现数据缺失。评判临床研究中数据缺失的影响大小, 应视缺失属性而定。缺失主要分两种, 一种称为随机性缺失, 如临床试验中试验组与对照组均可能出现缺失值, 缺失比例相近, 缺失与临床干预措施无关, 若缺失比例不超过 20%, 对结果影响不大; 另一种则称为非随机性缺失, 例如药物的毒副反应过大, 造成病人的大量失访, 此时试验组与对照组的缺失比例





会不同, 缺失一旦发现与干预措施有关, 会对结果造成较大的影响。

## (二) 如何发现与识别极端值、异常值

发现与识别极端值、异常值通常使用统计描述的方法, 如定量描述、统计图表等, 即可以清晰揭示数据资料的基本特征与变化趋势, 同时结合临床专业知识, 又可以发现与识别其中的极端值、异常值。若根据一般常识与临床病理生理学知识, 发现数据资料中最大值、最小值异常或不合常理, 则应高度重视。如在一项临床研究中, 病人的收缩压在400mmHg以上, 则应视为异常值。又比如某一个变量的标准差过大或者某些观察值偏离均数3倍标准差以上, 则说明观察单位间变异较大, 应进一步核实, 判断是否为异常值。另外, 使用统计图表也可直观地发现极端值。

## 二、设计统计分析框架图

在一个临床研究中, 若欲分析的内容较多, 为避免重复或遗漏, 不管是定量分析还是定性分析, 均要做详细周密的安排, 设计分析框架图(图11-1)。

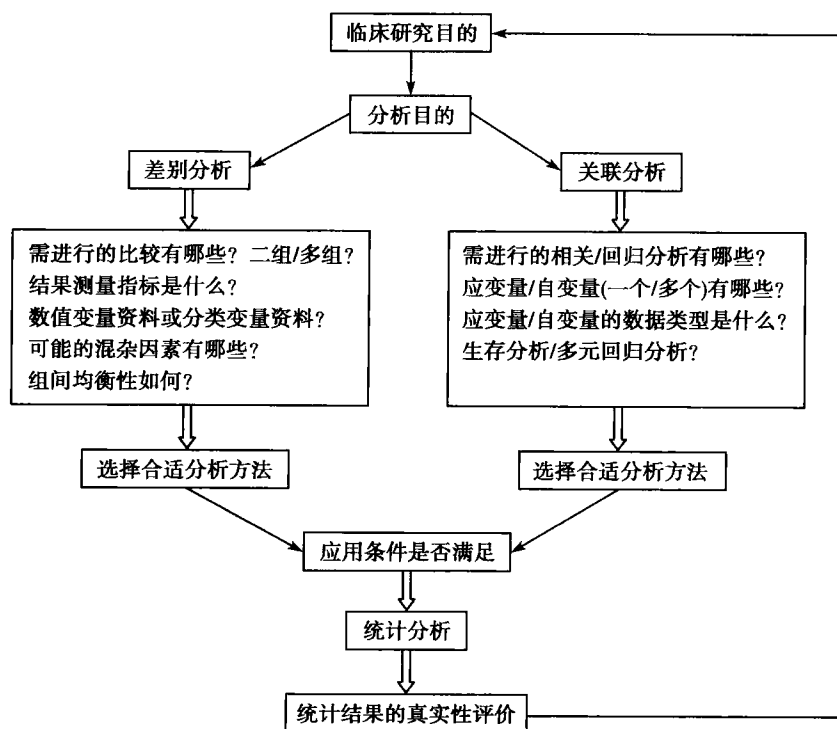


图 11-1 分析框架示意图

实例分析: 在一项颈动脉内膜切除术预防同侧脑卒中的临床试验中, 对照组使用阿司匹林进行内科治疗, 试验组则在阿司匹林治疗的基础上行颈动脉内膜切除术。研究目的是对颈动脉内膜切除术进行利弊综合分析。

则该研究的分析框架设计如下:

1. 研究目的 颈动脉内膜切除术的利弊分析。

2. 分析目的 两组利、弊差别比较。

(1) 颈动脉内膜切除术组的疗效是否优于对照组?

结果测量指标: 同侧脑卒中的发生率。

数据资料类型: 是否发生卒中, 为两分类变量资料。



(2) 颈动脉内膜切除术的副作用是否少于对照?

副作用测量指标: A 术中脑卒中发生率。

B 假性动脉瘤的发生率。

数据资料类型: 副作用事件是否发生, 也为两分类变量资料。

3. 根据分析目的与资料类型首选四格表  $\chi^2$  检验, 分析同侧脑卒中的发生率、术中脑卒中发生率以及假性动脉瘤的发生率等组间差别是否有统计学意义。

4. 考察应用条件是否满足: 若样本量大于 40 且最小理论频数大于 5, 则进行  $\chi^2$  检验与区间估计, 否则采用校正  $\chi^2$  检验或确切概率法。

5. 结果真实性的分析与评价: 结果真实性如何? 是否还有一些混杂与偏倚因素未被控制?

### 第三节 常用的统计分析方法

统计分析包括统计描述与统计推断。统计推断又包括假设检验 (如  $t$  检验、 $\chi^2$  检验、方差分析等) 和区间估计 (如均数、率、RR、OR 等参数的 95% 可信区间估计)。

#### 一、统计描述方法

根据数据资料类型选用相应的统计描述方法, 揭示数据资料的基本特征。如针对数值变量资料的描述可选用均数±标准差, 而分类变量资料则用率、比描述。

##### (一) 数值变量资料的描述

定量描述数值变量资料的基本特征有两类指标, 一是集中趋势指标, 用以反映一组数据的平均水平, 常用均数 (mean) 或称算术平均数 (arithmetic mean)、中位数 (median) 和几何均数 (geometric mean) 等; 二是描述离散程度指标, 用以反映一组数据的变异大小, 常用标准差 (standard deviation)、四分位数间距 (inter quartile range, IQR) 和变异系数 (coefficient of variation) 等。两类指标需联合应用才能全面揭示出一组数值变量资料的基本特征。如对于正态分布或近似正态分布资料, 可用均数±标准差来描述, 而对于偏态分布或未知分布资料的特征描述则使用中位数与四分位数间距。

定量描述时, 应注意两个问题, 一是要结合资料类型与适用条件, 选择恰当的描述方法。如不能用均数±标准差描述偏态分布、分布未知或两端无界的数据资料; 二是要注意同质性, 如不分性别和年龄地计算一组对象血红蛋白的均数和标准差, 既不能描述老年人、儿童血红蛋白量的集中趋势, 也不能反映女性的变异水平, 意义不大。

##### (二) 分类变量资料的描述

常用率 (rate) 和构成比 (proportion) 定量描述分类变量资料。由于其定义和计算都很简单, 是临床研究中常用的一类指标, 如病死率、治愈率、感染率等。

使用这类指标, 应注意避免两类错误: 一是以比代率, 即误用构成比描述某病发生的强度和频率, 如直接用某病的病人数除以就诊人数 (或人次) 得到“某病患病率”或“某病发病率”, 计算这些率需要人口数据。二是把不同率混用, 如将患病率与发病率、死亡率与病死率等同使用。

#### 二、假设检验方法

假设检验的方法很多, 应结合研究目的、资料性质、设计类型、样本含量等选择合适方法。



## (一) 数值变量资料的假设检验方法

1. 当比较两组小样本的数值变量资料时, 可以考虑选用  $t$  检验。 $t$  检验的目的是推断两组数值变量(计量)资料的样本所代表的总体均数是否相等。根据设计类型可大致分为三个类型: 单个样本均数与总体均数比较  $t$  检验, 配对设计  $t$  检验以及两个独立样本均数比较的成组  $t$  检验。

若比较两个独立组别的总体均数是否相等或其中一个组别总体均数是否大于另外一个组别总体均数时, 采用成组设计  $t$  检验。应用时要考虑三个基本条件: ①每组的观察值要服从正态分布, 否则将会造成  $P$  值低估; ②两组满足方差齐性, 假设检验中的无效假设为两总体均数相等, 也就是两组来自同一个总体, 两组方差相同。由于  $t$  检验对方差不齐敏感, 因此在  $t$  检验前要求做方差齐性检验, 现多数统计分析软件有此默认功能, 自动将方差齐性检验结果一并给出。方差不齐时, 可采用  $t'$  检验或秩和检验; ③独立性, 要求组内或组间的观察测量值相互独立。

例 11-1 在一项临床试验研究中, 探讨综合规范康复法对脑卒中患者预后的改善效果, 将研究对象分为两组, 一组患者(39 例)采用传统康复方法, 另外一组(45 例)采用综合规范康复方法, 半年后两组功能恢复评分依次为  $35.6 \pm 28.11$  分,  $51.41 \pm 28.45$  分。

显然两组的样本均数不等,  $35.6 \neq 51.41$ , 这个差异有可能由机遇造成, 下结论前, 我们需要知道, 由于机遇造成该差异的可能性有多大? 要回答这个问题需要进行假设检验。

本例每组例数均大于 30 例, 且方差齐, 故选用成组设计  $t$  检验。

步骤 1: 建立假设

$H_0$ : 两组预后功能恢复评分的总体均数相等,  $\mu_1 = \mu_2$

$H_1$ : 两组预后功能恢复评分的总体均数不等,  $\mu_1 \neq \mu_2$

检验水准:  $\alpha = 0.05$

这里  $H_0$  称为无效假设, 假设两组样本均来自同一总体, 样本均数不等是由抽样误差所致;  $H_1$  为备择假设, 即两组样本来自于不同总体, 不同康复法确实影响了患者的预后。检验水准是判断无效假设成立与否的界值点, 需要事先设定。 $\alpha = 0.05$  是指真实情况下两组的总体分值确实没有差别, 有 5% 的机会错误推断出两者总体平均分值有差别。

步骤 2: 计算统计量

$$t_{(n_1+n_2-2)} = \frac{(\bar{x}_1 - \bar{x}_2)}{SD_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \quad \text{公式 11-1}$$

$$\text{两组合并标准差 } SD_p = \sqrt{\frac{(n_1-1)SD_1^2 + (n_2-1)SD_2^2}{n_1+n_2-2}} \quad \text{公式 11-2}$$

这里  $n_1$ ,  $n_2$ ,  $\bar{x}_1$ ,  $\bar{x}_2$ ,  $SD_1$ ,  $SD_2$ ,  $SD_p$ , 分别代表两组的例数、均数、标准差以及两组合并标准差。

步骤 3: 确定  $P$  值, 下结论

$$\text{本例 } SD_p = \sqrt{\frac{(39-1) \times 28.11^2 + (45-1) \times 28.45^2}{39+45-2}} = 28.27$$

$$t_{(39+45-2)} = \frac{(35.6 - 51.41)}{28.27 \sqrt{[(1/39) + (1/45)]}} = -2.56$$

自由度为  $n_1 + n_2 - 2 = 39 + 45 - 2 = 82$

通过查表或直接计算, 得  $P < 0.01$ 。按照  $\alpha = 0.05$  检验水准,  $P < \alpha$ , 拒绝  $H_0$ , 接受  $H_1$ , 可以认为两组的总体功能恢复评分不等, 综合规范康复法效果好于传统方式。



成组设计  $t$  检验在临床研究中的使用频率较高, 其应用条件较为严格, 要满足正态性、方差齐性、独立性等。 $t$  检验只适合两组数值变量资料比较的假设检验。

2. 当比较组数不止两组, 而是三组及其以上时, 就不能用  $t$  检验, 否则会增大 I 错误 (假阳性) 率, 考虑使用方差分析 (analysis of variance, ANOVA)。

方差分析在临床研究中也较为常用, 可用于推断两个及以上总体均数有无区别, 分析两个或多个研究因素的纯效应与交互作用, 以及用做回归方程的线性假设检验等。

方差分析又称变异分析, 其中心思想就是将变异分解成两 (或两个以上) 部分: 组内变异与组间变异。组内变异与组间变异分别用组内均方和组间均方表示。利用组间均方与组内均方的比值, 构建统计量  $F$ , 再根据  $F$  分布理论、 $F$  值和自由度确定  $P$  值。

例 11-2 某医院对 31 名石棉矿工中的石棉肺患者、可疑患者、非患者进行了用力肺活量测定, 问三组石棉矿工的用力肺活量有无差别? 结果见表 11-1。

表 11-1 三组石棉矿工的用力肺活量 (L) 比较

	石棉肺患者	可疑患者	非患者	合计
	1.8	2.3	2.9	
	1.4	2.4	3.2	
	1.5	2.1	2.7	
	2.1	2.1	2.8	
	1.9	2.6	2.7	
	1.7	2.5	3.0	
	1.8	2.3	3.4	
	1.9	2.4	3.0	
	1.8	2.4	3.4	
	1.8		3.3	
	2.0		3.5	
均数	1.79	2.31	3.08	2.40
标准差	0.2	0.18	0.29	
例数	11	9	11	

上述资料为数值变量资料, 目的是推断三组样本所代表的总体肺活量均数是否相等, 因涉及三组均数间比较, 不能使用  $t$  检验, 应用方差分析。

用  $X$  表示个体观察值,  $\bar{X}_i$  表示第  $i$  组的均数,  $\bar{X}$  为所有观察值的合并均数。可将每个观察值与合并均数间的变异 ( $X - \bar{X}$ ) 分解成两部分, 个体观察值与组均数间的组内变异 ( $X - \bar{X}_i$ ) 以及组均数与合并均数  $\bar{X}$  间的组间变异 ( $\bar{X}_i - \bar{X}$ )。变异关系可写成:  $(X - \bar{X}) = (X - \bar{X}_i) + (\bar{X}_i - \bar{X})$ 。例如石棉肺患者组第一个患者的用力肺活量为 1.8, 观察值的合并均数是 2.39, 1.8 与 2.40 的差值为 -0.60, 可以写成:  $(1.79 - 2.40) + (1.8 - 1.79) = -0.61 + 0.01$ , 显然组间变异远大于组内变异。

同理, 将所有 31 个观察值的变异进行分解后求和, 为避免相加后正负抵消为 0, 采用离均差平方和表示变异累加。其中总变异为  $SS_{\text{总}} = \sum (X_i - \bar{X})^2$ , 即每个观察值与合并均数  $\bar{X}$  的离均差平方和。组内变异为  $SS_{\text{组内}} = \sum (X_i - \bar{X}_i)^2$ , 因  $SS_{\text{组内}}$  与样本例数有关, 为排除其影响, 用组内均方代替: 即  $MS_{\text{组内}} = SS_{\text{组内}} / (N - K)$ 。其中  $N$  为总例数,  $K$  为组别个数。组间变异为  $SS_{\text{组间}} = \sum n_i (X_i - \bar{X})^2$ , 为排除组数的影响, 同样用组间均方代替:  $MS_{\text{组间}} = SS_{\text{组间}} / (K - 1)$ 。三者的关系可表示为:



$$SS_{\text{总}} = SS_{\text{组间}} + SS_{\text{组内}}$$

$\nu_{\text{总}} = \nu_{\text{组间}} + \nu_{\text{组内}}$ ， $\nu_{\text{组间}}$ ， $\nu_{\text{组内}}$  分别为第一自由度和第二自由度。

$$F = MS_{\text{组间}} / MS_{\text{组内}}$$

当无效假设成立时， $MS_{\text{组间}}$  与  $MS_{\text{组内}}$  接近， $F$  值在 1 附近出现的机会多，而出现较大  $F$  值的几率小，但当小到某界值时，根据小概率事件原理，就反证出无效假设不成立，从而拒绝  $H_0$ ，接受  $H_1$ 。

步骤 1：建立假设

$H_0$ ：三组矿工用力肺活量的总体均数相等， $\mu_1 = \mu_2 = \mu_3$

$H_1$ ：三组矿工用力肺活量的总体均数不等或不全等

$\alpha = 0.05$

步骤 2：计算统计量  $F$  值与确定  $P$  值

按上述原则计算，结果见表 11-2。

表 11-2 方差分析结果

变异来源	SS	$\nu$	MS	F	P
总变异	10.8	30			
组间变异	9.266	2	4.633	84.54	<0.01
组内变异	1.534	28	0.0548		

通过查  $F$  界值表，得  $P < 0.01$ ，按照  $\alpha = 0.05$  水准，拒绝  $H_0$ ，而接受  $H_1$ ，可认为三组矿工的用力肺活量不同，表明三组工人中至少有两组用力肺活量总体均数间有差别，至于哪两组间有差别，哪些没有差别，还需进一步做两两比较。

步骤 3：两两均数比较的假设检验

如果三组总体均数间差别无统计学意义，不需做两两比较；但若三组总体均数间差别有统计学意义，则必须做两两比较。两两比较的假设检验方法很多，例如任两组比较方法有 Scheffe 检验，TukeyHSD 检验，Newman-keuls 检验（SNK 检验）等；若是多个试验组与一个对照组比较，则可用 Dunnett 检验。以 Scheffe 检验为例，简述两两比较的全过程。

第一步：建立假设

$H_0$ ：任两组矿工用力肺活量的总体均数相等，即  $\mu_A = \mu_B$

$H_1$ ：任两组矿工用力肺活量的总体均数不等

检验水准  $\alpha = 0.05$

第二步：计算均数间差值和  $S$  界值

$$S = \sqrt{(j-1) F_{\alpha, df} MS_{\text{组内}} \left( \frac{1}{n_A} + \frac{1}{n_B} \right)} \quad \text{公式 11-3}$$

$F_{\alpha, df}$  为  $F$  界值， $j$  表示组数， $MS_{\text{组内}}$  为组内均方， $n_A$ ， $n_B$  为比较两组的例数。

本例  $\alpha = 0.05$ ，第一自由度为 2，第二自由度为 28，其  $F$  界值 = 3.33。以石棉肺组与可疑组比较为例，其界值为：

$$S = \sqrt{2 \times 3.33 \times 0.0548 \times \left( \frac{1}{11} + \frac{1}{9} \right)} = 0.272$$

同理，石棉肺组/非患者组比较以及可疑组/非患者组比较的界值依次为：0.258 及 0.272。

所有两两比较结果，见表 11-3。



表 11-3 两两比较结果

	两组间差值	P 值
石棉肺组与可疑组	-0.52	<0.01
石棉肺组与非患者组	-1.29	<0.01
可疑组与非患者组	-0.77	<0.01

所有组间差值均大于相应的 S 界值, 因此均有统计学意义 ( $P < 0.05$ )。

#### 步骤 4: 结论

可认为三组矿工的用力肺活量不同, 进一步两两比较结果表明, 石棉肺患者/可疑患者, 可疑患者/非患者以及石棉肺患者/非患者间的用力肺活量差别均有统计学意义。以石棉肺患者为最低, 其次是可疑患者, 最高为非患者组。

应用方差分析时, 同样要求满足以下三个条件: ①各组来自正态总体, 服从正态分布或近似正态分布; ②各组为相互独立的随机样本, 各观察值间相互独立; ③方差齐性, 各组的总体方差相等或近似。

### (二) 分类变量资料的假设检验方法

进行两组或多组分类变量资料间比较的假设检验, 可选用  $\chi^2$  检验。 $\chi^2$  检验常用于分析两个或两个以上总体率或构成比有无差别, 两分类变量间是否相关以及拟合优度检验等。

#### 1. 两组率或构成比比较的 $\chi^2$ 检验

例 11-3 为探讨葛根素的疗效, 假设在一项临床试验中, 葛根素组治疗 100 例, 有效 40 例, 不加葛根素的对照组 100 例, 有效 10 例。问两组有效率有无差别? 结果见表 11-4。

表 11-4 两组有效率比较

处理	有效例数	无效例数	合计	有效率 (%)
葛根素组	40 (25)	60 (75)	100	40
对照组	10 (25)	90 (75)	100	10
合计	50	150	200	25

四格表内的数据 (40、60、10、90) 称为实际频数 (用 A 表示)。假如干预措施与结果无关, 两者相互独立, 那么葛根素组与对照组的有效率理论上相等, 均为 25%, 由此可以得到两组理论上的有效例数及无效例数 (见上表括号内数字), 称为理论频数 (用 T 表示),  $T = \text{组例数} \times \text{总体率}$ , 可简化为:  $T = (\text{行合计} \times \text{列合计}) / n$ 。

#### 步骤 1: 建立假设

$H_0$ : 两组总体有效率相等,  $\pi_1 = \pi_2 = \pi$

$H_1$ : 两组总体有效率不等,  $\pi_1 \neq \pi_2$

$\alpha = 0.05$

#### 步骤 2: 计算统计量

计算实际频数 (A) 与理论频数 (T) 差值的平方和  $\sum (A - T)^2$ 。若 T 值越大,  $\sum (A - T)^2$  值越大, 为消除其影响, 构造统计量  $\sum \frac{(A - T)^2}{T}$ 。在样本量足够大时, 该值服从自由度 ( $\nu$ ) 为 (行数 - 1)  $\times$  (列数 - 1) 的  $\chi^2$  分布。

$$\chi^2 = \sum \frac{(A - T)^2}{T} \quad \text{公式 11-4}$$

因此,  $\chi^2$  值实际反映了所有格子的实际频数与理论频数吻合程度, 当无效假设成立



时, 实际频数与理论频数接近,  $\sum \frac{(A-T)^2}{T}$  较小, 出现较大  $\chi^2$  值的概率  $P$  很小, 若  $P < \alpha$  (检验水准), 则怀疑假设成立, 因而拒绝它。

步骤 3: 确定  $P$  值, 下结论

本例  $\chi^2 = 6.48$ , 自由度为 1, 查  $\chi^2$  界值表, 得  $P < 0.05$ , 按照  $\alpha = 0.05$  水准, 拒绝  $H_0$ , 而接受  $H_1$ , 可认为两组有效率不同。

对于四格表资料, 还可用  $\chi^2$  检验专用公式。

$$\chi^2 = \frac{(ad-bc)^2 n}{(a+b)(c+d)(a+c)(b+d)} \quad \text{公式 11-5}$$

应用时同样需要考虑应用条件。当  $T \geq 5$  且  $n > 40$  时, 选用公式 11-5; 当  $1 < T < 5$  且  $n > 40$  时, 需用校正公式 (公式 11-6, 11-7); 当  $T < 1$ , 或者  $n < 40$  时, 选用确切概率法。

$$\chi^2 = \sum \frac{(|A-T| - 0.5)^2}{T} \quad \text{公式 11-6}$$

$$\chi^2 = \frac{(|ad-bc| - n/2)^2 n}{(a+b)(c+d)(a+c)(b+d)} \quad \text{公式 11-7}$$

两组率比较的假设检验还可以选用  $u$  检验 (或称  $z$  检验),  $u$  检验是一种近似检验方法, 要求每组事件发生率与样本含量的乘积 ( $np$ ) 以及未发生率与样本量的乘积 ( $n \times (1-p)$ ) 均大于 5。若条件不符, 可以选用二项分布方法或当  $np$  很小时, 直接采用 poisson 分布方法。

## 2. 多个样本率或构成比比较的 $\chi^2$ 检验

例 11-4 为观察两种药物治疗消化不良的效果, 进行一项临床试验, 一组给予药物 1, 一组给予药物 2, 剩下一组设为对照组, 观察三组消化不良的痊愈情况, 问三组消化不良未愈率有无差别? 结果见表 11-5。

表 11-5 三组消化不良未愈率的比较

组别	未痊愈	痊愈	合计	未痊愈率 (%)
药物 1 组	108	42	150	72.0
药物 2 组	119	45	164	72.6
对照组	130	34	164	79.3
合计	357	121	478	74.7

步骤 1: 建立假设

$H_0$ : 三组消化不良未痊愈率相等,  $\pi_1 = \pi_2 = \pi_3$

$H_1$ : 三组消化不良未痊愈率不等或不全等

$\alpha = 0.05$

步骤 2: 计算统计量

$$\chi^2 = n \left( \sum \frac{A^2}{n_{RNC}} - 1 \right) \quad \text{公式 11-8}$$

$$\chi^2 = 478 \left( \frac{108^2}{357 \times 150} + \frac{119^2}{357 \times 164} + \dots + \frac{34^2}{121 \times 164} - 1 \right) = 2.79$$

自由度  $\nu = (3-1) \times (2-1) = 2$

步骤 3: 确定  $P$  值与下结论

查  $\chi^2$  值表, 得  $P > 0.05$ 。按照  $\alpha = 0.05$  水准, 不拒绝  $H_0$ , 还不能认为三组的消化不良未痊愈率不等。



此类  $\chi^2$  检验要求理论频数不能太小, 不宜有 1/5 以上格子的理论频数小于 5 或有一个理论频数小于 1。一旦出现, 可尝试: ①增加样本例数以增加理论频数; ②删除理论频数太少的行或列; ③将理论频数太小、性质相近的行或列合并; ④或直接采用确切概率法。

### 3. 注意事项

(1) 应避免  $\chi^2$  检验的误用。如某外科手术后, 医生发现部分病人更容易出现并发症, 为探讨年龄与手术并发症的关系, 某研究者收集了相关资料, 将病人分成两组, 小于 45 岁组与大于 45 岁组,  $\chi^2$  检验表明两组并发症发生率无差别 ( $P > 0.05$ )。然而, 若直接采用  $t$  检验却发现并发症组年龄显著高于无并发症组 ( $P < 0.05$ )。显然, 以 45 岁作为分界点的  $\chi^2$  检验不合适。

(2)  $\chi^2$  检验有一定的局限性, 假设检验只能回答两组或多组率间是否有差别 (或差别有无统计学意义), 不能说明差别程度以及是否有临床价值等, 要回答上述问题, 需估计率差的可信区间。

### (三) 相关与回归分析

相关与回归分析是研究两个或多个变量之间相互关系的统计方法, 回归系揭示两个或多个变量之间的数量依存关系, 而相关则反映两个或多个随机变量间的密切程度与方向。

1. 直线相关分析 直线相关是描述两变量间是否有直线关系以及直线关系的方向和密切程度的分析方法, 用相关系数  $r$  表示, 两变量间相关关系可分为: 正相关 ( $0 < r < 1$ )、负相关 ( $-1 < r < 0$ )、零相关 ( $r = 0$ )、完全相关 ( $|r| = 1$ ) 等 (图 11-2)。

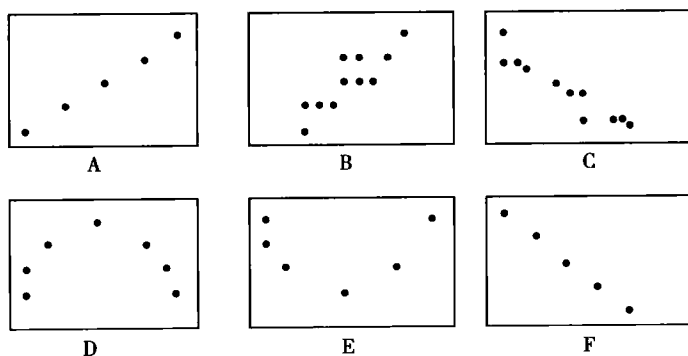


图 11-2 相关系数示意图

图中 A 为完全正相关 ( $r = 1$ ), B 为正相关, C 为负相关, D、E 为零相关, F 为完全负相关 ( $r = -1$ )。

进行直线相关分析, 要求双变量满足独立性、随机性及正态性等基本条件。

例 11-5 为探讨病人血糖与胰岛素水平的关系, 某研究者同时测量了 7 名糖尿病患者的血中胰岛素与血糖水平, 结果见表 11-6, 问血糖水平是否与胰岛素有关?

表 11-6 7 名糖尿病患者的血中胰岛素与血糖水平

患者编号	1	2	3	4	5	6	7
胰岛素 (mU/L)	24	17	17	12	14	12	9
血糖 (mg/dl)	142	170	194	213	214	238	249

步骤 1: 绘制散点图 (图 11-3)



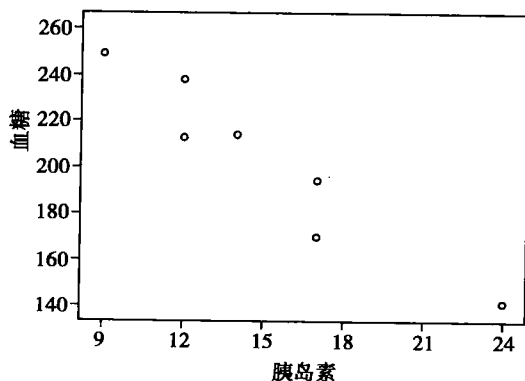


图 11-3 散点图

## 步骤 2: 计算相关系数

$$r = \frac{\sum (x - \bar{x})(y - \bar{y})}{\sqrt{\sum (x - \bar{x})^2 \sum (y - \bar{y})^2}} \quad \text{公式 11-9}$$

本例相关系数  $r = -0.956$ 。

注意不能只凭相关系数的绝对值大小来判断相关程度, 尚需进行假设检验及区间估计, 以推断两变量间是否存在总体相关关系。

## 步骤 3: 假设检验

## 第一步: 建立假设

$H_0$ : 胰岛素与血糖变化无关, 总体相关系数  $\rho = 0$

$H_1$ : 胰岛素与血糖变化有关, 总体相关系数  $\rho \neq 0$

$\alpha = 0.05$

第二步: 计算统计量  $t$  值

$$t = \frac{r \sqrt{n-2}}{\sqrt{1-r^2}} = \frac{(-0.956) \sqrt{7-2}}{\sqrt{1-(-0.956)^2}} = 7.29, \text{ 自由度} = 7-2=5$$

第三步: 确定  $P$  值, 下结论

查表得  $P < 0.001$ , 按照  $\alpha = 0.05$  的水准, 拒绝  $H_0$ , 接受  $H_1$ , 可认为胰岛素与血糖变化有关。

## 步骤 4: 计算直线相关系数的 95% 可信区间

由于直线相关系数  $r$  本身不服从正态分布, 需经 Fisher 转换为  $Z$  值, 由于  $Z$  值服从正态分布, 则其 95% 可信区间为:  $Z \pm u_{\alpha/2} / (n-3)^{1/2}$

$$Z(r) = \frac{1}{2} \ln \left( \frac{1+r}{1-r} \right) \quad \text{公式 11-10}$$

将上例相关系数转换  $Z$  值为:  $\frac{1}{2} \ln \left( \frac{1+(-0.956)}{1-(-0.956)} \right) = -1.897$

$Z$  值的 95% 可信区间:  $(-1.897) \pm 1.96/2$ , 其下限为  $-2.877$ , 上限为  $-0.917$ 。再将  $Z$  值逆转换后, 得到相关系数的 95% 可信区间为  $(-0.724, -0.994)$ 。

借助可信区间同样可以推断总体相关关系是否成立。若可信区间不包括 0, 表示总体相关关系成立; 若包括 0, 则表示总体零相关。本例相关系数的 95% 可信区间  $(-0.724, -0.994)$  不包括 0, 表明有统计学意义, 胰岛素水平与血糖水平存在负相关关系。

直线相关的注意事项: ①为方便理解相关系数的实际意义, 可人为将相关系数分为 4 个等级:  $0 \sim 0.25$  表示弱相关或无关;  $0.26 \sim 0.50$  表示有一定程度的相关关系;  $0.51 \sim$



0.75 为中等或较好的相关； $>0.75$  表示强相关，但当相关系数大于 0.95 时，反而要慎重。②相关系数的平方为决定系数，用以解释相关程度或关系密切强度。如上例的决定系数为  $(-0.956)^2 = 0.914$ ，意味着 91.4% 的血糖水平变化是由于胰岛素引起的，两者关系非常密切。③若变量资料为偏态分布或存在极端值时，可通过变量转换将其转化为正态分布后再进行直线相关分析，或者直接计算等级相关系数。

2. 直线回归分析 直线回归分析用于描述两个数量变量间的依存关系，属双变量分析 ( $X$ 、 $Y$ ) 的范畴。如研究胰岛素与血糖，体重与肺活量间的依从关系。直线回归要求应变量 ( $Y$ ) 独立且服从正态分布。

依存关系的定量描述可用直线回归方程： $\hat{Y} = a + bx$

其中  $x$  为自变量，在临床研究中一般为能精确测定和控制的变量； $Y$  为应变量，指在  $x$  规定范围内随机变化的量， $\hat{Y}$  为  $Y$  的估计值； $a$  为常数项，表示  $Y$  的基线水平； $b$  为回归系数，意为  $x$  每改变一个单位时， $\hat{Y}$  值的平均变化量。

仍以例 11-5 胰岛素与血糖关系研究为例，阐述直线回归分析过程。

步骤 1：绘制散点图

与图 11-1 相同。

步骤 2：计算直线回归方程参数

采用最小二乘法估计回归系数 ( $b$ ) 与常数项 ( $a$ )。

本例： $\hat{Y} = 312.8 - 7.33x$ ， $b = -7.33$ ， $a = 312.8$

步骤 3：在散点图上绘制回归方程的图示

在坐标轴上任意取相距较远的两点，连接上述两点就可得到回归方程的图示（图 11-4）。

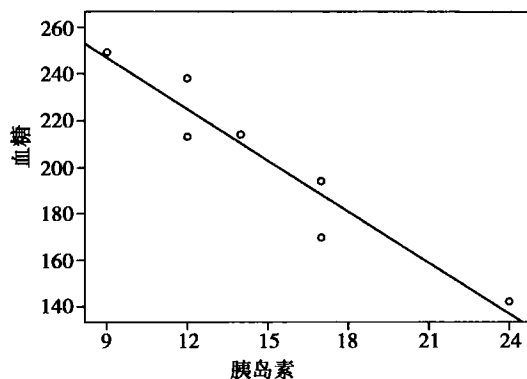


图 11-4 直线回归方程的图示

原则上要求连接的回归直线不应超过  $x$  的实测值范围，散点尽量分布在直线的两侧。

步骤 4：直线回归关系的假设检验

直线回归关系的假设检验又称回归方程检验，其目的是检验总体回归方程是否成立，有方差分析和  $t$  检验两种方法。

(1) 方差分析法：将变异分解为回归变异和剩余变异， $SS_{\text{总}} = SS_{\text{回归}} + SS_{\text{剩余}}$ ， $SS_{\text{总}}$  表示总变异， $SS_{\text{回归}}$  表示在总变异中能被  $x$  解释的变异， $SS_{\text{回归}}$  越大，表示回归效果越好。 $SS_{\text{剩余}}$  则表示排除  $x$  影响外的剩余变异（表 11-7）。

第一步：建立假设



$H_0: \beta=0$ , 即胰岛素与血糖之间无直线关系

$H_1: \beta \neq 0$ , 即胰岛素与血糖之间有直线关系

检验水准  $\alpha=0.05$

第二步: 计算统计量  $F$  值

表 11-7 方差分析结果

变异来源	SS	df	MS	F	P
回归变异	7729.3	1	7729.3		
剩余变异	723.5	5	144.7	53.42	0.0008
总变异	8452.9	6	1408.8		

第三步: 确定  $P$  值, 下结论

查表得  $P=0.0008$ , 按照  $\alpha=0.05$  的水准,  $P < \alpha$ , 拒绝  $H_0$ , 接受  $H_1$ , 可认为总体上血糖与胰岛素存在直线回归关系。

(2) 也可用  $t$  检验与区间估计推断总体回归系数  $\beta$  是否为 0, 与方差分析完全等价。如本例回归系数的 95% 可信区间为  $(-9.9, -4.75)$ , 不包括 0, 表明直线回归方程成立。

(3) 直线回归分析的注意事项

1) 直线回归分为两类, 若只有因变量  $Y$  服从正态分布,  $X$  为可以精确测量与严格控制的变量, 则称为 I 型回归; 若  $X$  同时满足正态分布, 则称 II 型回归, 可进行直线相关分析。

2) 直线回归分析前必须绘制散点图, 若散点图提示两变量间呈非线性关系, 可通过变量变换使其符合条件后, 再行直线回归分析; 若变量变换后, 仍不能满足条件, 则可采取分段直线回归、多分类回归或者非线性回归分析等。

### 三、可信区间

可信区间 (confidence interval, CI) 是按一定的概率  $(1-\alpha)$  去估计总体参数所在的范围。可信区间包括准确度和精度两种属性。准确度是指区间内包含总体参数的可能性, 如总体均数的 95% 可信区间, 准确度为 95%, 在该区间范围内有 95% 的可能性包含总体均数, 也就是说从总体中做 100 次随机抽样, 得到 100 个可信区间, 那么理论上有 95 个可信区间包含被估计的总体均数; 精度是指可信区间的宽度, 宽度越小, 则精度越高。精度与样本量和准确度有关, 在样本量固定的情况下, 准确度不能太高, 准确度越高, 精度越差, 反之亦然, 因此, 多数统计分析软件包常选用 95% 可信区间作为默认值。增加样本量可提高精度。

可信区间同样具有统计推断的功能, 与假设检验相比, 可信区间能够提供更多的信息。如两组样本均数比较资料中, 若两组均数差值的 95% 可信区间不包括 0, 说明差别有统计学意义, 反之, 则无统计学意义。同时可信区间能显示出差别的程度, 并由此判断差别有无实际价值或临床意义。但可信区间不能提供确切概率。

在临床研究中, 常易将 95% 可信区间与 95% 医学参考值范围混淆。95% 医学参考值范围是指同质总体中包含 95% 个体观察值的估计范围, 计算 95% 正常参考值范围用标准差, 而可信区间估计则要使用标准误。

#### (一) 单个总体均数与总体率的区间估计

##### 1. 单个总体均数的可信区间

当总体标准差  $\sigma$  已知时, 总体均数的  $1-\alpha$  可信区间为:  $\bar{X} \pm u_{\alpha} \sigma / \sqrt{n}$



若总体标准差  $\sigma$  未知,  $1-\alpha$  可信区间为  $(\bar{X}-t_{\alpha, \nu} s/\sqrt{n}, \bar{X}+t_{\alpha, \nu} s/\sqrt{n})$ 。

以例 11-1 中传统康复方法组 (39 例) 为例, 半年后其功能恢复评分为  $35.6 \pm 28.11$  (分), 计算该组功能恢复评分的 95% 可信区间。

查表得  $t_{0.05, 38} = 2.024$ , 其 95% 可信区间为:  $35.6 \pm 2.024 \times 28.11/(39)^{1/2}$ , 其下限是  $35.6 - 9.1 = 26.5$  (分), 上限为  $35.6 + 9.1 = 44.7$  (分)。

2. 单个总体率的可信区间 在样本例数 ( $n$ ) 与样本率 ( $p$ ) 和  $(1-p)$  的乘积均大于 5 的条件下, 采用正态近似法估计总体率可信区间。

$$p \pm u_{\alpha} \sqrt{p(1-p)/n} \quad \text{公式 11-11}$$

以例 11-3 中的葛根素组为例, 已知该组治疗 100 例, 其中有效 40 例, 有效率为 40%。同时  $np=40$ ,  $n(1-p)=60$  均大于 5。故采用正态近似法计算总体率的 95% 可信区间:  $0.4 \pm 1.96 (0.4 \times 0.6/100)^{1/2}$ , 其下限为 0.30 (30%), 上限是 0.50 (50%)。

## (二) 两组均数或率比较的可信区间

临床研究中常把有临床意义 (或实际价值) 的差值称为效应量 (effect size)。效应量的表达形式与数据类型有关。如在临床研究中, 对于数值变量资料采用均数差值作为效应量表达。

### 1. 均数差值的 95% 的可信区间

$$\bar{x}_1 - \bar{x}_2 \pm t_{\alpha} SE_{diff} \quad \text{公式 11-12}$$

$$SE_{diff} = \sqrt{\frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2} \times \left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$$

其中  $n_1$ ,  $n_2$ ,  $\bar{x}_1$ ,  $\bar{x}_2$ ,  $SD_1$ ,  $SD_2$ , 分别为两组的例数、均数、标准差。

以例 11-1 预后研究为例:  $t_{0.05, 82} = 1.99$ ,  $SE_{diff} = 28.27^* (1/39 + 1/45)^{1/2} = 6.18$

两组均数差值  $= -15.8$ , 其 95% 可信区间的下限是  $(-15.11 - 1.99 \times 6.18) = -28.11$ , 上限为  $(-15.8 + 1.99 \times 6.18) = -3.5$ 。该可信区间不包括 0, 说明两组间的差别有统计学意义。

2. 两组率比较的区间估计 分类变量资料中常以率的差值作为效应量, 关系强度效应量则采用相对危险度 (RR)、比值比 (OR)、NNT 等。

(1) 率差的区间估计: 率差定义为  $p_1 - p_2$ , 则其可信区间为  $(p_1 - p_2 \pm u_{\alpha} SE_{diff})$ , 其中  $p_1$ ,  $p_2$  为试验组与对照组的不良事件发生率,  $SE_{diff}$  为率差的标准误,

$$SE_{diff} = \sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}} \quad \text{公式 11-13}$$

以例 11-3 葛根素疗效研究为例, 已知两组有效率分别为 40% 和 10%, 则效应量为  $(0.4 - 0.1) = 0.3$ , 其 95% 可信区间为 0.19~0.41。区间范围不包括 0, 说明率差有统计学意义, 与  $\chi^2$  检验的结论完全一致。

(2) 关联强度的区间估计: 分类变量资料关联强度的效应量常用 RR、OR、NNT 等。相对危险度 (RR) 是试验组的不良事件发生率与对照组不良事件发生率的比值; 比值比 (OR, 又称机会比) 则是病例组的暴露与非暴露比和对照组的暴露与非暴露比间的比值, 为两个比值之比。

因相对危险度 (RR) 与比值比 (OR) 服从对数正态分布, 所以 RR 与 OR 的 95% 可信区间分别按下式估计:

$$\exp \left[ \ln (RR) \pm 1.96 \sqrt{\frac{1 - [a/(a+b)]}{a} + \frac{1 - [c/(c+d)]}{c}} \right] \quad \text{公式 11-14}$$

$$\exp \left[ \ln (OR) \pm 1.96 \sqrt{\frac{1}{a} + \frac{1}{b} + \frac{1}{c} + \frac{1}{d}} \right] \quad \text{公式 11-15}$$



以例 11-3 探讨葛根素治疗研究为例。已知葛根素组无效率为 60%，对照组为 90%，则  $RR = (60\%)/(90\%) = 0.67$ ，其 95% 可信区间为 0.56~0.79。

同样葛根素组中无效与有效的例数之比为 60/40=1.5，对照组无效与有效的例数之比为 90/10=9， $OR = 0.167$ ，其 95% 的可信区间为 0.07~0.37。 $RR$  与  $OR$  的 95% 可信区间均不包括 1，说明疗效差异与葛根素有关，且有统计学意义。

## 第四节 统计分析方法的选择

临床研究中应正确抉择统计方法，合理分析研究资料。统计学方法是一把双刃剑，用对了可以为临床研究服务，起到事半功倍的作用；用错了则会影响研究的真实性，甚至得到错误结论。因此在临床研究中，要正确使用统计分析方法，应充分考虑研究目的、设计方案、数据类型以及统计方法的应用条件等因素，避免统计分析方法的误用与滥用。统计分析方法的正确抉择应重点考虑以下几方面：

### 一、与研究目的及分析目的有关

进行统计分析之前，一定要明确研究目的。欲分析两种或多种干预措施间的效果有无差别时，常用  $t$  检验、方差分析、卡方检验、秩和检验等；若研究两个或多个因素间的关系，进行关联分析时，可选用相关分析来衡量各因素间的密切程度和方向，用回归分析揭示某个因素与一个或多个因素（变量）间的依存关系或者因果关系。

方法的选择同样与分析目的有关。如临床研究中常见的单向有序分类变量资料，像临床试验中，疗效按照痊愈、显效、有效、无效等分类汇总。若直接采用  $\chi^2$  检验，则只能回答两组在疗效分类构成上有无差别，而不能回答两组中何者的疗效更好，要达到此分析目的，只能借助于秩和检验。

### 二、与数据类型以及变量个数有关

数据类型不同，选用的统计方法不同。例如同样是组间差别比较，数值变量资料需选用  $t$  检验或方差分析，而分类变量资料要用  $\chi^2$  检验。

前面介绍了临床研究中常用的一些统计方法，这些方法均有一些共性，如都有应变量和自变量或者独立变量。自变量可以是一个，如只有组别因素，也可以是多个，除了组别因素，还可同时研究多个混杂因素。因此这些方法，实际上可归纳为分析一个应变量与一个或多个自变量的关系。

若只有一个自变量与一个应变量，可选择表 11-8 中的方法。如应变量为数值变量，选用  $t$  检验与单因素方差分析。应变量为分类变量资料时，选用  $\chi^2$  检验。

若分析多个自变量与一个应变量的关系，可选择表 11-9 中的方法。如果应变量为数值变量，可选用多元线性回归；若应变量为分类变量资料，可选用 Logistic 回归分析。

表 11-8 涉及两变量的主要分析方法

独立变量	应变变量	方法
分类变量	分类变量	$\chi^2$ 检验
分类变量（二分类）	数值变量	$t$ 检验
分类变量（多分类）	数值变量	单因素方差分析
分类变量	数值变量（有截尾值）	生存分析
数值变量	数值变量	直线回归



表 11-9 涉及三个及三个以上变量的分析方法

多个独立变量	应变量	方法
分类变量	分类变量	对数线性回归
分类变量或数值变量	二分类或多分类变量	Logistic 回归
分类变量或数值变量	二分类或多分类变量	判别分析
分类变量	数值变量	方差分析
数值变量	数值变量	多元线性回归
数值或分类变量	数值变量 (有截尾值)	COX 风险比例回归
分类变量伴混杂因素	数值变量	协方差分析
	分类变量	Mantel-Haenszel 法
有多个数值变量	无	因子分析或聚类分析

(详见有关统计参考书)

### 三、与设计方案有关

不同设计方案,对应着不同统计方法。若与设计方案脱节,也可能影响统计结果的真实性,甚至出现错误结果。例如数值变量资料中配对设计的两组差别比较,应选用配对设计  $t$  检验,若选用成组  $t$  检验则会降低检验效能。

### 四、与应用条件有关

统计分析方法是基于数理统计与概率论,并在一定假设条件下推导建立的。只有满足了这些应用条件,数理推导才成立。

例如,很多方法的应用条件都与样本含量有关。像四格表  $\chi^2$  检验要求样本量大于 40 且最小的理论频数大于 5,成组  $t$  检验要求样本含量不小于 30 例等。

## 第五节 统计分析结果的正确解释与评价

### 一、统计分析结果的正确表达

完整的统计分析结果应同时包括假设检验与区间估计结果。假设检验是以统计量的抽样分布为理论依据,根据统计量与自由度的大小来确定  $P$  值。 $P$  值则是在由检验假设所规定的总体中做随机抽样获得等于及大于(或等于及小于)现有统计量的概率,再通过与检验水准  $\alpha$  比较,作出组间总体参数是否有差异的结论。以  $t$  检验为例,若  $P \leq \alpha$ ,说明两总体均数间的差异有统计学意义,  $P > \alpha$ ,表明差异无统计学意义。若  $P$  值在  $\alpha$  附近时,应具体表明确切概率值。

假设检验只能表明差别有无统计学意义,但不能说明差别的程度以及是否实际意义。而可信区间却能提供更多的信息,既能表明差别有无统计学意义,同时又能显示差别程度,并由此结合临床专业知识判断有无临床价值或实际意义,但可信区间不能提供确切概率。

因此统计结果的正确表达应是  $P$  值与可信区间相结合,两者同时报告。

### 二、正确解释统计结果

1. 统计结论具有概率性,不能绝对肯定或否定,统计推断可能出现 I 型或 II 型错误(表 11-10)。



表 11-10 假设检验中的 I、II 型错误

真实情况	假设检验结果	
	拒绝 $H_0$	不拒绝 $H_0$
事实上 $H_0$ 成立	$\alpha$	推断正确
事实上 $H_0$ 不成立	推断正确	$\beta$

真实情况是总体参数间无差别 ( $H_0$  成立), 但统计推断出有差别, 推断结论与真实情况不符, 则犯了错误, 称之为 I 型错误, 大小用  $\alpha$  表示, 相反则推断正确; 若真实情况是总体参数间有差别 ( $H_0$  不成立), 统计却推断出无差别, 结论与真实情况不符, 则也犯了错误, 称之为 II 型错误, 大小用  $\beta$  表示。

因此当  $P \leq \alpha$ , 差别有统计学意义 (或称阳性结果) 时, 有可能犯 I 型错误; 当  $P > \alpha$ , 差别无统计学意义 (或称阴性结果) 时, 有可能犯 II 型错误。

I 型错误、II 型错误、样本含量三者密切相关。当样本含量固定不变时, I 型错误率降低, II 型错误率将增加, 反之亦然; 样本含量增加时, 可使 I、II 型错误率同时降低。

2. 无统计学意义结果 (阴性结果) 与有统计学意义结果 (阳性结果) 同样重要。特别是在样本含量偏小, 出现阴性结果时, 要格外注意, 有可能犯了 II 型错误, 得到假阴性结果。

因此在临床研究中, 若得到阴性结果 (无统计学意义) 时, 应评价结果的真实性, 方法之一就是考察检验效能 (power)。检验效能是指事实上总体参数间确实存在差别, 推断正确 (假设检验拒绝  $H_0$ ) 的可能性大小, 用  $1 - \beta$  表示。若检验效能为 0.8, 是指当总体参数间确有差别时, 做 100 次假设检验, 其中 80 次能检验出有差别。

检验效能的计算, 实际上是由样本含量估算公式演化而来, 下面列出了两类资料检验效能的计算公式, 有兴趣者可以验算。

#### (1) t 检验的检验效能估算方法

$$u_{\beta} = \sqrt{\frac{n_1 n_2}{n_1 + n_2}} \times \frac{(\bar{x}_1 - \bar{x}_2)}{\sigma} - u_{\alpha}, \quad \sigma = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}} \quad \text{公式 11-16}$$

#### (2) $\chi^2$ 检验的检验效能估算方法

$$u_{\beta} = \sqrt{\frac{2n_e n_c}{n_e + n_c}} \times \frac{(p_e - p_c)}{\sqrt{p_e(1-p_e) + p_c(1-p_c)}} - u_{\alpha} \quad \text{公式 11-17}$$

如果临床研究的检验效能在 80% 以上, 则可认为结果真实性较好。

### 三、统计学意义与临床意义的综合评价

临床研究的最终目的是创造最佳研究证据, 为临床实践服务。因此一个临床研究仅有统计学意义是不够的, 还应结合临床专业知识, 考察其临床价值。例如在高血压干预研究中, 结果测量指标为收缩压的降低值, 在样本量足够大 (如超过 20000 例) 时, 即使两组的差异仅为 2mmHg, 结果仍可能有统计学意义, 但 2mmHg 的临床意义却不大。又比如在一个临床研究中, 差异有临床价值, 即便无统计学意义, 也应重点关注, 必要时需要扩大样本量, 进一步研究。临床意义与统计学意义评价判断见表 11-11。

表 11-11 临床意义与统计学意义评价判断表

统计学意义	临床意义	结果评价
有	有	可取
有	无	临床价值不大
无	有	? 继续
无	无	不可取



#### 四、统计分析结果的真实性评价原则

1. 研究方案的设计是否科学合理？

对照设置是否合理？组间的均衡性是否满足？是否随机？随机方案是否隐匿？

2. 统计分析结果是否全面？

在临床研究中的利弊结果应同时报告，不能报喜不报忧。这里“利”主要指疗效，“弊”指副作用（或不良反应）与费用等，无论假设检验结果如何，都应如实报告；除报告主要研究结果外，还应报告失访和（或）未纳入分析的研究对象数量与原因。

3. 统计分析方法的选择是否合适？

数据资料是否满足应用条件？选择的方法与分析目的是否匹配？

4. 是否考虑了混杂与偏倚因素并纳入分析？

混杂与偏倚直接影响结果的真实性，造成真实效应低估及假阴性结果。对于能够准确测量的混杂因素，可采用多元统计分析方法加以控制；对于不能够准确的混杂因素，则需要通过严格的设计加以控制，如分层、配对或盲法及隐藏等（图 11-5）。

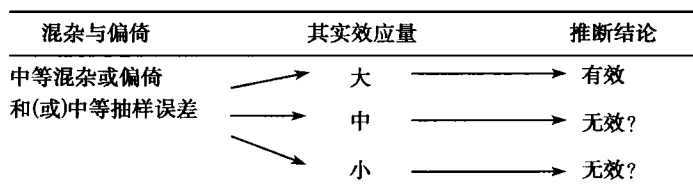


图 11-5 混杂与偏倚影响示意图

5. 结果解释是否综合考虑了统计学意义与临床价值？统计结果的适用性如何？

在临床研究中，统计分析结果实际反映了效应的平均水平，个体效应可能高于或低于平均水平，因此在推广应用时，忌生搬硬套，要注意分析总体与有关亚组情况。

（康德英 王滨有）



## 第十二章 临床研究中的伦理学

生物医学研究是生命科学技术的主体，是认识疾病本质及其防治规律而进行的一项实践活动，是控制和预防疾病、维护和增强人体健康所必需的。涉及人体健康的生物医学研究要求必须有人类受试者作为研究对象接受试验，其试验成果才能被推广应用。因此，在人类受试者参与的研究中，既有如何确保科学性的科学技术方面要求，又有如何确保伦理性的人文方面要求。医学研究如何最大程度地保障人类的安全、利益与公平，它给受试者和人类带来的是幸福还是灾难？医学研究者必须对此作出评价，这就涉及医学研究中的伦理问题。尽管科学研究对社会非常有益，但同时也引发了一系列的伦理道德问题。

### 第一节 概 述

西方“伦理”一词，是从希腊文 *ethos*，即风俗、风尚、性格演绎而来。在公元前 4 世纪，由希腊著名哲学家亚里士多德创立的一门以道德品质为研究对象的学问。中文中的“伦”指“人与人之间”，“理”指“道理、准则”，而“伦理”就是指人与人相处时应遵循的规则和道理。因此，伦理学（*ethics*）就是研究人与人、人与社会之间行为规范和原则的一门科学。它不仅让人们知道待人处事应当遵循的准则，更重要的是培养人们自觉地按照一定规范来支配自己的行动。随着现代社会的发展，人与人之间、人与社会之间的关系也越来越密切和复杂，同时不断显现出许多新的伦理问题，进一步促进了伦理学的发展。

医学伦理学（*medical ethics*）是在医学研究与医疗实践中逐渐形成和发展起来的，用以研究医疗卫生人员与服务对象以及医疗卫生人员之间行为规范的一门科学，是伦理学的应用发展，为提高医疗卫生人员的道德水平，推动医疗卫生保健事业的发展服务。例如，晚期肿瘤病人姑息照顾中的伦理原则？急危重病人上下级医院之间的会诊、转诊行为规范？等等。

研究伦理学作为应用伦理学的一个分支，是应用一般道德理论、原则和规范探讨科学研究领域中的伦理问题及其解决方法。其中，涉及人体的生物医学研究的伦理问题最为关注，生物医学伦理学的目的就是探讨生物医学研究和临床试验中的伦理问题。它除了具有一般科研共同的特征，如探索性、创新性、复杂性外，还有其自身的重要特点，即研究对象的生命属性。要求在医学研究的不同环节中，应做到最大限度地保护人类受试者的权利和利益。

### 第二节 生物医学研究中的伦理问题

涉及人的生物医学研究对象有临床中的药物、器械、手术、影像等试验，以及疾病的流行病学调查等。研究的样本可能是人体，也可能只是医疗记录或生物标本。因此，伦理问题涉及医学研究的不同环节（课题选择、试验设计、资料收集、干预试验、成果鉴定和论文发表等）。为此，我们需要了解与伦理问题相关的生物医学研究的主要特性。

#### 一、研究对象的特殊性

现代医学研究在经过实验室研究和动物实验之后，最终都将在人体身上进行验证。因



此,要求研究者遵循医学研究中的伦理要求,保障受试者健康。人的生命只有一次,受试者的健康与利益应高于医学研究本身。在进行科学研究时,有时会出现利益冲突,如受试者利益、研究人员利益、资助者利益以及社会利益之间发生利益冲突时,伦理问题随之产生。

其次,医学研究中的受试者除了生物属性外,还具有社会属性。医学研究的结果除了影响人体的健康(促进或危害),也将不同程度地影响社会关系。如现代诊疗设备的使用、人工生殖、基因工程、器官移植的发展,大大提高了人们诊治疾病的能力,但同时也带来了令人忧虑的伦理社会冲突,对医学研究的善恶评价显得更加复杂。

## 二、研究结果的不确定性

医学研究的目的是希望发现未知证据或者检验一种尚不能肯定其结果的假定。与许多科学研究一样,生物医学研究的结果或有益于人类,或给人类带来灾难?我们事前也不得而知。许多基础试验能在动物模型中成功,但不能把这些结果等同于人体的生理、药理和毒理反应指标。目前认为,所有创造性的干预措施,不管是诊断、预防还是治疗的,最终都要在人类个体上进行评估。因此,必须强调人类试验的安全性。例如,20世纪食道癌术后加放疗是否有利于提高患者生存率研究。由于众所周知的放疗副作用,加用放疗的安全性和预期效果就备受关注和争议,参加此次试验的受试者是否受益未知。但若进行研究,将有利于未来病人选择合理治疗方式,但伦理问题由此产生。如果试验结果显示生存率降低,那么参与研究个体受到双重伤害,另一方面,如果放疗提高了生存率,那么接受放疗的患者就等于丧失了一个较好的治疗机会,不公正现象由此产生,出现了医学伦理问题。因此,医学研究者应充分估计科研活动可能产生的人体损害和潜在危险,并事先提出相应的补救措施,减少对人体的伤害,并做到知情同意,以最大程度地保障人体安全。

因此,建立生物医学研究的伦理原则并遵循之是进行医学研究中一个非常重要的问题。

## 第三节 涉及人体生物医学研究的伦理基本原则

### 一、涉及人体医学研究的国内外伦理法规

涉及人体的生物医学研究史上,发生了许多漠视甚至侵犯受试者利益的丑闻事件,这些危害受试者健康的深刻教训受到各国及国际组织的关注,并以此为基础总结制定了维护受试者权益的国际伦理准则和法规,以保护受试者,使人体研究顺利进行。

第二次世界大战期间,纳粹德国曾迫使集中营中受害者参加许多惨无人道的“生物医学试验”,如众所周知的“低温冷冻”、“芥子气”试验等,造成了数百万人的死亡。这些受害者是在被迫的、完全不知情的情况下参加的,根本没有选择与拒绝的权利。二战结束后,这些不人道的人体试验才得以披露,并引起了公众的广泛关注。国际社会组织在德国纽伦堡审判战犯的国际军事法庭上,揭发了这些罪行,并形成了10点声明,由此诞生了第一部关于人体试验研究的国际伦理准则《纽伦堡法典》(Nuremberg Code, 1946年)。这部国际性法典的核心内容是:在没有得到“自愿同意”前不能进行人体试验。这一条在以后的各项准则中一直保留。

1964年世界医学会在世界卫生组织的协助下,以《纽伦堡法典》精神为模版,并经完善补充后,在芬兰赫尔辛基召开的第18届世界医学大会上正式通过了关于生物医学研究伦理准则,即《赫尔辛基宣言-指导医生进行人体生物医学研究的建议和伦理原则》



(Helsinki declaration), 并先后于 1975、1983、1989、1996、2000 年修订了 5 次, 共有 32 条细则, 主要内容包括: ①医学研究的目的, 即人体试验必须以促进疾病诊治、预防水平提高以及对疾病病因的认识为目的, 从根本上促进医学的发展和人类的健康; ②人体试验必须遵从公认的科学原则, 要有充分的实验室与动物实验基础, 以及进行科学的设计; ③人体试验必须以尊重和维护受试者的利益为根本前提, 进行利弊分析, 把可预见的对受试者及他人的益处与预测风险进行比较, 不能只顾及科学和社会利益而牺牲受试者的利益, 为此应密切观察试验反应, 一旦出现严重意外, 立即终止试验; ④进行任何人体试验都必须得到受试者的知情同意; ⑤涉及人类的试验设计应详细描述, 并由一个独立的委员会审查等。详见美国医学杂志中文版 2001 年第 1 期。目前已成为指导“以人体为对象的医学研究的伦理准则”, 为各国医学界普遍接受。

另外, 国际医学科学组织理事会 (Council for International Organizations of Medical Sciences, CIOMS) 于 1993 年制定了《涉及人体的生物医学研究国际伦理准则》, 探讨医学研究中具有普遍性的伦理问题, 在伦理评价原则的基础上, 还详细制定了伦理评价的具体步骤, 并于 2002 年 2 月再次在 CIOMS 专家会议上进行讨论修订。

由于各国社会、文化、政治、宗教信仰等背景差异, 决定了对待和处理医学研究中伦理问题的态度和方法会有所不同。因此, 各国以《赫尔辛基宣言》为基本准则, 结合当地现状相继制定了一些涉及人体医学研究的伦理法规和文件。我国卫生部于 1998 年颁发了《药品临床试验管理规范 (试行)》(good clinical practice, GCP)。国家药品监督管理局在 1999 年发文明确要求, 在应用药物的临床试验中, 应“遵守伦理道德, 保障受试者权益”, 临床试验方案及其修改必须经过伦理委员会审查, 并尊重伦理委员会的意见和建议, 试验过程中应严格按照 GCP 要求向受试者说明有关试验的详细情况, 获取知情同意书, 保证受试者依从性等伦理管理法规。2001 年 7 月卫生部伦理专家委员会颁布了《人体生物医学研究伦理审查办法》(草稿), 旨在规范涉及人体的生物医学研究和相关技术的应用, 尊重和保护人类受试者的合法权益, 保护人的生命和健康, 维护人的尊严。要求部级以上的人体试验必须按此要求进行伦理规范。运行 6 年后于 2007 年 1 月 11 日正式公布了《涉及人的生物医学研究伦理审查办法 (试行)》, 分五章, 共三十条, 主要包括: 涉及人体生物医学研究伦理审查原则, 伦理委员会的设置, 伦理审查的程序、方法, 以及审查的监督与管理等。2001 年 11 月卫生部还专门颁布了《人类胚胎干细胞研究的伦理原则和管理建议》, 进一步规范了一些特定领域研究中的伦理学管理问题。

## 二、涉及人体研究的基本伦理原则

上述法典或宣言所包含的伦理准则多为条款性, 在大多数情况下有一定的指导意义和参考价值, 但有时面对一些复杂的问题往往不适合, 常常很难理解或应用, 甚至互相矛盾。因此, 迫切需要更为广泛的基本伦理原则, 为制定具体的条规、评论及解释提供理论基础。如 1974 年 7 月, 美国国家科研法案 (公共法则 93348) 出台, 成立了保护生物医学研究人体试验对象的全国委员会, 主要任务之一就是为以人体为对象的生物医学研究确定基本的伦理原则, 并对有关科研进行指导监督。经过委员会专家组的多次讨论和审议, 于 1976 年 2 月在 Smithsonian 机构 Belmont 会议中心正式发表了“Belmont 报告”, 确定了所有涉及人体生物医学研究都应遵循的三条基本伦理原则, 即“Belmont 原则-尊重人、有利和公正”。为研究中伦理问题的解决提供一个指导框架。

### (一) “尊重人”原则 (respect for person)

“尊重人”是指研究者应尊重受试者的决定并保护他免遭伤害, 包括两个基本伦理要求。第一, 尊重自主权, 对每个个体都应看作是拥有自主权的道德行动者。这要求对那些



有能力考虑其个人选择的人，应该尊重他们的自我决定能力；第二，自主权受损的人理应得到保护。这要求对那些弱势群体主动提供安全保护，使其免受伤害。

以人体为对象的研究中，试验者和受试者对研究内容的认识以及相关医疗信息的掌握存在很大的不对称性，对试验研究的理解认识角度和程度都会有较大偏差；再加之研究者对研究科学价值的期待偏好等因素，可能会有意无意地对受试者隐瞒一些信息，以便让更多的受试者参加试验。从伦理角度看，违背了尊重原则。

所谓“自主权”就是每个个体独立、自愿做出决定的权利。尊重自主权就是承认有自主权个体的意见和选择。有自主权的个体能够熟悉个人目标，不管他是否身患疾病，均应享有选择决定自己行为方式的权利。有独立自主权的受试者对自己行将参加的临床试验，可选择接受或拒绝，不应受内在疾病因素和外界环境因素的干扰影响。即使试验可能会给受试者带来益处，研究人员也不得强迫其参加，只能耐心解释，使其自愿参加。在试验的过程中，受试者最终将承受试验的一切未知结果，他们有权了解试验的可能利弊，享有对试验的知情同意权，并可以随时选择退出。

然而一个人的自我决定能力并不是一成不变的，既可随年龄增长而成熟，也可能由于疾病、自由处境受限等原因，部分或全部丧失这一能力。易受伤害和缺乏自我保护能力的人群常被定义为“弱势人群”。自主权受损包括内在和外在两方面因素，最常见的内在限制包括儿童因为年幼，老年出现痴呆，精神出现智障等形成的“弱势”，使得他们不能准确的理解和表达对试验的认识；纳粹时期强迫使用犯人为受试者，剥夺了受试者自主权，则是典型的自主权外在限制。虽然历史不会重演，但应注意当今某些特殊环境下的研究也可能限制了受试者自主权。例如，前世界干细胞权威、韩国首尔大学黄禹锡教授的“胚胎干细胞”研究丑闻，就是利用实验室女工作人员的卵子并购买卵子做研究，捏造干细胞试验结果。研究团队成员做受试者，可能会受切身利益的过度影响，使研究的科学性、真实性大打折扣。其他的还有附属实验室和医院工作人员、制药工业的工人、军人作为受试者，以及专门选择一些国家或特定文化背景的人群进行试验等。

对弱势人群受试者应提供主动保护，维护他们的自主权，必要时可拒绝他们参加试验。如除非进行精神病或精神缺陷治疗性研究，有精神缺陷的受试者一般不应纳入试验，这样可大大减少他们在试验中受到伤害的可能性。对于不得不进行的试验研究，应自觉地维护他们自主权利，更加审慎的进行设计和研究。受试者的自主权能力可随不同场合而变，应注意定期审核。

总之，“尊重人”的原则就是最大限度地保证受试者免受伤害。

## （二）“有利”原则（beneficence）

“有利”是指研究者有伦理学义务帮助受试者以确保他们的健康利益。有利原则也包括了两个基本伦理要求。首先，做到“不伤害”，禁止对人体的故意伤害；第二，权衡利害，做到利益最大化，伤害最小化。

“不伤害”是“有利”原则的基本要求。临床试验要保证每一位受试对象都得到最佳效果是不可能的，重要的是不能对他们造成生理、心理和精神上的伤害以及经济上损失。风险是指研究时可能发生的潜在伤害，而伤害是指研究时实际发生的伤害。研究者必须十分审慎地思考风险并判断其大小，不管好处多大也不应造成人的伤害，有时必须放弃研究。如研究中安慰剂的选择问题，若病情较重的受试者，以安慰剂代替常规药物会带来不良后果，这种对受试者的故意伤害行为，严重违背伦理原则。但如果是病情较轻，药物疗效尚有分歧的疾病，选择安慰剂对照则不会对人体造成伤害。

如何避免伤害呢？首先，试验之前应充分评估可能的风险。即使以前的试验表明受试者的伤害风险很低，研究设计时也要认识到所有潜在的风险，建立严谨可靠的试验方案。



任何潜在的伤害都违背了“无害”原则。其次, 评估受试者接受试验后将承受的危险程度。《赫尔辛基宣言》中提到, 进行治疗性和非治疗性试验时, 受试者可接受的危险程度是不同的。如治疗性研究可以给受试者带来较大益处, 改善健康(和以前的治疗相比), 那么值得他冒一次危险参加试验; 反之, 如果是一个非治疗性的试验, 则没有必要去冒险, 以避免伤害。

如何保证利益最大化呢?“有利”是指试验中的任何行为、动机和结果均应有利于受试者。试验之前, 应仔细考虑研究目的, 研究风险是否合理以及预期效益, 明确谁是试验的获益者, 是受试者、科研人员还是社会。就科研本身来说, 应充分认识到知识进步以及医学发展而带来的长期好处。受试者本人常不受益, 但可让未来的病人和社会获益。如果研究能为他人提供更为有效的治疗方法, 受试者忍受一些并不严重、可逆的不适甚至最小程度的伤害, 是可以得到伦理学辩护的。因为现在的病人也是从过去的研究中获益, 他们同样有义务来让未来的病人获益; 但如果答案是否定的, 即使受试者忍受最小程度的伤害也都是不允许的。就具体课题来说, 科研人员必须事先筹划。如高血压的新药研究, 目前已有许多疗效明确的降压药物, 受试者使用新药获益不大, 还存在一定的风险, 新药的风险程度就需要评估。

为达到“有利”于受试者的目的, 试验选用的一切措施都应遵循最优化原则。设计应科学合理, 研究者应具备足够的研究能力和保护受试者福利能力, 在实验室和动物试验基础上选用预期利益最好的诊治措施, 试验对照也应选择现代最佳的诊治措施, 降低研究可能带来的危险。为满足研究需要, 有意选用效果不好的措施作对照, 是违背“有利”原则的。研究实践中, 关键在于时机的把握, 何时尽管危险也应追寻证据, 何时由于危险性太大而放弃试验。

总之, “有利”原则就是确保受试者健康获益, 免受伤害。

### (三) “公正”原则(justice)

“公正”是指任何群体或阶层都应公平承担研究相关的责任和义务, 公平享有研究相关的短期或长期收益。包括参与研究的直接获益, 以及研究所得的新知识利益。“公正原则”要求研究者按照适当的方式分配每个人的应得利益及参与研究的义务。不论受试者的地位高低、职业差异, 应“分配公正”, 同样对待。对于弱势人群, 还应主动地保护他们的权利, 避免受到不公正的待遇。

利益分科学利益、社会利益和受试者利益等。任何试验结果都会丰富知识, 促进社会进步与医疗水平的提高, 科学利益显而易见。社会利益与社会义务是紧密联系的, 需要综合比较, 才能明确谁是真正的受益者。当应得利益被剥夺, 或者不恰当地被强加风险, 不公正便产生了。例如 19 世纪末及 20 世纪初, 试验对象大多是贫穷的病人, 而医疗水平提高的受益者却大都是富有病人(即有支付能力的人), 不公正现象由此产生。因此, 利益应公平分担, 科学和社会利益的获得, 不能以伤害一部分受试者为代价。

弱势群体和特殊阶层(如福利救助病人, 贫困人群、特别是被限制人身自由的犯人等)因弱势地位常被选出受试, 公正原则要求不应过度使用这些弱势群体。当干预措施不能使受试者直接获益时, 必须慎用弱势人群为受试者, 以免加重负担。被过度使用的人群还包括一些年资较低的住院医师、研究者的学生及下属, 以及一些贫困社区居民等。

另外, 试验中的试验组和对照组, 不论何组受试, 都应得到同样水平的医疗措施和服务。

因此, 应用完善的制度来保护受试者权利就显得十分重要。在医学研究实践中, 有几条公认的责任和利益的分配标准可以借鉴, 如: ①每人平分; ②根据个人需要; ③根据各人的努力; ④根据每人对社会的贡献; 以及⑤根据每人的功绩等。



## 第四节 知情同意

在涉及人体的生物医学研究实践活动中，面对伦理问题，应用伦理学基本原则保护受试者权益时，通常是通过知情同意来实现。

### 一、知情同意的目的

“知情同意”（informed consent）是研究者提供相关的信息知识，使受试者了解自己在试验中的权利，经过与研究充分讨论，作出是否参加试验的决定。其目的为维护受试者权利和健康。对于一切涉及人体的生物医学研究，研究者在试验开始之前必须取得受试者自愿的知情同意。无自主能力者，则应由其法定监护人按照法律作出允诺，并在伦理审查委员会考虑和批准后方可进行试验。

知情同意包括两部分，首先是“知情”，即让受试者知晓临床试验有关信息并能理解。然后是“同意”，即受试者在无任何胁迫、不正当影响或恐吓下作出自愿参加试验的决定。“知情”与“同意”是相互联系缺一不可的，许多人往往仅注重获得书面的知情同意书，必须强调只有在受试者理解并同意参加试验后方能签署知情同意书。

### 二、知情同意的过程

一般说来，知情同意过程始于试验之前，研究者与受试者对试验中的有关问题进行交流讨论，达成共识之后，再开展试验。但也有在试验中二次进行知情同意，如试验方案变更，必须再次告知受试者，取得知情同意，受试者可以决定继续试验还是退出，并送交伦理委员会备案。因此，知情同意常常贯穿于整个试验过程。

公认的知情同意过程包括三个部分：提供信息，取得理解及自愿参加。第一步是向受试者提供研究相关信息，为知情同意的前提。研究者应简明扼要地撰写一份说明书给病人，内容包括：试验操作过程、目的、潜在危险和预计的好处、其他类似的操作，一定要避免故意截留信息，并且应声明受试者有提问题的机会且可随时退出试验。这些内容的描述最好使用通俗易懂、受试者能理解的文字或语言（如方言等），以口头或书面方式来传递信息。第二步是受试者对试验的理解。由于文化、社会、环境差异，研究者和受试者之间需要良好的互动沟通来确保对信息的充分理解。尤其要注意提供有关潜在危险性的资料并保证受试者对危险性充分理解。给予充足时间和机会保证提问质量，研究者有责任给予诚实、详细和清楚的解释，不能诱导患者参加试验。实施知情同意的环境应做到轻松舒适，避免受试者感到压力。有时，研究者可用口头或书面测验来决定信息的理解程度。最后是自愿同意参加试验。受试者如果同意参与试验，则由受试者或监护人签署知情同意书（informed consent form）并注明日期。当然同意也可以用其他方式，如口头同意等，但前者更多见。实施知情同意过程的研究者也需在知情同意书上签名。对于无阅读能力的受试者，在这一过程中应有一名见证人在场，同时也需见证人签名。对于无自主能力或自主能力不全的受试者，经过伦理委员会审查同意，并由监护人签署知情同意书后，也可进入试验。

需要特别提出的是弱势人群的知情同意。确保自主权受限的弱势人群知情同意，是出于真正的自主自愿，而不是被迫作出的决定；确保无行为能力的弱势人群必须有法律上承认的监护人代行知情同意，并在受试者能力所及的程度内取得其本人的同意或赞同。

知情同意书签署后一式两份，分别由研究机构和受试者保存。必须强调的是：知情同意的讨论过程远比签署一份知情同意书更加重要。



### 三、知情同意书的基本内容

知情同意书一般应包含的基本信息有：研究目的、研究过程、试验计划、潜在的危险和益处以及参加者的权利等。具体如下：

#### （一）项目介绍

告知受试者试验项目的名称、研究目的、试验过程、可能持续的时间（到中心来的次数）、实施程序以及预期的研究结果等基本信息。应解释研究和常规医疗的不同点，使受试者充分了解自己在试验中扮演的角色。

#### （二）危险描述

对受试者需交代试验中可能发生的副作用及其危害程度、副作用的发生率及相应的避免和终止措施。试验中任何可能预见的风险、痛苦、不适，包括对受试者配偶、胎儿的风险均应告知，任何致死、致残危险应详细说明并给于合理解释。一切实事求是，既不能把可能的危险扩大化，也不能无根据地化大为小，这样才符合伦理学要求，保证受试者客观地抉择是否参加试验。

#### （三）利益描述

不能夸大研究可能获得的预期益处，如降低致残率、延长寿命、改善生存质量等。如果研究对受试者没有直接益处，但对其他人或社会有益时，应明白地向受试者说明。这样有利于受试者权衡试验利弊，同意或拒绝参加试验，尊重受试者权利。

#### （四）替代方式

“替代方式”就是告知受试者其他可供选择的本研究外诊治措施等方法。当受试者了解到他们可以选择的所有诊治措施后，方能作出自由选择是否参加试验。研究者应如实告知各种替代方式及其益处和危险，供受试者判断收益风险比，最后作出选择。

#### （五）保密

受试者参加试验这一事实以及试验中的个人资料、有关记录均属保密内容。公开发表的试验结果也应对受试者的身份保密，以尊重受试者隐私权。但伦理委员会审查试验记录时，可以查阅受试者有关资料。如系试验赞助人，需要查看记录，应如实向受试者说明。

#### （六）赔偿

一旦发生损伤，受试者可以得到及时医治。对于较大风险研究，如可能致残应说明，保证提供经济赔偿及费用的来源。

#### （七）关于退出试验说明

受试者参加试验是完全自愿的，在任何阶段退出也可自我决定。拒绝参加或在任何时候退出不受任何歧视，不影响受试者应享有的医疗服务。但应按照试验要求程序逐步退出，以保证受试者安全。特别是当试验对受试者潜在危害存在时，研究者有责任和义务告知提前退出的后果，并定期随访、监测退出者，根据情况适时作出必要的处理，以确保受试者安全健康。

#### （八）关于终止试验的说明

知情同意书中还需说明在某些情况下研究者可以终止试验而不必得到受试者的同意。这些情况是：受试者不按试验方案要求的内容程序进行，受试者健康状况不适于继续参加试验，受试过程中出现严重并发症以及或试验结果提前结束等情况。同样，终止试验也应按照一定的程序进行，保证受试者的健康安全。

#### （九）试验费用

试验之前，应清楚地告诉受试者试验中接受的药物费用由谁支付，是否免费；受试者参加试验可能需要增加的额外费用（检查费、交通费等），由谁承担也应注明。以免发生



不必要的误会和矛盾。

#### (十) 关于联系人的说明

如有问题与谁联系非常重要。知情同意书中需明确的告知研究单位的名称、地址、联系人员的姓名、电话号码等。应保证能随时回答受试者有关试验的询问。

### 第五节 伦理委员会

由于存在利益冲突，单靠研究人员维护受试者权利的作用常常有限，而受试者本身又无法对保护是否得当作出判断。因此，第三方的监督就显得日益重要。1966年美国制定了第一个关于保护人类受试者的联邦政策，要求各伦理审查委员会（Ethics Committee，EC 或 Institutional Review Board，IRB）对每个由美国卫生部资助的研究项目进行伦理学审查。1974年美国卫生教育福利部正式修订了准则，并以联邦法规的形式公布。1991年再次修订，通称为《共同规则》。在美国，IRB 现已发展达到了 2000 余个。任何组织、机构只要符合 IRB 的组成要求并按规定行使其职能，均可组建相应的伦理审查委员会。伦理委员会可由卫生机构、大学、研究所、社会福利机构以及社区人员组成，但必须置于 FDA 的监管之下。在我国，卫生部和一些医学院校也建立了各自 IRB，国家药品监督管理局也规定了每个临床药物试验基地必须设立伦理审查委员会。

#### 一、伦理委员会的目的与作用

伦理委员会的工作目的是维护研究参与者的尊严、权利、安全与安康。WHO 认为研究目的虽然重要，但绝不允许凌驾于研究参与者的健康、福利与医疗关怀之上。

伦理委员会的作用包括两方面：一是保护受试者权利，即保护自主性、隐私、公正等；二是保护受试者的利益，达到利益最大化、风险最低化，风险利益比合理。监督试验方案及其实施过程对受试者是否会造成伤害。

#### 二、伦理委员会的审查内容

伦理委员会应遵循国际公认的伦理准则、遵守国家现行法律和法规，并符合社区的价值观和原则，以《赫尔辛基宣言》和各国当地的医学研究伦理法规指南为工作指导原则，遵循涉及人体研究的伦理基本原则。审查的内容主要是知情同意文件和研究计划书。需要提供的资料还包括研究者工作手册、受试者的有关赔偿措施、实验室和动物试验资料、安全性评价报告等。

所有涉及人体的生物医学研究除应对其科学性进行审查外，还必须进行伦理审查，以确保受试者和有关社区和群体的尊严、权益，并把参与研究的风险减到最低限度。具体细则有：是否把受试者的利益和安康放在第一位；是否遵循了公正原则；是否表现了充分的尊重，知情同意和保护隐私是否得到保证；是否对利益风险进行了认真分析，是否将利益增至最大而将风险降至最低；在科学上是否可靠；是否涉及利益冲突；是否符合现行法律和法规等。

#### 三、伦理委员会的组建与运作

为了从社会的不同层面保证受试者权利得到尊重，伦理委员会的成员组成应包括医学专业人员、科学家、伦理学家、法律专业人员以及有资格代表社区道德价值的非医务人员等。当受试者为特殊人群时，应考虑邀请他们作为委员或邀请他们参加会议，并充分表达他们的观点。这些人群有：教育程度低或为文盲者；某些特定疾病如 HIV 患者；以及受





试者中所包括的雇员、学生、老人儿童等。伦理委员会一般要求由 5 人以上组成，并至少应有一名非临床研究单位，包括男性和女性。

伦理委员会的审查和决策工作，必须独立于临床试验的组织和实施者，也不应受经济、政治、单位、行业等外界压力的干扰。对研究项目、申请书和知情同意书进行独立、有效、及时地审查，保障审查结果的公正和科学。

根据研究项目性质和要求，伦理审查都应该按照伦理学原则和规定的程序，在一个组织健全的正式伦理委员会中进行。可以是国家、地区或单位的伦理委员会，也可以是一些国际性伦理委员会。

伦理委员会工作贯穿于整个试验过程中。试验之前，对试验方案及知情同意等文件审核，批准同意后试验方可开始。伦理委员会可以决定临床研究方案是否需要修改以保证受试者伦理上得到最大程度的保护。有时甚至是否决试验。不过，伦理委员会更多地应是对研究者和资助者进行帮助教育，提高他们按照伦理要求进行研究的能力。在试验过程中，伦理委员会实时监督获批研究项目的实施及其进展情况，并向单位或政府当局报告严重的、持续的不遵守伦理标准的行为。任何方案或措施的变更同样需在得到伦理委员会的批准后才能继续进行。试验过程中出现了任何不良事件，也必须及时向伦理委员会报告，伦理委员会可以在认为必要时撤回对研究项目的批准。

(廖晓阳 林果为)

## 附例：知情同意书

### 给病人的信息

长期单独应用新药 TM 和 TM 与 LA 联合应用的多中心研究

#### (一) 简介

我们邀请你参加一个研究项目，请你阅读本研究的有关内容，了解你在研究中的作用对你决定是否参加本研究非常重要。本研究的参加者需患有冠状动脉疾病（如以前有心脏病发作、心绞痛、以前做过旁路手术、运动研究和冠状动脉造影证实有冠状动脉疾病），下肢血液循环不良，卒中，短暂性脑缺血发作（即发作时间小于 24 小时的短暂卒中）或有并发症的糖尿病。同时，你须了解本研究的情况以及参加本研究的风险，并签署知情同意，你是否参加本研究完全是自愿的。

#### (二) 研究目的

患有冠状动脉疾病、下肢血液循环不良、卒中或有并发症的糖尿病患者增加了进一步患心脏病发作、卒中、心力衰竭或因以上疾病而死亡的危险。最近证实，LA，一种血管紧张素转换酶抑制剂，在像你一样的高危人群中可以减少上述事件发生的危险。另外一种很有前途的新药是 TM，可能与 LA 同样有效，而副作用更少。此外，同时给予 LA 和 TM 可能比单独给予两者中一种更有效。但这需要证实。因此，我们开展本项研究，它将肯定地告诉我们 LA、TM 单独或合并使用是否可以减少心脏病发作、卒中、心力衰竭或因以上疾病而死亡的危险。研究将在全球范围内的大约 800 家医院共 25000 名患者中开展，并将持续 5 年半。在\_\_\_\_\_医院，本研究将由\_\_\_\_\_医生负责。本研究已被我国试验的牵头单位伦理委员会批准，确保不侵犯受试者的权益。

#### (三) 研究过程

如果你同意参加本研究，(a) 将询问关于你以前和目前的健康状况与治疗情况；(b) 进行简单的体检；(c) 你需提供清晨第一次尿标本；(d) 进行心电图检查；(e) 空腹采血进



行包括血糖、胆固醇水平和肾功能在内的血液学检查。同时用一种特定的试纸检查以诊断你是否患有糖尿病。我们获得这些资料将有助于将来比较你的病情变化。

完成上述检查后，你将每天服用 2 片研究药物。它们是 40mg TM 或外观相同的 TM 安慰剂（非活性的模拟药片）以及 LA（每天最多 5mg）。你将服用这些药物 3 周，以便了解你是否有可能出现药物副作用（仅极少数）。此后，我们将安排你进行肾功能和血钠、血钾水平的检测。

上述过程结束后，我们将对你进行进一步检查，包括有关记忆和思维能力的问卷检查。如你能耐受研究药物，你将被随机分配到不同的治疗组接受相应的药物治疗。你接受下列三组药物之一的机会相同：（a）LA（每天最大剂量 10mg）+ TM 安慰剂；（b）TM80mg + LA 安慰剂；（c）LA（最大剂量 10mg）+ TM80mg。这些药物每日服用一次。你和你的医生都不知道且不能够决定你将接受哪一组药物。这种盲法的作用是公正地验证药物所需要的。在紧急情况下，你的研究医生能够知道你正在服用的是哪一组药物。

在开始服用研究药物后的第 6 周、第 6 个月以及其后每间隔 6 个月直至 5.5 年，你需要来医院进行定期随访。我们将安排你在第 4 周和第 6 周之间进行肾功能和血钠、血钾水平测定以确保其正常。在每次随访时，将询问关于你的健康和治疗用药情况，并进行血压、心率的测定。此外，在一些随访中，我们还将安排你进行血液学和心电图检查。

请你持续服用研究药物并按要求定期随访直至研究结束。如果你在研究期间出现心脏病发作、卒中、糖尿病或其他疾病，你可以得到相应的治疗而不受到本研究的限制。

#### （四）参加本研究可能的风险

本研究将验证 2 种药物：LA 和 TM。LA 是一种血管紧张素转换酶抑制剂。在加拿大、美国、中国和许多其他国家已被证实批准用于高血压的治疗。它已经在像你一样的几百万人群中使用。LA 已被证实能减少心脏病发作和卒中的危险。在一些病人（大约 5%）、LA 可引起干咳，在极少数患者（1%）可引起轻度头痛、头昏、皮疹、皮肤瘙痒、血液异常、或肝肾功能的一些改变。罕见（发生率 0.4%）但严重的副作用是面部、嘴或咽喉肿胀，后者可能与呼吸困难有关（血管性水肿），此时需要立即采取医疗措施。所有副作用在停药后是可以恢复的。

TM 是一种血管紧张素 II 受体拮抗剂，其安全性已在超过 8000 名个体中做过验证。TM 不引起咳嗽，除偶有腹泻外，其他事件的发生率并不高于安慰剂（头痛的发生率甚至更低）。少数可观察到风疹、皮疹和瘙痒。

当采血时，你可能感觉有些不适和（或）采血处出现瘀斑，罕有轻度感染，在每次随访时询问你有关的副作用。你应及时按照下面所列的电话号码将上述的任何副作用报告给研究护士或医生，这一点非常重要。由科学家组成的独立小组将定期检查研究中的关键资料，以确保参加者在研究进行中的安全。

#### （五）参加本研究可能带来的益处

就你个人而言，接受研究药物治疗可能有益。整个研究汇总的资料对发现一种新的治疗方法意义重大，该治疗方法能减少有危险因素的病人发生死亡、心脏病发作、卒中、或心力衰竭的危险。

#### （六）关于补偿

参加本研究不需要交费。研究药物和每次的随访是免费的。对于你参加本项目研究没有其他的补偿提供给你。

#### （七）责任

在任何研究中，尽管非常谨慎，也可能出现副作用，这并非由于你或研究者的过失所致。所有药物的已知副作用已在上面描述。然而，不可预测的问题偶尔也可能发生。如果



由于你参加本研究而直接导致这些问题，将给你提供必要的和及时的医治。

### (八) 退出的权利

参加本研究完全是自愿的。在任何时候，你都有权利拒绝参加或从本研究中退出而不会受到惩罚或丧失你本该拥有的任何权利。在任何时候，如有下列情况，你也将从本研究中退出而不需要征求你的同意。(a) 你没有按研究方案服药；(b) 发生了不可预测的严重副作用；(c) 本研究已经结束。不论你是主动或被动地从研究中退出，都要求你来医院进行最后一次随访。

### (九) 保密原则

通过签署知情同意书，表明你同意研究人员、研究监察人员、政府注册部门和伦理委员会检查你的医疗记录。如你在其他的医疗机构接受治疗，这些记录也需检查。这种严格的检查是为了保证所有的研究资料的准确性。你的姓名将按照法律的要求严格保密。没有你直接的允许，你的个人资料将不会公开出现在与本研究有关的任何发言及报告中。当研究结果发表时，你的身份不会暴露。没有你直接的允许，你提供的或我们收集的关于你的更多资料将不会泄露给第三方。

你与研究人员联系将获得更多关于治疗的信息。对于你参加本研究的权益，如果你有任何问题，你可以与你的医生\_\_\_\_\_联系，电话：\_\_\_\_\_。

## 知情同意书

在签署知情同意书前，我已阅读或其他人已给我阅读上述信息。我参加这项试验完全是自愿的。我可以在任何时间从这项试验中退出，而不会感到不安或影响我的治疗。研究者也可以终止我参加这项研究，但须向我解释其原因。已给我充分的机会问问题并得到了令我满意的回答。我也清楚地知道如果发生了任何事件，我可以与下列人员联系。

我自愿参加这项研究，我已收到本知情同意书复印件。

\_\_\_\_\_  
参加者姓名（正楷）

\_\_\_\_\_  
签字

\_\_\_\_\_  
日期

\_\_\_\_\_  
见证人姓名（正楷）

\_\_\_\_\_  
签字

\_\_\_\_\_  
日期

获得知情同意书的人宣誓

下表所列一名或两名人员已仔细地向参加者解释了本研究的情况。在此我（们）保证，尽我（们）最大的努力让签署此知情同意书的人清楚地了解本试验的状况、要求，参加本试验所带来的益处及风险。他的签名是有效的。医学问题、语言或教育程度不妨碍对以上情况的了解。

\_\_\_\_\_  
研究协调员（正楷）

\_\_\_\_\_  
签字

\_\_\_\_\_  
日期

\_\_\_\_\_  
研究者姓名（正楷）

\_\_\_\_\_  
签字

\_\_\_\_\_  
日期

研究人员联系资料

	职务/职称	姓名	电话号码
研究协调员			
研究者			

## 第十三章 病因与危险因素的研究与评价

### 第一节 病因与危险因素研究的意义

疾病的病因学研究一直都是医学探索的一个重要领域。基础、临床和预防医学各科专家都致力于疾病病因的研究,但不同学科研究病因的方法、手段、考虑问题的角度各有不同,甚至对病因概念的理解及用于判断病因的标准也不一致。流行病学从群体的角度,应用概率论和逻辑推理的方法探索疾病的病因和疾病发生的影响因素,推动了病因概念的发展和病因研究的深入。随着医学科学的发展,病因学的研究及其评价已经成为每一位医学工作者必备的基本知识。

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效应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。其研究意义表现在以下几个方面。

#### 一、有助于疾病的临床诊断与治疗

通过病因学研究弄清病因,掌握其发病机制和转归,会帮助医师对患者进行正确的诊断和治疗。对疾病需要有针对性的治疗,力求治本,从而获得好的疗效。

医师在临床诊疗过程中,面对患者的实际,总要探讨可能致病的直接病因和危险因素,有针对性地进行相关的体检和化验检查,得到必要的信息,进行临床综合分析,力求获得准确的病因诊断,从而有利于实施更有效的治疗。

如果通过现代的医学手段而未能明确病因的某种疾病,特别是对于人民健康危害性显著者,为了有效防治,则务必首先从病因及发病的危险因素研究着手。例如:我国在20世纪50年代的某些地区于秋收季节,发生了钩端螺旋体病的流行,其中,有许多重症病例,发生肺大出血而死亡,当初怀疑“肺鼠疫”流行,采取了若干重大防疫措施,也造成了一些不必要的影 响,后来弄清了该病系钩端螺旋体感染所致。当病因弄明白后,采取了有针对性地对患者的临床治疗及人群预防,从而大大地提高了钩体病的防治能力,有力地保障了人民健康。

#### 二、有助于疾病的预防

病因学研究的结果,不仅有助于临床诊断与治疗,对于疾病的预防,同样需要了解 and 明确疾病的病因,以便采取针对性措施。

疾病病因研究结果,可以通过各种媒介,对社区人群进行健康教育,从而让社区人群了解疾病的病因和危险因素,提高他们对疾病病因和有效预防措施知识的认知程度,并有意 识地改变自己的生活方式和行为,通过采取去除疾病病因的各项措施预防疾病的发生。

对于某些传染性疾病,由于它们为特异性的病原生物体所致病,当其病因诊断明确之后,往往可以研究特异性高、免疫力强的疫苗,进行人群有效预防,如天花、麻疹、脊髓灰质炎、乙型肝炎等疫苗,都是病因成功研究对这些疾病有效预防的贡献。同时,病原生物体的确定,又可研制特异度高的诊断试剂,进而又提高了疾病的诊断水平。



## 第二节 病因与危险因素研究的基本概念

人类对疾病病因的认识是一个逐步发展的过程。现代医学研究对病因的认识主要经历了单一病因论及多病因论两个阶段。

### 一、病因与危险因素

#### (一) 疾病病因概念的发展

无论中国或外国，古代人常将疾病归因于鬼神、上帝及天意，平日靠求神拜佛或祈祷，以期消灾除病。公元前 5 世纪，由中国古代哲学思想中衍生出了阴阳五行学说——金、木、水、火、土，并用这一学说去解释人体生理现象和病理变化的规律，将疾病的发生与外环境的物质——金、木、水、火、土联系起来，从而产生了朴素唯物主义的病因观。公元前 5 世纪，在西方也出现了类似的观点，反映在 Hippocrates 所著的“Airs, waters and place”之中。他们还认为，与水、土、及风等有关，夏季有脾肿大、发热及腹泻，冬季常引起生痰及喉嘶哑。由此可见，在古代国内外医学家们都注意到疾病与环境有密切关系。我国古代早就认为山间的“瘴气”是疟疾的原因。19 世纪上半叶，Sydenham 等人关于疾病的“瘴气学说 (miasma theory)”在西方盛行。人们认为不洁的水和土壤里散发出来的污浊之气（瘴气）是使人发病的原因。故而强调应设法清除贫民窟和其他不卫生的地方的“瘴气”，以期减少疾病。

1. 单一病因论 16 世纪，意大利学者 Fracastoro 最早提出：特异的疾病与特异的“传染物”有关。拉开了特异病因论的开端。19 世纪，随着疾病微生物理论的发展，Henle 和他的学生 Koch 提出了推断独特的活微生物导致特异疾病的 Henle-Koch 原理，对推动人类病因学研究作出了巨大的贡献。该法则共有 4 条：①病原微生物必然存在于患病动物体内，但在健康动物体内不应出现；②从患病动物分离得到的病原微生物可以做纯培养；③将分离出的纯培养物人工接种易感动物，会出现该疾病所特有的症状；④从人工接种的被感染的动物体内可以再次分离出性状与原有病原微生物相同的纯培养物。Koch 于 1876 年证明炭疽病符合这一原理，随后许多传染性疾病的也得到了证实。该理论在病因学的发展史上是革命性的，为干预措施的实施铺平了道路。

尽管这个法则将有将问题绝对化的缺点，但却反映出当时人们在病因认识方面有不小的进步，即病原微生物是传染病的必要病因，而且每种传染病都有自己的特异的病原微生物。这是关于疾病的生物学病因的重要萌芽。随着 19 世纪末期微生物学的出现和发展，人们形成了这样的概念，即每一种疾病必定是某一种特异的病原物所致。这就是疾病单一病因论的“特异病因学说”。但是它并不能解释复杂的病因效应。因为即使是单一的病因，也可以引起多种疾病，更不用说绝大多数疾病的发生与多种因素共同发生作用有关的情况，因此，单一病因论的有其明显的局限性。

2. 多病因论 医学研究人员在长期的疾病防治实践中逐渐发现，疾病的产生并不完全依赖特异的病原物，还和环境及人体自身的多种因素有关。如在一些非传染性疾病的病因学研究发现，一种疾病的发生往往是多种因素综合作用的结果，而且多种致病因素的危害性要比其中单一因素存在时严重得多。这是由于它们在人体内的致病效应上，会产生交互作用。例如：一个多致病因素致冠心病的多病因交互效应研究，以单纯的收缩压水平为基础，分别与血清胆固醇/高密度脂蛋白浓度比、糖尿病史、吸烟以及左心室肥厚（心电图）等多种危险因素联合研究，观测 5 年以上，以研究这些因素对 55 岁男性发生冠心病的结果（发病率）（图 13-1），可见凡具 5 项危险因素者发生冠心病的危险性为正常者的



21.5 倍，具 2 种危险因素者也达 6~10 倍。因此，在病因学研究中，了解多病因及其交互效应是非常重要的。即使是传染性疾病的发生，也与多种因素共同作用有关。如肺结核，其发生固然需要有结核杆菌的存在，但个体的遗传易感性、营养状况、情绪状况、居住环境状况等均参与人体肺结核的发生过程。

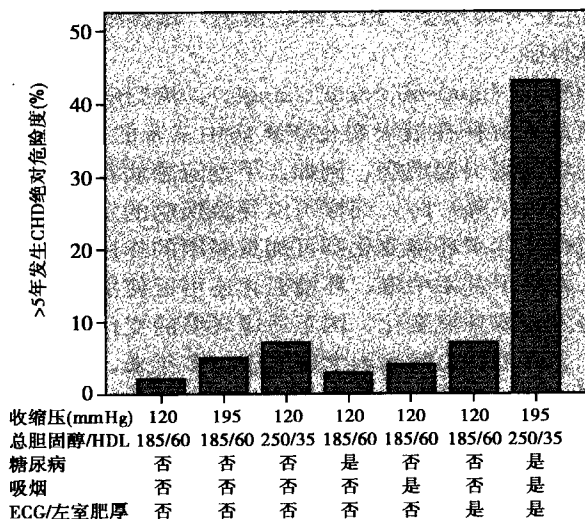


图 13-1 冠心病的多病因交互效应图

(引自 circulation; 1991; 83: 357-363)

3. 概率论因果观 人类对因果关系的认识一直处于发展之中。从古希腊学者亚里士多德 (Aristotle) 提出四原因说，到近代培根 (Bacon) 和休谟 (Hume) 提出决定论的因果观和经验论的因果观，走过了一个漫长的路程。然而，上述的因果论皆不能完美地解释生命现象中的因果关系。现代科学产生了概率论的因果观或称广义因果律 (law of causation)。概率论的因果观认为原因就是使结果发生概率升高的事件或特征，即一定的原因可能导致一定的结果。该观点为解释生命科学中的因果关系奠定了理论基础。

Lilienfeld 从概率论的因果观层面阐述了流行病学的病因概念，他在其所著的《流行病学基础》一书中给出的病因定义是：“那些能使人群中发病概率升高的因素就可以被认为是病因，当其中的一个或多个因子不存在时，人群中疾病频率就会下降。” MacMahon 也认为，流行病学的实际目的是发现能够预防疾病的联系，从这个目的出发，因果关联可以实用地定义为：事件或特征之间的一种关联，改变某一类别 (X) 的频率或特性，就会引起另一类别 (Y) 的频率或特性的改变，这样 X 就是 Y 的原因。因此，流行病学的病因观是符合概率论因果观的，流行病学层面的病因一般称之为危险因素 (risk factor)，这无疑体现了多病因论的思想，冲破了单病因论的束缚。概率论因果观的病因学定义不仅具有病因理论上的科学性和合理性，而且具有重要的公共卫生学意义。

例如：在 20 世纪 60 年代美国心脑血管病的死亡率居高不下，经研究发现高血压病为其主要危险因素，于是在全国开展了大规模高血压防治研究和人群防治运动。约 10 年后，高血压控制率大大提高，脑血管病死亡率大大降低，但冠心病死亡率下降不显著，于是又在全国开展了调节高脂血症的教育与防治，现在全民高脂血症得到了显著控制，心血管病死亡率呈现明显下降趋势，这些病因/危险因素研究符合概率论的因果观，结果表明了其对于疾病防治实践的重要指导意义。

## (二) 病因模型

病因模型以简捷的概念关系图表达因果关系，这种在已有理论和经验基础上构建的概



念关系图，为我们提供了因果关系的思维框架。由于对因果关系有不同的理解或不同的侧重，所以研究者构建了多种类型的病因模型。目前具有代表性的危险因素作用模式有 3 类，即生态学模型（流行病学三角模式与轮状模式）、疾病因素模型和网络模式。

1. 生态学模型 该类模型将机体与环境作为一个整体来考虑。代表性的有流行病学三角模型（又称病因-宿主-环境模型）以及轮状模型，该类模型给出了寻找病因的分类大框架，模型简明，整体性强。

(1) 流行病学三角模式 (epidemiologic triangle): 流行病学三角模式理论用病因、宿主和环境三者的平衡紊乱来解释在健康变化和疾病中三者的作用。若三者处于平衡状态，表现为健康；病因作用增强则引起疾病。例如：甲型流感病毒变异，发生流感流行；当宿主的易感性增强（或抵抗力下降）时，同样可以导致疾病的发生。例如：交通不便的农村相隔多年后发生麻疹流行。如果环境变化则可加重病因的作用。例如：夏季多蚊，乙型脑炎流行；松花江污染，发生甲基汞中毒。环境变化还可促使抵抗力下降，例如：战争时期的营养不良、结核病多发；或环境变化，因不适应环境的变化而发生疾病（筑波病）。

在病因学研究中，尤其是传染病病因学研究中，多年来，流行病学三角模式得到了广泛的认同。该模式（图 13-2）认为疾病的发生必须有病原物、机体（宿主）及环境三个要素的协同作用，任何一个因素的改变都会增加或降低疾病发生的频率。在工业革命以前疾病谱呈现以传染病为主的特征，医学研究关注的焦点集中在病原微生物，因此，三角模式特别强调病原体的作用，将其从环境因素中分离出来，单列为流行病学三角中的一个重要部分。但是，随着社会的进步和科学的发展，即使对于传染性疾病，流行病学家也更倾向于将病原微生物纳入到环境因素中，从而可以更全面地反映各因素之间的相互作用。更何况疾病谱已发生了很大变化，在以慢性非传染性疾病为主的今天，疾病的发生可能与社会经济、精神心理等多种因素有关，这些因素间的相互关系远非三角模式所能涵盖。

(2) 轮状模式 (wheel model): Susser 提出疾病病因的轮状模式用以表示机体（宿主）与环境的关系，相对于流行病学三角模式，疾病病因的轮状模式更强调环境及其与机体间的密切关系。轮状模型（图 13-3）以宿主为轮轴，模型由外环和内环两部分组成，外环指的是环境，包括生物、理化和社会环境；内环指的是机体，包括人的自然特征（如年龄、性别）、营养状况、免疫力、内分泌水平和遗传等，其核心是遗传基因。外环生物环境包括病毒、细菌等各种病原微生物以及传播媒介和作为食物、制药原料的动植物；理化环境指的是气候、水、大气、土壤、光、辐射和各种化学物质如农药、杀虫剂和职业污染物等；社会环境包括社会经济水平、文化水平、政治制度、职业、居住条件、精神因素、个人行为方式等。轮状模式各部分的相对大小随不同疾病而有所变化。

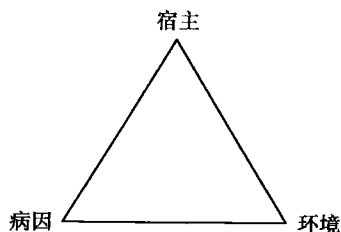


图 13-2 流行病学三角模式

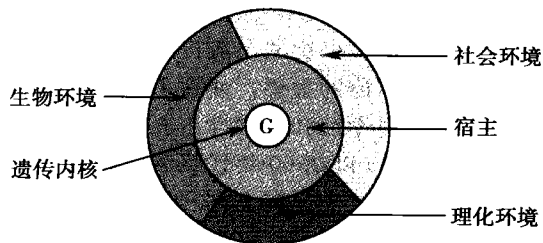


图 13-3 病因的轮状模式

2. 疾病因素模型 该模型（图 13-4）在病因分类上具有较强的可操作性和较强的实践指导意义。该模型将病因因素分为两个层次：外围的远因和致病机制的近因。外围的远因包括社会经济、生物学、环境、心理、行为和卫生保健因素。基础或临床医学的病因主



要是指致病机制的近因，临床流行病学病因学研究系以临床为基础，具有近因与远因相结合的特色。

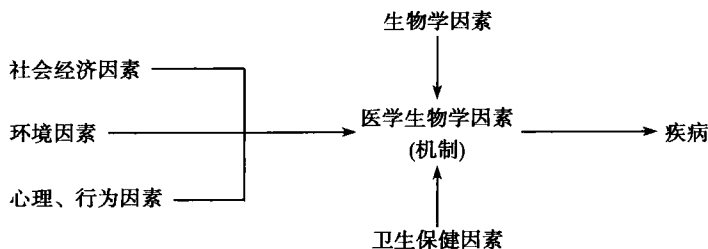


图 13-4 疾病因素模型

3. 病因链与病因网络模型 1970 年，MacMahon 等提出了病因作用的网络模型，即疾病的病因因素按时间顺序连接起来构成病因链（chain of causation），多个病因链交错连接构成病因网（web of causation）。MacMahon 等指出，任何结果都不是由单一的孤立原因所致，而是各种因素互相交错，各种因素又各有其前因后果，其复杂程度远超过我们的想象限度。例如，肝癌的病因网可看成是由乙型肝炎病毒感染、饮用沟塘水、食用黄曲霉毒素污染的食物、遗传倾向和过量摄入乙醇等多条病因链交错构成，其中每条病因链又由多个环节连接构成，如饮用沟塘水可能是由于水中的藻类毒素引起了肝细胞恶变。

病因网络模型的优点是表示直观、具体，为病因阐述提供了依据，具有较强的可操作性，但其分析的技术难度较大。

### （三）病因与危险因素

随着病因学研究的不断深入，多病因学说已被医学界所接受，已发现越来越多的疾病是多因性的，而且存在一病多因、一因多病和多因多病等复杂情况。按照病因与疾病间的作用方式、作用程度及传统哲学的观点，人们给病因以多种分类方法，本节仅介绍一下三种类型。

1. 必要病因与充分病因 按照传统的哲学观点，凡效应都有“必要条件”和“充分条件”之分，借助这种抽象的逻辑思维方式，我们可以认为任何疾病的发生都有相应的必要病因和充分病因。

（1）必要病因（necessary cause）：指有相应疾病发生以前，必定（概率为 100%）有该因素存在。如果缺乏该因素即不会引起某疾病，则称该因素为该病的必要病因。如没有结核杆菌感染就不会发生结核，因此，结核杆菌是结核病的必要病因。绝大多数传染性疾病、职业病等都有一个比较明确的必要病因，而大多数慢性非传染性疾病目前尚未发现他们的必要病因。由于大多数慢性非传染性疾病是多因性的，因此，这类疾病可能不存在一个必要病因，或是必要病因仍待探究。

（2）充分病因（sufficient cause）：是指若有该病因存在，必定（概率是 100%）导致某疾病发生。显然，充分病因即使针对传染性疾病也是非常少见的，因此，对充分病因的理解应对以下三点有清醒的认识：①对大多数疾病而言，充分病因的组成因素不是一个，而是一组。如上述的结核杆菌感染仅是结核病的一个必要病因，而不是结核病的一个充分病因。因为大多数的结核杆菌感染者可由于自身抵抗力的作用而不发生结核病，只有结核杆菌感染结合机体特异性和非特异性抵抗力的降低，才能构成结核病的一个充分病因。②对于大多数疾病而言，其充分病因目前并未明了，一般只证实或初步证实充分病因中的个别或几个因素。③对于大多数慢性非传染性疾病来说，目前认为其充分病因不止一个，有的可能多个充分病因，各充分病因的组成因素可能不同，因而这些疾病就可能没有必





要病因。如肥胖（超重）是高血压的一个“病因”，但有的高血压病人并不超重，提示导致这部分病人发病的充分病因中可能不包括肥胖。

在日常生活中人们发现开启开关则电灯发光，于是便认为电灯开关是“因”，电灯发光是“果”，只要启动因，则必然获得果（灯亮）。这时此因对其果来说是必要而且充分的原因。但在电灯开关与电灯发亮的因果关系中，实际上有些重要因素被省略了。例如电线、灯泡、灯头、电流等。这些环节的任何一個都与开关同样重要，任何一个环节的缺少都会影响结果的产生。因此可以认为，任何结果的原因必然是由一组作用相关的因素共同组成的，这就是充分病因。所以“充分病因”可以定义为：一组必然导致疾病的最低限度的状态或事件。这里的“最低限度”是指状态或事件的任何部分均是不可少的。这些组成充分病因的必不可少的部分称之为成分病因（component causes）。

2. 直接病因与间接病因 引起疾病的诸多因素有时可以连续按顺次起作用，即病因1导致病因2，最终导致疾病。可表示为：病因1→病因2→疾病。这里，病因2称为直接病因，病因1称为间接病因。直接病因是指只有该病因作用于人体才能够引起发病。例如：乙型肝炎病毒是乙型肝炎的直接病因；结核杆菌是结核病的直接病因。间接病因实际上反映了引起疾病的阶段性或中间性过程，指可以促成和加速疾病发生的某些因素，其存在与疾病的发生呈间接关联。例如：营养不良、居住条件差、机体免疫力低下、社会经济环境的恶化等都可能造成患病的易感性增加，这些因素被称之为间接病因。

3. 危险因素 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。如前所述，一般将流行病学层次的病因称为危险因素，它是指疾病的发生与该因素有一定的因果关系，但是尚无可靠的证据能够证明该因素的致病效应，但是当消除该因素时，疾病发生的概率也随之下降。在病因学研究中，将这类与疾病发生有关的因素即称为危险因素。危险因素的概念无疑体现了概率论因果观。

## 二、寻找病因的指南清单

上述病因模型及病因分类为我们指出了寻找疾病病因的大致方向或类别，在病因学研究实践中，我们可以从表13-1的寻找病因指南清单中得到启示。

表13-1 寻找病因指南清单

一、宿主因素	1. 先天的	基因、染色体、性别
	2. 后天的	年龄、发育、营养状况、体格、行为类型、获得性免疫、既往病史
二、环境因素	1. 生物的	病原体、感染动物、媒介昆虫、食入或接触的动植物
	2. 化学的	营养素、天然有毒动植物、化学药品、微量元素、重金属
	3. 物理的	电离辐射、噪声、振动、气象、地理（位置、地形、地质）
	4. 社会的	社会/人口（人口密度、居室、流动、都市化、交通、战争、灾害）、经济（收入、财产、景气）、家庭（构成、婚姻状况、家庭功能）、生活方式、饮食习惯、嗜好兴趣（烟、酒、茶、运动、消遣）、教育文化、医疗保健、职业（种类、场所、条件、福利、劳保设施）、政治、宗教、风俗

## 三、疾病自然史对病因学研究的意义

疾病自然史是指在不给予任何治疗和干预措施的情况下，疾病从发生、发展到结局的整个过程，疾病的自然史可以分为生物学反应期、亚临床期（或称临床前期，即症状出现



前阶段)、临床期和疾病相关结局(如痊愈、残疾和死亡)四个时期。不同疾病其自然史差异很大,各期经历的时间长短,以及各个阶段的表现也有很大的差异。了解疾病的自然史对病因学研究有重要意义,主要表现在以下两个方面:

1. 排除临床早期病例作为病因学的研究对象 为保证病因学研究结果的真实可靠,应避免由于将处于生物学反应期和亚临床期病例纳入研究对象(导致错分为“非病人群”)而造成选择性偏倚。这就要求在选择研究对象时,需要采用高敏感度的诊断试验,以发现处于生物学反应期和亚临床期病例,避免纳入可能的早期患者而影响研究质量。

2. 确定有效的随访期防止产生假阴性结果 对于若干慢性病发病后的自然病史,往往有一个较长的时间过程,其中的生物学反应期和临床前期经历的时间也许较长。因此,在病因学的研究中,从其暴露于可能致病因素之后,到临床表现,需要一个时间上的效应过程。故应设计合适的致病效应观测期。如果观测期短了,可能有些病例处于生物学反应期或临床前期而未能被确诊,于是就出现了“假阴性”结果而被漏诊,这就会影响研究结果的真实性和可靠性。从疾病自然史角度出发,确定合理的随访期,并采用特异度和敏感度高的诊断试验方法,以助早期或不典型病例的确诊。这些在病因学研究中应有充分的认识。

### 第三节 疾病病因与危险因素研究的基本过程与方法

#### 一、提出病因假设

##### (一) 假设的作用

在病因探索研究中,形成病因假设,是关键的一步。假设是科学研究中一种广泛应用的方法,它是根据已知的科学原理和科学事实,对未知的自然现象及其规律性所给出的假定性说明或推测性解释。

在临床医疗实践或人群流行病学研究中,当发生了某种疾病,根据其临床和流行病学史、临床体征、相关的实验室或特殊的检查资料,应用现代医学的知识和方法进行验证和推导,尚不明病因的时候,就会对正确的诊断和有效防治带来很大的困难,这就迫使医学工作者必须对这种疾病的发病原因进行研究。例如:我国2003年春季在广州始发的“严重急性呼吸窘迫综合征, SARS”。当初暴发流行,病情严重,病死率高,病因不明。对人群健康的威胁与危害十分严重,迫切需要弄清 SARS 的病因!

对于不明病因疾病进行病因学研究,首先必须依据该疾病在人群中的分布特点、临床表现、病理损害的定位及其损害的程度,现有水平的各种化验、检查结果,作出一系列的排除诊断,然后检索相关文献,在进行系统综合分析的基础上,作出可能的“假设”诊断,这就是要提出的病因假设。

例如: SARS 的暴发流行是通过呼吸道传染的。病因方面排除了细菌性感染、常见呼吸道病毒以及流感病毒感染。有人研究认为衣原体致病,其后被否定,于是提出了可能为一种毒力很强的未知病毒所致病因假设,通过国际协作研究,终于证实了 SARS 的病原体(病因)系变异型冠状病毒。

又如,我国华北地区大面积调查食管癌的流行特征后,通过资料分析与概括,发现该地食管癌死亡率较高的县、市,均集中在太行山脉南段三省(河北、河南、山西)交界处,并由此处向四周逐渐减低,与地势高低存在明显相关关系;而且发病率大体呈不规则的同心圆分布,最高和最低相差约 97 倍。研究者考虑到地理环境的可能具有重要作用,有人提出食管癌的水土病因学说。随后就有人根据这个假设,开展了地质化学方面的研究。



## (二) 形成病因假设的逻辑方法

前已提及事实依据和理论基础是建立假设的两个支撑点，但是，假设的形成不等于就事论事或事实与已有理论的混合。由事实和已有理论形成假设，必须经过一个较严密的逻辑思维过程。换句话说，假设形成的常用方法是逻辑方法。而假设形成过程中常用的逻辑方法主要是归纳演绎法。这种方法对于病因研究的因果假设具有重要指导价值。

1. 归纳法 (Mill 准则) Mill 是 19 世纪的哲学家，1856 年在他所著的《逻辑系统》一书中建立了数条准则，其中科学实验四法常被用于分析流行病学研究中形成假设、设计研究方案 and 进行病因推断。后人在科学实验四法的基础上将同异并用法单列，也即科学实验五法。

(1) 求同法 (method of agreement): 又称一致法或契合法或“异中求同法”，指对不同的事件或事物找出它们的共同点——共性。如在肝癌的病例对照研究中，肝癌病例组发现都有或相当部分有乙肝病毒感染标记；队列研究中，有乙型肝炎病毒持续感染者其肝癌发病率较高。提示乙肝病毒感染可能是肝癌的危险因素之一。食物中毒事件中，中毒患者大多有进食相同食物史，则可以提示由该食物引起中毒。

(2) 求异法 (method of difference): 又称差异法或“同中求异法”，指在相似的事件或事物之间找不同点 (重要的差别)。还是以肝癌研究为例，在病例对照研究中，对照组多数不饮用沟塘水；在队列研究中，不饮用沟塘水的非暴露组的肝癌发病率低于饮用沟塘水的暴露组的发病率。提示两组与肝癌发生有关的差异之一是有无暴露于沟塘水。再如在非肝癌病例中发现都没有或相当部分没有乙肝病毒感染标记，表明乙肝病毒感染是肝癌的危险因素之一。

(3) 同异并用法 (joint method of agreement and difference): 求同法和求异法并用，相当于同一研究中设有比较组，控制干扰因素。

如宫颈癌的病因问题，据报道：性生活越是混乱的妇女发病率越高，早婚妇女的发病率又高于晚婚者，这是求同。与此相反，修女、尼姑与独生主义妇女很少患宫颈癌，这是求异。因此，有人提出性生活中的某因素可能与宫颈癌的发病有联系。随后的研究表明：宫颈癌可能与性交时的 II 型疱疹病毒感染有关。同异并用法是比较性研究 (有对照组) 设计的逻辑学基础。

(4) 共变法 (method of concomitant variation): 可以看成是求同法的特例。指研究因素的暴露程度不同时，疾病的频率也发生相应的变动，即在研究中注意发现疾病的患病率 (有时是发病率) 波动时有哪些因素在变动。共变法的应用有一定的条件，只有当有关因素 (暴露) 不是定性的，而是等级或定量的，并与事件 (疾病) 效应成量变关系时，才可以应用共变法。

如在吸烟与肺癌的研究中，随着吸烟剂量 (等级) 的增加，肺癌的优势比 (OR) 或相对危险度 (RR) 也增加，即呈共变或剂量反应关系，故支持吸烟是肺癌的病因的假设。

(5) 剩余法 (method of residues): 剩余法可以看成是求异法的特例，指当人们已知某复合结局事件 (A、B、C) 的有关暴露因素在特定的范围 (a、b、c)，通过事先的归纳又知道 b 说明 B，c 说明 C，那么剩余的 a 必定说明 A。剩余法就像算术中的减法，即在一组复杂的现象中把已知联系的现象去掉，探寻其他 (剩余) 现象的联系。如在肝癌的病因研究中，肝癌的发病率除了乙肝病毒感染和黄曲霉毒素能解释的部分，还有未能解释的部分，这部分或可归因于暴露因素范围内“剩余”的因素，如饮水中的藻类毒素。

需要注意的是：如果病因假设清单中没有包括真实的病因，Mill 准则就不能提供任何帮助。遗憾的是，Mill 准则对列出病因假设清单并不能提供指导，我们也无法知道要寻找的那个因素是否在清单中。



2. 假设演绎法 演绎是从一般到个别的推理。它是根据已知的规律来推论未知事物的方法,故又称类推法。假设形成后,通过假设演绎法同检验假设的分析性研究相衔接。

例如,我国原发性肝癌高发区主要分布在温暖、潮湿的东南沿海地区,在这些地区进行的大量描述性研究所获结果提示,乙型肝炎病毒感染、饮用沟塘水、食用被黄曲霉毒素污染的食物等因素的分布与原发性肝癌的分布相一致,从而为其后的分析和实验流行病学研究提供了线索,并形成了相应的工作假设。这一形成假设的过程衔接了描述性研究和分析性研究,其原理本质上是假设演绎法(hypothesis-deduction method)。其整个推论过程为:从假设演绎推出具体的证据,然后用观察或实验验证这个证据。如果证据成立,则支持假设的成立。从逻辑上看,反推是归纳的。从一个假设可推出多个具体证据,检验证实的具体证据越多,或证据的条件越多种多样,则归纳支持这个假设的概率越大。如果由假设演绎出来的具体证据不成立,并不能简单地否定假设,还需要考虑其他影响因素的存在。

以上的所谓归纳法以及假设演绎法,在病因学的假设中,可联系具体实际参考应用。

## 二、验证病因假设的要素

如上所述,科学的假设是在一定的经验材料和科学理论上经过逻辑思维加工提出来的,因此,同一个论题,由于事实依据不同,对科学理论的理解不同,思维方式不同等,常常会出现不同的假设。不仅不同的科研工作者对某一论题会提出不同的假设,甚至同一学者对同一论题,在不同时期提出的假设也会有差别甚至是很大的差别,这在医学对未知病因的研究,如像肿瘤、心脑血管病等病因多元性就是如此,由此来看,在科学上出现不同的假设,促进对一个问题的全面认识是很重要的一种正常的现象,也是科学发展兴旺发达的重要标志。

当然,科学上各种不同的假设有时受着主客观认识水平的限制,总有正确和错误之分,也有完善与不完善之别。但是假设的提出应持科学求实、严谨、创新的态度,力争使自己对所研究问题假设,建立在较为可靠的科学理性基础之上,只有这样才有利于科学的发展。

1. 检验证实需要进一步深化 假设是否正确,需要通过实践(调查、观察和实验)来检验。检验的结果无非是证实或证伪。如果要证实了自己研究的假设(如未知病因),那就应该全面进行研究,以探索在哪些条件下是符合的,在哪些条件下是不符合的,找出它的适用范围和局限性,并且深入地研究它的本质性内在联系,找出它的规律与机制,争取由假设上升到理性认识并指导自己的工作实践。例如:SARS从临床流行病学研究深化到病原学的确定等等,这些研究证实了“假设”,其成果又有效地指导了SARS防治实践。

2. 检验证伪应予区别对待 如果检验结果与假设不符,甚至完全相反,这表明假设是不够正确或是错误的。对于科学的发展来说,证实和证伪都具有重要意义。因为没有证实,不能肯定正确的假设;没有证伪,就不能否定错误的假设。因此,对于任何假设的验证结果,应持科学态度,进行具体分析,区别对待。一般而言,凡实验结果或观察到的现象与假设截然相反,或面对检验结果即使补充假设也无法自圆其说时,则应当放弃原有假设。若虽屡遭失败,但检验结果并不能否定假设的核心,或虽难以证实,但无直接否定假设的证据时,则不应随便放弃原假设,仍应从不同的角度或侧面,对其进行检验。

3. 检验假设的注意事项 检验假设最忌的是主观偏性。对于任何研究的假设无非有



三种情况，一为通过验证，被确认为真理；一为受有关偏倚、混杂因素的影响得出虚假的因果联系；第三则为非科学的行为导致错误的因果联系。在病因学研究检验假设时，应避免或防止第二、三种情况的出现，这是要十分注意和坚持的，也是每位医学工作者务必养成尊重事实的思想作风与实事求是的工作态度。英国生理学家赫胥黎（Thomas Huxley）曾经说过：“我要做的是让我的愿望符合事实，而不是试图让事实与我的愿望调和。你们要像一个小学生那样坐在事实面前，准备放弃一切先入之见，恭恭敬敬的按照大自然指的路走，否则，就将一无所得”。只有与事实相符的假设，才有可能发展成为理论，进而促进科学的发展。

三、验证病因/危险因素假设的主要研究方法

对于病因/危险因素尚不清楚的疾病，如上所述，从临床或流行病学的角度，总归会提出病因的假设（假说），也许这种假设可能占有一些有说服力的证据，但都不能肯定或否定真正的病因，因此，必须通过科学的观察性或实验性的研究方可获得病因学的真实结论。

例如：原发性肝癌患者，从临床病例分析，多有乙型肝炎病毒感染—慢性肝炎的病史及其相应的临床特征和化验的证据。因而，临床医生从总结分析临床资料的角度，提出了肝癌的病因可能与乙型肝炎病毒感染、慢性肝炎病变有关。这个假说不经过科学的验证能肯定或否定吗？

又如：SARS 暴发流行，临床发现一些病例病情危急，肺部炎变严重伴急性呼吸功能衰竭，病死率高，显属急性呼吸性感染致肺部重症炎变。经临床及实验室检验，排除了细菌性、流感病毒及常见呼吸道病毒的病原学诊断，临床提出了多为一种毒力很强的某一特殊病毒感染的假设（诊断），同样如不经过病原（因）学的研究能肯定或否定这一假设吗？

因此，病因学研究往往需要多学科和多专业的协作和参与，才能获得病因学研究（病因假设的验证）的真正成功。

然而，从病因学研究的角度，临床流行病学兼纳了临床医学和流行病学的两大学科优势，经过科学评价，列出以下常用研究方案及其证据的论证强度（科学性）（表 13-2），供研究者应用时的参考。

表 13-2 不同病因研究设计类型的论证强度

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

1. 病例对照研究 病例对照研究是回顾性分析流行病学中最基本的研究类型之一。它是以确诊的患有某病的病人作为病例组，选择具有可比性的不患有该病的人群作为对照组。为排除混杂因素干扰，可采用配比设计模式，比如用年龄与性别作为配比因素来排除年龄与性别可能的混杂作用，选择不患研究靶病的患者，以 1：1 或 1：2 比例选择符合标准的对象作为对照组（参见本教材设计方案章节），然后进行病例组与对照组的比较分析，探讨因果效应。

例如：HBsAg 与肝癌相关性的非配比设计的病例对照研究，结果见表 13-3。



表 13-3 HBsAg 与肝癌关系研究

	病例组 (肝癌)	对照组 (非肝癌)	合计
HBsAg 阳性	35 (a)	8 (b)	43
HBsAg 阴性	15 (c)	42 (d)	57
合计	50	50	100

$OR=ad/bc=(35 \times 42)/(15 \times 8)=12.25$  (95%CI, 4.70~31.82), 结果显示 HBsAg 与肝癌发生有显著意义。

由于病例对照研究是从“果”到“因”的研究过程, 且存在未知的混杂因素, 论证效率相对较低。

2. 队列研究 队列研究是重要的、可行性良好的病因/危险因素的研究方法, 可以直接观察暴露于危险因素的不同人群的结局, 具有较强论证力度。其是在特定人群中按照目前或过去是否暴露于待研究的危险因素, 分为暴露与非暴露组, 随访观察一段时间后, 进而比较待研究疾病的发病率。由于为前瞻性研究, 因此能确定暴露与疾病的因果关系。

例如: 有一前瞻性队列研究, 观测高胆固醇血症致冠心病的危险性, 两个队列各为 1000 例 35 岁以上的观测对象, 追踪 5 年, 追踪率 90% 以上, 结果高胆固醇血症队列发现冠心病 83 例; 而正常血清胆固醇队列发现冠心病 23 例。结果见表 13-4。

表 13-4 高胆固醇血症致冠心病因果效应表

	冠心病		合计	发病率 (%)
	+	-		
高胆固醇队列	83 (a)	867 (b)	950 (a+b)	8.74
正常胆固醇队列	23 (c)	917 (d)	940 (c+d)	2.45
合计	106 (a+c)	1784 (b+d)	1890 (N)	

相对危险度 (RR) = 暴露组发病率/对照组发病率  
 $=8.74/2.45=3.57$  (95%CI, 2.27~5.62)

归因危险度 (AR) = 暴露组发病率 - 对照组发病率  
 $=8.74\%-2.45\%=6.29\%$  (95%CI, 4.22%~8.34%)

病因学分数 (EF) =  $\frac{RR-1}{RR}=71.99\%$  (95%CI, 55.95%~82.20%)

$NNT=\frac{1}{AR}=16$  (表示高胆固醇血症 5 年中每 16 人就发生 1 例冠心病患者)。结论: 高胆固醇血症系冠心病发病的危险因素, 其  $RR=3.57$  (95%CI, 2.27~5.62)。

3. 实验流行病学研究 实验流行病学研究是一种可以施加干预因素, 并且可控制研究条件的前瞻性的研究方法, 比较给予不同干预因素后实验组与对照组的结局, 借以判断干预因素对结局的影响, 以佐证病因学的结论。这种研究设计方案, 视具体情况, 可采用随机对照试验 (RCT), 也可采用非随机化的准实验研究设计模式。例如: 我国大部分地区对新生儿和适龄儿童, 采取了接种乙型病毒性肝炎疫苗以预防乙型病毒性肝炎的措施, 通过 RCT 或准实验研究, 追踪乙型病毒性肝炎的发病率, 进而观测若干年, 如肝癌发病率呈相应的下降, 则可反证 HBsAg 与肝癌的病因的关系。实验流行病学的因果论证的强度较强。

4. 实验病因学研究 实验病因学从微观的角度去检验病因, 借助生化实验、微生物学实验、动物实验等基础医学的研究, 阐述病因的作用机制, 为验证病因假设提供依据。



基础医学的微观研究对于认识疾病的本质以及对于疾病的进一步有效诊治具有重要价值。发病机制的研究成功能进一步肯定病因假设，加深对病因和致病过程的理解。但是同时应该注意到由于基础医学中模拟的人体内环境或者是动物实验均不可能代替人体本身，因此，实验结果外推到人这个过程需要谨慎。

#### 四、病因学因果关联的统计学指标

用于病因学因果相关性分析的指标常用者如下：

发病率（incidence）：即暴露有关可疑病因或危险因素后，发病人数占其总体人数的百分比等。

绝对危险度（absolute risk, AR 或叫归因危险度 attribute risk, AR）：即暴露可疑病因中人群和非暴露可疑病因组的人群，他们各自发病率的相减之绝对的差值，它表示危险特异地归因于暴露因素的程度。如：暴露组发病率 25%、非暴露组为 8%，则其绝对危险度为  $25\% - 8\% = 17\%$ 。

相对危险度（relative risk, RR）：即可疑病因接触组的发病率与对照组的发病率之比，常用来表示暴露与疾病联系的强度及其在病因学上的意义大小。

相对危险增高度（relative risk increase, RRI）或称病因分值（etiologic fraction, EF）是指暴露人群中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，即绝对危险度与因素暴露组发病率的比值。

需要多少例接触致病因素后才出现一例发病（number needed to treat, NNT）：用  $1/AR$  值表示，NNT 是临床和卫生决策十分有用且十分容易理解的指标。

上述这些评价因果之间关系的指标都可以计算各自的精确度（95%CI），至于因果相关性的强度和价值，则应结合具体的疾病情况和专业知识来确定。

#### 五、防止偏倚干扰，慎重分析病因效应的关联

##### （一）虚假关联

在研究过程中，由于偏倚的干扰，会导致研究结果不真实，可以表现为扩大或掩盖了暴露因素与疾病的关联强度，甚至出现完全虚假的关联。例如，一项应用利血平治疗高血压病研究。结果发现利血平治疗有致乳癌的高度危险关联的结论。于是人们就提出了利血平致乳癌的“假设”，经进一步地去验证这一假设，发现在对象选择方法上，对照组排除了心血管疾病的患者，而病例组并未排除，产生了偏倚的结果，导致利血平与乳腺癌的虚假联系。在之后的一项研究中避免了选择对象的不一致性，结果证明利血平与乳腺癌并无因果关联。从而纠正了这一“伪证”，“平反”了以上错误的结论。

##### （二）间接关联

当排除虚假关联后，不一定说明暴露因素与疾病肯定存在因果关联。当两类毫不相关的事件，如两种疾病都与某因素有关联时，这两种疾病会呈现明显的统计学关联，这种关联称为间接关联。为了避免与间接因果关联混淆，现在称其为继发关联（secondary association）。继发关联是由混杂偏倚引起的关联，即可疑的病因（暴露 A）与疾病 B 并不存在因果关联，但由于外部因素 C 的存在，而 A 和 B 与 C 均存在着关联，因此观察到 A 与 B 存在着关联。这也是在研究中应注意要控制的。比如：有调查发现，有伤寒病史者的痢疾发生率明显比无伤寒病史者高，提示伤寒病史可能与以后发生痢疾相关。这种相关经假设检验有统计学意义，可排除抽样误差的可能；同时经仔细分析研究设计和所获资料，认为偏倚得到控制，可以排除虚假关联。但以现代医学理论解释，伤寒病史与之后的痢疾发生毫无相关，但两者均受到卫生状况及个人卫生习惯的影响。因此，认为两者的关联属于



继发关联。再例如，高血清胆固醇是冠心病的危险因素，高血清胆固醇又可产生沉积于眼睑的黄色瘤，从而导致黄色瘤与冠心病的继发关联。这是一种纯粹由混杂偏倚产生的关联，即怀疑的病因黄色瘤与冠心病并不存在因果关系，而是由于两者均有共同的原因高血清胆固醇，黄色瘤和冠心病都与高血清胆固醇存在关联，从而导致黄色瘤与冠心病的继发关联。

当暴露因素与疾病 D 既存在直接关联，又存在继发关联时，暴露与疾病的直接因果关联的程度或方向将可能受到混杂干扰，即得到歪曲的关联估计值。例如：静脉吸毒（共用注射器）与性乱都是人类免疫缺陷病毒（HIV）感染的危险因素，吸毒者倾向于发生性乱行为，即吸毒同 HIV 感染既存在直接关联，又存在继发关联。在这种情况下，需要控制性乱的影响，避免性乱对吸毒和 HIV 感染的直接因果关联起混杂或歪曲作用。

当排除抽样误差、虚假关联和继发关联后，两事件间的关联才可能是因果关联，才能进行暴露因素与疾病的病因推导。（参见本教材第十章有关内容）

### （三）因果关联（causal association）

统计学关联是判断因果关联的前提，但只有少数统计学关联属于因果关联。因果关联可以有直接关联和间接关联，随着对因果关系研究的深入，直接关联和间接关联可以相互转化，原来认为是直接病因的可以被后来的研究证明是间接病因。我们可以用因果关系判断标准（病因学研究评价原则）来推断所研究的因素是否是疾病的病因。图 13-5 概括了进行病因推断前必须考虑的问题和步骤。

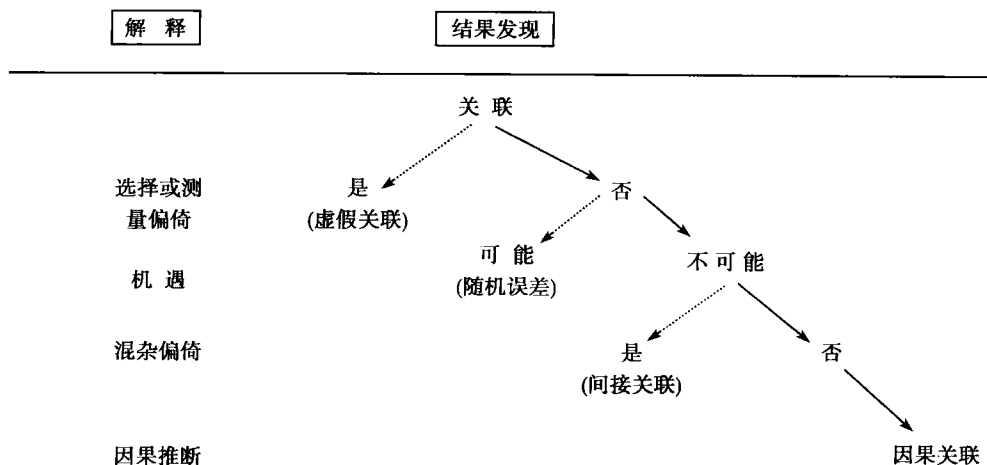


图 13-5 因果关联的判断进程

关于疾病的病因研究，一般言之，不能依靠临床观察的方法来研究病因。临床观察可以积累丰富的经验，它对病因问题可以提供某些疑点和线索，或经验性证据。例如，1959年至1960年代初期关于海豹肢体短畸与孕妇服用反应停（Thalidomide）的关系的研究，这种出生缺陷先在西德，后在许多欧洲国家流行，甚至在社会上引起了恐慌。尽管是由临床医生首先提出可疑之点——即孕妇服过缓解妊娠反应的“反应停”，并怀疑可能同新生儿短肢畸形有关，进而引起了公众关注。但这个重要的病因，却是通过有关国家疾病统计数字的累积，特别是经过分析性流行病学研究，才最后确定下来的。

人们往往有这样一种错误的看法，一提病因研究，就认为只是基础医学（即实验医学）的研究。其实不然，尽管实验医学的研究者可以利用现代科学技术，进行在有控制的条件下的某些动物实验、组织细胞实验以及分子水平的研究，以探索某因素与疾病的因果





关系。但这些毕竟只能解释关于人类疾病过程中整个生态系统的某个个别阶段的变化和规律。人类是在复杂的社会和自然环境中生活的，这与实验室所控制的条件、环境完全不同。况且人与动物还有种属差异。因此，实验医学是重要的，但其研究结果往往不可轻易地引申至人。再者，由于医德的原因，不能在人群中随意进行实验研究，因此，关于疾病病因问题的研究，很大程度上取决于以人为研究对象的流行病学研究或临床流行病学研究（大多数是观察性研究）。

## 第四节 疾病病因与危险因素研究的评价原则

疾病病因与危险因素研究结果是否能够确定病因，其研究的水平和价值多大，应对其进行严格的评价。表 13-5 所列是国际临床流行病学有关病因学研究的评价原则，其可以作为因果关系推断的参考标准。

表 13-5 病因学研究的评价原则

### 一、研究结果的真实性评价

1. 是否采用了论证强度高的研究设计方案？
2. 除研究的暴露因素外，试验组与对照组其他方面是否一致？
  - 包括 RCT、队列研究、病例对照研究
  - 其他已知的预后因素是否一致或经过了调整？
3. 试验组和对照组有关因果效应的测量方法是否相同？
  - 有回忆性偏倚、调查偏倚吗？
  - 是否采用了盲法？
4. 随访时间是否足够长？研究结果包含了所有随访病例吗？
  - 随访不完全的原因
  - 失访病例与未失访病例的危险因素是否相似？
5. 是否有因果效应的时间先后顺序？
  - 暴露先于结果
6. 有剂量反应梯度关系吗？
  - 随着暴露剂量、持续时间增加，结果的危险性增加
7. 病因学研究的结果是否符合流行病学的规律？
8. 病因致病的因果关系是否在不同的研究中反映出一致性？
9. 病因致病效应和不良反应发生的生物学依据是否充分？

### 二、研究结果的重要性评价

1. 暴露与结果联系的强度如何？
  - $RR$ 、 $OR$  或  $EF$  等
2. 危险估计的精确性如何？
  - 可信区间 (95% $CI$ )
3. 研究的样本量合适吗？

## 一、病因学研究结果的真实性评价

### （一）是否采用了论证强度高的研究设计方案？

不同的临床病因学研究方法，其因果论证强度是不同的，描述性研究论证强度最弱，病例对照研究论证强度不太强，队列研究论证强度较强，随机对照的临床试验最强，因为



它来自以人为研究对象的真正的人体实验。表 13-2 可作为评价不同的设计方案所获得的因果关系论证强度参考。

对于病因研究的证据是否来源于真正的人体实验,是指在研究中是否将人体置于暴露环境中,通过比较不同组别的结局来判断暴露因素的致病效应。临床上多使用随机对照试验,具有较强的因果论证强度。但如果在健康人体中进行施加危险因素的病因学实验,显然是不道德的,并且不具有可行性。但可以变通的采用去除可能的致病危险因素的方法来进行干预性研究设计。这种所谓的终止效应可以来自实验流行病学研究、自然实验或自发性改变的观察性资料。是否存在终止效应也是一个很关键的标准,也即如果对某些可能的致病危险因素或病因采取针对性的措施之后,能使被研究的疾病发生或流行随之减少,那么对病因或危险因素的肯定也具有非常重要的意义。

### (二) 除研究的暴露因素外, 试验组与对照组其他暴露是否一致? 是否存在混杂因素影响?

混杂因素可能是已知的,也可能存在未知者,其对病因研究影响颇大,因此,在拟探讨的致病因素之外,应该注意在试验组和对照组中是否存在混杂以及混杂的程度,是否采用适当的控制或消除办法。

### (三) 组间对因果效应的测量方法是否一致? 是否采用盲法?

在病因学研究中,对两组研究对象(暴露组与非暴露组)应采用同样的观察或调查方法、对暴露(不良反应)结果应使用相同的测量手段和指标。且应采用盲法来评价暴露结果,这样才有利于防止测量性偏倚干扰,使各项因果测试的方法,指标一致有利于对比,避免上述偏倚造成的影响,增加结果的真实性。

### (四) 随访观察时间是否足够长? 研究结果是否包括了所有纳入随访研究的病例?

任何致病因子引起人体发病都有一个致病的时间效应关系,急性及自然病程短的疾病则致病效应期短,如急性传染病。对于慢性非传染性疾病则其致病效应期较长,因此,研究慢性非传染性疾病发病危险因素的致病效应时,往往需要足够时间才能观察到结果的发生,观察期过短会获得假阴性结果。另外,在随访期间失访过多会影响研究结论的真实性(一般失访率不应 $>20\%$ ),因为中途退出的研究对象可能在某些特征上与仍然留在研究中的研究对象存在差别。

### (五) 因果效应的时间顺序是否确切、合理?

致病因素引起发病,必然是因在前,果在后,时序性是构成因果关系的基础。在评价某一病因学或不良反应研究时,如果能明确危险因素的出现早于疾病或不良反应的发生,则研究结果的真实性就高。对于实验性研究和队列研究可以确定因果效应的时间顺序。而横断面的、回顾性的以及叙述性研究则不能确定因果效应的时序性。因此在确定病因致病的因果顺序时应选用灵敏度高、特异性好的测量指标,制定明确的诊断标准,证明接触病因前被研究对象未患有所研究的疾病。

### (六) 是否存在剂量效应的致病关系?

剂量效应关系是指暴露因素的剂量、程度与疾病发生的进展与程度存在显著的相关关系。常见于化学性和物理性有害因素的损害作用,其致病效应的程度与接触的剂量存在相关,接触的剂量越大,以及接触的时间长,累积剂量大,致病效应越明显,损害越大。如果能够证明随着与可疑致病因素的接触增加(剂量增大、时间延长)导致更多更重的相关疾病的发生,则就能够肯定这一效应关系。

### (七) 病因学研究结果是否符合流行病学的规律?

流行病学在致病因素、机体和环境等方面宏观地进行探讨,为病因学研究常能提供重要的发现和有力的证据。研究中的暴露的分布,如与疾病的地理分布、时间分布及人群间



分布符合或基本符合,则意义更大。

### (八) 不同的研究,结论是否一致?

对于某一疾病的病因,在不同地区、不同人群中采用不同的或者是相似的研究方法,并采用相关的评价指标进行的病因研究,所得结论如果均一致,那么该病因致病的结论就较为可靠。它是病因推断中一个重要的指标,符合 Mill 准则中的求同法则。

例如,吸烟与肺癌关系的病因学研究,世界上至少已有数十项病例对照研究和 7 次以上较大规模的队列研究的结果均表明,肺癌与吸烟有很强的关联。多次研究的可重复性使因果关联的可能性增加,而少数或个别研究的不同甚或相反的结果并不能简单否定,需要仔细探究结果差异的缘由。由于临床病因学研究是一项较为复杂的过程,在研究设计及实施过程中不同的研究者会存在差异,因此所得到的结论并非都能反应真正的因果关联,因此在病因的一致性分析中要持谨慎的态度,权衡研究设计、实施过程及其结果的科学性。

### (九) 病因学效应的生物学依据是否充分?

如果病因学研究揭示的因果关系有生物学的可解释性,则可增加因果联系的论证强度。随着当今生命科学研究的飞速发展,把临床流行病学对致病因素的宏观研究结果与基础医学的分子生物学、细胞生物学、分子病理学、组织学水平、遗传学和免疫学等微观研究结果相结合,必将促进病因学研究的飞速、深入发展,将会对疾病有更加清晰的认识。

## 二、病因学研究结果的重要性评价

通过上述真实性的分析与评价,假设一个病因学研究的结果有着良好的真实性,那么我们应该进一步地评价这个研究结果有否具有重要的临床意义和价值,否则就没有必要再评价了。在评价重要意义方面则应有量化的相应指标。

### (一) 关联的强弱程度

在病因学研究的方法中有一些反应暴露与疾病的因果关联程度的量化指标。随机对照试验以及队列研究中,常用相对危险度 (RR)、归因危险度 (AR)、RRI、NNT 等来评价因果关联强度,病例对照研究则多用比值比 (OR) 进行评价。

在判断 RR 和 OR 的意义时,还要进行敏感性分析 (sensitivity analysis),这有助于对潜在的混杂因素进行“调整”或“修正”。如果调整后的 RR 或 OR 较调整前的 RR 或 OR 明显变小,则应该怀疑原来的结果。相反,如果调整后的 RR 或 OR 与调整前相比保持不变或比调整前明显增大,则可以更确信该因果关联的真实性。

### (二) 相对危险度 (RR) 或比值比 (OR) 的 95% 可信区间

除评价因果关系的点估计强度外,还需评价其精确度,方法是计算 RR 或 OR 的 95% 可信区间 (CI),如果 95% CI 的范围较狭小、下限和上限值不包括 1.0,则其精确度就高,有统计学意义。

以上所列病因学研究和评价的依据,实际应用中可以根据研究要求综合参考,灵活运用。可以根据不同的研究设计,有所侧重地运用其中的条款。另外,上述原则也可作为阅读和分析病因学的文献时,鉴别真伪的参考尺度。

(孙业桓 洪明晃)

## 第十四章 诊断试验的研究与评价

正确的诊断是临床医师进行有效治疗的前提。要达到这一目的，不仅需要高水平的诊断试验，而且需要对诊断试验的结果进行科学的诠释。诊断试验（diagnostic test）涉及临床采用的各种诊断手段和方法，包括实验室检查、影像学诊断和临床采集的症状和体征等。

### 第一节 研究和评价诊断性试验的意义

面对具体的主诉不适或异常体征表现的病人，临床医师为了作出诊断，可能花费大量的时间，他们凭借个人的经验、阅读文献介绍的方法而采用多种诊断试验以达到诊断的目的，这样做也许难以避免片面性、盲目性。实际上，在选择和采用诊断试验时，是有规律可循的。如果病情简单，则较易作出临床正确的诊断；如果病情复杂，则要作出早期正确的临床诊断，有时就不太容易。

早期正确的临床诊断，是有效地防治疾病的关键。然而，对于某些危害人民健康重大的疾病，如肿瘤、心脑血管疾病等，其诊断方法，恰恰有许多不尽人意！这就要求我们不断地去研究更加有效的诊断性新试验，去早期筛选和确诊病人，以提高防治疾病的质量。此外，由于受到科学发展及人类知识水平的制约，对于现存的一些敏感度及特异度或许不太高的诊断性试验，如果将有关多项诊断措施，针对不同的诊断目的加以综合应用，以分别提高诊断的敏感度或特异度，对于疾病的有效筛选或确诊，亦是十分重要的任务。

了解诊断试验的特征、属性和适用范围，对于我们选择合理的诊断方法和准确诠释试验的结果是非常重要的。一个准确的诊断试验才能较好的界定病人和非病人，并将漏诊和误诊给人们带来的损失减少到最小的程度。

学习和研究诊断性试验及其评价，一定要以临床医学和疾病防治的需要为基础，掌握与应用诊断性试验研究的基本理论、知识与方法，才可能提高疾病的诊断水平和效率。

### 第二节 诊断性试验研究的基本方法

诊断性试验的结果和疾病真实存在与否之间的关系可以用表 14-1 的简单的四格表表示。试验的结果只有两个，一个是阳性结果，另一个是阴性结果。而受试者也只有两类，要么是某种靶疾病的病人，要么是某种靶疾病的非病人。因此，诊断试验最终得到四个可能的结果，其中两个结果是正确的，另两个结果是错误的。诊断试验给出的正确答案是患某种靶疾病的病人的相应诊断试验结果阳性（真阳性）和非靶病人的诊断试验结果阴性（真阴性）。反之，诊断试验给出的则是错误答案，即非病人出现诊断试验结果阳性（假阳性）和患某种靶疾病的病人出现诊断试验结果阴性（假阴性）。

从表 14-1 可以发现，要研究或评价一个诊断性试验，首先要明确受试对象是否存在相应的靶疾病，而用于确定疾病是否真实存在的诊断方法就是诊断性试验研究中极为关键的“金标准”诊断，或称标准诊断。另外，受试对象的选择和足够的样本量在反映结果的真实性方面也是必需的。



表 14-1 诊断试验结果与存在疾病之间的关系

		疾病（金标准）	
		有靶疾病	无靶疾病
诊断试验	阳性	真阳性 a	假阳性 b
	阴性	假阴性 c	真阴性 d

## 一、确定诊断试验的金标准诊断

金标准（gold standard）诊断或标准诊断是迄今公认的最为准确和可靠的诊断疾病的方法，但从安全性、经济性和复杂性等角度综合考虑，金标准不一定是最好的方法。有些情况下，金标准本身是一个准确、简单和经济的诊断方法，如长期使用抗生素的病人发生菌群失调性腹泻或鹅口疮，通过大便涂片计数杆、球菌比例，口腔拭子涂片查真菌孢子或拭子培养出现真菌生长，即可证实临床医师的怀疑和判断。然而，多数情况并非如此，临床见到的用于确定疾病是否存在的诊断方法，大多可能是更为复杂、费用较为昂贵或为侵入性的方法，这类方法有活检病理学诊断、侵入性的影像学检查、手术探查或手术发现等，也包括尸检。

一些慢性非自限性疾病，现有条件下的诊断试验难以发现疾病早期阶段的病理改变，从临床开始怀疑到疾病表现明朗可能需要很长的时间，数月甚至数年，因此这些疾病的随访结果也可以作为金标准，如慢性胰腺炎等一些疾病即是如此。此外，一些疾病的原因尚不完全清楚或尚无肯定的诊断试验方法，因此，可以采用权威医疗机构颁布的诊断标准作为金标准，如 2003 年流行的急性重症非典型肺炎。一些慢性疾病，由于病因和病理尚不完全清楚，并缺乏有效的诊断试验方法，如慢性肾病等也可以采用这类的临床诊断标准。在有些情况下，采用准确的诊断方法来确立疾病存在与否可能带来较大的危险，临床医师和病人可能更倾向于采用风险较小的诊断方法来作为金标准，如临床怀疑病毒性肝炎的病人，通常采用肝脏生化指标和肝炎病毒的生物标志物作为标准诊断，而不常规采用风险较大的肝脏穿刺活检。对这种特殊的状况，我们应该清楚，错误分类的概率可能性是存在的，因此，研究中只有清楚存在错误分类的风险和这种错误分类的低概率可以接受时，这样金标准才可以采用。

## 二、试验对象的选择和样本量的确定

纳入研究的受试对象包括两组，一组被金标准诊断为患有某种靶疾病的病例组，一组为金标准确定为“无病”的对照组。这里的“无病”概念，是指没有诊断试验研究的靶疾病，并非无病的正常人。过去报告的一些诊断试验的评估结果非常令人鼓舞，而在实际的临床诊断中，却不尽如人意。造成这种过高估计诊断试验价值的原因，是由于评估常常在典型的病人和完全无病的正常人中进行，通常诊断性试验往往在这两种极端的状况下会有良好的表现，即能很好的区分有病和无病。例如，早期人们认为癌胚抗原（CEA）对诊断大肠癌具有重要价值，但后来，人们发现在很多其他疾病和情况 CEA 也升高，如大肠炎、溃疡性结肠炎等大肠疾病，甚至包括 20% 的无疾病的吸烟者。因此，受试对象的选择应有广泛的代表性，病例组包括典型的和非典型的病例、轻型的和重型的病例、无并发症的和有并发症的病例、病程中早、中、晚期的病例等，对照组包括非靶疾病的其他可能引起阳性结果的疾病病例和容易与靶疾病混淆的疾病病例。这样选择研究对象的结果，不仅仅有



利于各型靶疾病的诊断,而且有利于鉴别诊断,因此,完全无病的正常人一般不宜纳入对照组。例如针对胰腺癌诊断研究的试验,病例组纳入的研究对象应包括早中期、中期和晚期的胰腺癌病人,也应包括有和没有远处脏器转移的病人、手术可以切除和不能切除的病人。而对照组应纳入胰腺良性肿瘤、慢性胰腺炎等易与胰腺癌诊断相混淆疾病的病人,也应包括一些其他器官肿瘤的病人,如肝癌、胃癌等,而健康志愿者不应纳入。这样纳入的研究对象的诊断试验研究结果可以较为真实的反映胰腺癌临床诊断的价值,并具有临床实用性。

每个诊断性试验的敏感度及特异度均是较为稳定的指标,对于诊断试验样本量的估计,应依假设试验的敏感度、特异度以及容许的  $\alpha$  错误和试验误差水平而定。因此,诊断试验样本的估算,可采用计算率的样本量的方法。用敏感度估计值计算患所研究靶疾病的病例组的样本量;用特异度的估计值计算未患所研究靶疾病的对照组的样本量。如预期敏感度、特异度在 0.5~0.8 时,可用下列公式。

$$n_1 = \frac{(Z_\alpha^2) S_n(1-S_n)}{\Delta^2}, n_2 = \frac{(Z_\alpha^2) S_p(1-S_p)}{\Delta^2} \quad \text{公式 14-1}$$

试验靶病组样本含量  $n_1$ , 非靶病对照组的样本含量  $n_2$ ,  $\Delta$  为容许误差,  $S_n$  为敏感度,  $S_p$  为特异度。

例:超声对急性胰腺炎诊断的敏感度为 80%, 特异度为 75%。样本量多大时才能有统计学意义的研究结果?

设  $\alpha=0.05$ ,  $Z_\alpha=1.96$  (双侧),  $S_n=0.80$ ,  $S_p=0.75$ , 假定  $\Delta=0.10$

$$n_1 = (1.96)^2 (0.80) (1-0.80) / (0.10)^2 = 62$$

$$n_2 = (1.96)^2 (0.75) (1-0.75) / (0.10)^2 = 72$$

根据计算,病例组应有 62 例,对照组应有 72 例。

如预期敏感度或特异度  $\geq 0.8$  时,因其趋向偏态分布,采用平方根反正弦转换后计算,如下。

$$n_1 = \left( \frac{57.3 Z_\alpha}{\sin^{-1} [\Delta / \sqrt{S_n (1-S_n)}]} \right)^2$$

$$n_2 = \left( \frac{57.3 Z_\alpha}{\sin^{-1} [\Delta / \sqrt{S_p (1-S_p)}]} \right)^2 \quad \text{公式 14-2}$$

### 三、诊断性试验的数据简化

诊断性试验的常见的数据有三类,定量数据、多项分类等级数据和两项分类定性数据。不论诊断性试验给出的结果是哪一类数据,具有临床实用价值的数值只有两类,其中最为重要的、并且是临床医师和病人所共同关心的数据是两项分类定性数据,即结果的阳性(有病)和阴性(无病),这也是诊断的目的。如果存在疾病,是否需要治疗和采取何种治疗方法,则取决于疾病的严重程度,这是另一类临床医师关心的数据,即多项分类等级数据。因此,评估诊断性试验的过程,需要将数据进行简化。如一组受试者的血清清蛋白的测得值为 30.5~46.7g/L,从诊断低蛋白血症的角度看(诊断参考值  $\geq 35$ g/L), $\geq 35$ g/L 的值为正常(阴性), $<35$ g/L 的值为异常(阳性),因此诊断结果的数据需要将定量的血清清蛋白的测得值简化为有诊断意义的  $\geq 35$ g/L 和  $<35$ g/L 两类值。而从治疗的角度看,血清清蛋白  $<30$ g/L 的低蛋白血症病人需要治疗,而血清清蛋白  $\geq 30$ g/L 但  $<35$ g/L 的低蛋白血症病人通常并不需要特殊的针对性治疗,因此对低蛋白血症病人的血清清蛋白值简化为多项分类等级数据则有利于治疗的决策。



## 四、列四格表

四格表是诊断性试验分析和评估的基本形式，将一新的诊断性试验的诊断结果和金标准诊断结果列入四格表就可以对其进行较为直观的分析 and 评估，四格表的排列如表 14-2。

表 14-2 诊断试验结果与金标准诊断结果的关系

		金标准		合计
		靶病	非靶病	
诊断试验	阳性	真阳性 a	假阳性 b	a+b
	阴性	假阴性 c	真阴性 d	c+d
	合计	a+c	b+d	n

金标准诊断有病的病例为  $a+c$ ，无病病例为  $b+d$ 。诊断性试验结果阳性的病例为  $a+b$ ，阴性病例为  $c+d$ 。金标准诊断的有病病例中，诊断性试验阳性者为真阳性 (a)，阴性者为假阴性 (c)。金标准诊断的无病病例中，诊断性试验阳性者为假阳性 (b)，阴性者为真阴性 (d)。如果在阅读文献中的数据不能列出四个表，则不能对其进行评价。

## 五、盲法比较诊断性试验的结果

诊断性试验和金标准诊断应同期进行，这对于急性、自限性疾病尤为重要，因为疾病过程的不同阶段其病理形态和生理的发展或变化并不相同。同时，采用盲法比较和评估，也是非常重要的，诊断性试验的研究者不应知道金标准的诊断结果或临床信息。如果研究者知道金标准的诊断或了解了受试者相关的临床信息，可能发生疑诊偏倚，这在影像学诊断性研究中尤为突出。对有病者或疑病者病例的阅片太过于“仔细”，而产生过度诊断偏倚 (over-diagnosis bias)；而对无病者病例的阅片仔细程度可能不足，以致人为提高了诊断性试验的诊断结果与金标准的诊断结果的吻合程度，由此导致发生与实际情况不符或偏差。

## 第三节 诊断性试验的评价指标及其临床应用

### 一、敏感度与特异度

在金标准诊断的全部有病病例中，诊断性试验的结果为阳性的病例所占的比例即为敏感度 (sensitivity,  $S_n$ )，也称真阳性率。其中诊断性试验的结果为阴性的病例所占全部有病病例的比例为假阴性率，代表漏诊率。真阳性率和假阴性率之间的关系为：真阳性率 =  $1 - \text{假阴性率}$ 。可见诊断性试验的敏感度越高，则假阴性越少，故漏诊率越低。

$$\text{Sensitivity } (S_n) = a / (a + c) \quad \text{公式 14-3}$$

在金标准诊断的全部无病的受试者中，诊断性试验的结果为阴性的受试者所占的比例即为特异度 (specificity,  $S_p$ )，也称真阴性率。其中，诊断性试验为阳性的无病受试者所占全部无病受试者的比例为假阳性率，代表误诊率。真阴性率和假阳性率之间的关系为：真阴性率 =  $1 - \text{假阳性率}$ 。特异性越高，则假阳性者就越少，故误诊的比例越低。

$$\text{Specificity } (S_p) = d / (b + d) \quad \text{公式 14-4}$$

诊断性试验的敏感度与特异度是我们选择诊断试验的基础。如因漏诊造成严重后果，例如高度怀疑某些危险的，但又可以治愈的诸如结核、淋病等疾病时，应选择敏感度高的



诊断试验。另外，在疾病诊断初期，敏感度高的诊断也是非常有帮助的，如开始接触病人时可能考虑很多种疾病的可能性，为了逐一排除，可以借助敏感度高的试验，如我国对急性重症非典型肺炎所采取的初步诊断措施就是选择了检测体温这样敏感度较高的诊断方法，这种情况下的试验阴性结果更有价值，因为假阴性率低。因此对临床医师而言，一个敏感度高的试验，其阴性结果有助于排除相应的疾病。

特异度高的诊断试验，因其假阳性率低，特别有助于肯定诊断。对于假阳性结果可能对人的精神、身体有巨大伤害和带来巨额经济负担时，应选择特异度高的试验，以减少误诊的可能性。如在对怀疑恶性肿瘤、而要采取手术治疗或放、化疗前，通常需要活检来明确诊断，而不是依靠特异度不高的诊断试验做出的诊断。因此，特异度高的诊断试验的阳性结果对诊断更有价值。

敏感度与特异度均很高的理想诊断试验，在实际的临床诊断中通常是很少有的。一个诊断性试验的敏感度和特异度之间的关系是彼消此长的关系，其中一个升高时，另一个必然降低。下面是一个说明这种关系的例子，利用检测餐后 2 小时血糖水平来诊断糖尿病，不同浓度血糖的敏感度和特异度的变化见表 14-3。

表 14-3 餐后 2 小时血糖水平诊断糖尿病试验的敏感度和特异度

血 糖 (mg/dl)	糖 (mmol/L)	Sn (%)	Sp (%)	1-Sn	1-Sp
70	3.88	98.6	8.8	0.014	0.912
80	4.44	97.1	25.5	0.029	0.745
90	4.99	94.3	47.6	0.057	0.524
100	5.55	88.6	69.8	0.114	0.302
110	6.10	85.7	84.1	0.143	0.159
120	6.66	71.4	92.5	0.286	0.075
130	7.21	64.3	96.9	0.357	0.031
140	7.77	57.1	99.4	0.429	0.006
150	8.33	50.0	99.6	0.500	0.004
160	8.87	47.1	99.8	0.529	0.002
170	9.43	42.9	100.0	0.571	0

由表可见，随血糖浓度的升高，敏感度下降，而特异度升高，反之亦然。因此，敏感度和特异度中提高任何一个值都是以牺牲另一个值为代价的。选择其中一组的血糖浓度作为划分正常和异常的诊断参考值，称为临界值。如选用血糖 100mg/dl 作为诊断参考值时，糖尿病病人的漏诊率仅为 11%，但有 30% 的非糖尿病受试者被误诊。如果将诊断参考值定为 130mg/dl，误诊率仅为 3%，但漏诊率则达到 36%。当诊断参考值定为 110mg/dl 时，此时域值的敏感度和特异度之和为最高，即漏诊率和误诊率之和最低。

## 二、准确度

准确度 (accuracy, Ac) 是指该项诊断性试验的全部真阳性者和真阴性者占受试对象总和的比例。准确度又称为符合率，反映了诊断试验结果与金标准试验结果符合或一致的程度。准确度越高，诊断试验的漏诊和误诊比例越低。

$$Accuracy (Ac) = (a + d) / n$$

公式 14-5





### 三、预测值与患病率

当我们就一个诊断性试验是否采用而做出决定时, 这个诊断性试验的敏感度和特异度是我们考虑的重要因素。一旦得到诊断性试验的阳性或阴性结果, 借以判断受试者是否患病, 则与诊断性试验的预测值有关, 这种根据诊断性试验的结果来判断受试者的患病概率称为预测值 (predictive value, PV)。阳性预测值是在诊断试验结果阳性的受试者中, 标准诊断证实有病的病例 (真阳性) 所占的比例。阴性预测值是诊断试验结果为阴性的受试者中, 标准诊断证实无病的受试者 (真阴性) 所占的比例。预测值也称验后概率, 换个角度看, 具体到个人, 则指某人在得到诊断性试验的阳性结果后患相应疾病的概率或得到阴性结果后无相应疾病的概率。

$$\text{Positive predictive value (+PV)} = a/(a+b) \quad \text{公式 14-6}$$

$$\text{Negative predictive value (-PV)} = d/(c+d) \quad \text{公式 14-7}$$

患病率 (prevalence, P) 指在某特定的时间点、某一人群中患某种疾病的人的比例, 也称验前概率, 意即在获得诊断性试验结果前某一人群的有病者的比例。验前概率具体到个人, 则指某人在接受诊断性试验前的患病概率。通常, 有几种信息资源可以用来估计患病率, 包括医学文献和当地的调查统计数据, 而对个人的患病概率可以通过病史和体征做出判断。应用诊断性试验研究时, 患病率指的是被标准诊断的患者占全部受试者总和的比例 (率), 计算公式:

$$\text{prevalence (P)} = (a+c)/N \quad \text{公式 14-8}$$

预测值受诊断试验本身的敏感度和特异度的影响, 一般讲, 一个诊断性试验的敏感度越高, 其阴性预测值越好, 也使临床医师更有把握排除相应的疾病。相反, 特异度越高, 其阳性预测值越好, 更有把握肯定相应的疾病诊断。如图 14-1, 假定估计一个受试者患病的验前概率为 50%, 在敏感度 (95%) 和特异度 (95%) 均较高的试验中, 阳性预测值 (验后概率) 达到 95%; 阴性预选值低到 5%。而在敏感度 (85%) 和特异度 (85%) 均相对较低的试验中, 阳性预选值仅达到 80%, 而阴性预选值也只达到 20%。

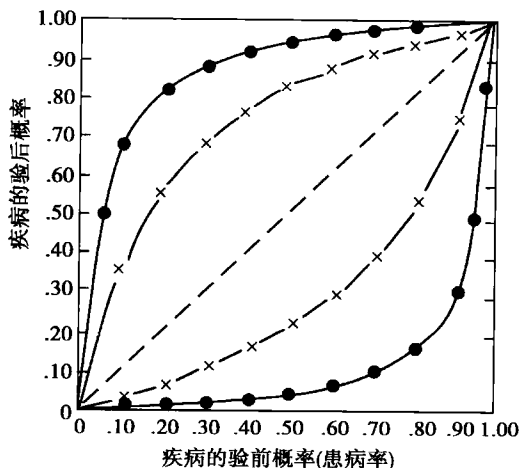


图 14-1 不同敏感度和特异度试验对验后概率的影响

除此之外, 预测值也受患病率的影响, 如果患病率很低, 即使高特异性诊断试验的阳性结果, 假阳性也会占很大比例, 同样, 高敏感性试验的阴性结果, 在低患病率下, 也会有较高比例的假阴性率。由 Bayes' 条件概率理论推导的预测值与敏感度、特异度、患病率



的关系如下。

$$+PV = Sn \times P / [Sn \times P + (1 - Sp) \times (1 - P)] \quad \text{公式 14-9}$$

$$-PV = Sp \times (1 - P) / [(1 - Sn) \times P + Sp \times (1 - P)] \quad \text{公式 14-10}$$

我们可以想象,如果我们面对一个无病的人群,不论诊断性试验的特异性如何高,所有的阳性结果都是假阳性。因此,当患病率为零时,阳性预测值也为零。反之,面对一个全部为有病的人群,即使采用高敏感性的诊断试验,其所有的阴性结果也都是假阴性。所以,当患病率为百分之百时,阴性预测值则为零。总之,预测值受诊断试验的敏感度、特异度和受试人群患病率的双重影响,当我们选定一个诊断试验时,即敏感度和特异度已经恒定不变,诊断试验的预测值则只随患病率而改变。患病率、敏感度和特异度与阳性预测值之间的关系可以用图 14-2 说明。

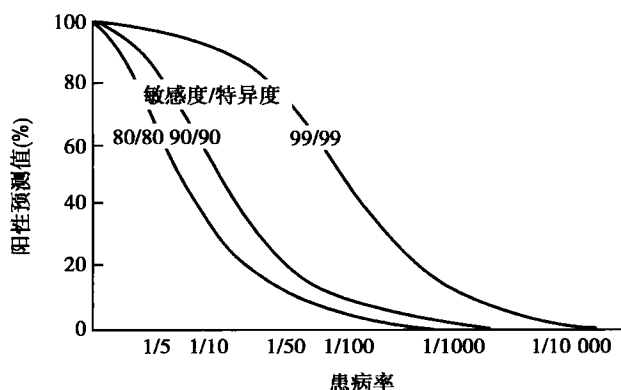


图 14-2 患病率、敏感度和特异度与阳性预测值的关系

受试人群的患病率高,则阳性预测值也高,诊断试验也越有价值。受试人群的患病率太低,阳性预测值也变得很低,甚至接近于零,此时诊断试验就失去了诊断价值。所以,在不同等级的医院和不同的门诊,面临的就诊人群的患病率差异很大,同一诊断试验的预测值结果也会有较大差异,如某一疾病的专科门诊会比普通门诊的这一疾病的患病率高,因此针对这一疾病的诊断试验的阳性预测值也必然高,故诊断试验使用价值也高。

#### 四、似然比

似然比 (likelihood ratio, LR) 提供同敏感度和特异度一类的信息,并用于计算阳性或阴性试验结果的患病概率。表示一个诊断试验结果出现在有病受试者和出现在无病受试者的比值大小,代表了一个诊断性试验区分有病和无病的能力大小。阳性似然比为出现在金标准确定有病的受试者阳性试验结果与出现在无病受试者阳性试验结果的比值大小或倍数。或表示试验结果阳性对一个受试者患病的概率与未患病的概率的比值或倍数,即真阳性率和假阳性率的比值。阴性似然比为出现在金标准确定的有病的受试者阴性试验结果与出现在无病的受试者阴性试验结果的比值。或试验结果阴性时,一个受试者患病概率与未患病概率的比值,即假阴性率和真阴性率的比值。

阳性似然比 (positive likelihood ratio, LR+) 与阴性似然比 (negative likelihood ratio, LR-) 的计算公式为:

$$LR+ = [a \div (a+c)] / [b \div (b+d)] = Sn / (1 - Sp) \quad \text{公式 14-11}$$

$$LR- = [c \div (a+c)] / [d \div (b+d)] = (1 - Sn) / Sp \quad \text{公式 14-12}$$



例：某医院心血管病内科病房，收治了急性心前区疼痛疑诊急性心肌梗死的患者共 360 例，在经冠状动脉造影诊断基础上，同步地检测了患者血清肌酸磷酸水平，在设定诊断异常界值  $\geq 80\text{u/L}$  条件下，获得了下列结果，结果见表 14-4。

表 14-4 血清磷酸肌酸激酶测定诊断急性心肌梗死

血清磷酸肌酸激酶 (u/L)	急性心肌梗死		合计
	有	无	
$\geq 80$	215	16	231
$< 80$	15	114	129
合计	230	130	360

$$Sn = a/(a+c) = 215/230 = 0.93$$

$$Sp = d/(b+d) = 114/130 = 0.88$$

$$LR+ = Sn/(1-Sp) = 0.93/(1-0.88) = 7.75$$

$$LR- = (1-Sn)/Sp = (1-0.93)/0.88 = 0.0795$$

在血清磷酸肌酸激酶诊断急性心肌梗死的例子中，阳性似然比为 7.6，是指急性心肌梗死患者出现血清磷酸肌酸激酶试验结果阳性的机会是非急性心肌梗死者出现血清磷酸肌酸激酶阳性机会的 7.75 倍；阴性似然比 0.0795，是指急性心肌梗死患者出现血清磷酸肌酸激酶阴性的机会是非急性心肌梗死者出现阴性结果的 0.0795 倍。

实际上，很多有病的受试者的试验结果并不能在诊断试验后得到非常“异常”的结果，而可能是接近临界值的“边缘异常”的结果，这常常使临床医师很苦恼。似然比摆脱了对一个试验结果简单的正常和异常分类，而是根据临界值的敏感度和特异度来描述诊断试验结果的准确性。在上面的这个例子中，如果我们来计算血清磷酸肌酸激酶的不同临界值的阳性似然比，见表 14-5，我们会发现随血清磷酸肌酸激酶的诊断试验的临界值增大，似然比的比值也增大，血清磷酸肌酸激酶  $> 280\text{u/L}$  在急性心肌梗死病人身上出现的概率比非急性心肌梗死的受试者高 55 倍，血清磷酸肌酸激酶值在  $80\text{u/L}$  至  $279\text{u/L}$  之间，在急性心肌梗死病人身上出现的概率比非急性心肌梗死的受试者仅高 4.4 倍。由此可见，阳性似然比越大，阳性结果反映有病诊断的准确性越高。

表 14-5 血清磷酸肌酸激酶测定诊断急性心肌梗死的分层分析

血清磷酸肌酸激酶 (u/L)	有急性心肌梗死		无急性心肌梗死		阳性似然比
	n	真阳性率	n	假阳性率	
$> 280$	97	$97/230 = 0.4217$	1	$1/130 = 0.0077$	55
$80 \sim 279$	118	$118/230 = 0.5130$	15	$15/130 = 0.1154$	4.4
$40 \sim 79$	13	$13/230 = 0.0565$	26	$26/130 = 0.2000$	0.3
$1 \sim 39$	2	$2/230 = 0.0087$	88	$88/130 = 0.6769$	0.01
	230		130		

似然比的另一个重要作用是能缩短诊断假设，即如果我们知道了或估计出验前概率（患病率），利用似然比可以计算出验后概率。在此之前，我们需要明确率和比的概念和二者之间的关系。例如，某一特定人群的患病率是指有病者占整个人群（包括有病者和无病者）中的比例，可以表示为患病率 = 有病者 / (有病者 + 无病者)。患病比则指这一人群有病者和无病者的比数，可以表示为患病比 = 有病者 / 无病者。如率 = 50%，则比 = 1；率 = 1/3，则比 = 1/2，因此，比 = 率 / (1 - 率)，反之，率 = 比 / (1 + 比)。利用似然比计算验后



比的公式如下。

$$\text{验后比 (posttest odds)} = \text{验前比 (Pretest odds)} \times \text{似然比} \quad \text{公式 14-13}$$

通过似然比计算验后概率，首先将验前概率（患病率）转换成验前比，计算得到的验后比再转换为验后概率。例如一位老年患者突发心前区疼痛，并测得血清磷酸肌酸激酶值为 170u/L，问发生急性心肌梗死的可能性有多大？

根据病史和体征，估计这位老年人患急性心肌梗死的可能性（验前概率）为 50%，根据表 14-5，170u/L 血清磷酸肌酸激酶的似然比是 4.4，因此，计算验后概率的步骤如下：

- (1) 验前比 = 1
- (2) 验后比 =  $1 \times 4.4 = 4.4$
- (3) 验后概率 =  $4.4 / 5.4 = 0.81$

这位病人患急性心肌梗死的可能性为 81%。

Fagan 绘出了验前概率和验后概率借似然比换算图（图 14-3），简化了换算过程，方便了临床医师在临床工作中的应用，可以随身携带。如前面在知道验前概率和似然比的情况下，借助米尺，连接验前概率 50%（左柱）与似然比 4.4（中间柱），可以得到验后概率大致为 80%（右柱）。

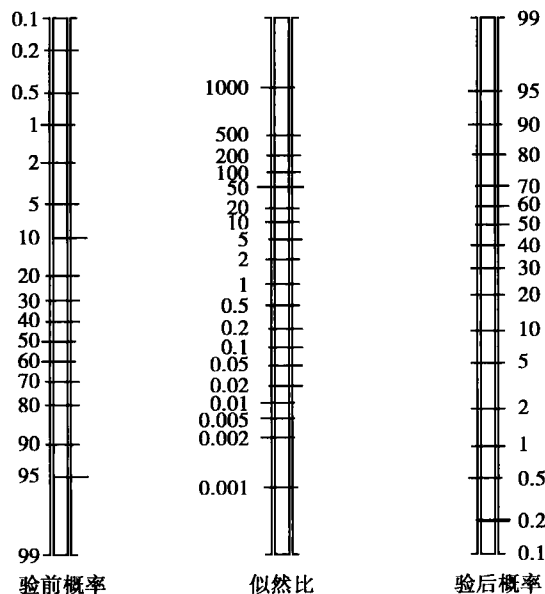


图 14-3 似然比与验前概率和验后概率换算图

## 第四节 受试者工作特性曲线

受试者工作特性曲线（receiver operator characteristic curve, ROC 曲线），在诊断性试验中，有两方面的重要作用，一是根据诊断目的确定参考值，二是对两个以上独立的诊断试验进行比较。

制图时以试验的敏感度（真阳性率）为纵坐标（Y），以 1—特异度（假阳性率）为横坐标（X），坐标轴上的率值由 0~1 或 0~100%。见前面表 14-3，这是一组采用餐后 2 小时血糖水平诊断糖尿病的数据。依照连续分组测定的数据结果（阈值），分别计算敏感度和特异度，并标于坐标。这样，每组（或每个阈值）的敏感度和相对应的特异度在坐标图



上都成为一个点，将这些点连成线，即为 ROC 曲线。根据每组血糖测定的数据演算出的敏感度和特异度划出 ROC 曲线，如图 14-4。

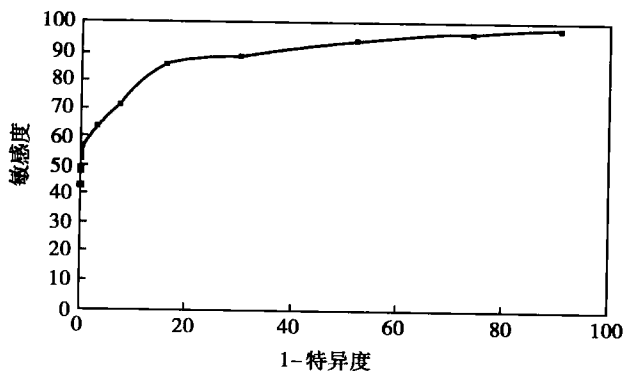


图 14-4 餐后 2 小时血糖水平诊断糖尿病试验的 ROC 曲线

## 一、确定诊断试验的参考值

理论上讲，一个完美的诊断试验的 ROC 曲线恰好通过左上角，左上角的敏感度（真阳性率）是 100%，假阳性率是 0（即是说特异度是 100%）。这样的诊断试验可以区分所有的病人和非病人，即受试对象中所有的病人均为阳性，而所有的非病人均为阴性，即无漏诊和误诊。如果一个诊断试验的 ROC 曲线恰是一条从左下角到右上角的、与 X 坐标轴呈 45°角的直线，则这个诊断试验完全没有区分病人和非病人的能力，即敏感度和特异度均为 0。因此，ROC 曲线越接近左上角，区分病人和非病人的能力越强，即准确性越高。

ROC 曲线可以更形象地表达敏感度和特异度在选择诊断参考值中的作用，由图 14-4 可见，离左上角距离最短的一点正是我们从表上探讨选择的诊断参考值，即血糖值为 110mg/dl，其敏感度和特异度之和最大，这一点或临近点常常被作为确定诊断试验阳性和阴性的诊断参考值，这些点称为最佳临界点，点上的值即为最佳临界值。最佳临界值的选择除考虑漏诊和误诊之和最小外，在实际的临床诊断中尚需要考虑漏诊或误诊给受试者带来的危害程度，如因漏诊而延误诊断并给受试者和（或）周围所接触的人群带来致命的危险，我们应选择敏感度更高而特异度相对较低的点作为临界值。反之，误诊给受试者带来严重的后果和更大的经济损失，诊断则应慎重，临界值应该选择特异性更高而敏感度相对较低的一点。因此，距离左上角最近的点，不一定是实际的最佳临界点，而应结合临床、综合考虑诊断性试验结果带来的利和弊。如果利和弊相当，当然视距离左上角最近的点为最佳临界点是比较恰当的，因为此点的误诊和漏诊之和为最小。

## 二、不同试验诊断同一疾病优劣的比较

ROC 曲线另一个重要的作用是比较同一疾病诊断的不同试验的优劣。ROC 曲线下的面积反映一个试验的整体准确性，面积越大，准确性越高。将 ROC 曲线下的面积用符号 A 表示，通常  $0.5 \leq A \leq 1$ 。如果 ROC 曲线恰是一条从左下角到右上角的、与 X 坐标轴呈 45°角的直线，这个 ROC 曲线下的面积则为 0.5，即  $A = 0.5$ ，则诊断试验不具备诊断能力，即完全不能区分病人和非病人。 $A < 0.5$  的情况在实际的诊断试验中是不存在的。如果  $A = 1$ ，则诊断试验非常完美，即能完全区分有病的病人和非病人，而无误诊和漏诊，这在现实中也是不可能的。因此，A 值越接近 1，则准确性就越高，诊断试验也就越完美。有关 ROC 曲线下的面积的计算通常采用两种方法，即非参数法和双正态参数法，前者包



括 Bamber 和 Hanley-McNeil 法。

图 14-5 显示对同一组病人进行的 CT 和放射性核素 (RN) 诊断脑肿瘤的 ROC 曲线, 显然 CT 的 ROC 曲线下面积较放射性核素的 ROC 曲线下面积大。同时, 我们可以对两种诊断方法的准确性通过 ROC 曲线下面积作检验来进行比较, 作出具有统计学意义的诊断准确度的差异来证明 CT 的整体准确性较高, 在诊断脑肿瘤中的价值更大, 具体的计算 ROC 曲线下面积和两样本 ROC 曲线下面积比较的检验方法参照本节提供的参考文献或者参照医学统计学的相关章节。

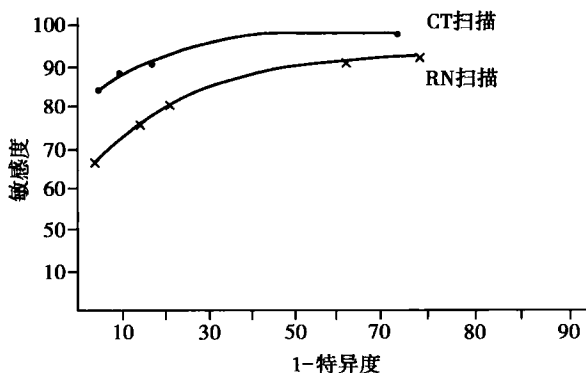


图 14-5 放射性核素 (RN) 与 CT 扫描诊断脑肿瘤的 ROC 曲线

## 第五节 参考值的确定

### 一、参考值的概念

参考值是反映健康状态下, 人的解剖、生理和生化等功能指标数据的波动范围。这些指标在检测后的数据结果如果是在确定的参考值的界限内, 临床上视为“正常”, 超出参考值的界限则视为“异常”。但需要注意, 有些“异常”的指标结果并非疾病的结果, 而有些患某种疾病的病人其相应的指标结果又可能在参考值的范围之内, 这是因为各项指标的“正常”和“异常”之间并无截然的界限, 健康状态下的指标波动范围和受疾病影响的指标波动范围存在交叉重叠, 因此确定参考值应能最大限度地减少漏诊和误诊的情况。

### 二、确定参考值的基本方法

#### (一) 均数加减标准差法

临床上常见的一些参考值, 如反映儿童发育指标的身高、体重等的频数分布是正态或接近正态分布, 其参考值的确定常采用均数  $\pm$  2 个标准差 (s) 表示, 这样 95% 测量值在参考值范围内, 两端各 2.5% 的测量值被定为“异常”。对某些指标, 只有过高或过低为非正常的状态, 则其单侧 5% 被定为“异常”。用这种方法确定参考值, 首先需要确定一个同质人群, 由获得的测量数据计算出均数和标准差, 确定出参考值。这种方法获得的参考值是一个统计学的界限, 并未充分考虑临床因素, 如无论什么疾病的诊断试验的结果超过 95% 即诊断为有病, 势必造成任何疾病的患病率都为 5%, 显然违背真实的情况。并且采用这种方法确定参考值, 需要假定疾病状态下测量值也有同样的正态频率分布。由于得到的参考值与临床疾病状态下的测量值之间存在交叉重叠, 因此, 需要临床实践的校正。

#### (二) 百分位数法

多数诊断试验的测定值的频数分布并非正态分布, 因此不能适用均数加减标准差法,

而采用百分位数法确定参考值。这种方法可以不考虑数据的分布，适用于偏态分布或分布类型尚不能确定的数据。通常确定为百分范围在 2.5%~97.5%的测定值为参考值，对于单侧指标可以确定第 5 百分位数或第 95 百分位数为参考值。由于此方法确定的参考值误差较大，因此测定的样本量要求较大，以减少误差。其缺陷同均数加减标准差法，也是统计学意义上的参考值，如果参考值包括 95%的健康受试者的结果，那么一个健康受试者接受两种相互独立的诊断试验的结果而诊断为正常的可能性只有  $95\% \times 95\% = 90\%$ ，经过  $n$  个诊断试验后，结果为正常的可能性依次递减 ( $0.95^n$ )。

**(三) 从临床治疗出发制定的参考值**

许多生理和生化指标达到或超出一定的范围即可能导致有严重后果的疾病发生，因而需要治疗使之维持在一个安全的范围内，这个即临床上称为的某项指标的危險水平，这需要流行病学研究来确定。如舒张压<90mmHg（12kPa）为安全界限，高于安全界限的病人发生脑血管意外的风险明显增加，而这个参考值的确定，就是基于长期的临床观察和流行病学研究得出的，对指导临床治疗有很大的实用价值。

**(四) ROC 曲线法**

本法确定诊断参考值具有重要的临床价值，颇为实用（见本章第四节）。

## 第六节 多项试验的联合诊断

临床工作中采用的诊断性试验并不总是完美的，实际上没有 100%特异度和敏感度的诊断试验，有些情况下，单一诊断试验的结果导致的漏诊率和误诊率可能是不能接受的。另一方面，对个体而言，单一的诊断试验的结果产生的验后概率可能不是很高或不是很低，例如，一个诊断试验后说某位病人的患胰腺癌的概率是 20%，对这种状况，医师和病人都不会接受手术或化疗，而是寻求新的诊断试验。即使是像肝脏血管瘤这样的非致死性疾病，在 CT 检查后给病人的结论是 70%的患病可能性，也是毫无意义的。因此，对群体，单一的诊断试验可能存在不能接受的漏诊率或误诊率；对个体，单一的诊断试验的结果给出的患病概率不是非常的高或非常的低，都需要联合多个诊断试验，以提高诊断的敏感度或特异度。当然，不是每种疾病都能实施多项试验，如无可以选择的备选试验或备选的试验有极大的危險性等。

多项试验的联合诊断有两种方式，一是平行多项试验的联合诊断，任何一项试验阳性即可以看作是疾病存在的证据。二是序列多项试验的联合诊断，在结束第一个诊断试验后，是否继续下一个试验，取决于前一个试验的结果，如果是阳性，则继续下一个试验，否则，则终止试验，只有所有的试验均为阳性，才被看作是疾病存在的证据。

### 一、平行试验

为了较快的作出诊断，我们往往需要同时作多个诊断试验，只要任一试验的结果阳性，即判断为患病者，这种方法称为平行试验（parallel test），见表 14-6。

表 14-6 平行试验的结果判断

结 果		判断
试验 A	试验 B	
+	—	+
—	+	+
+	+	+
—	—	—



通常平行试验增加了敏感度和阴性预测值,但降低了特异度和阳性预测值,即是说减少了漏诊、增加了误诊的机会。增加敏感度和阴性预测值的程度取决于一个诊断试验漏诊的病人能否被另一个诊断试验检出。举例来讲,用于一平行试验的两个诊断试验(A和B)的敏感度分别为80%和60%,敏感度较高的试验(A)检出了敏感度较低试验(B)检出的所有有病病例,这个平行试验的敏感度也只有80%。反之,每个诊断试验相互检出对方试验漏诊的有病病例,则这个平行试验的敏感度达到100%。假设用于平行试验的诊断试验相互间完全独立(即不是诊断同一疾病的类似试验或重复的试验),此平行试验的敏感度为92%。

$$\begin{aligned}\text{平行试验 } Sen &= SenA + (1 - SenA) \times SenB \\ &= 0.80 + (1 - 0.80) \times 0.60 = 0.92\end{aligned}$$

平行试验敏感度的增加是以牺牲特异度为代价的,特异度低于平行试验中的任一试验。试验A的特异度为80%,试验B的特异度为90%,假设用于平行试验的诊断试验相互间完全独立,此平行试验的特异度为72%。

$$\begin{aligned}\text{平行试验 } Spe &= SpeA \times SpeB \\ &= 0.8 \times 0.9 = 0.72\end{aligned}$$

平行试验特别适用于下面的情况,即当临床需要一种高敏感度的试验对某种疾病作出诊断,而现有的两种或两种以上的试验均不是特别敏感的试验,那么平行试验就是增加敏感度的唯一选择,这些情况在外科急腹症比较常见。而由于平行试验带来的误诊和误治应给予足够的重视,以及可能过高的花费也是评估平行试验考虑的重要方面。

## 二、序列试验

非紧急的情况下,并缺乏非常特异的诊断试验,而误诊将给病人带来无法挽回的损失,可以在一项诊断试验阳性的前提下,再考虑使用另一项诊断试验,如果也为阳性,我们才作出疾病诊断。或某项诊断试验的危险性很大或非常昂贵,首先使用较为安全或价廉的诊断试验并获得疾病可能存在的提示情况下,再使用这个诊断试验。这种序贯使用两个或两个以上的诊断试验,并均取得阳性的结果后再作出疾病存在的多项试验称序列试验(serial test)。

序列试验增加了特异度和阳性预测值,但降低了敏感度和阴性预测值,即是说,减少误诊是以牺牲可能增加的漏诊为代价的。因此,序列试验适于单一的诊断试验并不具有非常高特异度的情况时。同样用平行试验中诊断试验A和诊断试验B的例子,计算用于序列试验的特异度和敏感度。

$$\begin{aligned}\text{序列试验 } Spe &= SpeA + (1 - SpeA) \times SpeB \\ &= 0.8 + (1 - 0.8) \times 0.9 = 0.98\end{aligned}$$

$$\begin{aligned}\text{序列试验 } Sen &= SenA \times SenB \\ &= 0.8 \times 0.6 = 0.48\end{aligned}$$

如果用试验A和试验B做一个序列试验,仅就试验的工作效率(不考虑其他因素),是先用试验A,还是先用试验B?假设受试者1000人,患病率为20%,我们分别计算如下,见表11-7。

由表14-7得知,序列试验中,诊断试验无论用何种顺序,所得到的正确诊断和漏诊、误诊的受试者人数是一样的,但先使用特异度较高的试验,使接受第二个试验的人数大大减少,这样可以提高工作效率和降低费用。但是序列试验中的一个试验极其便宜并且安全性高,则宜先使用。





表 14-7 序列试验工作效率的比较

序列试验 1：先用试验 A，后用试验 B					序列试验 2：先用试验 B，后用试验 A				
A		疾病			B		疾病		
		+	-				+	-	
	+	160	160	320		+	120	80	200
	-	40	640	680		-	80	720	800
	合计	200	800	1000		合计	200	800	1000
320 例阳性受试者接受试验 B					200 例阳性受试者接受试验 A				
B		疾病			A		疾病		
		+	-				+	-	
	+	96	16	112		+	96	16	112
	-	64	144	208		-	24	64	88
	合计	160	160	320		合计	120	80	200

三、试验的独立性

以上讨论的多项试验的联合使用是基于参与联合试验的每个试验间是相互完全独立的假设，即是说用于其中一个诊断试验的信息不能被另一个诊断试验重复使用，否则联合试验的价值将被错误的估计。如果联用的多项试验中每一个诊断试验用于作出诊断的信息不同，则就有一个试验检出另一个试验的漏诊病例的可能（平行试验），或有下一个试验排除前一个试验误诊的可能（序列试验），这是增加敏感度或特异度的基础。如果联用的多项试验中每一个诊断试验并不是独立的，这个平行试验或序列试验的敏感度或特异度则增加非常有限，因此其价值可能会被夸大。

第七节 诊断试验结果的一致性评价

许多诊断试验研究是由多人组成的团队来完成，参加试验研究的不同测试者或观察者对诊断结果的判断可能存在差异。如对同一病人肝胆的超声诊断，不同的超声影像诊断医师可能会给出有差异的结果，即使同一位医师在不同的时间点所作的两次检查，也可能给出有差异的诊断。如果不同观察者对同一研究对象人群所作的观察和判断差异太大，势必影响诊断试验结果的可信度和研究的质量。因此，为了提高观察结果的一致性，应严格制定切实可行的诊断标准、稳定试验的环境因素、不同观察者对受试对象的观察或诊断应同期进行等一系列控制诊断一致性的措施，对观察者进行专业培训以提高诊断水平也是非常必要的。但即使实施了上述这些提高诊断一致性的办法，也不能避免诊断结果差异，那如何看待不同观察者之间的诊断结果的差异呢？不同的观察者给出有差异的诊断结果是难以避免的，但达到相对的一致性诊断性研究所追求的，即每个观察者间的诊断不符合率在可以接受的范围，这对于提高诊断试验的质量十分重要。

一、Kappa 值及其意义

实际的研究中，常借助 Kappa 统计来了解观察者间作出诊断的相互符合程度，以确保每个参加研究的观察者的经验和专业水平接近，使每个观察者间的诊断达到相对较高的一致性。Kappa 值是评价分类变量结果一致性的重要指标。



$$\text{Kappa } (\kappa) = (P_A - P_e) / (1 - P_e) \quad \text{公式 14-14}$$

不同研究者的实际的观察一致性或实际诊断符合率  $P_A$ ；不同研究者的机遇一致性或机遇诊断符合率  $P_e$ 。Kappa 值的取值为： $-1 \leq \kappa \leq 1$ 。即  $\kappa = -1$ ，提示不同研究者间的诊断完全没有一致性； $-1 < \kappa < 0$ ，提示不同研究者的实际诊断一致性小于机遇一致性，没有实际意义； $\kappa = 0$ ，提示两个研究者间没有一致性，不同研究者作出诊断结果的一致性完全是机遇的作用； $\kappa > 0$ ，不同研究者作出诊断的一致性增大、诊断结果的符合程度受机遇的影响减小； $\kappa = 1$ ，一致性最好，不同研究者作出诊断结果的完全符合，排除了机遇的影响。Kappa 值多大才不至于对诊断试验的结果产生严重影响，则需要针对具体的诊断试验的性质而定，一般来讲， $\kappa \leq 0.4$ ，表明一致性差； $0.4 < \kappa \leq 0.6$ ，表明中度一致性；Kappa  $> 0.6$ ，一致性较高，通常即可以接受；Kappa  $> 0.8$ ，说明一致性极高。

## 二、观察或诊断结果的一致性评价方法

两个研究者相互盲法对同一批参加诊断试验的研究对象进行诊断，其诊断的结果可能并不完全吻合，如 A 和 B 两名放射科医师对 88 例可疑胰腺癌的 CT 片进行盲法读片诊断，结果见表 14-8。

表 14-8 医师 A 和医师 B 对 88 例疑似胰腺癌 CT 片的盲法读片诊断

医师 A	医师 B		总计
	阳性	阴性	
阳性	54 (a)	4 (b)	58 (a+b)
阴性	2 (c)	28 (d)	30 (c+d)
总计	56 (a+c)	32 (b+d)	88 (n)

$$P_A = (a+d)/n \quad \text{公式 14-15}$$

$$P_e = [(a+b)(a+c) + (c+d)(b+d)]/n^2 \quad \text{公式 14-16}$$

这个例子中的实际诊断符合率， $P_A = (54+28)/88 = 0.9318$ ；

机遇符合率， $P_e = [(54+4)(54+2) + (2+28)(4+28)]/88^2 = 0.5434$ ；

Kappa 值， $\kappa = (0.9318 - 0.5434)/(1 - 0.5434) = 0.8506$

由此可见，A 和 B 两名放射科医师对胰腺癌的诊断有很高的符合率或是说一致性。

## 第八节 诊断试验的评价标准

任何新的诊断试验在临床应用之前，应该经过科学的严格评价，以对诊断试验有一全面地了解和对可适用性作出评估，下面是国际通用的评价原则。

### 一、诊断试验与金标准进行独立的盲法比较

金标准是保证受试者正确分类（有病和无病）的基础，诊断性试验只有与金标准比较才能确保其准确表达诊断价值。用盲法做比较和判断，可以避免观察者偏倚，以确保试验结果的真实和准确。

### 二、纳入受试者的合理性

病例组应该包括从轻度到重度、治疗与未治疗等各种类型的病人，对照组也应包括一些其他常见的易混淆的疾病病例，这样对总体有代表性的研究样本才能反映试验的诊断效



能,即诊断试验除能正确诊断疾病外,并且可以判断病情严重程度及鉴别类似临床表现的其他疾病,因此,这样的诊断试验更有临床实用价值。

### 三、充分告知研究的地点、环境、受试者来源

研究地点、环境、受试者来源不同,与诊断试验效能有密切关系的患病率会有很大差异,直接影响诊断试验的预测值。详细地报告这些情况,对读者判断该诊断研究是否能适用于自己所在地区和相应的环境及自己诊治的病人。

### 四、诊断性试验的重复性

诊断试验作为判别疾病存在与否和疾病严重程度的工具,本身应具有重复性,重复性又称稳定性和可靠性,即对同一疾病的多次测试的指标结果相同或接近,即是否作了“一致率”的分析。报告重复性应对试验的控制条件、观察者的熟练程度予以详细说明。

### 五、诊断性试验的参考值是否合理、可靠

参考值的含义直接影响参考值的数据,虽然正态分布数据的参考值用平均值加减2个标准差表示,非正态分布数据的参考值用百分位数表示,但需要明确的是,这种采用统计学方法确定的参考值可能与临床意义有一些差距,尤其在采用相互独立的诊断试验进行多项试验联合诊断时对结果的解释更应注意。

任何诊断试验的确定的参考值均有一定漏诊率和误诊率,ROC方法确定的参考值可以明确漏诊和误诊的比例,并可以根据诊断的目的的不同,确定敏感度高的或特异度高的参考值,或确定漏诊和误诊之和最低的参考值。

### 六、如果是联合诊断试验,是否每一个试验的敏感度与特异度等重要指标都进行了测量

对联合试验研究中每一个单独的诊断试验进行全面评价,才能准确判断联合试验的真实性,并为读者提供了解及应用这些试验的客观依据。

### 七、对诊断试验操作步骤、注意事项等是否有详细描述

新的诊断试验能被迅速普及应用才能产生社会效益,使更多的病人得到及时正确诊断和治疗,因此要详细介绍诊断试验的操作步骤、仪器设备、试剂、试验条件等,利于他人重复或应用。

### 八、诊断试验的实用价值是否予以描述

作者应对新的诊断试验的应用前景、可能带来的社会效益、安全性等作出实事求是的评价,以便推广使用。

(刘续宝 王家良)

## 第十五章 治疗性研究设计与评价

临床治疗性研究与评价是临床流行病学的重要内容之一。随着新药、新的治疗方法的不断问世,许多疾病的治疗药物和治疗手段已经趋于多样化。临床医生在做出正确的诊断之后如何在众多的治疗措施中选择安全、有效的措施和方案已成为临床医生的一项重要任务。

治疗性研究与评价是指在临床实践中以人为研究对象,应用医学科研的理论和方法,通过科学严谨的设计(design)、精确的测量(measurement)对所研究或选择治疗的效果进行客观的评价(evaluation),以达到提高治愈率,降低病残及病死率,提高生存质量,改善人体健康的目的。由此可见,临床治疗性研究与评价往往是通过临床试验完成的。

治疗性研究与评价的内容既可以包括药物、手术的疗效,也可以包括各种预防性措施的效果。尽管评价的内容不同,但都必须遵循一定的评价原则。

20世纪80年代以来,国内外在临床治疗性研究与评价方面有了较大的进展,越来越多的随机对照试验和多中心临床试验的研究方法被用于临床治疗性研究,不断取得新的研究成果,对于提高临床治疗或预防水平,有着相当重要的意义。因此,掌握正确的治疗性研究与评价方法对医务工作者具有重要的意义。

### 第一节 治疗性研究与评价的重要性及基本条件

#### 一、治疗性研究评价的重要性

传统的临床治疗性研究多属经验性总结,缺乏严格设计的前瞻性研究或回顾性研究颇多,因而研究结果真实性有许多质疑之处,临床应用的重叠验证也存在一些问题。近数十年来,由于临床流行病学的发展,催生了循证医学。两者均十分注重临床研究的随机对照试验,因此,随机对照试验的研究成果也随之日益增加。尽管一些报道的RCT结果并非均为“一流”,但却有一些高水平的RCT研究否定了某种(些)“传统有效药物或措施”的效力,如像冰冻疗法治疗溃疡性出血、乳内动脉结扎治疗顽固性心绞痛、抗心律失常药物预防急性心肌梗死致猝死等,极大促进了临床治疗的进步。

临床治疗性研究的对象是病人,因此,凡是人体外研究的基础医学研究结果,在未经临床治疗研究证实之前,是不允许直接用于临床病人治疗的。例如,体外实验证实:阿糖胞苷(Ara-c)有干扰嘧啶合成的作用,可抑制多种DNA病毒,能抑制播散性带状疱疹病毒。然而经临床试验发现,Ara-c不仅对播散性带状疱疹的临床治疗无效,反而有害。

面对着十分丰富的各类临床治疗性研究试验的结果,需应用临床流行病学关于治疗性试验原则、国际公认的评价标准,进行严格评价(critical appraisal),方能肯定有价值的研究结果,并用于提高临床治疗水平过程之中。同时,在这一评价过程中,也能提高自己的批判与接受能力,既能防止误导,又能提高自己的学术水平。

#### 二、治疗性研究评价的基本条件

临床治疗性研究评价一般要考虑下列条件:

##### (一) 临床治疗新试验要有充分的科学依据

任何一项临床疗效研究都要有其研究的背景和充分的立题依据,应充分掌握其历史与



现状,明确试验的目的或目标、特色及实际价值,而不能仅凭经验或不足的依据来做研究的假设或推断。

早在 1964 年于赫尔辛基召开的第 18 届世界医学大会通过,并经数次修改的人体试验所应遵循的赫尔辛基宣言(The Declaration of Helsinki)中就规定:“凡涉及人体的生物医学试验,必须遵循科学的原则,应建立在足够的实验室和动物试验及科学文献认识的基础之上。”强调用于人体治疗性试验的任何药物或措施,在针对疾病发病的机制或对某种环节的干预应有理论依据,同时还应该具备药物化学、药理学、毒理学、药动学以及药效学等基础医学研究的资料。在初步证明对患者具有安全性和有效性之后,方可投入临床治疗性试验。例如:治疗高血压的药物盐酸贝那普利(洛丁新),经临床前药理学研究发现,其作用机制是盐酸贝那普利水解成具有活性的贝那普利后,通过抑制血管紧张素转换酶(ACE),阻止血管紧张素 I 转变为血管紧张素 II,使全身外周血管舒张,降低血管阻力,从而产生降压作用。临床药理学研究发现,盐酸贝那普利服药后迅速吸收,30 分钟后原形盐酸贝那普利血浆浓度达到峰值,95%与血清蛋白结合;原形盐酸贝那普利在 4 小时内迅速从血浆中消除,贝那普利主要经肾和胆汁代谢。其药代动力学很少受年龄和轻、中度肾功能不全及肾病综合征的影响,肝硬化所致的肝功能不全者贝那普利的药代动力学和生物利用度均不受影响。在充分取得上述研究结果的科学证据后才可进行临床治疗性试验。

## (二) 明确研究拟达到的最佳目标

拟研究的药物或治疗措施有着不同的治疗效能。因此,在治疗目标的选择方面,除了要考虑治愈和根治外,还要考虑缓解症状,维持功能,预防复发或并发症等,故应根据所研究疾病的具体情况和试验药物本身的效能进行审慎的选择。

1. 临床治愈或根治 如果所观察的疾病属于可被治愈或根治的疾病,那么任何临床试验都应力求最大限度地实现这一治疗目标。如抗生素治疗某些细菌性感染性疾病的临床试验研究,外科手术根治早期肿瘤的临床研究等。

2. 预防复发或并发症 有些疾病在急性期控制后,幸存者或痊愈者在某种情况下有可能复发或发生某种并发症而引起更加严重的后果。对这类临床问题作研究决策的时候,其主要目标应是有效地预防复发或减少某些并发症的发生,以达到改善预后的目的。如对脑梗死得到控制的病人,临床治疗性试验的目标应是预防其再次复发。而对急性心肌梗死幸存的患者,临床治疗性试验的目标应是预防心肌再梗死和有关的并发症,如心衰、严重心律失常等,以降低病死率和提高生存质量。

3. 缓解症状、维持功能及改善生存质量 某些不能彻底治愈的慢性病患者往往存在一些临床症状,影响日常生活功能。对此,临床治疗试验的目标在于缓解其症状,最大限度地改善其功能状态及生存质量。如脑血管疾病致残后的康复治疗;类风湿疾病治疗控制关节疼痛,防止畸形,改善生存质量等。

## (三) 要筛选最新有效的试验药物或措施

在选择投入临床治疗性试验的药物或措施时,除了要有科学依据及安全有效性外,还应该善于从不同的药物或制剂中筛选最新有效的药物或措施进行疗效试验,这样可避免低效劳动或无价值的重复试验。尽管有些药物或治疗措施不能显著提高疗效水平,但其副作用发生率低或程度轻,或者成本低,也应是考虑的重要方面。

## (四) 确定研究的最佳治疗水平的终点指标

治疗效果的评价是通过一定的指标变化来体现的。这些指标包括终点指标和中间指标。终点指标代表疾病的最终结局,如某些肿瘤外科手术或化疗后不同时期的存活率或病死率;中间指标是指疾病发展变化过程中的一些结果,如某些药物在治疗乙型肝炎期间转氨酶的变化水平及其他实验室指标水平等等。治疗终点指标的最佳水平要有合适的标准,



过高或过低均可影响研究质量。应根据疾病的性质、病损程度、治疗后机体的病理损害和生理功能状况的可复性等选择而定。例如：高血压病治疗的最佳目标是防止心脑血管重要靶器官损害，并使血压下降到最佳治疗水平。

### （五）遵守临床试验研究的伦理学问题

临床治疗研究是以人作为研究对象，所以必须面对医学伦理学问题。应遵循赫尔辛基宣言的规定原则。具体要求详见本书第十二章。

## 第二节 临床治疗性研究的设计

### 一、立题的依据和研究的目的

一个治疗性试验研究，一定是针对临床实际中需要治疗并提高疗效的疾病。因此，研究的本身应具有科学依据、临床重要价值以及明确的研究目的。

立题的依据可以来源于基础研究的提示，也可以是动物试验结果在人体的进一步验证，更多的是来源于临床医生的实际观察和总结以及来自对人群流行病学的观察和研究。但无论来自于哪一方面，其创新性及实用性是必不可少的。

#### （一）研究的目的

临床治疗性研究，应根据被研究疾病的具体情况，以及所采用的试验干预措施（或药物）的治疗效力。明确课题研究所假设的预期目的，主要有三种情况：一是临床治疗或根治，目的是力求提高治愈率，降低病死率、伤残率；二是预防疾病并发症与复发，通过干预性研究达到降低并发症发生率、复发率，改善预后的预期目的；三是缓解症状，提高生存质量。不同的预期目的在试验方案的选择和研究实施方面都有所不同。

#### （二）干预措施（药物）的科学依据

治疗性试验的对象是人体（病人及健康人），因此，用于治疗性干预试验的措施或药物，务必要通过基础医学的有关实验研究（如药物化学、药理、毒理、药物动力学、病理学等等），证明具有治病的效力且无明显的毒效依据，然后再选择适量的健康人作Ⅰ期临床试验，被证明无明显不良反应者（具有若干科学的量化指标），并经行政主管部门审核批准，方可进行临床Ⅱ/Ⅲ期临床试验。

任何临床试验，一定要符合临床试验的伦理学要求，充分保障受试者的安全和人权（参考本教材十二章）。

如果有两种以上的干预措施（药物）可用于对同一疾病及其同一治疗研究目的者，则应从中比较并优选其一作临床治疗试验的对照性研究。

### 二、研究方案的选择

治疗性研究方案的选择多以试验性研究为主。按照随机、对照、盲法以及基线可比的临床试验设计的原则与方法，可分为随机对照试验、非随机对照试验、历史性对照试验、自身前后对照试验、交叉对照试验及序贯试验等，其次还可采用非试验性的研究方法，如队列研究。应结合研究中的实际情况加以选择，具体的设计方案参阅本教材第九章。现仅从治疗性研究的适用角度予以阐述几种主要的设计方案。

#### （一）试验性研究方案

1. 随机对照试验（randomized controlled trial, RCT） 随机对照试验是治疗性研究设计首选的金方案。RCT方案通过随机分组、设立对照、实施盲法，可有效防止若干混杂或偏倚因素的干扰，同时能确保研究对象具有一定的代表性以及基线可比性等特点，因



此, 获得研究结果的真实性最佳, 被誉为临床试验的金标准方案。

随机对照试验既可以用于某种药或措施的临床治疗性试验, 又可以用于群体性疾病的预防性试验, 如像麻疹疫苗、病毒性肝炎疫苗等人群预防性对照试验等。同时还可用于一些特殊情况下的病因学研究。

在较为复杂的 RCT 治疗性试验中, 以分层随机对照试验最为常用。首先选出对研究对象治疗结果有较大影响的重要特征 (如年龄、性别、病情、临床分型等), 然后按此将研究对象分为不同的层, 再在每一层内分别进行单纯随机, 将研究对象随机分配到试验组或对照组, 分别接受相应试验处理, 这样可基本维持某些重要临床特征在组间的均衡可比。分层因素应精选, 保持最小量, 以避免过多分组。因此, 在选择分层因素时要考虑该因素应与疾病或重要并发症的发生有重要关联, 同时对治疗效果及预后有明显影响。

例如: 对乳腺癌患者进行两种不同化疗药物疗效的随机对照研究, 考虑有些因素可能对结果造成影响, 故选择: ①闭经与否; ②肿瘤大小:  $\leq 2\text{cm}$  与  $> 2\text{cm}$ ; ③淋巴结转移情况: 1~3 个与 4 个以上作为分层因素。将入选的合格病例, 依照上述三种因素分层后, 再随机分配到试验组或对照组 (图 15-1)。这样, 上述三种因素在试验前的两组间分布是一致的。从而使得组间基线可比, 控制了混杂因素的影响, 会使试验结果更为真实可靠。

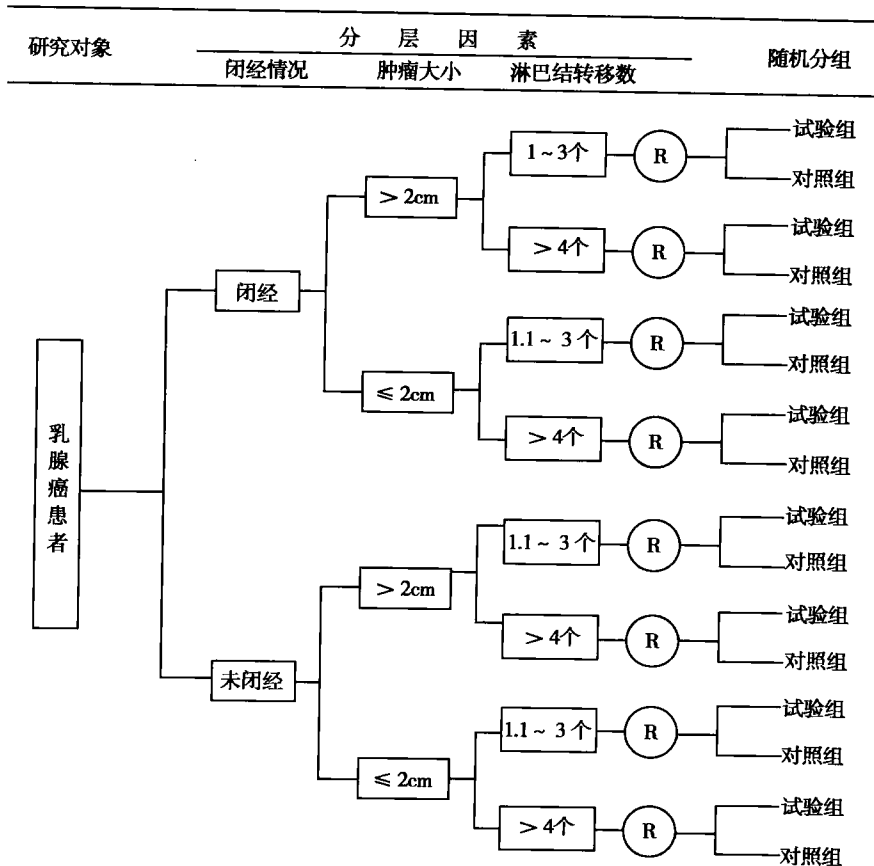


图 15-1 乳腺癌病人分层随机分配示意图

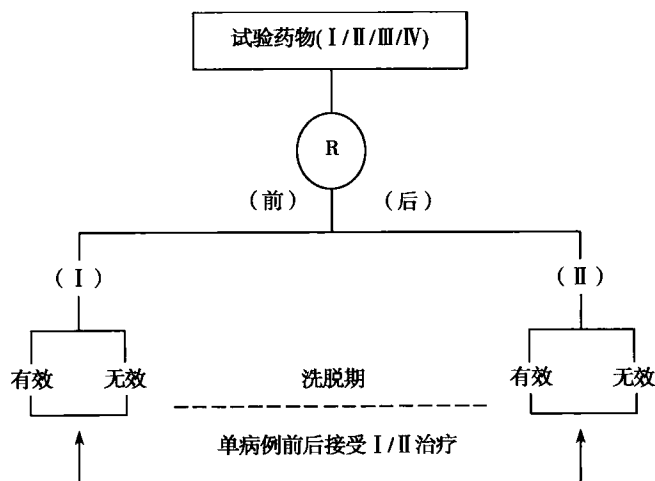
有些情况下, RCT 不需要分层随机, 如大样本随机对照试验。若研究的样本量足够大, 简单随机分组后, 两组研究对象基本能实现基线均衡。即使在试验终末阶段, 组间基线不一致、可比性欠佳, 由于总体样本量大, 故可在分析阶段采用分层分析或多因素分析



等方法，同样可以获得真实性较好的研究结论。

2. 单病例随机对照试验 (N of 1 RCTs) 尽管随机对照试验是研究药物有效性和安全性的最佳方案，但其结果往往仅反映研究对象对药物的平均效应水平，因此一项结果明明有效的随机对照试验，对于某个具体病人，其疗效可能低于平均水平甚至无效。此时可考虑选用单病例随机对照试验。

单病例随机对照试验是将随机对照的原理应用于单一病例所进行的试验。在试验过程中，受试者交替接受试验药与对照药。试验的目的在于明确哪一种药物对病人更有效。因此，其随机分配的对象是药物或干预措施（试验药物或对照药物），而不是病人。研究过程中要求采用双盲法。在每个观察期间及每轮试验间歇设有一段合理的药物洗脱期。当试验结果达到试验药物的预期研究目标时，则可终止试验。其试验模式见图 15-2。



注：在同一单病例的RCT设计，是将所用的各种药物进行随机对照试验，如 I : II, I : III, I : IV 等

图 15-2 单病例随机对照试验模式

单病例随机对照试验并非适合于所有疾病以及所有干预措施的研究，它仅适用于某个慢性病患者，因同时服用多种有效或无效药物、而需要进行的筛选抉择试验。因此，要充分考虑研究的需要性及可行性等。

3. 交叉对照试验 (cross over design, COD) 在临床研究实践中，对于某些慢性、非根治的疾病，特别是病人来源及研究时间有限时，若采用金标准的随机对照试验，尽管最佳，但往往就会遇到病人不足、研究期长等诸多困难。因此，综合考虑科学性与可行性，可采用交叉试验设计。

交叉设计分为两个阶段，在前一阶段（期），将合格研究对象随机分配至试验组或对照组，分别接受相应的干预治疗，疗程结束后分别统计分析疗效的结果及其差异。然后经过一定的洗脱期后，在后一试验阶段（期），试验组及对照组的患者，则与前期的试验或对照干预互相调换，即试验组患者接受对照组的干预措施，而对照组患者后期接受试验干预措施，疗程结束后，再分别统计分析疗效及其差异。最终可将两阶段治疗结果进行综合分析。具体的设计方案可参考本教材第九章。

由于这种设计方案采用了随机、对照及盲法，且试验本身又可消除个体内在环境的差异，尽管疾病前后两个阶段可能有程度的不同，但最后的综合性分析在一定程度上可弥补这一不足。因此，从总体上看，其科学性不逊于 RCT，且可行性更是其长处。所以在临





床治疗性试验的证据论证强度上,仍属于一级设计方案的范畴。

例如,地尔硫草治疗肥厚型心肌病的随机交叉试验,一组先服用地尔硫草,而另一组服用安慰剂,经过一段时间的治疗后,两组进行交叉,最后综合比较地尔硫草和安慰剂效果。

4. 自身前后对照试验 (before-after study) 自身前后对照试验的特点是仅设一组合格试验对象,他(她)们分别接受前后两个阶段(期)的药物干预试验治疗。这种前后两种不同的干预对照药物,通常用随机法分配药物,如随机法确定先用对照性药物(前期),那么后期则用试验药物,期间应有适当的停药洗脱期,最后将前后两个阶段的效果进行综合统计分析和评价。

自身前后对照试验的适用范围和应用指征与交叉试验相同,但科学性不及交叉试验。

例如,采用前-后对照研究评价地尔硫草治疗肥厚型心肌病的疗效,首先让全部研究对象服用安慰剂,观察其疗效。此阶段结束后,对研究对象停用一切药物约一周,然后让全部研究对象服用地尔硫草,继续观察疗效。最后对服用安慰剂和服用地尔硫草的两阶段疗效进行比较。

5. 非随机对照试验 (non-randomized controlled trial, NRCT) 作为临床前瞻性对照试验,非随机对照试验的设计,通常是不可取的。因为非随机化的样本分组,发生人为选择性偏倚的几率太大,故研究结果的真实性远不及 RCT。关键的缺陷就发生在随机化的分组环节,因此,从临床流行病学的角度,在临床科研设计方面是应予避免的。

然而,从医学杂志发表的文献看,非随机对照试验研究并不少见,故在分析评价其证据的论证强度方面,应列为“三级”证据。

6. 历史对照试验 (historical control trial, HCT) 是将现在患某病的病人作为试验组,对之采用新的干预措施。对照组不是在同时期确立的,而是将过去某一时期患同种病的病例作为对照组,这些患者患病时接受过传统疗法或干预措施,然后比较两组的疗效。可用文献资料作为对照,也可以将本单位的历史资料作为对照。本设计方案的主要优点是患者和临床医师均易接受采用的临床治疗措施,所以较易实施。该方案最主要的缺点是试验组和既往治疗组间的可比性较差,而且影响的偏倚因素太多,其研究结论的真实性,甚受质疑。

## (二) 非试验性研究设计方案

临床治疗性试验研究方案尽管丰富,但并非唯一。例如对某种疾病进行两种或多种疗法/药物试验时,有时病人要主动选择。鉴于尊重病人的选择权利和医德的原则,在某些特殊的医疗环境下,可以设计非试验性的研究方案,同样可以达到研究的目的,代表性的方法为队列研究 (cohort study) 和病例对照研究 (case-control study)。现以队列研究作为代表予以阐述。

在治疗性队列研究设计时,即将符合某病诊断标准及所设计纳入标准的病人,依他(她)自愿选择的某一治疗措施,分别纳入相应的队列,以接受相应的治疗,最后进行队列间的疗效分析与评价。例如,符合同一诊断标准的若干肺癌患者(病理诊断与临床分期均一致),均接受了手术治疗,他(她)们的主要临床基线可比性好。术后,有的患者愿意接受化学疗法及放射疗法;另一些患者考虑放化疗不良反应大,或不适于接受放化疗者,因此他(她)们或许用中药,甚者可能不愿意接受其他任何治疗。在这种情况下,研究者拟研究肺癌术后放化疗的疗效以及对远期预后的影响。于是可采用队列研究的设计,其中一个队列为术后接受放化疗者,另一队列进行同步观测,追踪两个队列的病死率及生存率,借以评价肺癌术后接受放化疗的患者是否优于对照队列(未接受放化疗)(图15-3)。

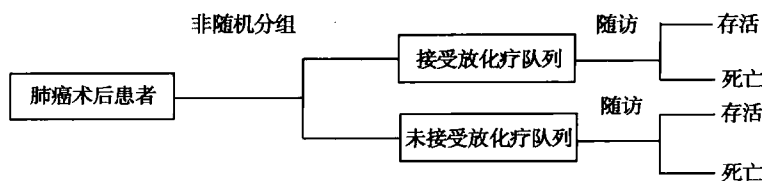


图 15-3 肺癌术后放化疗队列研究图

### (三) 多中心临床试验 (multi-center clinical trial)

是指由一个或几个单位的主要研究者总负责,多个单位的研究者合作,按同一方案进行的临床试验。这种研究方案的特点是收集病例快、病例多、试验规模大,因此完成临床试验需要的时间较短;研究的范围广,样本的代表性好,结论外推性强,但由于参加的单位多、人员多,故不宜进行质量控制和标准化,需要的研究经费也很多。

## 三、研究对象的选择

首先要求纳入的每个研究对象必须符合公认的临床及有关金标准诊断。为避免过多、过杂的临床病征的干扰,维持研究对象主要特点的相对均质性,从而获得较为准确的治疗观测结果,需要在符合诊断标准的病人中,根据研究目的,分别制订纳入标准及排除标准,以选择合格的对象进行试验。

例如:某一新型钙离子拮抗剂被用以治疗冠心病心绞痛的疗效研究。纳入的对象必须是冠心病伴有心绞痛的患者,但并不是所有的冠心病患者都符合研究目的的要求。所以必然有其纳入标准,如规定年龄在 65 岁以下,心电图有缺血型 ST-T 变化以及本人愿意参加试验等。为了保证纳入对象的标准确切,还需要规定排除标准,凡符合其中之一者,一律不予纳入研究。如该研究课题排除标准为:心肌梗死性心绞痛,心功Ⅲ级或以上者,严重心律失常者以及血压水平低于  $BP < 100/50\text{mmHg}$  者等。

有了明确的诊断、纳入及排除标准,就能确保研究对象的纳入准确可靠。但是,纳入标准的制订不宜过严,排除标准也不宜过多,否则就可能影响研究结果的代表性及适用性。有时也可能影响在研究期内获得足够的合格样本量而使研究达不到预期目的。

## 四、样本量的估计

样本量的计算是根据研究设计的有关类型(如 RCT 或队列研究等),以及有关研究假设的若干参数水平(如干预措施差异显著性水平,容许的Ⅰ型错误的水平等),科学地计算研究课题需要的最低样本量,以避免样本量的不足影响研究质量。通常在计算样本量时要考虑下述因素:

### (一) 试验组与对照组显著性差异的设定

样本量估算首先应当提出试验组与对照组两组间疗效的显著性差异水平的假设。计数资料以试验组和对照组有效率的差值作为有效假设的基础;计量资料则以试验组和对照组均值差作为有效假设的基础,分别作为组间疗效差异的参数来计算样本量。如显著性差异程度越大,则样本量就少,反之样本量就大。

### (二) Ⅰ类 ( $\alpha$ ) 误差和Ⅱ类 ( $\beta$ ) 误差的水平

1. Ⅰ类 ( $\alpha$ ) 误差 即试验设计所容许的假阳性错误水平,通常限定不超过 0.05 (5%)。该类误差越小,所需要的样本量越大。

2. Ⅱ类 ( $\beta$ ) 误差 即试验设计所容许的假阴性错误,通常限定为 0.1,不宜超过 0.2。 $1-\beta$  为检验效能 (power),又称把握度。此值越小, $1-\beta$  越大,要求的样本量也



越大。

(1) 两组率的比较：根据上面的这些基本设计参数，应用公式 15-1 计算出各组所需的试验样本量。

$$n = \frac{\pi_1 (100 - \pi_1) + \pi_2 (100 - \pi_2)}{(\pi_2 - \pi_1)^2} \times f(\alpha, \beta) \quad \text{公式 15-1}$$

其中  $n$  为计算所得一个组的样本量； $\pi_1$ 、 $\pi_2$  为试验组和对照组的事件发生率（如有效率）； $f(\alpha, \beta)$  为限定假阳性和假阴性水平时相应的数值，可以由表 15-1 查出。

表 15-1 常用  $f(\alpha, \beta)$  数值表

$\alpha$	$\beta$			
	0.05	0.10	0.20	0.50
0.10	10.8	8.6	6.2	2.7
0.05	13.0	10.5	7.9	3.8
0.02	15.8	13.0	10.0	5.4
0.01	17.8	14.9	11.7	6.6

例如：某一研究课题选用甲乙两种药物对糖尿病患者进行治疗，甲药的有效率为 70%，此作为有效药物的对照组；乙药为新药，假设有效率为 90%，此作为新药试验组，现要进行随机对照试验，设  $\alpha=0.05$ ， $\beta=0.1$ ，问每组至少需要多少病例？

已知  $\pi_1=70\%$ ， $\pi_2=90\%$ ， $\alpha=0.05$ ， $\beta=0.10$ ，查表 15-1 得， $f(0.05, 0.1)=10.5$ ，代入上述公式得：

$$n = \frac{0.7 \times (1 - 0.7) + 0.9 \times (1 - 0.9)}{(0.9 - 0.7)^2} \times 10.5$$

$$= 79$$

结果表明，每组需观察 79 个病例。

(2) 两组均数的比较：作为疗效差异显著性依据的研究课题，其样本量估计可采用下一个公式。

$$n_1 = n_2 = 2 \left[ \frac{(u_\alpha + u_\beta) s}{\delta} \right]^2 \quad \text{公式 15-2}$$

此式中的  $n_1$ 、 $n_2$  分别为两样本所需含量，一般需要相等， $s$  为两总体标准差的估计值，一般假设其相等或取合并方差的平方根， $\delta$  为两均数的差值， $u_\alpha$  和  $u_\beta$  分别为检验水准  $\alpha$  和第 II 类错误的概率  $\beta$  相对应的  $u$  值。

例如：观察两种药物治疗肌痉挛的疗效，其中 B 药使肌痉挛分数平均减少 2.16，L 药使肌痉挛分数平均减少 1.66，设两种药物疗效的标准差相等，均为 0.7 分，要求  $\alpha=0.05$ ， $\beta=0.1$ ，若要得出两处理差别有显著性结论，需要多少研究对象？

已知： $\delta=2.16-1.66=0.5$   $s=0.7$  双侧  $\alpha=0.05$ ， $\beta=0.1$  查  $u$  值表得： $u_{0.05}=1.96$ ， $u_{0.1}=1.28$  代入公式得：

$$n_1 = n_2 = 2 \left[ \frac{(1.96 + 1.28) \times 0.7}{0.5} \right]^2 = 41.2 = 42 \text{ 例}$$

故认为两个药物组各需 42 例患者，两组共需要 84 例。

在实际工作中，可以采用下述公式进行计算，较为方便且更为常用。

$$n = \frac{2\hat{\sigma}^2}{(\hat{\mu}_2 - \hat{\mu}_1)^2} \times f(\alpha, \beta) \quad \text{公式 15-3}$$

$n$  为每组所需的例数， $\hat{\mu}_1$ 、 $\hat{\mu}_2$  分别为两组的预期均数， $\hat{\sigma}^2$  为两组的合并标准差或对照组的标准差， $f(\alpha, \beta)$  可由表 15-1 中查出。上例用此公式计算结果相同。



由于试验中多半要丢失些病例,故常在计算量的基础上增加10%。

上述两类样本的计算方法简单且较为实用。但由于各种研究方案的不同以及设计的有关假设参数各异,因此,样本量的计算方法也就是多种多样的了。故在实际应用中,应参阅临床流行病学或临床统计学专著。

## 五、试验药物或措施的选择

治疗性试验所应用的药物或措施,首先要有科学的证据,要有临床前期的观察,证明其有效性和安全性,同时具有一定的创新性。如果没有这些最基本的科学依据,任何药物或措施是不容许用作临床治疗性试验的,因为临床治疗性试验的对象是病人,务必遵循赫尔辛基宣言的规范原则。对照试验如应用阳性对照药物或安慰剂,应该在外观、色泽、气味、制剂等方面与试验组药物一致,服用方法和疗程也需要一致,否则会影响结果的真实性。

## 六、试验效果测试指标的选择

任何药物或措施所呈现的治疗效应,包括疗效及药物不良反应,都要采用某种测量方法和指标加以度量,并将这些指标作为最后判断治疗效果的依据。总的要求是:

1. 灵敏度要好 对于治疗引起的客观反应,测试指标要能敏感地发现并能度量。如评价结核菌素免疫疗法治疗乙型病毒性肝炎的效果时,若用琼脂扩散法测试乙肝病毒标志,其敏感性比放射免疫法要差得多,如用该法作为测量治疗反应的方法和指标,必然大大地增加假阴性率。所以测试方法的敏感度要高。

2. 特异度要强 对治疗反应的结果,采用的测试方法和指标除了敏感之外还要特异。例如,冠心病急性心肌梗死应用溶栓疗法,在治疗的前后,采用冠脉造影,比较分析冠脉狭窄和闭塞改善的程度,以作为疗效的测量指标。这种指标的特异度强,有助于结论的可靠性。

3. 经济可行 在考虑敏感度和特异度的基础上,应从多种方法中,选择经济及可行性良好的测试方法和指标。如在评价输卵管堵塞病人复通术的效果时,为了解术后输卵管的通畅情况,可选用输卵管子宫碘油造影或腹腔镜下通美蓝。这两种方法的敏感度和特异度均较高,但前者较为经济,操作相对简便,对病人的创伤小,故应优先考虑。

4. 注意远期效应的测定 对于某些慢性病治疗措施的效果,除测试和评价近期效应外,更要追踪观察远期效果,这样有助于获得更为可靠的结论。例如,高血压患者的治疗不仅要观察用药后血压的控制水平,而且更要观察高血压所致的心、脑、肾不良事件的发生情况。这表明建立临床远期治疗效果的测试方法与指标是很重要的。

5. 测试的终点指标 测试指标的选择应该以治疗试验的终点目标而定。如治疗的终点目标是降低病死率和非致死事件发生,则测试指标定为病死率、生存率以及非致死事件发生率(如冠心病发生心肌梗死、心衰等)。验证治疗措施本身的有效性,则采用临床公认的有效或无效的效果判断标准,精选有关临床及实验室的定量及定性指标予以测量。如用血压计测量降压效果;以血糖水平测量评价糖尿病降糖药物的效果等。

6. 实验室观测指标的数量 对于同一观测目的所设计的实验室指标的数量要少而精,因为测试指标选择的越多,假阳性的发生概率就越大,有时甚至会影响到疗效测试的真实性。

## 七、干预实施方法及随访观察期的确定

在进行临床治疗试验时,需要详细说明该试验的用药方法和要求与疗程,标明试验药



物的原始资料和包装，建立药物的分发和供应制度；当试验需要设立盲法的时候，则应该建立保证盲法的制度；在实施盲法的过程中可能会有患者出现药物不良反应事件，因此应该建立遇到紧急情况的破盲制度；同时还要建立避免沾染和干扰制度以及保证依从性的制度等。

治疗性研究结果的判定需要有一个明确的观察期，因此需要建立随访观察制度。观察时期不宜过长，也不宜过短。观察期过长，会造成不必要的浪费，过短则可能会使药物或措施致假阴性效果。治疗观察期的选择必须根据研究的目的，并在基础研究的基础上，参考临床达到治疗最佳水平的时间而定。如骨质疏松的防治性研究应考虑到骨代谢的周期较长，少于1年很难得出结果。

## 八、试验结果的整理分析

治疗性研究结果的分析与其他研究一样，需在原始资料完整、准确的基础上，采用正确的统计分析方法分析处理资料。

### （一）资料的整理

对获得的原始记录包括病历、观察表、临床化验及各种功能检查结果要进行详细的核查，然后通过计算机建立数据库并将所有的资料输入到计算机保存以用于统计分析。

### （二）统计方法的选择

1. 不同性质的资料要用不同的统计学方法 治疗性研究资料中最常见的是计数资料与计量资料和等级资料。计数资料主要是试验组与对照组的各种百分率，如有效率、治愈率、病死率等，常用的显著性检测方法为卡方检验（ $\chi^2$ -test）；计量资料是测量所得的记录，如身高、体重、血压、各项血液生化指标的定量测定数值及体液内微量物质或药物测量数值等。计量资料需先求出均数±标准差，然后进行显著性检验，常用的显著性检测方法包括  $t$  检验（小样本）， $u$  检验（大样本）及  $F$  检验（多因素方差分析）、非参数检验；等级资料是将某一指标划分为若干等级，常用的显著性检测方法为 Ridit 分析及非参数检验等。

2. 多组间的比较 如治疗性研究本身有两组以上的结果比较，必须先做多组间的显著性差异分析，只有多组间存在显著性差异时，才能做多组间的两两比较。

3. 配对与非配对的比较 治疗性研究设计中，试验组和对照组的研究对象，有的是配对的，有的是非配对的。由于两种设计的原理不同，因此，分析处理的方法也不同，二者不可混淆。

4. 单侧检验或双侧检验 如肯定试验新药（或措施）疗效比对照老药（或措施）效果好，则用单侧检验法；如不能肯定，则采用双侧检验法。如果对照组采用的是阳性对照药物，则需要进行非劣效性检验，此时用单侧检验即可。

5. 治疗效果的多因素分析 任何治疗效果的产生，除了治疗措施本身的效力之外，还与患者的生理及病理状态以及诸多环境因素有关。例如年龄、营养状态、病情、药量、疗程、并发症等。它们与治疗反应几乎都有关系。为明确治疗措施和其他因素对疾病的影响，应在单因素分析的基础上，选择具有显著意义的有关变量做多因素分析，进一步评价疗效。

6. 意向治疗分析（intention-to-treat analysis, ITT）如果在试验过程中出现了退出和失访者，应该进行 ITT 分析。ITT 分析是将所有纳入随机分配的病人，不管最终是否接受到分配的治疗，在最后资料分析中都被包括在内。该方法可以保证结论更真实可靠。

在临床治疗性研究资料的分析中主要应包括：详细列出主要的和次要的数据分析方



法,详细列出各个亚组的分析方法;在分析中如何处理缺失的数据,如何解释结果的意义;详细比较进入试验组和对照组病人的基线特征;失访、退出和脱落病例的情况;试验结果的有效性,安全性等等。

不要夸大研究结论,尤其是对亚组分析更是如此;不管如何解释,都要如实的报告结果。

### 第三节 影响临床治疗性研究质量的常见因素及处理方法

临床治疗性研究中,往往存在很多研究结果真实性和可靠性的偏倚及机遇,如果在研究中不加以识别和控制,研究的结果将会受到歪曲。

#### 一、机遇

机遇(chance)即随机误差或抽样误差。机遇因素在治疗性研究中不可能消除,只能在研究设计中,通过限制Ⅰ类误差和Ⅱ类误差的容许水平,将机遇因素的影响控制在容许的范围之内,使假阳性及假阴性率减到可容许的最低程度。

#### 二、偏倚

在临床治疗试验中,存在选择性偏倚、信息偏倚和混杂偏倚。

##### (一) 选择性偏倚(selection bias)

选择性偏倚在治疗性试验中的出现,往往是研究者从被研究的目标人群中(如高血压病总体人群),人为的按其所愿,去选择自己感兴趣的研究对象进行治疗性“研究”,或者出现在被选择的研究对象作人为的主观分组,因此,其研究的结果自然不能反映出真实性与代表性,从而使研究结果缺乏临床价值。

避免治疗性试验中的选择性偏倚的方法,是采用真正的随机抽样与随机分组法。

##### (二) 测量偏倚(measurement bias)

测量性偏倚指在资料的观察、测量及收集过程中,在信息的准确性方面受到人为偏倚因素的影响而歪曲了研究的真实性。控制的方法是盲法测试,标准化方法以及测试一致率等。

##### (三) 霍桑效应(Hawthorne effect)

霍桑效应是指在研究过程中,研究者对自己感兴趣的研究对象较对照者往往更为关照和仔细;而被关照的患者对研究人员又极可能报以过分的热情,更多地向医生报告好的结果。这种人为地夸大客观效果的现象,称为“霍桑效应”。控制霍桑效应的最好方法是严格实施盲法。

##### (四) 干扰(co-intervention)

干扰是指试验组或对照组的对象额外地接受了类似试验药物的某种有效制剂,从而人为地夸大了疗效的假象。如果是试验组的对象接受了“干扰”药物,则会造成疗效提高,从而引起试验组与对照组疗效差异的增大;反之,如果是对照组的对象接受了“干扰”药物,则可引起对照组疗效率增高而使组间疗效差异减小。最好的控制方法就是保持实施规范化试验措施及改善依从性,尽可能减少其他药物的使用。

##### (五) 沾染(contamination)

沾染是指对照组患者额外地接受了试验组的药物,人为地夸大了对照组疗效的现象。在试验设计时应该加以限制,在试验过程中加强质量控制,并改善依从性。



### (六) 向均数回归现象 (regression to the mean)

有些测试指标如血压或某些生化指标在初试时有些患者可以在异常水平,然而,在未干预或无效治疗的条件下复试,可能有些回复到正常水平。这种现象表明两次测试值(高或低)都在向着均值的上或下波动,这或许属生理性波动,而非干预的结果,但这种情况可造成误以为治疗有效的假象。克服的办法是可以采取对同一个体的有关测试指标在相同条件下,不同时间内多次测定,取均值以排除其干扰。

### (七) 依从性 (compliance)

依从性是指纳入观察的对象按照研究设计要求执行医嘱的客观反应的程度。全面认真地执行医嘱,按规定的药物剂量和疗程接受治疗,称为依从性好;反则为依从性不好或不依从 (non-compliance)。依从性差的原因包括:简单的遗忘,误解药物使用方法,不能耐受药物的副作用,讨厌服药或费用不足等。治疗时间长(如几个月)或治疗方案复杂对依从性也有较大的影响。

解决依从性问题最主要的方法是使病人充分理解试验目的、要求及参加这项试验的意义,使病人在理解的基础上给予合作。此外,在增强研究人员的责任感,改善服务态度和 方法的基础上,还必须同时加强试验工作的管理,从客观上减少不依从的可能性。如建立检查制度,在复诊时采用清数患者服药量的方法计算患者的依从性(公式 15-4),如服药量 $\geq 80\%$ ,则依从性佳, $< 80\%$ 为依从性不佳。为了保证研究质量,不依从率应力争控制在 10% 范围内。同时还应建立药物、血、尿浓度的检测方法,必要时可进行体液内药物浓度测定来确保病人的依从性。另外,选用有某种疾病的高危人群做研究的对象对提高依从性也很有意义。

$$\text{依从性} = \frac{\text{实际用量 (应服药量 - 复诊时剩余药量)}}{\text{总设计用量 (应服药量)}} \times 100\% \quad \text{公式 15-4}$$

患者的依从性程度对研究的质量会有很大的影响。依从性越高,治疗组和对 照组间的差异越能反映治疗的真正效果。而依从性降低就不可能获得预期效果,使本来应该有的治疗效果消失或降低。所以在治疗性试验中,维持与改善患者的依从性是十分重要的。

## 第四节 临床治疗性试验研究的评价原则

治疗性研究结果的科学评价,可以概括为真实性、重要性和临床实用性三个方面。

### 一、真实性评价

#### (一) 治疗性研究是否为真正的随机对照试验

随机对照试验的设计要求最为严格,结果的真实性优于其他设计方案。这是因为设计严谨的 RCT 可有效地控制已知或未知的偏倚的干扰,确保研究结果真实可靠,故为治疗研究金方案。在评价 RCT 时要考虑以下问题:①是否采用了真正的随机方法?要注意是否交待了具体的随机方法,是否存在随机分配隐匿,是随机还是随意。对于分层随机对照试验,要注意分层因素的数量、试验组与对照组的样本量及分层后各亚组的病例分布是否一致;②是否采用了盲法?要注意是否交待了具体的“盲”法,实施的是单盲、双盲或三盲?阐明试验过程是否按盲法操作等;③组间基线状态的可比性如何?要注意是否描述了组间基线的状态并作了比较,如果存在组间基线状态的不一致性,则要注意是否作了分层比较或校正;④伴随的辅助治疗是否对结果有影响?在某些较为复杂疾病的治疗性研究中,有时是在“基础治疗”的基础上进行试验干预的,或许在试验中某些病例出现新征候需要辅以相关的临时用药。因此,组间的辅助或基础治疗,应注意差异,此外,还要注意



考证有无“干扰”及“沾染”的影响。

## (二) 治疗性研究是否报告了全部研究结果

凡在研究设计中涉及治疗效应—疗效及可能产生的不良反应，均应如实报道。

1. 疗效 应具疗效判断的客观标准，要有相应的测试指标及具体的数据表明疗效的程度，如有效率、其近期/远期的价值等，应具客观真实的全部研究证据。

2. 药物不良反应 研究资料应如实报道任何程度的不良反应内容，缺乏者则视为真实性差，因为任何试验药物不可能没有不良反应，包括安慰剂在内。

## (三) 治疗性试验是否对纳入对象全部完成了所有治疗

真实性佳的临床治疗性试验，应在报告中反映被纳入研究的全部病例，例如各组完成全部试验的病例数、不依从的失访者和试验中因不合格被剔除者的例数，以及发生这类情况的原因均应交代。

为了保证研究质量，对于丢失及不依从的病例应该尽量控制在10%以内。对于试验组和对对照组丢失的病例可分别采用以下方法处理：第一，可以将试验组的丢失病例全部作为“无效”病例计算。第二，将对对照组的丢失病例全部作为“有效”病例计算。这样两组试验有效率的差值将会减小。如果此时再作临床及统计学差异性检验与评价，如仍有显著性意义，则可下肯定的阳性结论，否则，需要进一步研究。

除上述方法外，还可以使用ITT方法进行分析，也可以对试验组和对对照组丢失的病例采用末次观察前推法(LOCF)处理。

## 二、重要性评价

临床治疗性试验的研究成果经过真实性评价被确认真实性好之后，则应分析评价它的临床重要性，即对提高临床治疗水平究竟有多大的意义和价值。其中应注重三个方面的内容：即临床价值，统计学意义和临床经济意义等。

### (一) 临床意义的评价

根据临床试验的对照观察结果，依设计的终点指标，常规者可能有治愈率、有效率、好转率、病死率、复发率、致残率等，以及某些生化或生理功能性计数及计量指标，以反映组间治疗效果及其差异，借以评价试验结果的临床价值。为了更能反映出试验结果的临床意义，临床流行病学研究设计的如下指标现为国际上广泛应用：

1. 绝对危险降低率(absolute risk reduction, ARR) 是试验组事件发生率与对照组发生率之间的绝对差值(公式15-5)。该值越大，说明治疗产生的临床效果越大。

$$ARR = | CER - EER | \quad \text{公式 15-5}$$

式中 CER：对照组的事件发生率，如病死率；EER：试验组的事件发生率。

2. 相对危险降低率(relative risk reduction, RRR) 是绝对危险降低率占对照组事件发生率的比例(公式15-6)。此值的大小表示试验组比对照组治疗后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%或以上才有临床意义。

$$\text{相对危险降低率(RRR)} = | CER - EER | / CER = ARR / CER \quad \text{公式 15-6}$$

式中 CER 与 EER 代表同公式15-5。

3. 需要治疗的人数(number needed to treat, NNT) 即避免发生1例临床事件所需治疗的病人例数。如应用干扰素治疗慢性乙肝患者，为促使1例患者的HBsAg阴转，需要治疗70例数患者，则  $NNT=70$ 。又如，应用β-阻滞剂治疗食管静脉曲张的肝硬化患者预防食管静脉出血，结果预防该病致食管静脉曲张出血1例，需治28例，那么  $NNT=28$ 。

NNT 是治疗性试验效果的良好量化指标，有重要的临床价值及经济价值。





$$NNT=1/ARR$$

公式 15-7

现举一例说明上述指标的计算及其意义。为观察干扰素治疗慢性乙肝的效果,进行了一项随机对照试验。结果发现,在干扰素治疗的 300 例研究对象中,有 105 例 HBeAg 转阴,而常规治疗的 300 例中,仅有 60 例转阴。

1. 试验组事件的发生率 (EER)

$$\begin{aligned} EER &= (\text{事件发生例数} / \text{试验组研究对象例数}) \times 100\% \\ &= 105 / 300 \times 100\% = 35\% \end{aligned}$$

2. 对照组事件的发生率 (CER)

$$\begin{aligned} CER &= (\text{事件发生例数} / \text{对照组研究对象例数}) \times 100\% \\ &= 60 / 300 \times 100\% = 20\% \end{aligned}$$

3. 绝对危险降低率 (ARR)

$$ARR = | CER - EER | = | 20\% - 35\% | = 15\%$$

4. 相对危险降低率 (RRR)

$$\begin{aligned} RRR &= | CER - EER | / CER = ARR / CER \\ &= | 20\% - 35\% | / 20\% \\ &= 15\% / 20\% = 75\% \end{aligned}$$

5. 避免一例临床事件发生需要治疗的病例数 (NNT)

$$NNT = 1 / ARR = 1 / 15\% = 7$$

该结果说明若应用干扰素治疗 7 例此类患者,能使 1 例患者 HBeAg 转阴。

上述指标中的 RRR 是个相对值,在评价临床治疗效果的意义时,如果仅应用 RRR 指标而不用 ARR、NNT 等指标则不够全面,有时会产生误导。因此应该综合应用各项指标进行临床重要性的评价。

## (二) 试验结果的统计学显著性检验

当治疗性试验的结果具有临床意义时,则应作统计学的显著性检验,以帮助判断临床意义差异的真实程度。假定组间的临床差异无意义,在一般的情况下则不必作统计学显著性检验。因为,当样本量增大很多时均会使  $P < 0.05$ 。因此,统计学的显著性差异,并不能对临床意义下结论。

1. 研究结果有临床意义,且样本量较合适,组间统计学显著性检验  $P < 0.05$ ,意味着假阳性错误水平  $< 0.05$ ,同时,如  $\beta$  错误  $\leq 0.1$ ,检验效能  $\geq 90\%$ ,结果自然有临床及统计学意义。但是,如果样本量过小 ( $< 30$  例),即使  $P < 0.05$ ,应防抽样误差的影响,结论应谨慎。

2. 组间有显著临床差异,样本量不够大,统计学差异不显著,应计算  $\beta$  错误水平,如  $> 0.2$ ,应扩大样本再试,防止对有临床价值的结果下假阴性的结论。

3. 组间没有显著的临床意义,样本量特大,即使统计学差异极显著,这种研究结果也没有多大的临床价值。

4. 无论样本量如何,当研究的结果既无临床意义,又无统计学意义,自然下否定的结论。

## (三) 临床经济学评价

试验研究的结果是否能被临床应用,除了它的临床价值之外,还要作经济学分析与评价(详见第十八章)。重点是成本-效果、成本-效用以及成本-效益的评价。最关注的是成本要低、效果要好,病人在经济上可承担,能在解决“看病贵”的方面显出实效。否则,即使是临床价值高,但是价格昂贵(如冠心病安置冠脉支架治疗),其临床意义就有限了。



### 三、实用性评价

一项研究具有了真实性和重要性，并不代表它可以用于自己的病人。因此还要进行可行性评价，看它是否适用于自己的病人和所处的环境。

#### (一) 研究对象的特点是否仔细描述?

纳入治疗对象的临床特点如病情的程度、病程的长短、并发症的有无等，社会及人口学的特点如社会经济水平、性别、年龄、种族等，要尽可能地详述，以便他人对照自己的病人状况及医疗环境，批判性地应用。

#### (二) 治疗方法和措施是否可行?

对于试验治疗的方法或措施，要作详细交代。例如药物治疗除了交代其药物代谢及药物动力学的生物学依据外，对于用药途径、剂量、疗程、药量的增减条件，可能出现的药物不良反应及其对策，以及中止试验的标准，都应清楚写明。

对于某种特殊试验治疗措施，例如冠心病经皮冠状动脉内成形术（PTCA），一定要交代病人的适应证及禁忌证，手术程序与方法，术中和术后注意事项以及某种意外事件的处理等。

只有试验治疗方法与措施交代清楚了，他人才可能根据这些具体的描述判断在自己的实践中是否可行。

研究的科学性（真实性和可靠性）及临床的应用价值，是重要基础。只有具备了真实性和可靠性，同时具有重要意义，方能考虑实用性的问题。

(时景璞 王吉耀)

## 第十六章 疾病预后的研究与评价

### 第一节 疾病预后的概念

#### 一、疾病预后及其研究的意义

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、五年生存率等。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。医生、患者及其家属都迫切需要了解该病的预后情况, 医生知道该病预后情况, 不仅对选择治疗方案有重要意义, 而且可以回答患者及其家属所提出的各种问题。然而, 要对预后作出客观估计与判断, 尽可能使预后的结果接近患者的实际结局, 有时有一定难度, 只有对疾病的预后进行了科学的研究, 掌握了大量的预后信息才能做到科学的预测。

疾病预后研究的意义在于: 了解某种疾病的发展趋势和后果, 从而帮助临床医师作出治疗决策; 研究影响疾病预后的各种因素, 有助于干预并改善疾病的预后。此外, 可以从疾病预后研究中, 来正确评定某项治疗措施的效果, 从而促进治疗水平的提高。例如治疗癌症方案 A 的五年生存率比方案 B 为高, 则说明方案 A 的治疗效果较好。因此, 疾病预后的研究是具有重要的临床意义。

#### 二、疾病的自然病史

疾病的自然史 (natural history) 是指在不给任何治疗或干预措施的情况下, 疾病从发生、发展到结局的整个过程。疾病的自然史包括四个时期:

1. 生物学发病期 (biologic onset) 指病原体或致病因素作用于人体引起有关脏器的生物学反应, 造成复杂的病理生理学改变, 此时很难用一般临床检查手段发现疾病已经存在;

2. 亚临床期 (subclinical stage) 是指病变的脏器损害加重, 出现了临床前期的改变, 患者没有明显症状, 自觉“健康”, 但如采用某些实验室检查或特异性高及灵敏度高的诊断手段, 发现疾病已经存在而被早期诊断, 可获得早期治疗;

3. 临床期 (clinical stage) 指患者病变脏器更加严重而出现解剖上的改变和功能障碍, 临床上出现了症状、体征和实验室检查的异常, 而被临床医师作出诊断, 并进行及时的治疗;

4. 结局 (outcome) 指疾病经历了上述过程, 发展到终末的结局, 如痊愈、伤残或死亡等。

不同疾病, 其自然史差别很大, 某些疾病自然史较短, 如急性感染性疾病, 短期内出现症状体征和实验室异常, 进展较快, 较短时期内即可出现结局。而某些慢性非传染性疾病的自然史较长, 甚至可达数十年之久, 如心脑血管疾病、糖尿病、高血压等, 这些疾病的自然史也比较复杂。

研究疾病的自然病史对病因和预后研究、早期诊断和预防、判断治疗效果都有重要的意义。



### 三、临床病程

临床病程 (clinical course) 是指疾病的临床期, 即首次出现症状和体征, 一直到最后结局所经历的全过程, 其中可经历各种不同医疗干预措施。临床医师可采取医疗干预措施来改变其病程。

病程的概念和疾病自然史不同, 病程可以因受到医疗干预 (包括各种治疗措施) 而发生改变, 从而使预后发生改变。在病程早期就采取积极医疗干预措施, 往往可以改善预后, 在病程晚期进行医疗干预措施的效果就不那么明显, 疾病预后就比较差, 因此, 临床医师十分重视疾病临床病程的估计。

### 四、预后因素

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。预后因素和危险因素不同, 危险因素 (risk factor) 是指作用于健康人, 能增加患病危险性的因素, 而预后因素是在已经患病的病人中研究与疾病结局有关的因素, 因此, 疾病的危险因素和预后因素是不同的概念。虽然有些疾病中某些危险因素也可能同是预后因素, 但多数是不相同的, 例如从图 16-1 可见急性心肌梗死的危险因素与预后因素。我们可以看出, 有些因素是相同的, 且作用相似, 如年龄和吸烟, 随年龄增大, 患病危险性增加, 预后也不好。但有些因素是相反的, 如性别, 男性发生急性心肌梗死危险性比女性增加, 但发生心肌梗死后女性的预后比男性差。又如血压, 高血压是危险因素, 发生急性心肌梗死后低血压则预后不佳。

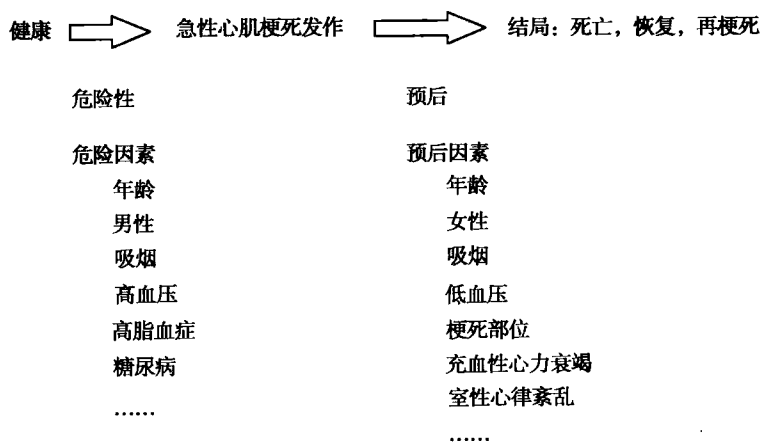


图 16-1 急性心肌梗死的危险因素和预后因素之间的差别

影响疾病预后的因素是复杂多样的, 概括起来有以下几个方面:

#### (一) 早期诊断、及时治疗

任何疾病能否得到早期正确诊断, 及时合理治疗, 是影响预后的重要因素, 尤其是恶性实体瘤, 如能早期及时诊断, 通过手术治疗, 常能获得较好的预后。而发现较晚, 已多处转移, 失去手术根治机会, 则预后很差。如胃癌, 通过胃镜发现的早期胃癌, 如微小胃癌术后五年生存率可达 100%, 原位癌术后 10 年生存率也可达到 80%, 如侵及黏膜下者, 10 年生存率 65%, 如侵及固有肌层者术后 5 年生存率 70%; 通过临床诊断的中晚期胃癌



术后5年生存率仅16.9%。

### (二) 疾病本身的特点

疾病本身的特点包括疾病的性质、病程、临床类型与病变程度等常是影响疾病预后的重要因素。某些自限性疾病如上呼吸道感染,不需要治疗也可自愈,预后良好,同样是病毒感染如艾滋病和重症肝炎,预后就很差;败血症虽然病情很重,但可采用有效抗生素治疗而痊愈,但运动神经元疾病肌萎缩侧束硬化虽发展缓慢,但无有效治疗,预后很差,最终都因呼吸麻痹并发肺部感染死亡;霍奇金病的预后和病理类型有关,淋巴细胞为主型预后最好,5年生存率为94.3%,而淋巴细胞削减型预后最差,5年生存率仅27.4%。

### (三) 患者的病情

通常病情与预后密切相关,病情重者,预后较差。例如,黄疸腹水型重症传染性肝炎的预后远比无腹水的轻或中型肝炎的预后为差。

### (四) 患者身体的素质

患者身体的素质是项综合指标,包括年龄、性别、营养状况、免疫功能等。同一种疾病,由于患者身体素质不同,预后差别可以很大。例如同一种病理类型的非霍奇金淋巴瘤,如患者身体素质较差,年龄大,营养状况差,不能耐受强烈化疗,因而,病情易进展,预后差,生存期短。而身体素质好的患者,经过正规强烈化疗,不仅可长期生存,甚至可以治愈。

### (五) 医疗条件

医疗条件的优劣,直接影响疾病预后。例如败血症可因抗生素选择不合理,疗效差;如果结合细菌培养、药物敏感试验合理选用抗生素,疗效可以提高,预后也就较好。又如急性心肌梗死在医疗条件差的医院,许多疗效好的治疗措施都不能实施,病死率较高;而条件好的医院不仅医疗设施好,患者早期的正确诊断几率高且有抢救经验丰富的专科医师及许多有效治疗措施如溶栓治疗、经皮冠状动脉腔内成形术、冠状动脉支架术、冠状动脉搭桥手术等都可以选择,从而可以降低病死率,改善预后。

### (六) 社会、家庭因素

如医疗制度、社会保险制度、家庭成员之间关系、家庭经济情况、家庭文化教养、患者文化教养及心理因素都会影响患者疾病的预后。

## 第二节 疾病预后评定方法及其指标

### 一、疾病预后的评定指标

1. 病死率(case-fatality rate) 见本书第十一章。

2. 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。

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

3. 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。

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

4. 复发率(recurrence rate) 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。

复发率 = (复发的患者数 / 接受观察的患者总数)  $\times 100\%$

5. 病残率(disability rate) 见本书第二章。



上述预后指标：缓解率、复发率、病残率等常用于长病程低死亡的疾病，许多慢性非传染性疾病都属于此类。这类疾病病情复杂，预后多样，可缓解、复发、好转、恶化、致残、死亡等。

6. 生存率 (survival rate) 从疾病临床过程的某一点开始，一段时间后存活的病例数占总观察例数的百分比。

$$n \text{ 年生存率 (nP0)} = (\text{活满 } n \text{ 年的病例数} / n \text{ 年内观察的总例数}) \times 100\%$$

式中 P 为生存率，前标 n 为随访时间长度，后标 0 为观察起始点。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 (1P0)，一般癌症用 5 年生存率 (5P0) 表示预后。

## 二、生存分析

生存率的计算有两种方法：直接法和间接法（见有关统计学专著）。生存率尚有总生存率 (overall survival, OS) 和无病生存率 (disease free survival, DFS) 之分。后者指患者治疗后进入疾病消失期的生存率。预后研究中还计算生存期的长短，即存活期。存活期的计标，可有平均生存期、中位生存期、无病存活期 (disease free survival, DFS)、无“事”存活期 (event free survival, EFS) 等。生存时间是一种连续变量，但其分布常非正态，因此计标中位生存期比较合理。生存时间数据中有一部分是完全的，即已经观察到某种事件的发生，如死亡，而一部分是不完全的，即经历了一段时间的观察，某种事件尚未发生，但由于种种原因而中止了观察，这种数据称为截尾数据，生存分析的统计特点就是包含了截尾数据。生存分析出上述数据值分析外，尚有图示分析及生存曲线，它是以随访时间为横坐标，生存率为纵坐标作图即为生存曲线，常用的 Kaplan-Meier 曲线，表示时间与生存关系的函数曲线。随访时间单位越小，则精密度越高，即生存期用日计比月计为佳。生存曲线分析能获得有关疾病过程任何时刻的生存率，提供的信息远远超过点估计值，例如几种疾病的 5 年生存率相同，但各种疾病的生存曲线却大不相同。小样本获得的曲线为阶梯形，而大样本就形成光滑的曲线。应用间接法进行生存率分析时必须注意，所绘生存率曲线纵坐标所示是一个假想队列的生存概率，而不是患者实际的生存率，因此曲线左侧的估计值较右侧可靠，因为左侧的观察病例数总是比右侧多。总生存率 (OS)、无病生存率 (DFS)、中位生存期、无病存活期及生存曲线分析常作为恶性肿瘤的预后研究评定指标（有关生存率分析方法，详见医学统计学专著）。

## 第三节 疾病预后研究方法

### 一、疾病预后研究常用设计方案

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、前瞻性队列研究等，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。所选用研究设计方案不同，研究结果可以相差很大，例如泌尿系结石的复发率可由 20%~100%，溃疡性结肠炎癌变的机会可从 3% 至 10%，相差数倍。这是因为研究方法不当造成的偏倚所致（详见第十章）。

疾病预后的评定指标，如描述疾病的病死率、治愈率、缓解率、复发率、致残率、生存率等，可以将研究对象进行长期随访，纵向调查获得，其基本设计方案是纵向的描述性研究。如要进行两组病例预后评定的比较，如两组生存率比较等，其基本设计方案是队列研究。



## 二、疾病预后研究设计中的若干注意事项

### (一) 队列研究的起始点

预后研究采用队列研究设计,其起始点称零点时间(zero time),该起始点在研究设计时必须明确规定,是在病程的哪一点起进行观察,在两个队列中的每一个研究对象都要用同一起始点,进行追踪和观察以及预后结局的比较。如研究对象不采用同一起始点,研究结果就会不正确。对于预后研究,要尽可能选择疾病的早期,如收集的队列其集合时间接近疾病初发时日,则称起始队列(inception cohort),采用起始队列则更好,但起始点必须有明确的标准,不能含糊不清。在队列研究中,不同病人在病程中,开始观察时间不同,则对康复、复发、死亡等时间的描述很难正确,假如零点时间有些病人是采用普查检出计算起,有的是按治疗开始日算起,这种杂乱的零点时间,难以评价真正的预后。

此外,如像研究肿瘤患者的预后,可根据临床病理分期作为起始点,同期或异期间相比,其对预后的评价也是可信的。

### (二) 研究对象的来源和分组

研究对象的来源要具有代表性,能代表目标疾病的人群。同一种疾病来自不同级别医院,其预后研究结果可能不同,如采用来自三级医院病例的结局评估该病目标疾病人群的预后,显然代表性较差,因为三级医院常集中病情较重,病程接近后期的病人,因而预后差。如采用来自某地区各种级别医院中该疾病的病例作为预后研究对象,通常包括了各种型别及其病情严重程度各异的病例,颇能反映出目标人群的特点,因而代表性就比较好。

研究对象的分组原则也必须遵循下述原则,即非研究因素在二组分布应相同,才有可比性。

### (三) 随访和失访

预后研究中随访工作十分重要,随访工作应组织严密,要尽量使所有研究对象都随访到,做到失访率越低越好。如失访率大于开始队列成员的10%应引起注意,如果大于20%则研究结果可能没有参考价值。因为失访的患者会使疾病预后的信息丢失,从而影响预后结果的可靠程度。防止病例失访应注意以下数点:加强对患者及其家属进行随访意义的宣传,以提高随访的依从性;建立健全随访管理制度,随访要有专人负责,并对失访者要及时采取随访措施;做到积极回答患者来信的要求,不失信于患者;以及改进随访信格式与内容,不用使患者及家属反感的措辞,要采用关心、体贴的语言等。

随访期限视疾病病程而定,原则上要有足够长的随访时间,以便能观察到疾病的所有结局。随访问隔时间的确定要合理,以便能观察到各种变化的动态过程。对于一般短病程的疾病,随访问隔时间要短些,病程长的病,随访问隔时间可以长些。随访过程中确定的各种结局,一定要有明确的定义和判断标准,在设计时就规定好,执行中不再变动,标准要有客观性,为了防止测量偏倚,最好用盲法观察。

对失访者的处理可采用两种方法,一是按死亡统计,另一种是从观察患者人数中删除,不予统计,这两种方法均可损失预后的信息。对失访病例不多的研究,可采用经验法则估计预后的范围,具体方法为先假定失访者均出现预定结局,得到结局因素“最高”发生率,然后假定失访者均不出现结局因素,得到结局因素“最低”发生率,比较“最高”和“最低”率,如两者相差不大,则结果可取,如两者相差很大,则研究结果不可靠。

这里宜将“最低”和“最高”值作统计学处理,如 $P < 0.05$ ,95%CI位于两值之间,其结果反而不可信,否则可信,这是因为该差值离真值太大。

## 三、疾病预后因素的研究方法

从前述已知影响疾病预后的因素很多,包括患者一般情况如性别、年龄、体质和营养



情况、社会经济和心理状况等,疾病本身的情况如病理组织学类型、病灶大小、病原体种类、临床分期等,治疗方法及患者、医护人员的依从性等。对疾病预后因素的识别和研究,并予以干预将改善疾病的预后,这是疾病预后研究中另一项重要内容。

预后因素的研究方法和疾病危险因素的研究方法相似。一般可以先从回顾性的临床资料中进行筛选,然后通过病例对照研究,及进一步进行前瞻性队列研究加以论证,从而确定是否为预后因素。分析方法可先从单因素分析开始,然后进行多因素分析。

在单因素研究中确定某因素是否系预后因素时,必须保证观察组(存在某预后因素)和对照组(不存在该预后因素)两组的临床特点和其他非研究预后因素都要相同,但在实际工作中常不易做到。为了尽量减少混杂性偏倚,使预后因素的研究获得比较正确的结论,可以采用下列方法,如限制、配比、分层及标准化等方法加以平衡(参见本章第四节)。由于疾病的结局和多种预后因素有关,各种预后因素可以互相影响,它们对结局的作用大小也不相同,为了全面正确地衡量预后因素的作用,近年来随着统计学方法的发展和计算机的应用,多因素分析方法,如多元回归、逐步回归、Logistic 回归及 Cox 模型等分析方法的应用,可以进一步筛选出与疾病结局有关的主要预后因素,以助建立该疾病预后函数或预后指数。其中尤以 Cox 模型应用最广。1972 年英国统计学家 D R Cox 提出一种能处理多因素生存资料数据的回归模型,称比例风险模型(proportional hazard model),简称 Cox 回归或 Cox 模型(详见有关统计学专著)。它可以允许“终检”(censoring)即截尾数据的存在,终检的原因可能是死于其他原因、失访或到资料总结时随访对象还活着但尚未发生所规定的事件,由于一部分病人“终检”,使得许多常规统计方法都不宜应用。且 Cox 模型还能有效处理随访迟早不一,随访时间长短不一及资料失访等临床预后研究中经常碰到的问题。前述生存率分析统计方法,基本上都属单变量的统计方法,以描述和分析一个因素对生存时间的影响,而 Cox 模型可同时分析众多的因素对生存的影响。在 Cox 模型作分析时,也可估计相对危险度。

## 第四节 预后研究中常见的偏倚及其处理方法

### 一、预后研究中常见的偏倚

#### (一) 集合偏倚(assembly bias)

集合偏倚或称分组偏倚、就诊偏倚。由于各医院的性质和任务不同,各医院收治患者的病情、病程和临床类型可能不同,就诊患者的经济收入在不同地区也可能有所不同。在集合成队列进行随访,随访结束时发现预后的差异是上述因素造成的而不是研究因素所致。这是一种选择性偏倚。

#### (二) 存活队列偏倚(survival cohorts bias)

从各医院收集病例组成队列进行预后研究,由于收集的队列不一定是起始队列(inception cohort),而是可供研究的病例,都是该病程中某一时点进入队列,且都是存活的病例,故称存活队列偏倚,那些未入院失访病例的信息丢失,造成预后判断的不正确。例如采用起始队列进行研究,集合队列 150 例,随访结果时,预后好的 75 例,预后不好的 75 例,各占 50%;如从医院可供研究的病例组成队列,共 50 例进入队列,其中预后好的 40 例,预后不佳 10 例(20%),前者占 80%(40/50),可是未被随访到的 100 例中预后好的 35 例,预后不佳 65 例,占 65%(65/100),可见这里不良预后差异竟是 20%:65%,而真值系 50%。从这里可以发现,同一例子因为形成队列不同,如采用起始队列,则 50%病例预后良好,如采用可供研究的病例集合成队列,结果有 80%病例预后良好。因





此,存活队列偏倚实际上也是集合偏倚的一种特殊类型。

### (三) 失访偏倚 (lost to follow-up)

这在临床预后研究中是经常遇到的。由于观察时间长,观察对象因迁移、外出、不愿继续合作、因药物不良反应而停止治疗或死于非终点疾病等原因脱离了观察,即失访,造成对研究结果的影响,称失访偏倚。例如一项预后研究是对100例病人进行随访观察,但最后有20%病人失访,在观察到的80%病例中,其疗效为80%。如果失访的20例中也同样有80%的疗效,则该研究的实际疗效为80%。若在随访的20例中疗效差的占80%,则实际疗效仅68%。可见失访偏倚可严重影响研究结果的真实性和可靠性。

### (四) 零时不当偏倚

系由于观察对象之间观察的起始时刻不在该疾病病程的同一起始时刻。例如肾结石的复发率,有人仅从患肾结石住院手术的病人中了解既往有无结石史,发现肾结石的复发率为30%。住院病人由于观察的零时刻不同,将初发与复发划入同一组去观察预后,显然过高估计复发率。

### (五) 迁移性偏倚 (migration bias)

见本书第十章。随访观察期间患者退出、失访或从一个队列移至另一个队列等各种变动引起的偏倚。变动的人数过多必然会影响结论的真实性。

### (六) 测量偏倚 (measurement bias): 详见本书第十章。

## 二、偏倚的处理方法

### (一) 随机化方法 (randomization)

从理论上讲,两个队列进行比较,应当除研究的预后因素外,其他因素最好两组均相同,即基线状况要相同,这样才能比较该预后因素在两组中有无差异。随机化方法进行分组是消除选择性偏倚最好的方法,真正的随机化是指每个研究对象都有同等的机会进入观察队列和对照队列,随机化分组使两组可比。

### (二) 限制 (restriction)

在选择研究对象时,限制在具有一定特征的对象中进行观察,以排除其他因素的干扰。例如研究年龄是否系急性心肌梗死的预后因素,如将研究对象限制在黄种人、男性、无并发症的前壁心肌梗死病人中进行观察,这样就可以排除种族、性别、心肌梗死部位和并发症等因素的干扰和影响,就能比较清楚地反映年龄对急性心肌梗死预后的影响。但用这种方法来控制偏倚所获得的预后结论常有很大局限性。

### (三) 配比 (matching)

配比就是为观察组的每一个研究对象匹配一个或几个具有同样因素的对照,然后比较两组的预后因素,配比的方法能消除这些因素对结果的潜在影响。许多研究者常以年龄/性别和种族作为配对条件,因为这些因素常是最常见的混杂因素,许多其他因素也可作为配对条件,如病期疾病严重程度和先前的治疗等。但千万不能把研究因素作为配对条件,否则就不能观察该研究因素在两组中的差异。

### (四) 分层 (stratification)

分层常是最常用的检出和控制偏倚方法之一,特别是有潜在的混杂偏倚时,应用分层方法控制偏倚主要是在临床科研资料的分析阶段。分层是指将科研资料按某些影响因素分成数层(亚组)进行分析,观察研究因素是否在每层内两组间均有差异,以明确该研究因素是否系独立的预后因素。有作者欲研究霍奇金病的预后和初诊时纵隔肿块大小有何关系,所有的研究对象均经过根治性淋巴结放疗,治疗后,无论纵隔肿块的大小,所有患者均进入缓解期。但是发现复发率和纵隔肿块大小有关,纵隔肿块大者的复发率为74%,显



著地高于肿块小者（27%）及无纵隔肿块者（19%）。此结论是否真实，有否混杂因素？众所周知，霍奇金病的预后和临床分期及有无症状有关。因此作者按临床分期及有无症状进行分层分析，从表 16-1 可见纵隔肿块大小和霍奇金病预后关系是独立存在的，并不是由于其他预后因素如临床分期和有无症状的影响而获得的假象。

表 16-1 纵隔肿块大小和霍奇金病预后关系的分层分析

		复发率例数（%）	
		纵隔肿块大组	纵隔肿块小或无组
分期	Ⅱ	10/14 (71)	6/32 (19)
	Ⅲ	4/4 (100)	7/13 (54)
症状	无	10/14 (71)	11/41 (27)
	有	4/4 (100)	2/4 (50)

### （五）标准化（standardization）

当我们比较两个率时，如果两组对象内部构成存在差别足以影响结论时，可用率的标准化加以校正，即使可能影响结果的因素受到同等的加权，则这两个率可比。这种方法称标准化（或校正）。

### （六）多因素分析方法

在临床预后因素研究中常比较复杂，可有多个预后因素相互作用，从而影响结局，应用单因素分析还不能将各预后因素对结局的影响分析清楚，此时应借助于多因素分析方法，多因素分析可以同时处理许多因素，可以从中筛选出与疾病结局有关的主要预后因素，及这些因素在决定预后中的相对比重。在预后因素研究中以比例风险模型-Cox 回归分析方法最为常用。

## 第五节 疾病预后研究的评价原则

对有关疾病预后研究的质量及其研究结论是否真实可靠，应进行评价。评价的原则和标准可归纳为七条，现分述如下。

### 一、观察预后的研究对象是否都处于同一起始队列

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，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。对人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）。

### 二、研究的对象是否能代表被研究疾病的目标人群

对纳入的研究对象应具有明确的诊断标准，纳入标准和排除标准。对研究对象的来源应作详细叙述，以便判断有无选择性偏倚，对进行预后研究的地区、医疗机构也应叙述，以便了解研究对象的代表性，判断选择研究对象时是否存在集中性偏倚（centripetal bias）、倾向性偏倚（popularity bias）、转诊偏倚（referral filter bias）和诊断条件偏倚（diagnostic access bias）。对研究对象的情况包括年龄、性别、疾病严重度和有否并发症存在等都应详细介绍，因为这些因素都和预后有关。



### 三、随访时间是否足够？随访是否完整

由于预后因素常常存在于不良结果时间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。如何来判断随访的完整性，可以从失访率的多少来估计，一般认为失访率 $<5\%$ ，其研究结果可以被接受，如失访率 $>20\%$ 则严重影响结果真实性，因此失访率宜控制在 $10\% \sim 20\%$ 以内，超过 $20\%$ 结果不可信。亦可通过前述的经验法来估计预后的范围，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。

### 四、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。有些预后容易确定如死亡，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，如一过性脑缺血、不稳定性心绞痛，则应采用盲法判断，以避免发生疑诊偏倚（diagnostic-suspicious bias），即研究者竭力去寻找观察组中存在被研究的预后因素的证据，而对待对照组则不然；以及预期偏倚（expectation bias），即凭主观印象判断预后。

### 五、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病人的脑卒中发生率为 $41/1000$ 人年，与非风湿性心脏病心房颤动病人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。但对患者的年龄、性别和高血压状态进行校正后，风湿性心脏病心房颤动病人脑卒中的发生率是非风湿性心脏病心房颤动病人的 $6$ 倍。由于治疗可以改变病人的预后，因此在分析预后因素时，应对治疗上的差别进行校正。校正的方法最简单的是分层分析，如各亚组有不同预后结果，说明具有混杂因素的干扰，如各亚组均获得相同预后结果，说明被研究的预后因素是独立的预后因素。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。

### 六、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。比如报告生存率有三种方法：①某一时点的生存率，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）可以了解预后可随时间而变化。例如所报告的 $1$ 年生存率都是 $20\%$ ，但二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，并且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。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% 可信区间。

## 七、研究结果的实用性和临床意义怎样

上述六条是涉及预后研究结果的真实性和科学性的评价，对预后研究结果的真实性的严格评价外，还须要对预后研究结果的实用性和临床意义作出评价。研究结果是否直接有助于治疗方案的取舍？例如在非风湿性心房颤动病人中应用华法林抗凝治疗，能降低缺血性脑卒中的发生率，但是在“孤立性心房颤动”病人（60 岁以下，无相关的心肺疾病）的研究中，15 年内脑卒中发生率仅为 1.3%，长期应用华法林治疗的危险性很可能超过得益，因此这类病人不必采用华法林抗凝治疗。该研究对确定华法林抗凝治疗的取舍有重要临床价值。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病人及其家属作出解释而使其放心。另一方面，一项质量高的研究结果显示疾病预后不良，就可以与病人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。

(林果为 王吉耀)

## 第十七章 临床决策分析与评价

### 第一节 概 述

决策 (decision making) 是为达到某一目标, 而从多个备选方案中选定最佳方案。在临床决策中, “最佳方案” 有两个层面的含义, 一是效果最佳的方案, 有可能的话应尽量选择该方案; 二是最合适的方案, 是根据患者具体情况而选择的最好方案。

临床决策分析可运用于临床各个方面, 但有必要强调的是: 虽然已奠定了决策的基本理论, 但对解决临床复杂问题仍有捉襟见肘的时候。决策理论应用于临床还面临许多困难: 由于临床医学涉及伦理、法律和经济等方面, 使医生的决策面临更多不确定性和复杂性; 许多临床状态或预后的定量仍很困难; 临床事件发生的概率往往缺乏准确的数据, 只能依靠经验判断。另外, 临床决策还需要和患者及其亲属进行合作, 患者及其家属的文化水平、社会经济状态以及对疾病的认识都影响着医疗决策。

临床医生需不需要决策分析? 事实上, 医生不可能回避“决策”问题, 每天对病人都得作出大量简单或复杂的诊治决策: 患者需马上收住院治疗吗? 是先做费用较高的某项检查还是立即进行有风险的治疗呢? 立即手术还是再观察一下呢? 对于一例肿瘤患者是单纯放疗还是综合治疗? 等等。可以说, 对于较复杂的临床问题, 任何高明的医生, 都不会只想到一个方案, 而是要考虑多个方案。但许多不确定性因素又困扰着决策, 如各方案都有不同的效果、不良反应及成本等。因此, 临床医生需要综合考虑各种影响因素, 总是自觉或不自觉地进行决策分析。有些时候由于病情危急, 必须马上做出决定, 有些时候还可以去查阅资料, 或请教其他专家, 来帮助正确决策。有时一个人可以决策, 有时还要经过集体会诊才能作出相应决策等。

由于决策的问题属于未来事件, 未来就必然包含着许多不确定性 (uncertainty)。医生面临的不确定性主要有:

1. 临床表现的多样性 同样的疾病在不同的人身上有不同的症状体征, 这对于诊断和判断预后带来困难。

2. 诊断方面 随着科学技术的发展, 可选择的诊断方法愈来愈多, 但每一种方法可出现不同的假阳性、假阴性、危险性和费用。

3. 治疗方案 不同的治疗方法有不同的效果、不良反应、并发症和后遗症甚至可造成伤残或死亡的严重不良后果。由于个体、环境、经济等因素的差异, 同一疾病用同样治疗方法也会出现不同的结果。

4. 临床资料 在分析过程中, 采用以往资料获得的概率等数据, 有时不能正确反映现在的实际情况。另外, 病人的回忆偏性、医生的疏忽大意、不准确的数据、不客观的研究结果, 都可能使医生作出错误的决策。

当然, 简单的临床问题依靠一个人或几个人的学识、技能即可解决, 而面临非常复杂的临床问题, 即使很有经验的医生也有困惑的时候。如果能运用已有的资料信息, 利用决策分析的技术, 结合患者的具体情况就很可能作出正确的决策。

正确决策的前提是掌握充分的、有价值的信息和准确的预测。通过大宗病例分析、临床研究可获取患者临床表现、诊断、预后等各个层次的资料。目前把决策分析应用于临床的主要困难是缺乏完备的临床信息。预测是对事件的发展趋势和未来状况作出正确估计,



预测是为决策服务,没有正确的预测就不可能有正确的决策。预测既要事件的结局作出定性分析,更要对其发生的时间、程度和概率进行定量估计。同一事件中,不同的决策者可能会得到不同的信息和作出不同的预测,其最后的决策也就不同,在临床的查房、会诊中,专家教授们的意见常常相左也就不足为奇。

## 第二节 临床决策的类型

从不同的角度可把决策分为不同的类型:如按决策的性质而分为程序化决策和非程序化决策;按决策目标的多少而分为单目标决策和多目标决策;按决策级别而分单级决策和多级决策;按决策范围而分为团体决策和个人决策。临床则多按决策的可靠程度分为:确定型决策、风险型决策和不确定型决策。

1. 确定型决策(decision under certainty) 指事件的结局已经完全确定,通过分析各方案的最后得失,从中选择一最佳方案,如成本分析法。

2. 风险型决策(decision under risk) 指事件的结局不能肯定是否发生,但其概率可以估计时的决策。它有三个基本条件:①存在两种以上的备选方案;②可估计各事件发生的概率;③各结局可定量估计。常用的方法有:期望值决策、贝叶斯(Bayes)决策等。

3. 不确定型决策(decision under uncertainty) 当各事件的结局不能肯定,且发生的概率也无法预测时的决策。这种决策的正确性很大程度上取决于决策者的经验与主观判断。有一些准则,如最大最小准则、赫维茨准则等可供决策者参考。

应该指出的是,决策是一个过程,而不是一种瞬间的活动,因此在作出决策后,还要对所作出的决策进行追踪,这在整个决策过程中占有十分重要的地位,因为决策过程中主客观情况不是一成不变的,要在实施中通过的不断信息反馈、方案调整,修正错误,使决策更加完善。

## 第三节 决策的步骤与原则

决策的问题千变万化,决策的方法多种多样,决策者多根据自己的经验和主观判断,可从不同的地方入手来进行决策分析,不要求统一的步骤或程序,但一般说来,决策分析的过程可分为以下四个步骤:

### (一) 解剖问题

对面临的决策问题进行解剖,使之分解为简单成分,逐一分析。一般包括四方面内容。

1. 备选方案 为了不遗漏最佳方案,在分析开始,就要考虑达到既定目标的所有方案,除非能证明某方案肯定不如另一方案;

2. 结局 上述备选方案干预后,患者可能会出现各种结局,如诊断试验的结局:阳性、阴性等;治疗的结局:痊愈、改善、恶化、死亡等。

3. 概率估计 对各种结局进行概率估计。

4. 结局的得失估计 对各种结局的损失与收益进行量化,可以用生存率、生存质量、成本等指标进行分析。

5. 其他考虑 有些因素对于决策者(包括病人)作出决策的时候是不能忽视的,如涉及社会效益、伦理道德等。

### (二) 组合问题

通过对问题的解剖,明确构成决策问题的各种成分后,进行有机地组合,可利用一些图表把问题的成分按时间顺序罗列出来,使决策问题和分析过程一目了然,并在图表中把



关键信息标明，有利于分析、讨论。常用的方法是决策树和决策表。

### (三) 分析问题

问题“摆”出来之后，就要进行分析。可选用一些分析技术来确定最佳方案。如计算期望生存率、期望效用或期望成本。对于一些概率估计和结局的效用值会有不同的意见，还要进行敏感性（sensitivity）分析和阈值（threshold）分析。

### (四) 解决问题

实施“最佳”方案过程中，还必须观察问题是否得到满意的解决。

## 第四节 决策树与期望值

决策树（decision tree）是按逻辑、时序把决策问题中的备选方案及结局有机地组合并用图标罗列出来，它如一棵从左到右不断分枝的树，包括一些节点与分枝。节点有两种：一种是决策节点（decision nodes），用小方格表示，此节点所引出的分枝叫方案分枝，为表明不同的方案，可在分枝上注明方案名称；另一种是机遇节点（chance nodes）或称事件点（event points），用小圆圈表示，由此节点引发的事件或结局不能由决策者随意控制，所引出的分枝叫概率分枝，其上标明事件或结局出现的概率。在决策树末梢，为各方案的最后结局（outcome），各种结局必须定量描述。决策树仅仅预测了各方案的中间及最后结局，并标明了必要的信息，要从中选出最佳方案，应计算、比较各方案的期望值（expected value 简称 EV）。各机遇节点的期望值为此节点各分枝概率（ $P$ ）与结局得失值（gain-loss value）乘积的总和。

下面通过两个例子说明风险型决策过程：

例 1：男性患者，50 岁，因发热、呼吸道症状疑为肺孢子虫病（pneumocystis）收入院。此病对免疫功能低下者威胁很大，可产生严重肺炎，出现发热、肺广泛浸润、低氧血症。在有类似临床表现的病人中，肺孢子虫病约占 1/5。现在问题是是否要做活检（open lung biopsy）以明确诊断？假如活检是唯一能正确诊断的方法（为简化计算设敏感性、特异性为 100%），但严重病人接受活检有 4% 的死亡率。根据文献资料肺孢子虫病不治疗，半年生存率只有 10%，如用 pentamidine 治疗，半年生存率达 50%；非肺孢子虫病人，半年生存率为 65%；如用 pentamidine 治疗，由于其毒副作用，生存率降为 63%。

为获得最大的生存率，是否要活检？如不活检，是否采用药物治疗？这时实际面临三个方案：活检；不活检，全部服药；不活检，均不服药。

如果活检，有两种结果：死亡与生存，而生存者又有两种情况，一是确诊的肺孢子虫病人，这些病人将服药，其生存率约为 50%；二是其他非肺孢子虫病人，将不服药，生存率为 65%。下面画出活检方案的决策树（图 17-1 至图 17-3）：

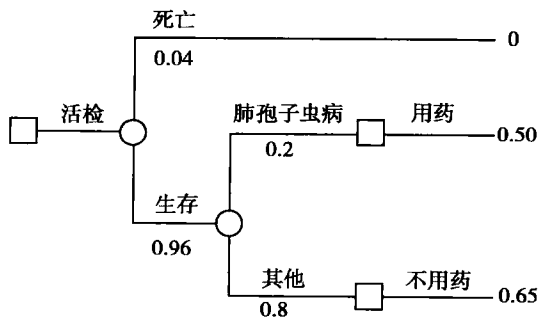


图 17-1 活检方案的决策树



如果不活检,就不能把肺孢子虫病患者与其他病人分开,只能要么全部服药,要么全部不服药。分析资料,不难画出如下决策树:

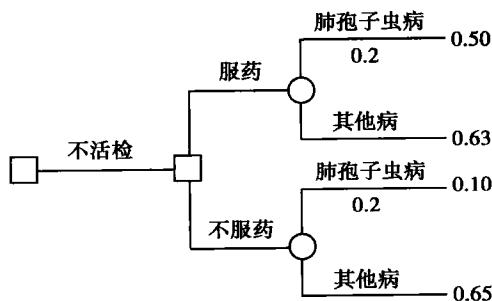


图 17-2 不活检的决策树

把两棵决策树合并则有:

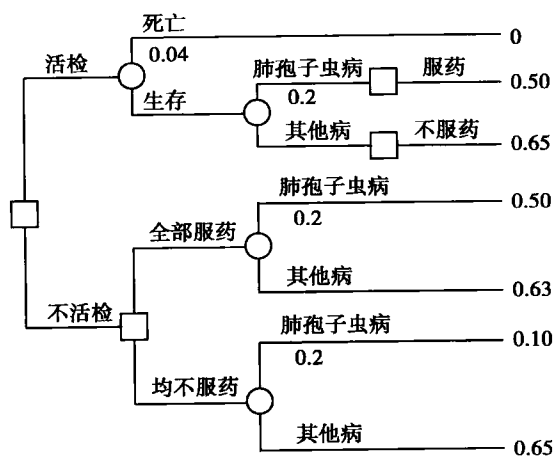


图 17-3 肺孢子虫病诊治的决策树

根据决策树中各枝的概率和生存率,可计算各方案的期望生存率。如生存节点的期望生存率为:  $(0.2 \times 0.50) + (0.8 \times 0.65) = 0.62$ , 以此数值作为生存率,可计算:

活检的期望生存率 =  $(0.04 \times 0) + (0.96 \times 0.62) = 0.595$

同样可计算:不活检、服药期望生存率 =  $(0.2 \times 0.50) + (0.8 \times 0.63) = 0.604$

不活检、不服药期望生存率 =  $(0.2 \times 0.10) + (0.8 \times 0.65) = 0.540$

比较三方案的期望生存率,不活检,服药的方案为最佳方案。

许多时候,决策者们对某项概率或生存率的估计会有不同的意见,例如有医生认为肺孢子虫病人服药后生存率可达 60%。解决的方法是进行敏感性分析,可把 60%代替 50%,重新计算各方案的期望生存率。利用计算机还可同时进行多项因素的敏感性分析。经分析后,如果决策的结果不变,说明对这些数值的改变不敏感,结论就比较可靠,否则,下结论时应慎重。对于有争议的概率或效用值都应进行敏感性分析,其上下限可根据临床经验或有关资料而定。

对敏感结果,还可进行阈值分析,如有医生想了解:活检死亡率降低到多少时,两方案的期望值相等?设活检死亡率为  $P$ ,那么活检生存率为  $(1-P)$ ,得方程

$$P \times 0 + (1-P) \times 0.62 = 0.604$$

$$P = 0.026$$





就是说,如果活检的死亡率能低于 2.6%,选择“活检”,否则,应选“不活检、服药”的方案为好。这是由于活检有 4% 的死亡的危险性,它的临床信息净期望值 (net expected value of clinical information) 为:  $0.62 - 0.604 = 0.016$ ,意思是,如果活检的危险性为 0,其平均生存率较不活检服药的平均生存率提高 1.6%,所以应选择活检。

例 2:某地区卫生部门拟对 65 岁以上人群进行流感预防接种,有资料显示:不接种流感发病率达 30%,平均每例治疗费用 \$50;流感所致死亡为 0.1%,每例死亡损失估计约 \$100000;如接种,每例接种费 \$3,可降低发病率 50%,接种不良反应需治疗者占 0.1%,其每例治疗费 \$300。

问题:从成本效益考虑,是否应该推行预防接种计划?

根据以上提供的资料,不难绘出以下决策树(图 17-4),并可标明相关事件的概率和成本,然后分析各节点的期望成本。

不接种的期望成本 =  $(\$100000 \times 0.001 + \$50) \times 0.3 = \$45$

接种的期望成本 =  $(\$100000 \times 0.001 + \$50) \times 0.15 \times 0.999 + \$300 \times 0.001 = \$22.8$ ,再增加每例接种费 \$3,共 \$25.8

按成本最小原则,应选择接种。

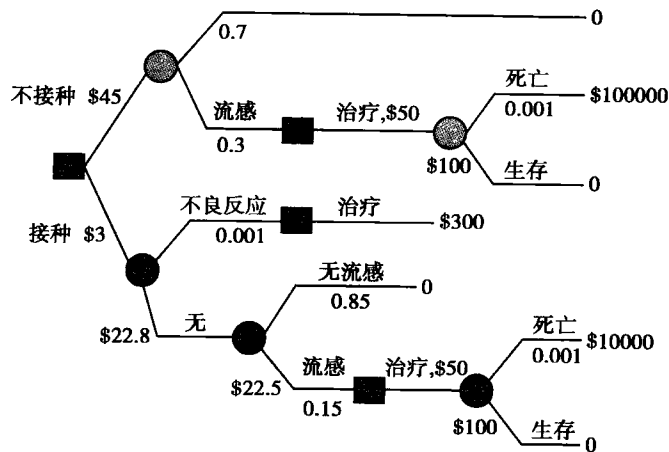


图 17-4 流感预防接种的决策树

## 第五节 概率估计

从上述的例子可看出,只要有足够的信息,临床决策分析并不困难。决策分析的是要素之一,也是应用于临床的难点之一是各种结局的概率估计。概率的估计有两种:一是客观估计 (objective estimates of probabilities),二是主观估计 (subjective estimates of probabilities)。

概率的客观估计可根据临床研究的结果,文献结果报告或其他资料,来自于大量病例的临床试验的概率最为可靠。从文献或记录中也可直接估计概率。例如,有文献报告文 1000 个病人接受某种手术治疗,手术中 23 人死亡,那么  $P(\text{死亡}) = 23/1000$  或 0.023。当然,把来自一个研究中估计的概率,用于另一个决策问题,应注意:

### (一) 处理背景不同

设备、仪器、诊断过程不同,会得出不同的概率。例如,即使同样的病人、药物,在初始用药时,治愈率较高,随着耐药性出现,治愈率可能下降;也可能随着药物质量改



善，治愈率提高。

## (二) 病人情况存在差异

如性别、年龄、病情等。因此，不必照搬其他研究中的概率，针对你的病人的实际情况，并注意引用文献中有否分层分析的结果与自己的病人情况相似，然后根据你的经验、知识或专家的意见，参考其他研究资料，修改的概率可能更符合实际要求。

有时在临床决策分析中，有一两个结局的概率得不到客观的估计，只能依个人的经验和判断作定量估计。它不是研究人群中观察到的概率，只是取决于决策者对某事件将发生的信念程度 (strength of belief)。这样估计的概率称为主观概率。在具体应用时，可先根据经验估计一概率范围，进行敏感性分析。如不敏感，结果可靠；如敏感，可作阈值分析，找出阈值，同时收集更多更准的信息，或进行针对性的专项研究。

对于主观概率的真实性，是存在争议的。心理学家发现，主观概率常带有偏性。例如，有一患者的临床表现与两种疾病的特征有相同程度的类似，有些医生认为该患者患两种病的可能性是一样的，这就把两种疾病有不同的患病率给忽视了。运用经验时，回忆还受概率以外的因素影响，近期的、有意思的、特殊的或严重的疾病更容易记忆，这样可能过高地估计其概率。另外，人们在估计概率时，喜欢从一初始值开始，然后再作调整，问题是这些调整往往是不够的。例如，把初始概率放在 0.5 附近，其调整后，罕见病的概率仍可能估计偏高；有些人喜欢趋于 0 或 1 的极端值，可能低估罕见病的概率而高估常见病的概率。有人认为，硬要清晰地估计未知事件的概率，可能影响医生的最佳判断；而持不同意见的人则认为，寻求量化的过程中，将促使临床医生深入观察研究，发现未知事件的本质与规律，最终对正确决策有利。一些研究表明主观概率在临床上能提供有用的信息。

在临床不得不依靠主观概率时（临床一些决策不能等到找出客观概率后才作出），有些减少偏性的方法可参考：

### 1. 改善个人概率估计的准确性

(1) 用定性语言描述各不确定事件，并与其他相似事件联系起来；

(2) 粗略地按可能性大小排列各不确定事件，如  $A > B > C$ ；

(3) 相互权衡各事件发生的可能性或可能性之比，然后作出定量估计，例如在三种事件中，A，B 事件的出现是 C 事件的 5 倍，那么 A 事件的概率为 5/11。每一事件还可与一些已知概率的事件比较，以帮助确定概率。

2. 采纳一组专家的概率估计，特别是主观概率敏感性分析的结果敏感时。具体做法是，在一组专家中，要求每位对一事件估计其概率，不记名写在纸条上，然后把第一轮的结果整理、统计，用匿名的方式反馈给专家们，如 5 位给的值在 0.20~0.25，10 位在 0.26~0.30 等，之后再进行第二轮的概率估计和结果反馈，连续几个轮回直到出现某些水平的一致。这种方法称为得尔斐 (Delphi) 法，原理是意见的趋同现象，各种意见在一起相互影响将导向一致，匿名也可消除心理因素的影响，缺点是意见的趋向中仍会有偏性，也不能保证意见都一致或是准确的一致。尽管得尔斐法不能取代设计良好的临床研究，但某些情况下也还有一定的参考价值。

## 第六节 效用分析

决策过程中结局的量化是决策分析的重要一环。前面介绍的例子是用生存来定量的，但临床的结局是错综复杂的，不妨先把各种结局分一下类。



### (一) 两项结局

这是最简单的情形，如生存与死亡，有效与无效等，可不用数量比，只计算备选方案的期望生存率或有效生存率即可得出结论，也可把“生存”或“有效”赋值为1或者100，“死亡”或“无效”则为0。

### (二) 单一属性的多项结局

这些结局可按程度分为几个等级。在最好与最差的结局之间是一些中间状态，如一些疗效的评价常用：完全缓解、部分缓解、稳定、恶化，如果完全缓解被定值为1，恶化为0时，部分缓解与稳定的量值是多少？还有一种结局，如用生存率表示，似乎不存在量的问题，决策分析中只要最大期望生存率的方案，但患者可能更喜欢肯定9年生存的方案，而不选择50%机会生存20年和50%机会立即死亡的方案，尽管后者的期望值均为10年。

### (三) 多种属性的多种结局

最为复杂，例如一方面生存时间的不同，另一方面生存质量也各异。问题的关键是怎样用同一尺度把这些结局量化出来，既能反映医生的判断和病人的意愿，又能进行各种决策分析。

面对各种各样的结局，决策者必须对每一结局进行量化，称为效用估计 (utility assessment)。

下面简要介绍效用分析的方法：

#### 1. 单一属性多项结局的效用分析

例：一位68岁的男性患者，患有数年的周围血管病，现左脚刺伤，出现感染和可能的坏疽，病人面临两种选择，一是及时截肢，保住性命，但失去左脚；二是保守治疗，但有感染和坏疽扩散的危险，也有两种可能的结局：痊愈或死亡。对于这些程度不同的结局，如何给予每个结局一相对值，才能真实反映结局的严重程度或患者的意愿？

(1) 直接度量法 (direct scaling method)：先设计有量度的直尺或“梯子”，最好的结局放在一端点上 (数值为1)，另一端点则表示最差的结局 (数值为0)，其他结局处于两者之间。请病人在直尺上标上失去左脚的位置，如0.9。许多患者对此法会感到困惑，觉得很难确定具体数字，因为说不出道理，也不好解释结果。

(2) 博弈法 (standard gamble)：上例中患者的决策问题也可用决策树表示 (图17-5)。如果告诉患者，如保守治疗，痊愈的可能性  $P=0.9$  时，患者选择保守治疗；如改变  $P$  值，当  $P=0.6$  时患者改选截肢，不断改变  $P$  值询问患者，当0.8时，患者觉得两方案差不多，此时  $P$  值称为均衡概率 (breakeven probability)，其数值就是A的效用值，即  $U(A)=0.8$ ，意思是，如痊愈视为1，把死亡看成0，失去左脚这种结局，患者认为它相当于0.8。

如果最差的状态不是死亡，而是一种较A更差的结局 (B)，如图17-6所示：

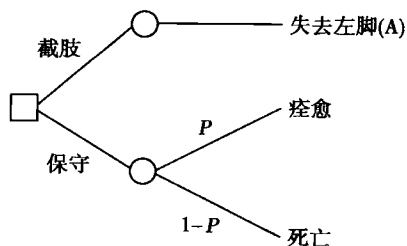


图17-5 当保守治疗可能面临死亡时的决策树

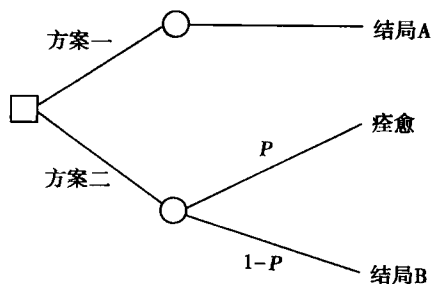


图17-6 面临不同结局时的决策树



同样用上述的方法帮助患者确定 A 的效用值:

$$U(A) = P + (1-P) \times B$$

即使结局为生存时间,也不能简单地认为每个患者都只看期望值,例如面临如下决策问题(图 17-7):

如根据期望值准则,将选择方案二,但实际上可能很多患者宁愿方案一。假设死亡的效用值为 0,生存 20 年为 1,那么生存 9 年的效用值是多少呢?这里也可以采用博弈法来帮助确定效用值。

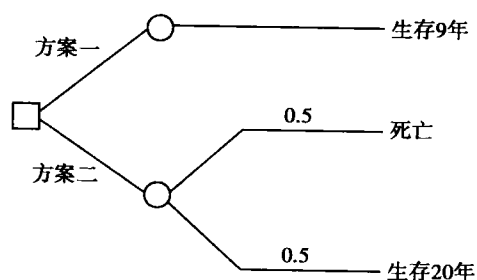


图 17-7 期望生存值不同时的决策树

2. 多种属性多项结果的效用分析 对于病人的诊断、治疗,特别是对于慢性疾病一方面要尽量延长生存时间,另一方面要考虑提高生存质量。但常常难以两全其美,如恶性肿瘤的

治疗,手术、放疗和化疗等可能延长生存时间,但严重的并发症后遗症又使生存质量下降,不同的人对生存时间与生存质量会有不同的认识与期望。对此应与患者一起进行效用分析。

(1) 时间权衡法 (time trade-off): 某种结局 (A) 被认为好过死亡,故患者面临两种选择:

1) 结局 A 生存  $t$  年后死亡;

2) 健康状态生存  $X$  ( $< t$ ) 年后死亡。图 17-8 中,横坐标表示时间,纵坐标表示健康状态:

不断移动  $X$ ,直到患者认为两种选择无差异时,此时结局 A 的效用值:

$$U(A) = X/t。$$

3) 如果患者最后不是死亡,两种结局维持一段时间后恢复健康,如像两种治疗方案都可以达到康复的目的,不过 A 方案的不良反应对健康影响较 B 方案小,但疗程长;后者不良反应严重一些,对健康影响也大一些,但能较快显效,对此患者两种选择是:

A. 暂时结局 A 延续  $t$  单位时间后恢复健康;

B. 暂时结局 B 延续  $X$  单位时间后恢复健康 (图 17-9)。同样变动  $X$ ,直到两种选择被认为一样时,结局 A 的效用值:

$$U(A) = 1 - (1-B) X/t。$$

(2) 质量调整生存年 (quality-adjusted life years, QALYs): 有时影响生存质量的因素很多,如因为患病,不但躯体功能、精神心理状态受影响,而且正常的工作、爱好、娱乐、交际、睡眠、家庭生活、饮食等也受影响。所以,首先应对患者的生存质量 (quality of life) 进行测量 (详见本教材生存质量章),然后把它与生存时间结合起来,生存质量一

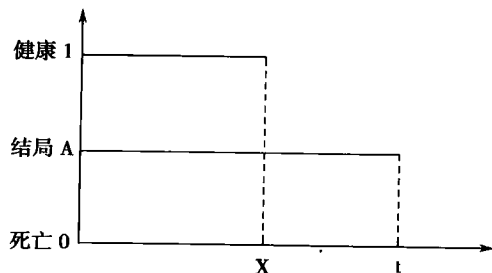


图 17-8 终点为死亡时的权衡

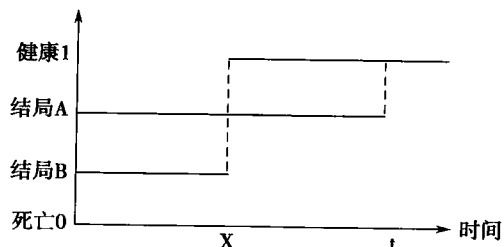


图 17-9 终点为健康时的权衡



般用 0 至 1 之间的数字表示, 如果严重心肌梗死为 0.7, 轻度者为 0.9, 一个患者最近 20 年中, 其中 15 年健康, 3 年有轻度心肌梗死发作, 2 年有严重心肌梗死, 那么 QALYs 为:

$$(15 \times 1.0) + (3 \times 0.9) + (2 \times 0.7) = 19.1$$

意思是此患者 20 年的寿命相当于生存了 19.1 个完全健康年。用 QALYs 来测量各种各样的结局, 使之有通约性, 一些复杂的临床决策问题得以简化。但不同的病人 (即使同一病人) 在不同时间、心境, 对生存时间与质量会有不同的认识。例如上面提到带病生活 20 年和没有任何疾病生活 19.1 年从 QALYs 的数值上看是一样的, 但很多人会认为是不同的, 会作出不同的选择, 这里不必太强调效用值的外部真实性。如果不用 QALYs, 可在决策分析时分别考虑生存时间与生存质量, 或摆出生存时间与生存质量, 由患者自行选择。

上述决策的效用分析, 关键在于医生要提供决策的科学依据以及患者对备选方案的选择, 因此医患的平等合作是十分重要的。

关于临床决策经济分析, 则可参考本教材第十八章的相关内容。

## 第七节 不确定型决策

此类决策是指结局及其发生概率均不确定的条件下的决策。现以一实例来说明解决此类决策的一些准则方法。

例: 某患者诊断为舌前 2/3 鳞状细胞癌, 经临床检查未发现颈部淋巴结转移或远处转移。临床对于有颈部淋巴结转移的舌癌, 常规进行原发癌切除术和颈部淋巴结清扫术; 对于未发现颈部淋巴结转移的舌癌患者, 要不要进行颈部淋巴结联合清扫术, 迄今仍有不同的意见。

该患者的主要不确定因素在于: 颈部是否存在微小的转移灶不明确, 处理时, 是否要行“预防性颈部淋巴结清扫术”? 据文献报道, 舌癌发生颈部淋巴结转移较多见, 转移率 60%~80%, 并有约 40% 在初诊时即已发生了颈部淋巴结转移; 同时临床未发现转移灶的患者中, 有一部分患者颈部淋巴结存在微小转移灶, 故单独治疗原发癌而不做颈部淋巴结清扫术是不够的。可是, 另一方面, 由于颈部淋巴结清扫术存在手术死亡的风险, 并有一部分患者因手术而出现并发症。如给确实未发生转移的舌癌患者进行颈部清扫术则可造成不必要的损失。因此医生面临着是否手术的决策困惑。

对于癌症治疗的评价, 主要是使患者的损失尽量减少, 其结局主要考虑“效用损失 (utility loss)”。表 17-1 中为四种结局的效用损失定量估计: 无颈部淋巴结转移也未做颈部清扫的效用损失为 0; 有颈部转移而未作颈部清扫的为 -100; 其余介于两者之间, 本例不论是否转移, 凡清扫者均初步定为 -40。效用的确定, 一般应考虑: ①手术致死率 (operation mortality); ②手术致病率 (the morbidity of the operation); ③手术费用; ④患者与家属由于住院、手术毁容等造成的情感创伤 (emotional trauma)。由于不同的年龄、条件以及病人与家属的态度其效用损失的估计也各异。

表 17-1 决策矩阵

	颈部清扫	不清扫 ( $S_2$ )
有颈部转移	-40	-100
无颈部转移	-40	0



由于不知道颈部转移出现的概率，所以就不能利用前面介绍的期望值原则来帮助决策，即属于不确定型决策问题。下面几种准则方法可供参考：

### (一) 最大最小准则 (max min criterion)

在每一备选方案中，找出各结局下的最小收益值，从中选择收益值最大的方案，也就是说，从最小收益中求最大。本例颈部清扫的最小收益值为-40，不清扫的为-100，那么，两者比较，收益最大的方案是颈部清扫，因此，将选择颈部清扫。也即在最糟的情况下，收益也不至于到-100，但也不期望最高收益0。这是一种相对悲观的准则。

### (二) 最大最大准则 (max max criterion)

它常被称为乐观准则。与悲观者相反，它将选择收益值最大的方案。本例颈部清扫效损失-40，不清扫为0，两者比较，收益最大的方案是不清扫。应用这一原则，将不做清扫术，它不管收益值会有-100的可能。

### (三) 拉普拉斯准则 (Laplace criterion)

它假设各结局的概率是相等的，即一半患者颈部有转移灶。该准则也称为等可能性准则。因此：

$$EV(S_1) = (-40 \times 0.5) + (-40 \times 0.5) = -40$$

$$EV(S_2) = (-100 \times 0.5) + (0 \times 0.5) = -50$$

$$EV(S_1) > EV(S_2), \text{ 据此应选择颈部清扫。}$$

### (四) 赫维茨准则 (Hurwicz criterion)

采用一个所谓乐观系数  $a$  (介于0至1之间)，用  $a$  乘以各方案中这最大收益，用  $1-a$  乘以各策略中最小收益，将其乘积相加，选择和为最大的方案，反映了决策者对于各种结果的概率的主观判断，也被称为乐观系数准则。

本例如确定  $a$  为0.65，则

$$EV(S_1) = (-40 \times 0.65) + (-40 \times 0.35) = -40$$

$$EV(S_2) = (-100 \times 0.35) + (0 \times 0.65) = -35$$

$$EV(S_2) > EV(S_1), \text{ 结果决定不做清扫术。}$$

由于得不到充分准确的信息，利用上述准则决策时，往往带有很大的主观性。因此，应用不同的准则得出不同的结果，也不足奇。胆大者喜欢用乐观准则；怕事者宁愿用悲观准则；而折中者善于用 Laplace 准则。当然，最好是去收集信息，即应尽量获得各种结局发生的概率，使不确定型决策转变为风险型决策。

决策理论应用的范围是非常广泛的，方法也是多种多样。决策理论应用于临床，还有许多地方有待于完善。临床的一些决策问题是相当复杂的，经常是几个问题交织在一起，非决策树所能描述清楚；病情也往往变化无常，加上个体差异，根据过去的经验、资料分析，不一定能说明新的问题；临床信息的获取也还不容易；效用分析、概率估计的方法也有待改进；正确的决策要求决策者应具备医学、概率、决策和逻辑多方面的知识。无论如何，临床决策分析的重要意义在于培养医生的决策分析思维 (decision analytical thinking)。它通过把复杂的问题分解成简单的成分，在时间上、逻辑上重新组合，而达到全面看问题的目的。通过寻找概率、效用值而进行定量分析，提高临床资料的利用率，也促使医生与病人的合作，一起决策，有利于进行敏感性分析，便于讨论不同的意见，活跃学术争鸣的空气，使临床决策民主化、科学化。

(洪明晃 华贻军)

## 第十八章 临床经济学分析与评价

### 第一节 概 述

#### 一、临床经济学的概念

卫生经济学 (health economic) 是一门研究卫生保健中的经济规律及其应用的学科。它是运用经济学的理论、概念和方法阐明和解决卫生 and 卫生服务过程中出现的现象和问题的一门科学。它研究卫生资源的筹措、配置和利用, 研究卫生服务的需求、定价与供给中的经济学问题及卫生经济的政策与策略, 是经济学在卫生保健领域中的应用。

临床经济学 (clinical economics) 是卫生经济学的一个分支, 它在卫生经济学理论指导下, 用技术经济学的评价方法对临床使用的药物、诊治方案、仪器设备等技术干预措施进行经济评价和分析, 其目的是探讨最佳的诊断、治疗和预防方案, 评价医疗效果, 以提高卫生资源的配置和利用效率, 同时为临床人员和政策决策者提供决策信息。

#### 二、临床经济评价的必要性

##### (一) 医疗卫生服务费用的不断攀升

由于医疗卫生服务需求的增加造成了医疗卫生服务费用的增加。而需求的增加是源于以下四个因素: ①人口老龄化: 由于健康服务对象中老年人口所占比例增加, 不仅会增加医疗服务和社会福利的需求, 也会促进医疗高新技术的发展和应用。例如心脏冠脉搭桥和支架的广泛应用。②新技术和新知识的不断涌现: 医疗技术向高科技发展, 新药不断出现, 高级仪器设备投入使用, 病人和医务工作者会对医疗卫生服务提出新的更高的要求, 新的有效防治方案的出现势必会增加医疗卫生服务需求。当病人和医生得知一种新的有效的防治方法时, 例如肝移植治疗终末期肝硬化, 他们会要求该服务, 从而增加了医疗费用。更重要的是, 有的病人根本不适合做肝移植如晚期肝癌, 也应病人或医生的要求做了, 不仅浪费了钱, 还缩短了生命。③病人对医疗卫生服务期望的增高: 即使没有疾病, 当人们希望提高自身健康水平时, 也会寻求医生的指导。患者从被动接受者向主动的消费者转变。④医务人员对服务质量期望值增高: 医生为了赶上医学进步而采用不断涌现的新技术和新药, 以期提高服务质量。有时为了避免不必要的纠纷, 医生会为了保护自己采取一些并不十分必要的措施, 过度诊断和过度医疗均造成医疗卫生服务需求的增加。

##### (二) 卫生资源的有限性

全社会面临的一个根本问题是资源的匮乏。匮乏意味着没有或不会再有足够的资源来满足全人类的需求。因此, 就会在可供使用的资源中出现不断的冲突, 在可选择的分配中不断地需要选择。政府或社会每年用于医疗卫生领域的投入在发达国家占国民总产值的10%左右, 发展中国家则更少。上述的各种原因使各国卫生总费用不断上升, 导致医疗保健的预算远远跟不上医疗费用上涨的速度, 出现医疗卫生的经济需要与提供资源之间的矛盾。当有限的卫生资源无法满足人们对健康的需求时, 社会对临床医生的临床实践便提出了更高的要求。

##### (三) 卫生资源的最优使用

如何使有限的卫生资源得到最大限度的利用成为临床医生每天面临的课题。由于资源



短缺，用于某一项目就不能用于另一项目，有限的经济资源的投入必须考虑投入后产生的效果。如果一项昂贵的治疗只对一些人有益而我们支付这项治疗，那我们将不能支付其他可以帮助更多人的相当或更有效的治疗措施。一共只有这么多钱，不可能一分钱掰成两半花。如果我们资助只有相对适应证的患者进行骨髓移植，那我们就不能支付高血压的监测，而这项监测可以指导临床治疗，使患者避免将来其他器官的移植（肾脏和心脏）。一次骨髓移植可以延长六年的寿命，但是将导致不能支付高血压随访以及治疗项目，而这些措施可以在六年里保护6个人免于因血压控制不佳而导致的死亡。每一种新药上市、新仪器、新技术的使用，如CT、MR、腹腔镜手术、新生儿监护病房、超声内镜、器官移植等均需要同时考虑其临床和经济学效果。人们需要有效率的、效果好的卫生保健服务，社会也要求最大限度地充分利用所花费的卫生保健费用。我国人口众多，资源匮乏，进行卫生经济学的研究和应用显得尤其必要。只有同时从资源的使用（成本）和临床结果（效果，包括有效性，危险性和生命质量）二方面比较不同措施，进行临床经济学评价，在此基础上做出决策，才能使有限的卫生资源对改进人类健康发挥最大的作用。

## 第二节 临床经济评价的基本要求及分析评价的步骤

### 一、临床经济学评价的要求

临床经济分析和评价就是从对社会是否有利的角度，用经济学的基本原理和方法确定、测定、评价和比较各种卫生措施的成本及获利，作出经济分析，提供经济学上的证据。

1. 进行临床经济评价，首先需要明确经济评价的目的或者出发点，是单纯从病人的角度出发，还是从医疗费用的实施者（医院）、提供者或者全社会的角度出发。

2. 一项完整的经济分析必须是对二种或者两种以上方案的临床疗效和成本同时进行比较，完整的经济学评价方法包括最小成本-分析、成本-效果分析、成本-效用分析、成本-效益分析四种类型（表18-1）。仅比较临床疗效（3A），或者仅比较成本（3B）都不是完整的经济评价。

3. 选择合适的经济学评价类型以及准确估计各种成本，采用合理的临床结果测定单位。

表18-1 完全和不完全经济学评价比较

是否同时检查了各种方案的成本与效果				
		否		是
		仅检查效果	仅检查成本	
是否对两种或两种以上的方案进行比较	否	1A 部分评价 效果描述	1B 部分评价 成本描述	2 部分评价 成本-效果描述
	是	3A 部分评价 效力或效果分析	3B 部分评价 成本分析	4 完整的经济评价 最小成本分析 成本-效果分析 成本-效用分析 成本-效益分析





## 二、临床经济学分析与评价的基本步骤

### (一) 确定所要分析的项目及要进行比较的措施

1. 确定研究问题时, 应该包括被研究对象的人口学特征并写在研究开始部分。
2. 在研究报告中应写明提出所选的需要比较项目的理由, 其合理性如何。
3. 详细描述所选的具体治疗方案以及在各种治疗方案中所占的地位。例如治疗方案药物的具体剂量、用法和疗程。应进行文献复习, 比较其在各种治疗方案中占的地位。

### (二) 确定经济评价的观点

前瞻性的研究大多是从社会观点出发进行评价的, 确定对社会各部分的影响作用, 包括对患者、医院、提供医疗者以及整个经济的影响。成本与结果应分开报道, 便于成本和结果的确定。

### (三) 按照研究目的, 确定使用的经济评价方法

包括 CMA、CEA、CUA 和 CBA, 并陈述为何选择这种评价方法。

### (四) 获得资料的研究方法

可选以下任一种或两者的结合, 如前瞻性随机对照研究、荟萃分析、观察性资料等, 陈述选择该方法的理由, 一般前瞻性研究经常采用临床随机对照研究, 在此基础上进行 CEA 或 CUA。

### (五) 增值分析

在作 CEA 时, 研究问题的测定除成本/效果比以外, 还应该包括测定及报告增值分析 (incremental analysis) 的结果, 即由于额外措施造成成本的增加时, 其相应增加的效果是多少。(详见本章第四节)

具体表示为一个项目比另一个项目多付的费用, 与该项目比另一个项目多得到的效果之比, 称为增值比 ( $\Delta C/\Delta A$ )。

### (六) 确定结果测定的方法

例如用 CUA 时就应报告质量调整生命年和效用值的具体测定方法, 是用标准概率法、时间交换法还是等级尺度法, 并阐明这些方法的具体测定步骤。

### (七) 成本的确定

所有相关的成本都应确定并收集及报告。成本的报告应反映机会成本, 包括人头费、管理费等。成本的种类主要包括直接成本和间接成本。

### (八) 对发生在将来的结果和成本作贴现计算

当某一医疗措施的实施需要数年完成时, 为了准确地估计成本和效果, 去除由于物价上涨因素而带来的影响, 必须对发生在将来的成本和效果 (或效益) 通过贴现 (discounting) 的方法, 换算为目前的实用价值。计算公式:

$$P = \sum_{n=1}^t F_n(1+r)^{-n}$$

式中  $P$  为成本或效果现在值;  $F_n$  为成本或效果在  $n$  年时的值;  $r$  为年贴现率;  $t$  为项目完成的预期年限。

贴现率一般相当于银行利率。

### (九) 敏感性分析

在已得出上述经济评价的结果后, 测定当其中几个主要的变量如价格、成本、贴现率、结果的判断标准发生变化时, 以及不同的经济分析类型对评价结果的影响程度, 称为敏感性分析 (sensitivity analysis)。如果稍微改变一下变量的数值, 其经济评价的结论就发生改变, 则表明其可靠性较差。



由于对将来发生的某些情况如工资、失业率、期望寿命、治疗费、年贴现率等不能肯定,因此必须将敏感性分析作为经济评价中一项必要步骤。研究中所有不肯定的结果都应报告,应该使用关键参数可信限及范围的概念。

#### (十) 推广及应用价值

在前面分析和复习文献基础上,得出结论,并加以说明,应与其他研究结果进行比较,特别注意方法学上的区别。例如对间接成本的处理以及不同的人群的差别,这对研究结果的推广应用十分重要。

### 第三节 成本的种类及其测量

#### 一、直接成本

直接成本(direct costs)为卫生服务成本,系将资源直接提供治疗,服务时所花费的成本。

##### (一) 直接医疗成本(direct medical costs)

指卫生服务过程中用于治疗、预防、保健的成本,常包括住院费、药费、诊疗费、实验室检查费、X线检查费、手术费、家庭病房费、康复费及假肢等费用。

##### (二) 直接非医疗成本(direct nonmedical costs)

患者因病就诊或住院所花费的非医疗服务的个人成本,如患者的伙食、交通、住宿、家庭看护、由于疾病所要添置的衣服,患者住院后家属探望的往返路费、外地患者家属的住宿费等。

需要提出的是,在计算直接成本时收费的多少并不完全等于成本。但是,从病人角度出发,用收费来代替成本使用还是能说明病人的医疗花费的,因此,在一定情况下,可以用收费的多少来代替成本。

#### 二、间接成本

间接成本(indirect costs)为社会成本,指由于疾病而丧失的资源。

##### (一) 与病残率(morbidity)有关的成本

由于病假和疾病引起工作能力减退及长期失去劳动力所造成的损失,如因病假损失的工资、奖金及丧失的劳动生产力造成的误工产值。

##### (二) 与死亡率(mortality)有关的成本

由于病死所造成的损失,例如,规定60岁退休,患者因病于50岁死亡,损失的10年工资、奖金都应作为间接成本计算。

间接成本的计算有一定困难,通常的方法包括人力资本法和意愿支付法。人力资本法是用工资率、失业率、期望寿命、退休年龄等计算由于病残或死亡引起的收入减少。本法测定容易,但问题很多,因不同人群的收入并不相同,同时对没有收入的老人、儿童、失业者很难测定。意愿支付法,又称支付意愿法,是将市场经济的原则用于人的生命价值。提出任何结果和生命的丧失,某一脏器的切除,如果用钱能将此挽回的话,用某人愿意支付的金额数来估计。即直接测定由于减少疾病的病残和死亡个人自愿支付的费用。虽然此法符合经济理论,但应用时存在不少实际问题,例如在调查人们愿意为改善健康付出多少时,富人与穷人的回答通常是不一致的。但是当人们生命处于,某一危险或知道付出的多少会直接影响结果时,他们就会愿意多付。



### 三、无形成本

无形成本 (intangible costs) 是指由于疾病所致疼痛和死亡给家属带来的悲痛等疾病和医疗上非经济的结果, 如精神损失等。

## 第四节 临床经济学分析与评价的类型

### 一、最小成本分析

最小成本分析 (cost-minimization analysis, CMA) 也可称为成本确定分析 (cost-identification analysis)。比较结果相似的各种医疗方法, 并根据成本提出最佳策略。该类型是假设不同医疗措施的治疗结果相同, 确定不同医疗措施所花的成本, 选择成本量低的措施, 测定结果以提供的每一项服务所花费的成本来表示。例如骨髓炎患者提早出院在门诊继续用抗生素治疗和常规住院相比, 前者花费 2271 美元, 比常规住院 2781 美元的费用低。最小成本分析结果显示早期出院方案每例患者可节约 510 美元。由于只能比较同一种疾病相同结果时的成本, 故使用范围较小。

### 二、成本-效果分析

#### (一) 定义

成本-效果分析 (cost-effectiveness analysis, CEA) 是将成本和效果结合在一起考虑, 不仅研究卫生规划及医疗措施的成本, 同时研究卫生规划的结果, 它测定某项措施的净成本以及成本消耗后得到的效果。其表示方法为每一效果单位所耗费的成本 (成本效果比), 或每一个增加的效果所需要耗费的增量成本 (增量比) 等。这就使两种不同的医疗措施在进行比较选择时, 有了相同的评价单位, 从而为临床决策单位提供科学的依据。

成本-效果分析是用来确定最有效地使用有限资源的一种分析方法, 也是目前在医疗保健领域的完整经济评价方法中最常见的一种。在已经发表的该领域的文章中占 50%~75%。

例如某透析中心进行的成本-效果分析显示, 对终末期尿毒症患者每延长一个寿命年的成本, 使用连续腹膜透析为 33400 美元, 而在医院做血液透析为 48700 美元, 由此可见从经济学角度分析连续腹膜透析优于血液透析。

#### (二) 成本-效果 (CEA) 分析中效果的表示指标

在 CEA 中, 效果可以同时或分别使用中间测量指标 (intermediate measures) 和健康测量指标 (health measures)。前者包括症状、危险因素或有关临床测定的结果, 例如溃疡的愈合率、乙型肝炎病毒 e 抗原的阴转率、血清胆固醇的下降程度等。后者包括病残天数、寿命年的延长、死亡数等。例如在高血压的治疗项目中, 血压下降的百分率为中间测量指标, 而通过降压治疗预防卒中造成的死亡则是最终健康指标。如果当最终结果的测定所需时间太长时, 可选用中间结果。

#### (三) 成本-效果 (CEA) 分析的表示指标

在成本-效果分析的研究中, 通常采用的指标是成本-效果比和增量成本-效果比。

1. 成本-效果比 (cost/effectiveness, C/E) 成本-效果比即每个生命年或每治愈一例病人或每诊断出 1 例新病例所花费的成本。成本-效果比值越小, 就越有效。通常单一的成本-效果比值是没有意义的, 主要用于两个或两个以上的项目比较, 并且是比较有相同结果单位的两个项目。



例如比较纤维结肠镜和乙状结肠镜加钡剂灌肠两种措施对治疗下消化道出血及诊断结肠癌的成本-效果分析,结果显示纤维结肠镜不仅诊断的敏感性(80%)、特异性(95%)均高于后者(分别为57%和80%),而且有更好的经济效果。成本-效果分析的结果是治愈一例下消化道出血的成本,前者为2319美元,后者为2895美元;诊断1例结肠癌成本,分别为2694美元和2896美元。

2. 增值比 我们希望病人得到最有效的治疗方法同时也是最省钱的。然而,我们时常面临的情况是一项治疗和另一项治疗相比花费较多,但是也更有效。因此成本效果的平均比例还不能充分显示两者的相互关系,故可进一步用增量成本效果比来表示,其回答的问题是“这多出来的效果到底要花费多少?”,即计算一个项目比另一个项目多花费的成本,与该项目比另一个项目多得到的效果之比。增值分析能充分说明由于附加措施导致成本增加时,其相应增加的效果是多少以及是否值得。

例如,有一项治疗措施成本为3 000美元,可以挽救2个生命年,成本效果比为 $3\,000/2=1\,500$ 美元/每个生命年。新措施成本为6 000美元,可以挽救3个生命年成本效果比为 $6\,000/3=2\,000$ 美元/每个生命年。简单地看,同样挽回1个生命年新措施比老措施成本高,但是老措施挽救的生命年数较少。因此如果想挽回更多的生命,可能会选择新措施,那么需要提出的是:你愿意多付出多少来挽回一个生命年?

$$\text{增值比} = (6\,000 - 3\,000) / (3 - 2) = 3\,000 \text{ 美元/提高1个生命年}$$

因此,决策者可以考虑宁愿花3000美元在这项目上提高1个生命年,还是花在其他项目上。通常当利益和成本均增多时需要做增值分析。此外有多种措施供选择时,增值分析能提供更多的信息,从而指导做出决策。

对于临床医生而言,经济学分析主要用于临床决策。应同时考虑候选治疗措施的效果、风险和费用。以表18-2说明。

表18-2A 效果、费用综合分析

		与对照组比较,新治疗措施的效果		
		更大	相同	更小
与对照组比较,治疗措施的费用	更大	7	4	2
	相同	3	9	5
	更小	1	6	8

表18-2B 临床决策抉择

对于决策,强烈支持	支持强度较弱	不能做出决定
1. 完全接受该治疗	3. 接受该治疗	7. 是否值得采用由于效果增加费用也增加的治疗?
2. 完全不接受	4. 不接受该治疗	8. 是否值得采用由于效果减少费用也减少的治疗?
	5. 不接受该治疗	9. 费用与效果都与对照组相同,有无其他理由采取该治疗
	6. 接受该治疗	

在上表中,第1种情况中,新治疗措施不仅其效果比对照组好,而且价格更低,因此强烈推荐用新治疗措施。例如用根除幽门螺杆菌的方法(1周疗法)治疗十二指肠溃疡与服 $H_2$ 受体拮抗剂(1年)比较,前者溃疡复发率低、总费用下降,因此毫无疑问选择根除幽门螺杆菌的方法。第2种情况中,新治疗措施费用比对照组高,效果反而差,肯定不采用。第3、4、5、6种情况是支持强度较弱的。例如第3种情况中,两种措施的费用相同,在效果上新疗法要好一些,提示可接受新疗法,但支持强度不如第1种情况。第4种情况则是效果相同,新疗法费用反而比对照组高,因此不采用新疗法,但支持强度不如第



2 种情况。第 5、6 种情况相同。第 7、8、9 种情况中,不能马上做出决定。第 7 种情况效果增加费用也增加,此时要做增值分析。例如用 tPA 与链激酶较其溶栓治疗急性心梗的效果就是这种情况。增值分析结果显示,每增加一个生命年要多花费 33000 美元。据此根据不同情况做出决策。

### 三、成本-效用分析

成本-效用分析 (cost-utility analysis, CUA) 是 CEA 分析的一种特殊形式。由于 CEA 不能比较两个完全不同的卫生项目,如肾移植治疗慢性肾衰竭与预防卒中的抗高血压治疗项目,因为两种措施干预的病种不同,而且对病残或病死率的影响也不同,因此无法应用 CEA 比较两者的经济效果。如果将其分母单位都化为质量调整生命年 (quality adjusted life year, QALY),进行成本-效用分析就可以对两者进行比较。结果显示肾移植项目每获得一个 QALY 花费的成本为 4710 英镑,而抗高血压治疗预防卒中为 940 英镑/QALY。结论是后者的经济效果好。表 18-3、表 18-4 分别显示不同健康状况的效用值及不同疾病的成本-效用分析结果。

表 18-3 不同健康状况的效用值

健康状况	效用值
健康	1.00
高血压治疗副作用	0.95~0.99
肾移植	0.84
中度心绞痛	0.70
家庭透析	0.54~0.64
严重心绞痛	0.50
抑郁	0.45
死亡	0.00
失去知觉	<0.00

表 18-4 成本-效用分析

项 目	成本/QALY (英镑)
胆固醇测定和节食疗法 (40~69 岁)	220
脑外伤神经外科治疗	240
劝导戒烟	270
蛛网膜下腔出血神经外科手术	490
预防卒中的抗高血压治疗 (45~64 岁)	940
装起搏器	1100
髋关节置换	1180
主动脉瓣狭窄换瓣	1140
冠状动脉搭桥 (左主干病变严重心绞痛)	2090
肾移植	4710
乳腺癌普查	5780
心脏移植	7840
胆固醇监测和治疗 (25~30 岁)	14150
家庭血液透析	17260



续表

项 目	成本/QALY (英镑)
冠状动脉搭桥 (单支病变、中度心绞痛)	18830
连续腹膜透析	19870
医院血液透析	21970
用促红素治疗血液透析患者贫血 (假设可降低病死率 10%)	54380
恶性颅内肿瘤神经外科手术	107780
用促红素治疗血液透析患者的贫血 (假设不增加存活率)	126290

(引自 Mason J, Drummond M, Torrance G 1993)

应用值测定方法有以下几种。

### (一) 时间权衡法 (time trade-off)

让患者在“接受某一特殊措施后,可维持好的健康状态,但是活的时间却要短些”与“不接受这一特殊治疗可维持目前的症状,但是活的时间要长些”之间做出自己的选择。例如告诉心绞痛患者,如果不治疗可带病再活 25 年,但假设有某一种治疗可使心绞痛完全缓解,可是寿命可能要缩短些,问他无病生存时间为多少年 (X) 时,他才宁愿选择这一治疗,这就需要患者决策。如果患者愿意能健康地活 15 年才选择这一治疗,否则就拒绝,于是没有心绞痛的 15 年就相当于有心绞痛的 25 年效用。心绞痛的效用值为  $15/25 = 0.6$  (即  $0.6 \times 25 = 15$  年),也即 X 年的健康 = 效用  $\times$  不健康年限。

### (二) 标准概率法

又称标准博弈法 (standard gamble)。这是一种风险选择法 (最坏和最好的结果),即在可选择的范围内做出的判断。例如某一疾病可以手术治疗,但要冒好或坏的风险,手术 (A) 的最坏结果是死亡,最好的结果是术后可以无病生活 25 年 (风险选择),其概率均为 50%;另一方面也可以进行姑息治疗 (B),而不冒手术的风险,但处在带病状态,效果比手术的最佳效果差。因此,可以在手术和非手术姑息治疗间作出选择。当问患者姑息疗法可生存 5 年时,选择 A 还是 B,患者回答选择 A。生存 6 年时仍选择 A。生存 7 年时患者改为选择 B,也就是说患者宁愿不手术以带病状态生存 7 年,也不愿冒有 50% 可能死亡,50% 可能治愈生活 25 年的风险,此时该病的效用值为  $7/25 = 0.28$  (图 18-1)。

### (三) 等级尺度法 (rating scale)

画一条线,由患者自己操作。每一条线两端写上描述性短语,线可划为 10 等份。0 为死亡,1 为健康,将疾病状态清楚地描述给患者后,要求患者在线段上某一点画一条竖

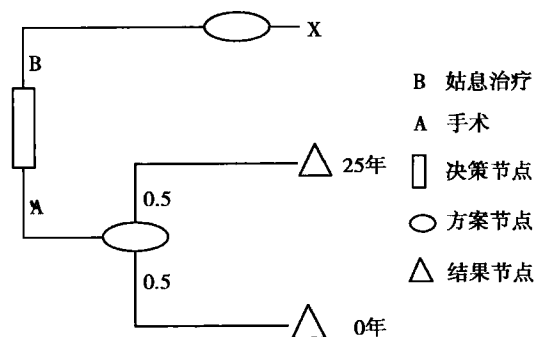


图 18-1 标准概率法



线,以表明自己目前的健康状态,画线处即为自己所得的效用值。等级尺度法是 CUA 评价的方法之一,它在信度和效度方面能比较客观地反映健康的效用值。例如:

0	1
死亡	健康

#### 四、成本-效益分析

成本-效益分析(cost-benefit analysis, CBA)是将医疗服务的成本和效果都用货币单位来表示,用相同的单位来分析所花的成本是否值得,常用效益成本比或净效益(效益-成本)来表示。以泌尿道感染的预防为例,以往每年有 2 次以上尿路感染的患者分为两组,一组给予复方 SMZ 预防,另一组给予安慰剂,治疗组发作次数 0.15 次/年;对照组感染发作为 3 次/年,每年抗感染的治疗费用为 126 元/人;而预防给药费用仅 85 元/人,以人为单位,成本-效益分析如下:

预防费用:85 元;

效益:  $126 \times (3 - 0.15) = 359$  元;

效益-成本比:  $359/85 = 4.2$

净效益:  $350 - 85 = 274$  元。

从结果可以看出这个预防项目有经济效益,是值得推广的。

另一项风疹疫苗预防项目提示如用于 12 岁以上女性,其效益/成本比为 25:1,而用于 2 岁以下的男儿童,则效益-成本比为 8:1,显然前者优于后者。

#### 五、临床经济分析类型的比较

表 18-5 将上述三种常用的临床经济分析方法进行了比较。

表 18-5 几种不同类型临床经济分析比较

	CEA	CUA	CBA
比较方式	C/E	C/U	B-C 或 (B/C)
成本单位	货币	货币	货币
结果单位	临床自然单位	QALYs	货币
相比较的措施	2 个或 2 个以上	2 个或 2 个以上	1~2 个或以上
可测定的目标	1 次测定 1 个	1 个以上	1 个以上
需测定结果资料	效果测定	效用测定	健康效益转为货币
测定方法	根据不同结果单位而变化	标准概率法 时间交换法 等级尺度法	人均收入法 支付意愿法
可比性	随结果测定而不同	理论上可比	理论上可比

### 第五节 临床经济学分析结果的真实性评价

#### 一、立场是否明确

是否有详细描述成本效果分析的出发点(一家医院,一个卫生部门或社会作为一个整体)。出发点应该由两方面观察,一方面谁来支付治疗费用,另一方面最后治疗或不治疗的决定将影响谁。



从不同的角度或立场进行的经济分析其成本和结果的评价常常是不同的。因此经济分析的目的是否明确,立场是否得到广泛认同,对评价结果的真实性起很大作用,也决定了你能否将这一经济评价结果应用于你的临床实践。例如在—项评价早期出院的经济分析研究中,单纯报告住院费用是不够的,因为病人早期出院后可能在社区中进一步处理。而这—部分费用并不包括在住院成本中,但却是费用支付者(病人或者第3方)所关心的。

## 二、是否比较了所有相关的临床措施

分析是否比较了明确定义的可替换的措施?需要详细列出各治疗措施之间的区别。典型的成本效果分析是采用两种治疗措施比较或者治疗与不治疗比较。治疗措施应该是较常用的。如果是选择不常用的治疗将导致结果的偏倚。

## 三、是否提供了完整的经济评价

完整的经济分析是比较2种或2种以上治疗,诊断或其他医疗措施,并且同时从临床结果和成本两方面评价。如果仅仅比较成本,则是成本分析,如果仅仅比较临床结果,如随机对照临床试验,那么仅仅是决策的部分评价。

## 四、成本和临床效果测量是否准确

是否分析确定了所有你认为的重要的相关的成本和效果?是否选择了可信的测量方法计算这些成本和效果?

成本方面,包括组织(新建—个项目)的实际成本和持续的成本,患者和家庭的附加成本,卫生系统以外的成本(丧失工作时间,减少的生产力等)并且包括无形成本(疾病所致的疼痛,丧失伴侣所致的悲痛等)。这些成本必须在干涉措施和替换措施中都进行比较。

在效果方面,分析应该包括客观的临床结果:死亡率、病残率、剩余的功能、调整的生命质量/调整的效用和将来的效果。措施的效果应该是有效的。这需要由设计好的随机临床研究中提供可靠证据,来显示干涉措施是有效的。如果事先没有这部分资料,那么需要先做系统综述或者荟萃分析作为经济学分析的一个部分。如果没有提供相关临床问题的RCT,那么也可以用观察性研究,但是要警惕这将产生更多主观倾向性的偏倚。

## 五、结果成本是什么,是否做了增量分析

在两种措施进行比较时,如果—项更有效的措施花费更少,那么无可厚非地选择花费更少的措施,除非是有严重的并发症。如果新措施在增加了临床疗效的同时也增加了成本时,尽管单位效果的成本新措施比老措施少,此时决定是否采用新措施就需要进行增量分析。

—个很简单的获得快速成本效果分析的方法是应用需要治疗数目(NNT)。例如,GUSTO研究心梗的溶栓问题,当t-PA替代链激酶时出现不同的结果:t-PA花费\$2000/次,死亡率6.5%,而链激酶花费\$200/次,死亡率7.5%。ARR是两者死亡率之间的区别,即1%。t-PA的NNT是100(1/ARR),这是用t-PA替代链激酶治疗多少例病人可以多预防—例死亡。每条生命的增益成本是\$ $(2000-200) \times 100 = \$180000$ 。

## 六、结论是否随着成本和结果的敏感性改变而变得不可靠

在经济分析研究中,常常存在很多不确定因素,或者这些因素在不同时间,不同场合,不同人群中变化较大,常常影响结果的推广应用。例如在不同人群中,某种措施的疗





效可能不同。在不同场合,不同时间,医疗成本的变化可能较大,人们的价值取向可以完全不同等。这些都可能影响研究的结果。作者需要采用敏感性分析的方法假想改变这些可能变化的因素,确定结论是否随着成本和结果的改变而变得不可靠,从而进一步证明结果的正确性。通常进行分析的数据只是近似值,附加95%的可信区间。所以,需要进行敏感性分析来检查分析中的假设。通过敏感性分析可以评价改变可信区间范围内的评估值是否会影响结果或结论的稳定性。

## 七、研究中评估得出的成本和结果是否和人群中的基线接近

人群中存在不同层次的风险。在一个亚组中有经济效果的可能在另一组中就没有了。研究应该试图确定人群中的各组群并且对他们进行亚组的成本效果分析。例如,我们研究治疗心力衰竭的同样有效的药物的成本效果时,会发现对严重心脏衰竭的患者是有经济效果的药对较不严重的病例中是没有经济效果的。在研究 $\beta$ 受体阻滞剂在心脏衰竭中的作用时发现在高危人群中比低危人群更有经济效果。然而,这都是建立在两种情况都能挽救一条生命的基础上。

## 八、是否讨论了读者关心的问题,包括伦理问题

一项效果好的措施同时也是最廉价时,作出决策是容易的,但是,很多时候是增加效果就会增加成本。我们到底愿意支付多少来挽救一条生命?每获得一个质量调整生命年所能接受的成本是多少?如当费用降低时,效果也减少(寿命的缩短或病死率的增加)。此时是否要采用该项措施。这时就涉及到伦理上的问题。

(王吉耀 林果为)

## 第十九章 健康相关生存质量的测量与评价

健康相关生存质量 (health related quality of life, HRQL) 对临床研究与医疗质量的评价均很重要。随着疾病谱与医学模式的转变, 临床科研与医学实践, 所面对的病人不单是纯生物学的, 而是具有生物、心理、社会等综合属性的“完整”病人。因而, 以病人为中心的健康相关生存质量的分析与评价, 将临床研究推向了一个更为广阔的范畴, 具有十分重要的现实意义。

### 第一节 健康相关生存质量的基本概念

#### 一、产生背景

随着医学模式从单纯生物医学模式向生物-心理-社会-环境医学综合模式的转变, 临床医疗实践与研究也应从单纯的生物学视角, 扩展到患者生理、心理、社会与环境功能的综合范畴。若再用传统的纯生物学指标来分析与评价这种医学模式下的综合效应, 显然有一定的局限性, 不符合以病人为中心的新医疗观, 因此, 迫切需要一套新的指标体系, 以全面反映患者的健康状态。

伴随着社会与经济的快速发展, 人们生活水平得以普遍提高, 平均期望寿命也随之延长。同时由于现代医学的进步与发展, 过去许多难治的疾病得以早诊早治, 病死率随之减低, 生存期也大大延长; 在生存时间延长的同时, 人们开始关注生存的质量, 是不是随着患者生存期的延长, 生存质量也相应提高了呢?

在以上背景下, 对健康相关生存质量 (health-related quality of life, HRQL) 研究与评价应运而生。

由于 HRQL 涉及了生理、心理、社会及疾病本身等诸多因素, 对其测试具有多指标、多时点以及主观隐匿等特点, 远非像生物学指标那样简单, 因而必须设计颇为完善的测量工具—量表。最早有关生存质量的测评量表, 当推 20 世纪 40 年代 Karnofsky 研制的 KPS 量表, 20 世纪 70 年代为 HRQL 引入和探索期, 医学研究领域开始正式引入通用量表测定病人的健康相关生存质量, 进入 20 世纪 80 年代后, 开始制定专用量表用于测量特定慢性疾病及肿瘤患者的 HRQL。

近十多年来, 健康相关生存质量已成为临床研究的热点问题之一, 越来越多的临床工作者以及卫生决策者开始关注生存质量研究, 并应用于临床研究以及指导对疾病的防治实践。生存质量文献的发表数量也呈几何倍数增长, 其中 2000~2005 年度有关生存质量的发表文献篇数与 1971~1975 年度相比, 增长了千倍之多! 由此可见对 HRQL 的研究与应用的重要性。

#### 二、健康相关生存质量的定义

要了解健康相关的生存质量定义, 应分两部分理解: 健康与生存质量。世界卫生组织 (WHO) 对健康 (health) 下了如下定义: “健康不仅意味着无病, 而且在生理、心理及社会功能等方面, 都要处于一种完全的良好状态” (WHO definition of health as a state of complete physical, mental and social well-being and NOT merely the absence of disease)。



可见健康既要在生物学上没有患病，又要综合生理、心理以及社会功能状态，这与新的医学模式相互对应。健康状态、功能状态、生存质量三个概念在用于描述健康状态方面常有交叉。

那么什么是生存质量呢？生存质量从本质上应该体现以下内涵①生存质量是一个多维概念，应测量生理、心理及社会功能状态；②生存质量是评价主观感受或体验，属于主观指标，应由被测试者自己评定；③生存质量是建立在一定的文化与价值体系之下，不同文化背景的观点不尽相同，对生存质量的理理解存在差异。

对此，世界卫生组织提出了一个能为广泛接受的定义，“生存质量是指个人处于自己的生存环境与文化价值体系下，对本身生存的一种自我感受，它与个人的生存目的、期望、标准及关注有关”（WHO definition of QOL as an individual's perception of their position in life in the context of the culture and value systems in which they live and relation to their goals, expectations, standards and concerns）。

可见生存质量既有人生哲理的内涵，也有人类生存相关的文化以及社会、心理诸多方面面对人们生存质量的综合影响。从哲学的角度看生存质量，宜将之称为“生命质量”，强调自身价值的实现以及对社会的作用；从社会经济学的角度，宜称之为“生活质量”，反映一般人群对其生活的自然、社会条件以及其自身状况的主观评价和体验；而从临床医学角度，衍生出健康相关生存质量（HRQL），用以评价病人对其疾病和相关临床干预所产生的在躯体、心理、社会地位上的作用及影响的主观认知和体验。

HRQL 涉及疾病、生理功能、心理功能、社会功能等多方面内容，而且各方面间是相互作用、相互影响的（图 19-1）。例如人们由于患病，机体必然出现病理损害而导致生理功能下降或出现功能障碍；在心理及精神上亦会产生压力，造成心理负担及精神刺激（如出现“戴帽子效应”等）；同时对于社会交往、人际关系及社会适应等，亦会受到不同程度的限制。因此，健康相关生存质量必然会随着心理、生理及社会诸功能状态的降低而下降；反之，也会因为疾病的痊愈，引起上述诸方面功能的好转或正常，从而使健康相关生存质量得以改善。

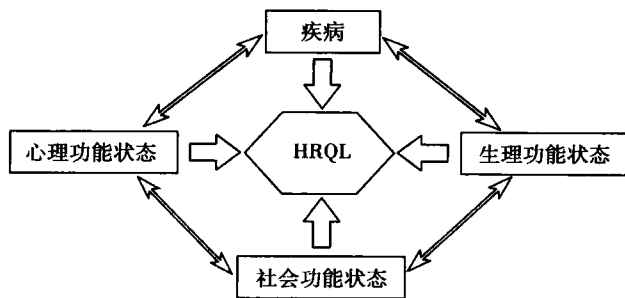


图 19-1 健康相关生存质量示意图

尽管收入低或不稳定、环境质量恶劣、人身自由受限等，也可能最终影响健康，降低生存质量，但对健康的作用多为间接，一般不列入健康相关生存质量（HRQL）的测试范围。

## 第二节 健康相关生存质量的测试工具（量表）

鉴于健康相关生存质量的测定较为复杂，需要借助专门测量工具（或称量表）。量表测试内容的变迁经历了三个发展时期，最早为“硬指标期”，早期研究多局限于测试所谓



的“硬指标”范畴，如生存时间、收入、教育程度、身体结构是否完整等。鉴于这些指标可以单独测量，后来逐渐加入一些主观指标，进入“主观感觉指标为主期”，重点测试评价对象的主观感觉以及健康和社会环境状态等。20 世纪 80 年代中后期，生存质量界定及测量愈来愈趋向于仅测量主观感觉指标。

目前用于测试 HRQL 的量表数以千计，按照测试目的、量表内容及适用范围大致可归为两大类：即通用量表（工具）和专用量表（工具）。

## 一、通用量表

HRQL 通用量表（generic instrument of HRQL）常用于病人健康状态的总体估价，测试内容一般不具有针对性，它包括了与健康相关的生理、心理及社会功能等诸方面内容。在现有通用量表中，较为常用的有 WHOQOL-100 型量表及其简化量表，以及 MOS-SF-36 量表等。

WHOQOL-100 型量表是由 WHO 牵头设计、全球 31 个国家（包括中国）研究者共同参与、协作完成的，共包括 100 个条目（问题），涉及了生理、心理、独立生活水平、社会关系、环境以及宗教信仰六大方面内容。其中生理领域包括疼痛与不适、精力与疲乏、睡眠与休息、行动能力、日常生活能力、对药物及医疗手段的依赖性、工作能力七大项，心理领域方面包括积极感受、思想/学习/记忆和注意力、自尊、身材与相貌、消极感受、精神支柱六大项。设置这些测试内容对估价患病个体以及一般人群的健康相关生存质量是颇为有用的。有研究表明，该量表具有较好的信度与效度。但在大规模临床研究中，因 WHOQOL-100 型量表条目过多，完成时间较长，可能出现依从性较差的问题，为此又相继开发了简化量表（详见本章附件 1）。

SF-36 量表涉及了生理功能（10 个条目）、生理问题对功能的限制（4 个条目）、心理问题对功能的限制（3 个条目）、心理健康（5 个条目）、精力疲惫或乏力（4 个条目）、疼痛（2 个条目）、社会功能（2 个条目）、健康总体评价（5 个条目）八个方面，外加一项测试健康状态前后变化的条目，条目数共计 36 个。该表亦较实用（详见本章附件 2），其信度及效度颇佳，简便易行，可以在 5 到 10 分钟内完成，敏感性高。同时也可以该量表为核心内容，与其他一些条目搭配使用。因此，在临床研究中已被广泛采用。

类似通用量表还有 EuroQoL 量表、Spitzer's QOL 指数、SIP 量表等。

根据通用量表的应用特点又可进一步细分为健康维度亚类和效用测量亚类。健康维度类既可按照不同维度分别比较生存质量，又可经降维后，直接比较综合评分；效用测量类量表融合了健康维度量表与专用量表的一些特征，可直接得到一个综合效用值，一般将死亡的效用值定义为 0，完全健康的效用值为 1，其他生存质量状态的效用值则介于两者之间。此类量表常用于成本效用评价以及临床决策分析。

## 二、专用量表

通用量表可用于不同疾病别、不同人群间 HRQL 的横向比较，但敏感性较差，若临床用以观测某些特定慢性疾病的患者生存质量或药物治疗中的某些反应时，可选用专用量表。专用量表（specific instrument of HRQL）具有针对性强、敏感度高的特点，专用量表形式和内容也多种多样，既可以是疾病别量表（如 COPD 量表、癌症量表）、年龄别或特殊人群量表，又可以是功能别量表（如睡眠或性功能量表）、症状别量表（如疼痛量表）等。其中又以癌症别专用量表较为多见，如欧洲癌症治疗与科研组（EORTC）设计了一套专用癌症别量表，每个量表均由核心问卷加上癌症别特殊问卷两部分组成。共同核心问卷称为 QLQ-C30，包括 5 个功能子量表（躯体、角色、认知、情绪、社会功能）以及 3



个症状子量表（乏力、疼痛、恶心呕吐）。并根据不同癌症的具体特点增加了疾病别问卷，目前有肺癌量表（QLQ-LC13）、乳腺癌量表（QLQ-BR24）、头颈部癌症（QLQ-H & N37）量表、食道癌（QLQ-OES24）量表、直肠结肠癌（QLQ-CR38）量表等，该套量表于 1993 年研发成功，并被翻译成 20 多种语言，在临床研究中已被推广使用。

### 三、量表选用的注意事项

量表的基本单位是条目（问题），由多个测量目的相似的条目组成一个子量表，再由多个子量表，共同构建一个完整的量表。量表的设计是一个相当复杂的过程，从设计条目池、条目筛选、词义表达、信度与效度测试等方面，均需要有关专业的专家共同参与，同时还需要在实践中反复修改、补充和完善，以达到可信度与效度俱佳，为 HRQL 的准确测试服务。

由于我国对健康相关生存质量的研究起步较晚，尚缺乏适合中国国情的高质量量表，大多临床研究是引用国外现成的生存质量测试量表。目前有大量的量表可供选择，例如有关乳腺癌的量表就有 21 个，因此，应结合临床研究目的，选择那些与 HRQL 密切相关、真实可靠、使用频繁的量表。

但即使是一种有价值的 HRQL 量表，在不同国家、不同民族以及文化和生活习俗各异的环境下应用时，HRQL 测试结果可能出现不一致。所以对任何生存质量的测试量表，特别是信度和效度尚不理想的量表，需要结合实际，在实践中不断加以检验、修改和完善。因此在引用国外量表时应特别注意：

1. 量表条目翻译的等价性 即条目的中文含义应符合原文。可先请多名专家背靠背地将原文翻译成中文，经汇总统一后，再请专人将译好中文量表重新翻译成外文，并与原文比较，确保量表译文简单易懂又不失原义。

2. 执行的等价性 调查方式与操作规范应统一，调查内容、范围、定义、对象等尽可能与原量表保持一致。

3. 测量尺度的等价性 如原量表测量尺度等间距，中文量表也应如此。以 SF-36 量表中第 8 个问题为例，“过去四周内，身体上的疼痛影响你的工作和家务事吗？”备选答案是“1. 完全没有影响 2. 有一点影响 3. 中等影响 4. 影响很大 5. 影响非常大。”上述 5 个备选答案反映了疼痛的不同影响等级，原文中相邻等级之间的尺度是等距的，因此 5 个备选答案的中文表达在尺度上也应与原文一致。

### 四、HRQL 量表测试的一般要求

在正式测量健康相关生存质量前，还要考查欲使用量表是否满足下面一些基本要求：

1. HRQL 量表的测试条目 应简单易懂，被调查者能正确理解和准确回答。
2. HRQL 量表要适合研究对象 处于疾病的急性进展期和慢性恢复期，对量表的要求不同。
3. 确定信息收集方式与量表形式 是患者本人还是相关人员（如家属、医生）提供信息？是自评还是访谈者相助？若是病人自评，要考虑病人的文化程度及宗教信仰；是开放式问卷还是封闭式或者两者兼而有之？完成量表的时间？这些问题在使用前均应加以考虑。
4. 采集 HRQL 信息的场合应合适 是住院、门诊还是两者兼有？测试场合不同可影响真实性。例如住院病人的依从性要高于社区患者，但在医院内测试，可能由于人情世故，得不到患者的真实感受或体验。



### 第三节 HRQL 量表的基本功能与质量测评

用量表测试健康相关生存质量，要如实反映被测试者 HRQL 的真实水平，就要求测试量表具有良好的预测、判别、评价功能以及一定的信度与效度。

#### 一、HRQL 量表的基本功能

HRQL 量表应具备以下三种基本功能，测试目的不同，可有所侧重。

1. 预测功能 即应用某种特殊的 HRQL 量表所获得信息，在尚无特殊事件发生的情况下，通过追踪观察健康相关生存质量的变化情况，具有预测某种特殊事件发生的功能，例如预测疾病的发展、转归、康复或死亡等。

2. 辨别功能 即应用 HRQL 量表同时测试不同健康状态的对象，其结果要能反映出测试对象间 HRQL 差异的真实水平，因而具有辨别健康相关生存质量差异性的功能（图 19-2）。

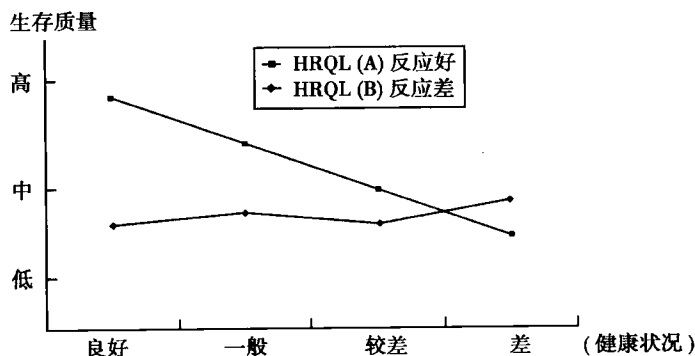


图 19-2 HRQL 量表辨别性能示意图

3. 评价功能 即应用 HRQL 量表，测试患者接受治疗或干预前-后的健康相关生存质量变化，应对干预或治疗效应具有评价功能。例如高血压患者血压值处于高水平时，HRQL 测试值为低水平，而如果患者接受了有效的降压治疗后，血压水平下降至正常水平，症状缓解，健康相关生存质量改善，量表的测试值也应随之上升。能反映治疗前后 HRQL 测试值变化的量表，就具有评价功能。量表评价功能的好坏与量表条目的敏感度（sensitivity）或称反应度（responsibility）密切相关（图 19-3）。

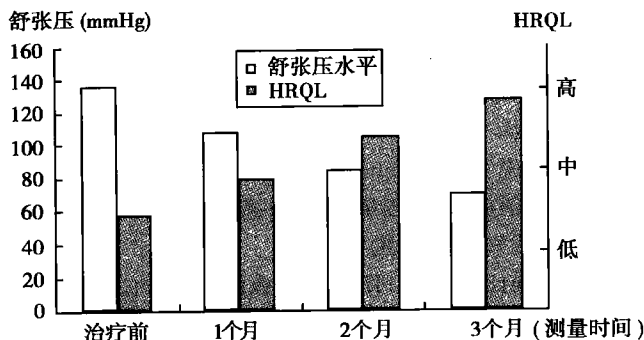


图 19-3 HRQL 量表评价性能示意图



## 二、HRQL 量表的信度与效度评价

### (一) 信度 (reliability)

信度又称为可靠性,是指 HRQL 量表所测试结果的可靠程度或可重复的程度。HRQL 量表应力求有好的信度,要能很好地被重复 (reproducibility),同时量表本身所含条目的内部一致性也要强。信度变化范围在 0~1 之间,在临床研究中,信度应在 0.70 以上,这样就能确保 HRQL 测试结果的稳定性。

量表的信度评价主要有两个方面:重测信度与内在一致性检验。

1. 重测信度 为了验证 HRQL 量表测试结果的可信度,可以采用两种方式:一是同一个调查员对相同研究对象,间隔适当时间后,重复测试,然后分别就两次的测试结果进行一致性分析 (intro-observation analysis),计算 Kappa 值或组内相关系数 ICC;或采用另外一种方式,即两个或多个调查员对同一批研究对象重复测试 HRQL (如随机抽取 20~30 个研究对象,由两个调查员分别测试 HRQL),然后将两次结果进行一致性分析 (inter-observation analysis),同样可计算 Kappa 值。如果 Kappa 值  $>0.75$ ,则表明信度高;0.4~0.75,信度较好;若  $<0.4$ ,则信度差。

2. HRQL 量表条目间的内在一致性检验 旨在检验有关 HRQL 量表中的条目,在测试整体 HRQL 的过程中,各条目间内在的一致性程度,通常用 Cronbach  $\alpha$  系数表示。Cronbach  $\alpha$  系数大于 0.8,表明条目间内在的一致性较好,结果较为可靠。但大多数研究者认为若该系数大于 0.6 即可接受。有研究表明, SF-36 量表及其各分量表的信度在 0.80~0.93 之间,信度较佳。

信度实际上反映了可重复性,信度好,重复性就好,可以使个体测量值趋于稳定。但是如果存在系统误差,会使测量值重复过高或过低,偏离 HRQL 的真实水平。因此,量表只有好的信度是不够的,还应考虑测试结果的真实性。

### (二) 效度 (validity)

又称真实性,效度是指 HRQL 量表所测试的结果符合被测者生存质量的真实程度。主要从表面效度、效标效度和结构效度等方面加以评价。

1. 表面效度 又称内容效度,是根据一般的主观感觉并借助病理生理学以及临床专业知识,对量表中条目进行的综合评价。从“表面上”评价测试内容与测试目的是否一致、适用性如何、是否简洁明了、是否有可能出现偏倚影响完整性、有无多余条目?从而确立量表的表面效度。

2. 效标效度 (标准效度) 选用某个常模量表或临床结局作为“金标准”,同步与待评量表测试研究对象,若两者测试结果高度一致,则说明待测量表,具有很好的效标效度。常采用相关系数表示,一般要求  $>0.7$  以上。如为验证 WHOQOL-100 型量表简化表的真实性,可选用 SF-36 量表为标准,进行效标效度分析。但在很多情况下,并无现成的金标准量表,应采用结构效度加以评价。

3. 结构效度 应用 HRQL 量表测出的生存质量水平,应与患者的病况相关联,反映患者 HRQL 的真实水平。即患者病况差时,测试的 HRQL 分值就低;而当患者病况好转或痊愈时,测试的 HRQL 分值就高。若量表各个条目的实际测试结果都反映出良好的真实性,那么量表的结构效度就好,反之则差。

评价结构效度,可用信噪比方法。HRQL 量表的实际测试结果包含两种成分,其一,是有用的信息,称之为“信号”(signal),反映了 HRQL 真实水平的信息量;其二,是无用的信息,称之为“噪度”(noise),表示 HRQL 量表中有关测试条目反映出的信息结果,与拟测试 HRQL 总目标无关或关联弱。将“信号”和“噪度”联合考虑,用“信噪比”



表示, 信噪比=信号/(信号+噪度), 自然是该比值越大越好。“信号”越强, “噪度”越弱, 其结构效度越佳。一般认为信噪比应在 0.7 以上, 表示结构效度较好。

### (三) 信度与效度的综合评价

如果 HRQL 量表仅有良好信度, 或者仅有好的效度, 都是不够的, 信度与效度两者缺一不可。只有信度与效度俱佳的量表, 方可称为好的 HRQL 量表。效度、信度综合评价见图 19-4。

有关信度和效度评价的统计分析方法, 较为复杂, 可参阅相关生物统计学专著。

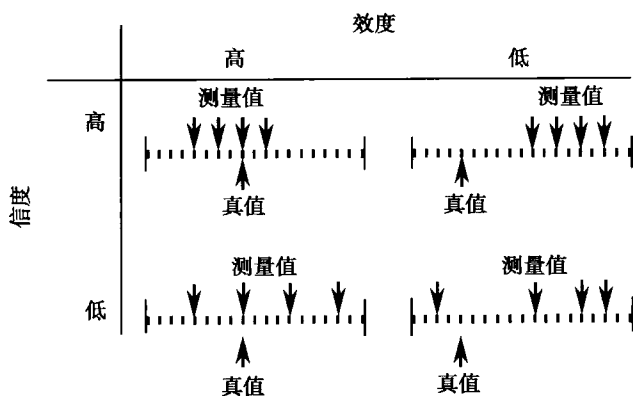


图 19-4 效度、信度综合评价图

## 第四节 HRQL 在临床研究及卫生政策决策中的应用

在传统的临床医学研究中, 对于临床效果判定, 通常只注重有关生物学指标, 而且都是由临床学家根据治疗反应自行决定。对生存者预后的健康状态及其 HRQL 自我评价却少有关注。近年来, 这种状况正逐渐改变, 越来越多的研究者开始重视 HRQL 的研究与评价, 并逐步应用于临床研究、卫生决策以及临床实践之中。

### 一、HRQL 用于慢性非传染性疾病的疗效测量与临床决策

HRQL 可直接用以测评一些疾病的临床变化与效果, 主要用于患慢性疾病且需要长期维持治疗者, 如患有心脑血管疾病、高血压、心肌梗死、肿瘤、糖尿病、风湿病、有关神经疾患以及药物成瘾者等。

借助健康相关生存质量还可用于临床决策分析, 如乳腺癌患者乳腺全部切除还是部分切除治疗方案的选用, 喉癌病人的手术与放疗选择等等, HRQL 均起到决定作用。

### 二、HRQL 用于药物临床试验以及药物流行病学研究

采用 HRQL 量表测试疗效与药物不良反应 (adverse drug reactions, ADRs), 可以帮助研究者获得更为全面的结论。如在长期维持治疗中, 发现患者的 HRQL 由好变差时, 则要警惕药物不良反应, 这种生存质量的变化有时可能早于生物学指标, 因此有重要的临床价值。

### 三、HRQL 用作评估社区人群和特殊人群的总体健康状况

如用 COPE 老年综合评价量表研究老年人的生存质量。





## 四、HRQL 还可用于效用评价与卫生经济学评价

HRQL 还可用于效用评价 (utility evaluation) 与卫生经济学评价效用测量对于临床医疗、卫生决策以及卫生资源分配等, 有重要的参考价值。效用综合考虑了生存时间和 HRQL, 一般采用质量调整寿命年 (quality adjusted life years, QALYs) 表示。如定义死亡的 HRQL 为 0 分, 完全健康为 1, 其余在 0 和 1 之间变化。患病时病人对自己病后的 HRQL 可在 0~1.0 之间评分定位, 然后来评价自身的生存效用。结合费用可进行卫生经济学的成本效用分析。QALYs 可为不同健康服务干预措施的结果评价与抉择提供重要信息。

## 第五节 健康相关生存质量研究的评价原则

健康相关生存质量研究的评价可参考以下标准。

### 一、研究结果是否真实可靠

#### (一) 主要标准 (I 级)

1. 对于 HRQL 量表, 是否作了信度和效度评价? 量表的辨别或评价功能如何? 如果 HRQL 量表的信度和效度好, 确有一定的辨别或评价功能, 则可进一步评价。

(1) 研究者测试患者 HRQL 的内容, 患者本人是否也认为是重要的? 医生考虑的 HRQL 一定要与患者自己认同的重要内容相统一, 而不能仅注重临床及生物学指标。例如慢性肺心病患者由于肺心病变, 气体交换障碍造成缺氧,  $O_2$  分压低和  $CO_2$  分压高, 出现呼吸困难, 全身乏力等患者特别难受的症状。经临床治疗后, 若医生只关注血气分析结果以及肺活量、心功能的变化, 忽视了患者生存质量变化与改善, 那么疗效评价就很片面。因此, HRQL 要采取医患相结合的原则与方法进行测评。

(2) HRQL 测试的结果是否有被省略的重要内容? HRQL 除测试有关体能、心理及社会功能等方面的重要内容外, 应特别关注疾病所致功能下降或障碍等内容。例如对风湿性关节炎患者, 如果 HRQL 的评价仅仅注重疼痛及其体能方面的变化, 而缺少了心理及社会活动功能方面, 评价就很片面。因此, 在研究 HRQL 时一定要避免有关重要 HRQL 内容的遗漏或省略。此外, 在临床研究中, 还要注意药物不良反应 (ADRs) 对生存质量的影响, 特别警惕是否有这方面重要内容的省略。

2. 研究者是否对 HRQL 分析方法作了评价? 是否全面报告了 HRQL 量表测试内容、分析方法以及相应结果; 有无地板效应和天花板效应? (在接近 HRQL 上下限时, 即使 HRQL 变化明显, 测试分值无变化或变化幅度很小, 出现钝化); 试验过程中是否有失访并影响了最终结果?

#### (二) 次要标准 (II 级标准)

研究者对生存质量是否作了质与量的转化以及卫生经济学评价? 是否采用了正确的方法作 QALY 分析以及成本效用分析。

### 二、HRQL 研究结果有多大的临床意义

重点从 HRQL 变化与疾病的严重程度、时间变化, 以及相关临床测试结果的关系等方面加以考虑。HRQL 量表测试出的分值以及由此确定的界值, 应考虑对临床疾病和健康的判断意义, 是否高于或低于某一分值标准就属于正常或异常? 或者多大的分值变化才有临床价值? 因此要作出相应分析与评价。



### 三、HRQL 的测试结果是否有助于自己的病人处治

HRQL 测定应与临床医疗实践相结合, 在应用时一定要考虑自己的具体条件和其可行性, 以及是否有助于自己的临床实践。因此, 需特别考虑“PROGRESS”等诸方面因素: 其中 P (place) 为地点, 考虑研究地点与证据应用地点的异同; R (race) 为种族, 应考虑不同种族以及基因类型; O (occupation) 为职业; G (gender) 为性别; R (region) 为风俗信仰; E (education) 为教育程度; S (social economics status) 为社会经济地位; S (support system) 为社会支持系统等。

### 四、HRQL 的研究能否促进临床实践, 有利临床决策分析

药物治疗疾病, 既可以消除或缓解病症、减轻病损或治愈疾病, 促进病人康复, 同时也可能产生 ADRs, 严重的药物不良反应事件甚至会引起病人的死亡或病残。现代临床药物治疗试验, 严格设计者多采用随机双盲对照试验, 使用安慰剂或有效的药物进行对照, 盲法测量, 试验中出现疗效或副效反应时, 执行者及患者都不了解所用真正的药物是什么。因此, 在临床药物治疗性研究设计中, 如增加 HRQL 相应的特殊项目, 同时进行治疗反应的观察, 通常疗效与生存质量的改善具有一定的相关性。对于治疗试验的执行者, 尤其要注重 ADRs 的发生, 它可以通过 HRQL 的有关指标去早期发现。一旦发现用药过程中出现 HRQL 变差, 要及早检测其他生物学指标, 必要时停止或更换治疗药物, 以避免发生意外。因此, HRQL 测试是可以促进临床研究和临床医疗质量提高的。

(康德英 王家良)

### 附件 1 世界卫生组织生存质量测定量表简表 (WHOQOL-BRIEF)

填表说明: 这份问卷是要了解您对自己的生存质量、健康状况以及日常活动的感觉如何, 请您一定回答所有问题, 如果某个问题您不能肯定回答, 就选择最接近您自己真实感觉的那个答案。所有问题都只是您最近两星期内的情况。

请阅读每一个问题, 根据您的感觉, 选择最合适您的情况回答。

#### 1. 您怎样评价您的生存质量?

很差	差	不好也不差	好	很好
1	2	3	4	5

#### 2. 您对自己健康状况满意吗?

很不满意	满意	既非满意也非不满意	满意	很满意
1	2	3	4	5

下面的问题是关于两周来您经历某些事情的感觉

#### 3. 您觉得疼痛妨碍您去做自己需要做的事情吗?

根本不妨碍	很少妨碍	有妨碍	比较妨碍	极妨碍
1	2	3	4	5

#### 4. 您需要依靠医疗的帮助进行日常生活吗?

根本不需要	很少需要	需要	比较需要	极需要
1	2	3	4	5



5. 您觉得生活有乐趣吗?

根本没乐趣	很少有乐趣	有乐趣	比较有乐趣	极有乐趣
1	2	3	4	5

6. 您觉得自己生活有意义吗?

根本没意义	很少有意义	有意义	比较有意义	极有意义
1	2	3	4	5

7. 您能集中注意力吗?

根本不能	很少能	能	比较能	极能
1	2	3	4	5

8. 日常生活中您感觉安全吗?

根本不安全	很少安全	安全(一般)	比较安全	极安全
1	2	3	4	5

9. 您的生活环境对健康好吗?

根本不好	很少好	好	比较好	极好
1	2	3	4	5

10. 您有充沛的精力去应付日常生活吗?

根本没精力	很少有精力	有精力	多数有精力	完全有精力
1	2	3	4	5

11. 您认为自己的外形过得去吗?

根本过不去	很少过得去	过得去	多数过得去	完全过得去
1	2	3	4	5

12. 您的钱够用吗?

根本不够用	很少够用	够用	多数够用	完全够用
1	2	3	4	5

13. 在日常生活中您需要的信息都齐备吗?

根本不齐备	很少齐备	齐备(一般)	多数齐备	完全齐备
1	2	3	4	5

14. 您有机会进行休闲活动吗?

根本没机会	很少有机会	有机会	多数有机会	完全有机会
1	2	3	4	5

15. 您行动的能力如何?

很差	差	不好也不差	好	很好
1	2	3	4	5

16. 您对自己睡眠情况满意吗?

不满意	满意	既非满意也非不满意	满意	很满意
1	2	3	4	5



17. 您对自己日常生活事情的能力满意吗?

不满意	满意	既非满意也非不满意	满意	很满意
1	2	3	4	5

18. 您对自己的工作能力满意吗?

不满意	满意	既非满意也非不满意	满意	很满意
1	2	3	4	5

19. 您对自己满意吗?

不满意	满意	既非满意也非不满意	满意	很满意
1	2	3	4	5

20. 您对自己的人际关系满意吗?

不满意	满意	既非满意也非不满意	满意	很满意
1	2	3	4	5

21. 您对自己的性生活满意吗?

不满意	满意	既非满意也非不满意	满意	很满意
1	2	3	4	5

22. 您对自己从朋友那里得到的支持满意吗?

不满意	满意	既非满意也非不满意	满意	很满意
1	2	3	4	5

23. 您对自己居住地的条件满意吗?

不满意	满意	既非满意也非不满意	满意	很满意
1	2	3	4	5

24. 您对得到的卫生保健服务的方便程度满意吗?

不满意	满意	既非满意也非不满意	满意	很满意
1	2	3	4	5

25. 您对自己的交通情况满意吗?

不满意	满意	既非满意也非不满意	满意	很满意
1	2	3	4	5

26. 您有消极感受吗? (如情绪低落、绝望、焦虑、忧虑等)

没有消极感受	偶尔有	时有时无	经常有	总是有消极感受
1	2	3	4	5

27. 家庭摩擦影响您的生活吗?

根本不影响	很少影响	影响	有比较大影响	有极大影响
1	2	3	4	5

28. 您的食欲怎么样?

很差	差	不好也不差	好	很好
1	2	3	4	5



29. 如果让您综合以上各方面（生理健康、心理健康、社会关系和周围环境等方面）给自己的生存质量打一个总分，您打多少分？（满分为 100 分）\_\_\_\_\_分

您是在别人的帮助下填完这份调查表的吗？            是            否

您花了多长时间来填完这份调查表？            （        ） 分钟

您对本问卷有何建议：

感谢您的合作！

填表日期：

## 附件 2 健康状况调查问卷（SF-36）及其使用说明

健康状况调查问卷（SF-36）

数据核对用

以下问题是询问您对自己健康状况的看法，您自己觉得做日常活动的能力怎么样。如果您不知如何回答是好，就请您尽量给出最好的答案，并在本问卷最后的空白处写上你的注释与评论。

请✓一个答案

- |                         |         |                       |                          |
|-------------------------|---------|-----------------------|--------------------------|
| 1. 总体来讲，您的健康状况是：        | 非常好     | <input type="radio"/> |                          |
|                         | 很好      | <input type="radio"/> |                          |
|                         | 好       | <input type="radio"/> | <input type="checkbox"/> |
|                         | 一般      | <input type="radio"/> |                          |
|                         | 差       | <input type="radio"/> |                          |
| 2. 跟一年前相比，您觉得您现在的健康状况是： |         |                       |                          |
|                         | 比一年前好多了 | <input type="radio"/> |                          |
|                         | 比一年前好一些 | <input type="radio"/> |                          |
|                         | 跟一年前差不多 | <input type="radio"/> | <input type="checkbox"/> |
|                         | 比一年前差一些 | <input type="radio"/> |                          |
|                         | 比一年前差多了 | <input type="radio"/> |                          |

健康和日常活动

3. 以下这些问题都与日常活动有关。请您想一想，您的健康状况是否限制了这些活动？如果有限制，程度如何？

请在每一行✓一个答案

- |                                   | 有限制<br>限制很大           | 有限制<br>有些限制           | 毫无<br>限制              |                          |
|-----------------------------------|-----------------------|-----------------------|-----------------------|--------------------------|
| (1) 重体力活动，如：跑步、举重物、参加剧烈运动等        | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="checkbox"/> |
| (2) 适度的活动，如：移动一张桌子、扫地、打太极拳、做简单体操等 | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="checkbox"/> |
| (3) 手提日用品，如买菜、购物等                 | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="checkbox"/> |
| (4) 上几层楼梯                         | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="checkbox"/> |



(5) 上一层楼梯	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="checkbox"/>
(6) 弯腰、曲膝、下蹲	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="checkbox"/>
(7) 步行 1600 米以上的路程	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="checkbox"/>
(8) 步行 800 米的路程	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="checkbox"/>
(9) 步行 100 米的路程	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="checkbox"/>
(10) 自己洗澡、穿衣	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="checkbox"/>

4. 在过去四个星期里，您的工作和日常活动有无因为身体健康的原因而出现以下这些问题？

对每条问题请回答“是”或“不是”

是 不是

(1) 减少了工作或其他活动的时间	<input type="radio"/>	<input type="radio"/>	<input type="checkbox"/>
(2) 本来想要做的事情只能完成一部分	<input type="radio"/>	<input type="radio"/>	<input type="checkbox"/>
(3) 想要干的工作和活动的种类受到限制	<input type="radio"/>	<input type="radio"/>	<input type="checkbox"/>
(4) 完成工作或其他活动困难增多 (比如需要额外的努力)	<input type="radio"/>	<input type="radio"/>	<input type="checkbox"/>

5. 在过去四个星期里，您的工作和日常活动有无因为情绪的原因（如抑郁或者忧虑），而出现以下问题？

对每条问题请回答“是”或“不是”

是 不是

(1) 减少了工作或活动的时间	<input type="radio"/>	<input type="radio"/>	<input type="checkbox"/>
(2) 本来想要做的事情只能完成一部分	<input type="radio"/>	<input type="radio"/>	<input type="checkbox"/>
(3) 干事情不如平时仔细	<input type="radio"/>	<input type="radio"/>	<input type="checkbox"/>

6. 在过去的四个星期里，你的健康或情绪不好在多大程度上影响了您与家人、朋友、邻居或集体的正常社会交往？

请√一个答案：

完全没影响	<input type="radio"/>	<input type="checkbox"/>
有一点影响	<input type="radio"/>	
中等影响	<input type="radio"/>	<input type="checkbox"/>
影响很大	<input type="radio"/>	
影响非常大	<input type="radio"/>	

7. 过去四个星期里，您有身体疼痛吗？

完全没有疼痛	<input type="radio"/>	
稍微有一点疼痛	<input type="radio"/>	
有一点疼痛	<input type="radio"/>	
中等疼痛	<input type="radio"/>	<input type="checkbox"/>
严重疼痛	<input type="radio"/>	
很严重疼痛	<input type="radio"/>	



8. 过去四个星期里，身体上的疼痛影响你的工作和家务事吗？

- 完全没有影响 ☐
- 有一点影响 ☐
- 中等影响 ☐
- 影响很大 ☐
- 影响非常大 ☐

☐

您的感觉

9. 以下这些问题有关过去一个月里您自己的感觉，对每一条问题所说的事情，你的情况是什么样的？请圈出最接近您的情况的那个答案。

请在每一条问题后√ 出一个答案

持续的时间	所有的 时间	大部分 时间	比较多 时间	一部分 时间	一小部 分时间	没有这 种感觉	
(1) 您觉得生活 充实	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="checkbox"/>
(2) 您是一个敏 感的人	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="checkbox"/>
(3) 您情绪非常 不好，什么 事都不能使 您高兴	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="checkbox"/>
(4) 您心里很 平静	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="checkbox"/>
(5) 您做事精力 充沛	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="checkbox"/>
(6) 你的情绪 低落	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="checkbox"/>
(7) 你觉得筋疲 力尽	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="checkbox"/>
(8) 您是个快乐 的人	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="checkbox"/>
(9) 您感觉厌烦	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="checkbox"/>

(10) 不健康影响了您的社会活动（如走亲访友等）	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="checkbox"/>
<hr/>							
总体健康情况							
10. 请看下列每一条问题，哪一种答案最符合您的情况？							
	绝对正确	大部分正确	不能肯定	大部分错误	绝对错误		
<hr/>							
(1) 我好像比别人容易生病	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>		<input type="checkbox"/>
<hr/>							
(2) 我跟周围人一样健康	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>		<input type="checkbox"/>
<hr/>							
(3) 我认为我的健康状况在变坏	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>		<input type="checkbox"/>
<hr/>							
(4) 我的健康状况非常好	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>		<input type="checkbox"/>
<hr/>							

如果您有注释或评论，请写在下面：

非常感谢您的合作！

请按要求将这份表填好后交还给我们。



# 第二十章 突发性公共卫生事件研究设计与应急处治

## 第一节 突发性公共卫生事件概述

突发性公共卫生事件是一种突然发生的社会灾难，严重危害到人群整体健康水平和生活质量，甚至会给经济造成巨大损失，影响社会的安定。近年来，随着全球人口的不断增长和资源的逐步耗竭，突发事件的发生频率和强度进一步增加，其危害日显突出。突发事件由于自身的难以预测性，其影响除造成大量的人员伤亡、社会危害和环境污染外，还能引发一系列的心理卫生问题和后期效应。因此，已日益成为社会普遍关注的热点问题，并被许多国家列为重要的公共卫生问题之一。

### 一、概念

突发事件（emergency events）指突然发生、发展速度快且出乎人们意料的事情，这些事件的发生常难以应对，造成或者可能造成严重社会危害，必须采取非常规方法来处理，可被广义地理解为突然发生的事情。在公共卫生方面则称为突发性公共卫生事件，是指突然发生，可造成社会公众健康严重损害的重大传染病疫情、群体性不明原因疾病、重大食物和职业中毒以及其他严重影响公众健康的事件，需要卫生机构联合多方面力量，采取行动以紧急救援和处理事件。

突发性公共卫生事件范围很广，包括甲类传染病在某地首次出现，乙类和丙类传染病暴发，多例死亡、罕见的或已消灭的传染病出现，新发传染病，临床及病原学特点与原有疾病特征明显异常的疾病，各类自然灾害过程中及灾后疾病的暴发流行，医院感染事件，疫苗接种产生的群体癔症，生物化学恐怖事件，放射恐怖事件，核事故，营养缺乏性疾病流行以及其他非常规性公共事件等。表 20-1 列举了近一个世纪以来有代表性的突发公共卫生事件。

表 20-1 20 世纪以来发生的重大突发公共卫生事件

年份	事 件	危 害	原 因
1910 年	中国东北鼠疫大流行	死亡 42000 人以上	鼠疫杆菌
1930 年	马斯河谷烟雾事件	近 60 人死亡，且千余人患呼吸系统疾病	二氧化硫粉尘
1932 年	中国霍乱大流行	23 个省传播，病人 10 余万例	霍乱杆菌
1952 年	伦敦烟雾事件	1 万多人死亡	烟尘、二氧化硫
1953 年	日本水俣病事件	大量居民出现中枢神经系统中毒症状，60 多人死亡	食用被汞污染的鱼、贝等水生生物
1968 年	日本米糠油事件	中毒患者超过 1 万人，16 人死亡	食用多氯联苯污染的米糠油
1977 年至今	埃博拉出血热	累计发病 1000 余人，死亡 700 余人	埃博拉病毒
1986 年	切尔诺贝利核电站事故	299 万人受大剂量照射，1 万人死亡，数万人残疾	核泄漏
1988 年	上海甲肝大暴发	310 746 人发病，21 人直接死于本病	食用不洁毛蚶和饮食卫生习惯不良
1995 年	东京地铁沙林事件	5500 人受害，死亡 12 人，1000 人紧急入院	沙林释放中毒
2000 年	英国疯牛病事件	18 万人牲畜感染，100 多人死亡	疯牛病病毒
2003 年	非典型性肺炎（SARS）	全球累计报告病例 8000 多例，死亡 800 多例	冠状病毒变异株



## 二、分类

突发公共卫生事件强调的是一种紧急状态,紧急状态为“一种特别的、迫在眉睫或危险局势,影响全体公民,并对整个社会的正常生活构成威胁”。

突发公共卫生事件的分类有以下几种。

### (一) 依性质分类

按照突发公共事件的性质可将其分为以下四类,各类均涉及公共卫生问题。

1. 自然灾害 自然灾害(natural disaster)主要包括水旱灾害,台风、冰雹等气象灾害,火山、地震灾害,山体崩塌、滑坡、泥石流、地面塌陷等地质灾害,风暴潮、海啸等海洋灾害,森林火灾和重大生物灾害等,及由此而产生的对居民的直接与间接灾害。如2002年全世界发生洪水44起,地震20起,台风、风暴16起,森林大火、生物入侵各3起,严重山体滑坡5起。

2. 事故灾害 火灾、化学事故及其他事故灾害。主要包括民航、铁路、公路、水运等重大交通运输事故,工矿企业、建设工程、公共场所发生的各类重大安全事故,造成重大影响和损失的供水、供电、供油和供气等城市生命线事故以及通讯、信息网络、特种设备等安全事故,核与辐射事故、危险化学物品事故,重大环境污染核生态破坏事故等。例如,印度Bhopal化工厂毒气泄漏;核电厂事故,如美国三里岛、前苏联切尔诺贝利核泄漏。

3. 公共卫生事件 主要包括突然发生,造成或可能造成社会公众健康严重损害的重大传染病疫情、群体性不明原因疾病、重大食物和职业中毒、重大动植物疫情以及其他严重影响公众健康的事件。

4. 社会安全事件 主要包括涉及公众安全的重大刑事案件、涉外突发公共事件、恐怖袭击事件、经济安全事件以及规模较大的群体性事件等。例如,恶意攻击暴力恐怖事件如2001年美国“9·11”事件;日本奥姆真理教“地铁沙林毒气”事件;我国南京地铁“毒鼠强”事件。

### (二) 依严重程度分类

按照引起危机的严重程度可以分为三类。

1. 一般性突发公共卫生事件 指对人身安全、社会财产及社会秩序影响相对较小的突发公共事件,由事发地所属市、县级人民政府处置。

2. 重大突发公共卫生事件 指对人身安全、社会财产及社会秩序造成重大损害的突发公共事件,由省人民政府处置。

3. 特大突发公共卫生事件 指对人身安全、社会财产及社会秩序造成严重损害的突发公共事件,由省人民政府处置或者省人民政府报请国务院、国务院有关职能部门协调处置。

## 三、特征

一般来说,突发公共卫生事件具备以下特征:

### (一) 突发性和意外性

突发公共卫生事件往往是在人们意想不到的时间、地点突然发生,不易预测,有的甚至不可预测,常常使人措手不及,难以防范。

### (二) 群体性

突发公共卫生事件常常同时波及许多人,甚至整个工作或生活的群体,在公共卫生领域发生,具有公共卫生属性。

### (三) 对社会危害的严重性

由于突发事件发生突然、涉及面广、损失巨大,往往引起社会的极大关注和惊恐不



安,影响经济、政治、军事等多方面。2003年SARS的发生,对中国和全球的影响就是一个典型的例证。

#### (四) 处理的复杂性

由于突发事件发生的原因在开始阶段多不明确,缺乏充足的有效的信息,对于现场抢救、控制和转运救治,原因调查和善后处理常涉及多个部门、系统,增加了综合协调处理的难度。

#### (五) 影响的深远性

突发事件发生突然,一般持续时间不长,但是影响不仅限于事发当时,许多突发事件还具有继发效应和远期效应。如洪水过后,肠道传染病流行;放射事故发生多年后,人群白血病发病率增高等。

## 第二节 突发性公共卫生事件调查研究

### 一、调查研究的意义

根据突发公共卫生事件的突发性、群体性及现场复杂多样性等特点,对其采用现场流行病学调查方法。现场流行病学调查是突发公共卫生事件研究的起点,也是处理好某一具体突发公共卫生事件的关键,决定着后续工作的成败。通过调查研究可以收集到较为完整的信息,提出有效的控制措施。

### 二、暴发调查

#### (一) 暴发调查的概念与目的

暴发(outbreak)是指在一个局部地区或集体单位的人群中,短时间内突然发生许多相同症状的病人称为暴发。针对暴发的流行病学调查称为暴发调查(outbreak survey)。暴发调查是突发事件调查的基本形式之一。疾病暴发的原因可能是通过共同的传播途径感染,或有共同的传染源,或有共同的致病因素的暴露,大多数病人出现在该病的最长与最短潜伏期内。暴发有传染病暴发(如呼吸道传染病),也有非传染病暴发(如皮炎暴发);有一次性集中暴发,也有连续蔓延的暴发,表现形式多种多样。对非传染病暴发调查的思路、方法及步骤与对传染病的暴发调查大同小异。其一般步骤见图20-1。

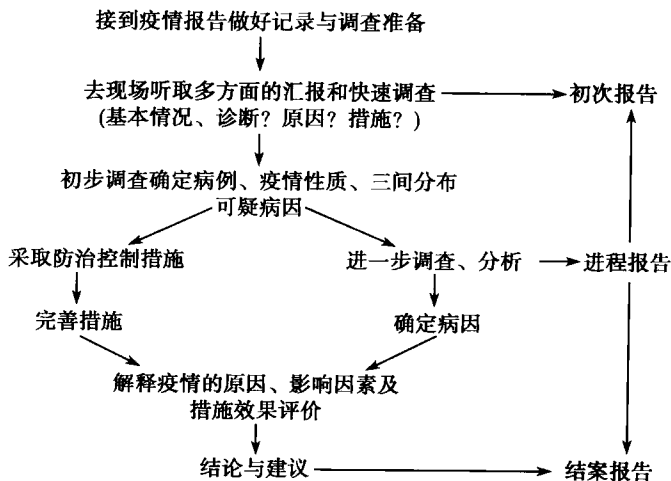


图 20-1 暴发调查步骤



暴发调查的目的主要是确定疫情性质，如传染病（甲类、乙类、丙类）、非传染病、新发疾病或病因不明疾病的暴发等；查清暴发危害程度即疾病三间分布；查明病因和暴发影响因素如传染病的传染源、传播途径等；确定高危人群，并予以有效保护；制定切实措施，控制疾病暴发和流行，并总结经验教训，避免此类事件再次发生。

## （二）暴发调查步骤

在暴发调查中，要遵循边调查边采取措施的原则。

1. 组织准备 组织现场调查组进行现场调查，应明确调查目的和具体调查任务。现场调查组应由相应的专业人员组成，一般应该包括流行病学、实验室和临床医学等专业人员，必要时还应增加其他卫生专业和管理人员。

现场调查前应准备必需的资料和物品，一般可以包括：相关调查表（有时需要到现场根据初步调查结果现场设计调查表用于调查）和调查器材、现场预防控制器材、采样设备和相应的采样试剂、现场联系资料（联系人和联系电话）、电脑、照相机和个人防护用品等。

2. 确定疾病暴发的存在 根据疾病发生的时间分布特点、波及范围、既往该类疾病的发病频率来判断是否发生了暴发。如果病例发生的时间和地区集中，数量明显多于过去同期水平，即可认为暴发存在。

3. 收集病例 通过有关媒体宣传，利用登记报告系统对病例进行监测外，还可采用医师询问调查、电话调查、入户调查、病原体分离与培养及血清学调查等方法主动发现病例。

（1）首例病人调查：主要是发病时间、发病前后的活动情况等。

（2）发病情况调查：登记每个病人的发病情况。

（3）标本收集、送验和保存：病人标本、接触者标本、可疑环境因素标本等要及时采集、迅速送验，并妥善保存。

4. 明确诊断 根据现场调查所发现的暴发性病例的病史、临床特征、实验室和某种（些）特殊的检查资料，并结合流行病学的资料综合分析，以作出疾病的诊断，为有效防治提供科学的依据。

通过现场调查，在疾病诊断方面通常有三种情况：

（1）确诊：有充足的可靠的确诊证据，如病原学依据，像霍乱、鼠疫……

（2）基本确诊：根据临床与流行病学特点，以及一些特殊的检查证据进行诊断，如像 SARS 初期诊断为急性传染性呼吸窘迫综合征，但无病原学依据。

（3）诊断不明：对于暴发疾病仅能依据临床表现和病损特点而诊断为某种系统的疾病，但病因不明，尚待进一步研究确诊，但应制定在当时条件下相对合理的诊断标准，且应不断完善，如 SARS 的诊断标准经多次完善。

无论诊断明确与否，作为突发公共卫生事件，从流行病防治而言，需要深入调查。

5. 疾病三间分布的描述与分析 理论上，要完成疾病的三间分布的描述与分析，需要等到流行终止以后，然而暴发调查时需要尽快提出假设及相应的预防控制措施，因此，应根据已收集到的资料及时进行

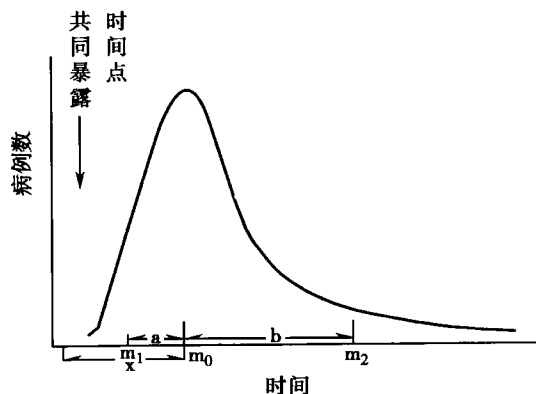


图 20-2 一次暴露发病日期示意图



分析。

(1) 时间分布：一般以  $1/8 \sim 1/4$  潜伏期为时间单元。绘制发病直方图或发病曲线。

1) 潜伏期的推算：如果疾病的病因和潜伏期是未知的，利用流行曲线可推算疾病的平均潜伏期。假设某次疾病暴发的流行曲线如图 20-2。

设  $m_1$  为发生 16% 病人数的时间， $m_0$  为发生 50% 病人数的时间， $m_2$  为发生 84% 病人数的时间， $m_0 - m_1 = a$ ， $m_2 - m_0 = b$ ；由共同暴露日期至  $m_0$  为  $x$ （平均潜伏期）。由于一次共同暴露的流行曲线类似对数正态曲线，故：

$$\begin{aligned} \lg x - \lg(x-a) &= \lg(x+b) - \lg x \\ \frac{x}{x-a} &= \frac{x+b}{x} \quad \text{化简得: 平均潜伏期} \\ x &= \frac{ab}{b-a} = \frac{(m_0 - m_1)(m_2 - m_0)}{(m_2 - m_0) - (m_0 - m_1)} \end{aligned}$$

该公式适用于潜伏期短，暴露为一次共同来源的暴露，流行只有一个高峰，病例集中在该病的最长与最短潜伏期的范围内的疾病流行。

2) 暴露日期的推算：暴露日期的推算对于传染来源的调查是至关重要的。暴露日期推算的依据是潜伏期。如果病原是已知的，暴发为同源性暴发，则暴露日期的推算有两种方法：一是从中位数病例的发病日期向前推一个平均潜伏期；另一种方法是从首例发病日期向前推一个最短潜伏期，再从最后一例病例向前推一个最长潜伏期，该两个时点之间即为同源暴露的可能时间。

3) 传播或流行特点：①同源一次暴露：病例常集中在一个最短、最长潜伏期。②同源多次暴露：同源多次暴露的发病曲线具有两个以上高峰；连续性暴露的发病曲线具有两个以上高峰或持续高峰。

(2) 地区分布：绘制标点地图或地区分布图；注意河流、公路、铁路、食品供应区等特征。有时可获得关于病原体的来源、传播途径及可能传播媒介的重要线索。19 世纪中叶，英国医师 John Snow 即利用标点地图的方法成功地查明伦敦宽街霍乱流行是由于宽街的一口水井被污染所致。

(3) 人群分布：按年龄、性别、职业、劳动情况等进行分析。注意早期病例（初发病例）的人群分布情况。通过分析不同人群疾病罹患率的差异，可能查找出危险人群及暴露因素的线索。如某职业人群的罹患率高，则危险因素暴露可能与该职业有关；在某食堂就餐者的罹患率高，则可能提示该食堂某种食品被污染等。

6. 提出假设与检验假设 对疾病暴发进行调查时，应及时提出假设并检验假设。提出假设要大胆，要具有流行病学的敏锐性，要善于从病例描述中找线索，从特殊事例中找原因；检验假设要严谨，要具有流行病学的逻辑性，应用分析流行病学的方法进行检验。

提出假设除了从典型事例中找线索外，最重要的是在初步描述性研究（即上述对病例的三间分布的描述）的基础上，仔细审核资料，结合分析临床、实验室和流行病学特征，提出可能的假设。假设应包括致病因子、传染源、传播途径及暴发原因等。由于暴发调查的特殊性，提出暴发原因的假设也可能是超常规的，可根据调查者的推测，少量的发现等随时提出假设。

检验假设一般采用病例对照研究，亦可采用病例对照研究的一些衍生类型，如病例—队列研究，累积病例对照研究等。在某些情况下，可采用队列研究，特别是历史性队列研究。有关暴发原因的假设应能解释本次发现的绝大多数的流行特征。比如应从病人的既往暴露史中找出假设的致病因子的暴露。



7. 明确高危人群 在对致病因子、传染源、传播途径及暴发原因提出假设并初步检验之后,即可提出继续受威胁的高危人群。及时对高危人群采取保护性措施,阻止流行进一步蔓延,是暴发调查的一项十分重要的任务。如初步调查结果提示:只有居住在某一个社区,拥有特别供水系统的居民才发病,暴发原因假设是该供水系统受到污染,则该供水系统所供居民即为发病的高危人群,同时要注意该供水范围内个别临时流动人口。在调查初期,对高危人群的准确划定往往是困难的,此时,宁愿将怀疑的对象广一点,以确保避免对高危人群保护的遗漏。

8. 验证假设 经初步调查后提出的假设,再经分析流行病学方法进行检验后,应进一步作最后验证。验证假设的根据是事实,是实践。常用的验证假设的方法有两种。一种是全面收集与本次暴发有关的临床、实验室和流行病学资料,分析本次暴发在这些方面的特点是否与假设应该有的特点相符。例如,如果假设本次暴发为一次细菌性食物中毒暴发,其原因是某食堂的某种食品被污染,则这些病人的临床表现应与细菌性食物中毒相符,病例集中在该种细菌性食物中毒的最长与最短潜伏期之间,所有病例均有进食该食物的历史,从病人呕吐物、排泄物及可疑食物中分离出相应的细菌或毒素等。如果满足这些条件则假设成立。另一种方法是做干预试验,可以是标准的流行病学实验,也可以是类实验。如在上述例子中,在对可疑食物采取措施后,经过一个疾病的潜伏期,流行即告终止,这就是验证假设的一个很好的事实,但也应注意流行是否自然终止。

9. 及时采取控制措施 在暴发调查中,制定与采取控制暴发的措施应与假设的提出同步,不必等待假设的验证,以免延误时机。假设一旦提出,则应针对假设立即采取措施,假设修改后,控制措施亦应同步修改。控制措施的采取,在暴发调查中有两大重要意义,其一是迅速控制流行的蔓延,保护受威胁人口的生命与健康,其二是通过评价控制措施的效果,可以反过来验证假设的成立。控制措施一般是针对疾病流行的三个环节,重点是针对致病因子的措施。对控制措施效果评价的一个最直接、最客观的指标就是日罹患率的下降。然而对该指标的解释要慎重,因为日罹患率的下降可能有多种原因:①控制措施开始发挥作用;②所有易感高危人群均已受感染,易感者人数减少;③诊断标准、发现病例的方法不一致;④某种自然现象减少了病原因子的来源,如气温下降可使蚊虫活动减少等。如果暴发的真正原因没有被消除,那么在稍后阶段将可能出现再次暴发。另外,在评价控制措施效果时,应考虑疾病的潜伏期。如果疾病的潜伏期长,或者病例诊断和病例发现方法得以改进,即使控制措施有效,日罹患率在短期内仍可能持续增加。

10. 总结报告 总结报告应成为疾病暴发调查的重要一环。要认真、全面、准确、实事求是。暴发调查报告的内容一般包括:①背景材料:地理位置、气候条件、人口统计状况(人群构成)、社会经济状况、卫生服务的组织、疾病监测系统、针对疾病流行的准备及平时的疾病患病率等。②历史资料:以前当地或其他地区同种疾病的流行情况,本次暴发首例病人发现的经过。③调查方法:病例确定的诊断标准,流行病学调查中所用的调查表、调查队伍、调查方法(个案调查、描述性调查、回顾性调查等)、实验标本的采集及所用的实验室技术。④资料分析:临床资料包括症状和体征的发生频率、病程、诊断与鉴别诊断、预后等;流行病学资料—包括疾病发生的方式、三间分布、显性感染与隐性感染的比例、传染来源、传播途径及影响传播的因素等;实验室资料—包括可疑病原的分离、血清学实验结果及实验结果的意义。并且应对资料进行合理的解释,包括暴发的综合描述、病因假设的形成与检验。⑤控制措施:控制暴发的策略与方法,控制措施的结果的评价,包括对费用效益及费用效果的评价,预防类似事件发



生的建议措施等。

### 三、研究设计的要点

由于突发公共卫生事件时限上要求急迫，因此，必须尽可能制定出完整的调查研究计划。

#### (一) 原则

1. 时效性原则 当突发公共卫生事件发生时，应及时到达现场调查。争取时间，救治病人，隔离密切接触者，及时采集到血清标本或可疑污染物标本。在很多情况下，现场调查的“迅速”比“完整”更加重要。如有毒化学品引起的急性中毒事件，出现死亡病例的急性传染病暴发流行等。现场调查应及时分析，为卫生行政部门的决策提供可靠依据。

2. 优先性原则 突发公共卫生事件调查的一项重要内容，就是提出现场干预措施及其实施。如对不明原因疾病流行进行调查时，救治病人，切断传播途径，以保护易感人群。有时需要及时杀灭蚊虫，立即中止可能污染的水体供水等。总之，流行病学现场调查要从实际出发，解决那些需要优先解决的问题。

3. 科学性原则 突发事件调查不是简单的收集资料、实施干预的活动。要根据突发事件的可能的性质，作出调查的设计与安排。

内容应包括采用的调查方式，选择调查对象及其项目，问卷的拟订，现场采样，优先实施的干预、调查资料的整理、统计分析的方法等。

调查结果的评价，可能获得的调查结论，以及尚需要进一步弄清不明因素等。在此基础上方能较科学地上报卫生行政机关或通报媒体。总之，现场调查的设计、调查的过程及现场调查报告都必须严谨科学的。

#### (二) 研究设计的要点

突发事件的调查研究设计的要点如下：

1. 调查目标及调查项目 突发事件调查有两个基本目的，一是搞清楚事件发生的原因——事件诊断。二是提出干预措施——事件处理。进行突发公共卫生事件调查一般是回应事件发生地组织或个人的报告，或上级有关部门指派的任务，而不是进行纯理论的流行病学研究。

事件诊断：对突发公共卫生事件进行现场流行病学调查以确定事件发生的原因。突发卫生事件有些发生原因较明确，如工厂某种有害气体泄漏，运送毒品的船只因交通事故致毒品污染水体；但在大多数情况下，突发公共卫生事件的原因不明确，要经过严谨、科学的现场调查及实验室证据，方能得出正确的结论，如调查某一集体食物中毒究竟是化学性污染还是细菌性污染，某人群突发原因不明疾病究竟是地方病还是传染病，这些都要通过现场流行病学调查，进行“事件诊断”，以便确定事件的原因，采取干预措施。

事件处理：对突发公共卫生事件进行现场流行病学调查，采取干预措施，控制事态发展，并检查其效果。例如对传染病突发疫情的处理，是针对传染源、传播途径还是易感人群采取干预措施。应急接种是否有效。

2. 确定调查现场 现场调查地点指发生突发公共卫生事件的地区，一般易于确认，但有时对于调查范围的大小难于确定。调查范围过小，可能不足以得出正确的结论，甚至由于防治措施的范围覆盖面较小而留下疾病再次暴发的隐患；调查范围过大，则可能浪费人力、物力和财力，特别是可能耽误处理时间。调查点的确认要做到三个有利于：有利于达到调查目标；有利于实施现场调查；有利于节约人力、物力、财力和时间。在对疾病暴



发作调查时，有些经验是值得重视的，如有时疾病发生在社区中一个独特的地区或农村某一特殊地形，如果你能发现，则可直接获得很多关于病原体和暴露特征的有价值的信息。供水系统、牛奶供应、常年风向、建筑物间的气流、地理特征及传播媒介的生态习性在确定现场调查的地点时扮演重要角色。

3. 确定调查对象 就疾病的现场流行病学调查而言，在基本明确了病人的特征（年龄、性别、职业），发病时间，发病地点及临床表现后，则可为调查对象的确定提供足够的信息。如有些疾病首先侵犯一定年龄组或种族；患某种疾病与某类职业有关；有些疾病只发生在居住同一社区、拥有同一供水系统或空调系统的人群中，或只有学校中某些学生，工厂中某些工人，接触水田的某些农民发病，或许只有那些到餐馆就餐的人才可发病。因此，要视发病者的共性情况确定调查对象。特别是具有传染性流行性的突发性疾病更要对其接触过的高危人群或易感人群扩大追踪调查。

4. 确定调查方法 流行病学现场调查的方法来源于传统的流行病学调查研究方法，其要点是迅速行动，制定正确的实施方案，尽快完成调查工作。调查方法应适应现场工作的实际需要，同一现场可采取不同的调查方法，同一调查方法也可以适用于不同的现场调查，因此，如何选择最有效的调查方法，就成为调查计划中的重要内容。

5. 制定调查计划表 制定调查计划表，涉及调查工作的整个内容及其总体进展的要求，详见表 20-2。

表 20-2 调查计划表

调查计划项目	内 容
调查目的	为什么调查，需要了解什么，调查结果有何用途
调查方法	描述法，分析法，（病例对照或队列研究）
调查地区	事件波及的范围
调查对象、样本	事件累及的人群，样本量的大小（抽样或普查）
调查时间、地点	调查起止时间，调查会时间、会址
调查项目	备考项目、特征项目，其他
实验室	理化分析、微生物分离与鉴定
分析方法	统计分析的项目、运用的软件
提交调查报告	报告的形式、份数、内容和对象
调查费用	各项开支细节、总开支
调查人员	调查人员、负责人姓名、资历

### 第三节 突发性公共卫生事件的预防和处理

突发公共卫生事件的预防与应急处理包括平时的预防和战时的处理两方面。平时的预防主要是做好对突发公共卫生事件的监测和预警；战时的处理主要是做好对突发公共卫生事件的就地处理和控制在蔓延。突发公共卫生事件的预防与应急处理是一项系统工程，它需要政府的统一领导和统一指挥，也需要各部门、各单位的广泛参与，各负其责，相互配合。其工作流程如图 20-3。



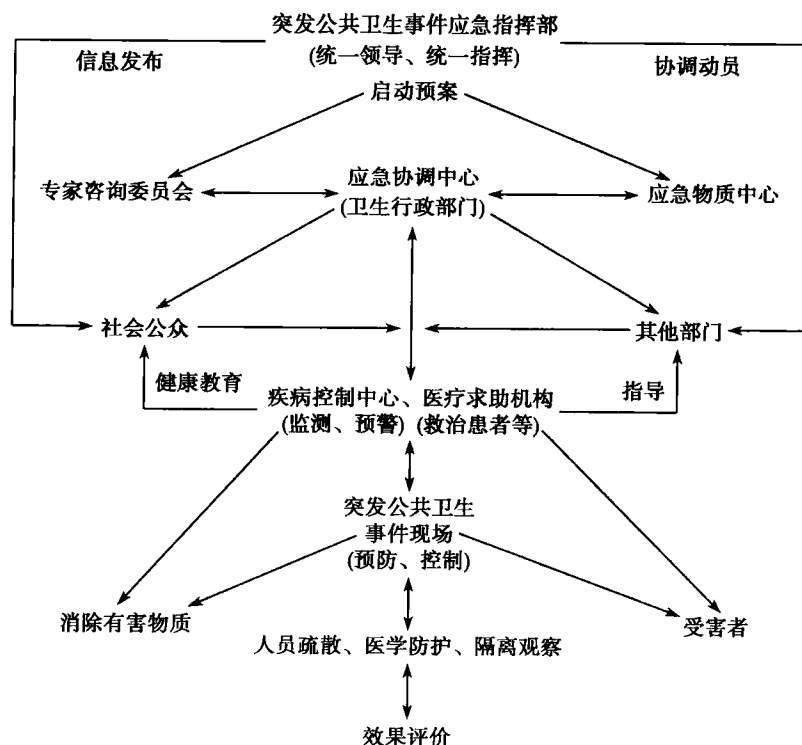


图 20-3 突发公共卫生事件的预防与应急处理工作流程

## 一、预防措施

### (一) 做好突发公共卫生事件的科学监测

通过各种监测网络收集突发公共卫生事件的相关信息，并运用科学的方法进行分析处理，及时掌握突发公共卫生事件的动态，为预防与应急处理突发公共卫生事件提供科学依据。

### (二) 拟定突发公共卫生事件应急反应计划

在科学监测的基础上，根据准备全面的监测信息对可能发生的突发公共卫生事件拟定应急反应计划。应急反应计划的拟定要遵循灵活性、协调性、开放性、预见性、权威性、可持续性、可操作性、可扩展性的原则，尽可能完善有效。

### (三) 制定突发公共卫生事件应急预案

突发公共卫生事件应急预案的制定，应结合当地实际进行。制定预案时应坚持政府负责、统一指挥，科学防治、技术主导，属地管理、职能保障，满足投入、讲究效率，危机处理、融入常态和维护社会稳定、保障经济发展的原则，使应急预案实用有效。应急预案一般应当包括以下主要内容：

1. 突发公共卫生事件应急指挥部的组成和相关部门的职责；
2. 突发公共卫生事件的监测与预警；
3. 突发公共卫生事件信息的收集、分析、报告、通报制度；
4. 突发公共卫生事件应急处理技术和监测机构及其任务；
5. 突发公共卫生事件的分级和应急处理工作方案；
6. 突发公共卫生事件预防、现场控制，应急设施、设备、急救药品和医疗器械以及其他物质和技术的储备与调度；



## 7. 公共卫生突发事件应急处理专业队伍的建设和培训。

### (四) 及时对突发公共卫生事件进行科学预警

对突发公共卫生事件及时进行科学预警,可以做到早发现、早预防,有利于避免突发公共卫生事件的发生和防止事态的扩大。虽然突发公共卫生事件是突发的,但是,在发生前一般总有迹象显现,只要我们不断积累经验,及时注意捕捉信息,并对此进行科学分析,就能够作出及时、准确的科学预警报告。

## 二、应对措施

突发公共卫生事件的应急处理包括组织领导、应急预案的启动和执行。与通常认为的相反,对突发事件的应急并不是一次性的临时突击活动,而是一项长期的任务,其目的是增强国家和社区有效处理各种灾难性事件的能力,使灾难对社会造成的损失降低到最低程度,并尽快使救援工作有序过渡到正常的持续发展。应急和恢复工作通常需要一系列的部门合作构成一个统一的系统,并迅速开展各项工作,包括通讯与警报、调查与救援、卫生与医疗、交通与运输、治安与安全、社区生活与服务设施、信息与资源管理、人员撤离与危险控制等。

### (一) 应急处理原则

1. 以人为本,减少危害 切实履行政府的社会管理和公共服务职能,把保障公众健康和生命财产安全作为首要任务,最大限度地减少突发公共事件及其造成的人员伤亡和危害。

2. 居安思危,预防为主 高度重视公共安全工作,常抓不懈,防患于未然。增强忧患意识,坚持预防与应急相结合,常态与非常态相结合,做好应对突发公共卫生事件的各项准备工作。

3. 统一领导,分级负责 在有关行政部门统一领导下,建立健全分类管理、分级负责,条块结合、属地管理为主的应急管理体制,并充分发挥专业应急指挥机构的作用。

4. 依法规范,加强管理 依据有关法律和行政法规,加强应急管理,维护公众的合法权益,使应对突发公共事件的工作规范化、制度化、法制化。

5. 快速反应,协同应对 加强以属地管理为主的应急处置队伍建设,建立联动协调制度,充分动员和发挥乡镇、社区、企事业单位、社会团体和志愿者队伍的作用,依靠公众力量,形成统一指挥、反应灵敏、功能齐全、协调有序、运转高效的应急管理机制。

6. 依靠科技,提高素质 加强公共安全科学研究和技术开发,采用先进的监测、预测、预警、预防和应急处置技术及设施,充分发挥专家队伍和专业人员的作用,提高应对突发公共事件的科技水平和指挥能力,避免发生次生、衍生事件;加强宣传和培训教育工作,提高公众自救、互救和应对各类突发公共事件的综合素质。

### (二) 应急处理措施

1. 现场急救 严重的突发事件会造成大量病员或外伤,因此在突发事件发生的最初几个小时,最紧迫的任务就是进行现场急救。面对突发事件引起的大量伤亡,营救人员首先应做的是对伤亡人员按伤势轻重、受伤类型及估计的疗效进行初步分类,通常的做法是分别在死亡、重伤、中度受伤和轻伤者的手臂上围一黑色、红色、黄色和绿色纱布标明,分开处理和救护。突发事件造成了严重伤亡时的处理主要包括3个阶段:搜寻、营救及急救治疗;依靠附近医疗机构建立现场急救站,进行就地治疗;根据需要,将部分伤员分类转往最近的医院或其他合适的医院。医院接到大量伤员后应尽快添加床位,做好医院间病人的重新分配。最初治疗的原则是:抢救生命,防治严重的并发症。由于人力、物力的有限,宜推广标准化的简单治疗方案,如大清创术,一级伤口的延期闭合和使用夹板而不用



石膏固定等。

2. 公共卫生管理 在搞好现场急救的同时,搞好紧急情况下的公共卫生管理,对防止疫情发生十分必要。主要工作包括:保证供水安全,建议增加余氯量和水压,当水源可能被有毒物污染时,应积极寻找备用水源;检测餐具、厨具,监督食品加工者的个人卫生,搞好食品原料的防鼠、防虫和防霉变工作;修建临时厕所、提供洗手、沐浴等基本卫生设备;设立临时垃圾处理场,清理废品、垃圾和各种散落在环境中的有毒物质,处理建筑瓦砾、渣土,焚烧或掩埋动物尸体;开展爱国卫生运动,使用杀毒剂消灭蚊、虫等传媒介质;加强疫苗接种,保护体弱多病者;尽可能快地将死难者的尸体运往停尸房,传染病患者的尸体,应在鉴定后,及时深埋或火化。

3. 通讯与运输 充足的紧急通讯设备对于与卫生机构和突发事件工作人员保持快速可靠的联系发挥着必不可少的作用,对于参与救援工作的政府、非政府组织、私人和国际机构也同样至关重要。突发事件中,当电话线路遭到破坏时,应充分利用好如下几种通讯方式:无线电通讯、无线电寻呼服务、卫星通讯、电子邮件服务和电报通讯等。在紧急情况下,卫生部门应该协调好与国家后勤运输部门的关系,明确运输工作的负责单位尤为重要。可被紧急调动的运输力量有:公众运输机构、军队和警察运输力量、国有和私人客运公司,货运公司,轮船公司,航空公司和各种非政府组织。

4. 安排定居点 水灾和地震等突发事件可导致大量房屋的倒塌,使得许多人无家可归,此时卫生工作者必须完成的一项重要任务——搞好营地和临时定居点的建立和管理工作。工作思路是,首先根据突发事件的性质选择合理的地点,然后对宿营地进行恰当的规划,再做好各项服务设施的建设。

5. 寻求援助 所有地区和国家发生突发事件时必须尽量依靠自己的力量来完成救援和重建工作,但当情况危急,而当地或本国又缺乏相应的物资和技术的时候,就应积极将信息提供给外界或国际人道主义,寻求援助和支持。

(王素萍 徐德忠)

## 第二十一章 临床科研设计报告书的撰写

### 第一节 基本概念

撰写科研设计报告书或称标书是临床科研工作的重要工序,是保证临床科研工作顺利进行和取得成功的必要条件。那种认为不需要撰写临床科研设计报告书即可以做好临床科研,或认为临床科研是滑到哪里做到哪里看法都是不正确的。科研拨款制度改革后,国家、部门和单位都引入竞争的机制招标、基金合同制,因此在申请科研基金之前,都必须撰写完整的临床科研设计报告,也称科研投标的标书。

科研标书通常有下列要求:研究的课题要具有重要科学价值或效益;学术思想新颖,立题依据充足,研究方法和技术路线先进、科学、可行;有良好的专业科研工作基础,并具有深入开展研究工作的基本条件;研究的内容先进,目标明确,可望在2~4年内取得预期的结果;经费预算适当。

临床科研设计报告的撰写要符合投标的要求,科研基金会的投标课题设计报告书都必须经过同行专家的评议和审批。一份完整的临床科研设计报告书应包括下列项目:立题依据、科研目的、科研假设、科研设计、科研实施的地点、研究对象、样本大小计算、干预措施、研究因素和研究方法、资料收集和分析、预期结果、研究质量的控制、预试验、时间进度、经费预算等。

### 第二节 研究设计报告书的基本内容与撰写方法

#### 一、立题依据

立题依据应提供课题的背景资料(research background)。这是临床科研设计报告书中的重要部分,是它的灵魂,占据篇幅最大。书写之前要有相当多的时间查阅文献资料,要利用各种检索工具,包括Cochrane光盘等,需要详细掌握该研究领域的国内外近年来的信息和研究动向,才能写好立题依据。

首先必须阐明所研究的疾病是否为常见病、多发病,为危害人民健康较大的疾病,这是选题的原则之一,因此需要将所研究疾病的疾病负担阐述清楚,可参考本书第二章。

第二,要阐明该病目前研究现状,国内外的研究动向,尚存在什么问题,从而可以了解你所研究的课题是处于国际上的先进水平还是国内领先水平,你所研究的课题和该病研究所存在的问题有何联系,是解决这些问题中哪些具体问题。

第三,需要讲清楚你所研究课题的目的、意义,讲清你所研究的临床问题是什么?是有关诊断方法、治疗方面或病因研究或预后的研究?文献上对该领域的论述如何?也就是研究该问题的理论意义和实践意义是什么?国内、外对该问题研究的深度和广度怎样?哪些方面已获得结论,哪些方面尚有争议,有待于进一步研究。本课题主要是解决哪方面的问题?这些问题的解决对临床上有何理论方面和实践方面的意义和价值?有何经济效益和社会效益?如果已有一定的科研假设,应当详细描写假设的科学依据等等,从而说明该研究课题有无研究的价值。

总之,立题依据这部分必须将该课题研究的目的意义、重要性、研究的理论意义和实



践意义，以及处于国际和国内的水平讲清楚，使人们了解研究该课题的必要性和重要性。

## 二、研究目的 (research objectives)

为了引人注目，必须将研究目的 (research objectives) 用最简洁的文字列出。例如：“调查再生障碍性贫血的发病率和发病因素”，“比较三种铁剂治疗缺铁性贫血的疗效和不良反应”，“评价血清运铁蛋白受体在铁缺乏症中的诊断价值”。亦可分主要研究目的和次要研究目的逐一系列写清楚。

## 三、科研假设

科研设计报告中科研假设的撰写非常重要，因为整个科研的过程就是论证科研假设的过程。科研假设 (research hypothesis) 的书写，可分无效假设 ( $H_0$ ) 和备择假设 ( $H_1$ ) 分别列行书写，例如，“雷公藤多苷治疗成人晚发自身免疫性糖尿病的疗效研究—随机、安慰剂对照、双盲临床试验”采用治疗前后 C 肽差值作为主要观察指标，其科研假设书写如下：

$H_0$ ：雷公藤多苷组治疗前后 C 肽的差值 = 对照组治疗前后 C 肽的差值

$H_1$ ：雷公藤多苷组治疗前后 C 肽的差值  $\neq$  对照组治疗前后 C 肽的差值，假设雷公藤的疗效优于对照组 30% (假设值)

## 四、设计方案 (design)

描写临床研究课题设计的基本方案常需应用下列名词：干预试验或临床试验，随机对照、非随机同期对照或历史性对照，前后对照、交叉对照试验或安慰剂对照，单盲或双盲，横断面研究或诊断试验研究，病例对照研究或队列研究 (详见本书第九章)，成本-效果分析或成本-效用分析等。

为了表达清楚，可将设计流程图列出，设计流程图的绘制应结合课题参照规定格式绘制 (详见本书第七章)。

## 五、研究地点

研究设计报告中必须将研究地点 (setting) 写清楚，是在教学医院进行的研究课题，或在市级医院、区级医院或街道医院进行研究，是在住院病人中进行研究或门诊病人中进行研究，是否系多中心研究，如是，则需写清楚是哪些单位参加。因为研究地点不同，可影响研究结果，因而在研究质量与水平方面会有某些差异。

## 六、研究对象

课题研究报告需要写明目标人群、样本人群、纳入标准、排除标准及病人入组时的一般资料等等，需要描写研究对象 (patients/participants)，即病人的来源，系从三级医院来或基层医疗机构中来，或是从人群中来？是门诊病人或是住院病人。这些病人是怎样选人作为研究对象的，是随机样本，或连续样本，还是随便选择的研究对象，或选用志愿者？选入病人的诊断标准是什么？是公认的标准或自己制定的。排除标准是什么？对人选的标准要做具体规定，包括性别、年龄、民族及一般临床特征。如设对照组，则需要描写对照组的来源和条件。如采用随机化方法分组，则要详细说明随机化的具体方法，是简单随机、区组随机或分层随机？是应用随机数字表或计算器上的随机数字？如何执行随机化方法。如采用配对方法来平衡观察组和对照组非处理因素，则需要说明配对条件和比例。若是人群中的抽样调查，则要描写该人群的人口资料，抽样人群占整个人群的比例等。最



后,尚需叙述为减少选择研究对象时的偏倚所采取的各项措施。

例如,前述课题“雷公藤多苷治疗成人晚发自身免疫性糖尿病的疗效研究”,其研究对象书写格式:

目标人群:湖南长沙地区处于非胰岛素依赖型糖尿病(NIDDM)的成人晚发自身免疫性糖尿病(LADA)患者。

样本人群:凡符合以下纳入标准,且志愿参加者。

纳入标准:①糖尿病诊断需符合1985年WHO糖尿病专家委员会提出的标准;②发病年龄 $>14$ 岁;③起病6个月内无酮症;④谷氨酸脱羧酶(GAD)抗体阳性或胰岛细胞抗体(ICA)阳性;⑤空腹C肽 $>200\text{pmol/L}$ 。

排除标准:①继发性糖尿病及线粒体基因糖尿病;②合并严重的系统性疾病及感染;③最近2个月服用免疫调节剂;④对雷公藤过敏;⑤白细胞减少和血小板减少;⑥肝、肾功能异常;⑦妊娠及哺乳期妇女。

## 七、样本大小的确定(sample size estimation)

正确估算样本量是临床科研设计中的一个重要问题。若样本量过少,检验效能低,往往容易得到假阴性结果,影响结论正确性;若样本量过大,会增加临床研究的困难,造成不必要的人力、物力、时间和经济上的浪费。样本大小计算就是在保证科研结论具有一定可靠性条件下,确定最小观察例数。

样本量计算方法可借助于公式或查表法。样本大小的估计取决于下列因素:①第一类错误概率 $\alpha$ ,有单侧与双侧之分, $\alpha$ 越小所需样本越大,一般取 $\alpha \leq 0.05$ 为宜;②第二类错误出现的概率 $\beta$ , $1-\beta$ 即为检验效能(把握度), $\beta$ 为单侧, $1-\beta$ 是证实备择假设 $H_1$ 正确的能力,一般 $\beta$ 定为0.1或0.2, $\beta$ 值越大,检验效能越低,样本数量也越小;③允许误差 $\delta$ 或差值,一般根据研究实际设计要求,由研究者自行规定,此规定应合理;④总体标准差或总体率 $\pi$ ,一般是查阅文献或做预试验所得,因此在研究设计报告中要将你的出处写清楚,还要将计算样本大小的公式列上。计算出的样本数量,还应考虑失访等因素,适当增加,写出需要的样本总数。

## 八、干预方法(intervention)

研究中的干预措施必需详细介绍,如治疗性研究中所使用的干预药物的药名,不仅要写出化学名,还要写出商品名,是何药厂产品,批号多少,有些中药还要写出产地。治疗方案要具体写清楚,如剂量、疗程、用药途径、注意事项都必须明确规定。治疗前的条件也要明文规定,如需要停用有影响的其他药物后多少时间才可进入试验等。在干预过程中遇到的不良反应如何处理,哪些药可以用,哪些不能用都要加以规定,以保证试验顺利进行。在什么情况下停止试验也要写清楚,以便于执行。

如用安慰剂对照,需要介绍安慰剂的制备情况,如何保证和研究药物一样,包括外形及味道等。采用盲法要讲清是单盲或双盲以及如何保证盲法的实施。

介绍干预方法时还需要介绍在干预试验过程中如何提高研究对象的依从性,拟采取什么措施来提高研究对象的依从性,采用什么测定方法衡量依从性等。

## 九、研究因素

研究因素(study factor)或称研究课题拟达到目标的因素或指标,应该明确无误。例如应用某一降血压的药物进行治疗高血压患者,研究的因素是治前治后的血压水平;防治心脑血管重要脏器的损伤;病残率或病死率等。又如研究吸烟可否引起肺癌的发病,研究因素是



吸烟与肺癌。所有研究因素一定要有准确的测量方法和指标。

所研究的观察指标有哪几项,为什么选择这些观察指标,多少时间随访一次,一共几次,如何记录观察结果。如观察指标是实验室项目应详细描写实验方法,包括所使用仪器的型号、生产厂家,以及试剂如何应用,多少剂量,实验操作方法,如果系成熟的实验方法,则应注明该方法的出处,如是研究者创造的或修改过的应写明操作步骤。如研究变量是暴露因素或危险因素,应写明这些研究因素的定义,如“吸烟”是采用 Doll 与 Hill 的标准,“月经过多”是采用自定的依据月经周期长短、持续时间、有否血块、应用卫生纸数量综合制定的标准,对这些研究因素应采用公认的定义。此外,还必须写明如何提高观察指标的准确度和可靠度,是否采用盲法判断结果,是否有质控措施。

对终点指标最好应用公认的判断标准。治愈、缓解、有效或生存率等都必须写出具体的判断标准。

## 十、资料收集和分析 (data collection and analysis)

资料 and 数据的收集方法应详细介绍,是通过医院已有的病史资料或制定调查表直接向病人询问调查?如是采集实验室数据,需要介绍标本采集方法和时间,以保证实验室数据采集的正确性,如被检对象某些情况可影响检测结果,则需要制定某些规定,控制这些影响因素。在科研设计报告书后应附有该课题研究所用的调查表和观察表,并且要附有填写这些调查表的须知,及计算机编码的说明。如进行面对面的询问调查,应注明培训合格调查员的方式与方法、考核标准,同时要制定如何避免询问者偏倚的若干措施,确保询问调查结果的真实性(效度)。如是临床经济学分析,需要说明这些费用数据的来源,直接成本和间接成本计算的依据和种类,研究结果的效果及效益和效用的指标以及计算成本-效果、成本-效益和成本-效用的方法等,对所有的数据处理和分析要注明采用什么统计学方法,包括两组变量的显著性检验是选用什么类别的单因素分析或多因素分析方法,使用的计算机类型以及有关统计学软件都应交代清楚。

## 十一、预期结果 (expected outcomes)

科研设计报告书中应阐明预期的研究结果,例如治疗的有效率、降低致残率或病死率的程度。如果课题完成后获得的研究结果和预期结果不相同,则需要分析其原因。预期结果还应包括能获得的经济效益和社会效益。对预期结果的评价标准也应写明。

## 十二、质量控制 (quality control)

科研设计报告书需要对课题研究的条件进行分析,包括研究人员的素质和经验,是否有良好的技术力量以保证能实施课题的研究。实验室的条件和仪器设备是否可以胜任此项研究。此外,尚需介绍为完成该项研究还有哪些薄弱环节会影响科研质量,特别是可能产生的偏倚有哪些?并提出克服这些薄弱环节和防止这些偏倚出现所采取的措施,从而保证科研实施过程的质量。

## 十三、预试验 (pilot study)

科研设计报告的实施,是否能获得预期结果,有许多问题在书写科研设计报告时往往难以准确预测,因此,为保证科研工作能按报告书上写的要求顺利进行,有必要安排一个预试验。对于预试验,在设计报告书中,亦要详细交代预试验的目的与要求、样本、方法以及期望解决的问题等。预试验是采用少量研究样本,按照设计报告书上所规定的要求进行操作,其主要目的是为了发现设计报告书上制定的各种实施项目是否切合实际,在预试



验中, 可对设计方案不完整的地方或不切实际的地方, 提供修改、补充和使之完善的信息依据。此外, 还可核实样本量的估计是否合适等。经过预试验发现问题, 然后再来修改科研设计报告, 使之更切合实际, 保证科研工作顺利进行, 以获得预期结果。

#### 十四、医学伦理道德 (ethics)

伦理道德问题是医务人员对被研究者的职业道德。所有以人为对象的研究必须符合《赫尔辛基宣言》和国际医学科学组织委员会颁布的《人体生物医学研究国际道德指南》的道德原则, 即尊重人格、公正和力求使受试者最大限度受益和尽可能避免伤害。伦理委员会与知情同意书是保障受试者权益的主要措施。人体试验的方案需经伦理委员会审议同意并签署批准意见后方能实施。人体试验的临床科研设计报告书应附有伦理委员会签署的批准同意书及知情同意书的全文, 后者应写明受试者参加试验是自愿的, 在试验任何阶段有权随时退出试验而不会遭受歧视或报复, 其医疗待遇与权益不受影响, 参与试验及在试验中的个人资料均属保密, 写明试验目的, 试验的过程与期限, 检查操作。受试者预期可能的受益和可能发生的风险和不便, 并应告知受试者可能被分配到试验的不同组别, 如发生与试验相关的损害时, 受试者将获得治疗和适当地补偿。如采用双盲法, 则需要有人监督整个临床试验, 在临床试验过程中病人发生的问题都需要及时报告和处理。对于对照组的患者, 除了试验因素之外, 其他任何治疗、护理及观察条件, 应保证与试验组患者享受同等待遇和关照, 要同等重视其安全。

安慰剂对照, 要视疾病的特点、性质而定, 勿滥用以防损害医德及危害患者。对于设计报告书的内容, 均应按要求实事求是撰写, 不能有任何虚假不道德行为。

#### 十五、管理和时间安排 (administration and time table)

在科研计划中, 对主研人员职责, 尤其是项目负责人的职责必须明文规定。对收集资料与管理资料, 记录科研实验日志等均应明确分工。如系协作课题, 协作部门应签科研协作合同书, 写明各自在科研实施中的任务, 课题进展中各自职责, 成果的享有及论文发表排名次问题。为保证科研工作顺利执行, 主持部门与承担单位主管部门都要签字盖章, 以承担责任。

在科研的内容、方法、指标明确后, 可根据工作量大小来安排进度, 总进度和年度计划进度, 对于阶段小结时间和总结的时间都必须有明文规定等, 以便主管部门检查。

#### 十六、经费预算 (budget)

科研经费是进行科研工作必不可少的条件, 我国是发展中国家, 各种科研基金经费都不充裕, 往往不可能购置较多大型的仪器设备。因此, 往往资助那些有条件的单位, 因此, 需要写明申请科研经费单位现有仪器设备, 特别是课题所需的各种仪器设备。如国家自然科学基金会要求将经费预算项目包括: ①科研业务费包括收集资料、统计分析和参加学术会议交流等支出; ②实验材料费, 包括试剂、购买动物和检验费等; ③仪器设备费, 只允许添置小型仪器及一些消耗品; ④实验室改装费, 即为完成本课题实验室要作某些改装所花费的费用; ⑤协作费, 和外单位或本单位其他实验室协作需要支付的协作费; ⑥管理费, 指院校科研管理部门所要提取的科研管理费。要将申请科研经费按上述各项逐一列出, 并分配合理。

(林果为 王吉耀)



## 第二十二章 临床科研论文的撰写原则和方法

为什么要撰写科研论文？科研论文是以文字的形式对所做科研工作的总结和表达，是提供新的科学信息和科学证据以及推广和交流科研成果与临床经验的主要方式。一篇科研论文的质量是科研水平的直接体现，也是影响作者及其所在单位、地区、甚至国家在相关学术界地位和威望的重要基础。理所当然，科研论文也成为对研究者科研水平和科研贡献进行评估的重要指标之一。

### 第一节 医学科研论文的质量基础

科研论文的质量基础主要决定于其学术水平与写作水平，如同任何一件好的高档物品，上等材质和精美做工是缺一不可的。

#### 一、学术水平

科学性、创新性和实用性是任何一项科研工作必须具备的本质。是医学科研论文质量基础的基础。很难想象一个设计不严谨、技术不规范、数据不准确、统计不合理的研究工作能总结出一篇有价值的科研论文来。科研论文的学术水平主要从以下几方面评估：

1. 立题依据充分 好的研究课题应具有创新性，有重要的理论意义或临床实用价值。能解决临床中所遇到的问题。
2. 严谨的科研设计 科学严谨的科研设计和正确的研究方法是基本保证。
3. 符合伦理要求 临床研究一定要符合伦理要求，须经所在单位伦理委员会的批准。
4. 准确可靠的资料 研究资料的收集应客观、求实、准确、可靠。不能主观，尽量减少偏倚。任何作假、抄袭都是绝不能容忍的。
5. 正确合理的统计分析 对研究数据的处理要得当，必须符合医学统计学的原则。使用不恰当的统计方法会得出完全错误的结论。
6. 实事求是及恰如其分的评价 因为任何研究都不可能是完美无缺的，对研究的结果下结论时一定要实事求是及恰如其分。一篇不能用临床流行病学的原则客观评价研究结果的局限性、推广性及其实际应用价值的文章，则没有资格被认为是高水平的科研论文。

#### 二、写作水平

虽然科研论文不是文学作品，但高质量的表达技能同样是科研论文写作的基本功。高质量的表达技能指清晰的逻辑思维，高度的归纳总结能力，一目了然的图表设计，简洁流畅的语句文字以及规范的撰写格式。不可想象一个优秀的科研工作者会将一篇思路零乱、语句啰唆、文法不通的科研文章展现给读者。对科研论文来说，写作水平的基本要求是：

1. 逻辑性 根据论文的格式要求，层次分明，条理清楚。
2. 语言 简洁、流畅 文字精练、简洁，语言通顺流畅是科研论文写作的基本素质。
3. 内容 文章中用词要准确，通俗易懂，但要具学术性，术语要标准化，不要自造术语或生字。
4. 清晰的图表 图表应用可增强可读性和易理解性。一定要注意规范化。表格不宜过多过大，要简明清楚，一目了然。



## 第二节 临床医学科研论文的撰写格式与内容

1997 年国际医学杂志编辑委员会 (international committee of medical journal editors, ICMJE) 发表了〈生物医学杂志投稿统一要求〉 (uniform requirements for manuscripts submitted to biomedical journals), 对生物医学论文撰写和杂志编辑两个方面进行统一的规范和指导。该要求发布在 [www.ICMJE.org](http://www.ICMJE.org) 网站上, 并定期进行更新。目前, 国外许多著名生物医学期刊都采用或参考了这个要求, 成为全球医学科论文撰写遵循的共同写作格式, 一般包括以下几部分: ①题目; ②作者姓名和研究机构; ③摘要; ④关键词; ⑤导言或引言; ⑥资料与方法或对象与方法; ⑦结果; ⑧讨论; ⑨结论或小结; ⑩致谢 (酌情); ⑪参考文献。

### 一、题目

题目应当是一语道破这是一篇关于什么的研究。因此要求文字精练、确切、醒目、对读者有吸引力。画龙点睛地将论文的内容、目的和意义清楚表达出来。

精练指题目不要过长, 文字切忌过多, 是全部工作最简洁扼要的概括。一般以不超过 20 字为宜。准确指客观恰当地反映论文的性质和信息范围。比如一篇题目为“中国人 XX 基因多态性的研究”中, 研究对象是某一小城市中的 100 例某一种族的妇女。不管结果如何, 实验如何漂亮, 如此小的样本, 如此单一的民族, 如此局限的地域, 怎能代表中国人呢? 很显然, 题目中提供的信息范围不准确。好的论文题目还应当醒目, 引起读者的兴趣。做到这一点, 一方面取决于研究内容的新颖与独特以及是否为受关注的热门话题, 另一方面需要善于在题目中将其新颖独到之处用高度概括的文字突出出来。此外, 还有一点要注意的是在题目中尽量少用缩略语、深奥的行话, 不一定需要语法结构完整的长句。

### 二、作者

作者署名的意义在于既体现研究者对该项工作的贡献, 也表明其对该项工作的学术责任, 甚至法律责任。许多国际医学杂志编辑委员会对论文的作者署名有明确的规定, 要求在论文发表前签署作者确认书。这些规定和做法在国内一些正规的医学杂志也已经开始执行。对署名作者的主要规定包括:

1. 参与研究的选题和设计, 资料的分析和解释者;
2. 起草或修改论文中重要和关键内容者;
3. 最终同意发表论文者。

凡署名的作者均需符合以上规定, 而且应对论文涉及的内容和结论负责。作者署名应当反映实际情况, 署名的顺序应根据对论文贡献大小排序, 而不应按学术威望的高低排名。一般的排序规则是这样的: 研究计划执行人 (第一作者)、论文协作人、课题设计人、研究项目负责人 (排名末位)。对仅参与筹措研究经费或收集资料, 项目一般管理人员或仅对论文进行修改、评价者不作为论文署名作者。对于这些人员的贡献, 应列入致谢部分。

作者姓名之下应列出研究机构或工作单位, 地址、邮政编码, 以便于读者和编辑部与作者联系或咨询。

对多中心合作的研究课题, 论文的署名作者可以是负责单位或课题组织的名称, 全部作者可依序附于文末, 但也必须符合以上规定条件。同时还必须明确 1—2 通讯作者, 负责对该论文的联系与解释工作。



### 三、摘要

正规的医学杂志要求在正文之前有一段约 200~250 字摘要。是对全部论文的高度概括,为的是让读者能快速便捷地了解论文的基本内容并决定是否需要或有兴趣阅读全文。因此,摘要常常被收录在文献数据库里以供读者检索和筛选。这就要求摘要虽然字数不多,却要内容明确,主要研究数据完备,要点突出,论点明确。许多国内外重要的医学杂志对摘要的书写有明确的结构格式要求,即四段式结构或类似的结构:目的、方法、结果和结论。

1. 目的 (objective) 用 1~2 句话简要说明研究的目的、要解决的问题。
2. 方法 (methods) 简述课题的设计方法、研究对象、资料收集方法、观察的指标和研究的内容。
3. 结果 (results) 简要列出主要的研究结果。通常要有数据资料并明确统计学意义和临床价值。最重要和最有意义的结果写在最前面。结果的表达一定要准确、具体、清楚。
4. 结论 (conclusions) 表达作者通过本研究最想阐明的观点。这些论点的意义和价值,是否有尚待解决和需要进一步研究的问题。

### 四、关键词

是专门为标引和检索医学文献而设计的人工语言,反映科学论文的主要内容和研究领域,故又称主题词或索引词。主要用于读者检索。国内期刊要求一般关键词 3~5 个,国外期刊有些可以 3~10 个。

关键词应规范化,为便于论文的收录,应参考 MeSH 和中国医学期刊索引。

### 五、导言或引言

是正式论文的起始部分,或叫开场白。最主要的目的是让读者了解我要研究的是什么以及为什么要进行这项研究。因此导言的内容应当包括论文所涉及问题的背景,国内外关于这一问题的研究现状和进展,研究思路的来源与依据,本研究要解决的问题是什么及研究的目的和意义。通过对上述内容简短而精辟的阐述,引起读者对这项研究的兴趣,很想知道研究是如何进行的,结果又是如何。从而顺理成章地引出正文。

引言写得好坏,体现了作者的思路是否清晰,研究目的是否明确,立题依据是否充分。要求开门见山,引用文献准确,语言精练,逻辑层次清楚,篇幅适中。

### 六、材料与方法

在材料与方法这部分是要告诉读者用什么方法开展这项研究和如何进行研究的。因为方法学的科学性和严谨性决定了结果的可靠性与合理性。这一段的最重要意义在让读者鉴别用这样的材料与方法研究出来的结果是否真实可信。若所得结果不能让读者信服,应当算是失败的研究,甚至是垃圾研究。因此材料与方法是体现作者研究水平的和科研素质的重要部分。高质量的科研论文都会对材料和方法作认真阐述而不是仅用两句话敷衍了事。

临床医学研究中的研究材料常常来自人体,所以这一部分也常常写成“对象和方法”。

#### (一) 研究对象

1. 对象的来源 如果临床研究中涉及了动物实验部分,则应说明动物的品种、性别、年龄、特殊的特征要求以及数量等。

如临床研究仅涉及病人,则应说明研究对象的来源,如住院、门诊还是社区等,是否是随机抽样的样本,年龄、性别等一般人口学资料。如果是来自随机抽样的样本,则应详



细交代随机抽样的方法，而不能仅提“随机抽样”四个字了事。

2. 对象的选择 纳入研究的临床病例一定要有明确的诊断标准和确诊方法，应当是该病诊断的金标准或当前学术界比较公认的标准。有时除疾病的诊断标准外还有其他的纳入标准和排除标准。若有对照组应明确对照的选择标准。这些标准一定要具体、严格，便于研究结果推广应用或重复性验证。不能被重复验证出来的研究结果是不易被认可的。

3. 样本量的计算 有些研究要阐明样本量计算的过程，计算公式中各参数的确定理由，以表明本项研究结果统计学意义的把握度。

4. 研究对象基线资料的齐性检验 如果研究设了对照组或多组比较，则要交代研究对象被分组的方法。在研究前应该对组间的基线资料进行比较，常包括人口学资料和主要的临床特点，并进行统计学分析，以检验所纳入研究的各组之间是否有可比性，也就是所谓的资料均衡性或齐性检验。

## （二）研究方法

1. 研究设计方案 这是对研究总体策略的表达。

诊断性研究应说明诊断试验的金标准以及靶病组和非靶病组的诊断及分组的条件、依据，测试的诊断指标等。

治疗性研究：常用的研究设计方案有随机对照研究、非随机对照研究、开放研究、配对设计、交叉设计等。应阐明是哪种设计，开放研究还是对照研究，是否有随机及如何随机，是否用盲法及如何进行盲法等。

病因学研究：回顾性还是前瞻性，队列研究要明确暴露因素的定义、结果指标的定义；病例-对照研究应明确对病例与对照组选择条件或配对研究中的匹配条件。

疾病预后研究：要对研究因素（疾病）和预后因素（结局）的各种类型、各种级别有明确的定义和判断标准。

2. 试验措施和实施方法 包括标本的采集，实验技术和操作方法的描述，所需的试剂和仪器的来源、批号或型号、测定的质控指标等。治疗药物的名称、剂量和给药方式，随访频率、疗程，观察指标、记录方法和停止治疗的指标等。比较严谨的科研论文常常较详细地阐述采用哪些具体措施以控制或减少在实施过程中可能出现的偏倚或干扰。比如，如何提高和准确记录研究对象的依从性，如何提高随访率等。这样做更会提高论文的科学性和可信度。

## （三）资料的处理和统计学分析

根据研究类型和所设计的数据性质进行合理的数据处理。阐明所选择的统计分析模型。用计算机分析资料者，应说明使用的统计学软件。

# 七、结果

结果是研究的收获，是作者最想告诉读者的东西，也是读者最想知道的内容，因此是论文的核心部分。

对于研究结果的报告首先要真实、客观、准确。不允许随意舍去与意料不符的结果资料或不好解释的矛盾结果，尤其是治疗的不良反应，治疗无效病例等均要如实报告。结果的展示应密切围绕研究目的，将在对象和方法部分中涉及到的病例、检测指标等全部结果资料认真归纳分析，以清晰的逻辑层次表达出来，以反映出其中的规律或可能存在的内在联系。研究结果的表达通常是文字、图表或照片等多种形式，以帮助读者更清楚、更直观地理解这些结果。

## （一）文字表达的要求

如实、具体、准确。描述结果时不用“多见”，“大部分”“可能有效”等不确切词语。



也切忌将原始数据用流水账式地罗列出来。值得提醒的是对数据统计学分析结果中  $P$  值的解释,  $P$  值的意义只代表事件概率的大小, 而不是事物差异的大小, 所以在描述结果时, 若  $P$  值小于 0.05, 对此更确切的解释应当说“二者差异有统计学意义”而不宜说“二者有显著差异”。在表达结果时除描述统计分析结果外, 无须添加任何评论和分析的语句和文字。

总之, 文字表达要突出两点: 第一, 观察的临床研究结果; 其文字的表达要用具体的数据以显示出临床的意义, 如像诊断性试验的敏感度、特异度等; 治疗性试验的有效率、痊愈率、病死率等; 这样读者能一目了然其临床价值; 第二, 要突出组间比较的统计学分析, 看看组间差异的显著意义。

(二) 表格的设计与要求

表格有助于将多组数字分类分层表达, 条理清楚, 一目了然。表格的设计要合理规范(表 22-1):

- 1. 题 列于表格顶线的上方。现在国内有些杂志要求标题用中、英文二种文字;
- 2. 主栏与宾栏下有关项目表达的相关性要清楚, 简明扼要, 项目不宜过多;
- 3. 多采用三横线表格, 表的顶线与底线用粗线, 两端及表内项目间不用纵线分隔。

表 22-1 三线表模式			(表序、标题)
主栏纵标目	总的宾栏纵标目		(表头)
	宾栏纵标目	宾栏纵标目	
主栏横标目			(表身)
主栏横标目			(表注)
注			

(三) 图像的展示与要求

结果用图像可起到更形象、更直观的效果。有关统计图形适用的资料及分析目的详列于表 22-2。

- 1. 标题 在图像下方。
- 2. 线形图常用于表达通过干预治疗后结果随时间推移所发生的动态变化, 如表达某种药物治疗 5 年中冠心病死亡率逐年下降的情况。
- 3. 条形图常用来表达各独立事件发生频率的比较。比如, 说明某种药物治疗中各种不良反应(头痛、恶心、腹泻)的发生率。
- 4. 原始图片或照片要尽量清晰。有时需用箭头在照片上指示病变部位或需要特别说明的部位。

表 22-2 统计图形适用的资料及分析目的		
图形	资料性质	分析目的
线形图	连续性资料	线段升降表达差值变化
半对数线图	连续性资料	线段升降表达发展速度
条形图	间断性资料	条形长短表达数值大小
百分条图	构成比	直条的各段长表达构成比
圆图	构成比	圆的扇形面积表达构成比
直方图	频数资料	直方面积表达各段频率
散点图	双变量资料	点的密度表达二变量间的相关关系



## 八、讨论

讨论是整篇论文的精华部分之一。讨论不是对结果的简单重复,而是对结果的逻辑延伸;它不是文献综述,而是对本领域研究现状和新发现的深入解析。所以,讨论是最显示论文水平的部分。讨论中通过对本研究材料、方法和结果特征的综合分析和评价,引申出一定的规律。评估结果的科学性和代表性以及重要的临床意义。此外可能存在某种研究设计方法上的局限性或许有着某种偏倚干扰因素对结果的影响等等。讨论写得好坏特别能反映作者的逻辑性和科学性,总结提高的能力以及对知识了解的深度和广度。讨论要注意重点突出,客观严谨,简单明了、观点明确,评价恰当。对尚不能充分证实的发现切忌盲目地作出肯定的结论,不能任意夸大其意义。通常建议科研论文的讨论部分有以下内容:

1. 高度概括研究结果,推论或归纳出其中的规律,阐述其意义。
2. 重点突出地分析本研究重要发现、独特之处、创新点,与他人的研究比较,解释其差异。切忌脱离本研究的目的而作出不相关的延伸和推导。
3. 阐明新发现的意义及局限,有时需分析材料和方法上的局限对结果的影响。
4. 提出有更深入探讨的问题及新的研究思路。

## 九、致谢

致谢的对象应当是对本研究作出了贡献,但又不符合署名作者条件的人员,如对研究提供经费、技术、病例、标本、材料、试剂和仪器者,文稿审校者,协助整理文稿、打字、图片制作者等。投稿前,应征得当事人的同意。

## 十、参考文献

论文的最后应当附上所引用的参考文献。被引用的参考文献应当是:

1. 作者所阅读过的,与本文直接有关的关键文献;
2. 前言、方法和讨论中涉及的问题、观点以及实验方法的出处。作为对读者提供参考或解释读者疑问的证据;
3. 引用的内容要准确,忠实于原意,不能随意断章取义或曲解原意。
4. 参考文献的序号应与在文章中出现的先后顺序一致。
5. 引用期刊的格式:作者、题名、刊名、年、卷、(期):起止页
6. 引用书籍的格式:主编、书名、版本、出版地、出版者、出版时间、起止页
7. 引用书籍中的部分章节:作者、题名、见:主编、书名、版本、出版地、出版者、出版时间、起止页

以上是参考文献格式的基本要求,投稿时还要根据各期刊的具体要求而有所变更。

学习科研论文写作的格式,最好方法是多阅读科学论文,在阅读中学习、模仿。不同的杂志对论文格式要求可能略有不同,拟投哪本杂志,事先要仔细阅读其要求,以该杂志中的论著为样本,按规定格式书写。

## 第三节 如何写好医学科研论文

一篇好的科研论文固然需要高质量的研究工作为基础,但论文撰写技能同样是不可缺少的。以下简述论文撰写过程中几点注意的问题:



## 一、酝酿准备

科研工作完成后，不急于马上写作。要反复思考充分酝酿后再下笔。思考的过程和内容包括：

1. 明确主题 思索中要理清楚什么是论文的主题，阐明主题的主线是什么，什么是核心内容，什么是基本观点。以后所做的一切都应当始终不离主题。

2. 消化资料 将研究的结果资料进行消化、提炼。在紧密围绕主题的前提下对资料进行去粗取精、去伪存真的消化过程，并将其加以提炼、组织。用最醒目、简明、准确的方式（包括语言或图表）将获得的资料表达出来。

3. 研究资料的整理 资料建库与质量评估，保证收集的数据的如实无误，这是极为重要的基础工作。

4. 参阅文献 尽量广泛阅读相关的文献，以发现自己的研究结果与文献报告中的相同与不同之处，从而避免重复，启迪思维，发现创新。参阅文献不仅仅是阅读，还应当学会按照临床流行病学有关文献严格评价的标准进行评价、分析、反复领悟，以便使自己所引用的参考文献具有较强的真实性和可靠性。同时还要注意不同的观点，这样才能提高学术水平。

## 二、拟定提纲

如果是前瞻性的临床研究，通常都有较为严格的研究方案，而研究的实施都是在该方案的指导下进行的，所观察的各项结果也是计划之中的，因此，所获得各项结果都是解决或回答的问题的。因而拟定撰写论文的大纲往往都以原研究方案为基础的。当然，通过实践后一定有新发现和新发展。

总之，经过上述反复思考、咀嚼、消化研究所获的结果和新知后，文章的框架已是水到渠成。有了明确主题思想后，列出写作提纲。构思文章的布局、段落，拉出框架。

## 三、完成草稿

按医学科研论文要求的格式，在写作提纲的基础上将具体内容和资料逐段充实，完成论文初稿。

## 四、最后定稿

初稿仍不是产品，要加工打磨。此时需要反复通读并默诵全文，认真思索、推敲与核对。对论文的题目、内容、格式、逻辑、语言、文字、数据、图表、参考文献等做通篇检查。局部进行修改后，还要看全篇的内容是否上下连贯、前后统一。特别要仔细核对数据和外文字母拼写，避免错误。最后，以自己最满意的内容和形式脱稿。这时，它应当是内容精彩，眉目清楚，数据准确，论据确凿，文字精练，语言朴实，逻辑性强，长短适中的既有学术性又有文学性的科研论文。

（徐 苓 李绍忱）

## 主要参考文献

1. 王家良. 临床流行病学-临床科研设计、衡量与评价. 第2版. 上海: 上海科技出版社, 2000.
2. R. Brian Haynes etc. Clinical epidemiology: How to do clinical practice and research. 3rd ed. Philadelphia: Lippincott Williams & Wilkins, 2006.
3. Fletcher RH, Fletcher SW, Wagner EH. Clinical Epidemiology-the Essentials. 3rd ed. Philadelphia: Williams & Wilkins, 2001.
4. Sackett DL. Clinical Epidemiology. A basis science for clinical medicine. 2nd ed. Boston: Little Brown, 1992.
5. BMJ: Clinical Evidence. 12 issue, Dec 2004. London: BMJ publishing Group, 2004.
6. WHO World Report on Knowledge for Better Health. Geneva, WHO, 2004.
7. Brian Haynes. Clinical Epidemiology: How to Do Clinical Practice Research. Philadelphia: Lippincott Williams & Wilkins, 2006; 49-58.
8. 世界银行. 1993年世界发展报告—投资于健康. 北京: 中国财政经济出版社, 1993.
9. C. J. L. Murray, Quantifying The burden of disease: The technical basis for disability-adjusted life year. Bulletin of the WHO. 1994, 72 (3): 429-445.
10. C. J. L. Murray, A. D. Lopez. Global and regional cause-of-death patterns in 1990. Bulletin of the WHO, 1994, 72: 447-480.
11. CJL Murray, AD Lopez, DT Jamison. The global burden of disease in 1990: Summary results, sensitivity analysis and future directions. Bulletin of the WHO, 1994, 72: 485-509.
12. 陈洁. 临床经济学. 上海: 上海医科大学出版社, 1999.
13. 梁万年. 医学科研方法学. 北京: 人民卫生出版社, 2002.
14. Global forum for health research: the 10/90 report on health research. Geneva, WHO, 2002: 48-55.
15. Guyatt G. Drummond R. Evidence-Based Medicine Working Group. Users' Guides to the Medical Literature: A Manual for Evidence-based Clinical Practice. Chicago, IL: AMA Press, 2002.
16. Straus SE, Richardson WS, Paul Glasziou, Haynes RB. Evidence-based Medicine: How to Practice and Teach EBM. Third Edition. Edinburgh: Churchill Livingstone, 2005.
17. Dixon RA, Munro JF, Silcocks PB. The evidence based medicine workbook. Critical appraisal for clinical problem solving. Oxford: Butterworth-Heinemann, 1997.
18. Higgins JPT, Green S, editors. Cochrane Handbook for Systematic Reviews of Interventions 4.2.6 [updated September 2006]. In: The Cochrane Library, Issue 4, 2006. Chichester, UK: John Wiley & Sons, Ltd.
19. Li J, Zhang Q, Zhang M, Egger M. Intravenous magnesium for acute myocardial infarction. The Cochrane Database of Systematic Reviews, 2007, Issue 2.
20. 王家良. 临床流行病学. 第2版, 北京: 人民卫生出版社, 2004.
21. 黄悦勤. 临床流行病学. 研究生规划教材. 第2版. 北京: 人民卫生出版社, 2006.
22. 李立明. 流行病学. 第6版. 北京: 人民卫生出版社, 2007.
23. 施倡元. 流行病学词典. 北京: 科学出版社, 2001.
24. Fletcher RH, Fletcher SW, Wagner EH. Clinical Epidemiology-the Essential. 3rd ed. Philadel-





- phia: Williams & Wilkins, 1997.
25. Fletcher RH, Fletcher, SW, Wagner, EH. Clinical Epidemiology. The Essentials. 3rd ed, Victor Graghics, Inc. , Baltimore, Maryland, 1996.
  26. Greenberg, RS, Daniels, SR, Flanders, WD. , et al. Medical Epidemiology. 2nd ed. Appleton & Lange. A Simon & Schuster Company, 1993.
  27. 曹家祺. 临床医学研究方法学. 北京: 北京医科大学中国协和医科大学联合出版社, 1993.
  28. 孙中行. 临床流行病学-临床科研基本方法. 沈阳: 辽宁科学技术出版社, 1989.
  29. 傅荫宇. 实用临床科研方法学, 设计、衡量、评价. 北京: 北京医科大学中国协和医科大学联合出版社, 1990.
  30. 黄悦勤. 预防医学. 北京: 北京医科大学出版社, 2004.
  31. 杨树勤. 卫生统计学. 第3版. 北京: 人民卫生出版社, 1996.
  32. Beth Dawson and Robert G. Trapp. Basic & Clinical Biostatistics. 3rd ed. Lange Medical Books/McGraw-Hill, Medial Publishing Division, 2001.
  33. 林果为. 临床流行病学-临床科研设计方法. 上海: 复旦大学出版社, 2005.
  34. 方积乾. 医学统计学与电脑实验. 第2版. 上海: 上海科技出版社, 2001.
  35. 孙振球、徐勇勇. 医学统计学. 北京: 人民卫生出版社, 2004.
  36. 倪宗瓚. 医学统计学. 北京: 人民卫生出版社, 1998.
  37. Catherine Anne Berglund. Ethics for Health Care. Melbourne: Oxford university press, 1998.
  38. Ethical Issues in International Health Research. Workshop on Program on Ethical Issues in International Health Research. Boston: Harvard School of Public Health, 2002.
  39. 陈亚新, 王大建, 冯照详, 等. 当代医学伦理学 (21 世纪高等医学院校教材). 北京: 科学出版社, 2002.
  40. 李定国, 黄红. 医院流行病学教程. 北京: 科学出版社, 2002.
  41. 谭红专. 现代流行病学. 北京: 人民卫生出版社, 2001.
  42. 孙贵范. 预防医学. 北京: 人民卫生出版社, 2006.
  43. 常存库. 揭开生命与疾病奥秘的钥匙: 医学科学方法学. 北京: 中国协和医科大学出版社, 2006.
  44. 贺石林, 陈修. 医学科研方法学导论. 北京: 人民卫生出版社, 1998.
  45. Haynes B, Sackett DL, Guyatt GH, et al. Clinical Epidemiology: How to Do Clinical Practice Research. 3rd ed. Philadelphia: Lippincott Williams & Wilkins, 2005.
  46. Fletcher RH, Fletcher SW, Wagner EH. Clinical Epidemiology: The Essentials. 3rd ed. Philadelphia: Lippincott Williams & Wilkins, 1996.
  47. Rothman KJ, Greenland S. Modern Epidemiology. 2nd ed. Philadelphia: Lippincott Raven Publishers, 1998.
  48. Dawes M, Davies P, Gray A, et al. Evidence-based practice. 1st ed, London: Churchill Livingstone, 1999.
  49. 王吉耀. 循证医学与临床实践. 第2版. 北京: 北京科技出版社, 2006.
  50. 王滨有. 临床实用流行病学. 哈尔滨: 黑龙江科技出版社, 2002.
  51. 林果为, 沈福民. 现代临床流行病学. 第2版. 上海: 复旦大学出版社, 2006.
  52. 苏炳华. 新药临床试验统计分析新进展. 上海: 上海科学技术文献出版社, 2000; 220-253.
  53. 李强. 循证医学——临床证据的产生、评价与利用. 北京: 科学出版社, 2001.
  54. 李良寿. 临床医学研究原理与方法——实用临床流行病学. 第2版. 西安: 陕西科学技术出版社, 2000.



55. 陈灏珠. 实用内科学. 第12版. 北京: 人民卫生出版社, 2005.
56. 董建. 突发事件公共卫生学. 苏州: 苏州大学出版社, 2005.
57. 杜金香, 梁万年. 突发性传染病防治教程. 北京: 北京大学医学出版社, 2004.
58. Michael B. Gregg. 现场流行病学. 李良成, 张顺祥, 译. 长沙: 湖南科学技术出版社, 1999.
59. Rothman KJ. Epidemiology: an introduction. New York: Oxford University Press, 2002.
60. 洪明晃. 临床决策分析. 见: 洪明晃. 临床科学研究设计、测量、评价. 第2版. 广州: 中山大学出版社, 2002.
61. Petitti DB. Meta-analysis, decision analysis, and cost-effectiveness analysis: methods for quantitative synthesis in medicine. 2nd ed. New York: Oxford University Press, 2000.
62. Weinstein MC, Fineberg HV, Elstein AS, et al. Clinical decision analysis. Philadelphia: W. B. Saunders Company, 1980.

## 中英文名词索引

### A

安慰剂对照	placebo control	73
安慰剂效应	placebo effect	72

### B

半随机对照试验	quasi-randomized controlled trial	92
半随机化法	quasi-randomization	70
暴发	outbreak	252
暴发调查	outbreak survey	252
暴露怀疑偏倚	exposure suspicion bias	120
背景资料	research background	261
被动监测	passive surveillance	21
比例风险模型	proportional hazard model	209
比值比	odds ratio	63
必要病因	necessary cause	161
变异系数	coefficient of variation	131
标准博弈法	standard gamble	231
标准差	standard deviation	131
标准化	standardization	124, 211
病残率	mobility rate	15
病例对照研究	case-control study	102, 194
病死率	fatality rate	15
病因分值	etiological fraction, EF	99
病因链	chain of causation	161
病因网	web of causation	161
博弈法	standard gamble	220
不接受测量偏倚	unreceptive measure bias	121
不敏感测量偏倚	non-sensitive bias	121
不确定型决策	decision under uncertainty	215

### C

测量	measurement	5, 7
测量偏倚	measurement bias	62, 74, 119, 199, 210
巢式病例对照研究	nested case control study	106
称病因分值	etiologic fraction, EF	168
称反应度	responsibility	239
称事件点	event points	216



称算术平均数	arithmetic mean	131
成本确定分析	cost-identification analysis	228
成本-效果	cost-effectiveness	9
成本-效果比	cost/effectiveness, C/E	228
成本-效果分析	cost-effectiveness analysis, CEA	24, 228
成本-效益	cost-benefit	9
成本-效益分析	cost-benefit analysis, CBA	232
成本-效用	cost-utility	9
成本-效用分析	cost-utility analysis, CUA	230
成分病因	component causes	162
成员偏倚	membership bias	117
充分病因	sufficient cause	161
传统文献综述	clinical review article	59
存活队列偏倚	survival cohorts bias	209
错误分类偏倚	misclassification bias	114

## D

单纯随机化	simple randomization	119
单个患者的随机对照试验	number of one randomized control trial, n-of-1 trial	93
单盲	single-blind	74
定量分析	quantitative synthesis	63
定量系统评价	quantitative systematic review	58
定性分析	non-quantitative synthesis	63
定性系统评价	qualitative systematic review	58
定性研究	qualitative research	60
队列研究	cohort study	96, 194
对照	control	71
对照组	control group	65
多级抽样法	multi-stage sampling	70
多因素分析	multivariate analysis	124
多中心临床试验	multi-center clinical trial	195

## E

二次研究证据	secondary studies	49
--------	-------------------	----

## F

发表偏倚	publication bias	53
发病率	incidence	13, 168
方差分析	analysis of variance, ANOVA	133
方法	methods	268
非等量随机对照试验	unequal randomization control trial	92
非随机对照试验	non-randomized controlled trial, NRCT	194
非随机同期对照	non-randomized concurrent control	73



非同期对照偏倚	non contemporary bias	118
分层	stratification	124, 210
分层随机抽样	stratified sampling	68
分层随机法	stratified randomization	68
分层随机化	stratified randomization	119
分层因素	stratifying factor	68
风险型决策	decision under risk	215

## G

干扰	co-intervention	199
干预方法	intervention	263
公正原则	justice	150
共变法	method of concomitant variation	164
构成比	proportion	131
固定效应模型	fixed effect model	63
观察偏倚	observation bias	119
观察性研究	observational studies	49, 60, 87
管理和时间安排	administration and time table	265
归因危险度	attribute risk, AR	168
国际疾病分类	international classification of diseases, ICD	23, 82
过度诊断偏倚	over-diagnosis bias	176

## H

HRQL 通用量表	generic instrument of HRQL	237
赫维茨准则	Hurwicz criterion	223
横断面研究	cross-sectional study	106
患病率	prevalence rate	14
患病率及发病率偏倚	prevalence-incidence bias	116
回顾性队列研究	retrospective cohort study	97
回忆性偏倚	recall bias	120
混杂	confounder	5
混杂偏倚	confounding bias	112
混杂因素	confounding factor	122
霍桑效应	Hawthorne effect	72, 199

## J

机遇	chance	7, 81, 199
机遇节点	chance nodes	216
极端值	extreme value	129
疾病的自然史	natural history	204
疾病负担	burden of disease	12
疾病监测	surveillance of diseases	79
疾病普查	census	80



集合偏倚	assembly bias	209
集中性偏倚	centripetal bias	211
几何均数	geometric mean	131
加权均数差值	weighted mean difference	63
家庭信息偏倚	family information bias	121
假设演绎法	hypothesis-deduction method	165
间接成本	indirect costs	227
检出征候偏倚	detection sign bias	86, 114
检验效能	power	144
简单随机法	simple randomization	65
健康相关生存质量	health related quality of life, HRQL	235
交叉对照	cross-over control	72
交叉对照试验	cross over design, COD	193
交叉试验	cross-over design	93
结果	results	268
结局	outcome	204
结论	conclusions	268
金标准	gold standard	82, 174
经费预算	budget	265
精确的测量	measurement	189
决策节点	decision nodes	216
决策树	decision tree	216
绝对受益增高率	absolute benefit increase, ABI	8
绝对危险度	absolute risk, AR	168
绝对危险降低率	absolute risk reduction, ARR	8, 201
均衡概率	breakeven probability	220
均数	mean	

## K

开放试验	open trial	75
科研假设	research hypothesis	262
可靠性	reliability	112
可信区间	confidence interval, CI	140
空白对照	blank control	73

## L

拉普拉斯准则	Laplace criterion	223
罹患率	attack rate	14
历史对照	historical control	73
历史对照试验	historical control trial, HCT	88, 194
临床的重要性	clinical importance	8
临床对照试验	controlled trial	60
临床经济学	clinical economics	224



临床流行病学	clinical epidemiology	1
临床期	clinical stage	204
临床信息净期望值	net expected value of clinical information	218
临床资料遗漏偏倚	missing clinical data bias	121
零点时间	zero time	208
流行病学三角模式	epidemiologic triangle	160
率	rate	131
伦理学	ethics	146
轮状模式	wheel model	160

## M

盲法试验	blind trial	74
敏感度	sensitivity, Sn	176
敏感性分析	sensitivity analysis	63, 172
目的	objective	268

## N

纳入标准	inclusion criteria	83
内部真实性	internal validity	54, 62, 112

## P

排除标准	exclusion criteria	84
排除偏倚	exclusive bias	118
配比	matching	124, 210
配比过度	overmatching	104
配比因素	matching factor	104
配对对照	matching control	73
匹配	matching	77
偏倚	bias	5, 110, 199
平行试验	parallel test	184
评价	evaluation	5, 8

## Q

起始队列	inception cohort	208, 209
迁移性偏倚	migration bias	118, 210
前瞻性队列研究	prospective cohort study	97
潜在减寿年数	potential years of life lost, PYLL	15
倾向性偏倚	popularity bias	211
清洗期	washout period	93
求同法	method of agreement	164
求异法	method of difference	164
区组随机法	block randomization	69, 119
全球疾病负担	global disease burden, GDB	12



缺失值	missing value	129
确定型决策	decision under certainty	215

## R

人群病因分值	population etiological fraction, PEF	99
人群特异危险度百分比	population attributable risk percent, PAR%	99
入院率偏倚	admission rate bias	113, 85

## S

三盲	triple-blind	75
森林图	forest plot	63
筛查	screening	80
伤残调整寿命年	disability adjusted life year, DALY	16
设计	design	5, 6, 189
社区诊断	community diagnosis	13
生存分析		207
生物学发病期	biologic onset	204
剩余法	method of residues	164
失访偏倚	lost to follow-up	116, 210
时间权衡法	time trade-off	221, 231
实施偏倚	performance bias	62
实用性	applicability	9
事件发生率	event rate	8
试验性研究	experimental studies	49, 87
试验组	experimental group	65
适用性	applicability	55
受试者工作特性曲线	receiver operator characteristic curve, ROC	181
双盲	double-blind	74
双向性队列研究	ambispective cohort study	97
顺序偏倚	sequence bias	121
死亡粗率	crude death rate	15
死亡率	mortality rate	14
四分位数间距	inter quartile range, IQR	131
似然比	likelihood ratio, LR	179
随访偏倚	attrition bias	62
随机抽样	random sampling	65, 84
随机对照试验	randomized controlled trial, RCT	88, 191
随机分组	random allocation	65, 84
随机化	randomization	65, 124, 210
随机误差	random error	110
随机效应模型	random effect model	63

## T

特异危险度或归因危险度	attributable risk, AR	99
-------------	-----------------------	----





特异度	specificity, Sp	176
同期随机对照	concurrent randomized control	72
同行评审	peer-reviewed	53
同异并用法	joint method of agreement and difference	164
同质	homogeneity	76
同质性检验	homogeneity test	63
突发事件	emergency events	250
推广性	generalizability	55

## W

外部真实性	external validity	55, 112
外推性	generalizability	81
外在真实性	external validity 或称 generalizability	62
危险度差值	risk difference	63
危险因素	risk factor	159, 205
卫生经济学	health economic	224
无形成本	intangible costs	228
无应答	non-response	84
无应答偏倚	non-respondent bias	84, 115

## X

系统抽样或机械抽样	mechanical sampling	70
系统评价	systematic review	57, 89
系统随机抽样法	systematic sampling	70
系统误差	systematic error	110
现患率研究	prevalence study	107
现患-新病例偏倚	prevalence-incidence bias	85, 86
限制	restriction	124, 210
相对受益增高率	relative benefit increase, RBI	8
相对危险度	relative risk	63, 168
相对危险降低率	relative risk reduction, RRR	8, 201
相对危险增高度	relative risk increase, RRI	99, 168
向均数回归	regression to the mean	72, 200
效度	validity	240
效用估计	utility assessment	220
效用评价	utility evaluation	242
效用损失	utility loss	222
信度	reliability	240
信息偏倚	information bias	112
需治疗多少例患者才能发现一例不良反应	number needed to harm, NNH	8
需治疗多少例患者才能获得一例最佳效果	number needed to treat, NNT	8, 92, 168, 201
序列试验	serial test	185
叙述性文献综述	narrative review	57



叙述性研究	descriptive study	108
选择性偏倚	selection bias/allocation bias	62, 112, 199
循证临床综述	evidence-based clinical review	59
循证医学	evidence-based medicine	3

## Y

亚临床期	subclinical stage	204
严格评价	critical appraisal	50, 58, 189
研究地点	setting	262
研究对象	patients/participants	262
研究目的	research objectives	262
研究因素	study factor	263
阳性似然比	positive likelihood ratio, LR+	179
样本大小的确定	sample size estimation	263
《药品临床试验管理规范(试行)》	good clinical practice, GCP	148
医学伦理道德	ethics	265
医学伦理学	medical ethics	146
医学文献综述	review	56
依从性	compliance	4, 85, 200
疑诊偏倚	diagnostic-suspicious bias	212
易感性偏倚	susceptibility bias	117
意向治疗分析	intention-to-treat analysis, ITT	62, 198
阴性似然比	negative likelihood ratio, LR-	179
隐匿	concealment	71
“有利”原则	beneficence	149
预测值	predictive value, PV	178
预后	prognosis	204
预后因素	prognostic factors	205
预期结果	expected outcomes	264
预期偏倚	expectation bias	212
预试验	pilot study	264
原始研究证据	primary studies	49

## Z

沾染	contamination	199
诊断	diagnosis	82
诊断标准	diagnostic standard	82
诊断怀疑偏倚	diagnostic suspicion bias	119
诊断机会偏倚	diagnostic enter bias	118
诊断试验	diagnostic test	173
诊断条件偏倚	diagnostic access bias	211
真实性	validity	3, 81, 112
整群随机对照试验	cluster randomized controlled trial	92



整群随机法	cluster randomization	70
知情同意书	informed consent form	83, 151
直接成本	direct costs	227
直接度量法	direct scaling method	220
直接非医疗成本	direct nonmedical costs	227
直接医疗成本	direct medical costs	227
志愿者偏倚	volunteer bias	116
质量调整寿命年	quality adjusted life years, QALYs	20, 221, 242
质量控制	• quality control	264
中国生物医学文献数据库	Chinese Biomedical Literature Database, CBM	40
中位数	median	131
终检	censoring	209
主动监测	active surveillance	22
专用量表	specific instrument of HRQL	237
转诊偏倚	referral filter bias	211
准确度	accuracy, Ac	177
资料收集和分析	data collection and analysis	264
自身对照	self control	72
自身前后对照试验	before-after study in the same patient	95, 194
最大最大准则	max max criterion	223
最大最小准则	max min criterion	223
最好信息资源	best single source	58
最佳证据	best evidence	9
最小成本分析	cost-minimization analysis, CMA	228
“尊重人”原则	respect for person	148