国家卫生健康委员会“十三五”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

第5版

临床流行病学与循证医学

Clinical Epidemiology and Evidence Based Medicine

主编 刘续宝 孙业桓

副主编 时景璞 王小钦 徐佩茹

仅限自学参考！严禁任何人网络传播或盗卖！后果自负！

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

国家卫生健康委员会“十三五”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

临床流行病学与循证医学

Clinical Epidemiology and Evidence Based Medicine

第5版

主　编　刘续宝　孙业桓
副主编　时景璞　王小钦　徐佩茹

仅限自学参考！
严禁任何人网络传播或盗卖！
后果自负！

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

图书在版编目（CIP）数据

临床流行病学与循证医学/刘续宝，孙业桓主编. —5 版.
—北京：人民卫生出版社，2018
全国高等学校五年制本科临床医学专业第九轮规划教材
ISBN 978-7-117-26678-9

Ⅰ. ①临… Ⅱ. ①刘…②孙… Ⅲ. ①临床流行病学-高等学校-教材②循证医学-高等学校-教材 Ⅳ. ①R181. 3②R499

中国版本图书馆 CIP 数据核字(2018)第 167295 号

人卫智网 www. ipmph. com 医学教育、学术、考试、健康，购书智慧智能综合服务平台
人卫官网 www. pmph. com 人卫官方资讯发布平台

版权所有，侵权必究！

仅限自学参考！
严禁任何人网络传播或盗卖！
后果自负！

临床流行病学与循证医学
第 5 版

主　　编：刘续宝　孙业桓
出版发行：人民卫生出版社（中继线 010-59780011）
地　　址：北京市朝阳区潘家园南里 19 号
邮　　编：100021
E - mail：pmph @ pmph. com
购书热线：010-59787592　010-59787584　010-65264830
印　　刷：北京铭成印刷有限公司
经　　销：新华书店
开　　本：850×1168　1/16　　印张：15　　插页：8
字　　数：444 千字
版　　次：2002 年 8 月第 1 版　　2018 年 8 月第 5 版
　　2018 年 8 月第 5 版第 1 次印刷（总第 17 次印刷）
标准书号：ISBN 978-7-117-26678-9
定　　价：46. 00 元
打击盗版举报电话：010-59787491　E-mail：WQ @ pmph. com
（凡属印装质量问题请与本社市场营销中心联系退换）

编　　者

以姓氏笔画为序

王小钦（复旦大学）
田文静（哈尔滨医科大学）
史晓红（山西医科大学）
刘晓清（北京协和医院）
刘爱忠（中南大学）
刘续宝（四川大学）
闫永平（空军军医大学）
孙业桓（安徽医科大学）
李　静（四川大学）
李晓枫（大连医科大学）
时景璞（吉林大学）
张　玲（首都医科大学）
张凤英（承德医学院）
张秀军（安徽医科大学）
徐佩茹（新疆医科大学）
康德英（四川大学）

秘　　书

康德英（兼）

仅限自学参考！
严禁任何人网络传播或盗卖！
后果自负！

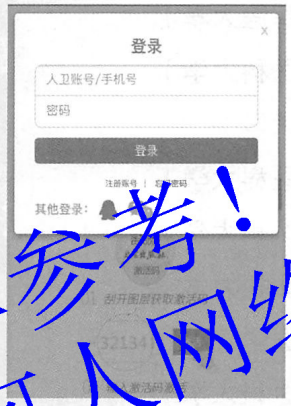
融合教材阅读使用说明

融合教材介绍：本套教材以融合教材形式出版，即融合纸书内容与数字服务的教材，每本教材均配有特色的数字内容，读者阅读纸书的同时可以通过扫描书中二维码阅读线上数字内容。

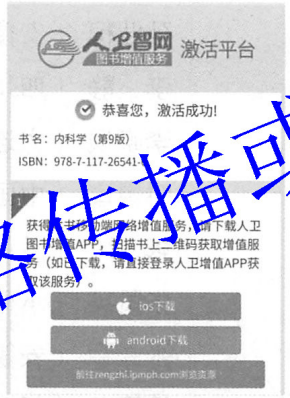
《临床流行病学与循证医学》（第 5 版）融合教材配有以下数字资源：

教学课件　案例　微课　本章小结　自测试卷　英文名词读音

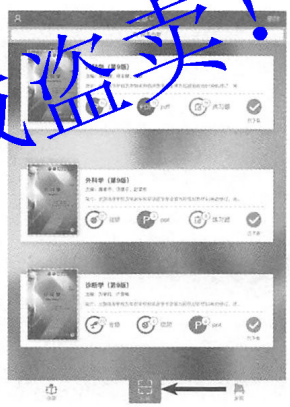
❶ 扫描教材封底圆形图标中的二维码，打开激活平台。



❷ 注册或使用已有人卫账号登录，输入刮开的激活码。



❸ 下载“人卫图书增值”APP，也可登录 zengzhi.ipmph.com 浏览。



❹ 使用 APP “扫码”功能，扫描教材中二维码可快速查看数字内容。

仅限自学参考！
严禁任何人网络传播或盗卖！
后果自负！

配套教材（共计 56 种）

全套教材书目

全套教材书目

《临床流行病学与循证医学》（第 5 版）配套教材

《临床流行病学与循证医学学习指导与习题集》（第 3 版）

主编：刘续宝、康德英

读者信息反馈方式

欢迎登录“人卫 e 教”平台官网“medu.ipmph.com”，在首页注册登录后，即可通过输入书名、书号或主编姓名等关键字，查询我社已出版教材，并可对该教材进行读者反馈、图书纠错、撰写书评以及分享资源等。

序　言

党的十九大报告明确提出,实施健康中国战略。没有合格医疗人才，就没有全民健康。推进健康中国建设要把培养好医药卫生人才作为重要基础工程。我们必须以习近平新时代中国特色社会主义思想为指引，按照十九大报告要求，把教育事业放在优先发展的位置，加快实现教育现代化，办好人民满意的医学教育，培养大批优秀的医药卫生人才。

着眼于面向2030年医学教育改革与健康中国建设，2017年7月，教育部、国家卫生和计划生育委员会、国家中医药管理局联合召开了全国医学教育改革发展工作会议。之后，国务院办公厅颁布了《国务院办公厅关于深化医教协同进一步推进医学教育改革与发展的意见》(国办发〔2017〕63号)。这次改革聚焦健康中国战略，突出问题导向，系统谋划发展，医教协同推进，以“服务需求、提高质量”为核心，确定了“两更加、一基本”的改革目标，即：到2030年，具有中国特色的标准化、规范化医学人才培养体系更加健全，医学教育改革与发展的政策环境更加完善，医学人才队伍基本满足健康中国建设需要，绘就了今后一个时期医学教育改革发展的宏伟蓝图，作出了具有全局性、战略性、引领性的重大改革部署。

教材是学校教育教学的基本依据，是解决培养什么样的人、如何培养人以及为谁培养人这一根本问题的重要载体，直接关系到党的教育方针的有效落实和教育目标的全面实现。要培养高素质的优秀医药卫生人才，必须出版高质量、高水平的优秀精品教材。一直以来，教育部高度重视医学教材编制工作，要求以教材建设为抓手，大力推动医学课程和教学方法改革。

改革开放四十年来，具有中国特色的全国高等学校五年制本科临床医学专业规划教材经历了九轮传承、创新和发展。在教育部、国家卫生和计划生育委员会的共同推动下，以裘法祖、吴阶平、吴孟超、陈灏珠等院士为代表的我国几代著名院士、专家、医学家、教育家，以高度的责任感和敬业精神参与了本套教材的创建和每一轮教材的修订工作。教材从无到有、从少到多、从多到精，不断丰富、完善与创新，逐步形成了课程门类齐全、学科系统优化、内容衔接合理、结构体系科学的立体化优秀精品教材格局，创建了中国特色医学教育教材建设模式，推动了我国高等医学本科教育的改革和发展，走出了一条适合中国医学教育和卫生健康事业发展实际的中国特色医药学教材建设发展道路。

在深化医教协同、进一步推进医学教育改革与发展的时代要求与背景下，我们启动了第九轮全国高等学校五年制本科临床医学专业规划教材的修订工作。教材修订过程中，坚持以习近平新时代中国特色社会主义思想为指引，贯彻党的十九大精神，落实“优先发展教育事业”“实施健康中国战略”及“落实立德树人根本任务，发展素质教育”的战略部署要求，更加突出医德教育与人文素质教育，将医德教育贯穿于医学教育全过程，同时强调“多临床、早临床、反复临床”的理念，强化临床实践教学，着力培养医德高尚、医术精湛的临床医生。

我们高兴地看到，这套教材在编写宗旨上，不忘医学教育人才培养的初心，坚持质量第一、立德树人；在编写内容上，牢牢把握医学教育改革发展新形势和新要求，坚持与时俱进、力求创新；在编写形式上，聚力“互联网+”医学教育的数字化创新发展，充分运用AR、VR、人工智能等新技术，在传统纸质教材的基础上融合实操性更强的数字内容，推动传统课堂教学迈向数字教学与移动学习的新时代。为进一步加强医学生临床实践能力培养，整套教材还配有相应的实践指导教材，内容丰富，图文并茂，具有较强的科学性和实践指导价值。

我们希望，这套教材的修订出版，能够进一步启发和指导高校不断深化医学教育改革，推进医教协同，为培养高质量医学人才、服务人民群众健康乃至推动健康中国建设作出积极贡献。

林蕙青

2018年2月

仅限自学参考！
严禁任何人网络传播或盗卖！
后果自负！

全国高等学校五年制本科临床医学专业
第九轮　规划教材修订说明

全国高等学校五年制本科临床医学专业国家卫生健康委员会规划教材自1978年第一轮出版至今已有40年的历史。几十年来，在教育部、国家卫生健康委员会的领导和支持下，以裘法祖、吴阶平、吴孟超、陈灏珠等院士为代表的我国几代德高望重、有丰富的临床和教学经验、有高度责任感和敬业精神的国内外著名院士、专家、医学家、教育家参与了本套教材的创建和每一轮教材的修订工作，使我国的五年制本科临床医学教材从无到有，从少到多，从多到精，不断丰富、完善与创新，形成了课程门类齐全、学科系统优化、内容衔接合理、结构体系科学的由规划教材、配套教材、网络增值服务、数字出版等组成的立体化教材格局。这套教材为我国千百万医学生的培养和成才提供了根本保障，为我国培养了一代又一代高水平、高素质的合格医学人才，为推动我国医疗卫生事业的改革和发展做出了历史性巨大贡献，并通过教材的创新建设和高质量发展，推动了我国高等医学本科教育的改革和发展，促进了我国医药学相关学科或领域的教材建设和教育发展，走出了一条适合中国医药学教育和卫生事业发展实际的具有中国特色医药学教材建设和发展的道路，创建了中国特色医药学教育教材建设模式。老一辈医学教育家和专家们亲切地称这套教材是中国医学教育的“干细胞”教材。

本套第九轮教材修订启动之时，正是我国进一步深化医教协同之际，更是我国医疗卫生体制改革和医学教育改革全方位深入推进之时。在全国医学教育改革发展工作会议上，李克强总理亲自批示“人才是卫生与健康事业的第一资源，医教协同推进医学教育改革发展，对于加强医学人才队伍建设、更好保障人民群众健康具有重要意义”，并着重强调，要办好人民满意的医学教育，加大改革创新力度，奋力推动建设健康中国。

教材建设是事关未来的战略工程、基础工程，教材体现国家意志。人民卫生出版社紧紧抓住医学教育综合改革的历史发展机遇期，以全国高等学校五年制本科临床医学专业第九轮规划教材全面启动为契机，以规划教材创新建设，全面推进国家级规划教材建设工作，服务于医改和教改。第九轮教材的修订原则，是积极贯彻落实国务院办公厅关于深化医教协同、进一步推进医学教育改革与发展的意见，努力优化人才培养结构，坚持以需求为导向，构建发展以“5+3”模式为主体的临床医学人才培养体系；强化临床实践教学，切实落实好“早临床、多临床、反复临床”的要求，提高医学生的临床实践能力。

在全国医学教育综合改革精神鼓舞下和老一辈医学家奉献精神的感召下，全国一大批临床教学、科研、医疗第一线的中青年专家、学者、教授继承和发扬了老一辈的优秀传统，以严谨治学的科学态度和无私奉献的敬业精神，积极参与第九轮教材的修订和建设工作，紧密结合五年制临床医学专业培养目标、高等医学教育教学改革的需要和医药卫生行业人才的需求，借鉴国内外医学教育教学的经验和成果，不断创新编写思路和编写模式，不断完善表达形式和内容，不断提升编写水平和质量，已逐渐将每一部教材打造成了学科精品教材，使第九轮全套教材更加成熟、完善和科学，从而构建了适合以“5+3”为主体的医学教育综合改革需要、满足卓越临床医师培养需求的教材体系和优化、系统、科学、经典的五年制本科临床医学专业课程体系。

仅限自学参考！
严禁任何人网络传播或盗卖！
后果自负！

其修订和编写特点如下：

1．教材编写修订工作是在国家卫生健康委员会、教育部的领导和支持下，由全国高等医药教材建设研究学组规划，临床医学专业教材评审委员会审定，院士专家把关，全国各医学院校知名专家教授编写，人民卫生出版社高质量出版。

2．教材编写修订工作是根据教育部培养目标、国家卫生健康委员会行业要求、社会用人需求，在全国进行科学调研的基础上，借鉴国内外医学人才培养模式和教材建设经验，充分研究论证本专业人才素质要求、学科体系构成、课程体系设计和教材体系规划后，科学进行的。

3．在教材修订工作中，进一步贯彻党的十九大精神，将“落实立德树人根本任务，发展素质教育”的战略部署要求，贯穿教材编写全过程。全套教材在专业内容中渗透医学人文的温度与情怀，通过案例与病例融合基础与临床相关知识，通过总结和汲取前八轮教材的编写经验与成果，充分体现教材的科学性、权威性、代表性和适用性。

4．教材编写修订工作着力进行课程体系的优化改革和教材体系的建设创新——科学整合课程、淡化学科意识、实现整体优化、注重系统科学、保证点面结合。继续坚持“三基、五性、三特定”的教材编写原则，以确保教材质量。

5．为配合教学改革的需要，减轻学生负担，精炼文字压缩字数，注重提高内容质量。根据学科需要，继续沿用大16开国际开本，双色或彩色印刷，充分拓展侧边留白的笔记和展示功能，提升学生阅读的体验性与学习的便利性。

6．为满足教学资源的多样化，实现教材系列化、立体化建设，进一步丰富了理论教材中的数字资源内容与类型，创新在教材移动端融入AR、VR、人工智能等新技术，为课堂学习带来身临其境的感受；每种教材均配有2套模拟试卷，线上实时答题与判卷，帮助学生复习和巩固重点知识。同时，根据实际需求进一步优化了实验指导与习题集类配套教材的品种，方便老师教学和学生自主学习。

第九轮教材共有53种，均为**国家卫生健康委员会“十三五”规划教材**。全套教材将于2018年6月出版发行，数字内容也将同步上线。教育部副部长林蕙青同志亲自为本套教材撰写序言，并对通过修订教材启发和指导高校不断深化医学教育改革、进一步推进医教协同，为培养高质量医学人才、服务人民群众健康乃至推动健康中国建设寄予厚望。希望全国广大院校在使用过程中能够多提供宝贵意见，反馈使用信息，以逐步修改和完善教材内容，提高教材质量，为第十轮教材的修订工作建言献策。

仅限自学参考！严禁任何人网络传播或盗录！后果自负！

全国高等学校五年制本科临床医学专业第九轮规划教材
教材目录

序号	书名	版次	主编	副主编
1.	医用高等数学	第7版	秦 侠 吕 丹	李 林 王桂杰 刘春扬
2.	医学物理学	第9版	王 磊 冀 敏	李晓春 吴 杰
3.	基础化学	第9版	李雪华 陈朝军	尚京川 刘 君 籍雪平
4.	有机化学	第9版	陆 阳	罗美明 李柱来 李发胜
5.	医学生物学	第9版	傅松滨	杨保胜 邱广蓉
6.	系统解剖学	第9版	丁文龙 刘学政	孙晋浩 李洪鹏 欧阳宏伟 阿地力江·伊明
7.	局部解剖学	第9版	崔慧先 李瑞锡	张绍祥 钱亦华 张雅芳 张卫光
8.	组织学与胚胎学	第9版	李继承 曾园山	周 莉 周国民 邵淑娟
9.	生物化学与分子生物学	第9版	周春燕 药立波	白英杰 汤其群 高国全 吕社民
10.	生理学	第9版	王庭槐	罗自强 沈霖霖 管又飞 武宇明
11.	医学微生物学	第9版	李 凡 徐志凯	黄 敏 郭晓奎 彭宜红
12.	人体寄生虫学	第9版	诸欣平 苏 川	吴忠道 李朝品 刘文琪 程彦斌
13.	医学免疫学	第7版	曹雪涛	姚 智 熊思东 司传平 于益芝
14.	病原学	第9版	步 宏 李一雷	来茂德 王娅兰 王国平 陶仪声
15.	病理生理学	第9版	王建枝 钱睿哲	吴立玲 孙连坤 李文斌 姜志胜
16.	药理学	第9版	杨宝峰 陈建国	臧伟进 魏敏杰
17.	医学心理学	第7版	姚树桥 杨艳杰	潘 芳 汤艳清 张 宁
18.	法医学	第7版	王保捷 侯一平	丛 斌 沈忆文 陈 腾
19.	诊断学	第9版	万学红 卢雪峰	刘成玉 胡申江 杨 炯 周汉建
20.	医学影像学	第8版	徐 克 龚启勇 韩 萍	于春水 王 滨 文 戈 高剑波 王绍武
21.	内科学	第9版	葛均波 徐永健 王 辰	唐承薇 周 晋 肖海鹏 王建安 曾小峰
22.	外科学	第9版	陈孝平 汪建平 赵继宗	秦新裕 刘玉村 张英泽 孙颖浩 李宗芳
23.	妇产科学	第9版	谢 幸 孔北华 段 涛	林仲秋 狄 文 马 丁 曹云霞 漆洪波
24.	儿科学	第9版	王卫平 孙 锟 常立文	申昆玲 李 秋 杜立中 母得志
25.	神经病学	第8版	贾建平 陈生弟	崔丽英 王 伟 谢 鹏 罗本燕 楚 兰
26.	精神病学	第8版	郝 伟 陆 林	李 涛 刘金同 赵旭东 王高华
27.	传染病学	第9版	李兰娟 任 红	高志良 宁 琴 李用国

仅限自学参考！
严禁任何人网络传播或盗卖！
后果自负！

序号	书名	版次	主编	副主编
28.	眼科学	第9版	杨培增　范先群	孙兴怀　刘奕志　赵桂秋　原慧萍
29.	耳鼻咽喉头颈外科学	第9版	孙　虹　张　罗	迟放鲁　刘　争　刘世喜　文卫平
30.	口腔科学	第9版	张志愿	周学东　郭传瑸　程　斌
31.	皮肤性病学	第9版	张学军　郑　捷	陆洪光　高兴华　何　黎　崔　勇
32.	核医学	第9版	王荣福　安　锐	李亚明　李　林　田　梅　石洪成
33.	流行病学	第9版	沈洪兵　齐秀英	叶冬青　许能锋　赵亚双
34.	卫生学	第9版	朱启星	牛　侨　吴小南　张正东　姚应水
35.	预防医学	第7版	傅　华	段广才　黄国伟　王培玉　洪　峰
36.	中医学	第9版	陈金水	范　恒　徐　巍　金　红　李　锋
37.	医学计算机应用	第6版	袁同山　阳小华	卜宪康　张筠莉　时松和　娄　岩
38.	体育	第6版	裴海泓	程　鹏　孙　晓
39.	医学细胞生物学	第6版	陈誉华　陈志南	刘　佳　范礼斌　朱海英
40.	医学遗传学	第7版	左　伋	顾鸣敏　张咸宁　韩　骅
41.	临床药理学	第6版	李　俊	刘克辛　袁　洪　杜智敏　闫素英
42.	医学统计学	第7版	李　康　贺　佳	杨土保　马　骏　王　彤
43.	医学伦理学	第5版	王明旭　赵明杰	边　林　曹永福
44.	临床流行病学与循证医学	第5版	刘续宝　孙业桓	时景璞　王小钦　徐佩茹
45.	康复医学	第6版	黄晓琳　燕铁斌	王宁华　岳寿伟　吴　毅　敖丽娟
46.	医学文献检索与论文写作	第5版	郭继军	马　路　张　帆　胡德华　韩玲革
47.	卫生法	第5版	汪建荣	田　侃　王安富
48.	医学导论	第5版	马建辉　闻德亮	曹德品　董　健　郭永松
49.	全科医学概论	第5版	于晓松　路孝琴	胡传来　江孙芳　王永晨　王　敏
50.	麻醉学	第4版	李文志　姚尚龙	郭曲练　邓小明　喻　田
51.	急诊与灾难医学	第3版	沈　洪　刘中民	周荣斌　于凯江　陈玉国
52.	医患沟通	第2版	王锦帆　尹　梅	唐宏宇　陈卫昌　康德智　张瑞宏
53.	肿瘤学概论	第2版	赫　捷	张清媛　李　薇　周云峰　王伟林　刘云鹏　赵新汉

仅限自学参考！
严禁任何人网络传播或盗卖！
后果自负！

第七届全国高等学校五年制本科临床医学专业教材评审委员会名单

顾　　问

吴孟超　王德炳　刘德培　刘允怡

主任委员

陈灏珠　钟南山　杨宝峰

副主任委员（以姓氏笔画为序）

王　辰　王卫平　丛　斌　冯友梅　孙颖浩　李兰娟
步　宏　汪建平　张志愿　陈孝平　陈志南　陈国强
郑树森　郎景和　赵玉沛　赵继宗　柯　杨　崔丹浩
曹雪涛　葛均波　赫　捷

委　　员（以姓氏笔画为序）

马存根　王　滨　王省良　文历阳　孔北华　邓小明
白　波　吕　帆　吕兆丰　刘吉成　刘学政　李　凡
李玉林　吴在德　吴肇汉　何延政　余艳红　沈洪兵
陆再英　赵　杰　赵劲民　胡翊群　南登崑　药立波
柏树令　闻德亮　姜志胜　姚　智　曹云霞　崔慧先
曾因明　雷　寒　颜　虹

仅限自学参考！
严禁任何人网络传播或盗卖！
后果自负！

主编简介

刘徭宝

1961 年 3 月出生于四川省成都市。 1983 年毕业于华西医科大学（现合并于四川大学），获医学学士学位并相继获医学硕士和博士学位，并于 20 世纪 90 年代在菲律宾大学医学院完成临床流行病学理学硕士学位学习和美国印第安纳大学医学院的博士后研究。 现为四川大学华西医院肝胆胰外科教授、博士研究生导师。 兼任中国外科医师协会第一届和第二届常务委员、中华外科学会胰腺学组委员、中国抗癌协会胰腺癌专业委员会副主任委员、中华临床流行病学会第 5 届和 6 届副主任委员、第 7 届常委，并为四川省医学会临床流行病学专业委员会主任委员、四川省抗癌协会胰腺癌和胃癌专业委员会主任委员。

主要从事普外科和临床流行病学的临床工作和教学，侧重于胰腺外科疾病的临床和相关基础研究，在国内、外杂志上发表 100 余篇研究论文，主编国家“十二五”规划教材《临床流行病学与循证医学》（第 4 版）。

孙业桓

1961 年 10 月生于安徽省寿县。 1983 年毕业于安徽医科大学，获医学学士学位，后分别在安徽医科大学和香港中文大学获医学硕士和博士学位。现为安徽医科大学研究生学院院长、循证医药学中心主任、公共卫生学院流行病与卫生统计学系教授、博士研究生导师。 现为《中华疾病控制杂志》副主编、《中华行为医学与脑科学杂志》《安徽医科大学学报》等编委；中国抗癌协会肿瘤流行病学专业委员会委员、中华预防医学会伤害预防控制学会委员、安徽省预防医学会常务理事、安徽省预防医学会卫生统计学专业委员会主任委员、伤害预防控制专业委员会主任委员。

从事教学工作 35 年，主要从事流行病学、临床流行病学、循证医学、医学科研方法学等教学工作，为安徽省学术和技术带头人、安徽省教学名师、省级重点学科、国家级精品课程“流行病学”学术带头人之一。 曾获省级科技进步三等奖 2 项、国家级教学研究成果二等奖、省级教学研究成果特等奖/一等奖/二等奖各一项；在国内外发表学术论文 200 余篇；主编、副主编、参编出版学术著作或教材 20 余本。 获评 2007 年和 2014 年安徽省优秀教师，2010 年安徽省优秀教学管理工作者。

仅限自学参考！
严禁任何人网络传播或盗卖！
后果自负！

副主编简介

时景璞

1956 年 11 月生于沈阳。 中国医科大学附属第一医院临床流行病与循证医学教研室教授，博士生导师，辽宁省教学名师。 现任中国医科大学及辽宁省循证医学中心常务副主任，中华医学会临床流行病学分会常委，中国医疗保健国际交流促进会循证医学专业委员会常委，辽宁省临床流行病学与循证医学分会前主任委员。 现担任《中华流行病学杂志》《中国实用内科杂志》《中国循证医学杂志》《中国公共卫生》常务编委、编委。 担任 *Chinese Medical Journal* 的审稿专家。

先后承担十几项包括国家自然科学基金、国家科技攻关、省部级基金、国际协作研究课题研究工作；并先后获得十几项各级科技进步奖；同时在各级杂志发表论文 200 多篇；先后参与近 20 多部国家级教材及专著的编写。

李小钦

1968 年 11 月生于上海。 教授，博士研究生导师。 现任复旦大学附属华山医院血液科副主任，复旦大学循证医学中心副主任，上海医学会临床流行病学与循证医学分会主任委员，中华医学会临床流行病学分会委员和工作秘书。 研究方向为血液学、临床流行病学和循证医学。 负责多项大型国际合作课题、国家级和省部级课题。 发表论文 120 余篇，其中 SCI 论文 20 余篇。 主编《循证临床实践指南的制定与实施》《现代临床流行病学》，副主编《循证医学》《中西医结合临床研究方法学》，参加编写专著和教材 10 余部。 科研成果获得上海市科技进步三等奖和上海市医学奖三等奖，主编教材《现代临床流行病学》获得上海市普通高校优秀教材奖。

徐佩茹

1963 年 9 月生于乌鲁木齐。 教授，博士研究生导师。 现任新疆医科大学儿科学院院长，新疆维吾尔自治区循证医学研究所所长；兼任中华医学会儿科分会委员，中华医学会儿科学分会呼吸学组儿童睡眠呼吸障碍协作组副组长，中华医学会儿科学分会儿科人文建设委员会副主任委员，中国女医师协会卫生发展与管理专业委员会常委，中国医师协会循证医学专业委员会常委，新疆医学会儿科分会主任委员；担任《中华实用儿科临床杂志》通讯编委、《临床儿科杂志》常务编委、《中国循证医学杂志》编委。 先后主持国家自然科学基金、自治区自然科学基金、国际合作课题、教育部多中心等 10 余项课题。 获宋庆龄儿科医学成果奖、自治区科技进步三等奖、新疆医学科技三等奖等多项奖励，先后发表论文 96 篇。

仅限自学参考！
严禁任何人网络传播或盗卖！
后果自负！

前　言

临床流行病学与循证医学是一门新兴的、前沿性的、多学科交叉的临床基础学科。旨在传授医学生以群体观点对疾病发生、发展、诊断、防治及预后等进行临床研究和开展循证临床实践的基本理论、基本知识和基本方法，使临床医学生从个体患者的特点，扩大到疾病的群体共性规律的认识，有助于培养医学生的科学素质、科研思维、临床研究与循证临床实践能力。

本教材第 4 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，是一本深受欢迎的优秀教材。在这次第 5 版的修订中，编委会扩增了有关高等医学院校并诚邀了学术造诣高的教授参与了编写，我们结合本学科的最新进展以及我国高等医学教育的实际和医学人才培养的总体目标，充分发挥了新的编委会的集体智慧，对第 5 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，进一步强调研究实践与动手能力的培养，新增“临床医学论文的撰写原则和报告规范”（第十八章）。

2. 鉴于经济学证据在循证决策中的作用和价值越来越重要，新增“临床经济学在临床科研中的应用与评价”（第十七章）。从经济学角度考虑权衡成本和健康获益，结合最佳证据、兼顾经济效益和价值取向，真正做到科学决策。

3. 同时结合学科发展以及一些新进展、新成果，进一步充实了“临床研究的常用设计方案”一章，新增“真实世界研究”和“外科临床研究的一些特殊性问题”等两节，力求与时俱进，紧跟现代科学发展的步伐，知识不过时，不落伍，及时反映临床医生的诉求，改善临床研究“重内科、轻外科”的现状。

我们共同的愿望是在全国一流教材的基础上，不断地吸收最新最佳证据，丰富发展本教材的科学内涵，使之永葆一流水平，从而更好地服务于我国高等医学教育和人才培养的使命。在本轮的修订中，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领导的大力支持，特致以衷心的谢意！

本教材的文字整理、编辑和排版以及部分内容的编写和图表绘制等，由康德英和洪旗两位老师负责完成，对他们辛勤的付出和奉献，致以诚挚的感谢！

刘续宝　孙业桓

2018 年 5 月

仅限自学参考！
严禁任何人网络传播或盗卖！
后果自负！

仅限自学参考！
严禁任何人网络传播或盗卖！
后果自负！

目　录

第一章　绪论　1

第一节　临床流行病学与循证医学的发展历程　1
第二节　临床流行病学与循证医学的概述　3
一、临床流行病学与循证医学的概念　3
二、临床流行病学与循证医学的关注对象　3
三、临床流行病学与循证医学的学科特点　4
第三节　临床流行病学与循证医学的共同方法学基础　5
一、设计　5
二、测量　7
三、评价　7
第四节　临床流行病学与循证医学的作用和价值　8
一、临床流行病学与循证医学的作用和价值　8
二、临床流行病学与循证医学的常见误区　9

第二章　临床研究与实践问题的构建　11

第一节　概述　11
一、提出临床问题的重要性　11
二、找准临床问题应具备的条件　12
第二节　循证问题的构建　12
一、临床问题的类型　12
二、临床问题的提出　13
三、循证问题的构建模式　15
第三节　临床研究的选题与立题　15
一、临床研究的特点　15
二、选题与立题的原则　16
三、研究选题和立题程序　18
四、研究立题的评价标准　19

第三章　临床研究设计的基本要求　20

第一节　随机化原则　20
一、概述　20
二、随机化方法　20
三、随机化过程中需注意的问题　23

仅限自学参考！
严禁任何人网络传播或盗卖！
后果自负！

第二节 对照原则 24
一、概述 24
二、设置对照的意义 24
三、对照的种类 24
第三节 盲法原则 26
一、概述 26
二、盲法的分类 26
三、非盲法评定 27
第四节 重复原则 28
一、概述 28
二、样本量的估计 28
第五节 临床研究的伦理学基础 28
一、概述 28
二、临床研究的伦理学原则 29
第六节 临床研究中的偏倚及控制 30
一、概述 30
二、选择偏倚 30
三、信息偏倚 31
四、混杂偏倚 32

第四章 临床研究的常用设计方案 34

第一节 随机对照试验 34
一、概述 34
二、设计模式 35
三、方案要素设计 36
四、资料的整理与分析 37
五、注意事项 38
六、优点和缺点 39
第二节 随机对照试验的一些特例 39
一、交叉试验 39
二、自身前后对照试验 41
三、基于单个患者的随机对照试验 42
四、非等量随机对照试验 43
五、整群随机对照试验 43
第三节 其他类型的临床对照试验 43
一、半随机对照试验 43
二、非随机同期对照试验 43
三、历史性对照试验 44
四、单臂研究 44
第四节 队列研究 44
一、概述 44
二、设计模式 45

仅限自学参考！
严禁任何人网络传播或盗卖！
后果自负！

三、方案要素设计　45
四、资料的整理与分析　47
五、应用范围　49
六、优缺点　49
第五节　病例对照研究　49
一、概述　49
二、设计模式　50
三、方案要素设计　50
四、资料的整理与分析　52
五、应用范围　53
六、优缺点　53
第六节　横断面研究　53
一、概述　53
二、设计模式　54
三、方案要素设计　55
四、资料的整理与分析　55
五、应用范围　55
六、优缺点　56
第七节　病例分析　56
一、概述　56
二、设计模式　56
三、方案要素设计　57
四、资料的整理与分析　57
五、应用范围　57
六、优缺点　57
第八节　真实世界研究　58
一、概述　58
二、真实世界证据的产生与应用　59
三、真实世界证据的评价　61
第九节　外科临床研究的一些特殊性问题　61
一、外科的特殊性容易导致的一些认识偏差　61
二、外科临床研究的难点与解决方法　62

第五章　医学研究证据的检索与收集　64

第一节　证据资源　64
一、资源分类　64
二、证据资源的入选标准　66
三、证据分级及其发展历程　66
第二节　证据检索和文献搜集的基本步骤　67
一、确定临床问题类型和构建临床问题　67
二、选择检索数据库　68
三、制订检索策略　68

仅限自学参考！
严禁任何人网络传播或盗卖！
后果自负！

四、评估检索结果 68
第三节 证据检索实例 68
一、提出并构建问题 69
二、制订检索计划 69
三、检索 69

第六章 医学研究文献的评阅与证据质量评价 71

第一节 阅读与评价医学研究文献的重要性 71
一、医学研究文献的特殊性 71
二、阅读和评价医学文献的重要性 72
第二节 阅读医学文献的基本步骤 72
一、医学文献检索 72
二、阅读医学文献的基本步骤 73
第三节 医学研究文献的评价要素与方法 75
一、确定医学研究文献评价的范畴和内容 75
二、评价医学研究文献的一般原则 75
三、常见临床研究类型的评价标准 76
四、研究质量评价工具 76

第七章 系统评价与 meta-分析 78

第一节 系统评价概述 78
一、基本概念 78
二、进行系统评价的必要性 79
三、系统评价与叙述性文献综述的区别与联系 80
四、系统评价分类 80
第二节 系统评价的方法 81
一、系统评价前的准备 81
二、系统评价流程 81
三、系统评价方法 81
第三节 meta-分析 89
一、meta-分析概述 89
二、meta-分析基本过程 89
三、发表性偏倚的识别与分析 94
四、meta 回归及其他方法学进展 95
五、慎重应用与评价 meta-分析结果 96
第四节 系统评价的评价原则 96
一、系统评价的结果是否真实 97
二、系统评价的结果是否重要 97
三、系统评价的结果是否能应用于我的患者 98

仅限自学参考！
严禁任何人网络传播或盗卖！
后果自负！

第八章　病因与危险因素的研究、评价与循证实践　99

第一节　病因与危险因素研究的意义　99
一、有助于疾病的临床诊断与治疗　99
二、有助于疾病的预防　99
第二节　病因与危险因素研究的概述　99
一、病因与危险因素　100
二、寻找病因的指南清单　104
三、疾病自然史对病因学研究的意义　104
第三节　疾病病因与危险因素研究的基本过程与方法　105
一、提出病因假设　105
二、验证病因假设　106
三、验证病因/危险因素假设的主要研究方法　107
四、病因学因果关联的统计学指标　109
五、慎重分析病因效应的关联　109
第四节　病因与危险因素研究的评价原则　110
一、真实性评价　111
二、重要性评价　113
三、适用性评价　113
第五节　循证实践及案例分析　114
一、提出问题　114
二、证据检索与评价　114
三、实践决策　115
四、小结　118

第九章　诊断试验研究、评价与循证实践　119

第一节　概述　119
一、基本概念　119
二、诊断试验研究与评价的意义　119
第二节　诊断试验研究设计　119
一、确定金标准　120
二、研究对象的选择　120
三、样本量的估计　120
四、评价指标与诊断效能判定　120
第三节　多项试验的联合诊断　125
一、联合诊断方式　125
二、独立性与联合试验的敏感度和特异度　126
第四节　诊断试验的循证评价原则与标准　126
一、评价原则　126
二、诊断试验的评价标准　128
第五节　诊断试验证据的临床实践及案例分析　129

仅限自学参考！
严禁任何人网络传播或盗卖！
后果自负！

一、构建循证诊断问题 129
二、检索证据 130
三、评价证据 130
四、应用证据 131

第十章 疾病治疗的研究、评价与循证实践 137

第一节 概述 137
一、治疗性研究与评价的重要性 137
二、治疗性研究与评价的基本条件 137
三、治疗性研究及其试验分期 138
第二节 治疗性研究的设计要素及方法 139
一、选择研究方案 139
二、选择研究对象和样本量估算 140
三、研究对象的随机化分配及设盲 142
四、选择试验药物或措施 143
五、试验观察指标的设置 143
六、确定干预实施方法及随访观察期 144
七、数据管理与统计分析 144
第三节 影响治疗性研究质量的常见因素及其处理方法 145
一、机遇 145
二、偏倚 145
第四节 治疗性研究的评价原则 146
一、单个原始研究评价 146
二、系统综述或 meta-分析证据评价 149
第五节 循证实践及案例分析 150
一、构建循证问题 150
二、检索证据 150
三、评价证据 150
四、证据应用与实践 152
五、后效评价与小结 152

第十一章 疾病预后的研究、评价与循证实践 153

第一节 疾病预后的概念 153
一、疾病预后及其研究的意义 153
二、疾病自然史 153
三、临床病程 153
四、预后因素 154
第二节 疾病预后研究中常用结局指标 155
一、率相关指标 155
二、中位时间 156
三、健康相关生存质量及其衍生指标 157

仅限自学参考！
严禁任何人网络传播或盗卖！
后果自负！

第三节　疾病预后研究设计方案　157
一、疾病预后研究常用设计方案　157
二、方案设计的注意事项　157
三、疾病预后研究的基本步骤　158
第四节　疾病预后研究的统计方法　159
一、生存率和中位生存时间分析　159
二、疾病预后因素的研究方法　159
第五节　预后研究中常见偏倚及其处理方法　160
一、预后研究中常见的偏倚　160
二、偏倚的处理方法　161
第六节　疾病预后研究的评价原则　162
一、真实性评价　163
二、重要性评价　163
三、适用性评价　164
第七节　循证实践及案例分析　164
一、提出问题　164
二、检索与评价证据　164
三、应用证据与临床实践决策　167
四、小结　167

第十二章　临床实践指南的制定与评价　168

第一节　概述　168
一、临床实践指南的概念　168
二、临床实践指南的作用和价值　168
三、指南与临床证据的关系　169
四、循证临床实践指南的网络资源　170
第二节　临床实践指南制定方法与流程　170
一、专家共识制定法　170
二、循证临床实践指南制定流程　171
第三节　临床实践指南的评价　173
一、临床实践指南的评价原则　173
二、临床实践指南的评价工具　173
第四节　临床实践指南应用的原则和方法　175
一、临床实践指南的应用原则　175
二、临床实践指南的应用技巧　175
第五节　循证临床实践指南应用案例　176
一、提出问题　176
二、检索临床指南　177
三、阅读和评价指南　177
四、应用指南　178
五、后效评价　178

仅限自学参考！
严禁任何人网络传播或盗卖！
后果自负！

第十三章 循证医学的个体化实践 179

第一节 最新最佳临床研究证据是循证个体化实践的基础 179
一、真实性 179
二、重要性 179
三、适用性 180
四、时效性 180
第二节 准确把握患者的个体化特征是循证个体化实践的基本功 180
一、生物学特征及相关依据 180
二、病理生理学特征及相关依据 181
三、社会-心理-人文-经济特征及其相关依据 181
第三节 个体化特征与证据的精准对接是循证个体化实践的成功关键 182

第十四章 临床决策分析 183

第一节 概述 183
一、临床决策的模式 183
二、临床决策的具体分型 184
三、临床决策过程 184
第二节 决策分析的实施 184
一、决策问题的确认 184
二、决策问题的结构化 185
三、相关证据及信息的检索与收集 186
四、确定最终结局的效用值 186
五、决策树分析 188
第三节 Markov 模型 189
一、Markov 模型概述 189
二、Markov 模型的分析步骤 189
第四节 敏感性分析 193
第五节 不确定型决策 194
第六节 决策分析的局限性 195

第十五章 循证医学自我评价 196

第一节 能力评价 196
一、评价“提出可回答问题”的能力 196
二、评价“寻找最佳外部证据”的能力 196
三、评价“严格评价证据质量”的能力 197
四、评价“整合外部证据与患者价值”的能力 198
第二节 效果评价 198
一、临床实践是否得以改善 198
二、多少临床实践有证可循 199

仅限自学参考！
严禁任何人网络传播或盗卖！
后果自负！

第三节　案例　199

第十六章　临床医学研究项目申报书的撰写　201

第一节　概述　201
一、科学研究的基本程序　201
二、临床医学研究的基本类型　201
三、临床医学研究项目申请的准备　202
第二节　研究项目申报书的基本内容与撰写方法　203
一、项目名称及摘要　203
二、立项依据　203
三、研究目标、研究内容及拟解决的关键科学问题　204
四、研究方案与可行性分析　204
五、项目的特色与创新之处　206
六、年度研究计划及预期研究结果　206
七、研究基础与研究团队介绍　206
八、医学伦理学问题　207
九、经费预算　207
第三节　研究项目申报书对照检查提纲　208

第十七章　临床经济学在临床科研中的应用与评价　209

第一节　概述　209
一、卫生经济学产生背景及发展历程　209
二、卫生经济学在临床科研中的运用与价值　209
第二节　临床经济学评价方法　210
一、成本和临床结果的测量与表达　210
二、成本与健康获益的测算　211
三、临床经济学评价的类型　213
第三节　临床经济学评价的基本步骤　215
一、确定要分析的项目及要比较的方案　215
二、确定经济学评价的角度　215
三、确定经济学评价的方法　215
四、确定资料获取的研究方法　216
五、增值分析　216
六、成本的确定　216
七、确定结果测定的方法　216
八、对发生在将来的结果和成本作贴现计算　216
九、敏感性分析　216
十、撰写报告　217
第四节　临床经济学评价研究文献的评价　217

仅限自学参考！
严禁任何人网络传播或盗卖！
后果自负！

第十八章 临床医学论文的撰写原则和报告规范 218

第一节 临床医学论文的格式和撰写原则 218
一、题目和作者 218
二、摘要和关键词 218
三、序言或引言 219
四、材料与方法 219
五、结果 220
六、讨论 222
七、致谢和其他 222
八、参考文献 222
第二节 临床医学论文的报告规范和投稿要求 222
一、论文撰写报告规范 222
二、论文投稿要求 224

附表 随机数字表 226

推荐阅读 228

中英文名词对照索引 229

本书测试卷

仅限自学参考！
严禁任何人网络传播或盗卖！
后果自负！

第一章 绪 论

临床流行病学(clinical epidemiology)和循证医学(evidence-based medicine,EBM)是先后始于20世纪的新兴临床医学基础科学(a basic science for clinical medicine),是临床医生进行临床科学研究和指导临床医学实践十分有用的理论与方法学,对促进临床医学的科学发展有着重要的价值。两者一脉相承,前者定位于临床研究的方法学,旨在提高临床研究质量,生产出更多、更好的临床研究证据;后者则是指导临床医疗进行科学诊治决策的方法学,即任何针对患者具体临床问题所作出的有关诊治决策,均应建立在最新、最佳的科学证据基础之上,这是与传统经验医学的最大区别所在。

第一节 临床流行病学与循证医学的发展历程

20世纪医学的快速发展促使人们需要采用从定性走向定量、个体研究转向群体研究的科学方法去认识和解决临床面临的、各种不同复杂的医学问题,以便从宏观角度更加科学的指导临床研究和临床实践。在20世纪30年代John R. Paul首先提出了临床流行病学的概念,后经几十年的探索和实践,特别是进入20世纪60年代以后,由David L Sackett、Alvan R Feinstein等创造性地将流行病学和医学统计学原理及其方法有机地与临床医学的研究和实践结合起来,并进一步拓展到与临床医学相关领域,如卫生经济学和社会医学等,极大丰富和发展了临床研究的方法学。在临床研究实践中,提高了对疾病的发生、发展和转归整体规律的宏观认识,深化了对疾病诊断、治疗和防治方法的科学观,有效地提升了临床医学研究及其实践的水平,为现代临床流行病学打下了坚实的学科基础。

从临床流行病学的提出、发展到普及推广,那些致力于从事临床流行病学研究和工作的临床医生、医学统计学家、方法学家等功不可没,同时也与国际上致力于医学发展的基金会和医学专业学会/协会等的鼎力支持密不可分。如在美国洛克菲勒基金会(Rockefeller Foundation)的大力支持下,1982年成立了国际临床流行病学网(International Clinical Epidemiology Network,INCLEN)。INCLEN的宗旨是:"在最可靠的临床证据和最有效地利用卫生资源的基础上,促进临床医学实践,从而致力于改善人类健康。为实现这一目标,本工作网内的各国临床医生、统计学家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"。INCLEN率先在加拿大、美国和澳大利亚等发达国家建立了七个一级国际临床流行病学资源和培训中心(The International Clinical Epidemiology Resource and Training Center,CERTC),为全球范围内的著名大学医学院,特别是发展中国家培训了大量临床流行病学专业人才。随后又相继在亚洲、非洲、拉丁美洲等地建立了多个二级区域性临床流行病学资源和培训中心(Regional Clinical Epidemiology Resource and Training Center,R-CERTC),这些资源和培训中心承担了全球、地区或所在国的人员培训和临床研究的指导工作。

我国临床流行病学的起步,始于改革开放的初期,在国际相关基金会的资助下,我国选派了一批优秀临床医生和流行病学家分赴美国、加拿大、澳大利亚等一些临床流行病学发展较为成熟的国家学习和进修。这些学者学成归国后为我国临床流行病学学科的建立、发展和普及做了大量辛勤的工作。如以原华西医科大学(现四川大学华西医学中心)王家良教授等为代表的一批临床流行病学的奠基人,在学科发展的初期,先是在本单位组织、参与和推动临床流行病学的研究和教学工作,继而在国家

卫生健康委员会(原国家卫生部)的大力支持下,于1983年获准在原华西医科大学(现四川大学华西医学中心)、原上海医科大学(现复旦大学上海医学院)和原广州中医学院(现广州中医药大学)建立了三个“设计、测量、评价”(design, measurement and evaluation, DME)国家级培训中心(the centre of DME on the medical research),简称DME中心,先后开设了针对临床医学本科生和研究生的临床流行病学课程(临床科研设计),同时又帮助其他医学院校建立临床流行病学教研室(或教研组)并开展相应地教学和研究工作。进而在1989年成立了INCLEN指导下的中国临床流行病学网(China Clinical Epidemiology Network, ChinaCLEN),鉴于临床流行病学在中国的成功,在INCLEN的支持下,又分别在当时的华西医科大学和上海医科大学成立了两个二级区域性的临床流行病学资源和培训中心(regional clinical epidemiology resource and training center, R-CERTC),随后又在中国协和医科大学以及中国中医科学院等多所院校单位建立临床流行病学组(clinical epidemiology unit, CEU),为国内培养了一大批临床流行病学的专业骨干,进一步促进了临床流行病学的普及和发展。在学科普及的基础上,1993年正式成立了中华医学会临床流行病学分会(现更名为中华医学会临床流行病学与循证医学分会),这为临床流行病学的进一步发展奠定了组织基础。近数十年来,临床流行病学的蓬勃发展直接推动了各临床学科的科学水平提高,尤其是在加强国际卫生研究能力、开展重大国际卫生问题的合作研究、促进发展中国家人民的健康水平和卫生资源的合理利用,以及为世界卫生组织和各国政府的卫生政策决策等,均作出了非凡的贡献或发挥了重要影响。世界卫生组织在其2004年的年度报告中,曾对临床流行病学的贡献给予了高度评价,指出“临床流行病学学科的建立,对在群体层面上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”。

临床流行病学的蓬勃发展,直接促进了大量临床研究成果的产生,人们认识到这些新的研究成果或称最佳证据(best evidence)应及时加以转化,指导临床实践,服务于社会,从而推动临床医学的整体发展和促进临床医学水平的提升。正是在此背景下,于20世纪90年代,在临床流行病学的基础上,循证医学应运而生。

循证医学的起源,从哲理上可以追溯到19世纪中叶,进入到20世纪80年代初期,在国际临床流行病学发源地之一的加拿大McMaster大学,以临床流行病学创始人之一、国际著名的内科学专家David L. Sackett为首的一批临床流行病学家,在该医学中心的临床流行病学系和内科系率先对年轻住院医生举办了“如何阅读医学文献的学习班”,在讲授临床流行病学原理与方法的基础上,进一步指导他(她)们联系患者的临床实际问题,检索与评价医学文献,并将所获得的新近成果(证据)应用于自己的临床实践。后又经过反复实践,不断完善这种基于证据的医学培训模式,取得了很好的效果。为此,同样来自加拿大McMaster大学的Gordon Guyatt等自1992年起相继在*JAMA*等杂志上发表了系列总结性文献,将这种对临床医生的新型培训模式和方法,正式冠以“循证医学”(evidence based medicine, EBM),自提出之日起就引起临床医学各界的广泛关注。另外,由Haynes和Sackett等倡议,美国内科医生学院(American College of Physicians)成立了一个杂志俱乐部(Journal Club),即ACPJC。从1991年起,由临床流行病学、临床有关学科及方法学专家组成的评审小组,对国际上著名的30多种医学杂志发表的论著,进行系统分析与评价,并将最佳的研究论文,以精练摘要加专家评述的形式,在*Annals of Internal Medicine*副刊发表。1995年David L. Sackett医生又受聘于英国牛津大学,斯时在英国组建了循证医学中心(Evidence-Based Medicine Center),开展循证医学教学和床旁循证实践,于1996年在英国医学杂志(*BMJ*)发表系列文章,提出了循证医学作为临床实践的新模式,相继出版了循证医学专著及由英国医学杂志和美国内科医生学院联合主办的循证医学杂志(*Journal of Evidence-based Medicine*)。为了全面地推荐国际上经过严格评价的最佳研究证据,自1999年起,他们还整理编辑并出版了*clinical evidence*专集,每年公开发行两期,将这些经过专家筛选、严格评价及评论后的最佳研究成果,推荐给临床医生,以指导临床医疗实践。

1993年成立的Cochrane协作网(Cochrane collaboration),又为循证医学的腾飞提供了一大助力和

组织保证。Cochrane 协作网的宗旨是在广泛地收集临床随机对照试验(RCT)的研究结果、严格评价其质量的基础上,进行系统综述(systematic review)以及 meta-分析(meta-analysis),并将这些有价值的研究结果推荐给临床医生以及相关专业人员,以帮助其实践循证医学。Cochrane 系统综述现已被公认为最佳的高质量证据之一。

循证医学在中国的发展历程基本与国际同步。我国最早于 1996 年在国家卫生健康委员会(原国家卫生部)的领导与支持下,在原华西医科大学附属第一医院(现四川大学华西医院)正式成立了中国 Cochrane 中心及循证医学中心,相继开展了循证医学国际协作研究与普训工作,陆续创刊了两种全国性的循证医学杂志,并率先在医学院校开设循证医学课程,同时编写出版了循证医学专著以及 5 年制、8 年制循证医学规划教材,对推动临床医学实践、提高医学水平产生了良好效果。该学科得以在全国迅速普及和健康发展,无疑会更好地推动临床医学各个学科的共同繁荣与进步。

总之,循证医学的发展方兴未艾,随着时代的前进步伐,它将日臻完善、为临床决策的科学性和临床医学的现代化作出更大贡献。

第二节 临床流行病学与循证医学的概述

一、临床流行病学与循证医学的概念

临床流行病学是将现代流行病学及统计学等原理和理论引入临床医学的研究和实践的一门方法学,它采用宏观的群体观点和相关的定量化指标,将科学严谨的设计、定量化的测量和严格客观的评价贯穿于临床研究全程,系统探索疾病的病因、诊断、防治和预后的共性规律,力求避免各种偏倚因素的干扰,确保研究结果的真实性,并对临床医学实践产生重要的循证价值,因此,重在产生最佳的研究成果(证据),促进临床医学的整体发展和临床医学水平的全面提升。

而循证医学则作为一门临床实践的科学,临床医生在自己的临床实践中,针对患者的具体临床问题(难点),谨慎、确切和明智地采用目前最佳的证据对患者的诊治作出科学的决策,从而取得最有效的治疗结果。这样做,既能有效地解决患者的临床问题,促进患者康复,同时也会推动临床医疗水平的提高。因此,重在应用最佳研究成果(证据),解决个体患者的具体临床实践问题。

可见,临床流行病学与循证医学一脉相承,是理论联系实际、高度统一的临床医学的一项系统工程,对推动临床医学进步、更好地服务于人民的健康事业,具有重要意义。

二、临床流行病学与循证医学的关注对象

临床流行病学与循证医学的关注对象是以医院为基础的患者及其相应的患病群体,这种特定疾病的群体性(specified population of disease)乃为本学科的“流行病学”特征。这种群体性的特征与传统的局限于医院的经验医学有很大的不同,它是从医院的患病个体延伸至社会层面的特定患者群,将医院内特定疾病的患者个体诊治和特定疾病患者群体的诊治研究相结合,跨越了医院或社会人群的界限,无疑对疾病的早诊早治、对疾病发生发展和转归规律的认识更加全面系统和深入,因此,对临床医学的整体发展有重要价值和意义。鉴于其临床基础和基于医院的患者人群特征,临床流行病学的理论体系、研究范畴、防治实践等诸多方面又有别于流行病学。

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。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影响因素,并应用研究的结果去有效地改善人民健康并预防、控制相关的疾病。而临床流行病学与循证医学则是临床学家学习与应用流行病学的原理与方法,并结合临床医学的实际,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、防治以及预后等等,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(证据),结合个体患者的具体情况,作出科学预测性的防治决策,以解决具体患者的健康问题,可见,它所

涉及的是具体的疾病及其在临床方面的诊治等临床问题，这与流行病学有所不同。然而，尽管两者在解决人类健康或疾病防治等问题上各有侧重，但在学科理论和方法学上却联系紧密。正如临床流行病学的提出者 J. R Paul 早年将“临床流行病学”定义为：将流行病学家在人群中用于研究疾病的定量化概念和临床医生日常用于对个体患者诊治决策之间的紧密融合。Fletcher RH 等在论证临床流行病学与流行病学两者关系时，颇为精辟地指出：临床流行病学与循证医学本身源于临床医学、流行病学两大学科，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”则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，这种学科间彼此交叉和融合，必将有利于学科本身的发展。

三、临床流行病学与循证医学的学科特点

临床流行病学和循证医学是以临床医学研究为基础、交叉融入了流行病学、卫生统计学、卫生经济学、社会医学等多学科的一门临床医学基础学科，立足于临床，旨在解决临床科研与临床实践问题。学科特点具体如下。

（一）力求临床研究结果的真实性与可靠性

临床流行病学的精髓还是在于强调临床研究结果（证据）的真实性与可靠性。在临床医学研究中采用科学的方法学，强化科学研究的设计和实施，力求消除各种偏倚及混杂因素对研究结果的影响，确保研究结果是真实情况的呈现，即研究结果真实反映了疾病的发生、发展和转归规律，使得研究结果满足真实性（validity）和研究结论具备可靠性（reliability）。这些真实可靠的研究结果和结论（证据），对指导临床的循证诊治决策有重要价值。

（二）立足于临床又服务于临床

循证医学本质上属于临床实践活动，当然其临床基础包括实践活动主体的临床医生、服务主体的患者、遵循的最佳证据、筛检和使用证据的理论与方法以及实践活动的医疗平台和外部环境。

1. 医生 高素质的临床医生是临床研究与临床实践的主体。良好素质表现在患者诊治时，既能善于利用个人临床技能和丰富的专业知识，发现临床问题，同时也善于使用最佳证据解决问题。如果没有精湛的临床技能，再好的证据也难用于患者的诊治。

2. 患者 患者是临床实践的服务主体。接受医生的治疗是患者出于对自己健康的渴望，除此之外，非常重要的也是出于对医生的信任。任何循证临床实践均需要患者的合作与配合，如果患者缺乏对诊治医生的信任，则依从性就不能保证。

3. 最佳证据 最佳证据乃是来源于现代临床医学的研究成果，这些成果经过严格评价而确认是真实的、有重要临床意义且又有实用价值的，方为最佳证据（best evidence）。最佳证据乃为循证临床实践的“武器”，是临床上解决患者问题的最新和最佳手段。而这种证据的获取，则是针对临床面临的具体问题，在全球范围内所有的相应最新研究成果中，应用科学的方法去检索、分析与评价，进而获得最新、最佳的证据成果，最终用于解决具体临床问题。

在获取的诸多研究成果（证据）中，由于研究设计、方法的缺陷和不足以及研究条件的限制，需要从中甄别出最佳的证据。而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以及临床研究质量的评价原则。因此，在循证临床实践中使用“最佳证据”时，一定要结合临床实际、批判性的加以采信。当对某种（些）证据有疑问时，应追溯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，这样方可避免因误导或误用，给患者带来伤害。

4. 医疗平台和外部环境 开展临床研究和循证临床实践都要在具体的医疗平台上实施。地区

笔记

经济的差异、医院级别的不同，设备条件和人员技术水平的迥异等，均会影响研究结果和实践效果。如某些治疗措施和方法的疗效很好，但需要借助一定的设备或掌握了相当难度技术的医生才能实施，若技术条件或人员缺乏，则难以进行。因此，开展临床研究和进行循证实践应结合具体的医疗平台，具体情况具体分析，不宜盲从或教条化。

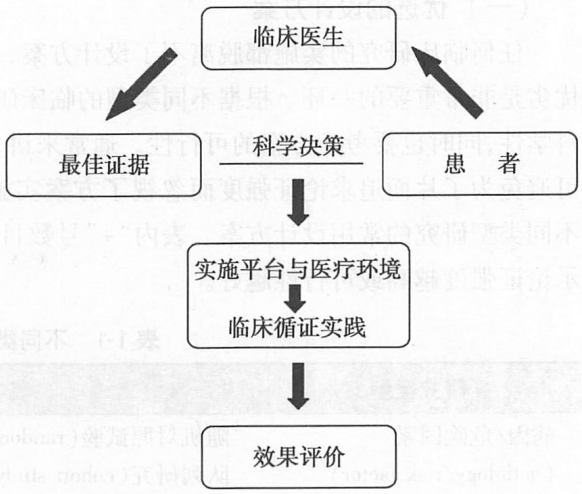


图1-1 循证临床实践示意图

综上所述，临床研究与循证临床实践首先必须准确地发掘患者存在并需要解决（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（证据）进行严格的评价，以找出“最新最佳证据”；再针对患者的具体情况，权衡利弊，将最佳证据用于循证医疗决策并付诸实践；最后分析与评价最终结果，及时总结成功的经验和失败的教训。这就是实践循证医学的“五步曲”（图1-1）。

第三节 临床流行病学与循证医学的共同方法学基础

临床研究的复杂性源于患者相关的临床特征，即疾病病程、心理状态和社会经济状况等千差万别，同时机体不同系统、器官病变间又相互影响，造成罹患同一疾病的不同患者，其临床表现大相径庭。面对患者复杂多样的临床状况以及临床资料采集的多源性，倘若没有科学的方法保障，则难以识别及防止各种已知和未知偏倚因素的干扰，临床研究结果以及疾病诊治的科学性和可靠性也就难以保证。倘若采信了这些研究结果指导临床实践，将会导致决策失误，错失最佳诊治良机。

特别是当前进入医疗大数据时代，基于回顾性医疗数据库的真实世界研究，异军突起，由于这些医疗数据库并非是为临床研究而专门建立的，在设计上存在先天缺陷，同时现有资料的收集，必然存在一些偏倚和混杂因素，因此，结果的真实性很受质疑。即使进行前瞻性研究，倘若存在严重的设计缺陷，偏倚和混杂将难以避免，使得真实世界研究结果并不一定是真实的。因此，针对临床医学研究中的复杂情况，临床流行病学家创造性地将临床研究资料的特殊性与流行病学和统计学的方法学结合，建立了以设计（design）、测量（measurement）、评价（evaluation）为三大核心的临床科研方法学（DME），应用于复杂的临床医学研究之中，成为临床流行病学与循证医学的共同方法学基础。

一、设计

临床研究设计主要围绕以下要素：①疾病患者/人群（patients/population，P）：即研究的目标疾病及其患者人群，来自何处的医院、采用何种诊断标准及有无明确的纳入/排除标准、需多大的样本量、受试者是否愿意参与研究、依从性又如何等都需要明确和说明。②干预措施（intervention，I）：设置是否有科学依据，特别是有效性和安全性的前期试验依据，具体的应用方法、时机、疗程如何，有否同时采用其他辅助疗法等。③对照（control，C）：对照组受试对象的入选标准和方法是否与试验组可比，同期或非同期对照，对照组接受的干预措施是安慰剂或其他有效干预措施，依据如何等。④结局（outcome，O）：反映有益（有效）及不良反应（有害）结果的指标依据，判断方法以及校正与否等等。⑤研究设计方案（study design，S）：是何种方案，该设计方案的科学性与可行性如何。⑥时间因素（time，T）：干预措施的疗程需要多长时间，是否能满足药效或不良反应显效的时间要求，即达到组间最小效应差异（minimally important difference，MID）所用的时间。对于某些疾病设定的追踪观察时间，应结合疾病自然病史的时间规律。“PICOST”是上述英文名词第一字母的缩写。牢记“PICOST”，对进行临床研究以及循证评价将是非常有益的。

（一）优选的设计方案

任何临床研究的实施都脱离不了设计方案，一个研究成功与否以及研究水平的高低，设计方案的优劣是非常重要的一环。根据不同类型的临床研究及其研究条件优选设计方案，不仅要注重方案的科学性，同时也要考虑方案的可行性。通常来讲，在可行性的基础上选择论证强度最强的设计方案，可避免为了片面追求论证强度而忽视了方案实施的条件，导致研究难以为继的局面。表 1-1 列出了不同类型研究的常用设计方案。表内“+”号数目表示论证强度高低及可行性大小，(+)数目越多则表示论证强度越高或可行性越好。

表 1-1 不同类型研究的备选方案

研究类型	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)	单入口诊断队列研究	++++	+++
	双入口病例-对照研究	++	+++
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)*	±	++++

* 含横断面调查(cross-sectional study)

（二）合理纳入的研究对象

需要采用临床公认的诊断标准对研究对象作出正确的诊断，确保研究疾病分类的准确性。同时按研究设计的具体要求，设置严格、合理明确的纳入标准及排除标准，保证研究对象的基线一致性和减少非研究因素对结果的影响。由于临床研究的场所多选在医院，研究对象也主要是医院就诊的患者，应根据研究设计要求，连续性纳入研究对象，对于患病率高的慢性疾病(如高血压的治疗研究)，研究对象也可从患病群体中随机抽样产生。

样本量是临床研究所需的、基本的研究对象(患者)例数。基于该样本量的研究结果基本可以代表研究对象总体的客观规律和效应特征，又不会出现研究结果与总体实际情况出入过大的风险。样本量过小，研究结果可能出现假阴性，样本量过大，不仅延长了研究时间，还可造成人力资源浪费和成本增高，研究难度加大、甚至意想不到的干扰。样本量通常根据有临床价值的效应值假设，容许的错误概率水平(包括Ⅰ型错误和Ⅱ型错误)，以及研究设计方案的性质，选择相应的样本含量估算公式。

（三）具有创新性的试验措施

有创新性的试验干预措施才具有推动相关医学进步的意义和价值；否则，低水平的重复性研究不仅价值有限，并可造成人力、物力等资源的巨大浪费。同时，试验干预措施还应具备有效性的科学依据，安全性也是需要考虑的重要方面，接受试验措施的对象应有安全保障。

（四）合理确定的试验观察期

临床试验观察期的时间长度要根据试验终点指标而设置。终点指标可以是痊愈、死亡、有效、无效等，大多数研究对象达到预期终点所需时间即为试验观察期，应结合疾病本身的自然病程和其他药物治疗反应的临床过程等而定。对于药物临床试验观察期，应基于生物学及前期临床试验依据，采用 MID 原则确定。观察期过短易致假阴性结论，过长则导致资源浪费。

笔记

（五）应设置防控偏倚的具体措施

由于临床研究的复杂性，各种已知和未知的偏倚因素不可避免地存在于研究的各个环节，偏倚不仅扭曲研究的结果，甚至颠覆研究结果，造成结果失真。因此在设计中要对研究中可能存在的偏倚因素进行评估，并相应建立具体的防控偏倚预案，确保研究结果真实可靠。

（六）合理选用的统计分析方法

应根据拟用的研究设计、设置的测试指标以及干预措施可能产生的预期结局及其资料类型，正确地选用统计分析方法。临床研究的资料有定量、定性的，也有等级的；有的是配对的，或者是非配对的；有的需要多组比较，也有的仅在两组间比较；有的需要作单因素分析，也有的需要作多因素分析。总之，不同类型的研究资料，宜采用不同的统计学方法，予以分析比较。统计学方法的正确选择对提高研究质量非常有帮助，选择不当，常可导致统计分析错误而误导结果。另外，所涉及的统计分析方法应在研究设计阶段根据研究的检测指标和资料类型预设，而不是在研究结束之后才对数据资料进行事后分析。

（七）科研道德和伦理要符合规范

临床研究的对象是患者，应享有知情权。因此，应将研究的目的、参加试验可能的获益和伤害包括不良反应以及研究者的责任和受试者的义务等告知受试者本人或其监护人，让其评估、理解并按自己的意愿决定是否参与。不应隐瞒受试者在试验中可能发生的风险或夸大可能的获益。要求在研究的设计和执行的各个环节都要充分考虑和尊重患者意愿，维护患者的切身利益。

这七个方面是临床研究设计方案时需要重点考虑的要素，也可以说是方案设计的基本框架。

二、测量

用某种方法或指标来发现、确定和计量患者接受药物治疗或者其他治疗后产生的效应，或者用某种方法或指标发现确定和计量人体对某种致病因素的反应，这些发现、确定和计量治疗效应或致病效应的方法统称为测量，测量的结果将作为治疗效果或诊断的依据。为了准确度量致病或治疗效应，需要注意下列几点。

（一）试验措施或暴露因素的效应性和可度量性

作为研究的暴露因素或治疗措施，其本身应具有产生致病或治疗的效应，且这种效应能被客观确定并被临床及实验室检查指标量度。

（二）测量方法的敏感度和特异度

测量方法应具有良好的敏感性，能够呈现致病或治疗效应，并且有良好的特异性对这些效应予以确定。否则，会发生漏测（诊）或误测（诊），造成测量结果失真。

（三）测量指标的临床意义或实际价值

临床效应的测量指标，包括定量、等级和定性的。定量数据如血液生化指标、身高、体重等，等级数据如病情轻、中、重，定性数据死亡、存活，有效、无效等等。上述数据类型中，尽管定量数据精确性最高，但在临床研究中，一般需要根据公认的临床标准，将其转化为具有重要临床价值的、也是患者最为关心的定性数据，如有效或无效等。同时在临床研究中，非常重视结果的临床意义与统计学意义的综合分析，如治疗效应量或治疗有效率是否同时具有临床意义和统计学意义，假如二者皆有，则结果有实用价值，倘若二者皆无，则结果价值不大。对临床研究而言，更看重其临床意义，若只具有统计学意义但缺乏临床意义，则研究结果的实用价值有限。

三、评价

临床研究结果是否真实可靠、临床意义和实用价值有多大、研究结果能否适用于临床实践及其适用程度如何等等，既是研究者对临床研究的自我评价需要回答的系列问题，同时也是临床医生能否采用他人研究成果作为循证决策证据、进行严格评价的系列问题。

NOTE

（一）研究结果的真实性（validity）

无论研究得到的是阳性结果、还是阴性结果，都需对研究结果作出科学的解释和结论，明确其真实性及其可信程度。

真实性评价贯穿于研究全程的各个环节，如设计方案的性质和论证强度如何以及是否存在缺陷；是否设置对照组以及设置是否合理；研究对象的诊断标准是否为金标准、纳入/排除标准是否合理恰当、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措施；受试者依从性如何；对相应的试验观测指标及资料所采用的整理、统计分析方法是否正确等等。

（二）临床重要性（clinical importance）

倘若研究结果的真实性良好，进而需要评价研究结果的临床意义和价值。具有临床价值的研究结果不仅提高人们对疾病及其治疗规律的认识，而且可作为具有指导临床实践的循证证据。当然临床重要性的评价需要通过具体量化的指标来呈现。

量化结果重要性的指标有多种，主要为事件发生率（event rate）及其差异变化，如病死率、生存率、治愈率等，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-益需治人数，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-害需治人数，NNH），基于这些指标还可进一步假设检验和区间估计。此外，用于诊断性试验评价亦有一系列量化指标（见本教材第九章）。这些量化指标都能清楚地表明有效性以及不良反应程度，便于临床评价。

这里要强调的是组间差值的大小究竟有无临床意义，如无临床意义，即使具有统计学意义的差异，实际应用价值不大。因此，统计学意义并不能完全代表临床意义。所以，对于任何临床研究的结果，务必要注重临床价值，倘若认为研究结果具有临床意义，则必须应用正确的统计学方法进行显著性检验，即肯定结果的真阳性、真阴性概率以及检验效能大小。当研究结果既有临床意义，又有统计学意义时，可下肯定性的结论；如仅有临床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型错误和检验效能水平。

临床研究的结果，必要时还应进行卫生经济学评价，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，分析它们可产出的健康效益，以肯定那些成本既低、效果又佳的研究成果，加大推广应用力度。

（三）研究成果的适用性（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，因而针对性强。因此，对于研究成果之实/适用性，要分析它们有无实用价值、有多大的实用价值，利弊比如何，在什么样的医疗环境和条件下可以采用或推广，宜作实事求是的评估。切不可脱离自己的环境、技术条件和患者的实际情况，盲目地接受或教条化地推行（见本教材第十三章循证个体化实践）。

上述三性评价也是循证医学“五步”中第三步（证据评价）以及第四步（证据应用）的核心和关键。

第四节 临床流行病学与循证医学的作用和价值

一、临床流行病学与循证医学的作用和价值

（一）临床流行病学为临床研究提供科学的方法学、催生高质量的成果

临床流行病学指导临床研究，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，BOD），以掌握危害人类健康的重大疾病；进而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；再进一步通过研究实践，

收集有关研究成果予以综合评价，再付诸实践，最后验证疾病危害人类健康之疾病负担减轻的程度。在这样一个螺旋式循环下，可不断地创造新方法、新经验、新措施、新成果，服务于临床医学。临床流行病学的诞生和发展弥补了传统临床研究方法上的缺陷和不足，促进了临床研究水平的进步和产生高质量的研究成果，更好地服务于循证医学实践和提升医疗质量。

临床研究的最新成果、特别是高质量的成果，已经作为指导临床诊断、治疗和预防的最佳证据（best evidence），诸如一些重要的临床诊治指南的推荐意见均基于最新最佳的研究成果（证据）。

（二）循证医学促进了临床实践的进步和医疗水平的提升

21 世纪科学技术的快速发展，特别是临床医学领域众多新药物、新方法、新技术不断地问世，临床医生如何选择这些新技术、新方法和药物，面临的不仅仅是机遇，同时也是挑战。一方面多数新技术、新方法、新药物，的确可以提高临床诊治水平，促进临床医学水平的进步，使患者获益；另一方面少数非但无效，并可能给患者带来严重的不良反应、并发症，甚至还有一些曾被认为有临床价值且被普遍接受的治疗措施或药物，后经严格的临床研究证明是无效或无益的。因此，临床医生应学会从众多新成果中去辨别真伪，以提高自己临床诊治水平，而临床流行病学与循证医学则为此提供了针对疾病病因、诊断、治疗及预后等一系列评价原则、标准与方法。

循证医学对临床工作的推进表现在两个方面，一是通过对现有的临床研究成果进行检索、系统评价/meta-分析等，获得对指导临床工作真正有益和有用的证据；二是善于从临床诊断、防治、预后、不良反应、经济分析、生存质量、卫生技术和医疗质量评估等具体的临床工作中入手，围绕具体问题，进行循证医学实践、以解决这些问题。

（三）临床流行病学与循证医学有助于培养高质量的医学人才

临床流行病学与循证医学的精髓在于培养医学生从事临床医学研究和临床医疗实践的能力，使其掌握临床研究与实践的基本理论、知识和方法以及养成严谨的研究和工作素养，形成批判性、创造性、科学性思维模式，为今后从事临床研究和临床实践打下坚实的基础，这对培养高质量、高素质的医学生，使之在毕业后的临床医疗实践中，更好地做好临床医疗和科研工作，对发展和提高临床医学的水平具有重要的战略意义。

二、临床流行病学与循证医学的常见误区

近几十年来，循证医学作为临床实践的一种新思维模式，日益得到普及和推广。循证医学的新理念正越来越为广大医务工作者所接受并付诸临床实践之中。当然，在此过程中，不可避免地出现了一些误区，如把“循证医学”神话，认为它是解决所有临床问题的“灵丹妙药”；再如，“言必称证据”，把证据神话，忽视了临床医生的经验、患者价值观、医疗环境和技术条件等在循证决策中的作用和价值。正如英国全科医生 Des Spence 所述，他认为“目前大部分的证据和研究都是由药企控制并且资助的，循证医学就像一把已上膛的枪，逼着临床医生‘最好乖乖的按照最佳证据去做’”。当然该说法比较偏激，但确实存在这种倾向，应引以为戒。

（一）证据至上论，证据缺乏时就不是循证医学

循证医学的精髓，是将最佳的外部证据、医生自身的经验和患者意愿结合起来，进行临床决策，三者同等重要。随机对照试验和系统综述/meta-分析不等于循证医学，而只是外部证据的体现。在缺乏可靠的外部证据时，医生的经验往往更加重要。循证医学的一个误区就是往往忽视专家意见和医生经验，而过分看重所谓的客观证据。在临床实践过程中，经常会遇到证据不足、缺乏甚至相互矛盾的情况。此时此刻，应充分发挥临床医生的主动性，临床医生的经验、心得同样是证据，对病患诊治仍具有重要的指导作用，同时，也应基于循证医学实践的理念，不满足于经验证据，开展临床研究、生产更高级别的证据，指导实践。

（二）循证医学就是系统评价/系统综述

系统综述是一种全新的二次研究方法，通过系统综述/评价可以实现证据的汇总和集成，成为循

证医学实践的重要证据源之一。但系统综述/评价不等于循证医学。系统综述尽管拥有诸多优势，但也有一定的局限性。倘若用于系统综述的原始研究质量差，则系统综述结果的真实性会大打折扣；另外大量的系统综述过度关注干预的有效性，而往往忽视安全性、实用性、接受度及效果成本等方面的评价，由此形成的决策较为片面、局限。

（三）大样本多中心随机对照临床试验就是循证医学

随机对照临床试验被尊为临床试验的金标准方案，是因为其研究方案严格遵循了随机、对照、盲法等科学原则。倘若在实施过程中采取质量控制措施，偏倚风险小，确实能为循证医学实践提供高质量证据。但要注意随机对照临床试验可能与临床实际脱节，如患者纳入/排除标准过于苛刻，研究时间有限、随访时间不长，结局指标为短效/中间替代指标，研究方案僵化等，导致结果外推受限。加之大部分临床试验结果来源于发达国家，在疾病负担、诊断标准、健康服务系统、医保模式等可能与发展中国家迥异，进一步加大推广应用难度，适用性更差。况且在外科领域，由于伦理原因难以进行随机对照临床试验，缺乏相应随机对照试验研究的证据，使循证医学成为“无米之炊”。

上述误区以及对循证医学的质疑，实际上真正质疑的并非循证医学体系本身，循证医学的理念是先进的，但它往往被不恰当地利用，甚至是误用。因此，实践循证医学，应还原其实践医学的本来面目，不宜将其神话、认为它是解决所有临床问题的“灵丹妙药”；同时要清醒认识到循证医学本身的局限性，循证医学同样是需要与时俱进的，目前来自真实世界的临床研究证据，业已成为循证医学证据的重要补充。最后，引用国际临床流行病学及循证医学创始人 David Sackett 对循证医学实践者的四项要求作为本章的结束语：①必须做踏实地临床基本训练，正确地收集病史、查体和检验，充分掌握患者的真实情况，方能发掘临床问题；②必须将循证医学作为终身自我继续教育的途径，不断丰富和更新知识、提高自己的理论水平；③保持谦虚谨慎，戒骄戒躁；④要有高度的热情、责任感和进取精神，否则就要成为临床医学队伍的落伍者。

（刘续宝　康德英）

笔记

第二章　临床研究与实践问题的构建

临床医师在临床实践中经常会遇到许多需要解决的实际问题，如针对某个患者应该选择哪项检查才能比较准确地诊断或排除某个疾病？应该进行何种治疗才能更好地改善这个患者的预后？实际上，临床医师对患者的诊治过程就是一个不断提出问题、寻找最佳的解决方法、直至最后解决问题的过程。这样的问题，被称为临床实践问题。对患者实践循证医学（evidence-based medicine，EBM）的第一步就是找出一个待解决的实际问题，将其构建成为一个需要回答的临床实践问题。因此，能否找准患者亟须解决的实际问题，对于循证临床实践至关重要。

另一方面，临床医师在进行临床研究时需要通过研究回答一个科学问题，如某个疾病的病因或危险因素是什么？影响这个疾病预后的因素有哪些？这样的问题，又被称为临床研究问题。构建一个恰当的科学问题是开展科学研究的第一步，它对临床研究也同样有十分重要的作用。

第一节　概　　述

一、找出临床问题的重要性

（一）实施循证医学的第一步

临床医师应善于在临床实践中对患者进行观察、发现问题并提出问题。只有提出问题后，临床医师才会带着问题去寻找相关的证据，而后根据可信度最强的证据并结合自己的临床经验、患者意愿等因素提出恰当、合理的解决方案，最终解决临床问题并使患者获益。因此，找出问题是循证临床实践的起点，找不准问题就不可能提出恰如其分的临床实践问题。倘若相应的证据缺乏或收集到的证据科学性不强时，临床医师可以将这个临床实践问题转化成为临床研究问题，从而进一步提出研究计划并加以实施，通过临床研究回答这个临床问题，为临床实践提供证据。

（二）医学发展的需要

有的临床医师认为自己在医学院学到的知识足以回答临床实践中的所有问题而不再需要继续学习，这种一劳永逸的想法是错误的。在临床实践中，如果临床医师不能结合患者的实际情况提出有意义的临床问题，不去查阅文献、寻找和评价最新的证据，不对自己的临床知识加以更新，仅凭既往经验，患者就难以得到最好的诊断和治疗。作为研究者，如果临床医师对临床实践不进行思考和总结，临床医学也难以取得进步。对一个临床问题的认识是在不断升华中逐渐逼近“真相”。

临床医学作为一门实践科学，临床医师只有在临床实践中不断地提出问题并通过临床研究回答所提出的问题，才能促进临床医学不断发展进步。

（三）循证医学所赋予的任务

循证医学以解决与患者罹患疾病相关的重要临床问题为宗旨，因此在循证临床实践中必须抓住患者身上关键的、待解决的临床问题。这些问题往往关系到患者的生死安危，而仅凭个人临床知识和经验常难以准确把握和回答这些问题。临床医师必须以这些问题为主线，按照循证医学步骤收集和评价现有的临床研究结果，选择最佳的证据回答和解决所面临的关键临床问题。可以说，循证医学的第一关键就是找准个体患者存在的、必须回答的临床重要问题。

二、找准临床问题应具备的条件

（一）临床责任心

对患者要有高度责任感和同情心，设身处地的以患者为中心思考问题，这样才能在与患者的交流中仔细观察、认真思考，从而发现重要的临床问题。

（二）丰富的医学基础知识和临床医学知识

人体各个系统的疾病都有其自身的客观规律，临床医师如果不了解疾病的病因、发病机制和临床表现，不熟悉各种诊断试验和辅助检查的特性和适应证，不了解各种药物的治疗机制、药理作用和可能发生的不良反应，那么当他（她）碰到一个具体的患者时就不可能提出恰当的临床问题。因此，具备系统扎实的医学基础知识和临床知识是找准临床问题的重要基础。

（三）具有一定的人文科学及社会、心理学知识

随着医学模式的改变，人们发现许多疾病的发生与心理、精神因素有关。也有一些疾病的发生虽与此关系不大，如慢性肝病、肿瘤等，但患者对疾病的认识和心态会影响疾病的发展和预后。因此，临床医师不仅要了解患者对疾病的想法、期望及忧虑，也要了解患者的社会经济状况和家庭负担等，这样做有利于发现这方面的临床问题。临床医师也只有掌握一定的人文科学、社会和心理学知识，才能与不同性格的患者顺利地进行沟通，发现患者的心理问题并及时加以疏导，其本身也是疾病治疗的一部分。

（四）扎实的临床基本技能

包括如何接触患者、采集病史、全面的体格检查以及对诊断试验选择和鉴别的能力。临床医师在临床实践中必须弄清患者的病史、认真详细地进行体格检查、掌握相关的辅助检查结果。在此前提下，临床医师才可能找出患者最迫切需要解决的临床问题。

（五）临床综合分析能力和判断能力

临床医师应用已掌握的医学理论知识和临床经验，结合患者的临床资料进行综合分析、逻辑推理，从错综复杂的线索中去伪存真、去粗取精，找出实践中的主要矛盾并加以解决，这反映出临床医师的综合分析能力和判断能力，是准确把握临床问题的必备条件。

上述五点是寻找和提出临床问题的重要必备条件，缺一不可。缺少其中任何一点都不利于找准患者的临床问题。

第二节　循证问题的构建

一、临床问题的类型

从医学生直至高年资临床医生都可进行临床研究以及参与循证临床实践活动，但鉴于他们的经验、阅历不一，视角与水平不同，即使在临床实践中面对同一个患者，其发现和提出的临床问题（clinical question）迥异。这些问题大致可归纳为以下类别：

（一）一般性问题

主要由初学循证医学的实践者提出，这些问题大多属于“背景性”问题。如：

1. 涉及患者的一般知识性问题　如患者的性别、年龄等。

2. 涉及所患疾病的基本问题　某个具体患者存在的临床问题，如在什么地方、何种环境下发病、何时发病、如何发病及发病的相关因素是什么、主要的临床表现又有哪些等。

（二）特殊的临床问题

这是临床医师在诊治患者的过程中，特别是充分掌握病史、体征、相关检查结果之后，通过临床综合分析从专业角度找到的问题，一般称为“前景性”问题。这些问题不解决则必然影响对患者的正确处理。为准确把握个体患者存在的特殊问题，应以患者为中心、结合个体患者的具体情况以及主观意愿，提出问题。例如，罹患同一个疾病但年龄不同的患者所关心的问题是不同的。一项有1012名乳腺癌妇女参加的研究发现，70岁以上患者最关心的是癌症治愈和转移的可能性，而小于50岁者关心

笔记

的是治疗对其性功能的影响；有阳性家族史者最关心的是该病是否有遗传性等。因此，临床医师应针对不同患者的具体情况提出临床需要解决的问题。这些问题与临床息息相关，涉及诊断、病因、干预和预后等方面。

1. 诊断问题　例如，一个肝硬化患者近期腹水明显增多，对其提出“腹水有无感染”就是一个十分重要的临床诊断问题，如果不能确定是否合并自发性腹膜炎，就无法对其进行正确的治疗。

2. 病因问题　如何进行干预往往涉及病因、危险因素的暴露等。例如，对一个消化性溃疡患者进行治疗时，必须先针对病因提出问题，包括患者有无幽门螺杆菌感染、有无服用非甾体抗炎药病史、有无应激状态等，这些都影响到治疗方案的选择。

3. 干预问题　同一个疾病的干预措施往往有许多种，每一种措施又各有利弊，这时就存在如何比较抉择干预措施的问题。例如，对恶性肿瘤患者是采取手术还是介入性治疗或放化疗，不仅要根据病情进行分析，将各种干预措施的利弊进行比较，还要考虑到患者的经济能力以及与患者及家属沟通的结果。

4. 预后问题　预后和最终结局如何，是循证医学的实践者追求最佳结局最感兴趣的问题。结局可以是症状、体征的改善或者生存率、死亡率和致残率，使用不同的结局指标找出的问题也不尽相同。

总之，以上四个环节是一个有机整体，循证医学的实践者在发现临床问题时一定要将其牢牢掌握。

二、临床问题的提出

（一）临床问题的提出形式

现在许多医学院校已开展“以问题为中心的学习（problem-based learning，PBL）”，这里的“问题（problem）”是指患者存在的一种症状或体征，如黄疸、腹水、胸痛等。而本章的“临床问题”是一个可以回答（answerable）的临床问题（question），如“为这个黄疸患者选择辅助检查时，腹部B超和CT哪一项更好？”。尽管“problem”和“question”均可翻译成“问题”，但两者的内涵完全不同。下面讨论临床问题（question）的提出形式。

1. 一般性的临床问题　一般性问题是与患者或患者罹患疾病有关的一般知识性问题，由以下两部分构成：

（1）问题由疑问词（谁、什么、何时、何地、怎么样、为什么）加动词组成：这些问题的答案在患者入院时常可通过询问病史和体格检查得到。例如，每一个主诉都应该包括症状发生的部位、严重程度、数量（如出血量）、起病情况（急性还是慢性、持续性还是进展性）、在什么情况下发生、加重和缓解的因素以及其他关联症状等，了解以往是否发生过与主诉相同的情况，曾经做过哪些检查，是否曾经做过治疗以及如何治疗，对预后有意义或对治疗有影响的既往史，相关疾病的治疗情况等。如以呕血为动词时，就必须弄清楚是谁呕血（患者的性别、年龄等特征）、呕血的性质（颜色、量、次数等）、何时何地发生呕血、呕血时患者有无其他症状以及什么是发生呕血的主因、诱因及其基本病变等。

（2）问题涉及一种疾病或疾病的某一方面：例如，“什么原因引起发热？”“急性胰腺炎通常在何时发生并发症？”等。

2. 特殊性的临床问题　在临床实践中，患者与医生都会在诊断、防治、预后、病因等各个方面提出许多需要解决的临床问题。例如，患者常常会问医生“我患的是什么病？”（诊断问题）、“我为什么会患这个病？”（病因问题）、“这个病应该用什么方法进行治疗？”（治疗问题）、“这个病对我的健康有多大影响，会不会影响我的寿命？”（预后问题）。医生在诊治不同疾病或同一个疾病的不同患者时，提出的问题可能不尽相同，临床医师可以对患者发生的每一个症状或体征提出问题。例如，上述呕血的患者在就诊时亟须解决的主要问题是及时止血并弄清呕血的原因；出血停止后患者出现计算力下降、昼夜颠倒、扑翼样震颤等表现，这时患者亟须解决的主要问题就变成了弄清是否出现肝性脑病并对此采取积极的治疗措施。也可以在准确采集病史和体检的过程中发现并提出新的问题。例如，一位中年男性因发现黄疸就诊，临床医师在体检时发现无痛而肿大的胆囊（Courvoisier征），提出的问题即可为“该体征对于梗阻性黄疸和肝细胞性黄疸的鉴别诊断是否有意义”，进一步可提问“其对于鉴

NOTE

别肝外梗阻的原因，即结石引起抑或肿瘤引起是否有帮助”等。

患者本次住院或门诊就诊时需要解决的问题以及住院后因病情变化而产生的新问题，归纳起来包括以下几个方面：

（1）诊断方面的问题：初学者在诊断方面常提出的问题是某个症状、体征或某项辅助检查对某个疾病的诊断价值，即关于诊断试验的敏感度、特异度及似然比等问题；而有多年临床工作经验的医师常提出的问题是某项检查对鉴别诊断的意义。通过询问病史和体格检查，临床医师会有一个诊断假设。为了证实该假设，医生可能会进行一些实验室检查或其他辅助检查来肯定或排除该诊断假设。此时可以针对诊断试验的敏感度、特异度、似然比等提出问题，针对其正确性、可靠性、可接受性、费用及安全性等方面也可提出问题。例如，对一个呕血的患者为了寻找出血的部位和原因是否应作急诊胃镜检查？仅此一点就可以提出许多临床问题，如“急诊胃镜检查对诊断上消化道出血的敏感度和特异度如何？”“急诊胃镜检查对该患者带来的风险有多大？”“对肝硬化患者和非肝硬化患者带来的利和弊有无差别？”“有无其他可供选择的诊断措施？”等。

（2）治疗方面的问题：如何选择利大于弊的治疗手段？如何从经济学成本-效果的角度选择治疗方案？特别有意义的是针对目前的常规疗法提出质疑，如“根据胰腺癌患者目前的病情可以采用什么治疗方法？”“该治疗方法的有效性如何？”“有什么不良反应？”“还有哪些替代的治疗手段？”“哪一种治疗方法更为有效而花费最少？”“患者对治疗的依从性和可接受性如何？”等。上述胰腺癌患者可以根据疾病的严重度选用外科手术或姑息疗法。若采用后者，为了减轻黄疸的危害及提高生存质量，可以通过内镜放置内支架引流减轻胆道梗阻。可以进一步提出的问题有“安置金属支架，还是塑料支架，何者为佳？”“为加强支持治疗是否给予患者静脉内营养？”等。

（3）病因方面的问题：包括“怎样识别疾病的原因及危险因素？”“其发病机制是什么？”等。例如，对于胰腺癌患者可以提出的病因问题包括“发病的原因是什么？”“有无遗传因素？”“发生胰腺癌的危险因素是什么？”“是否与喝咖啡或饮酒有关？”等。

（4）预后方面的问题：如何估计患者的临床病程？如何预测可能发生的并发症和最后结局？针对不同的结局测定指标可以提出不同的预后问题。例如，预防食管静脉破裂再出血的干预措施对“再出血的发生率”和“患者的生存率”两种预后指标是否有不同的效果？

（二）临床问题的提出策略

临床医师在临床实践中遇到患者存在的问题，想解决却存在知识、能力不足时，就要找准问题并记录下来，通过自己的临床思维对其进行整理和排序，先抓好关键问题并作出如何解决这个（些）问题的策略计划，有的放矢地查阅文献、进行文献评价并选择最佳证据来解决患者的问题。在提出临床问题的策略上，要注意的是：①涉及的问题一定与患者的诊断、治疗和预后密切相关；②涉及的问题一定与循证临床实践、提高医疗水平最为相关；③涉及的问题一定是临床实践中最有用、也是最重要的；④涉及的问题往往也是循证临床实践中最为常见的。

如果在平时的临床工作中对各种不同患者的临床问题能够做到日积月累并不断用最佳证据予以解决，终身坚持，必成名家。

（三）临床问题的梳理

在所有需要提出的临床问题（question）中，对学生而言，一般性的背景问题可能多一些；对有经验的临床医师来讲，前景问题可能多一些。有些问题不需要进一步查资料就可以回答，但就教学而言，有必要让学生了解如何针对患者的实际情况提出可回答问题。

1. 开具教育处方 为了进一步了解患者的情况（一般性问题）和对患者进行进一步处理（特异性问题），在上述工作的基础上可以开出教育处方（educational prescription），即提出患者亟须解决的临床问题（question），以进一步寻找答案。教育处方提出的可回答问题必须十分具体。例如，在讨论一位重症胰腺炎患者的治疗措施时，类似于“重症胰腺炎患者该如何治疗？”这样的问题，太过宽泛，如果据此去检索文献，会有成千上万篇文献，最终无法归纳总结出回答该问题的答案。因此，提出的问题必须要具体到某一项措施，如有人提出“对重症胰腺炎患者是否需要用静脉内营养或肠内营养”的问

笔记

题，结合患者实际就可以这样提出问题："全胃肠外营养和肠内营养对于急性重症胰腺炎在降低感染发生率、减少并发症发生率、缩短住院时间和降低死亡率方面，哪一种方法较好？"为了回答这个问题，可以用检索词"parental nutrition or TPN and enteral nutrition and acute severe pancreatitis"检索文献，寻找答案。由此可见，构建的循证问题必须包括对象（某种疾病、症状或患者）和需要比较的措施，这样查找出来的结果，才能对临床医师决策有所帮助。

2. **提出临床研究问题**　临床实践是临床科研选题的丰富源泉，在日常临床实践中临床医师无时无刻不面对许多关于诊断、治疗、病因、预后等方面的问题，倘若相应的证据缺乏或收集到的证据科学性不强时，临床医师可以将这个临床实践问题转化成为临床研究问题，从而进一步提出研究计划并加以实施，通过临床研究回答这个临床问题，为临床实践提供证据。

三、循证问题的构建模式

一个具体临床问题的构建，通常采用 PICO 格式。P 指特定的患病人群（population/participants），I 指干预（intervention/exposure），C 指对照组或另一种可用于比较的干预措施（comparator/control），O 为结局（outcome）。每个临床问题均应由 PICO 四部分构成。表 2-1 显示了 3 个临床问题的组成方式：①对于慢性肾衰竭尿毒症患者，肾脏移植与血液透析相比在生存率和生存质量上哪种方法更好？②ACEI 与 CCB 合用与单用 CCB 相比，在保护肾功能、降低血压和减少尿蛋白方面是否有更好的效果？③对于频发的尿路感染，长期小剂量应用抗生素是否能预防复发？根据 PICO 确定关键词，便于检索。

表 2-1　临床问题的 PICO 构建模式

Patient/Population	Intervention/Exposure	Comparator/Control	Outcomes
慢性肾衰竭尿毒症	肾脏移植	血液透析	能否改善患者生存质量和生存率
慢性肾病	ACEI+CCB	CCB	保护肾功能、降低血压、减少尿蛋白
频发尿路感染	长期小剂量抗生素	空白对照	预防复发

总之，一个好的循证问题的提出，需要具备系统扎实的临床基础与专业知识，深入临床实践，善于思考，跟踪本专业的研究进展，学会从患者的角度加以考虑，才能提出和构建良好的临床问题。

第三节　临床研究的选题与立题

临床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。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。

一、临床研究的特点

临床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）研究对象是患者

临床研究的对象是患者，患者是有关疾病的"载体"。罹患疾病的患者都饱受疾病的折磨，忍受着不同程度的痛苦，承受着肉体、精神及经济等方面的压力，同时由于各自的病理生理特点、心理情感状态、文化水平及所处自然与社会环境的差异，疾病的临床过程和情况十分复杂，即使同一种疾病的病理损害相似，患者的临床表现却往往迥异，这就形成了临床研究对象的复杂性。因此，临床研究对象务必要符合有关疾病的诊断标准，同时研究者还要制订恰当的纳入与排除标准，使研究对象具有相对匀质性，从而有利于研究的可比性。

（二）干预措施要安全有效

任何临床试验的治疗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，证明对患者具有安全性以及对疾病诊治的有效性，方可进行立题研究。例如，新药的临床治疗试验，一定要有药物化学成分、药效学依据、病理与毒理、药物代谢动力学等可靠的基础实验研究以及一期临床试验证据，证明其安全有效且符合相关的法律法规方可立题并启动临床试验。因此，对于缺乏科学依据且无疗效证据的药物或“措施”是不允许用患者做试验研究的。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也要保证患者的安全性。安慰剂应根据治疗性试验的具体情况和要求使用。

（三）医院是主要的研究场所

临床研究主要是在医院范围内进行，但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正常工作的患病群体进行研究。这种包括医院内和医院外患者的综合性群体研究，具有临床与流行病学研究的双重属性，也是临床研究的一个发展方向。

（四）遵循医学伦理的要求

按照世界医学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，保证无损于患者的利益。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、意义及可能发生的副作用。坚持自愿的原则，尊重患者的人格，不欺瞒研究对象。如果患者接受试验，则需要签署知情同意书，以防日后发生争议。即使在试验进程中，患者也有退出的权利。因此，任何涉及以患者为研究对象的临床试验，在正式立题研究时要高度注意医学伦理问题，必须向有关机构的伦理委员会申请，接受伦理委员会的审查，通过后方可进行研究。

二、选题与立题的原则

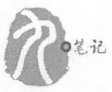
尽管国家各级机构确定的科研课题投标指南为研究者提供了研究方向和可选课题，研究人员仍然需要掌握临床研究选题、立题的原则，方能充分发挥科研人员的积极性和创造性。

（一）需要性

1. 选择重大疾病进行研究，满足社会和学科发展需要　首先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、波及地域范围广的疾病为重点课题进行研究，如恶性肿瘤、心脑血管疾病、急性传染病、人畜共患病、新生儿疾病等疾病负担十分突出的国家确定的重点防治疾病。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、高脂血症等患者为心脑血管病的高危人群；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的地区等。

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、行政管理人员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，方能作出符合中国国情的防病研究的科学决策。

2. 需要进一步明确研究重点和焦点　即具体和明确地提出期望解决的问题。例如，艾滋病的防治研究被选定为国家的重大专项课题，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、预后及预防研究等。从临床研究的角度出发，研究的重点应当是探索有效的防治措施和改善患者预后的方法，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，往往是不现实的。因此，一定要根据自己的基础、专业特长、兴趣爱好、工作条件、技术力量和经济支持等，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，否则极易造成困难的局面，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不同，可以根据研究的具体内容，以主课题为中心设立若干分课题（子课题）开展研究。分课题是主课题的一部分，与主课题密切相连。各个分课题的有机综合就构成了总课题的研究整体。

研究者要根据科学的依据对选择和立题研究的课题可能获得的科学成果、防病治病的效果和价值进行实事求是地预测，估计其成本-效果、是否会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。

（二）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，而不是跟着别人去盲目地重复或赶“时髦”。研究内容应力求居于该领域的研究前沿并具有领先水平，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有可能取得比以往组织、细胞水平的研究更为精确的结果。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，寄希望于有新的突破，否则就会造成人力、物力和财力的损失和浪费，且研究本身也往往没有什么意义。

（三）科学性

合理的研究设计方案是研究科学性的保证。根据立题研究的课题性质，抉择科学性及可行性良好的设计方案是保障研究课题获得成功的关键之一。

同时选择任何干预措施和诊断方法进行研究，都要确保其安全、有效。即使对目前尚无有效治疗的疾病进行治疗性试验，该试验药物或措施也要有科学依据，特别是药效学的实验依据。研究要有先进性和创新性，否则就没有研究的必要。

（四）可行性

选题时要根据实际情况充分考虑研究对象的来源和数量。如果一个地区或单位在规定期间内有足够的研究对象来参与研究，则该研究可行，否则需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，以保证该研究对样本量的需求。临床试验所采用的新措施或新药物都有研发时间限制，期望尽可能快地验证其临床效果和应用价值。如果因研究对象来源不足而导致试验时间延长，就有可能被更新的研究药物或措施所取代，造成原试验研究失去时效性甚至原有的价值。同时，对拟采用的研究措施、技术条件等也要考虑实施的可行性。

1. 执行研究措施的可行性

（1）仪器设备及实验室条件：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都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：要有一支具备科研设计、实施和观察、资料收集、统计分析及研究总体指导和管理等技术力量的较强专业队伍，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，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灵敏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、文化和宗教等的可接受性，临床试验主要是以患者作为研究对象，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对患者缺少足够的解释及关爱，则研究对象的依从性往往无法得以保证，研究者在研究课题立题的时候一定要充分注意这一点。

2. 技术条件、经费和伦理可行性　在选题和立题时，研究者务必要考虑经费的来源与保障，实事求是地预算研究经费，保证仪器设备、实验试剂、药品开支、劳务报酬和学术交流等必需的费用。不切实际（过高或过低）地预算经费是不利于所立课题的正常运作的。

任何临床研究课题的立题，务必要符合伦理学的原则，遵守国际公认的赫尔辛基宣言的精神。按照国际惯例，临床试验的研究课题必须经过有关机构的科研伦理委员会审查通过后方有资格立题研究。

（五）效能性

临床研究应面向广大人民群众亟待解决的健康问题。中国作为世界上最大的发展中国家，国内不同地区的经济发展水平存在着巨大差异，社会贫富差距也颇为显著，城乡医疗卫生和保健事业的水平也很不同。因此，临床研究课题的选择，应该结合我国的国情、地方特点及广大人民（不分贫富）对促进健康和有效防治疾病的要求进行考虑，以充分体现临床研究的社会公正性。另一方面，国家和地方的卫生研究经费投入是有限的，人力资源与水平也受某些客观因素的限制。因此在选择临床研究课题的时候，应该考虑将有限的资源投入到最需要解决的、关系大多数人民健康的方面，以体现卫生研究资源利用的社会公正性。

三、研究选题和立题程序

（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、基础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中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，艾滋病是威胁全人类健康的重大问题之一，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的研究，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症的解释和治疗提供了科学的依据。进一步研究有效治疗的药物挽救患者生命并改善预后。在预防研究方面，疫苗的研制、针对发病危险因素的干预等等方方面面。诸多未解决的问题，需要医学及社会学学者从中选择研究，一一突破和解决。

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，除了应用自己的聪明才智之外，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。

在复习文献时，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。这里要十分强调的是：好的文献并不一定都有阳性的结果，阴性的研究结果并不意味着研究质量差，标准取决于文献的真实性。

为研究选题而查阅文献，要充分地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

笔记

集高度相关、高质量的文献加以借鉴，切忌视野狭窄、主观片面、草率从事、结论武断。

（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、知识水平、研究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题研究。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。这里切忌贪多求全，企图通过一个研究课题来解决许多个问题，实践证明这样的想法和做法，都是脱离实际的。

立题后，应按一定规范格式和内容，规范撰写科研计划书（具体见本书第十六章）。

四、研究立题的评价标准

对于临床研究立题的评价，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

这里要考虑社会的需求性、注重国家或地区性疾病负担的现况，要重视对危害人民健康重大的疾病研究；要考虑社会的公正性，即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。

（二）研究的重点是否明确

一个科研课题应有明确的研究重点，即重点要解决的关键问题。例如：对 SARS 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？预期价值如何？这些问题应明确回答，方可论证研究的重点。

一个研究课题，若涉及的问题面宽，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和发展动态

立题研究与确立研究的重点，务必要通过文献复习，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，力求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，使立题研究更具时效性、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，照人家的方法在国内找一些患者采集一些标本，做人家做过的试验，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）是否具有可行性

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。

（七）医学伦理

临床科研务必要保障患者的安全及权益，尊重患者的人权和人格为最高伦理标准，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会所制订的赫尔辛基宣言的精神。

（刘晓清　康德英）

第三章　临床研究设计的基本要求

临床研究是以患者及患者群体为研究对象，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。与离体实验及动物实验不同，人体作为最高形式的生命体，除生物学属性存在个体差异外，在心理和社会功能属性方面也有不同；另外，就疾病本身而言，病程、病情、诊治手段等方面的差异，也进一步增加了临床研究的复杂性；同时以患者作为研究对象应考虑伦理要求。因此，临床研究的设计非常重要，掌握临床研究设计的随机、对照、盲法、重复等科学原则，将有助于生产高质量的临床证据。

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。在临床研究中，为防止来自研究者与研究对象两个方面主观因素的干扰，减少非研究因素对结果的影响而采用随机化。随机化是保证临床研究科学性的重要原则之一。随机化原则主要应用在选取样本和分配研究对象两个环节，包括随机抽样与随机分组。

（一）随机抽样（random sampling）

由于临床研究的目标人群常常是无限总体，如罹患某疾病的所有患者或发生该疾病的潜在人群，直接进行总体研究的可行性很小；另一方面，即使是有限总体，由于人力、财力、时间等因素的限制，也不可能研究所有对象，而只能从总体中抽取一定数量有代表性的样本进行研究。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中进入研究样本。随机抽样的目的是保证样本的代表性，避免选择偏倚。

（二）随机分组（random allocating）

随机分组又叫随机分配，是指在研究样本确定后，进一步采用随机方法，将研究对象以同等的机会分配进入试验组（experimental group）或对照组（control group）。随机分组是不受研究者和（或）研究对象主观意愿影响的一种手段，可以使组间的各种影响因素（包括已知的和未知的因素）分布趋于相似。因此，通过随机分组，可以提高组间均衡性，减少非研究因素的干扰，控制混杂偏倚；可以控制来自研究者分配研究对象时可能产生的选择偏倚，确保研究结果的真实性；另外随机分配结合盲法使用，可使研究结果更加客观，提高研究结果的准确性。但随机分组可能存在增加临床研究的难度，违背研究对象的个人意愿等问题。

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基本单元，即研究的独立观察单位。临床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，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，探讨家庭饮食结构与肥胖的关系，就常以一个家庭作为一个独立观察单位。另外一种情况是以个体的一部分器官或组织作为研究基本单元，如眼睛、牙齿、关节等。

（一）简单随机法（simple randomization）

简单随机法主要指采用抽签、掷硬币或随机数等方法实现的随机化，是最简单的随机化方法。抽签或掷硬币法简单易行，可用于小样本临床试验，但是因其存在随机过程不能重现、不适用于大样本研究等缺陷限制了其应用。在临床研究中常用随机数进行抽样或分组。

1. 随机数字表法　查阅随机数字表法是实现随机化的一种较为简单、实用的方法。统计学家基于概率论原理编制了随机数字表（附表），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，既可随机抽样，也可随机分组。

例如，欲将符合试验要求的20名患者，随机等分为试验组（T）和对照组（C）。首先将患者从1至20编号，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，如从附表第9行，15列开始，按从左到右，顺序获取20个随机数并依次列入表3-1，重复的数弃去重新抽取。进而按照随机数的大小排序，从排序后的随机数中，事先规定前10个随机数对应编号的患者进入试验组，后10个随机数对应编号的患者进入对照组，见表3-1。

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

2. 计算机或计算器随机法　该方法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通过随机数发生器产生随机数字。产生的随机数字是取值在0～1之间（不包括0和1本身）均匀分布的随机变量。得到随机数字序列后，可按照类似随机数字表法进行随机抽样或随机分组。

必须指出，将受试对象交替分到试验组和对照组的分配方法（TCTCTC……）不属于简单随机分组法。这是因为，研究者可根据规律预先知道下一位患者将分入哪一组，主观上可能会基于患者的病程早晚、病情轻重等进行取舍，导致研究对象选择上的偏倚。

（二）分层随机法（stratified randomization）

1. 分层随机抽样（stratified sampling）　先将研究对象按某一特征进行分层，然后在各层中采用随机的方法抽取研究对象组成样本。在进行分层时，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）。通常考虑的分层因素有人群特征、地区特征、疾病的临床特征等。

2. 分层随机分组（stratified allocating）　将研究对象按某一特征先进行分层，然后再在各层中按随机的方法将研究对象分配到试验组和对照组。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，如重要人群特征、临床特征或预后因素（包括性别、年龄、病情、有无合并症等）。在临床研究中，分层时可考虑以下几方面：①影响所研究疾病发生的危险因素；②影响所研究疾病预后的因素；③遵循最小化原则，将分层因素控制到最低限度，否则分层过多，会造成组内研究对象过度分散。

利用分层随机分组，可保证各“层”都有研究对象进入试验组或对照组，提高了组间的均衡性。例如，慢性心房纤颤复律后进行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（病因：瓣膜性心脏病与非瓣膜性心脏病；心脏大小：心胸比例≥0.50和<0.50；病程：心房纤颤病程≥6个月及<6个月）进行分层随机分组（图3-1）。

经图3-1所示的分层随机分组，可实现病因、心脏大小、病程等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）间的平衡。

（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若干区组，然后在每一区组

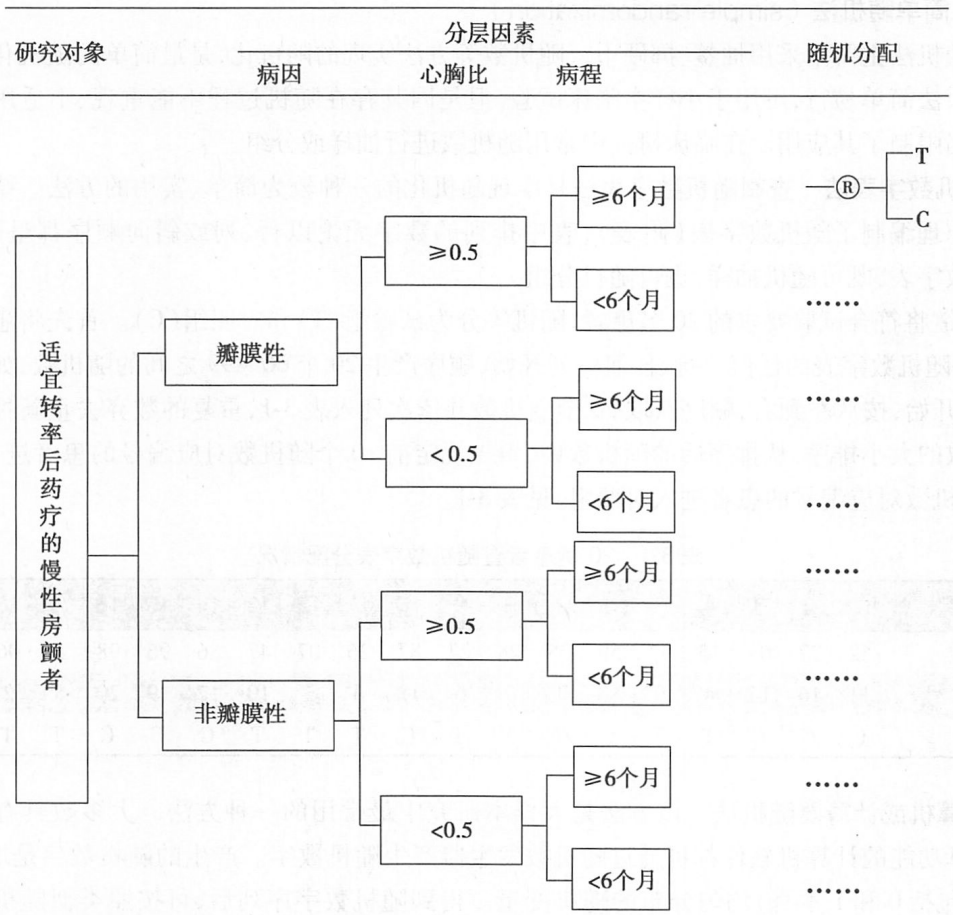


图 3-1　分层随机分组

中再进行随机分配，既可使组间的人数相同又保证了随机化。在临床研究中，每一区组的研究对象数一般按组别的固定倍数来确定。如研究分为试验组与对照组两组，则区组例数可选 2、4、6 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）两组，确定区组例数为 4。首先，按研究对象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进行随机分组，有 6 种组合：TTCC、CCTT、TCTC、CTCT、TCCT、CTTC。区组随机分组的特点是：分组后各组人数相等，同时区组随机分组也便于逐渐累积临床病例。

在临床试验中可将分层随机与区组随机结合应用，如多中心临床试验中，所有的研究对象先按中心分层，然后在各层中进行区组随机分组，分层随机与区组随机相结合可保证试验结束时各中心例数接近而且便于管理。更重要的是分层区组随机化可使组间的预后因素分布一致，提高检验效能，是大规模临床试验中比较适宜的一种随机化方法。

（四）系统随机抽样法（systematic sampling）

系统随机抽样也称为机械抽样（mechanical sampling）。是按照一定的顺序，机械地每隔若干单位抽取一个单位的方法。方法如下：设总体观察单位为 N，需要抽取的样本数为 n，抽样比例为 $k=n/N$，具体做法是先将总体的观察单位编号，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 10 000 个观察单位中抽取 1000 个组成样本，即抽样比例为 1/10，每 10 个观察单位为一部分，所有观察单位统一编号，可先从第 1～10 号中随机抽取一个号码作为起点，比如 8 号，之后抽取的号码为 18，28，38，……，以此类推，组成样本。

系统随机抽样的优点是简单易行。缺点是若总体各单位的排列顺序存在一定规律性或周期性，而该特征与研究结果有关联，以这种方法进行抽样则可能出现偏倚。例如，某年级小学生按照身高从低到高排列，按系统抽样的方法抽取 1/3 的学生进行身体发育状况调查，若抽样的起点为 1 号，之后

笔记

抽取的号码为4,7,10,……,这样抽取的样本,可能存在样本的平均身高明显低于总体的平均身高的情况,影响样本的代表性。

（五）整群随机法（cluster randomization）

整群随机法是以现成的群体(如以社区、街道、乡、村、医院、学校、班级等)为单位进行抽样或分组,可以采用简单随机、分层随机等方法进行。在整群随机抽样中,抽到的群体中的所有个体,都将作为研究样本。例如,采用整群随机抽样方法,从某市的30所幼儿园中随机抽取3个,并对这3个幼儿园的全部儿童进行牙齿检查,以了解该市幼儿的龋齿患病情况。同样,在整群随机分组中,被分到试验组或对照组的群体中的每个个体,都作为研究对象。

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。一般样本含量相同的情况下,整群随机抽样的抽样误差大,整群随机分组的组间可比性差。但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、从高到低分阶段进行的抽样方法。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(如省),再从抽中的一级单元随机抽取范围较小的二级单元(如市),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。

在一些大规模研究中,如全国肺结核病的抽样调查、全国高血压的抽样调查等,多级抽样成为唯一实用的抽样方法。具体实施时,多级抽样往往与上述其他抽样方法联合使用。

三、随机化过程中需注意的问题

1. 随机化要求分配的序列应具有重复生成的能力。抽签、抛硬币等方法尽管是随机的,但不能检查或重复,一般不常用。手工使用随机数字表、随机排列表虽可重复,但对样本量较大的研究较为繁琐。目前常利用统计软件实现随机化。

2. 为了保证临床试验的可靠性,研究中所用随机化方法、随机数等均应详细记录。倘若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;倘若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、起始列及获取的方法。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。

3. 随机化分组时,必须对随机化方案进行保密,该过程称为随机分配方案的隐匿(concealment)。从某种意义上讲,随机分配方案的隐匿比单纯产生随机分配序列本身更加重要。若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,试验执行者就可预知下一位研究对象将被分入哪一组。此时因研究人员主观因素造成的刻意选择研究对象甚至改变分配方案等所带来的选择偏倚将难以避免。因此,随机分配方案的隐匿就是让试验执行者事先不知道随机分配方案的具体内容,无法预知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:中央随机系统分配、药房控制随机分配、信封随机分配、编码容器随机分配等。中央随机系统分配法是研究者确定合格的研究对象后,通过电话或网络联系中央随机系统,中央随机系统根据该研究对象的基本情况确定其具体的分配。药房控制随机分配法是将事先产生的随机分配序列由药房保存并控制,药房根据研究者提供的合格研究对象情况,负责分配研究对象的组别。信封随机分配法根据产生的随机分配序列将记录有分配组别的卡片逐个放入按顺序编码、不透光、密封的信封中,当研究者确定研究对象合格后,依顺序拆开信封并将研究对象分配到相应的组别。为了避免研究者可能提前打开信封或一次同时打开几个信封,需要制定适当的防范措施。编码容器随机分配法常用于药物临床研究中,该法根据产生的随机分配序列,将药物放入外形相同并按顺序编码的容器中,之后将药物发放给相应的研究对象。

4. 半随机化法(quasi-randomization)是按研究对象的就诊顺序、入院日期、生日等的单双号交替分配入组的方法。该方法未按随机数字产生序列号,简单、机械,容易被破盲,因此并不是真正的随机化。当进行大样本研究时,如基于社区人群的试验,利用半随机化分配研究对象简便、实用,同时也能较好地实现组间均衡。但在临床试验中,由于研究对象往往偏少,半随机化难以达到组间均衡,因而

该方法的使用受到限制。动态随机化(dynamic randomization)是指在临床试验的过程中每个研究对象分到各组的概率不是固定不变的,而是根据一定的条件进行调整,保证各组间的例数和某些重要的影响因素的一致。动态随机化在临床试验中是一种有效的分组方法,国内外均有成功应用,但从其分配的原理和过程看,动态随机化也不是完全随机的。

第二节 对照原则

一、概述

对照(control)是指在临床研究过程中,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”,在对疾病的病因、诊断、治疗、预后等方面进行临床研究时,存在多方面已知或潜在未知因素的影响,因此需要合理设置对照,通过比较方可得出可靠结论。

二、设置对照的意义

在临床研究中,除了干预措施(处理因素)的效应外,还有很多因素可能影响临床结局:

1. **不能预知的结局(unpredictable outcome)** 结局的不确定性因病而异。像感冒、细菌性痢疾、病毒性肝炎等一些疾病都有明显的自愈倾向;而像高血压、消化道溃疡等慢性病,也存在病情自行缓解或加重现象;即使是恶性肿瘤,其生存时间长短、并发症的发生等情况也不尽相同。同时,由于个体的生物学变异和社会、心理因素的影响,导致患相同疾病的个体,其临床特征(临床表现、病情轻重、预后等)和结局迥异。

2. **霍桑效应(Hawthorne effect)** 指人们对部分研究对象特别感兴趣而倍加关注所产生的一种心理-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。但有时也可能因为不信任某医生或某医疗单位而产生负向效应。

3. **安慰剂效应(placebo effect)** 安慰剂为没有药理活性,但其色、形、味均与试验药物完全相同的制剂,如淀粉、乳糖、生理盐水等。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效果,其原因是由于某些研究对象对医药依赖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(平均水平)靠近的倾向。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。

5. **潜在未知因素的影响** 由于知识水平和认识的局限性,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或掌握。

因此,在临床研究中设立合理的对照很有必要,旨在消除非处理因素的干扰,鉴别试验效应与非试验效应。

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)** 按严格的随机方法将研究对象同期分配到试验组和对照组,以此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:①随机化分组,可以较好地保证各试验组之间的均衡可比,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,可以对各组同时、同步观察,有效避免了因试验先后顺序对结果产生的影响,使研究结果更有说服力;③鉴于多数统计方法是建立在随机样本的基础之上,采用同期对照更有利于资料的统计分析,一般不需要对数据进行校正。

但同期随机对照也存在一些不足:①需要有一部分研究对象充当对照,因此所需样本量较大;②在有些情况下可能涉及医德伦理问题。

笔记

2. 非随机同期对照（non-randomized concurrent control）　设置有同期对照，但未严格按随机化原则进行分组。例如在多中心临床试验中，将一所医院的对象全部作为试验组，而另一所医院的对象全部作为对照组进行研究。

非随机同期对照的设置简便易行，易为医生和患者接受；但研究对象非随机分配，组间基线情况可能不一致，可比性较差。

3. 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免前一阶段的携带效应对后一阶段产生影响。

自身对照主要适用于慢性反复发作性疾病的治疗性研究，这类疾病病程长，病情稳定，允许进行分阶段处理与观察。采用自身对照可消除个体差异，同时样本量减半，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，可能存在处理先后顺序对结果的影响。

4. 交叉对照（cross-over design control）　将两组研究对象分两个阶段进行试验。第一阶段，一组接受 A 措施，另一组接受 B 措施，间隔一段洗脱期后，在第二阶段，两组接受的 A、B 措施互换，最后对比 A、B 两种措施的效果。

交叉对照同样可以保证每个研究对象都接受有效治疗的处理，消除个体差异，节约样本量，提高统计效率，同时能避免试验先后顺序对结果的影响；但该方法也主要限于慢性复发性疾病。

5. 配对对照（matching control）　将可能对研究结果产生影响的混杂因素（如性别、年龄、病情等）作为匹配条件，为每一个试验对象选配一个或一个以上的对照，通常采用 1∶1或 1∶2配对。

配对对照可最大限度地保证组间均衡性，避免已知混杂因素对结果的干扰。

6. 历史对照（historical control）　试验中仅设一组研究对象接受新干预措施，并与过去的旧干预措施进行比较。因此，历史对照属于非随机、非同期对照。

历史性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，不存在伦理问题；但既往资料中的研究对象、条件、环境等都通常难以与本次试验一致、可比性差，一般不宜采用。但在特殊情况下，如对一些预后极差的疾病，采用历史性对照仍有一定说服力。例如，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，而随着化疗和放疗技术进步，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，证明现有的治疗效果明显提高、预后改善显著而得到公认。

（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）。主要分为：有效对照（标准对照或阳性对照）、空白对照、安慰剂对照等。

1. 有效对照　又称标准疗法对照或阳性对照，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）施予对照，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，研究对象随机分配到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，施予对照组的处理措施效果稳定，试验期间能保证对照组成员接受合理治疗，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，仅对其进行观察、记录结果，并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患者，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）　对照组在试验期间给予没有真正治疗或致病效应的安慰剂，称为安慰剂对照。主要用于上市前药物临床试验。

安慰剂对照不等同于空白对照。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，平衡对照组与试验组患者心理作用对疗效评价的影响，便于实施盲法。安慰剂对照主要用于病情轻、稳定，或目前没有有效治疗方法的疾病。同时，也应注意伦理学方面的问题。使用安慰剂要以不损害患者健康为前提，在没有可靠对照药物时才允许使用。

第三节 盲法原则

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形成等均在不知晓研究对象（或受试对象）所在的组别以及不知晓所接受的是何种措施的前提下进行。此外，盲法还可用于统计分析之中。

临床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与评价，否则将出现信息偏倚，导致结果失真。在临床试验中，信息偏倚既可来自研究者方面，也可产生于研究对象方面。一旦研究对象知晓自己的分组情况，可预知疗效或预后，会产生霍桑效应、甚至对治疗丧失信心而中途退出。另一方面，若研究者知晓研究对象的分组情况，则可能因希望研究得出阳性结论而有意或无意地暗示研究对象，诱导治疗组对象多报告有利结果而对照组对象则正好相反。盲法试验的目的是控制信息偏倚和测量偏倚。

二、盲法的分类

盲法试验的基本原则就是让研究对象和（或）研究者无法知晓哪些人接受处理措施，哪些人接受对照措施。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、统计分析者和论文撰写者等，根据“盲”的对象不同，一般可分为单盲、双盲、三盲等类型。

（一）单盲

1. 定义 仅研究一方（研究对象或研究者）不知晓具体分组和所接受的措施（如干预措施或对照措施）的具体内容，称为单盲（single-blind）。

2. 优缺点 单盲的优点是简便易行；能减少因研究对象主观因素对研究结果的影响；干预措施执行者和结果测量者知道受试对象的分组情况，便于更好地观察和掌握病情，一旦病情发生变化，可及时、恰当调整处理方案或采取其他干预措施，使受试对象免受伤害。单盲的缺点是不能消除研究方主观因素造成的影响。例如，主管医生可能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研究对象产生不同的反应。

（二）双盲

1. 定义 研究对象和试验执行者（干预措施执行者及结果测量者）双方均不知晓研究对象的分组情况和研究对象接受的是何种干预措施，称为双盲（double-blind）。

2. 实施方法及注意事项

（1）双盲试验往往与安慰剂和随机化分配的隐匿同时配合实施。随机化分配的隐匿是双盲试验的基础，否则将无法实现受试对象与试验执行者双盲，而设计科学、合理、逼真的安慰剂则是实现双盲所必备的条件。

（2）双盲试验要有严格的管理制度和方法，组织严密，操作规范。在双盲试验过程中，需要有“第三方”的监管，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，而研究对象和研究的执行者只知道每个研究对象的药物编号，如 A 和 B。待试验结束并完成数据分析后，才由密码控制者当着主要研究人员的面，公布隐匿的随机化方案并宣布 A 药和 B 药究竟是试验制剂还是对照制剂，该过程称为“揭盲”。

（3）双盲试验要求各组药物在外观的形状、大小、颜色，以及给药的途径、方法、次数上保持一致，若有疏漏，则可能导致破盲，失去“盲法”的作用。药物临床试验时，如果试验药与对照药的剂型不一样，可采取“双盲双模拟”方法：试验药+与对照药剂型相同的安慰剂，对照药+与试验药剂型相同的安

笔记

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”方案进行。例如，用双盲法评价两种抗高血压药的效果。一组用缓释片，每日一次；另一组用 β-受体阻滞剂，每日两次。在执行时可采用两组患者每天、晚各给药一次。药名标出 A 和 B，两组患者均为早上服 A 药一片，晚上服 B 药一片，而两组患者所用的 A、B 号药实质是不同的，两组的 A、B 号药分别为：每日服药一次组的 A 药为抗高血压的缓释片，B 药为安慰剂；每日服药两次组的 A 药为 β-受体阻滞剂，B 药也为 β-受体阻滞剂。

（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，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，需事先制定出明确的停药或更换指征，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，尽量减少因“破盲”带来的影响。

3. 优缺点

（1）优点：双盲是临床试验最常用的一种盲法形式，可有效避免研究对象和试验执行者主观因素对试验结果的影响。尤其对干预措施的评价指标是以主观指标（如恶心、呕吐、头痛、眩晕、呼吸困难、各种量表等）为主时，采用双盲试验能够有效减少信息偏倚的发生。

（2）缺点：在管理上缺乏灵活性；有特殊副作用的药物容易被破盲；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晓受试者接受的是何种干预措施，全部采用编号密封，它可避免双盲法在资料分析阶段的测量偏倚和报告偏倚。

2. 优缺点　可以使信息偏倚减到最小的程度，使评价结果更趋真实。但设计复杂，执行难度较大，常因医德、沾染、补偿、失密等问题而难以坚持；同时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。虽然三盲试验是减少信息偏倚最有效的方法，但在实际工作中使用并不普遍。

盲法与随机分配的隐匿是不等同的。盲法是为了避免干预措施实施过程中和结果评价时来自研究对象和研究者的信息偏倚，并不是任何随机对照试验都能实施的。随机分配的隐匿是为了避免研究者对受试对象分组时产生的选择偏倚，在任何随机对照试验中都能实施。例如，比较外科手术和内科药物治疗某种疾病的疗效，随机分配的隐匿是可行的，而盲法却难以实施。

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晓试验组和对照组的分组情况以及所给予的干预措施。

并非所有临床研究都适用盲法。例如，比较手术疗法和保守疗法的治疗效果，探讨功能训练或针灸疗法的疗效，评定生活习惯（吸烟、运动）改变对疾病的影响等。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，倘若主要结局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令人满意的结果。

（二）适用范围

1. 在一次研究中，需要对多种因素进行分析。例如，一项临床研究不仅希望观察药物本身的效果，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. 不足之处是容易产生信息偏倚。

第四节　重 复 原 则

一、概述

重复(replication)是指研究的各组(如试验组与对照组)应有一定数量的结果重复或再现。只有在相同条件下,对同一观察指标进行多次重复测定,才能用统计学方法估计研究对象的个体差异及其抽样误差。

重复原则的意义在于:①避免了把个别情况误认为普遍情况,把偶然性当成必然的规律,以致误导,将错误的研究结果推广到群体;②在同一研究条件下对同一观测指标进行多次重复观测,才能根据观测的结果,估计变异情况,进而描述观测结果的统计分布规律。这里的抽样误差属于随机误差,是客观存在的,只有样本量足够大时方能获得比较精确的统计量。因此,重复的作用在于控制和估计研究中的随机误差。

二、样本量的估计

样本量估计是重复原则的一种体现形式,是临床研究设计中一个极为重要的环节,直接关系到研究结论的可重复性,临床研究中要求样本量适当,以能够可靠地回答研究假设所提出的相关问题,同时又不至于样本量太大而造成浪费。

(一) 影响样本含量的因素

临床研究中估计样本含量大小时要考虑研究目的、研究设计、效应量、检验水准、把握度、单双侧检验、允许误差的大小等多方面因素的影响。同时还要适当考虑研究中存在的协变量、失访、退出和依从性等情况,对估计出的样本量进行调整。估算样本量时主要考虑以下几点:

1. 效应量　效应量是样本量估计所需的重要参数之一,不同类型的数据就有不同的效应量表达,其样本量估计方法亦不相同。常见的效应量有:均数差值、率差或比值(如 RR、HR)、OR、相关系数等。上述效应量的总体指标往往未知,通常以样本指标估计。

2. 检验水准与检验效能　检验水准也就是Ⅰ类错误概率,用 α 表示。检验效能用 $1-\beta$ 表示,β 代表Ⅱ类错误概率。在计算样本量的时候,必须考虑Ⅰ类错误和Ⅱ类错误(或把握度)的大小。Ⅰ类错误与Ⅱ类错误越小,所需的样本量越大。大多数情况下,α 设为 0.05,β 设为 0.2(即把握度 $1-\beta$ 设为 0.8),但也可根据研究目的进行调整。

3. 单侧和双侧检验　双侧检验比单侧检验更为稳妥,两者所需要的样本量不同,一般单侧检验所需样本量少于双侧检验。但要注意,使用单侧或双侧检验是根据研究目的事先确定的。例如,临床试验中的优效性检验、非劣效性检验为单侧检验,而等效性检验需同时在两个方向进行两次单侧检验。

4. 各组样本量是否相同和组数的多少　在其他条件不变时,各组样本量相同时所需总样本量最小。组数越多需要的总样本量也越大。

(二) 样本含量估计方法

样本含量的估计方法主要有查表法和计算法,查表法是按照研究条件直接查样本含量表来获得样本含量。计算法就是使用样本含量的计算公式来估算样本含量,公式计算比较复杂,目前有专门用于计算样本量的软件,如 Stata、R 和 PASS 等,不同的软件使用的公式不同,计算结果也不一定相同。样本含量估计的具体方法略,可参见统计学专著。

第五节　临床研究的伦理学基础

一、概述

笔记

以人体为研究对象的临床研究也极易引发一系列伦理学问题。历史上从 Tuskegee 梅毒研究到

Willowbrook 肝炎研究等违背医学伦理的案例来看，无论研究结果如何，终将会受到全人类的谴责。同时在当今大数据背景下的医学数据共享，也涉及伦理学问题。

医学伦理的基础是由“希波克拉底誓言”发展而来的《日内瓦宣言》，该宣言对医生的医疗行为提出了伦理学准则并在 1948 年世界医师协会大会上正式通过，而于 1975 年世界卫生大会上又提出了《赫尔辛基宣言》，针对医学研究中的伦理学问题作出了规范。20 世纪 90 年代后，我国国家药品监督管理局结合本国实际情况，发布的《药品临床试验管理规范（GCP）》，也对药品研究中临床试验阶段所涉及的伦理问题作出了明确规定。

因此，临床研究以人体（患者）为受试对象，在研究过程中不可避免地会涉及伦理学问题。要求每项临床研究都必须在获取科学结论与维护受试者利益间寻找一个最佳平衡点，使医学科学研究真正造福于人类。

二、临床研究的伦理学原则

临床研究中无论是探讨新的预防、诊断和治疗方法的效果，还是探索病因与危险因素研究都必须遵循“受益大于风险”原则，即有利于增进人类健康，尽可能避免各种潜在风险的基本原则。为实现上述目标，在研究立项前，应做大量的文献复习，并进行循证分析与决策。只有大量证据证明研究的必要性及预期获益超过其所带来的风险和损害时，研究才有了立项的基础。研究设计完成后，必须报伦理委员会审批。报送材料中应对所有涉及人体环节的研究过程、实施方法、潜在风险及控制措施等方案细节予以真实、完整的呈现，以便伦理委员会专家作出客观判定。同时，临床研究者必须由合格的科研人员承担，并至少在 1 名以上具有丰富临床经验的医生监督下进行，以确保患者的安全。临床研究中需遵循的伦理学原则包括：尊重原则、受益原则和公正原则。

（一）尊重（respect）原则

尊重原则是指临床研究中应充分尊重受试者的人格及其作为一个自主个体所应享有的权利。具体体现在以下几方面。

1. 知情（informed）　研究者应向研究对象阐明研究的目的、方法、实施过程、经费来源及研究可能带来的利弊风险，让研究对象有充分的时间了解加入研究项目对其自身的意义和价值所在，以便作出理性的决定。

2. 自愿（voluntary）　对于有自主行为能力的研究对象，必须在其完全自愿的情况下签署《知情同意书》（informed consent form），并告知他们有拒绝参加研究或无条件随时收回同意书的权利。对于自主性受限的对象（儿童、痴呆者、昏迷者等），研究者必须按法律规定从合法代理人处获取知情同意。同时，在临床研究中为避免研究对象（患者）因迫于研究者（医生）压力而被迫同意参与试验的情况发生，应由一位了解情况但不参与研究的独立第三方来征询研究对象的知情同意。

3. 保密（confidential）　研究者应对研究对象的个人信息、健康状况、行为习惯等资料严格保密，未经本人同意，不得将研究对象的个人信息公开，应充分尊重其隐私权。

（二）受益（beneficence）原则

受益原则主要体现在两个方面：

1. 在研究过程中，应尽量保证受试者获益最大化、风险最小化。如尽量避免侵入性操作，抽血量降至最低。医务人员应密切观察研究对象的病情变化，一旦发现潜在风险大于可能的受益或已经得出预期结果，应立即终止试验。

2. 研究得出确切结论后，应确保每个参加试验的患者能够利用研究所证实的最好的预防、诊断和治疗方法。如在疗效研究中，一旦证实试验组所用的新药疗效优于常规药物，则在试验结束后应该让对照组患者也使用新药。

（三）公正（justice）原则

公正原则要求研究者在确定研究对象时，应该遵循科学原理，设立合理的入选标准、排除标准和剔除标准，不应该只是针对弱势群体（老幼、残障、囚犯、有色人种等）。其次，在分配研究对象到不同

组时，应不分亲疏，一视同仁。分配前应告知对象可能被分配到的不同组别，并说明分组原则与方法。

第六节　临床研究中的偏倚及控制

一、概述

任何医学研究中，尤其以人（患者）为研究对象的临床研究中，因多种因素的影响，导致研究结果与客观实际不符合的情况，称为误差，可分为随机误差和系统误差。

随机误差（random error）没有固定方向且大小呈随机变化。主要来源于个体生物学变异、随机测量误差和抽样误差等，随机误差不可避免。

系统误差（systematic error）有明确的方向性、大小通常恒定，是由一些固定因素所致，如研究对象的选择不当、仪器未校正、试剂不统一、测量方法不规范以及来自于研究者和研究对象的主观因素等。系统误差可以通过周密的研究设计和测量过程标准化等加以控制和消除。偏倚（bias）是指因研究设计、测量、资料分析及结果解释等临床研究各个环节中所发生的系列错误，导致研究结果系统性偏离真实值的情况。偏倚属于系统误差，既可能是某种错误导致样本统计量（样本均数、率等）系统性高于或低于总体参数（总体均数、总体率等），也可能是系统性夸大或减弱研究因素与结果间的关联强度（OR），从而影响结果的真实性。偏倚是研究中的各种错误所造成的，按产生的原因可将偏倚分为选择偏倚、信息偏倚、混杂偏倚三类。

二、选择偏倚

（一）概念

选择偏倚（selection bias）是指在选取研究对象时，由于选取方法不当，使入选的研究对象与未入选者在某些特征上存在着差异而造成的一种系统误差。

选择偏倚主要发生在研究的设计阶段，当确定研究样本或选择比较组时，未严格遵循随机原则，或设置的纳入/排除标准不恰当。如探讨高血压与血脂异常的关系时，研究对象全部来自心内科住院患者，或者采用自身前后对照设计评价某减肥药疗效时，研究对象可能全部为自愿者，以上两种情况均可能夸大研究因素与研究结果间的关联强度，产生选择偏倚。

另外，在信息收集过程中也可产生选择偏倚，如无应答、失访的例数较多而又未采取有效的处理措施，若仅以资料完整对象进行分析，也可造成结果的失真。

（二）常见来源举例

1. 入院率偏倚（admission rate bias）　入院率偏倚又称伯克森偏倚（Berkson's bias），指当研究对象全部来自医院的住院或门诊病例时，可能因患有某种疾病或具有某种特征者入院机会较目标人群不同而导致的一种偏倚。例如，研究呼吸道疾病与骨及运动器官疾病的相关性，由于同时罹患呼吸道疾病与骨骼及运动器官疾病的患者入院或就诊机会较高，因此在医院选择研究对象时，会夸大两者之间的关联强度，见表3-2。

表3-2　呼吸道疾病与骨骼及运动器官疾病相关性研究

呼吸道疾病	一般人群 骨骼及运动器官疾病			住院6个月以上患者 骨骼及运动器官疾病		
	有	无	合计	有	无	合计
有	17	207	224	5	15	20
无	184	2376	2560	18	219	237
合计	201	2583	2784	23	234	257

$OR_{一般人群}=1.06$，$OR_{医院患者}=4.06$，可见，在一般人群中呼吸道疾病与骨骼及运动器官疾病间相关性很弱，而以住院6个月以上者为研究对象时，两者间显示有很强的联系。

笔记

2. 现患-新病例偏倚（prevalence-incidence bias）　现患-新病例偏倚又称奈曼偏倚（Neyman bias），主要发生在研究对象全部或大部分为现患病例时。现患病例与新病例相比，在病程、病情、病型等方面可能存在差异，可出现选择偏倚。

3. 检出症候偏倚（detection signal bias）　指在病因研究中，某种疾病的一些早期症状可能与研究的暴露因素相关，因此，出现这些早期症状者会及早就诊，并接受检查而导致该人群有较高的疾病检出率，夸大两者之间的关联。例如子宫内膜癌的病因研究，若研究对象来自医院门诊时，得出服用雌激素是主要病因的结论，但在普通人群中研究时，两者关联明显减弱。进一步研究发现，服用雌激素可引起阴道不规则出血，这部分人会及早地到医院就诊，因此，大大提高发现其患有子宫内膜癌的机会。

4. 排除偏倚（exclusive bias）　指研究中无应答、失访或不依从者较多，若直接剔除该部分人的资料，仅用余下资料进行分析而带来的一种偏倚。这是因为被排除对象与余下的对象可能存在明显差异。如在疗效研究中，疗效差或副反应较大者更容易失访而被排除在分析之外。

5. 易感性偏倚（susceptibility bias）　当研究对象过于集中在某一类特定人群，如以油漆生产车间工人为对象，研究油漆与皮肤过敏两者间的关系，可能会发现这一组人群过敏性皮炎的发生率并不比其他人群高。这里面应考虑纳入对象的特殊性，如因自然淘汰、能够留在该车间工作者可能多是对油漆刺激不易感者。

（三）选择偏倚的控制

在不同研究中因目的不同，产生选择偏倚的原因各异。控制选择偏倚的措施主要有以下几方面：

1. 当研究目的确定后，需广泛查阅文献，反复讨论可能出现选择偏倚的各种情况，从设计上加以避免。

2. 尽量在一般人群中选择研究对象，或设计多组对象以减少选择偏倚的发生。

3. 抽取研究对象及分组时，尽量做到随机化。

4. 尽量减少“无应答”“失访”或“不依从”情况，一般要求控制在10%以内。

三、信息偏倚

（一）概念

信息偏倚（information bias）又称为观察偏倚（observational bias），是指在收集、整理信息过程中，由于测量研究对象的暴露与结局的方法（或工具）存在缺陷，导致收集的信息与真实情况存在的系统误差。

信息偏倚主要发生在研究的实施阶段，其产生主要有三个方面：①来自研究对象。研究对象可能由于主观或客观的原因有意或无意提供不真实信息，如儿童隐瞒上网情况、孕妇有意隐瞒饮酒史、慢性病患者忘记曾经服用某种药物等。②来自研究者。研究者也有可能因主观原因有意或无意地选择性收集与接收某些信息，造成信息偏倚。如研究者一般都希望得到阳性研究结果，对暴露者相关暴露信息的挖掘和收集比对非暴露者更加深入、全面，从而造成夸大暴露与结局间的联系。③来自测量环节。如果研究使用的测量仪器或工具本身存在测量误差，如仪器未校正、测量方法不一致等，则必然导致研究结果的系统误差。

（二）常见来源举例

1. 回忆偏倚（recall bias）　在进行问卷调查时，被调查者可能因记忆不准确，对以往的疾病史、用药史或家族史等情况提供错误信息而导致偏倚。

2. 报告偏倚（reporting bias）　指研究对象可能出于各种原因故意隐瞒或谎报某些信息而导致的偏倚。如患者故意隐瞒吸烟、饮酒史等，造成某药物疗效评价失真。

3. 暴露怀疑偏倚（exposure suspicion bias）　易发生于病例对照研究中，主要来自研究者。例如，研究某暴露因素与一种疾病的关系时，研究者从主观上认为两者有联系并希望出现阳性结果，因此在收集信息时，会下意识的过度挖掘病例组的暴露信息而忽略对照组的暴露信息。

4. 诊断怀疑偏倚（diagnostic suspicion bias）　多见于队列研究和临床试验，指研究者先入

为主，在判断结局时对暴露组（或试验组）采取了比对照组更认真、详细的方法和态度，导致暴露组（或试验组）更易作出某结局诊断的情况。

5. 测量偏倚（detection bias）　测量偏倚指因检查、测量仪器误差或操作不当而导致的系统误差。

（三）信息偏倚的控制

根据信息偏倚产生的原因，控制信息偏倚主要从以下几方面进行：

1. 严密设计　研究者对拟进行的研究要事先制定科学的资料收集方法。如设计统一而明确的调查表，对调查内容或测量指标有明确的定义、客观的标准，避免模棱两可的提问，尽量采用定量指标收集信息等。

2. 严格培训调查员　研究正式实施前，让调查员充分了解每个问题的设置目的、内涵，掌握收集信息的要领和技巧。

3. 尽量选用客观指标，采用“盲法”收集信息　避免因研究者、研究对象主观因素对研究结果的影响。

4. 研究中的各种仪器设备应校正，试剂和检测方法应标准化。

5. 对研究对象做好宣传工作，使其知晓研究的目的、意义和要求，获得支持和配合。

6. 设计严格的质量控制手段，以及恰当实用的质量监测、控制措施（如调查员签名、复查、定期检查等）。

四、混杂偏倚

（一）概念

混杂偏倚（confounding bias）也可简称为混杂（confounding），是指由于有一个或多个潜在外部因素的存在，夸大或掩盖了研究因素与结局之间的真实联系，使研究结果系统性偏离真实结果的情况。这些潜在的外部因素称为混杂因素（confounding factor）。

混杂因素是导致混杂偏倚的关键，它具备三个基本特点：①混杂因素本身就是一个独立的、与研究结局有关联的因素；②混杂因素与研究因素之间也必须相关；③它不是研究因素与研究结局因果链上的一个中间环节。假设在病因研究中，欲探讨某暴露因素 E（Exposure factor）与疾病 D（Disease）的关联，若另有一个因素 F（Factor），它既是疾病 D 的一个独立危险因素，同时又与暴露因素 E 相关，那么 F 因素的存在意味着暴露组比非暴露组会有较高的疾病 D 发病率。因此，无论 E 与 D 之间是否真正有因果关联，都会得出暴露组的发病危险性高于非暴露组的结论。例如，某研究者采用队列研究探讨服用维生素 E 与心肌梗死间的关系，随访得到数据，见表 3-3。

表 3-3　心肌梗死与服用维生素 E 的队列研究

服用维生素 E	心肌梗死		合计
	发生	未发生	
服用	400	600	1000
未服用	600	400	1000
合计	1000	1000	2000

经分析显示，$P<0.05$，RR = 0.4/0.6 = 0.67。结论为服用维生素 E 与心肌梗死有关联，维生素 E 是心肌梗死的保护因素。但是这一结果是否受到“吸烟”这个外部因素的干扰，吸烟是否为本研究中的混杂因素呢？可通过以下分析来判断。

首先，分析心肌梗死发生是否与吸烟有关。通过随访的吸烟情况记录发现，2000 人中，吸烟者 1150 人，心肌梗死发病率为 71.30%，不吸烟 850 人，心肌梗死发病率为 21.18%，进一步结合服用维生素 E 的情况，得到的数据见表 3-4。

笔记

表 3-4　按是否服用维生素 E 分析吸烟与心肌梗死的关联

吸烟情况	服用维生素 E 组 心肌梗死			不服用维生素 E 组 心肌梗死		
	发生	未发生	合计	发生	未发生	合计
吸烟	240	30	270	580	300	880
不吸烟	160	570	730	20	100	120
合计	400	600	1000	600	400	1000

通过分析得到：$RR_{服用\ vitE}=4.06$，$RR_{不服用\ vitE}=3.95$，结果说明无论是否服用维生素 E，吸烟与心肌梗死都有很强的联系，吸烟是心肌梗死的独立危险因素。

进一步分析吸烟与服用维生素 E 的关系，见表 3-5。

表 3-5　吸烟与服用维生素 E 的关联分析

吸烟情况	维生素 E		合计
	服用	未服用	
吸烟	270	880	1150
不吸烟	730	120	850
合计	1000	1000	2000

计算得：OR=0.05，说明吸烟与服用维生素 E 这两者间存在负相关，即服用维生素 E 者吸烟比例低。

从表 3-4、表 3-5 可知，在该研究中，吸烟这个因素满足了作为混杂因素的基本条件，可能引起混杂偏倚。

判断某可疑混杂因素是否引起混杂偏倚，可通过比较包含该因素时研究因素与结局间的粗（crude）效应值（如 cRR 或 cOR）与排除该因素后的调整（adjusted）效应值（如 aRR 与 aOR）来判断。若数值差别过大，说明存在混杂偏倚，应以调整效应量作为研究因素的纯效应估计值。

（二）混杂偏倚的控制

混杂偏倚是在进行组间比较时，因混杂因素在组间分布不均衡而造成的。因此，控制混杂偏倚主要从研究设计与资料分析两方面考虑：

1. 设计阶段

（1）随机化分组：通过随机分组可以相对保证混杂因素在组间分布均衡，从而起到控制混杂偏倚的作用。对于已知混杂因素，可以采用分层随机的方式进行分组，以达到控制混杂偏倚的目的。

（2）匹配：将可疑混杂因素作为匹配因素，使对照组在这些因素上与研究组保持一致。

（3）限制：对可疑的混杂因素采取限制的手段。如考虑年龄、性别可能成为混杂因素，设计时可将研究对象的年龄限制在某一年龄段内、仅以男性或女性为研究对象，以控制年龄、性别对研究结果的影响。

2. 资料分析阶段

（1）分层分析：在资料收集之后，按可疑混杂因素进行分层，通过比较分层前后的效应值（cRR 与 aRR、cOR 与 aOR）来识别其混杂作用及其大小。如前例，可按是否吸烟进行分层，然后再在吸烟人群与非吸烟人群中分析服用维生素 E 与心肌梗死的关系，从而控制吸烟这个因素对研究结果的影响。

（2）利用标准化法（率的标化）进行调整：如欲比较甲、乙两地的死亡率，为消除年龄构成不同带来的影响，可对两地人口死亡率按年龄进行标化，以消除年龄因素对研究结果的影响。

（3）多因素分析：采用多因素分析方法，将混杂因素与研究因素一并纳入分析。通过数学模型把所有相关因素（包括多种混杂因素）对某一结局的影响进行定量综合分析。多因素统计分析方法包括多元线性回归、逐步回归、Logistic 回归、Cox 风险比例模型等。

（张凤英）

NOTE

第四章　临床研究的常用设计方案

临床研究结果要转化为高质量的循证证据，最为有效的方法就是选择论证强度高、可行性好的设计方案，开展高质量的临床研究。临床研究的设计方案通常分为两大类：观察性研究（observational studies）和试验性研究（experimental studies）。本章将分别介绍常见临床研究设计方案的概念、特点、设计模式、实施方案、资料的整理与分析、应用范围和优缺点，供大家在临床研究实践中选用。

第一节　随机对照试验

一、概述

（一）定义

随机对照试验（randomized controlled trial，RCT）是指研究者根据研究目的，按照预先确定的研究方案将符合条件的研究对象随机分配到试验组和对照组，进而分别接受相应的处理措施，并在一致的条件下或环境中，同步地进行研究、观测和比较组间效应，从而确认试验效果的一种试验性研究。

（二）历史回顾

早在17世纪培根（F. Bacon）就提出应该“通过试验来尝试或证实特定疾病疗法的效果”。18世纪G. Berkeley提出对照临床试验的观点。早期的流行病学实验主要是针对患者进行的，如18世纪J. Lind医生针对坏血病的治疗试验，19世纪高木兼宽的脚气病治疗试验，以及20世纪初J. Goldberg关于燥皮病的治疗试验。B. Hill被公认为随机对照试验之父，他所设计的链霉素治疗肺结核效果评价方案，成为临床试验的经典，于1948年发表在英国医学杂志（*BMJ*）上。

近60多年来，随着理论和方法学的日趋成熟，随机对照试验已被视为评价干预措施有效性的金标准方案。

（三）主要用途

随机对照试验属于试验性研究，无论是在预防医学研究中，还是在临床研究中可以解决很多问题，其在临床研究中的应用主要有以下几个方面：

1. 药物上市前的疗效研究　新药在上市前必须要经过Ⅰ～Ⅲ期临床试验，确认安全有效并通过新药审批后方能上市。

2. 对当前临床治疗方案进行评价　过去，某种治疗措施或药物能否应用于临床实践，主要取决于专家或顾问的意见。而临床医生选择治疗措施也主要基于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。鉴于疾病发病机制的复杂性和对疾病认知的局限性，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。随机对照试验能够科学评价某种治疗措施的真正疗效或副作用及其影响因素。

3. 药物上市后再评价及卫生技术评估　随机对照试验同样适用于真实世界研究，对上市药物或新的卫生技术在真实临床环境下的实际效果或长期安全性及其相关因素等展开全面、系统、科学地评估。

（四）主要原则

1. 对照原则　临床试验的研究对象因个体间的病理生理特征、心理状态、文化水平及所处的自然和社会环境等方面差异，导致出现一系列混杂因素，干扰因果推断，因此需设置对照组。对照的设

置可以：①抵消疾病自愈趋势的影响。倘若不设立对照，疾病好转或痊愈就很难区分是疾病自然康复还是治疗措施或者药物的作用结果。②消除研究对象的心理作用影响。一些患者因对某个医生或者某种治疗方案和药物的偏好或者反感，都会产生心理作用效应，这种效应往往会影响干预措施的效果，如霍桑效应。③消除不可预知效应或反应的影响。由于个体在年龄、性别、种族、免疫状态、遗传因素、精神心理状态等方面的差异，其临床表现和疾病特征各有不同，对干预措施的反应及其效应指标也会呈现差异性。④消除向均数回归现象的影响。这是临床上较为常见的一种生理现象，即一些极端的临床症状或体征有向正常水平回归的倾向，称为向均数回归现象。

2. **随机化原则**　随机对照试验中，将研究对象随机分配到试验组和对照组，意味着每个研究对象被分配进入各组的机会均等，使各种已知和未知的混杂因素在各组间均衡可比，从而减少系统误差（偏倚），确保研究结果更为真实可信。在分组过程中，还应注意对随机分配方案进行隐藏，同时对研究者和研究对象施盲，这样可减少信息偏倚。临床试验中常用随机分组方法有：简单随机、分层随机、区组随机等。详见本教材第三章。

3. **盲法原则**　为去除临床试验参与者（包括研究对象、观察者和资料整理分析者）的主观心理因素和行为对研究结果产生的某些干扰作用，建议采取盲法（blinding）。根据盲法的实施情况，可将临床试验分为开放试验、单盲试验、双盲试验和三盲试验，其中双盲试验最为常见。详见本教材第三章。

4. **重复原则**　重复是指样本含量的大小或重复实验次数的多少，可排除单个观察单位的观察或实验结果的偶然性。随机对照试验中由于抽样误差难以避免，足够大的样本量可将抽样误差控制在允许范围之内，从而获得更为接近真实的结果，重复原则是消除非处理因素影响的又一重要手段。

二、设计模式

（一）设计原则

随机对照临床试验的设计原则包括：①分配入组随机化；②组间具有可比性，特别是试验组和对照组基线资料要均衡可比，避免选择偏倚；③试验同步性和条件一致性，要确保试验组和对照组的干预措施在同时、相同条件下进行，以排除混杂的干扰；④试验实施一致性，在试验观察期间，对试验组和对照组的对待方式和态度等方面要保持一致性，采用盲法观察、记录，避免信息偏倚。

（二）设计模式

在试验开始实施前，应明确试验目的，确定试验对象的具体要求和来源，建立合理的纳入和排除标准，入选符合试验标准且知情同意的研究对象，入选方式既可从病患群体（目标人群）中随机抽样，也可连续纳入住院或门诊患者。再将入选研究对象随机分成试验组和对照组。明确规定研究因素，确定观察指标、随访观察的终止时间、间隔时间和资料收集方法，以及后期资料整理和分析的方法和标准等。在上述准备就绪后，方可进行试验。随机对照试验设计模式如图 4-1。

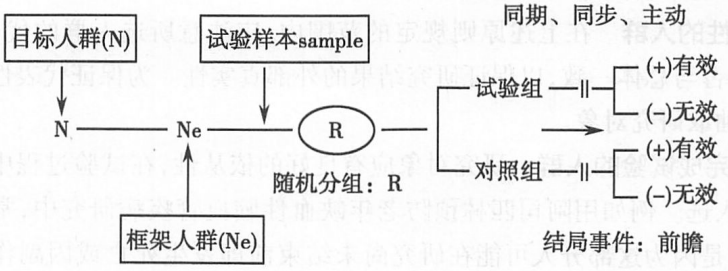


图 4-1　随机对照试验的设计模式

（三）结果分析模式

根据图 4-1 的设计模式，可列出四格表（表 4-1），将试验组和对照组的结果分别填入相应的表格内，分析比较两种干预措施的疗效。

NOTE

表 4-1　随机对照试验结果分析四格表

	结果		合计
	有效	无效	
试验组	$a(T_{11})$	$b(T_{12})$	$n_1=a+b$
对照组	$c(T_{21})$	$d(T_{22})$	$n_2=c+d$
合计	$m_1=a+c$	$m_2=b+d$	$n=a+b+c+d$

在 H_0 假设成立条件下，两组的率比相同。结合每一格的样本观察频数 A_{ij}（具体为表中的 a、b、c 和 d）与理论频数 T_{ij}（T_{11}、T_{12}、T_{21}、T_{22}）之间的差异，用下面的式(4-1)，式(4-2)进行 χ^2 检验。

$$T_{ij}=\frac{m_i n_i}{n}\quad (i=1,2;j=1,2) \qquad 式(4\text{-}1)$$

$$\chi^2=\Sigma\frac{(A-T)^2}{T} \qquad 式(4\text{-}2)$$

如果出现下列情况之一者，选用 Fisher 确切概率法：①样本含量<40；②理论频数 T<1；或 χ^2 检验后所得概率 P 接近检验水准 α。

三、方案要素设计

（一）研究对象的选择

基于不同的研究目的，选择与之相适应的研究人群，在收集目标人群的基线资料的基础上，按照事先制定的、合理的纳入和排除标准来选择，以避免某些外来因素的影响。应遵循以下原则：

1. 选用公认合理的纳入和排除标准　在明确的疾病诊断标准基础上，根据研究目的和具体条件，统一制定公认合理的纳入和排除标准。若纳入标准宽松，受试者代表性强，容易推广到临床实践，但入选受试者同质性较差且所需样本量大。若纳入标准严格，则受试者同质性好，但结果推广受限。因此，制定的标准要合理恰当。

2. 选择干预对其无害的人群　若拟研究的干预可能有潜在风险，这部分研究对象则不宜入选。如对胃有刺激性药物的临床试验中，有胃出血史者不宜选作研究对象；而可能出现较大副作用的对象也不宜入选，如老弱者或危重患者等。

3. 选择对干预措施有反应且可获益的人群　要结合受试者具体病情，让那些既能从中获益、风险又小的对象入选。如研究抗心律不齐药物疗效时，最好选择近期频繁发作者，而偶发者则不作考虑；又如评价某疫苗的预防效果，应选择高发区的易感人群为研究对象；再如评价强化补碘对轻度缺碘区居民的预防效果，宜选择孕妇和儿童。

4. 选择有代表性的人群　在上述原则规定的范围内，应注意所选人群的代表性，如性别、年龄、种族、职业等特征是否与总体一致，以保证研究结果的外部真实性。为保证代表性，在可能条件下，应从总体人群中随机抽取研究对象。

5. 选择预期能完成试验的人群　研究对象应有良好的依从性，在试验过程中有可能被剔除而不能完成试验者不宜入选。例如用阿司匹林预防老年缺血性脑血管疾病研究中，常将合并有癌症者或有胃病者排除，主要是因为这部分人可能在研究尚未结束前即发生死亡或因副作用过大而被迫停止试验。

（二）样本量估计

为控制Ⅰ、Ⅱ型错误率，在试验设计阶段，就应估计研究所需的样本量。一般情况下，样本量越大，Ⅰ、Ⅱ型错误率就越低；但样本量过大，不仅导致人力、物力、财力和时间的浪费，而且会给试验的质量控制带来诸多困难。

1. 样本量大小的主要影响因素

（1）干预措施的预期效应大小：干预措施的效应越大，试验组和对照组的目标事件（疾病）发生率的差异越大，所需样本量越小，反之越大。

（2）Ⅰ型（α）错误出现的概率：即出现假阳性错误的概率。确定的 α 越小，所需样本量就越大。通常将 α 定为 0.05。

（3）检验效能或把握度（1-β）：β 为出现假阴性错误的概率（Ⅱ型错误概率），1-β 为把握度。把握度越高，则所需样本量越大。通常将 β 定为 0.10。

（4）单侧检验或双侧检验：单侧检验比双侧检验所需样本量小。倘若凭借专业知识有充分把握排除两种可能中的一种（如试验组的效果优于对照组），就用单侧检验，否则就用双侧检验。

（5）资料类型：计数资料以治愈率、生存率、病死率等为分析指标时，基础发生率（p）越低，所需的样本量越大。而计量资料需要考虑总体的标准差，标准差越大，所需样本量越大。

2. 样本量的估计

（1）计数资料的样本量估计：如果结局变量是计数资料指标，如发病率、感染率、病死率和治愈率等，则可按下列公式计算样本量：

$$n=\frac{\left[Z_{\alpha}\sqrt{2\bar{p}(1-\bar{p})}+Z_{\beta}\sqrt{p_1(1-p_1)+p_2(1-p_2)}\right]^2}{(p_1-p_2)^2} \qquad \text{式(4-3)}$$

p_1：对照组某结局的发生率；p_2：试验组某结局的发生率；$\bar{p}=(p_1+p_2)/2$。

Z_{α}：为 α 水平相应的 Z 值；Z_{β}：为 β 水平相应的 Z 值。

n：为计算所得一个组的样本大小。

（2）计量资料的样本量估计：如果结局变量是计量资料指标，如血压、血糖等，则可按下列公式计算样本量：

$$n=\frac{2(Z_{\alpha}+Z_{\beta})^2\sigma^2}{d^2} \qquad \text{式(4-4)}$$

σ：为估计的标准差；d：为两组结局变量均值之差；Z_{α}、Z_{β} 和 n 所示意义同公式 4-3。

为快速征募病例并完成研究，一般会联合多家单位，一个单位为一个中心，各中心分别征募和治疗本中心的病例，在统一研究设计方案下可在不同地点同步进行。

（三）资料收集

资料收集包括基线资料和随访资料收集两个部分。基线资料指研究对象在研究开始时详细的基本情况，一般包括研究对象的疾病与健康状况、年龄、性别、文化、职业、婚姻等个人情况、家庭情况、个人生活习惯及疾病史等。随访资料收集指研究开始后采集试验组和对照组的相关资料。随访方法主要有面对面访问、电话访问、仪器检测、问卷调查、环境与疾病监测、病情记录等。随访方法在整个随访过程中应保持不变。

研究开始时要确定随访间隔和观察终止时间，其长短由试验本身的性质所决定。例如，降压药对血压的影响，其观察终止时间较短；研究加碘食盐对甲亢的影响，其随访时间较长。此外，预期结果的确定，即观察终点需事先确定好。

整个资料的收集过程要注意质量控制，使结果具有真实可比性。要求所有研究对象都要坚持到观察终止时间（病情严重等特殊紧急情况除外），尽量减少失访。在收集资料开始前应对资料收集者进行专业统一培训，保证资料收集的质量。

四、资料的整理与分析

（一）整理资料

收集的资料不一定都是完整、真实和规范的，需要进一步核实资料，核对无误后将资料录入、归

类，便于下一步分析。资料整理时需特别注意三种情况：

1. 不合格的研究对象，即不符合纳入标准，未参与试验者或无任何数据者，一般要剔除。但需注意的是，研究者往往对试验组的对象观察和判断更加仔细，剔除不合格人数多于对照组，造成结果出现误差。因此，有学者建议将“合格者”和“不合格者”两个亚组的基线资料分别进行分析，如果两者结果不一致，则在下结论时应慎重。

2. 不依从的研究对象，即研究对象不遵守试验所规定的要求。不依从的原因：①研究对象对试验不感兴趣；②试验或对照措施有副作用；③研究对象病情加重，不能继续试验；④其他特殊原因，无法继续进行试验研究。

对于这类资料，在整理资料时，可以根据研究对象的依从性进行分组并分析，有四种情况：①未完成 A 治疗或改为 B 治疗者；②完成 A 治疗；③完成 B 治疗者；④未完成 B 治疗或改为 A 治疗者。对此，一般有三种处理方法：

（1）意向性分析（intention-to-treat analysis，ITT）：比较①组+②组与③组+④组。它反映了原来的试验意向，该种分析往往会低估其效果，但如果 ITT 分析表明 A 干预措施有效，则基本可以确认 A 措施有效。

（2）遵循研究方案分析[per-protocol（pp）analysis]：只比较②组和③组，而不分析①组和④组。它只对试验依从者进行分析，能反映试验药物的生物效应，但剔除了不依从者，可能高估干预效果。

（3）实际接受干预措施分析：比较①组+③组和②组+④组。对实际接受了干预措施者进行分析，鉴于研究对象未遵守试验规程，偏离或违背方案，两组缺乏可比性，结果可能失真。

3. **失访的研究对象**　是指研究对象因迁移、与本试验无关的死亡等其他原因造成无法随访的现象。对于失访人群，要弄清楚失访的原因和各组的失访率，若两组失访率接近且原因和未失访者特征大致一样，其对结果的影响一般较小；若失访率不同，则会产生偏倚。

（二）统计分析

就疗效评价临床试验而言，其最终目的是比较两组或多组疗效之间的差异。若存在差异，则应检验其有无统计学意义，同时应评价其临床意义。临床意义的评价应包括利弊分析（效果、副作用和成本）以及治疗复杂性和方便程度等方面。无论是治疗性还是病因学的 RCT 研究，结果分析内容主要为两（多）组计数指标的比较、两（多）组计量指标的比较、相关性分析以及多因素分析等方面。具体参见相关统计学专著。

五、注意事项

（一）伦理问题

伴随现代医学快速发展，临床研究中的伦理学问题也日益突出。医者不仅要有高超的技术和职业操守，更为可贵的是要有一颗仁爱之心。无论是临床试验还是其他干预性研究，凡涉及人身安全问题，都应遵循医学伦理原则，一切以研究对象的健康安全为中心，切忌在试验中有意或无意出现有悖伦理行为。

（二）研究对象选择的问题

在设置入选和排除标准时，最好选择新发且发作频率较高的患者，这类患者对干预措施的效应反应明显，易于观察记录，且具有较好的反应性和敏感性。而对于孕妇、幼儿、老年人，则一般不宜选为研究对象，确有必要时，应结合这类人群的身体情况加以选择。

（三）试验过程的质量控制问题

开展临床试验的目的，就是为了探讨安全、有效、廉价、适宜的治疗方案，促进临床医学的进一步发展。在试验过程中，可能会出现研究对象不依从、失访等现象，对于这些研究对象的处理，需要一套严格的方案。此外，试验过程中的未知突发情况的处理，也需要事先设定好预案。

（四）可行性问题

一项研究能否得以成功实施，应进行可行性分析。特别是大型随机对照试验，需要先在小范围内做试点，以评价设计方案的可行性。

六、优点和缺点

（一）优点

1. 前瞻性的对照设计　在研究开始时，研究因素和研究目的已经确定，试验过程中可人为控制研究对象的条件和暴露情况，同时两组同步、同条件下观察，且两组观察指标相同，外部因素干扰较小，属于前瞻性研究，可以验证因果推断。

2. 组间可比性好　随机化可使每个研究对象入组概率相同，从而保持组间基线均衡可比，特别是分层随机化，控制某些影响大的因素，组间可比性更强。

3. 偏倚控制较好　采用随机分组可控制选择偏倚和混杂偏倚，采用盲法、标准化评价和测量指标等有效避免信息偏倚，采用对照可进一步排除非处理因素对试验结果的影响。

4. 病例诊断和干预措施标准化　标准化是随机对照试验结果质量的保证，采用严格一致的诊断/纳入和排除标准，有利于防止各种主观、人为因素干扰，确保研究结果的客观性，接近事实。

此外，高质量的 RCT 可作为系统综述的可靠资源和重要素材，一个高质量的 RCT 所提出的问题往往会成为该领域的研究导向。

（二）缺点

1. 成本高　随机对照试验需要严格的质量控制，实施条件苛刻，往往对研究对象质量、试验条件和观察指标等要求严格，耗费人时、物力、财力。

2. 外部真实性受限　随机对照试验往往是在小范围开展的，其研究对象一般是按照研究者拟定的标准控制入选，具有良好的同质性，内部真实性较高，但良好的同质性难以代表人群的差异性，其外部真实性相对受限。

3. 可能违背伦理　在随机对照试验中，研究者为了达到试验目的，往往会有意无意地使研究对象暴露于危险因素下，或隐藏试验信息，又或使病情加重者继续参与试验，给研究对象带来伤害，违背伦理。

（张秀军）

第二节　随机对照试验的一些特例

一、交叉试验

（一）概述

交叉试验（cross-over design）属于一级设计方案，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完成，即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。它属于随机对照试验的一种特殊类型，兼有随机对照试验和自身前后对照试验的优点，可以采用完全随机化和分层随机化来分配受试者。

在试验正式开始之前，符合标准的试验对象先进入筛选观察期，了解入选对象是否处于一个相对稳定的状态。经过一定时间观察后，进入第一阶段试验，将符合标准的研究对象随机分为试验组和对照组，如试验组首先接受方案 A，而对照组接受方案 B，经过一定时间干预后，观测并获得相应的结果；此后交叉试验进入洗脱期（washout period），旨在消除第一阶段的效应影响，即剩余效应（carry-over effects）或携带效应。经过洗脱期后方可交叉进入第二阶段试验，即两组干预方案交叉互换，试验组改为方案 B，对照组接受方案 A。按照第一阶段设置的指标和测量方法，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。

NOTE

交叉试验的设计要注意前后两个试验阶段的间隔期，即药物洗脱期设置要合理，其长短取决于药物在血清中的衰减度，一般要求不少于5个半衰期。

（二）应用范围

一般而言，交叉试验仅适用于不易根治并需要药物长期维持治疗的慢性疾病研究。如高血压病、冠心病、心绞痛、支气管哮喘等等，而对于那些一经治疗就能痊愈者则不能采用交叉试验设计方案。此外，在上市前新药临床试验中，为减少样本含量，Ⅰ期临床试验常采用交叉试验来观察药物的毒副反应，以减少或消除个体间差异的影响。

（三）设计模式及特点

交叉试验先将受试者随机分成两组，后进入两个试验阶段，第一阶段中甲组接受方案A、乙组接受方案B干预，两组同时进行随访观察。经过一定洗脱期后进入第二阶段，甲乙两组交叉互换A、B方案。最后对两个阶段观察到的结果进行综合比较（图4-2）。

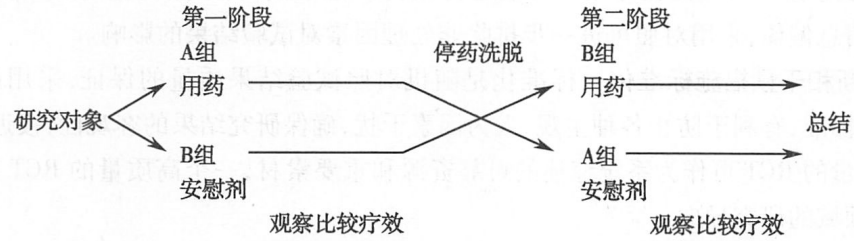


图4-2　交叉试验设计模式图

（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，必然得到两种方案的观察结果。由于系自身前后对照，对定性资料应采用配对χ^2检验（McNemar检验），而定量资料则采用配对t检验或符号秩和检验。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4-2），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4-2　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	$c+d$
合计		$a+c$	$b+d$	N

（五）优缺点

1. 优点

（1）具有随机对照试验的优点：交叉试验作为随机对照试验的特例，采用随机分组、盲法测量和同期对照的方法，有效控制了选择偏倚、信息偏倚和混杂偏倚，试验结果真实可靠。

（2）设计特点：每个受试对象先后接受两种方案处理，具有自身前后比较的特点，消除了个体差异，同时也获得组间比较的结果，可大大减少样本量。

2. 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，如过短则难以避免两种措施的沾染和干扰，过长影响试验周期，甚至使患者长时间得不到应有的治疗，影响预后。

（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，因而

笔记

降低了阶段之间的可比性，影响疗效评估。

（六）主要用途

1. 研究药物应用先后顺序对治疗效果的影响　交叉试验在试验组和对照组分别按不同顺序先后进行了不同的试验措施，可以探讨药物不同的顺序对治疗效果的影响。

2. 研究药物最佳配伍的方案　交叉试验中，一组研究对象至少先后服用两种药物，在此过程中既可观察单药物对研究对象的影响，也可发现多种药物组合使用后的研究结果，以此找出药物最佳配伍的方案。

二、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）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，最好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 随机地进入第一阶段研究，受试者先接受方案 A 的干预，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。然后进入消洗期，消洗期（洗脱期）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。如比较用药前、后体内某些指标的变化情况，以判断药物的疗效；或研究同一机体不同部位对不同药物的反应，分析何种药物疗效更佳。

（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 4-3）。

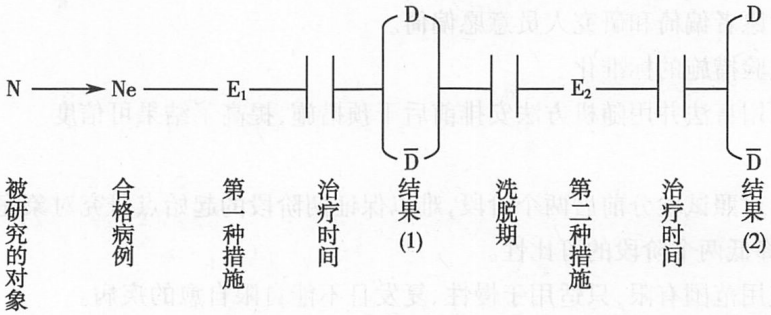


图 4-3　自身前后对照研究设计模式

N：符号公认的诊断标准的患者总数或人群总数；Ne：该人群或患者中符合纳入标准又不具备排除标准的人数；E_1：未暴露可疑致病因素或未接受防治措施的对照组；E_2：暴露可疑致病因素或接受防治措施的试验组；D：发病人数，有效人数，生存数等；$\bar{D}$：未发病人数，无效人数，死亡数等；-||-：为随访期或观察期，中间填时间

（四）结果分析

参与自身前后对照试验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的观察结果。受试对象可能有四种情况（表 4-3），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χ^2检验。

表4-3　自身前后对照研究结果分析四格表

第一种方案	第二种方案		合计
	阳性	阴性	
阳性	a	b	n_1
阴性	c	d	n_2
合计	m_1	m_2	n（固定值）

表4-3可得出：第一种方案阳性率$=\frac{n_1}{n}=\frac{a+b}{n}$；第二种方案阳性率$=\frac{m_1}{n}=\frac{a+c}{n}$

第一种方案的阳性率-第二种方案的阳性率$=\frac{a+b}{n}-\frac{a+c}{n}=\frac{b-c}{n}$

表中所示，两个方案阳性率的比较只与b、c有关，而与a、d无关。因此，要比较两种方案阳性率有无差别，只要对其中的b与c作χ^2检验即可。在假设两种检验方法的阳性概率相同的条件下，b与c两个格子理论频数都应是$(b+c)/2$，当$b+c\geqslant 40$，

$$\chi^2=\frac{(b-c)^2}{b+c}\quad v=1 \qquad \text{式(4-5)}$$

当$b+c<40$，需要对式(4-5)校正，公式为：

$$\chi_c^2=\frac{(|b-c|-1)^2}{b+c}\quad v=1 \qquad \text{式(4-6)}$$

（五）优缺点

1. 优点

（1）受试者以自身为对照，可消除个体差异，节省样本量、时间和成本。

（2）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）可实现试验措施的标准化。

（5）试验中采用盲法并用随机方法安排前后干预措施，提高了结果可信度。

2. 缺点

（1）自身前后对照试验分前后两个阶段，难以保证两阶段的起始点研究对象完全一致，难以排除时间因素的影响，降低两个阶段的可比性。

（2）试验的应用范围有限，只适用于慢性、复发且不能自限自愈的疾病。

（3）试验过程包括一定洗脱期，如过短则难以避免两种措施的沾染和干扰，过长影响试验周期，甚至使患者长时间得不到应有的治疗，影响预后。

三、基于单个患者的随机对照试验

单个患者的随机对照试验（number of one randomized controlled trial，n-of-1 trial）是基于罹患慢性疾病的单个个体，进行一种多轮、多阶段随机对照试验，以确定多种治疗措施中哪一种对该患者有效，从而避免服用多种药物，减少浪费以及避免误服某些无效甚至有害药物。

方法是将所有“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；进

笔记

而根据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，所有试验药物的观察期应保持一致；通常采用双盲法，以利于试验结果的评价。基于单个患者的随机对照试验适用于慢性复发性疾病，如偏头痛、心绞痛、支气管哮喘等慢性病，同样也适用于药物筛选，探讨何种药物对患者有效。

四、非等量随机对照试验

非等量随机对照试验（unequal randomized controlled 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，各组例数不等量。主要应用于新药疗效验证研究，特别是由于患者来源和研究经费有限而研究者希望尽快获得结果的情况。但需要注意的是试验组和对照组的例数不能相差过大，如超出1∶4或4∶1比例，检验效能反而会有所下降。

非等量随机对照试验虽能快速地获得试验结果，但在人数上试验组远多于对照组，可能会出现选择偏倚。

五、整群随机对照试验

整群随机对照试验（cluster randomized controlled trial）不同于多数随机对照试验以患者个体为随机分配单位，而是在某些特殊情况下，以多个个体组成的小群体作为分配单位，进行随机分组。

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人入选，有可能其中1人分入特殊膳食组，另2人分入普通饮食组。而在日常生活中，一家人不可能长期做两种不同膳食，即使做到了，沾染或干扰现象也难以避免，影响研究结果。显然单个体不宜作为试验的分配单位。于是就可以一个家庭，一对夫妇，一个小组甚至一个乡镇等作为随机分配单位，将其随机分入试验组或对照组，分别接受相应的措施，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，它是以群体为单位进行随机、干预、随访观察，虽易于组织、实施方便，节省人力、物力；但当以群体水平上得出的研究结果与该人群中的个体真实情况有出入时，特别是群体内个体变异大而群间变异小时，所需样本量较大。此外，资料的统计分析也应采用基于整群随机分组的统计分析方法。

（张秀军）

第三节　其他类型的临床对照试验

一、半随机对照试验

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，即按研究对象的生日、住院日或住院号等末位数字的奇偶数，将试验对象分配到试验组或对照组，接受不同干预措施。

半随机对照试验常因分配方式为不完全随机，容易出现选择偏倚，造成基线情况的不均衡，因此，虽然花费的时间、精力、财力并不亚于随机对照试验，但其结果的真实可靠性却不及随机对照试验。故在条件允许的情况下，尽量选择完全随机对照试验。

二、非随机同期对照试验

非随机同期对照试验（concurrent nonrandomized controlled trial）是一种传统的临床研究设计，与随机对照临床试验相似，唯一的不同在于受试对象不是随机分配的，或者由于研究数量大、范围广而实际情况不允许对研究对象实施随机分组的情况，而由患者或医生根据病情及有关因素人为的、非随机分配入组，并同期进行结果观察。

非随机对照试验属于前瞻性研究。常用于比较临床不同干预措施的实际效果。该试验在研究对

象的分组分配上，由于人为因素，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，使得许多已知/未知的偏倚影响观测结果的真实性。但在临床实际工作中，有些情况下不适合用随机对照试验，例如外科手术治疗、急重症患者抢救或贵重药物的选用等，因而只能根据具体情况将患者分入试验组或对照组。尽管非随机对照试验的论证强度不及随机对照试验，但在尚无随机对照试验结果或不能进行随机对照试验时，还是可取的，尤其是对于大样本的非随机同期对照试验结果，仍具有重要的临床价值。只不过在分析和评价研究结果的价值及意义时，应持审慎的科学态度。该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；但由于存在选择偏倚和测量偏倚，降低了结果的真实性，其论证强度也相应减弱。

三、历史性对照试验

历史性对照试验是将现有干预措施的研究结果与过去的研究结果作非同期比较，主要用于评价现有的治疗方案或干预措施的作用。

历史性对照试验为非同期对照试验，无法随机分配研究对象，其主要研究资料来源于过去的医学文献和病案资料，再与现有的试验研究结果作比较。历史性对照试验可节省时间和经费开支，不存在伦理问题；但由于临床医学内外环境随时间变迁而发生改变，其前后可比性较差。目前主要用于一些预后极差的疾病，如胎儿畸形过去归因于遗传因素，而现在添加叶酸后畸形发生率下降，则提示叶酸确有预防胎儿神经管畸形的作用。历史性对照虽可比性较差，不及同期对照，但在临床医学发展过程中，仍不失为一种选择。

四、单臂研究

单臂研究（single arm study）顾名思义，就是仅有一个组的研究，没有为试验组设计相对应的对照。但单臂研究并不是就没有参比的对象，实际上，它的参比对象是采用他人、或过去的研究结果，与试验组进行对照比较。单臂研究可以是多中心的临床试验，但肯定是开放的，也谈不上随机与盲法，所以论证强度较差。单组临床试验的优点是所有的受试者都接受同一种试验药物，试验设计、实施趋于简单、易行，但其缺点也是显而易见的，由于使用试验药物的受试者与外部对照的受试者不来自同一受试者总体，可比性差，仅适用于个别特殊情况。若所研究的疾病为非常严重的或罕见疾病，目前尚无一个满意的治疗药物，同时根据研究药物的作用机制、动物实验以及早期的经验，可使用单组临床试验开展探索性研究。

（张秀军）

第四节 队列研究

除上述试验性研究（experimental studies）外，最常见的就是观察性研究（observational studies），其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，比较的组别系自然形成。观察性研究可进一步分为分析性研究（如队列研究、病例对照研究）和描述性研究（如断面研究、病例分析）。其中队列研究（cohort study）是探讨预后与病因、研究疾病自然史和监测药物不良反应等常用的方法之一，其论证强度高，所获结果可靠性强，能较好地揭示两事物间客观存在着的因果关系。

一、概述

（一）概念

队列研究是将一群（组）研究对象（队列）按是否暴露（exposure）于某研究因素分为暴露组与非暴露组（对照组），随访观察适当长的时间，比较两组之间所研究疾病（或事件）的发病率（或发生率）或死亡率差异，从而判断这个（些）暴露因素与疾病之间有无关联及关联大小的一种观察性研究方法。

队列研究又被称为随访研究、定群研究、前瞻性研究(prospective study)等,但以队列研究这一名称最常用。“暴露”是指研究对象接触过某种因素(如重金属、药物),或具有某些特征(如年龄、性别、遗传)、行为(如吸烟)。暴露可以是危险因素,也可以是保护因素。

(二) 特点

1. 属于观察性研究。队列研究中所观察的暴露因素不是人为给予的,而是客观存在于研究人群中,研究因素的暴露情况及其变化是通过研究者观察获得。

2. 属于由因及果研究。暴露在前,结果发生在后。暴露开始时,研究的疾病尚未发生,或观察的临床结局尚未出现,随访一段时间后,才能观察到研究的结局是否发生。由“因”推“果”,符合先因后果的推理逻辑。

3. 研究对象按是否暴露于所研究的因素分为暴露组和非暴露组,其暴露状况已客观存在,研究者不能主动控制,如将其随机化分配。

二、设计模式

队列研究设计的基本原理见图4-4。队列研究按其研究时间的起止点(时序),又可分为三种设计模式:前瞻性队列研究、回顾性队列研究和双向队列研究。

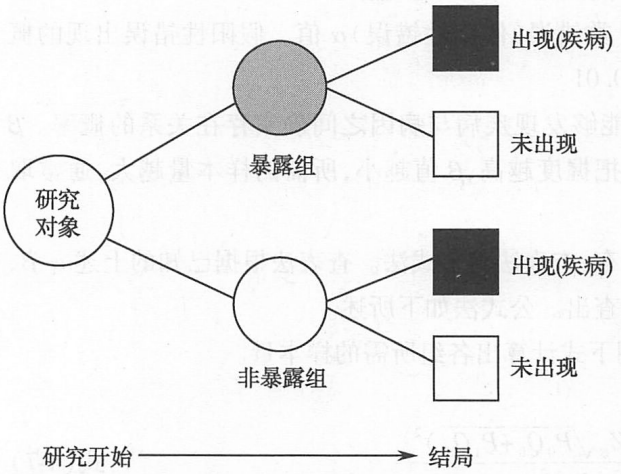


图4-4　队列研究基本原理示意图

1. 前瞻性队列研究(prospective cohort study)　是指观察时间从现在开始,追踪观察到将来某个时间,了解其发病或死亡情况,以确定某暴露因素与疾病发生发展间的关系。通常所说的“队列研究”就是指这种研究,属于队列研究的基本模式。

2. 回顾性队列研究(retrospective cohort study)　是指以过去某个时间为起点,收集基线和暴露资料,并按当时人群对研究因素的暴露情况自然分为暴露组和非暴露组,追踪观察到现在发病或死亡情况,以研究暴露与疾病发生发展的关系。这种设计模式又称历史性队列研究(historical prospective study)。回顾性队列研究的前提是过去有关暴露与结局的观察记录必须准确和完整。尽管收集暴露与结局资料的方法是回顾性的,但实质上仍是从因到果的研究方法。

3. 双向性队列研究(ambispective cohort study)　是指在回顾性队列研究之后,继续追踪观察到将来某个时间,又称为混合型队列研究,它融合了前瞻性队列研究和回顾性队列研究的优点,并在一定程度上弥补了相互的不足。在进行回顾性队列研究的过程中,如果从暴露到现在仍未达到足够的阳性事件发生数,还需继续前瞻性观察一段时间时,就可选用双向性队列研究。

三、方案要素设计

在确定开展某因素(如:吸烟)和结局(如:肺癌)因果关联的队列研究后,实施方案的制定和要素设计是研究成败的关键一环。

(一) 研究对象的选择

1. 研究现场　由于大多队列研究的随访时间较长,因此研究现场必须有足够的、相对稳定的人口基数,同时组织协调也很重要,当地也要具备较好的医疗条件,交通便利,便于随访观察。对于人口流动大的地区或单位,失访率会较高,对结果有较大影响,一般不适于作为研究现场。

2. 研究人群　研究人群包括暴露组和非暴露组,根据研究目的和研究条件的不同,研究人群有

不同的选择方式。

（1）暴露组：即对研究因素有暴露的人群。通常可以选择在某社区或地理区域内居住的全体人群（或医院住院患者），其中暴露于某研究因素（如：吸烟、药物）的人群即为暴露组。

（2）非暴露组：即对照人群。观察人群确定后，将其中暴露于所研究因素的对象作为暴露组，其余即为非暴露组（如：不吸烟者、未用药）。但非暴露组人群与暴露组要有可比性。对照人群除未暴露于所研究因素外，其他因素或一般人口学特征（年龄、性别、民族等）都尽可能地与暴露组人群相同。但是，当以特殊暴露人群（包括职业人群）作为暴露组时，不应在同一人群选择对照组，而应在与该暴露或职业无关的另一人群中选择对照。

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时，则对照组（非暴露组）应在无该因素的地区或人群中选择，谓“外对照”。

（二）样本量的估计

1. 样本量估计有关的参数 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。P_0越接近0.5，所需样本量越大。

（2）暴露人群中的发病率（P_1）：可通过查阅文献或预调查获得。如果已知预期相对危险度（RR），则可通过公式 $P_1=RR\times P_0$ 求得。P_1-P_0 差值越大，所需样本量越小。

（3）显著性水平（α 值）：是检验假设时的第Ⅰ类错误（假阳性错误）α 值。假阳性错误出现的概率越小，所需样本量越大，一般情况取 $\alpha=0.05$ 或 0.01。

（4）把握度（1-β）：又称为效力（power），反映能够发现疾病与病因之间确实存在关系的概率。β 是检验假设时第Ⅱ类错误（假阴性错误）的概率，把握度越高，β 值越小，所需的样本量越大，通常取 $\beta=0.10$ 或 0.20。

2. 样本量估算方法 常用的样本量估计方法有：查表法与公式法。查表法根据已知的上述 α、β、P_0 和 RR 四个基本参数，可从参考书的相应附表上查出。公式法如下所述。

在暴露组和对照组样本量相等的情况下，可用下式计算出各组所需的样本量。

$$n=\frac{(Z_\alpha\sqrt{2\bar{P}\,\bar{Q}}+Z_\beta\sqrt{P_0Q_0+P_1Q_1})^2}{(P_1-P_0)^2} \qquad \text{式(4-7)}$$

式中：P_1 和 P_0 分别代表暴露组和的对照组预期发病率，$\bar{P}=(P_0+P_1)/2$，$\bar{Q}=1-\bar{P}$，$Q_0=1-P_0$，$Q_1=1-P_1$，Z_α、Z_β 为标准正态分布下的面积。

在计算样本量时一般按 10% ~20% 估计失访率，所以应在原估计样本量的基础上加 10% 以上作为实际样本量。

（三）资料的收集与随访

1. 收集基线资料 队列人群确定后，应全面收集每个研究对象在研究开始时的基本信息，即基线资料（baseline information），这些资料是区分暴露组和非暴露组、判定研究结局的重要依据。资料收集的方法包括：①调查询问：调查的内容包括一般人口学特征（如性别、年龄、职业、民族、居住地、经济收入等）、研究因素的暴露情况（如吸烟、饮酒等）及身体健康状况等。②查阅现成记录：如医院病历、常规的出生死亡登记、各种人口与疾病统计、户口登记等。③体格检查和实验室检查：对于身高、体重、血压、血脂、血糖等研究相关指标需做相应的检查。进行现场流行病学研究时，还需环境调查和检测等，如探讨多囊卵巢综合征与环境激素暴露的关系时，需要检测水、土壤、大气中的环境激素。

2. 随访观察 暴露因素确定后，结局事件的发生往往需要一定的时间，在这个过程中，就必须对研究对象进行随访（follow-up），应结合研究目的和具体病种确定随访间隔时间和观察终止时间。

所有研究对象包括暴露组和非暴露组，应在相同时期内以相同的方式进行随访观察，不可有先有

笔记

后，中途放弃或漏访。随访方法与基线资料的收集方法相同，同时要求随访方法在整个随访过程应保持不变。

随访内容依研究目的和设计要求不同而各异。一般而言，应与收集基线资料的内容和方法一致。但是应重点观察以下内容：①暴露人群暴露情况及程度有无改变。如吸烟者吸烟量有无变化，是否戒烟等；对某些定量指标如血压、血糖等进行体格检查和实验室检查。②收集结局相关资料。如发病日期、入院时间、诊断方法、死亡原因、死亡时间和地点等。③人口变动的情况。如进入、退出、失访人数等。

3. 观察终点　随访观察终点（end-point）多为发生疾病或死亡，但必须是研究所限定的疾病或死亡。观察过程中发生其他疾病或死亡不能视为结局事件。如研究高血压和冠心病的关系，则观察终点是冠心病，随访过程中，研究对象发生了肿瘤或其他疾病，则不能认为是到了观察终点，应继续随访；倘若某研究对象死于肝癌，应作为失访处理，也不能视为结局事件。终点指标与测量标准都应在研究设计阶段明确规定，在观察中途不能改变，应按事先制定好的诊断标准统一进行评价，否则将造成结果的偏倚。

队列研究观察时间较长，在随访过程中应加强质量控制，以保证研究结果的可靠性和准确性。

四、资料的整理与分析

队列研究资料主要是计算暴露组和非暴露组的发病率或死亡率，检验其差异的显著性，分析暴露因素与发病或死亡之间是否有联系。如存在联系则进一步计算相对危险度等指标，分析联系的强度。

队列研究资料整理的基本形式见表4-4。表中暴露组的发病率为$a/(a+b)$，非暴露组的发病率为$c/(c+d)$，在控制各种偏倚后，如果暴露组的发病率显著高于非暴露组，则说明暴露因素与该疾病有一定关联，且很可能是因果关系。

表4-4　队列研究资料整理表

组别	病例	非病例	合计
暴露组	a	b	$a+b$
非暴露组	c	d	$c+d$
合计	$a+c$	$b+d$	$a+b+c+d$

1. 率的计算　队列研究计算的发病概率与普通发病率不同，根据队列人群的数量、稳定程度、发病强度和观察时间的不同，可以计算累积发病率、发病密度等指标。

（1）累积发病率（cumulative incidence，CI）：当研究对象人数较多，人口稳定，观察时间较短，资料比较完整且失访人数较少时，以该研究人群观察开始时的人口作分母，以观察期内某病的新发病例数作为分子计算累积发病率。累积发病率的量值变化范围为0～1，其数值高低受研究时间长短的影响，报告累积发病率时应说明观察时间的长短。

$$n\text{年的某病累积发病率}=\frac{n\text{年内的新发病例数}}{n\text{年内的平均暴露人口数}}\times 1000‰(100\ 000/10\text{万})\qquad 式(4\text{-}8)$$

（2）发病密度（incidence density，ID）：当研究人群观察时间较长，研究对象不断增减、队列人群难以稳定时，则需要以人时（person-time，PT）为单位代替平均人口数作分母来计算率。这时计算出来的是一种人口数有“变动”的发病率，即发病密度，而不是静止的累积发病率。计算发病密度时，分子为观察期间研究疾病的发病数，分母则不是普通人口数，而是用人时来表示。人时等于观察人数乘以随访时间，人时单位可为人年、人月等，通常用人年来表示。如1000人观察了1年或100人观察了10年均为1000人年。这种以人时计算的发病密度又称为人时发病率。在比较几项随访时间不等的前瞻性研究发病率时多用发病密度。

NOTE

$$发病密度=\frac{某人群在观察期内的发病数}{观察期内的观察对象人年数}\times 100\ 000/10\ 万 \qquad 式(4\text{-}9)$$

计算人年的常用方法有近似法和寿命表法。

2. 联系强度的估计 当统计学检验提示暴露与疾病具有显著的统计学联系时，应进一步计算联系强度。常用的指标有：

（1）相对危险度（relative risk，RR）：又称率比（rate ratio），是暴露组发病（或死亡）率（I_e）与非暴露组发病（或死亡）率（I_o）的比值。它表明暴露组发病或死亡的危险性为非暴露组的倍数。

$$RR=\frac{I_e}{I_o}=\frac{a/(a+b)}{c/(c+d)} \qquad 式(4\text{-}10)$$

RR 的 95% 可信区间 $=RR^{1\pm 1.96/\sqrt{\chi^2}}$

RR 值是反映暴露与结局事件联系强度的重要指标之一，即暴露于某因素所致的发病（或死亡）率是不暴露者的倍数，具有病因学意义。RR<1 说明暴露因素是疾病的保护因素，RR>1 说明暴露因素是疾病的危险因素，RR=1 说明暴露因素与疾病无联系。

（2）归因危险度（attributable risk，AR）：又称特异危险度或率差（rate difference）。是指暴露组的发病率或死亡率（I_e）与非暴露组的发病率或死亡率（I_o）之差。AR 表示暴露者中完全由某暴露因素所致的发病率或死亡率。

$$AR=I_e-I_o$$
$$或 \quad AR=I_o(RR-1) \qquad 式(4\text{-}11)$$

AR 的 95% 可信区间 $=AR^{1\pm 1.96/\sqrt{\chi^2}}$

AR 说明暴露人群比非暴露人群增加的疾病发生率，如果去除该因素，就可减少这个数量的人发病，因此更具疾病预防和公共卫生意义。

（3）归因危险度百分比（attributable risk percent，AR%）：表示暴露者中由暴露所致的发病率或死亡率（I_e-I_o）占暴露者发病率或死亡率（I_e）的百分比。AR% 与 RR 的大小有关。

$$AR\%=\frac{I_e-I_o}{I_e}\times 100\% \qquad 式(4\text{-}12)$$
$$或 \quad AR\%=\frac{RR-1}{RR}\times 100\%$$

（4）人群归因危险度（population attributable risk，PAR）：表示整个人群中，暴露因素所引起的发病率增高的部分，又称为病因分值（etiologic fraction）。是整个人群中该疾病的发病（或死亡）率（I_t）与非暴露人群发病（或死亡）率（I_o）之差。表示人群中因暴露于某因素所致的发病（或死亡）率。

$$PAR=I_t-I_o \qquad 式(4\text{-}13)$$

（5）人群归因危险度百分比（population attributable risk percent，PAR%）：表示人群中由于暴露所致的发病率或死亡率占人群总发病率或死亡率的百分比。PAR% 在卫生行政部门制订疾病预防策略时需加以考虑，也可用于疾病预防的宣传教育，它可向群众说明在完全控制该暴露因素后人群中某病发病（或死亡）率可能下降的程度。

$$PAR\%=\frac{I_t-I_o}{I_t}\times 100\% \qquad 式(4\text{-}14)$$

笔记

五、应用范围

1. 预后研究和检验病因假设　预后研究与检验病因假设是队列研究的主要用途和目的。一般说来，队列研究往往是明确因果联系最有力的方法。

2. 防治效果评价　如戒烟可减少既往吸烟者发生肺癌的危险，那么对戒烟效果的评价就是队列研究的目的。

3. 研究疾病自然史　队列研究可以观察人群暴露于某因素后，疾病逐渐发生、发展，直至结局的全过程，包括亚临床阶段的变化与表现，因此队列研究不仅可以了解个体疾病的自然史，而且可了解疾病在整个人群中发展和流行过程。

4. 新药上市后监测　新药通过三期临床试验上市后，还需对其不良反应等进行长期监测，队列研究可以更大样本和更长时间地观察该新药的各种不良反应事件。

六、优缺点

1. 优点　符合因果推断的时间和逻辑顺序，如验证病因与疾病之间的因果关系论证强度高；可以直接计算暴露组和对照组的发病率或死亡率，获得 RR、AR 等指标；一般不存在回忆偏倚；可以了解疾病的自然史；能对暴露因素所致的多种疾病同时进行观察，从而获得一种病因与多种疾病的可能因果关系。

2. 缺点　所需研究时间长，样本量大，人力、物力投入大，容易产生失访偏倚。一般不适用于发生率很低且潜伏期又很长的疾病的病因和防治效果评价研究。例如包皮环切术预防阴茎癌的效果研究，即使在未进行包皮环切的男性中，其终生累积发病率也仅有 0.16%，若想有 80% 的机会通过队列研究设计发现两者之间的因果联系，至少需要随访 6000 名研究对象 67 年方能得出结论。

（闫永平）

第五节　病例对照研究

病例对照研究（case-control study），也是临床上开展病因及预后研究最有实用价值的研究设计方案。它通过科学合理的对照设置，可以在一定程度上防止偏倚的干扰，是探讨病因及危险因素、评价预后等较为常用研究设计之一，对于防治效果评价研究也可选用，而且随着病例对照研究方法的不断完善，其应用范围也在不断扩大。

一、概述

（一）概念

病例对照研究是选择一组患研究疾病的人作为病例组，选择一组未患研究疾病的人作为对照组，回顾性调查这两组人在某个（些）因素的既往暴露情况，进而通过比较两组间暴露率或暴露水平的差异，判断该疾病与这个（些）因素的关系。因这种研究方法是比较病例组与对照组既往的暴露史，在时间上是“回顾性”的，故又称为回顾性研究（retrospective study）。

（二）特点

1. 按发病与否分成病例组与对照组，病例对照研究是在疾病（事件）发生后进行的，此时已有一批可供选择的病例。然后再选择一组无所研究疾病的人作为对照组。

2. 调查的暴露情况是由研究对象从现在对过去的回顾。研究开始之前，无论暴露还是结局事件已经存在或已经发生。

3. 由“果”推“因”。研究中是先有结果（结局事件），即已知研究对象患某病或不患某病，再追溯其可能与疾病有关的原因。

4. 病例对照研究受到回顾性观察方法的限制，不能按照由“因”到“果”时间顺序证实其因果关系（causation）。只能通过两组暴露率的比较来分析暴露与疾病发生发展是否有关联（association）。

二、设计模式

（一）病例对照研究

病例对照研究的基本原理见图4-5。若病例组某因素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有联系。病例对照研究可分为成组病例对照研究（group case-control study）和配对病例对照研究（matched case-control study）。

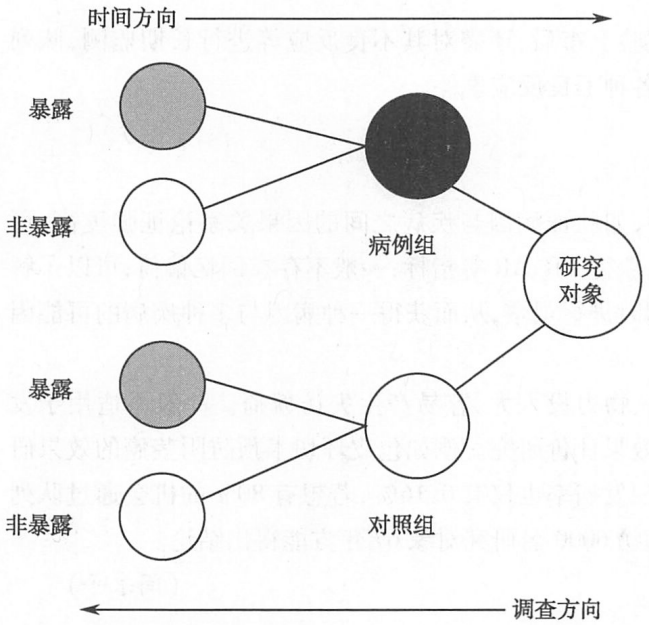


图4-5　病例对照研究设计基本原理示意图

1. 成组病例对照研究　在设计时对病例组和对照组人数无严格的配比关系，对照组人数可等于、多于或少于病例组人数。

2. 配对病例对照研究　要求对照组在某些因素或特性上与病例组保持相同，形成匹配关系，而且数量上也要是配比关系，如1∶1或1∶2等。

（二）巢式病例对照研究

病例对照研究与队列研究作为观察性研究的主要方法，已为广泛使用。由于两者各有特色，且优势与不足正好相互补充，因此在临床研究实践过程中产生了一扬长避短的新研究类型，称为巢式病例对照研究（nested case-control study）。其设计模式是：首先设计一项队列研究，收集基线资料，采集所研究生物学标志物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个或多个非病例作对照，再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的分析方法进行统计分析和因果推断。这种设计模式的主要优点是研究对象选择偏倚小，可较好地避免回忆偏倚，研究和统计检验效率高，论证强度明显强于传统的病例对照研究。

此外，近年来衍生的病例对照研究的其他设计类型还有：病例-队列研究（case-cohort study）、病例-病例研究（case-case study）、病例交叉研究（case crossover study）等，详见流行病学专著。

三、方案要素设计

基于提出的研究假设，需制定出一套科学合理的病例对照研究实施方案。

（一）研究对象的选择

研究对象选择的原则有二：一是具有代表性，即病例组应能代表目标人群中患该病的总体，对照组能代表目标人群中未患该病的总体；二是具有可比性，即病例组与对照组在年龄、性别、居住地、社会经济文化等人文特征方面应可比。

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”，否则会产生错误分类偏倚而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，探讨口服避孕药物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物者则不能选入；此外，应纳入新病例作为研究对象，以减少回忆偏倚等。

笔记

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”，必须确实排除患有所研究的疾病，否则，也会出现错误分类；其次，“对照组”的研究对象也应具有暴露于被研究因素之可能，并应与“病例组”同源（医院、社区等）。

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）配对法：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）。通过配对，可使病例组与对照组有可比性，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。应注意研究因素不能作为配比因素，同时配比因素不应过多，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。

在临床研究中，可设多组对照，如既选医院的患者，又选亲属或邻居作为对照。这不仅扩大了对照的来源，减少偏倚，增强代表性；同时还可研究疾病与暴露因素在不同水平之间的关系或发现另外一些病因线索。

（二）样本量的估计

病例对照研究的样本量估算需要预先掌握4种参数，包括病例组和对照组各自研究对象对被研究因素的暴露率；预期优势比（OR）；容许误差的α值和β值。病例组和对照组样本含量相等时统计效率最高。估计样本含量的公式如下：

$$n=\frac{\left[Z_{\alpha}\sqrt{2\overline{P}(1-\overline{P})}+Z_{\beta}\sqrt{P_1(1-P_1)+P_0(1-P_0)}\right]^2}{(P_1-P_0)^2} \qquad \text{式(4-15)}$$

或简便公式：

$$n=\frac{(Z_{\alpha}+Z_{\beta})^2\times 2\overline{P}(1-\overline{P})}{(P_1-P_0)^2} \qquad \text{式(4-16)}$$

N：样本含量；P_0：对照组暴露率；P_1：病例组暴露率；$\overline{P}=(P_1+P_0)/2$；Z_{α}和Z_{β}可根据标准正态差简表查出；P_1可根据下列公式推算：

$$P_1=\frac{\mathrm{OR}P_0}{1+P_0(\mathrm{OR}-1)} \qquad \text{式(4-17)}$$

（三）资料的收集

无论设计多么严谨，资料收集过程中若方法不恰当，将会产生无法纠正的系统误差，因而质量控制十分重要。

1. 资料来源　主要来源于设计良好的调查问卷，如果医院病历、疾病登记报告等能够满足研究所需，也可从中摘录，对调查表进行补充。

2. 收集方法　主要通过询问调查、查阅病历等方法收集资料，最常用的是访谈、信访及电话调查等，还可通过查阅资料来收集。

3. 调查表设计原则　调查表的设计绝非易事，一张好的调查表的设计需要临床医学、流行病学、统计学、心理学和社会学的专家共同讨论拟定，并经反复修订和预调查后，最终形成后才可以用于正

式调查。调查表设计的基本原则包括：①调查的项目要全而精；②每个项目都要有明确的定义；③所调查的每个因素要有量化标准；④项目中的问题要易懂，尽量口语化等。

四、资料的整理与分析

对原始资料进行整理、核查和计算机录入，在必要的数据转换后，首先要检验病例组与对照组是否具有均衡性，即在研究因素以外的其他主要特征方面有否可比性。计数资料常用χ^2检验，计量资料常用t检验。如果组间比较无显著性差异，提示两组的可比性较好。然后对所研究的暴露因素进行逐项整理统计，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。

（一）成组病例对照研究资料的结果分析

如果设计、资料的收集时按成组比较的方法进行的，结果分析时则要按照成组比较法的要求进行。成组法资料整理的四格表见表4-5。

表4-5　成组病例对照研究结果分析用四格表

暴露	病例组	对照组	合计
有	a	b	$a+b$
无	c	d	$c+d$
合计	$a+c$	$b+d$	$a+b+c+d=N$

1. χ^2检验　分析暴露与疾病之间是否有统计学关联。

$$\chi^2=\frac{(ad-bc)^2N}{(a+b)(c+d)(a+c)(b+d)} \qquad 式(4\text{-}18)$$

2. 比值比（odds ratio，OR）　病例对照研究计算的是两组暴露率之间的比值，也称为优势比。OR值的计算公式为：$OR=ad/bc$

OR值表示暴露者患某种疾病的危险性较无暴露者高的程度。当OR值大于1时，说明暴露因素与该疾病呈正相关，疾病的危险度增加，OR值越大，危险性越大；OR值小于1时，说明研究因素与该疾病呈负相关，疾病的危险度减少，OR值越小，保护作用越强；当OR值等于1或接近于1时，说明暴露因素与患病之间无联系。

（二）配对病例对照研究资料的结果分析

成组资料中的数字表示的是病例组和对照组的人数，而配对资料中的数字则表示的是对子数。对于1∶1的配对资料来说，表格中的数字表示1个病例和1个对照，若为1∶2的配对资料，则表示1个病例和2个对照，在分析资料时，以对子数为基础，不拆开进行分析，最为常用的1∶1配对的病例对照研究资料整理如表4-6。

表4-6　配对病例对照研究结果分析用四格表

对照	病例		合计
	有暴露史	无暴露史	
有暴露史	a	b	$a+b$
无暴露史	c	d	$c+d$
合计	$a+c$	$b+d$	$a+b+c+d=N$

χ^2检验和OR值计算公式如下：

1. χ^2检验　$$\chi^2=\frac{(|b-c|-1)^2}{b+c} \qquad 式(4\text{-}19)$$

2. OR 值　$OR=\frac{c}{b}$　式(4-20)

1∶2、1∶3和1∶4的资料整理表和 χ^2 检验、OR 值计算公式与1∶1配对的病例对照研究不同，参见相关书籍。

五、应用范围

病例对照研究主要用于疾病危险因素的探索，但也可用于临床筛检、治疗效果评价等的研究。

1. 探索病因和危险因素　从20世纪中叶开始已有大量针对病因的病例对照研究。如吸烟与肺癌的关系、孕早期服用沙利度胺(反应停)与婴儿短肢畸形等很多经典的病例对照研究，都为相关疾病的防治起到了决定性的作用。

2. 评价筛检试验效果　20世纪80年代以来，大量病例对照研究用于临床，如评价宫颈涂片检查宫颈癌、乳房照片检查乳腺癌等的效果。

3. 评价干预和治疗效果　病例对照研究对于发生率很低的某些疾病或目标事件很适用，此时很难开展随机对照试验(RCT)。

4. 研究药物的不良反应　当高度怀疑某种药物可能存在某些不良反应时，病例对照研究方法切实可行，此时 RCT 等试验性方法会因伦理问题无法实施。

六、优缺点

1. 优点　所需样本量小，研究对象易找，工作量小，人力物力也较少，因此易于进行，出结果快，可以对一种疾病的多种病因进行探讨。往往是罕见病、罕发事件病因研究的唯一设计模型。

2. 缺点　主要是容易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，论证强度不高；另外，病例对照研究无法计算发病率，只能推算出优势比(比值比)。

(闫永平)

第六节　横断面研究

横断面研究(cross-sectional study)常归类于描述性研究设计，但其研究对象的选择、影响因素的调查及其结果分析较其他描述性研究(如病例报告等)更为严密和规范；而且通过对患者与非患者的分布特征和影响因素分析还可进行一定程度的比较性研究。因此，横断面研究常是进一步开展病例对照研究与队列研究的基础，尤其是在疾病的患病率调查和人群、地区等分布特征研究中应用最为广泛。如：2010年开展的全国第五次结核病现况调查、2012年启动的全国精神障碍流行病学调查等，这些研究为我国相关疾病的防治和卫生决策制订提供了第一手资料。

一、概述

(一) 概念

横断面研究是指某一时点(或期间)内对某一特定人群中的疾病(或事件)患病(或发生)状况及其影响因素(暴露)进行的调查分析。由于在短的时间内，如一天、一周或一个月，且调查的是患病频率，因此又称现况研究，或现患率研究(prevalence study)。

(二) 特点

1. 在设计阶段一般不设立对照组，往往是对全部研究对象进行的调查，资料处理时，可根据患病与否或暴露状态进行分组比较。

2. 调查的时间是某一特定时间点或时间段，对于患病率来讲，这个时间应该是越集中越准确。

3. 在确定因果联系时受到限制，大多仅能提供病因线索。

NOTE

4. 在同一人群定期重复开展横断面调查可获得发病率、新发感染率、转归等资料。

二、设计模式

横断面研究主要是通过普查或抽样方式，研究目标人群疾病的患病率及其暴露状况，设计的基本原理如图 4-6 所示。

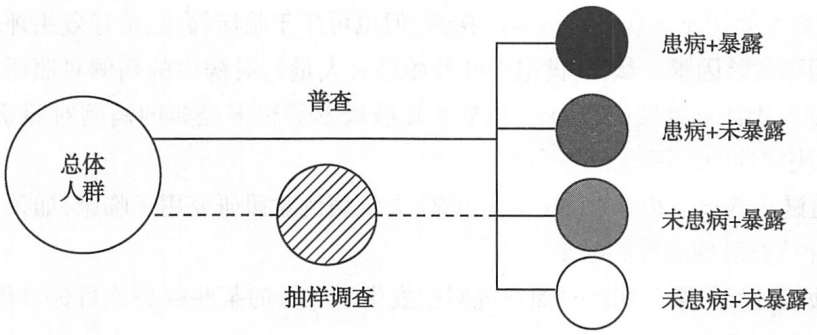


图 4-6　横断面调查设计原理示意图

横断面研究按研究对象选择的不同方式分成两种类型：普查与抽样调查。

1. 普查　普查（census）是指在特定时间对特定范围内的全体成员（总体）进行的全面调查。普查的目的主要是：①可早期发现、早期诊断和早期治疗患者，如很多地区开展的妇女宫颈癌普查；②可了解总体健康状况或某种疾病的患病率；③可了解人体各类生理生化指标的正常值范围，如各国开展的儿童身高、体重等发育和营养状况的普查。该种调查方式代表性较好，但相对费时费力，质量控制较难。

2. 抽样调查　抽样调查（sampling survey）简称抽查，是指在特定时间对特定范围内的全体成员（总体）中抽取的具有代表性的部分人群进行的调查。

基本原则是随机抽样（random sampling）。在实际工作中常用的随机抽样方法有下列几种。

（1）单纯随机抽样（simple random sampling）：又称简单随机抽样，是指从总体中抽取了若干个体，构成一个样本。抽样过程中不附加任何限定条件，在抽样前未进行分层或其他方式处理，保证总体中的每个个体被抽到的概率相等。抽样的基本方法是使用随机数字表法或计算机显示出的随机数字方法。单纯随机抽样是比较简单而可靠的随机化方法，但在总体与样本都比较大时，编制名单、号码、卡片或抽样都会耗费大量的人力物力。入选的个体如果散在分布于一个比较大的地区内，调查工作就会十分困难。因此调查中很少单独使用简单随机抽样的方法，往往将其作为其他抽样方法的基础。

（2）整群抽样（clustering sampling）：是以应调查对象的群组为单位而进行的随机抽样，对被抽中单位内的每个成员都进行调查。这种抽样和调查都比较方便，是在较大范围内的人员中进行横断面研究时经常采取的抽样方法。其优点是便于组织、节约人力和物力、抽样和调查都比较方便，在实际工作中容易被群众接受，适合于大规模的调查。缺点是抽样误差较大，分析工作量也较大。

（3）分层抽样（stratified sampling）：为了保证调查对象的同质性，可以按主要影响因素如年龄、性别和职业等先分成若干层，大型调查则可按行政单位或地区分层。各层内再作简单随机抽样。该方法的优点是分层后，层内各单元的特征比较齐同，变异较小，从而可以提高总体估计的精确度，减少样本量。

（4）系统抽样（systematic sampling）：系统抽样是指按照一定的顺序，机械地每隔若干单位抽取一个单位的抽样方法，又称机械抽样。优点是简便易行，样本的观察单位在总体中分布均匀，抽样的代表性较好，抽样误差与单纯随机抽样的误差比较相似或略小一些。缺点是如果总体各单元的排列顺序有周期性，抽取的样本就可能有偏倚。

（5）多阶段抽样（multi-stage sampling）：是指从总体中先抽取范围较大的单元，一般称为一级抽

笔记

样单元(例如省和市等),再从抽中的一级单元中抽取范围较小的二级单元(如县和乡等)……以此类推,最后抽取其中范围更小的单元(如社区、行政村)作为调查对象。

在实际的调查工作中,普查与抽查往往是相对的,如整群抽样时,在被抽中的基层单位内实际上是进行了普查,但总体上却是抽查。在规模较大的横断面调查中,各种抽样方法常常结合运用。

三、方案要素设计

横断面研究的规模一般都比较大,涉及的调查人员和调查对象也较多,因此有一个良好的设计实施方案是保证研究成功的前提。

(一) 研究对象的选择

要根据研究目的对调查对象的特点范围进行明确的规定,同时还要结合实际情况考虑在目标人群中开展普查或抽样调查的可行性。在设计时可将研究对象规定为某个社区或乡镇内的全体居民或其中的一部分,如儿童,即选择该社区内的≤14 岁者。

(二) 样本量的估计

样本大小与抽样的方法有关。简单随机抽样的样本量的大小与预期患病率和调查的精确度要求有关。预期的患病率(P)或感染率高,样本量则可以小一些;要求的精确度高,则样本量大。精确度又与容许误差(d)和显著性水平(α)有关。其样本量的大小可以用下列公式进行计算。

$$n=K\cdot Q/P \qquad \text{式(4-21)}$$

式中 n 为样本大小,P 为预期患病率,$Q=1-P$,K 为系数。在 $\alpha=0.05$ 水平上,当容许误差 d 为 $0.1P$ 时,$K=400$;当容许误差为 $0.15P$ 时,$K=178$;当容许误差的为 $0.20P$ 时,$K=100$。

(三) 资料的收集

所收集的资料信息因研究目的的不同而有所差异,通常主要包括:个人基本情况(如年龄、性别、文化程度、籍贯、职业及所在单位等)、人口学资料、生活习惯、环境资料等。相关资料一般可从临床和实验室检查、调查询问和常规资料记录中获得。

四、资料的整理与分析

在进行资料的整理和分析时,要根据不同的研究目的,选用不同的指标。除暴露因素外,对疾病或事件的频率描述,常见的有:患病率、抗体阳性率、抗原或病毒携带率、疾病诊治率、疫苗接种率、伤残率等。

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。

不同地区间患病率进行比较时,常采用率的标化方法。而相关影响因素的分析可采用单因素和多因素统计分析模型来完成。

五、应用范围

1. 描述疾病的分布　通过计算和比较所获得的患病率与感染率等指标,描述目标人群疾病的年龄、性别、种族、地区等的分布状况,从而为制订防治对策提供参考依据。

2. 提出病因线索　可以从疾病与研究对象特征以及环境因素的联系中研究疾病分布的不同影响因素,从而提出病因线索。

3. 确定高危人群　通过筛查,可以检出患者、可疑患者和病原携带者,从而确定高危人群,达到早发现、早诊断和早治疗的目的,提高疾病诊治效果。

4. 评价防治效果 在疾病监测、预防接种的实施过程中，通过不同阶段开展重复的横断面调查，比较患病率、感染率等相关指标的差异，评价防治方案的确切效果。

5. 医疗卫生服务的需求与质量评价 通过横断面调查可以评价社区卫生服务的近期和远期效果。近期效果如体检率、家庭和个人健康档案建档率、疫苗接种率、居民卫生知识知晓率等。远期效果如发病率下降、死亡率下降、人均寿命延长和医疗费用下降等。

六、优缺点

1. 优点 容易实施，科学性较强，研究对象代表性好，一次研究可观察多种疾病（事件）的患病状况及多种相关的可能影响因素。

2. 缺点 一次横断面调查无法获得发病率或死亡率，难以确定暴露与疾病之间的因果关系，尤其是在开展大规模调查时，需投入很多人力物力。

（闫永平）

第七节 病例分析

病例分析（case analysis）是临床医生最为熟悉和常用的一种研究方法。可发现临床试验中容易忽略的问题，产生新的研究思路，更好地服务于患者。同时也反映研究者临床经验积累的水平以及对当前学科发展和相关领域研究结果的关注度。

一、概述

（一）概念

病例分析是对现有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、分析、总结的一类研究。主要包括个案病例报告和系列病例分析等。它是临床医生日常工作中最为广泛应用和最容易掌握的一种描述性研究方法。

（二）特点

1. 以观察法为主要研究手段，主要通过观察、收集和分析相关病例数据，归纳和总结研究对象的特征。

2. 由于不设立对照组，仅能提供因素与事件之间的因果联系线索，但可为后续分析性研究打下基础。

二、设计模式

（一）个案病例报告

个案病例报告（case report）是对单个或少数几个病例的个人基本信息、临床和流行病学特征的描述、分析和总结。尤其在罕见和新发疾病的报道中最为有用，常常为新出现的疾病或药物副作用等提供第一手资料。

（二）系列病例分析

系列病例分析（case series）是对一系列或一组病例的人口学特征、临床和流行病学特征的描述、分析和总结。在设计时一般采用以下两种模式。

1. 连续系列病例分析（consecutive case series） 是将研究期间内所有符合条件的病例均纳入分析。

2. 非连续系列病例分析（nonconsecutive case series） 是将研究期间内部分符合条件的病例纳入分析。一般采用随机抽样、选择这部分病例，以代表总体病例。

笔记

三、方案要素设计

实施病例分析时，首先是要根据临床观察和资料报道提出拟分析的问题，进一步查阅、了解与该问题相关的文献，阅读相关的病历记录，制订调查和随访内容，然后确定病例范围和研究期限，明确纳入和排除标准，以保证结论的可靠性。

（一）研究对象的选择

一系列或一组病例的选择往往要求是具有同样的治疗方案，以便分析治疗效果和药物副作用等；或选择同样的疾病诊断病例分析不同治疗效果和影响因素等。

（二）样本量的估计

原则上样本量是越多越有说服力。但要注意的是：倘若样本量过大时，要控制好系统误差。

（三）资料的收集

需要收集详细的人口学特征（如年龄、性别、种族等）、主诉、症状、体征、体格和实验室检查、既往史、家族史、诊断、治疗、药物反应和随访等重要信息。除病历和随访记录外，还可设计简要的调查表收集资料，对记录不清楚、项目不完整的进行补充调查。

四、资料的整理与分析

一方面，可以观察干预措施对疾病的治疗效果和药物副作用等。如接受化疗的恶性肿瘤患者，观察其化疗的不良反应或并发症的发生率、发生时间、严重程度和恢复时间等。另一方面，还可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中评价降压药的作用和效果等。另外，要根据具体病例分析的研究目的进行相应的分析和总结，重点是分析研究中的新发现和对临床实践的应用价值。

结果分析中，计数资料常采用率、比表示，计量资料则用均数±标准差表示。并以95%可信区间来表示结果的可信度。

五、应用范围

病例分析在临床上应用较为广泛，几乎可以应用到临床的各个方面。如描述罕见或新发现疾病的临床特征、诊治方法和预后，描述新的手术方式和医疗革新，描述危及患者生命、少见的药物副作用，报告医疗事故、差错和经验教训，总结临床治疗和护理经验等。

病例报告是医学文献中最小的可发表单位。个案病例报告主要描述一种新的疾病或综合征。有部分疾病就是通过病例报告而得以认识和命名，如多发性骨髓瘤的发现。当一定数量的个案汇集到一篇文章中，就被称为系列病例分析。这可有利于发现疾病流行的端倪，了解药物的不良反应，如1981年洛杉矶发现多个艾滋病患者。个案病例可以简单地被考虑为偶然发生的事件，但系列病例发生时就要归因于某些潜在的可能原因了。如乳房硅胶植入与系统性红斑狼疮的可能联系。但这只能是线索，如果不做进一步检验和验证，往往会出现误判。

2003年10月，《中华结核和呼吸杂志》报道了世界首例严重急性呼吸综合征（SARS）的病例，为国内外SARS防治研究提供了宝贵的第一手资料。事实上针对一种未知原因疾病的研究，常常是以个案报道及系列病例分析为起点的，病例分析仍是临床研究不可或缺的设计方案。

六、优缺点

1. **优点**　研究容易实施，节省人力物力，短期易出结果，常常是很多临床分析性研究和试验性研究的基础。在个别情况下，设对照组可能没有必要，系列病例报告就能很好说明问题。例如，调查出生1个月内婴儿死亡的原因时发现有很多婴幼儿是由于头部跌落在地板上所致，此时就没有必要设置对照组、开展验证性研究。再比如，使用链霉素治疗结核性脑膜炎，不治疗是致命的，那么存活率的

提高就是显著效果的体现。

2. 缺点 缺乏对照组，大多情况下研究结果的说服力不强。不能控制选择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。

（闫永平）

第八节 真实世界研究

传统临床研究以随机对照临床试验（randomized clinical trials，RCTs）为代表，尽管 RCT 仍是临床前评价新医疗产品有效性的“金标准”，但由于受试人群高度均一化，干预措施过于简单单一，严重脱离真实的临床实践场景，因此其结果的外推性容易受到质疑，无法进行临床效果分析，更无法对近些年所倡导的个体化医疗提供指导。另外，传统 RCT 寻找合适的受试者入组变得愈发困难，因此仅依靠传统 RCT 来评价医疗产品的有效性和安全性，愈发难以满足现实需求。近些年，大型 RCTs 的费用逐渐上升，但获得的临床证据却并未相应增加。有数据显示：RCT 研究所获得的证据中，仅有 14% 在平均 17 年后广泛应用于临床。故从效率和成本角度分析，RCT 亦有其明显不足。

在此背景下，真实世界证据（real-world evidence，RWE）因其根植于真实的临床实践，具有外部效度高、无需严格设定、研究易于开展等优势，而愈发受到研究者及医疗产品开发商们的重视。此外，随着个人电子设备、应用软件的普及，以及各类电子数据库的建立，研究者可获取的与健康相关的真实世界数据正在呈指数增长，为真实世界研究的可行性提供了保障。

一、概述

（一）真实世界证据的定义

RWE 是指在现实医疗环境中，通过对来源于多种途径的数据优化分析获得的证据。其数据来源包括电子健康档案、理赔单和账单、药品和疾病登记表、个人健康设备所收集的信息等。

所谓的“真实世界研究”常被误解为不采用干预性试验和随机化试验设计，即在研究方法上与传统临床研究相区分。为此，美国食品药品监督管理局于 2016 年 12 月在 *The New England Journal of Medicine* 上发文，纠正这一错误观念，指出：RWE 与其他证据的本质区别不在于研究方法和试验设计，而在于获取数据的环境，即真实世界研究的数据来源于医疗机构、家庭和社区，而非存在诸多严格限制的科研场所。实际上，真实世界研究（real-world studies，RWS）不仅可以是观察性研究，还可以是干预性研究，甚至是采用类似 RCT 设计的随机对照研究。

（二）真实世界研究的特点

相比于 RCT，RWS 以真实临床场景为研究环境，不做过多人为限定，因此更适用于在整体人群中评价新疗法上市后的实际效果和安全性。由于所有干预可以在受试者知情选择下进行，且不存在接受无效治疗的情况，因此 RWS 容易满足伦理学要求，样本量和研究持续时间不受无伤害原则制约。通过对多种来源的电子信息进行收集和整合，RWS 的样本量可以非常巨大，因而具有更高的统计效能且更易发现罕见不良反应。RWS 纳入的研究对象异质性高，可进行亚组分析，建立特定人群的风险-获益模型。因研究的持续时间不受限制，故 RWS 可对具有更广泛临床意义的结局指标进行评价，如治愈率、病死率、复发率、伤残情况等，进而更好地反映新疗法的远期效应和安全性。

RWS 除以上优势外，亦存在诸多问题。首先，电子健康档案、疾病登记表等数据来源的建立并非以研究为目的，故其所提供数据的准确性、可靠性以及与所研究问题的相关性难以保证。另外，各数据库在收集和记录数据时标准可能存在差异，如何将这些数据标准化，也是需要研究者解决的难题。同时 RWS 立足于真实世界，对可能干扰结果的混杂因素未采取人为控制，因此选择偏倚、实施偏倚、检出偏倚和失访偏倚较多。由于许多数据来源于社交媒体，如不经过严格的知情同意，可能会存在暴露使用者隐私的风险。

RWS 及 RCT 的比较见表 4-7。RWE 和 RCT 证据的优势及不足总结于表 4-8。

笔记

表 4-7　真实世界研究与随机对照临床试验比较

	真实世界研究	随机对照临床试验
研究性质	效果研究，外部效度强	效力研究，内部效度强
研究时间	较长	较短
研究对象	无特殊限定，符合临床实践	一般纳入特定年龄段，无合并疾病者
样本量	大，尽量覆盖广泛的患者人群	研究前计算具有检验效能的最小样本量
干预	按照临床实际情况给予干预，可随机分组	随机分组后予以严格控制的干预措施，限制合并用药等干扰因素
盲法	可以不使用	常使用
结局测量	多种，临床意义明确	一种或几种，常为替代指标
药品审批	主要为上市后研究	主要为上市前研究
伦理	易满足	较难满足

表 4-8　真实世界证据及随机对照试验证据的优势和不足

	优　势	不　足
真实世界证据	①贴近临床实践，外推性好；②证据容易获得，研究效率高；③结局指标临床意义明确，可反映医疗产品的实际效果及安全性；④可行亚组分析，建立各人群的风险-获益模型	①数据常不完整、准确性差，可能与所研究问题的相关性不足；②偏倚多；③数据收集和结局测量未标准化；④样本异质性可能会掩盖疗效；⑤涉及隐私问题
随机对照试验证据	①数据完整、准确，与所研究问题的相关性好；②偏倚少；③数据收集和结局测量有明确标准；④反映效力	①只适用于特定人群及临床环境，外推性差；②获取证据的效率低，成本高；③常使用替代指标，临床意义有待明确

二、真实世界证据的产生与应用

如前所述，RWE 可产生于采用类似 RCT 设计的干预性研究。只是在真实世界的干预性研究中，除了随机分组后给予各组符合临床实际情况的不同干预外，不采取其他限制措施，称为实效性随机对照临床试验（pragmatic clinical trial，PCT）。PCT 以临床中有待解决的实际问题为研究方向，对多种临床场景中存在广泛异质性的患者进行纳入分析。例如 2015 年启动的阿司匹林心血管获益研究。虽然阿司匹林已广泛用于冠心病二级预防，但其最佳剂量尚缺乏科学依据。基于这一重要临床问题，阿司匹林心血管获益研究利用 6 个临床数据网络，纳入 2 万例动脉粥样硬化性心血管病住院患者。而后将其随机分为两组，分别予阿司匹林 325mg 每日及 81mg 每日，观察两组心血管二级预防效果的差异。由于所选剂量均为当前临床常用剂量，因此虽为随机分组，但仍符合真实世界原则。由此可见，PCT 研究可将 RCT 中用来减少偏倚的随机化设计与 RWS 中更贴近临床实践的研究环境相结合，在保证一定内部效度的前提下，大大提升了研究结果的外部效度，可在指导临床实践、制定医疗政策等方面发挥重要作用。

需要强调的是，虽然近些年 PCT 越来越受到研究者、医疗产品开发商及医疗政策制定者们的推崇，但不能武断地认为 PCT 在方法学上全面优于传统 RCT。首先，PCT 为了追求外部效度，势必会牺牲部分内部效度。以药物疗效评价为例。PCT 为了满足“真实世界”原则，所有入组患者均接受与病情及当地医疗现状相适应的治疗，故多药联合或用药剂量不同的现象普遍存在。除此之外，受试者异质性高，研究场所差异性大等均构成研究的混杂因素，虽然通过随机对照设计大大降低了这些因素的干扰，但偏倚仍然存在。其次，在大样本人群中进行随机分组干预需要耗费大量人力和财力。因此 PCT 目前多用于研究具有重大临床意义的问题，如脑外伤后糖皮质激素使用相关的 CRASH（Corticosteroid Randomisation After Significant Head Injury）研究、青少年戒烟相关的 ASSIST（A Stop Smoking In Schools Trial）研究。另外，干预性 RWR 并非仅限于 PCT 研究，还包括非随机对照试验。

在更多情况下，RWE 来源于观察性研究。其突出优势体现在样本量大及来源广泛，因此更有利于发现临床上罕见的情况。特别是在医疗产品安全性评价方面，RWE 的出现很大程度上改变了以往

医疗产品上市后的监管模式，即由出现不良事件后再采取相应措施的被动管理，转变为有计划地、实时地主动监察。以2012年底英国为评价孕晚期接种百日咳疫苗安全性而开展的大型前瞻性队列研究为例。鉴于2011年和2012年英国因感染百日咳致死的婴儿较前明显增多，英国疫苗及免疫接种委员会建议孕妇于孕晚期接种百日咳疫苗。然而此前该疫苗并未批准用于孕妇，缺少孕期安全性评价的相关数据，故英国对选择接种疫苗的孕妇开展了真实世界研究，以实时评价疫苗的安全性。英国的临床实践研究数据库（CPRD）收纳超过650个初级卫生保健中心的数据，覆盖超过1250万人口。通过该数据库，研究者得以获取孕妇接种该疫苗的信息，而后与未接种疫苗的孕妇进行配对比较，发现孕晚期接种百日咳疫苗不会增加孕妇或新生儿的死亡率及产科并发症的发生率。这一结论消除了人们对孕期接种百日咳疫苗安全方面的顾虑，对该疫苗在英国乃至全世界孕妇中使用提供了有力的证据支持。

除此之外，观察性研究还可用于药品上市后的剂量调整。如狂犬病疫苗。美国疾病预防控制中心（CDC）最初推荐5针的暴露后预防接种方案。在2007年疫苗供应紧缺后，研究者们开始重新评估该疫苗的推荐接种次数。由于狂犬病患者在出现症状后几乎100%死亡，因此出于伦理学考虑，不可能开展相关的随机对照研究。但是通过对临床实践中已有数据进行分析，研究者们发现所有患者在接种第4针疫苗后体内都出现了足够水平的病毒中和抗体，且在接种第5针后，抗体水平并无明显上升。另外，现实生活中很多患者由于种种原因，仅接种了4针疫苗，但这些患者最终均未患病。基于以上发现，CDC最终将原有的“5针方案”调整为了目前推荐的“4针方案”。

在临床实践中，常存在超适应证用药的情况，对这些数据进行收集，可探究药物在尚未获批的疾病及人群中的疗效，进而为上市后药物扩大适应证提供可能。2016年12月美国国会通过了《21世纪治疗法案》，其中也明确指出RWE可用于扩展已获批药物的适应证。然而，观察性研究无法消除未知混杂因素的影响，且常采用回顾性数据收集策略，容易产生偏倚，故在判断有效性时应慎重。最理想的情况是RWE用于发现新适应证的线索，RCT用于进一步确证及最终审批。

虽然RWE主要为上市后研究，但有时也可为药物上市前的研发提供方向。通过对真实世界的观察性研究，人们可以了解疾病的发病率、患病率、疾病负担、并发症、诊治情况等，从而获知目前亟待解决的重要临床问题。RWE还可能提供一些发病机制方面的线索，进而发现潜在的治疗靶点。

综上，从医疗产品研发到临床应用的各个阶段均可有RWE的参与（图4-7）。除此之外，RWS还能

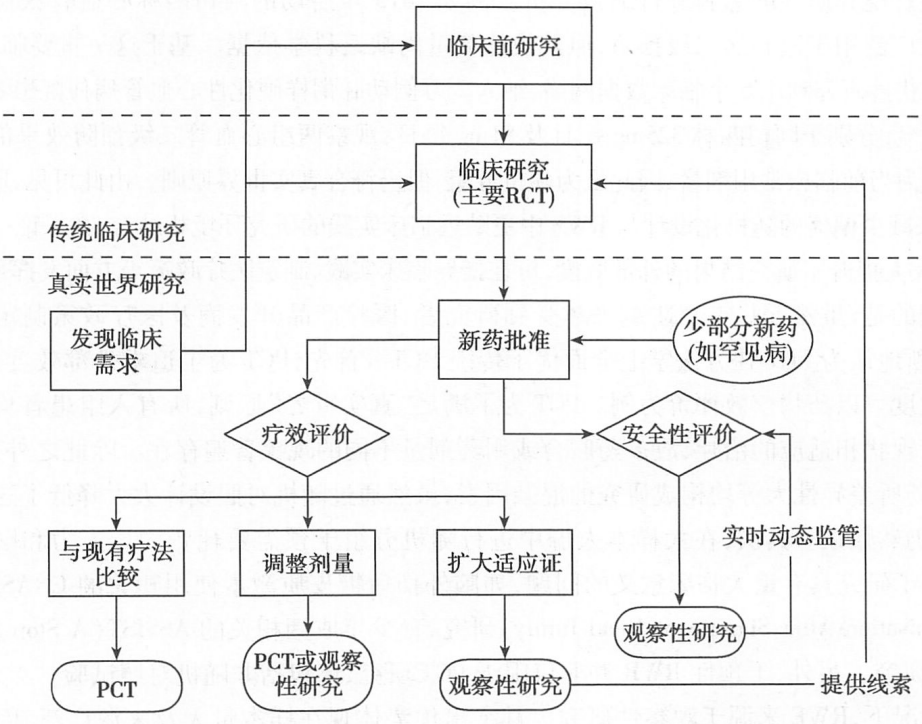


图4-7　真实世界研究在新药审批及药品上市后疗效及安全性评价中的应用

笔记

对医疗实践的过程进行评价，进而在提高患者就医体验、节约医疗花费等方面发挥作用。例如诺维本是一种治疗肺癌和乳腺癌的化疗药，有口服及静脉注射两种给药途径。临床研究者通过对接受不同给药途径的患者进行比较分析，发现口服给药可以大大缩短患者等待时间，提高化疗中心的接诊疗效。

三、真实世界证据的评价

很多人基于对“证据等级”的固有观念，认为 RWE 的可信度弱于 RCT 研究产生的证据。据此，英国制药工业协会（ABPI）在 2011 年明确提出 RWE 与 RCT 证据是相互补充的关系，两者在“证据等级”上处于平等的地位。实际上，研究问题决定研究设计。没有任何一种研究设计可以解决所有研究问题，因此也就不存在任何一种研究设计绝对优于其他设计。

虽然理论上 RWS 也可产生高质量的证据，但由于其数据来源广泛，容易出现数据质量差的问题，故在实际研究过程中，应着重对数据质量进行管理，以保证研究结果的真实性。主要包括以下两个方面：

（一）相关性

如何从浩繁庞杂的真实世界数据中选取与研究目的相契合的数据是开展 RWS 需要解决的一个关键问题。由于很多数据最初并非以研究为目的进行收集，且各中心数据的纳入标准不同，因此在统一命名的基础上，预先明确采集哪些数据，以及多次采集之间的时间间隔至关重要。

（二）可靠性

包括数据的准确性、完整性以及对偏倚的控制。如真实世界数据中很多信息是患者自填的，这部分数据是否可信，需要仔细评估。数据缺失或超出合理范围的情况也很常见，剔除这部分数据是否会影响数据对总体人群的代表性，进而产生偏倚，也应充分考量。很多时候，将研究数据与既往已被普遍认可的数据进行比较，可有助于我们评价数据的可靠性。

目前欧美国家普遍看好 RWE 的应用前景。2015 年英国医学科学院及制药工业协会联合会议上提出：希望在 2020 年，RWE 可以更广泛地应用于药品研发、审批、适应证调整等方面。美国更是在 2016 年底颁布法案，要求 FDA 将 RWE 纳入到现有的审批和监管程序中。相信在不久的将来还将出台相应的评价标准和指南，规范 RWS 的开展。

（刘晓清）

第九节　外科临床研究的一些特殊性问题

鉴于临床研究中的复杂性，以设计（design）、测量（measurement）、评价（evaluation）为三大核心的临床科研方法学（DME），已广泛应用于临床医学研究。但由于临床专业背景不同，可能会遇到一些特殊的问题，特别是临床治疗显然与药物为主要治疗手段的内科性质专业明显不同的外科专业。如何认识和解决这些专业上的特殊问题，消除临床研究中潜在的偏倚和混杂因素、确保外科临床研究结果的真实可靠，则是本节讨论的重点。

一、外科的特殊性容易导致的一些认识偏差

（一）重视手术技术但对手术指征缺乏严格的科学评价

外科强调应用医师个人的手术技术治疗疾病，医师之间或每个医师在成长至成熟的不同阶段，手术技巧和成熟程度是不同的，往往评价医师的技术水平是能做什么难度的手术，而不是临床的疗效。导致了外科医师特别对手术技术感兴趣，有些新技术、特别是手术技术在严格科学评估之前就被引入临床，而对于治疗的适应证缺乏科学的标准，造成手术技术的滥用。如为了达到实施腔镜胰十二指肠切除术的目的，而对于一些良性的胰头囊性疾病实施了此类手术，是典型的为了技术而手术。

（二）重视治疗过程而忽略了临床疗效的科学性评价

外科医师乐意展示自己的手术技术和技巧，并重视把这些技术用于临床的治疗过程中。结果是

做了多例高难度的手术，却没有重视用科学的方法对手术治疗的疗效作出客观评价，以至于这些手术技术难以推广或导致争议。这一情况可从发表的临床研究结果中见到，如描述性研究、回顾性研究在外科临床研究中比例过大，而高质量的前瞻性临床研究并不多见。

（三）重视手术安全但忽略手术安全性的科学评估

外科手术技术是用来治疗疾病，但手术也是有创的，可能导致创伤性损伤。外科医师特别重视规避手术创伤风险和手术并发症，但对规避风险的认识，多数是基于临床经验和积累，少有系统、科学的研究基础，鲜有对围手术期发生的病理变化所做的研究。

（四）重视手术治疗而忽略药物治疗

一些外科医师对药物药理作用缺乏深入、科学的认识，更愿意用手术解决各种临床遇到的问题，而忽略了药物或其他治疗的预防和治疗价值。外科医师实施的药物相关研究稀少，有些手术适应证长期并无大的变化，就是此类情况的一个反映。

二、外科临床研究的难点与解决方法

为了提高外科医师科学的临床研究意识，让一些疗效好、并发症少的手术技术得到推广和应用，除了掌握临床科研的设计、测量和评价总的原则外，根据外科临床特点，外科性临床研究还需要重视下面几个方面。

（一）干预措施（手术）的一致性

临床对照试验关键的一点是试验组所有受试对象接受的干预措施是相同的，同样，对照组所有的受试对象接受的干预措施组也应一致，即无论试验组还是对照组在组内接受的干预措施都应相同。但就外科性研究、特别是手术作为干预方式的疗效研究而言，情况可能复杂得多。要避免外科研究中常见的干扰和沾染偏倚，首先要解决外科性干预（手术）的一致性问题，即参与研究的外科医师之间无论完成的标准手术（对照组）或扩大手术（试验组）在组内都无明显的差异，这是实施外科性临床研究的前提。每位医师实施同一手术术式如何做到基本同质呢？即确保外科手术的一致性呢？

首先在确定手术术式后需要拟定手术规范，在规范中对手术的任何重要细节均应详尽描述，使参加试验的医师在手术实施中有证可查、有据可依；第二需要做试验前培训，对手术的每个细节都要详细讨论，做到参加试验的医生对手术的理解达到相对一致，同时要求术者均应在试验前达到试验对手术标准的要求；第三要通过手术观摩、录像等手段对拟参加研究的外科医师进行培训，并在现场或通过手术录像视频对手术者进行手术水平评估，以确定参与者是否达到技术要求，只有达到技术要求者方可参加试验；第四采用独立者对各参与试验的手术者进行一致性评估，要求参加试验的外科医师技术水平一致性良好（如 Kappa 值≥0.6），才可开始临床试验。一致性 Kappa 值的意义和评价方法可参考本书相关章节。

（二）外科研究中的盲法问题

盲法是减少测量性偏倚的技术手段，但涉及外科手术治疗的临床试验难以采用盲法，即多为开放式临床研究。主要原因是手术是一个高风险的治疗，新术式的采用必须获得患者及家属的知情同意，包括手术治疗的具体环节、可能的效果和并发症的发生等，否则有悖伦理和法律；另外试验的设计者和参与者往往也是干预措施（手术）的执行者和疗效的观察记录者。那么外科临床研究如何能避免非盲法带来的偏倚对临床试验结果的影响呢？首先试验的观察指标应采用客观数据，而不宜采用主观性指标。客观指标包括围手术期手术死亡和并发症发生率、术后生存时间、实验室数据（生理和生化）等，这些指标几乎不受主观因素的影响；第二进行病理检查和数据统计分析时采用盲法；对一些外科围手术期的重要的观察项目如引流液的性状和量、疼痛评分的量化等宜交由独立观察者完成。采取上述方法可以最大限度地减少非盲法带来的偏倚。

（三）统一的手术适应证与试验的可比性

外科临床常常将一些新技术与传统技术进行比较研究，如比较腔镜胰十二指肠切除术和开腹胰

笔记

十二指肠切除术治疗胰头肿瘤的效果是否相同或哪种技术治疗效果更好，虽然手术术式相同，治疗的疾病相同，但采用的技术手段不同。由于技术本身的限制，特别是技术尚未完全成熟阶段，较晚分期的患者采用腔镜技术实施胰十二指肠切除术可能存在较高的风险，医师更倾向于开腹手术。因此腔镜胰十二指肠切除术治疗胰头癌的适应证更为严格或保守，而开腹胰十二指肠切除术治疗胰头癌的适应证则相对宽松，因此，二者的手术适应证可能存在差异，若实施非随机对照研究，按各自手术适应证纳入的患者则存在疾病分期上的差异。这样的研究结果可能缺乏可比性。因此进行手术技术比较时，手术适应证作为纳入标准时应该严格统一，以保证所有进入试验的研究对象基本上是属于同一疾病阶段的患者，这样的临床研究中每组之间才具有可比性。否则腔镜组实施的治疗的胰头癌患者可能分期较早而开腹手术患者可能分期较晚，而得出错误的结论。在这些研究中，详细交代手术适应证是保证外科临床试验具有可比性和试验价值的重要内容。

（四）外科临床研究中的样本量估算

样本量是临床研究所需要的基本的研究对象（患者）的例数，基于这个样本量的研究结果基本可以代表研究对象总体的客观性质和效应特征，又不至于存在影响研究结果与总体实际情况误差过大的风险。样本量过小，研究结果可能难以反映总体的实际情况，样本量过大，不仅延长了研究时间，并造成人力资源浪费和成本增高。通常样本量根据具有产生有临床价值的效应值作为试验组与对照组的假设性差异水平值，以及容许的低水平概率错误水平（包括Ⅰ型错误和Ⅱ型错误），并依据设计的研究方案的性质，选择相应的样本含量估算公式进行计算。

现有条件下，某些恶性肿瘤临床治愈率仍然很低，甚至有的肿瘤治疗后5年生存率仍低于10%。例如胰腺癌由于特殊的生物学特性，长期生存率很低，总的5年生存率低于5%，治愈希望仍然是手术切除为主的综合治疗，若缺乏手术切除的基础，患者几乎丧失了长期生存的机会。在临床研究设计阶段发现，若进行一个前瞻性临床对照研究比较一个新技术手术是否比传统手术治疗胰腺癌能提高长期生存率，以华西医院近期统计的胰腺癌手术治疗的5年生存率8.8%为例，为了提高生存率，设计和实施一项联合血管和器官切除的扩大手术，期望5年生存率提高到15%。根据两样本率比较的估算公式计算的样本量 $n=(Z_{\alpha}+Z_{\beta})^2/[2(\sin^{-1}\sqrt{p_e}-\sin^{-1}\sqrt{p_c})^2]=(1.96+1.2816)^2/[2(\sin^{-1}\sqrt{0.15}-\sin^{-1}\sqrt{0.088})^2=567$，即试验组和对照组各需要567例胰腺癌患者。尽管胰腺癌死亡率很高，在北美和中国分别占肿瘤死亡率的第四和第六位，但其发病率并不高，短时间内募集到足够的样本量并非易事，整个研究周期可能很长，这对于临床研究来讲并不现实。这就提出了一个问题，这样的研究能不能做？如何做？目前临床研究解决这一问题最佳的方法是多中心协作研究，利用多中心优势在合理的研究时间内完成这一临床研究，得出科学的结果和结论。如果只是单个中心的临床研究，在研究期内未能募集足够样本量的研究对象，在研究期结束后，可能三个结果，一是这种扩大的手术的确大大提高了胰腺癌手术治疗后长期生存率并明显高于预期，并且与传统手术治疗后的生存率有统计学意义上差异。二是扩大的手术确能提高长期生存率，如达到15%的预期5年生存率，但由于样本量低于设计计算的样本量，并没有统计学意义的差异。三是与传统手术比较，扩大的手术治疗未能提高长期生存率，5年生存率与传统手术治疗的5年生存率基本上在一个水平上。第一和第三种结果都可以作出相应的有效和无效的结论。但如何看待第二种结果呢？从临床上来讲，胰腺癌是预后极差、治疗效果最不理想的肿瘤之一，扩大手术能将5年生存率由传统手术治疗的8.8%提高到15%左右，是非常重大的进步和具有重要的临床意义，但限制于样本量，差异并无统计学上的意义。对于这样的结果，应该根据疾病本身的特点作出科学的结论，正如上例中，对胰腺癌而言，生存率有6.2%的提高意义重大，很难用误差来解释，对此更应该看重其临床意义。

综上，外科性质的临床研究有一些特殊性，需要研究的设计者和实施者对这些特殊性在研究的设计、测量和评价时予以科学合理的处理，得出真实可靠、有价值的研究结果。

（刘续宝）

NOTE

第五章　医学研究证据的检索与收集

面对海量医学证据资源的不断涌现，繁忙的临床医生如何能够就自己感兴趣的问题（临床研究问题或循证问题），从中快速、精准获取当前可及的最佳证据，无论对开展临床研究还是进行循证临床实践，都是非常重要的。

第一节　证 据 资 源

一、资源分类

目的不同，资源分类也不同。根据临床问题，分为病因、诊断、治疗和预后证据；根据来源，分为一次证据和二次证据；根据用户需求，分为系统评价、临床实践指南、临床证据手册、卫生技术评估等；根据获取渠道，分为公开发表的文献、灰色文献、在研证据和网络信息等。

（一）文献资源

针对提出的临床研究问题，在临床研究设计阶段，需要进行大量的文献复习。根据文献来源不同，把来源文献分为一次文献，即原始研究文献；二次文献，即二次研究文献。

1. 原始研究文献源

（1）Medline 数据库：Medline 是生物医学领域最重要、最权威的数据库之一，成为全球生物医学科研人员及医务工作者不可或缺的文献信息资源。由美国国立医学图书馆（NLM）下属的国家生物技术信息中心（NCBI）研制开发，免费向全世界开放，Medline 因文献报道速度快、访问免费、使用方便、检索功能强大、查全率高、外部链接丰富、提供个性化服务等众多优点而备受青睐。

（2）Embase 数据库：Embase 文摘数据库是全球生命科学领域最重要的文献检索工具之一。是由荷兰 Elsevier 公司于 2003 年开发研制的一个生物医学和药物方面的文摘型网络数据库，覆盖了药理学、制药学、药剂学、药物经济学、基础生物医学、精神病学、传统医学、法医学、兽医学、口腔医学、护理学、卫生政策和管理等。

（3）中国生物医学文献服务系统：中国生物医学文献服务系统由中国医学科学院医学信息研究所开发研制，是对中国生物医学文献数据库（CBM）原有检索系统的全面继承和发展，为集检索、免费获取、个性化定题服务、全文传递服务于一体的综合性生物医学文献数据库。收录内容涵盖预防医学、基础医学、临床医学、药学等生物医学相关领域。

（4）中国期刊全文数据库：中国期刊全文数据库是中国学术期刊（光盘版）电子杂志社出版、清华同方知网（北京）技术有限公司发行，收录内容包括学术、技术、政策指导、高等科普及教育类文献。中国医院数字图书馆（China Hospital Knowledge Database，CHKD）是中国知识基础设施工程（China National Knowledge Infrastructure，简称 CNKI）的重要数据库之一，内容涉及基础医学、临床医学、中医学等医药卫生各个领域。

2. 二次研究文献源

（1）Cochrane 图书馆：Cochrane 图书馆由英国 Cochrane 中心委托 Wiley InterScience 公司出版，是一个提供高质量证据的数字化文献库。Cochrane 图书馆主要包括 Cochrane 系统评价全文库、Cochrane 有效评价文摘库、Cochrane 临床试验中心登记库、Cochrane 评价方法学文献库等数据库。其中，Cochrane 系统评价全文库收录了几乎覆盖医疗卫生所有领域的有关医疗卫生服务效果的系统评价

(systematic reviews,SR)。

(2) 最佳证据库:最佳证据(Best Evidence)库属于新型的、证据概要式杂志,是相关领域专家按照严格的评价标准,对全球已发表的最新系统评价和原始研究论文进行的评价和总结。收录了美国内科医师学会杂志俱乐部和循证医学杂志的文章。

(3) 循证医学评价库:循证医学评价库(Evidence Based Medicine Reviews,EBMR) 是由 Ovid 公司出版发行的、汇集临床实践和研究最佳证据的数据库,整合了 Cochrane 图书馆的 Cochrane 系统评价全文库、有效评价文摘库和最佳证据等三个数据库,并链接 Medline 和 Ovid 杂志全文,方便于读者同时获得二次文献和一次文献。

(4) 指南数据库:国立指南数据库(National Guideline Clearinghouse,NGC)和指南数据库(Guidelines)是循证临床实践指南数据库,收录了可信度较高的循证临床证据。国立指南数据库由美国卫生健康研究与质量机构(Agency Healthcare Research Quality,AHRQ)、美国医学会(American Medical Association,AMA)、美国卫生健康计划会(American Association Health Plans,AAHP)联合制作。指南数据库(Guidelines)由英国牛津医学科学研究院制作。

(二) 循证证据资源

开展循证临床实践时,对提出的循证问题同样需要在循证证据资源中先进行检索。按照"6S"证据分级标准,证据可分为计算机决策支持系统、证据综合、循证证据摘要、证据合成、研究摘要、原始研究类证据等。这里的"6S"证据是循证临床实践的现实版证据模型,考虑到临床医生工作繁忙、时间有限、工作条件和环境制约等诸多因素,将循证医学理念和临床实际状况相结合而形成的证据分类分级。开展循证临床实践时,原则上若从上一级数据库的证据检索已解决了提出的循证问题,则无需继续检索"6S"模型下一级数据库。

1. 计算机决策支持系统　计算机决策支持系统(systems)是循证医学证据资源的最高等级,是医学工作者将医学信息资源充分且准确地运用到临床实践工作过程而构建的完美模型。一个完美的计算机决策支持系统包含完整且精准的医学信息和证据链接。计算机决策支持系统勾勒了一个快捷方便、界面友好的人机对话系统。通过与计算机的简答互动,能够解答各种重要临床问题。同时,患者也可通过计算机决策支持系统,及时了解疾病诊疗情况,利于缓解因医疗信息不对称带来的医患矛盾。

2. 证据综合　证据综合(summaries)的主要形式是循证医学书库与循证临床实践指南。国外循证医学书库有:①临床证据(Clinical Evidence)是首推的循证医学书库,是临床医生必读的循证医学教科书,由英国医学杂志(*British Medical Journal*,*BMJ*)出版集团出版。临床证据中的医学信息以高质量的医学证据作为依据,并对每一个医学主题进行定期更新。②医师信息与教育资源库(the Physician's Information and Education Resource,PIER)是由美国医师学会(American College of Physicians,ACP)提供的在线循证医学书库,供临床医生方便查阅,明确提供了有关临床和初级保健证据及其级别。③UpToDate 以光盘和网络形式发行,每个季度更新,有广泛的参考信息,UpToDate 虽未对其纳入的文献进行质量评价,但收录了 Medline 摘要等关键信息。

3. 循证证据摘要　循证证据摘要(syntheses of synopses)是一种把原始研究和综述按固定格式整合提炼后所形成的摘要。临床医生通过循证题目检索和阅读相应的循证摘要,可方便快捷地获取相关循证问题的证据信息,即便在不清楚全文信息的情况下也能够及时采取相应措施,为患者提供最佳临床诊疗。

4. 证据合成　证据合成(syntheses)即系统评价,例如 Cochrane 系统评价等。若缺乏现成的计算机决策支持系统(system),查找到的摘要又不能提供更多信息时,可直接通过 Cochrane 图书馆查找系统评价。系统评价主要是针对当前疾病防治问题,通过广泛收集全球范围内已发表/未公开发表的、涉及卫生保健的干预措施效果文献,严格评价后再通过定量合成或定性分析得到的科学综述。

5. 研究摘要　研究摘要(synopses of studies)主要形式是循证摘要的杂志。倘若缺乏更好的途径找到所需的信息,则可在 Embase 中查找循证研究摘要。研究摘要是以高质量研究的结构摘要形式描

述，不仅符合严格的质量评级标准，并且以临床相关问题和有趣的形式汇编成册。

6. 原始研究 原始研究（studies）是指发表在各种杂志上的原始研究，可在检索 Medline、Embase 等原始文献网络数据库获取相关原始研究信息资源。

二、证据资源的入选标准

针对某一临床问题，要快捷、高效检索到相关信息，正确选择信息、资源或数据库十分重要。McKibbon 等针对如何选择信息资源给出了 4 条标准。

（一）循证方法的严谨性

在查找证据时，应精心构建临床问题、系统全面检索相关证据、严格评价文献的真实性与重要性以及适（实）用性，获取有代表性的高质量证据。循证医学信息资源在提出推荐意见时要充分应用已有的系统评价或自己制作系统评价，针对患者重要的结局指标，为不同干预措施的疗效和安全性提供最佳估计，并在充分考虑患者价值观和选择意愿基础上，采用恰当的分级系统对推荐意见进行分级。

（二）内容的全面性和特异性

理想的循证医学信息资源应该为临床实践中可能遇到的所有问题提供相关证据。其中，针对某一专业领域的信息资源可能会更有效地帮助查寻需要的证据。如要追踪专业的最新进展，可查寻某些循证摘要，如 Evidence based Cardiovascular Medicine，Evidence-based Mental Health 和 Evidence-based oncology 等。还有一些针对某一类临床问题的专门信息资源，如 Clinical Evidence 和 Cochrane 系统评价数据库，目前仅涉及治疗性临床问题，且后者仅包括基于临床对照试验的系统评价。

（三）易用性

某些信息资源的检索方便快捷，如美国内科医师学会杂志俱乐部（ACP Journal Club）为小型数据库，收集了内科领域约 140 种杂志上发表的与临床最相关的高质量研究的摘要，其优秀的检索引擎确保读者方便快捷查寻该领域的任何信息，包括疾病病因、诊断、治疗和预后等。

Medline 是检索全球生物医学文献资源的主要数据库之一，可借助隶属美国国立医学图书馆的国家生物技术中心开发的免费互联网平台 PubMed 进行检索。

（四）可及性

最可靠和最有效的信息资源往往需要付费，且价格比较昂贵。临床医生通常利用所在医院或医学院校图书馆免费检索网上资源，个人大多不会订阅昂贵的期刊或数据库。目前大多可及信息资源是免费的，如发展中国家的临床医生可通过学术机构的网络进入世界卫生组织的健康网络研究启动项目（Health InterNetwork Access to Research Initiative，HINARI）获取免费信息资源。除 Medline 外，对全球免费开放的检索资源还有：*Canadian Medical Association Journal*，*BMJ*，*JAMA* 和大多数 BioMed Central 的杂志。

三、证据分级及其发展历程

早在 1979 年由加拿大定期体检工作组（Canadian Task Force on the Periodic Health Examination，CTFPHE）首次提出研究证据分级标准并给出推荐意见，但应用范围有限。1986 年 David Sackett 提出的证据分级，虽简洁明了、操作性强，但证据研究类型局限。美国卫生保健政策研究所 1992 年提出的证据分级标准则将基于随机对照试验的系统评价作为最高级证据，但未考虑系统评价本身及其纳入临床试验的质量。2001 年 Scottish Intercollegiate Guidelines Network（SIGN）将系统评价及随机对照试验共同作为最高级证据。同年牛津大学循证医学中心将系统评价、随机对照试验、全或无系列病例分析共同作为Ⅰ级证据，并根据不同问题类型制作了不同的分级标准，分级也拓展到治疗以外的领域。2004 年由国际 GRADE 工作组制定出统一的证据质量分级和推荐强度标准，即 Grade 分级系统。此外，2004—2006 年中国循证医学中心把证据分为五级，并探索动物实验研究证据和循证决策领域证据分级。

Grade 证据分级系统明确定义了证据质量和推荐强度，清晰评价了不同治疗方案的重要结局，认

笔记

可观察性研究也是重要证据来源，对不同级别证据的升级和降级有明确的标准；从证据到推荐过程透明，综合考虑研究质量和研究结果的一致性和证据的直接性、利弊平衡、资源配置、结局指标的重要作用。Grade 把证据分为四个等级，先默认随机对照试验证据初始等级高，观察性研究证据初始等级低，再根据研究特征，对证据等级进行调整，在随机对照试验中找降低其质量的因素，在观察性研究中找增加其证据质量的因素，最后作出综合判断。

第二节　证据检索和文献搜集的基本步骤

一、确定临床问题类型和构建临床问题

临床问题分为病因、诊断、治疗和预后四种类型。针对每一问题，根据信息了解的程度又分为背景问题和前景问题。背景问题是关于某种疾病、某一诊断技术或某一干预措施的一般知识，如小儿肺炎的常用药物？一般通过查阅教科书或咨询上级医生就可获取答案；而前景问题涉及临床诊断、治疗的更深层次的具体问题，一般没有现成的答案。循证临床实践过程中提出的多数问题均为前景问题，如孟鲁司特与常规治疗相比能降低小儿哮喘的复发率吗？提出前景问题，需要明确问题涉及的 4 个要素：研究对象（patient）、干预措施（intervention）、对照措施（comparison intervention）和关心的结果（outcome）。按照 PICO 原则构建临床问题，有助于正确选择数据库资源、合理制定检索词和检索策略，提高证据检索的效率。如果是背景问题，可选择教材、专著和综述等，而在 Cochrane 图书馆就难以获得有关疾病病因、致病机制等方面的详细信息。反之像有关干预措施疗效的前景问题，Cochrane 图书馆则是主要信息来源。

运用 PICO 原则构建临床问题是将初始临床问题转变为可回答的临床问题（question），也是循证临床实践的第一步，有利于临床医生就某一临床问题进行证据检索。

临床问题的构建运用 PICO 原则，包括以下四个基本要素：①患者/群体（patient/population，P）患者或人群及其临床特征；②干预措施（intervention or exposure，I）关注的处理措施，在病因问题上是暴露因素，在诊断问题上是诊断试验，在治疗问题上多是药物、外科手术等；③对照措施（comparison，C）与干预措施或暴露因素相比较的措施，在治疗问题上，对照措施多是常规治疗、安慰剂，在诊断性问题上，对照措施常是诊断某种疾病的金标准；④临床结局（outcome，O）临床医生关注的暴露或处理措施导致的患者相关结局，多为远期指标，如：病死率、功能改善、不良反应发生率等。

【案例】 一患者 35 天前因跌倒外伤造成右手环指锤状畸形，X 光片检查可见Ⅰ区远端之间关节水平处有一小骨片。

原始问题提出：采用什么治疗矫正右手环指锤状畸形？

问题剖析：根据案例了解到治疗的对象是右手环指伸直肌断裂造成锤状畸形，治疗方案包括保守治疗，即夹板固定 6～8 周；手术治疗采用克氏针固定。通过查证得知Ⅰ区远端之间关节处指伸直肌断裂多采用夹板固定，但是患者和临床医生更多关注的是伤后采用夹板固定的时间和恢复环手指伸直功能有无关系？

问题构建：伤后 4 周后夹板固定与伤后 2 周内夹板固定相比，对右手手指伸直肌断裂造成锤状患者手指的伸直功能恢复情况有无差异（表 5-1）？

表 5-1　PICO 原则案例分析

PICO	信　息
患者（P）	右手手指伸直肌断裂造成锤状畸形的患者
干预措施（I）	伤后 2 周内夹板固定
对照措施（C）	伤后 4 周后夹板固定
结局（O）	手指伸直功能恢复情况

二、选择检索数据库

（一）证据检索时按照“6S”模型选择数据库

临床医师可以根据具体临床问题，按照“6S”模型依次检索计算机决策支持系统、证据综合、循证证据摘要、证据合成、研究摘要、原始研究类证据，从中获取最高级别证据资源，开展循证临床实践。

Cochrane 系统评价是基于高质量的临床随机对照试验文献的系统评价，属于证据等级金字塔中证据最高级别，因此，临床医生在循证临床实践时，应优先选择 Cochrane 图书馆检索 Cochrane 系统评价。

（二）文献复习时选择专业数据库

综合性文献数据库如 Medline、中国期刊全文数据库等虽然覆盖广，涵盖了理工、农业、医药卫生、文史哲、教育与社会科学综合等不同领域的文献，但有时却难以快速获得临床专业相关信息。因此，使用医学专业检索数据库更方便、快捷获取与医学专业直接相关的文献资料。但在缺乏专业数据库时，综合性文献数据库仍是常用的文献信息源。

三、制订检索策略

（一）确定检索词

根据 PICO 原则把循证问题分解为四个部分：人群 P、干预措施 I、对照措施 C、结局 O，进而确定相应的检索词。

（二）检索策略

检索策略的制定主要是运用逻辑运算符“AND”“OR”和“NOT”组合起来，不同数据库的检索策略稍有不同，但原理基本相通。在制订检索策略时，需要在检索实践过程中根据实际检索结果来扩大或缩减检索范围，从而实现查新、查全和查准。

1. 在扩大检索范围，提高查全率方面。当研究者检索实践中发现检索结果较少时，可扩大检索范围，采用方法包括选择主题词的上位词进行检索、多个主题词检索、用自由词检索、用“OR”运算符把多个检索词连接起来、运用截词符“*”和运用通配符“?”等。

2. 在缩小检索范围，提高查准率方面。当研究者检索实践中发现检索结果较多时，可缩小检索范围，研究者可运用下位主题词、应用限定字段检索和运用运算符如“AND”“WITH”“NEAR”“NOT”等对检索范围进行限定，从而缩小检索范围的目的。

四、评估检索结果

评估检索结果是否能回答患者提出的临床问题，应考虑检索策略制定是否合理、是否是当前可得到的最佳证据等。因此，严格评价文献检索流程以确定检索结果是否是当前最佳证据，确保为患者提供最佳的循证诊疗方案。即便是检索结果能回答患者提出的临床问题，也是当前可以得到的最佳证据，临床医生也不能单纯地依靠检索结果作出的诊疗决策，还需根据个体化原则，把检索出的证据和患者的实际情况以及患者意愿结合起来才能作出循证个体化决策。

第三节　证据检索实例

【案例】使用手机会不会增加患脑瘤的风险？

在现今社会中，很多人都在使用手机，由此引发了“长时间使用手机是否会增加得脑瘤的风险”的担忧，无论在门诊还是在病房都会被问起，经常使用手机会不会增加脑瘤的风险，看似是一个公共问题，实则也是一个医学问题。

一、提出并构建问题

根据 PICO 原则提出问题(表 5-2),使用手机会不会增加患脑瘤的风险?

表 5-2　PICO 原则构建临床问题

PICO	信息
患者(P)	可能暴露于手机的一般人群
干预措施(I)	使用手机
对照措施(C)	不使用手机
结局(O)	脑瘤的发生率

二、制订检索计划

证据等级由高到低依次为循证临床实践指南,基于病例对照研究的系统评价,单个病例对照研究,病例系列报告/个案报告,体外、体内试验/专家经验,因此检索根据证据等级高低来制订检索策略。见图 5-1 具体如下:

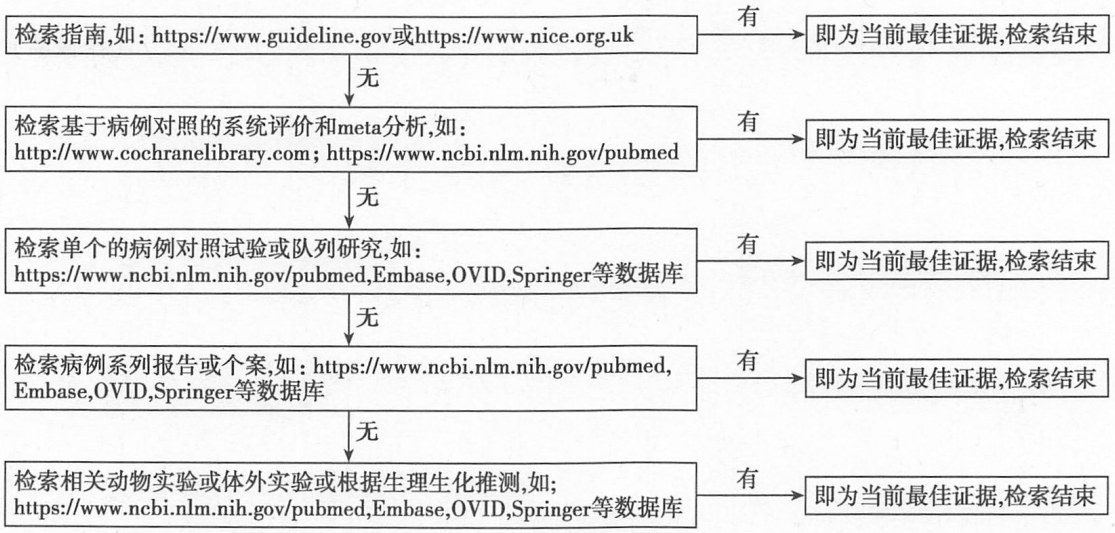


图 5-1　病因证据检索流程

三、检索

(一) 检索指南

首先检索“手机使用是否会增加得脑瘤的风险”的相关指南,依次在美国国家指南库、英国 NICE 指南库检索,结果未找到相应指南。

(二) 检索 Cochrane 图书馆

继续检索基于病例对照研究的系统评价,首先去 Cochrane Systematic Review Database 检索。在 Cochrane 图书馆也未检索到 Cochrane 系统评价和 meta-分析,继续检索 Medline(1966—2017.8)。

(三) 检索 Medline

本案例用主题词检索法,先找到“手机”的主题词为“cell phone”检索,然后找“脑瘤”的主题词为“brain neoplasm”检索,病例对照主题词为“Case-Control Studies”检索,“系统评价和 meta-分析”无主题词,只能将前三个主题词用逻辑与“and”连接再次检索,得到 91 篇文献,逐一阅读,找到题目为“MOBILE PHONE USE AND RISK FOR INTRACRANIAL TUMORS AND SALIVARY GLAND TUMORS-A

META-ANALYSIS”的基于病例对照系统评价和meta-分析，发表于2017年，即为当前最佳证据，本次检索结束（图5-2）。

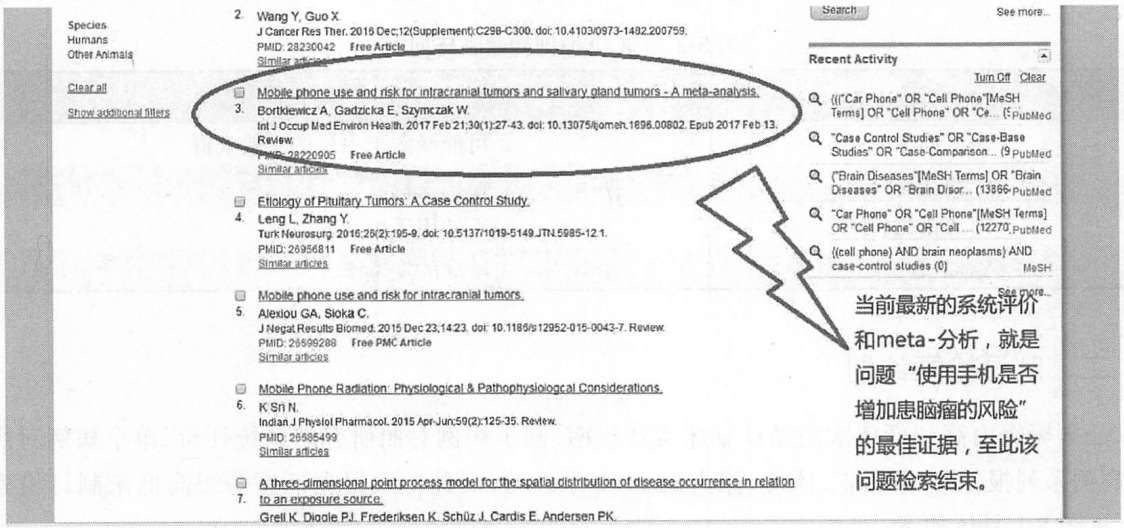


图5-2 PubMed检索

（徐佩茹）

笔记

第六章　医学研究文献的评阅与证据质量评价

掌握正确的医学文献阅读与评价方法，科学有效地管理好医学专业文献，将有助于临床实践、教学和科研，同时也是临床医生实现自我终身学习、保持与提高临床技能水平的必由之路。

随着循证医学的引入和推广，倡导使用最佳证据指导临床实践。如何获取证据，特别是高质量的证据，成为实践循证医学的关键，而有重要临床价值的、高质量医学文献则是证据的主要来源。

第一节　阅读与评价医学研究文献的重要性

一、医学研究文献的特殊性

（一）数量庞大

医学研究文献比较特殊，首先医学文献数量非常庞大且仍处于快速增长之中，已位居各学科文献量之首。同时医学文献还具有载体多、发表语种多的特点，以及重复发表、发表分散以及滞后发表等一系列问题。

（二）医学文献的分类体系复杂、质量参差不齐

文献是指以文字、图像、公式、视频与音频、代码等形式，将信息、知识记录或描述加以存储、传播的一切载体。医学文献就是记录有医学相关知识或信息等载体的总称。

1. 按照载体的属性可分为书写型、印刷型、电子型、微缩型和视听型等文献类型。其中书写型文献专指手工书写与抄写的文献，如病历、实验原始记录、设计草图等。印刷型文献特指纸制出版物，为图书馆收藏的主要类型，种类繁多，包括医药图书（专著、教科书、工具书）、生物医学期刊、学位论文、会议文献、研究报告或官方出版物（如卫生年鉴）、专利文献、医疗器械技术标准以及产品资料、医疗技术档案等。电子型文献是指医学信息以光盘、网络、软盘等形式存储和传播，目前已成为主要文献主体，且有逐步取代印刷型文献之势。

2. 按照证据质量分类　近年来，随着循证医学的引入和推广，如何获取证据，特别是高质量的证据，成为临床医生实践循证医学的关键。为此，加拿大临床流行病学与医学信息家 R. Brian Haynes 提出了 6S 证据模型。该模型是由最早的 4S 模型、5S 模型衍化而来。6S 类证据依次为系统类证据（systems，如计算机决策支持系统 CDSS）、指南类证据（summaries，如循证临床实践指南 CPG）、集成类证据（synopses of syntheses，如 DARE、提要类循证医学杂志系列等）、系统综述类证据（syntheses，如 Cochrane 系统评价/综述）、原始研究提要类证据（synopses of studies，如 handbook of clinical practice），原始研究级证据（studies，如 ACP JC+等）。这 6 类证据均由专业组织或机构分别从专业和方法学质量角度进行严格评价、后经筛选而来的证据，证据质量一般较高。需要注意的是，6S 类证据因入选标准较高、筛选制作要求较为苛刻，同时耗时耗力，生产数量比较少，临床实践中有关病因、诊断、治疗、预后问题，多数情况下并无现成的 6S 类证据可供参考，仍以传统的、未经严格评价的医学文献为主。

证据级别的分布一般呈现金字塔状，高级别证据是由低级别文献信息综合提炼加工而成，数量少，分布在塔尖；低质量文献或证据数量庞大，分布在塔底（图 6-1，左图）。例如，循证临床实践指南是指导疾病防治的高级别证据，综合萃取了系统评价及 RCT 等证据信息，在数量上要少于 RCT 等原

始研究文献。

二、阅读和评价医学文献的重要性

（一）医学文献与临床医疗实践

在日常的临床实践中，要求临床医生掌握临床诊治新技术与新方法，离不开对医学文献的阅读与评价。特别是现在倡导循证医学实践，要求高素质的临床医生结合医疗环境和患者的主观意愿、将最佳证据应用到具体的临床实践之中，从而创造一流的医疗水平、更好地为提高国民健康水平服务。最佳证据来源于高质量的诊断、防治、预后以及病因学/危险因素研究文献。另外，随着患者维权意识的提高，患者及家属也可能利用一些现成的检索资源和检索工具，查阅医学文献，希望掌握疾病诊治相关的知识与信息，并从中找出一些“证据”主动与医生探讨。鉴于医患关系紧张已是不争的事实，这也迫使临床医生及时掌握相关文献信息，以利于医患间的交流与沟通，创造和谐的医患关系。

（二）医学文献与临床科研

医学科研，选题立题是基础。当今社会是“知识爆炸”的时代，由于科技与信息科学的高度发展，信息量本已十分庞大，而新的海量信息又源源不断的涌现，这在医学相关领域表现得尤为突出。在此背景下，临床医生担负着临床医疗与临床科研的双重使命，临床研究要有所创新，需要进行大量的文献复习，以掌握最新的专业进展与研究动态，这也离不开对医学文献的评阅。

（三）医学文献与医学教育

在医学教育的不同阶段，对医学文献有着不同层次的需求。例如，处于培训阶段的临床医学生，对医学文献阅读、评价的需求不高，很多入门问题已由专人帮助解决；但过了该阶段，在独立进行临床实践活动中，所遇到一系列临床问题，这就要通过阅读和评价医学研究文献，大多由自己加以解决。为此，只有学会正确的阅读与评价医学文献，方能实现自我终身教育，使自己的知识水平永葆一流。

（四）医学文献与卫生决策

卫生政策的制定同样需要借助医学文献的阅读与评价，以从中发现决策所需的重要证据与依据。鉴于卫生服务资源的有限性与医疗卫生服务需求的无限性的矛盾将长期存在，“看病难、看病贵”问题日益突出，要实现卫生服务资源的最优分配以及医疗卫生服务的效率最大化，对卫生服务研究文献以及卫生经济学评价研究文献的评阅不可缺少。同时，在国家层面上形成的重大疾病攻关和支撑计划等决策，也是建立在大量的文献复习和调研的基础之上的。通过阅读医学文献，可以更好地确定重点疾病与研究重点，从而有针对性地制订招投标指南。此外，对于突发性公共卫生事件频发的现状，作为卫生政策决策部门，要提前制订处置预案，以防患于未然。而预案的规划与制订，需要学习和借鉴类似事件的处置经验和教训，同样也离不开大量的文献复习。

第二节 阅读医学文献的基本步骤

一、医学文献检索

文献检索分为手工检索和计算机检索。随着计算机技术和通讯技术的迅猛发展，计算机文献检索因其检索速度快、范围广、内容新、检索入口多、功能强、且用户使用方便，已成为医学生必须掌握的关键技能之一。

（一）医学文献来源与检索资源

医学文献具有来源广、种类多、分布广泛等特点，且有一定规律可循。例如，按照文献可及的数量以及分布密度顺序排列，一般呈现金字塔状分布（图 6-1）。其中电子文献检索数据库（电子文献索引或全文数据库），收录了包括指南、系统评价、临床原始研究以及其他计量研究等绝大多数医学研究文献。

尽管电子文献数据库是医学文献的主要来源，但仍存在一些不足。如大多只能收录摘要，未能收

录文献全文。其次可能存在发表性偏倚，阴性结果文献数量可能被严重低估。另外文献的收录也存在地域性发表偏倚。为弥补电子文献数据库的上述缺陷，常利用互联网（WWW）查阅一些灰色文献作为常规医学文献检索的重要补充。灰色文献是指由官方、NGO、学术机构等非营利单位出版发行的印刷品和电子出版物。形式多种多样，既可以是技术报告、研究报告，也可以是技术文件、技术说明与标准，以及一些非营利性的译文、著作、官方文件等。

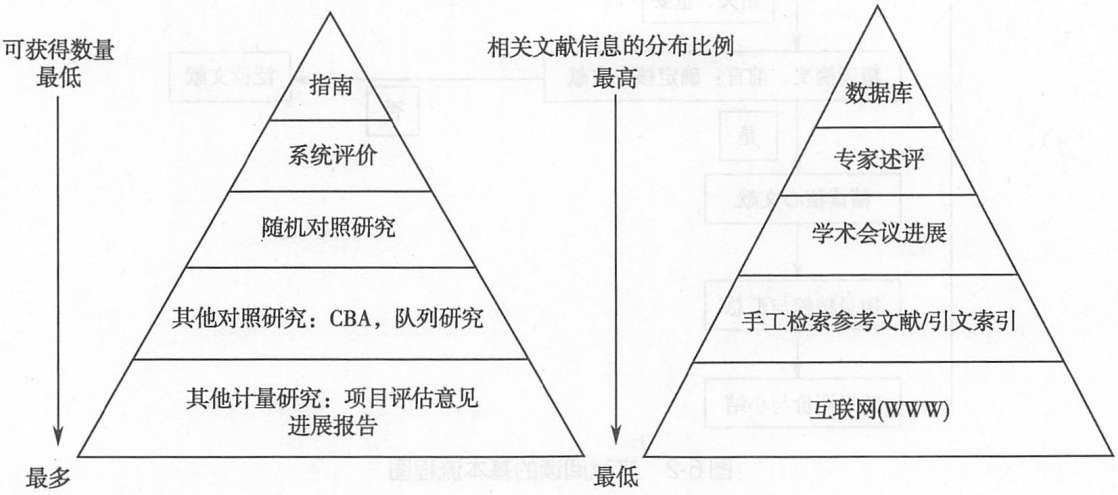


图 6-1　医学文献与信息资源分布金字塔图

（二）文献检索基本步骤

无论是人工检索还是计算机检索，都大致分为以下几大步骤：从分析检索题目入手，明确检索要求；进而选择检索工具，制定检索策略。如一般将那些文献类型全、数量大、时差短、途径多、著录标准的检索工具作为首选；最后选择检索途径，检索并获取原始文献。

1. 提出检索问题　围绕研究目的，将特定的研究问题进行规范化和标准化，形成结构化检索问题。

2. 制定检索策略　检索策略（search strategy）是指在解析相关问题的基础上，结合检索目的和信息需求，选择检索系统，确定检索词，构造检索式，从而制定出较为完善的检索计划或方案。

其中构造检索式需要使用检索系统规定或允许的符号（运算符），并用于连接已确定的检索词。检索系统中的运算符有位置运算符、逻辑运算符、限制运算符、截词符等。常使用的逻辑运算符有：①“AND”（逻辑“与”）：其作用为缩小检索范围，提高查准率；②“OR”（逻辑“或”）：其作用为扩大检索范围，提高查全率；③“NOT”（逻辑“非”）：其作用为缩小检索范围，提高查准率。

3. 选择检索数据库及实施检索　选择检索数据库，确定检索方法。一般将那些文献类型全、数量大、时差短、途径多、著录标准的检索数据库作为首选；同时为提高检索效率，在检索内容与顺序安排上，有一定的讲究。一般是先寻找可靠的三次文献，如指南、系统评价等，这些文献综合了大量相关的原始研究结果，且经过了加工和提炼。若无这样现成的文献，再寻找可靠的原始研究文献。

开始使用电子文献数据库时，最好寻求图书管理员的帮助，以尽快熟悉检索方法，提高检索效率。

二、阅读医学文献的基本步骤

医学研究文献具有两种属性，一是外部特征，比如题目、作者、作者单位、发表期刊、卷、期、页码、年限、语种等，这些信息由于具有唯一性和指向性的特点，常被用来标识文献，供文献的收集、整理、存储、传播与查询之用；二是内部特征。包括研究目的、科学假设、设计、研究方法、研究对象、主要结果、讨论和结论等方面内容，既是文献的核心，又是文献阅读的重点（图 6-2）。

（一）结合阅读文献的目的，筛选、甄别、获取相关医学研究文献

开始阅读之前，一定要弄清楚“为什么要阅读文献？”“要阅读哪些文献？”带着问题有针对性地筛

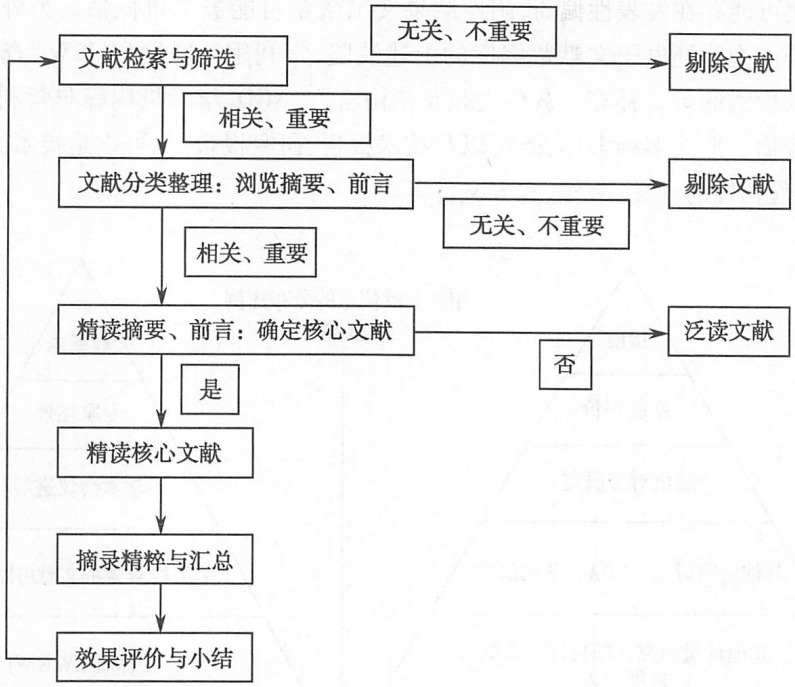


图 6-2　文献阅读的基本流程图

选、甄别、获取医学研究文献。筛选文献不能无的放矢，根据具体问题，一般先从阅读文献摘要入手，按照与阅读目的的关联性大小以及文献的时效性，依次安排文献阅读的先后顺序，从中遴选出最新、关联程度最高的文献，进而阅读全文。这一点对提高阅读效率尤为重要。鉴于候选文献中的摘要，无论是结构式摘要抑或非结构式摘要，提供的信息量有限，有时难以取舍，需要进一步获取全文，以确认是否最终纳入。

（二）熟悉文献的基本结构、选择阅读方式和阅读重点

对于纳入的文献，应首先熟悉其基本结构，确定阅读重点。因个人精力和时间有限，最好在正式阅读开始之前，对检索到的文献进行分类和整理，根据个人精力、时间以及与个人关注问题的密切程度，将纳入文献分为：精读文献和泛读文献。如，对于与个人关注问题密切相关的文献可采用选择性精读文献的方法；而对于与个人关注问题关系不密切的文献可以采用泛读的方法进行。在医学研究领域中，重要文献是指那些对学科发展具有里程碑意义的原始研究文献以及学术权威撰写的综述或述评等。对此类文献应设法获取全文并仔细阅读。其他文献以泛读为主，先阅读题目、摘要，在此过程中，若发现文献有价值可升格为重要文献，再精读全文。若同类文献较多时，考虑精读文献的顺序，新近且重要的文献，优先进行精读。

一篇完整的原始研究文献一般包括摘要、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。阅读目的不同，对同一篇文献的阅读重点有所侧重。如果想了解该文献的结论是否适合于自己的患者，可直接阅读“材料和方法”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。若为了解学科新进展，应重点阅读文献的“结果和结论”部分；若为启迪研究思路，则应以文献的研究方法为阅读重点，可取长补短、突破固有思维模式，发现创新点；若为临床科研的选题及立题提出依据，先阅读文献的“前言”部分，掌握该研究领域的历史与现状，同时在该文献“讨论”部分查找不足之处、方法学缺陷等，这些就是将来的研究方向与创新点。

（三）阅读文献、摘录文献精粹

对纳入的文献，即使一篇与自己研究目的高度相关的文献，也不可能全盘照搬，需要从中摘录出精粹部分，可以是文献中的某段话、某种方法，也可能是一幅图表等，宜精不宜多，过多内容会显得杂乱无章，对初学者无益。在摘录文献精粹的基础上，进一步加以系统总结。逐一将文献的精粹部分，加以汇总；为方便管理，需要创立一个文档文件，将相关内容作长久保存。同时，在汇总过程中，最好

笔记

能将文献的读后感，包括阅读后有哪些收获、新启发、新思路等提炼加工后，一并纳入。

一般的文献泛读可以在 5～15 分钟内完成，精读文献则需要 2 个小时到 1 周时间不等。相对文献泛读，精读文献时需要摘录的文献信息更为全面细致。特别是应弄清以下基本问题：

1. 研究目的是什么　通过阅读前言部分，明确该文献的科学假设、拟解决的关键问题、立题背景和依据是否充分等。

2. 研究方法与研究对象是什么　在这部分，需要了解具体的研究设计方案、研究方法是否新颖合理、样本来源与样本大小如何、设置哪些测量指标以及指标的实际价值和意义如何？

3. 主要结果和重要发现有哪些　全面熟悉文献的主要结果，明确文中的新发现、新贡献，用于结果报告的重要图表有哪些？

4. 讨论与结论　文献的结论是什么？该文的主要结果是否支持这些结论？文中的立题、设计、方法和讨论间是否有内在的逻辑关系？研究目的是否实现？主要不足有哪些？还遗留有哪些问题未解决？

5. 小结　精读后，最好能加以小结。如与同类文献相比，该文献有什么共同点和不同点？作者的整体思路有无创新之处？必要时，可进行类似 SWOT 分析，找出优势和不足。

第三节　医学研究文献的评价要素与方法

一、确定医学研究文献评价的范畴和内容

评价文献时同样需要带着问题，有针对性地进行。评价目的不同，决定了评价范畴和重点。文献质量评价涉及两个方面的内容，即报告质量和方法学质量。其中报告质量是指文献报告内容的全面性和完整性以及和相应报告规范的符合程度。方法学质量是指文献制作过程中遵循科学标准、有效控制混杂与偏倚、使研究结果达到真实可靠的程度，为文献质量评价的核心内容。

二、评价医学研究文献的一般原则

阅读医学研究文献的目的，主要是回答“what”类问题，以全面了解该文献中的研究目的、对象、研究方法、主要结果和重要结论等内容；而评价医学研究文献，则是回答“why”和“how”类问题，一般选在二轮精读时进行，要求评价者具备批判性思维并掌握一定的评价原则与方法。

评价内容主要包括真实性评价、重要性评价及适用性评价。要依次回答下列问题：该研究结果本身是否真实可靠？有多大临床意义和实用价值？用于临床实践的可行性及适用程度如何？其中研究结果的真实性和重要性是评价的重点，只有真实、可靠、重要的研究结果才有利用价值。

（一）真实性评价原则

真实性又称为效度，是指研究收集的数据、分析结果和所得结论与客观实际的符合程度。在进行科学研究的整个过程中由于设计、资料的收集、整理和分析、结果报告的各个环节均可能由于各种因素的影响而导致观察值与真实值不一致，产生误差（error），从而影响结果的真实性。

评价真实性的关键，在于考核研究过程中是否有效控制了混杂与偏倚对结果的影响。这些因素对结果的影响大小决定了真实性的程度。因此，评价真实性应综合考虑研究结果是来自何种设计方案、有无对照组以及设置是否恰当、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。

（二）重要性评价原则

临床重要性（clinical importance）是指研究结果在临床应用的价值，可以通过客观指标来评价研究证据的临床意义。评价临床研究结果的重要性常借助于一些定性或定量指标。重要性包括临床重要性与统计学意义两个方面，两者应相互结合，作综合评价。统计学意义的判定可以通过假设检验和区

间估计加以实现。临床重要性的判断，还应做卫生经济学的评价，进行成本-效果（cost-effectiveness）、成本-效益（cost-benefit）以及成本-效用（cost-utility）分析，使那些成本低、效果佳的研究成果，得以推广应用。

（三）适用性评价

适用性，即外部真实性（external validity，generalizability，applicability），指研究结果是否可以应用于研究对象以外的其他人群，即研究结果与推论对象真实情况的符合程度，又称为普遍性（generalizability），反映研究结果的适用价值与推广应用的条件。

适用性评价同样需要结合阅读文献的目的，是为了指导临床实践、教学，还是为临床科研所用。若将上述真实性好且有重要临床价值的文献结果在临床实践中加以应用和推广，应结合自己患者的实际病况和接受程度、现有医疗条件和知识技能水平，以及社会经济状况的承受能力等，对其临床适用性展开评价。

若在临床科研、教学过程中，为掌握学科最新进展或发展方向、拓展新思路、新视野之用，也要进行适用性评价，对文献结果是否具有外部真实性，即能否将结果推广应用到研究对象以外的群体或环境时，应考核研究人群与其他人群的特征差异、研究对象类型以及社会环境和经济等因素是否会影响适用性等。

三、常见临床研究类型的评价标准

临床研究常见的类型包括病因学研究、诊断性研究、治疗性研究和预后研究，针对不同的研究类型其采用的原始研究的设计方案和实施方法也不同，因此其评价标准与评价指标也各有特色。有关病因与不良反应研究的评价标准、诊断性研究的评价标准、治疗性研究的评价标准、预后研究的评价标准等参见本书相关章节。

四、研究质量评价工具

为方便评价，针对不同研究类型，已有一些现成的严格评价工具或量表，可供参考借鉴。评价工具大多由一些知名学术机构或组织研发。例如 *JAMA* 杂志发布的用户指导手册系列、CASP 严格评价技巧项目网提供的系列质量评价标准等，专门用于评估包括系统评价、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在内的不同类型文献。评价工具可分为清单（checklist）类和尺度评分（scale）类两种。不同的研究设计，其科学论证强度不同，评价方法及评价工具也有所不同。下面将分别介绍原始研究和二次研究的方法学质量评价工具和研究报告质量评价工具。

（一）原始研究评价工具

目前以随机对照试验的评价工具发展最快。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，相关文献发表也很多，成为临床证据的重要来源，相应的质量评价方法也发展很快。随机对照试验的方法学质量评价推荐使用 Cochrane 手册中的偏倚风险评价工具，Jadad 评分等，报告质量评价推荐使用 CONSORT（Consolidated Standards of Reporting Trials）；对于观察性研究（包括队列研究、病例对照研究等），方法学质量评价工具推荐使用纽卡斯尔-渥太华质量评价量表（NOS 评分），报告质量评价可以选择观察性流行病学研究报告规范（strengthening the reporting of observational studies in epidemiology，STROBE）；诊断试验研究方法学质量评价工具推荐使用 Cochrane 推荐的 QUADAS（quality assessment of diagnostic accuracy studies）清单，报告质量评价推荐使用诊断试验准确性研究的报告规范 STARD（standards for the reporting of diagnostic accuracy studies）。

（二）二次研究评价工具

1. 报告质量评价工具 PRISMA（Preferred Reporting Items for Systematic Reviews and Meta-

Analysis)声明主要针对干预性研究的系统评价和meta-分析的报告规范,该标准的制定对于改进和提高系统评价和meta-分析的报告质量将起到重要作用。

PRISMA的官方网站上可以下载到PRISMA详细的报告条目,PRISMA声明由一个27个条目清单和一个四阶段的信息收集流程图组成。清单包括7个方面的内容:①标题;②结构式摘要;③引言(基本原理和目的);④研究方法(方案与注册、纳入与排除标准、信息来源、检索策略、筛选研究、数据收集过程、数据项、单个研究偏倚的风险、结局指标、结果合成、不同研究之间的偏倚风险、附加分析);⑤结果(研究选择、研究特征、研究中的偏倚风险、单个研究的结果、结果合成、不同研究之间的偏倚风险、附加分析);⑥讨论(证据小结、局限性、结论);⑦项目资助情况。

PRISMA还发表了一系列相关报告清单:Beller等人于2013年发表了系统评价/meta-分析摘要的优先报告条目PRISMA-Abstracts,旨在规范摘要的报告;Moher等人于2015年发表了PRISMA-protocol,规范了系统评价/meta-分析研究方案的报告;Welch Vivian等编制了PRISMA-Equity,为报告关于健康公平性系统评价的透明报告提供了依据;PRISMA-IPD:2015年*JAMA*杂志上发表了,即单病例数据系统评价/meta-分析的优先报告条目,在标准的PRISMA声明的基础上补充了IPD的获取、核查、合成以及如何处理没有提供IPD的研究;PRISMA-NMA:为了完善网状meta-分析的研究报告,Hutton等人于2015年发表了PRISMA-NMA,用以指导和改善网状meta-分析的撰写和报告。PRISMA伤害清单(PRISMA harms checklist)一共包括27个条目,在报道不良事件时,能够确定一套最小化的评价项目,无论伤害作为主要结果还是次要结果,均能提高系统评价中伤害的报道,最终提高益处和伤害评价的均衡性。

1997年美国疾病预防控制中心组织专家小组制定了观察性研究的系统评价和meta-分析(meta-analysis of observational studies in epidemiology,MOOSE)的报告规范,该规范包含七大部分内容(研究背景、文献检索策略、研究方法、研究结果、讨论和研究结论),共35个条目。主要指导撰写纳入的原始研究是观察性研究设计类型,包括队列研究、病例对照研究、横断面研究进行的系统评价和meta-分析的研究报告。

2. 方法学质量评价工具　目前用于系统评价/meta-分析的方法学质量评价的工具量表推荐使用AMSTAR质量工具(A Measurement Tool to Assess Systematic Reviews),包括11个条目,每个条目均采用“是(Yes)”“否(No)”“不知道(Can't answer)”和“不适用(Not applicable)”进行判定。详见本教材第七章和第十章。

(张　玲)

NOTE

第七章　系统评价与 meta-分析

有专家提出，若对可能早产的孕妇使用激素，能促进胎儿的肺发育。你查寻了有关资料，发现有7个高质量的随机对照试验，其中5个试验结果为阴性（使用激素后未能减少早产儿的死亡率和呼吸窘迫综合征的发生率），2个试验结果为阳性，你将作何决策呢？科学回答这个问题就需要对这七篇文献进行系统评价，以助自己的科学决策。系统评价（systematic review，SR）和 meta-分析（meta-analysis）是一种科学合成医学信息的方法，本章将重点阐述系统评价及 meta-分析的方法及其相关的质量评价原则，以供读者参考和应用。

第一节　系统评价概述

一、基本概念

（一）系统评价

系统评价是一种全新的文献综合方法，针对某一具体医学及相关问题（如临床、卫生决策、基础医学、医学教育等问题），系统、全面地收集现有已发表或未发表的临床研究，采用临床流行病学严格评价文献的原则和方法，筛选出符合质量标准的文献，进行定性或定量合成（meta-analysis，meta-分析或荟萃分析），得出可靠的综合结论。系统评价可以是定性的（定性系统评价，qualitative systematic review），也可以是定量的（定量系统评价，quantitative systematic review）即包含 meta-分析过程。

系统评价非常明确和清晰的研究过程使其具有良好的重复性。系统评价可为某一领域和专业提供大量的新信息和新知识，多数是可信的。但是，由于是对原始文献的二次综合分析和评价，受原始研究文献质量、系统评价方法及评价者本人专业知识、认识水平和观点等的制约，因此，读者在阅读系统评价的观点和结论时，一定要持谨慎的态度，不能盲目被动地接受。

（二）Cochrane 系统评价

Cochrane 系统评价是 Cochrane 协作网的评价员按照统一工作手册（如 Cochrane Handbook for Systematic Reviews of Interventions），在相应 Cochrane 评价小组编辑部的指导和帮助下所完成的系统评价。由于 Cochrane 协作网有严密的组织管理和质量控制系统，严格遵循 Cochrane 系统评价者手册，采用固定的格式和内容要求，统一的系统评价软件（Review Manager，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每2年定期更新，有健全的反馈和完善机制，因此 Cochrane 系统评价的质量通常比非 Cochrane 系统评价质量更高，被认为是单一的、评价干预措施疗效的最好证据资源（best single source）。

Cochrane 协作网的56个系统评价小组是制作和保存系统评价的基本单元，根据具体的健康相关研究主题分组如高血压组、急性呼吸道感染组等。Cochrane 系统评价目前主要针对干预性措施疗效和安全性、诊断试验准确性、方法学、教育和公共卫生领域等相关问题开展系统评价。

（三）meta-分析（meta-analysis）

meta-分析由心理学家 Glass1976 年首次命名，国内翻译为荟萃分析、汇总分析。就其定义目前仍然存在不同的争议。Huque 及多数专家认为："meta-分析是一种统计分析方法，它将多个独立的、可以合成的临床研究综合起来进行定量分析"。因此，如果没有明确、科学的方法去收集、选择、评价临床研究资料，而仅单纯采用统计方法将多个临床研究进行合成并不能保证结论的真实性和可靠性。

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

（四）文献综述（review）

又称为叙述性文献综述（narrative review）或传统文献综述（traditional review），由作者根据特定的目的和需要或兴趣，针对某一领域、专业或研究专题，搜集大量相关资料，在广泛阅读和理解基础上，采用定性的方法，综合分析、归纳整理和提炼该领域的研究现状、最新进展、学术见解或建议，作出综合性介绍和阐述的学术论文，可为某一领域或专业提供大量的新知识和新进展，以便读者在较短时间内了解某一专题的研究概况和发展方向，解决临床实践中遇到的问题。但这种传统文献综述，往往受限于专家个人的知识和信念，缺乏客观方法，故存在一定局限性。在接受或应用这类证据时，宜持谨慎态度。

系统评价和传统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。确定一篇综述为叙述性文献综述，还是系统评价及其质量、价值，主要看其是否采用科学方法减少偏倚或混杂因素的影响。

传统文献综述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、诊断方法及预防、治疗、康复措施，也可仅涉及某一方面的问题如诊断、治疗等，有助于广泛了解某一疾病的全貌。系统评价或 meta-分析则为集中研究某一具体临床问题的某一方面如二甲双胍治疗 2 型糖尿病患者的疗效和安全性，具有相当深度，有助于深入了解某一具体临床问题，二者的区别和联系如表 7-1。

表 7-1　叙述性文献综述与系统评价的区别

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

二、进行系统评价的必要性

（一）应对信息时代的挑战

每年约有 400 万篇生物医学文献发表在 3 万多种生物医学杂志上，年增长率约为 7%。一个内科医师需要每天不间断地阅读数十篇本专业文献才能勉强掌握本学科的新进展和新研究结果，使得需要大量信息进行科学决策的临床医生、研究人员和卫生部门的决策者往往陷入难以驾驭的信息海洋之中。而系统评价采用严格的选择、评价方法，去粗取精、去伪存真，将真实、可靠而有临床应用价值的信息进行合成，可直接为各层次的决策者提供科学依据。

（二）及时转化和应用研究成果

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

用系统评价方法合成多个质量较高的同质临床试验结果可将有效措施及时转化和用于临床实践与决策。如采用累积性meta-分析回顾性分析了静脉滴注链激酶溶栓治疗急性心肌梗死(AMI)的临床试验,1973年前发表的8个RCT(2432例患者)的meta-分析即证明静脉滴注链激酶溶栓能有效降低AMI患者的总死亡率($P=0.01$);1977年前发表的18个RCT(4879例患者)的meta-分析,$P=0.001$(包括GISSI-1研究的11 712例);1986年前发表的26个RCT(18 938例患者,包括GISSI-1研究的11 712例)的meta-分析,$P=0.0001$;1988年前发表的33个RCTs(36 974例患者,包括ISIS-2的17 187例)的meta-分析,$P<10^{-15}$。至此静脉滴注链激酶才在传统综述和教科书中推荐常规用于治疗急性心肌梗死。临床应用比meta-分析结果整整晚了15年!试想如能早用可挽救多少AMI患者的生命?对此1994年Murphy等指出:1973年以后的大型临床试验,即使没有医德问题也是多余的,且耗费大量经费。

(三)提高统计效能

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

三、系统评价与叙述性文献综述的区别与联系

系统评价和叙述性文献综述均是对临床研究文献的分析和总结,目前多为回顾性、观察性的研究,也可为前瞻性系统评价。回顾性的系统评价因受到纳入的原始临床研究质量的制约,难以有效控制偏倚和混杂的影响。因此,确定一篇"综述"属于叙述性文献综述还是系统评价?其质量、价值如何?主要取决于是否采用科学的方法减少了偏倚、混杂因素的影响。

叙述性文献综述常常涉及某一问题的多个方面,如糖尿病的病理、病理生理、流行病学、诊断方法及预防、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。而系统评价或meta-分析的研究问题往往比较具体明确。因此,叙述性文献综述有助于广泛了解某一疾病的全貌,而系统评价则有助于深入了解某一具体疾病的某一具体方面。二者的主要区别如下(表7-1):

四、系统评价分类

系统评价本身只不过是一种研究方法,并不仅限于对随机对照试验或仅对治疗性研究进行系统评价。系统评价和meta-分析分类如表7-2:

表7-2 系统评价和meta-分析分类

分类方法	类 型
研究领域	基础研究、临床研究、医学教育、方法学研究、政策研究……
临床问题	病因、诊断、治疗、预后、卫生经济学……
原始研究类型	临床试验:随机和非随机对照试验 观察性研究:队列研究和病例-对照研究 定性/质性研究
纳入研究的方式和数据类型	前瞻性meta-分析/回顾性meta-分析、累积性meta-分析、网状meta-分析、个体病例资料meta-分析、系统评价再评价……
是否采用统计学方法	定性系统评价、定量系统评价

笔记

第二节　系统评价的方法

系统评价是一把双刃剑，一方面能够通过对多个有争议或甚至相互矛盾的小型临床研究采用严格、系统的方法进行评价、分析和合成，以解决争议或提出建议，指导临床实践、医疗决策以及明确今后的研究方向；另一方面，如果纳入的原始研究质量不高或进行系统评价/meta-分析的方法不当，会产生不正确的信息，造成误导。因此，系统评价方法和步骤，决定了其结果和结论的真实可靠性。

一、系统评价前的准备

（一）时间投入

完成 1 篇系统评价所需时间受多种因素影响，很难确切回答。一般针对中国的系统评价者，影响因素包括初筛的文献量、纳入系统评价的文献量、中英文文献的比例、评价者对系统评价方法的熟练程度等。完成 1 篇纳入 20 个研究的系统评价需要专职工作 2～3 个月。纳入研究越少、英文文献比例低，可能需要的时间相对少些，反之亦然。但完成 1 篇 Cochrane 注册的系统评价，不仅受系统评价者自身和文献量的影响，还受不同 Cochrane 评价小组工作效率的影响，需要专职工作 12～18 个月。

（二）人员组成

1 篇系统评价至少由 2 名作者完成，以保证在文献筛选、质量评价和数据提取过程中由 2 人独立完成，有不同意见时讨论后达成一致，增加发现问题的机会。1 篇系统评价的作者中应包括研究问题所涉及专业的人员、熟悉临床流行病学或流行病学研究方法及统计学方法的人员和信息专家，鼓励初学者与有经验的系统评价作者合作，保证研究的顺利进行。

（三）经费

制作系统评价和 meta-分析所需经费有限，主要用于人力资源、获取全文文献和必要的耗材。

二、系统评价流程

针对不同研究问题的系统评价其基本方法和步骤相似，但在文献检索策略、数据库选择、文献质量评价方法、原始文献中数据提取及统计分析等具体内容上有差异。生产系统评价的基本过程一般分 4 个阶段、9 个基本步骤（表 7-3）。

表 7-3　系统评价流程

4 个阶段	9 个步骤
第一阶段：确定系统评价题目	1. 确定题目
第二阶段：制定系统评价方案	2. 撰写系统评价研究方案和注册
第三阶段：完成系统评价全文	3. 检索文献
	4. 筛选文献
	5. 评价文献质量
	6. 提取数据
	7. 分析和报告结果
	8. 解释结果，撰写报告
第四阶段：更新系统评价	9. 更新系统评价

三、系统评价方法

尽管系统评价的种类繁多，但基本方法和步骤相似（图 7-1），仅在检索策略、文献质量评价、数据提取以及汇总分析等方面略有不同，本节将以 Cochrane 系统评价为例，简述其基本方法和步骤。

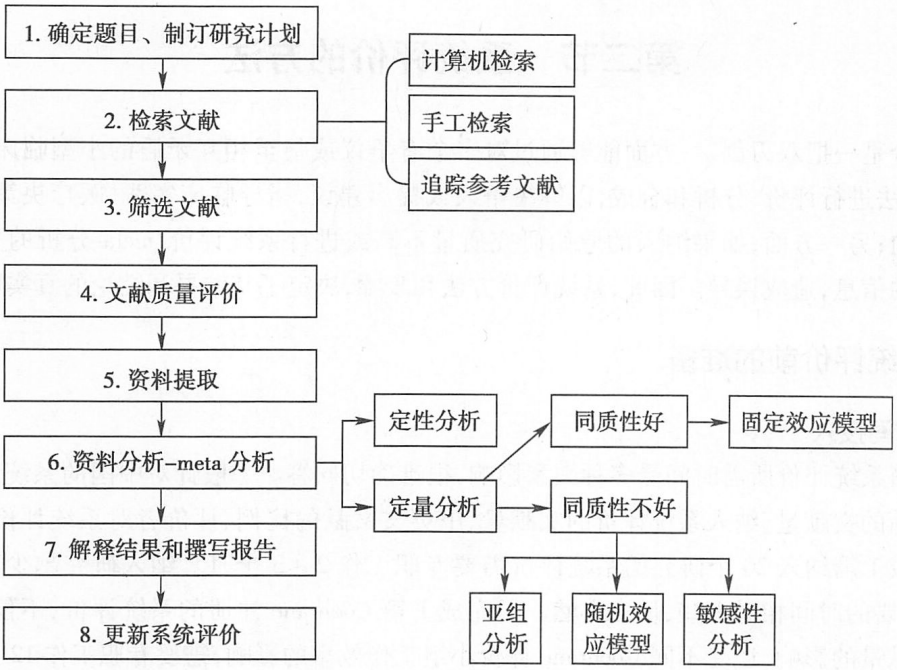


图 7-1　系统评价基本流程

（一）确立题目与制订系统评价计划书

制作系统评价的主要目的是为医疗和卫生决策提供依据，因此，系统评价的选题应遵循“三有一无”的原则。①有意义：所选题目应解决或回答医疗和卫生领域关注的重要问题，能改变我们对某些问题的认识、改变或更新当前临床实践指南、或者规范临床实践行为；②有争议：系统评价特别适合回答某些有争议或有疑虑的医疗和卫生问题，如针对同一临床问题的研究较多，但结论不一致，靠单个临床研究结果难以确定，或在临床应用过程中存在较大争议等问题的探讨。如：高危人群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继发心脏事件？③有研究：系统评价多数是对现有研究的再次分析、评价和总结，若没有针对某个问题的原始研究，如何开展评估？因此，所选题目应有一定数量、较高质量的原始研究；④无重复：这是一个相对的概念，是指要避免不必要的重复。若针对某一有争议的问题目前尚无相关系统评价，这样的选题当然最好。但某些热点、有争议的问题虽已有发表的系统评价，但因纳入研究数量有限、质量较差，当前证据尚不能明确回答，随着新研究的发表进行更新也非常必要。同样，某些系统评价并未全面回答某些有争议的问题，再重复也是有意义的。如评价联合使用质子泵抑制剂和氯吡格雷与心血管事件和死亡的关系，早期的系统评价因纳入研究有限，将质子泵抑制制作为同类药进行评估。是否所有质子泵抑制剂都有此不良反应呢？所以后期的系统评价不仅分析同类药的影响，还分析了不同质子泵抑制剂与心血管事件间的关系是否有差别，这对临床合理选择药物提供了更有价值的依据。因此，为避免重复，首先应进行全面、系统的检索，了解针对同一临床问题的系统评价/meta-分析是否已经存在或正在进行。若有，质量如何？是否已过时（如发表后有较多新的研究出现等）？若现有的系统评价/meta-分析已过时或质量差，则可考虑进行更新或做一个新的系统评价。

系统评价解决的问题很专一，涉及的研究对象、设计方案以及治疗措施需相似或相同。因此，确立题目时应围绕研究问题明确 PICOS 要素，如针对治疗性研究的 PICOS 要素包括：①P（participants/patients）：研究对象的类型，所患疾病类型及其诊断标准、研究人群的特征和所处环境；②I（intervention）：研究的干预措施及实施细节；③C（comparison）：进行比较的措施及实施细节；④O（outcomes）：主要研究结果的类型，包括所有重要的结果（主要结果和次要结果）及严重的不良反应；⑤S（study design）：研究的设计方案，如随机对照试验和/或非随机对照试验、队列研究、病例-对照研究。

这些要素对指导检索、筛选和评价各临床研究，收集、分析数据及解释结果的应用价值均十分重

要，必须准确、清楚定义。

一旦确立了系统评价的题目，应着手制订计划书（protocol），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案等。

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

若生产 Cochrane 系统评价，确定题目后需要在相关评价小组填表注册，以避免重复。注册需要填写所在系统评价小组的系统评价申请表（review proposal form），各系统评价小组申请表的内容和格式不完全一致，由各系统评价小组自行制定。内容主要包括：①立题依据；②系统评价目的；③研究入选标准（基于 PICOS 要素）；④研究团队成员的信息和制作系统评价的经历；⑤经费资助情况；⑥有无利益冲突问题；⑦预计完成计划书；⑧系统评价全文的时间等。完成申请表后提交给相应系统评价小组，能否成功注册由系统评价小组请相关临床专家和方法学专家讨论决定。

（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。为了避免发表偏倚（publication bias）和语言偏倚（language bias），应围绕要解决的问题，按照计划书中制订的检索策略（包括检索工具及每一检索工具的检索方法），采用多种渠道和系统的检索方法。除发表的论著之外，还应收集其他尚未发表的内部资料以及多语种的相关资料。

除利用文献检索的期刊工具及电子光盘检索工具（如 Medline、Embase、Registers of clinical trials）外，制作系统评价时，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，先后建立了 Cochrane 对照试验中心注册库（Cochrane Central Register of Controlled Trials，CENTRAL）和各专业评价小组对照试验注册库，既可弥补检索工具如 Medline 等标识 RCT 不全、漏录的问题，也有助于系统评价者快速、全面获得相关的原始文献资料。

为有效管理检出的文献，特别是当文献量较大时，有必要借助文献管理软件如 EndNote、Reference Manager、Procite 等管理文献题录、摘要信息、全文等，便于剔重、浏览、筛选和排序等，也有助于撰写文章时编写参考文献格式和引用参考文献等。

（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 天内的病死率，设计方案为随机对照试验（RCT），则所选临床研究必须符合上述条件。而像口服硫酸镁或静脉滴注硫酸镁与其他药物进行比较、结果为心肌梗死后 35 天以后的病死率或者非 RCT 的文献资料则不能纳入。

文献资料的选择应分三步进行（图 7-2）：①初筛：根据检索出的引文信息如题目、摘要筛除明显不合格的文献，对肯定或不能肯定的文献应查出全文再进行筛选；②阅读全文：对可能合格的文献资料，应逐一阅读和分析全文后，再确认是否合格；③与作者联系：一旦被排除的文献将不再纳入，因此，如果文中提供的信息不全而不能确定者，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。

文献筛选是以研究为单位，而不是以发表的研究报告为单位。同一研究可能被化整为零而发表

NOTE

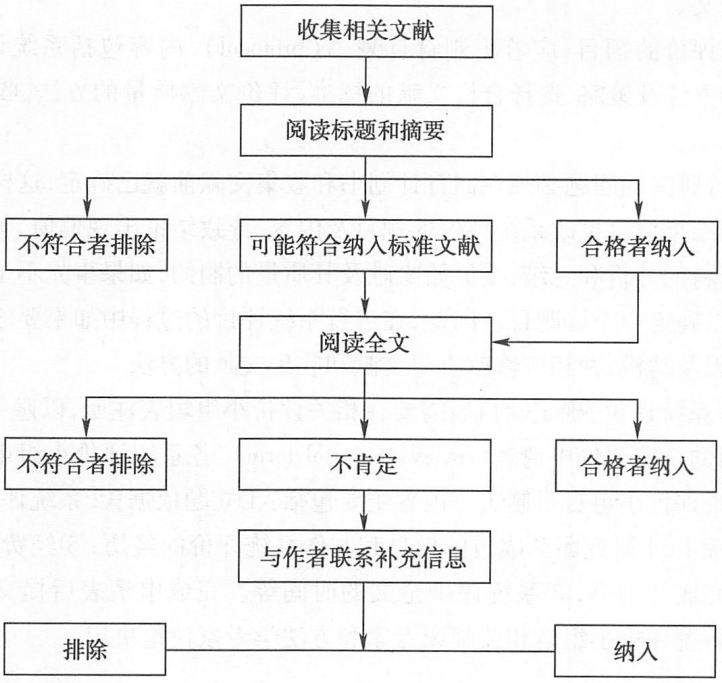


图 7-2　筛选文献的基本步骤

多篇文献(每篇文章报告部分结果,内容不重叠或部分重叠)、或者不同随访期(短期和长期)的结果分别报告、或者多重发表(内容和数据基本相似)。筛选时务必根据研究目的进行选择,而要避免将同一研究发表的多篇内容和数据相同的文献重复纳入,增加某一研究在系统评价中的权重。

文献筛选要设计筛选表和筛选说明,且由两人独立筛选,避免相关文献被排除的可能性。

(四) 评价偏倚风险

多数系统评价是针对已完成的研究进行二次评估,原始研究的质量直接影响系统评价结果和结论的真实性和可靠性。因此,评估纳入系统评价的原始研究在设计、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,以分析和解释纳入研究质量对结果的影响至关重要。研究质量评价应包括:①内部真实性(internal validity):指单个研究结果接近真值的程度,即受各种偏倚因素如选择偏倚、实施偏倚、失访偏倚和测量偏倚的影响情况;②外部真实性(external validity/generalizability):指研究结果是否可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及条件和结果的选择标准密切相关。

研究质量评价(assessment of quality)是评估单个研究在设计、实施和分析过程中,防止或减少偏倚或系统误差的情况,也称为"方法学质量评价(assessment of methodological quality)"。一直以来,研究质量评价和偏倚风险评价被认为是等同的,但 Cochrane 系统评价手册认为偏倚和质量有区别。例如,某研究不可能对研究对象、干预措施的实施者或结果评价者采用盲法,若由此认定该研究的"低质量",并不恰当。又如研究质量已达到了可能的最高水平,但并不是说该研究没有偏倚。另外,某些与研究质量相关的指标并不直接导致偏倚风险,如样本量估算、伦理审查和报告质量等。

评价文献质量和偏倚风险的方法较多,可采用单个条目、清单或一览表和量表评分,但缺乏共识。针对不同临床问题如治疗、病因、诊断和预后的系统评价,虽然进行系统评价的基本步骤相似,但其纳入研究的设计类型和实施方法并不相同。因此,纳入研究的质量评价工具和方法也有明显差别。

干预措施疗效和安全性研究从选择和分配研究对象、实施干预措施、随访研究对象、测量和报告研究结果的每个阶段均可出现偏倚(图 7-3)。因而根据研究阶段将偏倚分为以下五种:①选择偏倚(selection bias):发生在选择和分配研究对象时,因随机方法不完善造成组间基线不可比,可夸大或缩小干预措施的疗效。采用真正的随机方法并对随机分配方案进行完善的隐藏,可避免此类偏倚的发生。②实施偏倚(performance bias):发生在干预实施的过程中,除待比较的措施不同外,所提供的其

笔记

他措施也不一样。采用标化治疗方案和盲法干预可避免实施偏倚。③随访偏倚(attrition bias):指在试验随访过程中,试验组或对照组因退出、失访、违背治疗方案等造成人数或其他情况不一致而产生的系统偏差。对此,应尽量搜集失访者的信息以及对失访人员采用恰当的统计学方法处理,如意向性治疗分析(intention to treat analysis)等,可减少其影响。④测量偏倚(measurement bias/detection bias/ascertainment bias):试验组和对照组的结果测量方法不一致所造成的系统偏差,特别容易在主观判断研究结果时出现。采用统一、标化测量方法以及对研究对象和结果测量者实施盲法,可避免其影响。⑤报告偏倚(reporting bias):指文章中报告的结果与实际分析结果间存在的系统偏差,如有选择性地报告结果。

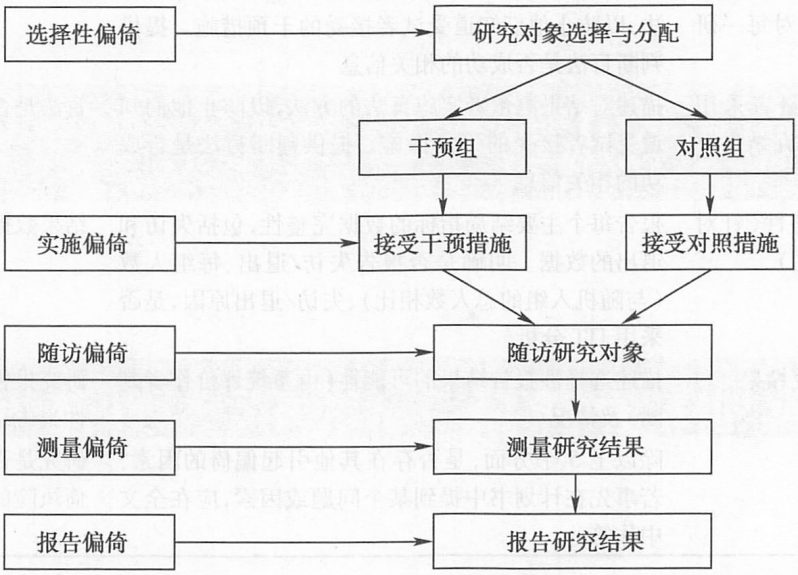


图 7-3　偏倚来源

治疗、预防、康复等干预措施疗效和安全性的系统评价多数纳入 RCT 为主,评价 RCT 质量的工具很多。1995 年 Moher 等鉴定出 9 种清单和 60 余种量表,分别有 3 ~ 57 个条目,需要花 10 ~ 45 分钟完成。由于这些评价方法易受文献报告质量和文献评估者的主观因素影响,Cochrane 手册 5.0 未推荐使用任何一种清单或量表,Cochrane 协作网推荐采用由相关方法学家、编辑和系统评价员共同制订的“Cochrane 偏倚风险评估”工具(表 7-4),包括 7 个方面:①随机分配方法;②分配方案隐藏;③对研究对象、治疗方案实施者采用盲法;④对研究结果测量者采用盲法;⑤结果数据的完整性;⑥选择性报告研究结果;⑦其他偏倚来源。利用上述 7 条标准可逐一比对每一个纳入研究,并作出“偏倚风险低”“偏倚风险高”和“偏倚风险不确定”的判断。其中,①②⑥⑦用于评估每一篇纳入研究的偏倚风险,其余 3 条则需针对每一篇纳入研究中的不同研究结果进行评估,强调同一研究中不同结果受偏倚影响程度不同。偏倚风险评价结果不仅采用文字和表格描述,还要求采用图示,更形象、直观反映偏倚情况。此评估工具对每一条的判断均有明确标准,减少了评估者主观因素影响,保证了评估结果具有更好的可靠性。

诊断准确性研究(diagnostic accuracy studies)质量评价的工具较多。2005 年 Whiting 等系统评价了诊断准确性研究的质量评价工具,鉴定出 90 种清单或量表,但均不是通用工具,涉及的条目数和内容各不相同。Cochrane 协作网的诊断准确性研究系统评价方法学组推荐采用改良的 QUADAS 清单评价诊断准确性研究的方法学质量。QUADAS 清单是基于已有的影响诊断准确性研究结果真实性、重要性和适用性的研究证据,采用严格的专家共识方法制订的通用评价工具,2003 年发表后先后于 2006 年和 2011 年修订(QUADAS-2)。2003 年发表的 QUADAS 清单共 14 个条目,针对诊断准确性研究的偏倚风险、可靠性和报告质量。Cochrane 协作网诊断准确性研究系统评价方法学组曾采用其中

11 个条目。2011 年修订的 QUADAS-2 包括 4 个维度，即①研究对象选择；②诊断技术；③金标准；④诊断技术和金标准进行的流程和时序，分别从偏倚风险和适用性进行评估。

表 7-4　Cochrane 协作网的偏倚风险评价工具

评价条目	评价内容描述	作者判断
①随机分配方法	详细描述产生随机分配序列的方法，以助于评估能否产生组间可比性	随机分配序列的产生是否正确
②分配方案隐藏	详细描述隐藏随机分配序列的方法，以助于判断干预措施分配情况是否可预知	分配方案隐藏是否完善
③对研究对象、治疗方案实施者采用盲法（针对每一研究结果评估）	描述对研究对象和试验措施实施者实施盲法的方法，以防止他们知道受试者接受的干预措施。提供判断盲法是否成功的相关信息	盲法是否完善
④对研究结果测量者采用盲法（针对每一研究结果评估）	描述对结果测量者实施盲法的方法，以防止他们知道受试者接受的干预措施。提供判断盲法是否成功的相关信息	盲法是否完善
⑤结果数据的完整性（针对每一研究结果评估）	报告每个主要结局指标的数据完整性，包括失访和退出的数据。明确是否报告失访/退出、每组人数（与随机入组的总人数相比）、失访/退出原因，是否采用 ITT 分析	结果数据是否完整
⑥选择性报告研究结果	描述选择性报告结果的可能性（由系统评价作者判断）及情况	研究报告是否提示无选择性报告结果
⑦其他偏倚来源	除以上 5 个方面，是否存在其他引起偏倚的因素？若事先在计划书中提到某个问题或因素，应在全文中作答	研究是否存在引起高度偏倚风险的其他因素

非随机研究（non-randomized studies，NRS）的设计方案有多种如非随机对照试验、队列研究和病例-对照研究等，受偏倚影响情况也有差别，因此尚无一种通用的非随机研究偏倚评价工具。Deeks 等系统收集了评价非随机研究的工具 193 种，鉴定出 6 种适用于系统评价的工具，但并非每种非随机研究方案均适合。目前已针对不同研究问题的非随机研究研发出相应的偏倚风险评估工具：

1. 干预性非随机研究的偏倚风险评估工具（Risk Of Bias In Non-randomised Studies of Interventions，ROBINS-I）　2016 年由 Sterne 等发表，适用于评估干预措施疗效和安全性的队列研究、病例对照研究、半随机对照试验和其他非随机同期对照研究。该工具包括 7 个维度，分别针对干预措施实施前（2 个：混杂导致的偏倚和选择研究对象的偏倚）、干预措施实施中（1 个：测量干预措施的偏倚）和干预措施实施后（4 个：违背原定干预措施、数据缺失、结果测量和选择性报告结果导致的偏倚），前 3 个维度与随机对照试验偏倚风险评估工具不同，后 4 个与 RCT 偏倚风险评估工具相似。

2. 病因学研究的偏倚风险评估工具　2 种工具最常用，分别是“Downs and Black instrument”和“Newcastle-Ottawa Scale（NOS）”。前者包括 29 个条目，需要具有相当的流行病学知识且费时，某些条目难以用于病例-对照研究。后者已被 Cochrane 协作网的非随机研究方法学组用于培训中，只包括 8 个条目，涉及研究对象选择、组间可比性和结果测量（队列研究）或暴露因素测量（病例-对照研究），简单易用，分别针对病例-对照研究和队列研究。

3. 预后因素研究的偏倚风险评估工具（Quality In Prognosis Studies，QUIPS）　由 Cochrane 的预后方法学组制订并于 2013 年发表，包括 6 个维度，分别针对研究对象、研究对象失访、预后因素测量、结果测量、混杂因素和统计分析和报告。

4. 预测模型研究的偏倚风险评估工具（Prediction model studies Risk Of Bias Assessment Tool，PROBAST）　2015 年的 Cochrane 年会上发布，但尚未正式发表，包括 7 个维度，分别为：①患者选择；②结果测量；③预测因素测量；④样本量；⑤缺失数据；⑥统计分析；⑦模型验证。

笔记

为避免选择文献和评价文献质量人员的偏倚，可考虑一篇文章由多人或盲法选评，也可采用专业与非专业人员相互配合、共同选评的办法，对选评文献过程中存在的分歧可通过共同讨论或请第三方协助解决。若多人选择文献时，应计算不同评价者间的一致性（Kappa 值）。此外，最好先进行预试验，以摸索经验，标化和统一选择、评价方法。

（五）提取数据

提取数据是指采用手写或计算机录入方式将需要提取的信息填入数据提取表，即从原始研究的全文或者研究者提供的资料中收集相关数据的过程。但此过程不仅是从原始文献中摘抄信息，还涉及数据的处理或换算（如有的研究中血压用 Kpa 为单位，而有的用 mmHg 为单位；当数据为连续变量时，需要提取均数和标准差，但有的研究中为标准误等），是系统评价结果分析的基础，研究数据提取的完整性和质量直接影响数据分析。因此在阅读全文提取数据前要精心设计数据提取表，以保证重要、有意义的信息和数据不被遗漏，否则反复修改提取表和反复提取信息会增加不必要的工作量。

数据提取可采用纸质表和电子表 2 种，各有优劣（表 7-5）。

表 7-5　纸质和电子提取表优劣比较

特征	纸质	电子表
设计	容易	复杂
储存和提取大量数据	难	容易
数据转换	不能	能
同时进行数据提取和录入	不能	能
找出错误和不一致	难	容易
表格修改	相对困难	容易
方便性	随时提取	需计算机
证据存档	永久记录，无法抹去删除	不留痕迹

不同题目的系统评价因涉及的研究问题不同，提取的数据信息不尽相同，要充分反映研究问题的独特性。虽然没有统一的标准，但需要提取的某些基本信息是一致的，包括：①研究基本信息：如纳入研究的题目和编号、引文信息、提取者姓名、提取日期等；②研究基本特征：如研究的合格性、研究的设计方案和质量、研究对象的特征和研究地点、研究措施或暴露因素的具体内容、结局指标测量方法等；③研究结果：如随访时间、失访和退出情况、数据资料如治疗性研究中计数资料应收集每组总人数及事件发生率、计量资料应收集每组研究人数、均数和标准差或标准误等。而诊断准确度研究中要收集敏感度、特异度或能计算相关指标的原始数据信息。

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。

1. 定性分析（non-quantitative synthesis）　定性分析是采用描述的方法，将每个临床研究的特征按研究对象、干预措施、研究结果、偏倚风险和设计方法等进行总结并列成表格形式，以便纵览所有纳入研究的结果、方法及研究间的差异，同时帮助制订定量合成计划及其结果解释，因此，定性分析是定量分析前必不可少的步骤。

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

（1）异质性检验（heterogeneity）：定量系统评价或 meta-分析是将多个研究结果合并成一个效应值，但不同研究结果间不可避免存在差异，即异质性。异质性分三类：临床异质性（clinical heterogeneity），指不同研究中研究对象、干预措施和结果测量等存在的差异；方法学异质性（methodological heter-

ogeneity)，指试验设计和质量在不同研究中存在的差异；统计学异质性(statistical heterogeneity)，指不同研究中干预措施的效应值存在的差异，是临床异质性和方法学异质性共同作用的结果。异质性检验是指对不同原始研究结果之间的变异程度进行检验。如果检验结果有统计学意义，应解释其可能的原因并考虑是否进行结果合成。确定各研究结果是否同质有两种方法：一是目测法，即通过绘制森林图观察各研究效应值的可信区间是否有重叠及其程度，如果可信区间差异太大，则放弃合成分析或采用随机效应模型；另一种方法是直接进行同质性检验(Q test，Chi-square test)，在此基础上借助 I^2 定量估计异质性大小，如 0% ~40% 表示异质性可能不重要，30% ~60% 表示有中度异质性，50% ~90% 表示有显著异质性，75% ~100% 表示有很大异质性。

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

(3) 敏感性分析(sensitivity analysis)：指改变某些影响结果的重要因素如纳入标准、偏倚风险、失访情况、统计方法(固定效应或随机效应模型)和选择不同的效应量(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，从而判断合并结果的稳定性及其程度。

(七) 解释系统评价的结果和作结论

系统评价的目的是帮助患者、公众、医生、管理者和决策者进行卫生决策，旨在提供信息和辅助解释结果，而不是作出推荐意见。因此，清晰陈述研究结果、深入的讨论和明确的结论是系统评价的重要组成部分。解释系统评价和做结论必须基于研究的结果，内容应包括：

1. 总结和解释结果 总结和解释 meta-分析结果时，干预性研究中应同时考虑干预措施的利和弊，结果的点估计值和 95% CI。点估计值主要表示效应值的强度和方向，而 95% CI 则反映效应值的变动范围和精确性，二者结合可提供更全面的信息，有助于解释结果的临床价值。

2. 评价证据的总体质量 Cochrane 协作网采用证据质量和推荐强度分级系统(Grading of Recommendations Assessment，Development and Evaluation，GRADE)分级和评估系统评价的总体质量。该系统是 2004 年由包括 WHO 在内的 19 个国家和国际组织、67 名专家(包括临床指南专家、循证医学专家、各个标准的主要制定者及证据研究人员)共同成立的 GRADE 工作组循证制定出的国际统一的证据质量分级和推荐强度标准，将证据质量分为高、中、低、极低 4 个等级，并根据纳入研究的总体偏倚风险、研究结果的一致性、证据的直接性、结果的精确性和是否存在发表偏倚 5 个因素降低随机对照试验的质量级别；根据效应值大小、是否存在剂量-效应关系和所有可能存在的、低估效应值的偏倚因素等 3 个因素升高观察性研究(如队列研究)的质量级别。

3. 证据的适用性 在确定系统评价结果的应用价值时，如治疗性问题，首先应考虑干预措施对患者的利弊，其次应考虑系统评价纳入研究中的研究对象是否与当前患者情况相似？是否存在生物学、社会文化背景、依从性、基础危险度、病情和价值观等方面的差异。

4. 系统评价的局限性 针对系统评价在文献检索的全面性、纳入研究质量、系统评价方法的可重复性、统计分析方法和是否存在发表偏倚等方面问题，阐述系统评价存在的潜在局限性。

5. 结论 系统评价的结论包括对①临床实践和②未来研究的意义两部分。在确定这两方面意义时，要考虑证据的质量、干预措施的利弊、患者的价值和喜好及卫生资源的利用，旨在帮助医务工作者和决策者正确选择和应用，为进一步的研究指明方向。

(八) 更新系统评价

系统评价的更新是指系统评价发表后，定期收集新的原始研究，按前述步骤重新分析、评价，以及

时更新和补充新的信息，完善系统评价。Cochrane 系统评价要求每 2 年更新 1 次，杂志发表的系统评价并不要求原作者定期更新。但若发表的系统评价无确切结论，或针对该题目的新研究不断出现时，也可考虑是否有必要更新系统评价。

第三节　meta-分析

一、meta-分析概述

meta-分析是一种定量合并方法。meta-分析通过对多个同类独立研究结果的汇总和合并分析，可以达到增大样本含量、提高检验效能之目的；同时也可提高效应量的估计精度；特别是当多个研究结果不一致或都无统计学意义时，用 meta-分析可得到更加接近真实情况的统计结果。

在一个系统评价中可以选用某个结局指标进行一次 meta-分析，也可选用多个结局指标实施多个 meta-分析。

二、meta-分析基本过程

meta-分析过程包括数据提取及汇总、异质性检验、模型选择、合并效应量估计及假设检验等。

（一）数据提取

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

（二）数据类型及其效应量的表达

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

不同数据类型决定了效应量的表达方式有所不同。效应量（effect size）常被定义为临床上有意义的值或改变量。当结局观察指标为二分类变量资料时，常用的效应量表达有相对危险度（relative risk，RR）、比值比（odds ratio，OR）、绝对危险度（absolute risk，AR）或 NNT 等；当结局观察指标为连续性变量资料、非罕发的计数数据、较多分类的等级资料时，效应量采用均数差值（mean difference，MD）或标准化均数差值（standardized mean difference，SMD）等表达方式。对于较少分类的等级资料或罕发的计数数据，可转化为二分类变量资料进行处理，并选用相应的效应量；对于类似发病密度的数据，可以使用 risk ratio，也简写成 RR。对于生存资料，效应量表达可用风险比（hazard ratio，HR）。

在此基础上，按照统一设计的数据提取表，系统收集所纳入研究的重要信息，如样本量、分析方法、主要结果变量、设计方案、发表年份、具体实施时间及地点、质量控制措施等。

（三）异质性检验（heterogeneity test）

meta-分析之前，应进行异质性检验，并根据异质性检验结果，来决定是否估计合并效应量。异质性检验又称同质性检验，旨在检验多个原始研究结果间的一致性。异质性检验方法主要有 Q 检验法与图形目测法等。若 Q 检验有统计学意义，则表明存在统计学异质性（statistical heterogeneity），需要探讨异质性的来源并进行相应处理。异质性来源主要从两个方面考虑：一是临床异质性（clinical heterogeneity），如纳入研究在研究对象、干预措施、结局观察指标等存在差异；二是方法学异质性（meth-

odological heterogeneity)，如纳入了不同设计方案、不同方法学质量的原始研究等。

1. Q 检验及 I^2 指数　Q 检验的无效假设为：所有纳入研究的效应量均相同(即 $H_0:\theta_1=\theta_2=\cdots\cdots=\theta_k$)，Q 统计量定义为：$Q=\sum w_i(\theta_i-\bar{\theta})^2$，进一步可表达为：$Q=\sum_{i=1}^{k} w_i\theta_i^2-\frac{(\sum w_i\theta_i)^2}{\sum w_i}$。

上式中 w_i 为第 i 个研究的权重值。θ_i 为第 i 个研究的效应量，$\bar{\theta}$ 为合并效应量，$\bar{\theta}=\frac{\sum w_i\theta_i}{\sum w_i}$。$k$ 为纳入的研究个数。Q 服从于自由度为 $k-1$ 的 χ^2 分布。若 $Q>\chi^2_{(1-\alpha)}$，则 $P<\alpha$，表明纳入研究间的效应量存在统计学异质性，可进一步计算异质指数 $I^2=\frac{Q-(k-1)}{Q}\times100\%$，用以定量描述异质程度。若 I^2 指数为 0% ~40%，表明异质性可忽略不计；I^2 = 30% ~60%，表明存在一定程度的异质性；若 I^2 = 50% ~90%，表明纳入研究的效应量存在较明显的异质性；当 I^2 = 75% ~100% 时，表明异质性明显，需探讨异质性来源，考虑进行亚组分析、meta 回归等，甚至放弃 meta-分析。

例 7-1　一项有关丹参治疗急性期脑卒中的系统评价，纳入了 5 个随机临床试验，结局观察指标为神经功能缺损改善率，资料见表 7-6。

表 7-6　丹参治疗急性期脑卒中的疗效观察

研究	干预组		对照组		合计(N_i)
	改善(a_i)	未改善(b_i)	改善(c_i)	未改善(d_i)	
Cao (1994)	48	1	45	2	96
Mao (2001)	33	3	25	9	70
Min (2004)	29	4	22	10	65
Pan (1992)	67	2	65	4	138
Zhang (2002)	36	4	29	11	80

以例 7-1 为例，Q 检验法的计算过程见表 7-7，其中 $\theta_i=ln\mathrm{OR}_i$，$w_i=1/Var(ln\mathrm{OR}_i)$，$Var(ln\mathrm{OR}_i)=1/a_i+1/b_i+1/c_i+1/d_i$。

表 7-7　Q 统计量计算过程(以 OR 为例)

研究	$ln\mathrm{OR}_i$	$Var_{(ln\mathrm{OR}_i)}$	w_i	$w_i*ln\mathrm{OR}_i$	$w_i*ln\mathrm{OR}_i^2$
CAO (1994)	0.758	1.543	0.648	0.491	0.372
Mao (2001)	1.376	0.515	1.942	2.672	3.678
Min (2004)	1.193	0.43	2.326	2.773	3.307
Pan (1992)	0.723	0.78	1.282	0.928	0.671
Zhang (2002)	1.228	0.403	2.481	3.047	3.741
合计	1.142		8.679	9.911	11.769

$Q=11.769-9.911^2/8.679=0.451$，$\upsilon=5-1=4$，$P>0.10$，异质性检验无统计学意义，还不能说明研究效应量间不同质，$I^2=0\%$。

需要注意的是 Q 检验的检验效能较低，如在纳入研究个数较少的情况下，有时不能检测出异质性，出现假阴性结果，可考虑提高检验水准，如 $\alpha=0.10$；相反，如果纳入研究过多，即使研究间结果是同质的，也可能出现 $P<\alpha$ 情况，即异质性检验有统计学意义。因此，对 Q 检验结果的解释要慎重，需要结合异质指数 I^2 以及森林图进行综合判断。

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

笔记

（四）合并效应量估计及其假设检验

在异质性检验的基础上，选用适当的方法进行分析。若异质性不明显，同时假定理论效应量为某一固定值，纳入研究效应量间的差异是由机遇造成的，可采用固定效应模型（fixed effect model）估计合并效应量；若存在一定程度异质性，且假定理论效应量不固定、服从于某种分布类型，如正态分布时，可用随机效应模型估计效应量；若异质性明显，可考虑亚组分析、meta 回归分析直至放弃汇总分析，只对结果进行简单描述。

以四格表资料为例，演示合并效应量估计及其假设检验过程。鉴于此类资料合并效应量的估计方法较多，诸如 Mantel-Haenszel 法、方差倒置法（inverse-variance methods）、Peto 法等。现仅以 Mantel-Haenszel 法为例（表 7-8）。

表 7-8　Mantel-Haenszel 法计算合并效应量

研究	a_id_i/N_i	b_ic_i/N_i	OR_i
Cao（1994）	1.000	0.469	2.133
Mao（2001）	4.243	1.071	3.960
Min（2004）	4.462	1.354	3.295
Pan（1992）	1.942	0.942	2.062
Zhang（2002）	4.950	1.450	3.414
合计	16.596	5.286	3.140

1. 估计合并效应量及其 95% 可信区间　$OR_{MH}=\frac{\sum a_id_i/N_i}{\sum b_ic_i/N_i}$，$OR_{MH}$ 对数的方差为 $Var(\ln OR_{MH})=\frac{\sum P_iR_i}{2(\sum R_i)^2}+\frac{\sum(P_iS_i+Q_iR_i)}{2(\sum R_i)(\sum S_i)}+\frac{\sum Q_iS_i}{2(\sum S_i)^2}$，其中 a_i，b_i，c_i，d_i 为四格表的实际频数，$R_i=\frac{a_id_i}{N_i}$；$S_i=\frac{b_ic_i}{N_i}$；$P_i=\frac{a_i+d_i}{N_i}$；$Q_i=\frac{b_i+c_i}{N_i}$，则 OR_{MH} 的 95% 可信区间为：$\exp(\ln OR_{MH}\pm 1.96\sqrt{Var(\ln OR_{MH})})$。

本例 $OR_{MH}=\frac{\sum a_id_i/N_i}{\sum b_ic_i/N_i}=16.596/5.286=3.14$，$ln OR_{MH}$ 的方差为：0.115，则 OR_{MH} 值 95% 可信区间为 1.62～6.10。

2. 合并效应量的假设检验　Z 检验。

$Z=\frac{\ln OR_{MH}}{\sqrt{Var(\ln OR_{MH})}}$，统计量 Z 服从于 u 分布（外文文献常表示为 Z 分布），用于检验合并效应量是否有统计学意义。本例 $Z=1.144/0.387=3.37$，则 $P<0.001$，表明合并效应量有统计学意义。

实际上，合并效应量估计以及异质性检验，可以借助一些现成分析软件来完成，方便快捷。这其中首推 RevMan 软件，图 7-4 则是利用该软件对例 7-1 资料的 meta-分析结果。

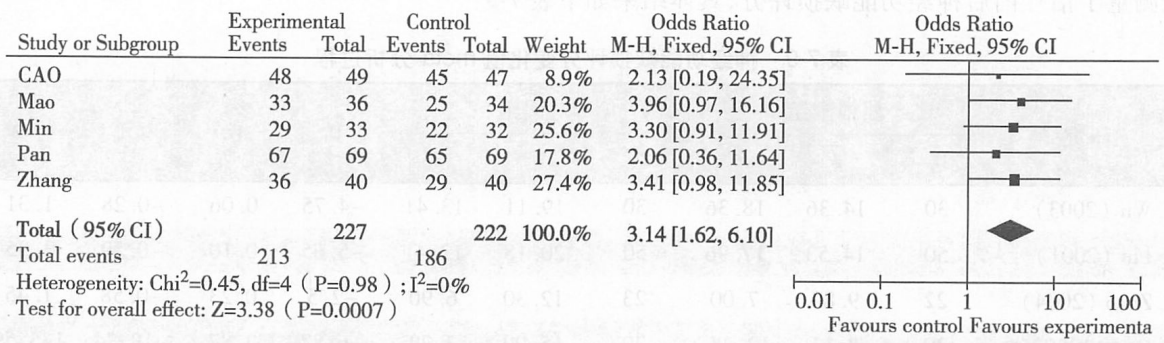


图 7-4　RevMan 计算结果

图中“♦”为合并效应量图示，“3.14(1.62,6.10)”表示合并效应量及其95%可信区间；“$Z=3.38, P=0.0007$”表示假设检验中的统计量及其P值。“$Chi^2=0.45, df=4, P=0.98$”，表示异质性检验的Q值及P值，异质指数$I^2=0\%$。

（五）固定效应模型与随机效应模型的选择

合并效应量的估计模型包括固定效应模型(fixed effect model)、随机效应模型(random effect model)以及最新提出的质量效应模型(quality effect model)等。模型的选择取决于异质性检验结果以及对效应量变异的理论假设。假如异质性检验无统计学意义且$I^2<40\%$，并假设总体效应量为一个固定值时，可认为理论效应量是固定的，原始研究间的效应量即使有差别，也是由于抽样误差造成的，合并效应量估计可选用固定效应模型；当异质性检验有统计学意义($P<0.10$)且$I^2>50\%$，若假设合并效应量不固定并服从于服从某种分布(常假定为正态分布)时，考虑选用随机效应模型，计算合并效应量。随机效应模型因将研究间的变异因子τ^2作为校正权重，其结果比固定效应模型结果更稳健，但可信区间的精度会有所降低；若异质性明显($I^2>75\%$)，考虑 meta 回归、亚组分析，探讨异质性来源；若临床异质性过于明显，则应放弃进行 meta-分析，仅作定性描述。

1. 固定效应模型　上例数据类型为典型的二分类变量资料，选用固定效应模型的 MH 法估计得到了合并效应量。若遇到数值变量资料(连续性变量资料)，且异质性检验无统计学意义时，同样可选用固定效应模型进行 meta-分析，具体过程与二分类变量资料相同，采用方差倒置法进行合并效应量估计。数值变量资料的效应量表达可以选择均数差值(mean difference，MD)和标准化均数差值(standardized mean difference，SMD)。当纳入研究的结果变量均采用相同方式测量与表达时，效应量表达可使用均数差值(MD)，因其带有自然单位，易于临床解释。而结果变量采用不同的度量衡单位或者效应量大小相差较大时，宜采用标准化均数差值(SMD)，但需谨慎解释这类结果。现以 MD 作为效应量，阐述固定效应模型的估计过程。

(1) 单个研究 MD 及其95%可信区间估计：MD 为均数差值，可表示为$d_i=\overline{x_i^t}-\overline{x_i^c}$，其中，$\overline{x_i^t}, s_i^t$为第$i$个纳入研究中干预组的均数和标准差，$\overline{x_i^c}, s_i^c$为第$i$个研究中对照组的均数和标准差，则$d_i$的标准误为$(s_i^t\times s_i^t/n_i^t+s_i^c\times s_i^c/n_i^c)^{1/2}$。

(2) Q检验及其I^2指数：$Q=\sum w_i d_i^2-\dfrac{(\sum w_i d_i)^2}{\sum w_i}$，进一步可计算异质指数$I^2$。

(3) 合并效应量(MD 合并值)及其95%可信区间：$d_{合并}=\dfrac{\sum w_i d_i}{\sum w_i}$，其95%可信区间为$d_{合并}\pm 1.96\sqrt{\dfrac{1}{\sum w_i}}$，其中$w_i$为$d_i$标准误的平方。

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

例 7-2　葛根素治疗急性脑梗死的系统评价中，纳入分析了4个随机临床试验，均使用 NIHSS 量表测量了治疗前后神经功能缺损评分，具体结果如下表7-9：

表 7-9　神经功能缺损评分变化值 meta-分析过程

研究	葛根素组			对照组			d_i	w_i	$d_i w_i$	$d_i^2 w_i$
	n_i^t	$\overline{x_i^t}$	s_i^t	n_i^c	$\overline{x_i^c}$	s_i^c				
Wu (2003)	30	14.36	18.36	30	19.11	13.41	−4.75	0.06	−0.28	1.31
Liu (2001)	50	14.53	17.96	50	20.18	12.41	−5.65	0.10	−0.59	3.35
Zhao (2004)	22	9.80	7.00	23	12.30	6.90	−2.5	0.23	−0.58	1.45
Cao (1999)	30	8.13	2.28	30	15.00	2.29	−6.87	2.87	−19.74	135.59
合计	132			133			−6.48	3.27	−21.19	141.70

笔记

本例合并 MD 为-6.48，其 95% 可信区间为-7.57 ~ -5.40。合并 MD 假设检验 $Z=11.72$，$P<0.00001$；异质性检验 $Q=4.39$，$P=0.22$，无统计学意义，$I^2=\frac{Q-(k-1)}{Q}\times100\%=31\%$。鉴于本例的异质性检验结果无统计学意义且异质指数<40%，考虑选用固定效应模型。同样，以该资料为例，可使用 RevMan 的固定效应模型分析，具体结果见图 7-5。

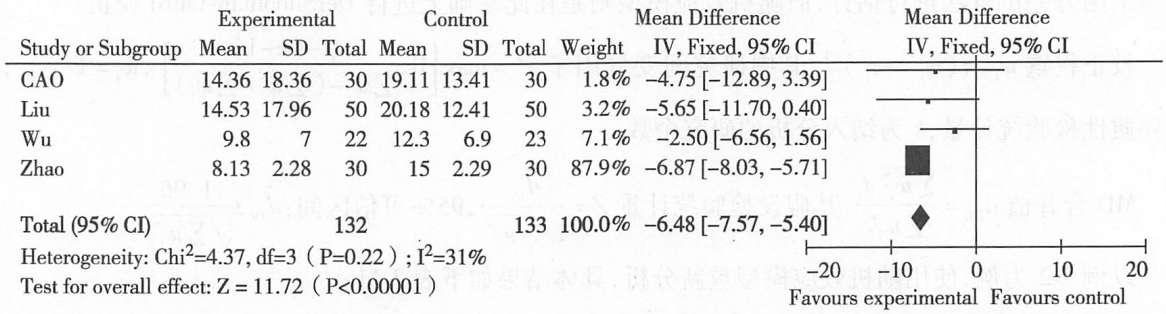


图 7-5　使用 RevMan 软件的 meta-分析结果

图中“◆”为合并效应量的图示，“-6.48(-7.57，-5.40)”表示合并效应量 $MD_{合并}$ 及其 95% 可信区间；“$Z=11.72$，$P<0.00001$”表示合并效应量的假设检验结果及其 P 值。“$Chi^2=4.37$，$df=3$，$P=0.22$”，表示异质性检验结果 Q 值及其 P 值。异质指数 $I^2=31\%$。由于 RevMan 在估计过程中对 MD 及标准误进行了校正，结果与人工计算略有差别。

2. 随机效应模型　当异质性检验有统计学意义且假定真实效应量不固定但服从正态分布时，考虑选用随机效应模型(random effect model)估计合并效应量。随机效应模型就是在固定效应模型分析的基础上采用了 DerSimonian-Laird 校正。两类模型的区别在于加权的方式不同，固定效应模型以每个研究内方差的倒数作为权重，而随机效应模型是以研究内方差与研究间变异之和的倒数作为权重，调整的结果就是样本量较大研究的权重适当降低，而样本量较小研究的权重则适当增大。

(1) 二分类变量资料的随机效应模型(random effect model)：该校正法是由 DerSimonian 和 Laird 于 1986 年提出，假设各原始研究的效应量不尽相同，以研究内方差及研究间变异之和的倒数为权重，并以此估计合并效应量。与固定效应模型相比，主要步骤相同，依次估计单个研究效应量、合并效应量及其 95% 可信区间，最后进行假设检验。唯一不同的是需事先计算研究间变异因子 τ^2。校正权重 $w_i^*=(w_i^{-1}+\tau^2)^{-1}$。其中 $\tau^2=\max\left[0,\frac{Q-(k-1)}{\sum w_i-(\sum w_i^2/\sum w_i)}\right]$，$Q$ 为异质性检验统计量，k 为纳入分析的研究个数。若 $Q<k-1$，$\tau^2=0$，若 $Q>k-1$，$\tau^2=\left[\frac{Q-(k-1)}{\sum w_i-(\sum w_i^2/\sum w_i)}\right]$。

仍以四格表资料为例，使用随机效应模型估计合并效应量及其 95% 可信区间：

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

以例 7-1 为例，$Q=0.451<4$，故 $\tau^2=0$，随机效应模型具体估计过程如下(表 7-10)：

表 7-10　随机效应模型的分析结果

研究	OR_i	$\ln OR_i$	w_i	τ^2	w_i^*	$\ln OR_i\times w_i^*$
Cao (1994)	2.133	0.76	0.648	0	0.648	0.491
Mao (2001)	3.960	1.38	1.943	0	1.943	2.674
Min (2004)	3.295	1.19	2.326	0	2.326	2.774
Pan (1992)	2.062	0.72	1.282	0	1.282	0.927
Zhang (2002)	3.414	1.23	2.480	0	2.480	3.045
合计	3.140	1.14	8.679		8.679	9.911

NOTE

其中 $OR_{合并}$ 为 exp(9.911/8.679)=3.133，其95%可信区间为1.62~6.10。由于固定效应模型以各研究内方差的倒数为权重，而D-L校正法则以研究内和研究间变异之和的倒数为权重，所以当异质性不明显时，两种模型估计结果完全相同；若存在较明显的异质性时，结果会有差异，利用随机效应模型估计可信区间明显宽于固定效应模型的估计结果。

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

校正权重 $w_i^* = (w_i^{-1} + \tau^2)^{-1}$，其中研究间变异因子 $\tau^2 = \max\left[0, \frac{Q-(k-1)}{\sum w_i - (\sum w_i^2 / \sum w_i)}\right]$，$w_i = 1/v_{(di)}$，$Q$ 为异质性检验统计量，k 为纳入分析的研究个数。

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

以例7-2为例，使用随机效应模型重新分析，具体结果如下表7-11：

表7-11　神经功能缺损评分变化值 meta-分析过程

研究	葛根素组			对照组			d_i	w_i	τ^2	w_i^*
	n_i^t	$\overline{x_i^t}$	s_i^t	n_i^c	$\overline{x_i^c}$	s_i^c				
Wu（2003）	30	14.36	18.36	30	19.11	13.41	-4.75	0.06	1.917	0.05
Liu（2001）	50	14.53	17.96	50	20.18	12.41	-5.65	0.10	1.917	0.09
Zhao（2004）	22	9.80	7.00	23	12.30	6.90	-2.5	0.23	1.917	0.16
Cao（1999）	30	8.13	2.28	30	15.00	2.29	-6.87	2.87	1.917	0.44
合计	132			133			-5.63	3.27		0.74

本例 $Q>k-1$，$\tau^2 = (4.39-3)/(3.27-8.322/3.27) = 1.917$，合并均数差值：-5.63；合并效应量假设检验：$Z=4.87$，$P<0.001$，说明有统计学意义；其95%可信区间为-7.90~-3.37。同样可以利用 RevMan 完成上述分析，相对于固定效应模型的结果（-7.57~-5.40），显然随机效应模型的区间范围变宽。

随机效应模型可以允许研究间效应量存在一定的异质性，并试图用特定的权重系数来解释这些异质性，所有纳入研究均使用相同的权重系数进行权重校正。但备受争议的是这些权重系数也许并无实际意义，所有研究都采用同一权重系数的做法也值得商榷。为此，Doi 和 Thalib 两位研究者提出的质量效应模型（quality effect model）可以较好地解决上述问题，他们认为异质性大多与原始研究的方法学质量有关，提出了偏倚风险概率（Qi），以其作为权重校正系数，Qi 取值范围为0~1，若 Qi 为0，表示该原始研究质量高，无偏倚发生风险。目前已有一款免费软件 MetaXL，可以利用质量效应模型估计合并效应量。

三、发表性偏倚的识别与分析

meta-分析为一种二次研究方法，即基于原始研究结果进行二次分析。纳入的原始研究是否全面无偏，将直接影响 meta-分析结果是否真实可靠。在可能影响 meta-分析结果真实性的偏倚中，发表性偏倚的影响程度较大且较难控制，因而备受关注。发表性偏倚可使 meta-分析过分夸大治疗效应量或危险因素的关联强度，导致临床个体治疗与卫生决策的失误。

发表性偏倚通常是指有统计学意义的研究结果比无统计学意义的研究更容易投稿和被发表，由此而产生的偏倚。对于无统计学意义的研究，研究者可能认为意义不大，不发表或推迟发表；作为杂志编辑则更有可能对这类论文退稿。因为存在发表性偏倚，即使具备周密的检索策略和手段（如与研究者个人联系），也不可能完全地纳入所有相关研究。

现有三类比较简单的分析方法：即漏斗图法、剪补法以及公式法可以用来正确识别与处理发表性

笔记

偏倚。其中以漏斗图法最为常用，它是基于样本含量（或效应量标准误的倒数）与效应量（或效应量对数）所绘制的散点图。效应量可用 RR、OR、RD 或者 RR、OR 的对数值等。漏斗图的前提假设是效应量估计值的精度随着样本量的增大而提高，其变化范围也随精度的增加而逐渐变窄，最后趋近于点状，其形状类似一个对称倒置的漏斗，故称为漏斗图。即样本量小的研究，数量多、精度低，分布在漏斗图的底部呈左右对称排列；样本量大的研究，精度高，分布在漏斗图的顶部，且向中间（合并效应量）集中。利用漏斗图可以直接观察原始研究的效应量估计值是否与其样本含量有关。当存在发表性偏倚时，漏斗图往往呈现不对称的偏态分布（图 7-6）。但绘制漏斗图，需要纳入较多的研究个数，原则上要求 5 个以上才能进行。

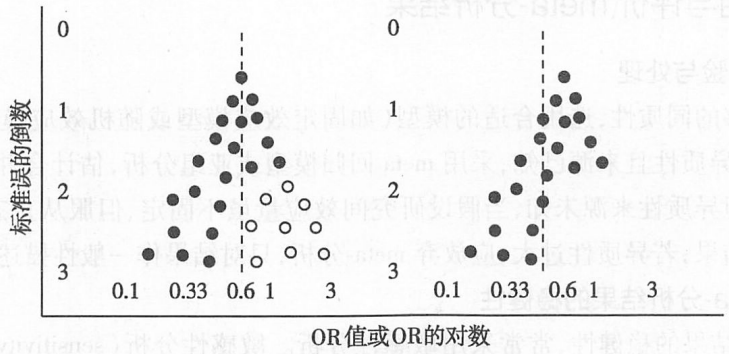


图 7-6　漏斗示意图

图 7-6 所示假设为漏斗图的两种情况，左图中所有研究围绕中心线对称排列，表明没有发表性偏倚，图中空心散点代表结果无效的小样本研究，小样本研究估计的效应量变异较大，出现效应量极端值机会要多于大样本研究；右图，呈不对称分布，表示存在发表性偏倚，所缺失部分恰恰为结果无统计学意义的小样本研究。基于例 7-1 资料，利用 RevMan5 绘制漏斗图如下（图 7-7）：

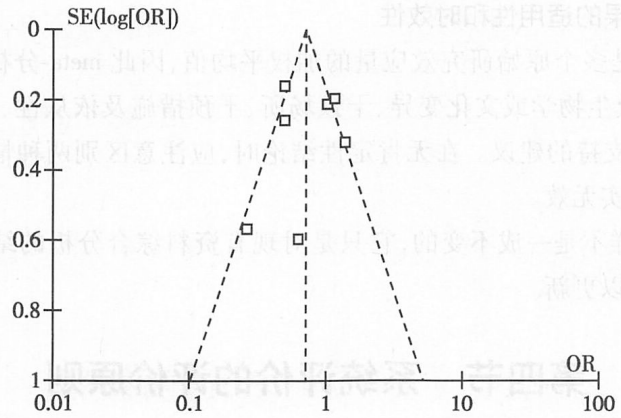


图 7-7　例 7-1 漏斗图

图 7-7 显示所有研究围绕中心线对称排列，表明发表性偏倚不明显，对 meta-分析结果的影响可以忽略。

除漏斗图法外，也可以进行 Egger 回归、Begger 分析以及计算失效安全数（fail-safe number）等，用以评估发表性偏倚。前者分析需要借助一些特定的软件，如 Stata 等。

四、meta 回归及其他方法学进展

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

及病例-对照研究等研究类型，也可用于敏感性分析。但 meta 回归容易产生聚集性偏倚，特别是当资料不齐或纳入分析的研究数目较少时，如小于 10 个时，不宜进行 meta 回归分析。

尽管上述回归模型中考虑了一些混杂因素，仍不能完全解释研究间的变异，可进一步在模型中加入随机效应项，那么该模型就成为混合效应模型。参数估计可采用加权最小二乘法或极大似然估计法，可以用来解释已知的异质性来源。但存在两大缺点：一是如果的研究个数目较少，如小于 10 个，则不能建立混合效应模型；二是不能进行剂量反应回归分析等。

除 meta 回归模型与混合效应模型外，其他相关的方法还有：累积 meta-分析、迭代随机效应模型、多水平 meta 模型以及贝叶斯 meta-分析等。

五、慎重应用与评价 meta-分析结果

（一）异质性检验与处理

若研究间有足够的同质性，选用合适的模型（如固定效应模型或随机效应模型，两者均可）估计合并效应量；若存在异质性且来源已知，采用 meta 回归模型或亚组分析，估计合并效应量。若异质性检验有统计学意义但异质性来源未知，当假设研究间效应量虽不固定、但服从正态分布时，应选择随机效应模型的估计结果；若异质性过大，应放弃 meta-分析，只对结果作一般性描述。

（二）考察 meta-分析结果的稳健性

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

（三）meta-分析结果的适用性和时效性

合并效应量实际上是多个原始研究效应量的加权平均值，因此 meta-分析的结果在推广应用时，应注意个体对象的特征及生物学或文化变异、干预场所、干预措施及依从性、有无辅助治疗等。不宜推荐没有 meta-分析证据支持的建议。在无肯定性结论时，应注意区别两种情况，是证据不充分而无定论，还是有证据表明确实无效。

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

第四节　系统评价的评价原则

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

系统评价/meta-分析的质量评价包括两方面：①方法学质量评价，评价工具包括 OQAQ（Overview Quality Assessment Questionnaire）、SQAC（Sacks'Quality Assessment Checklist）和在前 2 个工具基础上制订的 AMSTAR（Assessment of Multiple Systematic Reviews）等。AMSRAR 是 2007 年由来自荷兰 VU 大学（Vrije Universiteit，阿姆斯特丹自由大学）医学研究中心和加拿大渥太华大学的临床流行病学专家联

笔记

合发表，共 11 条，每一条评价结果表示为“yes”“No”“Can’t answer”和“Not applicable”。2017 年推出了 AMSTAR 2，共 16 个条目，评价了系统评价制作的选题、设计、注册、数据提取、数据统计分析和讨论等全过程。简化了对每一条的评价结果，只判断为“yes”“partial yes”“No”，且有全面的使用指南。②报告质量评价，评价工具包括 PRISMA（Preferred Reporting Items for Systematic Reviews and Meta-Analyses，主要针对干预性研究的系统评价特别是 RCT 的系统评价，也可用于其他研究类型的系统评价）和 MOOSE（Meta-analysis Of Observational Studies in Epidemiology，针对观察性研究的系统评价/meta-分析）等。

应用系统评价/meta-分析结果解决临床问题不仅要评估其方法学质量以明确结果的真实性，还要明确结果的临床重要性和适用性。因此，评价系统评价应包括①真实性；②临床重要性；③适用性 3 方面，评价干预性研究系统评价的基本原则如下所述。

一、系统评价的结果是否真实

（一）是否根据随机对照试验进行的系统评价

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

（二）是否采用广泛和详细的检索策略检索相关文献

从作者报告的文献检索方法中可明确收集的文献是否全面。由于标识不完整，一般的文献检索数据库如 Medline 仅能检出库中收录随机对照试验的 50%，而发表偏倚即阳性结果的文章更易发表的现象可能导致系统评价出现假阳性结果。因此，全面的文献检索应包括手检相关杂志、检索会议论文集、学位论文、厂家数据库和与已发表文献作者联系。此外，如果文献检索时限制语种，也可能影响系统评价结论。目前，多数杂志均要求系统评价作者按照 PRISMA 声明规范报告系统评价和 meta-分析全文，其中包含检索流程图，要求详细陈述检索结果和筛选流程，有助于读者判断检索的完整性和筛选的合理性。收集的文献越系统、全面，则结论受发表偏倚的影响就越小，可信度就越大。

（三）是否评估纳入的单个研究的真实性

由于系统评价多为对原始文献资料的再分析和总结，因此，除了进行系统评价的方法要严格外，原始文献的质量非常重要。所以文中应详细描述评价单个研究文献质量的方法，最好为多人独立评价并有良好的一致性。

（四）是否采用单个病例资料（或每个研究的合成结果）进行 meta-分析

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

二、系统评价的结果是否重要

（一）不同研究的结果是否一致

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

（二）系统评价的疗效大小，疗效的精确性如何

在进行结果合成时，不能通过简单比较阳性研究结果和阴性研究结果的研究个数来确定系统评价的结论，而应该根据研究的质量和样本含量的大小对不同研究给予不同的权重值，并采用恰当的指

标如比值比(odds ratio)、相对危险度(relative risk)、均数差值(mean difference)、防止一例事件发生需要治疗同类患者的例数(NNT)和统计方法如随机效应模型和固定效应模型等合成结果,同时计算相应的可信区间。

三、系统评价的结果是否能应用于我的患者

系统评价报告的结果是所有研究对象的"平均效应",而你主管的患者不一定在研究对象的范围内,因此在考虑系统评价的结果能否应用于你主管的具体患者时应从四个方面进行考虑:

(一) 我的患者是否与系统评价中的研究对象差异较大,导致结果不能应用于我的患者

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

(二) 系统评价中的干预措施在我的医院是否可行

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

(三) 自己的患者从治疗中获得的利弊如何

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

(四) 对于治疗的疗效和不良反应,自己患者的价值观和选择如何

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

总之,采用科学、严谨的方法生产的系统评价能为临床医疗实践、医学教育、医学科研和卫生决策提供真实、可靠的信息。作为最高级别的研究证据,系统评价对科学决策是必要的,但非唯一的参考。决策需要同时考虑当地实际情况、资源的可获得性、患者的具体特征、意愿和选择等,并在应用系统评价时严格评价其真实性、重要性和适用性。

(李静　康德英)

笔记

第八章　病因与危险因素的研究、评价与循证实践

病因学研究一直都是疾病探索的一个重要领域。基础、临床和预防医学各学科专家都致力于疾病病因的研究，但研究病因的方法、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完全一致，因此，本章在病因（etiology）、发病机制（pathogenesis）、致病机制（mechanisms）、危险因素（risk factor）等多个标题下均会出现有关病因的论述。在临床实践中的任何干预措施，包括手术和药物，都可能引起不同程度的不良反应，而不良反应的判定，实质上也是因果推断，这里的“因”是指临床医生在疾病预防、诊治过程中采用的各种措施和方法，如诊断技术、手术和药物等。

第一节　病因与危险因素研究的意义

病因学研究就是探索疾病发生的原因及相关因素间的相互效应以及各因素对疾病发生、发展的影响。

病因学研究属于因果关系研究，因此任何与结局（疾病）有关联的暴露因素都必在结局发生之前存在，但发生在结局（疾病）之前的暴露未必都是病因。病因学研究是正确认识疾病发生和流行规律的必要条件，也是临床正确诊断、有效防治的科学基础，可为临床决策提供依据，使疾病防治获益的最大化。其研究意义具体表现在以下两个方面。

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，探明可能致病的直接病因和危险因素，可以帮助医生对患者进行正确地诊断和治疗。疾病需要对因治疗，精准施治，从而获得好的疗效。

二、有助于疾病的预防

病因学研究结果不仅有助于临床诊断与治疗，同时对于疾病预防和控制也具有重要意义。通过各种传媒，对社区人群进行健康教育，让社区人群及时了解疾病的病因和危险因素，可提高对疾病病因和有效预防措施的认知程度，并有意识地改变自己的生活方式和行为，从而在群体层面上预防疾病的发生。

第二节　病因与危险因素研究的概述

人类对疾病病因的认识是一个循序渐进的过程，先后经历了单一病因论到多病因论，由机械决定论到概率论的发展过程。

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，中国古代哲学思想中衍生出了阴阳五行学说——金、木、水、火、土，这一学说被后人用来去解释人体生理现象和病理变化的规律，即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。他认为，疾病的发生与水源、土壤及风等因素有关，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，古代国内外医家们都注意到疾病与环境有密切关系。19世纪上半叶，Sydenham等提出的疾病“瘴气学说（miasma theory）”在西方盛行，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。

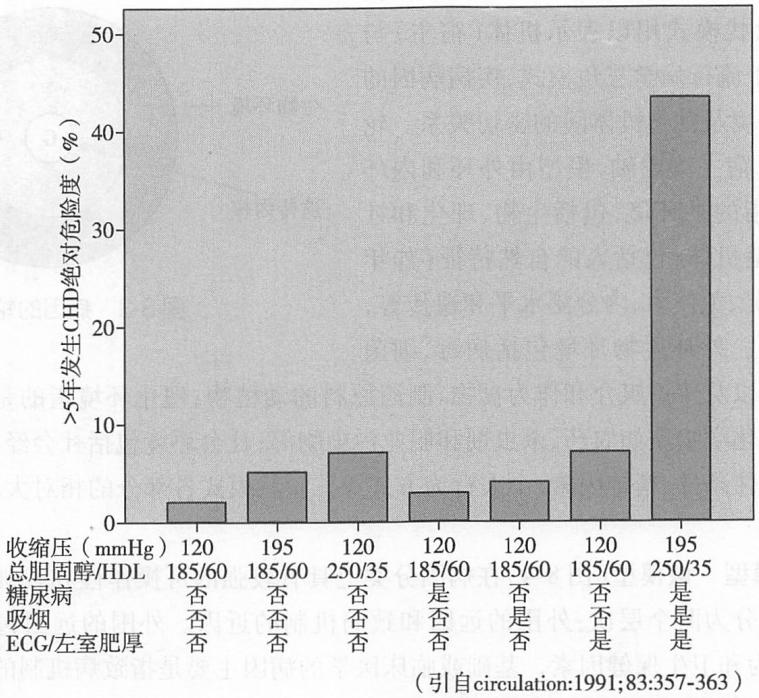
近代有关病因学理论，最早提出的是单一病因论，后续又出现了多病因论和概率论因果观等。

1. 单一病因论 意大利Fracastoro（1479—1553）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序幕。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“推断致病微生物导致特异性疾病”的Henle-Koch标准，奠定了人类病因学研究理论的基础。该法则共有4条：①病原微生物必然存在于患病动物体内，而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标准，随后许多传染性疾病也得到了证实。

尽管该标准有些绝对化，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各自特异的病原微生物。随着19世纪末期微生物学的出现和发展，逐渐形成了疾病单一病因论的“特异病因学说”。即每一种疾病必定是某一种特异的病原物所致。但该学说无法解释复杂的病因效应，因为即使是单一的病因，也可引起多种疾病，更不用说绝大多数疾病的发生是多种因素（包括社会和环境等因素）共同作用的结果，因此，单一病因论有其明显的局限性。

2. 多病因论 在长期疾病防治实践中人们逐渐发现，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中发现，一种疾病的发生往往是多种因素综合作用的结果，多种致病因素同时存在的危害要比单一因素更为严重，这是因为在致病效应上会产生交互作用。即使是传染性疾病，也是多种因素共同作用的结果。如肺结核的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、居住环境等因素均参与人体肺结核的发生过程。再比如冠心病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，探讨这些因素对55岁男性发生冠心病的结果（发病率）（图8-1），可见具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。

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，其间走过了一个漫长的历程。然而上述因果观仍难以完美阐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定律（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。



（引自circulation:1991;83:357-363）

图 8-1　冠心病的多病因交互效应图

Lilifield 从概率论的因果观层面阐述了流行病学的病因概念：即“那些能使人群中发病概率升高的因素就可以认为是病因，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”。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），这无疑又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。

（二）病因模型

病因模型以简洁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，展示了因果关系的思维框架。然而由于对因果关系有不同的理解或不同的侧重，出现了多种类型的病因模型，其中具有代表性的有 3 类，即生态学模型、疾病因素模型和病因网络模型。

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。常见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动因、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；若某个因素发生变化，三者失去平衡，则引起疾病。

在病因学研究中，尤其是传染病病因学研究，流行病学三角模式得到了广泛的认同。该模式（图 8-2）认为疾病的发生必须有病原物（动因）、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，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，流行病学专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。

图 8-2　流行病学三角模式

NOTE

（2）轮状模型（wheel model）：1973 年，Susser 提出了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 8-3）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。

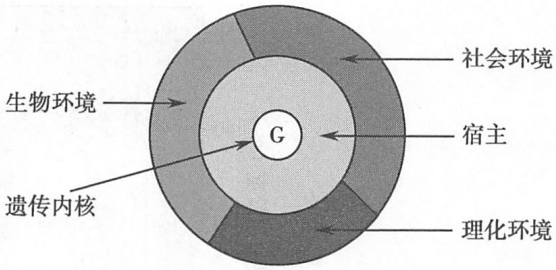


图 8-3　病因的轮状模式

2. 疾病因素模型　该模型（图 8-4）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，而临床流行病学的病因学研究则是以临床为基础，兼顾近因与远因。

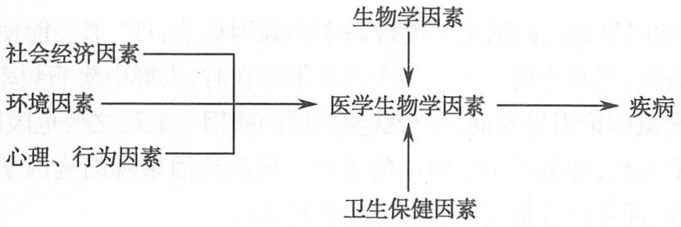


图 8-4　疾病因素模型

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），多个病因链交错联接构成病因网（web of causation），呈现了因果关系的完整路径。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，其复杂程度远超我们的想象。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，已发现越来越多的疾病是多病因性的，而且存在一病多因、一因多病和多因多病等复杂情况。多病因学说已被医学界所接受，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，病因有多种分类方法，常见的分类有：

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，因而任何疾病的发生也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。倘若缺乏该因素即不会引起某疾病，则该因素为该病的必要病因。如结核杆菌就是结核病的必要病因，无结核杆菌感染就不会发生结核病。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它们的必要病因。鉴于大多数慢性非传染性疾病是多因性的，此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，某疾病必定（概率是100%）会发生。对充分病因的理解应注意三点：①对大多数疾病而言，充分病因不是一个，而是一组因素集合（组分病因）。如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，大多数结核杆菌感染者因自身抵抗力强而不发生结核病，只有结核杆菌感染加上机体特异性和非特异性抵抗力的降低时，才成为结核病的一个充分病因。②对于大多数疾病而言，其充分病因目前仍未明了，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（组分病因）不止一个，有的可能有多个充分病因，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”，但有的高血压患者并不超重，提示导致这部分患者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打开开关则电灯发光，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，这些环节的任何一个都与开关同样重要，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协调的因素共同组成的，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。这些组成充分病因必不可少的部分称为组分病因或成分病因（component causes）。

2. 直接病因与间接病因　基于病因链和病因网络模型，引起疾病的诸多因素有时按顺次起作用，即病因1导致病因2，最终导致疾病。可简要表示为：病因1 ⟶病因2 ⟶疾病。这里，病因2称为直接病因（direct causes），病因1称为间接病因（indirect causes）。直接病因是指只有该病因作用于人体才能够引起发病，对应于上述疾病因素模型中的近因。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，对应于疾病因素模型中的远因。例如：营养不良、居住条件差、机体免疫力低下、社会经济环境恶化等都可能造成疾病的易感性增加，这些因素被称之为间接病因。

以世界卫生组织（WHO）2005年提出的主要慢性病的病因链为例（图8-5），高血压、高血糖、血脂异常和超重/肥胖这些因素就是所谓的直接病因或称之为“近端病因”，在病因链上距离疾病结局近，病因学意义相对明确。但值得注意的是越靠近疾病结局近端的因素，涉及的人群面越窄，预防的机会越少。而从个体层面来看，不合理膳食及过多的能量摄入、体力活动少、吸烟则是上述直接病因共有的、最重要的、可以改变的危险因素，有效干预这三种危险因素可以预防80%的心血管疾病、2型糖尿病以及40%的肿瘤；再往病因链更远端看，还有“病因的原因”，即根本的社会经济、文化、政治和环境因素，称之为“健康社会决定因素”（social determinants of health，SDH）。此类远端影响因素作为间接病因，与疾病的因果机制可能不是那么明确与直接，但涉及的人群面广，预防机会大，通过改善这些因

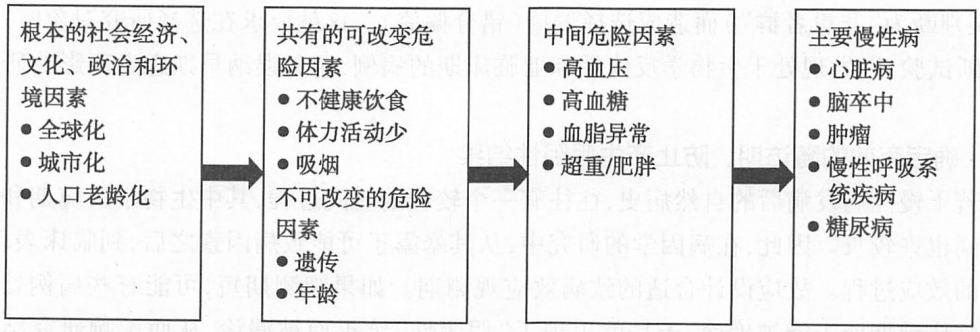


图8-5　主要慢性病病因链

素对于降低总疾病负担的预防效率会很高。这些关于病因的认识和探讨势必会对疾病防治策略的调整产生深远的影响。

3. 危险因素　目前，慢性非传染性疾病已成为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，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。美国流行病学家 MacMahon 指出，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。危险因素的概念无疑体现了概率论的因果观。

二、寻找病因的指南清单

临床医生在进行病因学研究之前，一般须考虑选择对人类健康威胁较大、有临床需求的疾病、病因不明的疾病进行研究，才更具有现实和科学意义。上述病因模型及病因分类有助于寻找疾病病因的大致方向、类别或因果联接方式（病因网），但这些模型相对比较抽象，在病因学研究实践中，为了增加寻找病因的可操作性，可借鉴表 8-1 的寻找病因指南清单。

表 8-1　寻找病因指南清单

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、既往史
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期的表现千差万别。了解疾病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：

（一）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（导致错误判断为“非患者群”）而造成选择偏倚（错分偏倚）。这就要求在选择研究对象时，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的病例，避免误纳早期患者而影响研究结果的真实性。

（二）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，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，从而影响研究结果的真实性。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，以助于

早期或不典型病例的确诊。

第三节　疾病病因与危险因素研究的基本过程与方法

病因与危险因素作为重要的临床问题之一，可以根据对疾病的认识和掌握资料的程度循序渐进、分阶段进行研究。基本过程和常用方法参见下图（图 8-6）：

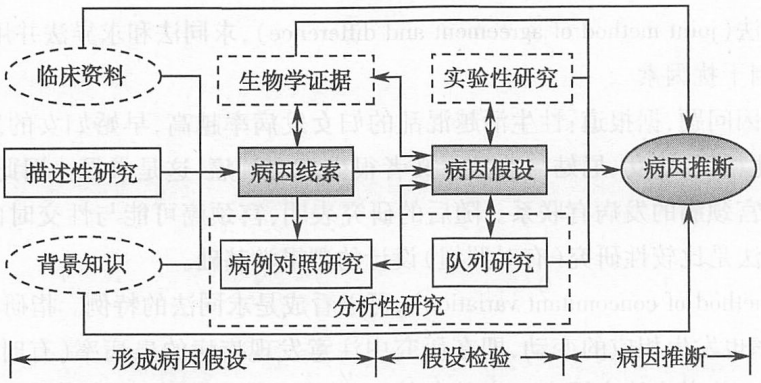


图 8-6　病因与危险因素研究的基本过程

一、提出病因假设

（一）假设的作用和特点

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，根据其基本背景知识、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。病因如果不明，会为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。

对不明病因疾病进行病因学研究，首先必须面对被研究的疾病，依据它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，以及现有水平的各种化验、检查结果，作出一系列的排除诊断，进而在检索相关文献、系统综合分析的基础上，提出可能的“病因假设”。

例如：我国华北地区大面积调查食管癌的流行特征后，通过资料分析与总结后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；同时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，有学者提出食管癌的水土病因学说，随后基于这个假设，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，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，需借助逻辑方法。假设形成过程中常用的逻辑方法主要是归纳演绎法，这对于病因研究的因果假设有重要理论和实际指导价值。

1. 归纳法（Mill 准则）　Mill JS 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，其中的科学实验四法常被用于分析流行病学研究中形成假设、设计研究方案和进行病因推断。后人在科学实验四法的基础上将同异并用法单列，发展为科学实验五法。

（1）求同法（method of agreement），又称契合法或“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。

(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、尼姑与独身主义者很少患宫颈癌,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)波动时有那些因素在变动。共变法的应用有一定的条件,只有当有关(暴露)因素不是定性的,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,肺癌的比值比(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),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:Mill 准则对列出病因假设清单并不能提供指导,如果病因假设清单中未包括真实的病因,Mill 准则就不能提供任何帮助,而更为遗憾的是,要寻找的"那个"因素是否在清单中也是难以判定的。

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结果表明,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,从而为后续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。这一假设的形成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),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,反推是归纳性的。从一个假设可推出多个具体证据,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,还需考虑其他影响因素的存在。

二、验证病因假设

科学的假设是在一定的经验材料和科学理论基础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,对科学理论的理解和思维方式不同等,常常会出现不同的假设。不仅不同学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,在不同时期也会提出不同、甚至差别很

笔记

大的假设，在科学上出现不同的假设，促进对一个问题的全面认识是一种很正常的现象，也是科学进步的重要标志。

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。

（一）假设的验证需进一步深化

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。如果是证实了自己研究的假设（如未知病因），那就应该进行全面研究，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。

（二）假设验证的结果应区别对待

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。对于科学发展来言，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，也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，实事求是，具体分析。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。如果出现虽屡遭失败，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。

（三）验证假设的注意事项

验证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，第一为通过验证，被确认为真理；第二是受有关偏倚、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，应避免或防止出现第二、三种情况。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，从背景知识或临床或流行病学的角度，总归会发现病因线索并提出病因的假设，也许这种假设已存在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，必须通过科学的分析性或试验性的研究方可获得病因学的真实结论。

病因学研究往往需要多学科和多专业的协作和参与，才能取得病因学研究（病因假设的验证）的真正成功。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 8-2）。

表 8-2　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	+
描述性研究	前瞻/回顾	好	±

（一）病例对照研究

病例对照研究是分析流行病学中最基本的研究类型之一，性质上属回顾性研究。它是以确诊的

患有某病的患者作为病例组，选择具有可比性、未患该病的人群作为对照组，然后比较分析组间差异，借助相应的统计学指标，探讨研究因素与疾病的关系。例如：HBsAg 与肝癌相关性的非匹配设计的病例对照研究，结果见表 8-3。

表 8-3 HBsAg 与肝癌关系的病例对照研究

	病例组（肝癌）	对照组（非肝癌）	合计
HBsAg 阳性	35	8	43
HBsAg 阴性	15	42	57
合计	50	50	100

OR=(35×42)/(15×8)=14，结果显示 HBsAg 与肝癌发生有联系。

由于病例对照研究是从“果”到“因”的回顾性研究，且存在未知的混杂因素，论证强度相对较低。

（二）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，可以直接验证因果关联，具有较强论证强度。该研究方法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，自然分为暴露与非暴露组，随访观察一段时间后，进而比较目标疾病的发病率。由于符合先因后果的时间顺序，可以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（表 8-4）。

表 8-4 高胆固醇血症致冠心病因果效应分析表

	冠心病		合计	发病率（%）
	有	无		
高胆固醇队列	83（a）	867（b）	950（$a+b$）	8.7
正常胆固醇队列	23（c）	917（d）	940（$c+d$）	2.5
合计	106（$a+c$）	1784（$b+d$）	1890（N）	

$$\text{相对危险度(RR)}=\frac{\frac{a}{(a+b)}}{\frac{c}{(c+d)}}=3.50(95\%\text{CI}:3.48\sim3.52)$$

$$\text{归因危险度(AR)}=8.7\%-2.5\%=6.2\%$$

$$\text{病因学分数(EF)}=\frac{\text{RR}-1}{\text{RR}}=0.71$$

$$\text{NNH}=\frac{1}{\text{AR}}=16$$（表示高胆固醇血症 5 年中每 16 人就发生 1 例冠心病）。

结论：高胆固醇血症系冠心病发病的危险因素，其 RR=3.5（95% CI：3.48～3.52）。

其中归因危险度也称特异危险度（attributable risk，AR），又称率差（rate difference，RD）。临床试验中常把 AR 称为绝对危险降低率（absolute risk reduction，ARR）。NNH（number needed to harm，可译为“害-需暴露人数”），具体概念详见本教材相关章节。

（三）随机对照试验

随机对照试验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。除随机对照试验（RCT）外，也可采用非随机试验的设计模式。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，追踪乙型病毒性肝炎

笔记

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 与肝癌的因果关系。

（四）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、分子生物学实验、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生物学证据。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，实验结果外推到人的这个过程需要谨慎。

四、病因学因果关联的统计学指标

用于病因学因果关联分析的常用指标如下：

1. **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。

2. **归因危险度（attributable risk，AR 或称绝对危险度 absolute risk，AR，又称率差 rate difference，RD）**　即暴露可疑病因组人群和非暴露可疑病因组的人群，他们各自发病率相减之绝对的差值。如 A 组发病率 25%、B 组 8%，则其归因危险度为 25% -8% = 17%。

3. **相对危险度（relative risk，RR）**　即可疑病因暴露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。

4. **相对危险增高度（relative risk increase，RRI）或称病因学分数（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，即绝对危险度与暴露组发病率的比值。

5. **NNH（number needed to harm，害-需暴露人数）**　系指需要多少例接触致病因素后才出现一例发病，用 1/AR 值表示。NNH 是临床和卫生决策十分有用且容易理解的指标。在疗效评价不良反应研究中，可称之为害-需治人数，其对应于 NNT（number needed to treat，益-需治人数）。

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。在此需要提醒的是，在不同的流行病学或循证医学教材或著作中各指标或有不同的称谓，读者应予注意。

五、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，可以表现为夸大或掩盖了暴露因素与疾病之间的关联强度，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危险的高度关联的结论。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，但由于两者有共同的原因 C，因此观察到 A 和 B 与 C 均存在着关联。这也是在研究中应注意要避免的。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，属于继发关联，两者均受到卫生状况及个人卫生习惯的影响。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。

当暴露因素与疾病既存在直接关联，又存在间接（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，又存在间接关联。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。

（三）因果关联（causal association）

统计学关联是判断因果关联的前提，但只有部分统计学关联属于因果关联。因果关联有直接关联和间接关联之分，随着研究的深入，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。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为病因。图 8-7 概括了进行病因推断前必须考虑的问题和步骤。

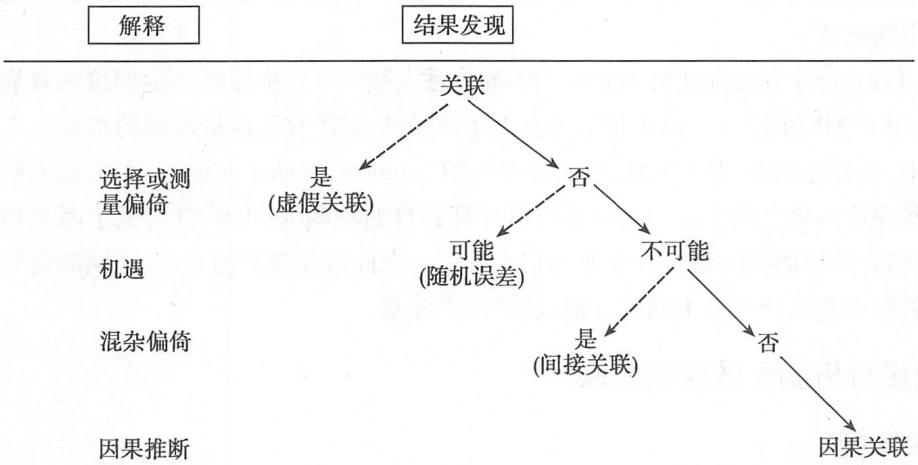


图 8-7　因果关联的判断进程

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 20 世纪 60 年代初期关于海豹短肢畸形与孕妇服用沙利度胺（Thalidomide）关系的研究，这种出生缺陷先在前联邦德国，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“沙利度胺”，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，才被最终认定。

第四节　病因与危险因素研究的评价原则

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 8-5 所列是病因学研究的评价原则，可作为因果关系推断的参考标准。

笔记

表 8-5　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、NNH 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?

三、研究结果的适用性评价
1. 你的患者与文献中的研究对象是否存在较大的差异,以至于文献的研究结果不能回答患者的疑问?
2. 你的患者发生疾病/不良反应的危险性有多大?
3. 确定你的患者喜好和希望解决的问题?
4. 是否应终止接触暴露因素或更改治疗措施?

一、真实性评价

(一) 是否采用了论证强度高的研究设计方案

不同的病因学研究方法,其因果关系论证强度是不同的,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 8-2 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率下降或流行随之减少,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，是否采用了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，导致以仍然留在研究中的研究对象的分析结果不能真实地反映客观事实。

（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、回顾性的以及描述性研究则很难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敏感、特异的测量指标，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及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相似的研究方法，并采用相关评价指标进行的病因研究，所得到结论如果均一致，那么该病因致病的结论就较为可靠，循证医学开展的系统评价所提供的信息是评价结论是否一致的重要依据，是病因推断的一个重要指标，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，世界上已有众多病例对照研究和至少 7 次以上较大规模的队列研究结果均表明，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映真正的因果关联，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、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，那么应该进一步地评价这个研究结果是否具有重要的临床或公共卫生学意义和价值，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、NNH 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”。若调整后的 RR 或 OR 较调整前的 RR 或 OR 有明显变小，则应该怀疑原有结果。相反，若调整后的 RR 或 OR 与调整前相比不变或明显增大，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）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 等的 95% 可信区间，如果 95% CI 的范围较狭小、则其精确度就高，下限和上限值不包括 1.0，表明有统计学意义。如果研究发现暴露因素与疾病、暴露因素与疾病/不良反应存在相关性，则可信区间下限表示相关强度的最低值，如果研究未发现两者的关系（没有统计学意义），则可信区间的上限表示相关强度可能达到的最高水平。

（三）害-需暴露人数

害-需暴露人数（number needed to harm，NNH）是指多少人暴露于危险因素/干预措施，就可能比非暴露组多发生 1 例疾病/不良反应。NNH 是一种更直观、更易被临床医生理解的表示因果相关强度的指标。NNH 为 AR 的倒数，即：NNH = 1/AR。在疗效评价不良反应研究中，则可称之为害-需治人数。在疗效评价研究中，其相对应的指标为 NNT（number needed to treat，益-需治人数）。

三、适用性评价

在评价了病因和危险因素研究结果的真实性和重要性后，如果这些研究证据具有良好的真实性，而且又具有临床或公共卫生学的重要意义，那么应联系患者的实际情况来探讨病因和危险因素，以助于解决患者的实际问题，所以还需考虑该研究证据的结果是否能够用于我们自己的患者或公共卫生实践。

（一）你的患者与文献中的研究对象是否存在较大的差异，以至于文献的研究结果不能回答患者的疑问

当考虑研究证据的适用性时，首先要比较你的患者与研究证据中的研究对象在年龄、性别、经济收入、文化背景等社会学特征及疾病严重程度、病程、用药剂量、依从性、对治疗的反应性等临床特征等方面是否有差异。如果存在影响结局的重要差异，可能研究证据就不能用于解决你的患者的问题。

（二）你的患者发生疾病/不良反应的危险性如何

针对单个患者的问题，需要根据研究证据提供的信息，估计你的患者发生疾病/不良反应的危险性。一种方法是在文献中寻找与你的患者各方面特征比较一致的亚组，参照该亚组的 NNH，但在这种情况下，一定要慎重，因为亚组的样本量常常较少，受机遇的影响较大。另一种方法是，计算患者发生疾病/不良反应的可能性与文献报告可能性的比值 F（decimal fraction），然后用文献结果中的 NNH 除以 F 值，获得你的患者发生疾病/不良反应的危险性大小。

（三）确定你的患者的喜好和希望解决的问题

医务人员在进行医疗决策时，除了参考研究证据、坚持利大于弊的原则外，同时还应邀请患者参与医疗决策。患者参与医疗决策是为了尊重患者的权利，不同的患者因其对自身疾病的关心程度，对医生所给予的诊治措施的期望值及对不良反应的耐受性等的不同，最终的选择会有差别。可以调查患者的喜好，提供一系列代表危险性和良好结局的价值尺度，了解患者在哪一点时会改变他们的决定，即接收或拒绝治疗措施（暴露）。

（四）是否应终止接触暴露因素或更改治疗措施

如果研究证据显示暴露因素与疾病/不良反应间具有较强的因果联系，即接触危险因素的危险明确而且巨大，应立即脱离接触危险因素或终止治疗措施。对于不良反应因果关系研究来说，即使证据显示暴露因素（治疗措施）与疾病/不良反应间因果关联不是太强，但有其他药物可选也容易作出决策。在临床实践中，同一种疾病常常有多种治疗措施可供选择，因此，医生应告知患者和与患者讨论每一种措施的利弊，最后决定采用哪一种治疗措施。

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，有所侧重加以应用。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。

第五节　循证实践及案例分析

疾病案例　患者，8岁，男性，不愿与他人交往和交流，具有社交障碍、交流障碍、兴趣狭窄、行为重复、刻板。临床诊断为“孤独症”，即“自闭症”。主治医生发现近年来就诊的自闭症患儿不断增多，鉴于自闭症给患儿及其家庭和社会带来的沉重负担，该医生希望了解自闭症患病的相关危险因素，从而有针对性地采取有效措施，降低自闭症的患病风险。现在，通过询问患儿父母，主治医生发现一部分患儿母亲在怀孕期间有服用抗抑郁药物史，因此医生提出了自己的疑问，母亲在孕期服用抗抑郁药物是否会增加子代患自闭症的风险呢？

一、提出问题

在本病例中，主治医生询问患儿母亲的孕期用药史，发现了一些线索，故希望了解母亲孕期服用抗抑郁药物是否会增加子代患自闭症的风险，这显然是一个药物不良反应的问题，是了解母亲孕期服用抗抑郁药物与其子代患自闭症之间是否存在因果关联的问题，从循证实践的角度，要明确这个问题就必须进行全面的文献搜索，以了解相关的证据。基于这个临床问题，为了能够准确检索相关的证据，需要进行相应的问题转换，构建相应的临床问题。

目前，构建临床问题通常采用PICO原则，即构建问题的四个要素：研究对象（patient）、干预措施（intervention）、与干预措施相比较的措施（comparison intervention）和产生的结局（outcome）。本例，依照该原则构建过程如下：

P：患者	患有抑郁症的孕妇
I：干预措施	孕期服用抗抑郁药物
C：比较措施	孕期未服用抗抑郁药物
O：结局	子代患自闭症

经过转换后的问题是：“孕期服用抗抑郁药物的母亲与未服用抗抑郁药物的母亲相比，其子代患自闭症的风险是否会增加？”

二、证据检索与评价

要快速获取所需要的证据（文献），需要掌握文献检索的方法和技巧，包括检索词以及文献数据

库的选择、检索策略的制订等。由于不同文献资料的质量不同，研究的设计和实施、数据的整理与分析、结果的解释和论文报告等方面存在差异，研究的真实性和可靠性以及适用性也不同。在循证医学实践中，建议首先检索二次数据库，即经他人评估和筛选过的循证医学资源，如 Cochrane 图书馆、Clinical Evidence 等，如未检索出所需要的信息，再进一步检索未经筛选的数据库，如 Medline（PubMed）、EMBASE、Springer link、CBM、中文期刊全文数据库等。

本例中以“Antidepressant”“Autism spectrum disorder”“Pregnancy”（中文数据库以“抗抑郁药物”和“自闭症”）为检索词进行检索，检索截止日期为 2017 年 10 月 8 日，例如在 Web of science 中以主题词为字段进行检索，检索式为：（Pregnancy OR Maternal exposure）AND（Offspring OR Childhood OR Children）AND（Antidepressant OR Antidepressants OR Selective serotonin reuptake inhibitors OR SSRIs）AND（Autism spectrum disorder OR Autism spectrum disorders OR ASD）。检索结果如下：

数据库	检索结果	与本例临床问题相关
Medline（PubMed）	39	39
中国生物医学文献数据库	3	1
中国学术期刊网	4	1
维普资讯	0	0
万方数据库	0	0
Web of Science	103	45
Springer link	287	18
Wiley online Library	1241	56
PsycINFO	5	5
The Cochrane Library	48	13

循证实践在回答临床问题时，应首先选择针对该问题的系统评价（systematic review），因为系统评价客观总结了所有相关的文献资料，其作为证据的论证强度最高。本例经去重后，共在国内外的 10 个数据库中检索到相关文献 35 篇。

本章分别选取队列研究和系统评价各一篇，分别是 Dheeraj Rai 等的队列研究“Antidepressants during pregnancy and autism in offspring: population based cohort study”（BMJ，2017，358：j2811）及 S. Andalib 等的系统评价“Maternal SSRI exposure increases the risk of autistic offspring: A meta-analysis and systematic review”（European Psychiatry，2017，45：161-166），参照表 8-2 相关内容分别对其真实性、重要性和适用性进行评价，以指导循证实践，回答提出的临床问题。

三、实践决策

（一）队列研究：Antidepressants during pregnancy and autism in offspring: population based cohort study（BMJ，2017，358:j2811）

1. 真实性评价

（1）是否采用了论证强度高的研究设计方法：在病因研究设计类型上，常用研究方案及其证据的论证强度以随机对照试验（RCT）最高，其次是队列研究。通过检索得知，目前关于此类临床问题大多为队列研究和巢式病例对照研究，相关 RCT 研究较少，而本研究属前瞻性队列研究，论证强度较高。

（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：本研究为队列研究，队列研究虽不能完全做到随机对照试验要求的严格控制，但暴露组与非暴露组的组间可比性在多个因素上是均衡的。

（3）试验组和对照组有关因果效应的测量方法是否相同（结果测量是否客观或采用盲法）：本研究中，母亲孕期接触抗抑郁药物的数据来源于瑞典的医学出生登记注册系统[the medical birth register

(since 1997)]和处方药物登记注册系统[the prescribed drug register (available from July 2005)],根据以往及当前的药物使用和诊断信息,将研究对象分配到抗抑郁药物暴露组和非暴露组。自闭症诊断信息几乎均源自瑞典健康照护系统(system of health and care),包括国家患者登记注册系统,斯德哥尔摩儿童和青少年精神健康登记注册系统以及小儿康复登记注册系统[ICD-9(299),ICD-10(F84)或DSM-Ⅳ(299)]中的诊断。本研究通过儿童或成人心理健康登记注册系统或国家患者登记注册系统,基于ICD 9(317-319),ICD-10(F70-79)和DSM-Ⅳ(317-319)来确定伴发的智力障碍(co-occurring intellectual disability)。为了评价登记注册系统诊断的准确性,研究者曾进行过两项验证研究:一项是全国性双生子研究孤独症谱系障碍诊断病例记录验证研究(case note validation study),一项是全国性双生子研究孤独症谱系障碍诊断的交叉验证研究(cross validation study),验证发现登记注册系统记录的诊断具有高度有效性,因此,本研究暴露组和非暴露组的测量方法是客观的。但本研究未提及是否使用盲法的问题。

(4) 随访时间是否足够长?研究结果包含了所有随访病例吗:文中表明,研究者使用了斯德哥尔摩市青年队列数据,数据包含2001—2011年度在斯德哥尔摩市生活的所有0~17岁青少年队列成员信息(n=735 096)以及全国和地区性性记录链接(record linkage)系统记录的队列成员一级亲属的信息。包括母亲在怀孕期间接触抗抑郁药物的3342名儿童,12 325例母亲患有精神障碍但在怀孕期间没有使用抗抑郁药的儿童,和母亲无患精神障碍史或妊娠期有暴露抗抑郁药物史相关记录的238 943名同龄儿童。

(5) 是否有因果效应的时间先后顺序:本研究属队列研究,该研究类型属于由因及果的前瞻性研究,其基本原理是依据该研究人群是否暴露于待研究的危险因素,将研究对象分为暴露组与非暴露组,然后前瞻性随访观察一段时间后,进而比较两组间所研究疾病的发病率,从而分析暴露于疾病之间的关联。本研究根据母亲怀孕期间是否使用抗抑郁药物分为暴露组与非暴露组,然后对随访结果进行分析,研究母亲孕期服用抗抑郁药物与其子代自闭症谱系障碍之间的关联。其因果时序关系是明确的。

(6) 是否存在剂量-反应关系:本研究在结果分析过程中并未涉及剂量-反应梯度关系。

(7) 病因学研究的结果是否符合流行病学的规律:关于暴露因素(母亲孕期间服用抗抑郁药物)与疾病/不良反应(子代谱系障碍)之间是否存在消长关系,未在本文中提及。

(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:本研究结果显示,在孕期暴露于抗抑郁药物会使子代自闭症谱系障碍风险增加0.45倍(OR值为1.45,95% CI:1.13~1.85)。与诸多研究如Dheeraj Rai(2013)及Croen LA(2011)等的结果是一致。然而,在Gardener H(2009),Yirmiya(2005)等研究中并未得到这一结论。因此,需进一步对此类研究进行系统评价,通过增大样本含量,减少随机误差所致的差异以提高统计效率;探讨多个研究结果间的一致性,实现多个研究结果的定量综合估计。

(9) 病因致病效应和不良反应发生的生物学依据是否充分:文中提及,所有抗抑郁药均可通过胎盘屏障作用于发育中的胎儿。而且在动物模型中,已有关于子宫内膜接触5-羟色胺能抗抑郁药(serotonergic antidepressants)与后代的自闭症样行为相关性的报道,据此推断人类具有类似的效果,故存在生物学合理性。但该因果关系的生物学依据至今尚未达成共识。

2. 重要性评价

(1) 关联的强度(OR或RR):在孕期暴露于抗抑郁药物会使子代自闭症谱系障碍风险增加0.45倍(OR值为1.45),对无并发智力障碍的自闭症而言关联性更强(OR=1.57)。

(2) 相对危险度(RR)或比值比(OR)的95%可信区间:与上述OR值相对应的95%可信区间分别为(1.13~1.85)和(1.21~2.04)。

(3) 害-需暴露人数(NNH):以孕期服用抗抑郁药物和未服用抗抑郁药物子代精神障碍患儿为例,其NNH=1/(136/3342-353/12 325)=83.33,即10年中每暴露83.33人出现1例自闭症谱系障碍

患儿。

3. 适用性评价

（1）你的患者与文献中的研究对象是否存在较大的差异：由文中的一般人口学描述中可知，本例中研究对象为0～17岁青少年，因此，研究结果适用于该患者。

（2）患者发生疾病/不良反应的危险性如何：根据患者的情况与文中对比，该患者母亲属于孕期暴露于抗抑郁药物的分组，NNH = 1/(136/3342 - 353/12 325) = 83.33。本研究中该亚组样本量足够大，且随访时间足够长，应可避免受机遇的影响。

（3）是否应终止接触暴露因素或更改治疗措施：研究证据显示该暴露因素与不良反应间具有一定的因果相关性，虽然关联强度较弱，但是存在潜在的自闭症治疗负担仍然不可忽视，所以在面对准备怀孕的抑郁症患者时，应与专科医生会诊后，结合患者的意愿进行决策，以决定孕期的抗抑郁药物使用方案，决定是否终止暴露因素，或更改治疗措施。

（二）系统评价：Maternal SSRI exposure increases the risk of autistic offspring: A meta-analysis and systematic review（European Psychiatry,2017，45:161-166）

1. 真实性评价

（1）是否是根据随机对照试验进行的系统评价：是否纳入高质量的原始文献，是评价系统评价作为证据的论证强度高低的重要标准。随机对照试验能够较好的控制各种偏倚对结局的影响，是评价干预措施疗效的最佳设计方案，因此依据随机对照试验产生的系统评价，其研究证据的论证强度被认为最高。但该方法多用治疗措施的有效性评价，而限于伦理学的要求，不良反应研究则较少应用。通过检索也得知，目前关于此临床问题的RCT研究较少，大多为前瞻性的队列研究和巢式病例对照研究，本系统评价研究为孕妇产前暴露于选择性5-羟色胺再摄取抑制剂（SSRI）是否会增加子代自闭症谱系障碍（ASD）的风险的多个研究的系统评价，是对多个队列研究和病例对照研究的定量综合。

（2）在系统评价的“方法学”部分，是否描述了检索和纳入所有相关研究的方法？评价单个研究证据的方法：文中清晰地说明研究者以“selective serotonin reuptake inhibitors，SSRI，autism spectrum disorder，ASD，pregnancy，childhood，children，neurodevelopment”相关词条为检索词，于2016年6月在PubMed and PMC，MEDLINE，EMBASE，SCOPUS和Google Scholar等数据库进行了系统检索，并进行了手工检索相关参考文献，但未公开发表的数据未纳入本研究。文中虽未提及评价单个研究证据的具体方法，但明确给出了3条纳入标准以及排除标准以保证纳入研究的质量。

（3）不同研究的结果是否一致：所纳入的原始文献中，关于长期服用选择性5-羟色胺再摄取抑制剂（SSRIs）是否会增加子代患抑郁症的风险，结论莫衷一是。但系统评价/meta-分析的主要作用即是对具有相同研究目的的多个研究结果进行综合定量分析的一种方法，其优点是通过增大研究样本含量，减少随机误差所致的差异以提高统计效率；探讨多个研究结果间的异质性，实现不一致的研究结果的定量综合估计。本系统评价共纳入了7项研究的数据，其中6项研究的结果为阳性，即母亲暴露于选择性5-羟色胺再吸收抑制剂（SSRIs）会增加子代患自闭症的风险，而另外1组数据未能得出此有统计学意义的结论。另外，本文中还采用异质性检验的方法来判断不同研究结果间的异质性。结果显示，I^2值为0%，表明研究间的异质性非常低。

（4）统计分析使用的数据资料是单个患者的资料还是单个研究的综合资料：系统评价的资料来源可以是原始研究中群体资料，也可以是原始研究中的个体资料，两种类型各有优缺点，应根据研究要求和已有条件进行选择。根据文中的内容可知，本系统评价统计分析中所使用的数据资料为单个研究的综合资料。

2. 重要性评价

（1）发生不良反应的强度大小如何（效应测量指标的大小）：本研究中通过对多个研究进行meta-分析后，直接给出了合并后的RR值为1.83，表明母亲孕期服用选择性5-羟色胺再摄取抑制剂（SSRIs）子代患自闭症的风险是不服用选择性5-羟色胺再摄取抑制剂（SSRIs）药物的1.83倍。

（2）发生不良反应的精确性如何（可信区间的大小）：与单个研究结果一样，系统评价结果的可信区间可以提供关于研究结果精确性的信息，若可信区间较宽，则表示精确性较低，反之则表示研究结果的精确性较高。本研究中得到的95%可信区间为（1.59～2.10），可信区间较窄，故本研究的精确性较高。

3. 适用性评价

（1）你的患者是否与系统综述中的研究对象差异较大，导致结果不可用：本研究所纳入文献的研究对象多为0～17岁的青少年，因此，该研究结果适用于该患者。

（2）规避不良反应的措施在你的医院是否可行：选择性5-羟色胺再摄取抑制剂（SSRIs）是临床治疗抑郁症的常规用药之一，但不是唯一治疗用药，因此可根据医生的经验和患者价值观选用其他替代药物。

（3）你的患者从治疗中获得的利弊如何：选用其他替代药物，在控制患者抑郁症的同时，又能够有效减少子代患自闭症等不良反应的发生，患者无疑将从中获益。

（4）对于治疗的疗效和不良反应，你的患者的价值观和选择如何：临床医生应该根据自己掌握的证据，耐心细致地向患者阐述更换治疗方法的利与弊，最后结合患者的价值观和选择，进行临床决策。

四、小结

综上所述，母亲暴露于选择性5-羟色胺再吸收抑制剂（SSRIs）会增加子代患自闭症的风险，应建议改用其他治疗抑郁症药物，并做好防范措施，防治子代自闭症等。

（孙业桓）

笔记

第九章　诊断试验研究、评价与循证实践

诊断性研究对于临床工作来说十分重要，正确的诊断是开展临床治疗的首要条件。但是任何一位临床医生在对疾病进行分析诊断时都可能受错误信息的影响，甚至作出包括错误信息的结论。为了避免受过多的错误信息的影响，需要对相关信息进行科学的分析和评价。

第一节　概　　述

一、基本概念

诊断试验（diagnostic test）是指应用各种实验技术、医疗仪器及其他手段对患者进行检查，以对疾病作出诊断，即通过应用某一诊断方法或多种诊断方法的综合运用将就诊者区分为患某病的患者和非患者，以便对确诊的患者给予相应的处理或治疗。

诊断试验可为疾病正确诊断以及鉴别诊断提供重要证据。诊断试验的实施，要涉及以下内容：①病史和体检所获得的临床资料；②各种实验室检查，如生化、血液学、细菌性、免疫学、病理学检查等；③各种影像学检查，如X线造影、B超、CT、核磁共振成像（MRI）及放射性核素检查等；④其他特殊器械检查，如心电图、内镜检查等；⑤各种临床公认的诊断标准，如各种自身免疫性疾病的联合诊断标准等。一项立项的诊断试验应该具有准确可靠、简便迅速、安全无损和成本低廉的特点。

二、诊断试验研究与评价的意义

诊断性研究证据的评价对于诊断疾病有着极其重要的作用。充分和准确的证据有助于帮助医生尽早明确诊断，及时提供临床治疗，对改善疾病的预后有着十分重要的意义。

（一）有助于诊断或排除某种疾病

在临床工作中，医生对任何一种疾病进行临床研究和治疗时都需要正确的诊断作为前提条件。要作出正确的诊断，首先需要收集患者的病史、体格检查、实验室检查等资料，然后将所获得的信息资源进行综合分析评价。此番临床论证的过程可能十分复杂，需要在各种相关的诊断性证据中寻找高质量的、有参考价值的并且具有临床指导意义的信息资源作为临床诊断的证据支持。目前国际上广泛认可的诊断方法在临床流行病学中被称为“金标准”。

诊断性研究以及对其证据的循证评价对于疾病的诊断和排除来说意义重大。

（二）有助于筛检无症状的患者

有些患者在患病初期，无任何自觉症状。若此时能采用合适的诊断试验对疾病作出正确诊断，及时进行干预和治疗，则可显著改善预后并对预防疾病的蔓延起到至关重要的作用。

第二节　诊断试验研究设计

评价一项新诊断方法的诊断价值，最佳研究设计是将待评价的诊断试验与金标准（gold standard）同步、盲法比较。首先确立疾病的标准诊断方法（最好是金标准）；其次是选择研究对象，并根据标准

诊断将这些对象划分为“有病组-病例组”与“无病组-对照组”；第三，再用新诊断方法同步测试这些研究对象，将获得的诊断结果与标准诊断比较，进而绘制四格表，计算敏感度、特异度、预测值、似然比等指标来综合评价该试验的诊断效率和价值。为减少偏倚，一般采用盲法评价。

一、确定金标准

金标准是指当前临床医学界公认的、诊断某疾病的最可靠方法，常用的金标准诊断方法有病理学诊断（组织活检和尸体解剖）、外科手术发现、特殊影像学诊断（如冠状动脉造影诊断冠心病），也可采用公认的临床诊断标准（如系统性红斑狼疮的 ARA 诊断标准等），通过长期临床随访所获得的肯定诊断，有时也可用作金标准诊断。

开展诊断试验时，金标准的选择应结合临床具体情况而定。例如肿瘤诊断的金标准是病理诊断，冠心病诊断最好选用冠状动脉造影。若金标准选择不当，就会造成“病例组”和“对照组”的划分错误，从而影响诊断试验的正确评价。同时，临床诊断试验还要遵循临床研究及伦理学的一些基本原则，如无伤害原则、知情同意原则等，变相增加了金标准操作的难度，有时要根据具体临床疾病，选择临床公认且得到广泛应用的诊断方法作为金标准。如诊断冠心病的金标准是冠脉血管造影，但造影有一定风险、操作存在一定难度，尤其是对照组的研究对象，此时可采用较为成熟的冠脉 CT 成像技术（CTA）作为金标准来使用。因此，临床研究中的金标准是相对的，但应注意：这些相对的金标准诊断，可能会带来一些偏差，需要一定的方法校正结果。

二、研究对象的选择

诊断试验的研究对象应具有代表性，其中病例组应包括各种临床类型和处于不同病程阶段的患者：如轻、中、重型，早、中、晚期，典型的和不典型的，伴有和不伴有并发症者，已治疗与未经治者等。对照组应选自确未罹患该病的其他病例，尤其应包括易与该病相混淆的病例，设置这样的对照才具有鉴别诊断的价值。研究对象应同期进入研究，可以是连续样本或者是按比例抽取的样本，但不能由研究者随意选择，否则就会出现选择偏倚，影响试验的真实性。

三、样本量的估计

诊断试验研究的样本量与下列因素有关：①对试验敏感度的要求，即假阴性率要控制在什么水平？敏感度高的试验一般用于疾病的筛选。②对试验特异度的要求，即假阳性率要控制在什么水平？特异度高的试验一般用于肯定诊断。③允许误差，一般取总体率 $100(1-\alpha)\%$ 可信区间宽度的一半。当敏感度和特异度均接近 50% 时，可用以下近似公式：

$$n=(Z_{\alpha}/\delta)^{2}(1-P)\times P \qquad \text{式(9-1)}$$

式中 n 为所需样本大小，Z_{α} 为标准正态分布中累积概率为 $\alpha/2$ 时的 Z 值，一般取 1.96，δ 为允许误差，一般定在 0.05～0.10 之间；P 指敏感度或特异度，可采用敏感度的估计值来计算病例组所需样本量，用特异度的估计值来计算对照组的样本量。

当待评价的诊断试验的敏感度或特异度小于 20% 或大于 80% 时，样本率的分布呈偏态分布，需要对率进行平方根反正弦转换，并用以下公式计算样本量：

$$n=[(57.3\times Z_{\alpha})/\sin^{-1}(\delta/\sqrt{p\times(1-p)})]^{2} \qquad \text{式(9-2)}$$

四、评价指标与诊断效能判定

诊断试验评价指标一般成对出现，如敏感度和特异度，阳性似然比和阴性似然比，阳性预测值和

阴性预测值等。

1. 敏感度与特异度　敏感度(sensitivity)，又称灵敏度或真阳性率，是指金标准确诊有该病的病例组中经诊断试验查出阳性人数的比例[$a/(a+c)$]。而病例组中诊断试验结果为阴性者，即为假阴性患者，占病例组的比率就是假阴性率，又称漏诊率，敏感度和假阴性率是互补的，即敏感度=1-假阴性率。

特异度(specificity)，又称真阴性率，是指金标准确诊无病的对照组中经诊断试验检出阴性结果人数的比例[$d/(b+d)$]。而对照组中试验结果为阳性者即为假阳性，假阳性例数占对照组的比率就是假阳性率，又称误诊率，特异度和假阳性率也是互补的，即特异度=1-假阳性率。

敏感度和特异度都为100%的诊断试验最为理想，即没有假阳性和假阴性出现，有病组和无病组中结果数据分布曲线没有重叠(图9-1a)，这在临床实践中往往难以实现，多数诊断试验结果的分布曲线是两组有部分重叠(图9-1b)。因此，诊断试验结果若为连续性变量时(如空腹血糖和血压等)，区分正常(阴性)、异常(阳性)的临界点(cut-off point)划分很关键，将直接影响敏感度和特异度，如选择A点(图9-1b)，该试验的特异度为98%，但敏感度仅为60%，造成许多患者漏诊；若要提高敏感度，可将临界点移向C点，敏感度可达95%，但特异度降低为60%，造成许多误诊病例。敏感度和特异度一般呈反比关系，临床实践中可通过以下途径选择临界点：第一，可利用受试者工作特征曲线(receiver operator characteristic curve，ROC曲线)寻找合适的临界点，以避免过多的假阳性和假阴性；第二，根据临床需要，通过权衡假阳性和假阴性造成的后果，选择临界点，以达到临床需要的高敏感度或者高特异度。高敏感度试验适用于：①疾病严重但又是可治疗的，疾病的早期诊断将有益于患者，而疾病漏诊可能造成严重后果者，例如结核病，梅毒等；②有几个诊断假设，为了排除某病的诊断；③用于筛检无症状患者而该病的发病率又比较低，因此当试验结果呈阴性时，高敏感度试验的临床价值最大。高特异度试验适用于：①凡假阳性结果会导致患者精神和肉体上严重危害时，例如诊断患癌，而准备实施化疗；②要肯定诊断时，高特异度试验阳性结果的临床价值最大。

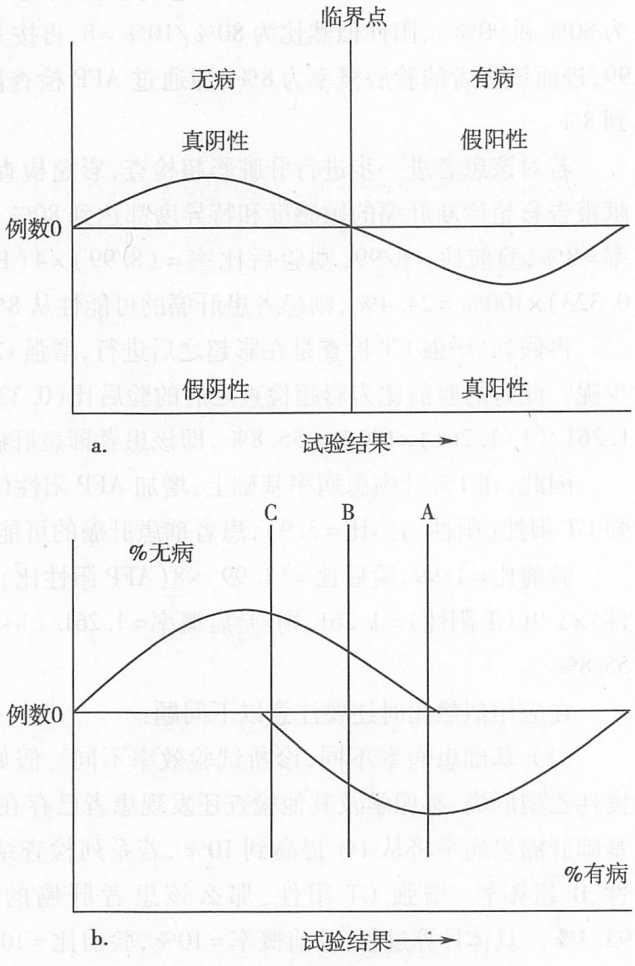


图9-1　敏感度和特异度的关系

2. 似然比　类似敏感度和特异度，似然比(likelihood ratio，LR)分为阳性似然比和阴性似然比，可用来反映诊断效能，即有病者得出某一试验结果的概率与无病者得出这一结果可能性的比值。既可计算连续性测量值不同区间的LR，又可避免简单地将试验结果划分为正常和异常，从而能全面反映诊断试验的诊断价值；似然比较敏感度和特异度更稳定，且不受患病率的影响，不足之处在于该指标是比而不是率，应用时需要在比值与率之间进行换算。

(1) 阳性似然比系真阳性率和假阳性率之比。一项诊断试验的阳性似然比为10，意味着当诊断试验结果为阳性时，判定患病的可能性是不患病可能性的10倍。

(2) 阴性似然比系假阴性率和真阴性率之比。若一项诊断试验的阴性似然比为0.01，意味着当

NOTE

诊断试验结果为阴性时，判定患病的可能性仅为不患病的1/100。因此，为肯定诊断之目的，应选择高阳性似然比的诊断试验，为排除某项诊断，则应选择阴性似然比更低的诊断试验。

似然比的应用步骤包括：先通过文献资料估计当前患者的验前概率，最初的验前概率常常是具有某一临床特征（年龄、性别、人群特征）的人群患病率，计算验前比[验前比=验前概率/(1-验前概率)]；确定待选择的诊断试验的敏感度、特异度和似然比；再按照诊断试验结果阳性或者阴性估计验后比[验后比=验前比×似然比]，最终得到验后概率[验后概率=验后比/(1+验后比)]。在连续进行多个诊断试验时，前一个试验的验后概率或者验后比就作为后一个试验的验前概率或者验前比。

例如：通过查阅文献获得中国人群乙肝患者40岁年龄段肝癌的患病率为1%。一例40岁有慢性乙肝病史的男性患者前来就诊，如果患者检查发现AFP阳性（AFP诊断肝癌的敏感度和特异度分别为80%和90%），阳性似然比为80%/10%=8，再按1%的验前概率，计算验前比为1/99，验后比为8/99，进而该患者的验后概率为8%，即通过AFP检查阳性这一结果，患者患肝癌的可能性从1%提高到8%。

若对该患者进一步进行肝脏彩超检查，彩超检查结果也是阳性（发现肝占位并且考虑肝癌），文献报告彩超诊断肝癌的敏感度和特异度都达到80%，阳性似然比为80%/20%=4。该患者的验前概率=8%，验前比=8/99，则验后比率=(8/99)×4（B型彩超阳性）=0.323，验后概率=0.323/(1+0.323)×100%=24.4%，则患者患肝癌的可能性从8%提高到24.4%。

再假如，增强CT检查是在彩超之后进行，增强CT检查为阳性，该患者患该癌的可能性升高到多少呢？此时的验前比为彩超检查之后的验后比(0.323)，而验后比为0.323×3.9=1.261，验后概率为1.261/(1+1.261)×100%=55.8%，即该患者罹患肝癌的可能性由24.4%提高到55.8%。

因此，在1%肝癌患病率基础上，增加AFP阳性（阳性似然比=8）、彩超阳性（阳性似然比=4）、增强CT阳性（阳性似然比=3.9），患者罹患肝癌的可能性提高到55.8%，如下计算过程：

验前比=1/99，验后比=(1/99)×8（AFP阳性比）×4（B型彩超阳性）×3.9（CT阳性）=1.261，则：验后概率=1.261/(1+1.261)×100%=55.8%。

在应用似然比时还要注意以下问题：

（1）基础患病率不同，诊断试验效率不同。假如该患者不仅为慢性乙型肝炎，影像学及其他检查还发现患者已存在早期肝硬化，其基础肝癌患病率将从1%提高到10%，若系列检查结果显示AFP阳性、B超阳性、增强CT阳性，那么该患者肝癌的可能性将达到93.3%。具体计算过程：验前概率=10%，验前比=10/90，验后比率=(10/90)×8（AFP阳性比）×4（B型彩超阳性）×3.9（CT阳性）=13.9，则验后概率=13.9/(1+13.9)×100%=93.3%。

不同人群的基础患病率不同，运用诊断试验得到的似然比不同，由此估计的验后概率也不同。临床上，确诊疾病需要寻找更多阳性诊断依据，而否定诊断则需要更多阴性结果支持。

（2）应用似然比应用图直接查找验后概率。如果事先测出一些诊断试验的似然比，依据文献资料获得人群的患病概率（验前概率），通过计算可推断患者检查后患某病的概率增加或减少了多少，有助于作出正确的诊断。除了运用上述公式进行计算外，还可利用似然比应用图（Fagan's nomogram）（图9-2），将直尺的一端放在验前概率对应点，再与该试验似然比所在点对齐，直尺另一端所指就是验后概率，此方法方便易行，临床实用性强。

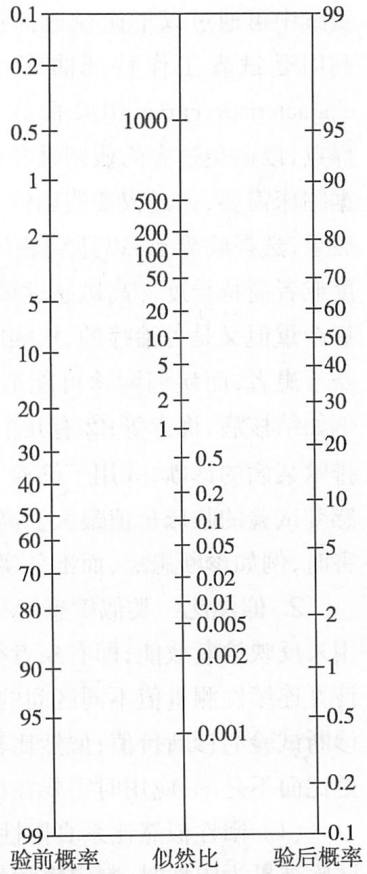


图9-2　似然比应用图

（3）当诊断试验结果为二分类变量时，可计算单个阳性或阴

性似然比；但当试验结果为连续性变量时，应分别计算不同区间对应的似然比。如试验结果范围是从0到100，则可分别计算诸如LR(0～10)、LR(11～20)、…LR(91～100)时的似然比。如式(9-3)：

$$LR(11\sim20)=\frac{\text{患者试验结果在11～20范围内所占概率}}{\text{无病组试验结果在11～20范围内所占概率}} \qquad \text{式(9-3)}$$

当试验测定结果为连续性变量时，诊断试验特征的最好表达方式是似然比，而不是常用的敏感度和特异度，似然比描述诊断试验的特征更为全面。

3. 预测值　在临床实践中临床医生和患者都会关心诊断试验的诊断价值，若结果阳性，患某病的可能性是多少？结果阴性时确未患病的可能性又是多少？这就涉及预测值(predictive value，PV)问题。①阳性预测值(positive predictive value，PPV)是指试验阳性结果者中真患该病者所占的比例[$a/(a+b)$]。②阴性预测值(negative predictive value，PV)是指试验阴性结果者中确未患病者所占的比例[$d/(c+d)$]。一般说来，越敏感的试验，其阴性预测值越高；反之特异度越高的试验，其阳性预测值越高。预测值的高低并不完全取决于试验本身，还与患病率[$(a+c)/(a+b+c+d)$]有关。不同临床情况下，患病率有时相差甚大。例如，AFP诊断肝癌试验的敏感度和特异度分别为80%和90%。肝硬化患者肝癌患病率大致为50%，如下表9-1，AFP诊断肝癌阳性预测值为80/(80+10)×100%=88.9%；如果在慢性乙型肝炎人群应用此诊断试验，人群中的肝癌患病率小于10%(假设为9.1%)，如表9-2所示，阳性预测值为44.4%；如果在携带乙肝病毒人群中应用该诊断试验，此时的人群患病率不足1%(假设为0.99%)，如表9-3所示，阳性预测值为7.4%；如果将AFP作为肝癌高发区普通人群(肝癌患病率100/10万人，0.1%)的筛查，如表9-4所示，阳性预测值降至0.8%。由此可见，患病率对预测值有着非常重要的影响。因此，在应用或者引用文献报道的诊断试验时，应考虑其患病率是否与本医院或者当地情况相似，在三级医院阳性预测值很高的试验，可能在基层医院就很低。

表9-1　AFP诊断肝癌的临床研究评价(人群患病率50%)

	金标准评价最终结果	
	肝癌	非肝癌
AFP(+)	80(80%)	10(10%)
AFP(-)	20(20%)	90(90%)
合计	100	100

人群患病率=100/(100+100)×100%=50%，敏感度=80%，特异度=90%，则：阳性预测值=80/(80+10)×100%=88.9%，阴性预测值=90/(20+90)×100%=81.8%。

表9-2　AFP诊断肝癌的临床研究评价(人群患病率9.1%)

	金标准评价最终结果	
	肝癌	非肝癌
AFP(+)	80(80%)	100(10%)
AFP(-)	20(20%)	900(90%)
合计	100	1000

人群患病率=100/(100+1000)×100%=9.1%；敏感度=80%，特异度=90%，则：阳性预测值=80/(80+100)×100%=44.4%，阴性预测值=900/(20+900)×100%=97.8%。

表 9-3　AFP 诊断肝癌的临床研究评价(人群患病率 0.99%)

	金标准评价最终结果	
	肝癌	非肝癌
AFP(+)	80(80%)	1000(10%)
AFP(-)	20(20%)	9000(90%)
合计	100	10 000

人群患病率=100/(100+10 000)×100%=0.99%;敏感度=80%,特异度=90%;则阳性预测值=80/(80+1000)×100%=7.4%,阴性预测值=9000/(20+9000)×100%=99.8%。

表 9-4　AFP 诊断肝癌的临床研究评价(人群患病率 0.1%)

	金标准评价最终结果	
	肝癌	非肝癌
AFP(+)	80(80%)	10 000(10%)
AFP(-)	20(20%)	90 000(90%)
合计	100	100 000

人群患病率=100/(100+10 000)×100%=0.1%;敏感度=80%,特异度=90%;则:阳性预测值=80/(80+10 000)×100%=0.8%,阴性预测值=90 000/(20+90 000)×100%=99.98%。

由此可见,预测值受患病率的影响,这就可以解释为什么一项诊断价值较高的诊断试验,用于普查时效果并不佳,主要是由于患病率相差过大所致。阳性预测值与敏感度、特异度及患病率之间的关系,可用式(9-4)表示:

$$\text{阳性预测值}=\frac{\text{患病率}\times\text{敏感度}}{\text{患病率}\times\text{敏感度}+(1-\text{患病率})(1-\text{特异度})} \qquad \text{式(9-4)}$$

前面所列的阳(阴)性似然比、阳(阴)性预测值等指标综合利用了敏感度与特异度的信息,但这些指标都与诊断界点(或阈值)的选取有关。同一诊断试验,不同的诊断界点就对应着不同的敏感度和特异度。为全面评价诊断试验的诊断价值,应分别计算不同诊断临界点下的敏感度和特异度。

4. **准确度**　准确度(accuracy)又称总符合率、粗一致性(crude agreement rate),表示观察值与标准值或真实值符合的程度。作为诊断试验的综合评价指标,它是真阳性与真阴性之和占受检总人数的百分率,反映正确区分患者与非患者的能力,准确度高,则真实性好。

5. **受试者工作特征曲线(receiver operator characteristic curve, ROC 曲线)**　是以真阳性率(敏感度)为纵坐标、假阳性率(1-特异度)为横坐标绘制而成的曲线(图 9-3)。ROC 曲线下的面积可以综合反映诊断试验的诊断价值。ROC 曲线可以用来确定诊断试验的截断点和比较不同诊断试验的诊断价值。

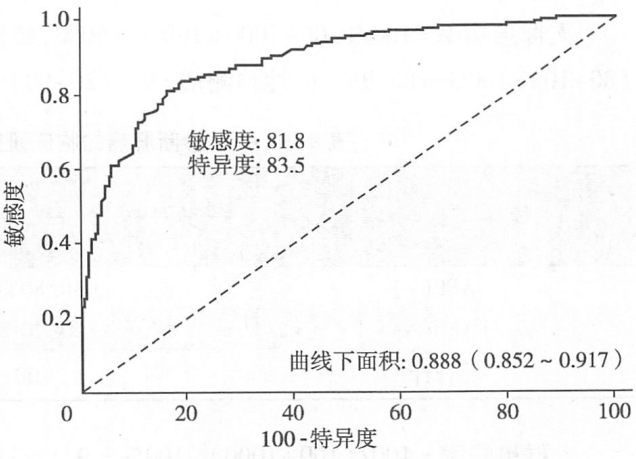


图 9-3　ROC 曲线

笔记

（1）确定诊断试验的截断点：ROC 曲线常被用来决定最佳临界点，如患病率接近 50% 时，最接近左上角那一点，可定为最佳临界点。如患病率极低或甚高，其最佳临界点可不在最接近左上角那一点。注意：仅靠一、二次试验难以找到一个敏感度和特异度俱佳的临界点。

（2）不同试验诊断诊断价值的比较：ROC 曲线是一种全面、准确评价诊断试验的有效方法，可用来比较两种或多种诊断试验的诊断价值，从而帮助临床医生正确选用诊断试验。图 9-4 中可以看到特异度相同的情况下，血清心肌酶 CK 诊断心梗的敏感度高于心电图 ST 段改变水平的诊断效果。

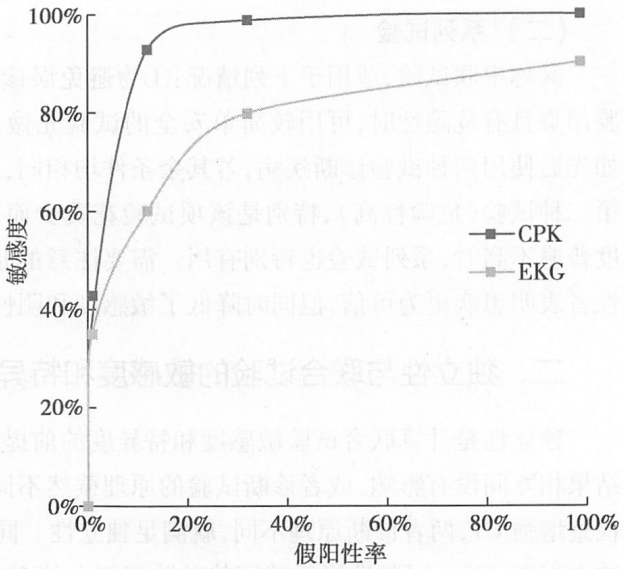


图 9-4　血清心肌酶 CK 水平和 EKG（ST 段改变水平）诊断心梗 ROC 曲线比较

除了上述目测方法外，还可计算 ROC 曲线下的面积（area under the ROC curve，AUC）来定量比较几种诊断试验的诊断效率。AUC 越大，越接近 1.0，其诊断价值越高；AUC 越接近 0.5，则诊断价值越低。AUC 的计算原理是将曲线下的面积分成多个梯形，计算每个梯形面积，最后将多个梯形面积相加求和，即为 AUC，亦可借助于计算机软件计算。

ROC 曲线具有如下优点：①方法简单、直观，通过目测就可判断和比较诊断价值；②可综合反映敏感度和特异度的相互变化关系；③ROC 曲线评价与基础患病率无关。

第三节　多项试验的联合诊断

鉴于敏感度和特异度俱佳的诊断试验不多，所以在不同的临床诊断需求的循证情况下，分别采取多项诊断试验的联合应用，以期提高循证诊断水平。

一、联合诊断方式

联合方式有两种：平行试验（parallel tests）和系列试验（serial tests）。平行试验指的是几个诊断试验同时进行，只要任一试验结果为阳性，联合诊断结果就为阳性。系列试验系按照一定顺序依次进行诊断试验，只有所有试验结果均为阳性者，联合诊断结果方为阳性。因此平行试验可以提高敏感度和阴性预测值，而系列试验相反，提高的是特异度和阳性预测值。平行试验和系列试验结果判定见表 9-5。

表 9-5　平行试验和系列试验的结果判定

试验 A	试验 B	平行试验	系列试验
+	+	+	+
+	−	+	−
−	+	+	−
−	−	−	−

（一）平行试验

平行试验又称并联试验，适用于下列情况：①对于住院或急症患者或外地的门诊患者复诊有困难时，需要迅速作出诊断；②为避免漏诊，但手中只有两项或两项以上不太敏感的试验时，组合成平行试验特别有用。需要注意的是：平行试验在提高了敏感度和阴性预测值的同时，特异度与阳性预测值会

降低，导致假阳性诊断增加。

（二）系列试验

又称串联试验，适用于下列情况：①为避免误诊情况的发生，可考虑使用系列试验。②当某些试验昂贵且有危险性时，可用较简单安全的试验先做，一旦提示该病可能存在之后，才使用这些试验。如先后使用两种试验诊断疾病，若其余条件均相同，应首先使用特异度高的试验使较少的患者暴露于第二种试验（危险性高），特别是该项试验花费少而又无危险性时，更应先用。③当单项试验的特异度普遍不高时，系列试验也特别有用。需要注意的是：系列试验可提高特异度和阳性预测值，试验阳性者表明患病更为可信，但同时降低了敏感度和阴性预测值，漏诊的可能性增加。

二、独立性与联合试验的敏感度和特异度

独立性是计算联合试验敏感度和特异度的前提条件：即诊断试验之间是互相独立的，诊断手段和结果相互间没有影响，或者诊断试验的原理截然不同。如诊断肝癌所用的血清学检查 AFP 和影像学检查增强 CT，两者诊断原理不同，就满足独立性。同样，冠心病的诊断，心电图、运动试验等反映心肌缺血程度，CTA、冠脉造影反映冠状动脉解剖上的变化，诊断试验间的互相影响也很小；而两种影像学检查如 CT 和核磁共振，都从占位角度诊断，CT 出现占位一般核磁共振也可以发现占位，两者结果就存在一定关联。

假设 A 试验和 B 试验为两个独立试验，其敏感度分别为 0.8 和 0.9，特异度依次为 0.6 和 0.9，并假设患病率为 20%。联合试验敏感度和特异度的计算方法如下：

平行试验的敏感度 = A 试验的敏感度 +B 试验的敏感度 *（1−A 试验的敏感度）

平行试验的特异度 = A 试验的特异度 * B 试验的特异度

系列试验的敏感度 = A 试验的敏感度 * B 试验的敏感度

系列试验的特异度 = A 试验的特异度 +B 试验的特异度 *（1−A 试验的特异度）

（1）平行试验方式：即两试验中一个阳性就作为阳性结果。根据计算公式可得：平行试验的敏感度 =0.8+0.9 *（1−0.8）= 0.98，特异度 =0.6 *0.9=0.54，由此可见平行试验提高了敏感度。

（2）系列试验方式：即两者均阳性才算阳性结果。根据上式得出系列试验的敏感度 =0.8 *0.9 = 0.72，特异度 =0.6+0.9 *（1−0.6）= 0.96，由此可见系列试验提高了特异度。

第四节　诊断试验的循证评价原则与标准

诊断试验证据要从真实性、可靠性和适（实）用性三个方面进行综合评价，以期得到真正对临床实践有意义的诊断性试验证据，供临床医生参考与应用。

一、评价原则

（一）真实性原则

真实性（validity）是指诊断试验所取得的结果与实际情况相符合的程度。主要评价指标包括敏感度、特异度、似然比和预测值等。循证医学对诊断性试验证据的要求，首先在于它的真实性，能及时对患者作出正确的诊断。在众多的诊断试验中，筛选具有真实性的试验，必须要有严格的条件，目前国际上通用的评估原则如下：

1. 是否用盲法将诊断性试验与参考标准（金标准）做过独立的对比研究　诊断性试验的研究过程中，对每个患者需进行两项试验，然后将新的诊断性试验结果与金标准诊断结果比较，才能判断该试验是否具有真实性。做这项新试验的技师（或医生），事先不应知晓金标准检测的结果，即在盲法之下进行试验，这样可避免人为的偏倚，使该试验结果更具真实性。最后列出四格表计算各项指标，根据 SEN、SPE 及 +LR 确定该项诊断性试验有无临床应用价值。

笔记

2. 该诊断试验是否包括了适当的病例谱　诊断性试验的受试者是否包括各型病例(轻、重、治疗、未治疗),以及易于混淆的病例?例如,测定血中T3、T4诊断甲亢、测定血糖诊断糖尿病,测定肝功能、肾功能判断肝脏或肾脏受损情况,这些都是较好的诊断性试验,当各型病例都包括在内时,这些指标既可诊断疾病,又可判断病情,还可以作鉴别诊断。

3. 诊断性试验的检测结果,是否会影响到参考标准的实施　倘若标准诊断(或参考标准)是确切可靠的金标准,那么不管新开展的诊断性试验结果如何,标准诊断结果都不会有所改变。倘若使用诊断标准不那么可靠,一旦发现新的诊断性试验结果与原有诊断有所出入,就会难以取舍,必须继续观察以明确患者的诊断,然后进一步判断原来的标准诊断是否存在的问题,以及新的诊断性试验是否真正可靠,这对提高临床诊断水平是有益的。因此,在评价时,一方面要考虑金标准的选择是否恰当,另一方面要考虑新的诊断性试验是否真有新的发现。

4. 若将该试验应用于另一组病例,是否也具有同样的真实性　一项可靠的诊断试验,在判断其真实性时,应考虑到该试验的重复性,如多次测定同一标本的结果接近,说明测定数值稳定、结果可靠。因此只要是疾病相同,不论在何处应用该项试验其结果都应是一致的,即使是在另一组病例中,对特定目标疾病的诊断也应具有同样的真实性。在评价中应了解作者是否用同样的试验对不同组的病例进行了检查,而且是否能够得到同样的结果。

(二) 可靠性原则

可靠性(reliability),是指在相同条件下用某测量工具重复测量同一受试者时获得相同结果的稳定程度。测量诊断试验可靠性的指标和方法如下:

1. 符合率(agreement/consistency rate),又称一致率,是重复试验判定的结果相同的人数占受检人数的比例。符合率$=(a+d)/(a+b+c+d)*100\%$

2. Kappa值及其意义　Kappa值是一致性评价的常用指标,用于判断重复试验结果的一致性。Kappa值的意义和推理举例说明如下:甲、乙两位临床经验相似的医生通过阅读相同的上腹部增强CT片诊断肝癌,共阅读100张CT片,结果见表9-6,两人均诊断肝癌的48例,两人均诊断非肝癌38例,观察一致率86%,此临床意见一致率较高,但根据一般常识,任何现象都存在偶然性(机遇)和必然性(非机遇),临床观察也不例外,假如排除了机遇因素后,实际一致率如何呢?实际一致率称Kappa值。Kappa值的计算步骤如下:

观察一致率(OA)$=(a+d)/N=(48+38)/100=86\%$;机遇一致率(EA)$=(r_1\times c_1)/N+(r_2\times c_2)/N=54\times56/100+46\times44/100=51\%$,实际非机遇一致率$=86\%-51\%=35\%$,则:Kappa值=实际非机遇一致率/理论非机遇一致率$=35\%/49\%=0.71$。

表9-6　甲乙医生阅读同样100张腹部增强CT诊断肝癌的一致率

乙医生	甲医生		合计
	肝癌	非肝癌	
肝癌	48(a)	6(b)	54(r_1)
非肝癌	8(c)	38(d)	46(r_2)
合计	56(c_1)	44(c_2)	100(N)

目前对判断Kappa值的意义尚有争议,但多数认为Kappa值在0.4~0.75有中度至高度一致性,>0.75时,一致性好。

观察结果的一致性评价方法依据数据类型而定,计量或连续性变量资料采用组内相关系数(intra-correlation coefficients,ICC);对于计数资料或分类变量资料,采用观察符合率与Kappa值表示。上述指标大于0.7,可认为一致性较好。

(三) 适用性原则

在评估诊断性试验的真实性与可靠性之后,应明确新的试验是否能够迅速地应用于临床,使更多

患者得到及时的正确诊断，并能够在取得社会效益的同时也取得经济效益，应对该试验的应用前景作出评价，以便推广应用：

1. 该试验是否能在本单位开展并进行正确的检测　在报道的资料中，是否明确地叙述了试验试剂、操作步骤与方法、检测对象与注意事项等，以便结合本单位情况，考虑是否有条件开展该项试验，有无经济效益等。例如：设有专科门诊的医院，开展肾动脉造影检查青少年高血压病，或血液专科门诊做血红蛋白电泳，检查长期贫血的患者，则阳性率较高，价值较大。如上述检查用于基层医院，去检查一般的高血压及贫血患者，则阳性率很低，开展后使用价值不大，经济效益也会明显受到影响。

2. 在临床上能否合理估算患者的验前概率　作好验前概率的估计，与下面三个因素有关，首先是医生本人的临床经验，其次是该病在当地的发生情况（群体的患病率），第三是诊断性试验检测的范围是否符合我们的病例。在临床医生掌握以上基本情况后，如果新的诊断试验确实可靠，能够及时对目标疾病进行检测，这样对患者的验前概率作出恰当的估计后，通过检测就可能达到早期诊断的目的，体现出诊断性试验的临床实用价值。

3. 检测后得到的验后概率是否有助于我们对患者的处理　利用阳性似然比，计算验后概率后，要了解该值是否已跨越诊断-治疗阈值，如已跨越说明诊断已明确，就应该及时治疗，才能使患者及早获得最佳效果。如果验后概率没有超越诊断-治疗阈值，仍在阈值 A-B 之间，则应进一步检查并给予适当治疗，以免延误病情，因此验后概率的计算，有利于患者及早达到治疗的目的，对患者的处理肯定也具有实用性。

二、诊断试验的评价标准

（一）是否将诊断试验与标准诊断法（金标准）进行了同步盲法比较

这是评价诊断试验最核心的一条。作为参照的金标准诊断，定义是否清晰明确，非常关键。在合理选择金标准的同时，待评试验必须同金标准诊断进行同步盲法比较。即要求试验结果的评价者预先无法获知哪些病例使用金标准判定为“有病”、哪些判为“无病”，同一患者诊断试验与金标准诊断结果要独立进行评价。

（二）研究对象的代表性如何

研究人群应包括两组：一组是用金标准确诊“有病”的病例组，另一组是用金标准证实为“无病”的对照组。病例组应包括各型病例：如典型和不典型，早、中与晚期病例，有无并发症等，以便使诊断试验的结果更具有临床应用的价值。最能体现诊断试验价值的是区分有病变的早期患者和易与该病混淆（症状、体征相同）的其他疾病。因此，诊断试验评价应该纳入那些临床实践中可能遇到的各种患者作为病例组，而对照组应选用金标准证实没有目标疾病的其他病例，特别是与该病易混淆的病例，以明确鉴别诊断的价值，正常人一般不宜纳入对照组。终末期患者检查时，试验结果常有明显的异常，因此待评诊断试验很容易将晚期患者与正常人区分开来。典型的例子是癌胚抗原（CEA）对结肠癌的诊断价值。最初报道中，36 例晚期结肠、直肠癌患者，有 35 例（97.2%）的 CEA 升高，对照组为未患结肠癌的其他患者，CEA 水平大多较低，据此作者认为 CEA 是一项筛选结肠癌的有用试验。而之后将诊断试验用于包括早期结肠癌以及有其他胃肠道疾病对象时，发现 CEA 诊断结肠癌的效率并不高，也不能将早期结肠癌患者与其他胃肠患者鉴别开来。

（三）样本量是否足够

诊断试验中样本量应该足够。过少，缺乏代表性；过大又增加工作量和研究费用。

（四）参考值选择是否合理

参考值或截断点（cut-off point），又称临界值（critical value），它的选择将直接影响试验的敏感度和特异度，因此，应交代截断点的选择方法并进一步说明其合理性。

（五）是否同时评价了真实性与可靠性

一项好的诊断试验，应该是既真实又可靠的。因此在评价时，不仅要计算反映真实性的指标，同

笔记

时还要计算评价可靠性的指标，两者缺一不可。

（六）是否交代了诊断试验的具体步骤

诊断试验方法（包括所用仪器、试剂、设备、实验条件等）是否具体详尽，操作步骤和注意事项是否明确，以便他人学习或验证，也有助于诊断试验在临床的推广和普及。

（七）是否对联合试验进行了正确的评价

在评价联合诊断试验的诊断价值时，不仅要计算联合试验敏感度、特异度等指标，各个单项诊断试验的评价指标也要一并报告。

（八）是否控制了偏倚

评价一项诊断试验时，还要考虑该诊断试验是否排除了各种偏倚（选择偏倚、错误分类偏倚、测量偏倚等）对结果的影响。

（九）临床意义及适用性如何

诊断试验经过效用分析后，还需用可靠的依据说明其临床意义和适用性，包括正确判断的收益和错误判断的可能后果等。

目前，为方便对诊断研究证据展开质量评价，陆续发布一些评价工具，如诊断研究的报告规范STARD（Standards for Reporting of Diagnostic Accuracy），用来评估诊断研究的报告质量；QUADAS系列评价标准（Quality Assessment of Diagnostic Accuracy Studies），可用于评价诊断研究的方法学质量。

第五节　诊断试验证据的临床实践及案例分析

循证诊断实践首先要将临床实践中有关诊断试验问题用一个可以回答的方式加以构建，其次证明有关该诊断试验正确性的证据是足够可信的，然后了解证据的重要价值和临床意义。最后考虑如何将这项诊断试验用于当前的患者，明确诊断。具体可参照表9-7。

表9-7　循证医学实践步骤

1. 根据临床问题（可回答的特殊临床问题）找出最恰当的、高度相关的研究文献
2. 评价诊断试验的科学性
   ①试验是否与金标准试验进行了同步盲法比较
   ②是否每个被测者都用参照试验进行评价
   ③所研究患者样本是否包括临床中的各种患者
   ④诊断试验的精确性如何
3. 估计临床应用的重要性
   ①估计疾病的验前概率
   ②说明和评估有关试验的敏感度、特异度、预测值和似然比资料
   ③评估应用该试验似然比估计的期望结果（验后概率）
4. 临床研究结果是否适用于自己的患者并取得预期的结果
   ①结果是否适用于我自己的患者
   ②诊断试验结果是否改变了对患病率的估计
   ③诊断试验结果是否改变了对患者的处理

一、构建循证诊断问题

案例：2013年3月5日，一位35岁乙肝病史10年的男性患者来某三甲医院就诊，实验室检查

HBsAg(+),HBcAb(+),HBeAg(+),ALT 正常水平,HBV-DNA 检测 5×10^3,血清透明质酸(HA)和Ⅲ型胶原(PⅢP)均不同程度升高,临床怀疑患者肝病程度已进展到早期肝硬化,进一步该如何处理?

首先基于临床案例,提出一个可回答的问题,将问题具体化。如根据现有的临床指南,慢性肝病患者接受抗病毒治疗的标准是不同的,而依据目前患者的资料,是否抗病毒治疗取决于是否进入明显肝纤维化(F3)或者肝硬化(F4)阶段。因此当前需要解决的问题是:"患者肝纤维化进展到何种程度?通过何种措施可以帮助我们进一步明确诊断"。

根据一般文献资料获知肝纤维化的金标准诊断为病理诊断,临床上常用肝穿刺活检病理组织学检查来明确诊断,但为有创性操作,尽管目前通过超声引导穿刺提高了安全性,但很多患者仍难以接受。非创伤性的诊断技术则成为研究和评价热点。血清学检查尽管简便,但临床诊断效率不高。近年来开展了超声弹性成像、磁共振弹性成像在肝纤维化诊断中的应用,那么我们将选择超声弹性成像还是磁共振弹性成像呢?

为此,进一步将问题具体化,转换为三个可回答的研究问题是:超声弹性成像诊断肝纤维化有效吗?磁共振弹性成像有效吗?磁共振弹性成像比超声弹性成像更好吗?

二、检索证据

检索证据通常分三大步:第一步,选择医学文献数据库,如最为常用的 Medline(PubMed,www. pubmed. com)、UpToDate 或 Embase 等。第二步,选择恰当的主题词,如针对本例患者的处理,主题词包括慢性肝病、肝纤维化(chronic liver disease,fibrosis),超声弹性成像(ultrasound elastography)或者磁共振弹性成像(Magnetic resonance elastography),诊断(diagnosis)。第三步,建立检索策略、实施检索。循证医学检索的基本策略是首先选择最新临床实践指南,然后选择指南未纳入的新近 meta-分析,最后选择未纳入 meta-分析的原始研究文献。这样可在最短时间内获得最佳证据。

检索式:(hepatic fibrosis) AND (ultrasound elastography OR Magnetic resonance elastography[Title/Abstract])

采用上述检索词,首先检索指南(文献类型 Practice Guideline),在 PubMed 中发现了 2 篇实践指南,最新一篇在 2011 年发表,为西班牙文发表的地方指南,不适合我们现在处理的患者。2010 年发表的另一篇意大利地方指南,主要针对非酒精性脂肪肝的诊断与处理,与目前慢性乙肝也不符合。

进而检索系统综述与 meta-分析(文献类型 Systematic Reviews),同样在 PubMed 上发现 16 篇文献,2 篇与指南重复,1 篇关于 HCV 肝纤维化,1 篇关于肝移植术后 HCV 复发肝纤维化,1 篇关于酒精性肝病肝纤维化,1 篇为药物性肝损害,1 篇非英文文献,1 篇与主题无关;6 篇为 2007—2010 年期间发表的 meta-分析,其中 1 篇 2011 年发表在 J Hepatology 杂志上的超声弹性成像 meta-分析和 1 篇 2012 年最新发表在 Hepatology 杂志上的磁共振弹性成像符合检索要求。但未发现比较超声弹性成像和磁共振弹性成像的诊断研究,需要在原始研究文献中继续检索。

继续检索原始研究文献。首先,采用上述检索式检索全部杂志,有 504 篇之多,若将检索式中超声弹性成像或者磁共振弹性成像中的 OR 修改为 AND,目标将缩小到 19 篇。从中很快找到 2008 年发表在 Gastroenterology 中的 1 篇原始研究文献,虽已被两篇 meta-分析纳入,但 meta-分析发表之后未发现更新的原始研究文献。

三、评价证据

当收集到了相关的文献资料后必须按照表 9-7 中第 2 点提出的标准对文献进行严格评价。通过者才可认为:设计合理、科学,其结果真实可信。以上述原始研究为例进行阐述。

(一) 试验是否与金标准试验进行了同步、盲法比较

判断诊断试验真实性的最好方法,是将所考核的诊断试验结果与"真实"情况进行比较。"真实"

情况是由金标准诊断来确定的。金标准的选择应结合临床具体情况而定，因此，当找到相关文章后，首先要检查每一位研究对象在进行诊断试验时，是否都采用了合适的金标准。前文中所举的例子，其金标准是病理组织学检查或者通过肝穿刺活检，若腹水或者凝血功能等影响肝穿刺，改用经颈静脉肝静脉活检获得组织学诊断。待评诊断试验为超声弹性成像和磁共振弹性成像，在方法学中都进行了描述，每种检查都是独立进行的，结果评价者不知道最终组织学诊断结果，也不知道另一项检查结果。

（二）是否每个被测者都用参照试验进行了评价

有些情况下，如金标准昂贵或具侵入性，可能不是所有的患者都进行了检测。研究者常常将试验结果阳性者，送去做金标准检测，而阴性者只抽一部分人去做，这样必然夸大了诊断试验的敏感度，造成偏倚。在上述例子中，对未完成金标准诊断的 13 例患者进行了剔除，这样的报告更客观和真实。

（三）所研究患者样本是否包括临床中的各种患者

上述例子中包括了临床上从无肝纤维化到明显肝硬化的各种患者，即便存在腹水也通过特殊检查获得组织学诊断纳入。

同时也应看到，文中部分患者由于种种原因，造成超声弹性成像（腹水 13 例，肥胖 10 例）或者磁共振弹性成像（幽闭恐怖症 3 例，血液病 3 例，肥胖 2 例）检查结果无法判定或者评价失败。这些患者实际就是将来诊断试验应用时，应限制的人群，在个体化选择诊断试验时将起决定作用。

（四）诊断试验的可靠性

上述例子中，作者虽没有描述变异系数和 Kappa 值，但全部测量结果都经过了前后两次评价，并且至少有两名高年资或者操作有丰富经验的专业人员评价，每种方法的评价都给出了具体的标准。在其他研究报道中两种诊断试验结果测量，也是可重复的。

如果表 9-7 中第 2 点标准中有至少一项不达标，就说明诊断试验可能存在严重缺陷，使用时应慎重。但要找到完全符合上述标准的诊断试验文献有一定的难度，为此，一些国际组织或学术机构开始出台一些临床研究报告规范，以促进研究质量和研究水平的提高。其中，对于诊断试验研究，2003 年初在 *Clinical Chemistry*，*Annals of Internal Medicine* 和 *BMJ* 等主流杂志上发布了诊断试验准确研究报告标准（Standards for Reporting Diagnostic Accuracy，STARD），为评定诊断试验提供了详细的可参考标准。STARD 包括 25 个条目，并采用了图标设计以提高评价诊断研究报告准确性的方法学和信息质量。

四、应用证据

（一）估计诊断试验临床应用的指标

1. 估计疾病的验前概率　上述案例中提示早期肝硬化的特征包括：35 岁，10 年慢性乙肝病程、血清学 HBsAg（+），HBcAb（+），HBeAg（+），血清肝纤维化指标阳性。临床判断早期肝硬化的概率在 50% 以上。相反，如果 25 岁患者，仅仅体检发现 HBsAg（+），其他病毒标志物阴性，肝纤维化指标均阴性，估计早期肝硬化的概率不足 1%。即使年龄 35 岁患者，母亲有乙肝携带状态，并且患者 HBsAg（+）来源于其母亲，早期肝硬化的概率也不超过 10%。

由于患病率与诊断试验的阳性预测值（试验结果阳性时，患者患病的概率，即验后概率）成正比，与诊断试验的阴性预测值（试验结果阴性时，患者不患该病的概率）成反比。正确地估计验前概率就显得十分重要。如果验前概率的估计来自他人报告的结果，应考虑自己患者的情况是否与其报告的一致，若不同，则需查阅更多文献以了解不同情况下的验前概率。

2. 说明和应用关于诊断试验结果的资料　诊断试验评价指标包括敏感度、特异度、预测值、ROC 曲线下面积、阳性和阴性似然比等指标，反映试验的诊断效率。通常特异度高的试验用于肯定疾病诊断。敏感度高的试验用于疾病筛查。相对于单个试验，meta-分析资料更有利于临床实践评估（表 9-8）。

表 9-8 磁共振弹性成像和超声弹性成像诊断明显肝纤维化和早期肝硬化的价值

	磁共振弹性成像	超声弹性成像	P 值
明显肝纤维化(F0～F1)vs.(F2～F4)			
敏感度	0.94(0.81～0.98)	0.79(0.74～0.82)	
特异度	0.95(0.87～0.98)	0.78(0.72～0.83)	
SROC	0.994(0.985～1.0)	0.837(0.756～0.918)	$P<0.05$
早期肝硬化(F0～F2)vs.(F3～F4)			
敏感度	0.92(0.85～0.96)	0.83(0.79～0.86)	
特异度	0.96(0.91～0.98)	0.89(0.87～0.91)	
SROC	0.985(0.968～1.0)	0.906(0.838～0.975)	$P<0.05$

综合受试者工作特征曲线(SROC)拟合技术主要用于诊断试验的 meta-分析。基本原理是针对同一指标的多个不同诊断试验所得到的多条 ROC 曲线,利用其诊断比值比,综合其诊断信息,最终绘制出一条光滑的 SROC 曲线。可见,SROC 曲线有机整合了多个独立 ROC 曲线的诊断信息,势必获得更高质量的诊断试验性能评价。根据相关计算公式可算出 SROC 的曲线下面积,SROC 的取值范围 $0 \leqslant SROC \leqslant 1$,SROC 越接近于 1,则诊断价值越高,表明诊断准确性越高。上表通过综合受试者工作特征曲线来判断磁共振弹性成像和超声弹性成像在诊断明显肝纤维化和早期肝硬化时的临床应用价值,数据表明,磁共振弹性成像和超声弹性成像的诊断效能均较高,并且在诊断时磁共振弹性成像的准确性比超声弹性成像更高。

3. 应用似然比 似然比的含义是试验的结果使验前概率提高或降低了多少。实际为验前比提高或者降低的倍数。表 9-9 为基于验前概率计算的验后概率。

表 9-9 磁共振弹性成像检查判断明显肝纤维化存在与否计算的验后概率

验前概率	验前比	试验结果	似然比	验后比	验后概率
50%	1	+	18.8	18.8	94.9%
50%	1	-	0.063	0.063	5.9%
10%	1/9	+	18.8	2.09	67.6%
10%	1/9	-	0.063	0.007	0.7%
1%	1/99	+	18.8	19.0	16.0%
1%	1/99	-	0.063	0.0006	0.06%

(二) 将临床研究结果用于自己的患者

如何将经过严格评价后的诊断试验结果,应用在自己救治的具体患者身上,应从以下三个方面去考虑:

1. 试验结果是否适用于自己救治的患者 首先要确定在本单位或者推荐的检查单位是否已开展或能开展该项检查,包括仪器、设备、试剂、人员的配备。还要考虑将该试验搬过来用于自己患者的测定时,一致性和重复性如何。其次,对比系统综述或文献报道的研究结果所报道的研究场所(如美国还是中国)其文化背景、人口特征等是否与自己所在单位患者的情况相同。两者的情况越接近,试验结果的重复性就越好。在本例中,肥胖患者的诊断无论超声弹性成像或者磁共振弹性成像都存在检查结果无法判定的问题,肥胖对超声弹性成像的影响更大。超声弹性成像同时受腹水影响,但对磁共振没有影响。

2. 如何合理估计自己患者的验前概率 可以根据个人临床经验、地区性和全国性资料或文献所报告的人群调查结果进行估计。对于一篇报道某一疾病患病率的文献可用如下标准进行评估:

(1) 有关验前概率的证据是否科学?具体包括:①被研究患者是否代表了有该临床问题的完整的疾病谱?②对于确诊的每一条标准是否明确和可靠?③诊断的程序是否全面并能持续应用?④对

笔记

于最初未诊断出的患者是否进行了足够长时间和完整的随访?

(2) 有关验前概率的证据是否重要? 具体包括:①诊断的是什么病? 其概率是多少? ②对于疾病概率估计的精确程度。

符合上述标准文献数据可以用于自己患者患病概率的估计。分两种情况:若诊治条件和患者特征与文献类似,可直接采用文献的验前概率;若诊断条件、患者特征与文献报道不同,则可以文献报告的验前概率作为基点,根据患者实际情况进行调整,观察验后概率的变化,最后确定验前概率。

3. 验后概率的结果是否改变了对患者的处理　在本例中,35 岁,10 年慢性乙肝病程、血清学 HBsAg(+),HBcAb(+),HBeAg(+),血清肝纤维化指标阳性。临床判断早期肝硬化的概率 50%,即验前概率为 50%,诊断试验阳性,肝纤维化或者肝硬化验后概率达到 94.9%,抗病毒治疗就非常必要,反之,若概率低于 5.9%,可暂时不选择抗病毒治疗。

倘若一名 25 岁就诊者,仅仅体检发现 HBsAg(+),其他病毒标志物阴性,肝纤维化指标均阴性,估计早期肝硬化的概率不足 1%。即使磁共振弹性成像诊断试验结果为阳性,验后概率仅 16%,抗病毒治疗的证据并不充分。

又假如一名年龄 35 岁患者前来就诊,母亲有乙肝携带状态,并且患者 HBsAg(+)来源于其母亲,早期肝硬化的概率估计 10%,磁共振弹性成像结果阳性,验后概率 67.6%,如何选择处理,是抗病毒治疗或者暂时不抗病毒治疗?

这里引出了"诊断阈值"和"治疗阈值"两个概念。决定一项诊断试验是否重要时,所关注的是该试验在鉴别有病和无病时的正确性及其改变验前和验后概率的能力,能较大地改变验前和验后概率的试验,对临床实践是重要和有用的。也就是说只有当某一项诊断试验能使验前和验后概率发生较大变化,而且这种变化会对是否要继续进行另一项检查或对治疗计划的改变有影响时,才能认为这项诊断试验的确对患者有帮助。

如果诊断试验的目的是为了决定是否对患者进行治疗,此时不仅要了解该诊断试验的有效性,更重要的是评估该项试验的阳性或者阴性结果是否影响了对患者治疗方案决策。如果新试验能增加已有的信息并导致治疗措施的改变,进而给患者带来益处,则该试验就有价值。因此,试验原则是使用有可能改变处理方案的试验,而治疗原则是选用利大于弊的治疗方案。下面以早期肝癌诊治为例说明行动点、诊断阈值和治疗阈值。

先来看看治疗的情况。早期肝癌的治疗,手术切除作为首选,但即使这样,手术仍存在一定风险。假设早期肝癌手术后 3 年存活率为 80%,疾病本身、肝功能状态以及手术创伤的病死率达到 20%,而如果采取其他措施包括积极观察病情变化而不立即采用手术治疗,可能的结果是:存活率下降到 50%。与暂时观察相比,及时手术显然利大于弊:80% −50% =30%,即存活率提高 30%,病死率下降 30%。若无其他选择,合理的决策一定是选择及时手术治疗(表 9-10)。

表 9-10　对于已确诊的疾病

是否有病	是否治疗	存活率	死亡率
有	是	80%	20%
有	否	50%	50%

上面的处理决策是在疾病诊断明确的情况下,手术方案利大于弊,有高于 30% 的存活率获益,但假如疾病诊断的概率不是 100% 呢? 出现误诊时,如果选择暂时观察,不会出现手术带来的风险,而即使择期手术,也几乎可完全治愈;相反,若选择及时手术,在患者一般状况未改善的情况下,手术本身的风险及病死率可能达到 10%。因此,与暂时观察相比,及时手术的结果是增加了 10% 的死亡风险。对不能确诊的疾病,临床决策还依赖于对疾病诊断的可能性大小,即诊断概率(表 9-11)。

表 9-11 对于尚未确诊的疾病

是否有病	是否治疗	存活率	死亡率
有	是	80%	20%
有	否	50%	50%
无	否	100%	0%
无	是	90%	10%

如果诊断该疾病的可能性，即概率为 90%，选择及时手术治疗：患者得到的好处为 27%，即 90%（患病可能性）×（80%-50%）（生存率增加），而可能出现的害处是 1%，即 10%（诊断错误可能性）×（100%-90%）（生存率减少），此时的合理决策仍是选择及时手术。

如果诊断疾病的可能性即概率为 10%，选择及时手术治疗：患者得到的好处是 3%，即 10%×（80%-50%），而可能出现的害处是 9%，即 90%×（100%-90%），合理决策是暂时不考虑手术。

倘若在 10% ~90% 之间，可以找到一个概率点，即行动点，在这点上，治疗的利弊相当。在这点之上，治疗利大于弊，选择该治疗；在这点之下，治疗弊大于利，暂时不选择该治疗。在本例中，患病可能性×（80%-50%）=（1-患病可能性）×（100%-90%），患病的可能性，即行动点为 25%。

在上述的案例分析中，前提条件是尚缺乏可以帮助进一步确定或者排除该疾病的诊断手段。因此，如果患者不是来源于肝硬化人群，也没有 B 型彩超结果阳性，依据患者有慢性乙型肝炎基础（基础概率 1%），AFP 阳性（阳性似然比 = 8），增强 CT 检查阳性（阳性似然比 = 3.9），该患者肝癌可能性达到 24.0%，如果没有其他试验可以进一步选择，可以直接决策，即选择手术切除治疗。

而实际上，肝癌的诊断还有很多其他方法。一方面，通过检查明确患者是否有肝硬化，增加肝癌患病可能性（即基础患病率），另一方面，接受如磁共振检查、血管造影检查、超声引导下穿刺活检病理组织学检查等帮助进一步肯定或者排除肝癌。治疗决策又会是怎样？是否需在术前选择其中一种或者多种检查？

处理（手术）与不处理（暂时不手术）的利弊已经明确，行动点在 25% 的概率水平。目前患者肝癌的可能性估计为 24.0%。若有更多诊断试验可以选择，是否选择取决于诊断试验的优劣。

首先考虑磁共振，除了增加费用之外，无过敏等副作用，安全性很好，医院可以提供（可获得性），患者接受程度也很高（可接受程度），是否选择、取决于其诊断效率，即敏感度和特异度。文献报告磁共振诊断肝癌的敏感度为 86%，特异度为 81%。目前患肝癌诊断可能性仍为 24.0%，试验结果阴性，疾病的可能性，即概率会减少，未改变随访选择。但若试验结果阳性，疾病的可能性，即概率增加，从 24.0% 提高到 25% 之上，则可能支持及时手术治疗。因此，应选择磁共振检查。

在上述案例中，如果患者 B 超和 CT 都发现占位考虑肝癌，患病可能性为 55.8%，那么是否还会选择 MRI 检查呢？试验结果阳性，直接手术，未改变决策，无需检查；试验结果阴性，根据试验的敏感度和特异度，推算：

诊断试验的阴性似然比为：（1-敏感度）/特异度 =（1-86%）/81% = 14/81；验前概率为 55.8%，验前比为：55.8%/44.2% = 1.26；验后比为：1.26×（14/81）= 0.218，则：验后概率 = 0.218/（1+0.218）×100% = 17.5%

这一结果已低于行动点（概率为 25%），所以应该为患者选择磁共振检查，因为检查结果若为阳性则更支持肝癌诊断，及时手术治疗，而阴性结果则可能改变手术治疗的选择。

另一种情况，假如知晓该患者基础患病的可能性为 1%，是否直接接受磁共振检查（如作为一种高危人群普查项目）而不先进行 AFP 检查、B 型超声检查、CT 检查呢？先来分析一下直接检查磁共振的结果。如果磁共振试验阴性，患病可能性即概率减少，不支持及时手术治疗，未改变选择；试验结果磁共振阳性，疾病的可能性即概率会增加。根据试验的敏感度和特异度，阳性似然比为：敏感度/（1-特异度）=（86%）/（1-81%）= 4.5；验前概率为 1%，验前比为：1%/99% = 1/99；验后比为：（1/99）×

笔记

4.5=4.5/99=0.046；则：验后概率=(0.046)/(1+0.046)100%=4.4%

这一结果低于行动点（概率为25%），无论该试验结果如何，仍然选择放弃手术治疗，诊断试验对治疗决策无影响，因此直接选择该诊断试验是不合理的。

倘若从1%基础水平逐渐增加患病的可能性，可以找到一点，在这点之下，无需进行该诊断试验，因为无论该试验结果如何，都要选择观察，在这点之上，进行诊断试验，若试验结果阳性，患病的概率可能上升到25%以上，将改变处理患者的决策，这一点称为诊断阈值（test thresholds）。

诊断域值比=行动点比/阳性似然比=(1/3)/[(86%)/(1-81%)]=1/45=0.074，则：诊断域值概率=0.074/(1+0.074)×100%=6.9%

从90%逐渐往下降低患病概率，同样也可以找到一点，在这点的患病概率上，无需进行该诊断试验（磁共振检查），因为无论试验结果如何，都会选择手术治疗；在这点之下，进行诊断试验，如果试验结果阴性，患病的概率下降到25%以下，将改变我们处理患者的决策，这一点称为治疗域值（treatment thresholds）。

治疗域值比=行动点比/阴性似然比=(1/3)/[(1-86%)/81%]=1.69，则：

治疗域值概率=1.69/(1+1.69)×100%=62.7%

行动点、诊断阈值、治疗阈值示意图如图9-5所示

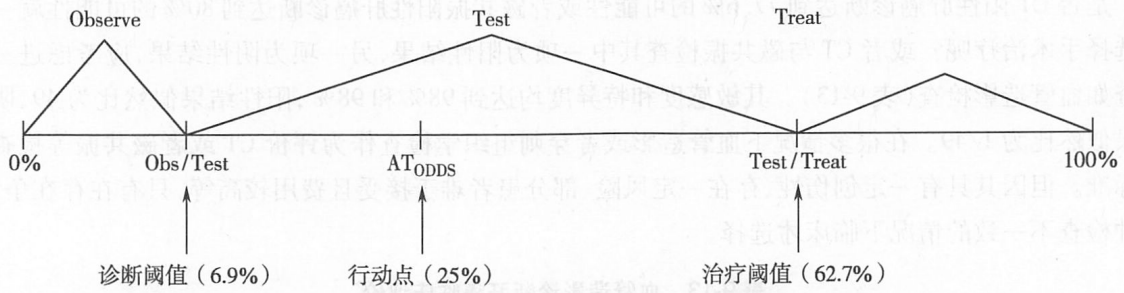


图9-5　行动点、诊断阈值、治疗阈值示意图

如果诊断概率出现下列情况，以下为合理处理：①诊断阈值之下，无需进一步诊断试验，直接决定暂时观察；②介于诊断阈值与行动点之间，需要进一步诊断；若没有其他诊断试验，这时采取的治疗害处大于好处，采取治疗要慎重；③介于行动点与治疗阈值之间，需要进一步诊断；若没有其他诊断试验，这时采取的治疗好处大于害处，可以采取治疗；④大于治疗阈值，无需进一步诊断试验，直接决定治疗；⑤介于诊断阈值与治疗阈值之间，如有其他诊断试验，需要进一步诊断，直到达到治疗阈值才开始治疗（理论上）。

因此，患者在基础状态下的肝癌患病率仅1%时，首先选择了AFP检查与B型超声检查，因为在肝癌患病率低于6.9%的人群中直接进行磁共振检查即使阳性也不适合直接选择手术处理。而直接选择磁共振检查也违背了诊疗路径所要的其他原则包括检查费用低、安全性好、可接受程度高、可获得性高（在一般医院都可以获得）等。

对于该临床案例，患者AFP阳性结果，肝癌患病率从1%提高到8%，又进行了增强CT检查并获得阳性结果。那么能否不选择增强CT检查而直接选择磁共振检查呢？

若直接选择磁共振检查，阳性结果似然比为4.5，患者阳性结果，肝癌患病概率从8%提高到28.1%，超过决策处理的25%，因此同样可以选择。两种诊断试验的差异见表9-12，鉴于磁共振的特异度更高且无造影剂过敏或者造影剂可能带来不良影响的疾病如肾功能不全，可优先选择。阳性结果提供了更高的患者肝癌诊断可能性（24.0% vs. 28.1%）。但无论首选何种试验，阴性结果可以就此终止，而阳性结果都需进一步选择另一个试验加以验证。

如果患者基础状态存在肝硬化（患病率从1%提高到10%，AFP阳性，患病率提高到30.7%），CT检查或者磁共振检查阳性，则患者肝癌的可能性分别增加到93%和87%，都不需要进一步选择另一项检查印证，而可以直接处理。因此可见，基础患病率即来源人群不同明显影响诊断试验对结果的判定。

NOTE

表 9-12　增强 CT 与磁共振诊断肝癌检查比较

比较内容	增强 CT 检查	磁共振检查
诊断基础	影像学特征，占位	影像学特征，占位
诊断效率		
敏感度	90%	86%
特异度	77%	81%
阳性结果似然比	3.9	4.5
阴性结果似然比	10/77	14/81
诊断试验特征		
可获得性	三级医院	三级医院
易接受程度	中等（相比抽血/造影或活检）	中等（相比抽血/造影或活检）
费用（元）	600～1000	600～1000
安全性	有造影剂过敏可能，可能带来肾脏影响	过敏可能小，体内存在金属物不适用
在该患者中的应用		
基础患病率 1%+AFP 阳性	24.0%	28.1%
交叉检查阳性结果	58.7%	58.7%
基础患病率 10%+AFP 阳性	77.6%	80.0%
交叉检查阳性结果	94%	94%

是否 CT 阳性肝癌诊断达到 77.6% 的可能性或者磁共振阳性肝癌诊断达到 80% 的可能性就一定能选择手术治疗呢？或者 CT 与磁共振检查其中一项为阳性结果，另一项为阴性结果，应考虑进一步检查如血管造影检查（表 9-13）。其敏感度和特异度均达到 98% 和 98%，阳性结果似然比为 49，阴性结果似然比为 1/49。在很多情况下血管造影或者穿刺组织学检查作为评价 CT 或者磁共振等检查的金标准。但因其具有一定创伤性、存在一定风险、部分患者难于接受且费用较高等，只有在存在争议、各种检查不一致的情况下临床才选择。

表 9-13　血管造影诊断肝癌临床评价

	金标准评价最终结果	
	肝癌	非肝癌
血管造影判断（+）	98（98%）	2（2%）
血管造影判断（-）	2（2%）	98（98%）

敏感度=98%，特异度=98%，阳性预测值=98/100×100%=98%，阴性预测值=98/100×100%=98%，阳性似然比=98%/2%=49，阴性似然比=2%/98%=1/49。

假设 CT 检查结果阳性，患者来源于肝硬化人群，肝癌可能性 77.6%。若血管造影结果阳性，疾病的可能性即概率增加，支持及时手术治疗，没有改变选择；若试验结果阴性，疾病的可能性即概率会减少。根据试验的敏感度和特异度，推算验后概率为 6.6%，低于 25%，仍应该选择进行这项检查。

阴性似然比为：（1-敏感度）/特异度=（1-98%）/98%=1/49；验前概率为 77.6%，验前比为：77.6%/22.4%=3.46；验后比为：3.46×（1/49）=0.071；则：验后概率=0.071/（1+0.071）×100%=6.6%

特别需要强调的是，利弊可以用临床疗效和不良反应评估，也可以包括生存质量、临床成本。假如知晓抗病毒治疗与暂时不抗病毒治疗的利弊，就可回答前面案例中患者诊断问题，通过磁共振弹性成像诊断早期肝硬化验后概率上升到 67.6%，是否需要选择抗病毒处理，或者是否有必要进一步采用创伤性检查即肝穿刺活检评估诊断。权衡利弊还需要考虑决策时间，临床上疾病常常处于动态变化过程中，临床疾病诊断的概率也在不断变化，对于不能等待的疾病或者状态，需要及时决策；对于可等待的疾病或者状态，可以选择暂时观察，但需要评估等待的风险。

（刘爱忠）

笔记

第十章　疾病治疗的研究、评价与循证实践

疾病治疗研究与评价是临床实践和临床研究工作的重要内容。目前由于一些疾病的治疗方法和手段日趋多样化，如何从中选择安全、有效的干预措施以及发现优质证据用于指导临床决策，已成为当前临床工作的一项重要任务。

治疗性研究与评价，既可进行原始研究与评价，也可进行二次研究与评价。内容较为广泛，涉及药物、术式等治疗手段，以及各种预防性干预措施等。

第一节　概　　述

一、治疗性研究与评价的重要性

首先，治疗措施在体外实验中观察到的效果须经严格的临床验证才能说明其是否有效。以动物为研究对象或体外实验完成的结论不能直接用于人体。如2型糖尿病动物模型的基因治疗研究中，发现有一定的治疗作用，但用于人体研究时，却未能获得相同的结果；一些肿瘤采用动物模型进行药物治疗性研究，也发现了一定的治疗作用，但用于人体研究时，结果却未能重复。

其次，治疗性研究的对象是患者，决定了研究的复杂性。患者的个体特征、疾病严重程度、心理及生理状态差异性，以及药物敏感性、医生诊治水平等因素的复杂作用，不可避免地影响研究结果的真实性及可靠性。临床试验过程中，特别是在设计、资料收集、分析评价等环节均可能受偏倚的影响，得出错误的结论。因此，科学严谨的科研设计尤为重要。只有经过严格的设计、测量和评价，才能确定干预措施的真实效果。

最后，临床经验并非都是科学证据，只有经过科学评价的经验才能被确认为证据。有些经验是经不起历史考验的。如：曾风行一时的胃冰冻疗法治疗消化性溃疡、乳房内动脉结扎治疗顽固性心绞痛以及国内曾盛行的鸡血疗法、甩手疗法、针灸哑门穴治疗聋哑等，后经科学证实均无效。

近些年来，随着临床流行病与循证医学的普及和发展，许多临床医生已逐渐认识到应用科学的研究方法对临床疗效评价的重要性，开始学习并接受有关临床研究的方法学知识，开展高质量的临床研究，如多中心的随机对照试验、登记注册研究、制定临床指南等等，从而不断获得高质量的临床研究结果，提高临床医学研究和临床工作的水平。

二、治疗性研究与评价的基本条件

（一）选题应有充分的科学依据

临床疗效评价研究的选题首先要具有科学依据，要清楚地阐明研究背景和立题依据，不能仅凭经验来做假设或推断。赫尔辛基宣言（*The Declaration of Helsinki*）目前仍然是人体试验应遵循的基本准则，要求用于人体治疗性试验的任何药物或措施，在疾病发病机制或对某个环节的干预方面应有明确的理论支持，同时还应具备药物化学、药理学、药动学以及药效学等研究基础。只有在初步证明具有安全性和有效性之后，方可投入临床试验。如非小细胞肺癌患者只有在EGFR阳性的条件下才能给予易瑞沙治疗，乳腺癌患者需在Her-2阳性时考虑曲妥珠单抗治疗。

（二）明确拟达到的最佳目标

不同疾病均具有各自特点和不同的结局表现，同时药物或治疗措施也都有着各自不同的作用机制，并非均可达到疾病根治的目的。因此，应结合疾病的具体特点和试验药物本身的效能设置预期目标。除要考虑治愈外，还应考虑症状缓解，功能维持，预防复发或并发症等。当然，如果所选择的疾病可根治，那么“治愈”始终应作为首选的最佳目标，如外科手术可根治早期肿瘤等。而其他一些疾病如果难以治愈，那么就要结合这些疾病预后的特点选择切实可行的目标，从而达到研究的目的。如对脑梗死恢复期患者，临床治疗的目标应是预防其再次复发；而一些不能彻底治愈的慢性病患者，临床治疗的目标主要在于缓解症状，最大限度地改善其功能状态及生存质量。如类风湿疾病治疗等，其目的就是控制疼痛、防止畸形、改善生存质量等。

（三）选择最有效的试验药物或措施

药物或临床治疗性措施的选择，除了要有科学依据及安全有效性证据外，还应考虑从同类药物或措施中选择最为有效者进行试验，避免低效或无价值的重复试验，如在多种降压药物中选择更安全、有效、经济的药物进行研究。此外，尽管有些药物或措施可能不会显著提高疗效，但其副作用发生率或者成本下降，也应优先考虑进行试验。

（四）选择最佳治疗水平的终点指标

评价治疗效果要选择合适指标，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总生存期，治疗高血压预防心、脑、肾等靶器官损害事件的发生率。中间指标是指疾病发展变化过程中的一些结果，如药物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。治疗终点指标的选择应基于疾病的性质、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而定。

（五）明确研究中伦理问题

在临床研究中，所有以人为对象的研究必须符合伦理学的相应规定，即公正、尊重人格、力求使受试者最大程度受益和尽可能避免伤害发生等。

治疗性研究中涉及的伦理问题应从以下方面考虑：即方案设计与实施、试验的风险与受益、受试者的招募、知情同意书告知的信息、知情同意的过程、受试者的医疗和保护、隐私和保密、涉及弱势群体的研究等。

三、治疗性研究及其试验分期

临床试验是治疗性研究中最常见的研究方案，其中以新药临床试验居多，按照新药的不同研发阶段可以分为以下四期：

（一）Ⅰ期临床试验

Ⅰ期临床试验为初步的临床药理学及人体安全性评价，是在大量实验室研究与动物实验基础上，将新疗法开始用于人类的试验，是进行人体新药临床试验的起始期。目的在于了解剂量反应与毒性，进行初步的安全性评价，主要研究人体对新药的耐受性及药代动力学，了解药物在人体内的吸收、分布、消除的规律，以提供初步的给药方案。受试对象一般为健康志愿者，在特殊情况下也选择患者作为受试对象，如肿瘤患者。Ⅰ期临床试验的样本量为20～30例。

（二）Ⅱ期临床试验

Ⅱ期临床试验主要对新药的有效性、安全性进行初步评价，以确定药物的疗效适应证，了解药物的毒副反应，确定给药剂量。一般采用严格的随机对照试验，以平行对照为主。通常应该与标准疗法进行比较，也可以使用安慰剂。需注意诊断标准、疗效标准的科学性、权威性和统一性。要根据试验目的选择恰当的观测指标，包括诊断指标、疗效指标、安全性指标。选择指标时，应注意其客观性、可靠性、敏感性、特异性、相关性和可操作性。参照临床前试验和Ⅰ期临床试验的实际情况制订药物的剂量研究方案。应有符合伦理学要求的试验中止标准和退出标准。对不良事件及不良反应的观测、

笔记

判断和及时处理都应作具体规定。应有严格的观测、记录及数据管理制度。试验结束后，对数据进行统计分析，评价药物的安全性、有效性、使用剂量等。Ⅱ期临床试验的对象为患者，试验组和对照组的例数都不得低于100例。

（三）Ⅲ期临床试验

Ⅲ期临床试验是Ⅱ期临床试验的延续，为扩大的、多中心随机对照临床试验，旨在进一步验证药品的有效性和安全性，评价利弊，最终为药物注册申请的审查提供充分的依据。可根据试验目的调整受试者的入选标准，适当扩大特殊受试人群，进一步考察不同对象所需剂量及其依从性。Ⅲ期临床试验的试验例数一般不低于300例，对照组与治疗组的比例不低于1∶3，具体数量应符合统计学要求。

（四）Ⅳ期临床试验

Ⅳ期临床试验也称上市后监测研究，为新药的上市后研究。旨在评估药物在更大范围、长期的实际应用中的疗效以及监测不良反应，特别是罕发、严重的不良反应事件。此外，还应进一步考察对患者经济水平与生存质量的远期影响。Ⅳ期临床试验应在多家医院进行，观察例数通常不少于2000例。

临床试验的样本量除满足上述最低标准外，还需要同时满足统计学要求。由于新药临床试验是应用尚在研究中的新药，在人体进行的安全、有效性评价，故研究者对受试者的安全负有重要责任。因此需要在国家批准认证的药物临床试验机构（GCP）进行。

新药临床试验在实施过程中多采用多中心临床试验的方式。除此之外，治疗性研究也可以采用社区干预试验或现场试验的方式，以上相关试验类型的特点及优势，请见本教材第四章。

第二节　治疗性研究的设计要素及方法

一项高质量的临床治疗性试验的前提，必须要有良好的选题和明确的研究目的。进而基于研究目的明确研究方案、设立指标、方法及统计分析手段。治疗性研究与其他临床研究设计内容类似，主要包括以下几个方面要素：

一、选择研究方案

治疗性研究有多种方案可以选择，但从研究的质量方面考虑，多选择试验性研究方案，如随机对照试验（randomized controlled trial，RCT）、单病例随机对照试验（N of 1 RCTs）、交叉试验（cross-over design，COD）、前-后对照试验（before-after study）、非随机对照试验（non-randomized controlled trial，NRCT）、历史对照试验（historical control trial，HCT）、序贯试验（sequential trial）等。但从临床实用性方面考虑，还可采用观察性研究，如队列研究、病例对照研究等。证据强度要弱于试验性研究。

有关上述研究方案的定义、特点及应用条件在本教材第四章中已有详细介绍，在此不再赘述。近年陆续有新的试验方案出现，如适应性设计、成组序贯设计、析因设计等。

1. 适应性设计（adaptive design）　是指在临床试验开始后，根据试验中已经积累的信息，动态修改试验设计的某些方面，但又不破坏试验有效性、安全性和完整性的一种试验设计。对试验设计的修改可在一个方面，也可在多个方面。

适应性设计可包括：试验组和对照组入组分配方式的改变，如由固定区组分配变更为动态随机入组分配；受试者入组数量的改变，如样本量的重新计算；试验终止条件的改变，如根据期中分析结果提示有效或无效性而提前终止试验；统计分析方法，临床终点的改变等。

适应性设计可加快研究的速度和提高研究的真实可靠性，但值得注意的是：试验的修改是否会引起Ⅰ类错误增大以及是否会导致研究结果难以解释。因此，对方案的任何修改都需谨慎，防止可能引入的偏倚而影响结论的准确性。

2. 成组序贯设计（group sequential design）　是把整个试验分成若干个连贯的分析段，每个

分析段病例数可相等也可不等，但试验组与对照组的病例数比例应与总样本中的比例相同。每完成一个分析段，即对主要指标[包括有效性和(或)安全性]进行分析，一旦可作出结论即停止试验，否则继续进行。如果到最后一个分析段仍不能拒绝无效假设，则作为差异无统计学意义而结束试验。

成组序贯设计主要用于：怀疑试验药物有较高的不良反应发生率时，可用此设计较早终止试验；试验药疗效较差时，用此设计可因药物无效而提前终止试验；或试验与对照的疗效相差较大而但病例稀少时或者临床观察时间过长也可适用。

成组序贯设计的优点是当试验药物与对照药物之间确实存在差异时，可较早地得到结论，从而缩短试验周期。但要注意Ⅰ类错误增加，同时由于需要进行多次期中分析，要全程保持盲态，以免引入新的偏倚。

3. 析因设计（factorial design） 是将两个或两个以上因素及其不同水平进行排列组合、交叉分组的试验设计，可探讨各因素作用以及各因素间的交互作用，又称交叉组设计。

在治疗性研究中，主要是通过试验药物剂量的不同组合，对两个或多个试验用药同时进行评价，这样不仅可检验每个试验用药不同剂量间的差异，而且可检验各试验药物间是否存在交互作用，或探索两种药物不同剂量的适当组合。

除上述方法外，近几年实效性研究在治疗性研究中也受到广泛关注，并逐渐在研究中被应用，如实效性 RCT，请参考本教材第四章第八节。

二、选择研究对象和样本量估算

（一）研究对象的入选

首先需要明确研究对象的来源，包括是哪一级医院，是门诊患者还是住院患者。明确符合公认的临床诊断标准，同时根据研究目的和要求，设置合适的纳入标准及排除标准，以保证入选对象具有较高的同质性和代表性，有利于减少偏倚的发生，获得准确的研究结果。在设置排除标准时，应特别明确不宜使用该药的人群，如心、肺、肝、肾功能不全者和小儿、孕妇、哺乳期妇女、计划近期怀孕者等；同时对该类药物过敏和其他不宜参加这项研究者，如依从性差、参与了其他药物临床试验的受试者也应排除。但要注意：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外推性及适用性。

此外，根据医学伦理学的原则，凡参加临床试验者，都要签署知情同意书。如某一新型抗生素治疗急性细菌性感染的疗效研究中，设置的纳入标准包括：年龄 18～65 岁的住院或门诊患者（慢性支气管炎急性发作部分患者年龄可放宽至 70 岁）、性别不限、经临床和实验室检查确诊为细菌感染、试验前未用过其他抗生素治疗或经其他抗生素治疗无效而细菌学检查为阳性并单独用此药治疗者、签署了知情同意书的受试者。并同时规定了排除标准：如对抗生素有过敏史者；严重心、肝、肾功能不全者；有精神、神经系统疾患以及晚期肿瘤患者；妊娠、哺乳期妇女；依从性差或病情严重，不能完成试验者等。

（二）样本量估算

临床治疗性研究需要有足够的样本数，以确保对所提出的问题给予一个可靠的回答。样本大小通常以试验的主要疗效指标来确定，如果需要同时考虑主要疗效指标外的其他指标时，应明确说明其合理性。

一般来说，在样本量的确定中应该明确以下相关因素：研究的设计类型、主要疗效指标的定义、临床上认为有意义的差值或者率、检验统计量、检验假设中的原假设和备择假设、Ⅰ类和Ⅱ类错误率以及可能脱落率等。较为常见的样本量计算方法如下。

1. 两组率的比较 是以率作为主要分析指标，可以根据式(10-1)计算出各组所需的试验样本量。

笔记

$$n=\frac{p_1(100-p_1)+p_2(100-p_2)}{(p_2-p_1)^2}\times f(\alpha,\beta)$$　　式(10-1)

其中 n 为一个组的样本量；p_1，p_2 为试验组和对照组的事件发生率；$f(\alpha,\beta)$ 为限定假阳性和假阴性水平时相应的参数值，可以由表 10-1 查出。

表 10-1　常用 $f(\alpha,\beta)$ 数值表

α	β			
	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：用两种药物对糖尿病患者进行康复治疗，经初步观察发现，甲药的有效率为 70%，乙药的有效率为 90%，现要进一步试验，设 $\alpha=0.05$，$\beta=0.1$，问每组至少需要观察多少病例？

假设已知 $p_1=70\%$，$p_2=90\%$，$\alpha=0.05$，$\beta=0.10$，查表 10-1 得，$f(0.05,0.1)=10.5$，代入上述公式得：

$n=\frac{70\times(100-70)+90\times(100-90)}{(90-70)^2}\times 10.5=79$，即每组需观察 79 个病例。

2. 两组均数的比较　其样本量估计可采用式(10-2)。

$$n_1=n_2=2\left[\frac{(t_\alpha+t_\beta)\times S}{\delta}\right]^2$$　　式(10-2)

此式中的 n_1、n_2 分别为 1 组、2 组的样本量，一般是相等的，S 为两总体标准差 σ 的估计值，一般假设其相等或取合并方差的平方根，δ 为两均数的差值，即最小效应量的估计，t_α 和 t_β 分别为检验水准 α 和第Ⅱ型错误的概率 β 相对应的 t 值，可以通过查 t 值表得到。

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，设两种药物疗效标准差相等，均为 0.7 分，要求 $\alpha=0.05$，$\beta=0.1$，若要得出组间差异有显著性结论，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 t 值表得：$t_{0.05/2,\infty}=1.96$，$t_{0.1,\infty}=1.28$，代入公式得：$n_1=n_2=2\left[\frac{(1.96+1.28)\times 0.7}{0.5}\right]^2=41.2=42$（例），故认为两组各需 42 例患者，共计 84 例。

在实际工作中，也可采用下述公式进行计算，较为方便。

$$n=2\hat{\sigma}^2\times f(\alpha,\beta)/(\mu_2'-\mu_1')^2$$　　式(10-3)

n 为每组所需的例数，μ_2'，μ_1' 分别为两组的预期均数，$\hat{\sigma}$ 为两组的合并标准差或对照组的标准差，$f(\alpha,\beta)$ 可由表 10-1 中查出。用此公式计算结果相同。

考虑试验中出现的病例失访和退出情况，故常在计算量的基础上增加 10%。

3. 非劣效性研究

（1）当主要结局指标为定量资料，且试验组与对照组的样本量相等时，样本量计算公式为：

$$n=2\left[\frac{(\mu_\alpha+\mu_\beta)S}{\Delta-\delta}\right]^2$$　　式(10-4)

式中 n 为每组的样本含量。μ 为标准正态离差界值，μ_α 和 μ_β 是检验水准 α 和Ⅱ型错误 β 对应的 u 值，S 为两组的合并标准差。Δ 为非劣效界值，δ 指容许误差，为两组的效应值之差。

例 10-1　为验证某种新型胶囊内镜的成像质量不低于市面上某种胶囊内镜，以两种胶囊内镜系统的成像质量评分作为主要评价指标，根据预试验测得新型胶囊内镜的成像质量平均得分为 10.4 分，标准差是 0.517，对照胶囊内镜的成像质量平均得分为 11.5 分，标准差是 0.529，计算得出两组合并后的标准差是 0.523，两组的效应差值 1.1 分，为并根据临床实际，设 1.5 分为可接受的非劣效性界值，取 $\alpha=0.025$（单侧），$\beta=0.1$。

将以上参数代入公式计算得到：$n=2\left[\dfrac{(1.96+1.28)\times 0.523}{0.4}\right]^2=35.89\approx 36$

每组需要完成有效病例 36 例，可以在把握度 90%，检验水准 0.025（单侧），非劣效性界值 1.5 的情况下验证非劣效性结论。

（2）当主要结局指标为分类资料，且试验组与对照组的样本量相等时，样本量计算公式为：

$$n=\frac{(Z_\alpha+Z_\beta)^2[p_t(1-p_t)+p_c(1-p_c)]}{(\Delta-\delta)^2} \qquad \text{式(10-5)}$$

式中 n 为每组的样本含量，Z_α 和 Z_β 是检验水准 α 和Ⅱ类错误 β 对应的 Z 值，P_t 是试验组的发生率，P_c 是对照组的发生率，Δ 为非劣效界值，δ 指容许误差，为两组的发生率之差。

例 10-2　为验证纳米银妇用抗菌剂治疗念珠菌性阴道炎的临床效果不劣于硝酸咪康唑栓，以两种药物治疗患者后的显效率作为主要评价指标，根据文献查阅以及预实验结果可知纳米银妇用抗菌剂和硝酸咪康唑栓治疗念珠菌性阴道炎的临床显效率均为 80%，根据临床实际设 15% 为可接受的临床非劣效性界值，取 $\alpha=0.025$（单侧），$\beta=0.1$。

将以上参数代入公式计算得到：

$$n=\frac{(1.96+1.28)^2[0.8\times 0.2+0.8\times 0.2]}{0.15^2}=149.29\approx 150$$

每组需要完成有效病例 150 例，可以在把握度 90%，检验水准 0.025（单侧），非劣效性界值 0.15 的情况下验证非劣效性结论。

三、研究对象的随机化分配及设盲

合格的受试者一旦入选临床试验，需进一步实施随机化分组，以保证每位受试者均有同等的机会被分配到试验组或对照组中，基本保证各种影响因素（包括已知和未知的因素）在处理组间的分布趋于相似。随机化方法包括简单随机、区组随机、分层随机等，一般多采用区组随机化法或分层随机化法。如果受试者的入组时间较长，区组随机化是临床试验所必需的，这样有助于减少季节、疾病流行等客观因素对疗效评价的影响，如果药物效应会受到一些预后因素（如受试者的病理诊断、年龄、性别、疾病严重程度、生物标记物等）的影响时，可采用分层随机化，但分层因素一般不宜超过 3 个。临床试验较为常用的是采用中央随机化系统，也可以通过计算机产生随机序列或查阅随机数字表产生随机数字。

此外，为避免临床试验中的测量偏倚，也需要设立盲法。根据设盲的对象和范围，可分单盲临床试验，双盲临床试验和三盲临床试验，其中以双盲试验最为常用。

随机分配方案的隐藏（allocation concealment），顾名思义就是对分配方案设盲，使研究参与者无法知晓某个患者所接受的是何种干预方案。随机分配隐藏的最佳方法是采用中央随机化系统管理受试者的入组，也可将产生的随机分配序列号用信封密封，同时还要保证试验参与者在试验过程中尽量处于盲态等。

四、选择试验药物或措施

试验组药物或措施首先要保证其有效性和安全性，并同时具有一定的创新性和经济性。对照组药物，可以是阳性药物对照，也可以是安慰剂对照，在选择阳性药物作为对照时，要注意判断是否满足作为阳性对照药物的条件，选择安慰剂作为对照，要注意选择研究的病种及是否违反医学伦理。无论是阳性对照药物还是安慰剂对照，试验、对照两组的药物在外观、色泽、大小等方面应相似，同时服用方法和疗程也要一致，否则会影响结果的真实性。试验药物或措施与对照药物或措施在剂量、用法、疗程、时机的设置和确定不能盲目，要有科学依据。

五、试验观察指标的设置

观察指标是指能反映临床试验中药物有效性和安全性的观察项目，必须在研究方案中有明确的定义和可靠的依据，不允许随意修改。在研究设计阶段，首先需要根据研究目的，严格定义与区分主要指标和次要指标，明确指标的性质（定量或定性）和特征（数量、主观或客观、终点指标或替代指标等）。

临床试验中选择的观察指标应具有良好的敏感度，较强的特异度，充分的客观性、清晰的终点指标，适当的指标数量、明确的可行性等。临床研究常用的几类指标如下。

（一）主要指标和次要指标

主要指标是研究的主要终点，能确切反映药物有效性或安全性。主要指标应根据试验目的选择易于量化、客观性强、重复性高，并且公认的指标。主要指标多为一个，但有时也可采用两个或多个主要指标。主要指标在试验进行过程中不得修改。次要指标是与次要研究目的相关的效应指标，或与试验主要目的相关的支持性指标。一项临床试验研究，可以设计多个次要指标，但不宜过多。

（二）复合指标

当单一的主要指标难以确定时，可将多个指标组合构成一个复合指标。临床上采用的量表（如神经、精神类、生存质量量表等）就是一种复合指标。

（三）替代指标

替代指标是指无法直接评价临床效果时，用另外的指标间接反映临床效果。例如对高血压患者的降压药物治疗，治疗的效果被认为是降低或延迟“终点事件”（心、脑、肾事件）的发生，但若采用此项指标评价，需要长时间的观察。而在临床研究实践中，降压药的治疗效果经常采用替代指标“血压降低值/血压达标”来评价药物的疗效，因为许多研究已经证实，将血压控制在正常范围内，可降低“终点事件”的发生。

选择替代指标为主要指标，可缩短临床试验期限，但也存在一定的风险，尤其是“新”替代指标。药物在替代指标上的优良表现并不一定代表药物对受试者具有长期的临床获益，相反，药物在替代指标上的不良表现也不一定表示无临床获益。因此在具体应用时需要仔细权衡。

（四）转化指标

在临床试验中，有时将根据临床研究的需要，将定量指标根据一定的标准转换为等级指标或将等级指标转化为定性指标，如：用药后血压降低到“140/90mmHg”以下的受试者比例（达标率）。这种指标转化的前提是要具有临床意义且已被公认。

六、确定干预实施方法及随访观察期

在研究设计中，需要说明试验干预方法和实施要求，建立试验药物包装、分发、转运、供应档案。设立盲法时，则有保证盲法的具体措施，以及当患者病情恶化或突发不良事件时，有紧急破盲的操作规范；同时还要建立避免沾染和干扰以及保证依从性的制度等。

临床试验中同样需要确立一个明确的随访观察期，观察期要适当、不宜过长或过短。过长会造成不必要的浪费，过短不但无法观察到远期效果，还无法观察到有效的治疗结局。观察期的确定应基于研究目的、前期基础研究结果以及临床达到治疗最佳水平所需时间等。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。

七、数据管理与统计分析

（一）数据管理

临床试验方案确定后，应根据病例报告表和统计分析计划书的要求制订数据管理计划，包括数据接收、录入、清理、编码、一致性核查、数据锁定和转换等，最好采用临床试验的数据管理系统进行管理。

（二）统计分析

与其他研究一样，临床试验也需在原始资料完整、准确的基础上，按研究目的、试验设计方案，以及资料类型选择正确的统计方法处理资料。内容包括：详细列出主要和次要指标的分析方法、亚组分析方法、中心效应；详细比较进入试验组和对照组患者的基线特征与协变量分析；失访、退出和脱落病例的情况；如何处理偏倚和缺失数据，如何解释结果的意义；如实报告试验结果的有效性、安全性等等。

1. 统计分析集的划分　首先要明确定义用于统计分析的数据集，确认每位受试者所属的分析集。一般情况下，临床试验的分析数据集包括全分析集（FAS）、符合方案集（PPS）和安全集（SS）。全分析集用来描述尽可能完整且接近于包括所有随机化受试者的分析集；符合方案集是全分析集的一个子集，是指符合研究方案要求且依从性好的受试者；安全集通常指包括所有随机化后至少接受一次治疗且有安全性评价的受试者。

2. 统计方法的选择

（1）根据资料的性质采用相应的统计学方法

1）治疗性研究资料中最常见的是计数资料、计量资料和等级资料。计数资料的描述采用率和比，如有效率、治愈率、病死率等，假设检验方法常为卡方检验（χ^2-test）；计量资料的描述常用均数±标准差，假设检验常用 t 检验（小样本），u 检验（大样本）及 F 检验、非参数检验等；等级资料是将某一指标划分为若干等级，常用非参数检验及 Ridit 分析等。

2）如治疗性研究本身有两组以上组别比较，必须先作多组间的显著性检验，差异有统计学意义时，才能进一步作两两比较。

3）治疗性研究设计可以是配对设计，或完全随机设计。两种设计的原理不同，分析处理的方法也不同，二者不可混淆。

（2）根据分析目的选用统计方法

1）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，还与患者人口学特征以及病理生理状态、疾病特点等有关。例如年龄、营养状态、病情、药量、疗程、并发症、合并症等，均会影响治疗反应。为此，应在单因素分析的基础上，选择具有显著意义的有关变量作多因素分析，进一步评价疗效。

2）单侧检验或双侧检验：如肯定试验药（或措施）疗效优于对照药（或措施），则用单侧检验法；若不能肯定，则采用双侧检验法。如果对照组用阳性对照药物，则需要进行等效或者非劣效性检验，此时用单侧检验即可。

3）对不完整资料的处理与分析：对于在研究中出现退出、失访的情况，可以采用意向治疗分析（intention-to-treat analysis，ITT）。ITT 分析是将所有纳入随机分配的患者，不管最终是否接受到分配的治疗，在最后资料分析中都应按随机分配方案统计，以保证结论更真实可靠。

笔记

第三节　影响治疗性研究质量的常见因素及其处理方法

治疗性研究中存在很多影响结果真实性和可靠性的因素，如偏倚、混杂和机遇因素等，若不加以识别和控制，临床研究结果将会失真。

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，只能在研究设计阶段，通过限制Ⅰ型错误率和Ⅱ型错误率，使机遇因素的影响控制在容许范围之内。

二、偏倚

偏倚（bias）又称偏性，是在整个临床试验过程中产生的并且能够干扰疗效和安全性评价的系统误差。由于偏倚会影响疗效、安全性评价结果，甚至影响临床试验结论的正确性，因此在临床试验的全过程中均须控制偏倚的发生。

偏倚在临床试验中呈多样化表现，除常见的选择偏倚（selection bias）、信息偏倚（measurement bias）和混杂偏倚（confounding bias）外，还存在一些临床治疗性研究特有偏倚。随机化和盲法是控制偏倚的重要措施。

（一）选择偏倚

该偏倚主要产生在设计阶段，入选研究对象与未入选者在某些特征上存在系统差别，同时除研究因素以外的其他因素组间分布也不均衡，导致研究结果出现偏差。选择偏倚可以通过随机、设立对照、严格纳入/排除标准等方法加以控制。

（二）测量偏倚

测量偏倚指在资料的观察、测量及收集过程中，由于受到人为因素的影响导致结果失真的一类偏倚。通过实施盲法、标准化测量、提高应答率等方法可以防止这类偏倚的发生。

（三）混杂偏倚

混杂偏倚是指由于混杂因素的存在，掩盖或夸大了研究因素与疾病间的关联度。可以通过限制、配比、随机、标准化、分层及多因素分析方法加以控制。

（四）临床治疗性研究中还存在一些属于测量偏倚的特殊偏倚，如有以下几种：

1. 霍桑效应（Hawthorne effect）　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。

2. 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实施规范化试验及改善依从性，尽可能减少其他药物的使用。

3. 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并在试验过程中加强质量控制，改善依从性。

4. 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，这种现象或许属生理性波动，而非干预的结果，但可造成治疗有效的假象。克服的办法是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。

NOTE

5. 失访　当大量患者失访时，研究也可以产生与结局测量有关的偏倚。因为失访患者如果有不同的事件发生率，仅仅依靠随访到的患者的数据资料将会得到与真实情况不同的结果。克服的方法是提高随访率或者采用 ITT 分析。

6. 依从性（compliance）　是指研究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；反之则为依从性不好或不依从（non-compliance）。依从性差的原因包括：简单地遗忘、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患者充分理解试验目的、要求及参加这项试验的意义，使患者在理解的基础上给予合作。此外，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，在复诊时计算患者的依从性（式 10-6），如服药量≥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患者的依从性。另外，选用高危人群作研究对象对提高依从性也很有意义。

$$\text{依从性}=\frac{\text{实际用药量（应服药量-复诊时剩余药量）}}{\text{总设计用药量（应服药量）}}\times 100\% \qquad \text{式（10-6）}$$

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，使本应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　治疗性研究的评价原则

治疗性研究结果只有通过科学地评价才能区别干预措施的利弊以及决定能否用于临床实践。无论是单个原始研究评价还是系统综述或 meta-分析等二次研究评价，都可以概括为真实性（validity）、重要性（importance）、适用性（applicability）三个方面。

一、单个原始研究评价

（一）真实性评价

1. 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，故为治疗性研究的金方案。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，是否交待了具体的随机方法，是否实施随机分配隐匿，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。

2. 所有纳入的研究对象是否随访完整？研究对象随访时间是否足够　任何观察病例的丢失，都会直接影响最后的结果和证据的真实性。例如，疗效差的患者退出，会导致治疗效果高估；若个别患者因药物或者干预措施的副作用从治疗组中退出，可能会低估药物的危害性。理想的情况是所有纳入的研究对象在研究过程中都没有失访，但这一点在实际临床研究中是很难保证的。

失访率一般应控制在10%以内，若超过20%，研究质量会受到很大影响，结果也许会不真实。对于一般失访率影响的估计，常用的方法是将试验组全部失访的病例，均计算在本组无效病例中，对照组丢失的病例全部计入有效病例内。如果仍然与原有效结论一致，则可以接受原来的结果。

同时，应该确保随访期足够长，以保证获得重要的临床效应结果。随访的具体时间取决于目标疾病的病程特点，通常临床观察的疗程至少数月，有的甚至需要1年以上方能充分显示防治措施的重要效果。

3. 是否对随机分组的所有研究对象进行了分析（是否采用意向治疗分析）　被随机分配入组的病例可因为各种原因出现失访。例如，因副作用发生出现中途停药者；患者依从性差而未认真按医嘱服药者以及发生沾染或者干扰者等。如果这部分研究对象不被纳入结果分析，必然影响原来的随机化原则和基线的可比性，最终影响结果的真实性。目前广泛采用意向治疗分析（intention to treat analysis）进行资料分析，即按最初随机分配入组的全部病例，无论其是否接受或未接受确切的治疗药物，都纳入进行最后分析。

4. 是否对研究对象、医生和研究人员实施了盲法　在治疗性资料的收集过程中往往也会产生较大的测量偏倚。如受试研究对象知道自己接受的是治疗措施、还是对照措施；研究者或结果测量者知道研究对象的分组情况，往往会夸大研究药物的效应等。实施盲法的目的就是为了减少测量性偏倚以维护观察结果的真实性。盲法可以是单盲、双盲或三盲。而对于无法实施盲法的试验（如外科手术），可以请另外的医生采用客观指标评价治疗效果。若使用临床记录，则应去除所有可能涉及破盲的治疗措施的信息，以保证真正的采用了盲法。

5. 除试验方案不同外，各组患者接受的其他治疗方法是否相同　如果研究对象除了接受规定的治疗方案外，还有意或无意采用了其他的具有类似作用的干预措施，必然影响结果的真实性。沾染和干扰即是常见的两种情况。前者是指对照组的患者接受了试验组的防治措施，使得试验组和对照组间的疗效差异减小；后者是指试验组或对照组接受了类似试验措施的其他处理，人为扩大或减小组间疗效的真实差异。

如果一项随机对照试验不符合上述五条标准的一条或多条，则需要考虑其研究结果的真实性。

（二）重要性评价

1. 治疗性证据的效应强度大小　疗效的强度通常用率表示，即有效率、治愈率、病死率、病残率，即使有些计量的疗效指标，亦多转换为“有效率”等。

当对RCT各组的有效率进行比较分析时，可以采用统计学方法判定各组间差异的显著性，然而，仅仅采用这些“率”对临床量化的表达是不足的。近些年来，随着临床流行病学与循证医学的发展，一些能够更好反映临床意义的试验结果指标被广泛应用，主要有以下几种：

（1）相对危险降低率（relative risk reduction，RRR）

$$RRR=\frac{CER-EER}{CER} \qquad 式(10\text{-}7)$$

注：CER=control event rate（对照组事件率）；EER=experiment event rate（试验组事件率）

（2）绝对危险降低率（absolute risk reduction，ARR）

$$ARR=CER-EER \qquad 式(10\text{-}8)$$

（3）需治疗多少病例数才获得一例好结果（number need to treat，NNT）或需治多少病例才发生一例不良反应（number needed to harm，NNH）

$$NNT\text{ 或 }NNH=\frac{1}{ARR} \qquad 式(10\text{-}9)$$

如一个应用他汀治疗预防脑卒中的RCT中，治疗5年后他汀组的脑卒中发病率为4.3%（EER），

对照组的脑卒中发病率为5.7%(CER);试验组不良反应率0.05%,对照组0.03%,根据以上公式,计算进一步量化的指标为:

$$RRR=\frac{5.7\%-4.3\%}{5.7\%}=25\%;ARR=5.7\%-4.3\%=1.4\%;$$

$$NNT=\frac{1}{1.4\%}=72;NNH=\frac{1}{0.02\%}=5000$$

由此可以看出,他汀治疗的不良反应发生率很低,相对而言是较为安全的。

2. 治疗性结果的精确度(precision) 研究中采用的各相关指标,仅表示着效应强度的大小,如有效率80%以及ARR、RRR、NNT的值等。但效应量的精度往往会因不同样本量大小而有所差异。因此,应进一步评估效应的精确度。

通常采用95%可信区间(95% CI)反映结果的精确程度,95% CI的范围越窄,则越精确。

(三) 适用性评价

上述证据的真实性与重要性经分析和评价获得肯定结论之后,紧接着要考虑这种有价值的证据是否可被应用于临床实践,需要结合自己患者的实际情况,评价其适用性。通常考虑以下几条规则:

1. 被评价的证据是否与患者情况不符而不能应用

(1) 整体证据:任何治疗性研究所产生的有价值的证据,应该审查疾病的诊断标准是否可靠,证据中的研究对象之纳入标准是否与拟引证的患者相符,其生理机能与病理学的依据、病情特点、年龄、性别以及社会经济状况是否存在显著差异等,假若以上特点一致或大体一致,则该治疗性证据就认为适用,否则就不可取。

(2) 亚组证据:有的证据在总体上可能缺乏适用性,但其亚组分析的疗效可能有显著意义,如果患者的病情与某亚组患者的病情相似,那么,这个亚组的治疗证据就有适用的价值。

对于有着显著适用的亚组分析证据,如能符合下述条件者,则有应用价值:

1) 确有生物学和临床依据者。

2) 确有统计学意义和临床价值者。

3) 亚组分析在研究设计阶段就已拟定,而不是在试验结束后有意为。

4) 仅属该研究中有限几个亚组分析的证据之一者(注意亚组不能过多)。

5) 该证据在另外的研究中可被证实者。

2. 拟采用的治疗证据是否可以在本单位医疗条件下采用 对于拟采用的有效治疗措施,需要在具有一定医疗水平、医疗条件的医院才能被采用,如医生的技术水平,医院的管理机制及设备条件、患者意愿以及经济的承受能力等。如像冠心病的介入治疗、风湿性心瓣膜病的换瓣治疗术等,即使这类治疗证明对患者有利并具颇佳效果,若不具备上述条件,也是不可行的。

3. 从该治疗性证据中,估计患者的利、弊 采用某一可行性好的最佳治疗措施一定是要保证利大于弊,且要有量化指标为依据。最直接的是用NNT(益处)及NNH(害处)进行评价。如果这两种指标不全,则有两种办法帮助解决。

一是确定患者预期事件发生率(patient's expected event rate,PEER),意思是如果患者不予治疗,其最终结局事件的发生率。这可以用治疗试验中的安慰剂对照组的事件发生率(CER)表示;如无CER证据,也可根据临床积累的未治或缺乏特效治疗观察结果值作为PEER参考值,如像急性心肌梗死患者PEER约15%;另外也可以亚组资料分层分析之相应的CER作为PEER等等。

当获得PEER、RRR等指标值后,用下列公式则可推算:

$$NNT\text{ 或 }NNH=\frac{1}{PEER\times RRR}\qquad 式(10\text{-}10)$$

笔记

其二是应用列线图，用已知的 PEER（或称不治疗的绝对危险率）与 RRR 数据，依据此两值直线连线的延伸，与 NNT 线上交叉点，即为该 NNT，十分方便（图 10-1）。

4. 考虑患者对于治疗措施的价值取向与期望

在循证治疗实践中，一定要尊重患者对治疗的价值取向，即愿意接受或者不愿意接受，或愿意接受哪一种备选方案（药物或有关治疗措施），而且要了解患者对治疗结局的有关期望。

首先，要保证不良反应最小化，同时还要注重疗效。因此，估计与掌握治疗的利弊比就十分重要了。通常应用治疗措施（或药物）的 NNT 与 NNH 计算其利弊比（likelihood of being helped vs harmed，LHH）。

$$LHH=\left(\frac{1}{NNT}\right)\Big/\frac{1}{NNH} \qquad 式(10\text{-}11)$$

如有关他汀药物治疗预防脑卒中的 NNT 为 72，NNT 为 5000，则：

$LHH=\left(\frac{1}{72}\right)\Big/\frac{1}{5000}=70$，这意味着有利面是不利面的 70 倍，是为安全和有效的治疗证据，显然 LHH 是越高越佳。

倘若同时存在几种备选药物，且它们的疗效与不良反应相似或有差异，但没有显著意义。对这些备选治疗措施（或药物），则应优先选择其成本（价格）低廉且疗效也好和安全的药物，在保证安全有效的基础上，尽可能地降低医疗成本。同时，对于任何治疗措施，一定要给患者尽可能的清楚解释，以保持良好的依从性。并在治疗过程中务必要认真观察治疗反应，关心与爱护患者，这有利于增进互信和睦关系，避免产生不必要的误解或纠纷。

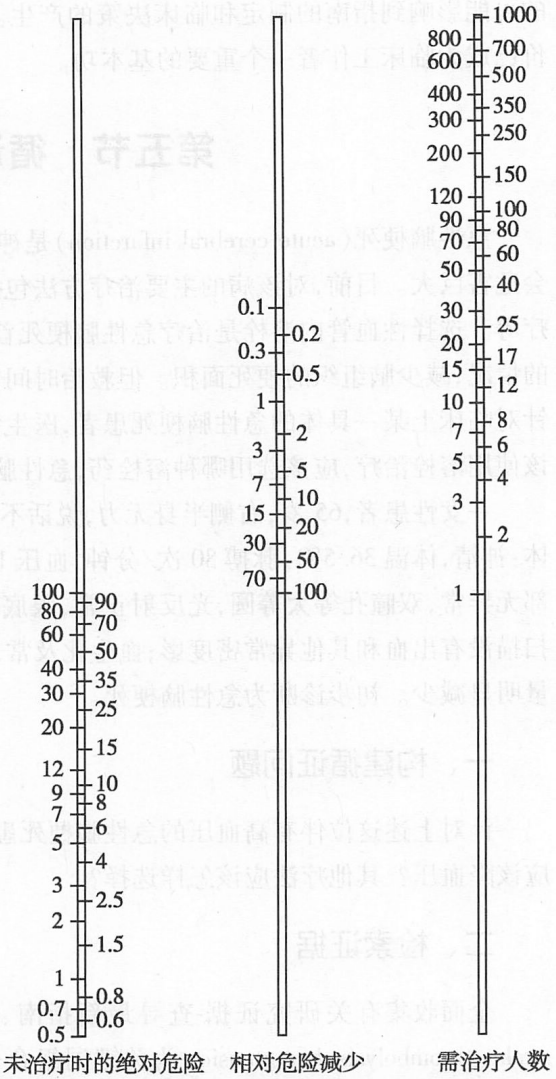


图 10-1　估计 NNT 的列线

二、系统综述或 meta-分析证据评价

系统综述是由多个原始研究结果综合而成，研究过程主要包括检索策略的制订，文献的收集、质量评价，证据结果的产生，偏倚的估计，敏感性分析等。尽管评价的主体内容仍然是围绕真实性、重要性、适（实）用性三个方面，但评价的内容与个体研究的评价比较存在一定不同，而且要借助评价工具完成。

AMSTAR（A Measurement Tool to Assess Systematic Review）量表，是系统评价/meta-分析质量评价的一种工具，用来鉴定系统综述的质量，帮助读者筛选出有价值的信息，理性采纳证据。

AMSTAR 1.0 量表是在 2007 年出版的，该量表由 11 个问题组成。经过几年的应用之后，发现了 AMSTAR 1.0 量表的一些不足，于是在 2017 年 9 月出版了 AMSTAR 2.0 改编量表，内容由原来的 11 条改为了 16 条。第二版与原版本比较，保留了原始版本中的十个方面，同时增加了部分内容。对有些条目的措辞也进行了不同程度的修改。判断评价的方法也有了一定的变化。在 AMSTAR 2.0 中每个方面问题只有一个“Yes”选项来表示积极的结果。去掉了原始版本中“不适用”和“不能回答”的选项，在某些情况下，提供了“partial Yes”，认为是确定部分遵守标准。这样更符合临床实际和实际

应用。

系统综述和 meta-分析位于循证医学证据金字塔的塔尖，其结果比任何其他类型的研究都有更大的可能影响到指南的制定和临床决策的产生。因此，能制作合格的系统综述/meta-分析并会正确评价已成为临床工作者一个重要的基本功。

第五节　循证实践及案例分析

急性脑梗死（acute cerebral infarction）是神经系统的常见病，致死率及致残率均较高，对人类及社会危害巨大。目前，对该病的主要治疗方法包括：溶栓、抗血小板、抗凝、脑保护剂、降纤治疗及手术治疗等。选择性血管内溶栓是治疗急性脑梗死首选方法之一，被认为能有效缓解局部脑组织缺血缺氧的情况，减少脑组织的梗死面积。但救治时间窗、患者基础情况不同会导致最佳的治疗措施不同，故针对临床上某一具体的急性脑梗死患者，医生常会遇到，如该患者最佳治疗措施是什么，患者是否应该使用溶栓治疗，应该选用哪种溶栓药，急性脑梗死患者如何处理血压等一系列问题。

一女性患者，65 岁，右侧半身无力，说话不清，4 小时收入某院神经内科，有高血压病史 15 年。查体：神清，体温 36.5℃，脉搏 80 次/分钟，血压 165/95mmHg，发育正常，营养良好，头颅无畸形，心肺腹部无异常，双瞳孔等大等圆，光反射正常，眼底阴性，颅神经阴性，有鼻唇沟变浅。检测：急诊脑部 CT 扫描没有出血和其他异常密度影；血生化及常规检查正常；全脑血管造影检查发现右侧大脑半球血流量明显减少。初步诊断为急性脑梗死。

一、构建循证问题

针对上述这位伴有高血压的急性脑梗死患者提出临床治疗问题：能否使用溶栓药物治疗？是否应该降血压？其他疗法应该怎样选择？

二、检索证据

全面收集有关研究证据-查寻最新指南、系统综述以及临床研究证据。英文以 acute ischemia stroke、thrombolysis、hypertension 为关键词组合分别检索 Medline（PubMed）、Web of Science、Cochrane Library 等英文数据库；中文以中国学术期刊网（CNKI）全文数据库、重庆维普（VIP）中文科技期刊全文数据库、万方科技期刊全文数据库作为检索数据库，以急性脑梗死、急性脑缺血、溶栓及高血压作为关键词进行检索。制定检索策略时，注意检索词的变换使用，如 acute cerebral infarction、acute cerebral ischemia、thrombolysis therapy、blood pressure 等。检索结果：关于溶栓治疗急性脑梗死的国内外研究相对较多，其中检索到 3 篇最新相关指南、近百篇系统评价及 meta-分析，包括 9 篇在 Cochrane 图书馆注册的相关系统评价。其中，2008 年/2009 年发布的欧洲指南一篇，于 2015 年更新；2013 年发布的美国指南一篇，于 2015 年更新；及 2017 年发布的中国缺血性脑卒中急性期诊疗指导规范。

三、评价证据

严格评价研究证据。根据具体的临床问题，寻找答案。指南及系统评价是在综合相应证据的基础上，具有较高的证据级别，其次是高质量的随机对照试验。针对本例中提出的问题，检索到了多个指南，对其给出的推荐予以重点考虑。

1. 能否采用溶栓治疗　有关指南和证据结论如下：

1）指南均建议对缺血性脑卒中发病 3 小时内（Ⅰ级推荐，A 级证据）和 3～4.5 小时（Ⅰ级推荐，B 级证据）的患者，应根据适应证严格筛选患者，尽快静脉给予重组组织型纤溶酶原激活（rt-PA）溶栓治疗。

2）在发病初期，尚无其他溶栓药物可以代替 rt-PA 的研究证据。

3）其他溶栓药物只能在临床试验的情况下使用。

4）中国指南建议如没有条件使用 rt-PA 且发病在 6 小时内，符合适应证的患者可考虑静脉给予尿激酶。

2. 血压的处理问题

美国指南建议：①对急性缺血性脑卒中（脑梗死）患者的高血压处理应十分谨慎，首先应密切监测血压，处理焦虑、头痛、处理高颅内压；②经上述处理后血压仍持续升高>220/120mmHg 时可用降压药，但须慎重进行，缓慢降压；③紧急情况下应使用非肠道给药，避免使用强作用的长效降压药物；④对有溶栓指征的患者，血压应降至≤185/110mmHg。

欧洲指南建议：①脑梗死血压>200～220/120mmHg，脑出血患者血压持续>180/105mmHg 时可用降压药；②有溶栓指征的患者，收缩压应降至 185mmHg 以下，舒张压应降至 110mmHg 以下；③对有心衰、急性心肌梗死、急性肾衰使用溶栓或抗凝情况的患者，应立即降压。

国内指南建议：①准备溶栓者，应使收缩压<180mmHg、舒张压<100mmHg。②缺血性脑卒中后 24 小时内血压升高的患者应谨慎处理。应先处理紧张焦虑、疼痛、恶心呕吐及颅内压增高等情况。血压持续升高，收缩压>200mmHg 或舒张压≥110mmHg，或伴有严重心功能不全、主动脉夹层、高血压脑病，可予谨慎降压治疗，并严密观察血压变化，必要时可静脉使用短效药物（如拉贝洛尔、尼卡地平等），最好应用微量输液泵，避免血压降得过低。③卒中后若病情稳定，血压持续≥140/90mmHg，无禁忌证，可于起病数天后恢复使用发病前服用的降压药物或开始启动降压治疗。④脑卒中后低血压的患者应积极寻找和处理原因，必要时可采用扩容升压措施。

3. 关于其他疗法

欧美的指南建议如下：①证据充分应广泛使用：卒中单元，阿司匹林 100～300mg/d（Ⅰ类证据，A 级建议）。②证据欠充分应限制使用：有高度深静脉血栓形成或肺栓塞风险的患者，建议应当考虑给予低剂量皮下肝素或低分子量肝素（Ⅰ类证据，A 级建议）。③目前不推荐：无选择使用抗凝剂、血液稀释疗法、多种神经保护剂等。

中国指南建议如下：

（1）抗血小板：①对于不符合溶栓适应证且无禁忌证的缺血性脑卒中患者应在发病后尽早给予口服阿司匹林 150～300mg/d（Ⅰ级推荐，A 级证据）；②溶栓治疗者，阿司匹林等抗血小板药物应在溶栓 24 小时后开始使用（Ⅰ级推荐，B 级证据）；③对不能耐受阿司匹林者，可考虑选用氯吡格雷等抗血小板治疗（Ⅲ级推荐，C 级证据）。

（2）抗凝治疗：①对大多数急性缺血性脑卒中患者，不推荐无选择地早期进行抗凝治疗（Ⅰ级推荐，A 级证据）；②关于少数特殊患者的抗凝治疗，可在谨慎评估风险、效益比后慎重选择（Ⅳ级推荐，D 级证据）；③特殊情况下，溶栓后还需抗凝治疗的患者，应在 24 小时后使用抗凝剂（Ⅰ级推荐，B 级证据）。

（3）降纤治疗：对不适合溶栓并经过严格筛选的脑梗死患者，特别是高纤维蛋白血症者可选用降纤治疗（Ⅲ级推荐，B 级证据）。

（4）神经保护神经保护剂的疗效与安全性尚需开展更多高质临床试验进一步证实（Ⅰ级推荐，B 级证据）。

（5）高压氧和亚低温的疗效和安全性还需开展高质量的 RCT 证实。

（6）中医中药中成药和针刺治疗急性脑梗死的疗效尚需更多高质量 RCT 进一步证实。建议根据具体情况结合患者意愿决定是否选用针刺（Ⅱ级推荐，B 级证据）或中成药治疗（Ⅲ级推荐，C 级证据）。

所查到的指南和系统评价报告已对证据进行了质量评价，故可以直接进入到下一步。在没有相应指南及系统综述的情况下，应对纳入的各研究进行仔细评价，其中随机对照试验是一种前瞻性研究，在各种临床疗效研究中的论证强度最高，所以对大型随机对照试验也应给予一定的重视，明确各

研究纳入的患者特征、各组干预措施的选择，从组间基线可比性、随访时间、随机化分组、盲法、失访率等方面评价各研究。

四、证据应用与实践

经过研究结果的整理和分析，应用于具体患者的处理上。针对该病例，可以参考上述指南的推荐意见。故医生可以参照指南，根据当地的习惯和经验进行处理：①可以考虑静脉使用 rt-PA；②若医生及患者均不愿冒早期颅内出血增加的危险以获得远期残废率降低的效益，也可不用溶栓治疗而立即使用阿司匹林、对症、支持和防止并发症及复发等措施，有条件则进入卒中单元；③患者的血压不是太高，故暂时不需使用降压药。医生将这些意见告知患者及家属，考虑到价格昂贵和担心出血的风险，患者方面选择放弃溶栓治疗，而选择了证据确凿的其他疗法。

五、后效评价与小结

本节具体应用循证实践的 5 个步骤，对急性脑梗死患者开展循证治疗。第一步，提出需解决的实际问题。第二步，检索相关证据，重点在于设计恰当检索策略，通过国内外各大权威数据库，全面检索出可信度高的证据。第三步，评价证据，根据适当的标准，对研究质量进行严格评价，明确高质量研究。第四步，制定决策，根据当前最好的证据，结合患者的实际情况及意愿，给予患者最好的临床治疗。第五步，后效估计，即治疗决策应用后患者对治疗的反应。治疗指南及相关的系统综述无疑是疗效评价的最佳证据，但缺乏相应的证据时，临床医生就应对现有的随机对照试验等研究进行评价，再结合患者的实际情况及治疗措施的利弊进行综合考虑，作出决策。

（时景璞）

第十一章　疾病预后的研究、评价与循证实践

预后是临床医生每天都要面对的问题，当患者确诊某疾病后，非常关注的问题是如何治疗？治疗效果好吗？容易复发吗？平均可以存活多少年？最长能够活多少年？注意哪些方面可以延长寿命？这些都是涉及预后的问题。

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。

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，从而帮助临床医生作出治疗决策；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，还可利用疾病预后研究正确评定某项治疗措施的效果，从而促进治疗水平的提高。

二、疾病自然史

疾病自然史（natural history）是指在不给任何治疗或干预措施的情况下，疾病从发生、发展到结局的整个过程。疾病的自然史包括四个时期：

1. 生物学发病期（biologic onset）　指病原体或致病因素作用于人体引起有关脏器的生物学反应，造成复杂的病理生理学改变，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”，但若采用某些实验室检查或特异性高及敏感度高的诊断手段，可早期发现、早期诊断。

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，而被临床医生作出诊断，并及时救治。

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。一些疾病自然史较短，如急性感染性疾病，短期内出现症状体征和实验室检查异常，进展较快，较短时间内即可出现结局。而一些慢性非传染性疾病，自然史较长，甚至可迁延数十年之久，如心脑血管疾病、糖尿病、高血压等。

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。临床医生可采取医疗干预措施来改变其病程。

病程和疾病自然史不同，前者可因接受医疗干预（包括各种治疗措施）而发生变化，预后随之发生改变。在病程早期就采取积极医疗干预措施，往往可以极大改善预后，而在病程晚期再进行医疗干

预，则效果不佳，疾病预后较差，因此，临床医生应要重视对临床病程的估计。如早期发现和诊断肝癌，积极治疗，肝癌患者的生存时间可大大延长，反之，晚期肝癌患者的预后就很差。

四、预后因素

凡影响疾病预后的因素都可称为预后因素（prognostic factors），若患者具有这些因素，可影响其病程发展过程、出现某种结局事件的概率就可能发生改变。预后因素研究有助于临床医生进行医学干预，包括筛检、及时诊断、积极治疗和改变患者的不良行为等，从而改善患者预后。

预后因素与危险因素不同，危险因素是指作用于健康人，能增加患病危险性的因素，而预后因素则是在已经患病的患者中与疾病结局有关的因素。虽然某些疾病的一些危险因素也可能同样是预后因素，但多数并不重叠。例如，急性心肌梗死的危险因素与预后因素中，有些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，预后也差。但有些因素则不尽然，如性别，男性发生急性心肌梗死危险性高于女性，但女性发生心肌梗死后的预后反而比男性差。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 11-1）。

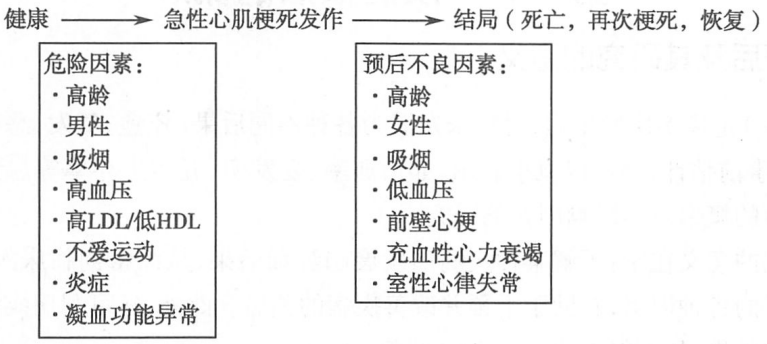


图 11-1　急性心肌梗死的危险因素和预后因素之间的差别

影响疾病预后的因素复杂多样，概括起来有以下几个方面：

（一）早诊早治

早期正确诊断并及时合理治疗对任何疾病而言都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，通过手术和放化疗，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。

（二）疾病自身特点

疾病自身特点包括疾病的性质、病程、病理类型与病变程度等，常是影响疾病预后的重要因素。如像上呼吸道病毒感染这些自限性疾病，不需要治疗也可自愈，预后良好。

（三）患者病情严重程度

通常病情与预后密切相关，病情危重者，预后较差。例如，黄疸腹水型重症传染性肝炎的预后远逊于无腹水的轻或中型肝炎。

（四）患者身体素质

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。

（五）医疗条件

医疗条件的优劣，也直接影响疾病预后。例如：条件好的医院不仅医疗设施好，患者早期的正确诊断概率高且有抢救经验丰富的专科医生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、家庭因素和其他因素

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、患者文化素养、依

从性及心理因素都会影响患者疾病的预后。

第二节　疾病预后研究中常用结局指标

一、率相关指标

（一）病死率（case-fatality rate）

患某病的总人数中，死于该病的患者所占的比例。常用于病程短且容易死亡的疾病，如各种传染病、急性中毒、心脑血管疾病的急性期和迅速致死的癌症。

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

（二）疾病别死亡率（disease-specific mortality）

某一人群在一定的时期内（通常指一年）因某病死亡的人数所占的比例，一般以1/10万或1/万为单位。死亡率与病死率是两个不同的概念，不能混淆。例如某种类型白血病的死亡率是2/10万，而病死率为30%。

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

（三）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人数的比例。

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

（四）缓解率（remission rate）

接受某种治疗后，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。例如，某种类型白血病经过正规化疗后，完全缓解率是60%，部分缓解率是10%，无自发缓解。

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

（五）复发率（recurrence rate）

疾病经过治疗缓解或痊愈后又再次发作的患者人数占总观察患者数的百分比。

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

（六）致残率（disability rate）

发生肢体或器官功能丧失者占观察患者总数的百分比。

$$致残率(\%)=\frac{致残的患者例数}{接受观察的患者总数}\times 100\%$$

（七）总生存率（overall survival rate，OS）

从疾病临床过程的某一点开始，经过一段时间后仍存活的病例数占总观察例数的百分比。生存率常用于长病程致死性疾病，如各种癌症，病程较短的癌症可用1年生存率，病程较长的癌症用5年生存率表示预后。

预后研究仅仅报道生存率是不够的，例如图 11-2 中 4 种情况的 5 年生存率均为 10%，但不能反映 5 年间生存率变化的全貌，所以还需要生存曲线等加以补充。

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

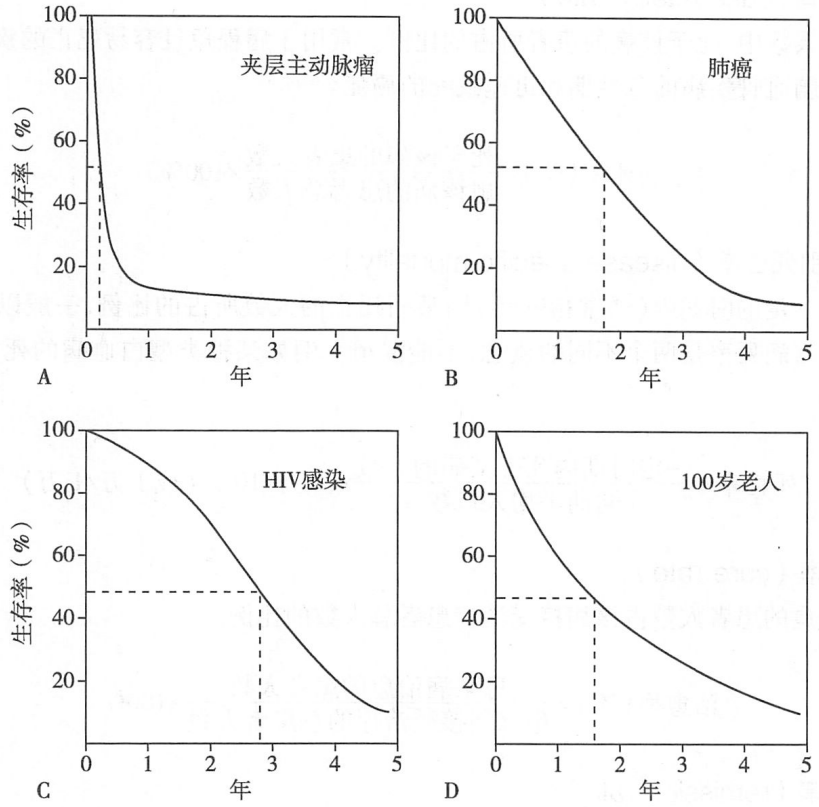


图 11-2　四种不同人群的生存曲线

注：四种人群的 5 年生存率均为 10%，但生存曲线明显不同，说明仅报道 5 年生存率有很大的局限性，虚线为中位生存时间

（八）无进展生存率（progression-free survival rate，PFS）

指疾病经过治疗达到病情稳定后，未出现临床疾病进展或死亡的患者占接受治疗患者总人数的比例。例如，比较某分子靶向新药和常规化疗作为一线治疗某种类型白血病的长期疗效，新诊断的白血病患者 200 例自然分为两组，一组为靶向药物治疗组（100 例），一组为常规化疗组（100 例），随访 5 年，观察 PFS 和 OS 等预后指标。PFS 定义为入组之日为观察起始点，疾病进展或各种原因的死亡为观察终点，靶向治疗组和化疗组 5 年 PFS 分别为 85% 和 40%。OS 定义为随机化之日为观察起始点，各种原因的死亡为观察终点，靶向治疗组和化疗组 5 年 OS 分别为 95% 和 80%（图 11-3）。

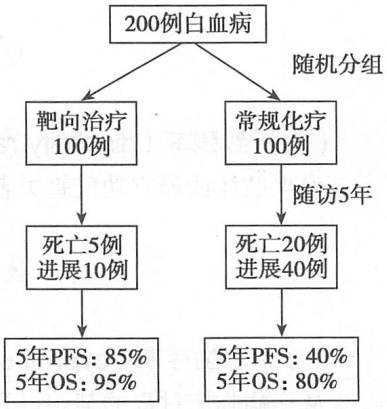


图 11-3　无进展生存率（PFS）和总体生存率（OS）

（九）无病生存率（disease-free survival rate，DFS）

常用于癌症的结局判断，指疾病经过治疗达到临床缓解后，没有临床疾病复发或死亡的患者占接受治疗患者总人数的比例。

二、中位时间

（一）中位生存时间（median survival time）

又称为半数生存期，即当累积生存率为 50% 时所对应的生存

时间，表示只有50%的患者可以活过这个时间。如AML的中位生存期为20个月，说明诊断为AML后只有50%的患者可以活过20个月。从生存曲线可以得到中位生存时间，图11-2显示夹层主动脉瘤中位生存时间不足半年，肺癌不足2年，HIV感染约为3年。

（二）中位无病生存时间（median disease-free survival time）

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

（三）中位无进展生存时间（median progression-free survival time）

指从规定的随访起始点开始，直至疾病进展或死亡的中位时间。在疗效评价中，常用分组之日为起始点。例如，靶向药物和化疗治疗某类型白血病，中位PFS时间分别为9年和5年。

三、健康相关生存质量及其衍生指标

生存质量（quality of life，QOL）也是常用的预后研究指标，一般用量表进行评定。主要包括生理功能、心理功能、社会功能和对健康状况的总体感受等方面。例如在肿瘤研究中，经常采用东部肿瘤协作组（Eastern Collaborative Oncology Group，ECOG）人体机能状态评分量表评估患者的生存质量（表11-1）。

表11-1　ECOG人体机能状态评分量表

人体机能状态	定　义
0	无症状
1	有轻微症状，但不需要卧床，生活可以自理
2	有症状，每天需要卧床的时间不到50%
3	有症状，每天需要卧床的时间超过50%
4	全天卧床
5	死亡

第三节　疾病预后研究设计方案

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、临床试验等，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。

二、方案设计的注意事项

（一）队列研究的起始点

预后研究多采用队列研究设计，其起始点称零点时间（zero time），该起始点在研究设计时必须要明确规定，是从病程的哪一时间点起开始观察，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。假如零点时间不一致，有些按出现症状之日，有些按照确诊之日，有些则按治疗开始之日算起，零点时间杂乱，难以评价真正的预后，甚至会得出错误的结论。

对于预后研究，要尽量选择疾病的早期，收集队列的集合时间接近疾病初发时点，则称为起始队列（inception cohort）。例如AML预后研究，采用确诊之日为研究起点，属于起始队列，结论更为可靠。

（二）研究对象的来源和分组

研究对象要具有代表性，能代表目标疾病的人群。来自不同级别医院，即便为同一种疾病其预后

研究结果也可能不同，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，这是因为三级医院常集中了病情较重、病程接近后期的患者，预后差。倘若采用来自某地区各种级别医院中该疾病的病例作为样本，通常包括了各种型别以及病情严重程度各异的病例，能反映出目标人群的特点，因而代表性就比较好，能反映疾病的真实预后。

研究对象的分组也应遵循可比性原则，即非研究因素在两组分布应相同，才有可比性。在评价治疗方案对预后的影响时，可采取控制、配比、分层等措施，尽量做到两组除了需要研究的治疗方案不同外，其他各种影响预后的因素能够相同。若部分因素组间确实不可比，则应在后续的统计分析阶段加以控制。

（三）随访和失访

预后研究中的随访十分重要，应组织严密，尽量使所有研究对象都随访到，做到失访率越低越好。如失访率小于5%一般可以接受，对结果的影响小，大于10%应引起注意，若大于20%则研究结果可能没有参考价值。因为失访患者的疾病预后信息丢失，从而影响预后结果的可靠程度。

防止失访应注意以下数点：①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；②建立健全随访管理制度，随访要有专人负责，一旦发现失访要及时补救；③改进随访方式，采用多种随访方式，如门诊、电话、邮件、微信等，方便患者。

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，以及观察到一些少见的不良反应。

随访间隔时间的确定要合理，以便能观察到各种变化的动态过程。随访间期不同，得出的结论可能也不同，例如同样一组肺癌患者，如果观察间期分别定为2个月和6个月，2个月的患者可能更容易被观察到肿瘤的进展，DFS时间可能短于6个月的患者。

随访过程中确定的各种结局，事先要有明确的定义和判断标准，设计时就规定好，执行中不再变动，标准要有客观性，为防止测量偏倚，最好用盲法观察。

对有失访病例的研究，可采用敏感性分析估计失访对研究结果的影响，具体方法可先假定失访者均出现预定结局，得到结局事件“最高”发生率，然后假定失访者均不出现结局，得到结局事件“最低”发生率，比较“最高”和“最低”率，若两者相差不大，则结果可取，倘若两者相差很大，则结果不可靠。

三、疾病预后研究的基本步骤

以上海市中美白血病协作组完成的623例AML预后研究为例，说明预后研究的基本步骤。

（一）确立拟研究疾病及研究目的

原发性、初治AML的预后。

（二）确定观察指标和研究因素

结局观察指标为完全缓解率，3年复发率，3年生存率，中位生存时间。分析影响生存的预后因素。

（三）确定研究方案

本研究采用前瞻性队列研究的设计方案进行预后研究。确定AML的诊断标准，定义OS为从诊断之日到死亡或最后的随访之日，确定随访终止时间为2008年5月1日。

（四）收集队列和登记资料

前瞻性收集2003—2007年的上海市白血病协作组24家医院的原发性、未治疗的AML患者623例，所有骨髓标本统一送到中美实验室进行统一诊断，以防误诊。登记病例信息和所有相关的实验室资料，也包括随访时需要的患者身份证号码、联系电话、地址、主管医生的联系信息等。

（五）随访

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

（六）统计分析

1. 病例一般资料分析　患者年龄、性别、WHO 亚型、染色体等资料的分析。

2. 完全缓解率、复发率的计算。

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

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

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

第四节　疾病预后研究的统计方法

一、生存率和中位生存时间分析

（一）生存率的计算

在生存分析中，生存率的计算有两种方法：直接法和间接法。

1. 直接法　如果病例数多，没有失访，则结果可靠，计算简单。例如 5 年生存率 = 活满 5 年的例数/总随访例数。但一般生存数据均存在删失值（censored data），也称为终检值、截尾值，需要用间接法计算生存率。

2. 间接法　截尾/删失值包括三种情况：①失访；②死于其他疾病；③观察到规定的随访截止时间仍存活。Kaplan-Meier 法和寿命表法为常用的间接法。寿命表法用于例数较多的研究，Kaplan-Meier 法可以用于小样本的研究，也可用于大样本的研究。目前有多种统计软件可以进行生存分析，不需手工计算。

生存曲线是以随访时间为横坐标，累积生存率为纵坐标作图，常用的是 Kaplan-Meier 生存曲线，表示时间与生存关系的函数曲线。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。例如四种情况的 5 年生存率虽相等，但其生存曲线却大相径庭（见图 11-2）。

（二）中位生存时间的计算

预后研究中还需要计算生存期的长短，即存活期。常用中位生存期、无病生存期（disease free survival，DFS）、无进展生存期（progression free survival，PFS）等来表示。

生存时间作为一种连续变量，常呈非正态分布，因此要计算中位生存期，不能计算平均生存期，鉴于有截尾/删失值的存在，也不能用普通方法计算中位数，需用生存分析计算中位生存时间。生存分析允许有截尾数据的存在。例如，有 12 例患者，8 例死亡，2 例失访，2 例截止到随访日仍存活，生存时间分别记录为 10、20、20、30、80、50、100、150、80、10、120、200 个月，按生存分析方法计算的中位生存时间为 80 个月，而按一般统计方法计算的中位数为 65 个月。

二、疾病预后因素的研究方法

（一）单因素预后分析

预后因素的分析一般先从单因素分析开始，进而多因素分析。常用的单因素分析方法是 Log-rank 检验，可以 Kaplan-Meier 生存曲线的形式直观展示结果。例如上述 12 例患者的分析结果如表 11-2 和图 11-4 所示。单因素分析中有统计学差异的因素或 $P<0.1$ 的因素可进一步多因素分析。

表 11-2　采用 Log-rank 分析进行单因素预后分析

		例数	死亡例数	中位生存时间（月）	χ^2 值	P 值
性别	男	7	5	30	6.55	0.01
	女	5	2	150		

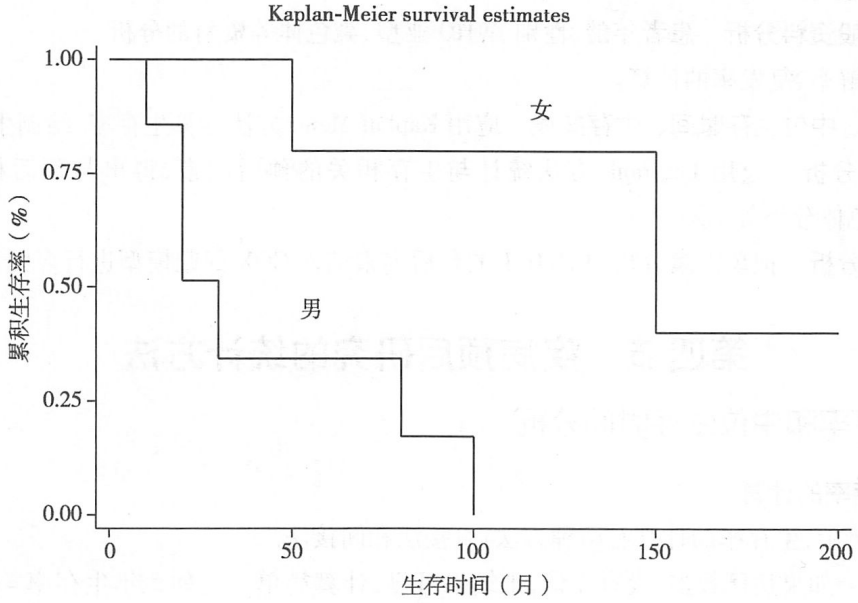


图 11-4　不同性别患者的 Kaplan-Meier 生存曲线

（二）多因素预后分析

由于疾病的结局和多种预后因素有关，而各种预后因素或独立或交互影响结局，为全面正确地衡量这些预后因素的作用，近年来随着统计学方法和计算机技术的不同更新，像多元线性回归、逐步回归、Logistic 回归、Cox 风险比例模型等多因素分析方法，开始广泛用于主要预后因素的筛选，进而建立疾病预后指数或预后模型。

第五节　预后研究中常见偏倚及其处理方法

一、预后研究中常见的偏倚

预后研究中存在多种偏倚，若不加控制，均可导致结论出现偏差（图 11-5）。

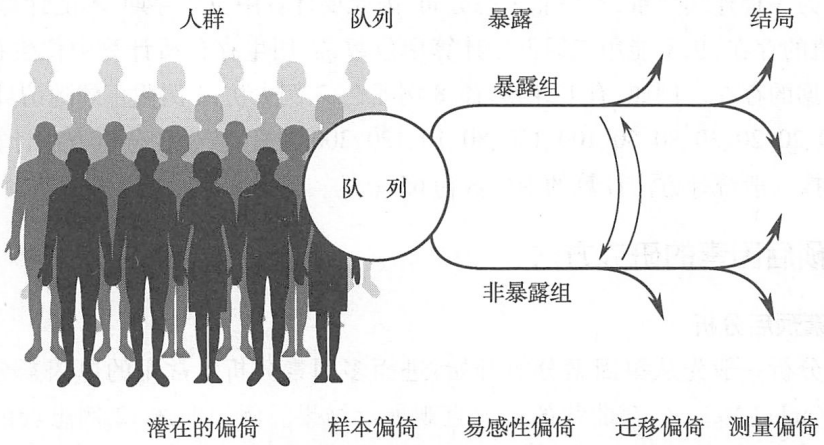


图 11-5　预后队列研究中常见的偏倚

（一）选择偏倚（selection bias）

选择偏倚包括样本偏倚（sampling bias）、集合偏倚（assembly bias）、存活队列偏倚（survival cohorts bias）、失访偏倚（lost to follow-up）、零时偏倚（zero time bias）、迁移偏倚（migration bias）等多种。

1. **样本偏倚**　研究队列一般是从总体人群中选择一部分进行研究，如果研究样本与总体的临床

特征有很大的差异，则说明样本的代表性差，存在样本偏倚。

2. **集合偏倚**　又称分组偏倚、就诊偏倚。例如，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，乃至就诊患者的经济收入在不同地区均可能千差万别，倘若一并集合成队列进行随访观察，即使发现预后不同有可能是上述因素造成而非研究因素所致。

3. **存活队列偏倚**　从各医院收集病例组成的队列，可能不一定都是起始队列，而是集中在该病病程中某一时点且都是存活的病例，故称存活队列偏倚，那些未入院失访病例的信息丢失，可导致结果高估。例如由心肌梗死住院患者组成队列，那些未入院者可能病情更为危重，在家里或转运途中就已死亡，入院患者队列均是存活患者，病情较轻，预后可被高估。

4. **失访偏倚**　在临床预后研究中经常会遇到。由于观察时间长，观察对象因迁移、外出、不愿继续合作、因药物不良反应或效果差而停止治疗或死于非终点疾病等原因脱离了观察，造成失访，使研究结果失真，故称失访偏倚。

5. **零时偏倚**　由于不是起始队列，处于病程不同阶段的观察对象进入研究队列，观察的起始点不尽相同而导致的偏倚，有时又称零时不当偏倚。例如，进入队列既有初发肿瘤患者也有复发肿瘤患者，初发患者从诊断之日开始观察，复发患者则从复发之日开始研究，两类患者的观察起点不一致，结果会出现偏差。

6. **迁移偏倚**　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所引起的偏倚，倘若变动的人数过多，必然会影响结论的真实性。

（二）测量偏倚（measurement bias）

观察与评定结局过程发生偏倚，有些如死亡、脑血管意外、某些肿瘤，诊断十分明确，不容易遗漏或误诊，但是有些结局，如特殊死因、亚临床疾病、不良反应、残疾等就不那么清楚，判断有出入，从而影响研究的结论。事先制定严格的、可行的结局判断标准，并采用盲法评定结局，可以减少测量性偏倚。

（三）混杂偏倚（confounding bias）

在研究某因素是否预后因素时，理论上要求观察组（暴露于某预后因素）和对照组（非暴露于该预后因素）两组的临床特点和其他非研究因素都要相同，但这在实际工作中常难以做到。这些与研究因素和疾病结局均有关的因素就是混杂因素，可以用下列方法处理。

二、偏倚的处理方法

（一）随机化（randomization）

两个队列进行比较，从理论上讲，除研究的预后因素外，其他因素最好组间均衡可比，即基线状况要相同，这样才能客观评价该预后因素的作用大小。

（二）限制（restriction）

研究对象入选时，可限制在具有一定特征的患者人群，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，可将研究对象限制在黄种人、男性、无并发症的前壁心肌梗死患者中进行观察，这样就可排除种族、性别、心肌梗死部位和并发症等因素的干扰和影响，客观反映年龄对急性心肌梗死预后的影响。但用该法来控制偏倚所获得的预后结论常有很大局限性，结论外推性差。

（三）配对（matching）

配对就是为观察组的每一个研究对象匹配一个或多个具有同样因素的对照，然后比较两组的预后因素。配对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，这些最常见的混杂因素作为配对条件，当然其他因素也可作为配对条件，如病期、疾病严重程度和治疗史等。但千万不能把研究因素当作配对条件，否则就失去了研究的意义。

（四）分层（stratification）

分层也是常用的偏倚控制方法之一，特别是有潜在的混杂偏倚时，应用分层的方法分别在设计阶

段(分层)和分析阶段(分层分析)来控制偏倚。分层分析是指将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。

例如A、B两个医院冠心病搭桥术后病死率比较(表11-3),总体粗病死率A医院为4%,高于B医院的2.6%,认为A医院手术死亡率高。但若以术前危险度分层,重新计算各层内的病死率,A和B医院结果相同,为什么会出现这样的结果呢?主要是因为两所医院患者的病情危重程度不同,A医院高危病例较多,而B医院低危患者较多,经过分层分析消除了疾病危险度这个混杂因素。

表11-3 2个医院冠心病搭桥术后病死率(以术前危险度分层)

危险度	A医院				B医院			
	患者数	死亡数	粗病死率(%)	标化病死率(%)	患者数	死亡数	粗病死率(%)	标化病死率(%)
高	500	30	6		400	24	6	
中	400	16	4		800	32	4	
低	300	2	0.67		1200	8	0.67	
合计	1200	48	4	3.6	2400	64	2.6	3.6

(五) 标准化(standardization)

当比较两个率时,如果两组对象内部构成存在差别且足以影响结论时,可用标准化率加以校正。表11-3医院A和B的标化病死率=(1/3×6%)+(1/3×4%)+(1/3×0.67%)=3.6%,每个危险度的权重均为1/3。

(六) 多因素分析方法

预后因素研究常比较复杂,可有多个预后因素相互作用,从而影响结局,此时应借助于多因素分析方法,既可同时处理多个影响因素,又可从中筛选出与疾病结局有关的主要预后因素,以及这些因素对预后的影响力。在预后因素研究中,以Cox比例风险模型最为常用。

第六节 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。文献评价包括三个方面:真实性、重要性和适(实)用性。评价的原则和标准可归纳为九条(表11-4),现分述如下。

表11-4 预后研究文献的评价原则

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

笔记

一、真实性评价

（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，如研究初发脑卒中的预后因素，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。

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）。如研究AML的生存时间，应该有统一的起点，最好是从诊断之日开始计算，而不能有的是诊断第一天为起点时间，有的是化疗第一天为起点时间。

（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，以便判断有无选择偏倚，对预后研究所在地区、医疗机构也应叙述，以便了解研究对象的代表性，判断是否存在选择性偏倚。

例如研究急性心肌梗死的预后，若采用入院病例作为研究对象，可能就把一些病情严重在急诊就死亡的病例排除在外，病例的代表性受限，会高估了预后。

（三）随访时间是否足够？随访是否完整

随访时间必须足够长，以便发现关注的研究结局。倘若随访时间很短，只有一小部分患者出现预期结局，如肿瘤发生、康复、复发或是不良事件的发生，这样就不能反映该疾病预后的真实情况。

同时要注意随访的完整性，在理想情况下，所有纳入的研究对象应当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，允许存在一定的失访率。判断失访对结论是否有影响，一般遵从“5和20”原则，失访率<5%，其研究结果偏倚少、结果可靠，若失访率>20%则严重影响结果真实性，5%和20%之间结果较可靠。亦可通过前述的敏感性分析来估计对结论的影响，比较“最高”和“最低”发生率，若两者相差不大，则结果可信，否则研究结果不可信。

（四）判断结局有无客观标准，是否采用了盲法

观察疾病预后的结局应有客观的标准。在研究设计阶段，就应明确定义结局指标，同时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；倘若结局属“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发生结局事件的蛛丝马迹，而对待对照组则不然；以及预期偏倚（expectation bias），即研究者先入为主、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加以校正。最简单的校正方法是分层分析，如各亚组有不同预后结果，说明存在混杂因素；而各亚组预后结果相同，说明研究因素是独立的预后因素。较为复杂的校正方法是Logistic回归、Cox模型等多因素分析方法，适用于有多个混杂因素的校正。

二、重要性评价

（一）报告预后研究的结果是否完整

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。①某一时间点的生存率，如1年、3年、5年生存率等；②中位生存时间，即观察到50%的研究对象死亡的时间；③生存曲线可了解预后随时间而变化情况。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。

（二）研究结果的精确性如何

除了报道生存率、生存时间、生存曲线，还应当报告预后估计的精确度，即预后结局概率的95%可信区间（95% CI）。95% CI较窄，说明样本量足够大，结果精确性高，结论外推性好，对总体预后的估计更精确。例如Cox回归分析得出的HR=1.3，但是95% CI不同时其含义也不同，比如HR为1.3（95% CI：0.8～1.6），说明差异没有统计学意义，该因素可能不是独立的预后因素。再如HR为1.3（95% CI：1.1～1.6），说明差异有统计学意义，该因素可能是独立的预后因素，暴露于该因素时发生不良事件（如死亡）的风险是不暴露时的1.3倍。

三、适用性评价

（一）我们自己的患者是否与文献报道的患者非常不同

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

（二）研究结果是否有助于治疗方案的制定和是否有助于对患者及其亲属作出解释

1. 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患者中应用华法林抗凝治疗，能降低缺血性脑卒中的发生率，但在一项“孤立性心房颤动”患者（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出血风险很可能超过得益，因此这类患者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。

2. 研究结果是否有助于对患者及其亲属作出解释　例如一项可信、精确度高的研究结果显示疾病具有良好的预后，则十分有助于向焦虑的患者及其家属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，也可以与患者及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。

第七节　循证实践及案例分析

一男性患者，40岁，确诊为急性髓系白血病（AML）M4亚型1个月，染色体正常核型，已经给予标准DA方案（柔红霉素+阿糖胞苷）诱导缓解（CR），缓解后准备给予缓解后的巩固强化治疗。缓解后的巩固强化治疗非常重要，正确治疗可减少复发，延长生存期。缓解后治疗可以采用多种标准剂量化疗药物组合的常规方案，也可用大剂量阿糖胞苷（Ara-C）单药治疗，到底何种治疗方案更好，复发率低，生存期长呢？即哪种治疗方法可以提高AML患者的预后？

一、提出问题

采用PICO方式提出临床问题：对于AML患者而言（P），缓解后应用标准化疗方案（I）和应用大剂量Ara-C化疗方案进行巩固强化治疗（C），复发率、生存率、生存时间何种治疗更佳（O）？不良反应是否可以耐受？

二、检索与评价证据

寻找证据的顺序按照先系统评价/meta-分析，后RCT研究，然后再观察性研究，从高级别的证据开始检索。首先以（acute myeloid leukemia）AND（high dose cytarabine）AND（overall survival）为检索式，限定在“Title/abstract”和“Meta analysis”，未检索到相关的系统评价/meta-分析。重新检索，把检索条件限定在“Title/abstract”和“Randomized Controlled Trial”和近5年的文献，检索到20篇相关文献。阅读题目和摘要，选择最新的、样本量较大、多中心研究、随访时间较长、发表在影响因子较高的文杂志，阅读全文，按照预后评价原则进行文献评价。

（一）文献基本内容

1. 文献题目　A randomized comparison of 4 courses of standard-dose multiagent chemotherapy versus 3 courses of high-dose cytarabine alone in post remission therapy for acute myeloid leukemia in adults: the JALSGAML 201 Study.

2. 文献来源　Blood,2011,117: 2366-2372.

3. 研究目的　在年龄<65 岁的 AML 患者获得第一次完全缓解(CR)后,比较 2 种治疗方案对患者预后的影响。一种方案为 4 个疗程的标准剂量、多种化疗药物组合的传统方案(CT),另一种方案为 3 个疗程的大剂量 Ara-C (HiDAC)。

4. 研究设计　为队列嵌套随机对照临床试验,研究流程图见图 11-6。2001 年 12 月—2005 年 12 月期间共有 1057 例新诊断的 AML 符合纳入标准,年龄 15～64 岁,来自 129 个研究中心进入研究队列。第一次随机分组,分别用 IDA+Ara-C 和 DNR+Ara-C 诱导缓解方案进行诱导治疗,共有 823 例患者取得 CR,CR 后再次随机分组,分别用 4 个疗程的大剂量 Ara-C 和 3 个疗程的标准传统方案,42 例患者因为各种原因未参加第二次随访分组。中位随访时间为 48 个月(5～78 个月)。第一次随机化后的观察指标是 CR,第二次随机化后的主要观察指标是 DFS,次要观察指标是 OS 和 3 级以上的不良反应。DFS 定义为从 CR 的第一天开始到白血病复发或各种原因死亡,OS 定义为从白血病诊断的第一天到各种原因死亡。Kaplan-Meier 法估计 DFS 和 OS,单因素预后分析采用 Log-rank 方法,多因素预后分析采用 COX 回归模型。

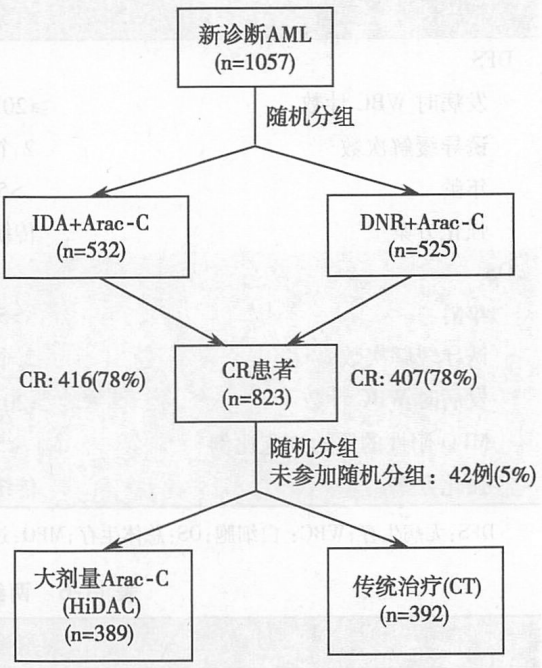


图 11-6　研究流程图

5. 研究结果　HiDAC 组和 CT 组的 5 年 DFS 分别为 43% 和 39% ($P=0.724$),5 年 OS 分别为 58% 和 56% ($P=0.954$)(图 11-7)。在染色体预后良好组,HiDAC 组和 CT 组的 5 年 DFS 分别为 57% 和 39% ($P=0.050$),5 年 OS 分别为 75% 和 66% ($P=0.174$)。在染色体预后中等和不良组中,两组的 DFS 和 OS 无统计学差异。

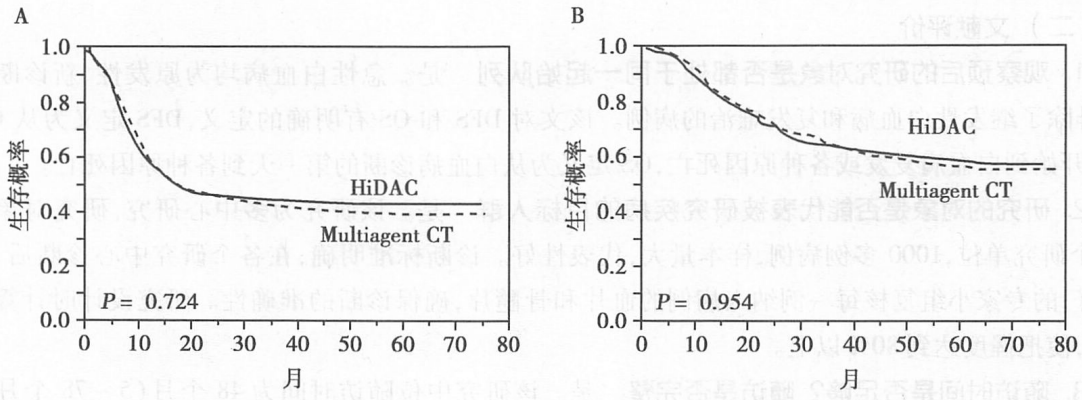


图 11-7　两组患者 DFS 和 OS 的生存曲线

A:大剂量 Ara-C (HiDAC)组的 DFS 是 43%,传统化疗 (CT)组为 39% (P=0.724);B: HiDAC 组和 CT 组的 5 年 OS 分别是 58% 和 56% (P=0.954)

应用 COX 回归模型进行多因素预后分析，对于 DFS 而言，发病时白细胞≥20×10^9/L、诱导疗程数 2 次、年龄>50 岁为独立的预后因素，校正了这些因素后缓解后的强化方案（应用传统化疗）并非独立的预后因素。对于 OS 而言，年龄>50 岁、诱导疗程数 2 次、发病白细胞≥20×10^9/L、原始细胞过氧化酶（MPO）阳性率<50% 为独立的预后因素，而校正了这些因素后缓解后的强化方案不是独立的预后因素（表 11-5）。但在 HiDAC 组，白细胞减少更明显，持续时间更长，感染率更高（表 11-6）。

表 11-5　多因素预后分析-COX 回归

变量	分类	HR	*P* 值
DFS			
发病时 WBC 计数	≥20×10^9/L	1.49	<0.0001
诱导缓解次数	2 个疗程	1.50	0.0006
年龄	>50 岁	1.33	0.0028
强化方案	传统化疗	1.04	0.7128
OS			
年龄	>50 岁	2.00	<0.0001
诱导缓解次数	2 个疗程	1.58	0.0033
发病时 WBC 计数	≥20×10^9/L	1.41	0.0070
MPO 阳性的原始细胞比例	<50%	1.42	0.0149
强化方案	传统化疗	0.96	0.7768

DFS：无病生存；WBC：白细胞；OS：总体生存；MPO：过氧化酶

表 11-6　两组不良反应发生情况比较

	大剂量 Ara-C 组	传统化疗组	*P* 值
第一次化疗后最低白细胞（$\times10^9$/L）	0.17	0.40	<0.0001
第一次化疗后白细胞<1.0×10^9/L 持续时间（天）	13（0～40）	12（0～36）	0.0005
感染发生率（%）	20.9	14.5	<0.001
出血发生率（%）	0.8	0.7	0.601
30 天死亡率（%）	0.9	0.6	0.389

6. **结论**　完全缓解后，传统化疗与大剂量 Ara-C 治疗的疗效相当。在染色体预后良好的患者中 HiDAC 的 DFS 优于传统化疗。

（二）文献评价

1. **观察预后的研究对象是否都处于同一起始队列**　是。急性白血病均为原发性、新诊断的病例，排除了继发性白血病和复发难治的病例。该文对 DFS 和 OS 有明确的定义，DFS 定义为从 CR 第一天开始到白血病复发或各种原因死亡，OS 定义为从白血病诊断的第一天到各种原因死亡。

2. **研究的对象是否能代表被研究疾病的目标人群**　是。该研究为多中心研究，研究对象来自 129 个研究单位，1000 多例病例，样本量大，代表性好。诊断标准明确，在各个研究中心诊断后，有一个专门的专家小组复核每一例纳入病例的血片和骨髓片，确保诊断的准确性。研究设计时计算了样本量，使把握度达到 80% 以上。

3. **随访时间是否足够？随访是否完整**　是。该研究中位随访时间为 48 个月（5～78 个月），从图 11-7 可以估计中位 DFS 约为 20 月，但尚不能得出中位 OS 时间，对于主要观察指标 DFS 而言随访时间足够，但对于次要观察指标 OS 而言，随访时间尚不够。该文没有直接报道失访率，根据文章描述，所有患者均有完整的随访结果。

4. **判断结局有无客观标准，是否采用了盲法**　是。观察结局为 CR、DFS、OS 和不良反应，均有

笔记

明确的定义。因为观察指标为客观的指标，不容易产生测量性偏倚，允许不采用盲法。

5. 是否对影响预后研究的重要因素进行了统计学的校正　是。应用了 COX 回归模型进行了多因素的分析，对影响预后的主要因素如年龄、白细胞计数、诱导缓解次数等进行了校正。

6. 预后研究的报告结果是否完整　是。报道了 CR 率、复发率、DFS、OS，采用 Kaplan-Meier 生存曲线形式展示了结果，也报道了 COX 回归分析后的 HR 值和 P 值。没有报道具体的中位生存时间，但读者可以根据生存曲线进行估计。

7. 研究结果的精确性如何　否。没有报道 95% CI，特别是表 11-5 中 HR 应同时写出 HR 的 95% CI。

8. 我们自己的患者是否与文献报道的患者相似或非常不同　该研究为多中心、大样本量研究，亚洲日本患者，与我们中国患者应该相似，不会有很大的人种差异。

9. 研究结果是否有助于治疗方案的制定和是否有助于对患者及其亲属作出解释　是。该结果对临床医生提供了治疗决策，有助于对患者进行解释。

总体而言，该研究结论可信，有临床应用价值，可以指导患者的治疗。

三、应用证据与临床实践决策

我们的 AML 患者染色体正常，属于染色体预后中等组，应用 HiDAC 和传统化疗可以取得相类似的长期预后，所以两种治疗方案均可以选择。与患者及其家属充分沟通后，患者采用 HiDAC 进行缓解后的强化治疗。但应用 1 个疗程后白细胞达到粒细胞缺乏、严重感染，以后换用了传统的化疗，直至缓解后强化治疗结束，患者持续缓解之中。

四、小结

该研究采用 RCT 的研究方法进行随机分组，保证了两组的可比性，再结合前瞻性队列的方法进行长期随访，随访率高，得出比较可靠的结论，指导临床决策。患者应用 HiDAC 后出现严重感染，因为有循征医学证据支持传统的常规化疗也可以取得相似的疗效，所以我们敢于换用常规化疗，事实证明患者取得了很好的效果。

（王小钦）

第十二章　临床实践指南的制定与评价

20世纪90年代初期，美国医学研究所正式提出了临床实践指南（clinical practice guideline，CPG）的概念，经过20多年的发展，高质量的CPG为临床医生从事预防、诊断、治疗、康复、保健和管理工作带来许多益处，已成为国际上用来规范医疗行为、提高服务质量、合理控制医疗费用等行之有效的方法。

第一节　概　　述

一、临床实践指南的概念

由于不同地区以及不同医生之间的医疗水平存在差异，有必要制定指导、规范医疗实践的文件。20世纪80年代开始制定诊疗常规，委托某一领域比较著名的专家撰写，由卫生管理机构统一发布。90年代后则开始由多人组成一个专家小组、经共同讨论后撰写，形成所谓的非正式专家共识。既往这些均被称作临床实践指南，但目前对实践指南已经有了明确和公认的定义，严格意义上讲，上述诊疗常规、非正式专家共识等不能算作临床实践指南。

从美国医学研究所（IOM）对临床实践指南定义的演变，可以发现临床实践指南的内涵已发生显著的变化。1990年IOM对临床指南的定义是：系统开发的多组指导意见，帮助医生和患者针对具体临床问题作出恰当处理和决策，从而选择合适的卫生保健服务。2011年IOM的定义是：通过系统综述生成的证据以及对各种备选方案进行利弊评价和权衡之后提出的最优推荐意见。该定义强调了循证医学的重要性，要求在寻找相应的证据和通过系统的文献评价、权衡利弊后提出推荐意见。

从上述国际公认的临床实践指南定义和著名指南文库的收录标准可以看出，现在的临床实践指南是基于循证医学方法制定的循证临床实践指南。

二、临床实践指南的作用和价值

临床实践指南的产生和发展主要与下列三个因素有关，其作用和价值也体现在促进和改善这几个方面上：

（一）临床实践的巨大差异

自20世纪80年代以来，人们开始注意到在临床诊治上的差异现象。对于相同一个临床问题，在不同国家或同一国家的不同地区甚至在同一个地区内的不同医疗机构之间其处理方法千差万别，差异极大。例如，曾有研究比较20世纪90年代中国与英国治疗急性缺血性脑卒中，发现19%～69%中国医生常规使用7种治疗方法（甘露醇、活血中药、阿司匹林、钙拮抗剂、低分子右旋糖苷、蛇毒、糖皮质激素），而英国医生除了使用阿司匹林外，基本不用其余6种疗法。

（二）医疗措施的不当使用

有研究表明在日常临床诊疗实践中，大约1/4至1/3的医疗措施使用不当，存在着滥用（overuse）、误用（misuse）或使用不足（underuse）等问题。例如，抗生素对病毒所致的普通感冒几乎没有益处，抗生素的使用不仅加重了患者的经济负担，而且长期使用还会增加副作用和耐药等问题。再

比如中国缺血性脑卒中和短暂性脑缺血发作的二级预防指南明确指出糖尿病患者合并高血压时，降血压药物以血管紧张素转换酶抑制剂、血管紧张素Ⅱ受体拮抗剂在降低心脑血管事件方面获益明显。而我国的调查研究则显示，实际的服药率不到50%，此为使用不足。

而基于证据的临床实践指南（CPG）同样可减少上述不当使用情况，避免误用和滥用，改善使用不足、使患者得到应有且合理的医疗服务。

（三）医疗费用的日益增长

目前，全世界面临的难题是有限的医疗资源与日益增长的医疗保健需求之间的矛盾。我国卫生总费用占国民生产总值（GDP）比重，1991年为4.10%，2013年达到了5.57%，而在医疗费用快速增长下，预计2020年将达到6.19%，可能演化成严重的社会负担。即使如美国这样的发达国家，庞大的医疗补助同样也让政府不堪重负。因此，明智地使用有限的医疗资源已经成为国际共识，而制定临床实践指南，或许可以减少不恰当的医疗行为，改善患者预后，提高和保障医疗服务，这对制定医疗费用补偿政策、合理高效地使用有限医疗资源具有重大意义。

临床实践指南的作用可归纳为以下六点：①减少不同医疗机构和不同医生间医疗实践中的不恰当差异；②提高医疗质量、改善临床结局；③概述研究发现的结果，并使临床决策透明化；④促进卫生资源的合理高效利用，从而减少患者的医疗费用；⑤区分医学研究结果的优先次序，作为医疗保险的凭证；⑥有助于医务人员的终身继续教育。

2011年哈佛大学研究人员在《医院医学杂志》上发表了一篇题为“Use of UpToDate and Outcomes in US Hospitals”的文章，对比了1017家应用综合性循证指南UpToDate的医院和2305家未采用UpToDate医院的医疗保险受益人数据，同时评估了这些医院的医疗质量和效率。研究结果也表明，应用UpToDate的1017家医院每年可减少约372 500住院日，在三年观察期中累计挽救了约11 500例患者生命；根据全美医院质量联盟（Hospital Quality Alliance）所制定的指标来评定，发现应用UpToDate的医院提供了更佳的服务质量；同时研究还提示：医生使用UpToDate系统，可帮助他们作出正确的即时诊断和治疗决策，缩短了住院日、降低死亡风险且提高了医疗服务质量。

三、指南与临床证据的关系

临床证据指的是循证医学中与临床实践密切相关的研究结果，这里的研究主要是针对人体的，包括探讨诊断性试验准确性和精确性的诊断证据、评估临床诊治安全性和有效性证据以及研究疾病预后及预后因素的预后证据等等。因为上述临床证据的来源、科学性和可靠性不尽相同，其论证强度各异。循证医学倡导在临床实践中应尽可能使用当前可及的最新最佳证据、结合临床经验和患者意愿进行诊疗方案的选择。

在临床实践中，推荐采用加拿大McMaster大学临床流行病学与生物统计学家Haynes教授提出的、支持循证决策的“6S”证据模型，按6S自上而下进行证据检索，将会提高效率。该模型将整合有CPG的计算机决策支持系统，置于模型的最顶层，可根据个体患者的特征（如电子病历）链接相关证据，提醒或告知医护人员治疗的决策；其次临床实践指南和证据总结为第五层；然后是系统综述的摘要（如对原始研究和系统综述的简要描述）作为第四层；系统综述（如Cochrane系统综述）作为第三层；原始研究摘要作为第二层；原始研究（如PubMed数据库的原始研究）作为最底层。遗憾的是目前处于6S证据顶层的计算机决策支持系统寥寥无几，且未能做到及时更新。但可以预见这将是CPG网络资源利用研究的一个重要发展方向。

“6S”模型充分体现了临床实践指南在循证医学证据体系中的地位及其重要的临床指导意义。同样临床指南也是循证医学证据的一部分，它与原始研究证据和系统综述的区别在于：指南是针对具体临床问题，分析评价证据后提出具体的推荐意见，是连接证据和临床实践的桥梁和纽带。原始研究证据和系统综述是客观地提供研究结果和对结果的解释，是循证指南的基石，证据越可靠、可信，推荐意见越可信。

四、循证临床实践指南的网络资源

21世纪的循证医学资源增长迅猛，临床实践指南作为循证医学资源的特殊成员，其网络资源也日益增多。为此现介绍主要的临床实践指南网络资源分布及检索方法。

（一）原始研究证据数据库

美国国立医学图书馆开发研制的PubMed是循证医学原始研究证据数据库的代表，也是查找临床实践指南的重要渠道和资源，收录了世界各国制作的指南，在“Article types”的下拉菜单中选中“guideline”即可检索临床实践指南。

此外，通过检索中国知识基础设施数据库（China national knowledge infrastructure CNKI）、中国生物医学文献数据库（China Biology Medicine disc，CBMdisc）等可以检索到各种中文临床实践指南。

（二）综合性临床指南和循证证据资源库

UpToDate是基于循证医学原则的临床决策支持系统，是图文并茂的综合性循证指南、患者手册以及药物间相互作用查询系统，面向医生、药师提供即时、实证的临床医药信息，并可快速解答临床专业人员提出的临床问题。数据库中附有图片，包括图表、X光片、相片、影像等及MEDLINE文献摘要。UpToDate主题包括内、外、妇、儿等二十个类别，每个主题之下尚有更专精的类别，其内容系由网罗众多相关的文献而成，可快速地获得临床上最前沿医学问题的答案。目前在国内已经有中文版数据库。

类似的资源库还有英国医学会开发的Best Practice，目前尚无中文版。

（三）临床实践指南网站

世界各国都有相应的专业网站专门提供临床实践指南检索。其中比较大型且权威的有美国国立指南数据库和英国国家临床指南中心网站等。

1. **美国国立指南文库（National Guideline Clearinghouse，NGC）** 由美国卫生健康研究与质量机构、美国医学会和美国卫生健康计划协会于1998年联合开发，是一个提供CPG和相关证据且功能完善的免费数据库。NGC还可对检索到的指南进行比较。

2. **英国国家临床指南中心** 由英国国家卫生和临床优化研究所（National Institute for Health and Care Excellence，NICE）于2009年创建的多学科卫生服务研究机构，为NICE制定循证临床实践指南，目的是改善国家卫生体系（National Health Service，NHS）的医疗质量，是全球最大的临床指南制定机构之一，可在NICE网站的“Guidance”栏目中进行检索和浏览指南。在“正在制定的指南（Guidelines in development）”栏目中还可查看正在制定的指南及其发展史。

3. **苏格兰校际指南网络** 建立于1993年，为苏格兰的国家卫生体系（NHS）提供临床实践指南。收录有已发表的和正在制定的指南。重点关注肿瘤、心血管疾病和心理卫生等领域，且免费提供指南全文。

4. **国际指南网络** 成立于2002年，该网络通过降低世界范围的不恰当的干预差异来支持循证卫生服务、改善医疗卫生效果。国际指南网络下设多个工作组专注于特定主题并交换意见以改进研究方法。

5. **WHO指南** 比较专注于公共卫生、重大疾病和常见疾病的指南研发。

第二节 临床实践指南制定方法与流程

临床实践指南是从非正式专家共识性声明、到正式专家共识性声明（或指南）、再到循证临床实践指南的一个逐步发展的过程，制定方法越来越规范，循证指南的占比越来越高，而非正式专家共识性声明开始淡出大家的视野。

一、专家共识制定法

分为非正式和正式的专家共识制定法（consensus guideline development）。非正式的专家共识制定

法是由一组专家开会讨论，将一次或多次开会讨论后达成的共识形成推荐意见，再由专业学会或政府机构进行发布。这种共识只包括推荐意见而缺乏形成推荐意见的证据基础和制定共识的背景及方法，易受参会人员的专业、优势、性格、组织和政治因素等影响，可靠性和质量较差。

正式的专家共识制定法，常用的有名义群体法、德尔菲法、共识形成会议法等。例如名义群体法，首先是应邀专家召开预备会，就某一治疗措施专家组提供相关研究证据的综述并形成可能的适应证清单。然后在第一次正式会议时专家组成员对每个适应证打分以评价其适用性，打分通常使用量表的方式，量表分为 1 分到 9 分，1 分为完全不适用，9 分为特别适用，5 分为可用或不用。再次开会时专家们将小组集体评分的情况与自己的评分相比较，讨论不一致的原因，然后再次重复评分，在会议充分讨论的基础上修改评分。最后的评分反映了专家组成人员的一致性程度。这种制定法虽考虑了研究证据，但仍未将推荐意见与相关证据的质量明确联系在一起，指南质量受专家小组个人影响较大。

二、循证临床实践指南制定流程

循证指南制定法（evidence-based guideline development），是目前国际推崇的指南制定方法，即将推荐意见与证据质量明确地联系在一起，依据现有证据来确定推荐意见的强度。这也是循证临床指南的明显特征，是保证指南的科学、公正和权威的方法。

国际通用的循证指南制定流程一般包括 10 个步骤（图 12-1）。

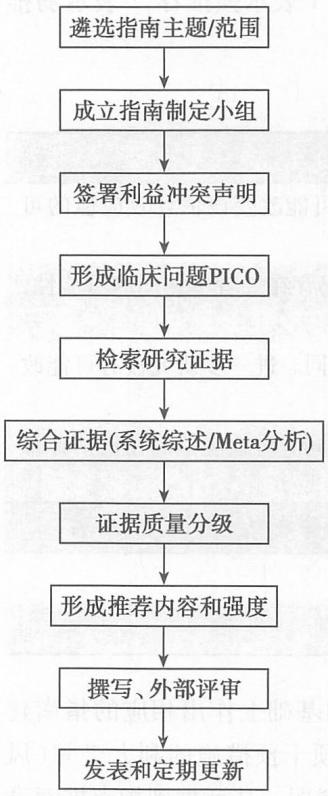


图 12-1　循证临床实践指南制定的主要步骤

1. 确定临床指南范围（scope of guideline）　指南范围可以明确界定该指南包含和不包含的内容、必须包括哪些重要的临床问题，提供工作框架。确定指南拟解决问题的重要性（如发病率、结局效果、经济费用）及制定指南的必要性、目的和使用范围。

2. 成立指南制定小组　由 13 到 20 人组成，成员主要包括：主席、临床医生（包括专科医生和全科医生）、护理专业人员、患者或其照护者、系统评价专家、卫生经济学家、信息学家等。一般每月召开 1 次会议，整个指南制定过程需要召开 10 ~ 15 次会议。一个临床实践指南的制定需要 10 ~ 18 个月。

3. 签署利益冲突声明　指南在发布前，指南制定小组成员均应对该指南中所涉及到的药物、器械等商业机构作出利益声明，任何受邀并切实参与到指南制定过程的人员都必须填写利益声明表，且必须同意在指南中发布。

4. 形成临床问题和系统综述问题　一般确定 15 ~ 20 个具体的综述问题，常见的是治疗、诊断、预后三方面的问题。每个问题均采用 PICO 方式。P（patients）：所要研究的人群是什么？I（interventions）：应该使用什么样的干预方法？C（comparisons）：目前所研究的干预措施有没有其他的替代方法？O（outcomes）：可能会出现什么样的结局？比如死亡率、发病率、复发率、副反应、短期结局、长期结局等。

5. 检索研究证据　文献检索的过程就是搜集证据的过程。由信息学家制定文献检索策略，尽量查全、查准。保证检索证据的过程是透明、全面、可重复的。每次应完整记录检索问题、检索日期、检索策略、使用的数据库、检索的结果等。

6. 综合证据　完成文献检索后需要阅读、筛选、评价所检索到的证据，采用系统综述的方法进行证据的综合，分别回答上述第 4 步提出的各个系统综述问题，保证指南的推荐意见是基于最佳的证据。

7. 证据质量分级　目前国际上常用的证据分级系统有两个：牛津循证医学中心证据分级系统和 GRADE 证据分级系统。

NOTE

（1）牛津循证医学中心（Oxford Centre for Evidence-based Medicine，OCEBM）的临床研究证据分级系统，对于诊断、治疗、预后、筛查等不同的循证问题有不同的分级标准，现以治疗为例来说明其分级方法（表 12-1）。

表 12-1 牛津循证医学中心临床研究证据分级（2011 年版）

证据分级	具体描述
1	以 RCT 为基础的系统综述，或以队列研究、巢式病例-对照研究为基础的系统综述，或效应值巨大的观察性研究
2	单个的 RCT，或效应值巨大的观察性研究（只有单独的研究显示有巨大的效应值，没有其他研究支持）
3	非随机、有对照的队列研究，或有足够的样本量、有足够随访时间的随访研究（如上市后研究）
4	病例-对照研究，或病例分析，或历史对照研究
5	有充分依据的机制研究

（2）证据推荐分级的评估、制定与评价（the Grading of Recommendations，Assessment，Development and Evaluations，GRADE）系统：把证据质量分为高、中、低、极低四个等级，推荐强度分为强、弱两个等级，现多数指南采用 GRADE 证据分级系统和推荐强度分级（表 12-2）。如以 GRADE 证据系统和推荐强度为依据的指南推荐意见用 1A，1B，1C，1D，2A，2B，2C，2D 等表示，数字 1 表示强推荐，2 表示弱推荐，证据由高到低的质量分别用 A，B，C，D 表示。

表 12-2 GRADE 证据质量与推荐强度分级的定义

证据质量分级	具体描述
高（A）	非常确信估计的效应值接近真实的效应值，进一步研究也不可能改变该估计效应值的可信度
中（B）	对估计的效应值确信程度中等，估计值有可能接近真实值，但仍存在二者不相同的可能性，进一步研究有可能改变该估计效应值的可信度
低（C）	对估计的效应值的确信程度有限，估计值与真实值可能大不相同。进一步研究极有可能改变该估计效应值的可信度
极低（D）	对估计的效应值几乎没有信心，估计值与真实值很可能完全不同。对效应值的任何估计都很不确定
推荐强度分级	
强（1）	明确显示干预措施利大于弊或弊大于利
弱（2）	利弊不确定或无论质量高低的证据均显示利弊相当

8. 形成推荐内容和强度 指南制定小组在证据分级和经济学证据的基础上作出相应的指南建议，权衡干预措施的利弊、健康获益和卫生资源后决定推荐的强度。当一项干预措施的利大于弊（风险和经济负担），强烈推荐；利弊均衡或者证据质量比较低时，推荐强度就减弱。从证据到指南推荐意见需要规范的方法，如共识形成会议、专家投票等。多数情况下，OECBM 的 1 级证据对应于指南中的 A 级推荐意见，2 和 3 级证据对应于 B 级推荐意见，4 级证据对应于 C 级推荐意见，5 级对应于 D 级推荐意见，但在制定指南或应用指南时还要权衡利弊，考虑经济学成本、患者价值观、医疗环境等因素，因此上述的对应关系并不是绝对的。同样道理，基于 GRADE 的 A 级证据不一定是强推荐意见，C 级证据也可以是强推荐。例如，妊娠妇女合并免疫性血小板减少症，治疗可首选静脉用大剂量丙种球蛋白（1C），虽无系统综述或 RCT 等高质量的证据支持，但因其安全性好，医生和患者可接受程度高，以及观察性研究发现有一定的疗效，虽然证据等级为 C，但强推荐使用。

9. 指南的撰写和外部评审 指南的基本结构包括以下几个部分：摘要，简介，制定方法，综述问

笔记

题，证据总结，比较详细的推荐建议和推荐强度，研究建议，参考资料，附录。指南初稿完成后，送外部评审（由非指南制定小组成员的专家组织外部审阅），反馈并提出修改意见。

10. **发表和定期更新**　当指南正式成文后，可制成各种版本发布，供不同的对象使用。除了全文外，形式上可以是摘要性的结论性建议，也可以是针对患者的教育手册。一般 3 ~ 5 年需要更新。

第三节　临床实践指南的评价

对同一种疾病，不同的国家或学术组织可能制定出不同的指南，指南的质量也参差不齐，甚至某些建议互相矛盾，这些都将给临床决策带来极大困扰。因此，在使用指南之前，临床医生应具有评价和鉴别指南质量高低的能力，进而判断该指南是否值得推荐，或者从众多的指南中挑选出最佳者用于指导临床实践。

一、临床实践指南的评价原则

首先，评价一个临床实践指南的好坏，应强调以证据为基础，即真实性。对指南的质量评价实际上主要集中在证据的收集、评价和合成，以及如何将推荐意见与相关证据进行综合的过程。

其次，当指南的质量真实性得到肯定后，就应对其重要性进行评价。该指南是否回答了临床需要解决的重要问题？这些问题必须是临床医生在工作中面对的问题。当然，由于临床所面临的问题相当复杂，一个指南不可能囊括所有的临床问题，只能是绝大多数重要问题。该指南是否包含了成本效果等患者关心的指标？除了疗效、安全性以外，成本费用也是非常重要的结局指标。

最后，应用指南前还要评估其适用性。一个指南的成功实施依赖于四个因素（4B），如下：

1. **疾病负担（burden of disease）**　如该病在本地区的发病率很低带来的疾病负担低，因而无需使用参考指南？我的患者是否不可能出现指南中所描述的结果？如果是这样，则应用指南不仅浪费时间金钱，还可能造成不必要的伤害。反之，则应积极使用指南。具体到个体患者时，还应综合考虑：①我的患者与研究证据中的对象情况差别大吗？②在现有的环境条件下这种治疗方案可以实施吗？③指南中推荐的治疗方案有哪些潜在的利弊？

2. 患者对治疗价值的看法（beliefs），对治疗方案的利弊评价是否有别于指南中推荐的内容？

3. 执行该指南所推荐的措施所需成本（bargain）有多大，机会成本如何？其他干预措施能否有更大的获益？

4. 对患者实施该指南是否存在不可克服的障碍（barriers）？这些困难包括地域性的（如该地区根本不能实施此治疗方法）、传统性的（如长期惯用另一种治疗方法）、权威性的（教授怎么说就得怎么做）、法律性的（医生惧怕因为舍弃了常用但效果不明显的疗法会遭到起诉）、行为性的（医生无能为力或患者不能服药）。若这些障碍明显，则无法实施指南。

总之，指南的推荐意见是原则性的，应在一般原则指导下，实施个体化诊治，同时还应结合患者（或其家属）的需求及价值取向。

二、临床实践指南的评价工具

许多国家和学术团体制定了 CPG 的评价工具以便科学客观地评价 CPG。目前，国际上应用比较广泛的指南研究与评价（appraisal of guidelines research and evaluation，AGREE）量表，2003 年由 13 个欧洲国家的研究者制定和首次发布，2009 年修改后称为 AGREE Ⅱ。

AGREE Ⅱ由用户手册，6 个领域（指南的范围和目的、参与人员的组成、指南开发的严谨性、指南的清晰性、指南的适用性和制定工作的独立性），23 个条目，以及 2 个总体评价条目组成（表 12-3）。每个领域针对指南质量评价中的某一方面。

表 12-3　AGREE Ⅱ评价工具

领域(维度)	条目	评分或意见等级*
领域 1. 范围和目的	1. 明确描述指南的总目的	1～7
	2. 明确描述指南涵盖的卫生问题	1～7
	3. 明确描述指南适用人群(患者,公众等)	1～7
领域 2. 参与人员	4. 指南开发小组包括了所有相关专业的人员	1～7
	5. 收集目标人群(患者,公众等)的观点和选择意愿	1～7
	6. 明确规定指南的使用者	1～7
领域 3. 严谨性	7. 应用系统方法检索证据	1～7
	8. 清楚描述选择证据的标准	1～7
	9. 清楚描述证据的强度和局限性	1～7
	10. 清楚描述形成推荐建议的方法	1～7
	11. 形成推荐建议时考虑了对健康的益处、副作用以及危险	1～7
	12. 推荐建议和支持证据之间有明确的联系	1～7
	13. 指南在发布前经过外部专家评审	1～7
	14. 提供指南更新的步骤	1～7
领域 4. 清晰性	15. 推荐建议明确,不含糊	1～7
	16. 明确列出不同的选择或卫生问题	1～7
	17. 容易识别重要的推荐建议	1～7
领域 5. 应用性	18. 指南描述了应用时的促进和阻碍因素	1～7
	19. 指南为如何将推荐意见应用于临床实践提供了建议和(或)配套工具	1～7
	20. 临床指南考虑了应用推荐意见时潜在的资源投入问题	1～7
	21. 指南提供了监督和(或)审计标准	1～7
领域 6. 独立性	22. 赞助单位的观点不影响指南的内容	1～7
	23. 指南开发小组成员的利益冲突要记载并公布	1～7
指南全面评价	1. 指南总体质量的评分	1～7
	2. 我愿意推荐使用该指南	是 是(修订后使用) 否

每个指南应该至少有 2～4 个评价者独立进行评价,评价人数最好是 4 人,便于进行重复性和可靠性评价。以上 23 个条目均采用 1～7 分的记分法,1 分表示很不同意,逐步递增到 7 分为很同意。在每一个条目下,都有相应的条目解释、信息查找以及如何评价等的说明。各领域的记分采用标准化百分法。计算方法如下:

(1) 该领域实际得分分值=该领域所有评价者每个条目记分之和。

(2) 该领域最高可能分值=评价者人数×该领域条目数×7。

(3) 该领域最低可能分值=评价者人数×该领域条目数×1。

(4) 该领域标准化分值=100%×(实际得分分值-最小可能分值)/(最大可能分值-最小可能分值)。

最后,AGREE Ⅱ还有 2 个条目用于对指南进行总体评价,包括对指南质量的总体评估,以及是否推荐在实践中使用该指南。应注意 6 个领域(维度)的得分是独立的,不宜合并成一个质量评价的总分值。

AGREE Ⅱ量表的不足在于缺乏具体标准来评价指南的临床内容,缺乏对推荐意见合理性的审查,因此还需要专家评议和临床试用来弥补。尽管如此,AGREE Ⅱ量表仍是评价 CPG 的首选工具。

第四节　临床实践指南应用的原则和方法

一、临床实践指南的应用原则

临床实践指南是为临床医生处理临床问题制定的参考性文件，是推荐应用而非强制执行，应避免不分具体情况盲目地、教条地照搬使用。应用时注意考虑以下原则：

（一）个体化原则

在应用指南时，医生应充分考虑该患者临床特征是否与指南中目标人群一致。面对具体的个体患者，临床医生应该在指南的指导下，综合具体病情和多方面的因素个体化地选择治疗方案。临床医生应具备以下两方面素质才能保证指南的正确使用：第一，快速判断患者病情状况和建立诊断的能力；第二，具有对患者接受干预措施可能获得的利弊进行评估的能力。例如冠心病合并糖尿病患者根据指南应强烈推荐使用阿司匹林，但若合并有严重的消化道溃疡病，阿司匹林需慎用甚至不用。

（二）适用性原则

自己患者的情况与指南的目标人群相似吗？如果相似，可以考虑应用指南推荐的干预措施。并结合本地区或医院目前的医疗条件，患者的经济状况，评估其对医疗费用的承受能力，以及当地医疗保健系统的覆盖支持能力等因素，评估该干预措施的可行性和费用-效益比。例如各国指南都推荐急性心肌梗死早期（3～12小时内）行经皮冠状动脉介入（PCI）治疗，但在我国绝大多数基层医院无条件开展此项技术，多数心肌梗死患者也不能承受高昂费用，此时的诊治就只能先采取指南建议的药物治疗方案，待病情适宜条件下转诊上级医院。

（三）患者价值取向原则

患者或其亲属的价值取向和意愿在临床决策中具有重要的作用。医生应事先与患者或家属沟通，了解他们期望的结局指标以及与指南中的结局指标是否一致，差距有多大。应充分尊重患者或家属的选择。例如对1例晚期肺癌患者，应用体能状况（performance statue，PS）评分标准进行评估，得到评分为3分（PS评分一般要求不大于2分才可能实施有效的抗肿瘤治疗），在这种情况下指南不支持使用细胞毒类药物，因为放化疗与姑息治疗相比对生存期没有改善，甚至严重副作用会影响患者的生存质量。但患者及家属坚持认为不积极治疗就等于放弃，最后还是尊重患者及家属的意愿，试行放化疗，并予以严密监测患者病情变化。

（四）时效性原则

随着医学的快速发展，基础和临床研究证据也在不断更新。过去认为有效的治疗手段可能被新的证据证明无效，而过去认为无效甚至禁忌的治疗手段可能被新的证据证明有效。例如既往认为慢性充血性心力衰竭是使用β-受体阻滞剂的禁忌证，但目前大型随机对照试验一致证实β-受体阻滞剂可以显著改善慢性心力衰竭患者的预后。因此，现有指南（Ⅰ类推荐，A级证据）β-受体阻滞剂治疗慢性充血性心力衰竭。

（五）后效评价原则

后效评价是指在患者接受根据CPG制定的诊治方案后，继续临床随访以评价患者病情的变化。后效评价在整个循证临床实践中具有重要作用，也可为指南修订和更新提供临床资料。

二、临床实践指南的应用技巧

在实际临床工作中，临床医生应用CPG应注意以下技巧：

（一）明确该指南制定的方法，一项真正的、以循证为基础的CPG较非循证CPG的可靠性更强。循证指南的核心是系统全面地收集证据，对证据质量进行评价和分级，根据证据的级别，并结合临床实际、医疗资源、成本效益、患者价值观等全面权衡利弊后提出指南的推荐意见及其强度。

（二）了解并分析指南中的证据水平与推荐建议强度，并明确它们之间的关系，以便判断推荐意

见的可靠程度。不同指南可能采用不同的证据分级系统，例如一项关于免疫性血小板减少症（ITP）的英国指南，采用的是牛津循证医学中心的分级方法，用“A级推荐，I类证据”来表达“如果儿童有广泛皮肤黏膜出血症状，应该应用大剂量泼尼松治疗，剂量为4mg/（kg·d）”这条推荐意见。而美国ITP指南采用的是GRADE证据分级系统，用“1B（强推荐，证据等级为B）”来表达“如果儿童ITP需要治疗，可以采用静脉用丙种球蛋白（0.8～1g/kg）或短疗程的糖皮质激素作为一线治疗”这条推荐意见。

（三）依据推荐意见强度确定是否应用于临床。如果患者的病情符合指南推荐，应该尽量采用指南的建议，特别是强推荐的意见，证据等级来自A级的指南意见更要优先考虑和应用，没有特殊的理由不应该拒绝应用。如果是弱推荐或证据等级很低，可以考虑不应用。

（四）当有多个指南的推荐意见不同时，要进一步结合本国国情、患者意愿、医疗条件等综合考虑。

（五）如何应用国外指南？国外许多指南制定严谨、质量可靠，深受广大医生的喜爱，但在应用时要注意国内的适用性。由于不同国家或地区间文化、组织的差异，即使是基于相同的证据也可能会导致推荐意见的差异，这意味着在一定环境下产生的指南，可能并不适合于另一环境，需要结合国情改编后应用。例如，美国输血指南提出对于慢性贫血成人患者的输血指征是血红蛋白<70～80g/L，但国内由于血源紧张和患者对长期输血有恐惧心理，一般都是<60g/L才输血，多数患者可以耐受，没有显著的不良反应。

（六）注意消除指南的实施障碍，避免指南使用不足。指南在实施过程中总会遇到一些障碍，如：①社会因素障碍，像某些新的治疗措施医保不予赔付；②个人因素障碍，如医生过度自信，盲目使用指南；③环境因素障碍，被医药宣传信息误导，脱离指南、习惯性给予“常规治疗”等。

在应用指南过程中，应充分重视个人沟通能力的重要性，加强与患者及家属的沟通交流，既可增加患者及家属对指南应用的理解，也是临床医生正确使用指南、进行个体化临床决策的基础。应用指南时应充分体现循证医学的理念，即根据患者的具体临床情况，将当前所获最佳证据与临床技能和经验相结合，考虑成本-效益比及当地卫生资源的实际情况，并在充分尊重患者及其亲属的价值取向和意愿的基础上，作出最佳的、综合性临床决策。

第五节　循证临床实践指南应用案例

一、提出问题

一女性患者，28岁，首次妊娠7个月，常规体检时发现血小板为34×10^9/L，多次复查血小板均在$(28\sim36)\times10^9$/L，皮肤黏膜无出血点，偶有刷牙后牙龈少量出血，胎儿发育正常。既往体健，无血小板减少病史，无其他疾病。经仔细检查，排除其他疾病引起的血小板减少，诊断为免疫性血小板减少症（immune thrombocytopenia，ITP）。妇产科医生请血液科医生会诊，协助治疗，血液科医生的意见不统一，有医生倾向于用泼尼松治疗，或大剂量静脉丙种球蛋白（IVIG）治疗，也有医生倾向于不给予任何治疗，只是观察血小板变化。

1. 临床问题的提出

（1）现患者妊娠7个月，合并ITP，无明显出血倾向，是否应该给予治疗？是否需要用泼尼松治疗？是否需要用IVIG治疗？

（2）分娩时血小板计数维持在多少以上可以减少分娩时母亲和胎儿的出血并发症？

（3）分娩时应该采取什么分娩方式，以保证母婴平安？

（4）新生儿需要哪些监护和治疗？

2. 把临床问题转化为PICO问题　为了更明确临床问题，方便检索，把临床问题转化为PICO问题。

（1）研究对象（population，P）：妊娠，ITP患者，新生儿。

（2）干预措施（intervention，I）：泼尼松，IVIG。

笔记

（3）比较措施（comparator，C）：不用药物治疗。

（4）结局（outcome，O）：出血发生率，分娩方式（剖腹产，经阴道产）。

二、检索临床指南

1. 在 PubMed 中检索　用检索式“purpura，thrombocytopenic，idiopathic”[MeSH Terms] OR immune thrombocytopenia[Text Word]，限定在 guideline，检索 PubMed 数据库，共检索到 8 个指南，阅读题目，排除非英文文献以及儿童 ITP 指南，发现与本案例密切相关的、最新的指南为 2011 年发表在 *Blood* 杂志的美国血液学会循证指南“The American Society of Hematology 2011 evidence-based practice guideline for immune thrombocytopenia”。

2. 在国外各种指南网站中检索　在国际指南网络（GIN）、英国国家医疗保健研究所（NICE）、苏格兰指南网络（SIGN）、美国国立指南文库（National Guideline Clearinghouse，NGC）中进行检索，也可以检索到相关的指南。例如 NGC 中可以检索到与免疫性血小板减少症（ITP）相关的指南 8 个，阅读题目发现相关性均不如上述美国血液学会制定的指南。由于 NGC 只收录近 5 年指南，所以 2011 年发表的这个美国指南并未被收录。

3. 在中文数据库中检索　在中文常用数据库进行检索，找到一篇 2016 年发表的专家共识“中国成人原发性免疫性血小板减少诊断和治疗专家共识（2016 年版）”，不属于循证指南，内容比较简单，其中也没有包括妊娠 ITP 的治疗，所以不采用。

三、阅读和评价指南

通过上述检索，找到较新、最相关的美国指南进行阅读和评价。

1. 指南的制定方法　该指南采用严格的系统评价的方法制定指南。指南小组检索了 EMBASE 和 MEDLINE 数据库，检索了 1996—2009 年间的相关临床问题。1996 年美国血液学会首次发表了有关 ITP 的指南，主要以 RAND 积分法和专家共识为主而制定指南。本次指南在 1996 年的基础上，主要检索 1996 年后有哪些新的文献证据和推荐有哪些新变化。

检索时如果没有找到相关问题的系统综述，则检索 RCT 文献。如果没有 RCT，则检索队列研究和病例-对照研究、病例分析。为了保证证据质量，剔除病例数小于 50 例的成人 ITP 的病例分析，剔除病例数小于 25 例的儿童 ITP 和继发性 ITP 的病例分析。该文献检索的方法是依据苏格兰指南网络小组（SIGN）的建议。

指南推荐意见经过多次指南小组会议投票和审查后决定。最后由一个外部同行评议小组来评价其外部真实性。所以该 ITP 指南属于循证指南，推荐意见可靠。

2. 评价指南　评价该指南可以从 2 个方面进行评价：真实性和适（实）用性。真实性主要包括该指南是否包括近 12 个月以来最新的、最全面的文献证据，并采用科学的方法如系统综述方法对这些证据进行分析、评价和分级；每一条推荐证据均有推荐分级、文献出处。该指南采用科学的方法对证据进行检索、合成，提出推荐等级。适（实）用性主要指是否包含了全部可能应用的人群，如儿童、孕妇等特殊人群；是否包含了生存质量、成本费用等各种患者关心的结局指标。该指南包含了各种特殊人群，结局指标中也包括了生存质量等多种结局，但未进行成本-效益分析，没有对费用进行分析。

同时也采用 AGREE Ⅱ对该指南进行评价，包括 6 个方面 23 个条目：指南范围和目的、参与人员、严谨性、清晰性、应用性、独立性。总体评价该指南质量较高，制定方法严谨，结论可信，适（实）用性强，可以采纳。

3. 指南推荐意见和等级　指南中有专门章节提到关于妊娠过程出现 ITP 的治疗意见（表 12-4）。

（1）案例的第一个临床问题为是否需要治疗？什么治疗最佳？指南中指出目前还没有 RCT 研究比较治疗和不治疗的后果有何不同，所有的推荐意见都是基于观察性研究。糖皮质激素如泼尼松和 IVIG 都没有致畸性，对孕妇是安全的，需要时可以应用。糖皮质激素有可能增加妊娠糖尿病和产

后精神障碍的发生。关于CD20单抗应用于这类患者尚无报道，只有少数关于Anti-D的个案报道。如果糖皮质激素和IVIG均无效，血小板持续减少有致命出血倾向，可以考虑脾切除，一般在妊娠3个月之后进行手术安全性较高，但引起胎儿流产的可能性增加。指南推荐：如果需要治疗则首选IVIG和糖皮质激素（推荐等级1C），IVIG剂量为每次1g/kg，泼尼松剂量为每天1mg/kg。推荐等级为强推荐，但证据等级为C。

表12-4　美国ITP循证指南中部分推荐意见

成人新诊断的ITP的一线治疗方法
• 血小板<30×10^9/L时需要给予治疗（推荐等级2C）
• 长期的糖皮质激素治疗疗效优于短期激素治疗或IVIG治疗（推荐等级2B）
• IVIG和激素联合应用可以更快速提高血小板计数（推荐等级2B）
• 如果激素有应用禁忌症时，IVIG或anti-D（合适的患者）可以作为一线治疗方法（推荐等级2C）
• 如果应用IVIG，初始剂量为1g/（kg·次），需要时可以重复应用（推荐等级2B）
妊娠妇女合并ITP的治疗
• 妊娠妇女需要治疗时可以选择糖皮质激素或IVIG（推荐等级1C）
• 生产方式由产科情况而决定（推荐等级2C）

血小板计数下降到多少应该进行治疗，目前还无具体标准，一般根据临床出血倾向、治疗获益等综合判断。成人初发ITP的治疗阈值为血小板<30×10^9/L（推荐等级2C），孕妇可以参照执行。推荐等级为弱推荐，证据等级为C。

（2）案例的第二个问题是产前血小板维持在多少以上才比较安全、减少出血并发症？关于血小板多少才是安全的麻醉和生产阈值，指南中对这个问题未明确推荐，因为现在没有明确的结论，也没有确切的证据支持。又在NGC网站检索到了一个荷兰血小板输注的指南，该指南指出硬膜外麻醉时血小板的安全阈值是血小板计数大于50×10^9/L（推荐等级C），可以作为参考。

（3）案例的第三个问题是分娩方式何种最佳？妊娠妇女合并ITP患者，分娩方式主要由产科情况而定（推荐等级2C），综合考虑产妇出血情况、麻醉、血小板计数等而决定，经阴道生产或剖腹产都是可行的。

（4）案例的第四个问题是新生儿需要什么处理？美国指南中认为ITP母亲生产的新生儿血小板计数与新生儿出血没有十分明确的因果关系，发生新生儿出血的概率非常低，因而没有明确的推荐建议。荷兰血小板输注指南中提出对于这类新生儿需要至少检测5天的血小板计数，如果<50×10^9/L需要治疗，首选的治疗是输注IVIG（推荐等级C）。

四、应用指南

根据上述证据，结合本例患者，制定治疗方案：

1. 因该患者血小板>30×10^9/L，所以目前不需要泼尼松治疗，应密切观察有无出血倾向，每周随访血小板计数，若血小板持续<30×10^9/L，可给予泼尼松或IVIG治疗。

2. 分娩前根据血小板计数，决定是否预防性输注浓缩血小板。若血小板<50×10^9/L，则输注一个单位的浓缩血小板，使血小板计数>50×10^9/L，减少麻醉和生产时的出血风险。

3. 分娩时根据血小板计数、产科情况和患者意愿选择分娩方式，经阴道分娩和剖宫产均可行。

4. 每天连续检测新生儿的血小板计数，连续一周，若<50×10^9/L需给予IVIG治疗。

五、后效评价

随访此患者，未予特殊药物治疗，血小板计数维持在>30×10^9/L，无出血倾向，根据患者及家属意愿，选择剖宫产，术后无明显大出血。新生儿健康，血小板计数正常，未予治疗。

（王小钦）

笔记

第十三章　循证医学的个体化实践

循证医学实践的基本步骤包括循证问题的提出，证据检索，证据评价，证据应用和后效评价。循证个体化实践就是应用证据，强调将最新最佳证据用于个体患者，以取得最佳临床诊治效果。

由于临床研究证据和成果是从基于个体特征的感性认识升华为对群体的共性规律认识，证据或研究结果就群体而言是真实的，其推广和应用对改善和提高患者群体的疾病防治效果会有积极的作用。然而就个体患者而言，可能与证据来源群体存在特征上的差异，比如所处地域、种族、年龄分布、疾病状态、社会经济和环境背景等方面不同，导致临床效果难以重现。因此，能否将从群体获得的最佳证据或研究结果用于解决个体患者的具体问题，还应结合患者个体化特征、意愿/价值观等，这是临床医生在实践循证医学时需要重视的问题，也是将最佳证据能否成功转化为最佳临床诊治效果的关键。

第一节　最新最佳临床研究证据是循证个体化实践的基础

循证问题无外乎涉及病因、诊断、治疗、预后等问题，当找到证据后，按照病因、诊断、治疗和预后证据的原则进行科学性评价。实践循证医学的临床研究证据，应该是最新最佳证据（current best evidence）。按照临床流行病学和循证医学的标准和原则，最新最佳证据应具备以下特征：

一、真实性

证据的真实性（validity）指研究结果与实际真实值的符合程度。由于研究设计、实施和评价过程有诸多影响因素，导致研究结果或多或少与实际真值之间存在偏差，这就是为什么同样的研究设计会得出不同结论的原因。为避免伪证据或夸大效果的研究结论对临床诊治方案的影响，对拟采用的证据首先要评价其真实性，应从以下几方面进行：①设计方案是否科学合理，对照组设立是否恰当；②研究对象的诊断标准及其纳入和排除标准是否明确；③组间基线是否可比、干预措施和方法是否科学有效和安全；④终点指标是否确切、恰当；⑤干预措施实施及结果测量、分析和报告是否采用盲法；⑥研究有无偏倚存在及是否采取了防控措施；⑦患者的依从性如何，随访是否充分；⑧资料收集、整理、统计分析方法是否合理等。

然而证据真实程度如何是无法直接度量的，只能间接推断。在任何一个临床研究中都会遭受选择偏倚、测量偏倚、实施偏倚、随访偏倚以及报告偏倚的影响，导致结果失真。因此评价证据的真实性，实际上是评价临床研究发生偏倚的风险大小，如果偏倚风险小，则可判定证据的可信度高，相反，如果发生偏倚的风险大，则证据可信度大打折扣。

二、重要性

证据的重要性（importance）指其临床价值和意义。重要性评价应建立在证据真实的基础上，否则不真实的证据就没有重要性可言。不同类型的研究证据，重要性有着不同的表达形式。采用定性和量化的指标将有助于判断证据的重要程度。如按照 GRADE 系统分级系统的评级方法，围绕证据指标本身的重要性及其临床实际价值等，采用三类 9 级法判断证据对患者的重要程度。第一类（7～9 级）

为影响决策的关键且重要证据指标;第二类(4～6级)为影响决策的重要但非关键证据指标;第三类(1～3级)为对决策者和患者影响不大的证据指标(表13-1)。

表13-1　证据指标重要性分级

分级	重要性	临床决策价值及意义
9 8 7	关键且重要结局指标(7～9级)	影响决策的关键结果
6 5 4	重要但非关键结局指标(4～6级)	影响决策的重要但非关键结果
3 2 1	不重要结局指标(1～3级)	对决策影响不大的结果

不同类别证据,所选用指标和临床重要性的判定标准各有所侧重。

三、适用性

获得真实、重要的证据后,还需要进一步评价其适用性如何。如证据能直接用于治疗我的患者吗?是否考虑了临床所有重要的结果?利弊分析如何?是否利大于弊?

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

来源于不同国家和地区的临床研究证据,由于人种及生物学上的差异,可能导致不同反应的出现;社会文化背景的差异可能影响证据的接受和推广;经济发展水平对新技术和新药物的应用有着明显的促进或制约;医务人员的知识水平和医疗条件在不同国家和地区的差异也会影响证据的使用。对来源于不同国家和地区的临床研究结果,其引用要充分考虑上述的差异,一定要结合个体患者的特点,具体分析相关证据的应用价值及可行性,不能盲目照搬。

四、时效性

不论是对病因的探索、对疾病诊断准确性的渴求,还是对疾病良好结局的期许都会促进临床研究者们不断地围绕临床需要解决的、新的热点难点问题展开深入研究,以进一步完善和拓展现有的证据,甚至否定现有证据。因此,还要考核证据的时效性,指导循证个体化实践的证据应是最新最佳的证据。

第二节　准确把握患者的个体化特征是循证个体化实践的基本功

应用最佳证据应准确把握患者的个体化特征,提高患者的依从性。患者的依从性受自身文化水平、社会经济状况、价值观甚至宗教信仰的影响。了解患者心理及所处的社会经济状态,倾听患者对治疗的意见,尊重他们的价值观,给予患者相关的医学知识教育,加强医患交流,消除患者可能存在的顾虑,尽可能的给患者提供方便的服务,并与患者建立良好医患关系,让患者愿意接受最佳证据指导的治疗,是循证医学实践成功的保障。患者个体化特征包括疾病生物学特征,病理生理特征,社会-心理及经济人文特征等等。

一、生物学特征及相关依据

任何临床研究的设计和实施都应有生物学依据(biologic evidence)。在证据应用中也要考虑临床

诊治的生物学依据。

从基于人群的流行病学研究或基于临床患者的观察总结，常常会发现不同人种或个体之间存在生物学上的差异。例如黑人的高血压发病率较其他人种高且危害重，临床证明有效的β-受体阻滞剂、血管紧张素酶抑制剂，对黑人高血压患者的治疗效果较利尿剂差；同时血管紧张素酶抑制剂引起的血管性水肿显著，较其他人种的发生率高。在丙型肝炎的治疗中也发现，黑人对聚乙二醇干扰素-加利巴韦林治疗的持续病毒应答（sustained virus response，SVR）率较白种人低。家族遗传性也体现在疾病的发病率和对药物的反应性上的差异。比如乳腺癌、胰腺癌和结肠癌有家族聚集的倾向，发病的风险高于普通人群。因此在疾病防治中要注意人种、民族和遗传因素的生物学特点。

临床治疗中还观察到，使用抗生素治疗感染性疾病过程中，同一种致病菌存在对同一种抗生素的敏感菌株和耐药菌株，或在抗生素不断使用中渐渐产生对抗生素耐药现象。在肿瘤患者的化疗治疗中，同一种组织来源的肿瘤，也观察到化疗敏感型或耐药型的差异。这些差异的存在和产生都需要研究者去探索其发生的生物学机制，从细胞甚至分子水平上予以解决，以期为患者制定个体化的诊疗方案。

随着人类基因计划研究的深入，对于个体基因遗传特征与疾病的临床表象的关系探讨，尤其是单核苷酸多态性与疾病发生及临床用药敏感性的预测，可能为今后临床研究和诊治提供更多的个体方面的生物学依据。

二、病理生理学特征及相关依据

临床实践中，疾病的发生、病情轻重、并发症发生、结局以及临床干预措施对疾病的影响都有相应的病理生理变化和特点。临床研究中若纳入不同病理生理状态的研究对象，不仅临床表现千差万别，对诊疗措施的反应也大相径庭。在证据的应用中要考虑患者的病理生理状态以及治疗过程中的病理生理反应。例如，曾在我国流行的急性传染性非典型肺炎（severe acute respiratory syndrome，SARS），部分中毒症状及肺部炎变损害明显、呼吸功能明显障碍者，适时适量的应用肾上腺皮质激素治疗，取得了较好的临床效果，挽救了不少患者的生命。但对于中毒症状以及肺部损害不严重、呼吸功能影响较轻的患者，应审慎应用激素，避免因给药带来不良后果（如股骨头坏死）。

任何最佳证据的应用，一定要仔细考虑患者由于病理生理损害不同而导致临床病情的复杂性。在评估疾病预后和对有关危险因素的干预，进而改善患者预后、提高患者的生存质量，这就要求将多种相关因素的研究证据，结合患者疾病的病理生理特点进行具体分析和评价，以估计可能发生某一事件的概率并提出有针对性的干预方案，从而防止或降低不良事件的发生，以达到最终改善患者预后的目的。

三、社会-心理-人文-经济特征及其相关依据

在宏观层面上，鉴于不同国家和地区，社会制度、人文和自然环境、经济发展水平和人群的受教育程度存在着极大的差异，这些差异不仅体现在临床研究的切入点和实施不同外，也会在证据的推广应用中存在差异。例如：宫颈癌是发展中国家妇女常见的恶性肿瘤之一，极大的危害着女性健康。宫颈脱落细胞学检查已经成为经典的宫颈癌筛查手段，并被全世界广泛接受和采纳。但在一些贫困国家和地区，由于资源匮乏和技术人员缺乏，醋酸试验或碘着色通过肉眼观察被用来取代细胞学筛查。而且宫颈癌筛查的宣传、人们对其重要性的认识以及传统观念影响等，使该方法的接受程度与发达国家相比也大大落后。若在这样的地区，强制推行 HPV 检测联合宫颈脱落细胞学筛查宫颈癌及癌前病变，就显得勉为其难了。相反，发展中国家根据当地的经济发展水平、宗教和文化现状，针对一些对当地人群健康危害大的疾病如急性传染病、呼吸道感染、腹泻、营养不良、地方病等展开研究，或者推广诸如上述宫颈癌肉眼筛查法等行之有效的证据，这些在发达国家就没有如此重要的价值。

NOTE

第三节　个体化特征与证据的精准对接是循证个体化实践的成功关键

证据应用时要精准对接患者的个体化特征(生物学特征、病理生理特征、社会-人文-经济特征等),具体情况具体分析,权衡利弊,不能搞一刀切。

最佳证据是否可以用于指导个体患者的医疗决策,一个显而易见的问题是:拟采用的诊治措施能给个体患者带来多大的利益,同时还要考虑它们被应用后可能产生哪些不良反应及其对个体患者造成的危害程度,即进行利弊分析。所谓"利"是指临床干预措施给个体患者带来的临床治疗获益,即疾病缓解的程度、治愈的可能性等,而"弊"则指不良反应的大小及危害程度。从临床经济学的角度还要考虑成本-效果比值。所以在证据的应用中要客观评估临床诊治措施对个体患者的利与弊,选择利大于弊的证据。

临床意义或价值有很多评价和表述的指标,一些量化的指标更有利于医生间或医患间的交流。例如,对个体患者的诊治作出决策,可根据利弊比(likelihood of being helped vs harmed,LHH)来衡量证据应用可能为患者带来的最佳效果和不良反应。比如某种防治措施的 NNT 为 10,NNH 为 30,即此种措施的利弊比值比(LHH)= 1/10∶1/30,约为 3∶1。意为该措施对患者带来的受益 3 倍于危害,但这一结果(利大于弊)并不是决策的最终依据,还要进一步结合专业及临床实际,考虑患者意愿及价值取向来校正 LHH,并最后决定是否采用该措施。表 13-2 可通过获得的 NNT 及 NNH 查得利弊比。

表 13-2　干预效果的利弊比(LHH)

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

医生的知识水平是循证医学实践不可或缺的要素,在充分运用临床基本知识和积累经验基础上,再结合具体患者的实际问题,搜寻和查找新近最佳证据为个体患者诊治服务,这是一个优秀临床医生的职责所在。医生是否具备循证医学实践的知识,能否在众多的研究结果中筛选到最佳证据,是否具有对先进研究成果的敏锐观察力,都关系着循证医学实践的最终效果。

临床研究的宗旨,就是要获得对防病治病有价值的成果。然后又将其应用于防病治病的具体实践,在实践中进一步评估临床效果,并从具体的患者个体应用中发现新问题,以此不断提高临床诊治的水平,达到保障人民健康的最终目的。循证临床实践应遵循个体化的原则,以最终解决患者个体的重要临床问题。

(刘续宝　康德英)

第十四章　临床决策分析

临床实践过程中的各个方面都可以应用决策分析，如病因、诊断、治疗、预后等决策，本章主要介绍治疗过程中的决策分析。

第一节　概　　述

决策（making decision）是基于不确定性的问题，通过一些方法与手段，从众多备选方案中选定最优方案的一个过程。决策分析（making decision analysis）则是通过决策模型再现问题、利用概率和结局估计值等，帮助确定最佳行动方案的一个过程。而临床决策分析（clinical decision analysis，CDA）是指由医务人员参与实施的、针对疾病的诊断和防治过程中风险与获益的不确定性，通过查阅文献资料，充分掌握证据，特别是在掌握最新最佳证据的基础上，结合以往临床经验和患者的实际情况，分析比较两个或两个以上可能的备选方案，从中选择最优者来进行临床实践的决策过程。

在临床实践中，当掌握了患者的全部情况、明确了需要解决的临床问题，以及找出了解决这些问题的最佳证据之后，那么就要对患者的诊治处理进行具体的临床决策分析，而后作出正确的决断。决策正确与否取决于3个前提条件：备选方案是否齐全、各事件的概率估计是否准确、结局的定量是否合情合理。无论是对概率还是结局的定量估计，都应来自证据充分的临床研究或文献资料。因此，决策前，应开展相应的循证医学研究，通过广泛调查、文献复习、科学分析，以充分掌握基础信息，进而选用论证强度高的结果评价备择方案，或应用准确的概率来描述事件的不确定性，以及使用合理的效用值来定量描述结局，然后计算各备择方案的期望值，从中选择收益最大者，应用于临床实践。

进行临床决策的难点在于缺乏各种决策证据，如干预后出现的各种结局事件常不确定，临床事件的发生不以人的意志为转移，虽可预测，但发生概率难以确定；同时鉴于各种备选方案都有优缺点，需要权衡利弊后加以取舍，也往往令决策者左右为难。

一、临床决策的模式

（一）患者作主模式

患者作主模式也称单纯知照模式（pure informed model），是指由患者自己作出选择。即医生只是提供各种方案的利弊等相关信息，患者根据自身的经验以及理解独立作出选择。在这里，患者是唯一的决策者，医生只提供客观信息而不参与决策。

（二）医生作主模式

医生作主模式也称家长作风模式（paternalistic model），是由医生为主导作出决策。即医生单独或者和其他医生一同考虑方案的利弊后而作出选择。在这种情况下，患者不参与决策，但前提是：假定医生知道哪种方案对患者最为适合。

（三）共同决策模式

共同决策模式（shared decision model）是指在决策过程中医生与患者保持互动、共同作出决策。这种模式的前提是信息充分对称，即医生向患者提供病情相关的信息，包括各种检查和可能的治疗方案以及各方案的利弊等情况，患者则如实提供自身的病情、生活方式以及意愿和价值取向等信息。然后医患双方再对相关的诊治方案进行讨论，并结合实际情况（如社会、家庭、医院技术条件等因素）以

寻求一个最优的诊治方案。这也是循证医学所倡导的最佳临床实践模式。在这种模式下，医患关系融洽，患者的意见得到充分的尊重。医患双方都是决策者，缺少任何一方，都是不完整决策。

二、临床决策的具体分型

（一）确定型决策

确定型决策是指事件的结局已经完全确定的情况下，只需通过分析各种方案的最后得失，作出选择。决策方法有增量分析法，即通过计算方案之间的增量收益与增量成本进行决策的方法。

（二）风险型决策

风险型决策是指事件发生何种结局不能肯定，但其概率可以估计，决策带有一定的风险性。风险型决策有3个前提条件：即两种以上的结局、自然状态下事件的概率能够估计、不同结局的利弊可以估算。方法包括期望值决策法、Bayes决策法。

（三）不确定型决策

是指事件的结局及其事件发生概率均不确定的情况下进行地决策。决策时可参考一些准则，如乐观准则、悲观准则、后悔值准则等，进行定性判断。

三、临床决策过程

（一）明确需决策的问题

首先要考虑，面临的是什么决策问题？要达到什么目标？有哪些备择方案？存在哪些影响因素？需要收集哪方面的信息，等等。总之，对涉及该问题的方方面面要全面把握，为最终正确解决问题打下良好的基础。

（二）决策问题的组织和构建

在选题的基础上，分析问题的组成要素，理清要素之间的关联，预测事态发展的趋势，通过一些直观的方法把各种方案实施的进程、路径、结局等形象地表现出来，如常用的决策树法。

（三）系统全面搜集信息

决策的前提是掌握足够的决策信息，这是决策中最为核心的一步。应系统全面搜集包括各种可能事件的概率、结局的量化指标（生存率、生存质量、成本等）等在内的决策信息。

（四）决策分析是解决问题的阶段

利用一些公式或模型进行分析和运算，得到各种方案的期望值供决策者参考。

（五）敏感性分析

鉴于各种事件的发生概率或结局效用值，有时难以确定，可以选用不同的概率或效用值重新进行分析和运算，分析当前决策结果的稳健性。

第二节　决策分析的实施

决策分析的基本思想就是将问题系统地分成几部分，确认每个部分之中的不确定因素，然后检索医学文献或者咨询专家意见得到各不确定因素的估计值，进而得到采取不同决策选项后，估计各种可能结局出现的概率。具体实施时，决策分析先要确认决策的问题并界定问题的范围；进而采用决策树（decision tree）的方法，将问题结构化；随后要系统搜集相关信息，明确各种事件出现的概率；同时对最终结局用适宜的效用值完成赋值；最后进行决策树分析，计算每种备选方案的期望值。鉴于临床实践中各种医学现象的复杂多变性，有时需要对结论进行敏感性分析。具体步骤如下：

一、决策问题的确认

决策分析的第一步是对所研究的问题给出精确的定义，并且根据问题的定义通常将其分解为三

个组成部分：第一部分一般是不同的决策备选方案，第二部分是不同决策方案实施后所有可能发生的事件，最后一部分是决策所关心的结局。

例 14-1　一名 60 岁男性患者，职业是卡车司机，吸烟、肥胖、有心脏病风险因素（包括高血压和高胆固醇）。因患有严重颈部疼痛，他希望医生给他开 COX-2 NSAID（非甾体抗炎药）治疗颈部疼痛（他听说这种药物适用于颈肩部疼痛），但被拒绝，这是因为最新的证据表明 COX-2 NSAID 能够引发严重心脏不良事件。医生建议采用颈部按摩疗法，而患者又不接受，因为他最近读到一个关于颈部按摩后中风瘫痪的报道，中风瘫痪的风险是他最担心的，此外，药物治疗更方便。医生指出，颈部按摩导致中风的可能性非常小，心脏病发作的危害同样严重。医患双方均陷入两难的困境：到底是颈部按摩治疗（治疗中可能发生中风瘫痪），还是 COX-2 NSAID 治疗（服药中可能发生心脏病）。

二、决策问题的结构化

绝大多数决策分析采用的是决策树的方法对问题进行结构化处理。决策树按逻辑、时序把决策问题中的备择方案以及相应结局有机组织起来并用图形展示，如同一棵从左至右不断分枝的树，包括一些结点与分枝：决策结点（decision node），用小方框“□”表示，由此结点发出的方案要求决策者从中作出选择，由决策结点发出的分支叫决策枝；机遇结点（chance node），用小圆圈“○”表示，由此结点发出的事件不受人的意志所控制，是随机的，但其概率可以估计，由它引出的分支叫机遇枝或概率枝。机遇节点的期望值为此节点下各分枝概率（P）与结局值乘积之和。在决策树末梢，为各方案的最后结局（outcome），各种结局必须定量描述。在每一个机会节点，其后相应事件的概率之后必须为 1，也就是说，每个机会节点之后的事件必须涵盖所有可能的情况，只有这样才能保证分析的有效性。对于结局而言，结局可以是生存或者死亡，也可以是其他治疗可能带来的任何收益或风险。

基于上述原则，例 14-1 可以通过绘制决策树加以解决。先绘制颈部按摩决策树（图 14-1）。

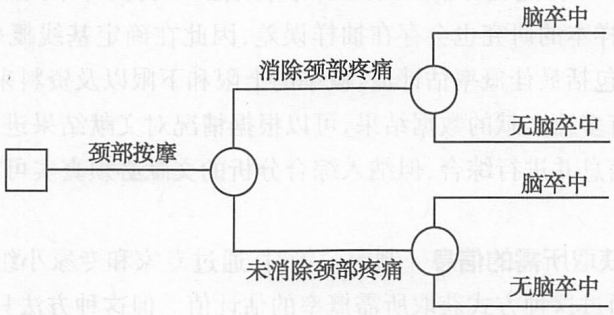


图 14-1　颈部按摩决策树

如果选择 COX-2 NSAID 治疗，该药物虽然有益于解决问题，消除颈部疼痛，但是却有服药期间发生心肌梗死的风险，选择 COX-2 NSAID 的决策树为（图 14-2）：

合并上述两种选择的决策树（图 14-3）：

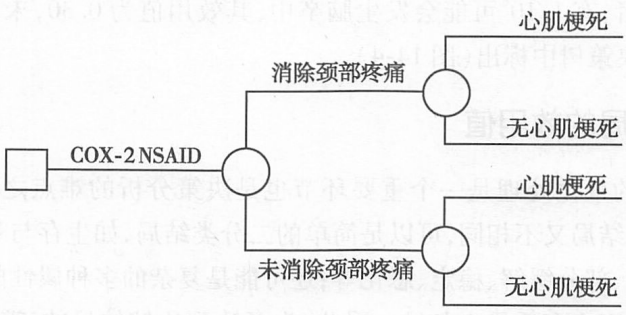


图 14-2　COX-2 NSAID 治疗决策树

NOTE

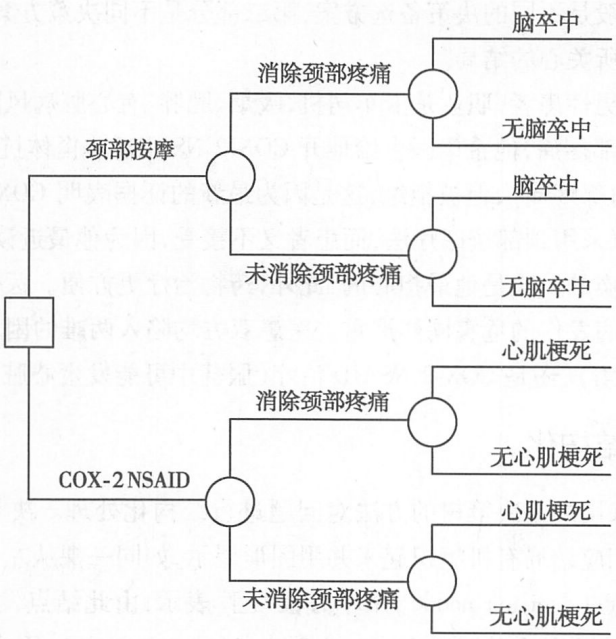


图 14-3　颈部按摩与 COX-2 NSAID 治疗的决策树

至此，决策树已清楚展示所要决策的问题，下一步就是收集信息并进行分析。

三、相关证据及信息的检索与收集

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

1. 借助文献估计概率　首先应评估所用文献的有效性，来自大样本的临床试验或随访研究的概率最为可靠，但即使是大样本的研究也会存在抽样误差，因此在确定基线概率时，应根据情况确定概率的可信区间，绘制一个包括最佳概率估计值、概率的上限和下限以及资料来源的表格。

如果某个决策问题有多篇文献的数据结果，可以根据情况对文献结果进行综合。推荐使用 meta-分析从多篇文献中获取信息并进行综合，但纳入综合分析的文献必须真实可靠，才能提供比较接近真实情况的分析结果。

2. 咨询专家意见以获取所需的信息　制定政策时，通过专家和专家小组讨论可以获取所需的信息。在决策分析中也可通过这种方式获取所需概率的估计值。但这种方法只能在其他获取信息的方法都不可行的前提下使用，而且任何通过这种方法获得的决策分析结果都必须进行敏感性分析。

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

例 14-1 中决策所需的基线估计值，假设如下：颈部按摩可有 80% 的机会消除颈部疼痛，其中有 $1/10^6$ 的患者可能会发生脑卒中，其效用值为 0.45，未发生脑卒中患者其效应值为 1.00；而选择颈部按摩但未消除颈部疼痛的患者，有 $1/10^6$ 可能会发生脑卒中，其效用值为 0.30，未发生脑卒中的效用值为 0.80。将这些估计值在决策树中标出（图 14-4）。

四、确定最终结局的效用值

决策分析中对结局的量化处理是一个重要环节也是决策分析的难点之一。因为备选方案有很多，每种方案对应的临床结局又不相同，可以是简单的二分类结局，如生存与死亡；也可以是单一属性的多项结局，如完全缓解、部分缓解、稳定、恶化等；还可能是复杂的多种属性的多种结局，如一方面是生存时间的不同，另一方面生存质量也各异。因此，为了达到比较的目的，需要预先将各种结局转换成同一度量衡单位的变量，这个变量就是效用（utility）。而效用值是一种表述结局相对优劣的数量化

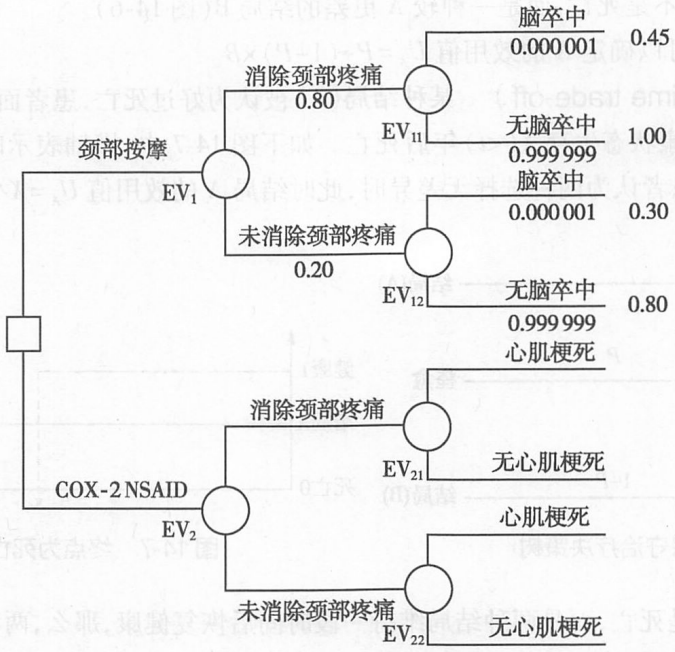


图 14-4　颈部按摩与 COX-2 NSAID 治疗的决策树

指标，通过对患者或公众进行调查获得，可测量出人们对不同健康状态的喜好程度，它是个体在不确定情况下的优先选择，表现出他们对某种健康状态的倾向和偏好，反映了个体的主观感受，并受年龄、经济收入、教育程度等多种因素的影响。效用值通常用 0 ~ 1 的数值来表示，1 代表完全健康，0 代表死亡，也可以为负数，表示比死亡更糟糕的疾病状态，如无意识或长期卧床伴严重疼痛等。

效用值有直接、间接两种测量方法，具体介绍如下：

（一）直接测量方法

效用值的直接测量方法有三种：划线法，标准博弈法，时间权衡法。下面举例说明：某男性患者，68 岁，有糖尿病病史多年，现因意外出现左脚刺伤，出现严重感染并可能出现坏疽。该患者面临两种选择，一是截肢，以保全性命，但失去左脚；二是保守治疗，但是存在感染和坏疽扩散的危险，从而有两种可能结局：痊愈或死亡。

1. 划线法（visual analog scale）　如定义死亡的效用值为 0，健康生存（痊愈）为 1，介于 0 与 1 之间的则代表各种不同状态的效用值，划一条直线如长 10cm，如下，让患者在认为可以接受的地方标记，表示可以接受某结局所对应的效用值。假如失去左脚时，该糖尿病患者在直线上画出相应的效应值（相当于正常生活能力的质量）为 0.8，倘若能以此状态生存 10 年，那么他的质量调整寿命年为 10×0.8 = 8QALYs。

0 ——————————————— 1.0

2. 标准博弈法（Standard gamble）　效用值还可以借用决策树表示（图 14-5），若告诉患者，保守治疗的痊愈可能性 $P = 0.9$ 时，患者可能会毫不犹豫的选择保守治疗，此时若改变 P 值，当 $P = 0.6$ 时，患者可能要选择截肢。不断改变 P 值询问患者，当达到某一数值如 $P = 0.8$ 时，患者认为两种方案差不多，此时的 P 值称为均衡概率（breakeven probability），其数值就是结局 A 的效用值，即 $U_A = 0.8$，意思是，如把痊愈视为 1，死亡为 0，患者把失去左脚认为相当于

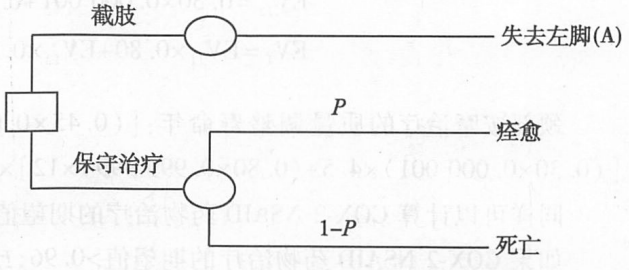


图 14-5　截肢的决策树

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（图 14-6）。

同样用上述方法可以确定 A 的效用值 $U_A=P+(1-P)\times B$。

3. 时间权衡法（time trade-off）　某种结局（A）被认为好过死亡，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。如下图 14-7 中，横轴表示时间，纵轴表示健康状态；不断变化 X，直到患者认为两种选择无差异时，此时结局 A 的效用值 $U_A=X/t$。

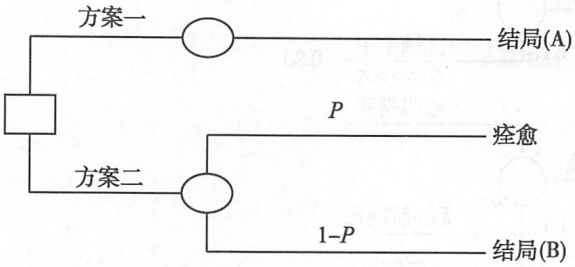


图 14-6　保守治疗决策树

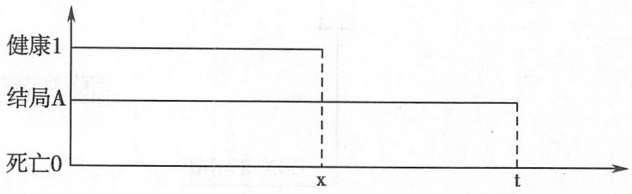


图 14-7　终点为死亡时的权衡

如果患者最后不是死亡，而是两种结局维持一段时间后恢复健康，那么，两种选择是：①暂时结局 A 延续 t 年后恢复健康；②暂时结局 B 延续 X 年后恢复健康。如下图 14-8 中，同样不断变化 X，直到患者认为两种选择无差异时，此时结局 A 的效用值 $U_A=1-(1-B)X/t$。

（二）间接测量方法

除了上述直接测量效用值外，还可用间接法测量。所谓间接法是通过间接方式测得某种疾病状态的效用值，如利用生存质量量表等，计算量表得分，再将量表得分转换为效用值，即效用值=（量表实际得分-量表的最低得分值）/量表的得分范围。此外，还可以计算质量调整寿命年（quality-adjusted life years，QALYs），用 QALYs 来测量各种各样结局，使之有通约性，简化复杂的临床决策问题。

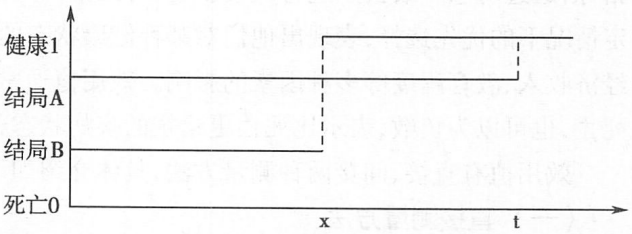


图 14-8　终点为健康时的权衡

例 14-1 应用了 QALYs 来衡量结果，设定患者无论是“颈部疼痛状态”下生存（无脑卒中）还是有病生存（伴脑卒中状态），期望寿命都是 15 年。上述四种状态的 QALYs 分别为 6.75 年、15 年、4.5 年和 12 年。获得这些资料之后就可以进行决策树分析了。

五、决策树分析

决策树分析也称决策树的折叠与平均，即指通过上述步骤所形成的决策树，计算每一种决策选项所对应的不同结局出现的情况。

例 14-1 中，根据上面的决策树可以计算颈部按摩治疗的期望值（EV_1）如下：

$$EV_{11}=0.45\times0.000\ 001+1.0\times0.999\ 999=0.999\ 999\ 45$$

$$EV_{12}=0.30\times0.000\ 001+0.80\times0.999\ 999=0.799\ 999\ 5$$

$$EV_1=EV_{11}\times0.80+EV_{12}\times0.20=0.96$$

颈部按摩治疗的质量调整寿命年：$[(0.45\times0.000\ 001)\times6.75+(1.0\times0.999\ 999)\times15]\times0.80+[(0.30\times0.000\ 001)\times4.5+(0.80\times0.999\ 999)\times12]\times0.20=25.92$QALYs

同样可以计算 COX-2 NSAID 药物治疗的期望值 EV_2 和期望质量调整寿命年。

如果 COX-2 NSAID 药物治疗的期望值>0.96，或者质量调整寿命年>25.92QALYs，说明药物治疗优于颈部按摩。倘若两个方案的期望值或质量调整寿命年相差不大，医生恐难以下定论。解决的方

笔记

法是进行敏感性分析(见本章第四节)以及选择合理的决策模式(见第五节)。

决策树模型一般用于近期效果的决策分析。如上例就仅分析了颈部按摩以及 COX-2 NSAID 治疗期间的短期效果。若要探讨各种干预措施的远期效果,又将如何进行决策分析呢?如本例中即使药物短期治愈的患者将来仍会有一部分人再次发生心肌梗死、死亡,颈部按摩治疗方案同样如此,评价这两种方案的远期效果,可选用 Markov 模型。

第三节　Markov 模型

传统的决策树方法需要固定分析期限,以便计算在此期间内各事件概率和效用的平均值。事实上疾病发展是一个动态过程,在疾病发展的不同阶段,各事件发生的概率和结局的效用值是随时间的变化而变化的。如慢性疾病的发展通常可分无症状阶段、轻度不适阶段、中度不适阶段、重症阶段。如果能够从疾病的发展阶段入手并结合各阶段转换的概率进行风险和效益的评估,其结果则更接近临床实际,也更有说服力。Markov 模型就是解决方法之一。

一、Markov 模型概述

Markov 模型的原理是将所研究的疾病按其对患者健康的影响程度划分为几个不同的健康状态(Markov state),并结合各个状态在一定时间内相互转换的概率,以及每种状态下的资源消耗以及健康结果,通过循环运算,估计出疾病的结局以及医疗成本。各个 Markov 状态通常按照临床规律进行彼此间的"互动"转换。Markov 模型多用于:临床干预措施的评价,临床试验结果的外推,药物经济学评价和疾病筛查措施的评价等。

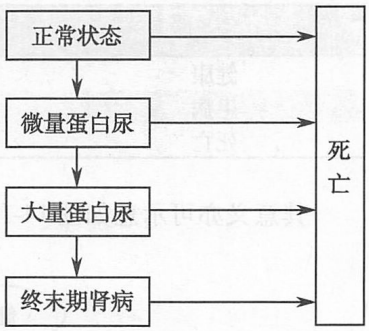


图 14-9　肾病综合征的 Markov 状态转移模型

根据研究目的和疾病的转归,将疾病的整个自然过程划分为不同的健康状态,即 Markov 状态。各种状态在一次转移后至下一次进行类似转移的相同时间间隔,称为 Markov 周期(Markov cycle),亦称转移周期,常根据不同疾病情况和临床意义而作具体的设定。如脑卒中、花粉症等发生的频率多以年为单位计算,以 1 年为循环周期;如果文献提供的是每年的发病率、死亡率或致残率等,则循环周期设置为 1 年。

以肾病综合征为例,可分为以下五个阶段:正常状态,微量蛋白尿期,大量蛋白尿期,终末期肾病,死亡。如图 14-9 所示。

对疾病进程的风险描述通常是按单位时间(一般是一个 Markov 周期)进行的。如:处于微量蛋白尿期的患者每年有 5% 进展到大量蛋白尿阶段,循环周期为 1 年。

二、Markov 模型的分析步骤

与决策树模型类似,Markov 模型也可以分为以下几步:

(一) Markov 状态的设定

将临床问题进行分解,根据研究目的确定各种疾病的状态,从而设定相关的 Markov 状态。对各种疾病状态的逻辑表现形式,可以借助 Markov 树(Markov tree)来表达。各种状态之间的转换必须符合逻辑和临床实际。

(二) 基础信息的搜集

即各种状态之间转换概率以及循环周期的确定。主要是从文献中获得,通常用矩阵按照事件发生的逻辑顺序列出来,如上述的肾病综合征的几个状态:正常状态(状态 1),微量蛋白尿阶段(状态 2),大量蛋白尿阶段(状态 3),终末期肾病(状态 4),死亡(状态 5)。

NOTE

表 14-1　各种疾病状态 1 年的转移概率

	正常状态	微量蛋白尿阶段	大量蛋白尿阶段	终末期肾病	死亡
正常状态	0.97	0.02			0.01
微量蛋白尿阶段		0.94	0.04		0.02
大量蛋白尿阶段			0.90	0.05	0.05
终末期肾病				0.80	0.20

表 14-1 中，假设肾病综合征自然状态从第 1 年开始，一个健康人群中（状态 1），每年有 2% 的人进入状态 2，1% 的人进入状态 5；而处于状态 2 的人中每年有 4% 的人进入状态 3，2% 的进入状态 5；处于状态 3 的人群有 5% 的进入状态 4，5% 的进入状态 5；处于状态 4 的每年有 20% 的人进入状态 5（死亡）。

（三）各种状态效用值的确定

如例 14-2。

（四）Markov 模型的综合分析

如例 14-2。

（五）敏感性分析

同决策树模型一样，Markov 模型也应进行敏感性分析，以判断分析结果的稳健性。

例 14-2　假设某种疾病只有 3 个状态：健康（状态 1）、患病（状态 2）、死亡（状态 3），其转移概率列于表 14-2

表 14-2　3 种状态 1 年的转移概率

	健康	患病	死亡
健康	0.75	0.20	0.05
患病	0	0.70	0.30
死亡	0	0	1

其意义亦可示意如图 14-10：

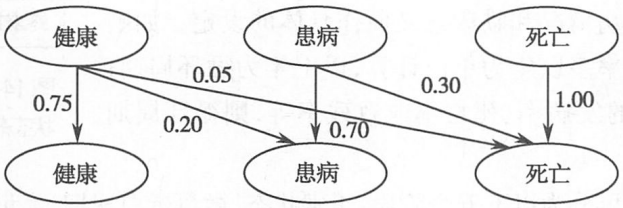


图 14-10　三种状态的转移关系及概率

设定 3 种状态的效用值分别如下：健康为 1，患病状态为 0.5，死亡为 0。构建 Markov 树如图 14-11：

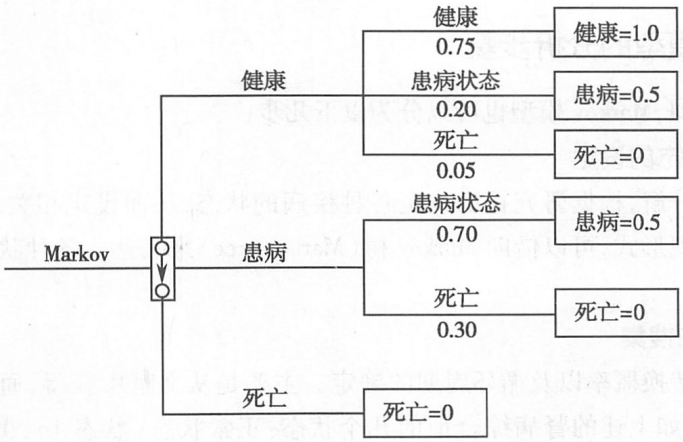


图 14-11　三种状态的 Markov 树

假设有1000名符合相关条件的健康人群，从某一时间点开始，依次进入这3种状态，据此计算出这一人群此后每年在这3个状态中的分布情况，进而计算每一循环周期或者是至研究结束时所有周期累计的资源耗费或各种结局的具体情况。

首先假设t为周期（年），a_t表示第t年（周期）时的“健康”人数，b_t表示第t年时的处于患病状态的人数，c_t表示第t年时总的死亡人数。计算如下：$a_0=1000$，$b_0=0$，$c_0=0$；$a_1=0.75\times a_0$，$b_1=0.20\times a_0+0.70\times b_0$，$c_1=c_0+0.30\times b_0+0.05\times a_0$；依次类推：$a_t=0.75\times a_0$，$b_t=0.20\times a_{(t-1)}+0.70\times b_{(t-1)}$，$c_t=c_{(t-1)}+0.30\times b_{(t-1)}+0.05\times a_{(t-1)}$，如表14-3所示：

表14-3　Markov模型模拟的计算结果

周期（t）	健康（a_t）	患病状态（b_t）	死亡（c_t）
0	1000	0	0
1	750	200	50
2	560	290	150
3	420	315	265
4	316	305	379
5	237	277	486
6	178	241	581
7	133	205	662
8	100	170	530
9	75	139	786
10	56	113	831
20	3	9	988
25	1	3	996
28	0	0	1000

进一步计算每一个循环末的结局，如第1个周期末的总质量调整寿命年（QALY）为：750×1+200×0.5+50×0=850QALYs，依次计算可以获得该模型模拟28年的总质量调整寿命年为5332.7QALYs，而人均质量调整寿命年为5.33QALYs。这是自然状态下未加干预手段的结果，如果对该人群施加一定的临床干预，则必然影响人群在不同状态下的分布，就可计算在该干预措施下的人均QALYs，如果有多个干预措施，同样可以计算不同措施下的QALYs，从而可以比较不同干预措施的效果。

临床干预措施既可影响患者在不同疾病状态的分布，也可影响各状态间的转换概率，此时可用Markov模型评价不同干预措施下的期望值，据此选择适当的干预手段，同时也可结合资源消耗估计，进行成本-效益分析等。

例14-3　某50岁男性患者，体检发现左颈动脉硬化，但目前没有任何临床症状。对于该患者的治疗，若不处理则有较高的发生脑卒中的风险，鉴于目前无任何症状，临床上可以有两种选择，一是暂时临床观察，二是行颈动脉内膜切除术。若只进行临床观察，虽可避免手术相关的短期危险因素（围手术期死亡，手术中发生脑卒中），维持无症状性颈动脉硬化状态（在一段特定时间内不发生脑卒中），但面临将来发生脑卒中的高风险。假设本例的基础估计值如下：选择临床观察的患者有11%的可能发生脑卒中，而选择做手术者有1%的围手术期死亡，以及有2%发生围手术期脑卒中，术后“不发生脑卒中”者将可能有5%再次发生脑卒中。

如图14-12所示，术后患者有三种结局，即维持无症状性颈动脉硬化状态、发生脑卒中和死亡。同时随着时间的推移，无症状性颈动脉硬化状态的患者将会有一部分人死亡，一部分发生脑卒中，余下的人维持无症状性颈动脉硬化状态。对于带病生存（伴有脑卒中）者，也会随着时间推移，部分人会死亡，而生存者继续这种循环，直至所有人都死亡。

用Markov模型来模拟他们的结局，并同其他的方案进行比较后作出选择。本例有三个Markov状

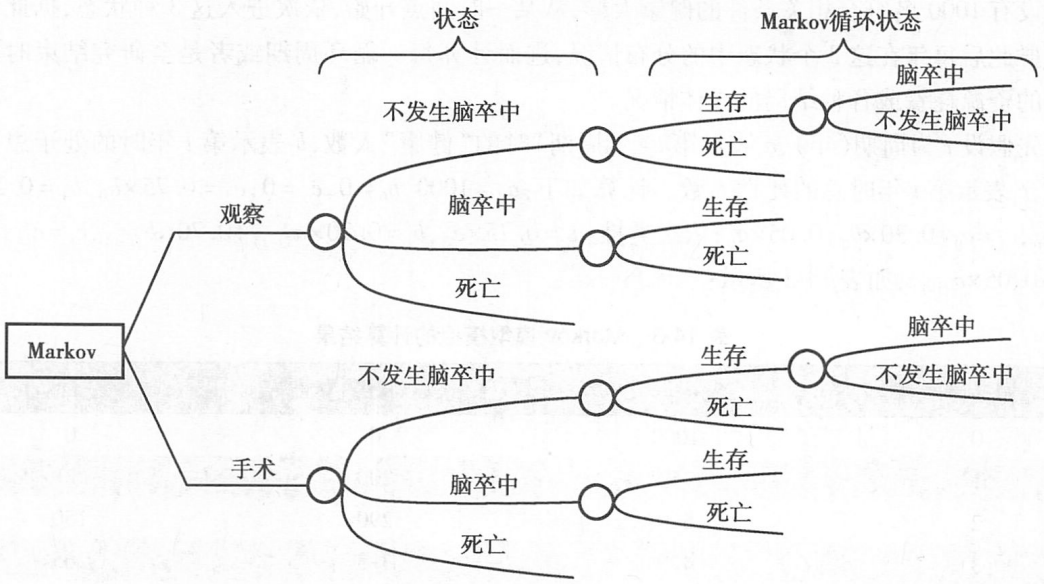


图 14-12　颈动脉硬化患者的 Markov 模型示意图（Markov 树）

态：无症状性颈动脉硬化（状态 1），伴有脑卒中生存（状态 2），死亡（状态 3）。其中状态 1 到状态 3 的死亡率为 5%，并不完全归结于颈动脉硬化这一疾病所致（要尽可能模拟实际情况，考虑有其他死亡原因）。它们之间的转换关系及转移概率见表 14-4 和图 14-13。

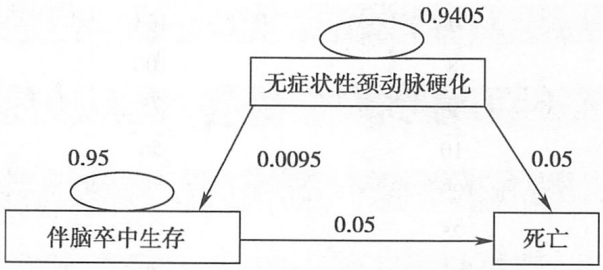


图 14-13　颈动脉硬化患者各状态间的转移关系

取 Markov 循环周期为 1 年，那么每一个周期，“无症状性颈动脉硬化状态”为 1 个 QALY，伴脑卒中生存则为 0.5 个 QALY，死亡则为 0。假设从手术当年起，维持无症状性颈动脉硬化状态占 97%，伴脑卒中生存的占 2%，死亡为 1%。此后人群在这 3 个状态的转移概率如下表（表 14-4）。

表 14-4　3 种自然状态每年的转移概率

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

计算原理同上，鉴于这一过程的计算比较麻烦，可以借助相关软件进行，如 decision maker，treeage 等，比较简单的计算也可运用 Excel，如本例计算结果见表 14-5，对该人群而言，选择手术治疗，其最终结果是人均获得 16 个 QALYs。同理，计算选择“临床观察”方案的人均 QALY，数值较大者的方案为优选方案。

表 14-5　手术决策枝的 Markov 模型计算结果

周期数	无症状性颈动脉硬化		生存（伴脑卒中）		死亡		总 QALY	累计 QALY
	例数	QALY	例数	QALY	例数	QALY		
1	97 000		2000		1000			
2	91 228	91 228	2822	1411	5950	0	92 639	92 639
3	85 800	85 800	3548	1774	10 652	0	87 574	180 213
4	80 695	80 695	4786	2093	15 119	0	82 788	263 001

续表

周期数	无症状性颈动脉硬化		生存(伴脑卒中)		死亡		总 QALY	累计 QALY
	例数	QALY	例数	QALY	例数	QALY		
5	75 893	75 893	4744	2372	19 363	0	78 265	341 266
6	71 377	71 377	5228	2614	23 395	0	73 991	415 257
7	67 130	67 130	5645	2822.5	27 225	0	69 953	485 210
8	63 136	63 136	6001	3000.5	30 863	0	99 137	551 347
9	59 379	59 397	6301	3150.5	34 320	0	62 530	613 887
10	55 846	55 846	6550	3275	37 604	0	59 121	672 998
11	55 523	52 523	6754	3377	40 723	0	55 900	728 898
50	0	0	0	0	10 000	0	0	1 599 600
总计：								1 599 600
人均：								16.00

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

当然，如果对来源基础数据的可靠性存在疑问，同样可以进行敏感性分析，以了解当前决策的稳健性如何。

第四节　敏感性分析

上述提到，如果认为选择做手术的死亡率(1%)过高或是过低，或者1%概率估计不准确，抑或不固定、是一个区间值，手术死亡率不同，将影响到最终的决策。此时可考虑进行敏感性分析。

当相关数值变化时，敏感性分析可用来判断决策结果的稳健性。这里就对第三节中的例14-3进行敏感性分析，当取不同的手术死亡率时，分别计算出不同的手术治疗方案的期望值，图14-14显示的就是当手术死亡率在一定范围内变动时、期望值相应变化情况。从图中可以看出，当手术死亡率为2.1%时，两种方案的期望值相当，此时的2.1%称为阈值，代表两种方案选择的分界点。当手术死亡率在小于2.1%的区间变化内，选择手术方案是合适的。

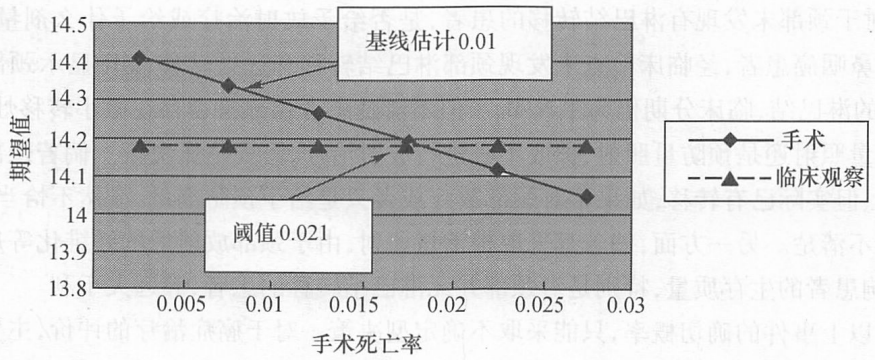


图14-14　对手术死亡率的敏感性分析

再如，对伴有脑卒中生存状态的效用值也存在疑惑，同样可以进行阈值分析和敏感性分析(图14-15)。

实际上对事件结局及其发生概率都可以进行敏感性分析。决策树模型如此，Markov模型也同样可以进行敏感性分析，计算出相关的阈值，方法同上。

NOTE

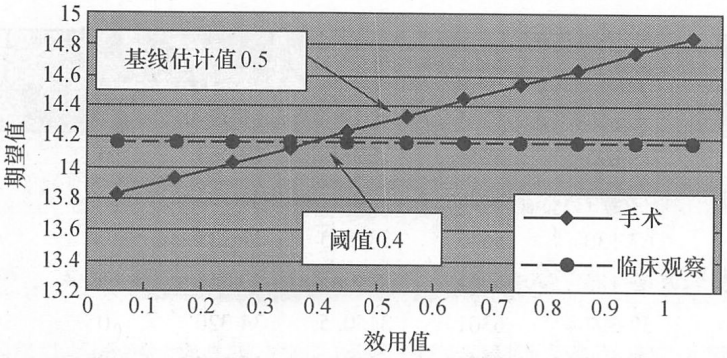


图14-15　对伴脑卒中生存效用值的敏感性分析

第五节　不确定型决策

假如临床事件结局及其发生概率明确并可量化，可选择决策树模型或Markov模型进行决策分析。但倘若有些临床事件结局及其概率不清楚且无最佳证据可循时，只能进行不确定型决策，即依赖于决策者的临床经验和主观判断进行决策。以下准则可供决策者参考：

（一）乐观准则

乐观准则也称大中取大法，即找出每种方案自然状态下的最大收益者，从中再取最大者。

（二）悲观准则

悲观准则也称小中取大法，即先找出每种方案自然状态下的最小收益者，从中再取最大者。

（三）后悔值准则

计算各方案在各种自然状态下的后悔值并列出后悔值表，找出每一方案在各种自然状态下后悔值的最大者，从中再取最小者所对应的方案，即为合理方案。后悔值法利用后悔值的计算作为最优方案选择标准。先将在不同自然状态下的最大收益值作为理想目标、再把各方案的收益值与这个最大收益值的差值，作为未达到理想目标的后悔值，然后从各方案最大后悔值中取最小者，由此确定选择方案。

例14-4　鼻咽癌以放射治疗为主，对于颈部有淋巴结转移的患者，颈部应给予根治量照射，而颈部没有淋巴结转移者，常规给予预防剂量照射，但有人也提出如鼻咽肿瘤不大，可不对颈部进行放射治疗。因此，对于颈部未发现有淋巴结转移的患者，是否给予放射治疗或给予什么剂量，目前仍有争论。假如有一鼻咽癌患者，经临床检查未发现颈部淋巴结转移和远处转移，MR显示颈部有两个最大径小于10mm的淋巴结，临床分期仍为$T_2N_0M_0$。但不清楚患者颈部是否存在微小转移灶，对该患者颈部是实施根治量照射还是预防量照射，拟或不必进行放射治疗，需要临床决策。倘若颈部检查虽未发现淋巴结转移、但实际已有转移，如果不对颈部放疗或者只是给予预防剂量，显然不恰当，但这种情况发生的概率并不清楚。另一方面，倘若都采取根治量照射，由于颈部放射后的纤维化等后遗症比较严重，将极大影响患者的生存质量，特别是对颈部并无淋巴结转移的患者，弊远大于利。

由于缺乏以上事件的确切概率，只能采取不确定型决策。对于癌症治疗的评价，主要是使患者的损失最小，其结局可用“效用损失”衡量。效用值的确定在这里需要考虑的有：①颈部淋巴结复发或远处转移的影响；②放疗后遗症对生存质量影响的程度；③治疗费用；④患者对治疗的期望值以及放射治疗的了解等情况。同时也应考虑到：不同年龄、职业、社会地位的患者，其效用值损失的估计也不尽相同。这时的决策应尽量做到个体化。

表14-6中为几种结局的效用损失的定量估计：无淋巴结转移也不对颈部进行放疗的效用损失为0，有颈部淋巴结转移而未进行放射治疗的效用损失为-100，有颈部淋巴结转移而只进行预防剂量放射治疗的效用损失为-30，无颈部转移而行预防剂量放射治疗的效用损失为-50，有颈部转移而行根治剂量放射治疗的效用损失为-20，无颈部转移而行根治剂量放射治疗的效用损失为-80。

笔记

表 14-6　鼻咽癌患者三种治疗方案的效用值

	不放疗	预防剂量放疗	根治剂量放疗
有颈部转移	-100	-30	-20
无颈部转移	0	-50	-80

由于不知道颈部淋巴结转移的概率，面对这一不确定性，依照前面提到的几种准则进行决策：

1. **乐观准则**　在每一备择方案中，找出各种结局下的最大受益值，从中选择收益值最大的方案，也就是从最大收益中取最大。本例中不放疗的最大收益为 0，预防剂量放疗为-30，根治剂量放疗为-20。按照这一原则，应该选择不放疗，这是一种相对乐观的准则。

2. **悲观准则**　与乐观者相反，找出各种结局下的最小益值，再从中选择收益值最大的方案，也就是从最小益值中取最大。本例中不放疗的最小收益为-100，预防剂量放疗为-50，根治剂量放疗为-80。最大者为-50，应选择预防剂量放疗。

3. **后悔值准则**　按照这一准则，两种自然状态下的理想值在有颈部转移时为-20，无颈部转移为 0，可以计算各状态下的后悔值，具体见表 14-7 后悔值矩阵。

表 14-7　后悔值矩阵

自然状态 / 治疗方案	有淋巴结转移	无淋巴结转移	最大后悔值（各方案）
不放疗	80[(-20)-(-100)]	0(0，理想值)	80
预防剂量放疗	10[(-20)-(-30)]	50[0-(-50)]	50
根治剂量放疗	0(-20，理想值)	80[0-(-80)]	80

可见在这一准则下应该选择预防剂量的放射治疗这一方案。

应该注意到，由于缺乏准确信息，利用上述决策准则时，往往带有很大的主观性。因此，决策准则及决策者不同，所得到的结果也不完全相同。当然，最好还是收集资料信息，尽量获得各种结局的概率，以改进不确定型决策，从而使选择的方案更趋合理。

第六节　决策分析的局限性

决策分析作为一种以证据为基础的临床应用技术，已开始逐步应用于临床实践中。但应注意，决策分析也存在一定的局限性，并不能保证每次都能作出正确的选择。以下几方面均可能影响决策结果：

首先，决策正确与否在很大程度上取决于数据是否充分、准确。对于决策分析所必需的基线数据，如事件概率和效用值，目前往往缺乏高质量的临床试验的结果，内部和外部偏倚会影响数据的准确性，由此可能造成决策失误。

其次，再复杂的决策模型也不能完全模拟临床实际，只是临床实际的简化版。患者病程进展往往复杂而变化无常，决策者必须考虑临床问题的方方面面，包括治疗方案、不良反应、成本，以及患者的个体差异、特殊要求等，决策中的影响因素之多远非一般决策树所能描绘。

最后，对决策分析的结果需慎重解释。决策分析的期望值为 0.96，并不能保证每个患者的实际效应值等于 0.96。它只是一个平均水平，即一组病情程度相似的患者，若选择颈部按摩后，其平均效应值为 0.96。因此，在临床实践中，应审视你的患者与决策分析中的研究对象之间的相似性，也可借助敏感性分析，为个体患者选择最恰当的治疗方案。

尽管如此，临床决策分析仍不失为一种有价值的临床实践工具，同时通过识别哪些变量（概率、效用值）是影响决策的敏感因素，决策分析也为下一步研究指明方向。

（李晓枫）

第十五章　循证医学自我评价

循证实践的第五个步骤是后效评价，即应用循证医学的理论和方法进行循证决策，其结果若是成功的，则可用于指导进一步实践；若不理想，则应具体分析原因，认真总结教训，并提出改进意见以指导新的循证研究和实践。后效评价的方式有自我评价和同行评价。前者是临床医生或其他卫生工作者在循证实践中对单个患者的结果进行评价，后者是通过相关专家根据统一的评价标准对群体患者的后效评价。因评估一例患者或一系列患者中证据应用的结果往往需要随访较长时间，影响因素复杂，难度大，通常不容易实现。因此常通过再评价循证实践全过程来完成自我评价。本章主要介绍针对循证自我设计、行动及效果的评价。

第一节　能 力 评 价

一、评价“提出可回答问题”的能力

提出“可回答问题”是循证医学实践的第一步，若问题提出的不恰当，循证医学的后续实施步骤均会受到影响。表15-1列出了5个有关“提出可以回答问题”能力的自我评价问题。首先是有没有提出问题？若没有提出问题，就根本谈不上实践循证医学。其次，提出的问题其陈述是否简洁明了，符合格式？有关背景问题的两个基本成分和核心问题的三到四个基本成分是否准确明了。接下来的问题是，在积累了一定的经验之后，可否进一步明确主要问题所在，明确自己还欠缺哪些知识，并对最初提出的问题进行修改？另一个问题是提出问题的过程中若遇到障碍，能否想办法有能力克服？最后一个问题是医疗实践过程中随时可能产生新的问题，有没有养成随时记录问题以待今后解决的习惯？是否将提出问题融入到每天的临床活动中？另外，随着循证医学实践活动的增加，可查看“提出可以回答问题”的成功率是否在提高，据此，可分析判断提出问题的能力是否在提升。如果提出问题的成功率很高，可以继续提出问题；如果成功率较低，可以通过咨询同行专家或参加继续教育培训来提升自己的能力。

表15-1　对“提出可以回答的问题”的自我评价

条目	具体问题
1	有没有提出临床问题
2	问题的陈述是否简洁明了，符合一定的格式
2a	有关“背景”知识的两部分问题
2b	有关诊断、治疗、管理等“核心”知识的三（或四）部分问题
3	有没有借助绘图法来明确自己还欠缺哪些知识，并对最初提出的问题进行修改
4	提出问题的过程中若遇见障碍，能否想法克服
5	有没有养成随时记录问题以待今后解决的习惯

二、评价“寻找最佳外部证据”的能力

可根据表15-2中所列条目对“寻找最佳外部证据”的能力进行自我评价。首先是有没有去寻找证据？其次，是否了解本领域内现有的最佳临床证据来源？若不知道最佳证据的来源，则很难能找到

最佳证据。同时是否能以找到最高级别证据为检索目的，并力求找到位于“证据金字塔”顶端的证据。第三点是能不能迅速寻找到临床实践所需的硬件、软件及最佳证据？从事循证实践人员、医院和相关其他机构是否将硬件、软件等资源合理配置？是否突出标记了最重要的证据来源？按照研究步骤如：确定资源、启动资源、键入问题、获得答案等进行时间分配时，可通过提高哪些环节的效率进而提高研究的整体效率？如果是独立开展研究，能否从庞杂的信息来源中寻找到最佳外部证据？研究能否更具效率？检索 MEDLINE 数据库时是否使用了医学主题词表（MeSH）、限制词以及智能检索（intelligent free text）等检索技巧？是否设置了有效的检索过滤器？

表 15-2　对“寻找最佳外部证据”的自我评价

条目	具体问题
1	有没有去寻找证据
2	是否了解本领域内现有的最佳临床证据来源
3	能否迅速寻找到临床实践所需的硬件、软件及最佳证据
4	能否从庞杂的信息来源中寻找到有用的外部证据
5	在寻找证据的实践过程中有没有逐步提高检索的效率
6	在 MEDLINE 数据库中检索时有没有使用截断、布尔逻辑算符（or 等）、MeSH 主题词、限制词及智能检索等检索技巧
7	与专业文献检索人员以及热衷提供最佳证据并受人们尊敬的同行检索结果相比，我们自己的检索结果如何

评价“寻找最佳外部证据”能力的另一个可行的方法是：请专业文献检索人员或其他专业人士对同一个问题进行重复检索，然后与我们的检索结果进行对比，比较两次检索在检索策略、检索结果、文献相关性及可用性等方面的差异。这样做有三个益处：一是能够完成自我评价；二是能从中学到更好的检索技巧；第三点是能够获得更多的、可以回答临床问题的外部证据。

如果我们在检索效率和效力方面仍有困难，还可以与附近的医学图书馆取得联系，参加他们举办的课程（如医学文献检索）或是索取有关的指南以便自我学习，达到实践循证医学所需要的文献检索能力。甚至还可以邀请文献管理员加入到我们的研究团队，更好地提高检索水平和检索效率。与循证医学有关的专业检索培训也可以有效地提高文献的检索能力。

三、评价“严格评价证据质量”的能力

对“严格评价证据质量”能力的自我评价可参照表 15-3 中的条目进行。首先是否对收集到的临床外部证据进行了质量评价？如果回答是否定的，那么原因是什么？能否找到并克服妨碍我们对临床外部证据进行质量评价的障碍？目前，在加拿大、美国等一些发达国家，临床医生可以加入到某个小组（如各种杂志俱乐部）中，不但能有助于循证医学实践，同时还能及时获得关于循证实践的反馈意见。

表 15-3　对“严格评价证据质量”的自我评价

条目	具体问题
1	是否对外部证据进行了严格评估
2	严格评估证据的指南是否易于使用
3	能否逐渐做到准确且熟练地使用某些严格评估的指标，如似然比、NNTs 等
4	有没有创建任何严格评价证据的题目（CATs），对严格评估进行总结

一旦开始评价，应考虑这些评估证据的指南是否容易应用。许多临床医生发现，随着时间的推移，他们能够从容评价大多数证据，但总有那么一两篇文献难以定论。这种情况在临床医生组成的专

业小组中也常会出现，但对于专业小组，问题提出的快，解决的速度也比较快。进一步需要考虑的是在严格评估质量时，对某些效应指标（如似然比、NNTs 等）的运用是否熟练而准确。将我们的评估结果与进行同样工作的其他专业工作组的评估结果进行比较，或从二次出版期刊所摘要的文献（6S 类证据）中提取原始数据，完成数据计算，然后将计算所得结论与摘要中的结论进行验证比较，可评价我们使用严格评估指标的能力。

最后，也是最高层次的要求，是否创建了任何严格评价证据的题目（Critically Appraised Topics，CATs）并加以总结。针对严格评价的证据进行总结评分，也可用牛津循证医学中心提供的软件 CATMaker 来完成。若用 CATMaker，比较粗略的计算结果是否与 CATMaker 软件对文献质量评估进行总结的结果一致。虽然 CATMaker 是很有用的评价工具，实际临床工作中使用较为麻烦，也可采用更为简便的方式来评价，包括研究引用、临床底线、研究方法的两行式描述以及结果总结简表等来简要记录评估结果。

四、评价“整合外部证据与患者价值”的能力

表 15-4 中列出了一些在自我评价“结合外部证据与自我临床经验”以及“将检索结果应用于临床实践”时需要考虑的因素。同样，首先需要问有没有将严格评价的证据与临床实践相结合。如果没有把已经严格评估的证据应用到实践中去，那我们在前面三个步骤的努力将是白费工夫。因此，若这一步骤没有完全实施，我们需要认真反思，时间和精力将相应调整。再次强调，与同行专家进行讨论或是参与专业小组工作将非常有助于我们克服这个“纸上谈兵”的问题。工作步入正轨之后，则需反思和总结：在将外部证据应用到个体病例的过程中我们能不能做得更加准确和高效？我们是否寻找到，或是总结出了外部证据用于临床实践的验前概率？

表 15-4　对“外部证据与临床经验的结合”的自我评价

条目	具体问题
1	是否将严格评价的证据应用到了临床实际中
2	能否逐渐做到准确且熟练地调整严格评价的指标（例如，验前概率，NNT/f 等）以适应具体的病例个案
3	能否解释（和解决）整合证据制定决策过程中出现的争议

我们是否能更加熟练地根据患者的“因素”（f）对治疗措施进行调整？能否对原有的争议进行解释（甚至是解决），是检验我们的整合能力是否有所提高的方法之一，这一评价方法尤其适用于教学医院的住院医生。

第二节　效果评价

对循证医学实践的评价不应仅仅是对知识技能的评价，更重要的还应包括对行为、态度的评价和对临床实践能力及临床结局改善的评价。

一、临床实践是否得以改善

自我评价之后，了解到自我开展循证实践的能力。在具备能力的基础上，进一步应评价是否将所具有的能力转化为开展循证实践的行动，进而使临床实践得以改善。因此，应进一步自问自己学到的知识有没有转化为更好的临床实践（表 15-5）。这方面可以借鉴在质量促进中，已被证实为有效的方法，即“计划—实施—检查—行动”循环法，来帮助我们回答能力是否转化为行动的问题。其中“计划”（plan）以提出问题、寻找证据、严格评价及证据应用等 4 个部分作为开始。同样需要考虑：运用此方法时可能会遇到哪些障碍，以及是否需要新的技能、设备、组织程序或提醒机制来防止遗忘而没有开展行动。

表 15-5　对“行为改变”的自我评价

条目	具体问题
1	当有外部证据表明要进行改变时，能否克服障碍进行相应行为调整
2	有没有进行检查，如对诊断、治疗及其他临床实践方面的审计

在拟订要开展那些工作的计划（plan）后，需要将计划付诸行动，亦就是实施（do）计划。计划试行一段时间后，可以看计划是否产生效果。然后检查（check）是否行动实施后有相应的效果变化。若试行后证明计划有效，则可推广，否则，需重新制定计划。检查包括对临床工作进行审计（临床审计为“一个质量改进的过程，旨在通过对卫生保健措施系统评价而对明确的标准实施变革，以改善患者卫生保健状况和疾病的结局”）。需要开展审计的原因有两个：一是审计结果可以显示出临床医生的绩效如何；其二，也是更重要的一点，审计通常会以多种方式进行，特别是个体反馈，这对改善临床工作会有极大的帮助。审计工作长期、持续、恰当地进行，会不断地改善临床实践。若审计结果显示临床工作有所改善，可以进一步考虑如何继续提高自身的水平。若审计显示临床工作停滞不前，除了自我批评之外，还须找到问题的根源并克服之，然后制定新的计划，进入下一循环。

二、多少临床实践有证可循

目前，许多临床工作者或专业小组研究现有的临床实践到底在多大程度上是有证可循的。实际从事临床实践的工作者也需要和有必要了解。1989 年英国著名的流行病学家 Iain Chalmers 发表了他的系统综述结果：当时在产科使用的 226 种方法中，临床试验或系统综述证明，只有 20% 是有效的且疗效大于不良反应；30% 是有害或疗效可疑；而 50% 缺乏随机试验证据。该研究结果公布后震惊整个医学界，继而引发了对循证医学的思考和关注。

1995 年由 David Sackett 领衔的审计小组，对牛津某临床机构某个月中的每一例死亡病例、出/入院病例的主要诊断（包括：疾病、症状、整体情况、存在多项诊断时导致患者就医的主诉症状）以及对患者实施的主要干预（包括：针对主要诊断所采取的治疗、操作及辅助治疗等）进行讨论，随后对干预进行追踪。审计小组基于当时的循证医学即时资源汇编及检索其他资源（如医学教科书、医学电子数据库、已发表的文献等），最后将干预措施分为三类：①已有一项或多项随机对照试验或有系统综述支持其有效性（或证明其无效）的干预；②有效性已经得到公认，审计小组认为再对其疗效进行随机对照试验是毫无必要的，且若设立安慰剂对照、不符合伦理学的干预；③已得到普遍使用，但不属于上述两种情况者。

在该月诊断的 109 名患者中，根据上述的分类，90 名（82%）患者接受的干预是有证可循的，其中 53% 患者所接受的主要干预措施得到了一项或多项随机试验或系统评价的支持，29% 患者所接受的干预措施有明确可信的、非试验性研究的证据支持。另有 18% 的患者，接受的对症治疗或支持疗法虽无有力的证据支持，但目前的干预优于其他干预或没有干预。这次审计的结果确认了住院患者所接受的干预应该是有证可循的。从此以后，世界范围内的多个临床机构都进行了类似研究，审计涉及了临床医学的不同领域，如：普外科、血液科、儿科、基础保健、麻醉科及精神科。事实上，临床中的绝大多数患者都只患有一种或两种常见病，而为常见问题寻找证据将会比为罕见问题寻找证据更有效率。上述研究还为各种常见问题累积了众多证据，从而保证开展循证医学实践的切实可行。

最后，需要强调的是：审计不仅仅是关注如何为某个临床问题寻找到高质量的证据，审计同时更是一个逐步学习的过程，它能帮助工作小组里的每一个成员进行知识更新。

第三节　案　　例

皮质类固醇产前应用于有早产风险的妇女加速胎儿肺成熟的循证实践及其效果。

早产儿是婴儿死亡的主要原因(65%),而肺脏发育不成熟致新生儿呼吸窘迫综合症是早产儿死亡的主要原因。1972 年 Liggins 等在动物研究的基础上,首次开展了皮质类固醇产前应用于有早产风险的妇女预防新生儿呼吸窘迫综合症的随机对照试验(RCT),至 1990 年又有 7 项 RCT 发表。然而,其效果不尽一致,1990 年 Crowley P,Chalmers I 等首次对产前用皮质类固醇的效果开展了系统综述和 meta-分析。

该分析表明,皮质类固醇总体上以及在所分析的所有亚组均能够降低新生儿呼吸窘迫综合症的发生。皮质类固醇可降低 30% ~50% 有早产风险的妇女新生儿死于早产综合症的概率。而新生儿呼吸窘迫综合症的降低,与脑室内出血(intraventricular haemorrhage, IVH)、坏死性小肠结肠炎(necrotizing enterocolitis NEC)和新生儿死亡的降低有关,同时没有强有力的证据表明皮质类固醇有副作用。

在该系统综述和 meta-分析发表之前,大多数妇产科医生没有意识到该治疗方法如此有效,而且该方法也没得到广泛使用。结果成千上万的早产儿可能罹患新生儿呼吸窘迫综合症且发生不必要的死亡(且需比实际需要更贵的治疗)。1991 年第一个 Cochrane 系统评价发表后,美国 NIH 发表共识,明确地推荐除少数以外,产前皮质类固醇治疗早产风险的妇女会大幅下降新生儿发病率和死亡率并可以节省大量的医疗费用。

在该系统综述发表前后收集的资料表明,综述发表时大约 20% 的有早产风险的妇女使用皮质类固醇。该综述发表后皮质类固醇的使用在 NIH 推荐之前开始稳步增长,到 1996 年使用率攀升到 65%;同期 1991—1996 年资料表明,随着皮质类固醇使用的增加,婴儿死亡率逐年降低。

(田文静)

第十六章　临床医学研究项目申报书的撰写

要做好临床研究，除具备一定的临床医学及科研方法相关知识外，临床研究思维的具体呈现形式——项目申报书的撰写也至关重要。

第一节　概　　述

一、科学研究的基本程序

科学研究应涵盖从提出问题、验证假说到得出结论等全过程。其基本程序包括：查阅文献、提出临床研究问题和理论假说、制订研究计划及研究方案、撰写项目申报书、获得资助后的项目启动与实施（观察或调查）、研究资料的整理与数据处理、统计分析、研究结果的归纳和总结、撰写研究报告以及成果转化等（图 16-1）。

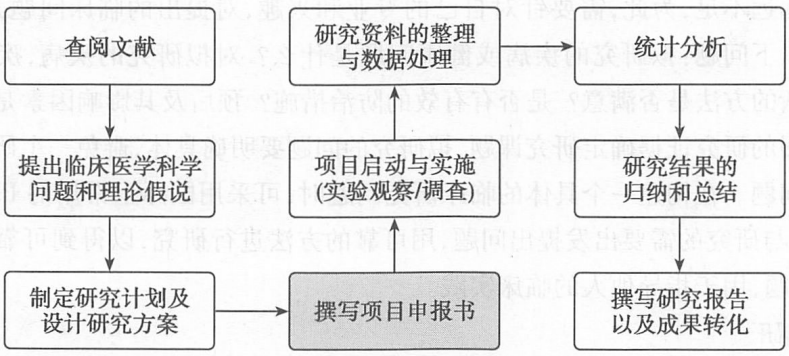


图 16-1　科学研究的基本程序

二、临床医学研究的基本类型

要撰写高质量的临床医学研究项目申报书，首先要了解临床医学科研的具体类型。按研究任务来源、科技活动类型可归纳为以下几方面：

（一）按任务来源分类

按任务来源可分为纵向科研任务、横向科研任务及自由选题项目。

1. **纵向科研项目**　是指各级政府主管部门下达的课题及项目，包括国家自然科学基金委员会设立的各类科学研究基金，政府管理部门科研基金，如科技部/卫生部的科学研究基金，以及单位科研基金。

2. **横向科研项目**　是以横向科技合同为依据的项目，主要由企业、事业单位及其他机构委托进行，研究经费一般由委托单位提供。

3. **自由选题项目**　是根据学科发展和科技人员的专长，并结合医疗卫生工作的实际需要、由研究者自己提出的研究课题。

（二）按科技活动类型分类

按科技活动类型可分为基础研究、应用研究及发展研究。

1. **基础研究** 又称基础理论研究，是以认识自然现象、探索自然规律、建立新理论为目的，未涉及或仅笼统涉及社会应用设想的研究活动。基础研究的特点是对研究手段要求高，研究结果是新观点和新信息，这类研究的未知因素多，探索性强，研究周期长，成果对广泛的科学领域产生影响，并常常说明一般的和普遍的真理，其成果成为普遍的原则、理论或定律，产生社会效益。

2. **应用研究** 是将基础研究的成果和有关知识应用到生产实践中，直接解决生产中具体实际问题的研究，也可对生产实践中的某些问题进行系统研究。应用研究主要是针对某个特定的实际目的或目标，使用基础研究获得的科学理论直接解决当前临床具有实际问题的研究。这类研究周期短，成果对科学技术领域的影响是有限的，但针对性强，产生社会效益。

3. **发展研究** 也称开发研究，是运用基础研究和应用研究的成果，推广新材料、新产品、新设计、新流程和新方法，或对之进行重大的、实质性改进的创造活动。这类研究是由科研过渡到生产的桥梁，是科研成果转化为生产力的必然途径。与基础研究和应用研究的区别在于，基础研究和应用研究是为了增加和扩大科学基础知识，而开发研究则主要是为了推广、转化和应用，其实用性和针对性更强。

三、临床医学研究项目申请的准备

（一）选题与立题

临床医疗与研究的实践中，会发现一些疾病或健康所面临的挑战性问题，而研究者对此进行解释的知识与能力却远远不足，为此，需要针对自己的专业和兴趣，对提出的临床问题开展相应的研究。研究者需要考虑以下问题：拟研究的疾病或健康问题是什么？对拟研究的疾病，疾病的病因是否清楚？早期正确诊断的方法是否满意？是否有有效的防治措施？预后及其影响因素是否清楚？在此基础上结合目前查阅的研究证据确定研究课题，拟研究的问题要明确具体，避免一个研究课题中去解决多个或多方面的问题。在构建一个具体的临床研究问题时，可采用国际上常用的 PICO 格式（见第二章）。从临床医疗与研究的需要出发提出问题，用可靠的方法进行研究，以得到可靠证据回答提出的问题，解决临床问题，用于指导他人的临床实践。

（二）文献调研

临床实践提出的问题是否能够成为研究课题，需要查阅文献。文献查阅和复习，既是课题选择的前提，也是产生或形成科学问题假设的前提。通过文献查阅，可以把握本学科的整体发展水平，跟踪相关领域新的进展，找出当前研究的热点和焦点问题，了解这些问题在国内外的研究现状；通过文献调研，还可展望本学科及相关学科的发展趋势以及未来的发展方向，进一步明确课题的立题依据，提出研究假设，帮助形成课题研究的基本思路。

进一步将拟定的研究题目与相关研究文献进行比较，精读文献，了解验证研究假设的技术方法、以及在本学科和本专业的应用现状，思考拟进行的研究可以探讨哪些临床问题？有哪些特色和创新之处？进一步构思研究方案。

（三）阅读申请指南

在申请科研项目之前，申请者应该仔细阅读申请指南，选定研究大方向和申报领域，如临床应用基础、疾病防治、新技术改进等，提出真正具有创新性、科学性，先进性、实用性的科学问题，从临床实践、基础研究以及文献调研过程中发现问题、提出问题。之后，通过相关领域文献的浏览或精读，把握本领域发展动态及最新研究成果。事先可以进行查重、查新的工作，以避免低水平重复。一般而言，临床的难点和基础的热点都可作为选题的方向，但申请者一定结合自身研究基础选题，创新立题，避免盲目跟风投机。

（四）研究基础与预试验

明确相关研究基础以及与本次选题的关系，若无现成的文献资料可供参考，也可进行预实验或小规模的预调查，以进一步理清思路、找准问题、选题立题，为撰写申报书做好铺垫。预试验是采用少量研究样本，按照项目申报书上所规定的要求进行操作，旨在发现项目申报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，然后修改项目申报书，使之更切合实际，保证科研工作顺利进行，以获得预期结果。

第二节　研究项目申报书的基本内容与撰写方法

研究项目申报是整个课题申请程序中最关键的一个步骤。申报书要充分表达出研究项目的必要性、科学性、创新性、可行性，反映出申请者及研究团队的科研作风、学术水平、科研基础以及综合能力。研究项目资助渠道不同，申报书的格式不完全相同，但内容大同小异。以国家自然科学基金为例说明申报书的内容与撰写方法。

一、项目名称及摘要

（一）申报项目名称

项目名称是对申请课题的研究方法、研究对象、研究目标的综合描述，要体现出特色、深度及创新，表达要简洁明了、语言修辞要正确、题目要有吸引力和可信度。对项目名称要字斟句酌，反复修改，使其能起到画龙点睛的作用，吸引评审者的高度关注。

（二）申报书摘要

摘要是申报书的核心部分，是申报书的高度概括，有固定的字数要求，一般在400字左右，要求所描述内容充满新意，阐述清楚条理，逻辑严密。摘要应介绍所申请项目的研究背景和现状、存在问题、立题依据，研究目的和重要意义，研究方法与主要技术路线，以及预期研究结果等。项目摘要要包含本项目的关键词，注意摘要与项目名称和正文的连贯性和一致性。

撰写摘要时，可参考如下撰写格式、填写相关内容："……是……防治的瓶颈，但其机制尚未明了。本课题拟采用……手段（模型、技术、方法），研究……机制（环节），特别是……的规律（作用）。应用……，进一步探讨……的新方法（技术）。为……提供科学的理论依据和安全有效地……方法，必将具有广阔的应用前景"。

二、立项依据

项目的立项依据是判定研究设计是否具有创新性、科学性、可行性、实用性、效能性的基础。该部分需要具体阐述研究意义、国内外研究现状及发展动态，结合科学研究发展趋势来论述科学意义或应用前景。同时应附主要参考文献目录。

立项依据一般可按以下流程撰写。

（一）介绍申报项目提出的重要科学问题

本项目所关注的一定是国内外尚未解决或者亟待解决的重要科学问题，以及临床实践中发现的、具有重大意义的研究问题。特别要突出研究对国民经济和社会发展的重要性、迫切性。

（二）描述国内外相关问题的研究现状及发展动态

重点阐述国内外相关问题的研究现状，特别是与本课题研究思路、拟采用的技术方法相似的研究现状，注意提炼、归纳、总结出学科研究中存在问题、热点和难点；同时还要对国内外研究的发展动态进行分析，这些都是形成立题思路和提出立题点的关键，用以充分说明这些问题是亟待解决、需深入探讨的重要科学问题。

（三）申报项目的假设和研究思路

提出本次申报项目的理论设想或假设，介绍拟通过什么研究思路、采用何种方法和技术手段、重点解决哪些问题。

（四）参考文献引用的注意要点

申报书的撰写，特别是对其创新性、科学性和应用价值等的阐述和论证，要有充分的依据，这些都需要有可靠权威的文献材料来支撑（包括引用相关文献中的论题、观点、概念、理论、方法、结果、结论或重要事实、数据等）。通过追踪和引述本研究领域近期发表的相关文献，阐明本项目的前沿性或新颖性，提出创新点。近3年的文献量应该占2/3左右，最好引用一定比例的当年参考文献，以展示申请者知晓国内外最新研究进展和发展动态；同时注意引用国内外重要的、高质量研究文献，以及引用与自己研究高度相关的参考文献，以论证或说明所申报项目的科学性和重要性。

三、研究目标、研究内容及拟解决的关键科学问题

（一）研究目标

指研究拟要达到的目的（最终目标），是对题目、研究对象、研究方法、成果和应用的高度概括和具体化。将研究内容中阶段性成果简单罗列在一起不能算作研究目标。撰写研究目标，可参考如下格式："通过……（技术和手段），解决……的问题、证实（验证、阐明）……的科学假设（假说），……环节？解决（揭示）……的关键科学（技术）问题，预期达到……目的"。

（二）研究内容

研究内容是对研究目标的深层次诠释，或者说是对研究目标的具体说明。撰写"研究内容"时，要围绕研究目标逐层、逐次展开，强调为了实现研究目标，研究分几个步骤进行，拟从何处入手，重点研究哪个方面，主攻什么方向，达到什么样的预期效果以及有哪些具体的考核和衡量指标等。

研究内容要具体、全面、完整和适度，重点要突出、层次要分明、用词要严谨、研究任务要明晰。切忌内容过多、研究周期内难以完成；或内容分散，研究目标无法逐一实现。

（三）拟解决的关键科学问题

拟解决的关键科学问题要反映申请者对课题总体目标实现的深刻理解和统筹解决的能力。该部分主要涉及研究过程中的一些关键技术问题、科学手段等，这些关键问题决定了研究能否有科学新发现、实验方法学有新突破、技术有创新等。注意技术和设备上或者课题合作时可能出现的困难等，不能当作拟解决的关键科学问题。此部分撰写时，要陈述对关键问题拟采用的解决方案及备用解决方案等。

四、研究方案与可行性分析

研究方案是申报书中文字篇幅最大的部分，是在研究目标和研究内容确定之后，对项目具体实施方案的详尽描述，撰写方法可因人而异，以时间顺序为主线或以研究内容为主线描述研究方案均可，但要分大小标题，突出逻辑关系，详细地写清楚每个具体步骤，具体阐述研究方法和技术路线，对复杂的技术路线和研究方法可采用流程图或图表来说明。研究方案撰写时要注意叙述清晰详细和上下文逻辑性，过于简单和过于繁杂都不可取。下面分几个方面重点介绍：

（一）研究因素

一项研究在决定了研究目标和内容后，就应该确定合适的研究因素，即研究变量。描述所选择的变量（因变量和自变量）及其规定，因变量也叫结局变量，一般指发病、死亡、伤残、治愈、有效、无效、某个观察指标变化等事件。自变量就是影响或者决定因变量的有关因素，如疾病史、家族史、疾病类型、基础疾病、药物种类、剂量以及行为生活方式（饮食习惯、食物种类等生物因素，体力活动等行为因素），一项研究中涉及的自变量和因变量可能很多，大多能通过病历、专门设计的调查表和记录表等途径加以测量。在选择变量，特别是确定自变量时一定要充分考虑到它与因变量之间的生物学关系及

逻辑学上的合理性。

申报书应介绍研究变量的规定和说明，将这些规定书写成文，例如，疾病的诊断及预后标准等等。

（二）研究方法

研究目的不同，研究方法也不尽相同，同一项研究也可用多种研究方法。研究方法的陈述主要包括所用的研究设计类型、生物学技术与方法、统计学分析方法、干预方法、资料收集方法等，均应说明，以体现研究设计的合理性和可行性，以及能够获得真实可靠研究结果的可能性。

1. 研究设计类型　应确定在研究中用人群研究的方法，还是进行动物实验或体外研究。人群研究的设计类型主要有观察性研究和试验性研究。可用观察性研究中的现况研究方法调查某一疾病在人群的患病率，如慢性肾病的流行病学调查；用病例对照研究和队列研究进行病因学和危险因素的研究，如中青年脑梗死危险因素的病例对照研究，妊娠期高血压对低出生体重影响的队列研究等；用临床试验方法进行药物疗效的评价，如某新药治疗原发性高血压的多中心随机对照临床试验等。

2. 生物学技术与方法　介绍项目所用的细胞生物学技术、蛋白分析技术、核酸分析技术等。如用流式细胞仪测定活细胞的某种蛋白，用酶联免疫吸附试验测定血清中某种蛋白，用实时定量 PCR 的方法扩增核酸等。

3. 干预方法　临床医学研究中的干预措施至关重要，是决定研究成败的关键之一。干预措施种类繁多，干预方法各有不同，要做详尽介绍。如某种药物的疗效评价，干预措施为给予研究对象某种药物，要确定并详细介绍干预组和对照组用药的药物名称、生产厂家、剂量、疗程、给药途径、疗效和不良反应的观察方法、不良反应的处理措施等。当然，疗效评价研究中所用干预措施还可以是一种治疗方案和策略或其他干预措施的疗效评价。对照组种类很多，标准对照即给予常规药物，安慰剂对照需给予安慰剂，常规药物和安慰剂用药情况同样应做详细介绍。另外，干预措施的依从性是影响药物疗效的重要问题，提高依从性和依从性监测的方法要有合理的安排。

4. 资料收集方法　资料的收集方法是临床研究设计的重要内容。获取资料的方法一定要介绍清楚，一般有两种，一种是指按研究设计方案通过专门调查或检查检验所收集的资料，如现场询问、信访或电话访问，医学检验、体检或直接观察中获得；另一种是从个人记录（病历等医学记录）或某些部门的资料（人口普查、生命统计、死亡登记、疾病登记、电子健康档案）中获得。如临床经济学分析，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。调查表是资料收集的最主要工具之一。

绝大多数的研究都是通过调查表来收集所需的资料或信息的。在申报书后要附所用调查表，并且应介绍保证调查真实性的具体措施。

要介绍生物学材料，如研究对象的血、尿、便或组织标本等的采集方法，储存条件，以及用采集的生物学标本进行何种实验室检验或其他检查，以及所用仪器设备的型号、生产厂家等。

5. 统计学分析方法　申报书中要介绍数据分析的方法，要列出分析的主要项目及指标、所应用的统计学方法，如治疗组与对照组血压值的比较方法，有效率及治愈率的比较方法，多因素模型的建立方法，对建模因素的选择、混杂因素的处理以及因素间的交互作用分析等。要介绍统计学分析方法的名称，如卡方检验、相对危险度及其 95% 可信区间、Logistic 回归、Cox 比例风险模型等等，具体何种研究采用何种分析方法，请参见有关医学统计学专著。

（三）研究对象及样本大小

根据研究目的、研究内容、研究方法选择研究对象，介绍选择研究对象的依据，研究对象来源、纳入及排除标准、随机抽样和分组的方法、样本大小计算的依据和方法等等。

（四）技术路线

技术路线是项目实施的流程图，是针对研究目标、研究内容、研究方法等所制定的合理可行的路径。用图文混排的形式表示，可能更为清晰、简单，一目了然。撰写时一定要与项目的研究目的、研究内容、研究目标等环节相互衔接，达到研究目标明确、研究内容充实、研究范围合理、实验设计较为完

善、实施过程可行的目的。

（五）可行性分析

可行性分析主要考察课题组是否有能力和条件完成所申请的项目，即已经具备的人员、设备、技术或者方法学上的能力和保证，内容包括：

1. 条件及资源　已具备项目所需要的实验环境条件、设备与物资，以及可以从医院或社区收集到的研究所需的病例及其他研究对象，所用的实验动物等。

2. 工作积累　课题组前期的相关工作积累已经为所申请项目创造了条件。

3. 研究团队能力　人员结构合理，如研究团队中有教授、副教授，或主任医师及副主任医师作指导，掌握关键技术的技术人员，研究团队成员的前期工作或研究背景与本次申请项目有关，能够胜任申请项目所承担的工作。具备了完成课题的能力。

（六）质量控制

一项临床医学研究的目的是为了展示临床医学中的客观现象，揭示其真实规律，任何研究者均希望将误差或偏倚控制到尽可能小，研究质量的控制是贯穿于整个研究过程中的，因此，应根据相关专业知识分析在整个研究过程中每个阶段产生误差或偏倚的可能性，如研究对象的选择过程、资料收集阶段、资料整理、资料分析等环节，制定正确的质量控制的对策或措施，有关在研究各环节可能出现的误差或偏倚的来源及其测量与控制方法，有关章节已有详细介绍，在制订质量控制对策与措施时，可根据不同的研究目的与研究方法参照有关内容，制定具体的切合实际的质量控制方法。

五、项目的特色与创新之处

特色与创新的凝练要从分析与国内外同行研究的不同点和项目自身的特点着手。应从对包括项目的立论依据、研究内容、研究方法与手段、技术路线及其实验方案上的研究与创新点进行概括、提炼并集中反映出来。主要包括理论的创新，技术或方法的创新。可以做如下描述："通过……实验、临床过程，提出……研究的假设，通过……手段和方法证明……；采用……实验方法，实验技术解决……中……关键问题；本研究的特色是从……方面入手，对……的问题进行研究，期望阐明……，揭示……机制。这些研究在国内外鲜有报告。此外，通过……的研究，为临床的应用提供……依据"。阐述项目的特色与创新要精练、到位、可信。特色与创新点不宜过多，一般2～4条较为适宜。

六、年度研究计划及预期研究结果

（一）年度研究计划

合理的划分年度计划进展，体现所制定的目标能够完成。此部分撰写的目标基本上是按照课题的研究内容、研究方法和技术路线所制定的方案进行。一般为准备或预实验预调查阶段，项目实施阶段，项目总结阶段。各阶段完成的内容，具体实施计划，以及不同阶段之间衔接要做清晰的介绍。

（二）预期研究成果

预期研究成果撰写要涵盖理论成果，应用成果等。预期研究成果与"研究特色和创新"基本上应该一致。即在某一理论方面有所建树或在某项关键技术、方法学上获得创新。

预期研究成果中应报告论文撰写的篇数和质量，包括国内外期刊、国内外学术会议论文等。也包括：拟组织或参加的国内、外重要学术交流活动、国际合作项目、交流计划、专利等。

七、研究基础与研究团队介绍

（一）研究基础

研究基础撰写时要介绍申请项目前期的工作积累。注意描述通过前期工作，所发现的与本次研究相关的问题与线索，前期工作成绩为本次申请项目奠定了基础，如：建立的某一方法学技术，建立的某种疾病预后研究的队列等。工作基础不仅仅指申报者个人的工作基础，要特别注意整个研究团队、学科或者学校的相关工作基础、人员和设备的资源。

（二）研究团队介绍

研究团队包括申请人及项目组成员。介绍内容包括申请人情况及主要成员情况。如申请人的学历、职称、从事研究和科研训练经历、主要研究方向，与本题相关的研究工作和课题（特别要说明本次申请项目是否为既往参加或主持的研究工作的延续），以及与项目有关或者相关的获奖和成果，代表性论文与级别等。主要成员一般为6～10名，年龄、职称、人员结构搭配要合理，一般有高级研究人员1～2人，中级研究人员2～3人，技术人员及研究生3～5人。各自所从事工作衔接良好。描述要切题，与此项目无关的内容要删除。

八、医学伦理学问题

医学伦理学问题是进行医学科学研究必须考虑的，应严格遵守《世界医学协会赫尔辛基宣言》（简称《赫尔辛基宣言》）和《人体生物医学研究国际道德指南》制定的涉及人体对象医学研究的道德原则，包括以人作为受试对象的生物医学研究的伦理原则和限制条件，人体生物医学研究政策和伦理标准以及已确立的伦理审查机制等。

九、经费预算

科研经费是进行科研活动的基本保障。应按照相关规定，根据“目标相关性、政策相符性、经济合理性”的基本原则，结合项目研究实际需要，认真据实编制项目预算表。支出预算项目直接费用和自筹资金。其中直接费用包括设备费、材料费、测验化验加工费、燃料动力费、差旅/会议/国际合作与交流费、出版/文献/信息传播/知识产权事务费、劳务费、专家咨询费以及其他支出。

根据科学基金项目资助方式的不同，项目预算表分为定额补助式预算表和成本补偿式预算表。重大项目和国家重大科研仪器研制项目填报成本补偿式预算表，其他各类科学基金项目填报定额补助式预算表。定额补助式预算表包括《国家自然科学基金项目直接费用预算表》和《预算说明书》。成本补偿式预算表包括《国家自然科学基金项目直接费用预算表》《预算说明书》《合作研究资金预算明细表》《设备费预算明细表》《测试化验加工费预算明细表》和《劳务费预算明细表》。

项目预算表中直接费用各科目如下：

1. 设备费　是指在项目研究过程中购置或试制专用仪器设备，对现有仪器设备进行升级改造，以及租赁外单位仪器设备而发生的费用。

对仪器设备鼓励共享、试制、租赁以及对现有仪器设备进行升级改造，原则上不得购置，确有必要购置的，应当对拟购置设备的必要性、现有同样设备的利用情况以及购置设备的开放共享方案等进行单独说明。

定额补助式项目要对单笔总额10万元（含）以上的设备费进行单独说明；成本补偿式项目要填报《设备费预算明细表》，对单笔总额10万元（含）以上的设备费进行单独说明。

2. 材料费　是指在项目研究过程中消耗的各种原材料、辅助材料、低值易耗品等的采购及运输、装卸、整理等费用。

3. 测试化验加工费　是指在项目研究过程中支付给外单位（包括依托单位内部独立经济核算单位）的检验、测试、化验及加工等费用。

成本补偿式项目要填报《测试化验加工费预算明细表》，并对单笔总额10万元（含）以上的测试化验加工费进行单独说明。

4. 燃料动力费　是指在项目研究过程中相关大型仪器设备、专用科学装置等运行发生的可以单独计量的水、电、气、燃料消耗费用等。

单独计量可以是单独装表计量，也可以根据仪器设备、科学装置等的能耗工时进行计算确定。要注意与间接费用中水、电、气、暖消耗的区别。

5. 差旅/会议/国际合作与交流费　是指在项目研究过程中开展科学实验（试验）、科学考察、业务调研、学术交流等所发生的外埠差旅费、市内交通费用；为了组织开展学术研讨、咨询以及协调项目研

究工作等活动而发生的会议费用;以及项目研究人员出国及赴港澳台、外国专家来华及港澳台专家来内地工作的费用。本科目不超过直接费用10%的,不需要提供预算测算依据。

6. 出版/文献/信息传播/知识产权事务费 是指在项目研究过程中,需要支付的出版费、资料费、专用软件购买费、文献检索费、专业通信费、专利申请及其他知识产权事务等费用。

7. 劳务费 是指在项目研究过程中支付给参与项目研究的研究生、博士后、访问学者以及项目聘用的研究人员、科研辅助人员等的劳务费用,以及项目聘用人员的社会保险补助费用。

项目聘用人员的劳务费开支标准,参照当地科学研究和技术服务业从业人员平均工资水平,根据其在项目研究中承担的工作任务确定。

成本补偿式项目要填报《劳务费预算明细表》。

8. 专家咨询费 是指在项目研究过程中支付给临时聘请的咨询专家的费用。

9. 其他支出 是指在项目研究过程中发生的除上述费用之外的其他支出。其他支出应当在申请预算时单独列示,单独核定。

所有项目支出计算应该分配合理,逐一列出,写明计算依据。仪器设备费,实验室改装费尽量少填,最好不填,不然会给人造成研究条件尚不具备的印象。

第三节 研究项目申报书对照检查提纲

一份项目申报书初稿撰写完毕,要反复审慎地审查、修改、再审查、再修改,项目组成员要多次讨论,直至认为基本没有什么纰漏后,可请相关专家学者审阅,提出批评建议,也可请从未涉足申请项目相关问题研究的老师和同学审阅,以评价所申报项目立论依据是否充分?研究目标是否明确?假设是否合理?研究内容是否围绕研究目标进行?方案合理可行否?项目的特色及创新性怎样?以下几点列出临床科研设计的检查提纲,可供参考:

1. 课题研究的意义是否阐释清楚,是否为重要领域的重要问题,是否具有重要的理论价值或应用前景。

2. 国内外目前的研究现状怎样,了解的广度和深度如何,近5年的主要研究进展是否掌握,是否介绍了目前研究存在的主要问题,以此提出所申请项目的切入点,进一步形成研究目标及思路。

3. 项目特色及创新性体现在哪方面,是否可能形成新的学说或理论,是否产生新的方法与技术,或改进及完善了以往的技术。

4. 尽可能引用国内外近3~5年参考文献。

5. 研究目标是否能够实现,是否目标过大或过于局限。

6. 研究内容范围是否合适,一般列出3~5个研究内容,要有1~2个突出的重点,把握1~2个关键问题。

7. 研究方案设计合理否,研究变量选择合理否,研究对象纳入排除标准是否明确,样本量是否足够,方法与技术是否成熟可靠,可重复性强,易于掌握。

8. 与本项目有关的工作积累有哪些,有无已发表的相关研究论文和工作基础,预实验的结果如何以及是否建立了实验模型。

9. 开展研究的条件是否具备,主要仪器和设备是否到位,病例积累是否足够等。

10. 项目组成员结构是否合理,主要成员组成是否合理,高、中级研究人员,技术人员以及学生的搭配比例如何。注意大学生创新项目的主要成员应为在校的大学生。

在了解临床医学研究的种类及研究基本程序的基础上,从临床研究与实践中提出研究问题,查阅分析目前有关研究证据,建立科学假说,按照临床医学研究项目申报书要领,把研究的各个要素组织在一起,逐一加以明确和完善,撰写出高质量的申报书,增加了研究的科学严谨性,提高了研究成功的概率,是学术水平和综合研究能力的具体体现,也是学科发展和临床医学工作者基本素质的标志。

(史晓红)

笔记

第十七章　临床经济学在临床科研中的应用与评价

从经济学的角度研究卫生保健领域资源的合理配置、有效安排、绩效评价等，可为临床科研以及循证医学实践提供有关经济效益方面的证据，从而帮助临床医生利用现有最佳证据、兼顾经济效益和价值取向，真正做到科学的循证决策。

第一节　概　　述

一、卫生经济学产生背景及发展历程

自从人类出现就有了医疗卫生活动，医疗卫生事业的发展与社会经济发展密不可分。进入20世纪中叶以来，医疗卫生事业在国民经济中的地位举足轻重。特别是在20世纪后期，由于医疗卫生保健体系变革、人口激增、疾病谱的改变等，直接推动了卫生总费用的急剧增长。如20世纪90年代美国用于医疗卫生及保健服务上的支出占美国GDP的14%～15%，预计到2018年，该比例将达到1/5。这其中，不合理医疗服务、过度医疗服务现象普遍，是造成卫生总费用急剧上升的主因。

鉴于医疗卫生服务领域存在卫生服务资源的稀缺性、而就医需求与愿望的无边际性，因此就产生了如何配置和使用资源以取得更高经济效率的问题——经济学问题。最早卫生经济学的研究文献，首推Arrow等在1963年发表经典论文“不确定性与卫生保健的福利经济学”。经过几十年的不断发展，卫生经济学理论体系与方法日臻成熟，Kielhorn. A在*The health economics handbook*一书中给出了卫生经济学（health economics）的定义，即综合应用经济学与医学中的理论和方法阐明和解决卫生及卫生服务中的现象及问题，是经济学在医疗卫生服务领域中的具体应用。旨在解决卫生资源的筹措、配置和利用，探讨医疗卫生服务的需求、定价与供给中的经济学问题以及制定卫生经济的政策与策略等。而其中，临床经济学（clinical economics）又成为卫生经济学的一个重要分支，是在卫生经济学相关理论和方法指导下，对临床使用的药物、诊治方案、仪器设备等干预措施进行经济学评价和分析，为临床决策和政策决策提供证据，进而明确最佳诊断、治疗、预防方案（安全、有效、经济），改善预后，乃至提高卫生资源的配置和利用效率等。

二、卫生经济学在临床科研中的运用与价值

运用卫生经济学评价可以解决许多临床研究与实践领域中的决策问题，如：①特定情况下的临床治疗方案选择。像肾衰竭的患者选择肾移植还是透析治疗；低位直肠癌患者的手术抉择，是低位括约肌保留切除术，还是选择传统经腹会阴切除术等。②最佳干预时机的选择。例如冠状动脉搭桥手术对象应选择中度心绞痛且伴有单支血管病变者，还是选择严重心绞痛伴有左主干支冠状动脉病变者。③选择提供医疗服务的最佳场所。例如糖尿病患者是选择医院、社区还是家中进行诊疗，门诊治疗还是住院治疗？

对上述医疗卫生服务措施的选择，应有令人信服的证据。临床经济学分析和评价就是从社会或其他特定的角度，比较不同医疗卫生措施的成本及临床效果（effects），形成经济学上的证据，以便进行循证决策。

第二节　临床经济学评价方法

经济学评价主要从经济学中的投入和产出进行分析。大体可分为两种类型：完整经济学评价以及不完整经济学评价。前者是同时从“投入”和“产出”两个方面进行综合分析，后者仅从“投入”或“产出”单个方面入手，进行分析。完整的临床经济学评价必须对两种或两种以上的干预措施同时分析其成本和临床效果，根据临床效果及其度量衡单位可分为：成本最小化分析、成本-效果分析、成本-效用分析和成本-效益分析。

一、成本和临床结果的测量与表达

（一）成本

经济学中的投入一般用成本（cost）表示，是指在从事某项生产、服务等过程中所消耗的物化劳动和活劳动的货币价值。医疗服务成本则是医院在提供医疗服务的过程中所发生的物化劳动和活劳动的货币价值总和。其中，活劳动是指医疗服务过程中消耗的医务人员的脑力和体力；物化劳动是指医疗服务过程中消耗一定的物质资料。因此，基于医院的医疗成本包括：

1. 直接成本（direct costs）　是直接提供医疗卫生服务过程中所花费的成本。包括直接医疗成本和直接非医疗成本。

（1）直接医疗成本（direct medical cost）：特指医疗卫生服务过程中用于治疗、预防、保健的成本。如住院费、药费、诊疗费、实验室化验费、大型仪器设备检查费、手术费、家庭病房费、康复费及假肢等费用。

（2）直接非医疗成本（direct nonmedical cost）：指患者因病就诊或住院所产生的非医疗服务成本，如患者的伙食费、交通费、住宿费、家庭看护费、由于疾病所要添置的衣服费用、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。

2. 间接成本（indirect costs）　间接成本又称生产力成本（productivity costs），为社会成本，指因罹患疾病而丧失的资源。包括：

（1）与病残率（morbidity）有关的成本：即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、奖金及丧失劳动生产力造成的误工产值。

（2）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，早死损失的10年工资、奖金都应作为间接成本计算。

间接成本的计算通常有一定难度，常用方法有人力资本法和意愿支付法。其中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合计。

3. 隐性成本（intangible costs）　是指因疾病或实施预防、诊断等医疗服务所引起的疼痛、忧虑、紧张等给患者和家属带来生理上和精神上的痛苦、不适合对生存质量的影响等。这部分成本也是难以估计的成本。

（二）临床效果

在完整的临床经济学评价中，临床效果的测定尤为关键。目前有关临床效果的表达方式有：效果（effectiveness）、效益（benefit）、效用（utility）等。

1. 效果（effectiveness）　效果是指医疗方案实施后所取得的一切结果，既包括好的效果，又包括不好的效果。临床经学评价中使用的效果是指有用的效果，它是由各种价值构成，具有满足人们需要的属性。效果指标包括中间指标和最终指标。中间指标通常是临床治疗期间效果的改善，如生化指标的转好、病灶的缩小以及其临床指标的恢复等。具体指标应根据各评价方案的内容与目的来决定，所选的效果指标应该是可以测量、具有客观性、灵敏性和特异性的临床指标。最终指标通常是指生存率、死亡率、生存年限等长期的效果。

2. 效益（benefit）　效益是指将有用的效果用货币的形式表现出来，即用货币值表示临床方案取得的有用结果。由于人的生命价值很难用货币形式来衡量，但实际生活中许多国家对涉及人的生命或健康的赔偿都是用钱来支付，因此效益的测量具有相当的难度，甚至涉及伦理问题。这是成本效益分析中的一个难题。

临床经济学评价中，可以将效益分为直接效益，间接效益和无形效益。

（1）直接效益：直接效益是指采用某项医疗方案之后所节省的医疗资源。如实施某项医疗项目后疾病发病率下降，患者就诊或住院次数的减少，从而减少了医院在药品、材料等方面的支出，也减少了人力、物力资源的消耗，这种比原来节省的支出或消耗就是该医疗项目的直接效益。

（2）间接效益：间接效益是指实施某项医疗方案之后所带来的其他方面的经济损失下降。如脑卒中患者进行系统康复治疗后身体功能恢复良好，减少了残疾的发生，避免和减少了患者家属陪护所导致的工资或奖金的损失或雇佣保姆给家庭带来的经济负担。

（3）无形效益：无形效益是指实施某项医疗方案之后避免或减轻了患者身体或精神上的痛苦以及康复带来的舒适和愉快等。

表 17-1　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

3. 效用（utility）　效用是经济学中最常用的概念之一，是指对于消费者通过消费或者享受闲暇等使自己的需求、欲望等得到的满足的一个度量。临床经济学中的效用是指患者或社会对特定的健康结果如特定的健康状况或通过一定时间到达的健康状况可能有的偏好（preference）或价值（value）。效用指标通常用质量调整生命年（quality-adjusted life years，QALYs）和伤残调整寿命年（disability adjusted of life years，DALYs）来表示。伤残调整生命年也被称为失能调整生命年。表 17-1 显示了不同健康状况下的效用值。

二、成本与健康获益的测算

（一）成本测算

医疗机构的成本核算是基于财务会计部门的成本核算数据资料来进行。根据核算对象不同，可以分为项目法、病种法和综合法。项目法是以医疗项目作为核算对象，定期核算每个服务项目的全部成本；病种法是按照不同的病种进行成本核算；综合法是指以各部门、科室作为成本核算对象进行成本测算的方法，也称为作业成本法或科室成本法。

1. 成本测算的内容　根据医疗服务的性质，我国通常将成本测算的内容分为以下六大类：

（1）劳务费：指支付给工作人员的的劳动报酬，包括职工的工资收入、奖金及各种福利、补贴等。

（2）公务费：包括办公费、差旅费、邮电费、公杂费等。

（3）卫生业务费：维持机构正常业务开展所消耗的费用，包括水、电、煤和设备维修、更新费等。

（4）卫生材料费：包括化学试剂、敷料、X 光材料、药品等。

（5）低值易耗品损耗费：包括注射器、玻片等。

（6）固定资产折旧及大修理基金提成：包括房屋、设备、家具、被服等各种固定资产的损耗。

2. 成本的时间调整　成本与效益具有时间效应，由于物价、医疗服务价格不断地发生变化，当某项医疗项目实施持续时间较长时，货币的时间价值也影响着成本，因此成本和效益需要随着变化不断进行调整才能反映其真实的资源消耗。贴现（discount）就是考虑货币的时间价值，把不同时间的成本/效益换算为现在时点的价值过程。贴现时使用的利率称为贴现率（discount rate），贴现后的值称为现值（present value，P_V）。

其计算公式为：

$$P_V = P_C \times (1+r)n$$

（P_V：现在的成本；P_C：过去的成本；r：贴现率；n：年限）

如果把现在成本折算成过去的成本，其计算公式为：

$$P_C = P_V \times 1/(1+r)^n$$

根据所计算成本的需要，可分别采用不同的计算公式进行计算，以保证成本具有可比性。

如某疾病2014年测算出的治疗成本为5000元，考虑到货币的时间价值，贴现率为3%，其2015年的成本为5000×(1+3%)=5150元，2016年的成本为5000×(1+3%)^2=5304.5元，2017年的成本为5000×(1+3%)^3=5463.6元。

（二）健康获益的测量

1. 效果的测算效果的测量就是产出测量 从广义上来讲效果是指相关卫生规划或卫生活动的方案实施后所取得的结果，可能是好的结果，也可能是不好的结果。以预防接种的效果评价为例：实施预防接种可以提高人群对传染性疾病的免疫率，从而降低传染病的发病率，这是该卫生规划的方案实施后取得的好结果。相反，如果在预防接种的过程中，由于某些原因而造成冷链中断或疫苗污染，由此引发不必要的传染病流行，这就是该卫生规划所造成的不好结果。

从狭义上来讲，效果是指好的有用的结果，也就是能满足人群需要，给人们带来好处或者满足感的结果。除了上述预防接种提高免疫率的例子，通过卫生规划或卫生活动的方案实施，各种健康指标、卫生问题指标和服务利用指标等的改善都属于这个范围。效果指标既可以是绝对指标，比如就诊人数的增加，早期诊断例数的增加，治愈患者数的增加等；也可以是相对指标，比如床位利用率的提高，发病率的下降，孕产妇死亡率的下降，婴儿死亡率的下降等。一般而言，在临床经济学分析中提到的效果如果没有特殊说明，都指的是狭义的效果。

2. 效益测算

（1）人力资本法（human capital method）：是将人个体当作有价值的商品资本来衡量，通过测量未来的生产潜力来体现其生产价值来进行评估。如测量医疗方案实施后患者因健康改善能够返回工作岗位从而在将来对社会的生产价值。人力资本法往往采用个人的平均收入，考虑到货币的时间价值进行贴现后进行分析。

（2）支付意愿法（willingness-to-pay method）：是一种用以测量健康改善，包括生命延长，疾病治愈，身体和精神痛苦减轻所带来的价值的方法，它是建立在健康效用理论基础上的。健康效用理论认为人的健康效用由两部分组成，一是人的健康状况，二是人的收入。人的健康状况决定了人的生命效益，所以有时我们可用人的生命效益来表示健康状况，人的生命效益包括未来的劳动力收入、非劳动力收入（包括资本收入、房产收入等）、非市场活动（如享受、感情）等。支付意愿法弥补了人力资本法对人的疼痛、悲伤等效用评价的不足。因此有学者认为支付意愿法比人力资本法在理论上更正确，在实践中更全面、更有价值。

（3）摩擦成本法（friction cost method）：摩擦成本主要指患者离开工作岗位到其他人接替其工作期间造成的生产损失或培训新人的上岗成本，也就是说疾病导致的损失取决于厂家恢复生产所花的时间。

3. 效用测量 效用结果的判断通常用取得的生命质量来表示，代表指标有生命年、质量调整生命年（QALYs）和伤残调整生命年（DALYs）。效用常用的测量方法有直接测量法和间测量接法。

（1）直接测量法：生存质量可采用等级衡量法（rating scale，RS）、标准博弈法（standard gamble，SG）和时间权衡法（time trade-off，TTO）来进行评估，最终结果以效用值的形式来体现，效用值的范围在0～1之间，效用值的大小反映出被评估者对生存质量、生命价值和失能程度的自我偏好。

笔记

（2）使用量表：除以上3种直接测量效用值的方法，还可以用量表测量法间接测量，量表包括通用量表以及疾病专用量表。通用量表主要包括：欧洲五维量表（EQ-5D）、SF-6D生存质量表、健康效用指数（Health Utility Index）、健康质量量表（Quality of Well-Being Scale）等。这些量表均由若干条目组成，提供一个总分或指数来表示生命质量，针对生活质量从多个维度，如躯体功能、生理功能、身体疼痛、精力、情感、精神健康以及社会功能等进行测量。

三、临床经济学评价的类型

（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）。最小成本分析是假定两个或更多临床医疗服务干预方案的结果相同，通过分析和比较每个干预方案的成本来进行方案的选择，以成本最小为最佳方案。

该类型适用于多种医疗措施的治疗结果相同或相似情况下，比较不同医疗措施所消耗的成本，以成本最低者作为首选。例如骨髓炎患者提前出院在门诊继续用抗生素治疗和常规住院治疗相比，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。由于该法只能比较同一种疾病且结果相同时的成本，故使用范围有限。

（二）成本-效果分析

成本-效果分析（cost-effectiveness analysis，CEA）是将成本和临床获益结合在一起考虑，主要是评价获得单位健康效益所消耗的医疗资源数量。表示为每一健康效果单元所耗费的成本（成本效果比）或每增加一个健康效果单元所消耗的增量成本（增量比）等。在比较两种不同的医疗措施时，因评价单位相同，可直接为临床决策者提供科学的依据。

成本-效果分析是完整经济评价方法中最为常用的一种，可用来防治方案的优选。例如一项成本-效果分析结果显示，对于终末期尿毒症患者，每延长一个寿命年的成本，连续腹膜透析成本为33 400美元，而血液透析的成本达到48 700美元。在效果相同（即延长一个寿命年）的情况下，连续腹膜透析成本低，成本-效果更佳，成为首选。

1. 效果评价指标　在成本-效果分析中，效果可以同时或分别使用中间替代测量指标（intermediate measures）和健康结局指标（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、血压下降程度等。后者包括病残、死亡、寿命年延长等。例如在高血压的治疗项目中，血压下降百分率为中间替代测量指标，而减少中风后死亡则是最终健康结局指标。

2. 成本效果比（cost-effectiveness ratio，C/E）　也可以称为效果成本比（ECR），是指每单位效果所花费的成本或每单位成本所产生的效果。如挽救一条生命、延长一个生命年所消耗的成本，或每消耗一单位成本所改善的血糖值等。除了少见病和个案报道外，单一的成本效果比意义不大，主要用于两个或两个以上结果指标相同的医疗方案的比较。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结肠癌的诊断价值。成本效果分析结果显示：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；诊断一例结肠癌成本，分别为2694美元和2896美元。结果还显示纤维结肠镜诊断的敏感度（80%）、特异度（95%）均高于后者（分别为57%和80%），最后选定纤维结肠镜。

3. 增量成本效果比（incremental cost-effectiveness ratio，ICER）　即每获得一个增加的效果单位所消耗的增量成本，它代表一个方案的成本效果与另一个方案比较而得到的结果。在研究中常以效果最低的方案作为参照，将其他方案与之相比而得，这一比值越低说明该方案增加一个效果单位所追加的费用越低，该方案的实施意义越大。

（三）成本-效益分析

成本-效益分析（cost-benefit analysis，CBA）是用相同的货币单位来分析比较成本与健康获益之间

的关系，成本和健康效果都用货币单位表示。成本-效益分析方法包括：

1. 净现值法（net present value，NPV）　即效益现值减去成本现值的差值。它是根据货币时间价值的原理，消除货币时间因素的影响，计算实施卫生规划方案期间，各年的效益现值之和与成本现值之和的差值。反映项目计算期内的获利能力。为了使不同时期的货币值可以比较，就要选定某一个时点作为基准点来计算不同时间的效益和成本的价值。

需要注意的是，贴现率（年利率）的选择对于净现值的计算及方案的评价结论影响很大；如果各个方案周期不同，净现值法难以直接判断方案的优劣；净现值没有考虑投入资金的利用效率。

计算公式：$$D_{NPV} = B - C = \sum_{t=1}^{n} [(B_t - C_t)/(1 + r)^t]$$

（NPV：净现值之和；B：效益；C：成本；r：贴现率；t：时间）

评价标准：单一方案，NPV>0，即效益>成本，则方案可行；NPV<0，即效益<成本，则方案不可行。多个方案，以 NPV 大的方案为优选方案。

2. 年当量净效益（net equivalent annual benefit）　将方案各年实际发生的净效益折算为每年平均的净效益值。它是净现值考虑贴现率时的年平均值。

计算公式：

$$A = CR \times NPV$$

（A：年当量净效益；NPV：净现值之和；CR：资金回收系数）

评价标准：单一方案，A>0，方案可行；A<0，方案不可行。多个方案，以 A 值大的为优选方案。

3. 内部收益率（internal rate of return，IRR）　指方案在计划期内使其净现值等于零时的贴现率。计算公式：

$$NPV = \sum_{t=0}^{n} \frac{B_t - C_t}{(1 + i)t} = 0$$

计算方法：

（1）试差法：用不同的贴现率反复试算备选方案的净现值，直至其等于零。

（2）插入法：使用两个不同贴现率试算方案净现值得到正负两个相反结果时，运用插入法计算内部收益率。

公式如下：

$$IRR = I_1 + (I_2 - I_1) \times (NPV_1 - NPV) / (NPV_1 - NPV_2)$$

（I_1、NPV_1为偏低的贴现率和相应为正的净现值；I_2、NPV_2为偏高的贴现率和相应为负的净现值）

评价标准：IRR 大于标准收益率，则方案可行；IRR 小于标准收益率，则方案不可行。

4. 效益-成本比率法（benefit-cost ratio）　它是医疗方案的效益现值总额与方案的成本现值总额之比，其计算公式为

$$B/C = \sum_{t=1}^{n} B_t/(1 + r)^t / \sum_{t=1}^{n} C_t/(1 + r)^t$$

评价标准：单一方案：B/C≥1，方案可行；B/C<1，方案不可行。多个方案：以 B/C 大的为优选方案。如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 岁以下的男女儿童，则效益/成本比为 8∶1，显然前者优于后者。

（四）成本-效用分析

成本-效用分析（cost-utility analysis，CUA）是通过比较项目投入成本和经质量调整的健康产出来衡量卫生项目或治疗措施效果的一种经济学分析方法。它是成本效果分析的一种特殊类型。特别是

笔记

在比较两个完全不同的卫生项目、效果指标又各异时，如肾移植治疗慢性肾功衰与抗高血压预防中风，因两种干预措施、干预对象不同，而且产生的临床结局也不同，此时，难以进行 CEA。倘若将其分母单位都统一为质量调整寿命年（quality-adjusted life years，QALYs），利用成本-效用分析就可以直接比较两者的优劣。如有研究结果显示肾移植项目每获得一个 QALY 花费的成本为 4710 英镑，而抗高血压预防中风为 940 英镑/QALY，显然后者更佳。

（五）临床经济学评价类型间的比较

三种常用临床经济分析方法的比较见表 17-2。

表 17-2　三种临床经济分析方法的比较

	CEA	CUA	CBA
比较方式	C/E	C/U	B-C 或（B/C）
成本单位	货币单位（元）	货币单位（元）	货币单位（元）
结果单位	健康效果自然单位	QALYs	货币单位（元）
可比较的措施数量	2 个或 2 个以上	2 个或 2 个以上	1 个或 1 个以上
可测定的目标	1 次测定 1 个	1 个以上	1 个以上
需测定的健康结果	效果测定	效用测定	健康效应转为货币
测定方法	随不同结果单位而定	标准概率法 时间交换法 等级尺度法 量表	人力资本法 支付意愿法 显示偏好法
可比性	随结果测定而变化	理论上可比	理论上可比

第三节　临床经济学评价的基本步骤

一、确定要分析的项目及要比较的方案

1. 确定研究问题时，不仅要阐明研究的方法，还要描述研究对象的人口学特征。例如一项不同抗幽门螺杆菌（Hp）方案治疗十二指肠溃疡的经济学评价研究中，应首先描述患者的基本特征，如年龄、性别、病程、溃疡大小、溃疡数目等。

2. 应写明欲比较方案的选择理由和依据。例如在上述研究中，拟比较质子泵抑制剂加阿莫西林与 H_2 受体拮抗剂加二个抗生素两种方案，应说明为什么选择这二种方案进行比较，其合理性和依据如何。

3. 详细描述所选的具体治疗方案及其在各种治疗方案中所处的地位。例如描述这两个方案中各自药物的具体剂量、用法和疗程，还应进行文献复习，比较其在众多候选治疗方案中的地位，以及是否为临床一线方案等。

二、确定经济学评价的角度

明确表明立足点（角度）是临床经济学评价的基础，这决定了在评价中有关成本和结果的定义、范围与内涵。选取何种评价角度主要取决于评价的目的，不同的评价目的、其评价角度不同。评价可从患者、医疗服务提供者、医疗费用支付方（如保险）或社会等角度进行。

三、确定经济学评价的方法

评价方法包括 CMA、CEA、CUA 和 CBA，根据研究问题及研究目的，选择适宜的评价方法，并说明选用该评价方法的理由。

NOTE

四、确定资料获取的研究方法

可选以下任一种或方法组合，如前瞻性随机对照研究、系统评价、观察性资料等，陈述理由。一般是在随机对照临床研究的基础上进行 CEA 或 CUA。

五、增值分析

卫生经济学评价时，除了比较成本/效果（效用、效益）比外，还应报告增值分析的结果，即在额外措施造成成本增加的同时，健康效果（效用、效益）相应增加的额度是多少。具体可表示为一个项目比另一个项目多付的费用，与其对应的效果（效用、效益）增幅之比，称为增值比（ΔC/ΔE、ΔC/ΔU、ΔC/ΔB）。

六、成本的确定

所有相关的成本都应确定、收集并报告。成本的测量应尽量反映机会成本的概念，包括人头费、管理费等。成本主要包括直接成本和间接成本。

七、确定结果测定的方法

如 CUA 分析时，应报告质量调整寿命年（QALYs）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步骤。

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数年完成时，为了准确估计成本和效果，应扣除因物价上涨带来的影响，对发生在将来的成本和效果（效益或效用）进行贴现处理，换算为当前的价值。

九、敏感性分析

在得到上述经济学评价结果后，进一步做敏感性分析（sensitivity analysis）。即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，或者采用不同经济学分析方法时对评价结果的影响程度，称为敏感性分析。倘若稍微改变一下变量参数，其经济评价的结论就发生改变，则表明结果的可靠性较差（图 17-1）。

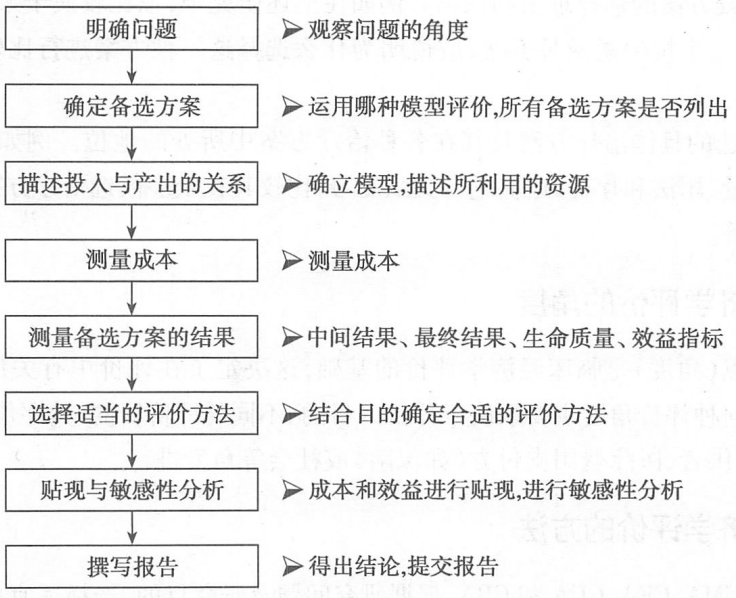


图 17-1　临床经济学评价步骤

笔记

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，因此应将敏感性分析作为经济学评价中一项必要步骤。研究中所有不肯定的结果都应报告，建议使用关键参数的可信区间等概念。

十、撰写报告

在上述分析的基础上，得出结论，还要进行文献复习，与其他研究结果进行横向比较，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，这对确定研究结果的推广应用价值非常重要。

第四节　临床经济学评价研究文献的评价

对临床经济学评价研究结果，建议使用下述的十条标准评价其真实性、重要性和适用性。

1. **研究是否回答了关于经济学评价的问题**　即是否同时比较了两种或两种以上不同措施的成本和结果（效果、效用或效益），采用的是何种临床经济学评价方法。同时，是否阐明评价角度，是从患者角度、社会角度、医疗服务提供者，如医院角度，还是支付方角度，如医疗保险公司等。

2. **对所要比较的方案是否作了详细的描述**　方案的描述内容应包括实施方案的时间、地点、对象、方法和分组情况等。

3. **有无健康结果测定的有效证据、结果的测定是否真实可靠**　可靠性最强的证据大多来自临床随机对照研究，其次是非随机同期对照的研究。历史对照研究结果的可靠性较差。如果是队列研究应该有相同的起点，还应考虑各项评价结局指标是否客观，各项指标的估算与预测是否科学等。

4. **是否对每一组重要的成本和结果都作了确定**　结果的测定是中间替代指标还是健康结局指标。成本是否包括直接医疗成本、直接非医疗成本及间接成本等。

5. **成本与结果的测定单位是否恰当**　各测定单位是如何确定的，有无科学性。

6. **成本结果估计的可信性**　如效用值是如何确定的，测定方法的可信度如何，是否作了信度与效度分析。计算成本的数据来源是否可靠和合理。

7. 对发生在将来的成本和结果是否作了时间校正，贴现率是多少，是如何确定的，贴现后经济学评价结果如何？

8. 有无进行增值分析。

9. **是否作了敏感性分析**　是否列出敏感性分析的各项参数及其变化范围，经济学分析结果是否发生改变，敏感性分析的结论是什么。

10. 研究报告中的结果和讨论是否包括了读者所关心的问题，是否作了伦理学上的讨论。本点涉及到研究结果的推广应用价值，在决策时应兼顾伦理学问题，特别是涉及一些与生命有关的问题，如当费用降低时，效果也随着变差（寿命的缩短或死亡率的增加），此时是否要继续采用该项方案。

（李晓枫）

第十八章　临床医学论文的撰写原则和报告规范

临床医学论文是临床科研工作的总结，通过对临床问题的研究，既提升理论认知水平，同时又获得一定的可推广的结论，能更好地指导临床实践。临床医学论文更关注临床研究和临床应用价值，本章将重点阐述。

第一节　临床医学论文的格式和撰写原则

国际医学杂志编辑委员会（International Committee of Medical Journal Editors，ICMJE）讨论制定了医学论文的统一格式和要求，包括题目、摘要、前言、材料和方法、结果（图和表）、讨论、致谢、参考文献。一般性论著的字数不超过4000字（不包括图表和参考文献）。下面将逐一介绍各部分的撰写格式和原则。

一、题目和作者

论文的题目必须醒目、简单、扼要，一目了然，字数不能太多，以20个汉字以内为宜，并且标题的表达方式要能吸引读者，同时文题要相符，题目不能太大超越了文章的内容。

标题不可出现缩写，但可有副标题（subheading），副标题常常是将主要研究方案列出附在主标题之后，如“氯霉素一线治疗儿童急性感染性结膜炎：随机双盲对照试验”，对随机对照试验的研究题目，规定副标题必须加上。在题目中最好体现出该研究的设计方案，如“出生体重与环境因素对儿童青少年超重肥胖影响的队列研究”。

作者姓名在文题下按序排列，作者单位名称及邮政编码脚注于首页左下方。作者应是：①参与选题和设计，或参与资料的分析和解释者；②起草或修改论文中关键理论或其他主要内容者；③能对编辑部的修改意见进行核修、学术答辩，并最终同意该文发表者。仅参与获得资金或收集资料，及进行一般管理的人员不宜列为作者，对这些人员的贡献应列入致谢部分。通讯作者对论文结论、修改负有主要责任。集体署名文章必须注明对该文负责的关键人物或执笔人。

二、摘要和关键词

摘要（abstract）放在论文正文的最前面，但应该在完成整篇论文后最后撰写。摘要分为结构式摘要（structured abstract）和非结构化摘要。

国外期刊大多采用Haynes RB等提出的格式，包括目的（objective）、设计（design）、研究场所（setting）、患者或其他研究对象（patients or other participants）、干预措施（interventions）、主要结果的测量方法（main outcome measures）、结果（results）及结论（conclusion）共8项（表18-1）。

我国大多数医学期刊和一些国外杂志将其简化为：目的、方法、结果和结论四部分，并采用第三人称撰写，不用“本文”等主语，文字要极其精练，不一定用完整句子，字数限于250字左右。

目的主要介绍研究背景和研究目的，一般1～2句话。方法主要介绍设计方案、研究对象、干预方法、主要和次要观察指标、对结局指标的测量方法和重要的统计分析方法等。结果部分包括纳入研究

的人数、纳入分析的人数、主要观察指标一定要有具体数据和可信区间、重要的不良反应。结论中不写具体数据，是对结果的总结和解释，一般也是1～2句话。若是临床试验，最后还需写出临床试验注册号。

表18-1　结构式摘要的内容

八项结构	四项结构
1. 目的	1. 目的
2. 设计	2. 方法　包括设计、研究场所、研究对象和干预措施，以及测量方法
3. 研究场所	
4. 研究对象	
5. 干预措施	
6. 主要结果的测量方法	
7. 结果	3. 结果
8. 结论	4. 结论

摘要下面标出3～5个关键词(key words)。标出关键词的目的主要是为了便于建立索引和文献检索，因此要求尽可能准而全地标出研究和讨论的重点内容，一般性内容则不予标出。尽量使用美国国立医学图书馆编辑的Index Medicus中医学主题词表(MeSH)内的主题词。

三、序言或引言

论文开始部分是序言或引言(introduction)，一般不将“序言或引言”列为标题，序言部分主要讲清楚所研究问题的来源、背景及本文的研究目的，字数不宜过多，一般300～500字。

序言要求清楚阐述论文撰写的目的，使读者能一目了然，弄清本文所研究问题的来源和重要性以及研究目的是什么。引言部分的内容不能和文中任何部分重复，初写者常错将引言部分内容和讨论部分重复、混为一谈。

引言的写作一般遵从“从大到小”“从面到点”的方式，例如比较急性髓细胞白血病的两种治疗方案的疗效，先简单介绍急性髓系白血病常用的治疗方法和治疗现状，再介绍存在的问题以引出本文研究的问题，说明为什么需要比较这两种治疗方法的理由，研究的必要性，最后介绍本研究的目的和假设，以及采用何种设计方案来进行研究，预期结果和达到目标，研究假设等。

四、材料与方法

论文正文的第二部分“材料与方法”(material and methods)，也有称“对象与方法”。这部分是论文的重要组成部分，约占正文的1/3篇幅，一般分析性和实验性研究大约需要1500字才能写清楚。需要详细撰写的理由是：使读者看了以后能重复，以及便于审稿者复核。内容如下：

1. 研究对象

(1) 研究对象入选的方法，即如何从目标人群选入样本人群，将研究对象的来源介绍清楚，主要目的除了估计样本代表性和抽样误差外，尚能帮助读者了解论文结论的适用范围。撰写时可使用下列术语：随机样本、基于人群的样本、转诊样本(referred sample)、连续样本、志愿者样本(volunteer sample)、方便抽取样本(convenience sample)等。

(2) 诊断标准和纳入/排除标准，如有公认诊断标准应写明出处，切不可笼统地冠以“全部研究对象符合全国统一诊断标准”。

(3) 入选研究对象的样本数，如有拒绝入选者应注明人数。同时简述计算样本量的具体方法和依据。

(4) 研究对象的分组方法，是否随机分配，采用何种随机分配方法：简单随机化、区组随机化或分

层随机化，切不可简单地写“随机分组”一句话。随机分配方案隐藏机制和如何实施随机化。

2. 研究方法

（1）要详细说明研究设计的方案是什么，具体如何设计的。如治疗性研究常用随机对照研究，要描述随机方法、如何选择对照组、是否盲法、样本量的依据。

（2）研究场所要写清楚，如“人群或社区”“医学中心”“基层医院”“门诊”“住院”等。

（3）试验的措施及执行方法应详细交代；用于患者的药物应写明化学名、商品名、生产厂名，中药还应注明产地，并详细说明每日剂量、次数、用药途径和疗程；试剂应写明生产厂家名，试验方法如是作者新建立的要详细介绍，已有的方法应注明出处，所采用的仪器须注明型号及生产厂名。

（4）资料收集方法。盲法的具体实施情况应交代，包括安慰剂的制作，保证盲法成功的措施等。写明施盲对象，是对研究对象或观察医师或统计人员实行了盲法，属于单盲或双盲临床试验。

（5）测量指标及判断结果的标准。主要观察指标和次要观察指标要明确，不能仅仅罗列出指标，分不清主次。确定暴露及疗效标准等都要有公认的标准，如缓解的标准、生存时间是从哪一天开始算起、无病生存的定义是什么等等。

（6）保证研究质量，控制偏倚发生所采用的措施。

3. 统计分析　数据分析方法包括统计软件。基于何种数据类型选用何种统计方法，如两组正态分布计量数据采用 t 检验，三组以上数据比较应采用方差分析；计数资料采用 χ^2 检验等；如何进行亚组分析和多因素校正分析。

五、结果

结果（results）是论文的核心部分，须将观察结果或试验结果实事求是地撰写清楚，用全文的1/4～1/3篇幅书写这部分内容。

1. 数据表达要完整　列出随机分组例数和纳入分析的例数，报告结果的例数与入选研究对象的例数应吻合，剔除例数与剔除理由应交代，失访数及因其他原因死亡数也应写清楚，如有数据不全应作解释。临床试验应该按照随机对照试验强制报告规范（CONSORT声明）的要求画出受试者流程图（图18-1）。

如进行两组或几组比较，应列出几组除研究因素以外的其他临床基线情况（baseline），并进行均衡性试验，各组间是否可比（表18-2）。

表18-2　铁剂治疗前3组患者的基线情况比较

	多糖铁复合物（n=31）	硫酸亚铁控释片（n=36）	琥珀酸亚铁（n=38）	F值	P值
年龄（岁）（中位数）	40	41	39		
性别比例（男/女）	2/29	3/33	6/32		
血红蛋白（g/L）	70±16.2	74.9±12.6	70±12.6	1.26	0.29
网织红细胞（%）	1.35±0.7	1.28±0.6	1.5±0.9	0.48	0.62
平均红细胞体积（fl）	71.76±5.1	74.6±13.6	68.6±13.0	1.69	0.19
血清铁蛋白（nmol/L）	8.61±10.6	9.5±6.1	9.8±7.3	1.63	0.44

2. 科研设计时确定的科研假设及主要测量指标，如在结果部分作了更动应作解释。全部的结果都应该报道，不能只报告有意义的阳性结果。

3. 诊断试验的研究应报告敏感度、特异度、预测值、似然比及受试者工作特征（ROC）曲线。

4. 统计处理注意事项　要同时报道百分比和绝对数，如10/20例，而不能只报告50%。应用的变量性质、均数、百分数、率、比例等应正确，选择的各种统计分析方法要合理，复杂的统计分析要作恰当解释。应同时报道95%可信区间（95% CI）和 P 值，不能只报告 P 值，应同时报告统计量，如 t 值、F

笔记

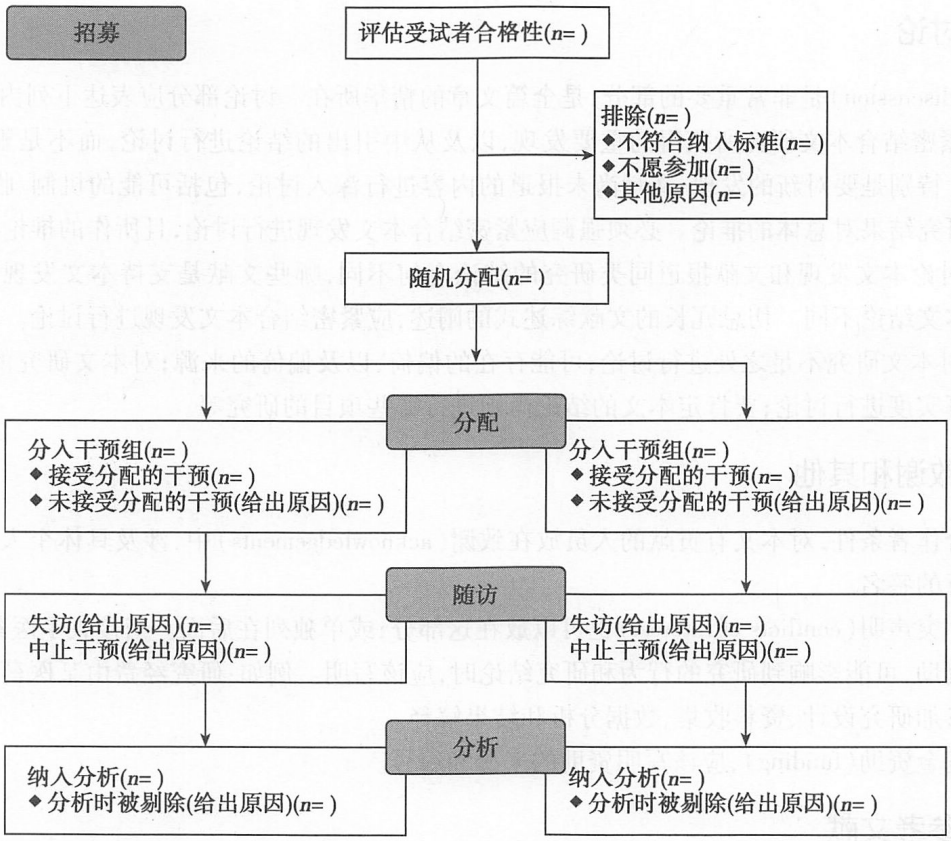


图 18-1　CONSORT 声明受试者流程图

值等。如差异无统计学意义，还应进一步说明有无临床意义。

5. 结果部分的表达形式可分文字部分和图表部分。文字表达和图表表达不要重复，能够用文字表达清楚的就不用图表。文字表达应当是要点式叙述，可分几项撰写，每一项报告一组数据，使读者看了一目了然。图表的表达应符合统计学的规定。

6. 统计表的结构应包括标题、横标目（表达研究和观察项目）、纵标目（表达横标目的各个统计指标），横标目列在表的左侧，纵标目列在表的上端。指标的计算单位须注明，表内数字必须正确，小数的位数应一致。线条不宜过多，表的上下两条边线可用较粗的横线，一般采用三横线表（顶线、表头线、底线），如有合计可再加一条横线隔开，但不宜用竖线。说明不列入表内，可用“*，†，‡，#”等符号标出写在表下面（表 18-3）。

表 18-3　急性髓系白血病非长期和长期生存患者发病时实验室指标的比较

组别（例数）	血小板（$\times10^9$/L）	白细胞（$\times10^9$/L）	骨髓原+幼（%）	血红蛋白（g/L）
长期生存组（$n=150$）	70.6±51.3	6.65（1～82.6）*	77.0±35.8	71.5±22.9
非长期生存组（$n=878$）	59.3±49.9	10.3（1.25～163）*	73.8±42.0	72.9±27.7
t	2.55	5.47	0.88	0.59
P	0.01	<0.01#	0.38	0.56

*：中位数（90% 区间），#：非参数检验

7. 统计图比统计表更便于理解与比较，但统计图中不能获得确切数字，所以不能完全代替统计表。图的标题应置于图的下端，图有纵轴和横轴，两轴应有标目，标目应注明单位，横轴尺度自左至右，纵轴尺度自下而上，尺度必须等距，数值一律由小到大，一般纵轴尺度必须从 0 点起始（对数图及点图等除外）。图的长、宽比例一般以 7∶5 为宜。

常用的统计图有直条图、圆形图、百分直条图、线图、直方图、散点图等。

六、讨论

讨论(discussion)是非常重要的部分,是全篇文章的精华所在。讨论部分应表达下列内容:

1. 应紧密结合本文研究所获得的重要发现,以及从中引出的结论进行讨论,而不是重复结果部分的内容。特别是要对新的发现、文献尚未报道的内容进行深入讨论,包括可能的机制、临床应用范围以及从研究结果对总体的推论。必须强调应紧密结合本文发现进行讨论,且所作的推论必须恰当。

2. 应讨论本文发现和文献报道同类研究的结论有何不同,哪些文献是支持本文发现的,哪些文献报道与本文结论不同。切忌冗长的文献综述式的阐述,应紧密结合本文发现进行讨论。

3. 应对本文研究不足之处进行讨论;可能存在的偏倚,以及偏倚的来源;对本文研究的内部真实度和外部真实度进行讨论;要肯定本文的结论尚须进行哪些项目的研究等。

七、致谢和其他

不符合作者条件,对本文有贡献的人员放在致谢(acknowledgements)中,涉及具体个人时,需要获得书面同意的签名。

利益冲突声明(conflicts of interest)也可以放在这部分,或单独列在后面。当接收了医药公司或其他来源的资助,可能影响到研究的行为和研究结论时,应该写明。例如:研究经费由某医药公司提供,但公司不参加研究设计、资料收集、数据分析和结果解释。

若有基金资助(funding),应该写明资助的来源和编号。

八、参考文献

依照其在文中出现的先后顺序用阿拉伯数字标出,尽量避免引用摘要作为参考文献,也不要引用未公开发表的文章及私人提供的个人信息,不要把相关文献中的参考文献不经阅读而“转引”。参考文献中的作者,1~3 名全部列出,3 名以上只列前 3 名,后加“,等”。外文期刊名称用缩写,以 Index Medicus 中的格式写准,中文期刊用全名。每条参考文献均须录起止页。

【举例】

[1] 王小钦,陈子兴,陈书长,等. 成人骨髓增生异常综合征难治性贫血生存率和预后因素分析. 中华血液学杂志,2008,29(7):491-493.

[2] You CH, Lee KY, Chey WY, et al. Electrogastrographic study of patients with unexplained nausea, bloating and vomiting. Gastroenterology, 1980, 79: 311-315.

第二节 临床医学论文的报告规范和投稿要求

一、论文撰写报告规范

为了提高研究报告的质量,目前不同研究设计的论文撰写均要遵循相应的报告规范和指南(可在相关网站中获取)。如随机对照试验的 CONSORT 报告规范,非随机干预研究的 TREND 规范,观察性研究的 STROBE,病例报告的 CARE 标准,诊断试验评价的 STARD 规范,卫生经济学评价的 CHEERS,基于随机对照试验的 meta-分析的 PRISMA 以及基于观察性研究 meta-分析的 MOOSE 等。

最著名的是临床试验的报告统一标准(consolidated standards of reporting trials, CONSORT)声明,分为报告清单(表 18-4)和受试者流程图(图 18-1)二部分。清单共有 25 个条目,对题目、摘要、方法、结果和讨论等需要撰写的内容都进行了详细的要求。在研究设计和论文写作之前应该认真阅读和理解,论文完成后也应检查是否都报告了这些内容。投稿时多数杂志要求作者逐条核对清单,并且标注报告的页码,说明是否均按照要求完成了相关的内容报告。CONSORT 声明尚有多种扩展版,如群组随机试验、非劣效性和等效性试验、非药物治疗、草药干预以及实效性试验的各种 CONSORT 扩展版,必要时可以参考。

表 18-4　CONSORT 2010 对照检查清单(Checklist)

论文章节/主题	条目号	对照检查的条目	报告页码
文题和摘要			
	1a	文题能识别是随机临床试验	
	1b	结构式摘要,包括试验设计、方法、结果、结论几个部分(具体的指导建议参见“CONSORT for abstracts”)	
引言			
背景和目的	2a	科学背景和对试验理由的解释	
	2b	具体目的和假设	
方法			
试验设计	3a	描述试验设计(诸如平行设计、析因设计),包括受试者分配入各组的比例	
	3b	试验开始后对试验方法所作的重要改变(如合格受试者的挑选标准),并说明原因	
受试者	4a	受试者合格标准	
	4b	资料收集的场所和地点	
干预措施	5	详细描述各组干预措施的细节以使他人能够重复,包括它们实际上是在何时、如何实施的	
结局指标	6a	完整而确切地说明预先设定的主要和次要结局指标,包括它们是在何时、如何测评的	
	6b	试验开始后对结局指标是否有任何更改,并说明原因	
样本量	7a	如何确定样本量	
	7b	必要时,解释中期分析和试验中止原则	
随机方法			
序列的产生	8a	产生随机分配序列的方法	
	8b	随机方法的类型,任何限定的细节(如怎样分区组和各区组样本多少)	
分配隐藏机制	9	用于执行随机分配序列的机制(例如按序编码的封藏法),描述干预措施分配之前为隐藏序列号所采取的步骤	
实施	10	谁产生随机分配序列,谁招募受试者,谁给受试者分配干预措施	
盲法	11a	如果实施了盲法,分配干预措施之后对谁设盲(例如受试者、医护提供者、结局评估者),以及盲法是如何实施的	
	11b	如有必要,描述干预措施的相似之处	
统计学方法	12a	用于比较各组主要和次要结局指标的统计学方法	
	12b	附加分析的方法,诸如亚组分析和校正分析	
结果			
受试者流程(极力推荐使用流程图)	13a	随机分配到各组的受试者例数,接受已分配治疗的例数,以及纳入主要结局分析的例数	
	13b	随机分组后,各组脱落和被剔除的例数,并说明原因	
招募受试者	14a	招募期和随访时间的长短,并说明具体日期	
	14b	为什么试验中断或停止	

续表

论文章节/主题	条目号	对照检查的条目	报告页码
基线资料	15	用一张表格列出每一组受试者的基线数据，包括人口学资料和临床特征	————
纳入分析的例数	16	各组纳入每一种分析的受试者数目（分母），以及是否按最初的分组分析	————
结局和估计值	17a	各组每一项主要和次要结局指标的结果，效应估计值及其精确性（如95%可信区间）	————
	17b	对于二分类结局，建议同时提供相对效应值和绝对效应值	————
辅助分析	18	所做的其他分析的结果，包括亚组分析和校正分析，指出哪些是预先设定的分析，哪些是新尝试的分析	————
危害	19	各组出现的所有严重危害或意外效果（具体的指导建议参见“CONSORT for harms”）	————
讨论			
局限性	20	试验的局限性，报告潜在偏倚和不精确的原因，以及出现多种分析结果的原因（如果有这种情况的话）	————
可推广性	21	试验结果被推广的可能性（外部可靠性，实用性）	————
解释	22	与结果相对应的解释，权衡试验结果的利弊，并且考虑其他相关证据	————
其他信息			
试验注册	23	临床试验注册号和注册机构名称	————
试验方案	24	如果有的话，在哪里可以获取完整的试验方案	————
资助	25	资助和其他支持（如提供药品）的来源，提供资助者所起的作用	————

二、论文投稿要求

论文撰写完成后投稿前应该仔细阅读各个杂志的投稿要求，按照其格式整理，然后网上在线投稿。目前国际上大多数杂志根据国际医学期刊编辑委员会（ICMJE）制定的生物医学期刊投稿的统一要求，稿约内容主要包括3个方面：①与研究实施和报告相关的伦理道德问题，如临床试验注册、伦理委员会、知情同意书、利益冲突问题等；②与在生物医学期刊发表论文相关的出版和编辑问题，如发表阴性结果的义务、重复发表、一稿多投、单子版等问题；③文稿准备与投稿的技术问题，如文章格式、参考文献格式、图表、统计要求等。

投稿国外杂志时需要按照稿约规定，依照下列顺序整理好再投稿，先注册后投稿。

1. 给编辑的信 Cover letter
2. 题目页 Title page
3. 摘要和关键词 Abstract and Key words
4. 引言 Introduction
5. 材料和方法 Materials and Methods
6. 结果 Results
7. 讨论 Discussion
8. 致谢 Acknowledgments
9. 参考文献 References

10. 表格 Table

11. 图 Figures

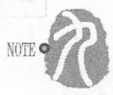
12. 图注 Legends for figures

给编辑的信(Cover letter)即投稿信,主要说明该文的题目和创新点,是首次发表还是重复发表,或已将其中某一部分投寄到别处。注明稿件属于该期刊的哪一类文章。声明可能导致权益冲突的经济关系或其他事宜。推荐的审稿人或需回避的审稿人。说明原稿已经所有作者阅读认可,所有作者均符合作者资格。以及是否愿意支付版面费用等。

题目页包括论文题目,所有作者的姓名和所在单位,有些杂志希望提供作者的最高学历。通信作者,包括姓名、通信地址、电话、传真号码、Email 地址。资助基金、药物、仪器和其他资助的来源、项目名称。提供页头题目(running head):有些杂志需要作者提供非常简短的放在页头或页脚的能概括该研究主要内容或研究方向的小题目,少于 40 个字符(包括字母和空格)。提供全文的字数,但不包括摘要、致谢、图的注解和参考文献的字数。注明文章中图和表的数量。

表格和图不放在正文内,放在参考文献后面,单独成页,图注放在图的后面,也要单独成页。文字用 12 号字体,用双倍行距。

(王小钦)



附表　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
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

续表

编号	1～10					11～20					21～30					31～40					41～50				
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 RA, Yates F. Statistical Tables for Biological, Agricultural and Medical Research. Oliver and Boyd, 1957:128

推荐阅读

[1] 刘续宝,王素萍. 临床流行病学和循证医学. 4 版. 北京:人民卫生出版社,2013.
[2] 詹思延. 流行病学. 8 版. 北京:人民卫生出版社,2017.
[3] 李立明. 临床流行病学. 北京:人民卫生出版社,2011.
[4] 彭晓霞,唐迅,译. 临床研究设计. 4 版. 北京:北京大学医学出版社,2017.
[5] 闫永平. 临床流行病学. 北京:人民卫生出版社,2009.
[6] 王家良. 临床流行病学. 3 版. 北京:人民卫生出版社,2008.
[7] 王家良. 循证医学. 2 版. 北京:人民卫生出版社,2011.
[8] 洪明晃. 临床科学研究设计、测量、评价. 3 版. 上海:上海科技出版社,2009.
[9] 方积乾. 卫生统计学. 7 版. 北京:人民卫生出版社,2012.
[10] 康德英,许能锋. 循证医学. 3 版. 北京:人民卫生出版社,2016.
[11] 孙凤. 医学研究报告规范解读. 北京:北京大学医学出版社,2015.
[12] 刘爱忠,黄民主. 临床流行病学. 2 版. 长沙:中南大学出版社,2010.
[13] 黄民主,刘爱忠. 临床流行病学. 2 版. 北京:高等教育出版社,2013.
[14] 黄民主,陈立章. 循证医学. 北京:高等教育出版社,2015.
[15] 黄悦勤. 预防医学. 北京:北京医科大学出版社,2004.
[16] 杨克虎. 循证医学. 北京:科学出版社,2007.
[17] 王小钦,王吉耀. 循证临床实践指南的制定与实施. 北京:人民卫生出版社,2015.
[18] 罗爱静. 医学文献信息检索. 2 版. 北京:人民卫生出版社,2010.
[19] 林果为,王小钦,陈世耀. 现代临床流行病学. 3 版. 上海:复旦大学出版社,2014.
[20] 王行环. 循证临床实践指南的研发与评价. 北京:中国协和医科大学出版社,2016.
[21] 谭红专. 现代流行病学. 北京:人民卫生出版社,2005.
[22] 孙贵范. 预防医学. 北京:人民卫生出版社,2006.
[23] 贺石林,陈修. 医学科研方法学导论. 北京:人民卫生出版社,1998.
[24] Straus SE, Richardson WS, Glaszion P. Evidence-based medicine: how to practice and teach EBM. 3rded. London: Churchill Livingstone, 2005.
[25] Sharon E. Straus. Evidence-Based Medicine: How to practice and teach it. 4th ed. Elsevier Ltd, 2011.
[26] Haynes B, Sackett DL, Guyatt GH, et al. Clinical Epidemiology: How to Do Clinical Practice Research. 3rd ed. Philadelphia: Lippincott Williams & Wilkins, 2005.
[27] Rothman KJ, Greenland S. Modern Epidemiology. 2nd ed. Philadelphia: Lippincott Raven Publishers, 1998.
[28] Fletcher RH, Fletcher SW, Wagner EH. Clinical Epidemiology: The Essentials. 3rd ed. Philadelphia: Lippincott Williams & Wilkins, 1996.

中英文名词对照索引

ROC 曲线下的面积　area under the ROC curve, AUC　125

A

安慰剂对照　placebo control　25
安慰剂效应　placebo effect　24

B

半随机对照试验　quasi-randomized controlled trial　43
半随机化法　quasi-randomization　23
保密　confidential　29
报告偏倚　reporting bias　31,85
暴露　exposure　44
暴露怀疑偏倚　exposure suspicion bias　31
比值比　odds ratio, OR　88,98,89
标准博弈法　Standard gamble　187
标准化均数差值　standardized mean difference, SMD　89
病残率　morbidity　210
病例对照研究　case-control study　49
病例分析　case analysis　56
病因分值　etiologic fraction　48
病因链　chain of causation　102
伯克森偏倚　Berkson's bias　30
不能预知的结局　unpredictable outcome　24
不依从　non-compliance　146

C

测量　measurement　5
测量偏倚　detection bias　32
巢式病例对照研究　nested case-control study　50
成本　cost　210
成本确定分析　cost identification analysis　213
成本-效果　cost-effectiveness　8
成本效果比　cost-effectiveness ratio, C/E　213
成本-效果分析　cost-effectiveness analysis, CEA　213
成本-效益　cost-benefit　8
成本-效益分析　cost-benefit analysis, CBA　213
成本-效用　cost-utility　8
成本-效用分析　cost-utility analysis, CUA　214
成组序贯设计　group sequential design　139
尺度评分　scale　76
重复　replication　28
抽样调查　sampling survey　54
粗一致性　crude agreement rate　124

D

单臂研究　single arm study　44
单病例随机对照试验　N of 1 RCTs　139
单纯随机抽样　simple random sampling　54
单纯知照模式　pure informed model　183
单个患者的随机对照试验　number of one randomized controlled trial, n-of-1 trial　42
单盲　single-blind　26
定量系统评价　quantitative systematic review　78
定性系统评价　qualitative systematic review　78
队列研究　cohort study　44
多级抽样法　multi-stage sampling　23
多阶段抽样　multi-stage sampling　54

F

发表偏倚　publication bias　83
发病密度　incidence density, ID　47
方法学异质性　methodological heterogeneity　87,89
非等量随机对照试验　unequal randomized controlled trial　43
非连续系列病例分析　nonconsecutive case series　56
非随机对照试验　non-randomized controlled trial, NRCT　139
非随机同期对照　non-randomized concurrent control　25
非随机同期对照试验　concurrent nonrandomized controlled trial　43
分层抽样　stratified sampling　54
分层随机抽样　stratified sampling　21
分层随机法　stratified randomization　21
分层随机分组　stratified allocating　21
分层因素　stratifying factor　21
符合率　agreement/consistency rate　127

G

干扰 co-intervention 145
个案病例报告 case report 56
公正 justice 29
共同决策模式 shared decision model 183
固定效应模型 fixed effect model 88
观察偏倚 observational bias 31
观察性研究 observational studies 34,44
归因危险度 attributable risk,AR 48
归因危险度百分比 attributable risk percent,AR% 48
国际临床流行病学网 International Clinical Epidemiology Network,INCLEN 1

H

赫尔辛基宣言 *The Declaration of Helsinki* 137
横断面研究 cross-sectional study 53
划线法 visual analog scale 187
坏死性小肠结肠炎 necrotizing enterocolitis NEC 200
患病率 prevalence 55
患者预期事件发生率 patient's expected event rate,PEER 148
回顾性队列研究 retrospective cohort study 45
回顾性研究 retrospective study 49
回忆偏倚 recall bias 31
混杂 confounding 32
混杂偏倚 confounding bias 32
混杂因素 confounding factor 32
霍桑效应 Hawthorne effect 24,145

J

机械抽样 mechanical sampling 22
机遇结点 chance node 185
急性脑梗死 acute cerebral infarction 150
家长作风模式 paternalistic model 183
检出症候偏倚 detection signal bias 31
简单随机法 simple randomization 21
健康结局指标 health measures 213
交叉对照 cross-over design control 25
交叉试验 cross-over design,COD 39,139
金标准 gold standard 119
净现值法 net present value,NPV 214
决策 making decision 183
决策分析 making decision analysis 183
决策结点 decision node 185
决策树 decision tree 184
绝对受益增高率 absolute benefit increase,ABI 8
绝对危险降低率 absolute risk reduction,ARR 8,147
均数差值 mean difference,MD 88,89,98

K

卡方检验 χ^2-test 144
开放试验 open trial 27
可靠性 reliability 4,127
空白对照 blank control 25

L

累积发病率 cumulative incidence,CI 47
历史对照 historical control 25
历史对照试验 historical control trial,HCT 139
历史性队列研究 historical prospective study 45
利弊比 likelihood of being helped VS harmed,LHH 149,182
连续系列病例分析 consecutive case series 56
临床经济学 clinical economics 209
临床决策分析 clinical decision analysis,CDA 183
临床流行病学 clinical epidemiology 1
临床实践指南 clinical practice guideline,CPG 168
临床试验的报告统一标准 consolidated standards of reporting trials,CONSORT 222
临床问题 clinical question 12
临床异质性 clinical heterogeneity 87,89
临床重要性 clinical importance 75
临界点 cut-off point 121

M

盲法 blinding 35
盲法试验 blind trial 26
美国医学会 American Medical Association,AMA 65
敏感度 sensitivity 121
敏感性分析 sensitivity analysis 88,96,113,216

N

奈曼偏倚 Neyman bias 31
脑室内出血 intraventricular haemorrhage,IVH 200

P

排除偏倚 exclusive bias 31
配对对照 matching control 25
偏倚 bias 30
平行试验 parallel tests 125
评价 evaluation 5

普查 census 54

Q

期间患病率 period prevalence 55
起始队列 inception cohort 157
前-后对照试验 before-after study 139
前瞻性队列研究 prospective cohort study 45
前瞻性研究 prospective study 45
潜在减寿年数 potential years of life lost, PYLL 16
清单 checklist 76
区组随机法 block randomization 21

R

人群归因危险度 population attributable risk, PAR 48
人群归因危险度百分比 population attributable risk percent, PAR% 48
人时 person-time, PT 47
入院率偏倚 admission rate bias 30

S

三盲 triple-blind 27
森林图 forest plot 88
伤残调整寿命年 disability adjusted life year, DALY 16, 211
设计 design 5
"设计、测量、评价" design, measurement and evaluation, DME 2
生物学依据 biologic evidence 180
剩余效应 carry-over effects 39
时点患病率 point prevalence 55
时间权衡法 time trade-off 188
实施偏倚 performance bias 84
实效性随机对照临床试验 pragmatic clinical trial, PCT 59
试验性研究 experimental studies 34,44
适应性设计 adaptive design 139
适用性 applicability 146
受试者工作特征曲线 receiver operator characteristic curve, ROC 曲线 121
受益 beneficence 29
双盲 double-blind 26
双向性队列研究 ambispective cohort study 45
死亡率 mortality 210
似然比 likelihood ratio, LR 121
随访 follow-up 46
随访偏倚 attrition bias 85
随机抽样 random sampling 20,54
随机对照试验 randomized controlled trial, RCT 34,139
随机分组 random allocating 20
随机化 randomization 20
随机误差 random error 30
随机效应模型 random effect model 88,93

T

特异度 specificity 121
贴现率 discount rate 211
同期随机对照 concurrent randomized control 24
统计学异质性 statistical heterogeneity 88,89

W

卫生经济学 health economics 209

X

析因设计 factorial design 140
洗脱期 washout period 39
系列病例分析 case series 56
系列试验 serial tests 125
系统抽样 systematic sampling 54
系统随机抽样法 systematic sampling 22
系统误差 systematic error 30
现患率研究 prevalence study 53
现患-新病例偏倚 prevalence-incidence bias 31
相对受益增高率 relative benefit increase, RBI 8
相对危险度 relative risk, RR 48,88,89,98
相对危险降低率 relative risk reduction, RRR 8
向均数回归 regression to the mean 24
向均数回归现象 regression to the mean 145
效果 effectiveness 210
效益 benefit 211
效用 utility 211
信息偏倚 information bias 31
需治多少病例才发生一例不良反应 number needed to harm, NNH 147
需治疗多少病例数才获得一例好结果 number need to treat, NNT 147
序贯试验 sequential trial 139
叙述性文献综述 narrative review 79
选择偏倚 selection bias 30,84
循证医学 evidence-based medicine, EBM 1,11
循证证据摘要 syntheses of synopses 65
循证指南制定法 evidence-based guideline development 171

Y

阳性预测值 positive predictive value,PPV 123
依从性 compliance 146
异质性检验 heterogeneity 87
易感性偏倚 susceptibility bias 31
意向性治疗分析 intention to treat analysis 85
意向治疗分析 intention-to-treat analysis,ITT 144
阴性预测值 negative predictive value,PV 123
隐匿 concealment 23
隐性成本 intangible costs 210
英国医学杂志 *British Medical Journal*,*BMJ* 65
语言偏倚 language bias 83
预测值 predictive value,PV 123
预后 prognosis 153
预后因素 prognostic factors 154

Z

增量成本效果比 incremental cost-effectiveness ratio,ICER 213
沾染 contamination 145
诊断怀疑偏倚 diagnostic suspicion bias 31
诊断试验 diagnostic test 119
真实世界研究 real-world studies,RWS 58
真实世界证据 real-world evidence,RWE 58
真实性 validity 4,126,146,179
整群抽样 clustering sampling 54
整群随机对照试验 cluster randomized controlled trial 43
整群随机法 cluster randomization 23
知情 informed 29
直接成本 direct costs 210
直接非医疗成本 direct nonmedical cost 210
直接医疗成本 direct medical cost 210
质量调整寿命年 quality-adjusted life years,QALYs 188,211,215
质量评价 assessment of quality 84
中国临床流行病学网 China Clinical Epidemiology Network,ChinaCLEN 2
中间替代测量指标 intermediate measures 213
中位生存时间 median survival time 156
重要性 importance 146,179
准确度 accuracy 124
自身对照 self control 25
自身前后对照试验 before-after study in the same patient 41
自愿 voluntary 29
总生存率 overall survival rate,OS 155
最小成本分析 cost minimization analysis,CMA 213
最新最佳证据 current best evidence 179
尊重 respect 29

28检

国家卫生健康委员会“十三五”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

1.	医用高等数学	第 7 版	28.	眼科学	第 9 版
2.	医学物理学	第 9 版	29.	耳鼻咽喉头颈外科学	第 9 版
3.	基础化学	第 9 版	30.	口腔科学	第 9 版
4.	有机化学	第 9 版	31.	皮肤性病学	第 9 版
5.	医学生物学	第 9 版	32.	核医学	第 9 版
6.	系统解剖学	第 9 版	33.	流行病学	第 9 版
7.	局部解剖学	第 9 版	34.	卫生学	第 9 版
8.	组织学与胚胎学	第 9 版	35.	预防医学	第 7 版
9.	生物化学与分子生物学	第 9 版	36.	中医学	第 9 版
10.	生理学	第 9 版	37.	医学计算机应用	第 6 版
11.	医学微生物学	第 9 版	38.	体育	第 6 版
12.	人体寄生虫学	第 9 版	39.	医学细胞生物学	第 6 版
13.	医学免疫学	第 7 版	40.	医学遗传学	第 7 版
14.	病理学	第 9 版	41.	临床药理学	第 6 版
15.	病理生理学	第 9 版	42.	医学统计学	第 7 版
16.	药理学	第 9 版	43.	医学伦理学	第 5 版
17.	医学心理学	第 7 版	44.	临床流行病学与循证医学	第 5 版
18.	法医学	第 7 版	45.	康复医学	第 6 版
19.	诊断学	第 9 版	46.	医学文献检索与论文写作	第 5 版
20.	医学影像学	第 8 版	47.	卫生法	第 5 版
21.	内科学	第 9 版	48.	医学导论	第 5 版
22.	外科学	第 9 版	49.	全科医学概论	第 5 版
23.	妇产科学	第 9 版	50.	麻醉学	第 4 版
24.	儿科学	第 9 版	51.	急诊与灾难医学	第 3 版
25.	神经病学	第 8 版	52.	医患沟通	第 2 版
26.	精神病学	第 8 版	53.	肿瘤学概论	第 2 版
27.	传染病学	第 9 版			

仅限自学参考！严禁任何人网络传播或盗卖！后果自负！

A180 0428 8107
激活码

扫描圆标二维码 或登录 jh.ipmph.com 享受增值服务

策划编辑 邹梦娜
责任编辑 肖宛凝 邹梦娜
数字编辑 马 双
整体设计 郭 淼 单 斯
郑 阳

人卫智网
www.ipmph.com
医学教育、学术、考试、健康，
购书智慧智能综合服务平台

人卫官网
www.pmph.com
人卫官方资讯发布平台

关注人卫健康
提升健康素养

ISBN 978-7-117-26678-9

定 价：46.00 元