

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

国家卫生和计划生育委员会住院医师规范化培训规划教材

# 医学科研方法

Medical Research Method

主 编 | 陈世耀 刘晓清



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

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

国家卫生和计划生育委员会住院医师规范化培训规划教材

# 医学科研方法

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

策划编辑 蔡向阳 刘 彬  
责任编辑 蔡向阳 张亚琴

封面设计 李 蹊  
版式设计 张亚楠



ISBN 978-7-117-20437-8



9 787117 204378 >

定 价：29.00 元

人卫社官网 [www.pmph.com](http://www.pmph.com) 出版物查询，在线购书

人卫医学网 [www.ipmph.com](http://www.ipmph.com) 医学考试辅导，医学数据库服务，医学教育资源，大众健康资讯



国家卫生和计划生育委员会住院医师规范化培训规划教材

# 医学科研方法

## Medical Research Method

主 编 陈世耀 刘晓清

副主编 吕 明 张宏家 肖志波

编 者(以姓氏笔画为序)

王伊龙 首都医科大学附属北京天坛医院

王春晖 四川大学华西医院

方卫纲 北京协和医院

吕 明 山东大学齐鲁医院

刘天舒 复旦大学附属中山医院

刘晓清 北京协和医院

李秀君 山东大学公共卫生学院

李春波 上海市精神卫生中心

李满祥 西安交通大学第二附属医院

肖志波 哈尔滨医科大学

张秀军 安徽医科大学第一附属医院

张宏家 首都医科大学附属北京安贞医院

张爱华 南京市儿童医院

陈世耀 复旦大学附属中山医院

贺 佳 第二军医大学

人民卫生出版社

图书在版编目 ( CIP ) 数据

医学科研方法 / 陈世耀, 刘晓清主编. —北京: 人民卫生出版社, 2015

国家卫生和计划生育委员会住院医师规范化培训规划教材

ISBN 978-7-117-20437-8

I. ①医… II. ①陈… ②刘… III. ①医学-科学研究-研究方法-医师-职业培训-教材 IV. ①R-3

中国版本图书馆 CIP 数据核字 (2015) 第 064400 号

|       |                                                  |                                 |
|-------|--------------------------------------------------|---------------------------------|
| 人卫社官网 | <a href="http://www.pmph.com">www.pmph.com</a>   | 出版物查询, 在线购书                     |
| 人卫医学网 | <a href="http://www.ipmph.com">www.ipmph.com</a> | 医学考试辅导, 医学数据库服务, 医学教育资源, 大众健康资讯 |

版权所有, 侵权必究!

## 医学科研方法

主 编: 陈世耀 刘晓清

出版发行: 人民卫生出版社 (中继线 010-59780011)

地 址: 北京市朝阳区潘家园南里 19 号

邮 编: 100021

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

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

印 刷: 中国农业出版社印刷厂

经 销: 新华书店

开 本: 850 × 1168 1/16 印张: 12

字 数: 330 千字

版 次: 2015 年 6 月第 1 版 2016 年 1 月第 1 版第 3 次印刷

标准书号: ISBN 978-7-117-20437-8/R · 20438

定 价: 29.00 元

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

(凡属印装质量问题请与本社市场营销中心联系退换)



## 出版说明

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

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

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

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

本套教材是在全面实施以“5+3”为主体的临床医学人才培养体系,深化医学教育改革,建设和一支适应人民群众健康保障需要的临床医师队伍的背景下组织编写的,希望全国广大住院医师培训基地在使用过程中提供宝贵意见。

# 国家卫生和计划生育委员会住院医师规范化培训规划教材

## 教材目录

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



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



# 全国住院医师规范化培养教材

## 评审委员会名单

---

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

### 顾 问 (按姓氏笔画排序)

冯友梅 武汉大学  
吕兆丰 首都医科大学  
闫剑群 西安交通大学  
步 宏 四川大学  
张英泽 河北医科大学  
张学军 安徽医科大学  
汪建平 中山大学  
郑树森 浙江大学医学院

赵 群 中国医科大学  
柯 杨 北京大学  
桂永浩 复旦大学  
陶立坚 中南大学  
黄 钢 上海交通大学医学院  
雷 寒 重庆医科大学  
腾皋军 东南大学医学院  
瞿 佳 温州医科大学

### 主 任 委 员 王 辰 陈贤义

### 副主任委员 (按姓氏笔画排序)

于晓松 中国医科大学第一附属医院  
万学红 四川大学华西临床医学院 / 华西医院  
王 虹 南京医科大学第一附属医院  
刘继红 华中科技大学同济医学院附属同济医院  
李 青 国家卫生和计划生育委员会医药卫生科技  
发展研究中心  
杜 贤 人民卫生出版社  
邵 洁 上海交通大学医学院附属瑞金医院

陈 红 北京大学人民医院  
陈旻湖 中山大学附属第一医院  
林建华 福建医科大学附属第一医院  
唐安洲 广西医科大学附属一院  
郭启勇 中国医科大学附属盛京医院  
高海青 山东大学齐鲁医院  
阎作勤 复旦大学附属中山医院



## 委 员 (按姓氏笔画排序)

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

## 主编简介



陈世耀

主任医师, 研究生导师, 现任复旦大学中山医院内科教研室主任, 消化科/内镜中心副主任, 复旦大学临床流行病学/循证医学中心副主任。兼任中华医学会临床流行病学分会候任主任委员; 消化分会临床流行病学协作组组长。上海医学会临床流行病学和循证医学分会主任委员; 食管胃静脉曲张治疗分会候任主任委员。

长期从事临床一线医疗和教学工作, 在临床医疗和教学实践中贯穿循证医学理念, 并开展临床研究。受国际临床流行病学工作网(INCLEN)资助接受临床研究方法等专业培训, 获得临床流行病学硕士学位。曾在德国Essen 大学医院、日本北里大学东病院、香港中文大学等接受临床培训, 在消化病及消化内镜基础与临床研究领域开展了大量工作, 在内科学、临床流行病学和循证医学教学工作中积累了丰富的经验。发表论文 130 余篇, 其中第一作者或通讯作者 60 余篇。参与或主持多项国家级和省部级研究课题, 获多项国家和省级教学与科研奖项和荣誉。



刘晓清

主任医师, 教授, 中国医学科学院北京协和医院感染内科副主任; 国际临床流行病学网(INCLEN)协和临床流行病学室(CEU)主任; 中华医学会临床流行病学分会常务委员、秘书长、循证医学学组副组长; 北京医学会临床流行病学分会副主委; 中华医学会肝病学会分会委员, 中华医学会北京肝病分会常务委员兼学术秘书; 欧洲呼吸病学会会员。

2002~2003 年在英国牛津大学作访问学者; 2006~2007 年在澳门仁伯爵医院(CHCSJ)作感染内科顾问医生; 北京市总工会“教学创新标兵”; 曾获北京协和医院医疗成果一等奖、北京协和医学院优秀教师。在国内外医学杂志发表论文 60 余篇, 编写 10 余部专著及 3 部译著; 主译《NIH 临床研究准则与规范—临床研究基础建设》; 主持及参与国家级、省部级课题 10 余项; “十二五”国家科技传染病重大专项课题“中国风湿免疫病人活动性结核病的流行病学调查和治疗效果及预后研究”(ETHERTB)总负责人。

## 副主编简介



吕明

主任医师、教授,博士生导师,现任山东大学齐鲁医院科研处处长,山东大学临床流行病学与循证医学中心副主任;齐鲁医院临床流行病学暨循证医学研究室主任;复旦大学现代人类学教育部重点实验室特聘教授;复旦-泰州健康科学研究院特聘研究员。兼任中华医学会临床流行病学分会委员,山东省医学会临床流行病学分会主任委员,中华医学会医学科管理分会委员,山东省医学会医院科研管理分会副主任委员。

长期在医院致力于临床研究方法的普及教育工作,组织实施多个大规模的临床研究项目。2007年开始组织筹建国内第一个大型的分子流行病学平台——泰州队列(国家队列)的建设。近五年来承担国家自然科学基金3项,省级课题4项,发表SCI收录论文20余篇。



张宏家

首都医科大学附属北京安贞医院副院长、心脏外科主任医师、博士生导师、教授;首都医科大学心脏外科学系副主任;中国医师协会心血管外科医师分会委员;2007年作为医学访问学者至美国宾夕法尼亚大学心脏中心学习半年。

先后负责或参加“九五”重点攻关课题、国家杰出青年基金科研课题、国家自然科学基金、北京市自然科学基金、北京市教委及科委多项科研项目。国内外发表论文20余篇。作为主要完成人曾获国家科学技术进步二等奖、北京市科学技术进步二、三等奖及中华医学会二等奖,2005年入选北京市卫生局“十百千杰出人才基金支助”,2010年获“全国优秀科技工作者”称号,2011年入选教育部“新世纪优秀人才支持计划”。



肖志波

哈尔滨医科大学整形外科教授、主任医师、整形外科副主任;兼任美国贝勒医学院外科学部 Research associates、韩国延世大学 Severance hospital 访问学者;中华整形外科学会黑龙江分会委员兼秘书;中国中西医结合学会整形专业学会青年委员;哈尔滨医科大学杰出青年拔尖人才。从事教学工作16年,作为第一作者或通讯作者发表SCI论文20余篇,主持国家自然科学基金、中国博士后基金及黑龙江省自然科学基金等多项课题,获中华医学科技奖1项、黑龙江省医药卫生科技进步奖2项、黑龙江省医疗新技术应用奖3项。



# 前 言

医学科研方法是临床医师开展临床研究的基础。临床医师在临床实践中,一方面需要借鉴他人的经验,应用证据解决临床问题;另一方面也需要总结自己的临床实践经验,应用医学科研方法开展临床研究,提高处理疑难、复杂病例的水平。住院医师从医学院走进临床实践,从规范化培养的第一天开始,不仅要具备诊断思维与循证实践的能力,还需要在临床实践中不断提高发现问题、分析问题、处理问题的能力,这就需要掌握医学科研方法,尤其是临床研究方法。

与以往教材或专著不同,针对住院医师规范化培训群体,全书通过医学研究案例剖析,介绍研究设计、测量、结果报告等,分析研究的优点、缺点,包括设计、测量、报告中的问题,让住院医师充分了解医学研究的基本过程。重点掌握临床研究基本设计类型,包括病例报告、系列病例分析、横断面调查、病例对照研究、队列研究、随机对照临床试验等。熟悉临床研究测量方法,生命质量评价方法、系统评价与 Meta 分析方法。了解卫生经济评价、决策分析、比较效果研究,了解临床研究统计分析原理。通过学习,在住院医师培训期间,在教师指导下,能独立完成一项基本临床研究设计,并通过具体方案实施获得数据,分析处理数据并最终完成论文书写与报告。

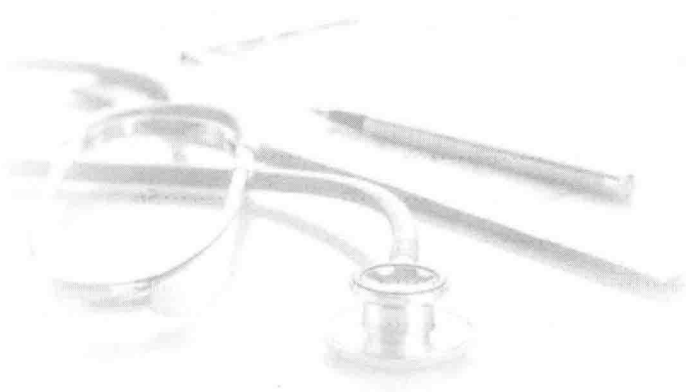
该书尽管面向住院医师规范化培训编写,但我们更期望对各级临床医师、研究生在开展临床研究或者教学时有益,或者作为临床研究方法学的案例教材。

感谢复旦大学医学院李莹硕士在编写过程中与各位编委的沟通与安排,李冰博士在最终稿件整理所做的工作。在成书过程中还得到各编委单位的领导和同仁的多方支持与帮助,上海交通大学附属精神卫生中心李伟,西安交通大学第二附属医院刘璐,北京协和医院王丽、朱慧娟,复旦大学附属中山医院王妍,首都医科大学附属北京安贞医院刘静,第二军医大学张新佑等在本书的编写过程中也付出了很多,在此一并表示由衷的感谢。

书稿完成后又经过反复推敲,同时也得到中华医学会临床流行病学专科分会的很多专家的指导,但由于我们的经验有限,书中仍有不妥与疏漏之处,期盼同行专家和读者们批评、指正。

陈世耀

2015 年 5 月





# 目 录

|                       |    |
|-----------------------|----|
| 第一章 医学科研方法总论          | 1  |
| 第一节 医学科研的基本概念         | 1  |
| 第二节 临床研究的重要性          | 2  |
| 第三节 我国临床研究存在的问题       | 2  |
| 第四节 临床研究的方法学          | 2  |
| 第五节 临床研究的伦理问题         | 6  |
| 第六节 临床研究注册            | 6  |
| 第七节 学习临床研究方法,开展临床研究实践 | 6  |
| 第八节 如何阅读本书            | 8  |
| 第二章 文献综述              | 9  |
| 第一节 文献综述的基本概念         | 9  |
| 第二节 撰写文献综述的方法与步骤      | 10 |
| 第三节 文献综述的问题及解决方案      | 14 |
| 第三章 系统评价              | 16 |
| 第一节 提出临床研究问题          | 16 |
| 第二节 系统评价的制作步骤         | 19 |
| 第三节 解读系统评价与 Meta 分析   | 24 |
| 第四节 文献阅读与评价           | 26 |
| 第四章 病例报告和系列病例分析       | 35 |
| 第一节 病例报告              | 35 |
| 第二节 系列病例分析            | 37 |
| 第三节 循证实践案例报告          | 39 |
| 第五章 横断面调查研究           | 44 |
| 第一节 如何获得临床研究问题        | 44 |
| 第二节 抽样调查研究设计与实施       | 47 |
| 第三节 研究报告与问题分析         | 49 |
| 第四节 临床问题的回答           | 51 |
| 第五节 横断面研究的其他应用        | 51 |
| 第六节 文献阅读与评价           | 51 |
| 第六章 病例对照研究            | 54 |
| 第一节 如何获得临床研究问题        | 54 |
| 第二节 病例对照研究设计          | 56 |
| 第三节 研究报告与问题分析         | 61 |
| 第四节 临床问题的回答           | 63 |
| 第五节 病例对照研究的其他应用       | 63 |



|                                   |     |
|-----------------------------------|-----|
| 第六节 文献阅读与评价·····                  | 64  |
| 第七章 队列研究·····                     | 66  |
| 第一节 将临床问题转化为科学问题·····             | 66  |
| 第二节 队列研究设计·····                   | 68  |
| 第三节 研究报告与问题分析·····                | 74  |
| 第四节 临床问题的回答·····                  | 77  |
| 第五节 文献阅读与评价·····                  | 77  |
| 第八章 随机对照临床试验·····                 | 80  |
| 第一节 如何获得临床研究问题·····               | 80  |
| 第二节 随机对照临床试验设计·····               | 82  |
| 第三节 研究报告与问题分析·····                | 89  |
| 第四节 临床问题的回答·····                  | 90  |
| 第五节 随机对照临床试验的其他应用·····            | 91  |
| 第九章 成本效果分析·····                   | 92  |
| 第一节 如何获得临床研究问题·····               | 92  |
| 第二节 成本-效果分析研究设计·····              | 94  |
| 第三节 临床案例结果描述与临床选择·····            | 96  |
| 第四节 临床经济学文献评价标准·····              | 97  |
| 第十章 比较效果研究·····                   | 98  |
| 第一节 如何获得临床研究问题·····               | 98  |
| 第二节 如何利用临床实际服务的二手数据开展 CER 研究····· | 100 |
| 第三节 比较效果研究中混杂因素存在的判定·····         | 102 |
| 第四节 比较效果研究中混杂变量的控制·····           | 103 |
| 第十一章 病因与危险因素研究·····               | 107 |
| 第一节 如何获得临床研究问题·····               | 107 |
| 第二节 临床问题转化为科学问题·····              | 108 |
| 第三节 病因学研究的模型与工具·····              | 110 |
| 第四节 病因与危险因素研究的基本过程·····           | 112 |
| 第五节 判断因果关系的原则·····                | 112 |
| 第六节 选择具体的研究方法·····                | 113 |
| 第七节 计算样本量·····                    | 116 |
| 第十二章 诊断性试验评价研究·····               | 118 |
| 第一节 临床研究问题的提出·····                | 118 |
| 第二节 诊断试验研究的设计思路·····              | 119 |
| 第三节 诊断试验研究中正常参考界值的选择方法·····       | 120 |
| 第四节 诊断试验研究的方法和评价指标·····           | 120 |
| 第五节 诊断试验的可靠性评价·····               | 121 |
| 第六节 诊断试验的评价原则·····                | 122 |
| 第十三章 临床疗效评价·····                  | 123 |
| 第一节 如何获得临床研究问题·····               | 123 |



|                           |     |
|---------------------------|-----|
| 第二节 临床疗效评价的基本研究方案·····    | 125 |
| 第三节 临床疗效评价设计(随机对照试验)····· | 126 |
| 第四节 研究报告与问题分析·····        | 128 |
| 第五节 文献阅读与评价·····          | 129 |
| 第十四章 预后研究·····            | 131 |
| 第一节 预后研究的基本概念及步骤·····     | 131 |
| 第二节 预后研究常用的临床研究设计·····    | 132 |
| 第三节 预后研究的队列设计·····        | 135 |
| 第四节 研究报告与问题分析·····        | 136 |
| 第五节 预后研究证据的评价·····        | 136 |
| 第六节 临床问题的回答·····          | 138 |
| 第七节 文献阅读与评价·····          | 138 |
| 第十五章 如何撰写研究计划书·····       | 140 |
| 第一节 撰写研究计划书前的准备工作·····    | 140 |
| 第二节 临床研究计划书的撰写·····       | 142 |
| 第十六章 数据收集、数据管理与统计分析·····  | 149 |
| 第一节 数据收集·····             | 149 |
| 第二节 数据管理·····             | 153 |
| 第三节 统计分析·····             | 159 |
| 第十七章 论文撰写、投稿与修改·····      | 169 |
| 第一节 医学论文的写作技巧·····        | 169 |
| 第二节 如何选择适合的期刊进行投稿·····    | 170 |
| 第三节 论文的修稿·····            | 170 |
| 参考文献·····                 | 172 |
| 附录·····                   | 175 |
| 中英文名词对照索引·····            | 176 |

# 第一章 医学科研方法总论

临床医师在临床实践中每天面对的是患者和疾病,需要不断回答患者的诊断问题、治疗问题与预后问题。阅读文献,学习和借鉴他人的经验是成长过程的一个方面,总结自己的临床实践经验是成长过程的另一个方面。医学的不断进步就是在从实践到理论,再从理论到实践的探索与应用的过程中。开展医学科学研究是每一位临床医师,尤其是住院医师不断提高自己的重要过程。

## 第一节 医学科研的基本概念

医学科学研究是在医学专业理论的指导下,以正确的观点和方法,围绕人类身心健康,探索医学领域的未知或未全知的事物或现象的本质及规律的一种认识和实践。

### 一、医学科研的类型

医学研究包括基础医学、临床医学、预防医学和卫生事业管理学研究。研究对象包括正常人、患者、动物(实验动物)和生物体赖以生存的自然和社会环境。研究类型分为:①调查研究:研究者为了了解人群的健康状况(疾病的分布、患病率、发病率、病死率和死亡率的水平和消长),研究环境因素的致病或保护作用,开展的调查设计。调查设计的特点即研究因素是客观存在的,如职业、地域、民族等;不能用随机化分组来平衡混杂因素对调查结果的影响,故重点是调查表与抽样方法设计。②临床观察:临床医学观察研究的内容很广泛,包括了病因学、诊断学、疗效和预后诸领域的研究。例如,疗效观察研究内容既有药物、手术、理化因素的效应,也有营养、护理等辅助措施与预防措施的作用,也可以是对完整的一组治疗方案或一种特定形式的治疗措施的评价。疗效研究的指标,可以是生存或死亡,痊愈或未愈,有效或无效,症状或体征的存在或消失,生理、生化指标的变化及副作用等。疗效观察研究同时要考虑设计的类型,受试对象的代表性及其诊断的正确性,评价指标的统一性和可靠性。临床观察性研究的特点是研究者不对观察对象实施主动干预。③实验研究:将若干随机抽取的实验对象随机分配到两个或多个处理组,观察比较不同处理因素的效应,这种研究称为实验研究。实验研究的特点是:研究者能人为设置处理因素;研究对象接受处理因素的种类或水平是由随机分配决定的。因此,实验研究能够更有效地控制误差,使多种实验因素包括在较少次数的实验之中。广义的实验研究包括动物实验、临床试验和社区干预试验。

#### 知识点

##### 医学科研特征

1. 医学研究三层次 ①群体水平;②器官组织水平;③细胞分子水平。
2. 医学研究对象 ①人,包括正常人和患者;②离体组织细胞;③动物。
3. 医学研究的方法 ①观察法;②实验法;③理论法。
4. 医学研究的场所 ①社区;②医院;③实验室。
5. 医学研究的三个基本环节 ①设计;②测量;③评价。



## 二、医学科研的基本步骤

1. 科研选题 确定所要探索的题目。选题是科研工作的起点,是科研工作基本内容和目标的高度概括。
2. 科研设计 运用专业理论知识和实验技术安排有关的科研顺序和步骤,为即将进行的科研实践制定出一个科学而合理的研究方案。
3. 基金申请 获取科研资助的过程。各类科研基金立项的主旨不同,资助的研究类型也不同。填写科研课题申请书是申请课题的一个途径。
4. 研究实施 医学科研实施方法按研究对象的属性和实验场所不同,一般可分为调查、观察和实验三大类型。这个阶段主要是资料的收集。
5. 总结归纳 根据研究数据与统计分析结果,运用综合、归纳和演绎等方法把感性认识上升为理性概念,得出科研结论。基本形式是撰写科研论文、课题鉴定总结、形成科研成果。

## 第二节 临床研究的重要性

临床研究以患者为主要研究对象,以疾病的诊断、治疗、预后、病因为主要研究内容,以医疗服务机构为主要研究基地,由多学科人员共同参与组织实施的研究工作。目的在于探索疾病发生、发展及转归规律,研究和创造新的疾病诊疗方法,提高临床治疗水平。临床研究为基础医学研究提供信息和需求,促进基础医学的可持续发展。临床研究有别于基础研究,高质量临床研究为临床决策提供循证医学的最佳证据。

国家对临床研究的重视与投入为未来我们开展临床研究提供了机遇。根据国家整体战略部署,国家临床医学研究中心建设工作列入“十二五”规划。为有效解决中国医学科技整体投入水平相对较低、尤其临床研究方面薄弱这一问题提供机遇。国家临床医学研究中心将承担起提出攻关方向、搭建公共平台、培育专业人才、优化管理机制、搭建协同网络、组织开展研究、推动转化应用、强化基层能力等任务。

《医学科技发展“十二五”规划》将建设一批临床/转化医学研究中心,打造一批跨学科、跨地域的专科临床协同研究网络体系,通过优化探索临床研究的组织模式,打造更加高效的临床转化平台,构建一个适合中国国情的医学技术研究普及平台。

## 第三节 我国临床研究存在的问题

我国临床研究质量相对较低,近10年(1998~2008年)《中华医学杂志》《中华内科杂志》《中华外科杂志》《中华妇产科杂志》《中华儿科杂志》5个期刊中收录与临床研究相关论文1721篇,其中前瞻性随机对照研究仅占11.5%,绝大多数(84.8%)为回顾性病例分析研究。国外同期10年在4个主要学术期刊 *New England Journal of Medicine*, *JAMA*, *Lancet* 和 *Canadian Medical Association Journal* 上发表的临床研究论文6353篇,其中44.5%为随机对照研究;尤为需要引起重视的是这6353篇临床研究论文中来自中国的仅占0.27%。因此,临床医师需重视掌握实施临床研究的方法,提高临床研究的质量。

## 第四节 临床研究的方法学

临床研究与基础研究最大的区别是研究条件受临床工作平台限制。临床研究的步骤包





括发现临床问题,形成研究假设,设计研究方案,实施研究项目,分析研究数据,合理解释结果,最终为了解决临床问题。根据研究目的不同,临床研究可以分为病因与危险因素研究,诊断试验评价,预防与治疗效果评价,预后及预后因素研究。临床研究的评价主要包括结果的科学性评价,结果大小的表达,结果是否可推广应用于当前患者。合理的临床研究设计,正确的研究实施与过程管理,科学评价临床研究结果是保证实施高质量临床研究的基本原则。

## 一、合理的临床研究设计

一个好的临床问题胜过无数平庸的研究,临床问题的提出多来源于日常临床医疗实践,要有创造性、先进性和科学性,要有临床意义和临床价值,并具有可行性。临床问题的提出需遵循 PICO 原则,P 指特定的患病人群 (population/participants);I 指干预或者暴露 (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)。例如:用抗凝剂和不用抗凝剂相比能否降低缺血性脑卒中患者远期死亡或残疾的风险。

根据不同的临床问题,其基本分类包括:病因与危险因素研究,诊断试验评价,预防与治疗效果评价,预后及预后因素研究。不同的临床问题需要不同的研究设计(表 1-1)。

表 1-1 临床问题与最佳研究设计

| 临床问题                 | 最佳研究设计      |
|----------------------|-------------|
| 疗效评价                 | RCT(随机对照研究) |
| 治疗的不良反应              | RCT(随机对照研究) |
| 诊断或筛查试验              | 与金标准进行盲法比较  |
| 预后评价                 | 队列研究        |
| 无法进行 RCT 或有伦理问题的疗效评价 | 队列研究        |
| 暴露不良环境的危害            | 病例对照研究      |

临床研究的设计类型有很多,归纳起来分为两大类:观察性研究 (observational study) 和试验性研究,后者即临床试验 (clinical study)。试验性研究设计者可人为地控制条件,随机分组,有目的的设置各种对照,从而探讨某研究因素与疾病的联系。因此,试验性研究的论证强度较高,结论比较可靠。常用的临床试验包括随机对照试验 (randomized controlled trial)、非随机同期对照试验 (non-randomized concurrent control studies)、历史性对照试验 (historical control studies)、前后对照试验 (before-after study)、序贯试验等。观察性研究和实验性研究最主要的区别是前者不能由研究者人为控制试验条件,只能尽量控制非研究因素的影响,以求得研究结论的真实性。因此,研究结论的论证强度不及试验性研究。临床上常用的观察性研究根据有无对照分为描述性研究和分析性研究,描述性研究包括病例报告、横断面研究等。分析性研究包括病例对照研究 (case-control study)、队列研究 (cohort study) (前瞻性和回顾性),其设立对照组进行比较性研究,进一步分析和推论、模拟试验性研究,其论证强度较观察性研究为强。临床研究分类图解(图 1-1)。2013 年 1~6 月不同学科等级杂志临床研究类型分布(图 1-2~图 1-4)。

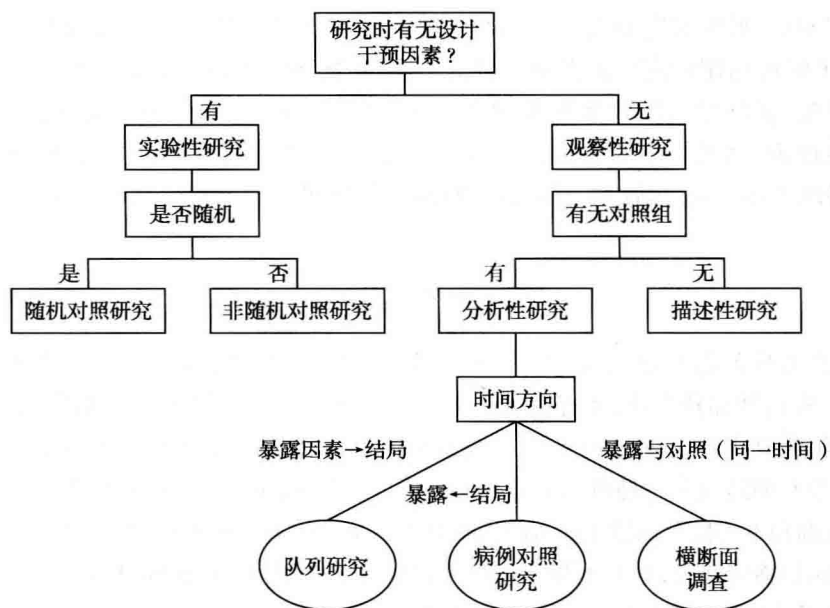


图 1-1 临床研究分类

临床研究论文数量 (n=103)

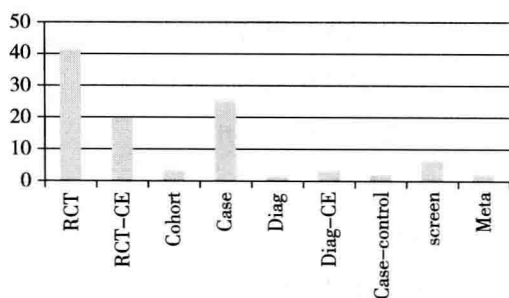


图 1-2 临床医学(一级学科)NEJM 中临床研究设计的基本类型

临床研究论文数量 (n=39)

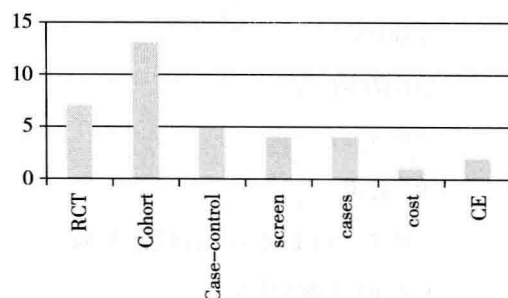


图 1-3 内科专业(二级学科)Ann Int Med 中临床研究设计的基本类型

临床研究论文数量 (n=49)

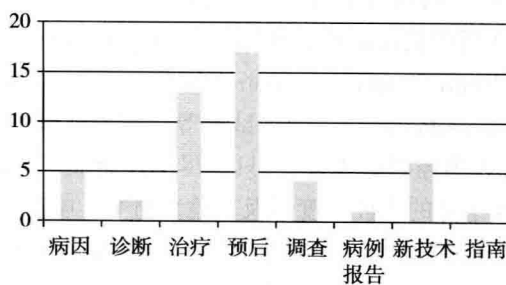


图 1-4 消化专业(三级学科)期刊 Gastroenterology 中临床研究设计的基本类型

## 二、正确的研究实施与过程管理

临床医师在临床诊疗过程中,不仅要善于发现问题,提出问题,尚需重视各类病例资料的收集,注意收集病例的相关数据,如病理、影像、超声、基因多态性等各项临床和实验室数据。临床研究的开展基于这些完整数据的保存。



### 三、科学评价与临床应用(代表性与应用范围)

临床研究质量评价包括证据的科学性、证据的重要性和证据的实用性。

科学性评价内容包括:有关研究的设计是否科学,提供的证据是否可靠;有效性指这些结果能否真正确定患病的危险因素,是否对疾病的发生、发展起到防治作用;实用性指结果能否用于自己的患者。

#### 知识点

##### 临床研究评价内容

1. 科学性评价 根据不同临床研究设计的评价标准进行。
2. 结果大小 包括灵敏度与特异度,疗效,生存时间,生命质量,成本效果。
3. 推广应用 人群代表性。

证据评价依据证据类型不同,有不同的评价标准。不仅临床指南、系统评价、各种原始临床研究有不同的评价标准,即使相同类型的临床研究,也根据用途不同,标准不同。

#### 知识点

##### 疗效比较研究标准

1. Jadad 评分量表 一般系统综述时文献质量评价标准。包括随机序列的产生,随机化隐藏,盲法,退出与失访。

2. Delphi 清单 为广泛认可的治疗性研究质量评价标准。包括:①是否采用了随机分组,分组过程是否应用了盲法;②各组严重影响结果的基线资料是否一致;③纳入排除标准是否明确;④结果测量是否应用盲法;⑤干预措施执行者是否实施盲法;⑥受试对象是否实施盲法;⑦主要结果的点估计及估计的精确性;⑧分析中是否包含了意向性分析。

3. Cochrane Reviewer's Handbook 推荐的偏倚风险评估工具 用于 Cochrane 系统综述评价标准。根据随机序列、隐藏分组、盲法、结果数据不完整、选择性报告结果及其他偏倚六项绘制出风险偏倚评估图。

4. CONSORT 声明 论文发表时的报告标准。PRISMA 声明包括流程图和清单(27 项)两部分内容,全文可以在网站 <http://www.prisma-statement.org> 上免费获得。

在实施某项预防性治疗措施前,应该将治疗措施的必要性、预期效果、风险、费用,以及如果不采取该措施,是否有其他替代措施,其相关效果、风险及费用等信息充分提供给患者,让患者或者医患双方共同作出决定。临床应用时还需要注意以下几点:

(1) 三个正确:使用证据时应在正确的时间、正确的地点使用正确的方法。

(2) 研究与实际的差异:查获的研究证据往往是较好的医生在较好的医院完成的。不同地区、不同医院的就诊患者具体情况及疾病谱存在很大差异。临床研究中,对于服药、治疗、检查等多个环节的环节都很关注,受试者具体情况及依从性也与实际情况存在差异。因此研究时可能会出现较理想的临床结果,而实际应用中,往往很多地方都会打折扣,最终效果可能也与研究结果不一致。

(3) 医师应该关注每个患者的具体情况:即每个患者本身的健康状况、社会经济状况、心理状况都不一样,可能会影响治疗措施的效果。

(4) 患者的愿望:医师应该增进医患沟通,了解并尊重患者的愿望。如果医生认为某方法效果确切,而患者顾虑重重,则医生利用所掌握的证据去说服患者接受治疗,这既是一门艺术,也体现了医师的责任感。



(5) 患者的依从性:依从治疗计划对疗效好坏十分关键。因此建立有效的方案鼓励患者遵照处理计划十分重要,可以根据具体情况开展各种健康教育,提供依从性。

## 第五节 临床研究的伦理问题

临床研究涉及患者,需要获得伦理委员会同意,即使回顾性研究,只要涉及人或者疾病人群信息、血液学及组织学标本等,都需要考虑伦理问题。《纽伦堡法典》和《赫尔辛基宣言》是伦理委员会同意的基本原则。

《纽伦堡法典》基本内容包括:①进行人体实验必须获得受试者知情和同意;②必须保证实验是对社会有益,又是非做不可;③人体实验前必须经过动物实验;④要避免给受试者造成精神或肉体上的痛苦及创伤;⑤实验的危险性不得高于人道主义的重要性;⑥不准进行有可能造成受试者死亡或残废的实验;⑦精心安排,采取一切必要的措施杜绝发生受试者伤残;⑧实验必须由受过科学训练的人来进行;⑨实验期间受试者有权停止实验;⑩实验过程中,发现有可能造成受试者伤残或死亡时,应立即停止实验。

《赫尔辛基宣言》的基本原则包括:①以人作为受试者的生物医学研究目的,必须旨在维护人体健康,用以增进诊断、治疗和预防等方面的措施以及为了针对疾病病因学和发病机制的了解;②实验必须符合普遍认可的原理,应该建立在足以胜任地履行实验室任务和动物实验法的基础之上。实验程序的设计和行动,应该作系统科学的说明,并得到有关专门委员会的评议和指导;③实验研究工作只有通过曾受过严格训练的有资格的人们和在临床上一个被认可的医生监督下,才可进行;④实验前应细致比较可以预见到的利益。对于受试者利害关系的重要性,一定要始终压倒对科学研究和人类社会方面的影响;⑤实验室工作正义性要始终服从于保护受试者身体与精神上的完整。在实验研究中应该特别注意受试者的外界环境和生活福利不受到影响。同时要采取一切预防措施,使实验对受试者的影响和冲击减少到最低程度;⑥实验过程中如果确已查明或者发觉有可能出现危险,该项实验就应停止进行;⑦医药卫生工作者对于保证研究成果的准确性负有极大的责任。违背宣言基本原则的实验报道,将不被接受发表;⑧无论哪项科学研究,对于每个可能的被实验者,这项研究的目标、方法、预期好处、潜在危险以及可能承担的不舒适与困难等,都必须足够充分的被告知,必须得到受试者自愿参加该项实验的承诺。受试者应该了解他们有权不参加这个研究,而且任何时候都可以撤销他们的承诺。

## 第六节 临床研究注册

开展临床研究,在完成了研究设计并获得伦理委员会批准同意之后,可以通过临床研究注册网站([www.clinicaltrials.gov](http://www.clinicaltrials.gov))进行临床研究注册。一方面可以规范临床研究,获得更多发表机会。另一方面,还可以通过注册网站了解他人开展的相关临床研究,取长补短。

## 第七节 学习临床研究方法,开展临床研究实践

### 一、在临床实践中发现研究问题

#### 临床案例

一位 55 岁的男性,临床诊断为恶性淋巴瘤,患者需要接受化疗,医生除了血常规、肝肾功

能等化验外还给他开出了 HBsAg、HBsAb、HBeAg、HBeAb 和 HBcAb 等项目的检查,患者想起



他有 HBsAg 阳性病史,担心是否会影响化疗,他向床位医生询问,想详细了解对他这样的情况如何更有利。化疗对乙肝携带是否有影响?或者乙肝是否会激活?是否需要药物预防?

预防相关的临床问题:与治疗证据不同,疾病的预防是疾病尚未发生,通过一定的处理措施,可以阻止或者减少疾病的发生。临床上通常有如下一些情况:①个体是某种疾病的高危人群,通过干预某种危险因素减少疾病的发生。如有胃癌家族史的个体,存在幽门螺杆菌感染这一高危因素,是否能通过根除幽门螺杆菌减少胃癌的发生风险(根除与不根除比较);②患者可能存在某些危险的并发症,通过干预减少并发症的发生。如肝硬化患者,可能出现门脉高压食管胃静脉曲张破裂出血,是否能通过口服普萘洛尔降低门脉压力或者内镜下套扎治疗预防首次出血的风险(预防与不预防比较),是药物预防好还是内镜下治疗好(两种措施的比较);③患者由于某种疾病需要接受特殊治疗,但这种治疗可能导致处理相关的问题,通过提前干预减少治疗相关的问题发生的可能性。如冠心病患者,服用阿司匹林和抗血小板药物治疗,可能出现胃黏膜损伤相关的消化道出血,通过预防性应用抑酸药物可以减少这一风险。本案例系淋巴瘤患者,由于合并乙肝携带状态,在进行化疗时,可能存在乙肝激活就属于这一情况。

## 二、将临床问题转换为科学问题或者研究问题

当需要了解某一疾病的进展或者处理方法,如对局部进展期胃癌的处理时,可以通过阅读一般综述。一般综述通常能提供针对该患者一般问题的更多信息,包括:①如何明确诊断并选择合适的方法(如增强 CT)进行正确的分期;②有哪些处理的办法(如直接手术治疗、术前辅助化疗+手术治疗、术前放疗等);③不同处理方法的优点与缺点是什么。

当我们需要为患者选择具体处理时,系统综述或者 Meta 分析能为我们提供针对性证据。特定的问题是:针对局部进展期胃癌患者,与单纯手术治疗相比,联合围术期化疗是否能带来更多获益即延长生存时间。PICOS 原则,包括:①患者和(或)疾病状况(population, patient and/or problem):局部进展期胃癌患者的处理问题;②治疗干预(intervention),有时候是暴露(exposure):联合术前化疗;③比较(comparison):与单纯手术治疗比较;④临床的结局(outcome):手术切除率,1~5 年生存率;⑤研究类型(study):可以查找的文献或者开展的研究包括 RCT 及 Meta 分析。

针对上述案例的问题,依据 PICOS 模式,P 为恶性淋巴瘤拟行化疗并乙肝携带患者,I 为抗乙肝病毒干预,C 为不给予抗乙肝病毒预防,O 为乙肝是否激活,表现为肝功能损伤以及 HBV-DNA 复制。因此,采用随机对照临床研究来回答这一问题就是最佳研究设计选择。

## 三、利用基础或者实验数据开展临床研究

临床研究中,无论是分组的暴露因素或者结果观察指标,都离不开实验室检查或者辅助检查这些客观资料,这些实验室数据或者特殊检查数据是临床研究数据的质量保证,也是基础研究与临床研究结合的重要内容。包括基因多态性在内的分子生物学检测指标在临床研究中的采纳与应用为临床研究增加了深度,为临床应用提供了个体化决策依据。

## 四、更多更新的临床研究设计与规范

多中心临床注册研究(multicenter clinical registries)是近年来国内外学术界兴起的一种新的临床研究形式,采用多目标、多任务、多中心研究的设计方案,与目前单一目标的临床研究方案完全不同,有很强的挑战性。临床注册研究的本质是利用信息科学技术将多家医院中具有某些同质性特征的病例资料整合起来,形成一个虚拟的多中心研究平台。临床注册研究得益于信息科学技术的发展,其出现为多中心临床研究提供了一种新的形式,为拓展临床研究的空间、时间





和范围提供了新的机遇。

除临床研究评估疗效与安全性之外,卫生经济评价、生命质量评价、比较效果研究等不仅提供包括花费成本之后获得的临床疗效和安全性、生命质量以及社会结果,同时可以获得多种措施之间的比较、真实临床实践状态下的比较、全体而非特定人群的比较、长期结果而非中间或者替代结果的比较。全新的临床研究类型为临床决策提供更多、更新并且更加客观的证据。

(1) 卫生经济评价:同时比较两种或者两种以上的措施,在资源消耗之后获得的效益或者效用差别。包括最小成本分析、成本效果分析、成本效用分析和成本效益分析。

(2) 比较效果研究:在真实临床情况下,对不同干预措施或干预策略(包括药物、手术或其他诊疗方法)在预防、诊断、治疗和疾病监控方面的利与弊进行比较,对实效资料进行系统评价和整合,总结证据。表 1-2 显示了比较效果研究与 RCT 研究特征的差异。

表 1-2 理论疗效研究(随机对照临床试验)和实际疗效研究的特征

|      | 理论疗效研究<br>(效率, efficacy) | 实际疗效研究<br>(效果, effectiveness)  |
|------|--------------------------|--------------------------------|
| 研究状态 | 临床试验—理想条件下               | 临床实践—实际条件下                     |
| 比较组  | 治疗组 vs. 安慰剂组             | 多种治疗选择, 互相比较                   |
| 研究人群 | 患单一疾病的患者                 | 全体, 包括合并症、联合用药等                |
| 应用人群 | 实际使用人群可能会被排除             | 一般不对研究对象做任何限制                  |
| 结果测量 | 通过测量替代终点、生物标志等评估短期效果     | 通过测量临床状况改善、生活质量、残疾、死亡等指标评估长期效果 |

学  
习  
笔  
记

除多中心研究之外,多学科合作研究越来越受到关注。对糖尿病治疗的关注,从内分泌科的药物治疗,到消化内科的内镜下胃内水囊治疗、小肠短路支架放置治疗,再到外科腹腔镜胃带治疗和胃肠短路手术治疗,甚至用粪便移植疗法探索治疗肥胖和糖尿病。而针对粪便移植治疗的手段也进入从分离粪便到粪便银行再到粪便胶囊的探索,再次回到内分泌科的药物治疗上来,只是彻底改变了药物成分,所有这些都体现了多学科合作的贡献。进入临床应用,仍然需要多学科合作,为不同患者选择合适的治疗手段或者策略。

## 第八节 如何阅读本书

临床医师更需要掌握临床研究方法,一方面在临床实践中阅读和评价他人的临床研究结果,为患者临床决策提供循证医学证据;另一方面总结和交流自己的临床实践经验,为更多患者提供帮助。

《医学科研方法》是一本临床医师尤其住院医师从事临床科研,开展临床研究的工具书。住院医师在学习临床研究案例中,可以模拟各类临床研究设计与过程,更好更快掌握临床研究方法。在学习阅读本书的同时,还可以结合临床面临的处理矛盾或者问题,阅读文献,不断思考创新问题。在阅读文献时,更需要客观科学地评价文献,思考研究设计方法。

全书基本写作格式是:介绍一个临床案例,提出临床问题;通过 PICO 模式将临床问题转化为可以通过研究回答的科学问题;进一步提出了临床研究设计类型及其优缺点。通过对不同类型临床研究论文的介绍和评价,让读者更熟悉临床研究设计,并在这一过程中介绍临床研究的基本概念,如暴露的测量、对照的选择、盲法的实施、随机化方法、偏移的控制与估计等。鼓励读者检索文献阅读部分的全文文献,按照思考题提供的思路,还原临床研究设计。

(陈世耀)

## 第二章 文献综述

### 临床案例

患儿男性,5岁,因反复蛋白尿1月余来院就诊。患儿1个月前无明显诱因出现眼睑水肿,同时伴有尿量减少。在本院住院治疗,24小时尿蛋白定量为58mg/kg,血清白蛋白25g/L,诊断为原发性肾病综合征。于4周前给予泼尼松2mg/(kg·d),24小时尿蛋白定量无明显好转。入院查体:血压85/65mmHg,眼睑无明显水肿,咽部无充血,双肺呼吸音清,心律齐,心音有力,腹软,肝脾肋下未及,神经系统检查正常。

### 第一节 文献综述的基本概念

文献综述,是在确定了选题后,在对选题所涉及的研究领域的文献进行广泛阅读和理解的基础上,对该研究领域的研究现状(包括主要学术观点、前人研究成果和研究水平、争论焦点、存在的问题及可能的原因等)、新水平、新动态、新技术和新发现、发展前景等内容进行综合分析、归纳整理和评论,并提出自己的见解和研究思路。“综”即收集“百家”之言,综合分析整理,“述”是对文献的观点和结论进行叙述和评论。文献综述既是医学科学研究的重要基础,也是研究成果之一,做好文献综述是进行学术研究的基本前提。

文献综述一般分成两类,即基本文献综述和高级文献综述两种。基本文献综述是较为宏观的,涉及的范围为整个领域、专业或某一大的研究方向,是对有关研究课题的现有知识进行总结和评价,以陈述现有知识的状况。高级文献综述则是较为微观的,这类综述可以涉及相当小的研究方向,谈的问题更为具体与深入,是在选择研究兴趣和主题之后,对相关文献进行回顾,确立研究论题,再提出进一步的研究,从而建立一个研究项目。前者立意高,范围广,易读易懂但不易深入,而后者则与此相反。虽然基本文献综述和高级文献综述所包含的任务有所不同,但是它们呈现知识和提出论题的方式是相似的。

文献综述在于高度浓缩了几十篇甚至上百篇散乱无序的同类文献之成果与存在问题或争论焦点,对其进行了归纳整理,使之达到了条理化和系统化的程度。其意义在于:①综述前人的研究成果,使读者熟悉现有研究领域中相关研究的进展和问题;②了解前人的研究工作,发现研究空白,确立自己的研究方向,从而避免简单重复;③改进与批判现有研究的不足,推出另类研究,发掘新的研究方法途径,验证其他相关研究。

#### 知识点

##### 文献综述与其他文体的区别

文献综述与“读书报告”、“文献复习”、“研究进展”等有相似的地方,它们都是从某一方面的专题研究论文或报告中归纳出来的。但是,文献综述既不像“读书报告”、“文献复习”那样,单纯把一级文献客观地归纳报告,也不像“研究进展”那样只讲科学进程,其特点是“综”,“综”是要求对文献资料进行综合分析、归纳整理,使材料更精练明确、更有逻辑



辑层次;“述”就是要求对综合整理后的文献进行比较专门的、全面的、深入的、系统的论述。文献综述是作者对某一方面问题的历史背景、前人工作、争论焦点、研究现状和发展前景等内容进行评论的科学性论文。

## 第二节 撰写文献综述的方法与步骤

写文献综述一般包括以下六步:选择主题、文献检索、展开论证、文献研究、文献批评和综述撰写。在前述的临床问题中,该患儿确诊为肾病综合征,给予糖皮质激素泼尼松治疗,泼尼松 $2\text{mg}/(\text{kg}\cdot\text{d})$ 治疗4周无效(24小时尿蛋白定量无好转),表现为激素耐药。临床上有哪些原因可导致糖皮质激素耐药?针对此类患者需要进行哪些检查?如何选择后续的治疗方案?下面以《Mechanisms of glucocorticoid resistance in idiopathic nephrotic syndrome》(Kidney Blood Press Res, 2013,37:360-378)为例(以下简称 Mechanism),通过综述原发性肾病综合征中出现激素耐药的机制来具体讲述如何进行文献综述。

该文章类型是综述,因此,在文章中要对肾病综合征激素耐药这一研究领域进行分析,描述前人已经做了哪些工作,进展到何程度,要对国内外相关研究的动态、前沿性问题做出较详细的分析,并提供参考文献。全篇以客观概括地反映事实为主,同时适当加入自己的评论和建议,少量但切中要害。

### 一、选择主题

选择主题前首先要思考临床上有什么问题需要解决或值得深入研究,也即“提出问题”。理论问题通常是隐藏在纷繁复杂的现实背后,而发现理论问题,则需要运用理论思维的能力。理论思维的训练是一个长期积累的过程。不过初学者也不必望而却步,大体上可以分“三步走”:第一步,先划定一个“兴趣范围”,可以结合自己从事的专业,如肾内科专业的可选择激素耐药性肾病综合征的治疗,激素能否预防过敏性紫癜肾损害的发生等,找到其中的“症结”或“热点”。第二步,总结以往的研究者大体从哪些理论视角来分析“症结”或“热点”、运用了哪些理论工具,如有研究者称C3基因多态性与过敏性紫癜肾损害的发生有关,可以提示基因多态性与疾病相关的研究方向。第三步,考察问题的可研究性,也就是研究的空间和可行性。最后,要进行的是凝练科学问题的过程,科学问题应该是来自于临床,服务于临床,而不是为了套理论而“削足适履”。在确定研究方向后,我们在筛选信息和资料时也就有了明确的标准,在这个信息冗余的时代,能够大大提高研究效率。以“Mechanism”为例,是综述原发性肾病综合征糖皮质激素耐药的机制,用“idiopathic”来限制了文章只综述原发性肾病综合征,而继发性肾病综合征则不作为综述对象。用“glucocorticoid”限制治疗药物只限于糖皮质激素,以免主体太大。

#### 知识点

#### 选题的原则与技巧

撰写文献综述通常出于某种需要,如学术会议的专题、科研课题的选题、文献资料积累等,因此,文献综述的选题首先需要明确。文献综述选题范围可大可小,大到一个领域、一个学科,小到一种疾病、一个方法、一个理论,可根据需要而定。选题的原则主要有:

1. 结合所学知识选自己专长的或有基础的主题,否则难以写出水平较高的综述。
2. 根据所占文献资料的质和量选题。
3. 选题一定要能反映出新学科矛盾的焦点,新成果、新动向。
4. 选题不宜过大、范围不宜过宽。



## 二、文献检索

综述不是写学科发展的历史,而是要搜集最新文献资料,获取最新内容,一般所引用的相关文献应当是5年以内的,除非是经典的文献,一般不引用超过10年以上的文献。因此,检索和阅读文献是撰写综述的重要前提工作。一篇综述的质量如何,很大程度上取决于作者对本课题相关的最新文献的掌握程度。如果没有做好文献检索和阅读工作,就去撰写综述,是不可能写出高水平综述的。

收集文献的方法主要有两种:一是通过各种检索工具进行检索。如:常用学术网如PubMed, EMBASE, CA, 中国知网等专业性门户网站;国内外专业期刊网站、中外文电子期刊数据库、国家图书馆以及搜索引擎等。目前国内有三种主要的全文数据库,即中国知识基础设施工程网、中国科技期刊数据库和万方数据资源系统。当然某些特定文献可能需要从论著或期刊中获得,可以利用国家图书馆检索到已出版的该领域的有关专著。也有一些传统的检索方法,如文献索引、文摘期刊检索等,这些传统检索方式现在已经很少使用了,但是,可以利用它们检索到一些较为早期的经典文献。二是根据自己所选定的题目,查找内容较完善的近期(或由近到远)期刊,包括综述性文章、专著、教科书、博士论文等,再按照文献后面的参考文献,去收集原始资料。以“Mechanism”为例,选择“Glucocorticoid resistance”及“Idiopathic nephrotic syndrome”为关键词,分别检索PubMed及EMBASE等搜索工具,检索出结果后,初步阅读,剔除相同文献及关联度不大的文献。

### 知识点

#### 文献搜索的技巧

技巧一:瞄准主流。主流文献,如该领域的专业核心期刊、经典著作、专业部门的研究报告以及有关综述性文章等。查阅主流文献,并注意其引用的参考文献,然后顺藤摸瓜,从而了解该领域的最新研究。其次还要瞄准该领域比较著名的研究者的研究成果。

技巧二:由近及远。检索文献时应由近及远,因为最新研究常常包括以前研究的参考文献,这样可以使研究者更快地了解该领域的研究现状。

技巧三:随时整理。在平时的学习过程中,看文献时养成好的习惯,将所看到的文献中好的观点摘要时,顺手就把文献的详细信息标注出来,这样会省去以后重新查找文献出处的麻烦。

## 三、展开论证

对于已经搜集到的文献,要在阅读文献的基础上做进一步的筛选与分类整理,组织好文献进行初步论证。文献应分为通读、细读、精读三种,这既是撰写综述的重要步骤,也是咀嚼和消化、吸收的过程。通过阅读文献摘要和文内结论,将与本研究无关的文献去掉。对与研究有关的文献,在阅读中,要详细、系统地摘要出文献中研究的问题、目标、视角、方法、结果和结论,以及简要评述该文献研究的不足与尚未提出的问题。然后根据研究需要,将搜集到的文献根据自己所要研究的问题按照研究领域、研究方法、研究视角、研究的问题等进行分类整理。

## 四、文献研究

所谓文献研究即是分析、概括、归纳与批判,综合别人的研究成果。综述不应是材料的罗列,而是对亲自阅读和收集的材料,加以归纳、总结,做出评论和评价,并依据提供的文献资料引



出重要结论。综述别人的研究成果要先读懂,这样,引用时才能忠实于原意,才可以避免错误引用,而不会出现断章取义,导致南辕北辙。对要综述的文献先概括(不是重复),然后进行分析、归纳、比较和对照,目的不是为了解释已有研究,而是为了了解、明确本研究相关的已有研究的主要内容。对前人研究的贡献、不同观点和研究的不足进行分析、评论与批判,在此基础上明确研究者的研究主题,这在文献综述中是非常重要的。由于综述是三次文献,不同于原始论文(一次文献),所以在引用材料方面,也可包括作者自己的实验结果、未发表或待发表的新成果。以“Mechanism”为例,从激素受体,足细胞分子,免疫系统等三方面横向论述在原发肾病综合征中激素耐药的主要机制,在足细胞分子这个子专题内又横向综述了各种分子的情况,而在各种足细胞分子的概述中则纵向地从发现这个分子开始叙述其研究历史。只有如此,文章才会占有大量素材,经过综合分析、归纳整理、消化鉴别,使材料更精炼、更明确、更有层次和更有逻辑,进而把握本专题发展规律和预测发展趋势。

## 五、文献批评

文献批评即在综合别人研究成果时,尚需兼述自己的研究见解,并提出问题。兼述自己的研究见解要力求正确、明了、清晰、简洁,但此处不必长篇大论,不喧宾夺主。特别要指出前人研究的边界,说明需要继续研究的问题。

文献综述的最终目的就是要指出已有研究没有解决的问题是什么,研究的不足之处在哪里,然后提出后续研究要解决的问题。文献综述就是将已有研究的主要观点综述归纳起来予以介绍之后加以评述,说明有哪些研究上的不足和空白点,而这些不足和空白恰恰是以后的研究需要解决的。如在后续的研究中将使用什么新方法或新视角来解决这些问题,或者提出什么新观念、新思想来弥补前人研究之不足等。通过文献综述,在回顾和分析的基础上,提出新的研究方向和研究建议。“Mechanism”在综述完一些主要的研究成果之后都会加上一两句简要但不可或缺评论,这些评论就是该研究方向可能的发展趋势和存在的问题。

## 六、文献综述撰写

文献综述通常包含六部分,即题目、摘要和关键词、前言、主体、总结和参考文献。香港大学建议的论文撰写模式见表 2-1。

表 2-1 香港大学建议的论文撰写模式

| 导论           | 开头段落                                | 简介主题,主题的重要性,理清首要问题,简介各篇文章作者,及其不同或互补之处 |                    |
|--------------|-------------------------------------|---------------------------------------|--------------------|
| 主文部分<br>三个示范 | 1.1 分析 A 的理论观点                      | 1.2 分析 B 的理论观点                        | 1.3 比较 A 与 B 的理论观点 |
|              | 2.1 找出 A 与 B 的共同性                   | 2.2 找出 A 与 B 的差异性                     | 2.3 探讨出一个中心议题      |
|              | 议题 1:探讨 A 和 B                       | 议题 2:探讨 A 和 B                         | 议题 3:探讨 A 和 B      |
| 结论           | 提出一个比其他更好的理论与立场<br>提出一个优于每一个理论部分的摘要 |                                       |                    |

1. 题目 题目应对综述内容起到概括和揭示的作用,要确切、简明、一目了然,且文题相符,切不可小题大做或文不对题。题目过大使文章非常空泛,难以落笔,而过小的题目可能造成文献量不足,不足以形成综述。题目字数一般控制在 20 字之内。

2. 摘要和关键词 综述的摘要主要概述这篇综述将从哪些方面论述某一问题以及最终可能得出的结论。摘要要具有独立性和自含性,不应出现图表、冗长的公式和非公知的符号、缩略





语。摘要后须给出 3~5 个关键词,中间应用分号“;”分隔。

3. 前言 主要包括以下内容:①首先要说明写作的目的,定义综述主题、问题和研究领域;②指出有关综述主题已发表文献的总体趋势,阐述有关概念的定义;③规定综述的范围、包括专题涉及的学科范围和时间范围,必须声明引用文献起止的年份,解释、分析和比较文献以及组织综述次序的准则;④扼要说明有关问题的现况或争论焦点,引出所写综述的核心主题,这既是广大读者最关心而又感兴趣的,也是写作综述的主线。以“Mechanism”为例,首先描述肾病综合征的发病率,其中激素耐药的比例,说明研究激素耐药对治疗肾病综合征的重要意义。其次,描述目前主要围绕哪几个方面对其进行研究,而机制研究结果并不明朗,从而说明综述有助于更好地理解这一问题。

4. 主体 是综述的主要部分,其写法多样,没有固定的格式。可按文献发表的年代顺序综述,也可按不同的问题进行综述,还可按不同的观点进行比较综述。不管用哪一种格式综述,都要将所搜集到的文献资料归纳、整理及分析比较,阐明引言部分所确立综述主题的历史背景、现状和发展方向,以及对这些问题的评述。主体部分应特别注意代表性强、具有科学性和创造性的文献引用和评述。主体部分的内容主要包括论据和论证两个部分,通过提出问题、分析问题和解决问题,比较不同学者对同一问题的看法及其理论依据,进一步阐明问题的来龙去脉和作者自己的见解。当然,作者也可从问题发生的历史背景、目前现状、发展方向等提出文献的不同观点。正文部分可根据内容的多少可分为若干个小标题分别论述。以“Mechanism”为例,从激素受体的改变,足细胞相关分子的变异以及免疫系统紊乱等三方面进行横向阐述。在撰写激素受体的改变时,采用历史发展的顺序,从激素受体的基础知识开始介绍,最后提及目前的研究成果。在介绍足细胞相关分子时,要全面概述目前发现的各种足细胞分子突变与激素耐药的关系,其中又要详细地从基因、分子组成开始叙述,最后介绍各种报道的典型病例,使读者对这一领域有全面的了解。

### 知识点

#### 主体部分的写作方法

主体部分的写作方法通常有两种。

1. 搭架填空法 所谓“搭架”,就是先提出论文综述所涉及的各种问题,先将这些问题按学科理论的逻辑关系排列,就像造房子一样按总体设计将房子各部分搭起来。然后,将与各“房架”有关的问题归并在一个标题下,对相关资料进行整理和分析写成小“综述”。其次,把这些小综述连接起来。这一写法的关键是各小“综述”(即小标题)的确定和连接,它关系到一篇综述的质量优劣,而抓住综述专题的主线和内容主题间的逻辑联系则是最关键的。

2. 火车接龙法 这是一种按学科进展的时间顺序进行论述的综述写法,形象化地称之为“火车接龙法”。作者必须对该主题的科学进展非常清楚,可先按历史阶段分别写成小“综述”,每个小“综述”内也可按搭架填空法写,最后将各小“综述”连接起来。这种写法在选择和引用资料时,要注意反映所综述的专题的理论发展进程和阶段性的成果。在实际写作过程中,往往是以某一种方法为主,同时穿插运用另一种方法,这取决于写综述的内容和所掌握的文献资料。

5. 总结 是对综述正文部分作扼要的总结,作者应对各种观点进行综合评价,提出自己的看法,指出存在的问题及今后发展的方向和展望。内容单纯的综述也可不写总结。

6. 参考文献 是综述的重要组成部分。列出参考文献除了表示尊重被引证者的劳动及表明文章引用资料的依据外,更重要的是使读者在深入探讨某些问题时,提供查找有关文献的线



索。综述性论文是通过对各种观点的比较说明问题的,读者如有兴趣深入研究,可按参考文献查阅原文。因此,必须严肃对待。一般参考文献的多少可体现作者阅读文献的广度和深度。对综述类论文参考文献的数量不同期刊有不同的要求,一般以30条以内为宜,以3~5年内的最新文献为主。参考文献建议用EndNote或NoteExpress进行编辑,建议边写边编辑参考文献,在软件里建立一个总的图书馆,可以按综述的版块添加不同的子文件夹。

### 第三节 文献综述的问题及解决方案

#### 一、综述的内容与所研究的内容相脱节,为综述而综述

文献综述要紧扣研究的问题,在做文献综述时,要始终明确要解决什么问题,哪些问题已解决,是如何解决的,解决的理论依据和实践基础是什么,有没有道理。其目的在于进一步明确自己的问题,寻找研究的切入点,避免重复研究,少走弯路。文献综述不是资料库,要紧紧围绕所研究的“问题”,确保所述的已有研究成果与要开展的研究直接相关。一般来说,文献综述采用由宽到窄的综述方法,从大处着手,逐渐把已有研究集中到要解决的问题上去,切忌避免把有关领域无目的地泛泛地进行综述。

#### 二、文献综述要文字简洁,尽量避免大量引用原文,罗列堆砌文章

很多初学者误认为文献综述的目的是显示对其他相关研究的了解程度,结果导致很多文献综述不是以所研究的问题为中心来展开,而变成了读书心得清单。在综述中需要用自己的语言将作者的观点讲清楚,从原始文献中得出一般性结论。

#### 三、综述选择的文献水平低,不全面,使研究的基础不牢固

克服方法:①围绕“研究的题目”和“关键词”确定综述的范围;②选择学术性强的媒介中的文献进行综述;③可以从本领域知名学者所开展研究的角度了解前沿成果,从这些知名学者的研究入手综述研究文献;④明确综述的目的,整理前人的研究成果,确定自己的概念和研究体例;⑤认真阅读前人的研究成果,避免出现误读或重复遗漏;⑥综述的框架要覆盖主要研究内容,表述体现逻辑性。注意引用文献的代表性、可靠性和科学性。在搜集到的文献中可能出现观点雷同,有的文献在可靠性及科学性方面存在着差异,因此在引用文献时应注意选用代表性、可靠性和科学性较好的文献。

#### 四、综述时对别人的观点理解不准确、不全面

在文献综述过程中,首先要全面、准确地“综”,也就是说,在查阅某一领域的已有的研究文献时,一定要全面,确保文献综述完整不偏,尤其是不能遗漏该研究领域的国内外知名学者的研究成果,切忌不能有选择的搜集文献并对其综述。否则,文献综述的结论缺乏可信性。为确保文献的综述准确,最好是查阅第一手文献资料,尽量避免使用别人对原始文献的解释或综述,即第二手文献。仅阅读第二手文献往往容易产生错误。其次是客观地“述”,即在评述已有研究成果时,一定要以已有研究的文献为依据,不能脱离文献进行评述,要客观地说明已有研究的研究现状。

#### 五、综述中只有别人的观点,没有自己的取舍和评价

文献综述不是已有研究成果的重复、罗列和一般性介绍,而应是对以往研究的优点、不足和贡献的批判性分析与评论。文献综述由“综”和“述”构成。“综”对所查阅的文献进行综合的归类、



提炼、概括。“述”则需要融入研究者自己理论水平、专业基础、分析问题、解决问题的能力,在对问题进行合情合理的剖析基础上,提出自己独特的见解。

### 思考题

1. 什么是文献综述? 文献综述与其他文体的区别。
2. 文献综述通常包含哪几部分内容? 每部分内容如何撰写?
3. 结合自己专业与研究方向,选择一个与自己研究或者临床实践问题相关的主题进行文献综述。

(张爱华)

## 第三章 系统评价

### 临床案例

门诊病历摘要:女性,72岁,因“进行性记忆和生活自理能力下降2年”来某院门诊就诊。患者2年前开始出现记忆力问题。初始表现为记不住客人的名字,记不住看过的新闻,做事情丢三落四等。近半年记忆下降逐渐明显,2个月前一次上街,找不到回家的路。记忆力检查提示近事记忆减退明显。头颅CT检查发现皮质性脑萎缩和脑室扩大。临床诊断为“阿尔茨海默病”。曾至多家医院就诊,但主治医师给予的治疗用药意见却不一致。

### 第一节 提出临床研究问题

问题1:面对高年资专家医师的不同诊疗意见,如何为患者选择最有效的治疗?

对于如何为患者提供最有效的个体化治疗,是向经验丰富的老专家请教,还是检索相关的临床RCT研究,还是查找相关的系统综述(systematic review),还是去查阅相关的病例报告,说到底,我们如何得到患者所需治疗的最佳证据(best evidence)。

首先,什么是最佳证据呢?最佳证据不是将所有查阅到的文献内容罗列,而是要通过科学质量评价、筛查和归纳,有时候需要采用统计方法,将资料合并,得出综合分析结果。最佳证据的一个重要来源是系统综述。

#### 知识点

干预研究的常用证据等级

| 证据性质分级 | 证据水平分级 | 证据种类                 |
|--------|--------|----------------------|
| A      | 1a     | 随机对照临床试验(RCT)的系统综述   |
|        | 1b     | 单个的RCT(95%CI 信度空间较窄) |
| B      | 2a     | 队列研究的系统综述            |
|        | 2b     | 单个的队列研究及质量较差的RCT     |
|        | 2c     | 结局研究                 |
| C      | 3a     | 病例对照研究的系统综述          |
|        | 3b     | 单个的病例对照研究            |
| D      | 4      | 系列病例分析及质量较差的病例对照研究   |
| E      | 5      | 没有分析评价的专家观点          |

具体临床问题证据的可靠性评估依据相应的标准,其基本要点均需以研究类型等原则来确定。根据上述5个级别进行证据分级后,仍必须根据自己的专业知识和临床流行病学知识等对证据的科学性、实用性和有效性进行评价。



由证据水平分级可以知道,基于 RCT 的系统综述是治疗性干预中证据等级最高的研究,是为临床决策提供最佳证据的重要来源之一。

系统综述(systematic review),国内亦称系统评价,是一种按照严格的纳入标准,全面地收集关于某一具体医疗问题的所有研究,对纳入研究进行全面的质量评价,并进行定性和(或)定量分析(多个研究间具有足够同质性时进行定量数据合并),以对该问题进行系统总结的研究方法。

系统评价是高质量证据的来源,其目的是获取所有关于某种特殊疗法或干预措施的真实证据,并由此对疗效进行更准确的量化评价。能为临床医生、患者及其他相关人员提供重要信息。加拿大的流行病学和生物统计学家 Brian Haynes 提出了“5S”模型(图 3-1)。“5S”模型有助于决策者查找所需的证据,从顶层开始逐层向下寻找相关证据。

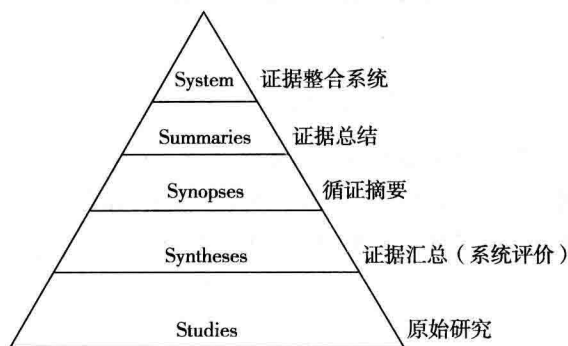


图 3-1 Haynes “5S”模型——循证信息资源分布示意图

常用的系统评价检索资源主要有:

- (1) The Cochrane Library (考科蓝图书馆): [www.thecochranelibrary.com](http://www.thecochranelibrary.com)
- (2) 最佳实践 Best Practice: [bestpractice.bmj.com](http://bestpractice.bmj.com)
- (3) 临床指南 Clin-eGuide: [www.clineguide.com/index.aspx](http://www.clineguide.com/index.aspx)
- (4) 循证医学指南 (EBM Guideline): [ebmg.wiley.com](http://ebmg.wiley.com)
- (5) 临床证据 (Clinical evidence): [www.clinicalevidence.com](http://www.clinicalevidence.com)
- (6) 英国临床指南 (NICE): [www.nice.org.uk](http://www.nice.org.uk)
- (7) Evidence-Based Medicine: [www.evidence-basedmedicine.com](http://www.evidence-basedmedicine.com)
- (8) 中国循证医学杂志: [www.cjebm.org.cn](http://www.cjebm.org.cn)
- (9) 医学文献专业数据库: 比如, PubMed, EMBASE 等。

问题 2: 所有的系统评价结果都可以作为高质量证据来指导临床吗?

经过数据库检索,你检索到了多篇相关的系统评价,但其结果也各不相同,那究竟该选择哪些系统评价作为临床最佳证据呢?

系统评价作为文献研究的一种方法,能够为临床实践提供依据,但使用不当也将影响研究结果的价值,所以系统评价研究的质量需要评价,而不是盲目接受。系统评价研究的质量可从方法学质量和报告质量两个方面进行评价。只有方法学质量(methodological quality)和报告质量(reporting quality)均较高的系统评价,才能作为最佳证据指导临床。

评估系统评价研究的真实性和质量,是考察研究方法学质量的重要指标,主要涉及以下方面:

- (1) 是否是对相关问题和随机对照研究的系统综述:首先,确定一个合理的主题,与需要解决的治疗问题有关。其次,切忌随机和非随机研究混杂,否则将可能产生错误的结果。
- (2) 是否描述了检索相关研究文献的方法:资料收集(data collection)是系统评价中非常重





要且较为困难的步骤,“全面”是资料收集的基本原则(包括不同语言,已发表和未发表的研究等),如果收集不全面,带来选择性偏倚,就会影响研究结果。电子数据库检索由于错误编码等原因可能使超过半数相关研究遗漏,手工检索是必要的补充手段。

(3) 是否有明确的纳入标准:根据研究的主题,需要制定明确系统评价研究的纳入和排除标准(包括研究的类型、受试对象的特征、干预措施和结局指标等),如果纳入标准制定的不合理,将导致资料收集的偏差。同时资料筛选过程也要注意对选择性偏倚的控制,一般至少需要两名研究者独立进行,再核实一致性。

(4) 不同研究间是否有同质性:系统评价常采用 Meta 分析(Meta-analysis)对数据进行定量的合并,合并的条件是研究的同质性(homogeneity)(包括临床同质性、方法学同质性和统计学同质性);如果纳入的研究存在异质性(heterogeneity),需要分析产生异质性的原因,而不是一定要合并。如果试验的异质性很难解释,那么系统评价结果的可靠度就会受到质疑。

### 知识点

#### 研究间的同质性和异质性

系统评价得到的结果来源于多项研究的效应合并,然而只有那些具有同质的研究才能合并,若研究间差异过大,且未采取一定的方法(严格统一的纳入和排除标准,统计学模型处理等)对其加以控制,结果就不可靠,其结论也不适合用于指导解决相应临床问题的证据。

报告质量(reporting quality)评价关注的是研究者对研究方法、过程和结果等内容的报告情况。Meta 分析报告的质量(quality of reporting of meta-analysis, QUOROM)是系统评价和 Meta 分析报告规范。近年来国际学术界在 QUOROM 基础上进一步发展,提出了系统综述和 Meta 分析优先报告的条目声明(preferred reporting items for systematic reviews and meta-analyses, PRISMA 声明),被用来评价系统评价的报告质量。此规范包括系统评价的标题、摘要、前言、材料与方法、结果、讨论、结论与资金支持的 7 个方面 27 个条目。

由此可见,只有当一篇系统评价的方法学质量和报告质量均较高时,才能很好地为临床决策提供依据。当临床医生将检索到的系统评价证据指导自身临床实践时,需仔细评价系统评价的质量,采纳研究质量较高的系统评价证据应用于临床。

此外,当临床医生或其他研究者进行系统评价写作时,应提高系统评价的方法学质量和报告质量,为其他临床医生、患者和医疗决策制定者提供高水平的循证证据。

### 知识点

#### 系统评价质量与原始研究质量的关系

系统评价是基于文献的二次研究,结果受原始研究质量的影响,如果纳入的原始研究质量较差,将影响结果的价值,故需要对纳入研究进行质量评价,以考察其结果的可靠性。研究结论要结合数据分析的结果、研究中存在的问题和纳入研究的质量客观地列述,如果不考虑纳入研究的情况和潜在的问题,可能得出不正确的结论。

但系统评价本身的质量与原始研究质量没有直接关系,系统评价的重要作用在于向读者提供全面和客观的信息,其质量的高低取决于研究过程中对偏倚的控制和报告。如果系统评价能按照规范制作,评价了纳入原始研究存在的问题,在讨论和结论中能结合具体问题,客观报告研究结果,那么系统评价多属于高质量,提供的证据强度水平也较准确。



## 第二节 系统评价的制作步骤

### 问题3:对于发现的临床问题,如何制作一篇系统评价?

当你在数据库中进行检索后发现,目前没有一篇系统评价能作为最佳证据来解答你临床的疑惑,那接下来该怎么办?你可以根据目前检索到的 RCT 研究,自己制作一篇系统评价,这样也为其他有类似疑问的临床医生提供最佳证据。

那如何制作一篇高质量的系统评价呢?

根据 2000 年 David Sackett 等对系统评价的定义“A summary of the medical literature that uses explicit methods to perform a thorough literature search and critical appraisal of individual studies and that uses appropriate statistical techniques to combine these valid studies”也就是说,系统评价是全面收集所有有关研究,对所有纳入的研究逐个进行严格评价,整合所有的研究结果进行综合分析和评价,必要时进行 Meta 分析,得出综合结论(如,有效、无效、应进一步研究以及安全性等),提供尽可能减少偏倚、接近真实的科学证据。

系统评价的制作包含以下步骤:

1. 提出问题,制订研究计划 制作系统评价首先应提出问题,进行科研设计并制订研究方案。明确研究目的、提出研究问题是最重要和最基本的第一步,而且提出问题的过程也是系统复习文献的过程。提出研究问题应包含 5 个要素,即循证医学实践中的 PICOS 模式,P:patient or population (研究对象),I:intervention (干预手段),C:comparison intervention (干预措施与安慰剂或标准治疗的比较),O:outcomes (结局变量或疗效判定指标),S:study design (研究设计,如随机对照临床研究)。

确定研究问题后,应制订详细的研究计划书(protocol),包括研究问题的背景材料,文献检索的途径和方法,纳入标准(inclusion criteria)和排除标准(exclusion criteria)。拟定纳入与排除标准时,除考虑研究设计的类型、报告发表的时间、地区、语种、文献形式外,对每个独立研究,研究对象(年龄、性别、疾病类型、疾病严重程度)的选择、对照组(空白对照、安慰剂对照、标准治疗对照、常规治疗)的设置、药物或暴露(剂型、剂量、用药途径、疗程)的定义、随访的长短、结果的判断标准等均应有明确的规定。此外,还包括统计分析步骤,文献质量评价和结果解释等。

我们就拿本章开始提出的临床问题进行举例。上级医师关于是否给患者使用银杏叶提取物的意见不一,经过检索相关数据库,并未发现有关银杏叶提取物对于痴呆治疗效果的临床证据,因此决定自己制作一篇系统评价。首先,提出研究问题,即银杏叶提取物对痴呆治疗是否有效。根据 PICOS 模式,P:痴呆患者,包括阿尔茨海默病、血管性痴呆以及混合性痴呆;I:银杏叶提取物;C:安慰剂;O:日常生活能力量表(ADL)、简易精神状态筛查量表(MMSE);S:选取随机对照临床试验的研究类型等。

其次,撰写一份研究计划书(protocol),包含研究背景;纳入研究类型:随机对照临床试验(randomized controlled trials, RCT);研究对象:由公认的标准(如 ICD、DSM、NINCDS/ADRDA、NINDS/AIREN、CCMD)诊断为阿尔茨海默病(Alzheimer's disease, AD)、血管性痴呆(vascular dementia, VD)、混合性痴呆;排除标准:①非人类研究;②综述、报告类以及未完整发表的研究;③重复发表、重复收录的研究;④合并其他促认知药物;⑤非安慰剂对照;⑥无对照人群研究等。

2. 检索相关文献 系统而全面地收集与研究问题相关的文献是系统评价有别于传统文献综述的重要特征之一,是完成一篇高质量系统综述的基础。在上一步我们就已经制定了文献检索策略(search strategy),即通过分析研究问题,将其分解为几个方面,写出相应的检索词,并确定



检索词与检索词之间的逻辑组配关系。为保证系统评价的质量,应尽可能地查找一切与所研究的主题相关的文献。文献检索的完整性会直接影响研究结果的可靠性。文献检索时最好能找到所有有关的文献(包括未发表的),以减少发表偏倚对研究结果的影响。因此,必要时可以咨询专业图书馆员或信息检索人员,尽量避免漏检和误检。检索文献时可对检索时间段、文章发表的语种、出版年限、出版类型进行必要的限制。很多时候,文献的检索需要专业情报检索人员协助。

一般可通过计算机或手工文献检索进行资料收集。常用的英文医学数据库包括 PubMed、EMBASE、ISI Web of Science、The Cochrane Library 及 EBSCO。常用的中文数据库包括中国期刊全文数据库、中国科技数据库、中国生物医学数据库及万方数据库等。有时需手工检索相关期刊与书籍,收集灰色文献(grey literature)(如与同事、专家、药厂联系获得未发表的文献;政府报告、会议专题论文、未发表的学位论文、个人通信等)。

首先,确定检索词,将英文检索词定为“ginkgo biloba”、“dementia”;中文检索词定为“银杏叶提取物”、“痴呆”。为防止漏检,同时使用关键词“GbE 761”或其商品名如“Tanakan, Tebonin, Rokan, Ginkoba”分别替换“银杏叶提取物”(“ginkgo biloba”)扩大检索范围。其次,在确定检索词后,需明确检索哪些数据库,以什么形式检索,手动检索还是计算机检索。通过计算机检索 PubMed、EMBASE、the Cochrane Library、ISI Web of science、中国生物医学文献数据库、中国期刊全文数据库、中国科技期刊全文数据库和万方数据资源系统,并辅以文献追溯方法。最后,确定检索年限,对 1982 年 1 月至 2012 年 9 月公开发表的关于银杏叶提取物对痴呆治疗的文献进行收集。

3. 纳入文献的筛选 检索到相关文献后,必须根据纳入和排除标准进行仔细筛选。由于收集的资料可能很多,可借助文献管理软件(如 EndNote 软件)进行文献筛选和管理。通常,研究的选择过程至少要求两名研究人员独立进行,如果有分歧可通过共同讨论决定是否纳入,必要时可由第三位研究者协助解决。

筛选文献一般分两步进行,首先进行初筛,通过浏览检索到文献的题目、摘要等信息可以剔除部分不合格文献。随后通过阅读全文,根据预先制订的纳入和排除标准,仔细甄别筛选,对于存在疑问的文献如经讨论仍无法统一意见,可先纳入,待联系原文作者获取相关信息后再作取舍。

在筛选过程中,为保证纳入文章的同质性(homogeneity)及可重复性(repeatability),不仅要重点关注前面提到 PICOS 模式的五个要点(研究对象、干预手段及对照、结局指标和研究设计),还要考虑文献的研究开展时间或文献发表的年代和语种、样本大小及随访年限、多重发表的处理及提供信息的完整性。

根据研究计划中指定的入排标准,对检索到的文献进行筛选,首先由两位研究者(JL,SL)独立通过文献的题目和摘要进行初筛,初筛后的文献通过阅读全文进行二次筛选,然后交叉核对筛选结果,如果有分歧则通过共同讨论决定是否纳入,必要时可由第三位研究者(LC)协助解决。如果文中信息不全或信息不清楚,与原始研究作者联系获取信息。并绘制具体检索和纳入流程图(图 3-2)。

4. 纳入文献的数据提取 资料提取(data extraction)是从符合纳入要求的文献中摘录用于系统评价的数据信息,所提取的信息必须是可靠、有效、无偏的。为保证数据收集的质量,在资料信息提取和计算机录入时应双人独立进行,核查过程中遇到不同之处应经过讨论决定。资料的提取至少应包括研究的文献来源(文章题目、第一作者、发表期刊、名称、发表年限)、研究的设计类型及方法学信息(如分组数、随机方法、盲法、样本量、研究场所等)、研究对象的基本特征(年龄、性别、种族、诊断标准、分期、病例来源等)、干预措施、结局或结果。对于一些纳入文献但原始



数据提供不完整的情况,应直接与原作者联系,如仍无法得到原始数据,则应排除此文献。

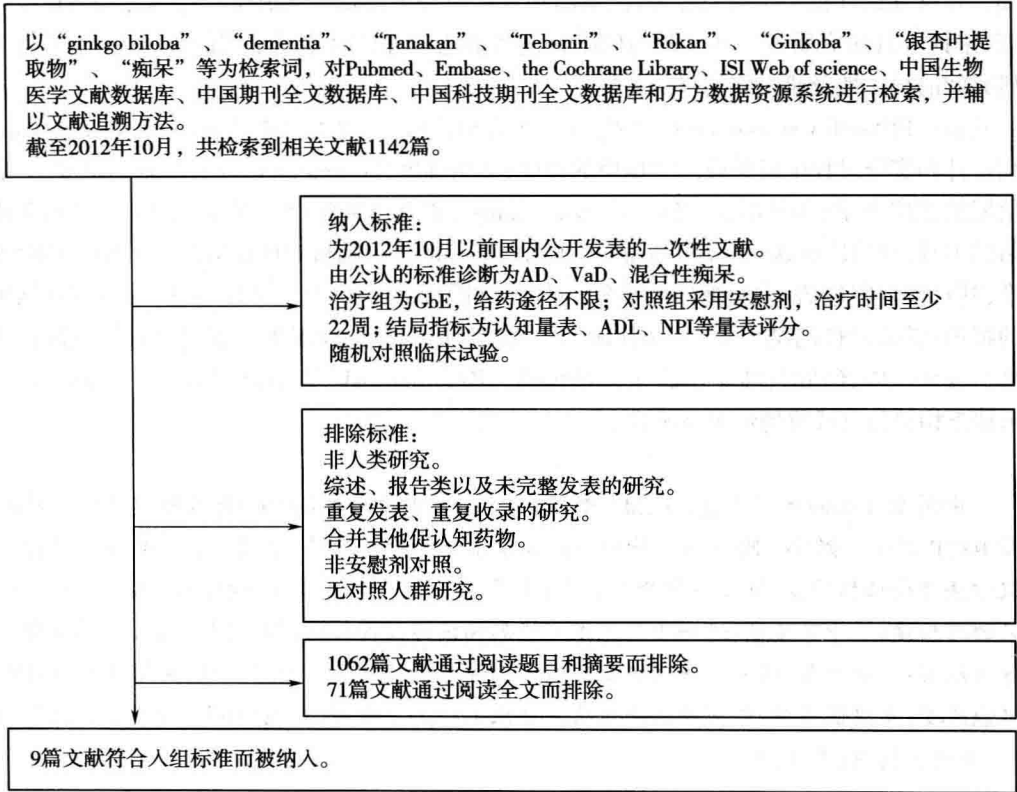


图 3-2 文献筛检流程图

对纳入的文献,应考虑采用哪些效应指标(effect size)进行合并,而合并的指标并非越多越好,而是看哪些指标具有代表性或临床意义重大,即主要结局指标(primary outcome)。通常两组间比较时,连续性变量(continuous variable)用加权均数差(weighted mean difference,WMD)、标准化均数差值(standardized mean differences,SMD)表示效应大小;二分类变量(dichotomous variable)用率差(rate difference,RD)、比数比(odds ratio,OR)、相对危险度(relative risk,RR)、相对危险度降低值(relative risk reduction,RRR)等表示效应的大小。

当文献筛选结束,则采用 EndNote X5 软件管理文献。制定数据提取表(data collection form)(表 3-1),由两名研究者对资料进行提取并核对,资料提取包括:①一般资料:研究标题、第一作者、发表年份、平均年龄、性别构成、纳入/排除标准等;②研究方法:干预措施、基线情况、试验时间、入组人数、主要结局指标和次要结局指标等。又由于结局指标为量表评分等计量资料,因此效应指标选取标准化均数差值。

表 3-1 纳入研究数据提取表

| 研究<br>编号           | 纳入标准                                                                  | 试验时间<br>(周) | 治疗方案                       | 样本量        | 平均年龄<br>(标准差)      | 性别构成<br>(M/F)  | 测量<br>指标                    | 基线分数值<br>(标准差)                                                               |
|--------------------|-----------------------------------------------------------------------|-------------|----------------------------|------------|--------------------|----------------|-----------------------------|------------------------------------------------------------------------------|
| Le<br>Bars<br>2000 | Age>45, ICD-10<br>and DSM- III -RAD<br>or MID, MMSE 9-<br>26, GDS 3-6 | 52          | GbE761<br>120mg<br>placebo | 155<br>154 | 69 (10)<br>69 (10) | 76/79<br>67/87 | MMSE<br>ADAS-<br>Cog<br>ADL | MMSE:<br>21.1 (5.8)<br>21.2 (5.5)<br>ADAS-cog:<br>20.0 (16.0)<br>20.5 (14.7) |

学  
习  
笔  
记



5. 纳入研究的质量评价 我们前面讲过系统评价的质量评价,对纳入研究的质量评价非常重要。常常通过评价一个研究在设计、实施和分析中防止和减少系统误差(偏倚)和随机误差的程度,来评价其研究质量。并以此为依据在进行敏感性分析时给予不同的权重,用于考察和解释研究间的异质性及研究间结果的差异。

质量评价(quality assessment)一般包括三方面内容:①方法学质量(methodological quality):研究设计和实施过程中避免或减少偏倚的程度;②精确度(accuracy):即随机误差的程度,一般用可信限的宽度表示;③外部真实性(external validity):研究结果外推的程度。国际上有很多质量评估的工具,目前广泛认可的是 Cochrane 协助网推荐的针对随机对照试验进行的偏倚风险评估标准,其具体内容包括:①随机分配方案的产生;②隐匿分组;③是否采用盲法;④不完整结果数据的报道;⑤选择性的结果报告;⑥其他影响真实性的潜在危险因素。我们对每篇随机对照试验进行偏倚风险评估时:“low risk”表示低偏倚风险;“high risk”表示高偏倚风险;“unclear risk”表示缺乏相关信息或偏倚情况不确定。

由两名独立的研究人员(JL,SL)根据 Cochrane 协作网推荐的偏倚风险评估方法对每篇 RCT 进行了偏倚风险评估。并进一步基于系统评价的结果,应用 GRADE 系统推荐分级方法评价证据质量,证据质量分级如下:①高质量:进一步研究不可能改变该疗效评估结果的可信度;②中等质量:进一步研究很可能影响该疗效评估结果的可信度,且可能改变该评估结果;③低质量:进一步研究极有可能影响该疗效评估结果的可信度,且该评估结果很可能改变;④极低质量:任何疗效评估结果都很不确定。最后,对偏倚风险评价结果进行编辑、分析和制图(表 3-2)。

表 3-2 文献偏倚风险评估表

|           | Le Bars<br>2000 | Dongen<br>2003 | Kanowski<br>2003 | Schneider<br>2005 |
|-----------|-----------------|----------------|------------------|-------------------|
| 随机分配方案的产生 | ?               | Low            | Low              | Low               |
| 隐匿分组      | ?               | Low            | Low              | ?                 |
| 盲法采用      | ?               | Low            | Low              | Low               |
| 结果数据的完整性  | High            | Low            | High             | Low               |
| 选择性的结果报告  | Low             | Low            | Low              | Low               |
| 其他可能偏倚    | ?               | High           | High             | High              |

6. 资料的统计学处理——Meta 分析 Meta 分析(Meta-analysis)作为一种定量的系统评价(quantitative systematic review)的统计学方法,它通过定量合并的方式得到综合的结果估计。对各独立研究结果合并进行统计学分析的基础是假定各独立研究的结果是同质的,即各研究间现有结果的不同仅仅是由于抽样误差造成的。综合各独立研究的结果进行合并,理论上因为增大了样本含量,从而使随机误差减小。但如果各研究结果的差异不仅仅是由于抽样误差造成的,Meta 分析有时就会导致错误的结论。因此,在对结果数据进行统计合并之前,应首先对其进行异质性检验。

进行异质性检验(heterogeneity test)时,异质性程度采用统计量  $I^2$  表示效应值变异大小。若同时符合  $I^2 < 50\%$  和  $P \geq 0.1$  时,纳入文献被认为是同质性,采用固定效应模型(fixed effect model)分析;反之说明研究间存在实际异质性,需要查找异质性的来源,之后采用随机效应模型(random effect model)分析。

通过对 ADL 量表得分的 Meta 分析,发现异质性较大,  $I^2 = 84\%$ 、 $P < 0.0001$ ,因此采用随机效应模型,得到合并效应值  $SMD = -0.28$ 、95% 可信区间为  $[-0.51, -0.05]$  (图 3-3)。



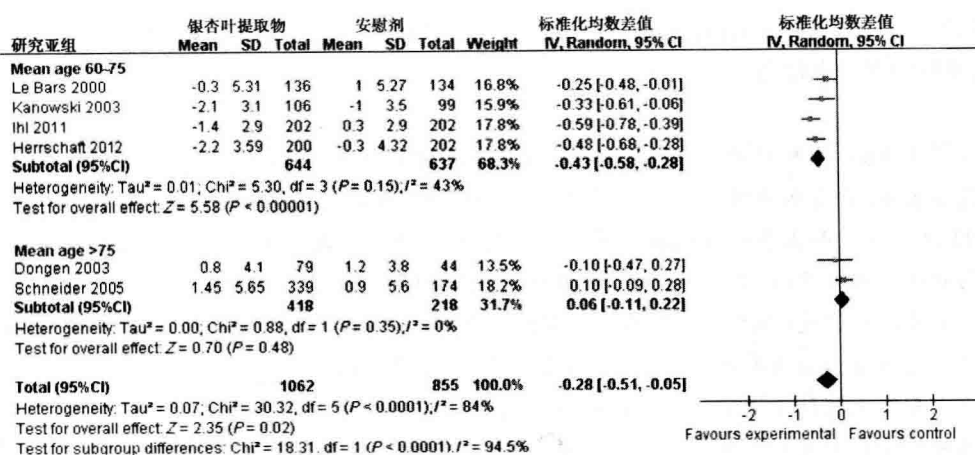


图 3-3 Meta 分析结果森林图

### 知识点

#### Meta 分析异质性的来源

Meta 分析的异质性来源可从三方面分析:临床异质性、方法学异质性和统计学异质性。临床异质性是指:受试对象的不同、干预措施的差异和研究的终点指标不同所导致的变异。方法学异质性:由于试验设计和质量方面的差异引起的,如盲法的应用和分配隐藏的不同,或者由于试验过程中对结局的定义和测量方法的不一致而出现的变异。统计学异质性:干预效果的评价在不同试验研究间的差异波动(变异),是不同研究间临床和方法学上变异联合作用的结果。临床异质性、方法学异质性和统计学异质性三者可相互独立又可相互关联的,临床或方法学上的异质,不一定在统计学上就有异质性表现,反之亦然。

7. 敏感性分析 敏感性分析(sensitivity analysis)主要是观察当选用不同模型、按研究质量评价标准从纳入文献中去除低质量文献、根据样本量大小对纳入研究进行分层分析、改变纳入和排除标准时,重新进行 Meta 分析,效应合并值点估计和区间估计的差异,考察结论有无变化,以保证 Meta 分析结果的稳健性。

作者对纳入研究进行敏感性分析,发现文献 Napryeyenko 2007 引起很大的异质性,剔除该文献后  $I^2$  从 91% 降至 0%,因此在 Meta 分析时剔除此文献,以增加结果的稳健性。有时候,可以采用亚组分析,图 3-3 中,按照年龄段分成两组(60~75 组, >75 组)分别进行 Meta 分析,此时,两组的异质性分别是 43% 和 0%,而总体异质性  $I^2$  是 84%。

8. 结果的分析和讨论 在进行一系列统计学合并分析后,应对所得结果作客观、科学、合理的解释,并写出相应的研究报告。在撰写研究报告时,应详细陈述分析的目的,文献查找方法及取舍标准,所综合的单个研究的特征;说明所应用的统计学方法;提供包含有各个研究统计结果的图表;结论可能遇到的偏倚及处理方法;讨论分析结果应用价值等。

系统评价的讨论部分是对评价结果的解释,其重点应当介绍有助于人们决策的几个方面:证据的强度、结果的可应用性、其他与决策有关的信息,干预措施的利弊,费用的权衡等。

将系统评价的结果应用到临床决策中,除了考虑方法学质量和报告质量外,还要考虑结果的临床重要性,包括纳入的是否为高质量的研究、结局指标是什么、结果是否精确以及合并效应



量等内容。若纳入的是高质量的研究,且数量充足,各研究结果同质性较好,那么结果精确度就越好,证据的强度也较高。

根据 Meta 分析结果,发现银杏叶提取物与安慰剂在改善 60 岁至 75 岁之间痴呆患者的日常生活能力水平方面有统计学差异[ $n=1281, 4 \text{ RCT}, \text{SMD}=-0.32, 95\% \text{CI}(-0.43, -0.21), P=0.00$ ],而银杏叶提取物与安慰剂在改善 75 岁以上痴呆患者的日常生活能力水平方面无统计学差异[ $n=636, 2 \text{ RCT}, \text{SMD}=0.00, 95\% \text{CI}(-0.16, 0.16), P=1.00$ ]。本研究结果证实了我们的假设:银杏叶提取物与安慰剂在改善 60 岁至 75 岁之间痴呆患者的认知功能和日常生活能力方面有统计学差异,而在改善 75 岁以上痴呆患者的认知功能和日常生活能力方面则无统计学差异。这对临床实践而言提供了较为明确和具体的客观数据。但限于纳入研究较少及 9 个研究的总体为偏低质量,故银杏叶提取物的有效性尚需大样本、长程的高质量随机对照临床试验进一步验证。而病例中患者 72 岁,因此可以用银杏叶提取物进行治疗。

以上就是制作一篇系统评价的完整过程简介,但在实际操作中,由于 Meta 分析的统计学方法较为复杂,因此在制作过程中,常常借助一些计算机软件来进行 Meta 分析和系统评价撰写。具体操作流程可以参考相应计算机软件的使用手册或相关书籍,我们同样以本章开头的病例为例,在附录中给出了 Cochrane 协作网推荐的 review manager 软件(RevMan)的操作流程。

#### 知识点

##### 常用的系统评价/Meta 分析软件

1. RevMan RevMan(review manager)是 Cochrane 协作网提供的一款免费软件,在 Cochrane 协作组织的官方网站可免费下载。目前最新版本为 RevMan5.2。由于 Cochrane 协作网的影响和免费的特点,其应用较多。
2. CMA CMA(comprehensive meta-analysis)是一款商业软件,其输出的森林图可以自定义编辑,并可输出为 ppt 和 doc 格式文件,是目前应用较广的 Meta 分析软件。
3. Meta Win 商业软件,由美国 Sinauer Associates 公司经销,带有效应计量计算器。近年应用较少。
4. Meta-Disc Meta-Disc(meta-analysis of diagnostic and screening tests),免费软件。界面友好,可进行异质性检验、线性回归和诊断试验的 Meta 分析,图形质量较高。
5. R R 软件是国际通用的免费统计软件。它是一套完整的数据处理、计算和绘图软件系统,通过 Meta 分析扩展包,不仅可以完成经典的 Meta 分析功能,一些新近出现的如网络 Meta 分析等,也可以在 R 中实现。目前用得越来越多。
6. 通用统计软件中的 Meta 分析模块 STATA、SAS 和基于 Bayes 方法的 BUGS 等统计软件都嵌入了 Meta 分析的模块,但是从界面、方便程度、灵活性、输出图形等方面不及上述软件。

### 第三节 解读系统评价与 Meta 分析

问题 4:对于系统评价的结果和临床意义应该如何做出正确的解释呢?

对于检索到的或自己制作的系统评价,如何看待评价得到的结果?对临床的指导意义有多



大? 这主要在于能否对系统评价的结果进行正确的解释。

1. 全面而正确地解释系统评价 对系统评价全面而正确的解释,主要包括以下五部分:

(1) 主要结果的总结:简单归纳所有重要结局指标的结果,包括有利和不利的结果,并给出重要结局指标的质量评价。

(2) 证据的总体完整性和适用性:明确说明证据的适用人群,重点解释证据在特定环境下不适用的原因:生物学差异、文化差异、对干预措施依从性的差异,探讨应怎样使用干预措施才能获得收益、风险和成本的平衡。

(3) 证据质量:重点从总体上客观评价纳入研究的质量。

(4) 可能存在的偏倚或局限性:只要包括检索策略是否全面、是否进行质量评价、研究的选择和纳入的可重复性、分析方法是否恰当、是否进行发表偏倚评估等方面进行描述。

(5) 与其他研究相同点和不同点的解释:与原始研究或其他系统评价进行比较,从中寻找相同或不同的结果,解释其可能发生的原因。

综合以上几点,才有可能得到一个全面而系统的解释,为临床实践提供证据,但如何解读具体的系统评价结果呢?

2. 解读具体的系统评价结果

(1) 效应值的解释和临床意义:我们会根据系统评价所基于的原始研究设计类型和数据资料类型,选择不同的效应值(effect size),如适用于计数资料的比数比(odds ratio, OR)、危险比(risk ratio, RR)、危险差值(risk difference, RD)、风险比(hazard ratio, HR),以及适用于计量资料的加权均数差值(weight mean difference, WMD)和标准化均数差值(standardized mean difference, SMD)。

在这些效应值中,RR 适用于前瞻性研究,包括随机对照试验(randomized controlled trial, RCT)和前瞻性队列研究(perspective cohort study);RD 适用于实验性研究及观察性研究中的队列研究;OR 主要适用于回顾性类型的研究;HR 适用于生存资料研究。对于计量资料,首选 WMD;当纳入研究的结局指标单位不同或均数相差较大的资料时,选择 SMD,但结果解释要谨慎。而具体的临床解释见表 3-3。

表 3-3 各效应值的解读

| 效应值        | 无效线 / 界值 | 临床解释                                                                   |
|------------|----------|------------------------------------------------------------------------|
| OR/RR/HR   | 1        | 包含无效线 / 界值:差异无统计学意义                                                    |
| WMD/SMD/RD | 0        | 位于无效线 / 界值左侧:对不利结局有保护意义,对有利结局有损害意义;位于无效线 / 界值右侧:对有利结局有保护意义,对不利结局有损害意义。 |

(2) 森林图的解释和临床意义:在森林图(forest plot)中主要包含以下内容:①每一条横线代表一个试验结果的可信区间(confidence interval, CI),可信区间是效应值可能存在的范围,反映结果的准确性,横线越长,说明样本量越小,结果欠准确可靠;横线越短,说明样本量越大,准确性越高,结果越可信;②横线中央矩形的中线对应效应值的点估计值,矩形面积大小表示对 Meta 分析的贡献度,即研究权重(weight)。一般来说,对于计数资料采用样本量作为权重的衡量依据,样本量越大,权重越大;计量资料则采用标准差作为权重的衡量依据,标准差越小,权重越大;也有以纳入研究的质量评分作为权重的衡量依据;③最下方的菱形代表多个纳入研究的合并效应值,即 Meta 分析的结果;④垂直线即无效线将森林图分为左右两半,用于判断结果差异是否有统计学意义,横线、菱形与无效线相交则表明差异无统计学意义;若不相交则表明差异有统计学意义,完全在无效线左侧表明对不利结局有保护意义,对有利结局有损害意义,而完全在无效线右侧表明对有利结局有保护意义,对不利结局有损害意义(图 3-4)。



图 3-4 森林图的解读

(3) 漏斗图的解释和意义: 漏斗图 (funnel plot) 是最常用于判断是否有发表偏倚的方法, 也是 Cochrane Handbook 推荐使用的方法。所谓发表偏倚 (publication bias) 是指“统计学上有意义”的阳性研究结果较“统计学上无意义”的阴性研究结果或无效结果更易被发表。在医学研究中是一种相当普遍的现象, 其主要来源于作者、研究赞助者和杂志社编辑。而发表偏倚对 Meta 分析结果的真实性和可靠性有

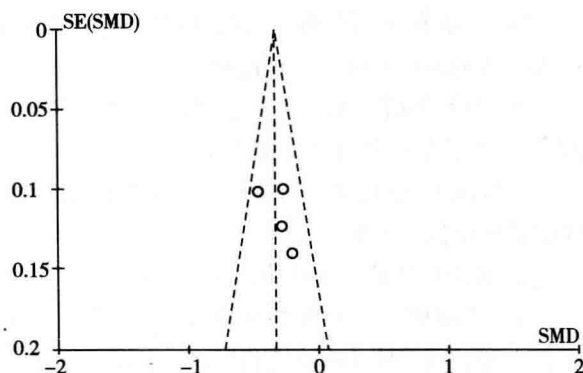


图 3-5 漏斗图

很大影响, 有可能会使本来没有统计学意义的 Meta 分析结果变为有统计学意义的结果。因此, 判断纳入文献是否存在发表偏倚也是一项十分重要的工作。

漏斗图是以效应大小为横坐标, 样本量或效应值标准误为纵坐标, 以“倒置漏斗”形状呈现的散点图, 以此来判断是否存在发表偏倚。(图 3-5) 如果没有发表偏倚, 漏斗图上的各个研究呈对称分布于中轴两侧, 对称轴即研究效应值的点估计值。小样本研究分布于“倒漏斗”的基底部, 随着样本含量的增加, 标准误降低, 大样本研究分布于“倒漏斗”的顶部。如果漏斗图不对称或不完整则提示可能存在发表偏倚。但也需注意, 导致漏斗图不对称的原因不仅仅是发表偏倚, 其他原因有: ①低质量小样本研究; ②真实异质性; ③假象; ④机遇。

漏斗图的最大优点是简单易行、直观, 缺点是当仅纳入几个小样本研究时, 漏斗图的检验效能受到极大限制, 因此, 制作漏斗图一般要求研究数在 10~15 个以上; 而且漏斗图只能使研究者认识到问题的存在, 但不能提供解决问题的办法。

## 第四节 文献阅读与评价

### 一、文献阅读与评价 1

Jiang L, Su L, Cui H, et al. Ginkgo biloba extract for dementia: a systematic review. Shanghai Archives of Psychiatry. 2013, 25 (1): 10-21.

1. 背景 痴呆的疾病负担不断增加, 而且缺乏有效的治疗方法, 因此有些国家就推荐使用银杏叶提取物 (ginkgo biloba extract, GbE) 来治疗痴呆, 虽然有关 GbE 疗效的研究结果尚不一致。

2. 目的 银杏叶提取物对痴呆患者认知功能和日常生活能力改善作用的研究进行 Meta 分析。方法: 检索国内外数据库, 找出 1982 年 1 月至 2012 年 9 月发表的关于银杏叶提取物治疗 (不少于 22 周) 痴呆患者的随机安慰剂对照研究的文献报告。根据 GRADE 系统推荐的方法进行文献质量评估并提取资料。采用 RevMan 5.1 软件进行异质性检验、敏感性分析并评估发表偏倚。



对连续性变量的合并效应值采用标准均差(standardized mean differences, SMD)表示,对分类变量则采用相对危险度(relative risk, RR)表示,Meta 分析的合并结果采用森林图显示。

3. 结果 有 9 项研究共计 2578 例患者符合入组和排除标准。其中 6 项研究共计 1917 例患者纳入 Meta 分析,结果发现仅在样本年龄相对较低(平均年龄 75 岁以下)的研究中 GbE 在延缓认知功能衰退和防止日常活动能力下降方面优于安慰剂。组间脱落率以及治疗中总的不良事件发生率均无显著差异。然而,不同研究结果间存在明显的异质性(主要是因为研究对象的年龄差异),文献存在可能的发表性偏倚(大多数是医药公司资助的),因此总体证据强度属于“低”。

4. 结论 Meta 分析表明,现有对此重要问题的研究证据依然极其薄弱。GbE 对 75 岁以下存在痴呆的人群可能有效。需要大样本、安慰剂对照的随机研究来验证上述结果,今后的研究应当聚焦在程度较轻的痴呆(包括轻度认知功能障碍)人群,比较不同剂量 GbE 的效果,并且随访更长的时间(至少 1 年)。

5. 评价 这是一篇比较系统、全面的系统评价研究,按照系统评价的原则和要求,完整、恰当地进行了结果报告,并分析可能存在的偏倚。因此,我们用这篇研究作为示例,在本章中进行讲解。临床问题是银杏叶提取物是否能够改善痴呆患者认知功能和日常生活能力。

## 二、文献阅读与评价 2

Li C, Xia J, Wang J. Risperidone dose for schizophrenia. Cochrane Database Syst Rev, 2009, 7(4): CD007474.

1. 背景 利培酮作为临床常用的治疗精神分裂症药物,而如何在最佳治疗效果和最少不良反应中获得平衡是一项十分重要的工作。

2. 目的 研究利培酮在治疗精神分裂症和类精神分裂症中的剂量反应关系。

搜索策略和纳入标准:搜索在 Cochrane 精神分裂组研究注册系统中的所有参考文献;纳入所有相关的随机对照研究。

3. 数据收集与分析 两位作者独立提取纳入文献数据,如遇分歧则同第三位作者讨论决定。当所需数据不完整时,与原文献作者联系索取。对于同质的二分类变量,则在意向性分析基础上计算固定效应的 RR 值及 95% 可信区间;对于连续型变量,则计算加权均数差值。

4. 结果 研究发现对于极低剂量组(<2mg/d)的利培酮与其他剂量组比较,会有更多的患者因为无明显效果而结束治疗[ $n=456$ , 1 RCT, 同标准低剂量(4~6mg/d)比较 RR 12.48, 95%CI 1.43~4.30]。效果不显著主要反映在对精神状态的评估中。而当低剂量组(2~4mg/d)同标准较高剂量组(6~10mg/d)和高剂量组( $\geq 10$ mg/d)比较时,仍会有更多的患者因为效果不显著而退出研究(4~6mg/d:  $n=173$ , 2 RCT, RR 4.05, 95%CI 1.09~15.07;  $\geq 10$ mg/d:  $n=173$ , 2 RCT, RR 1.92, 95%CI 1.36~2.70)。对于标准较高剂量组(6~10mg/d)的结果更偏向于“没有明显症状改善”(  $n=272$ , 2 RCT, RR 2.26, 95%CI 0.81~6.34),而此组的不良反应(如 EPS)较高剂量组少( $n=261$ , 2 RCT, RR unspecified EPS 0.56, 95%CI 0.31~0.99)。

5. 结论 目前仍缺乏足够的证据来确定最佳临床剂量。试验质量评价显示有高估临床疗效的可能。其中标准低剂量组(4~6mg/d)似乎有最佳临床效果和最低不良反应。微弱证据证明低剂量组(2~4mg/d)对首发病人效果较好,而高剂量组( $\geq 10$ mg/d)临床效果并没有优于其他组且引起较多不良反应,尤其是运动障碍。极低剂量组(<2mg/d)似乎无治疗效果。有待更多试验对特定人群进行研究,以及开展利培酮长期最佳剂量的研究。

6. 评价 此文献是发表在 Cochrane Library 并完全按照要求完成的一篇标准系统评价。主要研究的临床问题是探索不同剂量的利培酮对于精神分裂症的治疗效果以及不良反应的评价。





### 三、文献阅读与评价 3

Li C, He Y, Yu D, et al. Morita therapy for anxiety disorders (protocol). Cochrane Database of Systematic Reviews, 2010, Issue 7. Art. No.: CD008619.

1. 目的 搜集目前所有有关森田疗法治疗焦虑障碍的研究, 评价森田疗法治疗焦虑障碍的疗效和不良反应。

2. 评价 本文为一篇系统评价的研究计划书(protocol), 通过阅读此文献, 学习系统评价科研计划书的形式和内容, 为今后制作系统评价奠定基础。

#### 思考题

1. 系统评价包括哪些效应指标? 分别适用于哪些研究设计类型和数据资料类型?
2. Meta 分析时, 如何选择固定效应模型和随机效应模型?
3. 系统评价偏倚风险评估的标准有哪些? 分为几级?
4. 结合文献阅读, 试述“森林图”包含哪些信息?
5. 结合自己专业, 对于工作中遇到的临床问题, 尝试检索是否有相关的系统评价能够解决此问题。如果没有, 尝试按照阅读文献中科研计划书(protocol)的例子, 写一份临床问题的科研计划书。

(李春波 李伟)

#### 附: RevMan 软件操作示例

以第三章系统评价的临床问题, 应用 RevMan 软件进行系统评价制作示例。

通过在 Cochrane 协作网对 RevMan 软件进行下载, 目前版本为更新于 2014 年 3 月 28 日的 RevMan 5.2.10, 下载地址: <http://tech.cochrane.org/revman/download>

下载后按照指令对软件进行安装, 安装完毕后双击打开, 进入到 RevMan 5 程序的主界面(图 3-6)。

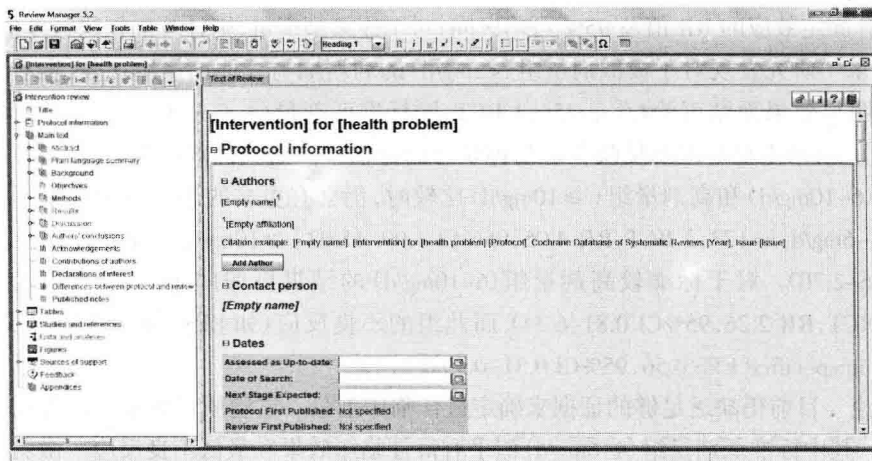


图 3-6 RevMan 5 的界面

#### (一) 创建新的系统评价

1. 选择菜单“File” — “New”, 弹出“New Review Wizard”对话框, 点击“Next”。
2. 在“Type of Review”中选择系统评价的类型(图 3-7), 最常用的为创建 Cochrane 干预评价。选择后, 点击“Next”。

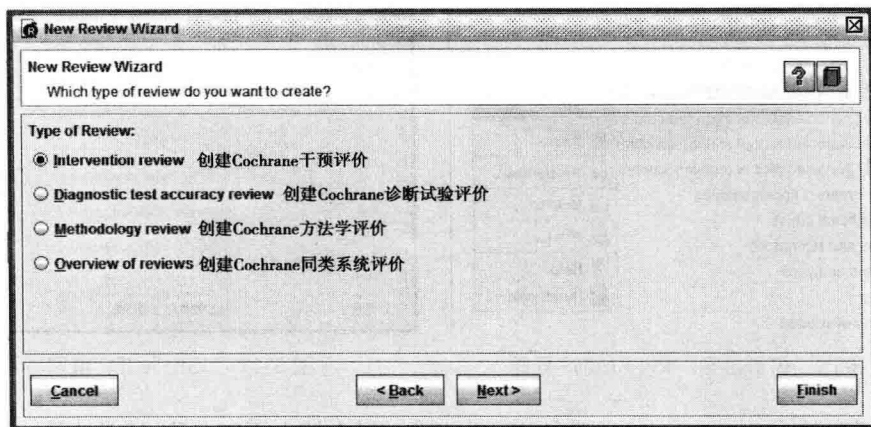
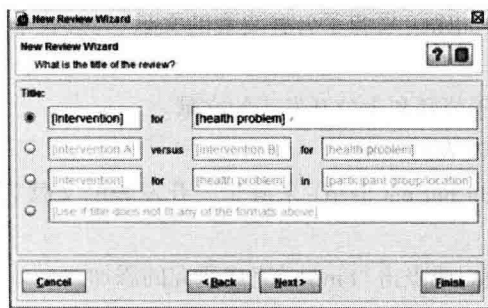


图 3-7 系统评价创建类型的选择界面

3. 在“Title”中输入系统评价的标题(图 3-8),输入完成后,点击“Next”。



某个干预措施对某个健康问题的影响  
A干预措施与B干预措施相对某个健康问题的影响  
某个干预措施对某个人群的某个健康问题的影响  
自定义

图 3-8 系统评价标题的输入界面

4. 在“Stage”中选择系统评价的阶段(图 3-9),一般选择全文阶段。选择后,点击“Finish”,“New Review Wizard”对话框关闭。

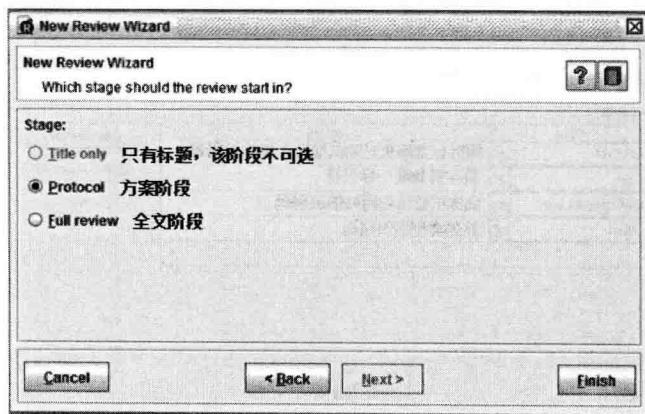


图 3-9 系统评价阶段的选择界面

## (二) 添加新的纳入研究

1. 在 RevMan 界面左侧的树形目录中点击“Tables”旁边的 ➤ 钥匙状图标,展开“Tables”下的 3 个子目录,再点击“Characteristics of studies”旁的 ➤ 钥匙状图标,展开其下的 4 个子目录。
2. 右键单击“Characteristics of included studies”,在弹出菜单中选择第一项“Add Study”(图 3-10),弹出“New Study Wizard”窗口。

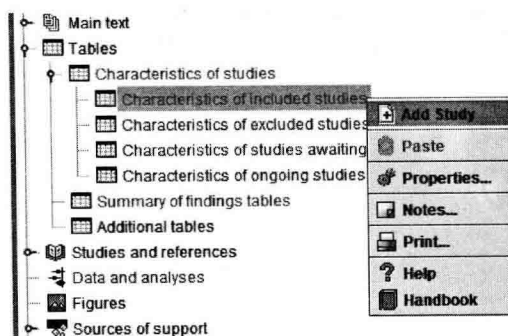


图 3-10 树形目录“Add Study”界面

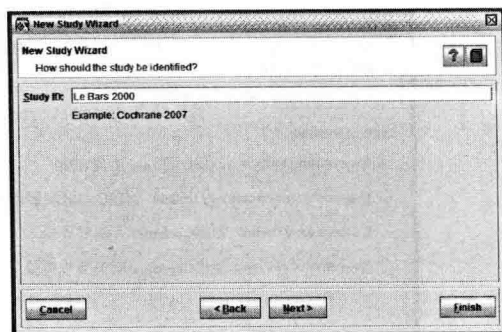


图 3-11 “Study ID”填写界面

3. 在弹出的“New Study Wizard”的“Study ID”一栏中填入研究名称,通常为第一作者名称加发表时间。输入完成后点击“Next”添加研究的特征信息,或者点击“Finish”完成研究的添加(图 3-11)。

4. 如点击“Next”,在“Data Source”中选择数据的来源

(1) Published data only (unpublished not sought):来源于公开发表的文献(为查找未公开发表的文献)。

(2) Published and unpublished data:来源于公开和未公开发表的文献。

(3) Unpublished data only:来源于为公开发表的文献。

(4) Published data only (unpublished sought but not used):来源于公开发表的文献(查找到未公开发表的文献,但未使用)。

选择后,点击“Next”添加研究的特征信息,或点击“Finish”完成研究的添加。

5. 如点击“Next”,在“Year”中填入研究的发表时间。输入完成后点击“Next”添加研究的特征信息,或者点击“Finish”完成研究的添加。

6. 如点击“Next”,在“Identifiers”中选择研究的识别码(图 3-12),添加完成后点击“Next”添加研究的特征信息,或者点击“Finish”完成研究的添加。

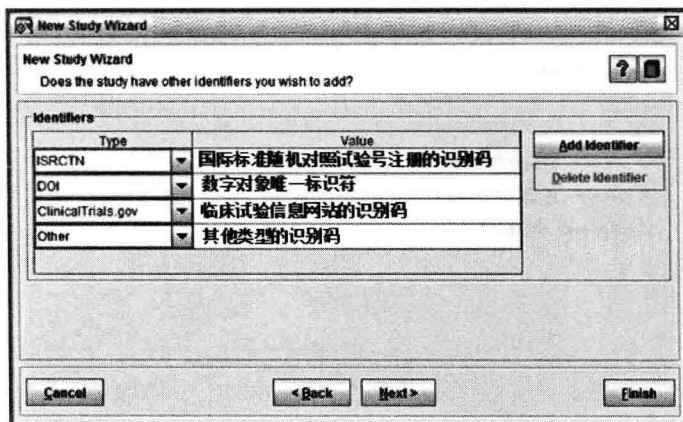


图 3-12 “Identifiers”研究识别码添加界面

7. 如点击“Next”,需要选择完成该研究添加后的下一步行动:选择“Nothing”后,点击“Finish”完成研究的添加;选择“Add a reference for the new study”后,点击“Continue”,在“New Reference Wizard”中完成参考文献信息;选择“Add another study in the same section”后,在“New Study Wizard”中继续按照以上的步骤纳入新的研究。



### (三) 添加新的比较

1. 在左边的树形目录中,右键单击“Data and analyses”,在弹出菜单中选择第一项“Add Comparison”,弹出“New Comparison Wizard”向导窗口(图 3-13)。

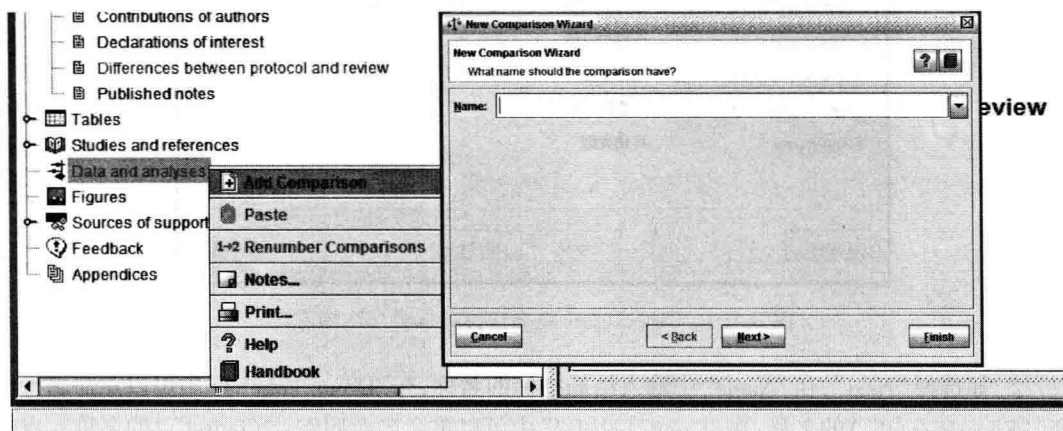


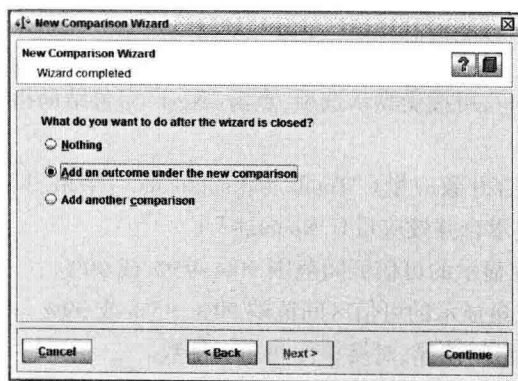
图 3-13 添加新比较向导窗口

2. 在“Name”中输入比较的名称,通常为治疗措施比对照措施,输入完成后,点击“Next”完善比较信息,或点击“Finish”完成比较的添加。

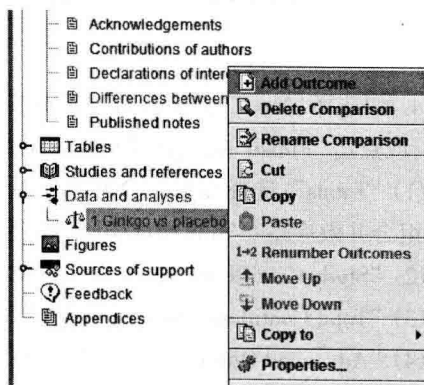
3. 如选择“Next”,需要选择完成该比较的添加后的下一步行动:选择“Nothing”后,点击“Finish”完成研究的添加;选择“Add an outcome under the new comparison”后,点击“Continue”,在“New outcome wizard”中完成结局指标的添加;选择“Add another comparison”后,在“New Comparison Wizard”中继续重复以上步骤的操作。

### (四) 添加结局指标

添加结局指标有两种方法:添加完比较后选择“Add an outcome under the new comparison”后,点击“Continue”(图 3-14A);在左边的树形目录中点击“Data and analyses”旁边的  钥匙状图标,展开其下的子目录,右键单击比较名称,在弹出菜单中选择第一项“Add Outcome”(图 3-14B)。均会弹出“New Outcome Wizard”向导窗口。



A



B

图 3-14 添加结局指标的方法

1. 在弹出的“New Outcome Wizard”向导窗口中,选择“Data Type”即结局指标的数据类型,最常用的数据类型为二分类变量和连续型变量。选择后,点击“Next”完善结局指标信息,或点击“Finish”完成该结局指标的添加(图 3-15)。

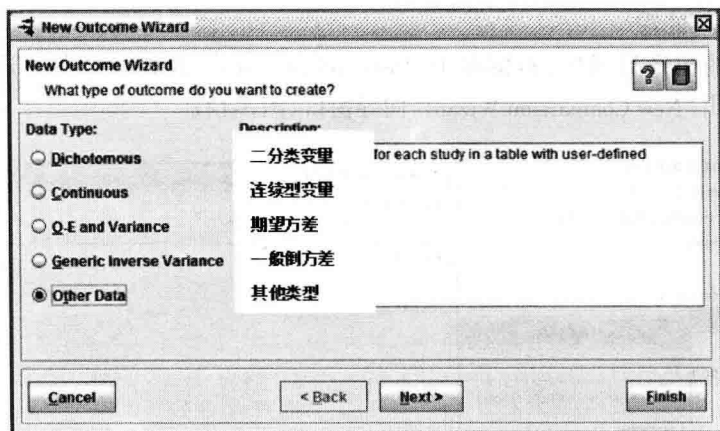


图 3-15 “Data Type” 结局指标数据类型选择窗口

2. 如选择“Next”,在“Name”中输入结局指标的名称,在“Group Label 1”中可以重命名干预组(“Experimental”)的名称;在“Group Label 2”中可以重命名对照组(“Control”)的名称。点击“Next”完善结局指标信息,或点击“Finish”完成该结局指标的添加。

3. 如选择“Next”,弹出分析方法选择界面(图 3-16),可接受默认选项,点击“Next”完善结局指标信息,或点击“Finish”完成该结局指标的添加。

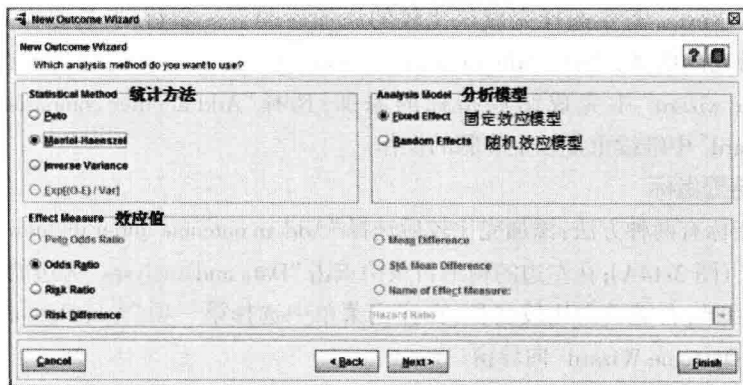


图 3-16 分析方法界面

4. 如选择“Next”,弹出分析方法细节选择窗口,可接受默认选项,点击“Next”完善结局指标信息,或点击“Finish”完成该结局指标的添加。

(1) “Totals”:选择显示总合并效应量和亚组合并效应量(“Totals and subtotals”)结果、只显示亚组合并效应量(“Subtotals only”)结果或不显示合并效应量(“No totals”)。

(2) “Study Confidence Interval”:选择纳入研究显示的可信区间范围 90%、95% 或 99%。

(3) “Total Confidence Interval”:选择合并效应量显示的可信区间范围 90%、95% 或 99%。

(4) “Advanced Options”:高级选项,可检验亚组间差异、对换事件和非事件数。

5. 如选择“Next”,弹出森林图细节窗口。可在“Left Graph Label”中更改森林图中线左侧图例,在“Right Graph Label”中更改森林图中线右侧图例,在“Scale”中更改森林图数据显示范围,在“Sort By”中更改纳入研究的排列顺序。选择完成后,点击“Next”完善结局指标信息,或点击“Finish”完成该结局指标的添加。

6. 如选择“Next”,需要选择完成该结局指标添加后的下一步行动:选择“Nothing”后,点击“Finish”完成结局指标的添加;选择“Edit the new outcome”后,点击“Finish”弹出结局指标编辑界面,可进行编辑;选择“Add a subgroup for the new outcome”后,点击“Continue”弹出“New





Subgroup Wizard”,可进行添加;选择“Add study data for the new outcome”后,点击“Continue”弹出“New Study Data Wizard”,可进行添加;选择“Add another outcome for the same comparison”后,点击“Continue”弹出“New Outcome Wizard”继续以上过程操作(图 3-17A)。

#### (五) 添加结局指标数据、绘制图表

添加结局指标数据有两种方法:添加完结局指标选择“Add study data for the new outcome”后,点击“Continue”(图 3-17A);在左边的树形目录中点击“Data and analyses”左侧的 钥匙状图标,展开其下的子目录,点击比较名称左侧的 钥匙状图标,展开其下结局指标子目录,右键单击结局指标名称,在弹出菜单中选择第二项“Add Study Data”(图 3-17B)。均会弹出“New Study Data Wizard”向导窗口。

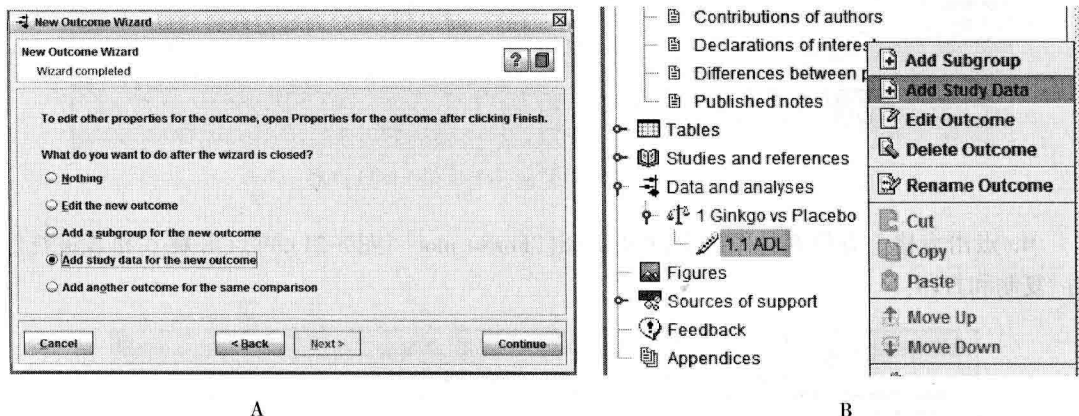


图 3-17 添加结局指标数据的方法

1. 在弹出的“New Study Data Wizard”向导窗口中“Included Studies”栏选中添加数据的名称。点击“Finish”将选中研究加入右侧编辑界面(图 3-18)。

2. 在右侧编辑界面的表格中输入相应的数据(图 3-19)。

3. 效应值和统计模型的选择

(1) 二分类资料的效应值:表格上方的效应值选择键,可在 **OR** **RD** 和 **RR** 中进行选择。

(2) 连续型资料的效应值:表格上方的效应值选择键,可在 **MD** 和 **SMD** 中进行选择。

(3) 统计模型:表格上方的统计模型选择键,可在 **FE**(固定效应模型)和 **RE**(随机效应模型)中进行选择。

(4) 在数据输入以及效应值和统计模型选择完成后,即可在表格中获得相应效应模型的并效应量(图 3-20)。

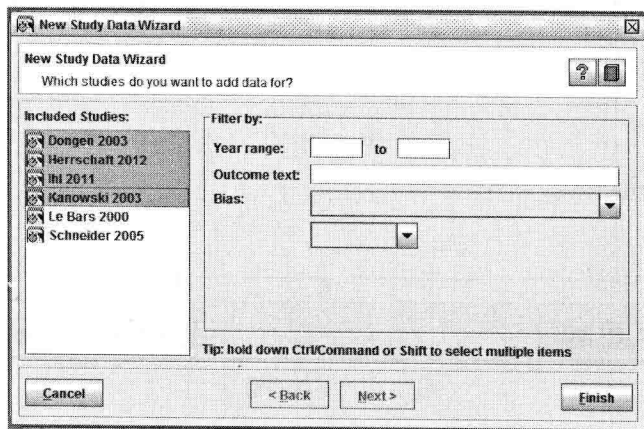


图 3-18 “New Study Data Wizard”纳入研究选择界面



| Study or Subgroup                                   | Ginkgo |    |       | Placebo |    |       | Weight | Mean Difference   | Year |
|-----------------------------------------------------|--------|----|-------|---------|----|-------|--------|-------------------|------|
|                                                     | Mean   | SD | Total | Mean    | SD | Total |        | IV, Fixed, 95% CI |      |
| <input checked="" type="checkbox"/> Le Bars 2000    | 0      | 0  | 0     | 0       | 0  | 0     |        | Not estimable     | 2000 |
| <input checked="" type="checkbox"/> Kanowski 2003   | 0      | 0  | 0     | 0       | 0  | 0     |        | Not estimable     | 2003 |
| <input checked="" type="checkbox"/> Ihl 2011        | 0      | 0  | 0     | 0       | 0  | 0     |        | Not estimable     | 2011 |
| <input checked="" type="checkbox"/> Herrschaft 2012 | 0      | 0  | 0     | 0       | 0  | 0     |        | Not estimable     | 2012 |
| Total (95% CI)                                      |        |    | 0     |         |    | 0     |        | Not estimable     |      |
| Heterogeneity: Not ap...                            |        |    |       |         |    |       |        |                   |      |
| Test for overall effect: ...                        |        |    |       |         |    |       |        |                   |      |

图 3-19 数据输入界面

| Study or Subgroup                                                                                       | Ginkgo |      |       | Placebo |      |       | Weight | Std. Mean Difference |  |
|---------------------------------------------------------------------------------------------------------|--------|------|-------|---------|------|-------|--------|----------------------|--|
|                                                                                                         | Mean   | SD   | Total | Mean    | SD   | Total |        | IV, Random, 95% CI   |  |
| <input checked="" type="checkbox"/> Le Bars 2000                                                        | -0.05  | 0.42 | 138   | 0.07    | 0.41 | 132   | 21.2%  | -0.29 [-0.53, -0.05] |  |
| <input checked="" type="checkbox"/> Kanowski 2003                                                       | -0.8   | 2.08 | 106   | -0.4    | 2.01 | 99    | 16.1%  | -0.19 [-0.47, 0.08]  |  |
| <input checked="" type="checkbox"/> Ihl 2011                                                            | -0.2   | 0.72 | 202   | 0       | 0.72 | 202   | 31.7%  | -0.28 [-0.47, -0.08] |  |
| <input checked="" type="checkbox"/> Herrschaft 2012                                                     | -0.11  | 0.36 | 200   | 0.04    | 0.29 | 202   | 31.0%  | -0.46 [-0.66, -0.26] |  |
| Total (95% CI)                                                                                          |        |      | 646   |         |      | 635   | 100.0% | -0.32 [-0.43, -0.21] |  |
| Heterogeneity: Tau <sup>2</sup> = 0.00; Chi <sup>2</sup> = 2.92, df = 3 (P = 0.40); I <sup>2</sup> = 0% |        |      |       |         |      |       |        |                      |  |
| Test for overall effect: Z = 5.73 (P < 0.00001)                                                         |        |      |       |         |      |       |        |                      |  |

图 3-20 随机效应模型合并效应量 SMD 界面

4. 点击表格上方的森林图绘制键,弹出“Forest plot”(图 3-21),可点击和进行保存、复制和打印。

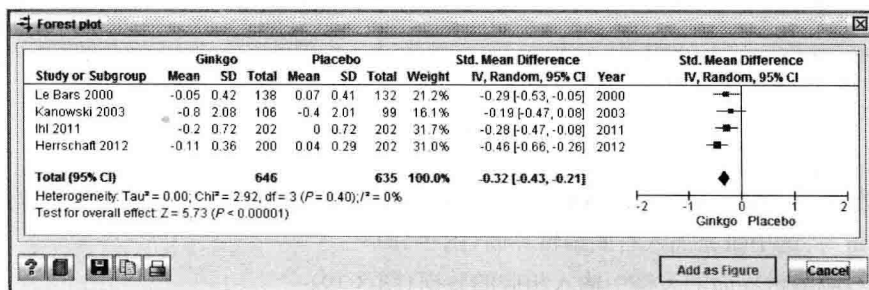


图 3-21 森林图窗口

5. 点击表格上方的漏斗图绘制键,弹出“Funnel plot”(图 3-22),可点击和进行保存和复制。

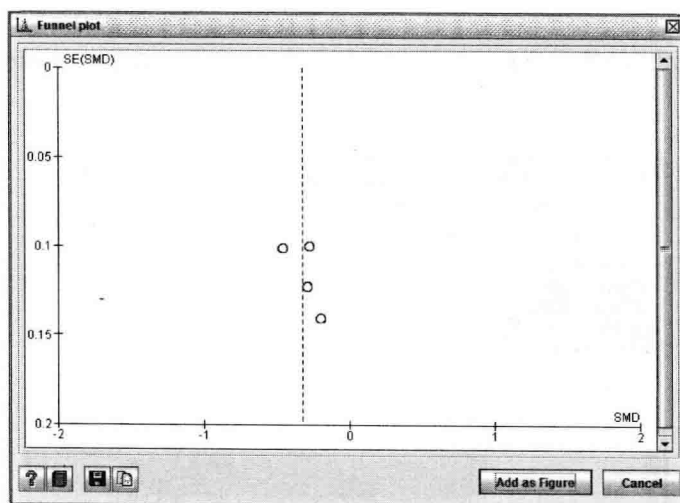


图 3-22 漏斗图窗口

(李春波 李伟)

## 第四章 病例报告和系列病例分析

### 第一节 病例报告

#### 临床案例

患儿女,9岁3个月,因“双下肢水肿1周”入院。患儿一周前无明显诱因开始双下肢水肿,并逐渐加重,尿量减少,尿液泡沫较多。在当地医院尿液检查:尿蛋白++++。入院查体:患儿贫血貌,双眼睑及双下肢凹陷性水肿。呼吸音清,肝、脾不大,全腹无压痛,移动性浊音阴性。左侧腹股沟处有一手术瘢痕。辅助检查:血常规示白细胞 $5.7 \times 10^9$ ,血红蛋白73g/L,平均红细胞体积65fl,平均红细胞血红蛋白含量17.4pg,平均血红蛋白浓度267g/L,CRP 170mg/L。尿常规:尿蛋白++++,24小时尿蛋白定量8.94g。血清总蛋白48.5g/L,白蛋白17.1g/L,球蛋白31.4g/L,胆固醇5.74mmol/L,三酰甘油(甘油三酯)1.89mmol/L,肝肾功能正常。胸腹腔CT:盆腔左侧一包块影,大小约66mm×54mm,右肾可见一圆形低密度影,边缘清晰,直径约6mm,肿块病理检查:经典型霍奇金淋巴瘤。手术后3天患儿尿蛋白转阴,血浆白蛋白逐渐升至正常,后续进行化疗。患儿出院诊断:1. 肿瘤相关性肾病综合征;2. 经典型霍奇金淋巴瘤。

#### 一、病例报告的基本概念

临床医生在工作中常常会遇到许多症状和体征十分不典型的疾病,甚至是少见病例或罕见病例,往往导致延误诊断和治疗。因此,临床医生在实际工作中需要不断总结这些个别病例诊断和治疗的临床经验和教训,这对于快速提高年轻医生的临床诊治水平十分重要。在前述临床病例中,该患儿确诊为霍奇金淋巴瘤相关性肾病综合征。通过文献检索,儿童霍奇金淋巴瘤相关性肾病综合征非常少见,国际上仅有数例报道,国内无相关的病例报道。因此,需要将该病例进行总结和整理,以提醒临床医生在诊断儿童肾病综合征时需注意排除是否为继发性,特别是继发性肿瘤。由于病例报告篇幅较小,写作相对容易,但却收到很好的效果,年轻医生均需掌握病例报告写作方法。

病例报告(case report)又叫个案报告,是临床工作中罕见的、特殊的病例,或是已知疾病的特殊临床表现、影像学及检验学等诊断手段的新发现、疾病的特殊临床转归、临床诊断治疗过程中特殊的经验和教训,值得临床医生深入研究讨论的病例,经归纳、总结和整理,以论文形式来报告单个或几个病例的一种文体。病例报告往往是识别一种新的疾病或暴露的不良反应的第一个线索,例如孕妇服用沙立度胺(thalidomide)引起新生儿先天畸形,口服避孕药增加静脉血栓栓塞的危险等。因此,病例报道对于认识临床上的少见病、罕见病,发现和掌握疾病诊治过程中的特殊性,以及为进一步研究这类疾病提供临床资料,都有其一定的意义。

#### 二、病例报告的写作步骤

1. 病例选择 个案顾名思义就是指个别病例,一例两例即可。但个案报道的病例选择



有一定范围,并非临床上所有的病例均值得分析报告。个案报道的病例可以是新发现的病例、对某种疾病的临床特点有新发现的病例、常见疾病出现异常现象的病例或者是虽难治但治愈的病例。如何发现这类病例,不仅需要有扎实的专业知识和丰富的临床经验,同时还需要临床医生详细询问病史,仔细查体,认真总结临床工作中的各种发现,不遗漏任何蛛丝马迹。

以下6类病例可撰写成病例报告:

①罕见及独特的疾病或综合征,且以往未曾被描述过;②未曾预料到的两种或两种以上无因果关系的疾病出现在同一病例并产生相互影响;③不同于某种疾病常规模式的、新发现的、具有临床重要性变化的病例;④以往未曾认识到的、有效的治疗方法,治愈一种顽症或过去认为“不治之症”的经验,取得前所未有的疗效或出现严重的药物不良反应;⑤可提供某种疾病发病机制线索的病例;⑥某种假象造成诊治失误教训的病例。

2. 临床资料 选择好病例后,应尽量在患者出院之前收集好资料,包括现病史、过去史、个人史和家族史,体格检查,影像资料,手术及病理、生化检查结果等。收集的临床资料应尽可能的详尽,对有意义的图片资料,如心电图、胸片、病理图像等,通过数码相机拍摄或扫描存储,以备日后查阅或发表时采用。通常配有少见、典型和清晰的图像资料的病例报告更容易被杂志社所采用。

3. 文献检索 收集好临床资料后,需要进行文献检索。通过文献检索,查阅该病例的国内外报道情况、临床价值、该病例与现有病例相比特别之处等,同时还要查阅了解当前该疾病的发病机制、诊断方法、治疗措施等方面的进展。可通过各种检索工具进行检索,如PubMed、EMBASE、CA、中国知网等专业性门户网站,国内外专业期刊网站,中外文电子期刊数据库,国家图书馆以及搜索引擎,如google、baidu等。目前国内有三种主要的全文数据库,即中国知识基础设施工程网、中国科技期刊数据库和万方数据资源系统。

### 三、病例报告的写作方法

病例报告的撰写也应像其他题材的生物医学论文的写作一样,除了按照经典病例报告的常用格式外,尚需按照所投稿杂志的投稿须知做相应的调整,同时,可以参考拟投稿杂志已经发表的病例报告格式。

病例报告的常用格式有两种,一种为病例报告的典型格式,即由引言(introduction)、病例报告(case report)、讨论和结论(discussion and conclusion)、参考文献(reference)四部分组成;另一种为研究论文的标准格式,即IMRAD格式,包括前言、方法、结果和讨论(introduction, methods, results, and discussion),主要用于病例观察伴有实验室研究的病例。考虑到大部分临床医师的需要,本章仅介绍典型病例报告的写作格式。

1. 引言 简要交待有无类似病例的报道,该病例在诊断和治疗上的困难和意义,该病的危害和预后,以及该病的特殊性等方面的内容。引言部分需尽可能简洁,切忌过多和过于详细。部分杂志也可以没有引言。

2. 病例报告 病例报告是论文的主体,因此临床资料应尽可能地详细。一般按以下顺序描述,主诉、病史、体检结果、实验室检查结果、其他检查结果、治疗及疗效。病例描述应集中与所报告主题切实相关的方面,而不是泛泛地交待患者的所有症状及体征。一般采取倒叙的方式描述病例,但应注意时间交待必须清晰无误。如果是遗传性疾病,则有关患者家庭成员的观察及检查资料也应在此交待。另外,应尽量避免使用可辨认患者身份的语言和照片,若确实需要使用,则必须获得患者或其亲属知情同意的书面材料。



病例报告的内容包括:

①一般资料,包括姓名、性别、年龄、住院号等以表明资料的真实性,但在期刊发表时,通常将姓名和住院号省略;②与该病有关的过去史、个人史和家族史;③重要的、特殊的或有鉴别诊断意义的临床症状、体征、辅助检查结果;④疾病的演变过程和治疗经过;⑤治疗结果及预后。

病例报告的内容中,除患者的年龄、性别、病程、症状、体征、检查及治疗情况等,必要时还要说明籍贯、住地(疾病的流行病学具有强烈的地域性,不同地域的实验结果可得出不同的科学结论,不介绍地名会影响论文的参考价值)、种族及职业等。有季节性的疾病尤其不要忽略发病日期及住院时间。病情材料的介绍要有所选择,无关材料不必常规地罗列。总之,要选择与诊断和治疗有关的特征性内容。当然也要包括虽与本病无直接关系,但对排除其他相关疾病有意义的其他症状和检查结果。临床治疗过程的叙述主要包括:①处理措施:治疗用药物的名称、剂量,超常规超常量使用的,尤其要表明;②病情变化:时间、程度、对处理的反应及是否出现并发症;③治疗结果:患者是好转或痊愈还是死亡,死亡的时间或存活者的随访情况,预后如何,还需接受何种处理。

3. 讨论和结论 讨论和结论是论文不可缺少的部分,讨论的内容根据报道的病例内容不同而有所不同。主要包括以下几个方面:①回答为什么该病例值得报告,并对该病例的罕见性及未预料性的论据进行讨论;②讨论该病例与以往已发表资料的关系,包括相反的观点;③对病例特征的可能解释;对临床研究及治疗的提示作用,包括不良的药物反应及药物的相互作用。一般认为“结果”部分是研究论文(即论著)的核心,而对于病例报告来说,其核心则是它的讨论和结论。这是因为一例患者的治疗结果并不决定文章的价值,病例报告的价值在于要通过对特殊病例的讨论,来产生对该病的新认识、提出新理论和展示诊治这种疾病的新前景。一般来说,国内外生物医学期刊均不允许病例报告单独列出“结论”,但临床医师可在“讨论”的最后部分对所讨论主题的未来研究及对临床实践的提示作用提出建议。

4. 参考文献 由于病例报道多为罕见或新发现的疾病,因此参考文献相对来说比研究论文少,一般1~10条不等。部分杂志会将参考文献省略,但在学术会议上发表这类论文时不宜省略。

病例报告写起来不难,难就难在确定该病例是否真“新”或具有什么新意。要确定这一点,除了报道者的丰富临床经验和阅历,还要充分而详尽地查阅文献。但即使没查到类似的报道,在写作过程中,也要注意用词不能过于武断,像“a new syndrome”、“the first report”之类的词,还是慎用为好。

## 第二节 系列病例分析

### 临床问题

C1q肾病是指肾小球系膜区有显著C1q沉积,并排除系统性红斑狼疮及Ⅰ型膜增生性肾炎的一种临床少见的免疫病理诊断的肾小球肾炎,1985年由Jennette等首先报道并命名。迄今为止,全世界文献报道的C1q肾病累计不足200例,国内仅有零星报道,对儿童C1q肾病的发病率、临床及病理特点、对激素的治疗反应、治疗效果和长期预后与临床及病理改变的关系等尚不清楚。我们回顾性分析了某院8年来经肾活检确诊的23例C1q肾病患儿的临床、病理特点及治疗方法,以进一步提高对该病的认识。





## 一、系列病例分析的概念

系列病例分析(case-series analysis)是回顾性的病例研究,一般来说其病例来源于一个医疗机构或由一位临床医师所观察并收集,就某一少见或较少报道病例,连续观察其病因,临床表现,治疗,预后等的变化,通过综合、分析、对比、研究上述病例相关资料,寻求某些方面的共同规律,总结撰写成文以指导、服务临床。如前述临床问题中(以下简称“儿童 C1q 肾病”),我们通过分析近年来确诊的所有儿童 C1q 肾病病例,探讨其临床、病理特点及治疗方法,使临床医生进一步提高对该病诊断和治疗的认识。

## 二、系列病例分析的格式

系列病例分析没有固定的格式,目前国内外生物医学期刊主要采用三种常用格式:病例报告的格式、研究论文的标准 IMRAD 格式、综述的格式。

病例报告格式的写作方式类似于病例报告,在此格式中必须对“系列病例”(也可包括文献中的相似病例)进行综合概括,且“系列病例”必须具有同样的明确诊断或均是采用同样干预治疗的患者,如外科手术患者或某一种药物治疗的数例患者。分析内容可包括干预治疗的不良反应、手术后并发症或其他异常反应。然而病例选择的偏颇、疾病发展阶段的不均衡性,以及缺乏对照数据则均可能限制“病例分析报告”的临床价值。

研究论文的标准 IMRAD 格式,包括前言、方法、结果和讨论。一般是临床医师为了回答就其所观察的病例及文献中所报告的病例提出问题,而进行系列病例分析研究,并且其病例分析研究遵循严格的设计程序(提出特定的问题或假设、有明确的命题、有严格的对照数据、且数据均经统计学处理),则可以采取研究论文的写作格式。目前,国内外期刊发表的系列病例报告大多为此种格式。

综述的格式,是在临床医师观察并收集的系列病例在他人已发表的文章中已被描述过,但临床医师想结合自己的病例进行联合分析,则临床医师可将其所观察收集的病例与文献上的病例综合起来,撰写综述性文章。综述性文章一般由前言、方法、主体部分、结论和参考文献 5 部分组成,较长的综述还需提供摘要。

## 三、系列病例分析写作方法

考虑到国内外期刊常见的系列病例分析形式为研究论文的格式,这里重点介绍该类型系列病例分析的写作方法。其结构包括标题、引言、对象和方法(含临床资料、治疗方法、疗效观察等)、结果、讨论、结论等几个部分。

1. 标题 标题要简短明了、开门见山、直扣主题。如“某药物治疗某疾病若干例疗效分析”、“某疾病若干例临床观察”等。在前述病例中,我们采用的标题为“儿童 C1q 肾病的临床病理特点及治疗探讨”,通过标题,读者即可了解文章的主要内容。标题写作中需注意:不要使用虚词,如“关于”、“对于”等;标题开头避免使用阿拉伯数字;避免或少用副标题;避免使用非公知的缩略语。

2. 引言 概括性地介绍本论文的研究背景,引起读者的关注。其主要内容包括:简明扼要的交代该研究的目的,即要解决的具体问题;研究的背景,如疾病的概述、国内外的相关进展、目前存在的问题等;该论文资料的来源及选择特定研究方法的理由。必要时还可以简单介绍研究的结果及其意义。在“儿童 C1q 肾病”中,前言是这样描述的:C1q 肾病是指肾小球系膜区有显著 C1q 沉积,并排除系统性红斑狼疮(SLE)及 I 型膜增生性肾炎的一种临床少见的免疫病理诊断的肾小球肾炎,1985 年由 Jennette 等首先报道并命名。临床主要表现为 NS 或无症状性蛋白尿,且多数激素耐药。迄今为止,全世界文献报道的 C1q 肾病累计不足 200 例,国内仅有零星报道,主要见于儿童及青少年。我们回顾性分析了本院 8 年来经肾活检确诊的 23 例 C1q 肾病患儿的



临床、病理特点及治疗方法,以进一步提高对该病认识。在前言中,简要交代了C1q肾病的概念、国内外的研究现状以及本文的研究目的等。

3. 对象和方法 包括研究对象、治疗方法、疗效观察等,这些是论文撰写的基础。

(1) 研究对象:包括病例采集的时间、数量、患者的性别、年龄、职业、诊断标准、排除标准、发病诱因、临床症状和体征、合并症等。这些内容可根据文章有针对性地进行选择,但与疾病相关的资料一定不能缺少。如与发病季节相关的疾病一定要交代病例采集的时间;与地域相关的地方病一定要交代居住地;职业病要交代患者的职业等。如在“儿童C1q肾病”中“研究对象”是这样撰写的:2003年3月~2011年5月本院肾内科收治的经肾活检诊断为C1q肾病的患儿共23例,均符合Jennette等制订的C1q肾病诊断标准:肾小球系膜区有显著C1q沉积,按免疫荧光0~++++积分,C1q的沉积强度必须 $\geq ++$ ,且呈弥漫性沉积。排除系统性红斑狼疮(SLE)及I型膜增生性肾炎。

(2) 治疗方法与疗效观察:包括治疗方案、药物剂量、给药方法、疗效评价标准、随访情况(包括随访时间、随访人数、失访人数、随访结果)、观察终止标准等。

4. 结果 结果是论文的核心,根据研究目的进行归纳、总结,以文字、图表的形式展现研究结果。需要注意的是:结果中的计量单位需使用法定单位;文字与图表避免重复;统计结果要正确标注;避免使用不确定性的词语,如“多见”、“少见”、“很少”、“绝大多数”等,而要给出具体数据;与研究目的相关研究数据与结果,不论其是否与作者的观点相符,均必须保留,不得按照作者的论点片面取舍数据,否则将影响文章的真实性、可靠性,使论文失去价值。

5. 讨论与结论 讨论与结论是论文的精华,通过对研究结果进行详细深入的分析和讨论,突出研究的目的和重点。该部分的写作要求:需大量阅读文献,充分了解该疾病的国内外研究现状;讨论必须紧紧围绕研究结果展开,对于研究结果中没有的内容,不得妄加结论,对不能肯定的观点,或因研究方法、研究手段、病例数、随访时间等限制而不能得出确切的结论时,可采用“有待进一步研究”、“尚需进一步随访观察”、“有待于更大样本的研究”等。

### 思考题

1. 什么是病例报告?如何进行病例选择?
2. 病例报告的撰写步骤和方法有哪些?
3. 哪些临床问题适合病例报告?
4. 系列病例分析包括哪几个部分?各部分写作有哪些具体要求?

(张爱华)

## 第三节 循证实践案例报告

### 临床案例

患者,男性,66岁,以“反复咳嗽、咳痰20年,活动后气短4年,加重1个月”之主诉入院。曾行肺功能检查明确诊断COPD(重度),并多次因急性加重住院治疗。入院1个月前,受凉后出现咳嗽、咳少量白色黏痰,伴气短,活动后明显。该患者无其他基础疾病,至今已有50余年吸烟史,约20支/天。体格检查:T 36.0℃,P 76次/分,R 20次/分,BP 124/89mmHg。口唇发绀,球结膜无水肿,气管居中,桶状胸,肋间隙增宽,双侧呼吸动度减弱。双肺叩诊呈过清音,双肺呼吸音清,可闻及干、湿性啰音,未闻及胸膜摩擦音。心律齐,剑突下未见异常搏动,双下肢无水肿。辅助检查:血常规示WBC  $11.45 \times 10^9/L$ ,中性粒细胞84.11%;血气分析示pH 7.413,PaO<sub>2</sub> 59.5mmHg,PaCO<sub>2</sub> 41.8mmHg。胸部正位片提示慢性支气管炎肺气肿。



## 一、提出临床问题

慢性阻塞性肺疾病(chronic obstructive pulmonary disease, COPD)以进行性发展的不完全可逆的气流受限为特征,伴有肺组织对有毒颗粒和气体的异常炎症反应,可引起肺脏及肺外器官的损害。

慢性阻塞性肺疾病急性加重期(acute exacerbation of chronic obstructive pulmonary disease, AECOPD)是指在呼吸道感染、空气污染、中断治疗等诱因下,患者呼吸系统症状(咳嗽、咳痰、气短等)恶化(超出日常变化范围)。

气道炎症加重是 AECOPD 的重要触发因素,糖皮质激素具有强大的抗炎作用,可有效抑制炎症细胞迁移、活化,阻止炎症介质释放,减轻微血管渗漏和黏膜水肿,减少气道黏膜腺体过度分泌,降低气道高反应性,并能增强  $\beta_2$  受体兴奋剂的支气管舒张作用,在 AECOPD 的治疗中被广泛应用,效果肯定。急性加重期全身应用糖皮质激素可促进患者恢复、改善肺功能( $FEV_1$ )和低氧血症,还可减少早期复发,降低治疗失败率,缩短住院时间。但全身应用激素的副作用较多,吸入性激素由于用量小,且不经血液循环直接作用于气道局部,与相应受体结合发挥强大的局部抗炎活性,对全身各器官的影响较小,所以近年来雾化吸入糖皮质激素在 AECOPD 治疗中广泛应用,并已达成共识。

根据该病例的病史及临床检查结果,确诊该患者属于 AECOPD,现呼吸系统症状明显,血气分析提示低氧血症,如果治疗不及时,效果不理想可能会出现呼吸衰竭等严重后果。

GOLD 强烈推荐应用糖皮质激素作为 AECOPD 的常规治疗,但对于该老年患者而言,全身应用激素的副作用不容忽视,如何选择适当的给药途径,抑制病情进展的同时将不良反应降至最低,是治疗的关键问题。

以患者实际病情为基础,可从以下几个方面综合考虑,提出具有实际意义的临床问题:

1. 查阅文献和相关治疗指南了解治疗现状 查阅 COPD 治疗相关指南,发现目前国内外最新指南均推荐应用糖皮质激素作为 AECOPD 的常规治疗,但全身应用激素存在消化道出血、感染、血糖升高、体重增加等不良反应,且对于 COPD 患者而言,随着病情进展,急性加重的次数逐渐增多,可能将接受更高剂量的全身糖皮质激素,发生不良反应的可能性也将大大增加。

2. 患者及家属的期望 通过及时的沟通与交流,患者及其家属表示,现阶段最关心的是:是否应该加用糖皮质激素缓解症状,改善血气及肺功能?加用糖皮质激素是否会产生严重的不良反应?

3. 现有的相关临床研究证据 已开展相关临床试验,部分结果表明,布地奈德雾化吸入可有效、安全地替代全身激素用于 AECOPD 治疗。

4. 其他相关研究证据 支气管哮喘(以下简称哮喘)与 COPD 同属呼吸系统炎症性疾病,且均存在气流受限,虽然哮喘患者的症状可自行或治疗后缓解,但随着病情发展,气道结构重塑,其可逆性逐渐下降,临床上表现出与 COPD 相似的特征。多项研究证实,吸入激素能够有效控制哮喘急性发作期的气道炎症,改善气道高反应性和肺通气功能,缓解症状。

结合患者的病情及其期望、治疗现状、相关临床证据,医生在治疗过程中可以提出以下问题:①布地奈德雾化吸入在减轻患者症状、改善肺功能等方面是否与全身糖皮质激素具有相似疗效;②布地奈德雾化吸入的不良反应是否比全身糖皮质激素少。

## 二、将临床问题转化成可以回答的形式

为快捷有效地检索到与临床问题密切相关的证据,根据 PICO 原则,将初始临床问题转换成便于检索证据的形式。问题的构建见表 4-1。



表 4-1 PICO 法构建的问题

| 患者特征      | 干预       | 比较     | 结局      | 问题类型 | 设计类型      |
|-----------|----------|--------|---------|------|-----------|
| AECOPD 患者 | 布地奈德雾化吸入 | 全身激素治疗 | 疗效、不良反应 | 治疗   | RCT 或系统评价 |

### 三、证据检索

本例临床问题属于证据性研究范畴,最佳证据来源是多个随机对照试验(randomized controlled trial, RCT)的系统评价、Meta 分析或经过验证的临床实践指南,在没有上述证据的情况下,可依次使用单个 RCT、病例-对照研究、系列病例观察、专家意见的报告,但其可靠性逐级降低。

1. 检索资源 可通过互联网检索各种资料库,包括:①已经过整理的医学文献数据库,如 Cochrane library, Up to date, Clinical evidence, Best evidence (ACP journal club and evidence based medicine), Sumsearch, Tripdatabase; ②未经过整理的医学原始文献数据库,如 PubMed, MEDLINE、CNKI 等全文数据库。

2. 关键词及检索策略 中文检索词“慢性阻塞性肺疾病急性加重期、布地奈德、随机对照试验”;英文检索词“(acute exacerbation of chronic obstructive pulmonary disease or AECOPD) and (anti-inflammatory agents or glucocorticoid or corticosteroid or steroid or prednisolone or methylprednisolone or dexamethasone) and (inhaled corticosteroid or budesonide or pulmicort) and (RCT or meta-analysis or systematic review)”。

3. 检索结果 AECOPD 患者布地奈德雾化吸入与全身应用激素疗效对比的相关文献共检索 27 篇中文文献,2 篇英文文献,2 篇相关指南。

### 四、证据评价

应用循证医学评价防治性研究文献的质量标准,对单个研究文献的真实性、临床重要性及实用性进行评价。

#### 知识点

#### 单个临床治疗性证据的评价

应用单个临床治疗性证据之前,要用科学的方法和质量评价标准对其进行分析与评价,首先评价研究证据的内部真实性,如果是真实可靠的,则进一步分析与评价该治疗措施(证据)是否有临床治疗价值,即临床重要性;假设其不仅真实而且重要的话,则要分析这种证据(措施)有无适用性。

1. Maltais F, Ostinelli J, Bourbeau J, et al. Comparison of Nebulized Budesonide and Oral Prednisolone with Placebo in the Treatment of Acute Exacerbations of Chronic Obstructive Pulmonary Disease. Am J Respir Crit Care Med, 2002, 165 (5): 698-703.

#### 知识点

#### 单个临床治疗性证据科学性分析与评价

1. 证据是否来源于真正的随机对照试验。
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。
3. 是否对随机分组的所有研究对象进行了意向性治疗。
4. 是否对研究对象、医生和研究人员采用了盲法。
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。



上述文献系 2002 年启动的多中心、随机、双盲、空白对照研究,由比利时、加拿大、法国 3 个国家的 34 个研究中心参加,共纳入 199 名 AECOPD 患者,随机分为对照组(标准疗法:  $\beta_2$  受体激动剂、异丙托溴铵、口服抗生素、氧疗 + 雾化吸入及口服安慰剂)、布地奈德雾化吸入组(标准疗法 + 布地奈德 2mg qid+ 口服安慰剂)、口服泼尼松组(标准疗法 + 泼尼松 30mg bid+ 雾化吸入安慰剂),目的是明确 AECOPD 治疗中,布地奈德雾化吸入是否可替代口服泼尼松。

入选标准:COPD;近 14 天内有急性加重史;需要住院治疗;年龄  $\geq 50$  岁;吸烟史  $\geq 20$  包/年。

排除标准:合并哮喘、过敏性鼻炎或其他特应性疾病;近 1 个月内系统性应用皮质类固醇;吸入激素量  $\geq 1500\mu\text{g/d}$  倍氯米松(或其等效物);可能突发急性呼吸衰竭需要机械通气或进入重症监护病房;由于特殊原因如肺炎、气胸、心力衰竭等急性加重者。

该研究设计终点结果:主要终点结果:治疗 3 天后,患者  $\text{FEV}_1$  水平较基线值改变量;次要终点结果:治疗 3 天后,患者血气参数较基线值改变量、平均住院日以及不良事件发生情况。

科学性评价:试验采用了随机的方法对受试者进行分配,利用安慰剂实施双盲,失访率较低(14.07%),结果中进行了 ITT 分析,各组受试者的初始基线资料具有较好的可比性。总的来说,该 RCT 质量好,真实可靠。

#### 知识点

##### 治疗性证据的重要性评价

1. 治疗性证据的效应强度大小。
2. 治疗性证据精确度的评估。

学习  
笔记

该研究主要终点结果显示:①布地奈德雾化吸入组与口服泼尼松组  $\text{FEV}_1$  改善值较对照组明显增多( $P < 0.05$ ),但两激素组间无明显差异  $-0.06\text{ L} (-0.14 \sim 0.02\text{ L})$ ;②次要终点结果显示:布地奈德雾化吸入组与口服泼尼松组动脉血  $\text{PaCO}_2$  下降  $\geq 5\text{ mmHg}$  的患者比例,显著高于对照组( $P < 0.05$ ),且口服泼尼松组优于布地奈德雾化吸入组( $P < 0.05$ );两激素组平均住院日组间无差异( $P > 0.05$ );不良反应:口服泼尼松组中 7 例(10%)出现高血糖,其中 2 例需要胰岛素治疗,而雾化组仅 1 例(2%)血糖偏高。

重要性评价:本研究最终证实,在未出现酸中毒的 AECOPD 患者治疗过程中,布地奈德雾化吸入与口服泼尼松相比,在改善肺功能及缩短住院日方面无明显差异,且可减少不良反应发生,将来可替代口服泼尼松用于 AECOPD 的治疗。

2. Gunen H, Hacıevliyagil SS, Yetkin O, et al. The role of nebulised budesonide in the treatment of exacerbations of COPD. Eur Respir J, 2007, 29 (4): 660 - 667.

该研究系 2007 年启动的单中心、随机、对照试验;共纳入 159 名 AECOPD 患者,随机分为对照组(仅接受标准支气管扩张治疗)、泼尼松静滴组(标准支气管扩张治疗 + 泼尼松 40mg)、布地奈德雾化吸入组(标准支气管扩张治疗 + 布地奈德 1.5mg qid),目的是明确 AECOPD 治疗中,布地奈德雾化吸入是否可安全、有效地替代泼尼松静滴疗法。

入选标准:COPD;急性加重需住院治疗。

排除标准:由于特殊原因如肺炎、肺栓塞、心力衰竭、气胸等急性加重者;可能突发急性呼吸衰竭需要机械通气或进入重症监护病房;近 1 个月内系统性应用皮质类固醇或有急性加重史者。

观察指标:血常规;生化分析;肺功能;血气分析;早期(10 天)、延迟(超过 15 天)出院率;出院 1 个月内复发率、再住院率以及不良反应。

科学性评价:试验采用了随机方法对受试者进行分配,利用安慰剂实施盲法,但未具体说明盲法类型,失访率较低(23.09%),结果分析时直接排除失访者数据,未进行 ITT 分析,各组受试者的初始基线资料具有较好的可比性。总的来说,该 RCT 质量好,真实可靠。





结果显示:治疗 10 天后,布地奈德雾化吸入组与泼尼松静滴组  $\text{PaO}_2$ 、 $\text{SaO}_2$ 、 $\text{FEV}_1$ 、FVC 均明显改善,但两组差异无统计学意义( $P>0.05$ );在动脉血  $\text{PaCO}_2$  改善方面,泼尼松静滴组优于布地奈德雾化吸入组( $P=0.046$ );该研究还报道了早期(10 天)与延迟(超过 15 天)出院率、急性加重次数、出院 1 个月内的再住院次数,但两组间并无差异( $P<0.05$ )。

不良反应:泼尼松静滴组在治疗 7~10 天过程中血糖水平有明显升高趋势,而布地奈德雾化吸入组血糖水平在治疗过程中逐渐下降。

重要性评价:本研究证实,布地奈德雾化吸入可有效、安全地替代全身激素用于 AECOPD 治疗。

指南意见:美国胸科学会/欧洲呼吸学会(ATS/ERS)的 COPD 指南(2004 年),2013 年全球慢性阻塞性肺疾病防治倡议(GOLD)提出对于 AECOPD 患者,可考虑应用吸入激素代替全身激素,其中 GOLD 明确提出可采用布地奈德。

## 五、实施决策

### 知识点

#### 治疗性证据适用性的分析与评价

1. 被评价的证据是否与患者情况不符而不能应用。
2. 在外面的医疗环境里对拟采用的治疗证据是否可行。
3. 基于该治疗性证据,对患者可能产生的利弊进行评估。
4. 患者对于治疗措施的价值取向与期望。

经过资料检索、整理、分析,结合该患者的具体临床状况(老年、重度 COPD、多次急性加重史、呼吸系统症状明显,听诊双肺可闻及干湿性啰音、血气分析提示轻度低氧血症),以及医院现有药物供应情况,结合患者意愿及经济条件,我们给予患者布地奈德雾化吸入 3mg,3 次/日,同时给予吸氧、特布他林雾化、多索茶碱解痉、抗感染化痰等对症支持治疗。

## 六、后效评估

治疗第 3 天,患者诉咳嗽、气短症状较前减轻,治疗第 10 天,未出现声嘶、口干等不良反应,症状缓解,口唇无发绀、双肺干湿性啰音消失,复查血常规、血气分析恢复正常。患者及家属对目前治疗效果满意,远期疗效还在进一步观察随访中。

## 七、循证实践案例论文小结

本节通过介绍一个老年 AECOPD 患者的循证治疗,利用循证实践方法检索证据并评价其质量,较为真实地提供布地奈德雾化吸入的疗效及可能出现的副作用,为临床决策提供依据。案例论文小结描述为:AECOPD 是 COPD 患者就诊及住院治疗的主要原因,以气道炎症加重为主要触发因素,可加快肺功能减退、影响患者生活质量,增加经济支出及病死率。糖皮质激素具有强大的抗炎作用,是急性加重期的基础用药之一,但全身应用激素的副作用较多,目前雾化吸入糖皮质激素治疗在 AECOPD 治疗中广泛应用,并已达成共识。

### 思考题

1. 什么是循证实践?在临床工作中具体如何实施?掌握循证实践有什么重要意义?
2. 结合自己专业,选择一种存在处理矛盾或者问题的疾病,应用循证实践方法解决临床上的诊断、治疗或预后问题。内容包括:案例描述、提出临床问题、检索证据、评价证据、实施决策以及后效评价。

(刘璐 李满祥)

## 第五章 横断面调查研究

### 临床案例

患者,男性,35岁,某省居民。因最近夜晚小便次数增多来院就诊,自诉无高血压和糖尿病史,入院查体:血压130/80mmHg,体重80kg,身高172cm,空腹血糖6.9mmol/L,该患者就诊的目的是想确诊是否患了糖尿病,或者处于危险状态。

### 第一节 如何获得临床研究问题

#### 一、明确临床问题

糖尿病具有较高的发病率、致残率和致死率,但是发病早期大多数患者没有特异性症状,发病较隐匿,所以导致人群中无症状的患者失去了早期进行确诊和治疗的机会。但是很多研究表明,人体的一些测量指标与糖尿病的发病密切相关,可通过这些测量指标来识别高危人群。目前的研究大多运用受试者工作特征曲线(receiver operator characteristic curve, ROC)下面积(area under the curve, AUC)和约登指数(Youden index)来确定这些人体测量指标筛检糖尿病的价值和最佳切割点。那么对案例中的患者,我们应该根据哪些人体测量指标来判断其是否为糖尿病的高危人群?这些指标的最佳切割点分别是多少?在不同的性别之间是否有差异?针对这些迫切需要解决的问题,我们可以进行相关的横断面研究。

#### 知识点

##### 横断面研究

横断面研究(cross-sectional study)是在某个时点(或期间)对某个特定群体的个体的疾病(或某种健康状况)以及有关变量(因素)进行调查描述,可进一步比较分析具有不同特征的暴露与非暴露组的患病情况或患病组与非患病组的暴露情况,为研究的纵向深入提供线索和病因学假设。从观察时间上来说,收集资料是在特定时间内发生的,一般不是调查过去或追踪将来的暴露或疾病情况,所以又称为现况研究。从观察分析指标来说,一般为在特定时间内调查群体的患病频率,故也称为患病率研究(prevalence study)。

#### 知识点

##### 横断面研究的特点

1. 横断面研究开始时一般不设有对照组,最后在资料处理与分析阶段,才根据暴露(特征)的状态或是否患病的状态来分组比较。
2. 调查时间是某一特定时点上或某一特定时期内,理论上,这个时间应该越集中越好。
3. 横断面研究在确定因果联系时受到限制,大多仅能提供病因线索。



4 对于诸如性别、种族、血型等这类不会因是否患病而发生变化的因素,横断面研究可以提示相对真实的暴露(特征)与疾病的时间先后顺序的因果联系。

5. 在同一稳定人群中,定期重复连续进行同样的横断面调查可获得发病率、新发感染率和转归等资料。

结合本章案例,如何从临床问题中获得有意义、具有探索性或者证据性研究课题,即形成科学问题?

1. 查阅文献阅读相关背景资料 糖尿病患者和中间高血糖患者(即患者的血糖水平介于正常人与糖尿病患者之间的一种中间代谢状态)会由于体内的代谢紊乱而导致体内脂肪重新分布,表现为人体的一些测量参数的变化,主要特征就是脂肪的向心性分布。其中较为常用的3个测量指标包括体重指数(body mass index, BMI)、腰围(WC)和腰臀比(WHR)。中国是个糖尿病的高发国家,所以采用这些人体测量指标对糖尿病进行筛检具有很重要的公共卫生意义,那么在中国人人群中这些人体测量指标的价值和最佳切割点是多少就成为大家关心的问题。

2. 已有的临床研究证据 目前有很多研究均在不同的人群中对这3个指标的筛检价值和最佳切割点进行了研究,但是结果不尽相同,而且在中国人群中的研究,其结果也不一致,比如有研究认为男性腰围 $>85\text{cm}$ ,女性腰围 $>80\text{cm}$ ,可作为中国人筛检糖尿病的界值点(危险人群),而另外的研究则提示中国人的腰围最佳切割点为 $\geq 80\text{cm}$ ,而且男女无性别差异。前者可能漏检,后者可能需要花费更多人力和物力,这就提示我们应该对某一地区的人群进行针对性的研究,得到指标的筛检价值和最佳切割点。

## 二、将临床问题转换成科学问题

针对上述临床问题,我们可以通过不同的临床研究设计来回答,这就需要将临床问题转换为科学问题。

转换思路1:为了得到一个可以适用于筛检当地成年人糖尿病状态的3个人体测量指标的最佳切割点,并分析其筛检价值,我们可以对当地所有的成年人进行调查和检测,得到他们的糖尿病患病情况以及BMI、WC和WHR的数值,然后利用适当的统计学分析,得到这3个测量指标筛检糖尿病的价值大小和最佳切割点,并分析是否存在性别差异。

这种方法属于横断面研究中的普查,普查(census)即全面调查,是指在特定时点或时期、特定范围内的全部人群(总体)均作为研究对象的调查。这个特定时点应该较短,特定范围是指某个地区或某种特征的人群。

### 知识点

#### 普查的目的及优、缺点

##### 目的

1. 了解人体各类生理生化指标的正常值范围。
2. 早期发现、早期诊断和早期治疗患者。
3. 了解慢性疾病的患病及急性传染性疾病的疫情分布。
4. 了解当地居民的健康水平。

##### 优点

1. 调查对象为全体目标人群,不存在抽样误差。
2. 可以同时调查目标人群中多种疾病或健康状况的分布情况。
3. 能发现目标人群中的全部病例,在实现“三早”(早期发现,早期诊断,早期治疗)预



防的同时,全面地描述疾病的分布与特征,为病因研究提供线索。

#### 缺点

1. 不适合患病率低且无简便易行诊断手段的疾病。
2. 工作量大而不易细致,存在漏查。
3. 调查工作人员设计面广,掌握调查技术和检查方法的熟练程度不一,对调查项目的理解往往很难统一和标准化,不能保证质量。
4. 耗费的人力、物力资源一般较大,费用往往较高。

普查可以对该地区的所有成年人进行调查,不存在抽样误差,可以最大可能的检测出该地区所有的成年人糖尿病病例。但是因为是对整个省进行调查,范围较大,工作量也会随着加大,同时耗费大量的人力、物力和财力,并且因为涉及面非常广,不能保证调查的质量。

#### 知识点

##### 将临床问题转化为科学问题:PICO

1. 人群、患者和(或)疾病状况(population, patient and/or problem) 该省成年人。
2. 治疗干预(intervention) 无干预。
3. 比较(comparison) 无对照组。
4. 临床的结局(outcome) 该省成年人糖尿病的患病率,以及 BMI、WC 和 WHR 的数值。

转换思路2:为了节省财力、物力,提高研究质量,我们也可以在该地区随机抽取一个有代表性的样本,对该样本进行检测,测量该样本人群中糖尿病的患病率以及这3种人体测量指标值,然后运用合适的统计学方法,分析其筛检价值,找出这3个测量指标筛检糖尿病的最佳切点,并分析是否存在性别差异。这种方法属于横断面研究中的抽样调查。抽样调查(sampling survey)指通过随机抽样方法,对特定时点、特定范围内人群的一个代表性样本进行调查,来推论其所在总体的情况。

#### 知识点

##### 抽样调查的原则及优、缺点

1. 抽样调查的原则 ①随机化原则:随机化原则是指整群中的每一个个体被选为样本的概率相等;②样本大小适当的原则:样本大小适当的原则是指样本应达到一定数量,不宜过大也不宜过小。
2. 优点 与普查相比,具有以下优点:①节省时间、人力和物力资源;②由于调查范围小,调查工作易于做得细致等。
3. 缺点 ①抽样调查的设计、实施与资料分析均比普查复杂;②资料的重复或遗漏不易被发现;③对于变异大的研究对象或因素和需要普查普治的疾病不适合用抽样调查;④患病率太低的疾病也同样不适合用抽样调查,因为需要很大的样本量,而如果抽样比大于75%,则不如进行普查。

抽样调查相对于普查来说,比较常用,如果严格按照抽样调查的设计要求进行调查检测,得到的样本统计量值结果是可以推论到整个整体的。而且在资料收集完成之后,可将样本按是否患病或是否暴露来分组比较,这样就有来自同一群体自然形成的同期对照组,使结果具有可比性。另外,一次横断面调查可同时观察多种因素,在疾病病因探索过程中,为不可或缺的基础工



作之一,比如,本次研究可对该省人群中的研究对象一般情况、生活行为方式、家庭饮食习惯、血糖值、血压值等很多资料进行调查。同时抽样调查又可以节省时间及人力物力,所以对该地区的人体测量指标筛检糖尿病的价值和最佳切割点进行分析时,应采用抽样调查的方法。

## 第二节 抽样调查研究设计与实施

### 一、研究目的

探索适合该省成年居民的人体指标适宜切割点,分析其筛检价值,为糖尿病的早期预防和诊断提供依据。

### 二、研究类型

横断面研究(抽样调查)。

### 三、确定研究对象

1. 年龄为 18~69 周岁。
2. 监测点地区居住 6 个月及以上的常住居民。

### 四、确定样本量

可通过公式法、查表法和估计法来确定样本量。本研究是对糖尿病的患病率进行调查,在估计样本量之前,必须明确一些参数:①预期现患率( $p$ ),可以通过查阅文献或者小规模的预调查得到,为了使样本量尽可能满足需要,查阅文献时可以选取其中较低的患病率;②对调查结果精确性的要求:即容许误差( $d$ ),容许误差越大,所需样本量就越小,一般采用  $d=0.1 \times p$ ;③要求的显著性水平( $\alpha$ ): $\alpha$  越小,即显著性水平要求越高,样本量要求越大。在得到以上参数之后,我们可以根据以下样本量估计公式进行样本量的计算:

$$n = \frac{pq}{\left(\frac{d}{Z_{\alpha}}\right)^2} = \frac{Z_{\alpha}^2 \times pq}{d^2} \quad \text{式 5-1}$$

其中  $p$  为预期患病率,  $q=1-p$ ,  $d$  为容许误差,  $Z_{\alpha}$  为显著性检验的统计量( $\alpha$  取 0.05 时,  $Z_{\alpha}$  为 1.96,  $\alpha$  取 0.01 时,  $Z_{\alpha}$  为 2.58),  $n$  为样本量。根据查阅文献的结果,  $p$  取 25.11%,  $d$  为 0.1 $p$ ,  $\alpha$  取 0.05,  $Z_{\alpha}$  为 1.96, 得出所需样本量大约为 1145 人, 本研究的设计效率  $deff$  取值为 2, 所以需 2290 人, 考虑分层因素及失访率(15%), 以及此患病率为 15 岁以上人群的患病率值(18 岁以上人群占 15 岁以上人群的 93.74%), 推算大约共需调查 15 133 名 18 岁以上的成人。

若抽样调查的分析指标为计量资料, 则应按计量资料的样本估计公式来计算, 公式如下:

$$n = \frac{4s^2}{d^2} \quad \text{式 5-2}$$

$n$  为样本量,  $d$  为容许误差,  $s$  为总体标准差的估计值。从上述公式可看出, 样本量大小与  $s$  的平方成正比, 与  $d$  的平方成反比, 故在实际应用中, 若同时有几个数据可供参考时,  $s$  宜取大一点的值, 这样不至于使估计的样本量( $n$ )偏小。

另外, 需要注意的是, 本研究中采用的率的样本量计算公式仅适用于  $n \times p > 5$  的情况, 如果  $n \times p \leq 5$  则宜用 Poisson 分布的方法来估算样本量, 可采用“Poisson 分布期望值的可信限简表”进行样本量确定。





## 五、抽样方法

为了抽取一个有代表性的样本对整体进行推测,需要采用大规模流行病学调查中经常使用的方法,即多阶段抽样(multistage sampling),此抽样方法是将抽样过程分阶段进行,每个阶段使用的抽样方法往往不同,即将各种抽样方法结合使用。其具体实施过程为:先从总体中抽取范围较大的单元,称为一级抽样单元(如省、自治区、直辖市),再从每个抽到的一级单元中抽取范围较小的二级单元(县、乡、镇、街道),依次类推,最后抽取其中范围更小的单元(如村、居委会)作为调查单位,对其符合条件的人群进行调查。多阶段抽样可以充分利用各种抽样方法的优势,克服各自的不足,并能节省人力、物力,但是在抽样之前要掌握各级调查单位的人口资料及特点。

根据以上多阶段抽样的规则,在掌握了该省的相关资料后,先将该省的140个县(市、区)作为初级抽样单元,并且根据其地理特征将其划分为3个地区类型。第1阶段抽样中,按照人口的比例情况,在3类地区中分别抽取5、8、7个县(市、区);第2阶段抽样中,在抽取到的这20个县(市、区)中,采用与人口容量成比例的概率(probability proportional to size, PPS)抽样方法抽取乡镇(街道);第3阶段抽样中,在抽取到的乡镇(街道)中再利用PPS抽样的方法抽取村(居委会);第4阶段抽样中,在抽取到的村(居委会)中,利用单纯随机抽样的方法抽取具体调查的对象。

### 知识点

#### 常用随机抽样方法

1. 单纯随机抽样 单纯随机抽样(simple random sampling)也称简单随机抽样,是最简单、最基本的抽样方法。从总体 $N$ 个对象中,利用抽签或其他随机方法(如随机数字)抽取 $n$ 个,构成一个样本。它的重要原则是总体中每个对象被抽到的概率相等(均为 $n/N$ )。主要用于总体不大的情形。

优、缺点:单纯随机抽样往往由于总体数量大,编号、抽样麻烦以及抽到个体分散而导致资料收集困难等原因实际应用的不多。但它是其他各种抽样方法的基础。

2. 系统抽样 系统抽样(systematic sampling)又称机械抽样,是按一定顺序,机械地每隔若干单位抽取一个单位的抽样方法。具体抽样方法为:设总体单位数为 $N$ ,需要调查的样本数为 $n$ ,则抽样比为 $n/N$ ,抽样间隔为 $K=N/n$ ,每 $K$ 个单位为一组,然后用单纯随机方法在第一组中确定一个起始号,从此起始点开始,每隔 $K$ 个单位抽取一个作为研究对象。主要用于按抽样顺序个体随机分布的情形。

优、缺点:优点:①可以在不知道总体单位数的情况下进行抽样;②在现场人群中较易进行;③样本是从分布在总体内部的各部分的单元中抽取的,分布比较均匀,代表性较好。缺点:假如总体各单位的分布有周期性,而抽取的间隔恰好与此周期或其倍数吻合,则可能使样本产生偏性。

3. 分层抽样 分层抽样(stratified sampling)是指先将总体按照某个特征分为若干次级总体(层),然后再从每一层内进行单纯随机抽样,组成一个样本。分层抽样又分为两类,一类叫按比例分配(proportional allocation)分层随机抽样,即各层内抽样比例相同;另一类叫最优分配(optimum allocation)分层随机抽样,即各层抽样比例不同,内部变异小的层抽样比例小,内部变异大的层抽样比例大,此时获得的样本均数或样本率的方差最小。分层抽样主要用于群间差异较小的情形。

优点:分层可以提高总体指标估计值的精确度,它可以将一个内部变异很大的总体分



成一些内部变异较小的层(次总体)。每一层内个体变异越小越好,层内变异则越大越好。分层抽样比单纯随机抽样所得到的结果精确度高,组织管理更方便,而且它能保证总体中每一层都有个体被抽到。这样除了能估计总体的参考值,还能分别估计各个层内的情况,因此分层抽样技术常被采用。

4. 整群抽样 整群抽样(cluster sampling)是将总体分为若干群组,抽取其中部分群组作为观察单位组成样本。若被抽到的群组中的全部个体作为调查对象,称为单纯整群抽样(simple cluster sampling);若再通过再次抽样后调查部分个体,称为二阶段抽样(two stages sampling)。整群抽样主要用于层间差异较大的情形。

优点:①易于组织、实施方便,可节省人力、物力;②群间差异越小,抽取的群越多,则精确度越高。缺点:抽样误差较大,故通常在单纯随机抽样样本量估算的基础上再增加 1/2。

## 六、资料的收集

根据研究目的来确定需要收集的资料,然后运用适当的方法进行收集,一般主要包括两种:①测定或检查,如血压的测定;②问卷调查,比如一般特征等。在资料收集的过程中要注意三点:①资料收集的方法确定下来之后,在整个科研过程都要前后一致,以避免研究资料的不同质性;②暴露(特征)的定义和疾病的标准均要明确和统一,比如本研究对糖尿病的检出定义;③所有参与检验或检测的人员以及调查员都必须经过统一培训,以统一调查和检测标准,避免测量偏倚的产生。

该研究主要收集的资料包括:调查对象的家庭一般情况、个人基本情况、生活行为方式、家庭饮食习惯、健康意识、疾病家族史等,这部分的内容采用问卷调查的方法进行收集;另外,还需要收集调查对象的身高、体重、腰围、臀围、血压和血糖等资料,这部分内容采用体格检查和实验室测定的方法进行收集。

## 七、资料的整理和分析

在资料收集后,首先要仔细的检查原始资料的完整性和准确性,填补缺、漏项,对重复的予以删除,对错误的予以纠正,然后采用各种数据录入软件对数据进行录入,如 EpiData、Excel、Access 等。对资料进行分析时,先对疾病或健康状态按已明确规定好的标准进行归类、核实,然后可按不同空间、时间及人群中的分布(三间分布)进行描述,进一步的将人群分为暴露组和非暴露组或不同水平的暴露组,比较分析各组间疾病或健康状况发生率的差异,也可将人群分为患病组和非患病组,评价各因素(暴露)与疾病的联系,具体统计分析应该结合研究目的选择合适的方法。

本研究在资料收集后,首先要检查一些关键数据是否完整,比如血糖值、腰围值等,数据是否正确,比如是否存在不合常理的数值,对所有资料进行核查后,可选择其中一种数据录入软件对资料进行录入,并可利用逻辑核查和双录入的方法对数据进行检查,发现问题再进行处理。统计分析主要通过 ROC 曲线和约登指数来分析该省筛检糖尿病的 BMI、WC 和 WHR 的价值和最佳切割点。

## 第三节 研究报告与问题分析

研究结果的报告包括研究背景与目的、方法、结果、讨论与结论四个方面。报告应当对研究结果进行详细的描述和分析,并应对研究设计和实施,资料的整理与分析中可能存在的偏倚以



及其他可能影响结论的因素进行恰当的分析。

### 一、研究案例 1

山东地区成人糖代谢异常患病率调查及人体测量参数对其预测效果研究(山东:山东大学, 2011)

1. 目的 探寻 BMI、腰围(WC)、腰臀比(WHR)、腰围身高比(WHtR)四个指标预测糖代谢异常的价值和最佳切点。

2. 方法 采用多阶段分层抽样的方法抽取了山东省济南、泰安、济宁、淄博四个地区的 3400 人进行问卷调查、体格检查和实验室检测。

3. 结果 在男性人群中,筛检价值大小分别为:WHtR、WC、WHR 和 BMI,其中 WHtR 预测糖代谢异常风险的最佳切点是 0.535,WC 为 91.5cm,WHR 为 0.905,BMI 为  $24.95\text{kg/m}^2$ ;在女性人群中,筛检价值大小分别为:WHtR、WHR、WC 和 BMI,其中 WHtR 预测糖代谢异常风险的最佳切点是 0.525,WC 为 84.5cm,WHR 为 0.855,BMI 为  $24.25\text{kg/m}^2$ 。

4. 结论 该人群中,4 种测量指标预测糖代谢异常的价值和最佳切点在不同性别之间不太一样,且男性 4 种测量指标的最佳切点均大于女性。

### 二、研究案例 2

三种人体测量指标对高血压和高血糖及高血脂的预测效果分析(中国全科医学,2012,15 (17):1917-1920)

1. 目的 探讨体质指数(BMI)、腰围(WC)及腰围身高比值(WHtR)对高血糖的预测效果,找出最佳的预测切点。

2. 方法 对重庆市在职成年人(18~59 岁)的健康体检数据进行分析。

3. 结果 在男性人群中,筛检价值大小分别为:WHtR、WC 和 BMI,其中 WHtR 预测高血糖风险的最佳切点是 0.52,WC 为 85.25cm,BMI 为  $25.38\text{kg/m}^2$ ;在女性人群中,筛检价值大小和男性人群中一样,其中 WHtR 预测高血糖风险的最佳切点是 0.48,WC 为 74.75cm,BMI 为  $23.81\text{kg/m}^2$ 。

4. 结论 该人群中,3 种人体测量指标预测高血糖的筛检价值大小顺序无差异,但最佳切点在男性和女性之间不一致,且男性 3 种测量指标的最佳切点均大于女性。

5. 评价 研究案例 1 研究对象的选择采用的是多阶段抽样,抽取了一个有代表性的样本,在对该样本人群进行调查和检测之后,得到的结论可以外推到该样本所来源的总体。而研究案例 2 中的研究对象不是对该地区人群的一个随机抽样,而是对进行体检的在职人群的数据进行分析,可能会使结果产生偏倚,不具有代表性。

#### 知识点

#### 偏倚产生的原因及控制

1. 偏倚产生的原因 ①主观选择研究对象;②任意变换抽样方法;③调查对象不合作或因种种原因不能或不愿意参加调查从而降低了应答率,此种现象称为无应答偏倚(non-response bias);④在横断面研究中,所调查到的对象均为幸存者,无法调查死亡的人,因此不能全面反映实际情况,有一定的局限性和片面性,此种现象又称为幸存者偏倚(survivor bias)。由以上原因导致的偏倚主要是选择偏倚(selection bias),其最终导致样本缺乏代表性而使研究结果不能外推;⑤询问调查对象有关问题时,由于种种原因回答不准确从而引起报告偏倚或调查对象对过去的暴露史或疾病史等回忆不清,特别是健康的调查对象由于没有疾病的经历,而容易将过去的暴露等情况遗忘,导致回忆偏倚(recall bias);⑥调查员有



意识地深入调查某些人的某些特征,而不重视或马虎对待其他一些人的这些特征而导致的偏倚,则成为调查偏倚(interviewer bias);⑦在暴露与疾病的测量中由于测量工具、检测方法不正确,化验技术操作不规范等可导致测量偏倚(measurement bias)。此外,在数据分析中,要注意有无混杂因素的存在及其影响程度。

2. 偏倚控制 ①严格遵照抽样方法的要求,确保抽样过程中随机化原则的完全实施;②提高研究对象的依从性和受检率;③正确选择测量工具和检测方法,包括调查表的编制等;④组织好研究工作,调查员一定要经过培训,统一标准和认识;⑤做好资料的复查、复核等工作;⑥选择正确的统计分析方法,注意辨析混杂因素及其影响。

## 第四节 临床问题的回答

本研究在运用适当的统计分析对数据进行分析之后,得出在男性和女性中,筛检指标的价值大小均分别为 WHR、WC、BMI,且男性和女性的 BMI 预测血糖异常的最佳切割点均为  $25\text{kg/m}^2$ ,但是 WHR 和 WC 在性别之间有差异,WC 男性为 91cm,女性为 83cm,WHR 男性为 0.92,女性为 0.87。

至此,可以得到结论,即不同人群中的人体测量指标对血糖异常预测的价值和最佳切点有一定的差异,故在利用这些人体测量指标对患者进行血糖异常预测的时候,应该结合实际情况进行判断。

## 第五节 横断面研究的其他应用

横断面除了确定高危人群的作用之外,还有以下几种用途:①掌握目标人群中疾病或健康状况的分布;②提供疾病病因研究的线索,通过描述疾病率在不同暴露因素状态下的分布差异、一致、趋同等现象,进行逻辑推理(如求同法、求异法、类推法等),进而提出该疾病可能的病因因素;③评价疾病监测、预防等防治措施的效果:在疾病监测、预防接种的实施过程中,通过在不同阶段重复开展横断面调查,既可以获得开展其他类型流行病学研究的基线资料,也可以通过对不同阶段患病率差异的比较,对防治策略、措施的效果进行评价。

## 第六节 文献阅读与评价

### 一、文献阅读与评价 1

Current status of contraceptive use among rural married women in anhui province of china. BJOG, 2009, 116(12): 1640-1645.

1. 目的 了解安徽省农村地区已婚妇女的避孕工具的使用现状及影响因素,采用了抽样调查对安徽省农村地区的一个随机样本进行问卷调查。

2. 方法 采用多阶段分层等概率抽样方法在安徽省农村地区随机抽取了 53 652 名 18~49 岁的妇女,问卷调查其一般人口学特征、性生活史、避孕工具的使用以及避孕知识等。

3. 结果 样本的总节育率为 95.2%,其中选择永久避孕方法的为 46.7%,选择临时避孕方法的为 48.5%。女性绝育手术是首选,使用率达到 43.6%,宫内节育器(IUD)的使用紧随其后,样本中有 41.1% 的女性选择此种避孕方法。单因素分析显示农村已婚妇女的避孕方法的选择与年龄、受教育程度、每个月性生活的频率、避孕知识、生殖道感染的症状以及最后一个孩子的性别有关。



4. 结论 安徽省农村已婚妇女的避孕措施的使用明显升高,其中主要是女性节育手术和IUD。在考虑了风俗习惯和信仰之后,通过正确使用和适当的管理作为一种辅助手段,计划生育应该被提倡。

5. 评价 为了解安徽省农村地区已婚妇女的避孕工具使用情况,需要对该地区的已婚妇女进行调查,但是如对所有已婚妇女进行调查,工作量较大,难以进行,所以采用了抽样调查的方法,抽取了该地区一个有代表性的样本,然后进行问卷调查,最终得到了该地区已婚妇女的避孕工具使用情况。

## 二、文献阅读与评价 2

Prevalence of diabetes among men and women in china. N Engl J Med, 2010, 362 (12): 1090-1101.

1. 目的 由于中国居民生活方式的迅速改变,可能会与糖尿病流行有关联,所以在 2007 年 6 月至 2008 年 5 月我们进行了一项全国性研究来估计中国成年人糖尿病的患病率。

2. 方法 全国有 14 个省、直辖市参与这项研究,采用多阶段分层抽样的方法从中选取了一个有代表性的样本,由 20 岁或 20 岁以上 46 239 名成年人组成。在一夜之间,参与者进行了快速口服葡萄糖耐量试验,测量禁食两小时后葡萄糖水平来识别未确诊的糖尿病或前驱糖尿病(即葡萄糖耐量降低或空腹葡萄糖降低)。预先确诊的糖尿病是建立在自我报告的基础上。

3. 结果 总糖尿病(包括以前诊断糖尿病和以前未确诊的糖尿病)和前驱糖尿病的年龄标准化患病率分别为 9.7%(男性 10.6% 和女性 8.8%)和 15.5%(男性 16.1% 和女性 14.9%),共有 9240 万成人糖尿病患者(男性 5020 万,女性 4220 万)和 1.482 亿成人前驱糖尿病患者(7610 万男性和 7210 万名女性)。糖尿病的患病率随年龄的增加而升高(年龄从 20~39 岁、40~59 岁和  $\geq 60$  岁时糖尿病的患病率分别为 3.2%、11.5% 和 20.4%),也随体重增加而升高(体重指数[体重(kg)/身高(m)<sup>2</sup>]为  $<18.5$ ~ $18.5$ 、 $24.9$ ~ $25.0$ 、 $29.9$  和  $\geq 30.0$  时糖尿病的患病率分别为 4.5%、7.6%、12.8% 和 18.5%)。城市居民的糖尿病的患病率(11.4%)高于农村居民(8.2%)。葡萄糖耐量降低患病率高于空腹葡萄糖降低(男性分别为 11.0% 和 3.2%,女性分别为 10.9% 和 2.2%)。

4. 结论 这些结果表明在中国,糖尿病已经成为一个主要公共卫生问题,旨在预防和治疗糖尿病的策略是必要的。

5. 评价 这也是一个大型的流行病学调查,临床问题是现在中国居民的生活方式改变了,那么人群的糖尿病患病率是否也随着改变,作者由此采用多阶段分层抽样的方法,抽取了 46 239 名 20 岁以上的成人进行了检测,最终得出了糖尿病的患病率。

## 三、文献阅读与评价 3

Sampling survey of disability in 0-6 year-old children in China. Biomed Environ Sci, 2006, 19(5): 380-384.

1. 目的 为了解中国 0~6 岁儿童的伤残率及其影响因素,同时了解伤残儿童的现况,为伤残儿童制定相关政策提供科学依据。

2. 方法 以社区为基础,采用多阶段分层非等概率的抽样方法调查了 60 124 名 0~6 岁的儿童。

3. 结果 共有 819 名儿童被确诊为残疾,总体残疾率为 1.362%,其中听力缺陷为 0.155%,视力缺陷为 0.160%,智力缺陷为 0.931%,肢体残疾为 0.424%,精神障碍为 0.101%,在农村地区,拥有 2 个或以上儿童的家庭,教育水平低的家庭,以及离婚的家庭,残疾率均较高。

4. 讨论 增强经济发展,改善卫生保健、人口素质和家庭关系,早期预防残疾,以及对残疾儿童进行学前教育可以降低儿童的残疾率。

5. 评价 这是一个典型的横断面研究,临床问题是中国 0~6 岁儿童的伤残率和影响因素,





作者采用了抽样调查的方法抽取了一个由 60 124 名儿童组成的样本进行了调查,最终得到了其伤残率,并分析了影响因素。

### 思考题

1. 什么是抽样调查? 抽样调查的基本要求是什么? 抽样方法有哪几种? 具体实施的方法是什么?
2. 针对上述 3 篇阅读文献(摘要,可以根据需要阅读全文),回答:该研究期望解决的临床问题是什么? 科学问题是如何描述的? 该研究采用了何种临床研究设计? 抽样方法是什么? 样本量怎么确定的? 你认为存在哪些偏倚,如何控制偏倚?
3. 结合自己专业,选择一种疾病,设计一项研究去测定该疾病在某地区的患病率,内容包括:科学问题与研究目的,研究设计与具体方法,结果测量方法及可能的结果。

(张秀军)

## 第六章 病例对照研究

### 临床案例

一名 35 岁女性,因为中上腹疼痛,抑酸药物(奥美拉唑)治疗无效,伴有消瘦,胃镜检查发现胃窦 2cm 溃疡病灶,幽门螺杆菌检查阳性,在接受标准 4 联(奥美拉唑 + 阿莫西林 + 克拉霉素 + 铋剂)10 天根除幽门螺杆菌治疗的同时,等待病理结果。一周后溃疡边缘活检病理证实为低分化胃癌。上腹部 CT 检查没有发现淋巴结转移,及时进行远端胃大部切除根治术 + 预防性化疗 6 疗程,一年后复查患者状况良好,无肿瘤复发。

### 第一节 如何获得临床研究问题

这是一例常见病临床处理。针对该患者和同类患者,仍然有很多争议问题,也是医学研究的问题来源。如该患者为什么会发生胃癌(肿瘤病因问题一直是我们探索的问题),是否与幽门螺杆菌感染有关?类似的患者,是否都需要检查幽门螺杆菌并接受根除幽门螺杆菌治疗?针对这一常见病和多发病,我们可以进行很多相关的研究。如何从临床问题中获得有意义、具有探索性或者证据性研究课题,即形成科学问题。

#### 一、明确临床问题

1. 阅读文献查阅相关背景资料 自 1983 年发现幽门螺杆菌以来,针对幽门螺杆菌与临床疾病的关系成为临床研究的热点。临床和基础研究证据已经明确,幽门螺杆菌是慢性胃炎的主要病因,在消化性溃疡的发病中具有重要的作用。根除幽门螺杆菌治疗溃疡并预防复发已成为消化性溃疡首选和规范治疗。中国人群中 50% 感染,同时也是胃癌高发人群,是否幽门螺杆菌感染与胃癌有关,有多大关系是大家关心的问题。

2. 其他研究证据提示的临床问题思路 炎症与肿瘤之间的联系同样密切。慢性胃炎-萎缩-肠上皮化生-异型增生-胃癌是胃癌发生的假说之一。

病原微生物引起的炎症与肿瘤一直是我们探索的主题。如肝炎与肝癌有大量证据表明两者之间的密切联系,宫颈炎与宫颈癌、炎症性肠病与结肠癌等。幽门螺杆菌感染是慢性胃炎的主要病因。幽门螺杆菌感染与胃癌发生之间的联系可能同样存在。

3. 基础研究证据 动物实验中,已建立了幽门螺杆菌感染致胃癌动物模型,幽门螺杆菌感染联合亚硝酸盐致胃癌前病变(胃黏膜上皮异形增生)和胃癌更明显。

4. 已有的临床研究证据 流行病学调查显示,高感染幽门螺杆菌地区(如中国、日本等亚洲国家),胃癌发生率也高,而西方国家胃癌发生率低,伴随的现象是人群幽门螺杆菌感染率也低。

#### 二、临床问题转换成科学问题

针对上述临床问题,我们可以通过不同的临床研究来回答,这就需要将临床问题转换为科学问题。

转换思路 1:幽门螺杆菌与胃癌的关系,我们可以转换为这样的问题:幽门螺杆菌感染的人



群,在经过若干年之后,发生胃癌的概率是否高于无幽门螺杆菌感染的人群,概率增高多少倍?这就需要前瞻性队列研究设计,通过幽门螺杆菌检测,将人群分为感染组和无感染组,观察若干年之后,分别检查两组人群胃癌的发生率,比较两组人群胃癌发生率高低。

这样的研究尽管通过幽门螺杆菌检测分组比较容易,但需要长期(5~10年)随访,在这过程中控制未感染人群不会获得感染,已感染人群不接受根除幽门螺杆菌治疗,并需要不断通过胃镜检测评估是否发生胃癌。这样的研究设计(自然队列研究设计)尽管结论会非常有利于回答临床问题,但研究本身费时、费力,需要很大的投入。

### 知识点

将临床问题转化为科学问题:PICO

1. 患者和(或)疾病状况(population, patient and/or problem),如胃癌。
2. 治疗干预(intervention),或者有时是暴露(exposure),如幽门螺杆菌感染。
3. 比较(comparison),如果有的话,幽门螺杆菌感染组与非感染组。
4. 临床的结局(outcome),有时候(如病因、生存分析等)时间因素也应考虑在内,胃癌、癌前病变,发生时间5~10年。

转换思路2:我们也可以转换为这样的问题:幽门螺杆菌感染人群,假如通过根除治疗之后,发生胃癌的概率,与未进行根除治疗相比,是否下降,下降多少倍?这就需要随机对照临床试验设计,对幽门螺杆菌感染阳性人群,随机化分为两组,一组接受根除幽门螺杆菌治疗,一组不接受根除幽门螺杆菌治疗,当然,最好有一直没有幽门螺杆菌感染的人群作为基础对照就更好,经过长期(5~10年)随访,通过胃镜评估两组胃癌发生概率,并进行比较。

这样的研究设计(一种比较效果研究设计,先是随机对照临床试验 RCT,然后队列研究长期随访)不仅需要随机对照临床试验设计,还需要随机化处理后通过幽门螺杆菌感染检测重新分组,建立两个队列,然后进行5~10年随访,不仅费时、费力,大量人群接受根除幽门螺杆菌治疗还可能产生不良事件也影响患者接受程度和依从性,因此需要更大投入。如何选一个更可行的研究进行幽门螺杆菌与胃癌关系的初步评价?

转换思路3:我们同样可以转换为这样的问题:与非胃癌患者相比,胃癌患者幽门螺杆菌感染的比例是否高于非胃癌患者?这样我们就可以首先设计一个花费少、耗时短的病例对照研究。对胃癌患者,追溯5年或者10年前是否存在幽门螺杆菌感染,感染的比例是多少,同时在胃癌来源的人群中选择合适的非胃癌人群作为对照,同样方法追溯5年或者10年前是否存在幽门螺杆菌感染,比较胃癌组与对照组之前幽门螺杆菌感染比例的差别。

### 知识点

病例对照研究设计思想

病例-对照研究(case-control study)系回顾性研究的一种,对一组患有某种疾病的患者,回顾追溯可能引起该病的因素,并通过与未患有该病的人群进行比较,评估暴露因素是否与疾病有关。在研究开始时,鉴别患病与未患病人员,并收集信息及时回顾来鉴别潜在暴露因素至关重要。

病例对照研究的基本原理是以现在确诊的患有某特定疾病的患者作为病例,以不患有该病但具有可比性的个体作为对照,通过询问,实验室检查或复查病史,搜集既往各种可能的危险因素的暴露史,测量并比较病例组与对照组中各因素的暴露比例,经统计学检验,若两组差别有意义,则可认为因素与疾病之间存在着统计学上的关联。在评估了各种偏倚对研究结果的影响之后,再借助病因推断技术,推断出某个或某些暴露因素是疾病的危险因



素,而达到探索和检验疾病病因假说的目的。

一般的策略是比较来自同一总体、具有代表性的病例组与非病例组间的潜在危险因素的频率或水平的高低。虽然病例对照方法有时用于常见病的研究,但对罕见病的病因学研究更有帮助。

### 知识点

#### 病例对照研究与队列研究比较

与队列研究相比(图 6-1),病例对照研究的特点在于:①研究在疾病发生后开始;②按发病与否分为病例组和对照组;③暴露测量是通过对研究对象过去暴露状态和程度的回顾;④结果已经发生,由果推因;⑤分析暴露与疾病的联系。

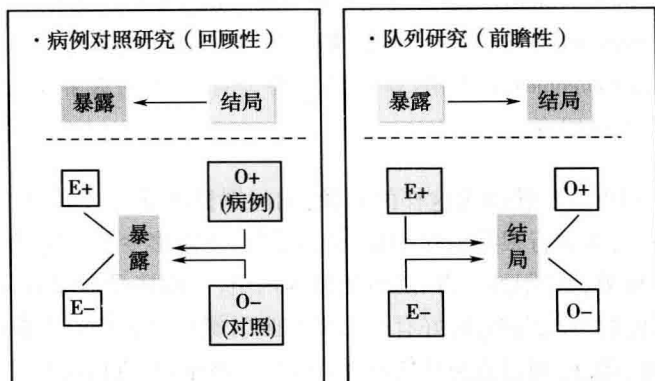


图 6-1 病例对照研究与队列研究设计比较

## 第二节 病例对照研究设计

### 一、研究目的

通过病例对照研究评估幽门螺杆菌感染与胃癌发生的关系。

### 二、研究假设

与非胃癌患者比较,胃癌患者幽门螺杆菌感染率更高。

### 三、病例的选择

病例选择:①诊断明确的胃癌患者:需要规定诊断手段与依据,如病理诊断,无论通过胃镜活检或者手术切除标本评价。淋巴结穿刺活检腺癌阳性标本是否纳入存在争议,因为有可能是其他来源的转移癌,除非免疫组织化学以及临床其他证据如增强 CT 等支持胃癌诊断,并需要排除其他恶性肿瘤;②胃癌患者的来源包括:综合医院或者专科医院患者来源,社区人群患者来源,他们代表了两种不同的人群,前者带有选择性,人群为到医院就诊患者,来自不同地域,结果可能带来选择偏倚。后者带有普遍性,但获得全部患者或者从全部患者中抽样有一定难度,获得他们的完整信息困难更大。人为选择获得完整信息的学生同样带来选择偏倚。因此,获得完整资料是选择病例的一项依据,权衡选择偏倚也是选择的一项依据;③确定来源之后,胃癌患者的具体入选标准和排除标准成为选择的第二个问题:包括是否一段时间内的全部患者,是否限



制患者地区,是否限制年龄与性别,胃癌不同部位、不同病理类型、不同病理分期的患者是否都包括,手术、化疗等治疗措施的使用是否影响入选;④纳入胃癌患者的时间需要规定,何时开始,何时结束,持续多久,入选时间是否以确诊时间为依据。

**案例病例选择 1:**假设我们拟在社区人群中开展一项幽门螺杆菌感染与胃癌的病例对照研究,我们需要有一个完整的胃癌登记系统,并记录完整患者资料。同时还能够从这一系统获得患者之前暴露的幽门螺杆菌感染资料。如已知有一个 60 年代开始的随访队列,共纳入 10 万多例人群,保留了人群纳入当时的一般资料,血清标本,跟踪随访资料包括是否发生胃癌。入选标准为从纳入随访到研究开始前首次诊断为胃癌患者。考虑到胃癌患者可能较多,可从其中按样本量随机抽取需要的胃癌病例数。排除标准包括不能获得血清样本患者,重新评估患者病理组织学标本不能确定胃癌诊断的患者。

**案例病例选择 2:**假设我们拟在医院就诊人群中开展一项幽门螺杆菌感染与胃癌的病例对照研究,我们可以采用医院住院患者信息系统(既往住院胃癌患者为纳入研究对象),或者内镜检查患者信息系统(既往经胃镜检查 and 病理活检明确诊断为胃癌患者)。我们同时需要得到患者幽门螺杆菌暴露资料,尤其之前暴露资料。根据需要的样本量,在规定的一段时间内在三级医院收集已诊断的胃癌病例。入选标准为新发胃癌行手术及病理诊断的胃癌病例,年龄 30 岁以上(幽门螺杆菌感染致胃癌发生需要一定时间,尽可能排除完全与遗传、化学接触等相关因素)。排除标准包括:不能获得组织学标本患者(我们采用组织学方法评估暴露因素,当然血清学 ELISA 方法检测抗体也可以,但需要病例与对照组一致),残胃癌。

### 知识点

#### 病例组来源与选择

病例来源可以是社区居民中的某病患者(基于人群),也可以是在医院就诊的患者(基于医院)。病例的选择首先要力求符合公认的诊断标准,保证病例的诊断准确无误,有时甚至要求疾病的病理分型也相同。对诊断有疑问的患者不应纳入病例组中。在选取病例时,尽可能选取病因学上同源的一组个体。病例一般可分为 3 种类型,即发病病例、患病病例和死亡病例。在病例对照研究中,研究危险因素的设计应该是发病病例,研究预后因素的设计可能选择患病病例或者死亡病例。发病病例有患病病例和死亡病例所不具备的优点:①发病时间更接近于病因暴露时间,病例能更好地回忆自己的有关经历和暴露史,并且容易得到就近的病历资料、职业暴露或其他记录;②发病病例能自己回答问题,比死亡病例靠亲友、家属回答要准确得多;③发病病例刚被确诊就接受调查,尚未受到各种决定生存因素的影响,而患病病例是以往确诊的大批发病病例中的幸存者,如果某种因素对生存有影响的话,则可能导致错误的结论;④如果发病病例收集完全的话,我们可以得到某种疾病的发病率。在选择病例时,所选病例对目标人群应有较好的代表性,所选病例应包括轻重各型病例。

## 四、对照的选择

**对照选择:**①需要确定非胃癌患者;②根据社区胃癌患者或者医院胃癌患者来源选择对照,对照的选择需要与病例来源一致;③无论医院内胃癌来源或者社区来源的胃癌采用性别、年龄等因素配对作为对照更合理。对医院胃癌患者,由于存在人群选择,同时采用医院对照和人群对照有时更合理;④对照选择时,同样需要有入选标准和排除标准。特别需要强调的是,消化性溃疡患者幽门螺杆菌感染率非常高,因为幽门螺杆菌感染在消化性溃疡发病中起重要作用。对这一组患者,作为胃癌研究的对照人群应该排除。

**案例对照选择 1:**社区人群研究,对照来源全体人群,成组对照采用随机抽样更加合理。在





社区人群研究中,更适合选择 1:1 配对或者 1:2 配对的对照。对照的入选标准包括:与病例组按性别、年龄、种族配对。排除标准包括:一般资料缺乏,不能获得足够的保留血清标本进行幽门螺杆菌抗体检测,胃切除后患者。

案例对照选择 2:在医院人群研究中,1:1 配对或者 1:2 配对的对照可以选择同时接受胃镜检查但没有胃癌,并性别、年龄配对。也可以增加与胃癌来源相同的一般人群的配对对照,但非胃癌的确认以及暴露因素的检测常常存在问题。对照入选标准包括:性别、年龄配对(如果不能找到合适的对照,年龄可以规定相差不超过 1 岁),胃镜检查在同一天(如果不能找到合适的对照,胃镜检查时间可以规定在前一天或者后一天)。排除标准包括:不能获得组织学标本,已行胃大部切除术后,胃镜诊断消化性溃疡患者。

### 知识点

#### 对照组来源与选择

与病例组相比,对照组的限定、选择和招募更困难。理想的对照组可能并不存在,病例对照研究最普遍的问题就是对照组存在潜在偏倚。应该依照入选和排除标准规范,像选择病例组一样,从一个明确定义的群体中选择对照组,并保证这些对照组无疾病。对照组原则上应与病例组有同一来源,即来自同一地区、同一社区中未患所研究疾病的居民(人群对照),或同一医院中未患所研究疾病的其他患者或健康体检者(医院对照)。当病例来源于某个人群时,选取一般人群对照可以保证与病例组高度可比,使研究结果有较高的普遍性。但这种人群对照所花费人力、物力较大,所选中的个体常常不予合作或不易找到,应答情况比其他类型的对照差。选取医院患者作对照时,通常较为可靠,他们有充裕的空闲时间并且能够合作。选择医院对照时,还可以使病例和对照组间有关决定住院的因素相似。医院对照的一个严重缺陷,就是对照可能因为某个与病因学特征有关的条件而有选择性地入院(如本案例中消化性溃疡病例入院,作为对照,同样与幽门螺杆菌感染有关,应该排除),并且这种选择性倾向和病例是不同的。这时,就会使结果产生偏倚。医院对照还可能因病种不同,对照和病例回忆时可能的思维内容也不一定相同。例如肺癌患者可能更侧重于回忆其吸烟史等,而胃癌患者可能更注重回忆其饮食方面的变化。对于某项具体的病例对照研究,是选择人群对照,还是选择医院对照,不能一概而论,要根据研究的具体情况,如病例的来源、性质、选择个体的方法等而定。

### 知识点

#### 选择多对照

缺少完美对照组的解决方法之一就是使用多类型对照。比如在医院选取研究对象时,可以从同一医院选取一个对照组,再从一般人群(如邻居)中选择另一个对照组,这两个对照组对于同一病例是“相邻”关系。这一做法在方法学上是先进的,因为它可以通过其他对照组来最小化组内偏倚,反之亦然。它们之间关联度的估计可以在两组中分别进行,如果两组对照中发现相似的关联度则结论将进一步得到论证。

### 知识点

#### 配 对

如果已知年龄、性别或者特殊的危险因素(如吸烟)和研究疾病相关,可以在每个人中准确的测量,并且研究目的是探索潜在病因时,对照组可以用这些因素来配对。在某一因素上将病例组和对照组配对的原因是为了消除这种因素混杂所要探求的联系。然而,重要



的是认识到一旦将某一因素进行配对,那么就不能对其进行分析,因为对于这一因素病例组和对照组在设计时是相同的。此外,发现匹配对照组的难度也随着需要匹配的因子数量增加而加大。配对也可能使得分析更加复杂化,因为分析方法必须将配对抽样过程考虑进去。一般来说,如果无法确定某一特定因素与疾病病因学相关,比较好的办法是不用它来配对,但在收集资料时详细记录这一因素,并在分析时考虑平衡。如果有多个对照组,其中一个可以配对,其他的可不配对。

配对的条件或变量应该是与疾病无直接关系的因素。如果配对变量中包括了疾病的危险因素或病因,就不能正确分析该因素与疾病的关系。常用的配对变量有年龄、性别、入院日期、血型、收入水平、文化教育水平等。具体某项研究的配对条件应视具体情况而定,但配对条件或变量不宜太多,否则将给对照的选择带来困难。病例与对照的配对比例一般可为 1:1,1:2,1:3,最多不宜超过 1:4。

## 五、样本量的计算

样本量计算:最好请专业人员帮助,通过公式法、查表法和估计法。但估计样本含量之前,必须明确一些参数:①对照组的暴露率  $P_0$ ,本案例中为一般人群幽门螺杆菌感染率。 $P_0$  可以通过查阅文献或小规模的前调查得到,如西方国家在 20%~30%;我国国家约 50%;②相对危险度(relative risk, RR),指暴露于某因素的人群中某病发病率或死亡率与无暴露人群中的某病发病率或死亡率之比,病例对照研究不能直接计算 RR 值,只能用优势比(OR 值)来近似估计,OR 值也可以通过查阅文献或前调查获得。另一种替代 OR 值的指标为病例组的暴露率;③  $\alpha$  值,即所希望达到的检验显著性水平,也就是第一类错误(假阳性错误)的概率,一般取 0.05 或 0.01;④把握度  $1-\beta$ ,  $\beta$  为第二类错误(假阴性错误)的概率,  $1-\beta$  即发现这种关系有多大把握度,一般取 0.90,也就是  $\beta$  值取 0.10。

根据计算,本研究案例可以估计样本量为 200 对。

## 六、结果(暴露)的测量

本研究案例中的暴露为幽门螺杆菌感染。幽门螺杆菌感染的检测有不同方法,同时可以分别反映现症感染或者既往感染。对本研究来说,这一点并不重要,一方面本研究的病因或者危险因素,幽门螺杆菌感染致胃癌需要很长时间,我们更希望了解以往感染资料;另一方面,幽门螺杆菌感染后,除治疗外,很少自愈,现症感染某种程度上能反映既往感染。病例组和对照组暴露测量可以采用相同的方法,案例 1 采用保留血清 ELISA 方法检测幽门螺杆菌抗体,案例 2 采用保留胃黏膜标本进行病理 Giemsa 染色。需要明确的是,我们能否获得相应的血清或组织,尤其对对照组人群,这是在病例和对照选择时就需要考虑的。如果不能获得配对病例与对照中的任何一个,都将成为无效病例。

### 知识点

#### 暴露因素测量

暴露因素的测量对于评价病例对照研究的优劣也是至关重要,最安全的方法是采用疾病发生前采集的完整、准确的信息记录。如药物相关研究的用药记录,外科手术并发症研究的手术记录,生物标记物研究储存的血清学标本。但有些重要的暴露因素测量却只能通过现场询问患病者和未患病者获得,这些因素包括运动量、饮食、生活方式、非处方药使用情况等,通过患者回忆之前状态,则可能产生回忆偏倚。



在病例对照研究中,暴露因素可以单个,也可以多个,即多因素暴露,如同一种暴露因素的不同剂量,一种疾病的多种早期症状等。

和其他观察性研究设计一样,无论是病例组还是对照组,待测因素和观测结局都应当是特定一致的,且要使用相同的方法。在研究过程中,病例组和对照组的测量技术的真实性和重复性都需要建立和评估。

病例对照研究的一个重要标准是病例组和对照组的信息收集采取相同的手段。这一点操作起来有点困难,特别是当病例状态对于调查者来说很熟悉或是很明显时。对于那些了解研究假设的调查者或资料收集者来说,他们会有意识地从病例组中寻找暴露信息。所以调查者或资料收集者必须接受一定的训练,不管疾病状态去询问问题,给予对阳性或阴性回答相同的反应。而且,不管面谈进行的如何好,被调查者往往用相同方法不能回想起当时的信息。如果我们的小孩生病了,我们可能会想很多可能已经发生的潜在暴露因素,但如果我们的小孩没有生病,我们可能不会思考这些信息,也不会很好地去回忆。这些是“回忆偏倚”的例子。同样地,医疗记录在病例组和对照组中在信息水平会不同。如果临床医生怀疑某个疾病可能存在,他们可能会更仔细寻找相关因素。从根本上来说,为了评估可能的病因,必须使用那些用同样的方法收集来的病例组和对照组的信息。

### 知识点

#### 暴露时间

从暴露到发病的时间也是评估暴露与结局之间联系的重要因素。在收集暴露信息时,应该同时考虑暴露出现与持续时间,并保持病例组和对照组均衡可比。

## 七、预期结果

预期结果:胃癌组幽门螺杆菌感染率高于非胃癌组幽门螺杆菌感染率,差异有统计学意义。对胃癌发生部位(非贲门区与贲门区)、不同病理类型(腺癌与非腺癌)等因素进一步进行分层分析。显示非贲门区胃癌、腺癌患者幽门螺杆菌感染率可能高于对照组。

在病例对照研究中研究结果以“ $2 \times 2$ ”表格表示,其中行代表暴露情况,列代表病例和对照情况(表 6-1)。a 代表病例组暴露例数,b 代表对照组的暴露例数,a+b 代表所有暴露例数。列总和 a+c 和 b+d 分别代表所有病例组和对照组例数。比较病例组暴露比例  $a/(a+c)$ ,和对照组的暴露比例  $b/(b+d)$ 。这两个构成比比较可以用自由度为 1 的卡方检验。病例组和对照组的均数或连续变量可以通过  $t$  检验或非正态变量的非参数检验来比较。

表 6-1 结果表达的  $2 \times 2$  表

| 患病情况 |     |      |     |
|------|-----|------|-----|
| 暴露   | 患病数 | 未患病数 | 合计  |
| 是    | a   | b    | a+b |
| 否    | c   | d    | c+d |
| 合计   | a+c | b+d  | N   |



## 知识点

## 比 数 比

比值(odds)与可能性有关,比值定义为发生某件事的次数和未发生某件事次数之比。如 $\text{odds}=p/(1-p)$ 。举个例子,如一匹马赢一个比赛的可能性为50%,那么它赢的比值为1:1,如可能性为25%,那么赢的比值为1:3,而不赢的比值为3:1。

如果某个人(患者)有某个特征或暴露因素的可能性为 $a/(a+c)$ ,从表6-1中我们得出暴露比值为 $[a/(a+c)]/[c/(a+c)]$ ,这个值等于 $a/c$ 。同样的,对于一个非患病的人暴露比值是 $b/d$ 。比较患病者和非患病者的暴露比值即产生了比数比(odds ratio, OR),在表6-1中指的是 $(a/c)/(b/d)$ 或 $ad/bc$ 。比数比被广泛用于各种病因学研究中,因为它在病例对照研究中作为一种相关度的指标仍是有效的。同时评估多个暴露因素的比数比或者平衡其他因素影响时,可以用logistic回归方法分析。由于在病例对照研究中我们已经确定了病例组 and 对照组的数量,因此我们不能评估研究中某个疾病的绝对危险度。在病例对照研究中,我们经常会设定病例组和对照组样本量一致,即1:1配对,因此在我们的样本中病例组的比例(1/2)不可能代表研究总体中患该疾病的风险。

## 第三节 研究报告与问题分析

病例对照研究的报告格式和要求可以参考描述性研究报告规范。研究结果的报告包括研究背景与目的、方法、结果、讨论与结论四个方面。讨论对研究设计、可能存在的偏倚以及对可能影响结论的因素进行恰当的分析。

## 一、研究案例1

幽门螺杆菌感染与胃癌风险病例对照研究[NEJM, 1991, 325(16): 1127-1131]

1. 目的 已有研究证实幽门螺杆菌感染与胃腺癌的癌前疾病——慢性萎缩性胃炎相关,本研究目的是采用病例对照研究来探索幽门螺杆菌感染是否增加胃癌发生风险。

2. 方法 人群选择来源于自19世纪60年代中期开始随访的一个健康维护组织的随访队列,共128 992人。186名患胃腺癌的患者被选为病例组,另外186名经过年龄、性别、种族等匹配的未患胃腺癌者作为对照组。采用ELISA方法检测初期储存的血清学标本中抗幽门螺杆菌——IgG抗体。在人群招募时已进行了问卷调查,记录吸烟史、血型、胃溃疡史、胃部手术史等。

3. 结果 自开始收集血清学标本至胃癌发生平均时间为14.2年。排除胃食管交界处肿瘤,病例组109例确诊为胃腺癌,其中84%既往有幽门螺杆菌感染,对照组仅61%既往有幽门螺杆菌感染( $\text{OR}=3.6, 95\% \text{CI } 1.8\sim 7.3$ )。而幽门螺杆菌感染与食管胃交界处肿瘤以及胃贲门部肿瘤都无关。分层分析显示,女性和黑人人群的胃癌发生与幽门螺杆菌感染相关性更加密切,OR值分别为18和9。对其他暴露因素的进一步分析显示,胃部手术史与胃癌的发生独立相关( $\text{OR}=17, P=0.03$ ),而胃溃疡史与胃癌发生负相关( $\text{OR}=0.2, P=0.02$ )。血型和吸烟史均未发现会增加胃癌的发生风险。

4. 结论 幽门螺杆菌感染与胃癌发生风险增加有关,可能是胃癌发生的共同致病因子。

5. 评价 研究案例1采用了巢式病例对照研究,病例和对照都选择了一个随访队列,幽门螺杆菌血清和患者信息资料都来源于人群招募时,即在胃癌发生前(平均14.2年)。克服了暴露时间与结果发生可能存在的问题。



## 知识点

## 巢式病例对照研究

巢式病例对照研究(nested case-control study)集中了病例对照研究和队列研究的优点。

首先,按照队列研究的方式进行:选择一个人群,收集基线资料,同时采集研究所需的生物学标志的组织或体液标本储存备用,然后开始随访,记录各种临床结果,或者研究者关注的临床结果。

其次,将这个队列中产生的病例作为病例组,按病例进入队列的时间、疾病出现的时间与性别、年龄等作为匹配条件,从该队列中选择1个或数个非病例作对照,同时抽取病例与对照的基线资料并检测收集的标本。

最后,资料的处理与分析按匹配病例对照研究进行。

## 二、研究案例2

幽门螺杆菌感染与胃癌的一项病例对照研究(胃肠病学和肝病学杂志,1998,3:227-229)

1. 目的 幽门螺杆菌(Hp)是慢性胃炎的主要致病因子,在消化性溃疡的发生中起重要作用。根据慢性胃炎→腺体萎缩→肠上皮化生→异型增生→胃癌发生学说,Hp在胃癌发生中的作用受到了人们重视。本研究目的是探讨Hp感染与胃癌发生的关系。

2. 方法 采用配对病例对照研究方法。自1994年1月至1995年12月中山医院共进行胃镜检查11 000例患者,经胃镜确诊为胃癌136例,组成病例组。性别相同,年龄相差在5岁范围内,相同或相近检查日期的胃炎患者作为对照组。Hp检测采用快速尿素酶法和病理染色相结合。Hp感染的诊断标准是:两种检测方法中任一阳性即定为阳性,两者均阴性定为阴性。统计方法采用Mantel-Haenszel卡方检验。

3. 结果 总的胃癌检出率为1.24%(136/11 000)。按照肿瘤生长部位,胃窦、胃体、贲门部所占比例分别为46.3%、34.6%、15.4%;按照肿瘤的病理类型,腺癌的比例最高(88例,占64.7%)。Hp感染率在病例组(74.3%)明显高于对照组(57.4%, $OR=2.35$ ,95%CI:1.33~4.15)。考虑到肿瘤发生部位及类型的影响,非贲门部肿瘤病人Hp感染率(89/115,77.3%)明显高于对照组( $OR=3$ ,95%CI:1.56~5.77),而贲门部肿瘤的Hp感染率(12/21,57.1%)与对照组无明显差别( $OR=0.5$ ,95%CI:0.09~2.73)。腺癌患者的Hp感染率(66/88,75.0%)明显高于对照组( $OR=2.42$ ,95%CI:1.23~4.47),而非腺癌患者的Hp感染率(33/48,68.6%)与对照组无差别( $OR=2.67$ ,95%CI:0.71~10.05)。

4. 结论 胃癌中非贲门部腺癌的发生与Hp感染有关。

5. 评价 研究案例2来源于医院,可能存在的问题:①研究的胃癌案例来源将非常容易,尤其在中国。可能存在选择性偏倚,案例不能代表一般人群来源的胃癌,如国内大多数是早期胃癌、或晚期胃癌,或有手术指征的胃癌,取决于该医院收治胃癌的患者状况。解决的办法是,尽可能纳入全部入院或者接诊的新发胃癌,联合多家医院选择不同来源胃癌患者作为病例组;②对照来源于医院也可能存在偏倚,尤其对照选择了接受胃镜检查的患者,这些患者人群幽门螺杆菌感染率可能高于一般人群,从而低估了研究结果,可以采取医院和社区人群分别对照的方法;③暴露测量,无论通过呼气试验或者血清学检测,都只能反映患者住院当时幽门螺杆菌感染状况,尽管幽门螺杆菌感染在未行治疗的人群中是持续存在,但我们仍然不清楚无论病例组或者对照组何时感染幽门螺杆菌,而感染时间到胃癌发生在两组的不同可能影响结果。





## 知识点

## 偏倚及其控制

混杂偏倚(bias)是指在病例对照研究中,由于一个或多个潜在的混杂因素的影响,掩盖或夸大了研究因素与疾病(或事件)之间的联系,从而使两者之间的真正联系被错误地估计,控制方法有匹配法,分层分析方法,多因素分析方法。

患病率或发病率偏倚是一种特殊的偏倚。对于极早暴露的病例,研究者的近期观察往往观察不到短期的或致命性不良事件。从这一方面来说,估计病例的发病年龄、病情持续时间或严重程度对研究都是有用的。

成员偏倚的发生是由于该组成员的健康程度与总体有系统上的差别。这是就业或称流动人口特有的问题,也常被称为“健康就业工人”或“健康外来人员”效应。可以通过从相同的就业工人或流动人口人群中抽取对照组来进行控制,但是研究结果要推广到整个人群中可能就会受限了。

在数据收集阶段可能发生其他的重要偏倚,包括诊断偏倚、暴露偏倚、回忆偏倚和家庭信息偏倚。

技巧:病例对照研究的三个重要标准可以用来最大限度地减少潜在偏倚。第一个标准是选取的病例要能代表患有该疾病的所有患者。当使用医院病例的时候,这一点可能较困难。这是由于三级医院的患者通常与在小医院治疗或不去医院治疗的患者差别较大。第二个标准是对照组要能代表没有发生该疾病的一般健康人群。一个相对简单但有潜在问题的途径是从人群中随机选择,然后剔除个别纳入的患病个体。另一个方法是用多组对照来避免一系列问题。第三个标准是病例组和对照组的信息收集采取相同的手段。

学习  
笔记

## 第四节 临床问题的回答

继病例对照研究显示幽门螺杆菌感染与胃癌相关之后,更多研究设计证明了幽门螺杆菌感染与胃癌之间的关系。

队列研究文献显示:幽门螺杆菌感染组 5~10 年后胃癌发生率明显高于无幽门螺杆菌感染组。

预防性随机对照临床研究文献显示:根除幽门螺杆菌感染,随访 7.8 年后,胃癌发生率与未根除幽门螺杆菌感染相比明显下降[N Engl J Med, 2001, 345(11): 784-789]。

治疗性随机对照临床试验文献显示:早期胃癌,无论内镜下切除,或者手术切除后,根除幽门螺杆菌感染组胃癌复发率明显低于未根除幽门螺杆菌感染的患者[Ref]。

至此,幽门螺杆菌感染作为胃癌重要的致病因素得到了临床研究的确认。

## 第五节 病例对照研究的其他应用

病例对照研究可以同时研究多个暴露因素,分别评估其 OR 值,评估相互作用与影响。病例对照研究方法可以用来调查某些急性病暴发的原因,还可以用于预后研究评估疾病预后因素。



## 第六节 文献阅读与评价

### 一、文献阅读与评价 1

A case-control study of screening sigmoidoscopy and mortality from colorectal cancer. N Engl J Med, 1992, 326(10):653-657.

1. 目的 乙状结肠镜筛选对于降低结直肠癌死亡率的效应尚不清楚。一项随机对照试验可能可以较好地解决这一问题,但实施较困难。病例对照研究也是估计这种效应的方法之一。

2. 方法 数据来自于凯撒医疗机构 1971 年至 1988 年间死于直肠癌或远端结肠癌的 261 名患者。分别分析病例组和以经年龄、性别匹配的 868 名对照组,在病例组癌症诊断前 10 年乙状结肠镜筛查情况。

3. 结果 病例组仅 8.8% 接受了乙状结肠镜筛选,而对照组为 24.2% (OR 为 0.3, 95%CI: 0.19~0.48)。校正潜在混杂因素后 OR 值为 0.41 (95%CI: 0.25~0.69)。最近一次结肠镜检查间隔 9~10 年或更近,这种负相关关联强度一致。反过来说,在乙状结肠镜筛选到达以上部位的 268 名结肠癌患者与 268 名对照相比,校正 OR 值为 0.96 (95%CI 0.61~1.50)。即使有一些为测量的混杂因素存在,乙状结肠镜检查范围内的结肠癌风险与乙状结肠镜筛选呈特异负相关。

4. 结论 乙状结肠镜筛选可以降低直肠至末端结肠癌症的风险。10 年一次或间隔时间更短的筛选效应可能一致。

5. 评价 这个研究的临床问题是,乙状结肠镜在无症状人群中筛查对于癌症和息肉检出有较高的敏感性,但是它是否能降低死亡率,却无相关的对照数据。因此本研究旨在探讨乙状结肠筛查是否能降低结直肠癌死亡率。考虑到随机对照研究的复杂、耗时等,采用了病例对照研究的方法。病例组选择了 261 名结直肠癌患者,且死亡原因为结直肠癌,这些患者癌症部位位于直肠或乙状结肠(或距肛缘 20cm 以下),可以被乙状结肠镜诊断。对照人群来自参加同一个医疗健康计划的人群,匹配因素包括年龄、性别、进入健康计划时间等。配对比例约 1:4。采集信息主要来自查阅既往的医疗记录。除了查阅乙状结肠镜筛选情况外,另外收集了可能影响乙状结肠镜检查次数的其他混杂因素,包括息肉史,家族性肿瘤史,溃疡性结肠炎病史等。

### 二、文献阅读与评价 2

Proton Pump Inhibitor Use and Risk of Colorectal Cancer: A Population-Based, Case-Control Study. Gastroenterology, 2007, 133(3):755-760.

1. 背景和目的 质子泵抑制剂(PPI)与血清中胃泌素水平升高有关,而高胃泌素血症可能会引起结直肠黏膜增生,与结直肠癌风险有关。所以进行了一项大型人群为基础(丹麦)的病例对照研究来探讨 PPI 是否与结直肠癌的发生有关。

2. 方法 病例组选择来源于 1989~2005 年间丹麦国家医院出院登记系统中的结直肠癌患者。对照选择来自丹麦公民登记系统,比例为 1:10,经年龄和性别匹配。根据处方来确定 PPI 使用情况。

3. 结果 最终确定了病例组 5589 例,对照组 55 890 例。对于从不或极少使用 PPI 药物的(观察时间内一共 <30 片),结直肠癌风险不增加(校正 OR=1.11, 95%CI: 0.97~1.27)。对于频繁使用 PPI 者(多于隔天使用),不管是短期使用(OR=1.07, 95%CI: 0.86~1.34)还是长期使用(大于 7 年, OR=1.09, 95%CI: 0.58~2.06),结直肠癌风险均未增加。

4. 结论 PPI 药物使用不增加结直肠癌风险。

5. 评价 这是一项大型病例对照研究。临床问题是 PPI 药物是否增加结直肠癌发生风险。



病例组选取的是 1989~2005 年间出院登记系统中登记的首次确诊结直肠癌的患者。对照组选择人群对照。

### 三、文献阅读与评价 3

Anti-inflammatory drugs and variceal bleeding: a case-control study. Gut, 1999, 44 (2): 270-273.

1. 目的 非甾体类消炎药(NSAID)有严重的消化道副作用,可能引起消化性溃疡出血。急性食管静脉曲张破裂出血是肝硬化的一个主要并发症。采用病例对照研究的方法研究 NSAID 是否对肝硬化食管胃静脉曲张破裂曲张首次出血的影响。

2. 方法 预先设计调查表调查 125 名肝硬化食管胃静脉曲张破裂出血患者,及 75 名肝硬化食管胃静脉曲张但未出血的患者。

3. 结果 因门脉高压出血住院的肝硬化患者与未出血者相比,1 周前使用 NSAID 的比例更高(31/125, 25%)和(8/75, 11%),OR 为 2.8,  $P=0.016$ 。单独使用阿司匹林或联合使用其他 NSAID 的比例在病例组(21/125, 17%)也高于对照组(3/75, 4%, OR=4.9,  $P=0.007$ )。Logistic 回归分析显示,NSAID 药物使用( $P=0.022$ , OR=2.9, 95%CI: 1.8~4.7)和静脉曲张的直径( $P<0.001$ , OR=4.0, 95%CI: 1.4~11.5)是静脉曲张破裂出血的独立危险因素。

4. 结论 阿司匹林,单独或联合其他 NSAID 药物使用,与肝硬化静脉曲张破裂首次出血相关。考虑到出血并发症的凶险,使用 NSAID 药物时需注意权衡这种出血风险。没有证据表明单独使用非阿司匹林的 NSAID 药物与静脉曲张破裂出血的相关性。

5. 评价 临床问题:NSAID 药物是否与肝硬化食管胃静脉曲张破裂首次出血相关。采用了病例对照研究的方法。病例组:选取因肝硬化门脉高压首次出血的患者,所有患者在入院 12 小时内接受急诊内镜检查。对照组:医院对照,选择同期入院的,经年龄、性别匹配的肝硬化患者。且内镜检查发现存在食管胃静脉曲张,但排除有出血史。信息采集采用问卷调查的方式,由一位经专业训练的医师完成。病例组和对照组基线匹配一致。

#### 思考题

1. 针对上述 3 篇阅读文献(摘要,可以根据需要阅读全文),回答:该研究期望解决的临床问题是什么?科学问题是如何描述的?该研究采用了何种临床研究设计?选择患者和对照来源在哪里?你认为存在哪些偏倚,如何通过改进设计等方法控制偏倚?

2. 收集 HBV 感染与肝癌发生的不同研究设计的临床研究文献,回答:该研究期望解决的临床问题是什么?科学问题是如何描述的?该研究采用了何种临床研究设计?选择患者和对照来源在哪里?你认为存在哪些偏倚,如何通过改进设计等方法控制偏倚?

3. 结合自己专业,选择一种病因未明的疾病,针对可能的危险因素(暴露)与该疾病之间的联系,设计一项回顾性临床研究,内容包括:科学问题与研究目的,研究设计与具体方法,结果测量方法及可能的结果。

(陈世耀)

## 第七章 队列研究

### 临床案例

68岁女性,一年来胸痛,体瘦,没精神,偶尔胃疼,经常失眠、咳嗽,以为是更年期综合征。胃镜活检发现:(食管)黏膜溃疡,幽门螺杆菌检查阴性,送检组织大部分区域被覆上皮脱落,局部被覆柱状上皮,考虑为 Barrett 食管(长节段)。现提倡对 Barrett 食管患者定期进行内镜和活检监测以早期发现癌变,于是每年做内镜和活检监测,4年后内镜复查发现食管腺癌,患者拒绝进一步治疗。

### 第一节 将临床问题转化为科学问题

这是临床常见的病例,作为临床医生,往往仅仅关注患病个体的发病情况 and 治疗方法,将此类问题仅当作临床问题来解决,从而忽略了从群体的角度去思考问题,也即将其转化为科学问题去思考,例如该患者为什么会发生食管腺癌,是否与 Barrett 食管有关? Barrett 食管患者有多大的概率发展成食管腺癌? 该类人群发生食管腺癌的概率是否高于正常人群? 定期对 Barrett 食管进行内镜和活检监测是否值得? 而流行病学就提供了解决这些问题的研究方法。

思路一:与非食管腺癌患者相比,食管腺癌患者中 Barrett 食管的比例是否高于非食管腺癌患者? 这就需要采用病例对照研究设计。在同一来源的人群中选择两组人群:一组患有食管腺癌的人群作病例组,一组未患该病的人群作为对照组,对病例组,追溯5年或者10年前是否患有 Barrett 食管,比例是多少,同时用同样的方法在对照组中追溯5年或者10年前是否患有 Barrett 食管,比较病例组与对照组之间患 Barrett 食管比例的差别。

这样的研究设计虽然花费少、耗时短,但研究在食管腺癌发生后开始,结果已经发生,由果推因,暴露与疾病的时间先后,有时难以判断,论证强度不够,并且无法获得暴露组(Barrett 食管)与非暴露组(无 Barrett 食管)的食管腺癌发生率。

思路二:患 Barrett 食管的人群,在经过若干年之后,发生食管腺癌的概率是否高于无 Barrett 食管的人群,概率增高多少倍? 这样我们可以采用前瞻性队列研究设计,通过病理检测,将人群分为暴露组(Barrett 食管患者)和对照组(无 Barrett 食管人群),观察若干年之后,分别检查两组人群食管腺癌的发生率,比较两组人群食管腺癌发生率的高低。

这样的研究设计尽管耗时较长,但资料可靠,一般不存在回忆偏倚;可直接获得 Barrett 食管患者和无 Barrett 食管人群的发病率,由于是由因到果的研究,因此检验假设的能力较强,一般可证实病因联系。

#### 知识点

#### 队列研究的定义及原理

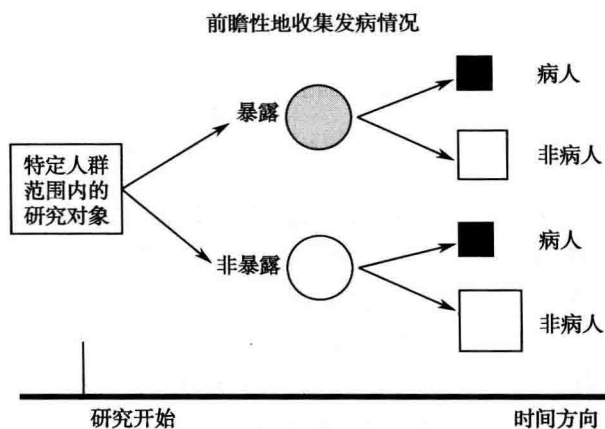
1. 队列研究的定义 队列研究(cohort study)是将人群按是否暴露于某可疑因素及其暴露程度分为不同的亚组,追踪其各自的结局,比较不同亚组之间结局频率的变异,从而



判定暴露因子与结局之间有无因果关联及关联大小的一种观察性研究(observational study)方法。这里观察的结局主要是与暴露因子可能有关的结局。

流行病学中的队列表示一个特定的研究人群组。根据特定条件的不同,队列可分为两种情况:一是指特定时期内出生的一组人群,叫出生队列(birth cohort);另一种是泛指具有某种共同暴露或特征的一组人群,一般即称之为“队列或暴露队列”,如某时期进入某工厂工作的一组人群。根据人群进出队列的时间不同,队列又可分为两种:一种叫固定队列(fixed cohort),是指人群都在某一固定时间或一个短时期内进入队列,之后对他们进行随访观察,直至调查终止,成员没有无故退出,也不再加入新的成员,即在观察期内保持队列的相对固定;另一种叫动态人群(dynamic population),是相对固定队列而言的,即在某队列确定后,原有的队列成员可以不断退出,新的观察对象可以随时加入。

队列研究的结构模式:



2. 基本原理 队列研究的基本原理是在一个特定人群中选择所需的研究对象,根据目前或过去某个时期是否暴露于某个待研究的危险因素,或其不同的暴露水平而将研究对象分成不同的组,观察随访一段时间,检查并登记各组人群待研究的预期结局的发生情况(如疾病、死亡,或其他健康状况),比较各组结局的发生率,从而评价和检验危险因素与结局的关系。如果暴露组(或高剂量暴露组)某结局的发生率明显高于非暴露组(低剂量暴露组),则可推测暴露与结局之间可能存在因果关系。在队列研究中,所选研究对象必须是在开始时没有出现研究结局,但有可能出现该结局(如疾病)的人群。暴露组与非暴露组必须有可比性,非暴露组应该是除了未暴露于某因素之外,其余各方面都尽可能与暴露组相同的一组人群。

### 知识点

#### 队列研究的优点及局限性

1. 优点 ①研究人群定义明确,选择性偏倚较小;②由于是前瞻性的,有可能使测量暴露的方法标准化,以减少观察者、对象和技术变异而引起的误差,又由于事先不知道谁将发病,信息偏倚较小;③可以直接计算暴露组和非暴露组的比率,从而计算出RR和AR等反映疾病危险关联的指标,可以充分而直接地分析病因的作用;④有可能观察到暴露和疾病在时间上的先后;⑤有助于了解人群疾病的自然史。有时还可能获得多种预计以外的疾病的结局资料;⑥可按暴露水平分级,从而有可能观察到剂量反应关系。

2. 局限性 ①不适于发病率很低的疾病的病因研究,因所需对象数量很大,难以达





到。即使是研究常见病,仍需大量对象,才能获得暴露组与对照组之间有意义的差异;②需要长期随访,对象不易保持依从性,容易产生各种失访偏倚;③研究费时间、费人力、物力,其组织与后勤工作相当艰巨;④研究者虽然可预先根据暴露与否进行分组,但有时难以控制暴露以外的其他特征在两组中的分布,而造成混杂偏倚。

## 第二节 队列研究设计

### 一、研究目的

通过队列研究探索 Barrett 食管与食管腺癌的发生关系。

### 二、研究假设

Barrett 食管患者发生食管腺癌的人较非 Barrett 食管患者多。

### 三、暴露人群的选择

暴露人群的选择:① Barrett 食管患者的来源:社区人群患者,但获得全部患者有一定难度,如由于选择条件受限或选择方法有问题,在研究开始时未能发现一些早期患者,从而误认成对照人群,会造成偏倚;②确定来源之后,需要规定统一的诊断标准来选择 Barrett 食管患者:诊断手段与依据,如内镜检查及活检,活检的部位等。

案例例选择:假设我们拟在社区人群中开展一项 Barrett 食管与食管腺癌的发生关系的前瞻性队列研究,我们需要有一个完整的病理登记系统,并记录完整患者资料。现某一国家有一套完整的病理登记系统和癌症登记系统,能收集到规定的一段时间内全国所有的 Barrett 食管患者。

#### 知识点

#### 暴露人群的选择

通常将暴露人群分为四类:一般人群、职业人群、特殊暴露人群和有组织的人群团体。

1. 一般人群 即一个范围明确的地区的全体人群或其样本,由具有不同暴露因素的个体组成;适用于同时观察多种暴露和多种疾病间的关系,若着眼于研究一般人群的发病情况,或暴露因素和疾病在人群中常见,不需要或没有特殊暴露人群,就可以选择一般人群作为暴露人群。如 Framingham 地区心脏病研究,该研究的主要目的,是在一般人群中前瞻性的观察冠心病的发病率及年龄、性别、家族史、职业、文化水平、国籍、血压、血脂、体力活动、吸烟、饮酒等因素在冠心病发生发展中的作用。实际工作中,常选择有组织的人群团体,如机关、团体、学校或详细可靠的人群资料作为一般人群的特殊形式,提高收集随访资料的效率。

2. 职业人群 某些职业中常存在特殊暴露因子,使职业人群的发病或死亡率远远高于一般人群,选择职业人群进行研究,便于证实暴露与疾病的联系。如研究联苯胺的致癌作用,选择染料厂工人;研究石棉致肺癌的作用,选择石棉作业工人等。

3. 特殊暴露人群 指具有特殊暴露经历的人群。如研究电离辐射的危险性选择原子弹爆炸后的存活者、铀矿工人或医疗过程中的暴露者(放疗后的人)。由于人们对某些职业暴露和某些特殊暴露的危险性多半不是一开始就认识到的,一旦认识到了,大多都



采取了防护措施以减少暴露,所以一般不易进行前瞻性队列研究,而常使用历史性队列研究。

4. 有组织的人群团体 该类人群可看作一般人群的特殊形式,如医学会会员,工会会员,机关、社会团体、学校或部队成员等。选择这样的人群的主要目的是利用他们的组织系统,便于有效地收集随访资料。如 Doll 和 Hill 选择英国医师会员以研究吸烟与肺癌的关系的人群。

#### 四、对照人群的选择

选择好一组研究人群建立队列后,其中患 Barrett 食管的患者组成暴露组,其余非暴露组即为对照组。队列研究结果的真实性依赖于是否正确选择了对照人群。选择对照组的基本要求是尽可能高的可比性,即对照人群除未暴露于所研究的因素外,其余各因素的影响或人群特征(年龄、性别、职业、民族、文化程度等)都应尽可能与暴露组相同,这称为齐同。

案例对照选择:选择全人群作为研究对象,其中符合规定的 Barrett 食管患者组成暴露组,其余非 Barrett 食管患者组成对照组,对照组的食管腺癌发病率资料来源于国家统计数据。

##### 知识点

##### 对照人群的选择

对照人群大致可分为四种:

1. 内对照 即先选择一组研究人群,将其中暴露于所研究因素的对象作为暴露组,其余非暴露者即为非暴露组。也就是说在选定的一群研究对象内部既包含了暴露组,又包含了对照组,不需到另外的人群中去找。这样做的好处是,除暴露因素本身外,其他因素可比性较强,研究偏倚较小;选取对照比较省事,并可以无误地从总体上了解研究对象的发病情况。

2. 外对照 选择人口学特征与暴露组相似的另一个非暴露人群作对照,称为外对照。在以职业人群或特殊暴露人群为暴露组时,常需选择外对照。如以放射科医生为研究射线致病作用的暴露对象时,可以不接触或极少接触射线的五官科医生为外对照。

3. 总人口对照 用暴露人群所在地区的一般人群的发病率、死亡率或其他结局与暴露组相比较。这种对照统计资料容易得到,但比较粗糙,有时暴露与疾病的联系会被低估。实际应用时,常采用间接标化比(即用暴露组发病或死亡数与用总人口率算出的期望发病或死亡数求标化比)来代替两组率的直接比较。

4. 多重对照 或叫多重对照,即用上述两种或两种以上的形式同时作对照,以减少只用一种对照所带来的偏倚,增强结果的可靠性。

#### 五、样本量计算

样本量的大小主要取决于四个因素:①非暴露人群的发病率( $P_0$ ), $P_0$  越接近 0.50,所需样本越小;②暴露人群的发病率( $P_1$ ),暴露人群与对照人群两发病率之差越大,所需样本量越小。若暴露人群发病率  $P_1$  不易获得,可设法得到相对危险度(RR)的数值,由  $P_1=RR \times P_0$  求得  $P_1$ ;③显著性水平  $\alpha$ ,即假设检验时的第 I 类错误。要求假阳性错误出现的概率越小(即  $\alpha$  越小),需样本量越大,通常  $\alpha$  取 0.05 或 0.01;④检验效能  $1-\beta$ ,即检验假设时能够避免假阴性出现的能力,



$\beta$  为检验假设时出现第 II 类错误的概率。若要求  $1-\beta$  越大, 即  $\beta$  越小, 所需样本量也越大。通常  $\beta$  取 0.10。

可用下列公式计算样本量 (N)。

$$N = \frac{(Z_{\alpha}\sqrt{2\bar{P}\bar{Q}} + Z_{\beta}\sqrt{P_0Q_0 + P_1Q_1})^2}{(P_1 - P_0)^2} \quad \text{式 7-1}$$

式中  $Z_{\alpha}, Z_{\beta}$  为  $\alpha, \beta$  所对应的标准正态差。  $\bar{Q}=1-\bar{P}$  为  $P_0$  和  $P_1$  均值。

通过查阅文献, 本案例中  $P_0$  取 0.0001,  $P_1$  取 0.001, 设  $\alpha=0.05$  (双侧),  $\beta=0.10$  (单侧), 根据计算, 本研究案例估计暴露组和非暴露组各需观察 14 247 人。

## 六、暴露测量

本研究案例中的暴露为 Barrett 食管。Barrett 食管是指在食管黏膜修复的过程中, 食管、贲门交界处的齿状线 2cm 以上的食管鳞状上皮被化生的柱状上皮所替代或伴有肠化。Barrett 食管诊断主要靠内镜及活检。

### 知识点

#### 暴露问题

1. 暴露的定义 队列研究是根据是否暴露于危险因素而对研究对象进行分组的, 因此弄清楚暴露的涵义才能准确把握队列研究。在流行病学研究中, 暴露 (exposure) 是指研究对象接触过某种待研究的物质 (如重金属)、具备某种待研究的特征 (如年龄、性别及遗传等) 或行为 (如吸烟)。暴露在不同的研究中有不同的含义, 暴露可以是有害的, 也可以是有益的, 但都是需要研究的。

2. 暴露因素的规定 暴露因素是泛指各种会影响人体健康的具体的物理、化学和生物因素。通常把导致疾病事件增加的暴露因素称为危险因素 (或致病因素), 把导致疾病事件降低的暴露因素称为保护因素。暴露因素的含义是相对的, 它既可以是某种疾病的致病因素或保护因素, 也可以是另一暴露因素的后果, 即疾病。例如, 高血压是脑血管病的暴露因素, 但它也可能是遗传或营养等其他暴露因素所产生的疾病事件。这种暴露因素的相对性取决于研究目的和研究者对暴露因素的认识水平。因此, 在研究开始前应详细了解所要研究的暴露因素, 并给予明确定义, 定义越具体越好。例如, 成年人高血压的标准是年龄  $\geq 18$  岁, 舒张压  $> 95\text{mmHg}$  (12.7kPa) 或收缩压  $\geq 150\text{mmHg}$  (20kPa) 连续半年。总之, 暴露因素须有明确的规定, 包括暴露因素的性质、暴露的时间、频率、强度等。若将暴露因素定量, 则应明确其单位。如不易获得准确的定量资料, 可将暴露水平粗略地分级。

## 七、随 访

对两组人群随访观察 6 年, 并记录该期间内两组人群食管腺癌发病人数 (食管腺癌的测量方法见预期结果部分)。

队列研究资料的收集包括两个主要方面, 即基础资料的收集和随访。随访期间由于种种原因某些研究对象脱离了观察, 研究者无法继续随访他们, 这种现象叫失访。失访会对研究结果产生影响。当失访率大于 10% 时, 应采取措施对其可能产生的影响做进一步估计。若失访过多, 如失访率达 20% 以上, 则研究的真实性会受到严重怀疑。因此保证随访成功是队列研究成功的关键之一。一般说来, 随访有三个目的: ①确定研究对象是否仍处于观察之中; ②确定研究人群中的各种疾病事件; ③进一步收集有关暴露和混杂因素的资料。由此可以看出, 随访的对象是



所有研究对象;随访内容应与取得基本信息时的完全一样,其具体项目可视研究目的与设计而不同。

由于涉及人时数和发病密度的计算,每个研究对象开始随访和终止随访的日期都应明确规定。随访期的确定应以暴露因素作用于人体至产生疾病结局的一般潜隐期为依据。在随访中会碰到两种情况:即某研究对象出现了预期的结果(称为观察终点),此时就不再对该对象继续随访。而有的研究对象没有出现结局疾病,则对其坚持随访,直到规定的观察期结束(观察终止时间)。另一个应该确定的指标是随访的间隔。如果观察时间较短,在观察终止时一次搜索资料即可。反之需多次随访,其间隔与次数视具体情况而定。如 Framingham 地区冠心病随访研究每两年随访一次,历时 24 年。英国以医生为对象进行的吸烟与肺癌的队列研究,历时 20 余年,分别于 1957 年、1966 年和 1972 年进行三次随访。

随访的方法有:①利用常规登记的人群和疾病资料随访。在某些发达国家,每个公民都有一个全国计算机联网的个人识别号,可查到有关就业、医疗、死亡等情况。在我国,可利用职工人事登记资料、肿瘤及传染病报告卡、死亡证明等;②进行特殊安排的随访,如定期家庭访视、电话访问或信访等。必要时也可以进行健康检查、采样检测。随访人员应经过严格培训和考核。

## 八、预期结果

患 Barrett 食管的人群,在经过若干年之后,发生食管腺癌的概率高于无 Barrett 食管的人群,差异有统计学意义。

原发性食管腺癌的最后诊断主要依靠病理组织学检查,其诊断标准为:①腺癌位于食管;②对于食管下端发生的原发性食管腺癌除手术记载肯定无贲门癌外,还需经病理证实癌灶上下缘有鳞状上皮存在,且不伴有胃癌或贲门癌;③ HE 染色切片镜下见癌组织为有一定分化程度的腺样结构;④黏膜组织化学染色阳性。

### 知识点

#### 结 局

研究结局变量(outcome variable)也叫结果变量,简称为结局,是指随访观察中将出现的预期结果,即研究者追踪观察的事件。结局就是队列研究观察的自然终点,与观察期的终点是不同的概念。结局不仅限于发病,还有死亡或者各种生理生化指标、生命质量的变化;结局变量既可是定性的,也可是定量的,如血清抗体的滴度、血糖、尿糖及血脂等。

结局判定,应给出明确统一的标准,并在研究的全过程中严格遵守。考虑疾病的诊断标准时要注意一种疾病往往有轻型和重型、不典型和典型、急性和慢性等多种表现。因此,应尽量按国际或国内统一的标准判断结局,还要记录下其他可疑症状或现象供以后分析。

队列研究的优点之一是一次可以同时收集到多种结局资料,研究一因多果的关系,故在队列研究中除确定主要研究结局外,可考虑同时收集多种可能与暴露有关的结局。

#### (一) 资料整理

在队列研究中研究结果以“2×2”表格表示,其中行代表暴露情况,列代表发病情况(表 7-1)。



表 7-1 累积发病率资料整理表

|      | 发病        | 未发病       | 合计          | 发病率     |
|------|-----------|-----------|-------------|---------|
| 暴露组  | a         | b         | $a+b=n_1$   | $a/n_1$ |
| 非暴露组 | c         | d         | $c+d=n_0$   | $c/n_0$ |
| 合计   | $a+c=m_1$ | $b+d=m_0$ | $a+b+c+d=t$ |         |

但由于观察时间较长,难以做到人口稳定,如观察对象进入队列的时间不一致;由于迁移、死亡或其他原因造成失访等,则应以人时为单位来计算发病率。以人时为单位计算出来的率带有瞬时频率的性质,因此区别于累积发病率而称之为发病密度(incidence density)。对于应计算发病密度的队列研究资料,其资料整理和率的计算,除了将每个观察对象折算成“人年”以代替“人”,其余均与累积发病率相同。其资料整理模式如表 7-2 所示。

表 7-2 发病密度资料整理表

|      | 发病数     | 人时数(人年/月)   | 发病密度    |
|------|---------|-------------|---------|
| 暴露组  | a       | $N_1$       | $a/N_1$ |
| 非暴露组 | b       | $N_0$       | $b/N_0$ |
| 合计   | $a+b=M$ | $N_1+N_0=T$ |         |

检验暴露组与对照组的发病(死亡)率是否有显著性差异可采用多种方法。

1. 若观察样本量较大,样本率的频数分布近似正态分布,可用  $u$  检验。
2. 如果率比较低,样本率的频数分布不符合正态分布,可改用二项分布或泊松分布检验卡方检验。此外,还可以用卡方来检验两组的率是否有显著性差异。其检验方法可参阅有关书籍。

### (二) 人时的计算

“人时”是观察人数与观察时间的综合指标。它是研究人群中所有个体暴露于所研究因素的时间的总和,即人数  $\times$  每人暴露时间 = 人时数,时间可以是日、月、年中任何一种单位,通常多用人年。计算人时的方法很多,步骤也比较复杂,这里只介绍以个人为单位计算人年的方法。该方法较精确,但费时间,如样本不太大时,可用此法计算,如表 7-3,表 7-4。现在已有专用于人年计算的计算机软件,如 PYRS、OCMAP 等。

表 7-3 3 个研究对象的出生日期与进出研究时间资料

| 对象编号 | 出生日期       | 进入研究时间     | 退出研究时间              |
|------|------------|------------|---------------------|
| 1    | 1927-03-21 | 1966-07-19 | 1977-09-14(迁居外地)    |
| 2    | 1935-04-09 | 1961-11-11 | 1973-12-01(死亡)      |
| 3    | 1942-11-12 | 1970-02-01 | 1981-01-01(观察结束时健在) |

表 7-4 3 例人年的计算

| 年龄组 | 对象 1<br>1927-03-21 出生 | 对象 2<br>1935-04-09 出生                             | 对象 3<br>1942-11-12 出生                                | 暴露人年 |
|-----|-----------------------|---------------------------------------------------|------------------------------------------------------|------|
| 25~ |                       | 1961-11-11~1965-04-08 共 3<br>年 4 个月 27 天即 3.41 人年 | 1970-20-01~1972-11-11<br>共 2 年 9 个月 10 天即 2.78<br>人年 | 6.19 |





续表

| 年龄组 | 对象 1<br>1927-03-21 出生                    | 对象 2<br>1935-04-09 出生                             | 对象 3<br>1942-11-12 出生                                | 暴露人年  |
|-----|------------------------------------------|---------------------------------------------------|------------------------------------------------------|-------|
| 30~ |                                          | 1965-04-09~1970-04-08<br>共 5.00 人年                | 1972-11-12~1977-11-11<br>共 5.00 人年                   | 10.00 |
| 35~ | 1966-07-19~1967-03-20<br>共 8 个月即 0.67 人年 | 1970-04-09~1973-12-01 共 3<br>年 7 个月 22 天即 3.65 人年 | 1977-11-12~1981-01-01<br>共 3 年 1 个月 20 天即 3.14<br>人年 | 7.46  |
| 40~ | 1967-03-21~1972-03-20<br>共 5.00 人年       |                                                   |                                                      | 5.00  |
| 45~ | 1972-03-21~1977-03-20<br>共 5.00 人年       |                                                   |                                                      | 5.00  |

(流行病学第四版教材,2000 年)

(三) 暴露与疾病关联强度的测量

队列研究的最大特点在于可确证暴露与疾病的因果联系。通常用以下几个指标来表示这种联系的强度。首先,将资料整理成表 7-1 的模式,然后计算下列指标。

1. 相对危险度(relative risk,RR)也叫危险度比(risk ratio)或率比(rate ratio),均以 RR 表示,它是说明暴露与疾病关联的强度及其在病因学上意义大小的指标。

$$RR=\frac{I_e}{I_0}=\frac{a/n_1}{c/n_0}$$

式 7-2

设  $I_e=a/n_1$  为暴露组的率, $I_0=c/n_0$ ,则 RR 表明暴露组发病或死亡的危险是非暴露组的多少倍。对于 RR 值的大小反映关联强度应根据的标准可参考表 7-5,工作中仍需根据实际情况 RR 值的可信区间来判断其意义。

表 7-5 相对危险度与关联的强度

| 相对危险度   | 关联的强度 |
|---------|-------|
| 0.9~1.0 | 无     |
| 0.7~0.8 | 弱     |
| 0.4~0.6 | 中等    |
| 0.1~0.3 | 强     |
| <0.1    | 很强    |

(Monson RA,1980)

相对危险度是估计暴露与疾病关联的一个点估计值,用它直接估计关联强度大小误差较大。考虑到抽样误差的存在,常按照一定的概率(一般为 95%)以区间来估计 RR 总体所在的范围。RR 可信区间上下限的数值即为可信限。其计算公式见公式 7-3,常取 95% 的 Z 值 1.96 计算 RR 的 95% 可信限。

$$RR_U RR_L=RR^{(1\pm Z\sqrt{x^2})}$$

式 7-3

2. 归因危险度(attributable risk,AR)又叫特异危险度,或叫率差(rate difference, RD),表明暴露组与对照组发病危险相差的绝对值,即暴露者单纯由于暴露而增加的发病概率。

RR 与 AR 同为估计危险度的指标,但其公共卫生意义不同。RR 说明暴露使个体比未暴露情况下增加相应疾病的危险程度,是比值;AR 则是暴露使人群比未暴露情况下增加超额疾病的数量。如果暴露因素消除,就可以减少这个数量的疾病。下面以表 7-6 为例说明二者的区别。

学  
习  
笔  
记



$$AR = I_e - I_0 = \frac{a}{n_1} - \frac{c}{n_0} \quad \text{式 7-4}$$

$$\text{或 } AR = I_0(RR - 1) \quad \text{式 7-5}$$

表 7-6 吸烟者与非吸烟者死于不同疾病的 RR 与 AR

| 疾病    | 吸烟者<br>(1/10 万人年) | 非吸烟者<br>(1/10 万人年) | RR         | AR     |
|-------|-------------------|--------------------|------------|--------|
|       |                   |                    | (1/10 万人年) |        |
| 肺癌    | 48.33             | 4.49               | 10.8       | 43.84  |
| 心血管疾病 | 294.67            | 169.54             | 1.7        | 125.13 |

(Lee, 1982)

它说明吸烟对每个暴露者来说,患肺癌的危险性比患心血管病的危险大得多。但就整个人群来看,吸烟引起心血管病的死亡率却比肺癌高。前者具有病因学意义,后者更有疾病预防和公共卫生上的意义。

3. 归因危险度百分比(AR%) Lilienfeld 等称它为病因分值 EFe,是指暴露人群中发病归因于暴露的成分占全部病因的百分比。

$$AR\% = \frac{I_e - I_0}{I_e} \times 100\% \quad \text{式 7-6}$$

$$\text{或 } AR\% = \frac{RR - 1}{RR} \times 100\% \quad \text{式 7-7}$$

学习  
笔记

4. 人群归因危险度 人群归因危险度(population attributable risk, PAR)说明人群由于暴露于某一危险因子而增加的发病率。PAR 与 AR 不同,因为 AR 仅仅是从抽取的人群资料中计算出来,而研究对象暴露与非暴露的比例不会与目标人群中两者的比例一致,若目标人群中暴露的比例低,尽管 AR 较高,人群中的实际发病者也不会很高,即人群中的归因危险度受人群暴露比例的影响。设  $I_t$  为全人群的率,  $P_e$  为全人群的暴露比例。

$$PAR = I_t - I_0 = AR \times P_e \quad \text{式 7-8}$$

5. 人群归因危险度百分比(PAR%) 类似地可以得到

$$PAR\% = \frac{I_t - I_0}{I_t} \times 100\% = \frac{P_e(RR - 1)}{P_e(RR - 1) + 1} \times 100\% \quad \text{式 7-9}$$

可见 PAR 和 PAR% 取决于暴露因子的流行率和相对危险度两个因素,可用于估计某危险因子对整个人群引起的疾病负担,说明在整个社会的卫生问题中哪些是重要的,在卫生保健工作及卫生管理上意义较大。

### 第三节 研究报告与问题分析

研究结果的报告包括研究背景与目的、方法、结果、讨论与结论四个方面。讨论对研究设计、可能存在的偏倚以及可能影响结论的因素进行恰当的分析。

#### 一、研究案例 1

Barrett 食管患者中的食管腺癌:被高估的风险[Gut, 1989, 30(1):14-18]

1. 背景与目的 Barrett 食管是食管癌的危险因素,因此人们提出每年对 Barrett 食管患者作内镜和活检监测以早期发现癌变。本研究的目的是探讨 Barrett 食管患者中食管腺癌的发病率并以期获得危险性稍高的亚组。



2. 方法 采用历史性队列研究方法。自 1973 年 11 月至 1986 年 5 月间确诊的全部 Barrett 食管患者组成暴露组,然后回顾性调查食管腺癌的发生情况,并计算腺癌发病率。对照人群为荷兰同年龄组的健康人群,其发病率数据来自中央统计局 1985 年的死亡统计报告。

3. 结果 166 名 Barrett 食管病例符合入选条件,平均年龄 62 岁,最终有 155 人进入分析,应答率为 93%。12 年随访结束后共 4 人发生食管腺癌,发生率为 1/170 人年,比正常对照人群增加 30 倍。但与年龄和性别匹配的对照人群相比,Barrett 食管患者的生存率并无差异。

4. 结论 我们不提倡对 Barrett 食管患者进行系统的内镜监测,但可以筛选出具备特别风险的亚组进行系统监测。

5. 评价 研究案例 1 采用了历史性队列研究,研究开始时暴露和疾病均已发生,研究对象的确定与分组是根据研究开始已掌握的历史资料,这种研究可在短时间内完成,不仅节约了时间,而且仍属于前瞻性的研究。

## 二、研究案例 2

Barrett 食管患者中食管腺癌的发病率研究[NEJM,2011,365(15):1375-1383]

1. 目的 采用人群为基础的队列研究获得 Barrett 食管患者中食管腺癌发病率的准确数据。

2. 方法 在丹麦开展一项全国性的、以人群为基础的前瞻性队列研究。自 1992 年到 2009 年丹麦全部 Barrett 食管患者(经内镜组织学检测,诊断标准依据 SNOMED 标准)被选为暴露组,数据来源于丹麦病理登记中心和癌症登记中心。计算腺癌的发病率(每 1000 人年中发病数)和相对危险度 RR。

3. 结果 暴露组共有 Barrett 食管患者 11 028 人,平均随访 5.2 年,在初次内镜检查后的第一年的时间里,131 例新发食管腺癌病例,在随后的几年中,有 66 例新发病例,食管腺癌发病率为 1.2/1000 人年(95%CI:0.9~1.5)。与正常人群相比,在 Barrett 食管患者组中,食管腺癌的相对危险度(RR)为 11.3(95%CI:8.8~14.4)。每年患食管腺癌的风险为 0.12%(95%CI:0.09~0.15)。

4. 结论 Barrett 食管为食管腺癌的危险因素,但本研究获得的绝对年度风险,0.12%,远低于现在监测指南的基线风险值,0.5%。本研究对现提倡的对未合并发育不良的 Barrett 食管患者定期作内镜和活检监测以早期发现癌变的做法提出质疑。

### 知识点

#### 历史性队列研究

历史性队列研究(retrospective cohort study)在研究开始时暴露和疾病均已发生,即研究的结局在研究开始时已从历史资料中获得,研究对象的确定与分组是根据研究开始已掌握的历史资料,这种设计模式即为历史性队列研究,也称为非同时性或非即时性(nonconcurrent)队列研究。这种研究方法无需等待疾病的发生,暴露和结局资料可在短时间内搜集完,并且可以同时进行,但应注意其观察性质仍属前瞻观察(图 7-1 的左半部分)。

历史性队列研究在研究开始时,暴露和疾病均已发生,可迅速得到研究结果,大大节省了时间、人力和物力。因此这种研究适宜于诱导期长和长潜伏期的疾病,并且也常用于具有特殊暴露的职业人群的研究。但因资料积累时未受到研究者的控制,内容未必符合要求,所以历史性队列研究仅在具备详细、准确而可靠的文字资料的条件下才适用。譬如具备医院的病历、出生记录、工厂的档案和车间的工作记录等资料。

另外,队列研究还包括双向性队列研究(ambispective cohort study),也称混合性队列研究,即在历史性队列研究之后,继续进行一段时间的前瞻性队列研究。这种研究方法兼有上述两种方法的优点,在一定程度上弥补了二者的不足,在实际工作中常常用到,适用范围较广。历史性与前瞻性队列研究的原理见图 7-1。

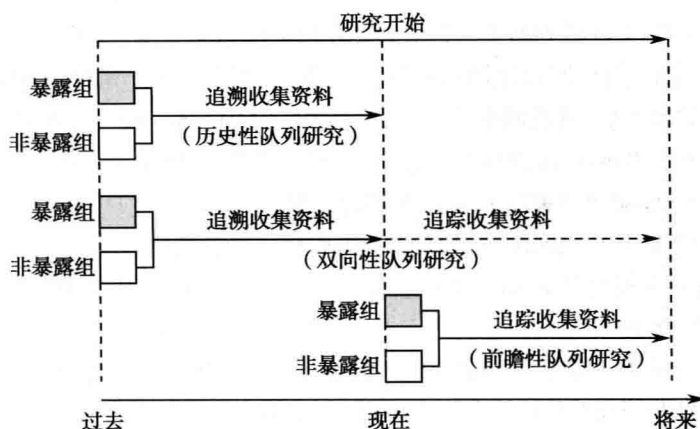


图 7-1 历史性与前瞻性队列研究示意图

5. 评价 研究案例 2 暴露组中大部分新发食管腺癌病例都出现在研究开始的第一年中,这就可能存在选择偏倚,原因是:在研究开始时未能发现早期腺癌患者。本研究解决的办法是计算暴露组发病率时删除第一年的病例,只用了随后几年的 66 例腺癌病例。从而避免了高估暴露组的发病风险。

### 知识点

#### 队列研究中的偏倚及控制

##### 1. 偏倚的类型

(1) 选择偏倚:由于选择研究对象的条件受限制或选择对象的方法有问题,而使研究人群中某个或某些非研究因素的分布与目标人群中该因素的分布不一致,造成研究结果偏离真实情况,就是产生了选择偏倚。选择偏倚发生的原因有:①最初选定参加研究的对象中有人拒绝参加了;②进行历史性队列研究时,有些人的档案丢失了或记录不全;③研究对象由志愿者组成,他们往往是较健康或具有某种特殊倾向或习惯的;④早期患者,在研究开始时未能发现,如肿瘤早期;⑤暴露与疾病的规定不明确,有时是执行得不严格等。在进行职业流行病学研究时,由于被选择作为暴露组的工人的健康状况优于一般人群,导致暴露组的发病率或死亡率低于一般人群,即发生了所谓的健康工人效应(health worker effect)。发生这种选择偏倚的研究常会低估暴露与疾病的联系。

(2) 失访偏倚:队列研究的研究方法决定了它不可避免地要发生失访偏倚,因为在一个较长的随访观察期内,总会有对象迁移,外出,死于非终点疾病或拒绝继续参加观察而退出队列。这种偏倚实质上与选择偏倚相同,即使研究人群与目标人群的人群特征发生了偏差,但它是在追踪随访过程中出现的。一般而言,一项研究的失访率最好不超过 10%,否则其结论的真实性值得怀疑。

(3) 信息偏倚:在收集和整理有关暴露和疾病的资料时所出现的系统误差称为信息偏倚。它主要取决于调查的内容、受调查者的素质和合作程度,以及资料收集过程中的质量控制好坏。引起信息偏倚最常见的情况有:①测量仪器不精确,检验技术不熟练;②被调查者故意谎答或不应答;③医生诊断偏严或偏松;④调查者询问技术不当而诱使被调查者做某一倾向性的回答;⑤长期随访时,使用的调查方法或诊断标准不一致,从而导致错误分类偏倚。

(4) 混杂偏倚:在对某病的病因学研究中,当对所关心的某种暴露因素与这种疾病之间



的联系定量估计时,由于其他外部因素的影响,致使暴露与疾病之间联系的真实性被歪曲,联系强度被放大或缩小,这种歪曲联系强度的作用被称为混杂作用(confounding effect),产生混杂作用的外部因素称为混杂因子(confounder 或 confounding factor)。混杂作用是在研究的设计阶段未对混杂因子加以控制或分析资料时未能进行正确校正所致,混杂偏倚在研究中可以避免和控制。混杂因子既是疾病的危险因素,又与所研究的暴露因素之间存在统计学联系,且它不是暴露因素与疾病因果关系链上的中间变量。正是由于混杂因子、暴露因素和疾病三者之间的内在关系造成了当混杂因子在暴露组与对照组中的分布不均衡时就会产生混杂偏倚。性别、年龄和吸烟是最常见的三个混杂因子。

## 2. 偏倚的控制

(1) 选择偏倚的控制:严格按照规定的标准选择对象,尽量使暴露组与对照组的人群特征相近,尽量使用敏感的疾病早期检查技术。

(2) 失访偏倚的控制:主要靠提高研究对象的依从性。在尽量减少失访的基础上,对失访者和已随访者的特征做比较分析,从各种途径了解失访者最后的结局,并与已随访者的最后观察结果做比较,有助于正确估计研究结果的正确性。

(3) 信息偏倚的控制:依靠精确的测量,同等地对待每个研究对象,提高调查诊断技术,明确各项标准,严格按照规定执行,可有效地减少信息偏倚的发生。

(4) 混杂偏倚的控制:在研究者有能力识别混杂因子的前提下,研究设计阶段可采用限制研究对象的选择条件和匹配的方法来控制;分析资料阶段利用分层分析、标准化和多因素分析对混杂偏倚加以控制。

## 第四节 临床问题的回答

队列研究及大型综述研究结果显示,Barrett 食管患者组若干年后食管腺癌的发生率明显高于无 Barrett 食管的人群,但是对 Barrett 食管患者每年进行内镜及活检监测的有效性 & 价值提出来质疑。

## 第五节 文献阅读与评价

### 一、文献阅读与评价 1

Analysis of risk factors for stroke in a cohort of men born in 1913. N Engl J Med, 1987, 317(9): 521-526.

1. 目的 探索父母死于脑卒中及其他一些潜在的危险因素与子代脑卒中发病风险的关系。

2. 方法 我们采用出生队列研究。首先选择一组 1913 年出生的男性组成出生队列,然后通过调查获得基本信息,进行随访并记录他们脑卒中发病情况。利用单因素和多因素方法进行分析。

3. 结果 出生队列共有 789 人,基线调查时他们都是 54 岁,随访 18.5 年后,57 人(7.2%)发生了脑卒中。单因素分析结果显示,高收缩压( $P=0.004$ ),高舒张压( $P=0.0001$ ),较大的腰围( $P=0.007$ ),较大的腰臀比( $P=0.0004$ ),增高的血浆纤维蛋白原( $P=0.01$ )以及较低的肺活量( $P=0.03$ )都与脑卒中有关系。若母亲死于脑卒中,其儿子患脑卒中的概率比常人增高 3 倍( $P=0.0005$ )。其他潜在可能的危险因素还包括:体重指数、血清胆固醇水平、血糖水平、吸烟、冠心病、心电图描记的左心室肥大的迹象和父亲因脑卒中死亡的家族史。多因素分析结果显示,血压升高、腹部肥胖,





增加血浆纤维蛋白原水平和母亲因脑卒中死亡的家族史仍会增加子代脑卒中的风险。

4. 结论 在中年男性脑卒中危险因素中应该加上母亲死于脑卒中家族史。

5. 评价 这个研究的临床问题是,具有母亲死于脑卒中的家族史是否会增加子代患脑卒中的风险。研究者采用了出生队列进行研究,分析同一年代出生的一批人在不同年代的发病或死亡水平随年龄的变化趋势。可以将年龄、所处时代及队列暴露经历三者对疾病分布的影响区分开来,从而可以确定疾病或死亡与年龄的真实联系。

## 二、文献阅读与评价 2

Long-term survival of a cohort of community residents with asthma. *N Engl J Med*, 1994, 331 (23): 1537-1541.

1. 背景和目的 研究证明哮喘已被列为美国人口潜在死因之一,报道也称与哮喘有关的死亡率不断增加。我们通过对一组以人群为基础的哮喘患者队列进行长期生存研究从另一角度探讨了这一问题。

2. 方法 我们选定了位于明尼苏达州的罗切斯特的所有居民建立队列,自 1964-01-01 至 1964-12-31,通过回顾所有哮喘患者的医疗记录和相关诊断,利用预先定义的哮喘标准对他们进行确诊。在最后一次随访时确认患者最后的生存状况,通过查阅医疗记录、死亡证明和尸检报告确定死亡原因。

3. 结果 根据诊断标准,2499 人确诊或可能患有哮喘。平均随访时间为 14 年(范围:0~29)。在 32 605 人年的随访中,有 140 人死亡。总体存活率与罗切斯特居民的预期生存没有明显不同,但 35 岁以上并合并另一肺部疾病(主要是慢性阻塞性肺疾病)的哮喘患者生存小于预期。研究对象中 4% 的死亡是由于哮喘,而且全部是成年人,生存率与哮喘发作的年龄无关。

4. 结论 没有合并其他肺部疾病的哮喘患者的生存与预期生存没有明显不同,但 35 岁以上合并慢性阻塞性肺疾病的哮喘患者生存比预期糟糕。本研究中,哮喘只引起 4% 的死亡,也无证据表明最近诊断为哮喘的患者死亡风险增加,说明以社区为基础的哮喘患者通常有一个良好的预后。

5. 评价 这是利用社区为基础的队列研究生存问题。临床问题是哮喘患者的生存是否比预期生存低。所有罗切斯特的居民组成研究队列,并且有长达 14 年的随访。这种研究的因果论证强度较大。本项目是研究患哮喘后的生存时间,可见,队列研究不仅可以检验病因假设,还可研究某种疾病的长期变动趋势。

## 三、文献阅读与评价 3

Statins are associated with a reduced risk of hepatocellular carcinoma in a large cohort of patients with diabetes. *Gastroenterology*, 2009, 136 (5): 1601 - 1608.

1. 背景 实验研究表明他汀类药物具有潜在预防癌症的效果。鉴于他汀类药物使用的日益普及与不断上升的肝细胞癌(HCC)发病率,他汀类药物和肝癌之间的潜在关联是一个重要问题。

2. 方法 我们在-组糖尿病患者队列中进行了匹配的病例-对照研究。进入队列 6 个月 after 发生的 HCC 患者组成病例组,利用发病率密度抽样抽取年龄和性别匹配的未患 HCC 者作为对照组。采用 logistic 回归计算 OR 值及其 95% 可信区间(95%CI)。

3. 结果 最终确定了病例组 1303 人,对照组 5212 人,平均年龄 72 岁,99% 为男性。病例组中(34.3%)至少有一个替代他汀类药物处方的比例显著小于对照组(53.1%)。HCC 和降低胆固醇或甘油三酯的非他汀类药物间没有关系。任何他汀类药物处方的未调整 OR 值为 0.46 (95%CI:0.40~0.517),调整的 OR 值为 0.74(95%CI:0.64~0.87)。为了减少潜在的混杂效应,我



们对没有记录肝病的患者进行了亚组分析,OR 值有轻微改变,但仍有显著的统计学意义 0.63 (95%CI:0.50~0.78)。

4. 结论 在糖尿病患者中,使用他汀类能显著降低肝细胞癌的风险。

5. 评价 临床问题:他汀类药物能否降低肝癌发生率。研究采用了巢式病例对照研究 (nested case-control study),在原有的糖尿病队列中插入一个病例对照研究,是 Mantel 于 1973 年提出的一种将队列研究和病例对照研究结合起来的方法,特别适用于研究因素包括有复杂生化分析的队列研究。其基本方法是首先进行队列研究,收集所有观察对象的暴露信息及有关的混杂资料,随访结束后以队列中的病例为病例组,并按一定条件在同一队列中选择对照,进行病例对照研究。巢式病例对照研究集中了队列研究和病例对照研究的优点。

### 思考题

1. 针对上述 3 篇阅读文献,回答:该研究期望解决的临床问题是什么? 该研究采用了何种临床研究设计? 该种设计方法的优缺点是什么? 暴露组和对照组来源是哪里? 你认为存在哪些偏倚,如何控制这些偏倚?

2. 结合自己专业,选择一种病因未明的疾病,针对可能的危险因素(暴露)与该疾病之间的联系,设计一项前瞻性临床研究,内容包括:科学问题与研究目的,研究设计与具体方法,结果测量方法及可能的结果。

(吕明)

## 第八章 随机对照临床试验

### 临床案例

患者男性,67岁,因阵发性左侧肢体无力2天来院就诊。2天前无明显诱因活动中突然出现左侧肢体无力,表现为上肢可持物,下肢尚可独立行走,持续30分钟左右后缓解,2天来上述症状反复发作。患者既往高血压病史10年。否认冠心病、糖尿病、脑梗死病史。吸烟30余年,每天10支,无饮酒嗜好。入院查体:血压160/100mmHg,四肢肌力、肌张力正常,神经系统查体未见明显阳性体征。查头CT未见出血灶,头MRI未发现明显病灶。

### 第一节 如何获得临床研究问题

短暂性脑缺血发作(TIA)和轻型脑卒中是神经内科常见的疾病,约占全部脑血管病的60%。对于轻型脑卒中和TIA患者的早期治疗一直是国内外医学界忽视的问题,并且仍然存在很多争议问题,需要开展相关的医学研究来解决。如对于类似的患者,是应该应用阿司匹林,还是氯吡格雷,还是联合应用阿司匹林与氯吡格雷?如果采用强化抗血小板治疗,应持续多长时间较为合适,会不会增加出血的风险?针对这一常见病和多发病,我们可以开展很多的临床治疗相关研究。如何从临床问题中获得有意义且具有探索性的研究课题,即将临床问题转化为科学研究问题?

#### 一、明确临床问题

1. 阅读文献查阅相关背景资料 查阅脑血管病治疗相关指南,发现目前国内外最新指南均只笼统针对缺血性脑卒中或TIA推荐使用阿司匹林作为急性期治疗和二级预防用药,使用脑动脉支架或其他血管成形术的极高危患者推荐联用氯吡格雷/阿司匹林治疗,并未专门针对轻型脑卒中和TIA作相关的指南推荐。轻型脑卒中和TIA患者早期发生脑卒中的风险很高,患者90天脑卒中风险为10%~20%,其中一半的病例发生在2天内。对于轻型脑卒中和TIA患者该如何有效地开展治疗,是目前医学界关注的一个重要问题。

2. 其他研究证据提示的临床问题思路 脑血管被认为是心血管在脑部的延续,很多脑血管病的治疗手段与药物均先在心血管病领域开展研究并得以应用,而后扩展到脑血管病领域。特别是轻型脑卒中和TIA被认为是更接近于冠心病的一种病理生理状态。

CURE试验表明在急性冠脉综合征人群中在服用阿司匹林同时使用氯吡格雷,3个月随访结束时联用氯吡格雷/阿司匹林组可降低脑卒中、心肌梗死、血管性死亡的风险达20%。联用氯吡格雷/阿司匹林组重大出血事件的风险略有增加,但致命性出血事件的风险与对照组并无差异。在急性冠脉综合征人群中实施的试验结果提示在轻型脑卒中和TIA人群中联用氯吡格雷/阿司匹林可能会有类似的效果。

3. 基础研究证据 查阅文献,未发现基于联用氯吡格雷/阿司匹林治疗脑卒中或TIA的动物实验研究证据报道。

4. 已有的临床研究证据 MATCH试验亚组分析提示,那些在确诊脑卒中或TIA后立即接受联用氯吡格雷/阿司匹林治疗的患者有获益增加的趋势。FASTER试验结果提示在TIA或轻



型脑卒中发生后 24 小时内联用氯吡格雷 / 阿司匹林较单独使用阿司匹林患者 90 天脑卒中风险有下降的趋势。这些已有的研究提示 TIA 和轻型脑卒中患者具有较高的脑卒中风险,并且表明联用氯吡格雷 / 阿司匹林可能存在更大收益。

## 二、临床问题转换成科学问题

针对上述临床问题,我们可以通过设计不同的临床研究来回答,这就需要将临床问题转换为科学问题。

转换思路 1: 抗血小板药物与轻型脑卒中或 TIA 患者预后的关系,我们可以转换为这样的科学问题:原先为轻型脑卒中或 TIA 而一定时间后(如 3 个月)发生脑卒中(包括缺血性或出血性脑卒中)复发的人群中联用氯吡格雷 / 阿司匹林治疗的比例是否低于原先为轻型脑卒中或 TIA 而一定时间后(如 3 个月)未发生脑卒中复发的人群?这就需要设计一个病例对照研究,通过追溯原先为轻型脑卒中或 TIA 而一定时间后(如 3 个月)发生脑卒中复发的人群,和原先为轻型脑卒中或 TIA 而一定时间后(如 3 个月)未发生脑卒中复发的人群,在发生轻型脑卒中或 TIA 后的抗血小板药物情况,比较两组人群联用氯吡格雷 / 阿司匹林治疗的比例。

这样的研究是属于花费少、耗时短的病例对照研究,属于观察性研究,但我们的临床问题针对的疾病状况是轻型脑卒中或 TIA,这就要求我们选择的研究人群应为一定时间前(如 3 个月)为轻型脑卒中或 TIA 的患者。由于轻型脑卒中或 TIA 患者经常因为病情较轻而未就诊,或 3 个月内已发生多次复发,使得研究的人群划分与界定会有一定的困难。患者在得了轻型脑卒中或 TIA 后应用抗血小板药物治疗的时间、类型和剂量都可能是多式多样的,而且,在患轻型脑卒中或 TIA 后可能变换抗血小板药物治疗的类型或剂量,如前几天单独应用阿司匹林治疗,因病情加重而改用联用氯吡格雷 / 阿司匹林治疗方案,造成暴露因素分组的困难。病例对照研究较多地应用于病因危险因素的研究。

### 知识点

#### 将临床问题转化为科学问题:PICO

1. 患者和(或)疾病状况(population, patient and/or problem) 如轻型脑卒中或 TIA。
2. 治疗干预(intervention) 如联用氯吡格雷 / 阿司匹林治疗。
3. 比较(comparison) 联用氯吡格雷 / 阿司匹林治疗组与单用阿司匹林治疗组。
4. 临床的结局(outcome),有时候(如病因、生存分析等)时间因素也应考虑在内 脑卒中(包括缺血性或出血性脑卒中)复发,发生时间如 3 个月。

转换思路 2: 抗血小板药物与轻型脑卒中或 TIA 患者预后的关系,我们可以转换为这样的科学问题:联用氯吡格雷 / 阿司匹林治疗的轻型脑卒中或 TIA 的人群,在经过几个月或若干年之后,脑卒中(包括缺血性或出血性脑卒中)复发率是否低于单独应用阿司匹林治疗的人群,复发率降低多少倍?这就需要前瞻性登记队列研究设计,通过记录患者的抗血小板药物情况,将人群分为联用氯吡格雷 / 阿司匹林治疗组和单用阿司匹林治疗组,观察几个月或若干年,分别检查两组人群的脑卒中复发率,比较两组人群脑卒中复发率的高低。

这样的研究尽管属于前瞻性研究,结论会非常有利于回答临床问题,但真实世界中往往病情较重的患者会采用联用氯吡格雷 / 阿司匹林治疗,而病情较轻患者会采用单用阿司匹林治疗,造成偏倚。由于这样的研究属于观察性研究,研究者无法对患者应用抗血小板药物治疗的时间、类型和剂量进行规定和干预。同样,在项目实施过程中患者可能变换抗血小板药物治疗的类型或剂量,如前几天单独应用阿司匹林治疗,因病情加重而改用联用氯吡格雷 / 阿司匹林治疗方案,造成研究人群分组分析的困难,较难分析联用氯吡格雷 / 阿司匹林治疗的净效益。



转换思路3:我们也可以转换为这样的问题:轻型脑卒中或TIA的人群,假如在发病早期(24小时之内)联用氯吡格雷/阿司匹林治疗一定时间(如3个月)之后,发生脑卒中(包括缺血性或出血性脑卒中)复发的概率,与单用阿司匹林治疗相比,是否下降,下降多少倍?这就需要随机对照临床试验设计,对于轻型脑卒中或TIA的人群,随机化分为两组,一组接受联用氯吡格雷/阿司匹林治疗,一组单用阿司匹林治疗,经过一定时间(如3个月)随访,评估两组的脑卒中(包括缺血性或出血性脑卒中)复发率,并进行比较。

随机对照试验(randomized controlled trial, RCT)的基本原理是将诊断为患有所研究疾病的同类患者随机分为两组,一组为试验组,给予待评价的新药或新疗法,另一组为对照组,给予常规治疗措施或安慰剂对照治疗,通过一定时间的随访,观察两组的结局(治愈率、有效率、复发率、病死率和存活率等)。经统计学检验,如果试验组结局优于对照组,则可认为待评价的新药或新疗法优于常规治疗措施或安慰剂治疗;如果两组结局没有差别,则可认为新药或新疗法与常规治疗措施或安慰剂治疗的疗效相同;如果对照组结局优于试验组,则可认为待评价的新药或新疗法差于常规治疗措施或安慰剂治疗。

### 知识点

#### RCT 的基本概念

RCT系实验性研究的一种,是以患者(包括住院和未住院患者)为研究对象,以个体为单位进行随机化分组,评价某种新药或新疗法对某种疾病的疗效。评价指标包括治愈率、有效率、复发率、病死率和存活率等。RCT常用于评价药物、治疗技术、疗法或其他医疗服务的效果和安全性评价。

### 知识点

#### RCT 与队列研究比较

与队列研究相比(图8-1),RCT的特点在于:研究的干预措施由研究者人为控制,研究对象的分组是随机的,能较好地排除混杂因素的干扰作用,检验能力较强。

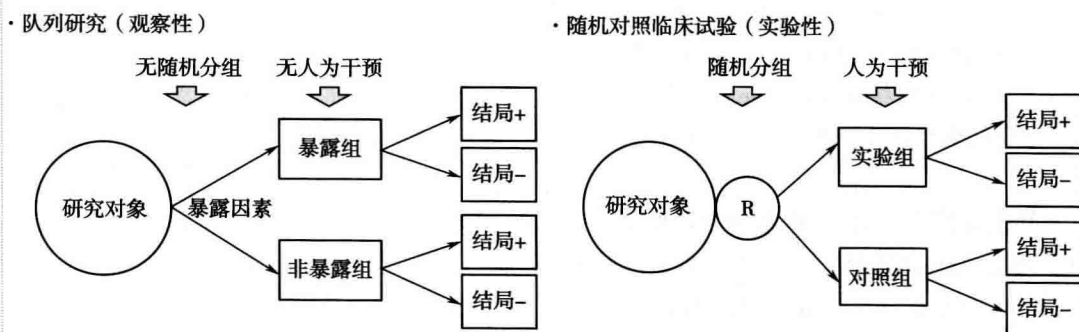


图 8-1 RCT 与队列研究设计比较

## 第二节 随机对照临床试验设计

### 一、研究目的

通过随机对照临床试验评估联用氯吡格雷/阿司匹林治疗轻型脑卒中或TIA的疗效。



## 二、研究假设

与单用阿司匹林治疗比较,轻型脑卒中或 TIA 的人群在发病早期(24 小时之内)联用氯吡格雷 / 阿司匹林治疗 3 个月脑卒中(包括缺血性或出血性脑卒中)复发率较低。

## 三、病例的选择

病例来源:①病例来源:因本研究需采集住院治疗相关信息,拟采用住院病例作为研究对象。研究拟在 114 家医院开展,为多中心研究,需制订相关标准操作规范(SOP),加强培训,确保各中心数据收集方式等方面规范、标准、一致,保证研究的质量;②诊断明确的轻型脑卒中或 TIA 患者:通过前面的科学问题提炼,该研究将研究对象锁定在轻型脑卒中和 TIA。首先应对纳入研究的患者严格按照国内外学术界统一、公认的诊断标准,明确诊断轻型脑卒中或 TIA 患者;③具体入选标准和排除标准确定:在明确研究对象的诊断之后,还应适合本研究的人选标准和排除标准做出具体规定,如只包含中高危脑卒中风险(随机化时 ABCD<sup>2</sup> 评分 $\geq 4$  分)的 TIA 患者,排除具有明确的抗凝治疗指征(怀疑存在心源性栓塞,如房颤、已知的人工心脏瓣膜、可疑的心内膜炎等)及严重的肝功能不全或肾功能不全的患者等。

## 四、对照的选择

对照(control)选择:设置对照组的目的是为了比较,我们想要比较联用氯吡格雷 / 阿司匹林治疗效果,就应选取另一组患者不采用联合氯吡格雷 / 阿司匹林治疗。对照组与试验组在其他方面应该是均衡一致的,唯一的不同是采取了与试验组不同的干预措施。对照组所采取的治疗措施可以是安慰剂治疗,也可以是单用阿司匹林治疗,还可以是单用氯吡格雷治疗。对照组采用的药品剂型、剂量和治疗时间又是如何? 对照组选择不同的治疗措施,可能使研究得到的结果也不同,研究结果所能说明的临床问题也不一样。合理选择对照组是影响试验成败的一个重要因素。

对照选择 1:假设我们采用安慰剂治疗作为对照,需选取一组患者给予以淀粉等无药效成分制成、但外形、颜色和味道与试验药物极为相近的安慰剂,观察并比较两组的疗效差异。因为目前临床已有被证明用于缺血性脑卒中和 TIA 的有效的抗血小板药,如仍用没有药效成分的安慰剂作对照就会涉及伦理问题,是违反伦理原则的,不可取。

对照选择 2:假设我们单用氯吡格雷治疗作为对照,需选取一组患者给予一定剂量的氯吡格雷,观察并比较两组的疗效差异。目前指南推荐和临床实践中氯吡格雷用于伴有动脉粥样硬化性动脉狭窄或其他重要危险因素(如糖尿病、冠心病等)高危或中高危的缺血性脑卒中患者,而我们的研究关注的是轻型脑卒中和 TIA,危险程度相对偏低,所以如采用单用氯吡格雷治疗作为对照,与我们的目标人群不符,而且得到的结论是“用氯吡格雷的患者如加用阿司匹林效果如何”,与我们的临床实践应用相脱节。

对照选择 3:假设我们单用阿司匹林治疗作为对照,需选取一组患者给予一定剂量的阿司匹林,观察并比较两组的疗效差异。因为用阿司匹林治疗是目前指南推荐用于低中危缺血性脑卒中和 TIA 的抗血小板药,也是目前临床实践的一线抗血小板药。选用单用阿司匹林治疗作为对照不仅符合伦理,而且与我们的临床实践相结合,得到的结论可直接应用于临床实践。在确定好采用单用阿司匹林作为对照后,下一步是要明确阿司匹林的剂型、用法、用量。结合目前临床上阿司匹林的用法,并考虑到可能的出血风险,我们选用阿司匹林 75mg/d,剂型是每片 25mg 或 50mg。





### 知识点

#### 对照的设置

实验研究设计的一个重要原则就是必须有对照。临床试验的目的就是评价某一新的干预措施(新药物、新疗法等)针对某一疾病状况的效果和安全性。这种新的干预措施的效果是根据比较组间效应差别来判定的,如果不设置对照,就得不到效应差值,也就无法评价新干预措施的疗效。

### 知识点

#### 对照的作用

设置对照可以抵消以下因素对效果判定的影响:

1. 抵消疾病自愈趋势的影响 某些疾病常有自愈的趋势,一些急性自限性疾病,如上呼吸道感染、胃肠炎等,即使医生未施加任何治疗,疾病本身也会自行好转。如果不设置对照,就很难区分疾病的好转或痊愈是由于疾病的自然康复还是治疗措施的效果。

2. 抵消安慰剂效应 安慰剂效应是指某些研究对象由于依赖医药而表现的一种正向心理效应。当患者接受了治疗措施,即使该治疗是无药效成分的(如淀粉、生理盐水等),因使其在心理上、精神上得到了安慰,而使所患疾病得到了改善,这种改善与其正在接受的治疗措施无关。

3. 抵消影响疾病预后的其他因素的干扰 疾病的预后除了受医生所给予的干预措施的影响,还受其他很多方面的影响,包括患者的基本情况(如患者的年龄、性别),试验开始时的疾病轻重和病程,患者本身所具有的疾病危险因素(如高血压、糖尿病等是影响脑卒中预后的危险因素),以及其他附加治疗措施(如护理措施、心理治疗等)的影响。通过合理的设置对照,可以均衡两组间这些干扰因素,自然也抵消了这些因素的干扰作用,显示出所评价干预措施的真实效果。

### 知识点

#### 对照如何选择

1. 安慰剂对照 安慰剂(placebo)通常采用乳糖、淀粉、生理盐水等成分制成,不加任何有效成分,但外形、包装、颜色、大小、味道与试验药物极其相近。在所研究的疾病在临床上尚无有效的防治药物,或某些慢性病自然病程反复波动,短期不治疗不至于明显影响疾病的预后时才使用安慰剂对照。

2. 标准治疗对照 当临床上已有明确的有效的治疗方法时,应采用常规或当前最佳的治疗措施作为对照。这种情况下采用标准治疗作为对照也比较符合伦理的要求,而且试验结果更贴近于临床,得到的结论有助于改善临床实践。

3. 空白对照 对照组不给予任何对照干预措施。空白对照一般仅用于尚无有效疗法又不便实施盲法的探索性措施疗效评价中使用。如试验的目的是评价一些预防性干预措施,如疫苗的效果,常采用空白对照。

## 五、样本量的计算

样本量计算:最好请专业人员帮助,通过公式法、查表法和估计法。但估计样本含量之前,



必须明确一些参数:①对照组结局事件的发生率 $p_1$ ,本例中为单用阿司匹林组3个月后脑卒中复发率。通过查阅文献及结合临床经验可知,在TIA高危患者(ABCD<sup>2</sup>评分>4)或轻型脑卒中患者症状出现24小时内给予阿司匹林治疗,90天脑卒中复发率约为14%;②试验组结局事件的发生率 $p_2$ ,指所评价的干预措施组的预计结局事件发生率,本例中为联用氯吡格雷/阿司匹林治疗90天脑卒中复发率。根据对现有相关研究的文献分析,我们认为TIA高危患者(ABCD<sup>2</sup>评分>4)或轻型脑卒中患者症状出现24小时内给予联用氯吡格雷/阿司匹林治疗,90天脑卒中复发的相对危险度降低将大于22%,即试验组的90天脑卒中复发率约为11%;③ $\alpha$ 值,即所希望达到的检验显著性水平,也就是第一类错误(假阳性错误)的概率,一般取0.05或0.01;④把握度 $1-\beta$ , $\beta$ 为第二类错误(假阴性错误)的概率, $1-\beta$ 即把握度,一般取0.90,也就是 $\beta$ 值取0.10。根据临床试验样本量计算公式计算得:

$$p = (0.14 + 0.11) / 2 = 0.125$$

$$N_{\text{每组}} = \frac{[z_{1-\alpha/2} \sqrt{2p(1-p)} + z_{1-\beta} \sqrt{p_1(1-p_1) + p_2(1-p_2)}]^2}{(p_1 - p_2)^2}$$

$$= \frac{[1.96 \times \sqrt{2 \times 0.125 \times (1-0.125)} + 1.28 \times \sqrt{0.14 \times (1-0.14) + 0.11 \times (1-0.11)}]^2}{(0.14 - 0.11)^2}$$

$$= 2550$$

根据计算,本研究案例估计样本量试验组与对照组各约2550例患者。考虑患者在随访过程中可能失访,故增加10%的样本量,得本研究需试验组与对照组各2805例。

## 六、随机化分组

本案例的目的是评价与单用阿司匹林治疗比较,轻型脑卒中或TIA的人群在发病早期(24小时之内)联用氯吡格雷/阿司匹林治疗的疗效。为更有效地比较联用氯吡格雷/阿司匹林治疗措施的真实效果,就要求其对照组除了治疗干预措施(试验组联用氯吡格雷/阿司匹林治疗,对照组单用阿司匹林治疗)不同外,其他因素(患者的基本情况、病情及其他附加治疗措施)都尽可能与对照组是均衡一致的。采用将参加研究的患者随机分组的方法能有效地帮助我们达到这个目的。由计算机系统自动生成随机顺序表,按1:1的概率比例分配治疗药物,将5610例病例,随机化(randomization)分为2805例为联用氯吡格雷/阿司匹林,另2805例为采用单用阿司匹林,药物将按照随机码表进行包装。受试者将按照入组的先后顺序,由小到大依次获得指定的随机试验药物编号,研究者将给予患者与号码相应的药物。

### 知识点

#### 随机化分组

影响预后的非研究因素在两组间均衡可比是准确估计干预措施效果大小的前提。随机分组是获得组间可比性的最可靠的方法。随机化是将研究对象随机分配到试验组和对照组,使每个研究对象都有同等的机会被分配到各组中去,使各种已知和未知的混杂因素在组间分布均衡,从而提高各组间的可比性,减少偏倚,增加试验结果的准确性,使研究结论更加可靠。随机化分组是随机对照临床试验的又一重要基本原则。常用的随机化分组方法有以下三种:

1. 简单随机分组 先将病例以个人为单位编号,如按入院顺序号或就诊序号,再利用掷硬币(正反面分别代表试验组和对照组)、抽签或使用随机数字表等方法,随机将病例分



配到试验组和对照组中去。随机表是试验前预先设置好的,一旦患者进入临床,只要符合纳入研究条件,就应按随机化分组方案接受治疗,不得随意更改。

2. 分层随机分组 当发现某混杂因素(如病情、年龄等)对疗效影响较大时,可根据混杂因素的不同类型将病例先分成若干层,然后在每一层内再按简单随机化分组的方法将患者分配至试验组和对照组。分层随机化分组可增加组间的均衡性,提高试验效率。

3. 区组随机分组 克服简单随机化带来的两组间样本量明显不等。可以与分层随机化结合使用。

### 知识点

#### 随机化的过程

随机化过程的关键是使每个研究对象都有同等的机会被分配到各组中去。一般采用随机数字表等方法求得随机分组的顺序。那种按照患者出生日期(奇偶年份)、患者参与试验的时间(单双日)、患者参与试验的顺序号按单双号交替分配等方法进行分组的方法无法使得受试者有相同的机会进入不同的研究,不是严格意义上的随机分组,是对随机化的误用。

### 知识点

#### 随机化的实施

在实施随机化分组的过程中,应做到参与研究的研究者和研究对象等均不知道随机分配的顺序,称之为随机分配隐藏(concealment)。随机分配的隐藏与随机化分配本身同等重要,如果随机分配的顺序泄露,则达不到控制偏倚的目的。随机分配的隐藏一般采用不透光的密封信封或中心随机化系统来实现。

## 七、盲 法

在确定随机分组方案后,下一步是具体给予研究对象进行药物干预。一种最简单的方法是根据随机分组方案分别给予受试对象相关分组的药物。如按照随机分组方案,对于对照组的患者给予阿司匹林药片,对于试验组的患者给予氯吡格雷和阿司匹林两种药片。但如果这样做的话,因为受试对象知道自己吃的是阿司匹林药片,还是氯吡格雷和阿司匹林两种药片,就会产生主观心理作用,影响试验的依从性或倾向性地报道一些病情相关信息。研究者也会因知道某患者是用的何种用药方案,而在收集病史、观察患者反应时产生倾向性,如对试验组,即联用氯吡格雷/阿司匹林组,更为关注,报道更为详细。这两方面都会使试验产生偏倚,而影响试验结果的真实性。另一种做法是采用盲法(blinding),利用双模拟技术,配制两种肉眼难以辨别的(其大小、形状、颜色、外观完全相同)氯吡格雷片剂(75mg 有效的氯吡格雷和氯吡格雷安慰剂)。对于试验组,给予有效的氯吡格雷药片和有效的阿司匹林药片,而对于对照组,给予氯吡格雷安慰剂和有效的阿司匹林药片。这种情况下,因研究对象和研究者均不知道具体患者是何种用药方案,使试验实施过程尽可能做到公正,避免信息偏倚,增加了试验结果的真实性。

对照、随机化分组和盲法是随机对照临床试验的三大基本原则。临床试验过程中的很多资料都是通过询问病史、观察患者的反应而收集到,如果研究对象或研究者事先知道了某患者的



具体治疗分组情况,就产生主观心理作用,对不同分组的患者产生不同的倾向性,而产生信息偏倚,影响试验结果的真实性。在数据整理和统计分析阶段,如果统计分析人员知道具体患者的分组情况,也会在数据整理、统计方法选择和结果解释方面产生倾向性,而影响实验结果的真实性。为避免这些偏倚,可采用盲法,使研究对象、研究者和(或)统计分析人员不知道具体的患者分组情况。

### 知识点

#### 盲法种类

1. 非盲法 又称开放试验,在研究过程中,研究者和研究对象都知道治疗的具体内容。这种方法优点是容易实施,但缺点是容易产生偏倚。但对于一些试验却只能是非盲的,比如手术治疗与药物治疗对某种疾病的疗效。

2. 单盲 研究对象不知道自己是试验组还是对照组。这种方法可避免来自研究对象的偏倚,但仍不能避免来自研究者的偏倚。

3. 双盲 研究者和研究对象都不了解试验分组情况,而是由研究设计者来和控制全部试验过程。双盲设计可以避免研究对象和研究者的主观因素所带来的偏倚。缺点是实施相对较复杂,难度大。

4. 三盲 不仅研究者和研究对象不了解分组情况,而且负责资料整理和分析的人员也不了解分组情况。三盲设计不仅可避免来自研究对象和研究者的偏倚,还可避免来自资料整理和分析人员主观倾向性所带来的偏倚。

## 八、结局测量

本研究案例中主要疗效指标为3个月脑卒中(包括缺血性或出血性脑卒中)复发。顾名思义,3个月脑卒中的复发情况应是在首次脑卒中发病后90天时的脑卒中复发情况。但是由于各种原因,很难正好在第90天当天能随访到患者,一般会设定一个时间窗(如 $90\text{天} \pm 7\text{天}$ ),只要在这个时间窗内完成随访都认为是可以的。在确定了结果观察指标及何时测量后,下一步是确定如何测量。对于脑卒中患者,可以事先与患者预约好,在第90天时来医院复查的同时进行医生与患者面对面的随访。另一种方法是事先记录患者电话,在第90天时由研究人员向患者打电话进行电话随访。对试验组和对照组在结果测量方法和终点判定上,做到公平一致。最后确定患者是否有脑卒中复发时,需经由脑血管专业领域专家组成的终点事件判定委员会在盲法状态下根据随访结果做出公平公正的判定。

### 知识点

#### 结局的确定

结局的确定与测量是临床试验成功的关键之一。临床试验中干预措施产生的结局可能有很多方面,而一项临床试验不可能测量所有相关的结局,一般会挑选一项关键的结局终点事件作为主要结局指标,而其他相关的指标作为次要结局指标。研究结局的选择需依具体研究的目标和研究角度来定,尽可能选择灵敏度高、特异度高的明确的终点事件(存活、病死、复发等)作为试验结果测量指标。除非研究有特殊需要,一般不采用生化检查结果等“中间结局”指标作为临床试验结果测量指标。



## 知识点

## 结局的测量

如果结局测量者知道试验的分组情况,就会在结果判定上产生倾向性,如倾向于将试验组的结果判定为阳性。为避免在结果测量过程阶段产生偏倚,和干预措施及其他指标的测量一样,结局指标的观察、测量和评价也尽可能采用盲法的形式,即结局测量者不知道患者的试验分组情况。无论是试验组还是对照组,在结果测量方法、终点判定上都应该特定一致的,且要使用相同的方法。如果有条件,应由建立在第三方基础上的终点事件判定委员会在盲法状态下对试验结果做出公平公正的判定。

## 知识点

## 结局事件出现的时间

从患者发病到结局事件出现的时间也是反映干预措施与结局之间联系的重要变量。在结局事件明确时,如能同时收集结局事件发生的时间将使结局事件信息更为充分,在统计分析时可以用生存分析的方法进行分析。

## 九、预期结果

与单用阿司匹林治疗比较,轻型脑卒中或TIA的人群在发病早期(24小时之内)联用氯吡格雷/阿司匹林治疗3个月脑卒中(包括缺血性或出血性脑卒中)复发率较低,差异有统计学意义。

## 十、随机对照临床试验结果的“2×2”列表

在随机对照临床试验中研究结果以“2×2”表格表示,其中行代表试验分组情况,列代表试验结局情况(表8-1)。a代表试验组结局事件发生例数,c代表对照组结局事件发生例数,a+b与c+d分别代表试验组与对照组的总例数。比较试验组结局事件发生率 $a/(a+b)$ ,和对照组结局事件发生率 $c/(c+d)$ 。这两个率比较可以用自由度为1的卡方检验。试验组和对照组的均数或连续变量结局指标可以通过t检验或非正态变量的非参数检验来比较。如果试验记录了发病到结局事件出现的时间的信息,可采用Cox比例风险模型评价两组间治疗效果的差异,并计算风险比(hazard ratio,HR)。

表8-1 结果表达的2×2表

| 试验分组 | 结局终点事件 |     | 合计  |
|------|--------|-----|-----|
|      | 发生     | 未发生 |     |
| 试验组  | a      | b   | a+b |
| 对照组  | c      | d   | c+d |
| 合计   | a+c    | b+d | N   |

## 知识点

## 意向性治疗分析

意向性治疗分析(intention-to-treat analysis,ITT)是指所有经过随机化分组的患者,应按照所分配到的处理进行随访、评价和分析,而不管其是否依从研究计划的治疗过程。该原则下



不是按患者实际所接受的处理组进行分析,强调只要是参与了随机分配的病例,都应纳入并按初始随机化分组进行疗效分析。因为随机分配的原则是确保各组间的可比性,如果将失访或退出的病例排除,只对符合试验方案者进行分析,就会破坏随机分配时设定的组间均衡性,甚至夸大试验治疗的效果。意向性治疗分析确保了随机化的原则,可使组间保持均衡性,减少可能的偏倚带来的影响,使结果更加真实,下结论趋于保守。对于缺失值的处理,一般采用将最近一次观察结果结转(last observation carried forward, LOCF)到终点的处理方法。

### 知识点

#### 绝对危险度减少

绝对危险度减少(absolute risk reduction, ARR)表示试验组事件发生率与对照组事件发生率之间的绝对差值。该值大小,表示相对于对照组,经试验组治疗后事件发生率的绝对下降水平。 $ARR = c/(c+d) - a/(a+b)$ 。举个例子,如一新药治疗脑卒中3个月的脑卒中复发率为8%,对照组治疗的脑卒中复发率为12%,则 $ARR = 4\%$ ,表示相对于对照治疗方法,新药治疗可使脑卒中的3个月复发率绝对下降4%。

### 知识点

#### 相对危险度减少

相对危险度减少(relative risk reduction, RRR)表示绝对危险降低率占对照组事件发生率的比例。该值大小,表示试验组比对照组治疗后事件发生的相对危险度下降的水平。 $RRR = [c/(c+d) - a/(a+b)] / [c/(c+d)]$ 。举个例子,如一新药治疗脑卒中3个月的脑卒中复发率为8%,对照组治疗的脑卒中复发率为12%,则 $RRR = 33\%$ ,表示相对于对照治疗方法,新药治疗可使脑卒中的3个月复发的相对危险度下降33%。

### 知识点

#### 需治疗人数

需治疗人数(number needed to treat, NNT)表示为避免1例临床事件发生,临床医师在一段时间内应用某一疗法需治疗的患者例数。

$NNT = 1 / [c/(c+d) - a/(a+b)]$ 。该指标是治疗性临床试验中反映治疗效果的良好量化指标,直观易懂,可操作性强,更贴近医生和患者的实际思维,同时具有临床价值以及经济价值。举个例子,如一新药治疗脑卒中3个月的脑卒中复发率为8%,对照组治疗的脑卒中复发率为12%,则 $NNT = 25$ ,表示为预防1例脑卒中复发,需治疗25例。

## 第三节 研究报告与问题分析

随机对照临床试验的报告格式和要求可以参考随机对照试验报告规范(CONSORT声明)。研究结果的报告包括研究背景与目的、方法、结果、讨论与结论四个方面。讨论对研究设计、可能存在的偏倚、可能影响结论的因素以及试验结果的临床应用等进行恰当的分析。

研究案例:氯吡格雷用于急性非致残性脑血管事件高危人群的疗效的随机对照临床研究





[ N Engl J Med, 2013, 369(1): 11-19 ]。

1. 背景与目的 短暂性脑缺血发作(TIA)或轻型脑卒中早期(最初几周内)复发风险极高。氯吡格雷与阿司匹林合用比单用阿司匹林在预防卒中的发生可能具有更好的效果。

2. 方法 本研究设计为全国多中心(114家中心)、随机、双盲、对照试验,5170例轻型脑卒中或TIA患者在发病后的24小时内被随机分配到氯吡格雷-阿司匹林联用组(氯吡格雷起始剂量300mg,随后75mg/d直至90天,在合并前21天使用阿司匹林75mg/d)和安慰剂-阿司匹林对照组(阿司匹林75mg/d,直至90天)。所有试验者在第一天服用公开标签的阿司匹林75~300mg。主要结局是基于意向性治疗分析的90天脑卒中事件(缺血或出血性);应用Cox比例风险模型评价治疗效果的差异,并考虑研究中心聚集的随机化效应。

3. 结果 氯吡格雷-阿司匹林组的8.2%患者发生脑卒中,而阿司匹林组的11.7%患者发生脑卒中(HR:0.68,95%CI:0.57~0.81, $P<0.001$ )。氯吡格雷-阿司匹林组7个(0.3%)患者发生中重度出血,而阿司匹林组8个(0.3%)患者发生中重度出血( $P=0.73$ );每组出血性脑卒中的发生率均为0.3%。

4. 结论 在发病24小时内给予治疗的TIA或轻型脑卒中患者中,氯吡格雷与阿司匹林联合治疗比单独应用阿司匹林更能降低90天脑卒中发生风险,且并没有增加出血风险。

### 知识点

#### 常见的偏倚

1. 霍桑效应 霍桑效应(Hawthorne effect)是指在研究过程中,研究者对自己感兴趣的试验组的研究对象较对照组更为关心,而被关照的研究对象由此产生心理变化,对研究者报以过分的热情,并改变了他们的行为,更多地向研究者报告好的结果,从而夸大了治疗效果。

2. 干扰与污染 干扰(intervention)是指试验组除接受研究措施以外,额外地接受了类似试验药物效果的附加措施,从而人为地夸大试验组的真实疗效。污染(contamination)是指对照组额外地接受了试验组的药物,从而人为地造成一种夸大对照组疗效的现象。

在临床试验实施过程可能发生其他的重要偏倚,包括向均数回归、安慰剂效应和失访偏倚等。

### 知识点

#### 偏倚的控制

控制霍桑效应、干扰和污染的最好方法是严格按照盲法的原则实施治疗方案,在实施过程中不轻易改变患者的治疗方案或改变药物的种类或剂量。

## 第四节 临床问题的回答

继随机对照临床试验显示联用氯吡格雷/阿司匹林治疗轻型脑卒中或TIA有效之后,更多关于轻型脑卒中与TIA防治相关的研究设计正在国内外开展。如美国一项正在进行的随机对照试验(POINT试验)旨在评估轻型脑卒中与TIA发病后12小时内联用氯吡格雷/阿司匹林强化治疗相对于单用阿司匹林治疗在预防90天脑卒中复发的疗效及安全性的差异[Int J Stroke, 2013, 8(6): 479-483]。



## 第五节 随机对照临床试验的其他应用

随机对照临床试验除用于药物、疗法或其他治疗性措施的效果外,还可用于针对疾病危险因素进行干预的病因学研究,评价两种以上检查方法真实性、可靠性和实用性的筛检研究,评价预后影响因素的预后研究等。

### 思考题

1. 什么是随机对照临床试验? 试验设计的基本原则有哪些? 具体如何实施?
2. 收集脑卒中治疗相关的不同随机对照临床试验文献,回答:该研究期望解决的临床问题是什么? 科学问题是如何描述的? 该研究采用了何种临床试验设计? 是否采用了随机分组? 是否采用了盲法? 盲法实施是否正确? 你认为存在哪些偏倚,如何通过改进设计等方法控制偏倚?
3. 结合自己专业,选择一种发病率较高的疾病,针对可能的有效治疗措施,设计一项随机对照临床试验,内容包括:科学问题与研究目的,研究设计与具体方法,对照设置,随机分组,盲法,结局测量方法及可能的结果。

(王伊龙)

## 第九章 成本效果分析

### 临床案例

男性,40岁,乙肝肝硬化病史10年,1周前因上消化道出血(呕血),胃镜检查证实为重度食管静脉曲张,经积极补充血容量、降低门脉压力、抑酸、止血等治疗后出血停止,目前状态稳定。肝功能 Child-pugh B级,为预防肝硬化门脉高压食管胃静脉曲张破裂再出血,临床医师为患者选择了内镜治疗,并在之后的随访中再次接受了内镜治疗,随访1年未再出现消化道出血。目前继续随访中。

### 第一节 如何获得临床研究问题

这是一例常见的肝硬化门脉高压食管胃静脉曲张出血患者的处理。针对该患者和同类患者,仍然有很多争议问题,也是医学研究的问题来源。例如:①该患者再出血的可能性有多大?是否需要采取措施预防再出血?②临床上目前有哪些措施可以用于预防再出血的处理?各自的优点和缺点是什么?③我们为选择不同的预防再出血的措施依据是什么?这样的选择对谁更有利?

针对这一常见病和多发病,我们可以进行很多相关的研究。如何从临床问题中获得有意义、具有探索性或者证据性研究课题,即形成科学问题?

1. 阅读综述文献了解相关研究背景资料 文献综述告诉我们,肝硬化门脉高压食管胃静脉曲张破裂出血是肝硬化严重并发症之一,直接影响患者的生活质量和生存时间,在这一过程中同时需要大量的费用,为患者或社会带来了巨大的经济负担。

肝硬化食管胃静脉曲张破裂出血患者1年内无干预情况下再出血的可能性在70%以上。需要采取有效措施预防再出血。

目前可以通过药物治疗(普萘洛尔)、内镜治疗(内镜下静脉曲张套扎治疗、硬化治疗和组织黏合剂注射治疗)、介入治疗(肝内门体分流TIPS术和经皮经肝门静脉穿刺胃冠状静脉栓塞术)、手术治疗(脾切除+断流术和门体分流术)达到预防再出血的目的。图9-1显示了文献报道的不同处理措施包括联合治疗预防再出血的不同疗效。无论哪一种措施的应用,都可以减少食管胃静脉曲张破裂再出血。干预治疗后,再出血的概率,从70%下降到10%~35%不等。

2. 阅读研究原著或者Meta分析了解临床疗效和安全性 阅读文献我们同时发现,分流手术疗效最好,但手术后发生肝性脑病的比例较高,并且需要患者一般情况以及肝功能状态比较好,能胜任手术。药物治疗,尽管方便服用,患者易于接受,但疗效较差,并且部分患者由于心率较慢或者存在哮喘等疾病不能耐受药物。

内镜治疗和介入治疗是临床医师为多数患者做出的选择。内镜治疗一年的再出血率在30%~35%,而TIPS治疗一年的再出血率小于20%。但TIPS治疗后肝性脑病的发生率较高(20%),TIPS治疗采用非覆膜支架一年后再狭窄的发生率较高。二者的死亡率相似。一次TIPS治疗的费用昂贵(相当于10~15次内镜治疗费用),而处理肝性脑病的费用也不便宜。那么在更低的再



出血率、肝性脑病的危险性、更高的费用之间,我们究竟为患者推荐哪一种治疗呢?或者如何权衡临床疗效、安全性、生命质量以及成本为患者选择预防肝硬化食管胃静脉曲张破裂再出血的预防措施?

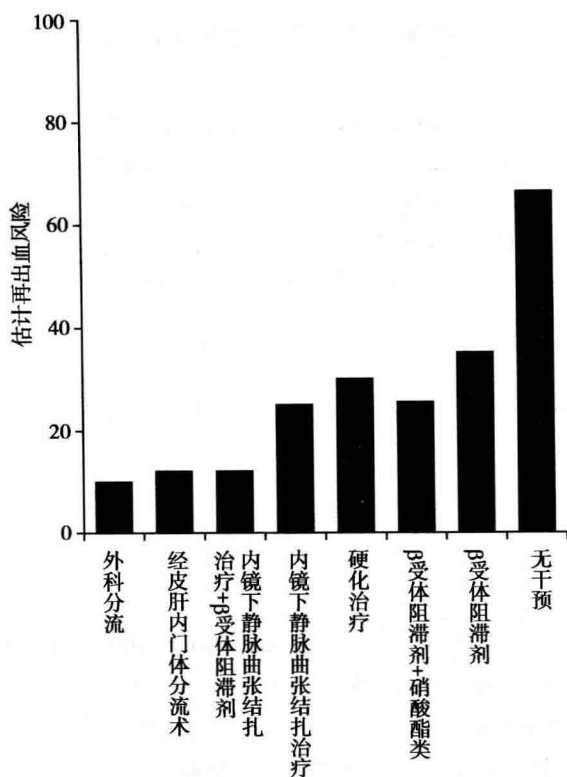


图 9-1 不同措施预防肝硬化食管胃静脉曲张破裂再出血疗效比较

我们可以从大量随机对照临床试验以及基于随机对照临床试验的 Meta 分析结果证实不同措施的疗效差异,同时也能获得不同措施不良事件或者安全性数据。但如何考虑成本?经济学分析为我们提供另一种解决问题的办法。

### 知识点

#### 经济学分析的基本概念

经济学分析是一种定量的分析方法,同时从资源的使用(成本)和临床结果(效果,包括有效性、危险性和生命质量)两方面比较不同措施。

3. 站在患者角度、群体与社会角度考虑临床问题 临床医生在为患者作出决定时,不仅要考虑到临床结果,即诊断或治疗措施的有效性和危险性,同时也要考虑到社会学效果(即患者的生命质量)和经济学效果(即卫生资源的消耗)。因为医生所面对的不仅仅是眼前的患者,日常有大量类似的患者,卫生资源的消耗是巨大的,而卫生资源又是有限的。

仅仅观察一种措施的疗效或者成本,只是一种描述性研究类型。或者描述了疗效,或者描述了成本,都只是部分评价了这一临床措施。即使同时评价这一措施的疗效和成本,也只是成本-效果的描述,没有比较,就不知道差别。

同时观察两种或者两种以上措施的疗效或者成本,尽管是一种分析性研究,但都是部分评价。两种措施的疗效评价,RCT 研究大都是这样一种类型,比较的是临床疗效和安全性。两种措施的成本比较,只有在疗效相同的情况下才会看出差别。因此只有同时比较两种或者两种以



上措施的疗效和成本,才是完整的经济学评价研究。

### 知识点

#### 临床经济评价完整性分析

| 是否同时检查了各种方案的成本与效果 |   |                    |                 |                                                            |
|-------------------|---|--------------------|-----------------|------------------------------------------------------------|
| 否                 |   |                    | 是               |                                                            |
|                   |   | 仅检查效果              | 仅检查成本           |                                                            |
| 是否对两种或两种以上的方案进行比较 | 否 | 1A 部分评价<br>效果描述    | 1B 部分评价<br>成本描述 | 2 部分评价<br>成本 - 效果描述                                        |
|                   | 是 | 3A 部分评价<br>效力或效果分析 | 3B 部分评价<br>成本分析 | 4 完整的经济评价<br>最小成本分析<br>成本 - 效果分析<br>成本 - 效用分析<br>成本 - 效益分析 |

### 知识点

#### 临床经济学分析的基本类型

1. 最小成本分析 最小成本分析(cost minimization analysis,CMA)也可称为成本最小化分析或成本确定分析(cost identification analysis)。测定不同医疗措施的成本并进行成本比较,假定这些措施的结果基本相同,成本低的措施经济效果好。

2. 成本 - 效果分析 成本 - 效果分析(cost-effectiveness analysis,CEA)是分析成本消耗后得到的效果。其表示方法为每一效果单位所耗费的成本(成本 - 效果比)或每一个增加的效果所需要耗费的增量成本(增量比)等。这就使两种不同的医疗措施,在进行比较选择时,有了相同的评价单位,从而为临床决策单位提供科学的依据。

3. 成本 - 效用分析 成本 - 效用分析(cost-utility analysis,CUA)是成本 - 效果分析的一种特殊形式,其结果的测定是以病残和病死为结果的综合指标,通常用质量调节生命年(quality adjusted life year,QALY)表示。将某项措施所能延长的生命年乘以效用值,就等于该措施实施后所能延长的经过调整的有生命质量的年数(QALY)。

4. 成本 - 效益分析 在比较完全不同的医疗措施时,由于所得结果截然不同,必须用一个共同的单位来比较,除了上述用 QALY 为单位外,将某一项目及医疗服务的所有成本和效果均用货币量为单位来表示,就是成本 - 效益分析(cost-benefit analysis,CBA)。

## 第二节 成本 - 效果分析研究设计

1. 确定临床经济学分析的出发点 进行临床经济学分析,首先需要明确经济分析的目的或者出发点。是单纯从患者的角度出发,还是从医疗费用的实施者(医院)、提供者或者全社会的角度出发。

2. 确定所要分析的项目和两种以上将要进行比较的措施 临床经济学分析的基本目的是从经济学角度,评估不同临床医疗措施或者健康服务的投入与产出,即不同措施的成本和结果。因此必须有两个或两个以上的备选方案进行比较,以供选择。

本案例中,第一种选择措施是内镜治疗,第二种选择措施是介入治疗(TIPs)。我们的目的是



需要回答哪一种措施更具有成本效果?或者,如果我们选择疗效更好的措施,需要增加花费的成本是否值得?而基本的选择措施是药物治疗或者不给予任何预防再出血的干预措施。由于不同治疗不仅临床疗效(再出血)存在差异,不良事件(如肝性脑病)、患者生活质量也存在差别,因此本案例中的问题选择经济学的研究类型采用成本效用分析更合理。

### 知识点

#### 成本效果分析的局限性

1. CEA 不能比较两种对病残或病死率有影响的不同措施。当比较两个完全不同的项目如新生儿注射乙肝疫苗预防乙型肝炎和治疗高血压预防脑卒中发生时,由于缺乏相同的单位,CEA 不能提供明确的经济学决策依据,解决的办法是应用成本效益分析。

2. CEA 只能用于在同一种疾病或条件下比较不同的干预措施,或结果都为延长生命年时,所采用的不同措施的比较。由于没有考虑到是否改善了生存质量,例如化疗对有些类型的恶性肿瘤患者可延长生命,但是降低了生存质量,此时就需要既能衡量数量(生命年),又能衡量质量(生存质量)的方法,这样就能更客观地比较以下两种措施:一种措施延长了生命,但没有损害生存质量,另一种措施比前一种措施延长生命时间更长,但降低了生存质量。成本-效用分析是一种特殊形式的成本效果分析。

3. 计算成本 第一,应该详细列出每一种方案中所有有关的成本项目。比如本案例研究中,除了内镜或者介入治疗的成本之外,住院费用、处理不良事件如肝性脑病的费用、随访费用等,每一项细目,都需要包括。第二,应该考虑花费成本的时间。需要注意这一时间应该与临床疗效的时间数据一致。如本案例研究中,可以选择1年疗效与成本的比较,当然也可以考虑2年或者更长时间的比较。

4. 各项措施实施后的效果和利益 包括增加生存率、挽回生命年、延长寿命、减少病残率、恢复正常健康、回到工作岗位、提高患者满意度以及由于治疗带来的不良作用及依从性带来的损失。在开始研究前需确定所实施的效果分析是有效的,最好是头对头临床随机对照试验,或者 Meta 分析数据,如果不是,应说明其有效程度。

### 知识点

#### 成本效果分析的特殊性

成本效果分析的计算包括成本的计算和效果的计算。成本和效果的计算有时很复杂,根据比较措施以及进一步处理的不同,可以引入决策分析模型包括决策树模型和 Markov 模型,分别计算投入成本和产出的临床结果如延长的生命年。

5. 对发生在将来的结果和成本作贴现计算 在进行某一项目投资时,一般的心理都希望早些取得效果,可以先享受由于该项目的实施带来的好处,而总想晚些付钱。因为随着物价上涨,钱的实际价值随着时间的推移是减少的,所以当某一医疗措施的实施需要数年完成时,为了准确地估计成本和效果,去除由于物价上涨因素而带来的影响,必须对发生在将来的成本和效果通过贴现(discounting)的方法,换算为目前的实用价值。贴现率一般相当于银行利率。

6. 经济学分析结果 根据不同经济分析类型,选择不同的表达结果。分析成本消耗后得到的效果。其表示方法为每一效果单位所耗费的成本(成本-效果比)或每一个增加的效果所需要耗费的增量成本(增量成本效果比)等。这就使两种不同的医疗措施,在进行比较选择时,有了相同的评价单位,从而为临床决策单位提供科学的依据。





## 知识点

## 经济学分析的结果评价指标

1. 成本效果比 成本效果比(cost/effectiveness, C/E)是 CEA 的一种表示方式,即每延长 1 个生命年、挽回 1 例死亡、诊断出 1 个新病例或提高 1 个结果单位所花的成本。通常 C/E 值越小,就越有经济效率,单一的 C/E 值是没有意义的,主要用于两个或两个以上项目的比较,并且是比较有相同结果单位的两个项目。

2. 增量分析 由于成本效果分析包含着对两种或两种以上的措施进行比较,因此成本-效果比还不能充分显示两者的相互关系,故常用增量分析(incremental analysis)来表示。增量分析计算一个项目比另一个项目多花费的成本,与该项目比另一个项目多得到的效果之比,称为增量成本效果比,能充分说明由于附加措施导致成本增加时,其相应增加的效果是多少以及是否值得推荐。

$$\frac{\text{新成本} - \text{旧成本}}{\text{新效果} - \text{旧效果}} = \frac{\text{增加的成本}}{\text{每一个增加的效果单位}} \quad \text{式 9-1}$$

$\Delta C$  表示两个方案成本之差,  $\Delta E$  为两个方案效果之差,  $\Delta C/\Delta E$  为增量成本效果比。

$$\frac{\Delta C}{\Delta E} = \frac{C_N - C_0}{E_N - E_0} \quad \text{式 9-2}$$

$C_N$  为新成本,  $C_0$  为旧成本,  $E_N$  为新效果,  $E_0$  为旧效果。

7. 敏感性分析 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,或者临床疗效结果本身存在一定的可变范围,故敏感性分析(sensitivity analysis)也是经济评价中的必要步骤,也是经济评价结果能否推广应用的依据。

8. 推广及应用价值 在前面分析基础上,得出结论,并加以说明,应在复习文献基础上,与其他同类研究结果加以比较,说明本结论的可推广性以及有关医德的问题,必要时作决策分析。

### 第三节 临床案例结果描述与临床选择

表 9-1 列出了 TIPS 治疗与内镜治疗预防食管静脉曲张破裂再出血成本效果评价结果。可以看到,按每人年计,再出血次数在硬化剂组、套扎组和 TIPS 组分别为 0.39、0.32、0.07。而每人年的费用则分别为 \$ 23 459、\$ 23 111、\$ 26 275。每预防一次再出血的费用在 TIPS 组优于其他组。与内镜治疗相比,TIPS 组疗效增加,但成本也增加了。选择 TIPS 治疗降低再出血率的代价必然是投入增加,包括介入治疗增加的手术费用投入、处理肝性脑病增加的投入。那么,我们需要知道,这种投入增加的程度能否接受。即每增加一次不出血机会,我们需要增加多少投入,是否值得。增量分析的结果显示,从较好的套扎治疗改为应用更好的 TIPS 治疗,增加一次不出血的机会,需要增加 \$ 12 660 元。也就是说,我们能否接受用 \$ 12 660 元换回一年中一次不出血。西方人群中(如美国), \$ 25 000 元是一阈值,依据是挽回一个生命年的可接受的代价是 \$ 50 000 元。

表 9-1 TIPS 治疗与内镜治疗预防食管静脉曲张破裂再出血经济分析

| 干预措施 | 出血次数 | 死亡数  | 肝性脑病次数 | 总成本      | 增量成本效果                                |
|------|------|------|--------|----------|---------------------------------------|
| 硬化剂  | 0.39 | 0.24 | 0.17   | \$23 459 |                                       |
| 套扎   | 0.32 | 0.21 | 0.17   | \$23 111 |                                       |
| TIPs | 0.07 | 0.20 | 0.26   | \$26 275 | Vs. 硬化剂: \$ 8803<br>Vs. 套扎: \$ 12 660 |

注:以上数值为每年每患者计



## 第四节 临床经济学文献评价标准

临床经济学分析的文献评价标准包括三方面内容,即经济分析的结果是否正确,结果是什么,结果是否适用于我的患者(表 9-2)。

表 9-2 临床经济学分析的文献评价标准

- |                       |
|-----------------------|
| 1. 结果是否正确             |
| ①是否提供了完整的经济分析         |
| ②是站在何人的立场上进行评价        |
| ③是否比较了所有相关的临床措施       |
| ④成本和效果的测量是否正确         |
| ⑤成本和效果资料是否进行增量分析      |
| ⑥是否进行了敏感性分析           |
| ⑦估计的成本和效果是否来源于干预人群    |
| 2. 结果是什么              |
| ①增量成本和效果是多少           |
| ②各亚组增量成本和效果有无不同       |
| ③允许变化的不确定结果是多少        |
| 3. 结果是否适用于我的患者        |
| ①治疗的收益? 益处? 是否超过成本或危害 |
| ②我的患者是否有相似的临床结果       |
| ③我的患者是否有相似的成本         |

### 思考题

1. 临床经济学评价的基本类型包括哪些?
2. 成本效果分析的结果有哪些表达的指标?
3. 经济学评价科学性评价包括哪些指标?

(陈世耀)

## 第十章 比较效果研究

### 临床案例

患者,男,农民,50岁,家庭年收入1万元,慢性乙型肝炎史20余年,HBsAg(+)/HBeAg(-)/HBeAb(+)/HBcAb(+),HBV-DNA 20 000IU/ml,ALT 100U/L,AST 60U/L,Tbil 60  $\mu$ mol/L,Dbil 30  $\mu$ mol/L。高血压病史25年,糖尿病史10年。以前从未使用过抗病毒药物。查体:巩膜轻度黄染,可见肝掌,心肺查体未见异常,腹部查体肝未触及,移动性浊音阴性,下肢无水肿。临床面临的问题是:该患者如何选择抗病毒治疗方案?

### 第一节 如何获得临床研究问题

乙肝病毒(hepatitis B virus, HBV)慢性感染是我国重要的公共卫生问题之一,目前仍有慢性HBV感染者9300万人,其中慢性乙型肝炎患者(以下简称慢乙肝)约2000万人。慢性HBV感染是导致我国肝硬化和肝癌发病和死亡的重要原因。慢性乙肝尚不能治愈,目前其治疗的总体目标是:最大限度地控制炎症活动,减轻肝细胞炎症坏死及肝纤维化,延缓和阻止疾病进展,减少肝脏功能失代偿、肝硬化、肝癌及其并发症的发生,从而改善生活质量和延长生存时间。其中抗病毒治疗是目前预防慢性乙型肝炎病情进展的最重要途径,其前提是患者能长期持续的抗病毒治疗,即保持良好的用药依从性。已有多种抗病毒药物用于治疗慢性乙型肝炎,主要包括两大类,核苷(酸)类药物和干扰素。其中核苷(酸)类药物有五种,阿德福韦酯(adeфовir dipivoxil, ADV)、拉米夫定(lamivudine, LAM)、替比夫定(telbivudine, Ldt)、恩替卡韦(entecavir, ETV)和替诺福韦酯(tenоfovir disoproxil fumarate, TDF);而干扰素则包括普通干扰素IFN和长效干扰素Peg-IFN。

本案例患者慢性乙型肝炎病史20余年,HBsAg(+)/HBeAg(-),HBV-DNA为20 000IU/ml > 2000IU/ml,ALT > 2 × 最高上限(upper limit of normal, ULN),伴发高血压,以前从未使用过抗病毒药物。依据我国2010年慢性乙型肝炎防治指南推荐,建议口服核苷(酸)类药物或使用干扰素。但是面临5种核苷(酸)类药物和2种干扰素这么多的备选方案,究竟首选推荐哪种药物是临床医师需要考虑的问题。

作为一个临床医生,我们可以通过学习慢性乙型肝炎的治疗指南来获取信息。亚太、欧洲和美国肝病协会等肝病协会均推荐ETV、TDF和Peg-IFN作为乙肝抗病毒治疗的一线药物,而我国最新的《2010年慢性乙型肝炎防治指南》提出“初始治疗时宜选用抗病毒作用强和耐药发生率低的药物”。但是上述的指南都是基于目前已经发表的Ⅲ期随机对照临床试验(randomized clinical trial, RCT)对于上述抗病毒药物的疗效(这里指的是效力, efficacy)以及低耐药性的比较确定的。但RCT的疗效距离真实临床的治疗效果经常有差异,其原因可能包括以下几方面:

(1) RCT通常严格限制研究对象的人选和排除标准。以评价核苷(酸)类药物治疗慢乙肝患者的Ⅲ期临床试验为例,通常入选标准包括:HBsAg阳性者,HBV-DNA  $\geq 10^5$  copies/ml; HBeAg阴性者,HBV-DNA  $\geq 10^4$  copies/ml; ALT值介于2~10倍ALT最高上限(upper limit of normal, ULN)



之间;肝组织病理提示为慢乙肝;同时排除合并艾滋病病毒、丙肝、丁肝感染,以及非代偿期肝硬化患者,孕妇以及其他并发症如糖尿病、高血压、肾功能不全等。但是在实际的临床工作中,就如本案例患者那样,对于一个伴发高血压和糖尿病的患者,上述抗病毒药物的效果如何尚不得知。

(2) RCT 通常采用标准化的治疗方案,严格控制患者治疗的依从性,随时解决受试者用药中出现的副作用,由此来保证治疗依从性。但在实际的临床工作中,医生每天要面对众多的患者,不能像 RCT 研究那样去关注每个患者的用药及副作用。

(3) 参加 RCT 的受试者的费用由赞助商赞助,患者不会因为费用的问题而影响用药。而现实临床工作中,药物的价格、患者的经济状况、医疗政策(包括医保政策的变化)都会影响到患者的药物选择和依从性。例如,LAM 虽然有较高的耐药性,但是由于其较低的成本在很多发展中国家仍然被作为一线药物使用。而本案例患者为农民,家庭年收入仅为 1 万元,无医保,是否能长期承担 ETV 这类一线药物的药价?如果不能承担长期的药物费用,则势必会影响到患者的治疗依从性,从而最终影响到患者的治疗效果。

因此,本案例慢乙肝患者虽然符合《我国 2010 年慢性乙型肝炎防治指南》进行抗病毒治疗的指征,但是由于其同时伴有高血压和糖尿病,目前的 RCT 没有提供抗病毒药物对于糖尿病和高血压患者的治疗效果。同时,该患者经济状况差,而不同的核苷(酸)类药物其药价相差巨大,如何权衡疗效与患者的经济承受能力也是需要患者和医生考虑的问题。上述这些我们在临床决策时需要考虑的内容都是基于“理想环境”的 RCT 所不能提供给的,我们只能用临床服务中积累的数据来回答我们临床中遇到的实际问题,也就是基于真实临床实践中收集的数据来比较现存的不同干预措施在治疗、护理、预防等方面的效果差异,即比较效果研究(comparative effectiveness research, CER)。

### 知识点

#### 比较效果研究定义

比较效果研究(comparative effectiveness research, CER)是在真实临床情景下,基于不同的数据库资源和测量方法,比较现存的不同干预措施和干预策略在预防、诊断、治疗和疾病健康方面的利与弊,最终回应患者、临床医生和其他决策者所表达的需求,从而协助患者、医生和决策者做出明智的决定,以提高个体和人群的医疗保健水平。

CER 是“以患者为中心的效果研究(patient-center effectiveness research),其最终产出是为了满足真实世界中患者、临床医生和其他决策者的决定需求,选择最适合于个体患者需求的最佳选择。CER 最常见的研究设计包括试验性研究方法[如 RCT、实用性对照研究(pragmatic controlled trial, PCT)、其他非随机对照实验]、观察性研究(如病例报告、横断面研究、病例对照研究及队列研究)及基于现有证据的系统综述等。

基于临床实际工作中积累的数据,由于其研究对象来源于日常真实临床实践,能更好地反映不同的诊断、治疗、预防策略、保健服务以及医疗政策对患者的影响,包括对干预或治疗后的长期效果。但其不足之处在,治疗或某种处理措施不仅仅是一种暴露因素,更是一种临床决策。此种临床决策会受到患者的症状、疾病严重程度、医疗政策、患者的社会经济状况等多种混杂因素的影响。因此很难保证不同治疗或处理组在处理因素以外的其他可能会影响到治疗效果的因素上的可比性。同时由于没有严格的质量控制标准,也很难保证针对每个个体的结局以及其他可能影响结局的资料收集的可比性。

我国临床案例居多,在既往的医疗实践中产生了大量的二手数据,如何利用这些二手数据开展比较效果研究,并从科学问题的提出、研究设计、研究对象的选择、暴露及结局资料的



收集以及统计分析方法等多方面进行控制,从而保证其结论的真实性是目前大家关注的热点之一。目前最常用于观察性研究的数据库包括病历数据库和管理数据库(如医保数据库),后者因为缺乏详细的临床数据而在使用过程中受到了一定的限制。本章将探讨如何以电子病历数据库来回答上述临床问题,包括如何确定科学问题、研究设计、分析以及如何控制偏倚等。

## 第二节 如何利用临床实际服务的二手数据开展 CER 研究

### 一、从临床问题到科学问题的转化

临床数据不是针对研究而产生的,面对海量的二手数据,我们需要先确定科学问题,并从现有的二手数据中提炼出能用于科学研究的数据。对于本案例患者,考虑到患者的病情、药物治疗的副作用、经济承受能力、药物的价格等,假设你首先考虑的是核苷(酸)类药物。截至 2013 年年底,在我国上市的核苷酸类药物主要有四种,ETV、ADV、LAM 和 Ldt,其中虽然 ETV 被推荐为一线抗病毒药物,但是目前我国初治慢乙肝患者中使用最多的仍然是 ADV、其次为 LAM,ETV 只排列第三,其中除了因为 ADV 和 LAM 早于 ETV 上市,很多患者已经习惯服用这两种药物之外,一个重要的原因是价格的问题。权衡患者的经济承受能力和疗效,想要在 ADV 和 ETV 之间做出抉择。虽然大量 RCT 已经证明了 ETV 和 ADV 治疗慢性乙型肝炎患者的效力,但在实际的临床实践中,患者的病情、使用药物的依从性、患者的经济能力(包括医保的类型以及药物报销比例等)都可能影响到患者使用核苷(酸)类药物的效果和副作用。由于该患者 HBeAg 阴性,复发率高,疗程宜长,而 RCT 报告了上述两种核苷(酸)类药物治疗慢性乙型肝炎的短期疗效,但其对患者的长期疗效并不清楚。因此,我们可以将科学问题定义为“真实临床实践下 ETV 及 ADV 治疗慢性乙型肝炎患者长期疗效比较”。

确定了科学问题之后,需要针对具体的科学问题和数据库中所涵盖的患者特征,对研究设计、研究对象的特征、处理因素和结局因素进行详细的定义,并在此基础上提取研究所需要的数据资源。

### 二、研究设计

目前基于二手数据的观察性研究的设计方法包括横断面研究、病例对照研究、队列研究、病例交叉设计等。不同的研究设计方法其适用条件不同。横断面研究是通过收集某一个时间点或某一个时间断面的人群的结局和处理因素来探讨处理因素和结局之间的关联。如评价不同诊断在真实临床环境下的诊断准确性研究。该设计方法由于不能确定处理因素和结局变量的时间先后而不适于时间相关的处理因素和结局研究。病例对照研究则通过选择一组发生了结局的患者为病例,以没有发生结局者为对照,比较两组人群发生结局之前接受的治疗措施的差别,从而来推断治疗措施与结局之间的关系,主要用于罕见结局的研究。队列研究是先选择两组或多组队列人群,也就是符合入选和排除标准的两组或多组接收不同治疗措施的患者,通过一段时间的治疗,观察不同治疗措施的患者的结局(包括疗效和副作用),从而比较不同治疗措施的结局的差异。该研究能很好的确定治疗措施和结局之间的时效关系,因此一项好的队列研究被认为是观察性研究的金标准。上述三种观察性研究方法都很难保证不同处理组研究对象的可比性。而病例交叉设计则通过以自身作为对照,比较前后两种不同干预措施的疗效差别,通常用于瞬时暴露导致急性事件的发生。综上所述,不同的研究设计方法有其自身的适用范围和优缺点,需要根据具体科学问题选择合适的设计方法。

对于“真实临床实践下 ETV 及 ADV 治疗慢性乙型肝炎患者的长期疗效比较”,假如我们现





在有  $\times \times$  传染病专科医院 2008-01-01 至 2013-12-31 时间段的门诊电子病例数据库,其中有近 6 万慢性乙型肝炎病毒感染患者的诊断、治疗及临床检查结果,则我们可以针对上述 6 万的慢乙肝患者建立一个抗病毒治疗及随访的队列人群。

### 三、研究对象的选择

虽然初步有 6 万个慢乙肝患者的治疗队列,但是上述 6 万患者特征各异。有的为在治患者,有的为初治患者;有的患者可能只有检查结果而没有取药记录;还有的患者在治疗过程中出现了换药等。为了保证研究结果的真实性和外推性,我们需要对研究对象严格的定义,一方面要尽可能保证研究对象能够代表所有到临床就诊和治疗的患者,另一方面需要针对你要回答的科学问题确定严格的入选和排除标准来提高研究对象的同质性。关于如何采用严格的入选和排除标准来提高研究对象的同质性并没有统一的原则,需要具体问题具体分析,包括排除现患;仅入选新发或初治患者;排除有用药禁忌证的患者以及是否依从性特别差的患者等。重要的原则是既要保证人群代表性和真实性,又需要两组或者多组可比性。例如在回答“真实临床实践下 ETV 及 ADV 治疗慢性乙型肝炎患者的依从性及其长期疗效比较”时,考虑到在治患者在疾病的严重程度、药物的疗效、治疗的依从性等和初治患者显著不同,我们可以只选择初治患者以尽可能的保证研究对象的同质性。同时排除 ADV 和 ETV 的用药禁忌的患者,以及合并 HAV、HCV、HDV、HEV、HIV 或 CMV 感染,或因怀孕及其他严重疾病(代谢综合征,其他肿瘤,严重心脑血管疾病,尿毒症,呼吸道感染,药物性肝损伤等)。但是仅依赖于二手数据的信息,上述问题如是否存在用药禁忌证有时很难断定。我们可通过倾向性评分的方法(混杂控制部分将做详细描述)计算出每个个体接受某种治疗的概率得分(介于 0~1 之间),将概率接近于 0 或 1 者分别视为具有用药禁忌证者或总是治疗者而删除。

### 四、处理因素(治疗措施)的确定及其偏倚

处理因素和结局状态的确定是比较效果研究的两个基本要素。如何利用二手数据确定患者的治疗或干预因素是一件看似简单但实际很复杂的事情。以上述科学问题为例,处理因素的确定包括是否使用了 ETV 或 ADV、用药量、用药时间、依从性以及在治疗过程中是否换药。用药信息在电子病历数据库中可以通过药房的发药纪录来收集,包括取药时间、药物名称、药物剂量及用药量等信息,从理论上推论出每个患者是否使用 ETV 或 ADV、使用的时间以及用药的依从性。

但实际情况是,由于电子病例数据库只能提供患者在该院的拿药情况,并不能判断患者是否在其他医院或药店购买药物;同时也不能反映患者的服药情况,即使用电子病历数据库来判断患者的用药状态并不能保证暴露分组的真实性,也就是可能带来暴露错分。暴露错分包括差异错分和无差异错分。当所有入组的患者出现暴露错分的概率相同时,称为无差异错分。例如,如果被评价的所有药物均为自费药,同时几种药物的价格相近,则所有患者在医院以外的药房自行买药的概率在不同药物治疗组中相同,此时利用电子病历数据库中取药的纪录来判断患者用药情况出现暴露错分的概率在不同用药组应该相同或相近,即为无差异错分。相反,当使用不同药物者出现错分的概率不同时,称为差异错分。例如,同样是评价药物疗效,如果 A 药纳入医保报销,B 药未纳入报销范围,则利用电子病历数据库来对暴露进行分类时,对医保患者在不考虑服药依从性的情况下,A 药发生暴露错分的概率要低于 B 药,此错分为差异错分。我们可以通过小样本的抽样,并采用问卷调查的方法来估计是否可能存在错误,并评价暴露错分对于研究结果真实性的影响。





## 五、结局的判定及偏倚

慢乙肝患者治疗的最终目标是预防肝硬化、肝衰竭和肝癌的发生。但如果随访时间不够长,通常只能用一些替代指标,包括病毒学应答和血清学应答等。以上述科学问题为例,由于患者最长的随访时间可达4年,我们可以评价使用ETV或ADV后第6、12、18、24、30、36、48周的维持病毒学应答、维持生化应答和HBeAg血清学转换(仅针对HBeAg阳性患者)率。维持病毒学应答定义为:在抗病毒治疗期间血清HBV DNA检测不到(PCR法)或低于检测下限(<500copies/ml)。维持生化应答定义为:治疗期间血清ALT恢复正常(40 U/L)。HBeAg血清学转换定义为:血清HBeAg转阴且anti-HBeAg转阳。

为了计算上述指标,我们需要收集每个患者在不同时间点的生化检查:ALT和病毒学检查(包括HBV DNA, HBeAg, anti-HBeAg)等指标。

对于肝癌或死亡等结局指标,如果国家有健全的肿瘤和死亡登记系统,则有可能通过数据库的关联实现对患者的远期结局的随访,如果没有,则很难通过常规的电子病历系统实现对这些患者的远期随访。

利用二手数据库来判定患者结局时,由于资料的收集没有严格的质控,也存在错分的可能性。首先,临床病历数据库对于结局的编码有时并不标准。以门诊电子病历数据库为例,大多数时候医生并不遵从标准的ICD-10或ICD-9进行编码;其次,不同患者结局的诊断所依赖的临床检查结果可能出自不同的临床检验人员之手,不同的临床检验人员其操作的信度有待评价。第三,如果数据库跨度多年,不同年代测量同一个指标的仪器及试剂的精度可能不同,相同的标本在不同的年代可能测定的结果并不相同,甚至同期在同一医院对同一疾病或指标有不同的检测设备。因此,在使用的二手数据进行科学研究时,需要判断上述可能的因素对研究对象结局状态判定的影响。

## 六、协变量的收集

除了收集处理因素及结局相关信息之外,还需要尽可能收集其他可能影响结局的因素,即协变量。例如,我们在比较ADV和ETV的疗效时,不仅要收集患者的治疗和结局信息,同时还需要收集其他可能疗效因素,如年龄、性别、疾病严重程度(包括基线的血生化检查以判断患者的肝功能、肾功能等;病毒学检查包括HBV-DNA、HBeAg等)、并发症及既往病史等。但对于与患者的用药及其依从性密切相关的一些因素,如受教育程度、经济收入等,如果是利用二手数据,实际很难收集到。

### 第三节 比较效果研究中混杂因素存在的判定

比较效果研究的最终目的是评价某个特定人群两个变量,即治疗措施和结局的之间的关系。而对于真实临床实践而言,患者被选择何种治疗措施会受到患者的症状、疾病严重程度、医疗政策、患者的社会经济状况等多种混杂因素的影响,如果不同治疗组的患者上述特征上存在差异,同时这些因素也与结局相关,则上述这些因素就构成了研究不同治疗措施结局差异的混杂因素,即可能歪曲不同治疗组的效果差异。例如,在评价目前的某种标准治疗措施和新药的疗效时,使用旧药疗效好的患者不可能使用新药,同时新药一般均较旧药昂贵,因此,在临床上使用新药的患者通常疾病严重程度都较使用旧药的患者严重。我们在分析××医院2008~2013年使用ETV和ADV的患者的特征时验证了上述假说(表10-1)。



表 10-1 2008~2013 年 × × 医院使用 ETV 和 ADV 的初治患者部分特征比较

|                           | HBeAg(-)       |                |         | HBeAg(+)       |                |         |
|---------------------------|----------------|----------------|---------|----------------|----------------|---------|
|                           | ADV<br>(n=581) | ETV<br>(n=765) | P       | ADV<br>(n=539) | ETV<br>(n=609) | P       |
| 年龄(y)                     | 34.8 ± 11.5    | 37.8 ± 12.1    | <0.0001 | 44.1 ± 11.8    | 46.8 ± 12.4    | 0.0002  |
| 男性(%)                     | 388 (66.8)     | 535 (69.9)     | 0.2171  | 357 (66.2)     | 452 (74.2)     | 0.0031  |
| HBV-DNA (log10 copies/ml) | 7.3 ± 0.9      | 7.5 ± 0.9      | <0.0001 | 5.7 ± 1.1      | 6.0 ± 1.1      | <0.0001 |

(数据摘自北京协和医学院邱倩博士研究生论文)

混杂因素可分为时间依赖性和非时间依赖性变量,或者测量和未被测量的变量。

虽然二手数据可能包含了丰富的患者的特征,如人口学资料、症状及体征等帮助我们有效地控制混杂,但是仍然不可避免的遗漏掉许多可能既影响临床决策又与临床结局有关的变量,即未被测量的潜在混杂变量。在可能的情况下,如果对未被测量的混杂变量进行控制可有效提高比较效果研究的真实性。

对于真实临床实践而言,患者选择何种治疗措施会受到患者的经济状况、症状及疾病严重程度等多种因素的影响,其中患者的经济状况在短时期内不会发生改变,为非时间依赖性混杂因素;而患者的症状及疾病严重程度会随着时间而改变,由此也可能导致临床治疗决策,即暴露分组的变化,为时间依赖性混杂变量。

## 第四节 比较效果研究中混杂变量的控制

对于混杂因素的控制可通过在研究设计阶段采用限制以及在统计分析阶段通过一系列的统计分析方法来实现。前面在研究对象的选择方面已对采用限制方法尽可能提高研究对象的内同质性进行了阐述,此部分主要针对统计分析方法控制混杂进行进一步阐述。

对于不同性质的混杂变量的控制所采用的统计分析方法并不相同。校正时间依赖性混杂变量的分析较为复杂,多采用边际结构模型或结构巢式模型等,在这里不做详细说明;下面仅针对非时间依赖性混杂变量(包括测量和未被测量的变量)进行详细论述。目前对于控制非时间依赖性混杂因素的统计分析方法可归为两大类,风险校正方法和工具变量分析法。其中风险校正法只能用于控制测量的混杂因素,其假设前提是未被测量的因素对不同处理组的结局影响可被忽略,包括分层、多元回归以及倾向性评分等方法。而工具变量法可用于控制非测量的混杂因素的影响。

### 一、分层分析

分层分析是最简单的用于控制混杂的方法,其原理是将所有的研究对象按照某种特征分成不同亚组,分别比较不同亚组不同治疗措施的疗效差别。分层分析是一种最直观的控制混杂的分析方法,但仅适用于只存在少数混杂变量的比较效果研究中。

例如,在评价 ETV 和 ADV 用于治疗慢性乙型肝炎的疗效时,考虑到 HBeAg(+)和 HBeAg(-)的患者其疗效可能不同,我们可以按照 HBeAg 的状态对患者进行分层,分别比较阳性和阴性者 ETV 和 ADV 的疗效。但是抗病毒药物疗效除了受 HBeAg 状态的影响外,还会受到患者的年龄、性别、基线 HBV-DNA 水平、ALT 水平以及用药依从性等多个因素的影响,如果按照上述每个变量进行分层,则可能导致每层下的样本量太小,此时分层分析不再适用。

### 二、多元回归方法

当存在多个混杂变量时,我们不再使用分层分析的方法来控制混杂,而是可以采用多元回



归的方法来校正混杂因素对于处理因素(如 ETV 和 ADV 抗病毒药物)与结局(如病毒学应答、ALT 复常或 HBeAg 阴转)之间关联的影响。基于所要研究的结局类型以及是否考虑发生结局的时间,可采用不同的回归模型对可能的混杂因素进行控制。

### 知识点

#### 多元回归方法选择

1. 当只关注结局而不关注发生结局的时间时,如果结局变量为数值型变量,则可采用多元线性回归的方法对可能的混杂因素进行控制。
2. 当结局变量为分类变量时,则可采用多元 logistic 回归等方法。例如,如果我们想要分析影响服用 ETV 24 周后 ALT 复常的影响因素,则可采用多元 logistic 回归方法。
3. 当不仅关注疾病发生结局(二值变量)同时也关注发生结局的时间时,则可采用 Cox 回归的方法对混杂因素进行控制。

### 三、倾向评分法

倾向评分法(propensity score, PS)是回归分析的进一步扩展,由 Rosenbaum 和 Rubin 在 1983 年首次提出。其原理是通过构建与结局相关的协变量与处理因素之间的函数关系,确定每一个研究对象划分到不同处理组的条件概率(即倾向评分, PS 得分),将 PS 得分相同或相近的个体在不同组间进行匹配,从而保证不同处理组间上述协变量在总体分布上的可比性。通常包括三个步骤,首先以大样本的观察性数据为基础,采用 logistic 回归或 probit 模型建立处理因素与混杂变量的回归模型;第二,计算每个个体的倾向性得分;最后采用合适的方法(如匹配、分层、回归校正和倾向评分加权等)保证不同处理组的均衡可比。

例如,前面我们看到,在  $\times \times$  医院就诊的 ETV 和 ADV 两组患者在基线,包括年龄、性别、HBV-DNA 以及 ALT 等水平均存在差异,服用 ETV 的患者其疾病严重程度明显较 ADV 患者要严重。而上述这些因素同时与患者的抗病毒治疗的预后有关。为了保证两组患者的疾病严重程度具有可比性,我们可以采用倾向性评分的方法对两组患者进行匹配。由于 HBeAg 阳性和阴性者患者特征以及抗病毒治疗效果显著不同,因此我们先以 HBeAg 状态分层,然后在每层下以所有接受 ADV 或 ETV 的初治患者为研究对象,以是否接受 ETV 治疗为因变量(接受 ETV 治疗者为 1, ADV 治疗者为 0),以患者的年龄、性别、医保类型、HBV-DNA 水平、ALT 水平等因素建立 logistic 回归模型,并计算出每个个体可能使用 ETV 进行抗病毒治疗的概率,即 PS 得分;之后按照每个个体的 PS 得分采用匹配、分层、回归校正或加权的方法来保证不同处理组的均衡可比。例如,我们可以以实际选择 ETV 治疗者为参照组,按照  $PS \pm 0.1$  标准差(指两组人群 PS 分布的标准差)实现 ETV 和 ADV 两组人群的匹配。表 10-2 为按照 PS 匹配后 ETV 和 ADV 两组人群基线特征的比较,我们可以看到,两组人群的基线特征具有可比性。

倾向评分法虽然能减少观察性研究在估计处理效应时的混杂偏倚,尤其适用于混杂因素很多而结局变量发生率很低的情况;但是此方法仅均衡了组间的已知因素,对未知混杂因素引起的偏倚不能进行校正;同时倾向评分法不适用于样本量较小的情况;而当处理组和对照组倾向评分没有重叠或者重叠很少时,无法进行合适的匹配。因此,虽然该方法在大样本的观察性比较效果研究中应用越来越多,但是应用时仍然要考虑其适用范围。

### 四、工具变量法

上述三种控制混杂的方法仅适用于已经测量的混杂变量。而工具变量的提出则为控制



表 10-2 2008~2013 年 × × 医院使用 ETV 和 ADV 的初治患者 PS 匹配后部分特征比较

|                          | HBeAg(+)       |                |        | HBeAg(-)       |                |        |
|--------------------------|----------------|----------------|--------|----------------|----------------|--------|
|                          | ADV<br>(n=460) | ETV<br>(n=460) | P      | ADV<br>(n=397) | ETV<br>(n=397) | P      |
| 年龄(岁)                    | 36.0 ± 11.7    | 36.2 ± 11.2    | 0.8025 | 45.7 ± 11.8    | 45.9 ± 12.1    | 0.8145 |
| 男性(%)                    | 310(67.4)      | 303(65.9)      | 0.6245 | 276(69.5)      | 274(69.0)      | 0.8777 |
| HBV-DNA(log10 copies/ml) | 7.4 ± 0.9      | 7.4 ± 0.9      | 0.8261 | 5.8 ± 1.1      | 5.8 ± 1.1      | 0.7945 |

(数据摘自北京协和医学院邱倩博士论文)

未测量的混杂因素带来可能。在探讨不同处理因素的疗效差异研究时,我们可以通过寻找一种与患者采用哪种处理因素密切相关,但是与患者的健康状况无关(包括测量和未测量的因素)的变量作为工具变量;并以工具变量作为分组依据,通过不同工具变量水平下(而非真实的不同处理组)患者的疗效差别来评价不同处理因素的疗效差别,由此避免了因不同处理组研究对象的不可比性所带来的系统误差。例如,在一项利用全国医疗保险健康管理数据对美国 1994~1995 年因发生急性心肌梗死入院的 122 124 例 65~84 岁老人的观察性研究中,试图通过上述二手数据探讨心导管术对发生心梗患者死亡率影响的研究。由于选择心导管术的患者大多较非手术者年轻、健康,发生急性心肌梗死的程度也较轻,因此很难判断选择心导管术的患者较低的死亡率是否因为该措施所致。进一步的资料分析显示,该数据涵盖了全国不同地区的患者,不同地区的心导管手术率不同,但患者心肌梗死的严重程度并无差异。由此可以将不同地区的心导管手术率作为一个工具变量,该变量与患者是否选择心导管术有关,但与疾病严重程度无关。通过比较不同心导管手术率的患者(而非实际采用心导管术与未采用心导管术的患者之间的比较)死亡风险来推断心导管术对于发生急性心肌梗死患者死亡率的影响。

对于同一观察性研究数据,为了减少混杂因素对研究结果真实性的影响,通常建议采用上述多种统计分析方法对可能的偏倚进行校正。但经常面临的问题是,针对同一数据采用不同的统计分析方法其结论并不完全一致。例如,Stukel TA 等分别采用 Cox 回归、倾向评分回归校正、倾向评分匹配以及工具变量等四种方法同时来校正混杂因素对于评价心导管术对发生心梗患者死亡率的影响。结果发现,采用上述三种风险校正方法,行心导管术的急性心梗患者发生死亡的风险均较没有进行心导管手术的患者降低 50%;ORs 分别为 0.51(95%CI:0.50~0.52)、0.54(95%CI:0.53~0.55)及 0.54(95%CI:0.52~0.56)。而采用工具变量方法进行校正后,OR 为 0.84(95%CI:0.79~0.90),相对于风险校正模型更趋向于保守。上述出现不同结果的原因可能与工具变量校正了未测量的因素有关。但工具变量与风险校正模型之间究竟是何种关系依然有待探讨。Fang G 等的研究显示,当某种治疗措施在受试人群中的治疗效果具有同质性,或即使具有异质性但是否采用某种治疗措施与治疗效果无关时,同一数据基于风险校正模型和工具变量分析方法所获得的结果应该可比;相反,当治疗措施在受试人群中的治疗效果不同,且治疗措施的确定与治疗效果密切相关时,则同一数据基于两种不同的校正方法所得结果不具有可比性。由于 Fang G 等的研究结论是基于模型基础上产生的,而在实际的研究中,控制变量的选择与确定不仅与研究者的专业知识有关,也与数据的可获得性有关,随着选择控制的变量不同,采用风险校正模型以及工具变量校正的结果也不相同。

总之,虽然大量的二手数据的存在为我们开展以真实临床实践为基础的比较效果研究提供了可能性,但需要基于所要回答的科学问题,从研究设计、研究对象的选择、暴露及结局资料的收集以及统计分析等多方面入手,对可能存在的选择、信息及混杂偏倚进行控制,从而尽可能地保证研究结果的真实性。



### 思考题

1. 什么是比较效果研究?
2. 开展比较效果研究的意义。
3. 阅读一篇已经发表的比较效果研究文献,回答文献的科学问题是什么? 研究设计如何? 如何收集信息(包括处理因素、结局因素以及潜在的协变量)? 应用什么方法控制混杂变量? 研究结果的临床意义。

(刘晓清 王丽)

## 第十一章 病因与危险因素研究

### 临床案例

患者男性,47岁,机关干部。体检发现血清尿酸 $563\mu\text{mol/L}$ 、空腹血糖 $5.4\text{mmol/L}$ 、肌酐 $83\mu\text{mol/L}$ 、甘油三酯 $2.8\text{mmol/L}$ ,BP 120/75mmHg,身高172cm,体重90kg,腹围103cm,B超发现中度脂肪肝。平时吸烟20支/天 $\times$ 20年,饮白酒4两/次 $\times$ (2~3)次/周,很少参加体育活动,无关节炎、肾脏病以及心脑血管疾病的病史,但父亲有高血压和冠心病。就诊的目的是了解其血清尿酸水平升高是否需要治疗。

### 第一节 如何获得临床研究问题

尿酸是人嘌呤代谢的最终产物。大多数哺乳动物体内有尿酸酶,可将尿酸分解为易溶于水的尿囊素并经肾脏排出体外,所以其血清尿酸水平一般较低。人类在进化过程中由于基因突变而失去尿酸酶的活性,只能将微溶于水的尿酸经肾脏排出体外,所以人的血清尿酸水平比多数哺乳动物都高。尿酸在血清中的饱和度大约是 $416\mu\text{mol/L}$ ,超过此值即可析出尿酸钠结晶,引发急性痛风性关节炎、尿酸性肾病、形成痛风石或肾结石而危害健康,故一般将高尿酸血症定义为血清尿酸水平 $>416\mu\text{mol/L}$ 。

患者有明确的高尿酸血症、高甘油三酯血症、肥胖( $\text{BMI}=\text{体重}/\text{身高}^2=30.4\text{kg/m}^2$ )、脂肪肝,有心血管疾病家族史,生活方式不健康(吸烟、饮酒、缺少运动)。他提出的问题是“高尿酸血症需不需要治疗”,这在临床上是一个治疗决策问题,如果想回答这个问题,我们需要知道高尿酸血症有哪些危害、治疗高尿酸血症有哪些获益和风险,综合考虑得失利弊之后,才决定是否需要治疗。在此,我们只针对“高尿酸血症的危害”这一点展开讨论。

人们对高尿酸血症危害的认识源于痛风。痛风是人类独有的疾病,主要由于尿酸生成过多和(或)排泄减少所致,临床上从关节炎的角度分为无症状高尿酸血症期、急性发作期、发作间期和慢性期4个阶段。高尿酸血症是痛风明确的危险因素,痛风发病的风险随血清尿酸水平的升高而增加。临床研究和实验研究均证实尿酸盐晶体是导致急性痛风性关节炎、尿酸性肾病和尿酸性肾结石的原因。美国、欧洲、日本和我国关于痛风治疗的临床指南都指出,急性痛风性关节炎发作 $\geq 2$ 次/年、慢性痛风性关节炎、尿酸性肾病、肾结石或痛风石都是降尿酸治疗的指征。降尿酸治疗分为生活方式干预和降尿酸药物两个方面,如果生活方式干预不能使血清尿酸水平达标( $\leq 300\sim 360\mu\text{mol/L}$ ),则需要加用降尿酸药物治疗。从关节的角度出发,无症状高尿酸血症本身并不需要药物治疗。

然而,无症状高尿酸血症是否会对心血管系统造成危害目前存在很大的争议。我们知道,心血管疾病是现代社会的常见病和多发病,在我国疾病负担很重,是人口最主要的死亡原因之一。心血管疾病的防控一直是国家公共卫生的工作重点,近30年来心血管疾病研究的主要成就就在于明确了若干主要的危险因素并对其进行积极有效的干预,使心血管疾病的发病率和死亡率显著下降。在此基础上,寻找新的危险因素对进一步提高心血管疾病的防控效果,其重要意义是不言而喻的。因此,高尿酸血症是否是心血管疾病的危险因素近年逐渐成





为研究的热点。本案例患者,他多个冠心病的危险因素,因此我们最关心的临床问题是无症状高尿酸血症会不会促进冠心病的发生,即“无症状高尿酸血症是冠心病的病因或危险因素吗”。

这是一个病因学研究的问题。病因学研究是临床医学研究中的一个重要内容,因为只有真正理解导致疾病发生的原因,才能有效地对疾病进行预防、诊断和治疗。一直以来,人们都对疾病发生的原因抱有浓厚的兴趣,发现病因也是医学科研工作者的目标。

### 知识点

#### 病因和危险因素

1. 病因 病因(cause)是导致疾病发生的原因。
2. 危险因素 危险因素(risk factor)是流行病学层面的病因,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。当该因素存在时,人群中疾病发生的概率随之上升;当消除该因素时,疾病发生的概率也随之下降。

针对这个临床问题,我们首先需要对现有的临床指南进行回顾,目前医学界对无症状高尿酸血症是否是心血管疾病的危险因素尚无一致意见,但是我国和日本却都提倡在经过饮食控制血清尿酸水平仍高于 $9.0\text{mg/dl}$ (即 $535\mu\text{mol/L}$ ),或合并高血压、缺血性心脏病、糖尿病和代谢综合征血清尿酸水平不低于 $8.0\text{mg/dl}$ (即 $476\mu\text{mol/L}$ )的患者,进行药物降尿酸治疗。另一方面,美国风湿病学院(ACR)和欧洲抗风湿病联盟(EULAR)却在其指南或建议中“因为缺少前瞻盲法对照临床试验,所以对无症状高尿酸血症的药物治疗不予置评”。

不难发现,由于高质量的研究证据很少,尿酸是否是冠心病的病因或危险因素目前还没有公认的结论,是否需要进行药物治疗也存在一定的争议。为了加深对这个问题的认识,我们还需要查阅和分析原始的科学研究论文,为设计和开展创新性研究做准备。

## 第二节 临床问题转化为科学问题

临床问题往往比较笼统,而科学问题则十分具体,如果将临床问题比喻成一个面,那么科学问题要落实为一个点,所以我们说将临床问题转化为科学问题是一个抽丝剥茧、具体细化的过程,在这个过程中需要根据可行性(feasible)、趣味性(interesting)、创新性(novel)、是否符合伦理(ethical)以及重要性(relevant)等几个方面(简称FINER)综合考虑并做出决定。

在将临床问题转化为科学问题的过程中,我们需要从以下几个方面对临床问题进行提炼。

1. 研究人群 研究人群选自来源人群,来源人群来自目标人群,所以首先需要确定的是:研究的目标人群是什么?因为我们关心的是最一般意义上“高尿酸血症对冠心病发生的影响”,所以目标人群应该是现在全世界的人,不分国家、地区、种族、年龄、性别等,这也是我们希望研究结果能够外延的人群。实际上,这样的目标人群是我们的研究所不可及的,我们可及的人群往往是某个地区、某个时间段、具有某种特征的人群,即研究的来源人群,如目前北京市东城区的常住居民。从来源人群中,我们再根据研究的具体要求(如纳入标准、排除标准以及研究对象知情同意的情况等)确定真正参加研究的具体对象,即研究人群。

在本案例中,我们从来源人群中获得研究人群的方式可以不尽相同,如在研究经费充裕的情况下,可以考虑在某些地区采取整群随机抽样的方法确定社区,再通过组织社区居民体检筛选出符合条件的研究对象,如果预算比较有限,也可以考虑与体检中心合作选取符合条件的体检人群作为研究对象(方便人群)。一般情况下,随机抽样获得的人群比方便人群的代表性要好一些。



2. 暴露因素 / 干预措施 在临床研究中,暴露因素或干预措施的选择取决于研究类型。在横断面研究、病例 - 对照研究或队列研究等观察性研究中,不涉及对可能的危险因素进行主动干预,而只对暴露因素进行测量;在本案例中,暴露因素是血清尿酸,具体的变量形式可以是二分变量(有 / 无高尿酸血症)、等级变量(如将血清尿酸水平按四分位分为4组)或连续变量(血清尿酸水平)。在对照试验等干预性研究中,试验组和对照组其暴露因素的暴露水平应该是大致一样的,试验组采取干预措施,而对照组不采取干预措施(安慰剂对照或常规治疗对照);在本案例中,试验组和对照组的血清尿酸水平应该大致相同,干预措施可选降尿酸药物治疗,具体的药物种类(如别嘌醇、苯溴马隆或非布司他等)、剂量和用法也需要进一步明确。

3. 对比 在横断面研究或病例 - 对照研究中,是将出现结局的研究对象与未出现结局的研究对象进行对比。在队列研究中,是将暴露于可疑致病因素之下的研究对象与未暴露的研究对象进行对比。在随机对照试验中,是将除干预措施外各方面均十分相似的试验组(药物干预组)与对照组(非药物干预组)进行对比。

4. 结局 在这个例子中,“冠心病”仍旧比较模糊而不明确。我们知道,冠心病是由于冠状动脉发生粥样硬化而引起血管腔狭窄或阻塞,造成心肌缺血、缺氧或坏死而导致的心脏病,WHO将冠心病分为无症状心肌缺血(隐匿性冠心病)、心绞痛、心肌梗死、缺血性心力衰竭(缺血性心脏病)和猝死5种临床类型,而临床中又常分为稳定性冠心病和急性冠状动脉综合征。那么,在实际研究中,我们如何定义“冠心病”,这个结局指标的选择十分重要,需要从科学性、重要性、可行性等几个方面加以考虑。一般而言,在临床研究中,有比较“硬”的结局指标,也有比较“软”的结局指标。“硬”和“软”都是相对而言的,“硬”通常是指结局指标比较直接、客观和准确,“软”通常是指结局指标比较间接、主观和模糊。举例来说,同样是“冠心病”的病因 / 危险因素研究,有人选择冠状动脉造影的诊断结果作为“冠心病”的判断标准,也有人选择问卷调查中研究对象回答的结果作为“冠心病”的判断标准,显然前者的结局指标比较“硬”(直接、客观和准确),后者比较“软”(间接、主观和模糊),和前者相比后者的缺点是可能带来比较明显的诊断偏倚,优点是获取信息的成本低、有效应答率高等。除了直接选择“冠心病”作为结局指标(比较“硬”)外,也有人选择一些中间指标如颈动脉粥样斑块或脉搏波传导速度(PWV)(比较“软”)替代“冠心病”作为结局指标进行研究,和前者相比后者的优点是出现所谓“结局”所需的时间较短,缺点是它并不是真正的结局“冠心病”。

在表 11-1 中,我们将临床问题转化为科学问题的几个要点进行了总结。

表 11-1 将临床问题转化为科学问题:PICO

|           |                            |
|-----------|----------------------------|
| 研究人群      | 社区人群 / 体检人群                |
| 干预措施或暴露因素 | 降尿酸药物治疗 / 高尿酸血症            |
| 对比        | 试验组与对照组 / 暴露组与非暴露组         |
| 结局        | 冠脉造影发现斑块 / 颈动脉超声发现斑块 / PWV |

根据 FINER 标准,我们可以将上述临床问题转化为不同的科学问题。例如:

(1) 在某地区三甲医院心脏内科的住院患者中,冠脉造影诊断冠心病者与冠脉造影正常者相比较,以往高尿酸血症的患病率是否有差别(病例 - 对照研究设计)。

(2) 在某社区既往无冠心病史的人群中,经体检化验发现高尿酸血症的患者与血清尿酸水平正常者相比较,随访出现临床冠脉事件的比例是否有差别(队列研究设计)。

(3) 在若干家三甲医院体检人群中,经化验发现高尿酸血症而无其他疾病的患者,随机分为2组,一组予以别嘌醇降尿酸治疗,另一组予以安慰剂作对照,1年后两组患者脉搏波传导速度



的变化是否有差别(多中心随机对照试验设计)。

### 第三节 病因学研究的模型与工具

对于某些疾病,导致疾病发生的因素很多,有些是明确的危险因素,有些是正待研究的可疑危险因素,我们怎样理清这些因素之间以及各因素与疾病发生之间的关系呢?从实用分析的角度,我们可以把病因分为直接病因和间接病因。

#### 知识点

##### 直接病因和间接病因

1. 直接病因 直接导致疾病发生的某些因素,称之为直接病因(direct cause)。
2. 间接病因 并不直接导致疾病发生而是通过一些其他因素的作用间接导致疾病发生的因素,称之为间接病因(indirect cause)。

在图 11-1 中,因素 X 是疾病 D 的直接病因;但是对于疾病  $D_1$  而言,因素  $M_1$  是直接病因,因素  $X_1$  是间接病因, $X_1$  完全通过  $M_1$  发挥其致病作用;对于疾病  $D_2$  而言,因素  $M_2$  是直接病因,因素  $X_2$  既是直接病因也是间接病因, $X_2$  除了有直接的致病作用以外,还部分通过  $M_2$  导致疾病  $D_2$  的发生。

学  
习  
笔  
记

将病因分为直接病因和间接病因的目的,主要是为构建病因模型时能够较清晰地勾勒出各个因素之间的关系及其对疾病发生的影响。这种以临床为基础,将直接病因与间接病因相结合的模型称为疾病因素模型,该模型在病因研究方面具有较强的可操作性和实践指导意义。

有一种帮助构建病因模型的工具,叫做有向无环图(directed acyclic graph, DAG),DAG 能够比较清晰地展现与疾病发生有关的各因素之间及其与疾病发生之间的关系,有利于在此基础上进行因果关系的建模、分析与计算。

#### 知识点

##### 有向无环图

有向无环图:由若干个线段、箭头将各种因素与疾病发生链接而成,箭头方向代表因果关系,没有箭头的线段代表因果不明的相关关系,箭头还代表发生时间上的前后关系,因此箭头不会形成回路。有向无环图是构建病因学研究模型的一种实用工具。

图 11-2 示当研究因素 L 对疾病 Y 的致病作用时,还需要考虑其他因素(U 和 A)的影响,其中 L 是 A 的病因,A 是 Y 的病因,即 L 除了对 Y 有直接的致病作用以外(图中由 L 直接指向 Y 的弧线),L 还通过 A 对 Y 有间接的致病作用,即 A 是 L 对疾病 Y 致病作用的中介因素(mediator),而 U 既是 L 的病因,也是 Y 的病因,所以 U 是 L 对疾病 Y 致病作用的混杂因素(confounder)。当计算 L 对疾病 Y 的致病作用时,需要对混杂因素 U 进行校正(adjustment),而不需要对中介因素 A 进行校正。

在了解病因学研究的模型和工具之后,我们再回到之前的临床问题。我们知道冠心病是一个多因素致病的疾病,年龄、性别、家族史、不良生活习惯如吸烟、不合理膳食(高脂肪、高胆

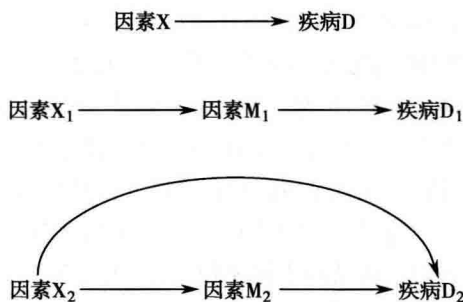


图 11-1 直接病因与间接病因



固醇、高热量等)、缺少体力活动、过量饮酒及高血压、血脂异常(包括总胆固醇过高或低密度脂蛋白胆固醇过高、甘油三酯过高、高密度脂蛋白胆固醇过低等)、超重/肥胖、高血糖/糖尿病等都是冠心病的危险因素。这些危险因素之间的相互关系及其对冠心病的致病作用十分复杂(图 11-3),其中年龄、性别、家族史、不良生活习惯属于比较上游的危险因素,高血压、血脂异常、超重/肥胖、高血糖/糖尿病等属于比较下游的危险因素,前者(间接病因)主要通过后者(直接病因)发挥其致病作用,但前者也有一部分直接致病作用(直接病因)。

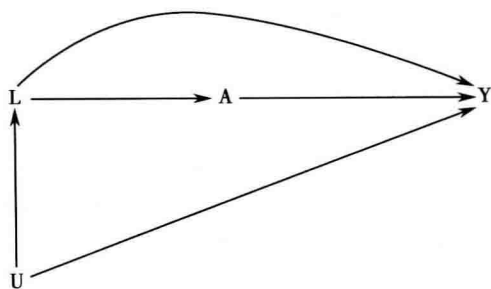


图 11-2 有向无环图

高尿酸血症作为一个可能的新的冠心病危险因素,它与传统的冠心病危险因素之间的关系如何?它是一个直接病因还是一个间接病因?抑或既有直接作用,又有间接作用?这个问题我们需要查阅文献,在现有科学认识的基础上,用 DAG 等理论图的形式提出假设。

目前有多个研究表明,高尿酸血症与高血压、血脂异常、超重/肥胖、高血糖/糖尿病之间存在相关性,高尿酸血症可能是这些冠心病危险因素的危险因素。以高血压为例,18 个前瞻性队列研究的荟萃分析表明,高尿酸血症使高血压的发病风险增高 41%(校正后 RR 1.41, 95%CI: 1.23~1.58),血清尿酸每增加 1mg/dl(约 60 $\mu$ mol/L),高血压发病风险增加 13%(校正后 RR 1.13, 95%CI: 1.06~1.20)。在动物实验中,通过抑制尿酸酶活性提高小鼠血清尿酸水平,导致小鼠血压升高,而黄嘌呤氧化酶抑制剂或促进尿酸排泄药可降低小鼠血清尿酸水平,从而阻断了血压升高。在高血压患者中进行的随机对照试验也提示别嘌醇可以轻度降低血压。

另一方面,性别、年龄、家族史、不良生活方式等冠心病的传统危险因素,也是高尿酸血症的危险因素。因此,我们在将图 11-3 中危险因素合并简化的基础上,构建了一个新的 DAG 来代表高尿酸血症与传统的冠心病危险因素之间的相互关系及其对冠心病致病作用的假设(图 11-4)。性别、年龄、家族史、不良生活方式既是高尿酸血症的危险因素,也是冠心病的危险因素,所以它们是高尿酸血症对冠心病致病作用的混杂因素,当计算高尿酸血症对冠心病的致病作用时,需要对这些混杂因素进行校正。高血压、血脂异常、超重/肥胖、高血糖/糖尿病等因素与高尿酸血症之间存在相关性,这些因素有可能是高尿酸血症的危险因素,也可能反过来高尿酸血症是这些因素的危险因素,或者二者之间存在非因果关系的相关性,所以在图 11-4 中高尿酸血症与这些因素之间是用不带箭头的线连接起来的,仅表明其相关性而并非因果相关。当计算高尿酸血症对冠心病的独立致病作用时,无论这些相关因素是混杂因素还是中介因素,均要对它们进行校正;如果只计算高尿酸血症对冠心病的总体致病作用,则对于在高尿酸血症之后出现的高血压、血脂异常、超重/肥胖、高血糖/糖尿病等因素(中介因素),则不需要对其进行校正。

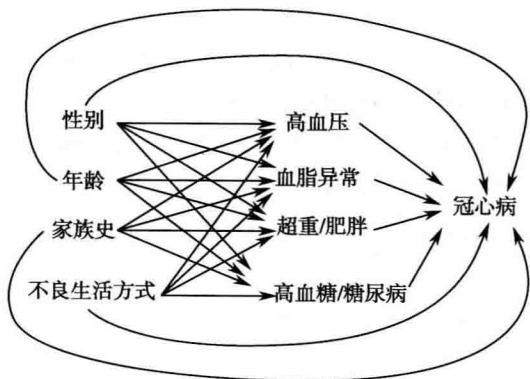


图 11-3 冠心病的危险因素

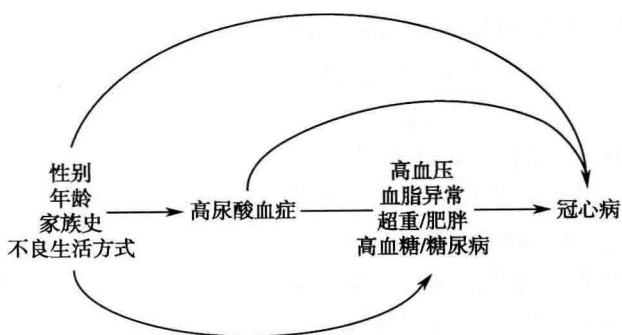


图 11-4 高尿酸血症与传统的冠心病危险因素

## 第四节 病因与危险因素研究的基本过程

病因与危险因素研究的基本过程经历以下三个阶段:①发现线索,提出病因假设;②初步验证病因假设;③综合判断,确定因果关系。临床研究以患者人群为主要的研究对象,涉及的研究方法有描述性研究(如个案报告、病例系列)、分析性研究(如横断面研究、病例-对照研究、队列研究)和干预性研究(如随机对照试验)等,其科学性和论证强度依次渐强。病因研究的第一阶段以描述性研究为主,临床观察和专家意见可以为病因研究提供一些线索和经验性证据。第二阶段以分析性研究和干预性研究为主,它们能够对病因假设进行有效的推断,是检验其成立与否的主要手段。因为干预性研究(如随机对照试验)所需要的条件(如伦理)较高,所以多数验证病因假设的临床研究以病例-对照研究和队列研究为主。在病因研究的第三阶段,如果存在较多的证据,而各个证据之间的结论并不一致时,需要对证据的科学性进行严格的评判。一般而言,高质量的集合研究(如 Meta 分析)可以为病因研究提供更高级别、说服力更强的证据。

另外,单纯的人群研究对确定疾病的病因是不够的,病因研究还需要其他学科(如基础医学)的广泛参与和支持。动物实验、组织细胞实验以及分子水平的研究都能够为探索某些因素与疾病发生之间的因果关系提供重要的证据,揭示人类疾病在发生过程中某个阶段的变化及其规律,但由于这些研究比较极端地控制了研究条件及可能的种属差异,其研究结果往往不可轻易地引申至人。因此,病因研究很大程度上需要以人为研究对象的临床研究和其他学科的研究互相支持和补充,经过反复分析和验证,确定疾病发生的原因。

## 第五节 判断因果关系的原则

判断疾病的病因或危险因素,我们通常参考因果关系的判断原则(Hill's Criteria,表 11-2)。在这些原则中,最值得强调的是时间顺序,即致病因素必须出现在疾病发生之前。这一点很容易理解,但是在病因学研究设计或数据分析时又很容易被研究者所忽视。在有向无环图中,箭头的方向不仅代表了因果关系,也代表了时间顺序;因此,致病因素应该位于箭头的起点,而疾病发生应该位于箭头指向的地方。其他主要的原则还包括关联强度、剂量-反应关系以及不同研究之间的一致性。

表 11-2 因果关系判断的原则(Hill's Criteria)

| 主要原则 |                                                                                                                  |
|------|------------------------------------------------------------------------------------------------------------------|
| 时间顺序 | 可疑的致病因素必须出现在疾病发生之前,这是因果推断中唯一一个绝对必需的条件                                                                            |
| 关联强度 | 常用相对危险度(relative risk, RR)、优势比(odds ratio, OR)或风险比(hazard ratio, HR)等来表示可疑的致病因素与疾病发生之间的关联强度,关联强度越强,该因素是病因的可能性就越大 |





续表

| 主要原则      |                                                                      |
|-----------|----------------------------------------------------------------------|
| 一致性       | 不同的研究者在不同的条件下用不同的方法进行不同的研究,发现可疑的致病因素与疾病发生之间的相关性是一致的                  |
| 剂量 - 反应关系 | 可疑致病因素暴露的水平(剂量和时间)越高,疾病发生的风险也越高,或可疑的致病因素与疾病发生之间的关系符合预期的理论形态(如 U 型关系) |
| 次要原则      |                                                                      |
| 试验证据      | 在人群的试验研究中,对可疑的致病因素进行干预能够预防或减轻疾病发生的风险                                 |
| 生物可能性     | 可疑的致病因素与疾病发生之间的关联能够用已知的生物学、生理学、病理学等相关的生命科学知识进行解释                     |
| 符合性       | 可疑的致病因素与疾病发生之间的关联与已有的理论和知识相一致                                        |
| 特异性       | 可疑致病因素会产生特异性的效果,如某个疾病的出现而不是多个疾病的出现                                   |
| 类比性       | 相似的因素会产生相似的结果(疾病发生)                                                  |

## 第六节 选择具体的研究方法

在进行病因和危险因素研究时,几乎所有的研究方法都可以选用,如描述性研究(如个案报告、病例系列)、分析性研究(如横断面研究、病例 - 对照研究、队列研究)和干预性研究(如随机双盲对照试验)等,但是它们各自的目的、科学性以及提供的证据级别存在一定的差别。

描述性研究以描述临床现象为主,它能够为病因假设提供一些宝贵的线索,但因为缺少对照组,所以对因果关系的论证强度很弱,如在 20 例急性非致死性心肌梗死的患者中发现 15 例在入院的时候血清尿酸水平都大于  $420\mu\text{mol/L}$ ,并不能说明二者之间存在因果关系。分析性研究和干预性研究是检验病因假设成立与否的主要手段。下面,我们结合这个例子,对分析性研究和干预性研究的几种主要研究方法做一个扼要的分析和比较,帮助大家根据特定的研究问题来选择合适的方法。

1. 横断面研究 Gur 等人以择期冠脉造影鉴别缺血性心脏病的患者(均有造影指征:典型或不典型的胸部不适和(或)运动试验异常)为研究对象,顺序入组 495 例冠脉造影诊断冠心病患者和 356 例冠脉造影正常者,并于冠脉造影前(空腹 12 小时)即刻抽血化验血清尿酸和其他生化指标,冠心病组和正常对照组血清尿酸水平分别是  $(5.31 \pm 1.73)\text{mg/dl}$  和  $(4.94 \pm 1.67)\text{mg/dl}$ ,两组之间有显著性差异,  $P=0.002$ 。按性别分层,在男性( $n=505$ )中,冠心病组和正常对照组血清尿酸水平分别是  $(5.55 \pm 1.56)\text{mg/dl}$  和  $(5.29 \pm 1.53)\text{mg/dl}$ ,  $P=0.082$ ;在女性( $n=346$ )中,冠心病组和正常对照组血清尿酸水平分别是  $(4.86 \pm 1.92)\text{mg/dl}$  和  $(4.66 \pm 1.73)\text{mg/dl}$ ,  $P=0.297$ ,均未见显著性差异。

在这个例子中,病例组是冠脉造影诊断“冠心病”的患者,对照组是来自同一个医院冠脉造影正常的人,比较病例组与对照组冠脉造影前 12 小时血清尿酸水平是否存在差别,以了解血清尿酸水平与冠心病的发生之间是否存在一定的关联。更多的此类研究是用优势比或比值比(OR)来表示高尿酸血症与冠心病之间的相关程度,即比较诊断“冠心病”与未诊断“冠心病”两组中高尿酸血症者与血清尿酸水平正常者之间的比值比( $OR_1$ ),或比较高尿酸血症组与血清尿酸水平正常组中诊断“冠心病”者与未诊断“冠心病”者之间的比值比( $OR_2$ ) (注意:在同一个研究人群中,  $OR_1=OR_2$ )。

这是一个比较典型的横断面研究(cross-sectional study),研究者同时对可疑危险因素和疾病进行了测量。需要特别指出的是,在横断面研究中,“冠心病”患者是现存患者(prevalent case)





而非新发患者(incident case),我们并不知道“冠心病”是何时出现的,血清尿酸水平或“高尿酸血症”也是目前的状况,我们同样不知道“高尿酸血症”是否出现在“冠心病”之前,更不知道“高尿酸血症”是否在“冠心病”出现之前已经持续作用了足够的时间,因此无法对二者之间的因果关系进行判断。横断面研究无法认定可疑危险因素与疾病发生之间的时间顺序,是其在病因学研究中最为严重的缺陷。

我们知道,性别是冠心病的危险因素,且男性的血清尿酸水平一般高于女性,即性别是尿酸的决定因素(危险因素)而不是相反,所以性别是研究高尿酸血症是否为冠心病危险因素的一个混杂因素,应该予以校正。分层分析是校正混杂因素的一种手段,在这个例子中,按性别分层以后再分析,就没有再看到冠心病患者和冠脉造影正常者的血清尿酸水平存在统计学差异。

2. 病例-对照研究 针对这个科学问题,我们也可以尝试设计病例-对照研究(case-control study)来回答,如病例组选择社区中新近临床诊断“冠心病”的患者,对照组是来自同一个社区临床诊断非“冠心病”的人,比较病例组与对照组若干时间以前高尿酸血症与血清尿酸水平正常的人之间的比值比(OR),以了解高尿酸血症与冠心病的发生之间是否存在一定的关联。

病例-对照研究的主要优点有:①非常适合少见病和慢性病的研究,因为病例-对照研究在工作伊始即收集病例,避免了队列研究随访多年而不出现结局的风险;②当疾病的发病率较低时,如果病例组是新发患者,对照组是来自同一个来源人群的健康人,则OR值可以比较准确地估计和替代RR值,但如果病例组是现存患者而不是新发患者,则不能用OR值来估计和替代RR值;③因为病例-对照研究是对病例组和对照组的暴露因素进行回溯,即在研究开始的时候结局已经出现,这大大缩短了研究的进程;④所需要的样本量较小;⑤研究成本较低。

病例-对照研究的主要缺点有:①不适合少见暴露的研究,因为在那种情况下,病例组和对照组可能都不曾暴露过,或是否有过暴露容易受偶然因素的影响;②对照组的选择容易偏离目标人群而出现选择性偏倚,这一点通常是病例-对照研究的技术难点之一;③关于暴露的信息容易因缺失或不准确而带来信息偏倚。

在此例中,不论新近诊断“冠心病”的患者(病例组)还是来自同一个社区诊断非“冠心病”的人(对照组),很多人可能都会因为没有测定或找不到检查结果而无法提供以前血清尿酸水平的情况,由此带来严重的信息偏倚,这是此例无法选用病例-对照研究的主要原因,实际上我们也未在文献数据库中检索到有关的病例-对照研究。

3. 队列研究 在Framingham心脏研究中,无心血管疾病的6763名当地居民按性别分层,测量基线(1971~1976年)血清尿酸水平,随诊117376人年(至1994年),共出现617例冠脉事件、429例冠心病死亡和1460例全因死亡。对年龄、BMI、糖尿病、收缩压、利尿剂、其他降压药、心电图左心室肥厚、总胆固醇水平、吸烟、饮酒以及是否绝经(仅限于女性)等因素进行校正后,男性血清尿酸水平每升高 $60\mu\text{mol/L}$ ,冠脉事件、冠心病死亡和全因死亡的风险比(HR)分别为0.91(95%CI:0.83~0.99)、0.95(95%CI:0.86~1.06)和0.97(95%CI:0.91~1.03),女性则分别为1.05(95%CI:0.94~1.17)、1.01(95%CI:0.90~1.13)和1.03(95%CI:0.97~1.09)。

在这个例子中,所有研究对象均来自同一个社区且在基线水平没有诊断“冠心病”,基线测量血清尿酸水平(可疑危险因素),在随访观察足够时间以后,比较基线可疑危险因素暴露程度不同者之间出现“冠心病”风险的比值,如可用发病率比(incidence rate ratio,IRR)、风险比(HR)或相对危险度(RR)等指标来表示,以了解血清尿酸与冠心病的发生之间是否存在一定的关联。

队列研究(cohort study)的主要优点有:①暴露因素和结局在时间上的先后顺序非常清楚,当二者存在关联时,对因果关系的判断比较有利;②非常适合少见暴露的研究,队列研究在工作伊始即收集暴露组,避免了病例-对照研究对暴露因素进行追溯时一无所获的风险;③可以比较准确地估算疾病发生的风险,并直接用相对危险度(RR)表示暴露因素与疾病发生之间的关联强度,而不必像病例-对照研究那样用OR去估计RR;④暴露组与非暴露组来自同一个来源人



群,除暴露因素以外,在许多方面都有很多相似性,可以有效减少病例-对照研究中常见的选择偏倚;⑤因为队列研究是对暴露因素和疾病发生进行追踪,所以测量比较完整、准确,可以有效减少病例-对照研究中常见的信息偏倚。

队列研究的主要缺点有:①不适合少见病和慢性病的研究,因为暴露组和非暴露组可能经过很长时间的追踪都不出现结局,从而导致研究失败;②疾病的发生往往需要较长的一段时间,队列研究的进程也相对较长;③所需要的样本量较大;④研究成本较高;⑤失访率过高会影响结论的真实性;⑥如果队列研究的时间跨度较大,而在此期间疾病的诊断标准发生了变化,则可能带来一定的诊断偏倚。

整体来讲,在观察性研究中,队列研究对建立暴露因素与疾病发生之间的因果关系力度最强,但时间成本和经济成本也是最高的。在此例中,对混杂因素进行校正之后,我们并没有看到血清尿酸水平与冠心病发病之间存在有意义的关联。当然,这个研究的结局指标是临床诊断的冠脉事件、冠心病死亡和全因死亡,这些情况往往是当冠心病发展到一定程度的时候才会出现,不能敏感地反映“冠心病”发病,因而可能给结果带来一定的误差。此外,基线血清尿酸水平实际只反映了可疑危险因素在一个时间点上的暴露水平,用它来代表研究全过程中血清尿酸的水平会带来一定的误差。如果我们选用前瞻性队列研究的方法来探索高尿酸血症与冠心病之间的因果关系,需要对上面提到的这些问题加以改进,如定期测量血清尿酸水平,采用更为敏感和准确的检测手段对“冠心病”发病进行判读等,有利于得到更加准确的结果,提供更加有力的证据支持或反驳前人的观点。

4. 随机对照试验 Waring 等人以 8 名 1 型糖尿病患者、8 名吸烟者和年龄匹配的 8 名健康人为研究对象(血清尿酸水平平均  $<400\mu\text{mol/L}$ ),进行随机安慰剂对照交叉试验,干预组静脉输注 1000mg 尿酸溶液,对照组静脉输注安慰剂,通过测量前臂血流反应(forearm blood flow responses),比较血管内皮功能的变化,结果发现静脉输注尿酸溶液能够提高 1 型糖尿病患者( $P<0.001$ )和吸烟者( $P<0.005$ )前臂血流对乙酰胆碱的反应,即提高了前臂血管床的内皮功能,但在正常人无效。

这是一个很有意思的试验,我们看到一次性静脉输注 1000mg 尿酸溶液能够改善某些具有冠心病危险因素者(1 型糖尿病患者、吸烟者)的血管内皮功能,但对正常人无效。这个试验对回答我们所关心的研究问题有一定的意义,但也存在一定的差别,冠心病的发生不是一朝一夕的事,尽管血管内皮功能异常可能是冠心病发病机制中较早的一环,但血管内皮功能对静脉输注尿酸溶液的即时反应是否与高尿酸血症对血管内皮功能的长期影响一致,目前也还缺少足够的科学证据。如果希望观察长期的效果,则静脉输注尿酸溶液十分不便而难以持久,且这样做很可能有失医学伦理的规范。

如果以长期静脉输注尿酸溶液为干预措施的随机对照试验(randomized controlled trial, RCT)不可行的话,一个替代的方法则是选择基线没有诊断“冠心病”的高尿酸血症者为研究对象,随机分为试验组和对照组,试验组服用降尿酸药物进行干预,而对照组服用安慰剂进行治疗,在随访观察足够一段时间以后,比较试验组与对照组出现“冠心病”风险之间的比值(RR),以了解高尿酸血症与冠心病的发生之间是否存在一定的因果关系。

Canbay 等人以 72 名高尿酸血症患者(无痛风、冠心病、心功能衰竭、糖尿病、高血压病史,不吸烟,不服用 ACE-I、ARB、他汀类降脂药和维生素,空腹血糖正常)为研究对象,进行随机对照试验,干预组( $n=32$ )连续 4 个月口服别嘌醇 300mg/d,对照组( $n=40$ )不接受任何治疗,通过测量血流介导的血管舒张反应(flow-mediated dilation, FMD),比较血管内皮功能的变化,结果发现干预组的 FMD 有明显改善( $P=0.003$ ),而对照组的 FMD 无明显变化( $P=0.52$ )。

在这个试验中,研究者特意将糖尿病、高血压、吸烟等排除在外,这是控制混杂因素的另外一个方法。一般来说,如果样本量较大,经随机分组后混杂因素会均匀地分布在干预组和对照



组中,从而消除混杂因素的影响,但在样本量较小的试验中,随机分组可能还不能充分地使其在两组中均衡,还要用排除或校正的方法去除混杂因素的干扰。

从上面两个例子,我们看出随机对照试验与前瞻性队列研究有许多共同点,但最主要区别在于随机对照试验对研究对象进行随机分组,而前瞻性队列研究根据暴露因素存在与否进行分组。随机对照试验和前瞻性队列研究相比,其最主要的优点是科学性强,即通过对研究对象的随机分组,不仅消除了试验组和对照组之间已知混杂因素的影响,而且消除了两组之间未知混杂因素的影响,这是其他类型的研究难以企及的,而最主要的缺点是可能面临医学伦理的挑战。

在上面的例子中,我们选择高尿酸血症人群进行这样的临床对照试验时,要求研究对象在基线没有心血管系统损伤,但可以有冠心病的其他危险因素(如糖尿病、吸烟等)。尽管从心血管研究的角度,我们可以选择已经有痛风或尿酸性肾病的高尿酸血症人群作为研究对象,但是从痛风或尿酸性肾病角度出发则存在较为严重的伦理问题,因为按照目前的临床指南和常规,多数痛风或尿酸性肾病患者都是需要接受降尿酸药物治疗的,所以研究对象选择仍以无症状性高尿酸血症患者为宜。此外,当随机对照试验的样本量较小时,混杂因素在干预组和对照组中的分布可能并不均衡,我们仍然需要采取一些措施(如排除、分层或校正等方法)消除混杂因素对研究的影响。

从上面的例子中我们还看到,两个试验均未选择真正的“冠心病”作为结局指标,这是因为冠心病的发病需要危险因素的长期作用,而受研究周期、人力、物力以及研究对象依从性等因素的影响,临床试验不可能过于长久。因此,临床试验需要选择一个合理的、可实现的结局,即需要考虑用比较“软”的指标来替代比较“硬”的结局。“硬”结局往往是指真正的结局,如出现冠脉事件等,获取这样的结局指标通常需要长时间和高投入;“软”结局往往不是真正的结局,而是疾病发生发展过程中的一些中间指标或标志物,与真正的结局密切相关,在危险因素的作用下能在较短的时间内出现。

在冠心病病因和危险因素研究中,人们曾经采用过许多不同的结局替代指标,如反映血管内皮炎症反应的C-反应蛋白,代表血管内皮细胞功能的前臂血流反应、血流介导的血管舒张反应,反映动脉僵硬程度的脉搏波传导速度,颈动脉内膜中层厚度或动脉粥样斑块等;即便采取较为直接的结局指标,也有不同的选择,如冠脉造影发现的动脉粥样斑块、临床冠脉事件、冠心病死亡等。不同的结局指标或结局替代指标,是疾病发生发展过程中不同阶段的标识,对于研究而言,其诊断疾病发生的敏感性和特异性不同,耗费的时间、人力、物力的成本也不同,需要根据研究问题和实际具备条件综合考虑进行选择。

## 第七节 计算样本量

下面我们以脉搏波传导速度(PWV)为例,说明“软”指标的选择及样本量的计算。PWV是指脉搏波在一定距离动脉节段上的传导速度。PWV取决于动脉壁的生物力学特性、血管直径、血管壁厚度以及血液的黏滞性等,与动脉僵硬程度密切相关。在各种心血管传统危险因素的作用下血管僵硬程度增加、弹性下降,脉搏波被血管壁吸收减少,因此脉搏波传导速度加快,PWV升高。PWV值越大,表示动脉弹性越差,僵硬程度越大;反之,动脉僵硬程度低,血管弹性好。PWV检测简单、无创,且具有较好的可重复性。荟萃分析显示,PWV作为反映动脉僵硬度的指标,可以预测未来心血管事件及死亡的风险。PWV作为“冠心病”随机对照试验临床观察结局的替代指标,能够有效缩短研究所需要的时间、降低投入的成本。

样本量是指在一定的条件下,如果希望研究结果出现显著性差异,则最少需要多少研究对象。样本量的计算在研究设计中有很重要的实用价值,通过计算样本量,我们才能够知道在所设定的条件下,能不能看到有意义的研究结果。比如根据前期研究结果,经过计算,发现要想使研究结果出现显著的统计学差异,需要很大的样本量,是实际研究所不能承受的,则需要及时修



改研究方案,使之贴合我们的实际。

样本量的大小受以下几个因素的影响:①效差(effect size):是指暴露组和非暴露组/试验组和对照组之间结局指标之间的差值,效差越大,所需样本量越小,效差越小,所需的样本量越大;②组内标准差(standard deviation):组内数据离散的程度,标准差越大,所需样本量越大,标准差越小,所需样本量越小;③检验效力(power):是指一个研究发现关联存在的能力,它等于1减去Ⅱ类错误的大小( $1-\beta$ ),检验效力越大,所需样本量越大,检验效力越小,所需的样本量也越小,检验效力一般取0.8或0.9(即将Ⅱ类错误控制在0.1或0.2);④统计学意义(statistical significance):Ⅰ类错误的大小( $\alpha$ ), $\alpha$ 值越小,所需样本量越大, $\alpha$ 值越大,所需样本量越小,Ⅰ类错误一般控制在0.05或0.01的水平。此外,研究方法不同、变量性质不同,样本量的计算方法也不同。

样本量可以通过公式进行计算,也可以通过一些软件获得结果,常见的统计软件一般都能够进行样本量的计算。这里介绍一个专门计算检验效力和样本量的小软件,下载地址是 <http://biostat.mc.vanderbilt.edu/wiki/Main/PowerSampleSize>,简单实用,供大家参考。

在本例的样本量计算时,我们需要对效差和组内标准差进行一定的估计,估计的根据主要来源于前人的研究(文献)、实践经验以及合理的推测等。比如,Kostka-Jeziorny 等研究发现,将66例轻中度高血压的患者随机分为两组,分别接受ACEI或利尿剂治疗,8周后两组均加用别嘌醇150mg/d,观察一段时间,两组加用别嘌醇后的PWV均明显下降,利尿剂组由( $10.7 \pm 1.4$ )m/s下降至( $10.0 \pm 1.2$ )m/s,ACEI组由( $11.5 \pm 1.7$ )m/s下降至( $10.4 \pm 1.5$ )m/s。

因此我们做出以下假设:在无症状高尿酸血症组中,加用别嘌醇300mg/d $\times$ 1年可以使其PWV下降幅度不低于Kostka-Jeziorny研究的ACEI组(PWV下降1.1m/s),对照组PWV因无特殊干预而无明显改善,应用上述软件进行计算,研究方法选非配对 $t$ 检验,效差是1.1,组内标准差是1.5,检验效力0.9,显著性水平0.05,可知试验组和对照组各需40例患者,可以达到90%的统计效力和 $P$ 值0.05的统计学差异。

总之,临床实践中遇到的疑问,经过提炼形成科学问题,结合实际情况,比较不同的研究方法,最终形成一个相对实用、完整的研究计划(表11-3)。按照计划组织力量进行研究,得到的最终结果有望能够对提出的科学问题有一个初步的解答。

表 11-3 研究计划概要

| 要素   | 内容                                  |
|------|-------------------------------------|
| 研究问题 | 高尿酸血症是冠心病的危险因素/病因吗                  |
| 研究方法 | 随机双盲安慰剂对照试验                         |
| 研究对象 | 社区无症状高尿酸血症患者                        |
| 研究参数 | 药物干预:别嘌醇300mg/d或安慰剂<br>结局指标:脉搏波传导速度 |
| 样本量  | 试验组和对照组各40例                         |
| 统计方法 | 非配对 $t$ 检验                          |

### 思考题

1. 在病因学研究中建立疾病因素模型和运用有向无环图的目的是什么?
2. 判断疾病病因/危险因素的主要原则有哪些,其具体含义是什么?
3. 可以用于病因学研究的方法有哪些,其各自的优缺点是什么?
4. 结合自己的临床工作选择一个疾病,运用所学知识设计一个比较科学、完整的临床研究方案,用以探索该疾病可能的病因/危险因素。

(方卫纲)



## 第十二章 诊断性试验评价研究

### 临床案例

患者男性,42岁,因口干、多饮和多尿就诊,无多食、体重明显下降。查体:体重指数(BMI)37.4kg/m<sup>2</sup>,腹型肥胖体型,BP 140/94mmHg。血生化检查发现空腹血糖 6.9mmol/L,糖化血红蛋白 6.6%。

### 第一节 临床研究问题的提出

随着全球经济的飞速发展和人类老龄化问题的日益加重,糖尿病逐渐成为危害人类生活质量和寿命的重要健康问题。糖尿病的诊断是根据血浆葡萄糖水平的测定确定,由于血浆葡萄糖的浓度是变化的,所以糖尿病诊断标准的设定是基于对糖尿病并发症的估计。评价血糖水平和糖尿病并发症的重点是糖尿病视网膜病变。大量的循证医学证据为制定糖尿病诊断标准提供了依据。虽然当空腹血糖 $\geq 7.0\text{mmol/L}$ 、随机或餐后2小时血糖 $\geq 11.1\text{mmol/L}$ 的患者经过重复血糖验证可以被诊断,但是对于血糖偏高未达到诊断标准,而临床上伴有或不伴有相关临床表现的患者需要通过2小时口服75g葡萄糖耐量试验(2小时OGTT)来诊断是否患有糖尿病。2小时OGTT是诊断糖尿病的金标准,具有重要的临床价值。但是OGTT操作的复杂性限制了其在临床上应用,特别是不适合用于糖尿病患者的筛查。

糖化血红蛋白是机体循环中红细胞内的血红蛋白与葡萄糖结合的产物,葡萄糖和血红蛋白结合生成糖化血红蛋白是不可逆反应,并与血糖浓度成正比。糖化血红蛋白通常可以反映患者近8~12周的血糖控制情况。糖化血红蛋白不受抽血时间、是否空腹、是否使用胰岛素等因素影响,因此糖化血红蛋白是国际公认的糖尿病监控“金标准”。糖化血红蛋白稳定、测定简单易行,该指标是否能够替代OGTT作为诊断糖尿病的指标呢?

该患者中年男性,因出现轻度多饮、多尿的症状就诊,查体发现有腹型肥胖和血压升高,实验室检查空腹血糖超过正常6.1mmol/L,但未达到糖尿病7mmol/L的诊断标准。而且未测定餐后2小时血糖,未行75g口服葡萄糖耐量试验(OGTT)这一诊断糖尿病的金标准。但患者糖化血红蛋白超过正常,是否能诊断糖尿病?由于OGTT试验需要口服葡萄糖溶液,并且需要2小时试验过程,多次抽血,临床上是否能够用糖化血红蛋白这一较为稳定的间接反映血糖水平的指标作为诊断标准。针对这一发病率接近10%的最常见内分泌代谢疾病,我们可以进行相关的临床研究,形成科学问题,中国人群中,糖化血红蛋白水平是否能够作为糖尿病诊断标准?以及糖化血红蛋白作为诊断指标的正常界值是多少?

#### 知识点

1. 诊断性试验 诊断性试验(diagnostic test)是对疾病进行诊断的试验方法,广义的诊断性试验既包括病史、查体等临床资料,也包括各种实验室检查、影像学检查和特殊器械检查结果。
2. 金标准 金标准(gold standard)即标准诊断(standard diagnostic test),是指当前临床



医生公认的诊断该疾病最可靠、准确的方法。拟评价或研究的诊断性试验在诊断疾病中的价值必须以金标准为根据进行“患病”或“无病”的判断。金标准不仅包括病理诊断,还有手术发现、影像学检查和公认的临床诊断标准等。

3. 受试者工作特征曲线 受试者工作特征曲线(receiver operator characteristic curve, ROC)是诊断性试验中常用于诊断界值判断的方法,同时也能够比较两种诊断性试验的诊断价值。制图时以诊断试验的敏感性(真阳性率)作为纵坐标,1-特异度(假阳性率)为横坐标,根据分层或分组测定的数据计算敏感度和1-特异度,分别在坐标上标出相应的点,最后描绘曲线。曲线上距离坐标图左上角最近的一点即为正常值的最佳界值。比较不同诊断试验的ROC曲线下面积可以用于判断不同诊断试验的价值。

4. Kappa值 分析计数资料时判断不同观察者间校正机遇一致率后观察一致率指标,即实际符合率和最大可能符合率之比。

5. 约登指数 约登指数(Youden index, YI)也称正确诊断指数,该指数常用来比较不同的诊断试验。 $YI=(\text{敏感度} + \text{特异度}) - 1$ ,用以判断某项诊断试验正确诊断患病或无病的能力。

## 第二节 诊断试验研究的设计思路

诊断试验的研究主要用于评价某种诊断方法(包括生化指标的测定、影像学检查、临床诊断标准等)在疾病诊断中与金标准相比,它的真实性、可靠性和临床应用价值。为保证研究科学性,研究设计更需要强调下面三个关键问题。

首先,诊断试验的研究需要确定金标准,被研究或评价的诊断试验必须和目前疾病诊断的金标准进行比较,才能应用相关指标来评价被研究的诊断试验的诊断价值。金标准,又称参考标准,是当前临床医学界公认的诊断该病最可靠的诊断方法。常用的金标准有:病理学标准、外科手术发现、特殊的影像诊断、长期临床随访结果、公认的综合临床诊断标准等。金标准是相对的,金标准的选择应结合临床具体情况决定,否则研究实施困难或者存在伦理问题。

其次,根据该诊断方法所针对的疾病以及容易混淆的其他疾病选择具有代表性的研究对象。研究对象要包括“无病”者和“患病”者,“无病”的对照组可以是无病的正常人,通常会选择在临床上和该病患者具有类似的症状、体征等需要鉴别诊断疾病的患者,而“患病组”除了选择典型的患者,还应当选择不同病情程度(轻、中和重)的患者,这样的诊断试验的结果更符合临床实际情况,具有更高的真实性和临床实用价值。

最后,是所有患者同时接受金标准检查和诊断试验检查并实施盲法评估和比较诊断结果。所研究的诊断性试验应与金标准进行独立的盲法比较。所研究的诊断试验判断某人“患病”还是“无病”时,不能受金标准检查结果的影响;同理,在金标准检查及判断结果时,亦不可受该诊断试验结果的影响。因为在某些情况下了解金标准试验的结果往往会影响到对被考核试验结果的解释。有时,当金标准检查结果模棱两可时,如果检查评定者知道新诊断试验的结果可能会有倾向性结果,从而引起偏倚。

结合本例临床科研问题,糖化血红蛋白是否能够作为糖尿病的诊断标准。首先应该确定本案例研究中要采用的糖尿病诊断金标准。目前国际上公认美国糖尿病学会(ADA)的糖尿病诊断标准,把空腹血糖、75g OGTT服糖后2小时血糖作为诊断标准,因此如果要评价糖化血红蛋白作为诊断试验诊断糖尿病的价值需要用75g OGTT作为金标准。入选的研究对象应该包括不同病情程度的糖尿病患者和无糖尿病者,可以是正常人,也可以是临床上存在多饮、多尿、消瘦症状需要和糖尿病鉴别诊断的患者,如甲状腺功能亢进的患者、尿崩症患者等。由于糖尿病是发病率较高的慢性代谢性疾病,因此本案例研究的患者入组可以在某医院的内分泌科门诊就诊的





患者。也可以在某个社区进行该项研究,入组所有非已知糖尿病患者进行 75g OGTT 的检查和糖化血红蛋白的测定。75g OGTT 的结果可以将受检对象分为糖尿病组、糖调节受损组(也称糖尿病前期,包括空腹血糖受损和糖耐量减低)、正常糖耐量组(非糖尿病组)。

有关糖化血红蛋白作为糖尿病诊断试验的研究已有发表。2011 年美国印第安纳州的研究者 Pinelli NR 等开展了相关的临床研究。该研究的主要研究目的是评价在阿拉伯裔美国人中使用糖化血红蛋白这一指标诊断糖尿病的价值。因此研究者在美国阿拉伯裔居住的聚集区底特律随机抽取 20~75 岁的非妊娠期阿拉伯裔成人(对阿拉伯裔有明确的定义),并且排除了 53 例已患糖尿病和贫血(贫血状态可能对糖化血红蛋白的测定准确性产生影响)的研究对象。所有的研究对象均接受空腹取血测定空腹血糖和糖化血红蛋白,然后进行标准 75g OGTT 试验。葡萄糖和糖化血红蛋白的测定均采用获得公认的标准测定方法。

通常在诊断试验评价的研究中样本量的计算需要请医学统计专业人员帮助,按照估计总体率的样本含量估算方法,分别计算“患病”组样本量  $n_1$  和“未患病”组样本量  $n_2$ 。即根据被评价诊断试验的灵敏度和特异度使用公式分别计算研究所需要的患者数和无病者数量。样本量的大小与特异度、灵敏度和显著性水平  $\alpha$  及允许误差  $\delta$  相关。通常取  $\alpha=0.05, \delta=0.05$  或  $0.1, \mu_\alpha=1.96$ 。

$$n=\mu_\alpha^2 p(1-p)/\delta^2$$

$n$  样本量;  $p$ : 预期敏感度或特异度

例如:血清甲胎蛋白诊断肝癌的敏感度约 50%,特异度约 75%,患者组和对照组各需要多少样本量?

$$n_1=1.96^2 \times 0.5 \times (1-0.5)/0.05^2 \approx 385$$

$$n_2=1.96^2 \times 0.75 \times (1-0.75)/0.05^2 \approx 289$$

即该诊断试验需肝癌患者 385 例,非肝癌患者 289 例。

### 第三节 诊断试验研究中正常参考界值的选择方法

诊断试验是用于临床医生判断患者是否患病,因此需要设定试验结果的正常与异常的界值,也称为参考值。由于诊断试验的结果通常在患者与无病者之间存在重叠和交叉,因此界值的确定会影响诊断试验的特异度和敏感度等。对于连续变量确定界值的方法有四种:① ROC 曲线:即受试者工作特征曲线,根据不同诊断试验的界值计算出一系列的敏感度和特异度,以“1- 特异度”为横坐标,敏感度为纵坐标,绘制出连续的曲线。ROC 曲线下面积体现了诊断试验的准确性,曲线下面积越大反映诊断试验的诊断效率越高。而 ROC 曲线的左上角的拐点是特异度和敏感度最优点,通常被选定为诊断试验的界值点;②测量值正态分布时,确定正常和异常的方法是 95% 的正常人测量值为正常区间,双侧常用“均数  $\pm 1.96$  标准差”表示;③百分位数法:诊断试验的测定值为偏态分布时,可采用百分位数界定正常与异常,双侧用  $P_{2.5} \sim P_{97.5}$ ;④临床判断:根据大规模人群调查确定试验的测定值达到什么水平需要干预。

本例临床研究糖化血红蛋白作为糖尿病的诊断试验的价值,糖化血红蛋白作为诊断指标它的界值可使用 ROC 曲线法,所有受试者都接受了金标准的判断:糖尿病、糖耐量异常或正常糖耐量,然后根据不同糖化血红蛋白的诊断界值可以绘制本人群的 ROC 曲线,将曲线的左上角的拐点作为糖化血红蛋白的诊断界值,这时该诊断试验的诊断效率最高。研究结果显示糖尿病的诊断界值为 6.2%,糖尿病前期的诊断界值为 5.1%。

### 第四节 诊断试验研究的方法和评价指标

诊断试验研究的方法主要是通过和金标准盲法比较,获得诊断试验的真实性、可靠性以及



临床应用价值的评价结果。绘制四格表是诊断试验研究结果表达的核心,是各种评价指标计算的基础(以糖化血红蛋白作为诊断糖尿病的研究作为实例)(表 12-1)。

表 12-1 诊断试验四格表

|                  |    | 金标准(OGTT) |            | 合计         |
|------------------|----|-----------|------------|------------|
|                  |    | 患病        | 无病         |            |
| 诊断试验<br>(糖化血红蛋白) | 阳性 | 真阳性 a(10) | 假阳性 b(2)   | a + b(12)  |
|                  | 阴性 | 假阴性 c(42) | 真阴性 d(428) | c + d(470) |
|                  | 合计 | a + c(52) | b + d(430) | N(482)     |

“( )”内为该临床诊断性试验研究中的患者例数

1. 敏感度 实际患病且诊断试验结果阳性的比例,敏感度越高代表漏诊率越低。

$$\text{Sen} = a / (a + c) = 10 / (10 + 42) = 19\%$$

2. 特异度 金标准诊断无病的例数中,诊断试验结果为阴性的比例,特异度越高代表误诊率越低。

$$\text{Spe} = d / (b + d) = 428 / (2 + 428) = 99.5\%$$

3. 准确度 诊断试验中的真阳性和真阴性占总检例数的比例。

$$\text{Acc} = (a + d) / N = (10 + 428) / 482 = 91\%$$

4. 阳性预测值 诊断试验结果阳性例数中,真正“患者”所占的比例,即诊断试验结果阳性中真正是“患者”的可能性。

$$+PV = a / (a + b) = 10 / (10 + 2) = 83\%$$

5. 阴性预测值 诊断试验结果阴性的例数中,真正“无病”所占的比例,即诊断结果阴性时真正“无病”的可能性。

$$-PV = d / (c + d) = 428 / (428 + 42) = 91\%$$

6. 阳性似然比 诊断试验的真阳性率与假阳性率的比值。诊断性试验结果阳性时患者患病和无病的概率比值,该比值越大说明患病的概率越大。

$$+LR = \text{Sen} / (1 - \text{Spe}) = 19\% / (1 - 99.5\%) = 38$$

7. 阴性似然比 诊断试验假阴性率和真阴性率的比值。诊断试验结果阴性时患者患病与无病的概率比值。

$$-LR = (1 - \text{Sen}) / \text{Spe} = (1 - 19\%) / 99.5\% = 0.8$$

本例临床研究显示在阿拉伯裔人群用糖化血红蛋白作为诊断糖尿病和糖尿病前期诊断标准时存在大量假阴性。

## 第五节 诊断试验的可靠性评价

可靠性即重复性,指诊断试验在相同条件下,重复试验获得结果的稳定性,也是诊断试验评价重要指标。计量资料的稳定性用标准差和变异系数(CV)表示, $CV = \text{标准差} / \text{平均值} \times 100\%$ 。标准差和 CV 值越小,说明该诊断试验的稳定性越好。而计数资料通常使用观察符合率与 Kappa 值表示,即两名观察者对同一事物的观察(观察者间)或同一观察者对同一事物的两次观察(观察者内)结果一致的百分率。Kappa 值是实际符合率与最大可能符合率的比值(以两名医生对同一组患者甲状腺进行超声检查诊断本甲状腺炎为例)(表 12-2)。



表 12-2 两名医生对同一组患者甲状腺进行超声检查诊断桥本甲状腺炎的四格表

| 甲医生诊断结果 | 乙医生诊断结果 |       | 合计    |
|---------|---------|-------|-------|
|         | 患病      | 无病    |       |
| 患病      | A       | B     | $R_1$ |
| 无病      | C       | D     | $R_2$ |
| 合计      | $C_1$   | $C_2$ | N     |

观察的符合率  $(p_o) = (A+D)/N \times 100\%$

机遇符合率  $(p_c) = (R_1C_1/N + R_2C_2/N)/N \times 100\%$

实际符合率 = 观察符合率 - 机遇符合率

最大可能符合率 =  $1 - \text{机遇符合率}$

Kappa 值 = 实际符合率 / 最大可能符合率 =  $(p_o - p_c) / (1 - p_c)$

## 第六节 诊断试验的评价原则

目前新的诊断试验层出不穷,无论是自己开展的诊断试验研究还是参考其他研究中心的诊断试验研究结果,都要进行科学的评价,才能客观的了解该诊断试验的诊断效率、诊断准确性以及临床实用价值等。通常遵循下述原则进行评价:

1. 诊断试验需采用盲法与金标准作比较 诊断试验必须和金标准进行比较,而且为了消除偏倚,应当采用盲法比较,得到客观科学的结论。
2. 纳入研究病例的分析和评价 诊断试验研究纳入的病例应包括轻、中、重各型患者,对照组要包括容易混淆临床上需要鉴别诊断的病例。
3. 叙述病例的来源和研究工作的安排 不同的病例来源对评价诊断试验的临床应用价值等有影响。
4. 诊断试验的重复性及其临床意义是否明确 临床诊断试验必须具有良好重复性即测定稳定性。而且应当明确叙述和评价诊断试验的临床诊断价值。
5. 诊断试验中使用的正常值是否合理、可靠 正常参考值的确定能够影响诊断试验的特异度和敏感度,应当采用科学合理的方法界定参考值,例如 ROC 曲线法使误诊率和漏诊率最低。
6. 在一系列试验中该诊断试验是否最正确 比较不同诊断试验的特异度、敏感度和准确度,选择最佳试验进行疾病诊断。
7. 诊断试验的操作步骤、注意事项和结果判断方法是否详尽,能否被重复 有价值的诊断试验研究应当就操作步骤和注意事项进行详细的描述,能够让同行进行类似诊断试验研究的重复。
8. 诊断试验的实用性如何 针对诊断试验的临床意义、真实性、稳定性等综合判断诊断试验的临床实用价值。

### 思考题

1. 诊断试验研究的原则是什么?
2. 诊断试验的研究中常用的评价指标有哪些?
3. 结合自己专业领域,选择 1~3 篇诊断试验研究相关文献,按照诊断试验评价原则进行逐条评价。

(刘晓清 朱慧娟)

## 第十三章 临床疗效评价

### 临床案例

女性,28岁。反复右上腹疼痛1年,疼痛向背部放射,伴恶心、呕吐,多在脂餐后出现,门诊腹部超声提示:胆总管0.8cm大小结石。普外科就诊后,建议外科手术取石或内镜下ERCP取石。1周后在内镜中心接受ERCP取石治疗,完整取出结石,术后出现急性胰腺炎,经治疗4天后痊愈出院。

### 第一节 如何获得临床研究问题

疗效分析和评价是临床科研工作的一个重要组成部分,包括某种措施对疾病的预防效果和治疗效果。临床工作的最终目标是治愈疾病,让患者最大限度获益。一般临床医生常常根据自己的经验积累做出治疗方法选择,而这些经验主要来源于对个别案例的认识,其中可能有偶然机遇和偏倚。为了提高临床工作的可靠性和科学性,对患者采取切实有效的治疗措施,临床医生需要经常进行疗效分析和评价。

上述案例中对患者采取的治疗措施是临床上较为常见的处理方案,患者有腹痛症状和明确的胆总管结石,有治疗指征。目前针对这类患者的处理有不同的看法,是采用外科手术还是ERCP治疗?如何预防ERCP术后并发症?针对这一临床问题,可以进行很多相关的研究。

#### 一、明确临床问题

1. 查阅文献及相关背景资料 从20世纪70年代开展ERCP以来,其应用价值已经从单纯的胆胰疾病诊断发展到治疗领域。尤其在治疗胆总管结石、胆源性胰腺炎、无法外科切除的胆道及胰腺肿瘤等方面都具有肯定的价值。但是,ERCP术后并发症包括:术后胰腺炎、出血、穿孔及胆道感染等,其中ERCP术后胰腺炎是最常见的并发症。在临床上选用某种治疗手段时,有两个十分重要的关注点,即有效性和安全性。如果能找到这样一种方案:既能利用ERCP治疗胆总管结石的优点,又能预防ERCP相关并发症,无疑有重要的临床意义。

2. 相关研究提供临床思路 急性胰腺炎是内科常见疾病之一,ERCP术后胰腺炎(post-ERCP pancreatitis, PEP)与一般急性胰腺炎发病机制方面有没有相似之处呢?急性胰腺炎发病机制中有一个十分重要的环节,即胰管压力增加导致的胰酶激活及相关炎症介质释放。以此类推,ERCP术后如果导致十二指肠乳头水肿及胰液排泌不畅,一样可以导致急性胰腺炎发生。那么,如果能避免胰液排出不畅或抑制炎症相关因子释放可能会预防ERCP术后胰腺炎。

3. 已有的临床研究证据 已有的研究表明,PEP发生主要与下列因素有关:①患者因素:青年女性、Oddi括约肌功能障碍、既往有PEP发生;②操作者因素:操作不熟练、反复胰管插管或注射造影剂、预切开等。预防PEP措施,包括放置胰管支架、药物(生长抑素、乌司他丁、硝酸甘油及非甾体类抗炎药等)。在这些预防方法中,我们不可能对每种方法都进行研究,需要了解目前哪些方法是可能有效,哪些方法是无效,哪些方法是研究不深入、尚无定论的。



## 知识点

## 疗效评价的定义、范围及伦理

1. 疗效评价的定义 疗效评价是对临床治疗效应所产生的效能和效力,按已确定的标准进行定性、定量和综合判断的过程。疗效评价应该包括5个关键环节,评价主体即评价什么,谁来评价即评价者是谁,评价标准即评价尺度如何制订,怎样评价即评价技术与方法和对评价结果的解析。
2. 疗效评价的范围 药物、手术、预防措施、治疗方案(如肿瘤化疗)、特定形式的治疗单元的评价(如监护病房的作用)。
3. 临床疗效评价的伦理特征 临床疗效评价研究是以人体为对象的医学研究,应遵从赫尔辛基宣言的道德原则。其宗旨是在涉及人体对象的医学研究中,应优先考虑人体对象的健康幸福,其次考虑科学和社会的利益。

## 二、临床问题转换成科学问题

针对上述临床问题,我们可以通过不同的疗效评价来回答,这就需要将临床问题转换为科学问题。

转换思路1: PEP与操作本身的关系,我们可以转化为这样的思维:熟练程度高的ERCP操作医生治疗的患者是否能减少PEP发生?这就需要前瞻性随机对照研究设计,将操作医生按操作成功例数分为熟练组和非熟练组,比较他们操作相同例数ERCP发生PEP的比例,明确能否通过提高操作技术预防PEP。

这样的研究尽管比较容易设计,但对于患者利益最大化原则这一伦理而言,那些被随机分配到非熟练医生组承担的PEP风险更大。而且,操作者熟练程度的提高也并非短时间能够实现。因此,这种RCT方案并不理想。非随机回顾性疗效观察比较研究设计可能更适合回答这一问题,但需要在熟练与非熟练组间具有可比性。

转换思路2: 我们也可以转化为这样的问题,PEP与胰管压力增高有关,在完成治疗后如果能想办法降低胰管压力,能否预防PEP发生?目前已有的办法是放置胰管支架,同样可以设计随机对照临床研究(RCT),比较放置支架组和不放置支架组PEP发生率,从而明确放置胰管支架能否预防PEP发生。

这样的设计尽管比较理想,在临床上的效果也可以得到证实,但在具体实施过程中仍然存在一定问题。主要是PEP本身发生率低,放置胰管支架增加费用,同时放置胰管支架后,患者需要再次接受ERCP取出胰管支架,除了费用之外又增加了患者痛苦并降低了依从性。基于PEP低发生率问题,在PEP发生的高危人群中进行这项RCT研究可能更合适。

转换思路3: 我们同样可以从另外一个角度考虑问题,能否从胰腺炎发病机制方面通过药物预防PEP发生?对于PEP高风险因素患者随机分组,观察组患者术前接受预防性药物,对照组患者术前接受安慰剂,观察并比较两组患者PEP发生率。这样的设计花费时间更少、成本低、患者接受程度高的预防PEP疗效评价研究。前提是预防性药物方便给药、副作用少、费用不高。

## 知识点

## 疗效评价设计的原则

1. 基于研究目的,确定观察人群,明确入选标准和排除标准。
2. 正确设立研究组和对照组并进行随机化分组。
3. 制订干预措施、步骤、时间、终止治疗原则,最好采用盲法原则。



## 知识点

## 临床疗效评价的设计思想

1. 临床疗效评价属于临床干预研究,其目的是确定一种干预措施是否能改变疾病的自然病程,或者是否可以提高目前有效治疗方法的疗效,或与其他治疗相比有什么不可接受的副作用。
2. 临床试验(clinical trial) 是以临床患者为研究对象随机分为实验组与对照组,实验组给予新药或新疗法,对照组给予标准(经典)疗法或安慰剂或不给予任何措施,追踪并比较治疗组与对照组的結果而确定某项治疗或预防措施的效果与毒副作用的一种前瞻性研究。
3. 临床疗效评价的基本研究方案 随机对照临床试验(安慰剂、空白对照)、非随机同期对照研究、历史对照研究、自身前后对照研究等。
4. 临床疗效评价的三要素 设立对照、随机分组、盲法评定。

## 第二节 临床疗效评价的基本研究方案

1. 随机对照试验 随机对照试验(randomized controlled trial, RCT)用于临床疗效评价时大致有两种情况:①用于新疗法与标准疗法的比较,应用的前提是目前尚不能肯定新疗法的疗效比旧疗法好。对照组患者及试验组患者均接受治疗,通常要包括较长时间的随访期;②用于暂且不治疗不影响预后的疾病,通常将试验对象随机分为对照组及试验组,对照组采用安慰剂或空白对照。对于这类疾病,一定是目前尚无一种肯定疗效的方法,采用新疗法患者可能受益,也可能有一定危害,通常观察时间不宜过长。

2. 非随机同期对照试验 非随机同期对照试验(non-randomized concurrent control trial)是临床传统采用的一种研究设计,试验组和对照组不是采用随机的方法分组,而是由患者或医生根据患者病情及相关因素人为的纳入试验组或对照组,并进行同期的对照试验。在临床实际工作中,有些情况下不适宜做随机对照试验,例如外科手术治疗、急重症患者抢救和贵重药物的选择等。因此,只能根据具体情况将患者分为试验组及对照组。其优点是研究方案的可行性好,易被临床医生和患者接受,缺点是由于选择性偏倚和测量性偏倚,导致其研究结果的证据不如RCT。在尚无RCT结果或不能获得RCT结果时,还是应予重视,尤其是对于样本量大的非随机同期对照试验研究,仍然有重要价值。

3. 历史对照研究 历史对照研究(historical control study)又称为非同期对照,也属于非随机对照试验。该方法是将现时给予干预措施患者的临床结果与既往未给予该干预措施的另一组同种疾病患者的结果进行比较,以评价该干预措施的疗效。作为历史对照的患者,或是没有进行治疗,或是只接受了常规治疗。历史对照因省时、省力、容易得出结果,临床应用较多。需要强调的是,两组之间要具有可比性,除了治疗因素外其他影响结果的因素在两组之间应尽可能的相似。由于历史性对照研究是非随机、非同时的对照研究,事先未经过严密的科研设计,产生各种偏倚的机会较多。

4. 自身前后对照研究 自身前后对照研究(before-after study in the same patient)是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的临床研究。受试者在接受第二阶段试验时,需要经过一定的洗脱期。优点:①每例受试者均以自身为对照,可以消除个体差异,减少样本含量,节约时间和成本;②研究过程中,每一例患者接受措施均等,具有公平性;③减少了偏倚,可以实现试验措施的标准化;④可以采用盲法,结果可信度高。缺点:①本研究方法分为前、后两个阶段,很难保证两阶段的起始点完全一致,可能影响两阶段的可比性;②试验的应用范围





有限,只适用于慢性复发性疾病;③如洗脱期过长,可能影响患者的及时治疗。

### 第三节 临床疗效评价设计(随机对照试验)

#### 一、研究目的

评价非甾体类抗炎药物(NSAID)对 PEP 的预防效果。

#### 二、研究假设

非甾体类抗炎药能减少 PEP 发生,对 PEP 具有预防作用。

#### 三、病例的选择

病例选择:①有临床指征接受 ERCP 治疗患者:需要符合 ERCP 治疗适应证,如胆总管结石、急性胆管炎、SOD、胆道蛔虫、胆道梗阻等。与此同时,患者不愿接受外科手术或无外科手术指征;②纳入对象的来源包括:单中心来源或多中心来源,他们各有优缺点。前者病例来自同一医院或研究中心,优点是具有很好的同质性,治疗方案容易实施,ERCP 操作医生相对固定。缺点是受病源数量的影响,可能需要花费较长时间才能完成足够的样本量。

多中心试验由一个或几个单位的主要研究者总负责,多个单位的研究者合作,按同一方案进行的临床试验。优点是规模大、耗时短,选择偏倚小,适用面广;广泛合作提高研究水平。但需要注意:强有力的领导,充足的财物支撑;严密组织,分工明确;充分合作,协调统一;应答率 >85%~90%。确定来源之后,接受 ERCP 治疗人群中具体入选标准和排除标准成为选择的第二个问题:是否所有准备接受 ERCP 治疗的患者、不论何种疾病包括?是否限制年龄与性别?有药物过敏史者是否纳入?因此,设计时还应规定排除标准,即有些患者虽符合诊断标准,但仍不能入选作为研究对象。首先,该患者同时患另一种可影响本试验效果的疾病时,就不宜选作研究对象。其次,选中的患者也不宜同时患其他病情凶险的疾病,因为这样的患者可能在研究过程中死亡或因病情恶化而被迫退出。最后,已知研究对象对药物有不良反应时也不应将之选入。

案例病例选择 1:假设我们将符合 ERCP 指征的患者均纳入研究,获得的病例数自然较多。在获得知情同意后,将他们随机分为治疗组和对照组,分别接受 NSAID 预防用药或给予空白对照,观察 PEP 的发生率,进行统计学分析并得出结论。这样,我们能够比较全面、客观地评价 NSAID 是否能预防 PEP 发生。但是,PEP 的发生率很低,同时在不同患者之间存在差异,如果忽略了这一重要因素,得到的研究结果就不能反映 NSAID 药物的真实效果。

案例病例选择 2:我们也可以换一个思维角度,PEP 更可能在部分高危患者中发生。通过查阅文献不难发现,PEP 发生的高危人群包括:①临床怀疑括约肌痉挛(SOD)患者;②既往有过 PEP 发生;③括约肌预切开;④反复插管(>8 次)不成功;⑤年轻女性患者。那么,是否可以只在 PEP 高风险人群中进行研究?从理论和实践上,这一设计思路是完全可行的,如果预期能降低这些高风险人群 PEP 发生率,就能说明 NSAID 能预防 PEP,而且研究需要的样本量明显减少。

#### 四、试验分组

我们选取符合纳入标准的病例后,即可按照临床试验设计的基本原则进行随机化分组,设立对照组和试验组。设立对照的意义一方面是排除疾病自然变化和非处理因素的干扰,另一方面鉴定处理因素的作用和效果。设立对照的基本原则包括:

1. 一致性原则 对照组与实验组之间必须具有可比性,要求除处理因素外,其他条件在理



论是应完全一致。在临床试验中,实际上做不到完全一致,但应在防止和减少主要混杂因素的影响上达到一致。组间样本的年龄、性别、职业、出生地、烟酒史等应一致;病情特点、预后因素以及同时接受的其他治疗措施相同;试验条件如仪器、试剂及操作人员和方法等都应相同,不能一组用一套;组间尽量同步进行,以消除时间因素造成的影响。

2. 对等原则 从统计学角度考虑,组间人数相等的组间合并误差最小,因此,对照组的例数应与实验组例数相等,一般对照组例数不应少于实验组例数。本案例研究中,PEP 高风险人群随机分配到试验组和对照组,随机是进行选择或治疗安排的过程。在此过程中,伴随每个合理的结果都有一个已知概率。在同一随机策略下,意味着每个受试者被安排到 NSAID 预防组或对照组的机会是相等的。最简单、最普通的设计是双盲试验,即一半受试者被安排接受某种治疗或预防措施,另一半给予其他治疗或安慰剂。

### 知识点

#### 盲法的应用

任何临床试验都希望结果的真实性不受外来因素的干扰。疗效判定者、治疗医生、患者在了解试验目的和方法后都可能各种因素干扰疗效的评价。

1. 单盲法 一般是患者一方不知道,医生是知道的。
2. 双盲法 评定疗效者和患者双方都不知道。
3. 三盲法 评定疗效者、治疗医生和患者三方都不知道。

## 五、干预方法

临床试验中,选择恰当的干预方法同样十分重要。比如,药物干预试验有静脉给药、皮下给药、口服给药、经黏膜给药等,选择时一方面要结合干预药物本身的特性及有效性,另一方面也要具体分析在此试验中何种途径最恰当。在本案例研究中,我们首先了解 NSAID 类药物可以口服或经肛门直肠给药,那么哪种途径更适合本项试验呢?这就需要相关知识,一是 NSAID 药物经直肠吸收更快、生物利用度更高,二是接受 ERCP 治疗患者术前、术后均需要禁食,因此口服给药不如经直肠给药好。结合上述背景知识,我们就比较容易做出选择,经直肠给药方便易行,患者依从性高。对照组我们可以采用安慰剂或空白对照,安慰剂的给予途径应该与试验组一致。

### 知识点

#### 安慰剂

1. 安慰剂是一种惰性或无药理作用的物质,对人体无害,如乳糖、淀粉、生理盐水等,也可以用对疾病没有直接影响的药物,如维生素、开胃药等。要求在外形、颜色、大小、制剂等方面与试验药物尽量做到一致。
2. 使用安慰剂的好处是可以防止可能产生的心理或主观因素干扰试验结果。其不足之处是对照组可能没有得到相应治疗,涉及伦理问题。因此,适于轻症患者。
3. 避免伦理问题可采用试验药 + 常规治疗与安慰剂 + 常规治疗的研究设计。

## 六、评价指标

本临床试验中,首先要定义 PEP,即有客观标准。国际指南对于 PEP 的标准是:ERCP 术后患者出现腹痛,同时伴有血淀粉酶/脂肪酶增高超过正常值上限 3 倍。需要明确的是,相当一部分患者 ERCP 术后有血清淀粉酶升高,如果没有胰腺炎的临床表现,称作 ERCP 术后高淀粉酶



血症,而不是 PEP。在临床试验中,定义主要结果和数个次要结果非常普遍。前者基于样本量和招募策略,而研究者对后者可能也很感兴趣,但不代表首要的研究问题。本试验中,患者是否发生术后胰腺炎是研究的主要结果,而其他问题同样值得关注。比如,两组患者胰腺炎的轻重程度、患者住院时间、有无消化道出血等并发症等。就 NSAID 而言,因为有消化道出血的副作用,如果试验组消化道出血的风险明显增加,即使它有预防 PEP 的作用,也应该慎用或不用。

### 知识点

#### 效应指标类型

1. 客观指标 如痊愈、病残、死亡等,疗效评价性研究应根据情况尽量采用这类客观指标。另外由实验或仪器测定的指标也是常用指标,但应注意实验或测量的条件一致和人员培训及质控工作。
2. 患者主观描述性指标 关节痛、头痛、乏力、腹胀等指标,不确切而且可靠性差,一般应谨慎选用。
3. 按反应的性质,效应指标大致分为定性的计数指标和定量的计量指标,前者如阳性、阴性;痊愈、好转、无效、恶化;X 线片变化、细胞坏死程度、症状是否出现等。后者如身高、体重、血压、体温、细胞计数等。

### 知识点

#### 效应指标选择原则

1. 客观性 定量指标一般以客观记录为主,定性指标也应尽量用客观方法记录,避免主观心理因素造成偏倚。
2. 特异性 要与研究目的密切相关,能确切反映处理因素的效应,因而要尽可能选用特异性指标,以防止非处理因素的干扰。
3. 灵敏性 对处理因素要能灵敏地反应出来。由于医学实验方法日新月异,故应根据专业知识、研究目的和要求,选用新的灵敏指标。
4. 稳定性 任何实验指标都要求稳定性好,能被不同时间、地点和操作者重复证实,误差应在允许范围之内。
5. 结局性 尽可能选用结局性指标,少用中间性指标。

## 第四节 研究报告与问题分析

### 研究案例

吡哌美辛预防 PEP 的随机对照研究[N Engl J Med, 2012, 366(15): 1414-1422]。

1. 背景 以往的研究提示非甾体类抗炎药能预防 PEP 发生。

方法:采用多中心、随机、安慰剂对照双盲试验,对 PEP 高危人群分别在 ERCP 术后立即直肠给予吡哌美辛或安慰剂。高危患者判定依据患者情况及操作相关因素。主要结果为 PEP,其标准为 ERCP 术后 24 小时内新出现的上腹痛,且血淀粉酶超过正常值高限 3 倍以上,住院时间超过 2 天。

2. 结果 共纳入 602 例患者且完成了观察。绝大部分(82%)患者临床考虑有 Oddi 括约肌功能障碍。吡哌美辛组 295 例患者中有 27 例(9.2%)发生 PEP,安慰剂组 307 例患者中有 52 例



(16.9%)发生 PEP( $P=0.005$ )。吡哌美辛组中~重症胰腺炎 14 例(4.4%),安慰剂组 27 例(8.8%)发展为中~重症胰腺炎( $P=0.03$ )。

3. 结论 PEP 高危患者中,经直肠给予吡哌美辛栓剂明显降低 PEP 的发生。

4. 评价 研究案例采用的多中心、随机、安慰剂对照双盲试验,样本纳入量大,试验设计可谓完美。但试验过程中是否仍然存在一些问题呢?仔细阅读全文不难发现,纳入对象中 80% 以上患者 ERCP 术后放置了胰管支架,该措施本身具有预防 PEP 的作用。因此,有必要重新审视和分析研究数据。分析发现:57 例仅仅接受安慰剂处理的患者中 12 例(21%)发生了 PEP,250 例放置胰管支架的安慰剂组中 40 例(16%)发生 PEP,49 例单纯接受吡哌美辛患者中 3 例(6.1%)发生 PEP,胰管支架联合吡哌美辛组 246 例患者中 24 例(9.7%)发生 PEP。虽然统计学分析仍然发现吡哌美辛具有预防 PEP 作用,但有必要进行胰管支架置入与吡哌美辛两者单独在预防 PEP 中的作用。另外,吡哌美辛的达峰时间为 30~60 分钟,而 ERCP 操作时间有长有短,到底是 ERCP 术前给药还是术后给药更好也值得探讨。

### 知识点

#### 临床试验中的伦理问题

临床疗效评价或预防效果评价均属于临床试验范畴,不可避免地涉及医学伦理学问题。综合各种伦理法规、指导方针和文献,人们提出了可以继续应用于所有临床研究的系统性伦理学原则的框架。根据这个框架,临床研究必须同时满足以下要求才是符合伦理的:①具有社会或科学价值;②研究的正确性;③公平选择受试者;④良好的风险受益比;⑤独立审查;⑥知情同意;⑦尊重受试者意愿。

RCT 试验需要注意以下问题:①选择对照组不仅是一个简单的科学上的需要,使用安慰剂作为对照需要经过伦理上的论证;②随机分配不允许自主选择;③受试者同意暂不知晓所接受的干预方法,在处理一些临床问题时,盲法研究可能需要破盲;④研究监察员、独立数据和安全监察委员会有助于决定何时终止研究或终止与受试者的信息分享。

## 第五节 文献阅读与评价

依维莫司对进展期胰腺内分泌肿瘤的作用[N Engl J Med, 2011, 364(6):514-523]

1. 目的 依维莫司是一种口服哺乳动物靶向雷帕霉素(mTOR)抑制剂,2 期临床试验显示其对进展期胰腺内分泌肿瘤具有抗肿瘤作用。本研究通过前瞻性、随机对照试验对其疗效进行 3 期临床试验。

2. 方法 在过去 12 个月纳入 410 例影像学表现为进展期、低度或中度胰腺神经内分泌肿瘤患者,随机分为依维莫司治疗组(207 例,依维莫司 10mg,1 次/日)及安慰剂组(203 例),所有患者均一同接受相同支持治疗。主要终点为无肿瘤进展生存,采用意向治疗分析。在研究期间,如患者有影像学进展表现应揭盲,如果该患者之前是被随机分配到安慰剂组,则将其提到依维莫司开放组。

3. 结果 依维莫司组中位无进展生存期为 11.0 个月,安慰剂组为 4.6 个月(疾病进展或死亡危险比在依维莫司组为 0.35,95% 可信区间),表示降低了 65% 的病情恶化或死亡风险。依维莫司组无进展生存达 18 个月的患者为 34%,而安慰剂组为 9%。药物相关副作用主要为 1 级或 2 级,包括口腔炎(依维莫司组 64%,安慰剂组 17%);皮疹(49% vs. 10%);腹泻(34% vs. 10%);乏力(31% vs. 14%);上呼吸道感染(23% vs. 6%)。药物相关副作用 3~4 级在依维莫司组比安慰剂组发生率更高,包括:贫血(6% vs. 0%)、高血糖(5% vs. 2%)。依维莫司组中位暴露时间是安慰剂



组的 2.3 倍(38 周 vs.16 周)。

4. 结论 依维莫司明显延长进展期胰腺内分泌肿瘤患者肿瘤无进展生存时间,且较少发生严重的副作用。

### 思考题

1. 针对上述阅读文献(摘要,可以根据需要阅读全文),回答:该研究期望解决的临床问题是什么?科学问题是如何描述的?该研究采用了何种临床研究设计?患者纳入与排除标准是什么?是否存在选择偏倚?是否采用随机化分组?对照组采用安慰剂是否存在伦理问题?如何克服?其他支持治疗是否影响两组疗效?观察终点如何确定?合理吗?如何评估该研究的科学性、重要性和实用价值。

2. 结合自己专业,选择一种干预措施进行评价。设计一项疗效评价研究,内容包括:科学问题与研究目的,研究设计与具体方法,结果测量方法及可能的结果与比较。

(王春晖)

## 第十四章 预后研究

### 临床案例

一名 32 岁女性,因扪及右乳肿块就诊。患者已婚未育,既往月经周期正常规律。入院后查体示右乳 9 点钟方向可见 1cm×1.5cm 肿块,质硬、边界清,活动可,无压痛、无乳头凹陷、无橘皮样变,左乳正常。乳腺钼靶示右乳占位。患者行右乳改良根治术,病理示:(右乳)浸润性导管癌(Ⅱ~Ⅲ级),神经与脉管未见癌累及;右腋窝下淋巴结清扫 20 枚,其中未见癌转移。免疫组化:瘤细胞:ER(+)(>80%,强),PR(+)(>80%,强),Her2/neu(++),Ki-67(+,约 20%)。FISH 法检测 *Her-2* 基因无扩增。一般临床问题是,该患者预后如何?哪些因素会影响患者预后?是否需要接受化疗或者靶向治疗等进一步处理?通常指南可以回答这一类问题,但患者年轻未育是其特殊性,就需要特殊研究证据回答预后问题。

### 第一节 预后研究的基本概念及步骤

对于这样一个年轻尚未生育的早期乳腺癌病例,患者及家属可能特别关心的问题包括:这种疾病是否会复发?大概何时复发?日后能否怀孕及生育?妊娠会不会增加乳腺癌复发的概率?实际上,他们提出的仍然是一个有关乳腺癌预后以及影响预后因素的问题,即乳腺癌复发率如何?妊娠对乳腺癌预后有无影响等。

预后(prognosis)是指疾病发生后,对疾病未来病程和结局(痊愈、复发、恶化、伤残、并发症和死亡等)的预测或估计。预后研究是疾病自然病程研究的一部分。疾病自然病程是研究疾病的发生、发展及转归,而预后研究只研究疾病被确立诊断后的临床进程和影响这一过程中的诸多因素。疾病预后的研究指对疾病各种结局发生概率及其影响因素的研究,至少包括率的估计和研究探讨影响预后的因素这两个方面。率的估计包括并发症发生率、生存率、治愈率、复发率、死亡率等。对于病程短,可治愈的疾病,预后指标多用治愈率;对于病程长、不易治愈的疾病,多用复发率、缓解率;而对于严重疾病,多用病死率(或存活率)、致残率等;评价疾病防治的近期效果,可用治愈率、有效率及病死率。研究肿瘤或其他预后较差疾病的治疗效果时,需要长期随访观察,多用生存率来评价。慢性病常增加生活质量评价。而影响预后的因素是指根据疾病不同亚型和一些临床指标,判断疾病的预后,以及筛选影响预后的指标。

#### 知识点

##### 疾病预后的判断指标

1. 病死率 病死率(case-fatality rate)指某病患者中因该病而死亡的频率。一般用于在短期内可见结局的疾病。表明疾病的严重程度,也可反映医疗水平和处理能力。

病死率=(某时期内因某病死亡人数/同期患某病的病人数)×100%

2. 致残率 致残率(disability rate)指发生肢体或器官功能丧失者占观察患者总数的百分比。





3. 治愈率 治愈率(cure rate)指患某病治愈的人数占该病接受治疗患者总数的百分比。

治愈率 = 患某病治愈的患者人数 / 患该病接受治疗的患者总人数 ( $\times 100\%$ )

4. 缓解率 缓解率(remission rate)指给予某种治疗后,进入疾病临床消失期的病例数占总治疗例数的百分比。

缓解率 = 治疗后进入疾病临床消失期的病例数 / 接受该治疗的总病例数 ( $\times 100\%$ )

5. 复发率 复发率(recurrence rate)指疾病经过一定的缓解或痊愈后又重复发作的患者数占接受观察患者总数的百分比。

复发率 = 复发的患者数 / 接受观察的患者总数 ( $\times 100\%$ )

6. 生存率 生存率(survival rate)是指从疾病临床过程的某一点(如发病、确诊、开始治疗或手术时间等)开始,到某时点的生存概率。

$n$  年生存率( ${}_n p_0$ ) = 活满  $n$  年的总病例数 /  $n$  年观察的总病例数 ( $\times 100\%$ )

$p$  为生存率,  $n$  为随访时间长度,一般以年为单位,0 为观察起始点,病程较短的癌症可用 1 年生存率,一般癌症可用 5 年生存率。

医师和患者都希望对所患疾病将来的情况做出客观的估计与判断,尽可能使预测结果接近正式的结局,这就要求医师在回答患者问题时将专业知识、患者疾病特征和预后文献提供的科学研究证据,进行综合性分析。在疾病预后判断的循证实践中,包括发现问题、寻找证据、判断证据的真实性、估计证据的重要性和结合具体病例进行相应的临床实践等 5 个主要环节。

### 知识点

#### 预后的循证步骤

1. 发现临床预后问题,并转化为可以检索的、易于回答的问题。
2. 根据预后问题找出相关的预后研究文献。
3. 评价文献,确定预后证据的真实性。
4. 估计预后研究结果的重要性。
5. 估计研究结果的适用性。

针对这样一个具体的乳腺癌病例,我们首先必须把患者提出的临床问题转化为可以检索(即包含关键词)并易于回答的问题,需兼顾临床预后问题所涉及的各个方面。如与不妊娠相比,妊娠的乳腺癌患者复发率有无差别?妊娠是否为乳腺癌复发的危险因素?

### 知识点

#### 临床预后问题涉及的方面

1. 疾病会发生何种结果 定性。
2. 发生不良结局的可能性有多大 确定概率。
3. 何时会发生不良结局 定时。
4. 疾病不良结局受哪些因素影响 筛选出影响预后的指标。

## 第二节 预后研究常用的临床研究设计

预后研究常用的临床研究设计包括队列研究(前瞻性和回顾性)、病例-对照研究、纵向描述



性研究、系列病例分析、专家意见和个案报道。不同研究设计所提供的预后研究证据质量水平完全不同。因此,在获取预后研究证据及应用证据前,掌握预后研究的不同设计方法及评估不同方法的优缺点对判断研究结果的真实性非常重要。

## 一、队列研究

队列研究是研究者将一组或多组没有经历过相关结局,但所有的人都有可能发生其结局的人群进行一段时间的随访,并对其组间结果进行比较。队列研究由病因(暴露、非暴露)推导结局(发病率、治愈率、缓解率、复发率和生存率等),可分为前瞻性队列研究和历史回顾性队列研究,是预后研究主要的研究策略。前瞻性队列研究指研究开始时,暴露因素已存在,但结局尚未发生,在随访过程之中,可获取暴露因素变动情况。而历史回顾性队列研究指研究开始时,暴露与结局均已发生,即研究结局在研究开始时已经发生,从而探讨过去的暴露因素与目前已发生某种疾病结局之间的因果关系;历史资料的完整性和真实性将影响研究的可行性和结果可靠性。

运用队列研究进行疾病预后因素分析时,首先要确定患者是否存在某因素(暴露与否),然后随访一段时间,再确定患者是否出现阳性结局。因此,首先要确定研究因素和结局因素,并记录随访的时间,这是队列研究的三大基本要素。

确定研究因素(暴露因素)是预后因素研究的主要内容,暴露因素必须有明确规定(性质、时间、频率、强度、方式)。最好将暴露定量,明确划分暴露与非暴露人群的界限以及暴露因素分级标准。可影响疾病预后的因素很多,不同的疾病不尽相同,应结合专业知识,将各种可能与预后有关的因素,均纳入研究因素,这样预后因素的分析才不会遗漏。一般来说,与疾病关系密切的临床指标,往往比较受研究者的重视。预后因素的研究须注意下列几个方面的因素:人口学和社会学因素,如性别、年龄、种族、职业、受教育程度、经济状况;生活习惯与嗜好,如烟、酒、茶、饮食习惯等;疾病临床特征,如疾病的亚型、症状、实验室检查和其他辅助检查结果;各种治疗措施;各种并发症等。

时间是指从研究起点至研究终点(出现阳性结局)或截尾(随访结束未出现阳性结局)的时间。预后研究需要一定的随访时间,因为时间代表着预后。随访时间必须足够长,使大部分患者有足够时间出现阳性结局,得出结论。一般需要根据疾病自然史或病程了解疾病在一段时间发生结局的概率来确定随访时间长短。由于队列研究的病例,入组时间不一致,各个病例到达研究终点的时间也不一致,尽管结束研究的日期是固定的,但由于入组时间不一致,使各个病例从起点到达终点的时间也就不同。因此在队列研究,各研究病例间的时间是不一致的。也正因为如此,才需要运用生存分析的方法。

生存分析是预后研究常用的统计学方法。在临床研究中,我们不能简单地将截尾理解为阴性结局,因为我们不知道他会不会在结束研究后,马上出现阳性结局。在计算时间生存率、半数生存期等时,由于截尾病例包含有不确定性的结局,所以生存分析的原理是将整个随访期分成许多小的时段,根据每个时段内总的人数、终点人数和截尾人数等,推算截尾资料中有多大的比例归入终点,以校正该时段内的死亡率。生存分析围绕着这样一个精神,有多种计算方法,如寿命表法、Kaplan-Meier法、指数模型法等可以计算生存率,并可画出生存曲线,一目了然地估计不同时间的生存率。除了计算生存率,临床研究更重要的是比较各组间的差异,生存分析可以比较两条或多条生存曲线间的区别,如采用Log-rank检验等。

确定研究的结局(outcome):即随访的终点(endpoint),又称阳性结局。根据具体的研究,确定结局的定义。最客观的结局是死亡,其他的结局可包括致残、脏器功能衰竭、疾病的缓解等。只要研究的结局与研究的主题相呼应就可以,关键在于阳性结局必须有一个明确和客观的定义,不能模棱两可,以免出现测量偏倚。如果阳性结局的判断受主观因素的影响,必须采用盲法。



结局包括研究终点和截尾。在结束研究时,尚未达到随访终点者,称为截尾。截尾原因必须与阳性结局无关,多数情况是在结束研究时还未出现阳性结局者为截尾,不称为“阴性结局”。一项足够长时间的随访研究,100% 会出现阳性结局。截尾代表不确定性结局,而不是阴性。

当研究对象因各种原因从原定的研究计划中退出,称为失访。队列研究难免会有失访病例,尤其是随访时间较长的研究。我们需要尽量减少失访,以减低研究的偏倚。失访偏倚影响研究真实性,其影响程度取决于两方面,一是失访人群的质,即失访人群与未失访人群预后因素是否相似;如果失访和不失访者的人口学特征和临床特征较为相近,偏倚较小。二是失访人群的量,如果失访量小于观察人群总数的 5%,可认为所产生的偏倚不大;如果失访率大于 20%,严重影响结果真实性;如失访在 5%~20% 之间,根据情况而定,需叙述失访原因。如果失访的原因与结局无关,如患者搬迁或移民,可以记录从随访起点到失访的时间,并将失访归入截尾病例。但是如果失访与结局有关,如因疗效欠佳而失访者,住院患者因病重而放弃治疗者等,不宜轻率地将其归入截尾,否则会导致偏倚。临床预后因素的研究常比较复杂,可有多个预后因素互相作用,从而影响结局。应用单因素分析不能将多个预后因素对结局的影响分析清楚,此时应采用多因素分析方法。多因素分析可以从多个研究因素中筛选出影响疾病转归和预后的主要因素,以及这些因素在决定预后中的相对比重。同时在预后因素中存在不少混杂因素,应该同时对这些混杂因素进行校正处理。多因素分析可以清除有关混杂因素的影响,从而使得出的结论较前可靠。最常见的多因素分析方法是用于前瞻性队列研究的 Cox 比例风险模型和用于历史性队列研究的 Logistic 回归模型。Cox 回归与 Logistic 回归的主要区别在于:前者与时间有关,后者与时间无关。前者计算 RR 值,后者计算 OR 值。要先对各个研究因素进行单因素分析。对于连续性的数值变量和有明确等级关系的有序分类变量,可以直接进行回归分析;无序分类变量和等级关系不太明确的有序分类变量,则须采用分层回归分析的方法。单因素分析一方面可初步筛选出可能与预后有关的因素;另一方面去除那些根本不可能相关的因素。在建立多元回归方程时,待选变量的标准应该放宽一些,以免遗漏那些由于混杂的影响,导致在单因素分析中被“埋没”的有意义的因素。

### 知识点

#### 单因素和多因素分析

1. 单因素分析 单独研究每个因素对疾病结果的影响,忽略联合作用的因素,有局限性。须保证观察组和对照组的临床特点和其他非研究的预后因素都相同。为减少混杂性偏倚,可采用限制、配对、分层及标准化等方法加以平衡。
2. 多因素分析 预后因素研究中,往往涉及多个因素,结局的产生不是孤立的,各因素之间往往相互影响,可以相互协同或拮抗;需要从许多临床指标中筛选出若干真正与预后相关的因素。虽然可以用单因素分析将各个因素逐个分析,但如果只用单因素分析,往往会产生混杂偏倚。可以采用多因素分析方法,校正各因素之间的相互混杂。

队列研究的优点:队列研究是研究因素和结局变量之间有一段连续的时间,使之建立清晰的因果关系,为临床的因果关系提供强有力的信息。可以直接测量某些暴露因素发展成结局的危险性;可以避免暴露因素测量时偏倚,因为在测量时尚不知结局因素;也可以同时测量已知的混杂因素,以便在统计分析时加以避免。队列研究还可以用于检测多个结局因素,也可以同时研究一系列的暴露因素。

队列研究的缺点:需要花费很大的人力、物力和时间,尤其对阳性结局发生率较低者或从因至果的周期较长者。由于对列研究的暴露与非暴露是顺其自然,而不是人为干预,因此研究初期确定的暴露组和非暴露组可能经过一段时间后会发生转变,如有些吸烟者戒了烟,而有些原



来不吸烟者却开始吸烟了,这常常造成偏倚。对列研究中失访也常常导致研究结果的偏倚。

## 二、病例对照研究

病例对照研究是以同一疾病的不同结局(死亡与痊愈、并发症有无)作为病例-对照研究的病例组和对照组,作为回顾性分析,追溯产生该种结局的有关因素,是由结果倒推至原因的研究方法。病例-对照研究仅能提供预后因素的研究证据,而不能对疾病预后进行评价,即不能提供生存率研究的证据。研究结果只提供是否暴露与结局事件出现的比值比(odds ratio, OR),而非相对危险度(relative ratio, RR)。因此,病例-对照研究在作疾病预后因素分析时,只能用在那些时间不重要的研究。

## 三、横断面研究

横断面研究策略难以胜任疾病预后研究。因为它是一次性获取全部研究资料,不能计算时间生存率,一般来说也难以验证暴露因素与结局之间的关系。只有当暴露因素一旦出现,长久不变者,如血型、基因位点、某些微生物感染后出现的终身抗体等。而且是以患某一不可治愈性疾病为结局的研究,方可以用横断面研究策略来推断暴露因素与结局之间的关系。否则患病死亡者无法被纳入研究样本,患病已痊愈者也难以准确归类。在推断研究因素与结局的关系时,不能顾及时间因素。

### 知识点

#### 几种临床预后问题研究策略的区别

1. 横断面研究 顾名思义横断面,即一次性获取所需的资料,这种资料不适宜作预后研究。
2. 病例-对照研究 由结局(病例/对照)推导病因(暴露率),预后研究价值有限。
3. 队列研究 由病因(暴露/非暴露)推导结局(发生率),是预后研究主要的研究策略。

学习  
笔记

## 第三节 预后研究的队列设计

1. 研究目的 通过队列研究评估妊娠是否对乳腺癌术后患者的预后产生不良影响,比较乳腺癌诊断后妊娠和非妊娠患者人群的无瘤生存率(DFS)。

2. 病例选择 ①研究组为在非转移性乳腺癌诊断明确后任意时间发生妊娠的患者,流产或引产者均包括在内;需要明确患者 ER 状态、淋巴结状态、术后分期、辅助治疗、年龄及诊断年龄;②妊娠期发生乳腺癌或妊娠前复发的患者排除在外。

3. 对照选择 ①对照组为诊断明确的没有妊娠的非转移女性乳腺癌患者,与病例组来源于相同乳腺癌患者登记中心;②因是否手术、术后分期、激素受体状态、发病年龄等也已证实为乳腺癌强而有力的预后因素,所有对照组乳腺癌患者均需采用(雌激素受体)ER 状态、淋巴结状态、辅助治疗、诊断年龄等因素配对作为对照。需要特别指出的是,由于妊娠乳腺癌患者有特定年龄范围,对照组乳腺癌患者的年龄需限定在一定范围。同时,研究中每一个不妊娠患者必须有一个无病时间窗,这个时间窗口的长度不应少于与之匹配的妊娠女性诊断-受孕时间窗口长度。

4. 样本量计算 估计样本含量之前,必须明确一些参数:①  $p_0$ :非暴露人群或全人群中被研究疾病在研究随访期间的结局事件发生率;②  $p_1$ :暴露人群在研究随访期间的结局事件发生率,  $p_1 - p_0$  差值越大,样本量越小,可以通过查阅文献或预调查获得;③  $\alpha$  值,即所希望达到的检



验显著性水平,也就是第一类错误(假阳性错误)的概率,一般取 0.05 或 0.01;④把握度  $1-\beta$ ,  $\beta$  为第二类错误(假阴性错误)的概率,  $1-\beta$  即发现这种关系有多大把握度,指能够发现疾病结局与暴露之间确实有关系的概率,把握度定得越高,样本量越大。可通过查表法或公式法确定样本量。需注意的是,确定样本量时应考虑失访问题,通常增大 10% 的样本量;非暴露组样本量不宜小于暴露组。

5. 暴露因素 本研究中的暴露因素为妊娠,无论是足月妊娠,还是流产或堕胎患者均包括在内。妊娠的发生有不同时间段,本研究限定在乳腺癌诊断明确后任意时间段发生妊娠的患者。妊娠期发生乳腺癌或妊娠期复发的患者排除在外。研究中每一个不妊娠患者必须有一个无病时间窗,这个时间窗口的长度不应少于与之匹配的妊娠女性诊断-受孕时间窗口长度。

6. 结局因素 本研究为比较妊娠组与非妊娠组无病生存率(DFS)的差异。DFS 定义为自诊断之日起,至肿瘤局部复发或远处转移,或出现第二原发肿瘤,或任何原因所导致的死亡。

## 第四节 研究报告与问题分析

队列研究结果的报告包括研究背景与目的、方法、结果、讨论与结论四个方面。讨论对研究设计、可能存在的偏倚以及可能影响结论的因素进行恰当的分析。

雌激素受体状态已知的乳腺癌患者诊断后妊娠与预后关系:一项多中心回顾性队列研究 [J Clin Oncol, 2013, 31(1): 73-79]

1. 目的 此前有荟萃分析显示诊断为乳腺癌后妊娠的女性患者复发率低,但这一结论很可能是由选择性偏倚造成的,因为患者往往在无复发后才选择妊娠。此外,尚无针对 ER 阳性的特殊患者人群探究妊娠效应的研究出现。本研究目的是通过队列研究比较 ER 阳性乳腺癌诊断后妊娠和非妊娠患者人群的无病生存率(DFS)。

2. 方法 研究者开展了此项多中心回顾性队列研究,研究组为已知 ER 状态且在乳腺癌诊断后任意时间发生妊娠的患者,对照组为无妊娠的女性患者,以 1:3 的比例在 ER 状态、淋巴结状态、辅助治疗、年龄及诊断年龄的基础上予以匹配。

3. 结果 研究共有 333 例怀孕患者和与其配对的 874 例非怀孕患者被纳入分析,其中 686 例患者的雌激素受体呈阳性。研究发现,雌激素受体呈阳性的怀孕患者与非怀孕患者在无病存活期方面不存在差异( $HR=0.91, 95\%CI: 0.67\sim 1.24, p=0.55$ ),同时雌激素受体呈阴性的两个患者队列间也不存在差异( $HR=0.75, 95\%CI: 0.51\sim 1.08, p=0.12$ )。不考虑 ER 状态的前提下,妊娠组女性的总体生存率明显更优( $HR=0.72, 95\%CI: 0.54\sim 0.97, p=0.03$ )。此外,193 例诊断为乳腺癌 2 年或者更长时间后发生妊娠的患者较匹配的对照组 DFS 亦无明显差异。怀孕结果与乳腺癌-怀孕间隔也并未对复发风险产生影响。

4. 结论 不管患者的雌激素受体(ER)状态如何,乳腺癌诊断后发生妊娠都可被认为是安全的,妊娠不会给她们的预后带来不良影响。

## 第五节 预后研究证据的评价

对于预后研究的证据,需评价其研究结果的真实性、重要性和适用性。

### 一、预后研究的真实性

研究对象的来源需作详细描述,并具有代表性。准确定义研究对象,有公认的诊断标准。明确研究对象的纳入和排除标准,以便判断研究时是否有选择性偏倚。叙述研究对象来源的研究地区或医疗机构人口学特征、病情轻重、并发症等,以便了解研究对象的代表性和局限性以及选





择研究对象时可能出现的几种偏倚(集中性偏倚、倾向性偏倚、转诊偏倚、诊断条件偏倚)。

预后队列研究要求纳入研究对象时,起始点要十分明确,因为患者处在不同病期和病程都将直接影响预后结局。只有保证研究对象纳入时在大致相同的阶段,其预后结果才具有真实性。研究起始点最好在病程早期,至少应在相同病程阶段。

任何疾病都需要经过一段时间才发生最后的结局,因此随访时间的长短直接影响是否能够发现足够样本的结局,而不会因为过多的假阴性结果影响研究结局的真实性。随访时间必须足够长,使大部分可能会出现阳性结局的患者能够达到研究终点。失访率在 5% 以下,偏倚较小;失访率超过 20%,严重影响结果真实性;失访在 5%~20% 之间,根据情况而定,需叙述失访原因。

研究时要根据疾病特点,对疾病终点进行准确定义,预后指标要明确,判断标准要客观,以避免临床医师在判断预后结局时发生意见偏倚。研究终点也称为阳性结局。结局包括研究终点和截尾。研究终点必须有一个明确的定义或标准,不能模棱两可,以免出现测量偏倚。

如果预后结局是用“死亡”或“病残”等硬指标,可不用盲法判断;但如果阳性结局的判断受主观因素影响就必须采用盲法,以避免偏倚。应该由不知情的其他医生判断,以避免疑诊偏倚和期望偏倚。

### 知识点

#### 疑诊偏倚和期望偏倚的定义

1. 疑诊偏倚 疑诊偏倚(diagnostic-suspicious bias)是指检查者竭力去寻找有关证据来证实假定因素对预后的影响。
2. 期望偏倚 凭主观印象判定预后所产生的偏倚为期望偏倚(expectation bias)。

如果研究亚组中有不同预后结果,该结论是否可靠?对重要的预后因素是否进行了调整?被观察者在亚组中是否具有真实性?

本研究为多中心回顾性队列研究,纳入病例及对照选择均来源于欧洲五家不同研究中心。纳入研究对象为 ER 状态已知的非转移性原发乳腺癌,起点从明确术后病理诊断开始,至以下任一终点:局部复发,远处转移,发生第二原发肿瘤,或任何原因所致死亡。为评估妊娠作为独立预后因素对结局的影响,将以下其他与预后相关因素进行分层分析:ER,淋巴结状态,辅助化疗,辅助激素治疗,年龄(<35 或 ≥35 岁)。

## 二、估计预后研究结果的重要性

是否报告了整个病程的预后结局,而不是某一时时间点的结局;需要了解在某一段时间内发生的结果事件数,生存率常要求有 3 种方法:①某一时时间点的生存百分率:从疾病临床过程的某一点开始,一段时间后存活的病例数占总观察例数的百分比(比如 1 年生存率、5 年生存率);②中位生存时间:观察到 50% 的研究对象死亡的随访时间;③生存曲线:在每一个时间点,研究样本中没有发生该结果的比例。

预后估计的精确度如何?是否报告了预后结局概率的 95% 可信区间。从统计学角度判断预后研究结果估计的精确性,区间越窄,可信度越高。如果 95% 可信区间不包含 1,表明研究结果具有统计学意义。

本研究共有 333 例怀孕患者和与其搭配的 874 例非怀孕患者被纳入分析,其中 686 例患者的雌激素受体呈阳性。中位随访时间妊娠组和非妊娠组分别为 4.7 年(3.1~6.9 年)和 4.7 年(2.5~7.2 年)。最终,共 354 名患者发生复发,两组复发率无差别。研究发现,雌激素受体呈阳性的怀孕患者与非怀孕患者,在无病存活期方面不存在差异( $HR=0.91, 95\%CI:0.67\sim1.24, P=0.55$ ),同时雌激素受体呈阴性的两个患者队列间也不存在差异( $HR=0.75, 95\%CI:0.51\sim1.08, P=0.12$ )。不考





考虑 ER 状态的前提下,妊娠组女性的总体生存率明显更优( $HR=0.72, 95\%CI:0.54\sim0.97, P=0.03$ )。

### 三、估计研究结果的适用性

该研究与自己患者相似吗?是否差异过大?如无较大差异,可以使用其结论。作者是否将研究对象情况介绍清楚?真实患者与研究对象比较,临床特征、诊断、纳入和排除标准及研究场所是否接近。如果越接近,应用结果越有把握。

研究结果是否有助于对临床治疗作出决策和有助于对患者及其家属进行解释?结论若提供给此例患者,是否会造成临床上重大冲突?如证据显示不治疗预后也极佳,需讨论是否应该开始治疗;证据显示不治疗预后不好,倾向给患者治疗;证据无法决定要不要治疗,这些研究证据也可以提供患者及其家属有关其未来可能的资讯。

本研究是首次尝试在 ER 状态已知的患者人群中探讨妊娠与预后关系,同时根据 ER 状态设定样本规模,这使得本研究得到足够的统计学数据来检验妊娠对患者结局的影响。本研究最终提供的数据进一步支持有生育意向的乳腺癌患者,妊娠不会给她们的预后带来不良影响。

## 第六节 临床问题的回答

除了本研究之外,类似队列研究显示了相同结果。美国西雅图 Fred Hutchinson 癌症研究所的研究人员对 438 例确诊后生育的乳腺癌患者进行了研究,同时选取 2775 例诊断后未生育的患者作为对照人群。结果发现,与未生育者比较,确诊 10 个月后生育的患者死亡率降低了 46%。但是,这可能并不是生育所带来的益处,而是由于“健康母亲”效应,因为健康母亲似乎更容易怀孕。

除了队列研究之外,还有荟萃分析的结果也回答了乳腺癌患者妊娠与复发率的关系。该研究显示,诊断为乳腺癌后妊娠的患者复发率低,但这一结论可能是由选择性偏倚造成的,因为患者往往在无复发后才选择妊娠[*Eur J Cancer*, 2011, 47(1):74-83]。

总体而言,确诊后生育的乳腺癌女性患者的复发率及死亡率与未生育者基本一致。

## 第七节 文献阅读与评价

Prognosis of women with primary breast cancer diagnosed during pregnancy: results from an international collaborative study. *J Clin Oncol*, 2013, 31 (20): 2532-2539.

1. 背景和目的 本研究旨在对妊娠期确诊的乳腺癌患者(BCP)预后情况进行确定。

2. 方法 本队列研究对 2003~2011 年间,多个中心(来自妊娠期癌症、勒芬、比利时及 GBC 29/BIG 02-03 登记中心)的 BCP 患者登记情况进行了前瞻性或回顾性整理,并将其与非妊娠患者进行了对比,本研究限制年龄为 45 岁。本研究排除了产后阶段得到确诊的患者。经过对年龄、分期、级别、激素受体状态、人表皮生长因子 2 状态、组织学情况、化疗方式、曲妥珠单抗使用情况、放疗及激素治疗情况进行校正后,本研究主要通过 Cox 比例风险回归对暴露情况下的无病生存率及整体存活率进行了研究。

3. 结果 最终登记结果包括 447 例女性 BCP 患者,其中多数来自德国及比利时,311 例(69.6%)为分析合格患者。非妊娠组有 865 例女性患者。妊娠组患者的中位年龄为 33 岁,非妊娠组患者为 41 岁。中位随访时间为 61 个月。妊娠组 DFS 风险比为 1.34(95%CI:0.93~1.91,  $P=0.14$ ),OS 风险比为 1.19(95%CI:0.73~1.93,  $P=0.51$ )。Cox 回归估测结果表明,BCP 患者 5 年 DFS 率为 65%,假使这些患者未发生妊娠,则估测 5 年 DFS 率为 71%。与之类似,5 年 OS 率也会从 78% 升至 81%。



4. 结论 本研究表明,非妊娠患者与 BCP 确诊患者 OS 结果类似。在为患者提供建议时,该信息具有重要意义,并且该结论支持了在选择继续妊娠的前提下进行的治疗。

### 思考题

1. 针对上述阅读文献(摘要,可以根据需要阅读全文),回答:该研究期望解决的临床问题是什么?科学问题是如何描述的?该研究采用了何种临床研究设计?选择患者和对照来源是哪里?你认为存在哪些偏倚,如何通过改进设计等方法控制偏倚?

2. 结合自己专业,选择一种诊断明确的疾病,探索可能的预后因素(暴露)与该疾病结局之间的联系。设计一项前瞻性或者回顾性临床研究,内容包括:科学问题与研究目的,研究设计与具体方法,结果测量方法及可能的结果。

(刘天舒 王妍)

## 第十五章 如何撰写研究计划书

撰写研究计划书是临床科研工作的第一道工序,对研究的成败起着举足轻重的作用。通过撰写科研计划书,研究者可以在课题开始之前把研究的各个要素组织在一起,逐一加以明确和完善,从而增加了研究的科学严谨性,提高了研究成功的概率。科研计划书也是各级科研基金管理部门对项目的优秀程度进行评审,从而决定是否予以经费支持的书面依据。如果计划书被批准,课题得以正式确立,科研计划书则可以作为开展研究的指导性方案,对今后的研究工作产生直接的影响。因此,撰写研究计划书是所有从事临床研究人员必备的能力。

### 第一节 撰写研究计划书前的准备工作

#### 临床案例

38岁男性,肥胖,患糖尿病2年,不吸烟,无高血压、血脂异常病史。体检:P 70次/分, BP 130/80mmHg。化验检查:TC 5.3mmol/L, TG 1.28mmol/L, LDL-C 3.4mmol/L, HDL-C 1.0mmol/L;空腹血糖 6.2mmol/L,餐后血糖 8.1mmol/L, HbA1c 6.0%。用药:二甲双胍 0.5g Tid。该糖尿病患者是否应服用他汀类药物预防冠心病?

#### 一、确定临床研究问题

1. 从临床实践中提出研究问题 临床实践是产生临床研究问题的根本来源。临床医生应该通过积极的观察和思考,在日常的诊疗实践中不断提高自己发现问题的能力。此病例中“糖尿病患者是否应服用他汀类药物预防冠心病”就是一个与治疗相关的问题。大量研究显示糖尿病患者发生冠心病的风险显著增加。90年代一项芬兰的流行病学研究发现有无糖尿病但无冠心病的患者与无糖尿病但已患冠心病的患者发生心脏事件的风险相当,从而提出了糖尿病是“冠心病等危症”的概念。2001年美国成人胆固醇教育计划(ATP III)建议有冠心病或冠心病等危症(包括糖尿病)的高危患者,LDL-C $\geq$ 3.36mmol/L时应开始采用他汀类降脂药物治疗,以积极预防冠心病的发生。据此,本例糖尿病患者应该考虑服用他汀类药物。然而,医生在实践中可能会对此处理产生疑问:该年轻患者除患糖尿病外,心血管病总体风险评估为“低危”,LDL-C属于边缘升高水平,确实需要服用他汀来预防冠心病吗?有关糖尿病是“冠心病等危症”的证据充分吗?是否需要进一步研究来回答“糖尿病患者是否应服用他汀类药物预防冠心病”这个临床问题呢?

2. 通过文献阅读提出研究问题 临床医生应该通过不断阅读文献动态跟踪相关领域新的进展。阅读系统综述和Meta分析是在短时间内了解该领域现状的最佳方法。针对上述病例引出的临床问题,一项2009年发表的Meta分析汇总了13项研究(45 108例研究对象)的结果,发现有糖尿病但无心肌梗死的患者发生冠心病的风险只有既往有心肌梗死但无糖尿病患者的43%,并不支持糖尿病是“冠心病等危症”的概念。因此,“糖尿病患者是否应服用他汀类药物预防冠心病”仍是一个尚待研究的临床问题。



3. 通过临床指南提出研究问题 临床实践指南是专业学术团体针对特定的临床情况,收集、综合和概括各级临床研究证据,系统制定出来以帮助医师做出恰当处理的指导意见。随着循证医学的推广和完善,标准的临床指南对每一种治疗或预防措施均给出了证据级别,并会指明哪些方面还缺少证据(证据级别为B级,或者是C级),对下一步的研究具有指导作用。同样以上述病例引出的临床问题为例,2001年ATP III的指南就糖尿病是“冠心病等危症”的论点仅引用了三项研究,未进行证据分级;2013年最新的美国血脂治疗指南对糖尿病患者降脂治疗进行了更详细的区分,认为大量证据已显示40~75岁糖尿病患者 $LDL-C > 1.8\text{mmol/L}$ 即应开始中等强度的他汀治疗(A级证据),但对40岁以下和75岁以上糖尿病患者是否需开始他汀治疗的证据级别只有C级,说明该问题尚待研究。

4. 构建临床研究问题 构建研究问题就是把一个临床问题转化为一个科学问题。一些标准的格式,如PICO格式,提供了构建研究问题的逻辑框架和基本思路,有助于使研究问题的定义、层次、范围和相关影响因素更加清晰明确。具体内容可详见第6章(病例对照研究)。结合上述临床问题,按照PICO格式可将研究问题凝练为“年轻糖尿病患者他汀治疗对预防冠心病的作用”,各部分内容具体为:

- (1) 患者和(或)疾病状况(population, patient and/or problem):40岁以下的糖尿病患者。
- (2) 干预(intervention)或暴露(exposure):他汀类药物治疗。
- (3) 对照或比较(comparison):安慰剂。
- (4) 临床的结局(outcome):冠心病。

## 二、制订研究计划

选定研究问题之后,接下来的工作就是要制订研究计划。在正式撰写研究计划书之前制订一个明确的研究计划有助于进一步理清研究的思路和框架,确定科研计划书的基本内容。制订研究计划时一般应遵循下列几项原则:

1. 明确研究目的和目标 制订研究计划的第一步就是将已经初步选定了的研究问题进一步具体化,明确地指出研究的目的,确定研究的具体目标。一个研究往往有多个层次的目标,制订研究计划时应明确各项研究目标之间的内在联系,从不同层次、不同角度回答研究问题,以达到研究的目的。

2. 确定研究设计 临床研究设计的基本原则是“代表性(representativeness)、真实性(validity)、可比性(comparability)和显著性(significance)”。在撰写计划书之前,研究者应该对研究有清晰的思路或构想,应该遵循上述原则确定研究的设计。首先应根据不同的研究类型选择适当的抽样方法保证研究对象的代表性。在研究的设计、资料收集和分析阶段采取适当的措施有效控制各种偏倚(bias),保证收集资料的真实性。可比性是临床研究的重要原则,贯穿于研究设计的始终,在研究对象的选择和分组、研究数据的收集方法、观察指标的选择、研究数据的分析等过程中均需体现可比性的原则。此外,由于临床研究的对象多为抽样所得的样本,因此,必须经过显著性检验(significance test),才能认为样本中研究对象和被研究因素之间的关联是客观存在,而非抽样误差(sampling error)所致。

3. 保证研究的可行性 制订研究计划时必须考虑在要求的研究期限内和研究条件下,研究内容是否可以完成,研究目标是否可以达到。研究的可行性(feasibility)不仅包括技术方案、设施和人员的可行性,也包括操作上的可行性、经济上的可行性和时间上的可行性。研究者需要综合考虑,逐条落实,才能保证研究的成功。通常可以通过预试验(pilot study)和开题论证过程增加研究的可行性。

4. 研究计划的稳定性和灵活性 研究计划的主体和核心部分应该是在充分论证和认真分析的基础上制订的,应该具有相对的稳定性。例如,研究目的是研究的核心纲领,一经确定后尽



量不变。研究的总体设计类型,例如,是病例对照研究,还是前瞻性队列研究,也是研究的基本设想,一经确定后,在研究过程中尽量不变或少变。研究计划毕竟是在研究开始时制订的,在实施过程中难免出现计划外的情况。因此,研究计划也应具有一定的灵活性,特别是一些具体方法,可能在研究过程中根据实际情况进行修订,但最终应保证达到研究的目的。

## 第二节 临床研究计划书的撰写

在完成上述准备工作之后,就可以正式动笔开始撰写科研计划书了。临床科研计划书主要需阐明四个问题:要做什么研究?为什么要做这项研究?如何做这项研究?以及是否有能力完成这项研究?具体而言,一份完整的临床科研计划书一般应包括题目、内容摘要、研究目的和目标、立项依据、研究内容、研究对象和方法、可行性分析、研究创新性、年度计划、预期结果和考核指标、研究基础和工作条件、经费预算和伦理问题等。

### 一、题 目

题目是对研究问题的高度概括。研究问题的选择是临床科研中最重要的一个过程。本章第一节已经论述了如何确定研究问题,本节则重点介绍如何将一个研究问题凝练为一个好的研究题目。题目是一份科研计划书的窗口和灵魂,它给了评审者对计划书的第一印象。此外,题目还应有助于检索,甚至确定计划书的学科分组。撰写科研计划书的题目时需要注意几个问题:

1. 题目要能够反映研究问题,准确表达研究的关注点 题目要包含构成研究问题的基本要素,如研究目的、研究对象、研究类型等。例如上述案例中的“年轻糖尿病患者他汀治疗对预防冠心病的作用”一题,由题目可知本研究的问题是他汀类药物治疗是否可降低年轻的糖尿病患者冠心病的危险,并可推测出研究对象是年轻糖尿病患者;研究目的是评价他汀治疗对年轻糖尿病患者发生冠心病风险的影响;研究类型是临床试验。

2. 题目要简洁明了 题目字数不宜过多,且应该便于理解。应避免使用缩写。必要时可以辅以副标题。副标题和正标题是一种功能互补关系,正标题用于表达研究的主题,副标题则补充说明研究的对象、范围,突出研究重点。例如“颈动脉壁形态学与心血管病的关系:心血管病高危男性的超声研究”一题,正标题说明研究的主题,副标题说明研究的对象和方法。又如,“血管钙化与动脉粥样硬化的关系:冠状动脉与颈动脉电子束 CT 检查的比较”一题,正标题说明研究主题,而副标题突出研究的重点。

### 二、内 容 摘 要

内容摘要是对研究计划书中除了经费预算以外各项内容的概括介绍。摘要虽然只有几百字,但却是计划书中举足轻重的部分。撰写摘要时需注意以下几方面。

1. 摘要必须准确反映研究中最重要内容,主要包括:

1) 介绍研究的大背景,通常为所研究疾病的危害。

2) 指明目前该领域研究的现况和不足。

3) 简要介绍既往的研究基础或预试验的结果。

4) 明确提出本研究的目标或研究假设。

5) 简述研究方法,包括研究设计、评价指标、样本量等。

6) 介绍将要研究的主要内容。

7) 说明你的研究将如何弥补目前该研究领域不足,将怎样促进该领域的发展,指明研究的重要意义。

2. 文字要言简意赅、便于理解 多数计划书的格式对摘要要有明确的字数规定,一般为 250~



500 字左右,因此摘要的撰写需字斟句酌、言简意赅。摘要的文字还应便于理解,因为读者不仅包括本专业的同行专家,还包括科研管理人员和一些其他领域的大同行专家,因此需避免在摘要中使用过多晦涩的专业词汇。

### 三、研究目的和目标

研究目的和目标一般为科研计划书正文的第一个部分,为全文设定了基调,是研究计划书的精髓,计划书的其他部分都应该围绕这部分进行组织。此部分撰写时需注意以下几点:

1. 目的和目标明确 研究目的即为什么进行这项研究,是研究的最终宗旨,通常是方向性的、定性的。与研究目的相比,研究目标更为具体,是为了最终达到研究目的而需要完成的具体、可操作、可评估、可量化的任务。例如上述的“年轻糖尿病患者他汀治疗对预防冠心病的作用”的研究,目的是“评价他汀治疗对年轻糖尿病患者发生冠心病风险的影响”;研究目标为“评价对 40 岁以下 2 型糖尿病患者进行他汀类药物治疗与安慰剂相比对未来 5 年冠心病发病风险的作用”。有时也可以设立多个研究目标。

2. 目标之间要有内在联系 一般一项研究只有一个研究目的,但可以下设几个具体的研究目标。如果有多个目标时,不同的目标应该是从不同角度、不同阶段回答研究问题,从而最终从整体上达到的预定的结果,实现研究的目的。

### 四、立项依据

立项依据是整个研究的立论基础,为整个计划书设定了方向,为计划书的其他部分构建了框架。立项依据部分应说明为什么要进行这项研究,研究有哪些意义,即研究的重要性和必要性,通常包括论述和参考文献两个部分。

1. 立项依据的论述,可以从三个方面进行阐述

1) 研究的需求:临床研究的意义在于解决临床实践中的问题,或增加相关的知识。因此,立项依据部分应首先说明下列问题:该研究问题是临床实践过程中需要解决的诊断、治疗、预后或病因学问题吗?需求的对象是谁?需求的普遍性和迫切性如何?该领域的国内外研究进展如何,既往有关该问题的研究有哪些不足,还有哪些方面尚需进一步研究,或需要加以改进,或提供更多的知识?例如,“老老年高血压患者降压治疗试验”(hypertension in the very elderly trial, HYVET)在立项依据部分指出:随着人口老龄化的进程加快,80 岁以上的老老年人的健康问题日益重要。高血压在 80 岁以上的老老年中普遍存在,其中 50% 以上的女性偶测收缩压高于 160mmHg。既往针对高血压的临床试验显示 80 岁以下高血压患者降压治疗可显著降低心血管病危险,然而 80 岁以上老老年高血压患者降压的疗效却不肯定,多数临床试验都排除了 80 岁以上的患者或样本例数不足以显示该组患者的获益。因此,80 岁以上老老年高血压患者降压治疗的获益和可能存在的危险尚需进一步研究。这是临床医生面对的临床问题,是临床防治指南需要回答的问题,也是制定相关社会医疗保健政策时必须考虑的问题。

2) 研究的获益:重点阐明研究问题解决后,将有哪些益处,获益范围有多大,获益者是谁。例如,研究的结果可以带来哪些诊断、治疗、预后或病因研究方面的改变,可以增加哪些知识?是可以提高对疾病的诊断和治疗水平,还是可以改善对疾病发病机制的认识,或是可以为临床指南或卫生政策的制定提供依据?

3) 如果该问题得不到解决将有什么危害:与研究结果可能带来的获益相对应的是,如果研究问题不解决将产生哪些后果。例如,是导致了误诊或漏诊,还是导致患者未能获得有效治疗,增加了患者不良预后的危险,或是从更广泛的角度而言,对患者的家庭和社会有何影响?

2. 参考文献 立项依据部分对国内外既往研究进行总结时必须列出主要参考文献,包括作者、题目、期刊名称、年份、卷(期)、起止页。参考文献引用需注意以下几点:





1) 参考文献的引用要有逻辑性和针对性:引用的文献必须与本研究问题密切相关,对文献的回顾应该自然地引出本研究要解决的问题,要突出本研究的起点和在当前研究中的位置、优势和突破点。应尽量引用本领域最有影响力的研究文献、系统综述或 Meta 分析、该领域最早的研究、或本研究团队既往发表过相关的文章。总之,参考文献的引用要体现出申请人对目前该领域的研究现状和发展趋势有充分的认识。

2) 参考文献的引用要避免偏倚:不能只引用支持自己观点的文献,或断章取义。对于与本研究论点不一致的文献或有争议的文献也要引用,并对产生不同结论的可能原因加以分析。

3) 参考文献要反应研究领域的最新进展:参考文献中多数应为最近几年的研究,务必包括与本研究问题密切相关的最新文献,这一要求在强调创新性的研究中更为突出。

## 五、研究内容

研究内容陈述的是课题研究的范畴,是对研究目标的具体阐述。可以根据每一项目标来确定内容。相对于目标而言,研究内容要更加具体、详实。同一目标可通过几方面的研究内容来体现;有时可将一个课题分为几个子课题,每个子课题针对一个研究目标和相应的研究内容,这也是一种简单明了的陈述方式。研究内容的撰写需注意几项原则:

1. 研究内容应紧扣研究目标,研究内容完成后应当可以实现研究目标。

2. 研究内容的表述应该具体详实、明确中肯,切忌笼统、模糊;如果有多项研究内容一定要加小标题,然后再把内容细化。

3. 研究内容要在预期结果中予以体现,通过完成研究内容可达到预期结果。

4. 研究内容的设定要适当,确保研究预算可满足需要,研究期限内能完成。

5. 要注意区分研究内容和研究方法,不能把研究方法作为研究内容。

## 六、研究对象与方法

研究对象与方法部分要说明研究的对象、研究类型、资料收集方法和内容、统计分析方法、质量控制措施和总技术路线等内容,即说明如何进行这项研究。这通常是科研计划书中篇幅最大、读者最关注的部分。研究对象与方法部分的缺陷通常是科研计划书未被批准的重要原因。如果计划书被批准,研究方法部分则可作为制订研究操作指南的依据。撰写研究对象与方法部分的总体原则是合理、可行、有针对性。具体而言,即要求研究方法能够解决研究问题,达到研究目标;在研究期限和现有的经费、人员和设备的条件下,研究方法是可行的。此部分的撰写原则细述如下:

1. 研究类型 研究方法部分首先应说明该研究的类型。这种说明不仅有助于读者评估本研究的价值,还有助于对具体资料收集方法和技术路线的理解,临床研究常用的类型包括横断面研究(cross-sectional study)、病例-对照研究(case-control study)、队列研究(cohort study)、临床试验(clinical trial)等。其中临床试验类的研究还需将重要信息在公开的临床试验注册机构进行登记,以便向公众、卫生从业人员、研究者和赞助者提供可靠的信息,使临床试验的设计和实施更加透明化。

2. 研究对象 包括样本来源、入选标准、入选方法和样本量的计算等内容。

1) 样本的来源:样本来自什么地区,是社区人群还是患者,是住院患者还是门诊患者、来自哪个或哪些医院。这些问题直接关系到样本的代表性,因此必须说明。

2) 样本的入选标准和排除标准:①入选标准:入选研究对象时,必须明确规定哪些人可以入选。通常要考虑的是年龄、性别、地区、种族等一般特征。研究对象为某种疾病的患者时,必须对疾病的诊断标准做明确说明。诊断标准最好选择金标准、指南规定的标准或临床公认的标准。无法采用上述标准,而必须加以修改或自行制订时,则要说明修改或制订的依据;②排除标准:



研究方案中要预先说明哪些人不能进入研究,已进入研究的对象在研究过程中出现哪些情况时应该退出等标准。

3) 样本的入选方法:首先应说明样本的入选是采用随机化方法(methods of randomization)还是非随机化方法,抽样(sampling)框架是什么,随机抽样采取的是哪一种随机方法。随机化分组的研究应说明随机数字是如何产生的,随机化是如何执行的。涉及匹配(matching)的病例对照研究应说明匹配的条件和比例。

4) 样本量:样本量即在保证科研结论具有一定可靠性的前提下,所需研究对象的最小例数。不同类型的研究有不同的公式用来计算样本含量。研究者可直接按照公式计算,也可采用查表法,或借助专门的统计软件来计算。样本量的估计主要取决于4个因素:第一类错误(type I error)的概率 $\alpha$ 、第二类错误(type II error)的概率 $\beta$ 、允许误差 $\delta$ 或差值、总体误差或总体率 $\pi$ 。样本量的估算方法应在统计方法部分详细说明。

3. 资料收集方法和内容 这一部分要说明为实现研究目标需要收集哪些数据、如何收集。资料收集的方法包括研究中使用的调查方法、干预方法、实验室检测和辅助检查方法等。方法的介绍要详尽,具体,层次清楚,应使读者了解研究者的技术水平。注意此部分的叙述要与可行性和创新性分析相一致,既让读者相信这些方法是可以实现的,同时又强调和论证了创新的方法。此部分还应说明每一种资料收集的方法要观察或测定的具体指标。指标的选择要求是可测量的、稳定的、客观的、特异的。指标的测量方法和定义应该是经过验证、有据可依的,可引用相关指南或既往研究发表文献。

4. 统计方法 统计方法部分是在综合考虑研究目的、研究方法和收集数据的性质等因素的基础上,对数据进行分析的计划和具体分析方法的描述。在统计方法中研究者还应说明样本量估算的方法,以及研究数据出现特殊情况应如何处理,如患者退出或失访,采用了其他治疗措施等。资料分析所使用的软件名称和版本号也应加以说明。

5. 质量控制措施 质量控制(quality control)的过程就是要找出可能影响研究质量的环节,并采取措施加以预防。通用的质控措施包括制订工作手册和标准操作规程、人员培训、质量考核等。此外,资料收集和分析的不同过程也有不同的质控重点。

现场调查或临床研究过程的质控:此部分涉及的质控内容包括:调查员的培训和考核、预调查、调查问卷和量表的信度和效度分析、调查资料的完整性和准确性评价等。临床试验,特别是盲法的临床试验还需对药品的标签、分配、储存、发药和回收等各个环节进行质控。

检查和检测过程的质控:实验室检测和心电图、超声等辅助检查过程涉及的质控内容包括:仪器设备的校准、检测人员盲法测定(检测人员不知道研究对象的分组和其他临床检查的结果)、使用内部和外部质控血清并绘制质控曲线、重复样本的盲样检测、检查者自身和不同检查者之间的质控等。

数据整理过程的质控:整理原始数据的过程需要遵循4项原则:①真实性原则:即根据经验和常识判断真伪,及时进行核实;②标准性原则:审查每项指标是否按规定要求收集;③准确性原则:检查资料是否有填写错误或逻辑性错误;④完整性原则:各项指标是否收集齐全。将原始数据录入计算机的过程常用的质控措施包括由两名录入人员分别录入,进行两次录入的对比查错,还可设计一定的计算机程序对已录入的数据进行逻辑查错等。

6. 技术路线 技术路线即研究的流程,具体而言,即研究者通过文字、流程图等形式对研究时间和研究步骤之间的内在逻辑关系的描述。流程图比文字说明更为一目了然,有助于读者在短时间内了解研究的总体框架。

## 七、可行性分析

可行性分析就是要说明本研究设计能够确保内容的完成和目标的实现。可行性分析一般



包括四部分:

1. 技术上的可行性 技术上的可行性评价的是解决研究问题的具体技术方案是否可以实现,以及研究设施和人员的可及性。技术可行性分析是对技术手段、设施和人员的总结分析,具体技术细节和人员、设备的详细介绍则一般在“研究基础和工作条件”部分进行描述。

1) 研究技术:现有的技术是否可以解决研究问题? 这些技术是否成熟,是经过验证的吗? 如果将采用新技术,其创新点体现在何处? 新技术的可行性是否经过预试验的验证?

2) 研究设施:研究所需的场地、设备等条件是否已具备? 如果尚不具备,有何解决措施?

3) 研究人员:申请人本人和研究团队是否具有相关的经验,掌握相关的技术? 研究团队在经验和专业上的搭配是否合理,是否有技术上的互补? 研究人员是否有充足的时间参与本研究?

2. 操作上的可行性 操作上的可行性评价的是在现有组织形式下研究计划是否可行。此部分应回答的问题包括:研究方案是否可以被相关组织和个人接受而推广实施? 是否符合医学伦理? 例如,如果计划在某大学进行有关 AIDS 的调查,首先应获得研究机构的伦理委员会的批准,还要征得学校管理部门的同意,也要获得被调查者的知情同意。

3. 经济上的可行性 经济上的可行性评价的是研究计划的费用效益。此部分应回答申请课题的经费支持是否可满足研究的需要? 有些计划书会要求在“经费预算”部分对此做详细分析。此外,对于推广应用性的研究,此部分还应说明研究将有哪些有形的或无形的获益,研究的获益是否超过对研究的投入?

4. 时间上的可行性 时间上的可行性评价的是研究计划是否能在截止日期前完成。此部分应回答的问题包括:研究计划是否可以按期完成? 影响研究进程的关键步骤是什么? 有何措施保证该步骤可以顺利实施? 如果不能顺利实施而影响了研究进程,对完成研究的总体目标有何影响?

## 八、研究创新性

创新性往往是科研计划书中撰写难度较大的部分。研究者常常感到无从落笔,因此只能空洞地套用“国内首次……”、“填补……的空白”等词句。其实,科学发展至今,几乎所有的科学研究都不是凭空产生的,都基于已知科学原理、技术和方法。因此,创新性应该是在对既往研究进行充分分析的基础上提出的。任何有别于既往研究的研究问题、研究方法,或在既往研究的基础上提供了新的知识,推动了临床问题的解决,都具有不同程度的创新性。具体而言,创新性可以体现在以下三个方面:

1. 研究问题的创新 一个有创新性的研究问题不一定是既往从未被研究过的问题,但一定是既往研究尚未完全解决的问题。它可以是一个全新的研究问题,也可以是对既往研究问题从深度和广度上的进一步补充,或者是既往有争议的研究问题。例如,已经有大量研究证实降压可以减少脑卒中的发病危险,但这些研究的对象多为 80 岁以下,而 80 岁以上老老年高血压患者降压治疗是否可以获益则尚无定论。因此,专门研究 80 岁以上老老年高血压患者降压治疗是否可以降低脑卒中的危险就是一个有创新性的研究问题。

2. 研究方法和材料的创新 采用一种全新的方法或材料,或是对现有方法或材料的改进,或是现有的方法或材料在不同领域的应用,都属于创新。例如抗凝药华法令的用药剂量有很大的个体差异,临床通常根据年龄、体表面积、种族、性别等因素来预测用药剂量。有研究在此基础上进一步加入了影响用药剂量的基因多态性的信息,建立了新的模型来预测华法令的剂量,这就是方法上的创新。航空航天材料在医学生物工程领域的应用则属于现有材料在新领域应用的创新。

3. 预期结果的临床价值 临床研究不同于基础研究,临床研究主要是为了解决临床实践中



的问题,因此,其创新性还应该体现在研究结果是否将在不同程度上推动临床问题的解决,是否将带来诊断、治疗或预后的改变。

## 九、年度计划、预期结果和考核指标

1. 年度计划 年度计划部分重点说明研究的进度,即研究方案的具体时间安排。年度计划不仅规划了研究的实施进度,也常被科研管理机构评价研究进展情况的指标。因此,年度计划应该具体、可操作,而且合理、可行。

2. 预期结果和考核指标 预期结果即研究的可能产出,考核指标是对研究产出的量化标准。二者都是科研管理机构对课题进行监督、考核和评价的依据。

预期结果通常包括:完成研究内容并达到研究目标、预期成果的应用前景和将产生的社会效益和经济效益、学术交流和人才培养等。

考核指标即针对预期结果制订的具体、量化的指标,具体包括:①完成研究内容:将研究内容转化为具体、可测量的指标,与研究目标相呼应;②研究成果:介绍本课题计划产出多少论文、专利、成果奖等;③国内外学术交流:包括参加国内外学术会议、以发言或壁报的形式汇报本课题的研究结果、出国培训学习、邀请国外专家来华讲学等内容;④人才培养:包括可培养多少研究生、主要研究人员的职称和业务水平如何提高等。

## 十、研究基础和工作条件

说明课题组主要成员既往主要工作基础和实验室支撑条件,并进行客观的自我评价。此部分是对可行性分析中有关技术、设备和人员的分析给出具体的说明资料,一般包括以下几个内容:

1. 工作基础 应说明课题组主要成员既往从事的与本课题相关的研究工作和已取得的研究工作成绩。工作基础的介绍应该注意:

1) 着重介绍与本课题相关的工作,而不是罗列所有既往从事的研究。

2) 应说明既往工作与本课题之间的联系,如为本课题奠定了方法学的基础,还是提供了研究人群,或是收集了前期数据等。

3) 如果进行了前期研究或预试验,即使尚未发表文章,也应该介绍取得的初步结果,并评价这些前期工作对本课题的作用。

2. 工作条件 研究方法中要求的条件都应具体说明,包括已具备的实验条件,尚缺少的实验条件和拟解决的途径。提出利用国家重点实验室或部门开放实验室,但应说明计划与落实情况。

3. 申请人和课题参与者简介 简要介绍申请人和课题主要参与者的学历和工作简历,说明申请人和课题主要参与者在本项目中承担的任务,提供近期(一般为近3年)已发表与本项目有关的主要论著、获得学术奖励情况和承担科研项目的情况,论著目录要求按标准的参考文献格式,科研项目应注明项目的名称和编号、经费来源、起止年月、本人承担角色等。

## 十一、经费预算

经费预算部分是要说明完成本研究的目标需要的费用,并给出相关依据。如果是申请科研基金,在撰写计划书之前就要知道该课题资助力度有多大,有时资助力度在课题申报指南中会有明确说明,有时则需参考既往该类课题的资助力度。制订预算需注意以下几点:

1. 经费预算要与研究内容相符 目前科研经费的预算基本采用零基预算,即不考虑过去的预算项目和收支水平,以零为基点编制的预算。因此,制订预算时要根据研究需要和客观实际情况,对各个项目逐条分析,按照成本效益原则,确定预算支出项目和数额。预算相对研究内容



过少,可能无法保证课题的完成;如果预算相对研究内容过多,漫天要价,则可能导致研究计划被否决。

2. 经费预算要清楚分类 研究的总预算一般包括设备费、材料费、测试化验加工费、燃料动力费、差旅费、会议费、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、管理费。确定各项开支的定额是编制零基预算的基本要求。多数科研管理部门要求不同类别间的费用不能大比例地转移(一般不超过该项预算的10%),因此每一类的预算都要有一定的依据。一些基金申请时还要求提供设备和试剂的报价单、测试化验加工的委托合同等。

## 十二、其他内容

计划书一般还要求包括伦理委员会批准书、单位意见、合作协议和推荐信等文件,通常作为附件资料。

1. 伦理委员会批准书和知情同意书 所有以人为对象的研究必须经伦理委员会审议同意并签署批准意见。有些项目在申请时还要求提供知情同意书的全文。

2. 单位意见 研究者所在单位应提供对课题的审查意见,通常内容包括对本人的评价、对本课题的评价、是否可以提供人、才、物等方面的支持等。

3. 合作协议书或意向书 如涉及与其他单位的合作,计划书中还要附有双方的合作协议书或意向书,对合作的形式、双方的责任、权利、义务,以及经费使用和知识产权等内容做出详细约定。

4. 专家推荐信 研究生和初级职称研究人员申报课题时常需要两位具有高级职称的同行专家推荐,推荐者应实事求是地对申请人的业务能力和研究基础做出评价。

总之,撰写研究计划书是一项具有挑战性的工作。从研究问题的凝练到具体的目标、从整体的构思到具体的方法、从可行性分析到对结果的预期,研究计划书整合了科学研究的方方面面。一份研究计划书是否优秀,不仅取决于研究本身的价值,也取决于研究者的撰写水平。因此,撰写研究计划书是一种至关重要的科研能力。本章介绍了研究计划书的主要内容,并分析了各项内容撰写过程中应注意的问题,希望对大家有所启发。

(张宏家 刘静)



## 第十六章 数据收集、数据管理与统计分析

临床医师通过多年的临床实践,积累了大量的临床经验,而不同的医师,经验可能不同,因此如何对不同的临床技能进行评估,促进临床诊治,是迫切需要解决的问题。而临床经验的总结,需调查和收集大量的数据,然后对数据进行整理分析,通过合理的统计学方法进一步验证。严格控制原始数据的收集质量、熟悉收集方法、数据来源和数据性质、不同类型数据间相互转换、完善数据管理,确保数据的完整、准确、一致,是临床研究的基本要求,也是临床科研工作者及初入临床的住院医师要面对的问题。

### 第一节 数据收集

#### 一、数据收集的意义和对收集数据的要求

数据收集(data collection)过程,就是按照研究设计所拟定的方法与过程,通过对研究对象的观察及试验,测量并记录其结果,以形成研究的效应的原始统计数据。

原始数据的收集,不仅构成了统计分析的基础,也是决定调查研究成败的关键,是调查研究至关重要的一步。因此作为临床医生,熟悉掌握临床数据的来源与收集方法非常重要。对疾病而言,数据收集的意义主要体现于利用所获数据描述疾病的特征,如地区分布、时间分布、人群分布等,以及对人类健康的危害程度等,为有效治疗疾病积累宝贵的临床经验。

无论采用何种研究方法,所需资料均应具备真实性、完整性、可靠性和可比性等条件,而且数据收集要遵循及时、准确、完整的原则。

#### 二、数据来源与收集途径

医学领域的统计资料主要来自三个方面。

1. 统计报表 如医院工作报表、法定传染病报表等。统计报表由国家统一设定,由指定的医疗卫生机构定期上报的反映居民健康和卫生服务需求状况的主要数据。

2. 经常性的工作记录 日常医疗卫生工作记录及有关的报告卡,日常医疗卫生工作记录如门诊病历、住院病历、健康检查记录、病理检查、各种生化指标及影像学检查指标等,这些资料均来自于医院的病案室或相应的各科室;有关的报告卡如传染病或非传染病报告卡、职业病报告卡、出生报告、死亡报告等。此类资料原始记录由于涉及多人,需要注意以下几方面的问题:

资料可能存在选择性偏倚。限于医院的水平、规模、性质、专业性等不同,收治的患者也各有差异,如三甲医院收治的患者大多为病情严重或疑难病例,最终统计数据会出现病情偏重,治愈率偏低的现象,由于医院所处地理位置不同或医疗水平不同,导致即使从不同医院中收集到的资料有年龄、性别、职业的差别,但也不能代表一般情况,因此在进行不同医院资料比较时,注意这种偏倚造成的影响。

医院的病史资料不能代表所有病例的情况。一个地区医院病史资料的总和也不能反映该地区总的发病全貌。

不同时期对医院病史资料的影响。时期不同,对医院的病史资料往往也有影响。如随着医





疗水平和诊断水平的提高,疾病的诊断标准也随着变化,势必影响该地区发病率高低的判断;另外医院的床位数增减、医疗设备更新、医务工作者业务水平、服务态度、护理质量等对治愈率均有影响。

3. 专题调查[实(试)验] 是指专门通过专题调查或实验获得的数据,如疾病的病因学研究。临床干预措施效果评价、疗效分析等。这类数据难以从常规的医疗工作中获得,必须根据研究目的,通过专题调查、现场调查或实验研究获得,需要花费大量的人力、物力和财力,而且对调查或实验数据,要特别注意指标的说明,数据的准确度与精确度。在动物实验中会出现动物中途死亡,在临床试验中,有的患者会中途退出试验,对这些情况都应该详细记录,以便分析和判断时间的发生是否与试验药物有关。

### 知识点

#### 资料收集要点

资料的收集方法应详细介绍,是通过医院已有的病史资料或制订调查表直接向患者询问调查。如果是采集实验室的数据,须介绍标本采集的方法和时间,以保证实验室数据采集的正确性,如被检对象某些情况可影响检测结果,则须制订相应的措施以控制这些影响因素。在科研设计报告中应附有该课题研究的调查表和观察表,并且要附有填写这些调查表的须知及计算机编码说明等。如进行询问调查,应注明调查者,对调查者是否经过培训等。

## 三、调查表

调查表(questionary)是临床流行病学研究中收集可靠信息、资料,进行统计分析的重要工具。它是记录调查内容的原始表格,通过调查表收集的信息质量直接影响到整个调查研究的质量。一份好的调查表需要调查者不断地修改和完善。

调查表又称调查问卷或询问表,是以问题的形式系统地记载调查内容的一种格式。问卷可以是表格式、卡片式或簿记式。设计问卷,是询问调查的关键。完美的问卷必须具备两个功能,即能将问题传达给被调查者和使被调查者乐于回答。要完成这两个功能,问卷设计时应当遵循一定的原则和程序,运用一定的技巧。

### 知识点

#### 调查表的设计原则

1. 有明确的主题 根据主题,从实际出发拟题,问题目的明确,重点突出,没有可有可无的问题。
2. 结构合理、逻辑性强 问题的排列应有一定的逻辑顺序,符合应答者的思维程序。一般是先易后难、先简后繁、先具体后抽象。
3. 通俗易懂 问卷应使应答者一目了然,并愿意如实回答。问卷中语气要亲切,符合应答者的理解能力和认识能力,避免使用专业术语。对敏感性问题采取一定的技巧调查,使问卷具有合理性和可答性,避免主观性和暗示性,以免答案失真。
4. 控制问卷的长度 回答问卷的时间控制在20分钟左右,问卷中既不浪费一个问句,也不遗漏一个问句。
5. 便于资料的校验、整理和统计。

1. 调查表的种类与结构 一般来说,调查问卷的问题有两种类型:开放式问卷和封闭式



问卷。

(1) 开放式问卷:临床资料的信息有些是开放式信息,如患者体征、主诉、图像信息(如X线片信息、B超、CT等),有些是封闭式信息如调查问卷(CRF)。

开放式问卷又叫无结构问卷,是问卷设计者提供问题,由被调查者自行构思自由发挥,从而按自己的意愿回答问题。

开放式问卷的特点:项目的设置和安排没有严格的结构形式,所调查的问题是开放性的,只是结构比较松散或较少。这种类型的项目较少作为单独的问卷使用,往往是在对某些问题需要做进一步深入调查时与结构型问卷配合使用,或者是用于研究者对某些问题尚不清楚的探索性研究中。

在采用开放式问题时,应答者可以用自己的语言自由地发表意见,在问卷上没有已拟定的答案。

例如:您抽香烟多久了?您胃不舒服多长时间了?有何症状?显然,应答者可以自由回答以上的问题,并不需要按照问卷上已拟定的答案加以选择,因此应答者可以充分地表达自己的看法和理由,并且比较深入,有时还可获得研究者始料未及的答案。通常而言,问卷上的第一个问题采用自由式问题,让应答者有机会尽量发表意见,这样可制造有利的调查气氛,缩短调查者与应答者之间的距离。可以收集到范围比较广泛的资料,可以比较深入的发现和探究一些特殊问题,探寻到特殊群体的意见和观点。

然而,开放式问题亦有其缺点。例如调查者的偏见,因记录应答者的答案是由调查者执笔,极可能失真,或并非应答者原来的意思。如果调查者按照他自己的理解来记录,就有出现偏见的可能。但这些不足可运用录音机来弥补。开放式问题的第二个主要缺点是资料整理与分析的困难。收集到的资料很难量化,难以进行统计分析,要求研究者有较强的资料分析问题的能力。不适合文化程度不高,文字表达有困难的研究对象。由于各种应答者的答案可能不同,所用字眼各异,因此在答案分类时难免出现困难,整个过程相当耗费时间,而且免不了夹杂整理者个人的偏见。因此,开放性问题在探索性调研中是很有帮助的,但在大规模的抽样调查中,它就弊大于利了。

(2) 封闭式问卷:封闭性问答题,又称有结构的问答题,它规定了一组可供选择的答案和固定的回答格式。

封闭式问题与开放式问题相反,它规定了一组可供选择的答案和固定的回答格式。

封闭式问题的优点包括:

- 1) 答案是标准化的,对答案进行编码和分析都比较容易。
- 2) 回答者易于作答,有利于提高问卷的回收率。
- 3) 问题的含义比较清楚。因为所提供的答案有助于理解题意,这样就可以避免回答者由于不理解题意而拒绝回答。
- 4) 封闭式问题也存在一些缺点。
- 5) 回答者对题目不正确理解的,难以觉察出来。
- 6) 可能产生“顺序偏差”或“位置偏差”,即被调查者选择答案可能与该答案的排列位置有关。研究表明,对陈述性答案被调查者趋向于选第一个或最后一个答案,特别是第一个答案。而对一组数字(数量或价格)则趋向于取中间位置的。为了减少顺序偏差,可以准备几种形式的问卷,每种形式的问卷答案排列的顺序都不同。

2. 调查技巧 敏感问题调查技术与方法:被调查者“不愿答”的问答题,有几种情况,其一是答卷人要花大力气来提供资料;其二是调查的某些问题与调查的背景不太符合(例如普通商品的消费与个人隐私问题放在同一问卷中就不相衬);其三是调查目的的合理性,被调查者不愿意提供没有合理目的的调查;其四是敏感的问题。鼓励被调查者提供他们不愿提供信息的方法,



包括:

- 1) 将敏感的问题放在问卷的最后。此时,被调查者的戒备心理已大大减弱,愿意提供信息。
- 2) 给问答题加上一个“序言”,说明有关问题(尤其是敏感问题)的背景和共性——克服被调查者担心自己行为不符合社会规范的心理。
- 3) 利用“第三者”技术来提问题,即从旁人的角度涉入问题。

### 抗结核药物不良反应调查问卷

#### 一、患者基本信息

1. 患者登记号:□□□□□
2. 患者姓名:\_\_\_\_\_
3. 患者住址:\_\_\_\_\_省\_\_\_\_\_地/市\_\_\_\_\_区/县\_\_\_\_\_ (社区)村
4. 联系电话:\_\_\_\_\_
5. 性别: (1) 男 (2) 女
6. 出生年月:\_\_\_\_\_年\_\_\_\_\_月
7. 民族:(1) 汉族 (2) 回族 (3) 藏族 (4) 蒙古族 (5) 壮族 (6) 其他(请注明)\_\_\_\_\_
8. 文化程度:(1) 小学及以下 (2) 初中 (3) 高中 (4) 大学及以上
9. 职业:\_\_\_\_\_
10. 婚姻状况: (1) 未婚 (2) 已婚 (3) 离异或丧偶 (4) 其他
11. 2011 年家庭总收入:\_\_\_\_\_元,家庭实际(一个家庭实际居住的)人数\_\_\_\_\_个
12. 既往病史:(1) 无 (2) 有(请注明)\_\_\_\_\_ (包括心脏病、高血压、血液病、肝炎等)
13. 过敏史:(1) 无 (2) 有(请注明药品名称)\_\_\_\_\_
14. 饮酒(指白酒)情况:(1) 从不 (2) 过去饮,现在不 (3) 现在饮(平均每天\_\_\_\_\_两)
15. 吸烟情况:(1) 从不 (2) 过去吸,现在不吸(烟龄\_\_\_\_\_年,以前每天平均吸\_\_\_\_\_支)  
(3) 现在吸(烟龄\_\_\_\_\_年,每天平均\_\_\_\_\_支)
16. 是否因为其他疾病(非结核病)使用过下列药物?  
(1) 链霉素 (2) 利福平 (3) 卡那或阿米卡星 (4) 氟喹诺酮类 (5) 以上均未使用
17. 是否因使用上述药品出现不良反应?  
(1) 否 (2) 是(请注明)\_\_\_\_\_

#### 二、本次诊疗情况:

18. 肺结核发病时间:\_\_\_\_\_年\_\_\_\_\_月
19. 本次症状:(1) 咳嗽或咳痰 (2) 发热 (3) 盗汗  
(4) 食欲下降 (5) 其他(请注明)\_\_\_\_\_ (6) 无症状
20. 确诊时间:\_\_\_\_\_年\_\_\_\_\_月
21. 确诊机构:(1) 县医院 (2) 县疾控中心结防门诊 (3) 市级及以上综合医院  
(4) 结核病专科医院 (5) 其他(请注明)\_\_\_\_\_
22. 诊断:(1) 原发性肺结核 (2) 血行播散性肺结核 (3) 继发性肺结核  
(4) 结核性胸膜炎 (5) 其他肺外结核
23. 是否有肺结核家族史:(1) 否 (2) 是
24. 是否有肺结核接触史:(1) 否 (2) 是
25. 本次医疗费用支出类型:(1) 自费 (2) 公费医疗 (3) 医疗保险  
(4) 新农村合作医疗 (5) 其他(请注明)\_\_\_\_\_

调查员姓名

调查日期

年 月 日



## 第二节 数据管理

从病历、调查表中收集到医学研究的相关数据后,数据的管理工作十分重要。数据管理的目的是保证数据的准确性、完整性、逻辑性与一致性。通过对数据进行整理与清洗,使其系统化、条理化,以保证数据质量、满足分析需求、便于数据的保存和传递并提高工作效率。

对于临床试验来说,其数据管理常通过专门的数据管理系统进行,并且要求研究者、监察员和专业数据管理员的密切配合。完整的临床试验数据管理流程一般包括病例报告表的接收,病例报告表的跟踪或记录,数据库的建立,数据的录入与校对,数据核查,数据清理与疑问表管理,数据更改和数据库的锁定等内容。整个过程临床试验数据管理过程受到严格的审查/检查或质量控制,需要根据《药物临床试验质量管理规范》、《临床试验数据管理工作技术指南》等标准进行操作,读者可参考相关的标准和资料,本书中不再详细介绍。

在日常工作中,临床医师也常常采用观察性研究的方法进行科学研究。对于观察性研究来说,其数据管理过程虽然难以完全按照临床试验数据管理要求进行,但是也应尽可能地保证数据的质量。为保证其研究数据质量,一般需要考虑以下问题。

### 一、如何建立数据库

1. 选择数据管理软件 数据管理软件有很多,如 Excel、Assess、EpiData、Oracle,还有许多自行开发的临床电子数据采集系统(electronic data capture,EDC)等。对于临床医生来说,Excel 和 EpiData 具有简单、好用、实用的特点,十分方便在日常的科研工作中进行简单的数据管理。

(1) 利用 Excel:Excel 是一种常见的办公软件,也可以利用它制作简易的数据文件。直接将收集到的各个指标录入到 Excel 中,就可形成一个简易的数据库。但是为了便于数据的分析,要注意数据的格式和数据编码。

(2) 利用数据管理软件 EpiData:EpiData 3.1 数据管理软件是一款免费的数据录入与管理软件,广泛应用于医学研究中需要数据采集、录入与管理的许多领域,如实验室研究数据的录入与管理、新药临床试验数据管理等。

一个典型的 EpiData 数据库包含 3 个核心文件:QES 文件,用来定义数据库的结构;REC 文件,用于录入数据、存放数据以及已经定义好的编码;CHK 文件,用来定义数据录入时进行字段控制的有效性规则,可以在一定程度上保证录入数据的质量。利用 EpiData 软件建立数据库的过程如下:

通过简单的文本来定义数据库结构(QES 文件),然后将其转化为录入界面(REC 文件),最后通过对录入设置严格的条件(CHK 文件)来进行质量控制。

对于所建立的数据库来说,准确地储存数据和导出数据是对数据管理系统的最基本的要求。此外,建立数据库需要综合考虑的其他因素有:

- 1) 符合研究方案的流程,易于数据的录入。
- 2) 数据导出的样式全面且内容完整,易于统计分析。
- 3) 数据在系统内可进行较为完整的检查,以及时发现“问题”数据。
- 4) 符合数据库应用软件的要求。

2. 确定数据库结构 确定数据库结构时应充分考虑研究的目的,符合临床操作的次序,数据录入的便利和统计分析的要求,尽可能减少数据录入中错误的发生,并保证其完整和正确。如果数据是利用调查表收集的,根据调查表来构建数据库结构比较方便。

确定 Excel 数据库的结构:在 Excel 中,通常在第 1 行中输入变量名,每一列表示一个变



量,从第二行开始,每一行表示一个观测,即一个研究对象的信息。如果不同模块的信息存储在不同的工作表(sheet)中,要注意同一研究对象用相同的ID标识,以方便查找信息与分析数据。

### 知识点

#### 变量名的设定

为了方便数据管理和分析,建议采用简单的英文变量名并保证变量名的可识别性。如血红蛋白的英文全称为 hemoglobin,因此可命名为 HGB。

变量名不要太过简单或者没有任何含义,如  $A_1, A_2, A_3$ 。也不宜赋予过长或过于复杂的变量名。为了方便数据处理,变量名不宜以数字开头,也不宜在包含空格、短横“-”、斜杠“/”等特殊字符。

可通过 Excel 的“设置单元格格式”功能对变量的类型及格式进行定义(图 16-1)。

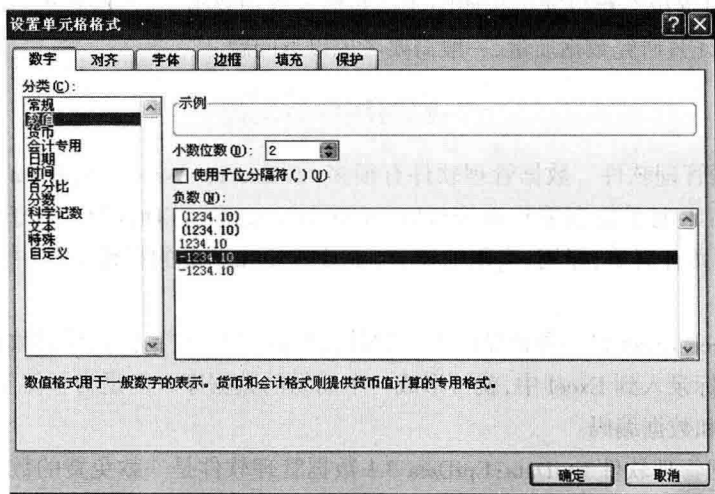


图 16-1 Excel 定义变量的类型及格式

对于数字型变量,如果是计量资料,可以按原始资料直接录入,但是要注意单位的统一,小数点位数等;如果是无序分类资料或者等级资料,则在录入前需进行编码,转化为便于计算机分析的数据,再录入到数据库中。对于字符型变量,需要设置其长度,以保证全部信息都能录入数据库。

### 知识点

#### 编 码

编码是指将信息从一种形式或格式转化为另一种形式的过程。例如,性别为无序分类资料,收集的原始资料为“男性”或“女性”。在编码时,可将“男性”转化为数值“1”,“女性”转化为“2”。通过编码,可方便数据的处理和分析。在编码时,需要对可能出现的各种情况考虑周全,例如妊娠检查结果,通常将“阴性”编码为数值“0”,“阳性”编码为“1”。但是需要考虑到绝经期妇女和男性是不适用的,因此,可将“不适用”编码为“2”。否则无法区分“不适用”与真正的缺失情况。

## 知识点

## 资料的类型

医学研究中通常将资料分成三种类型:计量资料、无序分类资料和有序分类资料。

1. 计量资料 计量资料(measurement data)又称为定量资料(quantitative data),是检测每个观察单位某项指标的汇总结果,表现为数值大小,具有度量衡单位。

根据指标(变量)的类型不同,分为连续型和离散型两种:连续型资料是连续型变量(continuous variable)的测量结果,在一定区间内的取值为无限个。如身高值(cm)、体重值(kg)。离散型资料是离散型变量(discrete variable)的测量结果,只能取整数,在一定区间内的取值为有限个,并且可以列举。如家庭人数,可以是1个、2个、3个等。

2. 无序分类资料 无序分类资料(unordered categorical data)是指变量值为某种属性,其取值无次序关系,相互独立。分为两分类和多分类两种:二分类资料(binary categorical data)的属性为两个相互对立、互不相容的类别,如性别(男性、女性)、过敏史(无、有)。多分类资料(multiple categorical data)的属性为多个互不相容的类别,如婚姻状况(未婚、已婚、离异或丧偶、其他)。

3. 等级资料 等级资料(ranked data)又称为有序分类资料(ordinal categorical data),是指变量值为某种属性,且其取值存在次序关系,具有半定量性质,表现为等级大小或程度。如文化程度(小学及以下、初中、高中、大学及以上)、疗效评价(痊愈、显效、有效、无效)、检查结果(-、+、++、+++)。

确定 EpiData 数据库的结构:在 EpiData 数据库中,通过 QES 文件构建数据库的结构。运行程序后,进入软件主界面。该界面除有菜单可供进行功能选择外,还专设了流程栏引导,以方便用户使用。要建立一个新的 QES 文件可通过调用主菜单“文件”下的“生成调查表文件(QES 文件)”子菜单项或直接点击流程栏“1 打开文件”下的“建立新 QES 文件”(图 16-2)。

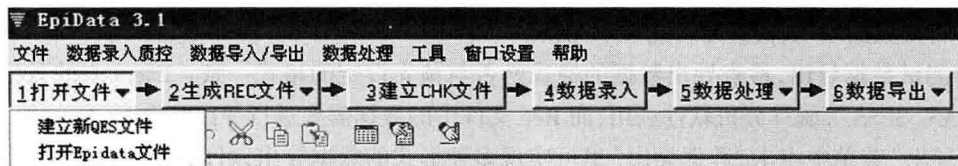


图 16-2 建立新 QES 文件

一个新的 QES 文件就类似一个空白的 txt 文件,用户可以在其中定义字段(即数据库中的变量),其格式可考虑采用以下形式:字段标签 - 字段名 - 字段类型及长度 - 注释。例如:性别这个变量的定义格式为:性别 {SEX} # (男 =1, 女 =2)。

其中,“性别”表示字段标签,即对应变量的所代表的含义。“{SEX}”表示定义字段名,即变量的名称,主要用于统计分析用。EpiData 软件只识别英文变量名,不识别中文变量名,因此,在实际工作中,对以中文为语言的调查表,最好直接指定英文字段名,而将中文变量名作为变量标签(label)。

字段类型的定义可应用软件提供的“字段编辑器”按钮实现。将鼠标置于字段名之后,在进行数据录入的位置插入字段类型。调用主菜单“编辑”下“字段编辑器”,打开“字段编辑器”对话框(图 16-3)。

软件提供的常用变量类型有数字型、文本型、日期型和特殊型等,数字型变量只允许变量值为数字,字符型变量允许变量值为中、英文字母,各种符号和数字,此时的数字被当做字符处理,





无法进行计算。“字段编辑器”可对字段的长度和形式等内容进行设定。如上文中的“#”表示“SEX”这个字段为数值型,且长度为1个字节。

同样地,对于数字型变量,需要对其中的无序分类资料或者等级资料进行编码,转化为便于计算机分析的数据。对每个编码所代表的含义可在字段注释中说明,方便数据录入和核查。如“(男=1,女=2)”说明如果录入数值为“1”,则表示为男性,如果录入数值为“2”,则表示为女性。除编码含义外,注释中还可以包括很多其他内容,如数值型变量的单位,数据录入的特殊要求等。

设计好 QES 文件后,可调用主菜单“REC 文件”中的“生成 REC 文件”子菜单项或点击流程栏“2 生成 REC 文件”下的“生成 REC 文件”(图 16-4),命名并保存,即可生成与“QES 文件”完全匹配的 REC 文件。

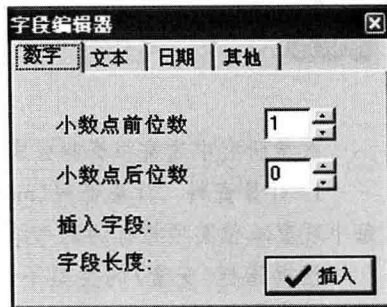


图 16-3 定义 EpiData 数据库中的字段



图 16-4 建立新 REC 文件

利用 REC 文件可录入并存储数据,但在录入数据之前,需要对所构建的数据库进行测试,以保证其中数据文件结构符合要求。

3. 数据库测试 建好的数据文件结构经过测试运行正常后才能使用。首先对数据录入界面进行检查,确认数据结构中包括了所有的变量。录入几个测试数据后,确认数据库采集数据的准确性,比如,变量的类型是否与事先设定的一致。如果是数字型变量,检查录入测试数据的位数是否符合要求;如果是字符型变量,检查事先设定的长度是否足够,可以将收集到资料中较长的字符型数据录入,检查是否能显示完整。

将测试数据导出,检查导出数据的格式是否与预先设定的格式一致。Excel 中的数据可直接被 SAS、SPSS 等统计分析软件利用,而 Rec 文件中的数据需要进行导出。在 EpiData 的“数据导入/导出”菜单中点击“数据导出”里可实现多种形式的导出数据,可以选择将数据输出为文本文件、dBase III 文件、Excel 文件、Stata 文件、SPSS 文件、SAS 文件等(图 16-5)。

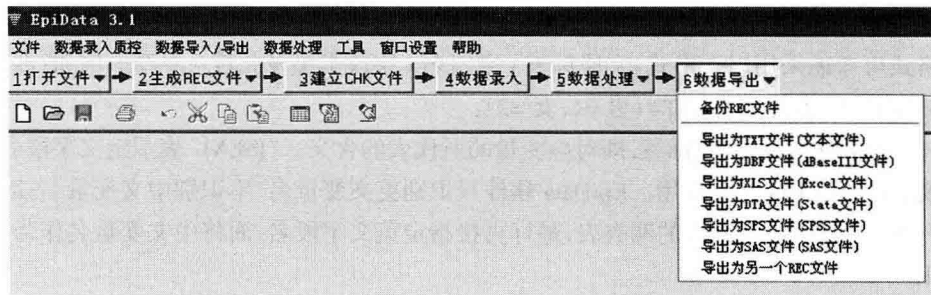


图 16-5 导出 EpiData 中的数据

## 二、如何录入数据

数据录入的基本要求是准确性,即“看见什么就录入什么”。



数据录入的方式有一次录入、二次录入等。从质量的角度来说,建议使用二次独立录入,即由两个录入人员对相同的数据分别独立地录入。双份录入比对是减少错误发生的有效方法,即用软件对由两名数据录入人员分别录入的数据文件进行比较,并打印出不一致的部分。由于两人对同一数据录错并错得一样的概率极小,因而可以保证数据录入的准确性。EpiData(validate 程序)和 SAS(proc compare)等软件具有对双份录入数据进行比较并打印出不一致部分的功能。不一致的部分将由录入人员对照原始的资料分别进行修改。

### 三、如何保证研究数据的准确性

对于 Excel 数据库,可通过 Excel 自带的“数据有效性设置”在数据录入时进行简单的质量控制(图 16-6)。

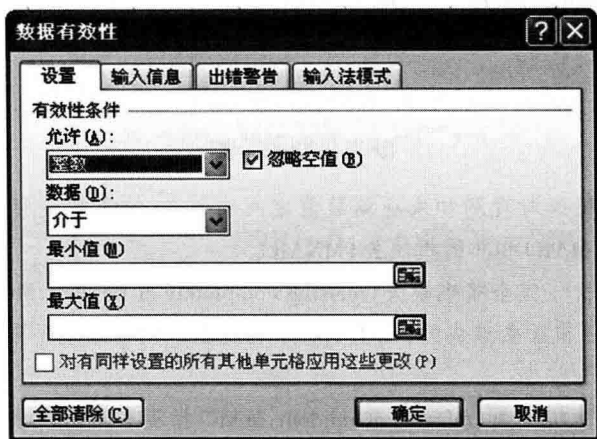


图 16-6 Excel 中的数据有效性设置

对于 EpiData 数据库,可通过 CHK 文件在数据录入时对输入的内容进行质控,如果输入内容不满足条件就无法录入,以初步保证数据的质量。调用主菜单“数据录入质控”中的“添加/修改录入质控程序”或点击流程栏的“3 建立 CHK 文件”,选择需要进行核查的数据文件,即可进入编写数据录入核查文件(CHK 文件)编辑界面(见图 16-2)。

此时会弹一个“编辑核查文件”对话框(图 16-7)。在这个对话框里,实现对该字段的输入范围、合法值、跳转、自动重复输入、设置数值标签等的定义。

利用 Excel 和 EpiData 自带的功能虽然可以在录入数据时做简单的质控,但有些复杂的数据核查无法通过上述功能实现。因此,为了保证研究数据的质量,在数据录入完毕后,常常需要做进一步的核查,核查的内容通常包括有:

**缺失数据。**检查是否有些项目没有填写,尤其注意变量在某些条件下不能缺失情况,如既往病史若填为“有”,则某些与既往病史相关的变量则不能缺失。

**极端值核查。**有些项目数据是否超出范围。例如,性别规定编码为“0,1”时写了“2”;年龄或血压等是否误写,如年龄为 655 岁,舒张压为 8mmHg 等。

**逻辑检查。**各项目的回答之间有无逻辑性错误,如舒张压大于收缩压等。妊娠检查是否有男性填为阳性。



图 16-7 利用 CHK 进行录入时质量控制



#### 四、如何处理缺失数据

缺失值是临床研究中的一个潜在的偏倚来源,因此,在研究设计和资料收集阶段应尽量避免缺失值的产生。研究中观察的阴性结果、测得的结果为零和未能测出者,均应有相应的符号表示,不能空缺,以便与缺失值相区分。

在绝大多数的研究中,数据缺失是无法避免的问题。但是在研究的计划、执行、分析和报告过程中忽略数据缺失的问题会造成潜在的偏倚。大量的缺失值将使得研究结果无法解释。在分析中直接排除有数据缺失的受试者可能会影响受试组间的可比性,破坏研究样本对于目标人群的可代表性(representativeness)。以上影响都会对研究结果造成偏倚。除此之外,对缺失值的直接排除还可能降低研究的把握度或增加变量的变异性。

处理缺失的方法目前并没有一个通用的方法,每一种处理方法都建立在对缺失机制不同的假设基础上。

##### 知识点

##### 缺失机制的分类

根据变量缺失概率与观测和未观测数值之间的关系,缺失机制可分为完全随机缺失(MCAR)、随机缺失(MAR)和非随机缺失(MNAR)。

1. 完全随机缺失 完全随机缺失(missing completely at random, MCAR)指观察对象的数据缺失完全是由随机因素造成的,既不取决于已观察到的数据也不取决于未被观察到的数据。

2. 随机缺失 随机缺失(missing at random, MAR)指观察对象缺失的概率取决于已有的观察结果,不取决于未观察到的结果。

3. 非随机缺失 非随机缺失(missing not at random, MNAR)指观察对象的缺失概率与当前尚未观察到的结果有关。

常用处理缺失数据的方法有:

1. 忽视缺失值 即只分析有完整记录的数据,只有当缺失值属于MCAR时才可以忽视这些缺失值,否则会得出有偏倚的结论。而且即使当MCAR成立时能产生无偏估计,但精确度和效率仍降低;除非理由充足合理,否则该方法不能用作主要分析。

2. 简单/单一填补 简单/单一填补(single/single imputation)法是将缺失值仅按某个填补方法结转一次,但该方法的不足是低估了数据的变异。常用的简单数据填补方法有末次访视结转(last observation carried forward, LOCF)、基线访视结转(baseline observation carried forward, BOCF)、最差病例填补(worst case imputation, WCI)、非条件均数填补(unconditional mean imputation, UMI)和条件均数填补(conditional mean imputation, CMI)。

末次访视结转是用末次观察应答填补缺失,而基线访视结转采用基线观察应答填补缺失,两者都适用于MCAR假设,倾向于得到保守的结论。最差病例填补是采用最差的结局填补缺失,例如:主要指标为有效率,临床结局表现为二分类的定性指标(有效或无效),一旦缺失,则该例患者按“无效”进行填补;如果主要指标为定量指标,则按照被观察人群中最差结果进行填补。非条件均数填补采用变量的均数来代替该变量中的每一个缺失数据,其缺点在于低估了变量的变异程度,且低估填补变量与其他变量的关联程度。条件均数填补则根据预测变量将总体交叉分层(如根据症状、年龄等分层),用该观察个体所在层的完整数据的均数来替代缺失数据,其变异程度较非条件均数填补法有所改进,但由于没有得到残差的依据,这种方法仍然低估了该变量的变异程度。



3. 多重填补 多重填补(multiple imputation, MI)为每一个缺失值构造  $m(m>1)$  个填补值, 这些值反映了缺失值的不确定性, 这样就产生  $m$  个完整数据集。然后用分析完整数据集的统计方法对这  $m$  个数据集分别进行统计分析, 再把得到的结果进行综合推断, 最终得到对目标变量的估计。多重填补方法考虑到缺失数据的不确定性, 从而结果更为可靠, 但 MI 方法仍假设数据缺失机制为非 MNAR(主要为 MAR), 对于 MNAR 的情况仍然无法处理。常用缺失数据多重填补方法可通过 SPSS 等软件实现。

此外, 还有基于似然的方法、逆概率加权法等。由于实际操作中, 既不能肯定缺失值和未被观测的结局变量之间的相关性, 也不能判断缺失数据是否能从已测值中得到很好的预测, 因此不能确定是否应将其视为 MCAR 还是 MAR。另外, 要想明确区分 MAR 和 MNAR 也很困难。研究者只能对同一份有缺失数据的资料分别进行 MAR 和 MNAR 的假设, 并在各自的假设下进行数据分析, 然后进行敏感性分析, 以比较所得结论是否一致, 结果是否稳定。

## 五、如何处理离群值

离群值(outlier)是指样本中个别值, 其数值明显偏离所属样本的其余观测值。离群值的存在将导致分析结果误差增大, 甚至可能造成相反的结果。抽样调查技术有问题, 疏忽大意记错, 或人为的虚报数据等都可能造成离群值, 这直接影响了数据的质量。因此, 出现离群值时, 首先应认真检查原始数据, 看是否存在数据质量问题或者是否从医学专业上加以合理解释。如果数据间无明显逻辑错误而原始记录又的确如此, 在实际资料分析时, 可将离群值删除前后各做一次统计分析, 进行包括和不包括离群值的两种结果比较, 研究结果是否不一致以及不一致的直接原因。注意, 不可以将离群值用其他数据替代, 如用均数替换等。

## 第三节 统计分析

使用适宜、正确的统计方法是统计结论真实可靠的重要保证。统计分析主要包括两方面: 一是统计描述, 对于数值变量资料, 常用的统计描述指标有算术均数、中位数、几何均数及标准差等; 对于分类变量资料, 主要有率、构成比、相对比(如相对危险度)等。二是统计推断, 主要目的就是用样本信息推断总体, 包括参数估计和假设检验, 这也是统计学主要目的。统计方法的正确选择常常与研究目的、资料类型、设计类型、样本大小、特定的条件综合分析等有关; 这也是大多数临床医师感到比较困惑的问题。

### 一、资料的统计描述

基本特征描述, 包括研究对象的人口学特征: 如性别、年龄、文化程度、婚姻状况、职业分布特点、民族构成等; 比如糖尿病, 一般需要描述研究人群的年龄构成、性别比、职业特征等。

描述某种现象或事物的特征: 如糖尿病的类型、病情、病程等构成情况, 另外还有临床生理生化指标, 包括一些饮食习惯等。

### 二、主要研究项目的描述性统计分析

调查研究或实验研究都需要对主要观察项目进行描述性统计分析(descriptive statistical analysis), 以便使读者了解主要观察指标的高低及分布情况。

根据研究目的不同, 通常描述性统计分析多从疾病或健康状况、研究因素等方面进行描述。

若研究的是某种疾病, 计算某病的发病率、患病率、死亡率、病死率、病型构成等(发病率和患病率是两个完全不同的概念, 研究者要加以区分, 而且它主要用于计算某种疾病在人群中的发病率和患病率)。



若观察的是心理、行为、生命质量等指标,则可以考虑计算心理或行为问题的发生率;若利用标准的心理行为问卷进行评价,某项心理或行为特征的得分又呈正态分布,则可以计算得分的平均值和标准差,如果不呈正态分布,则计算中位数和四分位数间距。

对健康指标还可以进一步描述其分布特征,如描述某种疾病的时间、空间分布、不同特征群体间的分布特征等。

通过对这些研究项目的分析可使研究者了解疾病、健康状况及其分布情况,为病因探索提供依据。

### 三、资料的推断性统计分析

各种分析性研究的结论往往是从组间比较中得出的。按照实验设计的基本原理,只有当各比较组人群之间研究因素以外的各种与研究结果有关的因素在组间都均衡可比时,所得研究因素在组间有无差异的结论才有说服力。因此需要进行资料的组间均衡性检验。

临床上常需对某些特征(一般多为主要的混杂因素,如平均年龄、性别构成、病情构成等)进行比较组间的均衡性检验(差异显著性检验)。比如要研究新药治疗糖尿病是否高于旧药,在研究设计时,我们要遵循统计设计的基本原则,要求被比较的两组患者要具有可比性,但是否可比,有时候仍然需要通过假设检验进一步验证,这样才保证两组样本具有可比性,结论也更有说服力。通常,单因素显著性检验方法即可满足均衡性检验的需要。

例如:某研究者对某种违法行为的人进行预防复发的干预试验,评价研究结果时,首先对试验组和对照组的人口学特征进行比较。从表 16-1 中可以清楚地看到试验组和对照组的性别构成、平均年龄和婚姻构成方面是均衡可比的。

表 16-1 资料的组间均衡性检验

| 变量类型           |    | 试验组 (n=54)  | 对照组 (n=56)  | t 或 $\chi^2$ | P     |
|----------------|----|-------------|-------------|--------------|-------|
| 年龄(数值变量)       |    | 31.24 ± 4.6 | 32.86 ± 5.8 | 1.612        | 0.198 |
| 性别<br>(分类变量)   | 男性 | 32 (59.3%)  | 35 (62.5%)  | 0.121        | 0.728 |
|                | 女性 | 22 (40.7%)  | 21 (37.5%)  |              |       |
| 婚姻状况<br>(分类变量) | 已婚 | 11 (20.4%)  | 11 (19.6%)  | 0.376        | 0.829 |
|                | 未婚 | 29 (53.7%)  | 33 (59.8%)  |              |       |
|                | 离异 | 14 (25.9%)  | 12 (21.4%)  |              |       |

对于参数检验统计分析思路,一般情况下要求所研究的资料符合参数检验的要求。如进行两个独立样本的  $t$  检验,要求两个样本均来自于正态分布的总体,而且要方差齐;当不满足正态性要求时,一般是先进行变量变换,使其满足正态性和方差齐的要求,再不满足条件,则可以考虑用非参数检验的方法做假设检验(表 16-2),任何一个参数检验方法,在非参数检验中都可以找到对应的方法。

表 16-2 不同资料类型单因素统计分析方法比较

| 资料类型                 |      | 适用条件 |        | 两组      | 多组    |
|----------------------|------|------|--------|---------|-------|
| 定量资料(两样本或多样本总体均数的比较) | 正态分布 | 方差齐  | $n$ 较大 | $u$ 检验  | 方差分析  |
|                      |      |      | $n$ 较小 | $t$ 检验  |       |
|                      | 偏态分布 | 方差不齐 |        | $t'$ 检验 | 非参数检验 |
|                      |      |      |        |         | 秩和检验  |

学  
习  
笔  
记





|               |                              | 续表              |    |
|---------------|------------------------------|-----------------|----|
| 资料类型          | 适用条件                         | 两组              | 多组 |
| 分类资料(两样本率的比较) | $n \geq 40$ 且 $T \geq 5$     | $\chi^2$ 检验     |    |
|               | $n \geq 40$ 且 $1 \leq T < 5$ | 校正的 $\chi^2$ 检验 |    |
|               | $n < 40$ 或 $T < 1$           | Fisher 精确概率法    |    |
| 等级资料          |                              | 秩和检验            |    |

四、比较性研究的数据分析

1. 标准比较法 标准比较法,即把研究对象的某一特征与公认的,或者正式颁布的标准进行比较,即样本与总体的比较。比如,空气、水、食物中有害物质的浓度是否超过标准;某个人群的身高体重是否与一般人群不同等(表 16-3)。

表 16-3 不同资料类型适用条件及其统计分析方法

| 资料类型 | 适用条件                        | 统计分析方法      |
|------|-----------------------------|-------------|
| 计量资料 | 正态分布 $n$ 较大或 $\sigma$ 已知    | $u$ 检验      |
|      | $n$ 较小且 $\sigma$ 未知         | $t$ 检验      |
| 计数资料 | $n$ 较小                      | 直接计算概率法或查表法 |
|      | $n$ 较大, $p$ 与 $(1-p)$ 均不太小时 | $u$ 检验      |

2. 两两比较法(表 16-4)

表 16-4 不同研究类型适用条件及其统计分析方法

| 研究设计类型 | 适用条件      | 统计分析方法    |
|--------|-----------|-----------|
| 配对设计   | 正态分布, 方差齐 | 配对 $t$ 检验 |
| 完全随机设计 | 正态分布 方差齐  | $t$ 检验    |
|        | 方差不齐      | $t'$ 检验   |
|        | 偏态分布      | 秩和检验      |

3. 多组比较法 根据研究需要,有时要求同时比较某个变量在多个组之间的差异。如比较某种药物对多个不同群体的疗效。

对于计量资料,若为完全随机设计,当任何观察值都是独立地来自具有等方差的正态总体,采用完全随机设计的单因素方差分析;若为随机区组设计,则采用随机区组设计的方差分析。方差分析只能检验出各组均数间的差别的总的信息,为了得到哪几个组均数之间的差别具有或不具有统计学意义,需要进行多个样本均数间的两两比较,常用的方法有 LSD- $t$  检验、Dunnnett- $t$  检验、SNK- $q$  检验等。

重复测量资料是同一个受试对象的同一个观测指标在不同的时间点进行多次测量所获得的资料,常用来分析该观测指标在不同时间点上的变化特点。这种类型的资料在临床研究中是比较常见的。如临床上为指导脑血管病患者的治疗和预后,某研究人员对不同类型脑血管患者的酸性磷脂(AP)在不同时间点上的变化进行了观察,随机选取三种不同类型的脑血管病患者(短暂性脑缺血发作(TIA)、脑血栓形成、腔隙性脑梗死)各 12 例,分别于脑血管病发作的第 24 小





时、48 小时、72 小时、7 天时采血,测量不同时间点血液中 AP 的数值,看是否有不同?该资料是重复测量资料,可以考虑重复测量资料的方差分析。两因素的重复测量资料既可以分析两因素(处理和时间)各水平间是否不同,还可以分析处理因素和时间因素的交互作用。重复测量资料分析要求满足球形对称的要求,不满足时,应采用球对称系数  $\epsilon$  对自由度进行校正。详细分析方法可以参考相关的统计书。

### 知识点

#### 数据分析统计方法选择

一般情况下,参数检验方法需要满足一定的条件才可以应用相应的统计分析方法,当不满足的时候,有时候还想用参数检验的方法,此时可以考虑变量变换方法。变量变换是应用最广泛的分析思路,其结果利于理解,如果找到合适的变换方式,则随后能够使用的分析模型非常丰富。但是在实际分析过程中,这种合适的变换方式有可能是不存在的。另一方面,变量变换改变了变量间的数量关联趋势,特别是在采用过于复杂的变量变换方法时,有可能严重扭曲了数据原本蕴含的信息,因此一般不推荐使用过于复杂的变量变换方法,且相应的分析结果在解释的时候需要相当谨慎。

对于分类变量资料多组率或构成比的比较,可采用“ $2 \times K$ ”表或“ $R \times C$ ”表的卡方检验资料的卡方检验,但应注意理论数小于 1 的格子数不能超过 1/5,否则应采取增加样本含量等措施。

## 五、关联性研究的数据分析

确定事物间的关系,定量表述事物间关系的方向、大小或强弱。如分析儿童的年龄与身高之间的关联性,吸烟与肺癌之间是否相关,结核病人耐药是否和服药时间,服药依从性有关等。两个变量关联性分析的目的在于推断从某一总体中随机抽取的同一份样本观测出的两个变量间是否存在关联性,以及这种关联性的密切程度如何?但需要注意的是这种关联并不表示专业上的因果关联。

对于计量资料,如果两个变量都服从正态分布,若想研究两个变量间是否存在相关性,应首先做散点图,如果散点图呈线性趋势,则可以计算 Pearson 相关系数,并检验相关系数的显著性;如果变量不服从正态分布,可先进行变量变换,使其满足正态性的要求,再计算 Pearson 相关系数,如果不满足正态性要求,则采用 Spearman 秩相关系数来分析其关联性;如果研究一个自变量对一个应变量影响的数量依存关系则用直线回归分析方法,首先也需要先做散点图,如果有线性趋势,则计算回归系数与截距,建立回归方程并进行显著性检验;若通过绘制散点图发现两者的关系不呈直线关系,可考虑拟合非线性回归(秩回归或曲线回归)。

对于两个反映属性的分类变量,可以做交叉分类的频数表,如果两个变量为双向无序的两分类或多项分类资料( $R \times C$  表),除了采用卡方检验推测有无关联,还可以利用卡方值计算 Pearson 列联相关系数反映关联密切程度;如果要检验 ABO 血型系统(A、B、O、AB 型)和 MN 血型系统(M、N、MN 型)有无关联,为双向无序分类变量,可以用卡方检验检验两种血型系统有无关联,如果有关联,进一步计算 Pearson 列联(关联)系数  $r_p, r_p$  取值范围在 0~1 之间,愈接近于 1,关系愈密切。对于双向有序即等级资料,可以采用 Spearman 等级相关或 Kendall 等级相关分析;有序分组资料可采用线性趋势检验。

## 六、多因素分析

多因素分析又称为多变量分析或多元统计分析方法,可一次观察多个变量分别与因变量之间的独立关系以及变量之间的交互作用,在临床研究中被广泛应用于危险因素的筛选、诊断试



验、防治效果、疾病的预后及风险评估等。由于医学现象的复杂多样性,如疾病的发生、发展往往受到多种因素的综合作用,而各种因素间也存在着各种或多或少的关联,因此需要从多个角度出发,应用多因素分析,探讨疾病的发生、发展过程,并进行疾病的预后评估。

通常先将所有变量做单因素分析,筛出可能的影响因素( $P$ 值小于某个检验水准的变量)纳入多因素分析。此时的单因素分析不要求比较组间的均衡,因为混杂因素的干扰可在多因素分析中被排除,但要求放宽纳入变量的多因素分析标准,如规定 $P$ 值小于0.10或0.15者可进入多因素分析,以防负混杂的干扰。该类分析方法多种多样,具体使用时一定要根据资料的特点、研究目的来选择适当的方法。主要包括:

1. 研究多个变量之间的相互关系 多个自变量与一个因变量的相关关系,可以考虑多元相关分析,计算复相关系数和偏相关系数并进行假设检验。但在医学研究中,常常会碰到两组变量之间的线性相关性问题;如研究患者的各种临床症状与所患各种疾病之间的线性相关性,如果是一组自变量(体质指标)与一组因变量(体能指标)之间的关系,则考虑典型相关分析。

2. 研究多个变量之间的依存关系 多元线性回归在医学领域中的应用广泛,如确定多个指标变量与一个反应变量之间的线性关系。如温度、湿度以及大气污染物与发病率的关系;筛选疾病的危险因素和有利于健康的健康促进因素,如肿瘤、冠心病等疾病的病因研究中,应用多元线性回归分析,可以从众多的相关因素中筛选出主要的危险因素,并估计出这些危险因素的相对重要程度,同时通过已经建立的回归方程可以预测其他同类型患者患病风险。多元线性回归分析要求因变量是连续的正态分布变量,且自变量和因变量呈线性关系;当因变量是分类变量,且因变量与自变量无线性关系时,则线性回归模型的假设条件被破坏,这时最好的回归分析模型是 Logistic 回归模型;Logistic 回归主要用途是筛选危险因素、分析各危险因素作用大小、调整或校正混杂因素及预测和判别。医学研究者经常关心的问题如哪些因素导致了人群中有的人患胃癌而有的人不患胃癌?哪些因素导致了手术后有的人感染,而有的人不感染?哪些因素导致了某种治疗方法出现治愈、显效、好转、无效等不同的效果?这些因素的作用如何?这时候 Logistic 回归模型可以回答上面的这些问题。详细分析过程可以参考相关的统计书籍。

3. 研究多个变量间的内在结构 对于涉及多因素的医学研究,常常需要了解所研究的多个变量之间的内在联系。通过变量内在结构的分析,就可以把这些变量依据它们内在的关系,客观地归纳为若干类。如聚类分析和判别分析。聚类分析是将一批样本或变量按照他们在性质上的亲疏程度加以分类;其实质是按照距离的远近或相似性大小将数据分为若干个类别,以使类别内数据“差异”尽可能的小,类别间的“差别”尽可能的大。判别分析则是用于判别样品所属类别的统计分析方法,它所解决的问题是在一些已知研究对象用某种方法分成若干类的情况下,建立分类规则用以确定新的样品归属于已知类中的哪一类。例如,某研究者为了进行体质研究,从12岁男生中随机抽取200人,分别测量了8个形态指标,他们是身高( $x_1$ ),体重( $x_2$ ),肩宽( $x_3$ ),骨盆宽( $x_4$ ),胸围( $x_5$ ),大腿围( $x_6$ ),上臂围( $x_7$ ),和小腿长( $x_8$ ),将样本个体或指标变量按其具有的特性进行分类,可以用聚类分析。聚类的结果可能将8个指标分成两大类或三大类。再如临床医师需要根据患者的一系列症状、体征及检查结果来判断某患者究竟患了哪种疾病,可以考虑判别分析。判别分析在医学研究中有非常广泛的用途,主要用于疾病的计算机辅助诊断系统研究中。方法是从有足够临床经验的老专家、老教授那里搜集一批已经确诊的患者和正常人的有关资料,如临床疾病诊断、化验诊断、X线片诊断、心电图、脑电图、B超、CT诊断等诊断指标建立判别函数,然后利用建立的判别函数为尚未确诊的患者病情提供一个有效的参考意见;除此之外,判别分析还可以用于慢性疾病的早期预防、手术预后估计、病因研究等。

4. 研究变量之间的关系网 事物的特征通过不同的变量体现出来。因此客观事物之间的关系就体现为多变量之间的、无比复杂的网状关系。对于这类分析,可以借用路径分析方法和结构方程分析方法。



## 七、资料的重复性评价

临床医生常常根据患者的临床症状或各种特殊检查结果对疾病或预后做出判断。无论采用单指标或多个指标(综合指标)作为诊断或分析的依据,都可能出现判断不一致的情况。例如,阅读同一个人的X线片,诊断患者是否患某种疾病,不同的医生可能出现诊断不一样的情况,即使是同一个医生,两次观察或多次观察都有可能出现判断结果不一致的情况。如果要评价不同医生对同一批患者的判断结果,或同一个医生先后两次或多次判断是否一致,或者在调查问卷中,同一批人针对同一个问卷,两次调查应答结果是否一致,常常使用 Kappa 统计量,Kappa 值愈大表示一致性越好。其主要用于分析计数资料(定性或多项分类)和等级分组资料一致性的方法。计量资料,则可用相关系数表示两次调查结果的符合程度。相关系数越大,一致性越好。

## 1. Kappa 系数及其假设检验

(1) Kappa 系数( $K$ ):适用于无序分类变量资料,包括两项分类和多项分类资料。用两种方法分别对  $N$  个观察对象逐一判断其属于  $C$  类别中的哪一类,其判断结果常以  $C \times C$  列联表的形式表示,见表 16-5。

表 16-5  $C \times C$  列联表

| 方法(1) | 方法(2)    |          |          |       |          | 合计率   |       |
|-------|----------|----------|----------|-------|----------|-------|-------|
|       | 1        | 2        | 3        | ..... | $C$      |       |       |
| 1     | $A_{11}$ | $A_{12}$ | $A_{13}$ | ..... | $A_{1c}$ | $A_1$ | $a_1$ |
| 2     | $A_{21}$ | $A_{22}$ | $A_{23}$ | ..... | $A_{2c}$ | $A_2$ | $a_2$ |
| 3     | $A_{31}$ | $A_{32}$ | $A_{33}$ | ..... | $A_{3c}$ | $A_3$ | $a_3$ |
| ...   | ...      | ...      | ...      | ..... | ...      | ...   | ...   |
| $C$   | $A_{c1}$ | $A_{c2}$ | $A_{c3}$ | ..... | $A_{cc}$ | $A_c$ | $a_c$ |
| 合计率   | $B_1$    | $B_2$    | $B_3$    | ..... | $B_c$    | $N$   |       |
|       | $b_1$    | $b_2$    | $b_3$    | ..... | $b_c$    |       |       |

根据表 16-1, Kappa 系数的计算公式为,  $\kappa = \frac{p_o - p_e}{1 - p_e}$ , 其中,  $p_o = \frac{\sum A_{ii}}{N}$ ,  $p_e = \sum a_i b_i / N$ ,  $a_i = \frac{A_i}{N}$ ,  $b_i = \frac{B_i}{N}$ 。式中,  $p_o$ 、 $p_e$  分别为两种测定方法结果的观察一致率和机遇一致率;  $p_o - p_e$  为实际一致率;  $1 - p_e$  为非机遇一致率;  $A_{ii}$  为  $C \times C$  列联表中主对角线上的实际值;  $N$  为总例数;  $A_i$ 、 $B_i$  分别为第  $i$  行、第  $i$  列的边际值;  $a_i$ 、 $b_i$  分别为第  $i$  行、第  $i$  列的边际频率。因期望一致率是假设两次调查或两种测定方法相互独立的前提下所期望的一致率,故按概率乘法定理计算。

(2) Kappa 值的假设检验:根据实际资料计算的  $\kappa$  值只是一个样本的统计量,存在着抽样误差,因而,所计算的  $\kappa$  值是否来自  $\kappa$  值为“0”的总体(即两者之间的一致程度是由于机遇造成的),应当经过假设检验( $u$  检验),检验公式为:

$$u = \frac{\kappa}{Se(\kappa)}, Se(\kappa) = \sqrt{\frac{p_o(1-p_o)}{N(1-p_e)^2}}$$

式中,  $u$  为标准正态分位数,  $Se(\kappa)$  为  $\kappa$  的标准误。

一致性强度的参考判断指标。

Kappa 值取值范围是  $|\kappa| \leq 1$ 。  $\kappa = -1$  时表示完全不一致,  $-1 < \kappa < 0$  时表明观察一致性小于机遇一致性,无意义;  $\kappa = 0$  时表示一致性完全由机遇造成;  $\kappa = 1$  时表示分类结果完全一致; Kappa 值究竟多大才具有实际意义,需要根据具体情况而定。Landis 和 Koch 将 Kappa 系数的大小划分了六个区段,分别代表一致性的强弱程度。当  $\kappa < 0$  时,一致性强度极差; 0.0~0.2,微弱; 0.21~0.40,



弱;0.41~0.60,中度;0.61~0.80,高度;0.81~1.00,极强。

例1 时间分辨荧光免疫分析(TRFIA)是近年发展起来的免疫测定方法,为评价其在检测HBV血清标志物中的临床应用,采用与国际上认可的微粒子酶免分析法(MEIA)对比的方法,对86例HBV感染者和50例正常人血清进行了抗2HBC检测,结果如下:TRFIA法阳性95例,阴性41例,MEIA法阳性96例,阴性40例,两种法均阳性92例,均阴性37例(表16-6)。试问两种方法检测结果一致性如何?

表 16-6 TRFIA 法和 MEIA 法检测 136 例血清抗 2HBC 结果

| TRFIA | MEIA  |       | 合计  | 率     |
|-------|-------|-------|-----|-------|
|       | +     | -     |     |       |
| +     | 92    | 3     | 95  | 0.699 |
| -     | 4     | 37    | 41  | 0.301 |
| 合计    | 96    | 40    | 136 |       |
| 率     | 0.706 | 0.294 |     |       |

$$\begin{aligned} p_o &= \frac{\sum A_{ii}}{N} = \frac{A_{11}+A_{22}}{N} = \frac{92+37}{136} = 0.949, \\ p_e &= \sum a_i b_i = a_1 b_1 + a_2 b_2 = 0.699 \times 0.706 + 0.301 \times 0.294 = 0.582 \\ \kappa &= \frac{p_o - p_e}{1 - p_e} = \frac{0.949 - 0.582}{1 - 0.582} = 0.878 \\ H_0: \kappa &= 0, H_1: \kappa > 0, \alpha = 0.05 \\ Se(\kappa) &= \sqrt{\frac{p_o(1-p_o)}{N(1-p_e)^2}} = \sqrt{\frac{0.949(1-0.949)}{136(1-0.582)^2}} = 0.0451 \\ u &= \frac{\kappa}{Se(\kappa)} = \frac{0.878}{0.0451} = 19.47 \end{aligned}$$

因为  $u$  为 19.47, 大于 95% 标准正态分位数 1.96, 故  $P < 0.05$ , 拒绝  $H_0$ , 接受  $H_1$ , 可认为两种测定方法结果具有一致性, 且根据参考判断指标, 其一致性强度为极强。

例2 为评价 UF100 尿沉渣分析仪在筛检尿路细菌感染中的可靠性, 以中段尿定量细菌培养为金标准, 用该仪器对 100 份中段尿标本进行细菌检测, 结果见表 16-7。

本例  $\kappa = 0.334$ ,  $Se(\kappa) = 0.0970$ ,  $u = 3.44$ ,  $P < 0.05$ 。而如果单纯以符合率(即观察一致率)来计算,  $p_o = \frac{\sum A_{ii}}{N} = 0.680$ , 即有 68.0% 的符合率; 若以配对资料的  $\chi^2$  检验来进行统计分析, 则  $\chi^2 = 0.281$ ,  $P > 0.05$ , 两者差异无统计学意义。以上两种统计结果, 似乎均可认为 UF100 的检测结果较可靠, 但前者究竟有多少是由于机遇造成的呢? 后者究竟存在多少假阳性和假阴性呢? 可用 Kappa 检验加以解决, 本例  $\kappa = 0.334$ , 提示有真正一致的成分, 但一致性强度较弱, 故可认为 UF100 尿液分析仪的细菌检测指标不够可靠。

表 16-7 UF100 尿液分析仪和细菌培养对 100 份中段尿的细菌检测结果

| UF100 | 细菌培养  |       | 合计  | 率     |
|-------|-------|-------|-----|-------|
|       | +     | -     |     |       |
| +     | 24    | 18    | 42  | 0.420 |
| -     | 14    | 44    | 58  | 0.580 |
| 合计    | 38    | 62    | 100 |       |
| 率     | 0.380 | 0.620 |     |       |

学  
习  
笔  
记



2. 加权 Kappa 系数及其假设检验 加权 Kappa 系数(weighted kappa,  $K_w$ )对于有序分类变量,应引入被第一种方法判为第  $i$  类,被第二种方法判为第  $j$  类的权重  $W_{ij}$ 。线性权重时,  $W_{ij}=1-\frac{|i-j|}{c-1}$ ,平方权重时,  $W_{ij}=1-\frac{(i-j)^2}{(c-1)^2}$ ,加权后,  $p_o(W)=\frac{\sum \sum A_{ij}W_{ij}}{N}$ ,  $p_e(W)=\frac{\sum \sum a_i b_j W_{ij}}{N}$ ,  $K_w=\frac{p_o(W)-p_e(W)}{1-p_e(W)}$

假设检验

$$u=\frac{K_w}{Se(K_w)},$$

$$Se(K_w)=\sqrt{\frac{\sum \sum a_i b_j (W_{ij}-\sum a_i W_{ij}-\sum b_j W_{ij})^2-p_e(W)^2}{2N[1-p_e(W)^2]}}$$

式中,  $Se(K_w)$  为  $K_w$  的标准误。

例 3 甲乙 2 名检验师同时对 34 份痰液标本做抗酸杆菌涂片镜检,结果见表 16-8。问 2 人镜检结果一致性如何?

表 16-8 甲乙 2 名检验师对 34 份痰液标本的抗酸杆菌检测结果

| 甲检验师 | 乙检验师  |       |       |       | 合计 | 率     |
|------|-------|-------|-------|-------|----|-------|
|      | -     | +     | ++    | +++   |    |       |
| -    | 4     | 0     | 1     | 0     | 5  | 0.147 |
| +    | 0     | 5     | 1     | 0     | 6  | 0.176 |
| ++   | 0     | 1     | 14    | 2     | 17 | 0.500 |
| +++  | 0     | 1     | 2     | 3     | 6  | 0.176 |
| 合计   | 4     | 7     | 18    | 5     | 34 |       |
| 率    | 0.118 | 0.206 | 0.529 | 0.147 |    |       |

学  
习  
笔  
记

本例为有序分类变量资料,故应用加权 Kappa 系数来进行统计分析。采用线性权重。具体计算步骤较繁琐,此处略,有兴趣的可参考相关的参考文献。

## 八、研究结果的解释

$P$  值( $P$  value)是假设检验中的一个概率,它只能解释该统计分析结果支持无效假设成立的概率有多大。 $P$  值越小,说明当前样本的证据越趋向于拒绝  $H_0$ ,当  $P$  值小于等于事先规定的检验水准  $\alpha$  时,就拒绝  $H_0$ 。推翻无效假设的概率就越大,即接受备择假设的概率就越大。

$P$  值的大小不仅与总体参数间的差别有关,而且与抽样误差大小有关。不能认为  $P$  值越大,总体参数间的差别就越大。 $P$  值越小,说明实际观测到的差异与  $H_0$  之间不一致的程度就越大,越有理由拒绝  $H_0$ ;假设检验只能做出拒绝或不拒绝  $H_0$  的定性结论,但不能给出总体参数间差别大小的结论。

无论在哪个水平上接受无效假设(null hypothesis)还是备择假设(alternative hypothesis),都意味着要犯第一类错误(type I error)或第二类错误(type II error)的风险。拒绝  $H_0$ ,有可能犯第一类错误(type I error),用  $\alpha$  表示,不拒绝  $H_0$ ,有可能犯第二类错误(type II error),用  $\beta$  表示。检验水准  $\alpha$  是研究者根据研究目的,在研究初期就已经确定好的。比如检验水准  $\alpha$  确定为 0.05,那么下结论时就按照  $\alpha=0.05$  的水准拒绝或不拒绝  $H_0$ ;如果在  $\alpha=0.05$  水平上拒绝  $H_0$ ,也就是说,在  $\alpha=0.05$  水平上认为某治疗方法有效的同时,该疗法还有 5% 的可能是无效的。有些研究者为了得到有统计学意义的结论,随意篡改检验水准,统计学上是绝对不允许的。但可以通过增加样本含量的方法减小抽样误差,增加检验效能,使假设检验的结论更可靠。因为在检验水准确





定后,随着样本含量的增加,抽样误差减小,结论有可能从小样本时的不拒绝  $H_0$ ,变为大样本时的拒绝  $H_0$  的情况。还有可能出现在对同一份资料做假设检验,双侧检验不拒绝  $H_0$ ,而单侧检验则有可能拒绝  $H_0$  的情况。因此在  $P$  值大小和  $\alpha$  值大小非常接近时,下结论尤其要慎重。因此在报告结果时,要同时给出检验水准  $\alpha$  的大小,单双侧检验的选择,样本量大小等的信息,以便研究者判断结论的可靠程度。

1. 统计学意义与临床意义 统计学意义的标准是明确的,通常情况下,如果  $P \leq 0.05$  就认为差异有统计学意义,但对于实际意义的判断,不同专业背景的人从不同的角度给出的结论可能不同。例如通过控制体重和合理膳食降血压是一种有益的方法,临床实验结果平均降低 4mmHg,经检验有统计学意义。对于临床医生来讲,降低 4mmHg 没有任何明显的临床意义,因为临床上一般降低 5mmHg 才具有实际的临床意义,但降低 4mmHg 可能具有公共卫生的意义。

2. 如何正确选择分析模型 对于一个临床问题,可能用不同的方法解决。“不同的分析方法”指的是方法学体系上的确有差别的方法,而且很多统计方法虽然形式不同,但其本质是等价或基本是等价的,如方差分析模型、线性回归模型等都可以被统一在一般线性回归模型框架之下;在设定正确的时候,可以得到完全等价的结果。当方法不同的时候,这些方法可能得到不一致的分析结果,此时应当首先根据数据的特征及研究目的,考虑哪一种方法更合适,或者说根据数据结构特征和研究目的,从而判断哪种分析结果更切合实际。

3. 数据分析对样本量的要求 正确估计样本含量是临床科研设计中的一个重要内容,任何一个统计模型都会有样本量要求的问题。若样本含量过少,所得结果往往不稳定,检验效能过低,结论缺乏充分依据;若样本量过大,又会增加临床研究的难度,造成不必要的人力、物力、财力和时间上的浪费。样本含量就是在保证科研结论具有一定可靠性条件下,确定的最少观察例数。

估计样本含量可借助公式计算,而且根据研究目的不同,设计类型不同,样本量估计方法也不同。样本量的计算有专门的软件和计算公式,有兴趣的读者可以参考相关的软件。一般情况下,设计越复杂,方法越复杂,样本量计算公式就越复杂。根据经验,在建立回归模型时,样本量应当在希望加入模型的自变量个数的 10~20 倍以上为宜。在样本量不足的情况下,数据信息无法满足建模需求,必须在数据信息量、模型的复杂程度和结果的精确度之间加以权衡。因此研究者只能充分利用信息,不可能凭空制造信息,最好的办法就是放弃过于复杂的模型,采用相对简单、参数较少的模型得到更为稳妥的分析结果。因为简单的模型需要估计的参数相对较少,可以保证模型估计时有较多的冗余信息可以利用,使各参数的估计更为稳健,不至于因为个别案例的增减而导致模型出现显著地变化。

### 知识点

#### 样本量估计注意事项

1. 多组设计时,一般要求各组间的样本含量相等,只在某些特殊情况下才考虑各组样本含量不等。

2. 多种样本含量估计方法相结合,如确定医学参考值范围时:①要求  $n$  应大于 100 例;②同时  $n$  是指标个数的 20 倍;若采用计算方法进行估计时,可多做几种估算方案,以便选择,如粗估样本率可以取几种不同的值作估算。

3. 由于估算的样本含量是最少需要量,在受试者中可能有不合作者、中途失访、意外死亡等情况出现,都会减少有效观察对象,故进行试验时尚须增加 10%~20%。

4. 提高试验效果(检验效能)的一般方法 一是设法缩小总体范围,减少个体变异,如比较吸烟与不吸烟的肺功能时,采取同年龄、同性别比较等;二是选择客观指标,如数值变





量(计量指标、多变量综合指标等);三是选择较优设计方案,严格控制试验条件,如配对设计、交叉设计、随机区组设计等。

5. 根据研究目的严格选择相应的样本含量估算方法,如单、双侧不同,估计参数假设检验不同;样本含量的估计要与使用的统计方法的条件相结合;如方差分析要求方差齐,无序分类资料(计数资料)一般要求  $np \geq 5$  ( $p$  为样本率),或  $n(1-p) \geq 5$ 。对现有  $p > \alpha$  ( $\alpha=0.05$ ) 的假设检验,欲加大样本含量进行再试验,设计时应注意检验效能及以前的样本含量,如  $n \geq 400$  或  $1-\beta \geq 0.80$  时建议终止试验。

6. 同时有几个结果指标时,以主要结果为主;当有多种方法估计的结果时,取最大者为最终样本含量估计值。

### 思考题

1. 一般临床资料来源有哪些?
2. 介绍关联性研究统计分析思路。
3. 如何正确理解  $P$  值的含义?
4. 调查表分开放式和封闭式,各有何优缺点?
5. 如何做关联性研究的数据分析?

(李秀君 贺佳 张新佑)

## 第十七章 论文撰写、投稿与修改

### 写作案例

为了分析 TGF- $\beta$ 1 与瘢痕疙瘩的关系,某医生对瘢痕疙瘩组织标本进行检测。取 19 例瘢痕疙瘩患者(A 组)和 19 例正常个体的真皮组织(B 组)进行分析。采用免疫组化方法检测各组 TGF- $\beta$ 1 的表达。发现 A 组,TGF- $\beta$ 1 标本阳性 16 例,阴性 3 例;B 组,TGF- $\beta$ 1 标本阳性 2 例,阴性 17 例,比较其差异有统计学意义( $P<0.01$ )。说明 TGF- $\beta$ 1 在瘢痕疙瘩呈高表达,TGF- $\beta$ 1 与瘢痕疙瘩发病相关。以此为例,我们剖析医学论著的写作技巧、投稿及修稿的注意事项。

### 第一节 医学论文的写作技巧

医学论文是科技论文的重要组成部分,是报道医学科学研究创新性成果的阵地,也是阐述原始研究结果并公开发表的书面报告,以论著最为重要。医学论著的结构一般由题目、摘要(中、英)、引言(前言)、正文、参考文献和致谢等部分组成。

1. 确定论文的题目 本例文的题目拟定为:“TGF- $\beta$ 1 在瘢痕疙瘩组织中的表达及其意义”较为适宜。通常情况下,题目应该用简洁、准确的词语反映文章最重要的特定内容,最好不超过 20 个字(知识点)。

2. 论文摘要的要求 论文摘要一般包括目的、方法、结果和结论四部分,本例文摘要可书写为,目的:探讨 TGF- $\beta$ 1 在瘢痕组织中的表达情况及其意义。方法:应用免疫组化技术检测瘢痕疙瘩组织及对照真皮组织中 TGF- $\beta$ 1 表达。结果:TGF- $\beta$ 1 在瘢痕疙瘩组织中的表达水平明显高于正常真皮组织。结论:TGF- $\beta$ 1 在瘢痕疙瘩组织中呈高表达,提示 TGF- $\beta$ 1 可能参与瘢痕疙瘩的发病机制。从中可以看出,论文的摘要应该着重反映论文的创新点和独到见解。通常由目的、方法、结果和结论组成;以第三人称书写,不列图标,不引用参考文献,避免主观评论和解释(知识点)。

3. 论文引言(前言)的写作特点 论文前言(亦称引言)是论文的开头一段,也是论文的开场白,要求开门见山,简明扼要,以介绍研究背景、范围和研究目的为主,力图说明目前研究的热点及存在的难点问题,引出本文主题(知识点)。引言顾名思义给读者以引导,切忌不要写成讨论、综述和回顾,不宜过长。本例文的引言可如下书写: TGF- $\beta$ 1 是具有调节细胞生长、分化的一类多功能细胞因子,在创伤愈合、胚胎发育、细胞外基质形成、骨的重建、免疫调节等过程中发挥重要作用(交待出宏观知识背景);由于瘢痕疙瘩属于创伤愈合异常所致,因此推测 TGF- $\beta$ 1 可能参与瘢痕疙瘩的发病机制,为了证明 TGF- $\beta$ 1 与瘢痕疙瘩的关系,本文应用免疫组化方法,检测了 TGF- $\beta$ 1 在组织中的表达,并分析两者的关系,为探索瘢痕疙瘩的发病原因提供一定的理论依据(解释本研究的必要性)。

4. 论文正文书写的注意事项 正文是科技论文的核心部分,占全文的主要篇幅。如果说引言是提出问题,正文则是分析问题和解决问题。这部分是作者研究成果的学术性和创新性的集中体现,决定了论文写作的成败和学术水平的高低。正文的写作和编排没有统一的规定,一般



包括材料与方法、结果、讨论等几个部分(知识点)。临床资料收集、实验观察、数据处理与分析、研究结果的呈现是正文的主要部分,应该给予重点地详细论述。尊重事实,在资料的取舍上不应掺入主观成分,或妄加猜测,也不应忽视偶发的现象和可能性。适当应用图或表来加以说明。图或表的信息要表达清楚。引用的资料,尤其是引用他人的成果应注明出处。本例文的研究方法可以叙述为:样本组织分别来源于 19 例瘢痕疙瘩患者和 19 例正常个体皮肤组织,组织取材后进行免疫组织化学染色以检测 TGF- $\beta$ 1 表达,并进一步分析 TGF- $\beta$ 1 与瘢痕疙瘩的关系;这样就阐明了研究是怎样进行的。本例文的研究结果可以描述为:瘢痕疙瘩组织中 TGF- $\beta$ 1 表达明显高于正常皮肤组织,根据需要可以具体描述每组表达阳性率的大小;结果的描述主要是告诉读者研究发现了什么。研究讨论的作用则是解释和评价结果,阐明研究的系统性,起到画龙点睛和深化主题的作用。

5. 参考文献的注意事项 在正文后面列出的参考文献必须是作者亲自阅读的一次文献,按照在正文中出现的先后顺序依次列出(知识点)。著录格式要符合所投期刊的具体要求。参考文献主要包括作者、论文题目、期刊名称、发表时间及起始页码等信息。

## 第二节 如何选择适合的期刊进行投稿

医学论文写好之后,就要尽快投稿,必须追求时效性,争取早日发表。目前国内医疗系统晋级一般要求有国家级或省级正式期刊发表的论文,对于工作在基层的广大住院医师来说是个不小的挑战,如果投稿时期刊选择不当,则可能导致发表周期过长,甚至被拒稿的可能。笔者发现很多住院医师的论文完成后不知道投向哪一本期刊最合适;其实,只要掌握了投稿的技巧,就不会在投稿时遇到太多困难,甚至可能会取得事半功倍的效果。

投稿前要对相关学术期刊的分布特征和期刊的学术水平有一个系统的了解。例如,往中华系列期刊投稿比较困难,因为中华系列期刊的稿源充分,办刊水平高,自然对所投稿件的要求也高;其他核心期刊相对容易一些,发表的概率也大一些。

通过比较不同期刊的办刊宗旨、学术水平、关注领域和发表周期之后,初步锁定欲投期刊,然后选看一下近几期欲投期刊发表过的论文,从中发现该期刊发表论文的共性,如每期重点或集中发表哪些主题的文章等,一旦确定你的论文主题和学术水平与之切合,则尽快投稿(知识点)。例如,题目为“TGF- $\beta$ 1 在瘢痕疙瘩组织中的表达及其意义”的论文所涉及的内容属于皮肤瘢痕领域,而《中华整形外科杂志》恰恰有专栏主要发表瘢痕相关研究的论文,所以该例文投给《中华整形外科杂志》有被刊登发表的可能性存在。

特别强调,投稿时切记要遵守必要的原则:①不要一稿多投,一稿多投现象损害作者自己的学术信誉;②时间要尽快,有些研究当前是热点,以后可能不是,有一定的时限性,要争取时间,不要等其他人的相关论文发表刊载,你才开始审稿,这样肯定会失败的。还有就是进入修改阶段的时候要尽快完成修改,虽然一般都有多余的修改时间,但是越早越好;③关于作者,自己是第一作者,其他有贡献或提出重要建议的人最好也写上,体现团队合作,同时也尊重他人的知识产权。

## 第三节 论文的修稿

论文修稿的过程,就是如何科学地回答审稿问题的过程。审稿意见是编辑人员和审稿专家对论文的综合评定,通常在审稿意见中包含若干个问题,这些问题往往需要论文作者高度关注,准确回答;对审稿问题的回答水平直接关系到论文是否被接受发表,必须重视。例如,某作者的论文《2009 年浙江地区手足口病的病原学鉴定》投稿后,收到如下审稿意见:“根据论文题目确



定,作者研究目的是探讨2009年浙江地区手足口病的病原学特征及相关病原。可论文的结论却为:2009年浙江地区手足口病的病原性特征与香港地区相似。该论文的题目说明本文是研究浙江地区手足口病的病原学特征,并不是比较浙江地区与香港地区手足口病的区别。所以该论文的目的与结论不统一”。作者根据审稿意见将论文的结论修改为:“EV71毒株是2009年浙江地区手足口病的主要病原体之一”。经过二审后,该论文顺利发表。再例如,某作者最近进行了一项东北地区大气中PM<sub>2.5</sub>与哮喘发病的关联研究。结果发现,东北地区大气中PM<sub>2.5</sub>的浓度较低,且与哮喘的发病无关。这与大多数空气污染的研究结果相反。该论文投出后,作者收到这样的审稿意见,“如何解释与其他研究相悖的结果?”作者分析了相关论文后做出如下回答,“PM<sub>2.5</sub>是空气中广泛存在的微小粒子,只有其浓度达到一定数值后才能致病,而本文的取样均在我国东北地区,PM<sub>2.5</sub>的阈值仍处于临界值以下,尚不至于造成哮喘发病率增加的结论客观可信”。该回答获得编辑人员和审稿专家的认可,论文得以顺利发表。

1. 学术观点的修改(知识点) 写论文的主旨是表达自己的观点,如果认知不够深刻,甚至有错误。修改论文时首先要推敲主题和论据是否正确,结果是否客观,文章是否有创新性;对于论文中出现的主观、片面、空泛的地方,必须改写。

2. 论文材料的修改(知识点) 材料是论文的“血肉”,是证明观点的依据。要保证引用的材料确凿有力,并且相互印证强化论点,要把陈旧的文献加以剔除或调换。

3. 文字表达的修改(知识点) 语言是表达的工具,要使论文写得准确、简洁,就要在语言运用上推敲修改。使论文表达清楚简练,把含混、笼统的句子改为清晰、具体、客观的阐述。

总之,世界上不存在毫无缺陷的科学设计,所以写出来的论文也不可能一气呵成。优秀的论文要经过千锤百炼般地修改,就如一个雕塑家细细打磨雕像一样,是一锤一凿修出来的,经过耐心科学地修改,相信自己的论文一定能够如你所愿,顺利发表。

(肖志波)

## 参考文献

1. Che R, Zhang A. Mechanisms of glucocorticoid resistance in idiopathic nephrotic syndrome. *Kidney Blood Press Res*, 2013, 37 (4-5): 360-378
2. Higgins JP, Thompson SG, Deeks JJ, et al. Measuring inconsistency in meta-analyses. *BMJ (Clinical research ed.)*, 2003, 327 (7414): 557-560
3. Oxman AD, GRADE Working Group. Grading quality of evidence and strength of recommendations. *BMJ*, 2004, 328 (19): 1490-1494
4. Haynes RB. Of studies, syntheses, synopses, summaries, and systems: the "5S" evolution of information services for evidence-based healthcare decisions. *Evidence Based Medicine*, 2006, 11 (6): 162-164
5. Elamin MB, Flynn DN, Bassler D, et al. Choice of data extraction tools for systematic reviews depends on resources and review complexity. *Journal of clinical epidemiology*, 2009, 62 (5): 506-510
6. 方积乾, 陆盈. 现代医学统计学. 北京: 人民卫生出版社, 2002
7. 刘鸣. 系统评价、Meta- 分析设计与实施方法. 北京: 人民卫生出版社, 2011
8. 吴文源. 社会精神医学. 北京: 人民卫生出版社, 2011
9. Jiang L, Su L, Cui H, et al. Ginkgo biloba extract for dementia: a systematic review. *Shanghai Archives of Psychiatry*, 2013, 25 (1): 10-21
10. Maltais F, Ostinelli J, Bourbeau J, et al. Comparison of nebulized budesonide and oral prednisolone with placebo in the treatment of acute exacerbations of chronic obstructive pulmonary disease. *Am J Respir Crit Care Med*, 2002, 165 (5): 698-703
11. Gunen H, Hacievliyagil SS, Yetkin O, et al. The role of nebulised budesonide in the treatment of exacerbations of COPD. *Eur Respir J*, 2007, 29 (4): 660 - 667
12. 王巧灵, 姚何, 宋文丰等. 三种人体测量指标对高血压和高血糖及高血脂的预测效果分析. *中国全科医学*, 2012, 5 (6B): 1917-1920
13. Zhang XJ, Wang GY, Shen Q, et al. Current status of contraceptive use among rural married women in Anhui Province of China. *BJOG*, 2009, 116 (12): 1640-1645
14. Yang W, Lu J, Weng J, et al. Prevalence of diabetes among men and women in China. *N Engl J Med*, 2010, 362 (12): 1090-1101
15. Zhang H, Bo SH, Zhang XT, et al. Sampling survey of disability in 0-6 year-old children in China. *Biomed Environ Sci*, 2006, 19 (5): 380-384
16. Parsonnet J, Friedman GD, Vandersteen DP, et al. Helicobacter pylori infection and the risk of gastric carcinoma. *N Engl J Med*, 1991, 325 (16): 1127-1131
17. 陈世耀, 王吉耀, 刘天舒. 幽门螺杆菌感染与胃癌, 一项病例对照研究. *胃肠病学和肝病学杂志*, 1998, 3 : 227-229
18. Selby JV, Friedman GD, Quesenberry CP Jr, et al. A case-control study of screening sigmoidoscopy and mortality from colorectal cancer. *N Engl J Med*, 1992, 326 (10): 653-657
19. Douglas JR, Larsson H, Friis S, et al. Proton pump inhibitor use and risk of colorectal cancer: a population-based, case - control study. *Gastroenterology*, 2007, 133 : 755 - 760
20. De Lédinghen V, Heresbach D, Fourdan O, et al. Anti-inflammatory drugs and variceal bleeding: a case-control study. *Gut*, 1999, 44 (2): 270 - 273
21. Van der Veen AH, Dees J, Blankensteijn JD, et al. Adenocarcinoma in Barrett's oesophagus: an overrated risk. *Gut*, 1989, 30 (1): 14-18
22. Hvid-Jensen F, Pedersen L, Drewes AM. Incidence of adenocarcinoma among patients with Barrett's esophagus. *NEJM*, 2011, 365 (15): 1375-1383
23. Welin L, Svärdsudd K, Wilhelmsen L, et al. Analysis of risk factors for stroke in a cohort of men born in 1913. *N Engl J Med*, 1987, 317 (9): 521-526





24. Silverstein MD, Reed CE, O'Connell EJ, et al. Long-term Survival of a Cohort of Community Residents With Asthma. *N Engl J Med*, 1994, 331 (23): 1537-1541
25. El-Serag HB, Johnson ML, Hachem C, et al. Statins Are Associated with a reduced risk of hepatocellular carcinoma in a large cohort of patients with diabetes. *Gastroenterology*, 2009, 136 (5): 1601-1608
26. 李立明. 临床流行病学. 北京: 人民卫生出版社, 2011
27. Wang Y, Wang Y, Zhao X, et al. Clopidogrel with aspirin in acute minor stroke or transient ischemic attack. *N Engl J Med*, 2013 (1), 369 : 11-19
28. Johnston SC, Easton JD, Farrant M, et al. Platelet-oriented inhibition in new TIA and minor ischemic stroke (POINT) trial: rationale and design. *Int J Stroke*, 2013, 8 (6): 479-483
29. Biecker E. Portal hypertension and gastrointestinal bleeding: diagnosis, prevention and management. *World J Gastroenterol*, 2013, 19 (31): 5035-5050
30. O'Brien BJ, Heyland D, Richardson WS, et al. Users' guides to the medical literature. XIII. How to use an article on economic analysis of clinical practice. B. What are the results and will they help me in caring for my patients. *JAMA*, 1997, 277 (22): 1802-1806
31. Liaw YF, Leung N, Kao JH, et al. Asian-Pacific consensus statement on the management of chronic hepatitis B: a 2008 update. *Hepatol Int*, 2008, 2 (3): 263-283
32. Zhuang H. Updates of EASL Clinical practice guidelines: management of chronic hepatitis B virus infection. *Chin J Hepatol*, 2009, 50 (2): 227-242
33. Golub RM, Fontanarosa FB. 2012 Theme issue on comparative effectiveness research-call for papers. *JAMA*, 2011, 305 (7): 717
34. Wu AW, Snyder C, Clancy CM, et al. Adding the patient perspective to comparative effectiveness research. *Health Aff (Millwood)*. 2010, 29 (10): 1863-1871
35. Ratner R, Eden J, Wolman D, et al. Initial national priorities for comparative effectiveness research. Washington, D.C.: National Academies Press, 2009
36. Cox E, Martin B, van Staa T, et al. Good research practices for comparative effectiveness research: approaches to mitigate bias and confounding in the design of non-randomized studies of treatment effects using secondary data sources. *Value Health*, 2009, 12 (8): 1053-1061
37. Johnson ML, Crown W, Martin BC, et al. Good research practices for comparative effectiveness research: analytic methods to improve causal inference from nonrandomized studies of treatment effects using secondary data sources; the ISPOR good research practices for retrospective database analysis task force report. *Value Health*, 2009, 12 (8): 1062-1073
38. Stukel TA, Fisher ES, Wennberg DE, et al. Analysis of observational studies in the presence of treatment selection bias effects of invasive cardiac management on AMI survival using propensity score and instrumental variable. *JAMA*, 2007, 297 (3): 278-285
39. Fang G, Brooks JM, Chrischilles EA. Apples and oranges? Interpretations of risk adjustment and instrumental variable estimates of intended treatment effects using observational data. *Am J Epidemiol*, 2012, 175 (1): 60-65
40. Kanbay M, Huddam B, Azak A, et al. A randomized study of allopurinol on endothelial function and estimated glomerular filtration rate in asymptomatic hyperuricemic subjects with normal renal function. *Clin J Am Soc Nephrol*, 2011, 6 (8): 1887-1894
41. Gur M, Yilmaz R, Demirbag R, et al. Relation of serum uric acid levels with the presence and severity of angiographic coronary artery disease. *Angiology*, 2008, 59 (2): 166-171
42. Culleton BF, Larson MG, Kannel WB, et al. Serum uric acid and risk for cardiovascular disease and death: the Framingham Heart Study. *Ann Intern Med*, 1999, 131 (1): 7-13
43. Waring WS, McKnight JA, Webb DJ, et al. Uric acid restores endothelial function in patients with type 1 diabetes and regular smokers. *Diabetes*, 2006, 55 (11): 3127-3132
44. Kostka-Jeziorny K, Uruski P, Tykarski A. Effect of allopurinol on blood pressure and aortic compliance in hypertensive patients. *Blood Press*, 2011, 20 (2): 104-110
45. Pinelli NR, Jantz AS, Martin ET. Sensitivity and specificity of glycated hemoglobin as a diagnostic test for diabetes and prediabetes in Arabs. *J Clin Endocrinol Metab*, 2011, 96 (10): E1680-1683
46. 约翰 I. 加林, 弗雷德里克 P. 奥格尼本. 临床研究原理与实践. 张玉峰, 译. 北京: 科学出版社, 2008
47. 王家良. 临床流行病学: 临床科研设计, 测量与评价. 第 3 版. 上海: 上海科学技术出版社, 2009



48. Yao JC, Shah MH, Ito T, et al. Everolimus for advanced pancreatic neuroendocrine tumors. *N Engl J Med*, 2011, 364 (6): 514-523
49. Elmunzer BJ, Scheiman JM, Lehman GA, et al. A randomized trial of rectal indomethacin to prevent post-ERCP pancreatitis. *N Engl J Med*, 2012, 366 (15): 1414-1422
50. Azim HA Jr, Kroman N, Paesmans M, et al. Prognostic impact of pregnancy after breast cancer according to estrogen receptor status: a multicenter retrospective study. *J Clin Oncol*, 2013, 31 (1): 73-79
51. Azim HA Jr, Santoro L, Pavlidis N, et al. Safety of pregnancy following breast cancer diagnosis: a meta-analysis of 14 studies. *Eur J Cancer*, 2011, 47 (1): 74-83
52. Amant F, von Minckwitz G, Han SN, et al. Prognosis of women with primary breast cancer diagnosed during pregnancy: results from an international collaborative study. *J Clin Oncol*, 2013, 31 (20): 2532-2539
53. 颜虹, 徐勇勇, 赵耐青. 医学统计学. 第2版. 北京: 人民卫生出版社, 2013
54. 孙振球, 徐勇勇. 医学统计学. 第3版. 北京: 人民卫生出版社, 2010
55. 赵仲堂. 流行病学研究方法与应用. 第2版. 北京: 科学出版社, 2005
56. 王家良. 临床流行病学 - 临床科研设计、测量与评价. 第3版. 上海: 上海科学技术出版社, 2013
57. 张文彤, 钟云飞. IBM SPSS 数据分析与挖掘实战案例精粹. 北京: 清华大学出版社, 2013
58. 李晓松. 医学统计学. 第2版. 北京: 高等教育出版社, 2008

## 附录

---

临床研究中,很多论文的书写依据研究设计有一定的书写格式,必须包括某些重要的内容,同时也是评价研究论文科学性的依据。读者可以从网上查询到更新版本。

临床随机对照试验(RCT)的报告标准(CONSORT 声明,2010 修订版)

[www.consort-statement.org](http://www.consort-statement.org)

诊断试验正确性的报告标准(STARD 2008 修订)

[www.stard-statement.org](http://www.stard-statement.org)

观察性研究的报告标准(STROBE)

[www.strobe-statement.org](http://www.strobe-statement.org)

系统综述或 Meta 分析报告标准(PRISMA)2009

[www.prisma-statement.org](http://www.prisma-statement.org)

# 中英文名词对照索引

Meta 分析 (Meta-analysis) ..... 18, 22  
Meta 分析报告的质量 (quality of reporting of meta-analysis, QUOROM) ..... 18

## A

安慰剂 (placebo) ..... 84

## B

报告质量 (reporting quality) ..... 18  
比数比 (odds ratio, OR) ..... 21, 61  
标准差 (standard deviation) ..... 117  
标准化均数差值 (standardized mean differences, SMD) ..... 21  
标准诊断 (standard diagnostic test) ..... 118  
病例 - 对照研究 (case-control study) ..... 55, 114, 144  
病例报告 (case report) ..... 35  
病因 (cause) ..... 108

## C

测量偏倚 (measurement bias) ..... 51  
巢式病例对照研究 (nested case-control study) ..... 62  
成本 - 效果分析 (cost-effectiveness analysis, CEA) ..... 94  
成本 - 效益分析 (cost-benefit analysis, CBA) ..... 94  
成本效果比 (cost/effectiveness, C/E) ..... 96  
抽样 (sampling) ..... 145  
抽样调查 (sampling survey) ..... 46  
出生队列 (birth cohort) ..... 67

## D

单纯随机抽样 (simple random sampling) ..... 48  
第二类错误 (type II error) ..... 145  
第一类错误 (type I error) ..... 145  
调查偏倚 (interviewer bias) ..... 51  
定量的系统评价 (quantitative systematic review) ..... 22  
动态人群 (dynamic population) ..... 67  
队列研究 (cohort study) ..... 66, 114, 144  
对照 (control) ..... 83  
多阶段抽样 (multistage sampling) ..... 48  
多中心临床注册研究 (multicenter clinical registries) ..... 7

## E

二分类变量 (dichotomous variable) ..... 21

## F

发表偏倚 (publication bias) ..... 26  
发病率比 (incidence rate ratio, IRR) ..... 114  
发病密度 (incidence density) ..... 72  
方法学质量 (methodological quality) ..... 17, 22  
非随机同期对照试验 (non-randomized concurrent control trial) ..... 125  
分层抽样 (stratified sampling) ..... 48  
风险比 (hazard ratio, HR) ..... 88, 112

## G

干扰 (intervention) ..... 90  
固定队列 (fixed cohort) ..... 67  
固定效应模型 (fixed effect model) ..... 22  
观察性研究 (observational study) ..... 67  
归因危险度 (attributable risk, AR) ..... 73

## H

横断面研究 (cross-sectional study) ..... 44, 113, 144  
患病率研究 (prevalence study) ..... 44  
灰色文献 (grey literature) ..... 20  
回忆偏倚 (recall bias) ..... 50  
混杂因素 (confounder) ..... 110  
混杂因子 (confounder 或 confounding factor) ..... 77  
混杂作用 (confounding effect) ..... 77  
霍桑效应 (Hawthorne effect) ..... 90

## J

加权均数差 (weighted mean difference, WMD) ..... 21  
间接病因 (indirect cause) ..... 110  
检索策略 (search strategy) ..... 19  
检验效力 (power) ..... 117  
健康工人效应 (health worker effect) ..... 76  
结局变量 (outcome variable) ..... 71  
金标准 (gold standard) ..... 118  
精确度 (accuracy) ..... 22



绝对危险度减少 (absolute risk reduction, ARR) ..... 89

## K

可信区间 (confidence interval, CI) ..... 25

可重复性 (repeatability) ..... 20

## L

历史对照研究 (historical control study) ..... 125

历史性队列研究 (retrospective cohort study) ..... 75

连续性变量 (continuous variable) ..... 21

临床试验 (clinical trial) ..... 125, 144

漏斗图 (funnel plot) ..... 26

率比 (rate ratio) ..... 73

率差 (rate difference, RD) ..... 21, 73

## M

盲法 (blinding) ..... 86

敏感性分析 (sensitivity analysis) ..... 23, 96

## N

纳入标准 (inclusion criteria) ..... 19

## P

排除标准 (exclusion criteria) ..... 19

匹配 (matching) ..... 145

偏倚 (bias) ..... 63

普查 (census) ..... 45

## Q

倾向评分法 (propensity score, PS) ..... 104

权重 (weight) ..... 25

## R

人群归因危险度 (population attributable risk, PAR) ..... 74

## S

森林图 (forest plot) ..... 25

受试者工作特征曲线 (receiver operator characteristic curve, ROC) ..... 119

数据提取表 (data collection form) ..... 21

双向性队列研究 (ambispective cohort study) ..... 75

随机对照试验 (randomized controlled trial, RCT) ..... 82, 115, 125

随机分配隐藏 (concealment) ..... 86

随机化 (randomization) ..... 85

随机化方法 (methods of randomization) ..... 145

随机效应模型 (random effect model) ..... 22

## T

同质性 (homogeneity) ..... 18, 20

统计学意义 (statistical significance) ..... 117

## W

外部真实性 (external validity) ..... 22

危险度比 (risk ratio) ..... 73

危险因素 (risk factor) ..... 108

污染 (contamination) ..... 90

无应答偏倚 (non-response bias) ..... 50

## X

系列病例分析 (case-series analysis) ..... 38

系统抽样 (systematic sampling) ..... 48

系统综述 (systematic review) ..... 16, 17

系统综述和 Meta 分析优先报告的条目声明 (preferred reporting items for systematic reviews and meta-analyses) ..... 18

现存患者 (prevalent case) ..... 113

相对危险度 (relative risk, RR) ..... 21, 59, 73, 112

相对危险度减少 (relative risk reduction, RRR) ..... 89

相对危险度降低值 (relative risk reduction, RRR) ..... 21

校正 (adjustment) ..... 110

效差 (effect size) ..... 117

效应值 (effect size) ..... 25

效应指标 (effect size) ..... 21

新发患者 (incident case) ..... 114

幸存者偏倚 (survivor bias) ..... 50

需治疗人数 (number needed to treat, NNT) ..... 89

选择偏倚 (selection bias) ..... 50

## Y

研究计划书 (protocol) ..... 19

异质性 (heterogeneity) ..... 18

异质性检验 (heterogeneity test) ..... 22

意向性治疗分析 (intention-to-treat analysis, ITT) ..... 88

优势比 (odds ratio, OR) ..... 112

有向无环图 (directed acyclic graph, DAG) ..... 110

预试验 (pilot study) ..... 141

约登指数 (Youden index, YI) ..... 119

## Z

增量分析 (incremental analysis) ..... 96

诊断性试验 (diagnostic test) ..... 118

整群抽样 (cluster sampling) ..... 49

直接病因 (direct cause) ..... 110

质量控制 (quality control) ..... 145



|                                 |     |                                                |     |
|---------------------------------|-----|------------------------------------------------|-----|
| 质量评价 (quality assessment) ..... | 22  | 自身前后对照研究 (before-after study in the            |     |
| 中介因素 (mediator) .....           | 110 | same patient) .....                            | 125 |
| 主要结局指标 (primary outcome) .....  | 21  | 最佳证据 (best evidence) .....                     | 16  |
| 资料收集 (data collection) .....    | 17  | 最小成本分析 (cost minimization analysis, CMA) ..... | 94  |
| 资料提取 (data extraction) .....    | 20  |                                                |     |