

《柳叶刀》

临床研究基本概念

THE LANCET
Handbook of Essential Concepts in
Clinical Research

主编 Kenneth F. Schulz David A. Grimes

主译 王吉耀



人民卫生出版社

《柳叶刀》 临床研究基本概念

THE LANCET

Handbook of Essential Concepts in
Clinical Research

人民卫生出版社

图书在版编目 (CIP) 数据

《柳叶刀》临床研究基本概念/王吉耀主译. —北京:
人民卫生出版社, 2010. 6

ISBN 978-7-117-12873-5

I. ①柳… II. ①王… III. ①临床医学 IV. ①R4

中国版本图书馆 CIP 数据核字 (2010) 第 064695 号

门户网: www.pmph.com 出版物查询、网上书店

卫人网: www.ipmph.com 护士、医师、药师、中医师、卫生资格考试培训

版权所有, 侵权必究!

图字: 01-2010-2621

《柳叶刀》临床研究基本概念

主 译: 王吉耀

出版发行: 人民卫生出版社 (中继线 010-59780011)

地 址: 北京市朝阳区潘家园南里 19 号

邮 编: 100021

E - mail: pmph@pmph.com

购书热线: 010-67605754 010-65264830

010-59787586 010-59787592

印 刷: 中国农业出版社印刷厂

经 销: 新华书店

开 本: 787×1092 1/32 印张: 8 字数: 200 千字

版 次: 2010 年 6 月第 1 版 2010 年 6 月第 1 版第 1 次印刷

标准书号: ISBN 978-7-117-12873-5/R · 12874

定 价: 39.00 元

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

(凡属印装质量问题请与本社销售中心联系退换)

中文版序

During the past few years, The Lancet has been glad to develop very warm relations with Chinese colleagues in medicine and medical science. We have seen a huge increase in research papers submitted from China. And we now publish many more Chinese authors. Through strong collaborations with leading Chinese universities, we have published series and theme issues on current predicaments in China's health system. We want to encourage these welcome trends, this short book might help.

Richard Horton

February 9, 2010

在过去几年中,《柳叶刀》很高兴能够逐渐加强与中国同行在医学和医学科学之间更加亲密的联系。我们已经看到极多数量从中国递交的研究论文。并且我们现在已经出版了许多中国作者的文章。通过与名列前茅的中国大学的坚定的合作,我们还出版了系列和专题专刊来反映当前中国医疗系统的状态。我们想要鼓励这些受欢迎的趋向,希望这本书会帮助大家。

理查德·霍顿

《柳叶刀》主编

2010年2月9日

主译前言

本书是一本简洁、实用的参考书,有助于临床医师及临床研究者批判性阅读文献和设计临床研究。我作为《临床医学研究发展战略研究报告》课题专家组成员,在起草该报告过程中,对我国临床研究的现状有了更多的了解。正如手册前言中提到的许多临床研究者 in 研究方法学方面很少经过正规的培训,我国的情况也如此。本人从事临床流行病学和循证医学教学 20 多年,深知在我国开展临床研究方法培训的重要性;更感到一本好的培训教材对培训起到事半功倍的效果。当我从爱思唯尔出版社看到本手册后,就爱不释手。这本手册文字简洁、重点突出。每章起始部分陈述该章要点,在章末有结论,可读性强。同时具有权威性(引用的文章均经国际同行评议)、先进性(首次提出一些较先进的概念)、实用性(每章均有例子说明)。因此有将它翻译成中文、让我国更多的临床医师学习的想法。

本书译者为复旦大学附属中山医院临床医师,均在国内外获得了临床流行病学硕士学位,并有多年教学经验。其中一批年轻的硕士,在荷兰 Erasmus 大学刚接受了临床流行病学专业培训归来。在翻译过程中他们反复推敲、修改,保证了本书的翻译质量;袁源智医师为中译本的编排付出辛勤劳动;从事翻译专业的儿媳

陆茸女士在德国客串了序的翻译。在此一并表示感谢。特别要感谢我国统计学大师金丕焕教授对本书在统计学相关内容翻译时的耐心指导和修改。最后更要感谢《柳叶刀》主编理查德·霍顿百忙中特地为我们的中译本写序,以及《柳叶刀》编审郭麟博士将它译成中文。

本手册翻译如有疏漏不妥之处,尚祈读者指正。

主译 王吉耀

2010年5月

原版序

绝大多数医师都会赞同“临床医师对临床研究评估的技巧掌握得并不好”这个观点。如有个别医师持有不同观点,我们有大量的证据可以反驳他们这种乐观的想法^[1]。随机试验对现今临床实践的巨大影响使我们熟知对现代医学的基础专业设计需要进行评估的理念。我们不难得出一个结论,医生不能掌握评估临床研究的技能就不能做好临床诊治。我们应该正确阐释试验,这不仅能确保重要的研究成果快速应用于临床,而且能促进对模棱两可的试验结果进行进一步探索。“随机试验”已经过多地作为“高质量和权威性研究”的代名词。但正如所有科研设计一样,随机试验的设计,有好的,也有差的。临床医师应该能明辨其中的差别。

1990年我加盟《柳叶刀》编纂之后,当时很快就确信了临床流行病学是临床医学的基础学科。而今天我没有完全摒弃这个看法,过去的十年里,我们已经看得很清楚,如果不加批评地接受流行病学研究中那些模棱两可的试验结果会严重地误导我们步入歧途。领会临床试验是将新医疗技术引入临床治疗的关键,同时,看清临床流行病学中的误区对正确掌握风险意识和追索病因来说同等重要。

本书中,KEN SCHULZ 和 DAVID GRIMES 提供了一份出色的关键指南,指引忙碌工作的医师们如何阐释科研。所有谙熟此原理的临床医师都可以从此更安全有效地医治患者。

理查德·霍顿

《柳叶刀》主编

参考文献

Perneger TV, Ricou B, Boulvain M et al. Medical researchers evaluate their methodological skills. J Clin Epidemiol 2004; 57: 1323-1329.

原版前言

本书是为忙碌的临床医师和活跃的研究者而设计的。临床医师是经过训练的病因寻找者。他们根据对每位病人身上观察到的症状、体征,积累起能找到其发病原因的临床线索(诊断)。临床医师在诊断过程中会应用那些来自于已在医学杂志上发表的文章中重要信息,这些文章与他们所处理的病人有相同的病因探索过程。对这些文献中因果关系进行批判性地阅读将会对解决诊断难题带来满意的临床结果。

我们希望本书能在研究方法上为临床医师提供评估文章的有用工具,并从临床科学中得到满足,感到有趣。更重要的是从更具有批判性和考虑周到的研究工作中涌现出更好的实践者。

本书的目标是为临床医师提供能批评性阅读医学文献的工具,并易被其理解吸收。当然,我们希望这些文章还有其他更崇高的目标。医学文献在研究方法包括研究的执行过程和报告方面均存在许多缺陷。不足为奇的是许多临床研究者在研究的方法学方面很少经过正规的培训。事实上,缺少科学性方面的培训成为医学教育上的一个主要问题。因此,对于临床研究和研究者在方法学基础上的支撑可提高研究成果的质量、甚至于数量,研究者和临床医师在这方面的需求是一致的。

当然,这些简短的章节并不能覆盖临床研究的广度,即包括许多研究设计和各门学科。我们的临床研究角度是将观察性和实验性研究的设计结合起来,正像 Austin Bradford Hill 在 20 世纪 40 年代末和 50 年代初应用回顾性的病例对照、长期的前瞻性队列和随机对照的设计一样,我们感到研究者将根据他们自己的计划从较广范围的工具中获益。与流行病学最初的观察性研究设计的较为狭隘的观点相反,我们亦会进一步鼓励与其他能使研究更为丰富的学科进行交叉取经。流行病学家通过寻找并与生物统计、医师、护士、微生物学家和行为科学家合作而变得强大。本书关注临床研究的要点,并不是所有的内容均在导言水平,对于某些主题如队列研究(第 4 章),将其主要精华作为基本重点,这也可在其他书中找到。相反另一些题目像如何产生随机序列(第 10 章和第 12 章)、分组隐藏(第 11 章)主要涉及一些较先进的概念,许多内容在其他书上不一定能找到。进而,我们对其他一些主题陈述了一些基本观点,也提到目前尚有的争论。例如低把握度的研究是否符合伦理(第 9 章)。所有情况下,我们都全面简明地传达在过去 60 年中与我们研究经验密切相关的概念。

本书并非是生物统计、医学统计或者流行病学方面的介绍性读物。关于这些内容已经有很多非常好的书了,我们无意重复这些内容。本书着眼于偏倚的消除,而非统计检验方法的选择。

每一章都用一篇从 2002 年到 2005 年在《柳叶刀》上由同行评议而发表的文章进行表述,在这意义上,与大多数的书不一样,本书是经过同行评议的。

本书内容包括描述性研究、队列研究、病例对照研究、在病例对照研究中的对照如何设定、偏倚如何避免、筛查试验和在筛查过程中的似然比,其中有 8 章是关于随机对照试验的。不成比例的安排是因为在临床研究中随机试验是金标准,随机试验有利于消除偏倚,研究确立了试验中避免偏倚的方法学因素。最后,由于临床试验是如此重要,以致该试验结果对临床医师起到的作用远远大于那些观察性研究所起的作用。因此研究者需要保证这些试验不但做了,而且做得很好,希望研究者可以从我们在随机试验中陈述的细节中获益。

我们感谢《柳叶刀》主编 Richard Horton 的建议,让我们的文章拿出来发表,他鼓励我们拿起笔将关键内容书写成章。我们也感谢 Berlex 基金会支持基金会成员组织了我们在过去 2 年中所开的课程,这本手册中的许多章节都是为这个课程所写。

本书所有章节都经过 Ward Cates 和 David Sackett 审阅,部分章节由 Doug Altman、Iain Chalmers 爵士、Vance Berger 和 Rosalie Dominik 审阅,使本书得以改进。当然任何的错误和疏漏均由我们负责。我们非常感谢“爱思唯尔”的 Heidi Harrison 在出版经验上进行指导,我们还衷心感谢《柳叶刀》的 Abigail Pound 和 Sarah Passey,不但感谢他们华丽的编辑和出版帮助,而且感谢他们的耐心和热情。

Kenneth F. Schulz & David A. Grimes

绪 论

1. 临床研究概述:着陆点

章节内容

临床研究的分类 2

结局的测量 10

研究可以做什么,什么不能做 5

结论 12

许多临床医生称他们不能准确地阅读医学文献。为解决这个难题,我们为临床医生和研究者提供了一本初级读本。根据研究者是否分配暴露因素将临床研究分为两大类:实验性和观察性。实验性研究也可再分为两类:随机和非随机研究。观察性研究可以是分析性的,也可以是描述性的。分析性研究的特点是有一个比较(对照)组,而描述性研究则没有。在分析性研究中,队列研究追踪人群的时间方向是从暴露到结局。病例-对照研究与之相反是从结局回溯到暴露。横断面研究像一次快照,同时检测暴露和结局。描述性研究,如个案报告则没有对照组。在这类研究中,研究者不能检测两者的因果关系,这是一个常常被遗忘或忽略的事实。对因果关系的衡量,如相对危险度或比值比是常用的二分结局(如患病和健康)表达结果的方法。这些参数的可信区间提示了结果的精确性。可信区间的因果关联测定显示了作用的强度、方向和可能的范围以及机遇发生的概率。与此相反, P 仅表示机遇的大小。以 P 为0.05检验无效假设是没有医学基础的,不应该鼓励。

今天的临床医生处境尴尬。繁忙的临床工作使他们没有机会与文献的发表同步前进,并批判性阅读。几项研究表明,他们掌握的当前的诊疗知识与从医学院毕业后的时间成反比^[1,2]。在很多司法体系中,参加特定数量的医学继续教育课程是保持执业执照的强制性要求。然而,这些课程并不能促进患者医疗水平^[3,4],因而要强调通过阅读自我学习的重要性。但是,许多执业医师称他们没有能力批判性地阅读文献^[5]。科学无知是医学教育的主要问题^[6]。

我们针对忙碌的医师和活跃的研究人员撰写了这一系列关于研究方法的短篇,这种需要在临床医师中是突出的,阅读这本书很可能产生更有批判性和思考全面的研究者和更好的执业医师。在整本书中,医师和研究者的需要是相同的,这种重叠的需要随机对照试验的讨论中最为明显。对希望准确评价随机研究的读者来说,他们应该理解从方法学研究中得到的实施试验的相关指南,而向临床医师展示的这些讨论,也同样有望帮助研究者做好随机试验。

我们用8章来写随机对照试验(randomised controlled trials)。这种不成比例的安排是故意的;随机对照试验是临床研究的金标准。《柳叶刀》发表了大量的随机对照试验。随机对照试验能帮助消除偏倚,研究已经明确了在试验中哪些是能减少偏倚的重要的方法学成分^[7,8]。最后,因为研究的重要性,临床医师更可能根据这些结果而不是观察性研究来采取行动;因此,研究者应该确保研究正确完成并准确报道。我们在这里对研究设计进行简要叙述,并讨论一些常用方法。

临床研究的分类

临床研究的分类(taxonomy of clinical research)和生物学分类

类似, 可以用一种简单的层次来归类大多数的研究(表 1-1)^[9]。但这样做, 一定要知道研究的设计。同生物学一样, 解剖决定生理。研究的解剖结构决定了什么能做, 什么不能做。读者面临的一个困难是作者有时并不报道研究类型或者提供足够的细节来明确研究的类型。与之相关的问题是作者有时把研究的类型写错了。例如称呼不是随机对照试验的研究为随机试验^[10], 或者把非

表 1-1 临床证据的分级 (rating of evidence): 美国预防服务工作组 (US Preventive Task Force) (US Preventive Task Force) 评估系统

证据质量	
I 级证据	来自至少一个设计良好的随机对照临床试验中获得的证据
II-1 级证据	来自设计良好的非随机对照试验中获得的证据
II-2 级证据	来自设计良好的队列研究或病例对照研究(最好是多中心研究)的证据
II-3 级证据	来自多个时间序列的、带有或不带有干预的研究得出的证据。重要的非对照试验(例如 19 世纪 40 年代青霉素的引入)的结果有时也可作为这一等级的证据
III 级证据	来自临床经验、描述性研究或专家委员会报告的权威意见
推荐强度	
A 级推荐	良好的科学证据支持该干预行为
B 级推荐	尚可的证据支持该干预行为
C 级推荐	没有足够的依据推荐或反对该干预行为, 但在其他场合可能会推荐
D 级推荐	尚可的科学证据反对该干预行为
E 级推荐	良好的证据反对该干预行为

同期队列研究称为病例-对照研究^[11-13]。“病例对照的”这个形容词也有时用来(不正确的)指有对照组的任何研究。

生物界有动物王国和植物王国。同样,临床研究也有两个大的领域:实验性研究(experimental research)和观察性研究(observational research)。图 1-1 显示,只要根据是否由研究者来分配暴露因素(如治疗)还是研究者对常规的临床实践进行观察等情况就能够迅速定位研究所处的领域^[14-18]。对实验性研究来说,需要分辨的是暴露因素(exposure)的分配是否真正采用随机化方法(对于相关人员来讲,下一例入组的分配是隐藏的)或者其他分

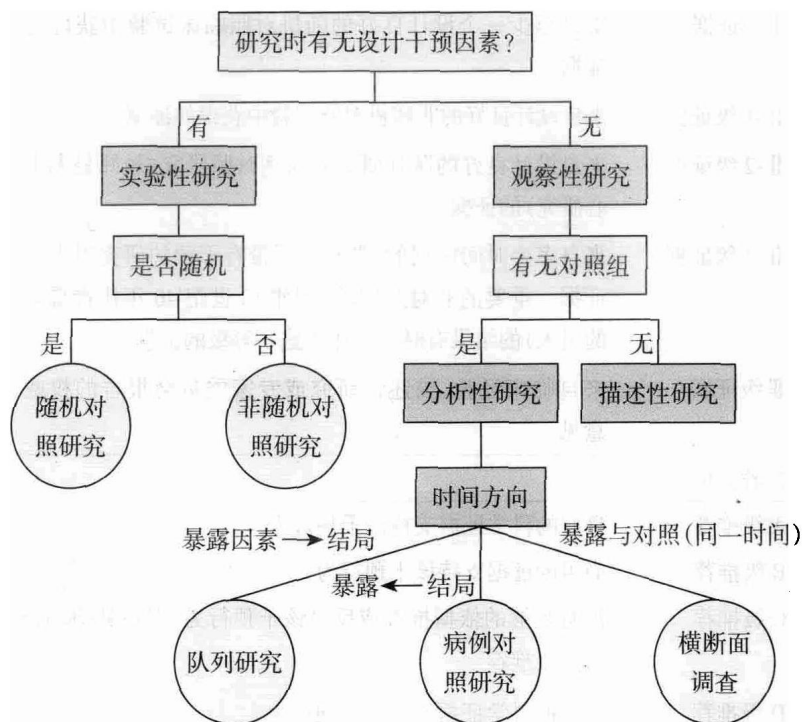


图 1-1 临床研究分类的法则

配计划,如交替分配(alternate assignment)^[19]。后者如一项试验中采用月份交替的方法来分配产妇的电子胎儿监测情况(一个月放开检查,另一个月限制检查)^[20]。

对文献中大量存在的^[21]观察性研究来说,下一步是确定研究有没有比较(或者说对照组)。如果有,此研究就称为分析性研究(analytical study);如果没有,就是描述性研究(descriptive study)(图 1-1)。如果研究是分析性的,需要明确研究的时间方向(temporal direction)。如果研究同时确定暴露和结局(outcomes),学术上称为横断面研究。例如,检测心肌梗死男性患者收治入院时的血清胆固醇和他们的隔壁邻居的血清胆固醇。这种类型的研究同时提供了人群的瞬时疾病和健康状况。

如果研究是从一项暴露开始(如口服避孕药),然后随访女性数年来评价其结局(如卵巢癌),那它肯定是队列研究。队列研究可以是同期的,也可以是非同期的(译者注:即前瞻性和回顾性队列研究)。与其相反的是,如果分析研究从结局(如卵巢癌)开始,回过去寻找暴露因素,如是否应用避孕药,那么研究就是病例-对照研究。

没有对照组的研究称为描述性研究。在其研究等级中,最下端是个案报告(case report)^[22]。当报告的患者超过一个时,就称为病例系列报告^[23]。

研究可以做什么,什么不能做

研究设计能够回答这个问题吗?

描述性研究位于研究等级的最下端,常常是进入一个医学新领域时首先进行的研究。调查者进行描述性研究来阐述发病率(incidence rate)、自然病程和可能的决定因素^[14,16,17]。这些结果显

示多少人在一定时间患该病或发生某种状态,描述疾病和患病者的特征,并产生关于疾病的假设。这些假设可以通过更严格的研究来评价,如分析性研究或随机对照试验。较早时关于军团病^[24]和中毒性休克综合征^[25]的报道就是描述性研究的一个例子。一个重要的告诫(经常被遗忘或故意忽略)是描述性研究没有对照组,不能用来评估因果关系。只有对照研究(分析性研究和实验性研究)才能够评估可能的因果关系。

横断面研究:特定时间的快照

横断面研究(cross-sectional study)有时又称为频率调查(frequency survey)或现况研究(prevalence study)^[26],用来检测特定时间疾病的存在与否和一项暴露因素存在与否。这样,焦点是患病,而不是发病。因为结局和暴露在同一时间被确定,这两个的时间关系可能不清楚。如假设横断面研究发现有关节炎的女性比没有关节炎的女性肥胖更常见。是增加的体重负担导致关节炎还是有关节炎的女性不得不减少活动然后出现肥胖呢?这种问题在横断面研究中是无法回答的。

队列研究:从暴露到结局

队列研究(cohort study)是一个有逻辑性的过程:从暴露到结局(图 1-2)。因此,这种类型的研究比病例-对照研究更容易理解。调查者确定一组暴露于感兴趣的因素的人群和一组或多组不暴露的人群,然后随访暴露组和非暴露组一段时间来确定结局。如果暴露人群比不暴露人群有较高的发生率,那么暴露因素与该结局的高危险相关。

队列研究有其优缺点,因为暴露因素在一开始就确定,可以假设暴露因素比结局早发生,就不必像在病例-对照研究中那样考虑回忆偏倚。队列研究可以计算真正的发病率、相对危险