

2011年度国家出版基金资助项目



中华医学统计百科全书

徐天和 / 总主编

医学研究与临床统计设计分册



徐勇勇 / 主 编

 中国统计出版社
China Statistics Press

2011年度国家出版基金资助项目



中华医学统计百科全书

◆ 徐天和 / 总主编 ◆

医学研究与临床统计设计分册



徐勇勇 / 主 编

 中国统计出版社
China Statistics Press

图书在版编目(CIP)数据

中华医学统计百科全书. 医学研究与临床统计设计分册 / 徐勇勇主编. —北京: 中国统计出版社, 2013. 5

ISBN 978-7-5037-6807-1

I. ①中… II. ①徐… III. ①医学统计—中国—百科全书②医学—研究—医学统计③临床医学—医学统计
IV. ①R195.1-61

中国版本图书馆 CIP 数据核字(2013)第 090208 号

医学研究与临床统计设计分册

作者/徐勇勇

责任编辑/陈悟朝

装帧设计/杨超 李雪燕

出版发行/中国统计出版社

通信地址/北京市西城区月坛南街 57 号 邮政编码/100826

办公地址/北京市丰台区西三环南路甲 6 号 邮政编码/100073

网 址/<http://csp.stats.gov.cn>

电 话/邮购(010)63376907 书店(010)68783172

印 刷/河北天普润印刷厂

经 销/新华书店

开 本/787×1092mm 1/16

字 数/485 千字

印 张/23.25

版 别/2013 年 5 月第 1 版

版 次/2013 年 5 月第 1 次印刷

定 价/56.00 元

中国统计版图书, 版权所有。侵权必究。

中国统计版图书, 如有印装错误, 本社发行部负责调换。

《中华医学统计百科全书》 专家指导委员会

主 任 方积乾
总 主 编 徐天和
委 员 (以姓氏笔画为序)
万崇华 方积乾 王广仪 田小利 田考聪
苏为华 苏颀龄 周燕荣 柳 青 赵耐青
饶绍奇 唐 军 徐天和 徐勇勇 徐端正
景学安 程 琮 颜 虹

《医学研究与临床统计设计分册》 编 委 会

主 编 徐勇勇
副 主 编 刘玉秀 陈长生
主 审 徐端正
编 委 (以姓氏笔画为序)
尹全焕 王 彤 王 玫 石德文 刘玉秀 刘延龄
孙 高 成 琪 闫丽娜 张 风 张中文 张罗漫
张晋昕 李 康 李克均 李炽民 杨振明 陈 林
陈长生 陈平雁 孟 虹 易 东 易雁青 林 林
郑俊池 姜 明 相 静 胡良平 赵清波 党少农
夏结来 徐天和 徐勇勇 贾改珍 高 永 高 辉
高纯书 曹秀堂 黄先敬 韩春蕾 缪华章 颜 艳
魏庆琤
秘 书 长 高 永
学术秘书 王 玫 石德文

序 言

国家统计局局长 马建堂

随着时代前进和科学技术的进步,我国的统计科学和医学统计工作的发展进入了一个崭新的阶段。统计科学既是认识社会现象与自然现象数量特征的手段,又是获取信息和进行科学研究的重要工具,历来为人们所重视。自20世纪20年代起,统计学理论与方法日益广泛地被应用于医学领域。近些年来,随着基因组学、蛋白质组学、药物开发、公共卫生、计算机和信息等学科的迅猛发展,统计学与医学学科的交叉融合不断深入,统计科学在医学领域中的应用与发展提高到了一个水平。

医学统计是统计科学的重要分支,也是国民经济和社会发展统计的重要组成部分,它关系到人民健康水平的提高和国家的长足发展。医学是强国健民学科,医学研究的对象是人及人群的健康,具有复杂性、特殊性、变异性等特点,这无疑需要全面系统的统计分析方法的支持与帮助。随着统计科学的迅猛发展,一些新的统计方法如遗传统计、多水平模型、结构方程模型、健康量表等不断涌现。一方面这些新的统计方法和理论亟需在医学科学领域内推广应用,为医学发展提供支持和帮助,另一方面,医学科研工作者为了科学研究工作的需要也迫切要求了解和掌握一些最新的、全面系统的统计方法和理论。因此,对当代医学科学研究中的统计分析方法进行全面系统的研究与介绍,是十分重要的一件事情,《中华医学统计百科全书》正是在这样的背景下编纂而成的,它满足了当前医学科学发展的需要,不失为一部好的大型医学统计参考书。

《中华医学统计百科全书》自2009年1月开始编写,由国内外著名医学院校的统计学教授和专家担任主编和编委,可谓编写力量强大,在编写过程中,他们本着精益求精的精神,精雕细琢,采百家之所长,融国内外华人统计学专家之所成。历时三年,终成其册。本套书内容浩繁,共八个分册,包含描述性统计分册、单变量推断统计分册、多元统计分册、非参数统计分册、管理与健康统计分册、医学研究与临床统计设计分册、健康测量分册和遗传统计分册。各

分册在内容上相互衔接并互为补充,贯穿“从简单到复杂”,“从一般、传统到先进、前沿”的循序渐进的编纂思路,一改目前医学统计著述中普遍存在的方法之间或评价指标之间缺乏相互联系、过于分散和单一的状况,使医学统计理论与方法更加具备了系统性、完整性与时代前沿性。本套书结构严谨,层次分明,科学性强,既突破了传统的辞典式编撰方法,又吸取了辞典的某些特点,在实用性、知识性、可读性、可查性等方面均具独到之处。

《中华医学统计百科全书》适应了我国医学科学研究发展对统计分析方法的需要,本书的出版,势必会大大促进我国现代医学的发展。本书既是我国医学统计工作者、医疗卫生统计信息工作者、高等医学院校师生以及广大医务工作者必备的大型医学统计参考工具书,也适合于医学各不同层次和不同专业的读者阅读。我相信本书的出版,不仅对于促进我国医学统计发展,促进我国与国际生物医学统计间的交流,繁荣社会主义先进文化具有重要意义,而且该书也必定会成为广大医学科学研究工作者的良师益友,故欣然为之作序。

编者的话

近年来,医学统计科学发展迅速,如遗传统计、多水平模型、结构方程模型、健康量表等新的统计理论与方法不断涌现,并被应用到医学科研实践中。这些新的统计理论与方法在医学科学研究中的不断拓展应用,要求广大的医学科技工作者在工作中必须学习和掌握这些新知识。所以,怎样使这些新的统计理论与方法易于被广大的医学科技工作者接受和使用,以提高医疗卫生工作质量,成为统计学专家的首要解决的任务。为此,组织编纂一部适合于广大医学科技工作者学习和使用的工具书,成为当前形势之必需。《中华医学统计百科全书》(下文简称“全书”)正是基于这样的背景而孕育产生的。

编纂“全书”的想法一经提出,就得到了国内高等医学院校和科研院所的统计学专家们的赞同。专家们云集一堂,进行商讨,达成共识——要集全国高等医学院校和科研院所的统计学专家之力,编纂出一部内容全面、概念精确、表述完整、接近世界医学统计学先进水平、编辑形式简洁的大型医学统计学工具书。2008年,“全书”开始酝酿筹备,几经讨论,搭成框架条目,确定编写格式,并开始全面着手编写,终于于2011年初编纂出初稿。值得欣喜的是,在中国统计出版社的大力支持下,“全书”项目先后成功申报了国家出版基金(项目编号2011C₂-003)和全国统计科学研究(计划)课题(立项编号2011LY080),皆荣获批准。有了国家出版基金和全国统计科学研究(计划)课题的支持,“全书”的编纂工作如虎添翼,更上台阶。

通过国内外数十所大学、医学院校与医学科研院所近百位统计学专家教授的共同努力,“全书”终于能够付梓成册,得以与广大读者见面,编者倍感欣慰。“全书”既全面介绍了医学统计学的基本理论、基本知识与方法,又介绍了大量新的统计理论与方法,对生物医学统计的传统方法及最新进展进行了全面梳理,同时还改变了目前医学统计著述中普遍存在的统计方法或指标之间缺乏相互联系,过于分散与单一的现象。这就形成了“全书”的特点:全面、系统、实用、前沿。

“全书”共8个分册:描述性统计分册、单变量推断统计分册、多元统计分册、非参数统计分册、管理与健康统计分册、医学研究与临床统计设计分册、健康测量分册、遗传统计分册,均由著名高校医学统计学教授担纲主编,同

时聘请国内外知名医学统计教授担任顾问。可谓举全国名校之力,集百家精英之长。在编写过程中,专家们严谨认真,精益求精,在注重科学性、知识性、先进性、可读性的前提下,紧紧把握医学科学研究与医疗卫生工作的特殊性和复杂性,精心研究论证各种统计理论与方法在医学领域的适用性与应用条件。为了便于读者学习和理解应用,书中不仅有理论分析,还提供了实例运用,并把计算机软件程序应用于其中,对统计方法或体系的科学性与可行性进行检验,使统计理论与医学实际得到紧密结合。在每一分册的内容安排上,遵循从简单到复杂、从一般到先进、从传统到前沿的原则,使各分册在内容上既相互衔接补充,融为一体,又能各自独立成册。为方便读者查阅,书中各条目层次分明,结构严谨,醒目易读,是广大医学科学工作者学习和使用、必备案头的大型医学统计工具用书。

“全书”在编写过程中,引用了相关专著及教材的部分资料,在此对引用资料的原作者表示衷心感谢! 引用资料中多数已在书中注出,也有部分没有一一注出,对于没有注出的部分,在此敬请原作者给予谅解! 中国统计出版社教材编辑部和滨州医学院的领导及同仁们为“全书”的编辑和出版付出了大量心血,在此致以诚挚感谢!

由于编者水平有限,书中难免会存在错误和不足之处,恳请广大读者提出宝贵意见。

最后,感谢您学习和使用“全书”,希望它能使您开卷有益。

总主编 徐天和

前 言

20世纪20年代,英国统计学家 R. A. Fisher(1890~1962)在伦敦附近的 Rothamsted 农业实验站创立了实验设计理论以及以实验设计为基础的现代统计方法。1948年,《英国医学杂志》报告了世界上第一篇采用随机对照设计的临床试验结果。医学研究统计设计的理论和方法体系的建立尚不满百年,且还在不断发展。近年来,国内外医学期刊和新药评审对统计设计内容的审查越来越严格,越来越多的医学研究工作者需要了解、学习和掌握医学研究统计设计的基本概念、基础理论、设计方法和研究进展。作为《中华医学统计百科全书》的一个分册,《医学研究与临床统计设计分册》的主要内容包括统计设计概述、随机分组方法、随机抽样方法、实验设计方法、调查设计方法、临床试验设计方法等。本书围绕医学研究与临床研究中的基本统计学问题,用简洁的文字描述取代复杂的公式表达,大多数条目内容浅显易懂,不要求读者具有很深的统计学知识,既可以作为医学科研人员、临床医生、卫生防疫人员从事科研工作的工具书,也可以作为在职人员、在校学生学习医学统计学课程的参考书。

在编写过程中,许多国内外学者给予了大力支持,加拿大著名统计专家曼尼托巴大学的王熙逵教授推荐易雁青教授参加本书的编写;复旦大学著名学者徐端正教授担任本书主审,提出了许多宝贵的修改意见,并献出自己的最新英文专著供本书编写参考。在此一并感谢!

本人水平有限,加之医学研究中关于统计设计的规范化要求还在不断变化,统计设计的理论和方法的研究不断有新发展,尚有一些关于医学和临床研究统计设计的内容没有收入本书,希望广大读者批评指正。

徐勇勇

2012年10月

目 录

医学研究设计	(1)
实验对象与实验单位	(6)
随机数	(7)
随机化	(7)
随机排列表	(11)
干预与对照	(12)
样本大小与检验效能	(13)
实验因素与水平	(14)
实验效率	(15)
实验效应	(16)
数据误差与数据精度	(17)
数据效度	(18)
测量信度	(22)
预试验	(28)
质量控制	(29)
主效应	(32)
交互作用	(33)
关联与因果	(34)
完全随机设计	(34)
配对设计	(36)
随机完全区组设计	(37)
不完全区组设计	(39)
平衡不完全区组设计	(41)
部分平衡不完全区组设计	(44)
格点设计	(46)

约登方设计与行列设计	(48)
交叉设计	(52)
拉丁方设计	(55)
希腊-拉丁方设计	(58)
析因设计	(60)
正交设计	(64)
均匀设计	(71)
系统分组设计	(74)
裂区试验设计	(76)
序贯试验设计	(79)
单侧序贯 t 检验	(84)
开放型定性资料单侧序贯试验	(88)
开放型定性资料单侧配对序贯试验	(95)
开放型定量资料单侧序贯试验	(100)
开放型定量资料单侧配对序贯试验	(106)
双侧序贯 t 检验	(110)
Armitage 设计	(114)
Schneiderman-Armitage 设计	(118)
重复测量设计	(121)
调查设计	(133)
普查	(140)
典型调查	(141)
横断面调查	(142)
前瞻性调查	(145)
追踪性调查	(149)
回顾性调查	(152)
随机抽样	(156)
抽样总体	(157)
抽样框	(158)
简单随机抽样	(158)
分层抽样	(161)
系统抽样	(164)
整群抽样	(167)
多阶段抽样	(168)
快速评价抽样	(169)

临床试验分期	(172)
盲法试验	(174)
试验预案	(176)
完全随机分组	(179)
有限随机分组	(181)
区组随机分组	(181)
分层随机分组	(182)
适应性随机分组	(183)
试验监测	(186)
期中分析与成组序贯试验	(187)
生存质量研究的设计	(189)
临床决策分析	(198)
诊断试验	(200)
试验清洗期	(209)
随访与失访	(210)
试验终结	(211)
终结后随访	(212)
实验报告	(213)
纳入标准与剔除标准	(216)
知情同意	(216)
预后因素	(217)
多中心试验	(218)
样本含量估计的条件	(219)
总体均数区间估计样本含量	(220)
总体率区间估计样本含量	(221)
样本均数与总体均数比较样本含量估计	(222)
样本率与总体率比较样本含量估计	(223)
两样本均数比较样本含量估计	(225)
两样本率比较样本含量估计	(226)
多组样本均数比较样本含量估计	(227)
多个样本率比较样本含量估计	(229)
前瞻性(队列)研究样本含量估计	(230)
回顾性(病例对照)研究样本含量估计	(231)
直线相关分析样本含量估计	(232)
魔方设计	(233)

优化设计	(235)
准实验设计	(237)
临床试验适应性设计	(243)
反应适应性医学临床试验设计	(247)
靶向临床试验	(252)
靶向试验随机停药设计	(253)
靶向药物Ⅲ期临床试验设计	(255)
单组临床试验目标值法样本量估计	(259)
动态随机化	(262)
非劣效临床试验	(267)
富集设计	(269)
富集纳入随机退出设计	(271)
基于专业特长的随机对照试验设计	(274)
疾病修饰作用评价的设计	(280)
临床均势原则	(283)
非劣效界值	(285)
非劣效试验样本含量估计	(287)
临床试验统计师	(289)
临床试验统计学管理规范	(290)
临床试验效应指标	(292)
群随机试验	(293)
数据管理计划	(294)
双盲双模拟	(295)
随机化分配隐蔽	(296)
统计分析计划	(298)
统计分析数据集	(299)
以个体反应为条件的换组设计	(300)
以均势为入选标准的设计	(303)
意向处理原则	(305)
附录一 统计用表	(307)
附表1 随机数字表	(307)
附表2 随机排列表($n=20$)	(308)
附表3 随机排列表($n=10$)	(308)
附表4 平衡不完全区组设计参数表	(309)
附表5 平衡不完全区组设计表	(310)

附表 6	质反应单向序贯试验边界系数表	(314)
附表 7	量反应单向序贯试验边界系数表	(316)
附表 8	小样本翼型设计的边界点坐标表	(317)
附表 9	开放式定性资料单侧配对序贯试验的边界系数表	(319)
附表 10	不同 θ_1 错误水平分别为(0.05,0.05)与(0.01,0.05)的截断点、 边界参数以及等价样本量	(319)
附表 11	错误水平为(0.05,0.05)双侧序贯 t 检验边界点	(320)
附表 12	错误水平为(0.01,0.01)双侧序贯 t 检验边界点	(321)
附表 13	成组序贯的名义水准与显著性界值	(322)
附表 14	成组序贯设计参数 Δ 值表	(322)
附表 15	患病率 P 、诊断符合率 e 、灵敏度 Se 及特异度 Sp 变化对信息量 (1.4427 bit)的影响	(323)
附表 16	平均数抽样调查时不同 s/δ 所需样本含量($\alpha=0.05$)	(326)
附表 17	平均数抽样调查时不同 s/δ 所需样本含量($\alpha=0.01$)	(326)
附表 18	总体率区间估计所需样本数表	(327)
附表 19	样本均数比较时所需样本例数表	(330)
附表 20	样本率与总体率比较样本含量估计(单侧)	(331)
附表 21	样本率与总体率比较样本含量估计(双侧)	(334)
附表 22	两样本均数比较所需样本例数	(337)
附表 23	两样本率比较所需样本例数(单侧)	(338)
附表 24	两样本率比较所需样本例数(双侧)	(339)
附表 25	ψ 值表	(340)
附表 26	多组样本均数比较时样本含量估计用表	(342)
附表 27	λ 值表	(344)
附表 28	单组临床试验目标值法样本含量精确估计($\alpha=0.05, 1-\beta=0.80$)	(345)
附录二	英汉医学统计学词汇	(346)
附录三	汉英医学统计学词汇	(350)
本书词	条索引	(354)

医学研究设计

生物的变异性是医学研究中普遍存在的现象。医学研究中的许多事例证明,如果没有统计学的支持,将一次或几次试验的偶然成功看作必然现象,并将其作为研究成果加以推广,不仅使科学研究失去其应有的意义,而且可能会给人类的健康带来灾难性的后果。

1 医学研究类型与统计学要求

医学研究可分为描述性研究和分析性研究两大类,前者主要以研究者的真实观察结果为依据,描述新发现;后者主要是以研究者的试验为基础,研究和推论试验以外的特定总体是否具有同样的特征。在医学研究中,有时这两类研究并不能严格区分,在有的研究中还会两者兼而有之。以观察对象的选择方法为判断标准,医学研究还可区分为以下几种类型:

1.1 单一样本研究

按一定标准选择一个或若干个生物个体或生物资料作为研究对象,主要用于描述性研究。如阐明新的生命现象、形态结构,研究新的检测方法、诊断和治疗技术。在临床上,最多见的单一样本研究是Ⅰ期临床试验,即以志愿者作为观察对象,进行药代动力学和药物安全性的观察。由于单一样本的研究不设立对照组,在基础研究中,一定是新发现的科学事实,在临床研究中,则是Ⅱ期、Ⅲ期临床试验所必须经过的试验阶段,但Ⅰ期临床试验不能作为判定治疗的依据。

1.2 观察对比研究

观察对象人为选择,数据资料以观察的方法收集,通过观察结果的前后对照、历史对照或相互对照得出研究结论。常见的观察对比方法有以下3种。

1.2.1 无对照研究

通常是回顾性的,也可以是前瞻性的,以治疗前后病情变化判断疗效。多见于临床病例总结。

1.2.2 历史对照

通常是前瞻性研究,但由于缺乏足够的病例(如罕见病)或已有公认的历史对照,在研究中不设立对照组,以文献报道或以往常规治疗方法的治疗结果作为对照。

1.2.3 非随机化同期对照研究

前瞻性的研究,在研究中设立对照组和试验组,但受试者的分组按非随机化的方式

进行,如接诊医生根据病情、患者意愿选择治疗方法,或按患者经治医生、所属病区(医院)分组等。在涉及环境等自然因素的较大规模的现场试验中,准试验的分组方法应属于非随机化同期对照研究。出于伦理学的考虑,有些临床研究只能采用观察对比的研究方法,如某病患者与正常人的对比,吸烟者与非吸烟者的对比等。在观察对比研究中,通常样本量大(除罕见病),时间跨度长,由于诊断、治愈标准的不同以及对照组和试验组患者的病情程度、年龄、病程、病情、用药史的不同,在进行数据分析时要用统计学方法调整和控制这些因素对观察结果的影响。

1.3 随机对照研究

按照实验设计重复、对照和随机化的3个原则进行分析性研究。所谓重复就是同样的试验条件下要有符合统计学要求的观察例数;对照就是在试验内安排同期对照组,而且必须是专业上公认的可靠对照方法;随机化就是保证每个观察对象(生物材料)分到试验组和对照组的机会均等。只有随机对照的研究结果,如II期临床试验,才能使研究的推论具有普遍性的科学意义。

1.4 调查研究

通过观察部分个体或全体了解特定总体(人群)的一般特征,如健康状况、生育模式、疾病流行因素、老年人医疗保障情况等。调查研究要有明确的抽样框架(“目标人群”)和调查时点,调查部分个体时要采用随机抽样的方法。

2 干预试验的设计问题

干预试验的主要设计问题是,如何在实验设计和试验过程中遵循重复、对照和随机化的原则。

据文献记载,有控制的一次医学干预试验发生在1747年。1747年5月20日,英国的J. Lind医生将12名病情相类似的患者带到一艘船上。患者的主要症状是牙龈溃烂,皮肤有出血点,双膝无力。Lind将12名患者分为6组,分别给予下列主要干预:

A组:每天饮1夸脱(1.136L)苹果汁;

B组:服25滴硫酸丹剂(elixir vitriol),每天3次;

C组:服2匙醋,每天3次;

D组:每天饮约半品脱海水,服缓和的泻药;

E组:每天食2个桔子,1个柠檬;

F组:每天服大蒜、芥子等成分组成的干药。

当6月16日船返回英国Plymouth港时,所有患者的病情都有好转,其中E组恢复的最快、最好,其中1人到第6天就可以工作了。B组也有1人比登船时健康。Lind医生的试验并没有得出明确的结论,直到160年后,通过动物实验才真正从科学意义上解释了E组疗效最好的真正原因。

根据现代医学知识,Lind医生描述的病例是坏血病患者的常见症状,但目前已无从考察其中真正坏血病患者的人数。从干预效果看,E组中病情好转最快的患者可能是真正的坏血病患者(坏血病的对症治疗是补充维生素C)。用现代统计学的观点看,Lind医生的干预试验至少有2点符合统计学要求。一是设有对照组,6个干预组相互比较;二是

进行了质量控制,所有患者在同一条船上,便于监督服药和观察病情。关于无对照和质量控制不严而导致研究结果“失真”的例子,在近代医学的研究中也没有完全避免。

例 1 门腔分流手术治疗肝硬化门静脉高压症,前后有 51 次研究(表 1)。无对照研究支持门腔分流手术的比例是 24/32(75%),而按统计学原则进行的 4 次随机对照研究,认为手术几乎没有价值。无对照的研究夸大了门腔分流手术的作用。

表 1 关于门腔分流手术的 51 次研究结果

对照方式	门腔分流手术的价值			小计
	非常支持	支持	不支持	
无对照	24	7	1	32
非随机对照	10	3	2	15
随机对照	0	1	3	4
合计	34	11	6	51

Journal of Gastroenterology. 1966,50:646—691.

例 2 一项随机对照的双盲试验,受试者为 8341 名有心脏病的中年男子,干预是可能防止心脏病突发的 5 种药物和安慰剂(相同的药物胶囊装入乳糖),用 5 年追踪观察的死亡率评价 6 种干预的效果。服用安妥明(Clofibrate)的患者中有 357 人没有达到应该服用药量的 80%,其死亡率达 24.6%;而坚持服用达药量 80%的患者,其死亡率为 15.0%,而所以不坚持服药是增加死亡率的主要原因,不能归究于安妥明的预防效果。再看表 2 中的失访人数(尽管在慢性病研究中失访不可避免),合计死亡率 18.2%,是假定失访的 38 人(失访率 3.6%)全是病情轻的患者,且无一例死亡。如果失访者全是重症患者,或失访率达到 20%、30%或更多,你还能信任表 2 中的合计死亡率吗?

表 2 服用安妥明的死亡率

分 组	人 数	死亡人数	死亡率/%
服药 \geq 80%	708	106	15.0
服药 $<$ 80%	357	88	24.6
失 访	38	—	—
合 计	1065	194	18.2

受试者是否遵从医嘱,并使干预措施得到充分落实的程度,称为患者的依从性(compliance)。依从性和失访是干预效果评价的 2 个“陷阱”。

如果今天重复 Lind 医生的干预试验,在统计设计和统计分析上还应做到:

- 1) 所有受试者均为经临床和实验室检验确诊的坏血病患者。
- 2) 有足够多的试验病例。样本估计方法详见样本含量估计有关条目。
- 3) 采用随机分组方法,保证各组干预效果的可比性。Lind 医生的试验将病情最重的 2 名患者分到 D 组(泻药是当时流行的治病方法),D 组的干预效果不及 E 组和 B 组也可