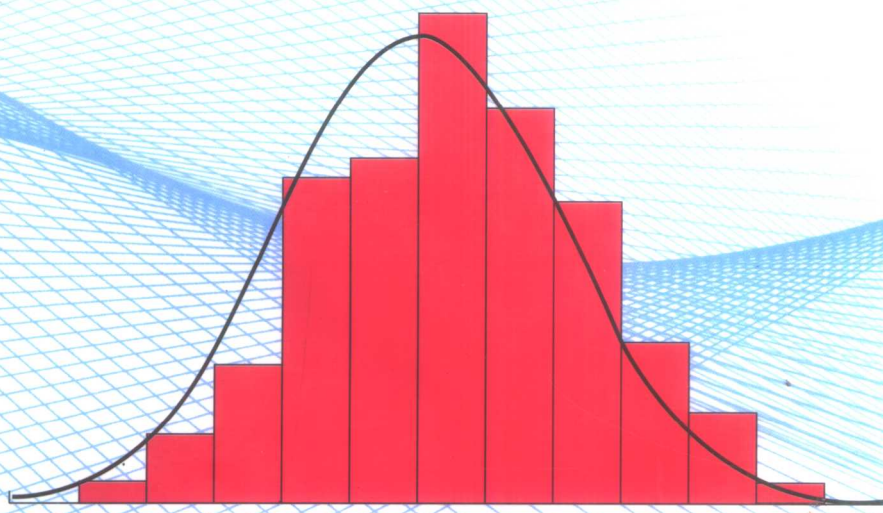




面向 21 世纪 课 程 教 材
Textbook Series for 21st Century

医学统计学

徐勇勇 主编



高等教育出版社
HIGHER EDUCATION PRESS

面向 21 世纪 课程教材
Textbook Series for 21st Century

医学统计学

徐勇勇 主编



高等教育出版社
HIGHER EDUCATION PRESS

内容提要

本书是教育部“高等教育面向 21 世纪教学内容和课程体系改革计划”的研究成果,是面向 21 世纪课程教材。

本书内容分绪论、统计描述、两组资料均数的比较、多组资料均数的比较、两个率或多个率的比较、非参数统计分析方法、回归与相关、实验设计、调查设计与资料分析、生存分析、统计软件简介、医学论文统计表述的基本要求等十二章,其中带 * 号的章节供学有余力的学生课外阅读。此外书末附有 6 个附录,其中统计用表 20 种,社会医学中常用的生命、健康、疾病的统计指标,国家执业医师医学统计学考试模拟试题,供进一步学习的主要参考文献等。

本书主要供五年制临床医学专业学生使用,也适用于非临床医学专业的本科生和研究生,临床医生也可作为医学统计学的参考书阅读。

图书在版编目(CIP)数据

医学统计学/徐勇勇主编. —北京:高等教育出版社,
2001(2003 重印)

ISBN 7-04-009188-7

I. 医… II. 徐… III. 医学统计 IV. R195.1

中国版本图书馆 CIP 数据核字(2000)第 85948 号

医学统计学

徐勇勇 主编

出版发行 高等教育出版社

社 址 北京市东城区沙滩后街 55 号

邮政编码 100009

传 真 010-64014048

经 销 新华书店北京发行所

排 版 高等教育出版社照排中心

印 刷 北京民族印刷厂

开 本 850 × 1168 1/16

印 张 17.75

字 数 360 000

购书热线 010-64054588

免费咨询 800-810-0598

网 址 <http://www.hep.edu.cn>

<http://www.hep.com.cn>

版 次 2001 年 2 月第 1 版

印 次 2003 年 2 月第 4 次印刷

定 价 19.40 元

本书如有缺页、倒页、脱页等质量问题,请到所购图书销售部门联系调换。

版权所有 侵权必究

《医学统计学》编写人员

主 编	徐勇勇	(第四军医大学)
副主编	孙振球	(中南大学湘雅医学院)
	颜 虹	(西安交通大学医学院)
	何大卫	(山西医科大学)
	周燕荣	(重庆医科大学)
	詹绍康	(复旦大学医学院)
编 委	骆福添	(中山医科大学)
	孙 高	(中国医科大学)
	颜 艳	(中南大学湘雅医学院)
	王 彤	(山西医科大学)
	陈平雁	(第一军医大学)
	贺 佳	(第二军医大学)
	熊林平	(第二军医大学)
	易 东	(第三军医大学)
	夏结来	(第四军医大学)
	曹秀堂	(第四军医大学)
	赵清波	(第四军医大学)
	姚 晨	(军医进修学院)
	张高魁	(军医进修学院)
	钟晓妮	(重庆医科大学)

学术秘书 赵清波(兼)

(高等学校骨干教师资助计划资助)

前 言

本教材是中华人民共和国教育部指导下,在所有高校医学专业课程中遴选出来的具有教学效果好、符合 21 世纪人才培养需求的优秀课程教材。

医学统计学是国内外临床医学专业的一门重要的基础课程,它是 21 世纪临床医生在从事临床工作和科学研究过程中必须掌握和了解的基本知识,统计学方法被医学界比喻为整个医学大厦中的一个支柱。

我国医学统计学的奠基人之一,带我步入医学统计学术之园的启蒙老师——郭祖超教授(1912—1999)编著的《医学与生物统计方法》早在 1948 年正式出版,并被当时的教育部定为大学用书。从那时候算起,我国为医学生开设医学统计课程已有 50 多年的历史。进入 21 世纪以后,随着信息技术的飞速发展,计算机随处可见,各种数据处理软件和统计软件唾手可得,统计学基本知识和基本原理的教学更显迫切。否则,一方面有人不懂得选用正确的统计方法,使大量的信息和统计数据得不到有效的利用,另一方面又盲目使用计算机和统计软件,不管是什么研究类型的数据都简单地交给计算机处理,用计算机取代统计,势必造成大量统计方法的滥用和误用。医学研究的许多数据关系到病人的治疗、转归,甚至生死,统计方法的误用会导致严重的伦理问题。计算机、统计软件、多媒体教学课件也同时引发了 21 世纪医学统计教学方式的深刻变革,使课堂教学摆脱了大量的繁琐演算的束缚,在单位时间内讲授的信息量将会大幅度增加。今后,统计学教学给人们的印象不应该再是数据的罗列和公式的堆砌,而应该是研究设计的艺术和数据表达的艺术,并能使学习者体会到统计思维和推理的乐趣。本教材在教学内容的编排上努力顺应了这一变化趋势。除了第一章至第三章外,其他各章都安排了“窗口”内容(目录中标有*),教师可根据不同的教学对象灵活选择,也可以指导学生课外阅读。

本教材的主要教学对象是五年制本科临床医学专业学生,也适用于其他非临床医学专业的学生。如果把部分“窗口”内容也纳入课堂教学,也可作为七年制临床医学专业学生、医学硕士研究生的医学统计学课程教材。

本教材的编写得到高等教育出版社、第四军医大学和中南大学领导和有关部门的大力支持。2000 年 4 月在长沙召开的医学统计学教学研讨会上,25 所医学院校的统计教师对本教材的初稿提出了许多宝贵意见。我两位尊敬的老师,四川大学的杨树勤教授(原华西医科大学)、中南大学的黄镇南教授(原湖南医科大学)始终关心着本教材的编写,他们的意见和建议使本教材大为增色。

感谢所有参加编写的老师为本教材付出的辛勤劳动。共同的劳动和创作使

我们增进了友谊,教材的字里行间里也留下了一段段美好的记忆。

感谢第四军医大学卫生统计学教研室姜建辉同志制作的图表,赵瑞松同志承担的全部稿件的整理和打印工作。

由于本人能力所限,教材中肯定存在许多不足之处,诚恳希望广大师生提出进一步修改意见。

徐勇勇

2000年7月于西安

目 录

第一章 绪论	1
第一节 统计学与医学统计方法	1
第二节 统计学基本概念	2
第三节 医学中常见的统计学问题	4
练习题	9
第二章 统计描述	11
第一节 计量资料的统计描述	11
第二节 正态分布	20
第三节 计数资料的统计描述	22
第四节 统计表与统计图	24
练习题	31
第三章 两组资料均数的比较	33
第一节 均数的抽样误差	33
第二节 t 分布与可信区间	36
第三节 均数差别比较的 t 检验	40
第四节 假设检验的步骤及有关概念	47
练习题	50
第四章 多组资料均数的比较	52
第一节 方差分析的基本思想及其应用条件	53
第二节 完全随机设计资料的方差分析	54
第三节 随机单位组设计资料的方差分析	56
第四节 均数间的多重比较	58
第五节 析因设计资料的方差分析 *	62
第六节 重复测量设计资料的方差分析 *	65
练习题	68
第五章 两个率或多个率的比较	70
第一节 率的抽样误差与可信区间	70
第二节 两个率的比较	72
第三节 行 $\times$ 列表资料的 χ^2 检验	78
第四节 率的多重比较 *	81
练习题	83
第六章 非参数统计分析方法	85

第一节	两组资料差别的秩和检验	86
第二节	配对设计资料的秩检验	88
第三节	完全随机设计多组差别的秩和检验	90
第四节	随机单位组设计的秩和检验 *	92
练习题		93
第七章	回归与相关	95
第一节	直线回归	95
第二节	直线相关	102
第三节	等级相关	107
第四节	多元线性回归与逐步回归 *	109
第五节	其他多元统计分析方法 *	114
练习题		119
第八章	实验设计	122
第一节	实验研究的分类及其基本要素	123
第二节	实验设计的基本原则	125
第三节	随机化分组方法	127
第四节	样本含量估计方法	130
第五节	临床试验设计 *	134
第六节	临床试验的评价 *	138
第七节	病案首页信息与统计分析 *	141
第八节	多个临床试验结果的综合分析 *	145
练习题		146
第九章	调查设计与资料分析	148
第一节	调查设计的内容	148
第二节	现况调查	155
第三节	队列研究 *	160
第四节	病例对照研究 *	165
第五节	标准化率 *	170
练习题		172
第十章	生存分析	174
第一节	生存分析基本概念	174
第二节	生存率的 Kaplan - Meier 法与寿命表法估计	177
第三节	生存曲线的比较 *	180
第四节	Cox 回归模型简介 *	182
第五节	寿命表简介 *	185
练习题		188
第十一章	统计软件简介	190
第一节	原始数据的记录与预处理	190

第二节	SPLM 中文统计软件 *	191
第三节	SPSS 统计软件 *	198
第四节	SAS 统计软件 *	202
第十二章	医学论文统计表述的基本要求	208
第一节	医学论文中的统计表达	208
第二节	数字表达 *	213
第三节	外文字母的表达与用法 *	215
第四节	法定计量单位的使用 *	216
附录 I	生命、健康、疾病统计常用统计指标	218
附录 II	附表目录	224
附录 III	国家执业医师(卫生学)医学统计学考试模拟试题	253
附录 IV	主要参考文献	257
附录 V	英汉名词对照及索引	260
附录 VI	中英文名词对照及索引	266

第一章

绪 论

不明于计数而欲举大事,犹无舟楫而欲经于水,险也。
举事必成,不知计数不可。

管仲(? —645BC)

第一节 统计学与医学统计方法

提起统计,非常容易使人联想到数据汇总,如常常能听到这样的说法:统计一下每个班的学生人数,统计一下食管癌手术的患者一年后还有多少人存活,等等。从统计发展的历史看,最初的统计确实主要是数据汇总,《史记》上就记有“禹平水土,定九州,计民数”的记载。但统计发展到今天,统计是一种对客观现象数量方面进行的调查研究活动,是收集、整理、分析、推断、判断等认识活动的总称,数据汇总仅仅是统计工作的一小部分内容。

统计学(statistics)作为一门学科的定义是:关于数据收集、表达和分析的普遍原理和方法。

统计方法是建立在现代科学方法之上,由统计学理论指导的数据收集、表达和分析的方法。现代科学方法可以概括为以下几点:

1. 问题的识别与表达。
2. 搜集有关资料。
3. 通过归纳得出假说:因果联系及重要的模式。
4. 从假说作出演绎:进行实验或收集更多的资料。
5. 推理:结果与演绎相符,假说得到加强,但不是被证明。

现代科学方法适用于一切科学研究过程,但在许多研究领域,特别是在生物医学研究领域,实验或观察结果往往是不确定现象。描述不确定现象,通过重复观察发现不确定现象背后隐藏的统计学规律,是统计方法的显著特征。

医学统计学是用统计学原理和方法研究生物医学问题的一门学科。医学统计方法在医学研究中的运用主要有三个方面:

1. 以正确的方式收集数据,如实验设计、调查设计等。
2. 描述数据的统计特征,如数据化简、统计指标的选择与计算、统计结果的表达等。
3. 统计分析及得出正确结论,如根据概率分布,对实验和观察结果存在的差异和关联作出统计推断。

第二节 统计学基本概念

一、随机变量

随机变量(random variable)指取值不能事先确定的观察结果,通常简称为变量(variable),统计上用大写的拉丁字母表示。例如 10 只试验动物在一定时间内死亡的只数记为 K ;某地男性健康人群中,各个个体血清血红蛋白量的测量值记为 X_i 或 X 。随机变量的实际观察结果用小写的拉丁字母表示,例如各用 10 只动物进行 2 次试验,死亡的动物数分别为 $k_1 = 7, k_2 = 9$;两名健康男子血清血红蛋白量的测量值分别为 $x_1 = 23.4 \text{ mg/L}, x_2 = 24.6 \text{ mg/L}$ 。另有一些现象并不直接表现为数量,例如某病患者治疗结果的有效或无效、试验结果的阳性或阴性等,但我们可以规定有效为 1,无效为 0;阳性为 1,阴性为 0,则非数量标志也可以用变量来表示。变量的具体内容虽然是各式各样的,但共同的特点是不能用一个常数(constant)表示,而且理论上讲,每个变量的取值服从特定的概率分布。

随机变量的取值可以分为两种基本类型:① 离散型变量(discrete variable),即在一定区间内变量取值为有限个,或数值可以一一列举出来。例如某地区某年人口的出生人数、死亡人数,某药治疗某病病人的有效人数、无效人数等。② 连续型变量(continuous variable),即在一定区间内变量取值有无限个,或数值无法一一列举出来。例如某地区男性健康成人的身高值、体重值,一批传染性肝炎患者的血清转氨酶测定值。

二、误差

误差(error)是指实际观察值与观察真值之差、样本指标与总体指标之差。误差可分为系统误差(systematic error)和随机误差(random error)。

1. 系统误差 在实际观察过程中,由于仪器未校正、测量者感官的某种偏差、医生掌握疗效标准偏高或偏低等原因,使观察值不是分散在真值两侧,而是有方向性、系统性或周期性地偏离真值。这类误差可以通过实验设计和技术措施来消除或使之减少。观察性研究由于组间不可比性产生的系统误差称为偏倚(bias),如吸烟组的平均年龄大于不吸烟组,两组死亡率的差异包含年龄偏倚。

2. 随机误差 或称偶然误差,是指排除了系统误差后尚存的误差。它受多种因素的影响,使观察值不按方向性和系统性而随机地变化,误差变量一般服从正态

分布。

随机误差中还包括重复测量误差。它是由于对同一受试对象或检样采用同一方法重复测定时所出现的误差。如用天平称同一个物体的重量,重复测定多次,其结果会有某些波动。控制重复测量误差的手段主要是改进测定方法,提高操作者的熟练程度。

三、数据类型

医学测量结果按其性质分为计数资料(count data)、计量资料(measurement data)和等级资料(ordinal data)三类。

1. 计数资料 定性观察结果,分二类反应和多类反应,如某患者临床治疗结果的有效、无效是二类反应(两者必居其一),某人的血型是四类反应(A、B、AB、O四种血型必居其一)。计数资料的常见形式是各类反应的绝对数,如治疗一批患者后的有效、无效人数,调查一批少数民族居民A、B、AB、O四种血型的人数。

2. 计量资料 用仪器、工具或其他定量方法获得的定量结果,一般带有计量单位,如某一患者的身高(cm)、体重(kg)、血压(kPa)、脉搏(次/分)、红细胞计数($10^{12}/L$)。

3. 等级资料 半定性或半定量的观察结果,如对急性病毒性肝炎患者作麝香草酚絮状试验,结果为-、+或++;患者的治疗结果评定为治愈、好转、有效、无效或死亡,各种结果既是定性分类,也有顺序和等级差别,但这种差别却不能精确量化。统计上可以将其视为计数资料,比较各个类别分布差异,但更好的处理方法是将各个类别的等级数量化,如转换为平均秩号,再做进一步分析。

4. 资料的转换 为了研究的需要,有时把计量资料转换为两分类或多分类的计数资料,如成年男子的血清总胆固醇按是否小于6 mmol/L划分血脂正常和异常两类,年龄划分为儿童、少年、青年、中年、老年五类。

四、总体与样本

特定研究对象中所有观察单位的测量值称为总体(population)。总体通常很大,或者根本无法得到,如某地区18岁男青年的身高测量值,全部高血压患者的血清总胆固醇测量值。当总体中所有观察单位都能够标识时,如某城区全部住户的门牌号码、公民的身份证号等,称为有限总体(finite population);反之,称为无限总体(infinite population),如某地野生老鼠数量。来自总体的部分观察单位的测量值称为样本。有限总体在获得抽样框架(sampling frame)后,可以实现随机抽样(random sampling),即总体中的每个观察单位都有同样的机会被选作样本。总体的统计指标,如均数、率,称为总体参数(parameter);样本的统计指标称为统计量(statistic),如样本均数、样本率等。多数情况下不容易知道总体参数,但可以通过随机抽样抽取有代表性的样本,用统计量估计未知的总体参数。总体参数是固定的常数,统计量是在总体参数附近波动的随机变量。统计量与总体参数的差别,称为抽样误差(sampling error)。在随机抽样的情况下,大样本的样本均数和样本率服从正态分布。

五、概率

概率(probability)又称几率,是度量某一随机事件 A 发生可能性大小的一个数值,记作 $P(A)$ ($1 > P(A) > 0$)。 $P(A)$ 越大, A 事件发生的可能性越大,反之亦然。当 $P(A) = 1$ 时或 $P(A) = 0$ 时, A 是必然发生或必然不发生的非随机事件。

实践证明,虽然随机事件 A 在一次试验或观察中可能出现也可能不出现,但在大量重复试验中它却呈现出明显的统计规律性。假设在相同条件下,独立地重复做 n 次试验, A 在 n 次试验中出现了 m 次,则比值 m/n 称为随机事件 A 在 n 次试验中出现的频率(frequency)。当试验重复很多次时,有

$$P(A) \approx \frac{m}{n} \quad (1-1)$$

由此可见,频率是就样本而言的,而概率总是从总体的意义上说的, m/n 是 $P(A)$ 的估计值。试验次数越多,估计越可靠。

抽样误差的概率也可用公式 1-1 估计。设 X 为抽样误差,事件 A 为“ $|X|$ 大于给定常数 ϵ ”,在已知总体独立地重复做 n 次抽样试验, m 是 A 发生的次数。但在只有一个样本的情况下, $P(A)$ 的估计要借助概率分布,如正态分布、 t 分布等。概率分布是在同一总体中抽样的假设下, X 在 $(-\infty, +\infty)$ 任一区间分布的概率。当 $P(A)$ 很小时(如小于 0.05 或 0.01),称 A 为小概率事件。如果一次抽样试验恰巧碰上小概率事件,就有理由怀疑 X 不是抽样误差所致,或者说“在同一总体中抽样”的假设不正确。

$P(A)$ 有时也不能用频率解释,如某罕见病的治愈概率、某幢大楼倒塌的概率、某地发生地震的概率等。这种不依赖客观频率对一个事件真实性的主观判定称为主观概率。经典统计方法对概率的解释是以频率为基础的,所以历史上形成了经典学派和对概率做主观解释的贝叶斯学派。本书介绍的主要是经典的统计方法,其中许多经典的统计方法亦可以用贝叶斯方法导出,只是在说法和解释上有所不同。而且在一些情况下,贝叶斯的解释可能更符合思维逻辑。例如,投掷硬币试验,经典学派通过实践,认为随着试验次数 n 逐渐增大,硬币“正面朝上”的稳定频率是 0.5。但我们知道了硬币质地均匀,不通过试验也可以主观估计出同样的概率值。

第三节 医学中常见的统计学问题

个体差异是生物医学研究领域里普遍存在的现象。严格地说,在自然状态下,任何两个患者或两个研究群体间都存在差异,其表现结果为各种生理测量值的参差不齐,如身高、体重、血压、胆固醇、血糖、免疫球蛋白、心功能、肺功能、肝功能,等等。即使同一个个体,许多生理测量值也不是恒定的常数。这种在自然状态下测量结果的差异,统计上称为变异(variation)。

变异的存在,导致我们对实验结果判断的不确定性。例如,用两种方法分别治

疗两例入院时病情相似的急性出血热患者,结果一例治愈,一例死亡。由于存在变异,我们不能简单地肯定是治疗方法不同产生的差异。因为,即使用同样的方法治疗这两名患者,由于患者病情的变化,也可能出现一例治愈,一例死亡的结果。因此,当存在着变异时,同样条件下的重复实验,结果不会完全相同,医生和医学研究人员必须学会用概率的眼光看待生命现象中的因果联系。

一、变异的描述

表 1-1 是医学中最常见的数据类型,必须用统计方法进行整理后,才能看出其变异规律。据图 1-1,这些数据的变异特征至少有以下两点:①变异的范围在 $3.2 \sim 6.2(10^{12}/L)$;②有明显的统计分布规律, $4.7 \sim 5.0(10^{12}/L)$ 的人数最多。根据表 1-1 的数据,统计上还可以进一步推论出正常成年男子红细胞计数的正常值参考范围。

表 1-1 120 名正常成年男子红细胞计数值($10^{12}/L$)

5.12	5.13	4.58	4.31	4.09	4.41	4.33	4.58	4.24	5.45	4.32	4.84
4.91	5.14	5.25	4.89	4.79	4.90	5.09	4.64	5.14	5.46	4.66	4.20
4.21	3.73	5.17	5.79	5.46	4.49	4.85	5.28	4.78	4.32	4.94	5.21
4.68	5.09	4.68	4.91	5.13	5.26	3.84	4.17	4.56	3.52	6.00	4.05
4.92	4.87	4.28	4.46	5.03	5.69	5.25	4.56	5.53	4.58	4.86	4.97
4.70	4.28	4.37	5.33	4.78	4.75	5.39	5.27	4.89	6.18	4.13	5.22
4.44	4.13	4.43	4.02	5.86	5.12	5.36	3.86	4.68	5.48	5.31	4.53
4.83	4.11	3.29	4.18	4.13	4.06	3.42	4.68	4.52	5.19	3.70	5.51
4.64	4.92	4.93	4.90	3.92	5.04	4.70	4.54	3.95	4.40	4.31	3.77
4.16	4.58	5.35	3.71	5.27	4.52	5.21	4.37	4.80	4.75	3.86	5.69

最大值 = 6.18, 最小值 = 3.29, 极差 = 2.89。算术均数 = 4.72, 标准差 = 0.57。

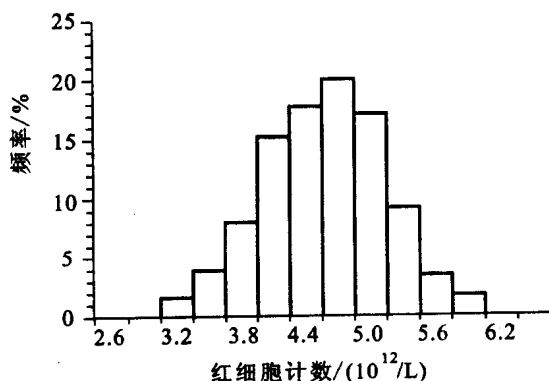


图 1-1 120 名正常成年男子
红细胞计数直方图



图 1-2 是变异还是有病?

二、观察对比

早期的观察对比是比较不同总体的统计指标的差异,需要获得大量的统计资

料。1839 年,英国就有了覆盖全人口的死亡登记制度。当 1848—1849 年伦敦地区霍乱流行时,在死亡登记处工作的统计学家 Farr W 发现,居民居住的地势越高,霍乱的死亡率越低,对比结果支持“瘴气”学说(混浊空气致病)。当 1853—1854 年霍乱再次流行时,Farr 和流行病学家 Snow J 一起分析死亡资料。Snow 对比了伦敦自来水厂供水范围的霍乱死亡人数(表 1-2)。Lambeth 公司在污染轻的泰晤士河上游取水, Southwark 和 Vauxhall 公司在污染重的泰晤士河下游取水。显然,表 1-2 的统计结果支持 Snow 水污染的假说。1883 年, Koch R 在水中发现霍乱弧菌,最终证实了 Snow 的假说,否定了“瘴气”学说。值得一提的是,作为一名统计学家,Farr 进行死亡统计时还致力于医学统计的术语统一和科学分类,他对疾病分类有突出贡献,并且最终形成了国际通用的疾病分类、创伤分类和死因分类标准。

表 1-2 1853—1854 年伦敦霍乱死亡率(1/万户)

水 源	用户数	死亡人数	死亡率
重污染(Southwark 和 Vauxhall 公司)	40 046	1 263	315.4
轻污染(Lambeth 公司)	26 107	98	37.5
伦敦其他地区	256 423	1 422	55.5
合计	322 576	2 783	86.3

观察对比的方法也是病因流行病学最常用的方法。1964 年 Doll R 和 Hill A B 将接受调查的 4 万名英国注册医生分为吸烟和不吸烟两组,通过以后的肿瘤统计发现,吸烟组和不吸烟组肺癌的年平均发病率分别为 1.66% 和 0.07%,强烈提示吸烟的致癌作用,为吸烟致癌更深入的研究提供了形成研究假说的基础。

尽管观察对比的方法对 19 世纪、20 世纪初的许多生物医学的重大发现有重要贡献,但毕竟是不同总体间的比较,系统误差可能导致对比结果的偏倚,甚至谬误。Farr W 用对比结果支持的“瘴气”学说就是一例。下面还有一例观察对比常见的数字“陷阱”。根据英国某年全人口的统计资料,比较男性某病的发病率(表 1-3)。如果简单地比较总发病率,英格兰和威尔士的发病率为 132.1/10 万,移民的发病率为 65.7/10 万,结论是英国本土居民比移民更容易得病。如果将人口按年龄分组,明显看出上述结论是错误的,因为每个年龄组都是移民的发病率高于本土居民,对比组年龄构成的差异导致总发病率比较的偏倚。

表 1-3 英格兰和威尔士男性与移民男性的发病率(1/10 万)

年龄分组	英格兰和威尔士			移民		
	人口(千人)	发病数	发病率	人口(千人)	发病数	发病率
0~	1 900	1 406	74.0	26	21	80.8
5~	3 100	186	6.0	30	2	6.7
15~	9 400	1 786	19.0	127	27	21.3
45~	4 900	7 350	150.0	25	42	168.0
65~	2 000	17 400	870.0	5	48	960.0
合 计	21 300	28 128	132.1	213	140	65.7

值得提醒的是,像吸烟与肺癌这样的强关联观察对比结果,在现阶段慢性病流行病学研究中已比较少见。对于多个致病因子综合作用引起的慢性病,某个致病因子的单独作用可能很微弱(“弱关联”),组间不可比性所带来的偏倚足以改变整个研究结论。因此,对比研究资料的统计分析需要更加精细的方法和更加小心翼翼地解释。

三、干预效果评价

干预效果的评价本质上是一种实验的方法。统计上假定干预对象属于同一总体,或者是来自同一总体的样本。西方医学中有记载的干预试验发生在 1747 年。1747 年 5 月 20 日,英国的 Lind J 医生将 12 名病情相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 个橘子,一个柠檬;
- F 组:每天服由大蒜、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 组恢复得最快、最好,其中一人到第 6 天就可以工作了。B 组也有一人比登船时健康。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两点符合统计学要求。一是设有对照组,6 个干预相互比较;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。但存在的缺点也很突出:一是各组的病例太少;二是他把病情最重的 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 组也可能是病情所致,未必是干预的真实效果。

到 19 世纪,现代科学思想和科学方法在医学研究中的应用初见端倪。代表性的人物是法国医生 Louis P C A。1835 年,他对当时流行的“放血”疗法治疗肺炎的效果进行了比较,发现“放血”的疗效不像预期的那么好,而且早期“放血”组和晚期“放血”组相比较,患者的诊断、病情、病程、年龄等方面的差异很大,比较平均治愈时间意义不大。晚期“放血”组虽然平均治愈时间长,但该组患者病情重、病程长、年龄大。Louis 对医学研究的方法学作出了很大贡献:用数字的方法表示不同疾病

患者的特征和预后;提出了临床疗效对比的前瞻性原则;提出抽样误差和混杂(confounding)概念。其中前瞻性研究和减少混杂,已成为当今干预试验必须遵循的原则。后来,在著名法国数学家 Poisson S D (1781—1840)的学生 Gavarret J 的协助下,Louis 的数字方法发展为“概率框架”,Gavarret 也于 1840 年在巴黎出版了世界第一部医学统计学的教科书。



图 1-3 是治疗无效还是病情太重?

20 世纪 20 年代,英国统计学家 Fisher R A 爵士(1890—1962)在伦敦附近的 Rothamsted 农业实验站,创立了实验设计方法和统计分析技术,奠定了干预试验的统计学基础。1948 年,英国发表了评价链霉素治疗肺结核疗效的随机对照的临床试验报告,第一次采用生物统计方法进行临床干预试验。生物统计方法在我国医学界的传播与运用始于 20 世纪初。1948 年,郭祖超教授(1912—1999)编著出版的《医学与生物统计方法》,是我国第一部医学统计方法的教科书。

干预试验成败的关键是研究设计。研究设计包括专业设计和统计设计两个方面。专业设计是针对专业问题进行的研究设计,如选题、形成假说、干预措施、实验对象、实验方法等,统计设计是针对统计数据收集进行的设计,如样本来源、样本数量、干预措施分配、统计设计类型、测量指标的选择等。统计设计是统计分析的基础,任何设计上的缺陷,都不能寄希望于在统计分析阶段弥补和纠正。

四、医学论文中的统计

对临床医生和医学科学工作者来说,医学论文起到两个作用。一是撰写论文,报告自己观察或实验的研究结果,二是阅读论文,吸收新知识,了解学术进展。撰写论文,对自己研究结果的数据资料要能够正确地表达和分析;阅读论文,要能够理解和评价别人的研究成果的可信程度,这两个方面都需要一定的统计学知识。1996 年,我们对申报科技成果的 4 586 篇科研论文的统计结果,统计方法使用率为 76%。医学论文中统计运用错误,除了影响论文的科学性,还可能导致严重的伦理