

LINCHUANG LIUXINGBINGXUE
YU TONGJIXUE

临床流行病学 与统计学

主编 段广才

郑州大学出版社

高等医药院校本科教材

(供临床医学类专业用)

临床流行病学与统计学

主编 段广才

郑州大学出版社

图书在版编目(CIP)数据

临床流行病学与统计学/段广才主编. —郑州:郑州大学出版社,2002.3

ISBN 7-81048-567-9

I. 临… II. 段… III. ①流行病学—高等学校—教材②卫生学:统计学—高等学校—教材 IV. R195.1

中国版本图书馆 CIP 数据核字 (2002) 第 000051 号

郑州大学出版社出版发行

(郑州市大学路 40 号)

出版人:谷振清

全国新华书店经销

郑州文华印刷厂印制

开本:850 mm × 1 168 mm

印张:25.5

字数:752 千字

版次:2002 年 8 月第 1 版

邮政编码:450052)

发行部电话:0371-6966070

1/16

印次:2002 年 8 月第 1 次印刷

书号:ISBN 7-81048-567-9/R·508

定价:35.00 元

本书如有印装质量问题,由承印厂负责调换

编 委 名 单

主 编 段广才

副主编 李颖琰 王凯娟

编 委 (按姓氏笔画为序)

王凯娟	代 敏	代丽萍	史元峰
田庆丰	朱 伟	张卫东	李炽民
李颖琰	施学忠	段广才	

秘 书 代丽萍

主编简介

段广才,男,1958年生。郑州大学医学院副院长兼公共卫生学院院长、教授、硕士生导师、博士生导师。国务院特殊津贴专家,河南省优秀专家。1984年毕业于河南医科大学获医学学士,1989年和1993年毕业于中国预防医学科学院分别获医学硕士和医学博士,1996年晋升为教授。曾获中国预防医学科学院优秀博士生(1993)、优秀共产党员(1994),首届郑州十杰青年教师(1998),河南省优秀教师(1998)等称号。河南省跨世纪学术带头人培养对象(1996年),国家“百千万人才工程”第一、二层次人选(1997);中华流行病学会教学与方法学学组副组长,中国抗癌协会肿瘤流行病学专业委员会委员,中华医学会河南分会常务理事、副会长。主要研究方向为分子流行病学,获国家专利1项,省部级科技进步二等奖1项,发表学术论文40余篇,10余篇被“SCI”、“BA”、“CA”等国内外权威刊物收录;出版学术专著6部,译著1部,教材5部。获省级以上教材著作奖4项。

编写说明

医学是研究人类疾病和健康的学问,其目的是防治疾病、促进健康。根据目前的学科分类,医学可以分为基础医学、临床医学和预防医学,但其研究的对象都是人,只不过研究的角度和任务不同而已。流行病学和统计学是研究群体和随机事件的学问,学科分类上属于预防医学,但对于临床医学工作者来说,观察一个病人是个体,而个体多了,达到一定数量就是群体。要解决好个体的问题,不了解群体是难以想像的,就像只见树木不见森林。因此,流行病学与统计学原理和方法不仅为预防医学研究和实践所必备,近年来也被广泛地应用于临床医学等医学的各个领域。国际临床流行病学网的成立是流行病学应用于临床医学研究与实践的重要标志。

教育部 1997 年对高等院校进行学科调整时,将流行病学和卫生统计学合并为一门学科。近年来,我们对临床医学类专业的预防医学类课程也进行了相应改革,即将原编入《卫生学》的“卫生统计学”内容与原来单独开设的课程《流行病学》合并为《临床流行病学和统计学》,对其内容进行了较大幅度的调整。同时在调整内容的基础上,开设新的《卫生学》和《社会医学》等预防医学类主体课程。几年的实践表明,这些改革和调整是符合学科发展和临床医学需要的。但到目前为止,尚未有这方面的配套教材。因此,我们组织本院的骨干人员在参考国内外有关教材和著作的基础上,结合自己的教学实践编写了这套教材,供临床医学、口腔医学、影像医学、医学检验、药学等专业本科学生使用,也可以作为临床医学类专业研究生和专科生的参考教材。我们认为,本书虽然是教材,但对广大医务工作者、教师和科研人员也有一定的参考价值。

根据目前学科的发展和临床医学实践的需求,本书的编写框架和书中的许多内容都具有较大创新,在重点突出流行病学与统计学基本原理和方法学的同时,也侧重构建临床流行病学与统计学教材的结构体系,对于一些重要的概念、定理、定义等作了较规范的阐述。限于编者水平,书中可能会有不妥或错误之处,诚望广大读者提出宝贵意见。

郑州大学教务处、郑州大学出版社对本书的编写和出版给予了极大支持,老一辈专家和许多教师对本书的出版给予了关心和鼓励,在此一并致谢。

段广才

2001 年 10 月于郑州

目 录

第一章 绪论	(1)
第一节 医学研究的对象及其特征	(1)
第二节 流行病学与临床流行病学	(3)
第三节 统计学与医学	(6)
第四节 学习临床流行病学与统计学应注意的问题	(9)

医学统计学

第二章 统计学基本知识	(13)
第一节 概率	(13)
第二节 随机变量	(20)
第三节 抽样	(22)
第三章 统计描述	(25)
第一节 数值变量的统计描述	(25)
第二节 分类变量的统计描述	(34)
第三节 统计表和统计图	(43)
第四章 概率分布	(49)
第一节 离散型分布	(49)
第二节 连续型分布	(55)
第三节 χ^2 分布、t 分布和 F 分布	(59)
第四节 随机变量的数字特征	(61)
第五节 中心极限定理与抽样误差	(64)
第五章 参数估计	(67)
第一节 总体均数的估计	(67)
第二节 总体率的估计	(69)
第六章 假设检验	(71)
第一节 假设检验的基本思想	(71)
第二节 u 检验	(73)
第三节 t 检验	(75)
第四节 F 检验与 q 检验	(80)
第五节 χ^2 检验	(89)
第六节 秩和检验	(95)

第七章 线性回归与相关	(103)
第一节 直线相关	(103)
第二节 直线回归	(107)
第三节 多元线性回归概述	(112)
第八章 统计学方法的正确应用	(115)

临床流行病学

第九章 疾病分布	(121)
第一节 疾病频率测量指标	(121)
第二节 疾病分布	(124)
第三节 疾病的流行强度	(130)
第十章 病因及其推断	(131)
第一节 病因的概念	(131)
第二节 病因推断	(134)
第十一章 临床流行病学常用研究方法	(139)
第一节 概述	(139)
第二节 现况调查	(140)
第三节 病例对照研究	(144)
第四节 队列研究	(155)
第五节 实验性研究	(161)
第六节 偏倚及其控制	(169)
第十二章 疾病自然史与疾病预防策略	(174)
第一节 疾病自然史	(174)
第二节 疾病预防控制的策略和措施	(176)
第三节 疾病监测	(180)
第十三章 疾病诊断与筛检	(184)
第一节 疾病诊断	(184)
第二节 疾病筛检	(194)
第十四章 临床疗效判断	(197)
第一节 试验性研究	(197)
第二节 观察性研究	(204)
第十五章 预后与临床决策	(207)
第一节 临床预后分析与判断	(207)
第二节 临床决策分析	(210)
第十六章 医院感染	(219)
第一节 概述	(219)
第二节 医院感染的发生和影响条件	(222)
第三节 医院感染的预防控制	(231)
第四节 医院感染的流行病学调查	(232)

第十七章 药物不良反应调查及其监测	(234)
第一节 概述	(234)
第二节 药物不良反应研究方法	(235)
第三节 预防措施	(239)
第十八章 传染病预防控制	(242)
第一节 传染病的流行过程与流行特征	(242)
第二节 传染病预防与计划免疫	(251)
第三节 传染病的控制	(254)

医学科研方法

第十九章 科研选题	(261)
第一节 课题的概念和特征	(261)
第二节 选题的意义和原则	(263)
第三节 选题的程序	(266)
第二十章 科研设计的总体要求	(269)
第一节 科研设计概述	(269)
第二节 科研设计的主要内容及科研设计书的基本格式	(270)
第三节 实验设计的基本要素	(275)
第四节 实验设计的四原则	(278)
第五节 样本例数估计与几种常用实验设计类型的随机化	(287)
第二十一章 动物实验设计	(295)
第一节 动物实验设计的意义	(295)
第二节 动物实验设计的种类	(296)
第三节 动物实验结果分析及注意的问题	(298)
第二十二章 人群研究实例	(304)
第一节 现况研究实例	(304)
第二节 病例对照研究实例	(315)
第三节 临床实验实例	(322)
第二十三章 研究结果的分析、表述与应用	(326)
第一节 研究结果的统计学分析	(326)
第二节 研究结果的疾病意义及经济意义分析	(328)
第三节 研究结果的表述	(330)
第四节 研究结果的传播和应用	(332)

课堂实习

实习一 数值变量的统计描述	(337)
实习二 分类变量的统计描述	(339)
实习三 数值变量的统计推断	(340)

实习四	分类变量的统计推断	(342)
实习五	临床流行病学常用指标	(344)
实习六	病例对照研究资料分析	(347)
实习七	诊断与筛检试验的评价	(352)
实习八	医学文献评价	(354)
实习九	临床科研设计	(361)
中英文名词对照		(366)
主要参考文献		(375)
附表		(376)

第一章

绪论

第一节 医学研究的对象及其特征

医学是研究人类疾病和健康的学问,其研究对象是人,研究内容是疾病和健康状态发生发展规律、影响因素以及防治疾病、促进健康的手段,目标是通过医学和社会实践,保障和促进人类健康。根据目前的学科分类,医学可以分为基础医学、临床医学和预防医学。基础医学的主要任务是通过对机体细胞、组织系统等研究,区分正常与异常,阐明其演变机制。临床医学的主要任务是通过对病人个体的观测研究,进行准确地诊断和合理有效的治疗。预防医学的主要任务是通过对人群的研究,进行疾病的群体诊断、预防和控制。流行病学和统计学从学科分类上属于预防医学,但近年来已被广泛地应用于医学的各个领域。因为对于临床工作者来说,一个病人是个体,而个体多了,达到一定数量就是群体。要解决好个体的问题,不了解群体是难以想像的。因此,临床流行病学与统计学是临床医学研究和实践中必不可少的理论知识和方法。

一、个体特征与群体特征

人类是由个体组成的。随着人类的繁衍,遗传和变异使人类不同群体和个体的绝大多数特征个性化,像人体的身高、体重、血压等。对于这些充满个性化的机体特征,我们如何来描述和衡量其是否正常呢?比如描述某地区16岁以下儿童的身高状况,我们不能把每个人的身高实测值列一表格就算完成任务;如果要描述全国16岁以下儿童身高状况呢?再比如,有一个12岁儿童,其身高为120 cm,如何判断其身高是否正常呢?这就需要流行病学和统计学的理论知识和方法。个体特征是群体特征的基础,群体特征是个体特征的一般化概括。一个个体的特征不仅受到遗传因素的影响,也取决于机体状态和环境等因素,如一个人的血压值在早晨、中午和晚上是不同的;一个人的机体代谢状态在不同季节也是不同的。只有通过大量个体特征的研究和概括,才能获得合理的群体特征描述指标体系,反过来,再用这些群体指标评判个体特征。对于生理健康状态的特征是这样,对于疾病状

态也是一样。

由于环境致病因素与个体遗传因素和机体状态等的差异,一个人在发病时也表现为具有许多自身特征;换句话说,找到两个发病状况完全相同的病人是不大可能的。但另一方面多个病人或一类疾病的所有病人构成一个群体,他们具有一些共同的或相似的特征。如流感病人的发热、头痛、流涕等;细菌性痢疾病人的发热、腹痛、腹泻等。只有通过大量个体病人的研究才能得出一类病症的共同特点和诊疗要点,才能对个体有更深刻的理解,也才会在疾病诊断和治疗中不犯错误或少犯错误。这也是在临床上为什么病人都愿意找“老大夫”看病的原因,因为他们观察得多,发现的规律也多,所以他们诊断和治疗的正确性就高。如何正确获取临床“经验”并对其进行科学的总结和分析,使“年轻大夫”尽快变成“老大夫”呢?这些都是临床流行病学与统计学要解决的问题,也是目前循证医学(evidence-based medicine)迅速发展的重要原因。

二、人与疾病

(一)生物与人类疾病

人类作为生物群落的一部分,在长期的进化过程中,与整个环境形成了有机的统一和平衡。但在整个生物界或整个环境中,平衡是相对的,而不平衡是绝对的。比如,人与微生物的关系即是如此。首先,人类的生存离不开微生物,因为微生物可以为人类提供很多营养物质,在许多部位还是人体的卫士,如肠道的正常菌群。其次,病原微生物对人是有危害的,尤其是一些新的病原微生物的感染可以使人类的健康和生命受到很大的伤害,甚至对人类是致命性打击,如霍乱、鼠疫、艾滋病等。因此,人在受到病原微生物攻击时,免疫系统就会发挥作用,抵抗病原微生物的侵袭,逐渐形成新的平衡。再者,病原微生物是可以变异的,变异的微生物对于人体来说又是“新的”病原体;而且,除少数“绝对”病原微生物外,大多数病原微生物和非病原微生物是相对的。比如,非病原微生物在一定条件下也会成为病原微生物,如条件致病菌、抗原性变异等;而有些病原微生物也可以成为非病原性的,如致病因子基因丢失等。所以,人类任何时候都不可能逃避病原微生物的袭击,感染性疾病永远不会消失,人类与生物的共存与斗争将永远进行下去。由于生物的多样性和人群的多样性,所以感染性疾病的发生、发展在不同个体和不同群体也是多样的。

(二)非生物性物质与人类疾病

环境中除了生物性物质以外,就是非生物性物质。这些非生物性物质的静止是相对的,运动和变化是绝对的,例如近年来人类活动造成的环境污染以及天体变化等因素的影响等,因此人类任何时候都处于对环境的适应、平衡→不适应、不平衡→适应、平衡这样的循环之中。在这样的循环中,就群体来说是长期循环的,但就个体来说,并不是所有的人都能很好地完成这个循环。所以疾病甚至死亡的发生是经常的,在所难免的。同时也必须认识到,不同的环境可以导致不同的疾病,但在相同的环境中,不同个体的发病也可以是不同的。

(三)社会心理因素与人类疾病

在传统的生物医学模式中,没有认识到社会心理因素与人类疾病的关系。但现在我们已经强烈地意识到,社会心理因素在疾病发生发展中的重要作用,因此医学模式就发生了从生物医学模式→生物、心理、社会医学模式的转变。

(四) 疾病与健康

所谓疾病(disease)是指个体表现出能够被患者个人感知或他人观察、测量到的异常状态。而所谓健康(health)不仅是没有病和不虚弱,而且是身体、心理和社会三方面均处于良好状态。对于医学生来说,很容易理解疾病与健康概念都是相对的,因为一个人从健康→疾病→健康,很多情况下是没有截然界限的。比如说,你很难准确判断一个冠心病患者是从哪一时刻开始患病的,因为疾病发生和临床症状、体征的出现是渐进的,即使用最先进的技术进行亚临床前的诊断,也是无法准确判断具体的健康与疾病分界点。再比如恶性肿瘤,其发生具有多因素、多阶段、长潜隐期的特点,不同阶段还可以逆向发展,因此,再先进的手段也无法确定健康与肿瘤的准确分界点,只是随着我们对肿瘤认识的逐步深入,加上技术的进步,可以更早地发现肿瘤罢了。即便是感染性疾病,也是无法准确界定健康与疾病的。所以,一个人从健康→疾病→健康或一个群体从健康→疾病→健康都可以认为是一个连续的过程,即健康疾病连续带(health-disease continuum, HDC)。这一观点在现代医学实践中非常重要,对预防医学实践是这样,对临床医学实践也是如此。

由于人类和生物的遗传变异性与环境的不断变化,疾病和健康状态在人群中表现出很大的差别。因此,正确理解个体特征和群体特征的关系、环境与人类的关系、健康与疾病的关系是学习临床流行病学与统计学的基础。对于临床医生来说,这些贯穿于疾病诊断、治疗、预防和预后判断等各个阶段、各个实践环节。

第二节 流行病学与临床流行病学

一、流行病学

(一) 流行病学的概念

定义 1.1 流行病学(epidemiology)是研究人群中医学事件的分布及其影响因素,并研究防治疾病及促进健康的策略和措施的科学。简单地说,流行病学是研究群体中医学事件的分布及其影响因素和调控对策的科学。

“epidemiology”一词来源于希腊词 epi(在……之中)和 demo(人群),直译即为“研究在人群中发生(事情)的学问(或学科,ology)”。当然,对于医学来说,这些“事情”主要是指疾病,也可以包括健康状况、意外伤害等。流行病学是医学的一个分支,其不仅是预防医学的应用学科,更是医学各学科的方法学,目前广泛应用于医学各领域,因此也有人称流行病学是医学的基础。

上述定义中的“医学事件”是指与疾病和健康有关的事件,目前主要指疾病、伤害和健康状态。这个定义明确了流行病学的研究对象是人群;研究内容是疾病、伤害和健康状态的分布及其影响因素和调控对策;研究阶段首先是描述或揭示分布(或流行)的现象,其次是阐明造成这种分布(或流行)的原因及规律,最后是制定合理有效的策略和措施防治疾病、促进健康。“流行”一词常常用于传染病,而“分布”则可用于传染病、非传染病、伤害以及健康状态等。以上可以归纳为:一个对象,三项内容,三个阶段。

任何一门独立的学科都有自己的理论、方法和应用范畴,流行病学(一门学科)的全部内涵也可以概括为理论、方法和应用三部分;其理论主要有疾病分布论、病因论、疾病预防控制论等(三大理论),方法主要有观察描述法、观察比较法、干预实验法、数理模拟法等(四类方法),应用的领域主要有疾病的预防控制、伤害的预防控制和健康促进(三个领域)。以上可以概括为:一门学科,三大理论,四类方法,三个领域。也有学者将观察描述法和观察比较法统称为观察法。

(二) 流行病学发展简史

流行病学的发展大致可以分为三个时期,即学科形成前期或萌芽期、学科形成期、学科发展期。

1. 学科形成前期(简称前期) 指有人类文明以来到 18 世纪的一个漫长时期。有史以来人们对疾病发生发展的认识和防治实践逐步深入,这一时期流行病学的萌芽逐步产生。例如:描述疾病尤其是传染病流行的,如《内经素问》中“黄帝曰:余闻五疫之至,皆相染易”;认识到疾病与环境因素有关,如希腊医生 Hippocrates 著《空气、水和地点》对这一关系进行了阐述;采取措施预防控制疾病流行,如我国隋朝开设“病人坊”以隔离麻风病人,宋朝创立人痘预防天花等;以上这些都是流行病学的萌芽,为学科的形成奠定了基础。

2. 学科形成期(简称形成期) 大约 18 世纪末到 20 世纪 40 年代的 200 多年。这一时期西方开始工业革命,城市化发展迅速,人们大量聚集,因而各种疾病尤其是传染病的广泛袭击使人类的健康和生命受到极大威胁。因此预防控制这些疾病的流行,成为医学的迫切任务和重要使命,流行病学也就应运而生。在这一时期有许多流行病学研究和应用的典范。如 1796 年英国 Jenner 医生发明了接种牛痘预防天花;1846 年丹麦 Panum 医生对法罗群岛麻疹大流行进行调查;1848 年至 1854 年英国 Snow 医生关于伦敦霍乱流行的调查等。这些都从不同角度揭示了传染病的流行(分布)特点和流行规律,并采取有效措施进行疾病预防控制。1850 年国际上首次在伦敦成立了流行病学学会,标志着流行病学学科的形成。此后流行病学不断发展,逐渐形成了以传染病预防控制为主要任务、以独特的调查研究方法为特点的独立学科——传统流行病学(traditional epidemiology)。

3. 学科发展期(简称发展期) 主要是指 20 世纪 40 年代至今,也有学者称其为现代流行病学(modern epidemiology)时期。其主要特点是:研究内容由传染病扩展到一切疾病、伤害和健康状态,研究方法由简单的描述和分析扩展到一整套科学规范的方法即前述的四类方法,学科的分支层出不穷、应用范围空前广泛。目前流行病学已应用于医学的各个领域,其学科分支按方法学可分为:描述流行病学、分析流行病学、实验流行病学、理论流行病学、分子流行病学、血清流行病学、遗传流行病学等;按学科应用可以分为:环境流行病学、职业流行病学、临床流行病学、药物流行病学、管理流行病学、营养流行病学等;按研究的事件可分为:心血管病流行病学、糖尿病流行病学、肿瘤流行病学、传染病流行病学、伤害流行病学、健康流行病学、老年流行病学等。

二、临床流行病学

(一) 临床流行病学的产生和发展

临床流行病学这一概念最早是 Paul(1938)提出来的,他定义临床流行病学为:流行病学工作者研究人群中疾病这一现象时所用的量的概念和临床医生日常工作中诊治单个病例过程之间的结合。但在此后的几十年间,由于流行病学方法的发展和临床研究、实践的需求未能很好结合,因此临床流行病学没有大的发展。到 20 世纪 70 年代后期及 20 世纪 80 年代,临床流行病学发展迅速,专著、研究论文不断涌现,医疗实践中的应用愈加广泛。为此,世界卫生组织(World Health Organization, WHO)和洛克菲勒基金会(Rockefeller Foundation)(1982)支持建立了国际临床流行病学网(International Clinical Epidemiology Network, INCLEN),在美国、加拿大等国先后建立了五个国际流行病学资源和培训中心,负责为部分国家的著名医学院校培训临床流行病学高级专业人才。这一机构先后在 18 个国家的几十所医学院校建立临床流行病学组(Clinical Epidemiology Unit, CEU),承担本地区的人才培训和研究工作。

中国 1989 年建立了临床流行病学网(CHINACLEN),1993 年成立了中华医学会临床流行病学分会。许多院校相继开设临床流行病学课程,为提高临床诊疗和科研水平,培养高质量的医学人才做出了贡献。

(二) 临床流行病学的定义

根据目前临床流行病学的发展和应用,我们暂定义如下:

定义 1.2 临床流行病学(clinical epidemiology)是应用流行病学原理和方法,通过对临床人群(主要是患者)的研究解决临床诊断、治疗、预后和医院管理等方面问题的一门学科。

由此可以看出,临床流行病学是流行病学的一个分支,其相同点是都应用流行病学的原理和方法;其不同点是:流行病学以普通人群(包括患者和非患者)为研究对象,而临床流行病学主要是以临床患者为研究对象,当然有时也可包括部分普通人群;流行病学要解决的问题是普通人群中的疾病预防控制和健康促进,而临床流行病学要解决的是临床诊断、治疗、预后和医院管理等问题。因此,二者既存在有机的联系,也有各自的特点。有学者将临床流行病学的主要任务概括为 DME,即设计(design, D)、测量(measurement, M)和评价(evaluation, E)。

三、流行病学研究方法

(一) 观测描述法

观测描述法就是通过观察和测量来描述疾病、伤害和健康状态的人群、时间和地区分布。这一类方法主要包括:横断面研究(如抽样调查、普查等),纵向研究(如疾病监测等),生态学研究(或称相关性研究)。横断面研究和纵向研究是以个体为单位进行观测和分析,而生态学研究是以群体为单位进行观测和分析。

(二) 对比分析法

对比分析法就是通过设立研究组和对照组进行观测和对比分析,以找出它们的异同。这一类方法主要包括:病例对照研究(即结局对照研究)和队列研究(即暴露对照研究或因素对照研究)。前者又可分为病例对照研究、病例病例研究等多种;后者又可分为同时性队列研究、历史性队列研究等。

(三) 干预实验法

干预实验法是指通过对暴露因素采取干预(或实验)措施来研究干预(实验)组与对照组的差别。这一类方法主要包括:临床试验、现场试验和社区试验等。前两者是以个体为单位实施干预(或实验)措施,而社区试验是以群体为单位实施干预(或实验)措施。

(四) 数理模拟法

数理模拟法是在研究疾病分布规律及其影响因素的基础上,应用数学原理和方法对疾病的流行过程及其流行预测等进行模拟。

第三节 统计学与医学

一、统计学有关概念

(一) 医学统计学的概念

统计学(statistics)是一门较古老的学科,它是建立在概率论和数理统计基础上的应用学科,主要研究数据的搜集、整理、分析与推断。近年来由于计算机的应用和大量新的统计分析方法的出现,使统计学有了巨大的发展,其应用领域更加广泛。可以说统计方法是我们认识社会和自然现象数量特征的重要工具。正确的统计分析能够帮助我们正确认识事物客观存在的规律性。20 世纪 20 年代以后,统计学在医学领域的广泛应用形成了一个新的学科分支,即医学统计学。

定义 1.3 医学统计学(medical statistics)是运用概率论和数理统计的原理、方法,结合医学实际,研究数字资料的搜集、整理分析和推断的一门学科。

医学的研究对象是人,主要研究内容是与健康有关的各种因素和现象。如上所述,由于生物现象的变异较大,各种影响因素又错综复杂,故需要运用统计学方法透过偶然现象来探测其规律性。在临床医学研究和医疗实践中,统计学的原理和方法得到了广泛的应用,因而产生了医学统计学。由于临床流行病学与医学统计学在原理、方法和应用上的密切联系,都是研究群体事件的,本教材将其融合为一体,简称为临床流行病学与统计学。

可以说,统计学的知识几乎无处不用,把概率论与数理统计的原理和方法应用于哪个领域或学科去揭示其内在规律,就称为该领域或学科的统计学,如社会统计学、人口统计学、管理统计学、教育心理统计学等。与医学密切联系的有卫生统计学(health statistics)和医学统计学,二者并没有截然的界限,只是前者侧重于公共卫生方面,后者侧重于临床方面。

(二) 统计学的几个基本概念

1. 同质与变异 宇宙中的事物千差万别,即使是性质相同的事物的同一观察指标,各个观察单位(也称个体)之间,也各有差异,这种差异称为变异(variation)。对于人类来说,个体之间的差异更是明显的。如儿童的身高、体重对不同个体都是不同的。即便是对于同年龄、同性别的儿童,其身高也是不同的,这种不同称为身高变异。这里同年龄、同性别统计学上称为“同质”,这些同年龄、同性别的儿童称为“同质”观察单位或“同质”个体。由此可见,所谓同质是一个相对概念,是根据研究和实际工作需要确定的。

2. 总体与样本 总体(population)是指性质相同的研究对象中所有观察单位某种变量值的集合,有时也简称性质相同的研究对象所有观察单位的集合为总体。如研究某地 2001 年正常成人的血压,则研究对象是该地 2001 年的正常成人,观察单位是每个人,变量值为血压值。该地 2001 年全部正常成人的血压值就构成了一个总体。这个总体的同质基础是同一地区、同一年份、同为正常成人。由于这个总体的观察单位有限,也称为有限总体(finite population)。有时总体是设想的,如研究高血压患者使用某药物后的血压,它的同质基础是高血压患者、同用某药治疗。这里的总体将包括设想用该药治疗的所有高血压患者,其观察单位是不确定的,这样的总体称为无限总体(infinite population)。

从总体中随机抽取部分观察单位,某变量的实测值构成样本(sample)。有时也简称这些随机抽取的部分观察单位为样本,样本中观察单位数称为样本含量或样本大小,医学上也称样本例数。

3. 参数与统计量 参数(parameter)是指总体的统计指标,如总体均数 μ , 总体率 π 等;参数常用希

希腊字母表示。统计量(statistic)是指样本的统计指标,如样本均数 $\bar{X}$,样本率 P 等;统计量常用拉丁字母表示。

4. 误差 严格地说,观察值(X)与真值(μ)之差称为误差,而观察值与样本均数之差称为偏差。但在实际工作中,真值常常是未知的,而用样本均数来估计,因此偏差也称误差。统计上则将调查或实验的原始数据与真实值之差,统计量与相应参数之差通称为误差(error)。

误差分为三种:一种称为过失误差或粗差、人为误差,是由观察者的错误造成的。如点错小数点,抄错数字,计算错误等。一种是系统误差或偏倚,其有一定来源且具有方向性和规律性。如血压计不准,所测血压值均偏高或偏低。还有一种误差称为随机误差,是在排除上述两种误差之后尚存的误差,是由多种尚无法控制的因素引起的,其值无一定方向和大小,随机变化,如抽样误差。

5. 抽样与抽样误差 从总体中抽取有代表性的一部分(样本)的过程,称为抽样(sampling)。抽取有代表性的一部分观察单位和随机抽取部分观察单位的意义是相近的,但前者还包含足够的样本含量。由于总体中各个观察单位之间存在个体差异,在抽样研究中,抽取的样本只包含总体的一部分观察单位,因而样本指标不一定恰好等于相应的总体指标。例如从某市1992年7岁男童的总体中随机抽取110名(样本),调查他们的身高均数为119.95 cm,这个数值不一定恰好等于该市7岁男童的总体均数。这种样本指标与总体指标的差别称为抽样误差(sampling error)。

6. 变量与变量值 在统计学研究中,由于观察单位(个体)之间的差异,其观察指标的取值是不相同的,通常把观察单位的观察指标称为变量(variable),如身高、体重等,把观察单位的观测值称为变量值,如身高170 cm、体重60 kg等。统计学中把变量分为数值变量和分类变量,由此而获得的数据资料称为数值变量资料和分类变量资料。如身高、体重为数值变量,性别、职业为分类变量。

7. 有效数字 一个测量值中决定数量大小的数字称为有效数字,用以正确表达测量结果。如0.012 34 g中有四位有效数字1234。在1234前面的0.0是确定小数点位置的,决定于采用的单位,如0.012 34 g又可以写成12.34 mg,因此0.0不是有效数字。

有效数字与测量仪器的灵敏性有关。如天平的灵敏性为0.1 mg,那么在称重结果12.34 mg中,12.3 mg为准确数字,0.04 mg为估计数字,两项合在一起组成有效数字。有效数字只保留一位估计数字,因此末位上0的有无其涵义是不同的,如12.34 mg与12.340 mg,前者指准确到0.1 mg,误差为0.005 mg,而后者准确到0.01 mg,误差仅0.000 5 mg。

二、医学统计学主要内容

(一) 统计设计

任何研究都要先进行设计。在进行调查设计或实验设计时,除了要从专业上考虑外,还必须从医学统计学的角度考虑,使调查或实验结果能够科学地回答所研究的问题。一个好的设计可以用较少的人力、物力和时间获得更多的较可靠的资料。

(二) 统计描述

在医学研究中,常常要对群体调查或实验结果进行概括性描述,如身高的平均水平(集中趋势)、变异程度(离散趋势)等,这些都属于统计描述。

(三) 总体估计

在医学研究中,除上述进行统计描述外,更重要的是通过样本信息来估计总体中相应的统计指标,即参数估计。