

YUFANGYIXUEZONGHESHISHIYANXU

预防医学综合实验学

主编 周建伟 沈洪兵 张绮



东南大学出版社

预防医学综合实验学

主 编 周建伟 沈洪兵 张 绮

编 者 (以姓氏笔画为序)

王克波	王惠娟	刘起展	张 绮
张 雪	沈洪兵	陈宇炼	徐济达
周建伟	唐玲芳	倪春辉	莫宝庆
教淑清	高素琴	萧 黎	喻荣彬
路 滢	蔡云清		

东南大学出版社

图书在版编目(CIP)数据

预防医学综合实验学/周建伟,沈洪兵,张绮主编. —
南京:东南大学出版社,2004.11
ISBN 7-81089-420-X

I. 预... II. ①周...②沈...③张... III. 预防
医学—实验—高等学校—教材 IV. R1-33

中国版本图书馆 CIP 数据核字(2004)第 098016 号

预防医学综合实验学

出版发行	东南大学出版社
社 址	南京市四牌楼 2 号
邮 编	210096
电 话	(025)83793328
印 刷	溧阳市晨明印刷有限公司
开 本	787 mm×1092 mm 1/16
印 张	21
字 数	581 千字
版 次	2004 年 11 月第 1 版 2004 年 11 月第 1 次印刷
印 数	1—3000
定 价	32.00

* 未经本社授权,本图书内任何文字图片不得以任何方式转载、演绎,违者必究。

序

无论是自然科学还是社会科学,不同学科间知识、技能的交叉和整合是新世纪科学发展的新的显著特征。高等院校要培养适应经济社会发展需求的高素质、创新性人才,必须对传统的单科设置的理论、实验和实践教学模式进行改革。预防医学是一门理论与实践并重、宏观与微观结合、知识与技能融通的学科,强化实验和实践教学环节,对学生全面系统地掌握本课程的知识和技能具有举足轻重的影响。遗憾的是,迄今为止国内尚缺乏适应这一教学环节需求的专门教材。

我感到由衷高兴的是,由周建伟教授等主编的《预防医学综合实验学》教材恰好弥补了这一遗憾。该教材新颖独特的风格完全打破了传统的实验教材的编写模式,具有以下特点。特点之一是从实际工作出发,坚持以课题为主线,以问题为中心,打破学科间壁垒,改变传统“验证性”实验教学模式,将实验教学、理论教学与社会实践相结合。特点之二是注意保持学科发展的连续性和教学秩序的稳定,体现了科学性、先进性和实用性。特点之三是实验教材中既包含了综合实验的内容,也涵盖了当今典型的公共卫生突发事件案例的处理;既注重传授基础理论、基本知识,训练基本技能,更注重培养学生创新能力和实际运用能力。

本教材不仅适用于预防医学、基础医学和临床医学相关专业的本科生,同样适用于公共卫生硕士(MPH)、七年制临床医学生、疾病预防控制中心(CDC)和卫生监督所及基层预防医学工作者。希望本教材在教学实践中不断加以修正以日臻完善。

王心如

2004. 8. 31



前 言

为适应 21 世纪公共卫生事业的发展及卫生改革的需要,培养 21 世纪预防医学高层次人才,预防医学教学尤其是实验教学的形式和内容有必要作相应的调整 and 改革。本实验教材是为开展预防医学综合实验改革而编写的。

编写本实验教材的总体思路是:从实际工作出发,以问题或课题为主线,整合原有单科实验教学内容。教材中的综合实验分为两个层次,第一层次是在单学科内的整合,本层次的特点是维系学科课程,在重逻辑与学术的同时,力图打破知识壁垒,强化知识融合,更新教育理念,改变知识组织模式,突破单一实验的封闭和分隔特性,使学生走向丰富多彩的社会实践。如环境卫生学这一学科,传统的环境卫生学分别开设大气中二氧化硫、二氧化碳测定,交通噪声测定,水中“三氮”测定等实验,重新整合后以专题为主,如开设公共场所(如医院、车站、图书馆等)空气质量调查与评价,交通污染状况分析等专题,这些综合性的实验既涵盖了上述这些实验的基本技能,又能使学生在环境卫生学方面的知识得到融合和灵活应用。第二层次是相关学科与预防医学领域其他学科交叉的综合实验,整合预防医学专业学生所学的各专业知识以及基础医学、临床医学等,在形式上以某一问题或案例为中心综合相关的课程进行实验,将预防医学的知识高度融会贯通。如对某一不明原因疾病进行讨论,融合课题设计、现场调查采样、实验室检验、调查数据的处理和分析,调查报告(论文)的撰写等内容于一体,相当于带领同学完成一项完整的课题研究。

本教材打破了原有传统“验证性”实验教学模式,将实验教学、理论教学与实际工作相结合,并在编写中注意保持学科发展的连续性和教学秩序的稳定性,尽可能体现科学性、先进性和实用性。实验教材中既包含了综合实验的内容,也涵盖了当今典型的公共卫生突发事件案例的处理;既注重传授基础理论、基本知识,训练基本技能,更注重培养学生创新能力及实际运用能力。

本教材编写过程中始终得到校、院领导的关心、支持,雷跃老师为此书的编写做了大量协调工作,王惠娟、魏世敏、张雪、陆晓梅等同志参与了本书文稿的文字处理,在此一并表示衷心的感谢。



本实验教材可供预防医学、基础医学和临床医学本科生以及七年制临床医学学生和公共卫生硕士等使用,也是基础医学、临床医学和部分医学工作者的参考书籍。

教材中的大部分内容已经过多年的教学实践并不断修改、补充和调整,但由于时间和知识、认识水平的限制,教材中仍难免有疏误,恳请学生和老师在使用中批评指正。

南京医科大学公共卫生学院 周建伟 沈洪兵 张绮
2004. 8. 20



目 录

前言	(1)
第一部分 综合实验部分	(1)
实验一 预防医学现场调查研究设计和数据统计分析原则	(1)
实验二 公共场所环境质量调查与评价	(14)
实验三 汽车尾气对城市大气环境质量影响的调查	(23)
实验四 水源水质监测与评价	(29)
实验五 牛奶的营养与食品卫生学评价	(53)
实验六 鱼类的卫生学检验	(66)
实验七 蔬菜、水果的营养价值及卫生管理	(73)
实验八 保健食品的卫生监督管理	(80)
实验九 某乡镇企业铅作业劳动卫生基本情况调查	(88)
实验十 采石作业环境劳动卫生基本情况调查	(101)
实验十一 农药安全性毒理学评价	(121)
实验十二 外源化学物所致机体损伤及机制探讨	(148)
实验十三 儿童少年健康状况调查与评价	(164)
实验十四 学生视力及有关影响因素的调查	(183)
实验十五 儿童少年碘营养状况的调查与分析	(191)
第二部分 公共卫生事件案例与讨论	(209)
实验十六 原因不明疾病的病因研究	(209)
实验十七 1988 年上海市甲型肝炎暴发调查	(213)
实验十八 一起教堂晚宴引起胃肠疾病的暴发调查	(217)
实验十九 某省糖尿病流行特征的抽样调查	(222)
实验二十 口服避孕药与卵巢癌关系的研究	(226)
实验二十一 Logistic 回归模型及其在流行病学研究中的应用	(230)
实验二十二 化学物质急性中毒事故调查处理	(235)
实验二十三 不明原因中毒事件案例分析	(244)
实验二十四 职业卫生案例——氯乙烯与癌	(253)
实验二十五 抗疟工人中的农药中毒案例讨论	(260)
实验二十六 一宗投诉食物中毒引出的一氧化碳中毒事故	(264)
实验二十七 细菌性食物中毒调查处理	(265)
实验二十八 化学性食物中毒案例讨论	(269)
实验二十九 突发性水源污染事故应急处理	(272)



实验三十	一起因自来水污染引起的甲型肝炎暴发流行·····	(275)
实验三十一	水污染对农作物危害案例·····	(277)
实验三十二	环境砷污染对居民健康影响的调查研究·····	(279)
附录 ·····		(282)
附录一	中华人民共和国传染病防治法·····	(282)
附录二	中华人民共和国职业病防治法·····	(292)
附录三	中华人民共和国食品卫生法·····	(301)
附录四	突发公共卫生事件应急条例·····	(307)
附录五	实验方法索引·····	(313)
附录六	英汉常用预防医学词汇选·····	(316)
主要参考文献·····		(328)



第一部分 综合实验部分

实验一 预防医学现场调查研究设计和数据统计分析原则

一、现场调查研究设计与数据统计分析概述

(一) 调查研究设计思路与数据管理分析步骤的关系

研究设计是研究工作的起始步骤,也是最重要的环节。首先提出研究设想,确定要回答或解决的问题,明确研究目的;根据研究目的,确定相应的分析指标,通过调查和实验,收集和研究数据。而调查研究的步骤则和设计思路相反。在严格设计的基础上,首先开展调查和研究工作,收集相关的数据资料,并对数据进行整理和统计分析,结合归纳、推理,最终验证所提出的假设。

数据整理和统计分析贯穿于研究设计和实施的整个过程中。在研究设计时应该明确所要收集的数据类型、测量的方法、统计分析方法和分析的指标。实际上,研究实施过程也可以被认为是收集、整理、统计分析数据并解释分析结果的过程(图 1-1)。

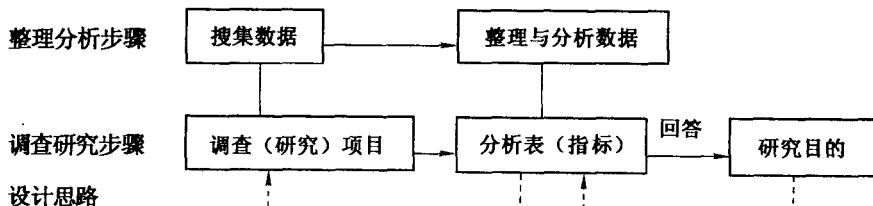


图 1-1 调查研究设计思路与数据管理分析步骤的关系

(二) 调查研究计划的制定和数据管理分析

1. 明确调查和研究的目的 各项调查和研究的目的可能不同,但从统计学角度来说,无非是:

(1) 了解参数,说明总体。如通过抽样调查了解某地学龄儿童的身高和体重水平。

(2) 研究变量间的关系,通过确立统计学联系来验证因果联系。如探索病因或评价疗效等。

这些都需要通过具体的指标来说明。确定调查和研究目的是选定研究指标的依据,而研究指标又是研究目的的具体体现。

2. 确定观察对象和观察单位 根据调查和研究的目的和指标,确定观察对象和观察单位,即划清调查和研究总体的同质范围。如评价某药物对高脂血症的治疗效果,研究总体则包含所有患高脂血症的个体,研究对象则应该是该总体的一个代表性样本。

3. 调查和研究设计类型的选择 研究设计类型的选择主要取决于研究目的和客观条件的限制。如评价某药物的疗效,可采用临床试验研究设计;探索某罕见病的危险因素,可



采用病例对照研究;为了解某病预后的影响因素,则可采用队列研究设计类型等。

(1)调查和研究设计的常用类型:

- 1)描述性研究:包括现况调查(普查和抽样调查)、生态学研究、个案调查等。
- 2)分析性研究:包括病例对照研究、队列研究。
- 3)实验性研究:包括临床试验、现场实验和社区干预等。

(2)常用的实验设计类型:

- 1)完全随机设计(completely random design):又称单因素设计。
- 2)配对设计(paired design)。
- 3)配伍组设计:也称随机区组设计(randomized block design)。
- 4)析因试验设计(factorial design)。
- 5)拉丁方设计(Latin square design)。
- 6)正交设计(orthogonal design)。
- 7)序贯设计(sequential design)。

4. 确定调查研究项目和调查表

(1)调查和研究的项目:包括分析项目、备选项目。

(2)调查表和原始记录表的设计。

5. 样本含量的估计 样本含量的估计是医学研究设计的一个重要内容。样本含量的大小至少满足“统计学效率”。样本含量大小主要取决于研究单位的变异大小、两组或多组可能差异的大小、精确性的要求、第一类错误和第二类错误的设定。不同研究设计可用各自样本含量计算公式来估计,也可采用专门的软件来估算,如 Epi Info、EpiCalc 等。

(1)单纯随机抽样的样本量估计公式

1)均数的抽样: $n = \left(\frac{u_{\alpha} \sigma}{\delta} \right)^2$, $\delta = \bar{X} - \mu$, 对有限总体来说,校正公式为: $n_c = \frac{n}{1 + n/N}$

2)率的抽样: $n = \frac{u_{\alpha}^2 \pi(1-\pi)}{\delta^2}$, $\delta = p - \pi$

(2)实验设计的样本量估计公式

1)样本均数与总体均数的比较: $n = \left[\frac{(u_{\alpha} + u_{\beta})s}{\delta} \right]^2$

2)两样本均数比较: $n = 2 \left[\frac{(u_{\alpha} + u_{\beta})s}{\delta} \right]^2$

3)两样本率比较: $n = \frac{(u_{\alpha} + u_{\beta})^2 2p(1-p)}{(p_1 - p_2)^2}$

4)配对分类资料: $n = \frac{[u_{\alpha} \sqrt{2p} + u_{\beta} \sqrt{2(p_1 - p)(p_2 - p)/p}]^2}{(p_1 - p_2)^2}$

6. 收集原始资料的方式和数据的类型

(1)原始资料的收集方式

- 1)直接观察法:包括体检、标本检测等。
- 2)采访法:包括访问、调查会及信访、电话访问等。

(2)医学研究的数据类型:医学研究数据可分为三种类型:定量数据、等级数据和名义



数据。

1) 定量数据:用定量的方法测量每个观察单位的某项(或几项)指标,所得的数据资料称为定量数据,也称连续性数据或计量资料。如身高、血压、体重指数等。

2) 等级数据:将观察单位按某种属性的不同程度分组,所得的各组观察单位数为等级资料,又称有序分类数据。这类数据一般无单位,但组与组之间有大小之分,且组内不分大小。例如临床疗效痊愈、显效、好转和无效。

3) 名义数据:各类数据之间没有顺序或等级关系。如白细胞分类的中性粒细胞、淋巴细胞、嗜酸性粒细胞、嗜碱性粒细胞等。

等级数据和名义数据也被称为定性数据或计数资料。有些等级数据或名义数据可以分成两类。如生、死,男、女,阳性、阴性,有效、无效,暴露、不暴露,患病、未患病等,属两分类数据。等级数据和名义数据也可以是多分类的。不同类型的数据,在选择研究方法时有所不同。实际工作中,根据统计分析的需要,对这三类数据可进行适当的变换。

7. 调查和研究的质量控制 质量控制是决定医学研究结果真实性和可靠性的关键。调查研究的数据只有准确地代表客观现实,通过归纳、比较、推理所获的结果才具有科学性,否则就会产生系统误差(systematic error),即偏倚(bias)。偏倚包括选择性偏倚、信息偏倚和混杂偏倚三类。医学研究质量控制即是控制这三类偏倚对结果科学性的影响。

8. 数据资料的管理

(1) 录入前的核对:在调查研究开始时,应采取措施保证原始数据的准确性。通常通过制订规范的质量控制系统来避免或减少调查研究中的信息偏倚。录入前的核对包括检查调查表中有无漏项、填写错误,及时纠正等内容。

(2) 录入计算机,建立数据库(database):可以通过统计分析软件或数据库管理软件录入。常用的软件有 Epi Info、EpiData、SPSS、FoxBase、FoxPro、Visual FoxPro(VFP)、Access、Excel 和 Lotus 等。录入软件的选择,取决于数据的大小(包括记录数、变量数)和录入的效率等。记录数和变量数较大的数据文件最好采用专门的数据录入软件进行。推荐应用 EpiData 或 Epi Info 软件录入数据。

(3) 录入后处理

1) 逻辑核对:在数据库软件中查看极大或极小值,再重新核对这些异常值,以决定取舍或修改。

2) 数据的编码和转换:有时需要根据连续性资料的值对个体进行分类,如根据血压的值判定是否为高血压患者,或根据既往有无糖尿病病史或 OGTT 血糖值综合判定其是否为糖尿病患者或分型,则需要重新编码。

3) 建立新变量:数据编码和转换的结果赋值给新变量。如新建立“DM”变量,“1”表示糖尿病患者,“0”表示非糖尿病患者;又如建立“BMI”变量表示体重指数,根据体重“weight”和身高“height”两个变量对其赋值。

(三) 统计分析

预防医学研究数据的统计分析,可以借助于各种统计软件在微机上进行。常用的统计分析软件有 SAS、SPSS、STAT、Epi Info、EpiCalc 等,其中 SAS 和 SPSS 适用于数据库数据的统计分析,EpiCalc 适用于表格数据的统计分析,Epi Info 则两者均可。

1. 统计描述



(1)定量数据的描述 均数(几何均数) $\pm$ 标准差($\bar{X} \pm s$)、标准误、百分位数、变异系数、极差以及分布的描述(正态性检验:偏度系数、峰度系数)等。

1)集中和离散趋势

$$\text{均数: } \bar{X} = \frac{\sum X}{n} \text{ 或 } \bar{X} = \frac{\sum fX}{\sum f}$$

$$\text{标准差: } s = \sqrt{\frac{\sum (X - \bar{X})^2}{n-1}}, s = \sqrt{\frac{\sum X^2 - (\sum X)^2/n}{n-1}} \text{ 或 } s = \sqrt{\frac{\sum fX^2 - (\sum fX)^2/n}{\sum f - 1}}$$

$$\text{变异系数: } CV = \frac{s}{\bar{X}} \times 100\%$$

$$\text{标准误: } s_{\bar{x}} = \frac{s}{\sqrt{n}}$$

2)正常值范围和总体可信区间的估计

95%正常值范围的估计: $\bar{X} \pm 1.96s$

总体 95%可信区间的估计: $\bar{X} \pm t_{0.05, n} \times s_{\bar{x}}$

(2)定性数据的统计描述:各种相对指标,包括率和比,如发病率、病死率、5 年生存率、治愈率、缓解率、相对危险度、比值比、标化死亡比等。

(3)统计描述方法小结(表 1-1、1-2)。

表 1-1 定量数据统计描述方法小结

定量数据	中位位置	正态分布	均数	
		非正态分布	对数正态分布	几何均数
			其他分布	中位数
	离散度	标准差		
		变异系数	均数相差大	
			单位不同的比较	

表 1-2 定性数据统计描述方法小结

定性数据	二分类	率	
	多分类	名义变量	构成比
		等级变量	构成比、平均得分
	联系	比值比 OR	各种设计
	对比	相对危险度 RR	队列研究设计
		患病比	横断面研究设计

2. 统计推断

(金丕焕, 2000 年)

(1)假设检验:包括定量数据分布类型的假设检验(正态性检验)、定量数据方差的假设检验(方差分析)、定量数据均数的假设检验(t 检验和方差分析)、定性数据分布情况或位置的假设检验(χ^2 检验)等。

1)定量数据



t 检验:

样本均数与总体均数比较: $t = \frac{\bar{X} - \mu_0}{s/\sqrt{n}}$,

配对设计资料均数比较: $t = \frac{\bar{d} - 0}{s_d/\sqrt{n}}$

成组设计两样本均数比较: $t = \frac{\bar{X}_1 - \bar{X}_2}{s_{\bar{X}_1 - \bar{X}_2}}$

方差分析(ANOVA):

成组设计多个样本均数的比较(one-way ANOVA);

配伍组设计多个样本均数比较(two-way ANOVA);

多个样本均数的两两比较(q 检验, Student-Newman-Keuls 法);

多个实验组和一个对照组均数间的两两比较(LSD 法和 Duncan 法)。

2) 定量数据差别的统计意义检验小结(表 1-3)。

表 1-3 定量数据差别的统计意义检验小结

配对比较	正态	配对 t 检验	
	非正态	数据转换	
		非参数成对比较	符号检验 符号等级检验(Wilcoxon 法)
两组比较	正态	成组比较 t 检验	
	非正态	数据转换	
		非参数成组比较	两样本等级秩和检验 Wilcoxon Mann and Whitney 法
配伍组比较	正态	随机区组 ANOVA	多重比较
	非正态	数据转换	
		非参数配伍组比较-M 检验 (Friedman 法)	
多组比较	正态	完全随机设计 ANOVA	多重比较
	非正态	数据转化	
		非参数多组比较-H 检验 (Kruskal and Wallis 法)	

(2) 变量之间的关系分析

(金丕焕, 2000 年)

1) 相关分析: 以直线相关为例, 用于双变量正态分布资料。



计算公式:
$$r = \frac{\sum (X - \bar{X})(Y - \bar{Y})}{\sqrt{\sum (X - \bar{X})^2 \sum (Y - \bar{Y})^2}}$$

假设检验:
$$t = \frac{r}{\sqrt{\frac{1-r^2}{n-2}}}, \nu = n-2$$

2) 回归分析: 包括直线回归、多元线性回归、COX 回归和 Logistic 回归等分析(表 1-4)。

表 1-4 回归分析小结

一般定量数据	线性回归
等级或名义数据	Logistic 回归
时间变量	COX 回归

3) 定性数据的关系分析: 列联表分析小结(单层或多层)(表 1-5)。(金丕焕, 2000 年)

表 1-5 列联表分析小结

行-名义变量 列-名义变量	一般联系-Pearson χ^2 检验(χ^2_a)
行-名义变量 列-等级变量	行平均得分差检验(χ^2_b)
行-等级变量 列-等级变量	相关分析(χ^2_R)

(金丕焕, 2000 年)

二、预防医学研究数据统计分析应遵循的原则

(一) 重视研究数据资料的处理过程

数据处理(data processing)是统计分析前数据管理中必不可少的步骤,应给予足够的重视。数据资料处理的目的是:一是保证所分析数据的正确性,获得的客观结果尽可能保持一致,控制信息偏倚;二是使原始数据经过编码、转换、重新赋值后符合统计分析的需要。

数据处理过程往往会花费研究者大量的时间,尤其在涉及较大规模的人群调查研究时。由于现今的统计分析软件大都具有较好的功能模块,一旦研究数据处理充分,统计分析过程就会大大简化。

(二) 选择合适的统计分析方法和指标

统计分析方法的选择主要取决于数据的类型,定量数据、定性数据的统计分析方法是不相同的。同时,描述和统计分析方法的选择又取决于数据的分布类型,大多数统计分析方法要求符合正态分布或近似正态分布。

1. 数据转换 选择统计分析方法时,必须遵循科学和客观的原则,只能根据研究数据的类型和分布特点选择,并要求最大限度地利用数据的“统计学信息”。不能满足正态分布的条件时,可以通过适当的数据转换(如对数转换、平方根转换等)达到要求。避免主观地选择统计分析方法和指标,以迎合自己的假设(表 1-6)。



表 1-6 常用的数据转换类型及方法

数据类型	转换方法	举例
Poisson 分布	平方根转换 $x'=\sqrt{x}$	水中细菌数、单位时间放射性计数等
二项分布	反正弦函数转换 $x'=\arcsin\sqrt{p}$	非传染病患利率、白细胞百分数等
标准差与均数呈正比关系	对数转换 $x'=\log\sqrt{x}$	发汞含量等

(金丕焕,2000 年)

2. 正态性检验 医学研究数据分析中常用的 t 检验和 ANOVA 是统计学家根据数据为正态分布且各组总体方差相同的条件下推导出来的,因而用以分析的数据应该是正态的而且样本方差间差别无统计学意义(方差齐)(表 1-7)。

表 1-7 常用正态性及方差齐性检验的方法

检验	方 法
正态性	用直方图或正态概率纸进行观察;用矩法、W 法或 D 法进行统计检验
两组方差齐性	F 检验
多组方差齐性	Bartlett 检验

(金丕焕,2000 年)

一般来说, t 检验和 ANOVA 是比较“稳健”的(robust)。当上述前提条件有所违反时对结果影响不太大。因而在一般情况下还是可以用的,不必太多顾虑。只有在与这些前提条件要求相距过远时才会有重大影响。

3. 非参数统计法 当 t 检验或 ANOVA 的前提条件不能满足而对数据的总体分布不能确定或没有适当的转换方法时,可以用一种不依赖于某一专门的总体分布因而也与参数无关的方法,称为非参数统计法。非参数统计法往往也适用于等级数据。非参数统计法与参数法在无效假设是正确时,其效率相同。当无效假设不正确时而分布为正态时其效率稍差;当分布为非正态时,其效率优于参数法(表 1-8)。

表 1-8 常用非参数统计方法小结

设计方法	参数统计方法	非参数统计方法
配对比较	配对 t 检验	符号检验*、符号等级检验(Wilcoxon 法)
两组比较	成组比较 t 检验	两样本等级秩和检验(Wilcoxon Mann and Whitney 法)、中位数检验*
配伍组比较	随机区组 ANOVA	M 检验(Friedman 法)
多组比较	完全随机设计 ANOVA	H 检验(Kruskal and Wallis 法)

* 效率较差的方法

非参数方法在配伍组设计或多组比较时也有多重比较的方法可用。具体参见相关统计学书籍。



对不同的研究设计类型来说,应选择合适的描述结果的指标。分析指标的选择主要取决于研究的目的、设计的类型和所获数据信息。统计分析的指标包括各种率(rate,如发病率、患病率、死亡率、病死率、治愈率、缓解率、有效率、保护率、 n 年生存率、累积发病率、发病密度等),比(ratio,如构成比、比值比、相对危险度、标化死亡比、灵敏度、特异度、阳性预测值、阴性预测值等),均数(mean,算术均数、几何均数),以及其他特定的指标(如遗传度、分离比、伤残调整寿命年、潜在减寿年数、正确诊断指数等)。

值得注意的是,由于绝大部分医学研究的对象只能是同质总体的一个样本,因而不仅要分析每个指标的点值(point value),而且要计算其95%的可信区间(confidence interval, CI),以估计总体的95%可能范围,用样本估计总体,由“特殊”推导出一般情况。

(三)忠实于原始数据是必须具备的科学精神

科学研究必须遵循客观现实。医学研究的本质即是通过观察、描述或模拟疾病和健康状态的人群现象,经过科学的归纳、分析和逻辑推理得出普遍性的规律。只有客观地记录、复制原始数据,才能使所获结果接近真实的情况,才能再现出客观规律。通常,医学研究的结果和客观现实之间,总会存在或多或少的不一致(即误差,error),包括随机误差和系统误差。医学研究过程中,应尽量通过科学的设计和严格的质量控制措施,控制系统误差。任何篡改研究数据的行为,不管出于什么目的,都违背科学精神。

三、不同研究设计类型的统计分析方法和指标选择

下面概括介绍两种常见的研究设计的统计分析思路、方法和指标,供参考。

(一)抽样调查

1. 分析思路和指标

(1)描述抽样方法、设计样本量、调查样本量、应答率等。

(2)计算有关统计指标:包括患病率、感染率、暴露率等频率指标,有关计量资料的均数、标准差、标准误、95%的可信区间等。

(3)描述疾病或健康状态的分布:结合频率指标,按时间、地区和人群描述疾病或健康状态的分布。可先按某个因素分层描述分布,再按多个因素描述分布,如按时间—地区、地区—人群、时间—人群等不同分层方法描述。计量资料的比较采用 t 检验或ANOVA,率的比较采用 χ^2 检验或 u 检验。

(4)分析性描述:对某些因素和疾病或健康状态之间的关系,可作初步的相关或关联分析。也可选择适当的自变量和因变量进行多因素分析。关联的统计推断采用 χ^2 检验或Logistic回归分析。关联强度的指标包括比值比(odds ratio, OR)或相对危险度(relative risk, RR)、人群归因危险度百分比(population attributable risk percentage, PARP 或 PAR%)等。

2. 结果解释 一般先说明样本的代表性、应答率等情况,然后分析调查中有无偏倚及其来源、大小、方向和调整方法,最后归纳疾病或健康状态分布规律的正确性及可能提供的病因线索。

3. 注意事项 现况研究资料分析所获结果的价值,取决于资料质量的优劣和分组、归纳方法是否恰当,尤其是偏倚的影响至关重要。在分析某因素与某疾病或健康状态之间的



关系时,要注意现况研究通常无法确定暴露和疾病之间的时间先后顺序。在和其他地区进行疾病频率比较时,必须考虑到年龄等非研究因素的影响,需将疾病频率按标准人口构成标准化后才具有可比性。

(二) 病例对照研究

1. 分析思路和指标

(1) 首先描述病例和对照的来源、代表性和人口学特征(如性别、年龄等)。

(2) 病例组和对照组的均衡性检验。比较两组研究因素之外的其他因素是否均衡可比,计量指标的比较(如两组平均年龄的比较)可采用 t 检验、ANOVA;计数指标的比较(如两组性别构成比较)可采用 χ^2 检验。

(3) 描述和比较病例组和对照组之间研究因素暴露率的差异,推断暴露和疾病之间是否存在统计学关联(χ^2 检验),计算关联强度(OR)。

$$\text{成组设计: } \chi^2 = \frac{(ad-bc)^2 t}{m_1 m_0 n_1 n_0} = \frac{(ad-bc)^2 t}{(a+b)(a+c)(b+d)(c+d)}$$

$$\text{校正公式: } \chi^2 = \frac{(|ad-bc| - \frac{t}{2})^2 t}{m_1 m_0 n_1 n_0} = \frac{(|ad-bc| - \frac{t}{2})^2 t}{(a+b)(a+c)(b+d)(c+d)}$$

$$OR = \frac{a/c}{b/d} = \frac{ad}{bc} \quad OR95\%CI = OR^{1 \pm 1.96/\sqrt{\chi^2}}$$

$$\text{配对设计: } \chi^2 = \frac{(b-c)^2}{b+c}, df=1$$

$$\text{校正公式: } \chi^2 = \frac{(|b-c| - 1)^2}{b+c}, df=1, (b+c < 40)$$

$$OR = b/c \quad OR95\%CI = OR^{1 \pm 1.96/\sqrt{\chi^2}}$$

2. 结果解释和注意事项

(1) 病例对照研究设计是医学研究中常用的一类设计方法,结合其衍生出来巢式病例对照研究(nested case-control study)、病例-病例研究(case-case study)、病例交叉设计(case crossover design)、病例系列研究(case series study)等,可广泛应用于疾病危险因素探讨、分子流行病学研究、遗传流行病学研究、疾病预后因素的研究等方面。

(2) 对病例对照研究来说,由于病例组和对照组一般不是来自同质总体,且该研究容易产生选择性偏倚、信息偏移和混杂偏倚,影响其验证假设的能力,因而下结论时要慎重。

(3) 目前,病例对照研究设计的资料通常直接采用 Logistic 回归模型进行分析,以控制和分析混杂作用和交互作用。Logistic 回归分析对数据库结局变量的设定、自变量的编码、结果的解释有一些特殊要求,请参见相关资料。

(4) 表格数据可用 Epi Info 软件或 EpiCalc 软件分析,可计算 χ^2 值、OR 值、OR95%CI,也可进行分层分析、分级分析、趋势性 χ^2 检验,非常方便、直观。

四、预防医学研究数据分析结果的解释

(一) 数据分析结果解释应遵循的原则

遵循科学的思维逻辑;了解不同研究设计的效率;遵循概率论的基本原则;树立对比的

