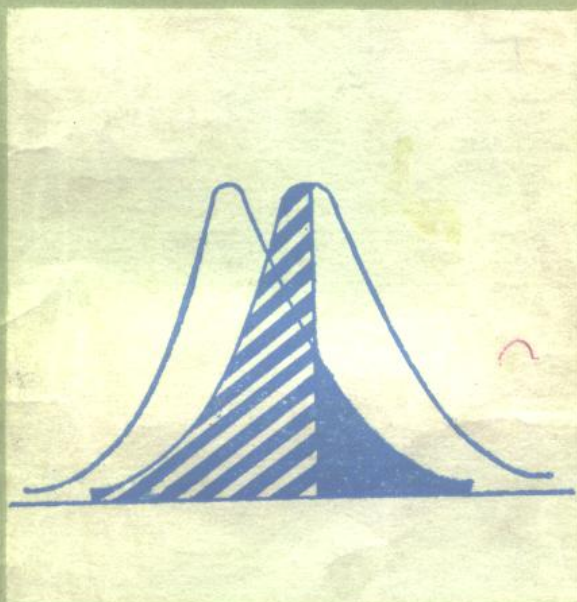


卫生统计应用丛书

临床试验设计与分析

丁道芳 凌瑞珠 编著



人民卫生出版社

81016

卫生统计应用丛书

临床试验设计与分析

丁道芳 凌瑞珠 编著

汤旦林 审

人民卫生出版社

卫生统计应用丛书编委会

主任委员 田凤调
委 员 (按姓氏笔划排列)
 丁道芳 田凤调 李天霖
 汤旦林 杨树勤 顾杏元
秘 书 金水高

临床试验设计与分析

丁道芳 凌瑞珠 编著

人民卫生出版社出版

(北京市崇文区天坛西里10号)

河北省遵化人民印刷厂印刷

新华书店北京发行所发行

787×1092毫米32开本 4 $\frac{1}{2}$ 印张 98千字

1989年11月第1版 1989年11月第1版第1次印刷

印数: 00,001—10,000

ISBN 7-117-01084-3/R·1085 定价: 1.80元

[科技新书目200—148]

卫生统计应用丛书编写说明

为了提高我国卫生统计知识水平，促进卫生统计工作的发展，更好地适应我国四化建设的需要，经过较长时间的同行酝酿，并经与人民卫生出版社协商，决定编写出版这套丛书。

本套丛书以介绍卫生统计基础知识、基本方法为主，注意实用性、科学性。既照顾到读者实际接受的可能性，又要求反映出时代的特点，介绍新的内容。

主要读者对象是：卫生统计专业工作者；医务人员和卫生防疫人员；医学界有关专业的科研与教学工作者；也可作为医学院校学生与研究生的参考书。

卫生统计应用丛书的选题包括下列几个方面：医学统计方法，居民健康统计，卫生资源统计，卫生业务统计，计算机应用技术，卫生统计工作改革等。每册一般为10~15万字，分批出版。

卫生统计应用丛书编委会

1988年12月

目 录

第一节 临床试验的基本知识	1
一、处理因素的选择	1
二、受试对象的选择	3
三、临床试验程序	5
四、临床试验的评价	8
第二节 临床试验原理	13
一、设计原理	14
二、对照	18
三、随机化	22
四、试验例数的估计	29
第三节 随机对照试验	33
一、什么是随机对照试验	33
二、统计分析方法	34
三、需注意的几个问题	40
第四节 配对试验	42
一、什么是配对试验	42
二、优点	43
三、适用条件	45
四、统计分析方法	48
第五节 配伍组试验	57
一、配伍的方法	57
二、统计分析方法	60
第六节 交叉试验	62
一、什么是交叉试验	62

二、优点及适用条件	66
三、统计分析方法	67
四、实例与评论	72
第七节 正交试验	75
一、什么是正交试验法	75
二、常用正交表的种类及其特点	75
三、正交表的选择方法	78
四、统计分析方法	79
五、实例与评论	80
第八节 拉丁方试验	89
一、什么是拉丁方试验	89
二、实例分析	90
三、统计分析方法	92
四、优点及适用条件	93
第九节 析因试验	94
一、什么是析因试验	94
二、统计分析要点	95
第十节 序贯试验	101
一、什么是序贯试验	101
二、序贯试验的种类及条件	103
第十一节 盲法试验	103
一、什么是盲法试验	103
二、双盲法实例	109
三、优点及适用条件	110
第十二节 病例对照研究	111
一、什么是病例对照研究	111
二、方法与步骤	113
三、统计分析方法	114
四、实例分析	116

五、优缺点及要求	117
附录	119
附表1 随机数字表	119
附表2 10个自然数的随机排列表	122
附表3 20个自然数的随机排列表	123
附表4 F 值表(方差齐性检验用)	124
附表5 t 界值表	127
附表6 χ^2 界值表	130
附表7(1) F界值表	132
附表7(2) F界值表	135
附表7(3) F界值表	138
附表7(4) F界值表	139

第一节 临床试验的基本知识

临床试验的目的是用比较经济的人力、物力和时间,得到较为可靠的结果。这就必须借助科学的试验设计和分析方法来达到。本节拟从临床的特点,说明临床试验的基本要素和基本程序以及必要的知识,作为试验设计和分析的基础。

一、处理因素的选择

试验一般是由处理因素、受试对象和试验效应所组成。如用某降压药治疗高血压病人,观察其血压的下降情况,这里所用的某降压药即处理因素,高血压病人即受试对象,血压下降值即试验效应。

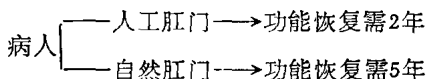
处理因素一般是指外部施加的因素(如某种药剂、某种手术方法等),然而受试者本身的某些特征,如性别、年龄等也可成为被试因素,例如研究冠心病的防治措施,性别、年龄就成为所要分析的因素。确定处理因素时,要注意以下几个问题。

1. 要抓住试验中的主要因素。任何试验都是多种因素作用的结果,而我们不可能将所有有关因素都放在一次或几次试验中,只能抓住主要的、带有关键性的一种或几个因素。

临床试验最常遇到的是治疗方法的选择问题。即医生需从几种疗法中选择一个或几个最好的方法。例如直肠癌的手术用人工肛门还是用自然肛门,就要决定用哪一种方法。

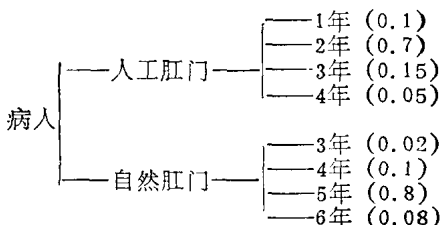
决定方式有两种,如施术后的状态可以肯定只有一种可能性,即有100%的把握,据此决定哪种疗法,称为确定性

决定。如施术后的状态有几种可能性，即非100%的把握，据此决定哪种疗法，称为非确定性决定。如果病人在采用各种疗法后，功能恢复效果如下：



而且两者的概率均为1.0 (100%)时，应选择人工肛门为佳，这种选择即确定性决定。

如果病人采用不同疗法后，功能恢复效果如下：



括号内为各年的概率，哪一种疗法最差的可能性最小，应当选择哪种疗法。人工肛门功能恢复最长时间是4年，其可能性为5%，平均时间为：

$$1 \times 0.1 + 2 \times 0.7 + 3 \times 0.15 + 4 \times 0.05 = 2.15$$

自然肛门功能恢复最长时间是6年，其可能性为8%，平均时间为：

$$3 \times 0.02 + 4 \times 0.1 + 5 \times 0.8 + 6 \times 0.08 = 4.94$$

人工肛门比自然肛门功能恢复平均时间可能快2.79年 (4.94 - 2.15)。因此，也决定选择人工肛门为合理。

2. 要确定处理因素与非处理因素。试验设计时，首先要确定什么是处理因素，什么是非处理因素。如用某降压新药治疗高血压病人，观察其疗效，该新药就是处理因素，但在治疗过程中，除了新药起作用外，其他非处理因素如饮食、

居住条件、心理状态等也必然对疗效发生一定影响。所以，在设计时必须明确划分这两类因素，不得混淆。对非处理因素应加以控制，控制的主要方法是使新药组和对照组的非处理因素均衡一致。

3. 要使处理因素标准化。处理因素标准化就是如何保证处理因素在整个试验的过程中保持不变。如果一种药物在实验的不同时间使用的批号不同；或者试验的某阶段用了超过有效期的药物；或者手术开始阶段不熟练，后期熟练；这实际上等于使用了不同的处理因素，所得的结果是“不纯”处理因素之所致。这样的结果的重现性是极差的，甚至会得出错误的结论。因此；处理因素的标准化，以及制订保证标准化的具体措施是试验设计中的十分重要的工作。

二、受试对象的选择

临床试验受试对象有时是正常人，有时是病人，对他们都应当有严格的规定和具体的要求。

1. 关于正常人的选择 研究正常值方面的问题当然需要在正常人身上进行，研究预防措施和治疗效果，有时也需要在正常人身上观察。如研究某种利尿药是否能治疗某种原因引起的水肿，当然应该在病人身上观察。但是，如果要进一步做药理、生化、生理作用机制试验，就应以正常人为受试对象。关于选择“正常人”的条件问题，不能以“完全健康”为条件，而要以接受试验的处理因素是正常的为其条件，如研究肝功能指标的正常值，只要是没有引起肝功能异常的疾病，即或是病人也可作为受试对象。

2. 病例选择 病例选择包括数量和质量两个方面。数量就是样本含量的大小，质量就是病例构成的均衡性或齐同

性。

临床试验指标多为定性的计算指标，病人个体差较大，试验条件控制也较差。这样，就不是单纯用样本含量大小公式估计的问题，还要采取一些正确的病例数设计来完成。采用以下一些办法较为实际有效。

(1) 用纯化样本的方法来保证样本的同质性。如去掉影响试验因素的病例。

(2) 用配对或配伍的方法求得组间的均衡。常用性别、年龄、病情、病程等作为配对或配伍的条件。

(3) 用分层法使比较组间保持均衡一致。即将受试者根据能影响试验效应的某些因素进行分层。再按相同比例将受试者分入各组中去。

(4) 用最省样本含量的序贯试验进行设计。它是用边试验边作统计检验办法，可以节省病例。

为保证病例的均衡性、对病例的基本要求是正确诊断、正确分期和病情的正确判断。一般是用国内外公认常用的诊断标准。我国对许多常见病多发病和职业病均已制定出诊断标准，如感冒与流感、慢性气管炎、冠心病、高血压、肺心病、阻塞性肺气肿、病毒性肝炎、矽肺、铅中毒、汞中毒、苯中毒等。如果研究者提出自己的或其他人的诊断标准，无疑在理论上和实际上都应是合理可行的。只做到有诊断标准是不够的，还应进一步排除对试验有较大影响的非处理因素，以真正达到受试对象的标准化。例如，研究慢性气管炎患者细胞免疫与机能防卫功能时，除病人按全国规定的诊断标准选择外，还必须排除长期使用某药物的病人，特别是用过免疫制剂和免疫抑制剂的病人，以及还患有肺结核、支气管扩张、肺心病等肺部疾患的病人，尽可能做到“纯化”。纯

化的程度越高，试验齐同性越好，试验效力也越强。

3. 关于志愿受试者 志愿受试者常为研究者所欢迎，因为他们能主动配合医生的安排，但是用志愿受试者会带来较大的偏性。由于他们清楚地了解医生所期望的结果，同时也是病人的期望，有时就不可避免地偏向于期望的结果。这种偏性使试验的灵敏度下降，假阳性增多，即使采用随机化方法，也难以消除。因此，应当知道志愿受试者并非最佳的受试对象。用时必须慎重对待，并应采取防止发生偏性的具体有效的措施。

4. 关于脱离观察的病人 脱离观察的病人应当合理地加以处理。对脱离观察者一般是从数据中剔除或作无效处理。然而同是脱离观察也有两种不同情况：一是伴有病情恶化的，例如死亡，改用了其他疗法，转到其他医院治疗等，而认为是试验疗法无效；另一是自然脱离观察，例如病情好转，移居他地等，试验疗法是否继续生效不能断定。所以，脱离观察的发生应根据具体情况决定处理办法。

(1) 因病情恶化脱离观察时，作无效放入非治愈组处理；

(2) 自然脱离观察：①如因好转而脱离观察，按有效的程度放入治愈组；②移居时，应对病人进行追踪，判定是否有效；③确实不能判定的少数病例应予剔除。

三、临床试验程序

(一) 临床前试验

这种试验主要是新药试用过程中有关药物疗效和安全性的鉴定试验。试验是按健康动物到患病、幼弱以至妊娠动物这样的顺序来进行的。试验的内容是药效和安全性这两项。

1. 药效：由于动物实验的实验条件不受限制并能得到充分控制，所以可以根据生物检定法或实验设计法进行实验。检验药物的疗效和估计有效量，进而可以对药物的吸收、分布、代谢和排泄作实验研究。

2. 安全性：安全性鉴定包括检验药物短期、中期和长期使用的安全性；药物作用的种类及其强度；急性、亚急性、慢性的一般毒性，特殊毒性（致畸性、依存性、遗传性、致癌性、过敏性）；中毒量、致死量的试验和调查。

由于药物的反应方法因动物不同有所差异，所以至少要用两种以上的动物，特别是要有小动物和大动物作受试对象；采用个体差异小的纯系动物；对实验环境统一管理；应重视中毒症状、致癌性、致畸性、遗传性等这些补充性的实验中这些现象出现的次数，并以较高的可信度估计相应现象出现率的可信区间。为了求得可信区间，需要增加实验例数，延长用药时间，进行长期观察。

临床前实验的任务是从动物过渡到人，对已进行过临床前实验的药物具有相似化学结构式的新药，也应进行严格的临床前实验。

（二）导入试验

1. 第一步：小儿药物不一定必须在小儿身上观察，一般是先在健康成人身上观察。试验的目的是检查药物的安全性，以安全性与剂量关系为中心，研究它的吸收、分布、代谢和排泄。旧药的用法、用量、剂型改变等试验，可省去第一步直接进入第二步。

2. 第二步：试验的目的是证明药物对特定疾病的效果和安全性。应在严格控制下进行少数病例（一般为10例）观察，孕妇应除外。以孕妇治疗为目的的试验，对由乳汁排出

药物与新生儿的调查十分重要。

用药应从小剂量开始，增加药量时要注意各种作用的出现。此时，为避免反复用药所带来的危险性和习惯性，受试者要逐渐加以更换。在尚未达到由动物实验预测出的出现某作用的有效量时，要考虑到个体变异，以及有可能出现的特殊反应——分布、代谢、结合、排泄等异常情况的发生。

(三) 正式治疗试验

正式治疗试验即第三相试验。此种试验是在大量病人身上进行观察，试验目的是对试药与对照药进行安全性和疗效的比较。

1. 临床试验的特征

(1) 个体差大：就受试对象的生物性而言，有很大的个体差。个体对试验因素的反应，由于个体间存在很大差异，即或疾病诊断、病情相同，对试验的反应也有很大差异。这种差异称为个体间变异（个体差）。同一个人血压，无论测量时多么精密，反复测出的观察值总要产生波动，这是许多因素造成的，是不能完全控制的变异（误差），这种变异称为个体内变异。治疗试验的一个显著特点是上述变异的存在。

对个体间变异应注意以下两点：①个体差对药反应大时，应把个体因素单独提出；②个体差对药效反应不大时，可以与试验操作等误差一起作为误差（残差）处理。多例数试验可随机分成两组，少例数试验可进行配对，每对又可用随机化或用交叉方法。总之，变异性大小的评价必须在相同条件下重复进行。

(2) 偏性大：就受试对象的社会性而言，有很大的心理偏性。病人之所以产生这种偏性是由于特定的环境和特殊的状态而造成的。如对自身疾病的认识，对医护人员的看法，

对得病后产生的身心、经济、家庭的变化等就形成了病人所具有的偏性。这种偏性反过来又影响疾病的未来状态。

上述这两个特征研究者必须给予充分的注意，在试验设计方法上，要有适于临床试验的特殊方法。

2. 试验方法 对照药与试验药比较，即所谓试验药物的对照试验。对照试验是指在随机分组、局部控制、分层化等条件下进行的试验。

对照药分为现有药的对照和假药对照两种。前者是新药的有效性和安全性与过去的药物作比较，是以现有药中最优药或标准药，即所谓积极安慰剂为对照。后者是使用乳糖、淀粉为成分，不加任何药物的假药，即非积极性安慰剂为对照。

试验的方式多采用双盲法，是用来避免病人和医生产生的偏性。即病人和观察疗效的医生都不知道受试者是属于试验组还是对照组。这是目前临床试验最通用的方法。

试验设计方法常用的有“不同患者的比较”（组间比较设计）和“同一患者的比较”（前后对照、交叉设计等）两种方法。此外还有序贯设计、拉丁方设计、析因设计、正交设计等。这些方法将在以后几节作详细介绍。

四、临床试验的评价

试验后对疗法和治疗效果的评价，病人、医生、护理人员常从各自的立场出发，所以评价出的结论也不尽相同。即或同是医生之间也有一定的差异。因此在评价时必须有统一的标准和科学的方法。

（一）评价标准

评价标准必须满足以下几项要求：

1. 有效性：评价有效性的标准应根据研究内容来确定。它应作到：

(1) 能把一种疾病和别的疾病区分开来：即用什么指标才能很好地表现该疾病。可以通过分析指标的体系类型，用相关、因素分析和其他多因素分析方法，将相关密切的指标合并一起进行分析。

(2) 能按病情轻重排列顺序：即指标要表现出等级、序列的问题。例如，以临床病状为尺度作评价，使之既有病状的顺序，也有各类病状的表现，可以进行双重评价。预后验证手段可使用逐步回归和判别分析的统计方法。

(3) 能区别改善与恶化的程度：上述②是有序的，而能区别改善与恶化同样也可认为是一个有效性的评价标准。

2. 重现性：无论何时、何地、何人都要得到相同评价的结果，这就是试验的重现性。几个评价者评价同一病人，医生间独立评价时的变异称为评价者间可信性。同一所见，同一人自己前次与后次评价也有时不同，称此为评价者内可信性。

3. 客观性：试验指标可分为客观指标和主观指标两种。客观指标是测量、化验等所表示的结果，借助仪器来回答。主观指标则是由受试者回答或医生判断。如检查痛觉用针刺深度表示定量化，但什么深度发生疼痛还是由受试者来回答，这也属主观指标。X光胸片上的影像是病人机体客观实像，但影像还需要医生判断，因此，也带有很大主观成份。在科研试验中无疑应尽量选用客观指标，并尽量避免用一些笼统的、不确切的主观指标。在对客观指标进行主观性判断时，如X光胸片结果的判断，常常会有主观偏性的存在，这就要求我们采取措施，如用交叉观片法、打分法等来避免偏

性产生。

临床疗效鉴定常用“临控”、“显效”、“好转”、“无效”等指标。这些都是综合判断指标。采用时必须对各项内容有明确的判断标准和具体方案，即符合哪几项标准判为“临控”，又符合哪几项标准判为“显效”，只有在明确标准的前提下，进行比较分析才是有意义的。有了明确标准，在个人掌握上也存在一定的偏性，所以最好采用分项交叉判定的方式才更为合适。

4. 精确度：它包括两层含意，即指标的准确度和精密度。前者是表示观察结果的真实程度如何，它是指观测值或均值与“真”值的接近程度，主要受系统误差所影响。后者是表示指标观察结果精度如何。试验的精密度属于偶然误差。所以，实验既要准确又要精密，首先是准确，其次才能谈到精密。但如果单纯从相对比较的意义上讲，即使用准确度较差，而精密度较好的相同的仪器测得的两组数据仍是有一定的可比性的。

5. 灵敏性：病情只有微小变化，试验指标就能敏锐测出，称此敏锐程度为好，即灵敏性高。它是增强试验效应的一个重要方面。在试验中应选用灵敏性较高的指标，如血压、脑电、心电等。

(二) 病状类型

从病状发生和持续时间来看，大致有以下四个类型：

1. 突发性病状：它是在某种条件下，突然出现而又很快消失的病状，如麻醉后的呕吐，晕船时的呕吐，遇冷后的抽筋等均属此类。此类的疗效指标可用出现率或病态的严重程度和持续时间来评价。

2. 短期病状：此类病状均可在短期内治愈，如急性肺