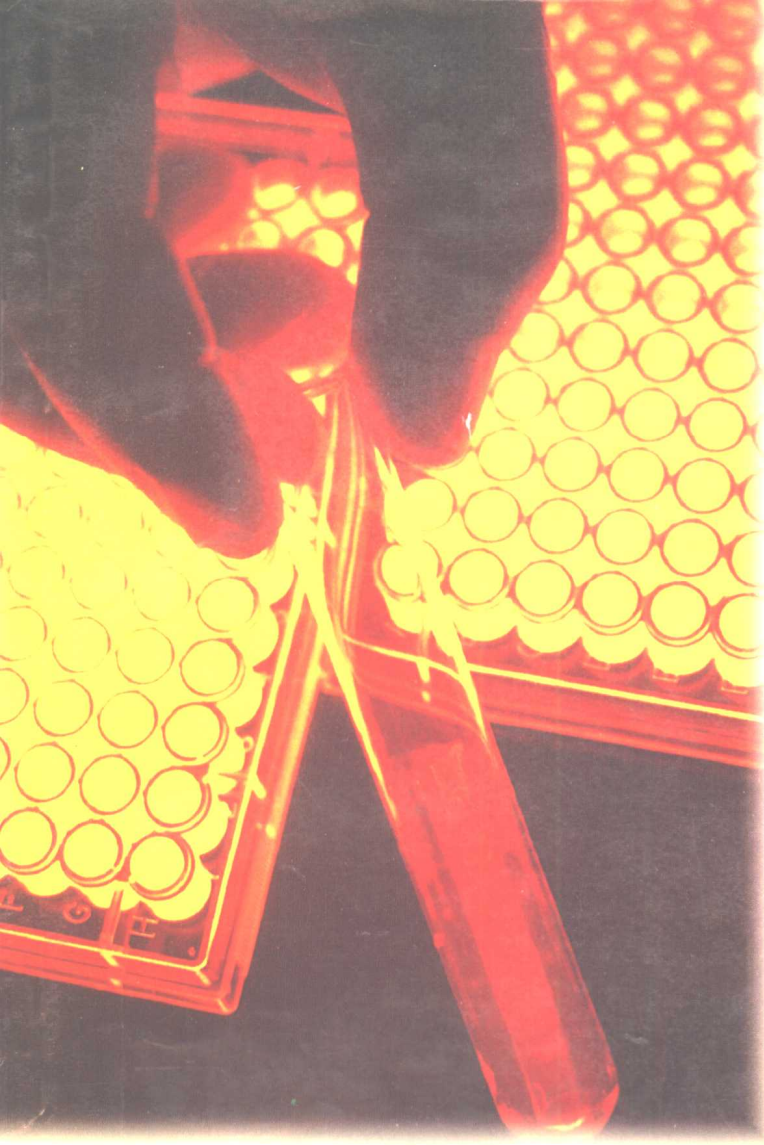


研究生
系列参考书

临床试验——设计与统计分析

LIN CHUANG SHI YAN SHE JI YU TONG JI FEN XI



金丕焕 苏炳华 编著

上海科学技术文献出版社

研究生系列参考书

临床 试 验

——设计与统计分析

金丕焕 苏炳华 编著

(此书由上海市研究生教育专项经费资助)



上海科学技术文献出版社

责任编辑：蔡 平
封面设计：何永平

研究生系列参考书

临 床 试 验

——设计与统计分析

金丕焕 苏炳华 编著

*

上海科学技术文献出版社出版发行

(上海市武康路2号 邮政编码 200031)

全国新华书店经销

上海科技文献出版社昆山联营厂印刷

*

开本 787×1092 1/16 印张 5.5 字数 137 000

1997年12月第1版 1997年12月第1次印刷

印数：1—2 000

ISBN 7-5439-1190-6/R·326

定价：15.00元

前 言

人类在与疾病的斗争中不断积累经验,使医学得到了发展;祖国的传统医学在其发展过程中保证了中华民族的繁衍生长;经验医学在人类的发展中起到了重大的作用。但是,由于人类机体的复杂性,以及个体差异,其病理生理变化受到十分复杂的外界和内在因素的影响,加之不少疾病有自愈的倾向,使得有些治疗显示出对有些病人有效,有些病人无效;有时有效,有时无效。这种不确定性导致医疗经验的积累变得十分困难。这不仅使经验积累需要很长的时间,而且也会使一些糟粕混入精华之中。因而,对于新疗法的研究,需要十分科学、严格的方法——从而也表明了这门临床试验学科的必要性。

目前,医学界对于临床试验中的一些原则,如随机、对照等概念已经逐步普及。国外许多医学院校已开设了临床试验课程。国内虽然有许多医师和研究生在进行临床试验工作,但一般并不真正了解临床试验的原理和方法。往往花费了不少人力物力却得不到正确可靠的结论。由此也使一些无效甚至有害的“医疗方法”得到推广。这是十分遗憾的。为此,笔者于1976年首次为我院的临床博士生开设了临床试验课。但当时国内尚无一本中文的参考书可供医师和研究生们参考,这就促发了笔者撰写本书的动机。

本书的编写是一种尝试。在编写中参考了国外一些书籍,结合了国内外一些临床试验的实例,以及笔者参加的一些工作。临床试验中涉及到的伦理道德问题是十分复杂的,临床试验中为了更有效地得出科学结论而采用的严格的科学方法与对个别病人的最佳治疗方案有时是会有矛盾的,这就涉及到临床试验中的道德问题。笔者在讨论中尽可能使之得到统一,但在国外的个别例子中仍不免会有不够合理的做法,这些是值得进一步讨论的。

本书得到在临床试验方面颇有造诣的苏炳华教授的合作,充实了本书的内容,感到十分欣慰。

由于本书是在国内首次系统地介绍临床试验这一课题,限于笔者的水平,在诸多方面会存在一些不足,恳切希望读者提出意见和建议。

金丕焕

1997年9月

目 录

第一章 绪论	1
概述.....	1
经验医学及其局限性.....	1
没有证据的一过性医疗方法.....	1
科学的临床试验.....	1
医师和临床试验.....	2
国际上关于药物试验的分期.....	2
我国关于药物试验的分期.....	3
一期临床试验.....	3
二期临床试验.....	4
三期临床试验.....	5
临床验证.....	5
临床试验的基本要求和步骤.....	5
文献检索与阅读	6
试验目的	6
试验设计	7
试验的实施	8
数据分析	8
结论.....	9
第二章 试验计划	10
研究方案	10
研究方案的内容	10
第三章 随机对照试验	17
概述	17
对照	17
随机化	17
历史对照	18
非随机化对照	19
随机化的必要性和合理性	20
第四章 随机化方法	21
接纳和随机安排病人的规程	21
随机安排表的编制	22
用随机排列表作随机排列区组	24

分层随机化	24
不均等随机化	25
随机排列区组的计算机程序	26
第五章 盲法与安慰剂	27
双盲试验的必要性	27
安慰剂	27
双盲法的实施	28
其他双盲试验	28
辐照食品的试验	29
双盲试验的可行性	29
第六章 临床试验的道德问题	31
赫尔辛基宣言	31
伦理委员会	32
道德问题	32
试验是否需要征得病人同意	33
随机化征求意见设计	33
第七章 交叉试验设计及其统计分析	35
病人内比较的试验设计	35
简单的两阶段交叉设计	35
交叉试验设计中的一些问题	35
交叉设计的随机化	36
交叉设计的应用	36
交叉设计试验结果的分析	36
交叉设计资料的非参数检验法	39
多于两种处理交叉设计及属性资料交叉设计的分析	40
第八章 临床试验的样本大小	41
决定试验样本大小要考虑的问题	41
定性结果样本大小的计算方法	41
单侧检验的样本大小	42
定量结果样本大小的计算	42
长期随访研究样本大小的计算	43
“阴性试验”样本大小的计算	43
小样本阴性试验的问题	44
多中心临床试验	44
临床试验中的处理数	44
析因设计	44
第九章 表格和数据管理	46
表格	46
数据管理	48

计算机分析	49
第十章 临床试验的质量控制	52
不符合标准的病人	52
符合标准的病人参加试验的比例	52
病人的依从性	52
退出治疗的病人	53
数据缺失	53
其他方面的质量控制	53
第十一章 临床试验的基本统计分析方法	54
试验计划与统计分析	54
数据类型	54
反应变量的类型	54
定性数据的统计描述与统计推论	55
定量数据的统计描述与统计推论	58
基线测定数据的比较	60
关于标准差、标准误和可信区间	60
第十二章 预后因素及生存分析	61
多元回归分析	61
多元 Logistic 回归分析	62
生存分析	64
寿命表生存曲线	64
时序检验	65
Cox 回归	67
统计检验结果的正确理解	68
第十三章 临床试验的总结报告	69
试验报告	69
试验报告的内容	69
论文的书写	70
临床试验中的假阳性与假阴性	70
附录一、随机数字表及随机排列表	72
附表 1-1 随机数字表	72
附表 1-2 20 个数字的随机排列表	73
附录二、统计用表	74
附表 2-1 t 值表	74
附表 2-2 Wilcoxon 秩和检验表(配对比较用)	75
附表 2-3 Wilcoxon 秩和检验表(成组比较用)	75
附录三、自动编制随机区组排列表的 C 语言程序	76
参考文献	78

第一章 绪 论

概述

经验医学及其局限性

随着人类历史的发展,医学也在发展。人们在生产和生活中免不了发生疾病和外伤。人类在与疾病的斗争中不断积累经验,使医学得到发展。人们积累行之有效的医疗经验,扬弃经不住考验的医疗方法。各种传统医学的发展,大致都经过了这样一个过程。祖国的传统医学在其发展过程中保证了华夏民族的繁衍生长。

但是,由于人类的机体是十分复杂的,其病理生理的变化受到十分复杂的外界和内在因素的影响。同样的治疗方法在有些病人会很有效,而对患同样疾病的另一些病人又会无能为力;不少疾病有自愈的倾向,在这种病人身上使用了毫无作用的药物病愈了,其实是自愈的效果;有些疾病随季节的变化而加重或减轻,因而某种药物会显示出有时很有效,有时则无效。凡此种使医疗经验的积累十分困难,这种情况不仅使经验积累需要很长的时间而且也会使一些糟粕混入精华之中,很难清除。

没有证据的一过性医疗方法

在医疗方法(包括新药)没有走上法制化管理的年代,曾经涌现出许多“十分有效”的治疗方法。

50年代在前苏联出现了活质学说,说苏打水洗澡可以长命,但很快就被淘汰了。

70年代早期,鸡血疗法风行一时,出现了许多宣传鸡血可以治百病的小册子,一时间到处是公鸡啼叫之声。但曾几何时鸡血疗法已烟消云散。

70年代晚期,卤碱疗法治百病的传说又风行一时,甚至有医师用这种杨白劳用以自杀的东西治疗心脏病病人,其结果当不难想象。这种经不住考验的方法不久就无声无息了。

这些昙花一现的方法也正是由于上述的疾病病理过程受外界和内在因素影响的复杂性,一种毫无作用甚至有害物质的应用在疾病的好转或痊愈上偶然的巧合,加上人们的传说和夸大造成了一时的轰动。其实,从医学的观点来看,“能治百病”本身就是不科学的。而“可以长命”又由谁来验证?

目前,传媒中有许多从未经过试验的“非常有效”的“营养品”、“药品”、“治疗仪”甚至“保健帽”、“保健鞋”,以及从几千里之外发功治疗病人的“气功”等等,如果这些都是可信的话,看来医师和医院都是多余的了。这些虚假广告至今仍充斥于市,但随着人们文化知识的日益提高,虚假广告不久都会消声匿迹,但是否又会有许多“非常有效”的“新事物”出现呢?因此,进行科学的临床试验是十分必要的。

科学的临床试验

因疾病病理生理过程的复杂性及个体差异,要决定一种新的治疗方法是否对某种病有效,或要比较哪种疗法更有效,并不是一件简单的事情,而是需要非常认真的态度,科学地进行严格的试验和分析。

临床试验(Clinical Trials)的定义:为了阐明对未来患某种疾病(或某种医学情况)的病人最合适的处理,而进行的有病人参加的试验。

动物研究显然不属于临床试验的范畴,而对健康志愿者的试验属于一期临床试验,是探讨一种新疗法的极为重要的一步。个案研究(Individual Case Study),一般不能归入临床试验。临床试验的特点是基于一定数量的病人(样本)的结果来推论在将来应当如何处理同样的病人(总体)。

这里“处理”是指对病人所采取的措施。绝大部分的处理是药物治疗。有许多情况是药厂资助试验新药。此外还有各种处理。例如在肝癌病人的治疗中,是否手术即为一种处理,手术开始前是否采取局部组织冷冻,是否采用介入疗法,是否用免疫治疗,如白细胞介素Ⅰ, LAK 细胞等,都属处理范畴。此外,各种医学建议,如心肌梗塞后病人的饮食和活动方案,以及病人的管理方案,如白内障手术后病人应当住院还是回家休养等。这些处理都需要由严格的临床试验加以评价。但后面几种非药物治疗的处理常常得不到严格的临床试验的评价。

医师和临床试验

医师在对病人进行各种日常处理中,根据学到的医学理论知识及上级医师的指导、自己的经验、其他医师的介绍以及各种文献上的报道,药房推荐的新药及其说明书对病人进行治疗。但这些治疗,尤其是传媒上大量广告所推荐的药物和治疗仪器,是不是都有效?所用的几种疗法那种方法更有效?这实际上是医师经常遇到的问题。如果能以临床试验的观点来加以考虑,就可对这些问题有一比较正确的看法,而不致轻信盲从,以不断改进处理方法,提高疗效,造福于病人。

在麻疹流行的年代,有人根据祖国医学文献的记载,应用紫草根预防麻疹,引起了一阵轰动,出现中药房紫草根售罄的情况。但紫草根预防麻疹究竟是否有效?有人对此产生了疑问。上海医科大学儿科医院的医师们所进行的严格的试验否定了它的预防作用。

我们有时候也看到医师之间对处理效果的不同意见。曾有两位医师对输卵管结扎所用的两组针麻穴位的止痛效果意见不同。但一组严格组织的临床试验结果显示,两组穴位的效果差异并无统计学意义。

国际上关于药物试验的分期

对药物的试验首先要通过动物试验,并经过有关部门的批准才能进入临床试验。本书仅涉及临床试验,有关动物试验可参考其他有关书籍。国际上比较通用的药物临床试验一般分为以下四个期:

第一期:临床药理学毒理学研究

第一期是解决药物对人的安全问题而非研究药物的疗效。实验一般在志愿者身上进行,他们常常是参加研究的人员自身。但有时,如放射治疗等,则必需是在病人身上进行。

第一期试验的主要目的是确定单一剂量。通常采用药物递增试验,即受试者按预先确定的计划接受一种递增剂量。第一期也进行药物代谢和生物利用度的研究。后期则进行多次给药(multiple doses)试验以供第二期试验参考。

第二期:疗效的初步临床研究

第二期是小规模的药物效果和安全性研究,这一期试验需要对每一个病人进行严密观察。目的是筛选出有效药物,排除无效或毒性过大的药物,为第三期试验打基础。

第三期:全面的疗效评价

在证明了药物有相当的效果之后,就需要与当前的标准治疗方法进行比较。这种比较应在条件相同的情况下进行,并需要较大数量的病例。

第三期试验是最全面的严格的新药临床科学研究。有些人把临床试验作为第三期临床试验的同义词。

第四期:销售后的监测

在研究结果经有关部门批准,新药在市场上销售后,仍然要进行不良反应的调查,以及长期的病死率和死亡率的研究。

我国关于药物试验的分期

我国于1984年9月20日第六届全国人大常委会第七次会议通过“中华人民共和国药品管理法”,并于1985年7月1日起施行。国务院又于1989年1月7日批准卫生部的“中华人民共和国药品管理法实施办法”。卫生部下达了“新药审批办法”。其中规定我国新药按药品管理要求分成以下几类(以西药为例):

第一类:我国创制的原料药品及其制剂(包括天然药物中提取的及合成的新的有效单体及其制剂);国外未批准生产,仅有文献报道的原料药品及其制剂。

第二类:国外已批准生产,但未列入一国药典的原料药品及其制剂。

第三类:西药复方制剂,中西药复方制剂。

第四类:天然药物中已知有效单体用合成半合成方法制取者;国外已批准生产,并已列入一国药典的原料药品及其制剂;改变剂型或改变给药途径的药品。

第五类:增加适应证的药品。

新药申请临床研究需取得卫生行政部门的批准,研制单位要与卫生行政部门指定的医院签订临床研究合同。新药的临床研究按照新药的分类,分为临床试验和临床验证。第一、二、三类新药进行临床试验,第四、五类新药进行临床验证。每一种新药的临床研究医院不得少于三家。

我国新药的临床试验分三期进行:

一期临床试验 在人身上进行新药研究的起始期。

二期临床试验 对照治疗试验期及扩大的对照治疗试验期,此期包括了国际通用分期的第二、三期。

三期临床试验 新药试产后的安全性考察期。

下面将介绍“中华人民共和国药品管理法实施办法”中有关这方面的内容。

一期临床试验

目的意义 研究人对新药的耐受程度,并通过研究提出新药安全有效的给药方案。

一期临床试验的设计 必须由有经验的临床药理研究人员和有经验的医师根据临床前药理研究结果进行周密的试验设计和观察,并在有必要的设备条件下进行这一期试验。

受试者 选择正常成年人(体格检查无严重的心、肝、肾、血功能障碍者)及少数适宜的病人,均以自愿为原则,男女数量最好相等,例数可在10到13例。怀孕妇女和儿童(除非儿科方面的特殊需要)不作为受试者。

必须自始至终对受试者的安全与健康负责,必须准备好应付意外的急救措施,对用药后出现的不良反应要给予有效的治疗。应给受试者必要的报酬。

初试剂量的确定 确定新药的初试剂量必须十分慎重,以保证安全为原则。应由有经验的临床药理研究人员和临床医师参考动物实验的剂量(如半数有效量、半数致死量、慢性毒性剂量,以及药代动力学参数等),共同讨论估计出一个预测剂量,然后以该预测剂量的分数量(一般不超过预测剂量的十分之一)作为人用的起始剂量,并应事先规定耐受性试验的最大剂量,一般以临床应用该类物质单次最大剂量为限。从起始量至最大量之间用几个剂量级别,需视药物的安全范围大小,根据需要而定。在达到最大剂量仍无毒副反应一般即可终止试验。如在剂量递增过程中出现了不良反应,虽未达到规定的最大剂量,亦应终止试验。耐受性试验时,每名受试者只能接受一个剂量的试验,不得对同一受试者进行剂量递增与累积耐受性试验,以确保受试者安全。对有明显药理活性的新药,起始剂量还应更小。切不可机械地按动物的剂量折算为人用剂量,因动物与人存在着明显的种属差异。

给药途径 根据新药的药理及用药目的选择给药途径(常用口服法,根据需要也可采用注射或其他给药途径)。为安全起见,无论采用何种用药途径,均须准备好相应的抢救措施。

药代动力学研究 应创造条件尽量采用灵敏的检测技术测定药物的浓度,以研究单剂给药的药代动力学。药代动力学研究对于指导临床安全有效用药具有特殊重要的意义。

结果的观察和记录 给药后必须仔细记录,尽量采用灵敏的及性能良好稳定的实验仪器,由技术熟练的实验人员操作,以便得到可靠结果。凡是在新药毒理研究中发现该新药对某些器官有较明显的毒性时,应对有关器官的功能作相应的化验检查。有条件的单位可在本单位进行。

各项实验结果均应进行统计学处理。必须有完整详细的实验记录和写出正式书面报告。

二期临床试验

此期是新药临床评价最重要的一期,应严格要求。可分两个阶段进行。

第一阶段

目的意义 在有对照组的条件下详细考察新药的疗效、适应证和不良反应。

试验设计 必须由有经验的临床药理研究人员和有经验的临床医师共同拟订周密细致的治疗试验计划。计划中应包括:病例选择标准、对照组的设置、各组检查指标、剂量与疗程以及给药方法、疗效标准和统计处理方法等。

病例选择 确定病例选择及排除病例的标准。受试病例必须有明确的临床诊断和必要的化验检查及其他检查结果。试验应选择一个条件较好的医院且在住院病人中进行,有时也可采用门诊病人,但必须能满足试验中规定的各项要求。

对照组 在本阶段临床试验中必须强调设对照组的重要性,并取得病人的密切合作。该组病人在数量上以及病情轻重程度上都应受试的新药组近似。要做到随机分组。由于病人和医师的主观因素都可能对药物的判断产生不小影响,因此,为了能有效地排除病人和医师对药物的主观偏见,尽量采用双盲法进行临床观察。这就要求对照药的外形与受试新药完全一致,使病人或进行具体观察药效的医师无法区分,只是在结束试验时才揭晓进行统计学分析。对照组有两种:已知药对照组和安慰剂对照组。已知药采用常用的有效治疗药,安慰剂一般采用乳糖或淀粉。使用安慰剂对照和双盲法试验是为了排除非药物本身的治疗效应。但用安慰剂对照必须十分慎重,以对病人不造成损害为前提,通常只用于慢性轻症患者,急、重症病人不允许采用,应选用已知有效药物为对照。

剂量与疗程 药物的剂量可根据一期临床试验结果而定,一般都采用一种固定剂量。如须多次给药,可根据其血浆半衰期来确定给药的间隔时间。用药疗程视病情而定,急性疾病用药数天至1~2周,慢性疾病用药数周至数月不等。

药代动力学研究 由有经验的临床药理医师或从事临床前药理研究的人员与临床医师密切合作进行药代动力学研究,观察体液中药物浓度与疗效及毒性关系。

结果的观察和记录 在用药期间应详细观察和记录各项指标,定期作化验及其他检查,采用灵敏的仪器并由技术熟练的实验人员进行操作。此外,还应密切注意是否出现一些未意料到的不良反应,对此应做好记录并不断追踪观察。

疗效判断 一般分四级:痊愈,显效,进步,无效。各级的判断标准应根据不同病种具体制订。

第二阶段

这是第一阶段试验的延续,目的是在较大范围内对新药进行评价。要求在第一阶段基础上除增加临床试验的病例数之外,还应扩大临床试验单位使之不少于三个。对此阶段的各项要求与第一阶段基本相似,但一般不要求双盲法。

二期临床试验所需病例数 为证明新药对常见病、多发病的疗效所需的二期临床试验的病例数,一般应不少于300例(其主要病种不少于100例)。应严格按照统计学要求,不得随意剔除受试病例。必须另设对照组,其病例数根据专业和统计学要求而定。避孕药要求不少于1000例。每例观察时间不得少于12个月经周期。特殊病种及其他特殊情况所需病例数可视情况而定。

为了获得新药安全性的初步概念,对那些可能在临床长期使用的新药,在二期临床试验中,应积累用药半年至1年的病例50~100例。

一、二期临床试验的总结

医院应将二期临床试验的所有结果进行统计学处理和评价,并作出相应的结论,与一期的结果一起,写出正式的新药临床试验总结。

三期临床试验

新药得到卫生部门批准试产之后,即应进行三期临床试验,目的是对该新药进行社会性考察与评价,重点了解长期使用后出现的不良反应,以及继续考察新药的疗效。

临床验证

目的: 考察新药的疗效和毒副反应,同原药品对照组进行对比验证,在原药品无法解决时,亦可与同疗效的药品进行对比。

验证设计 必须由有经验的临床医师拟订验证计划,计划应包括:病例选择标准、对照组设置要求、各项检查指标、剂量与疗程以及给药方法、疗效标准和统计处理方法等。

病例选择: 应选择合适的病人,有明确的临床诊断和必要的化验检查及其他检查结果。一般对象为住院病人,在某些情况下也可作为门诊病人,但必须能满足验证设计中所规定的各项要求。所需病人例数一般应不少于100例或视情况而定。必须另设对照组,其病例数根据专业及统计学要求而定。

临床试验的基本要求和步骤

临床试验涉及到病人,因而应该是十分慎重、严格的。

首先,它应当符合道德的要求,不能因此而给病人带来危害。其次,由于机体受到内外因

素的影响,变异性大,因而试验必需是严格科学的,否则会得不出结论,甚至得出错误的结论。此外,良好的组织工作也是保证试验正常顺利进行的十分重要的一环。

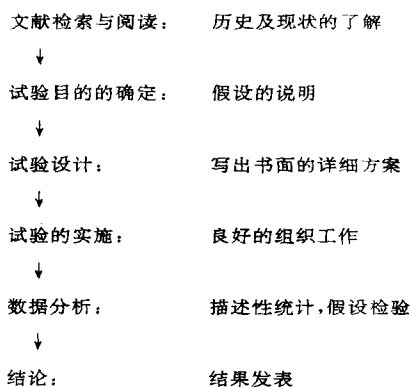


图 1-1 临床试验的基本步骤

文献检索与阅读

为了使试验做得更有效,必须了解这方面内容的历史和现状。这样才能有针对性地提出题目,充分吸取他人的经验和教训,避免重复出错误或失败。

试验目的

在任何一个临床试验中,试验的目的决定了整个试验设计,因而必需十分明确。

试验目的一般可有一个总目标,如某种药物对某种病治疗的效果。例如,研究“噻替派是否能延长乳腺癌手术后病人的生存期”。但对这一目的,应该有更加详细的定义,包括四个基本内容:①确定参加试验和不能参加试验的病人的标准;②对照组病人的设定;③治疗方法;④确定评价治疗反应的指标等。

参加试验的病人

对于参加试验的病人应有明确规定。例如乳腺癌可有各种手术方式及淋巴结转移情况。可定义为组织学上确定有淋巴结转移的乳房根治手术后的病人。同时也要确定哪些病人不应包括在内。如规定,有严重的心、肝、肾疾病患者,年龄大于70岁、妊娠或哺乳的病人等应排除在外。

对照组

对照组病人以标准的治疗方法治疗。即要有另外一组病人,同样符合参加试验的要求,同样进行乳腺癌根治术但不用噻替派治疗。然后以同样的方法随访以便比较治疗组和对照组无复发的时间长度。

治疗方法

治疗方法应有明确规定。同样的药物不同的剂量和使用方法可产生不同的效果。这里可定义为隔天注射一次,每次10mg,以120mg为一疗程。当病人表现出血液学的损害时即作剂量修改:如血小板计数小于 $100 \times 10^9/L$ 或白细胞计数小于 $4.0 \times 10^9/L$ 时则可减少一半剂量;血小板小于 $75 \times 10^9/L$ 或白细胞少于 $2.5 \times 10^9/L$,则中断噻替派用药。在三个连续疗程后若病人无毒性反应则剂量增加到每天一次。

治疗反应的指标

评价治疗的反应有各种指标。主要有定性和定量指标。如毒性反应(血液学毒性反应以

及恶心、呕吐)等。临床反应还有一种指标是从开始观察到某一情况发生的时间长度。如生存期(从手术到死亡)。上例则可用无病生存期(从手术到首次发现肿瘤复发)为主要判断标准。

至此可提出所研究的主要的假设为:噻替派(如以上所定义的)可以延长原发性乳腺癌病人(上述所定义的)的无病生存期。

若可能,则其他次要的假设,如病人的生存期、毒性以及无病生存期的延长是否局限于某一特殊病人组(如绝经前期病人)等也可进行检验。

试验设计

任何临床试验都必需有一份书面的方案(Protocol),方案中应详细说明许多设计的细节,如试验的目的、设计和执行过程等。下述内容也应在方案中加以说明。

病例数的预计

试验设计要事先计算为了得到对主要假设的结果需要多少病例。如果需要例数较多,如上例预计要几百例,而这种病又很少见,则不是在一个医院短时间内所能完成的。这时必须由多个单位合作完成,称多中心临床试验。多中心临床试验的组织工作要比单中心复杂得多,因而也更难保证质量。如果病例容易得到,如感冒等,似无必要进行多中心临床试验(参见 P. 5)。病例数的设计需要有一些关于本试验的基本数据,必要时需要作一个小规模的预试验,此内容将在第八章中进一步讨论。

病人的随机安排

对所有合乎参加试验条件的病人,要用随机化的方法分到试验组或对照组。对照组有时采用安慰剂。所谓安慰剂是为了避免心理作用影响试验结果而制作的一种外表和味道与试验药一样但无作用的“药物”。

分层随机化

在多中心研究中,为了使属于各研究单位(中心)的试验组和病人组中的例数相近,从而避免因中心情况不同而影响试验结果的对比,故需按中心分层进行随机化。即在每个中心中的病人随机化地分配到试验组和对照组,然后把各中心的病例相加起来进行分析,这时试验组和对照组分属于各中心的病例数基本相同,由此可避免因中心不同而带来的影响。

同样,病人也可按照年龄(小于或大于 50 岁),淋巴结情况(1~3 或 4 个阳性腋窝淋巴结)来分层,以避免这些因素对试验结果的影响。

双盲试验

为了避免病人知道自己服用的是试验药还是安慰剂所造成的情绪方面对试验的影响和医师知道病人吃的那一种药而引起在判断方面的影响,故对医、病家均不告知病人所服用的是哪一种药,这就叫双盲试验。不应告知的对象还应包括其他所有与病人医护有关的人员。这样,只有试验的组织者才知道那一位病人用的是哪一种药,其可将两种药分别放在标有 A 或 B 的封套内,并且告诉医师那一位病人用 A 或 B 封套内的药。

随访

所有的病人都要定期随访,并在一定时间检查血象及肝肾功能等。其他各种检查如胸部 X 线摄片等应在一定时间间隔进行。如此,一致、客观地对所有病人进行评价。

常见的缺陷是随访次数过多。随访和化验次数应当根据需要审慎决定。

试验的实施

病人的知情同意(Informed Consent)

临床试验应该征求病人的同意。因此事先必需对病人说明情况(让病人知情)。但有时实施比较困难,因而各个国家的做法也不一致。病人的同意可以是书面的,口头的,也有不必征求病人同意的。

方案的违反

在临床试验中,尤其是大规模的复杂的试验不可避免地会对原定方案有所违反。对于违反方案的各种情况的处理方法应在方案中说明。例如,在试验开始后,发现有些病人不符合参加试验的条件;有些未按照计划治疗等。这些病人应当从试验中除去。

失访

有些病人在治疗中途拒绝继续治疗;有些病人发生了与原来研究的肿瘤无关的其他肿瘤;有些因发生与原来研究的疾病无关的情况如心肌梗塞或车祸等而造成的死亡,都应作为失访进行分析。

试验委员会和中心协作办公室

对于大规模的试验,尤其多中心试验,有必要成立一个有效率的试验委员会,并有一位主席定期召集会议,对进程评估并在必要时修改计划。如在试验进行了一个阶段后,发现原来规定的仅包括受累淋巴结在4个及以下的病人数量过少,于是可将方案修改为包括有一个及一个以上淋巴结受累的病人。

在多中心研究中,成立中心协作办公室,可以在统计分析之前随时监察数据的收集和整理。由数据管理员保证及时收集所有病人数据的研究表格,及时检查、修改错误及补充缺失数据,并进行计算机处理。在统计分析前保证表格中数据的正确完整是十分重要而艰辛的工作。

数据分析

在正确收集必须的数据之后,要对数据进行正确的分析。一个好的统计分析要求符合统计学原理,按照资料的特性采用正确的统计方法,并尽可能地利用资料中的信息和使用效率最高的统计方法。

在上面提到乳腺癌的临床试验中,如果仅仅以达到某一年(如5年)时已复发的人数的比例作两组间的比较是效率不高的。因为未考虑到资料所提供的各组已观察满5年的病人无病生存时间的长短。如实验组有一位病人于观察4年5个月时复发与对照组中一位在观察1年4个月时复发的病人,虽然在观察时都是属于已复发者,但两者的无病生存期长短并不相等。应用寿命表分析法可以充分利用资料提供的信息,包括已失访者在失访前的信息,因而效率更高。

在数据分析中,用图表来表示结果可以更清晰地描述情况。图1-2显示了两组生存曲线的不同。但这种差异是由于噻替派与安慰剂的效果不同所致还仅为偶然现象?这就需要统计学检验作出评定。

统计学检验的逻辑推理如下。

1. 假设噻替派和安慰剂对无病生存期的效果实际上是相等的(称为无效假设)。
2. 用统计原理计算出,如果这一假设是真的,那么得到图1-2所显示的那么大的差异可能性(P)有多大?

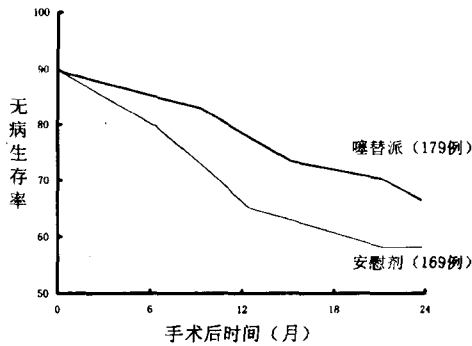


图 1-2 噻替派与安慰剂无病生存期的比较

3. 如本例的答案为 $P=0.009$, 则可预期如果这一假设是真的, 在 1000 次试验中会出现 9 次这么大或比它更大的差异。由于这一比例太小, 可以否定这一假设而认为两种药的疗效是不同的。

用统计学的术语来说, 可以说在 1% 的水平上 (即 $\alpha=0.01, P<\alpha$) 处理的差异是有高度统计意义的。

若答案为 $P<0.05$, 可以说在 5% 的水平上 (即 $\alpha=0.05, P<\alpha$) 处理的差异是有统计意义的。如答案为 $P>0.05$, 可以说在 5% 的水平上 (即 $\alpha=0.05, P>\alpha$) 处理的差异是没有统计意义的。

通过差别的统计意义检验, 可以认为噻替派延长了无病生存期。然而, 在任何差别的统计意义检验中, 永远不能绝对证明一个处理间的差异。仅仅是判断处理有差异的证据有多强, 而这也只是通过“如果没有差异, 出现这种情况的可能性 (P 值) 有多大”来推论的。这就是科学方法局限性的现实。

除了总的比较之外, 可以检验影响噻替派效果的各种预后因素。如年龄、闭经和阳性淋巴结数等情况均可能影响噻替派的效果。这可以用 Cox 回归分析法来分析 (详见第十二章)。

鉴于某些药物的毒性反应, 副作用及由此而停止治疗的数据也应进行统计分析。

结论

在数据分析之后, 应以图表方式表达所得结果, 并从这些结果中作出结论。

例如上述例子得出了如下主要结论: 噻替派对原发性乳腺癌已经有淋巴结转移的病人在根治术后使用是有效的, 而对年轻的闭经前病人效果更好。

研究结果结论当以论文在医学杂志上发表。相对于图 1-1 中列出的目的、设计、实施、分析和结论的论文有引言、方法、结果和结论等部分。所有在结论之前的内容只是集中在事实的说明, 结论则是作者基于事实的经验解释的主观意见。

第二章 试验计划

计划一个临床试验的重要步骤之一是写出一份详细的研究方案(Protocol)。它是说明研究目的及如何进行试验的细节的文件。

研究方案

研究方案的内容主要有两个方面。一方面是有有关对于每个病人的试验步骤的详细说明,如病人选择、处理、评定以及数据收集步骤等。这样可以使每个参加研究的人员明确该如何正确处理每一个病人。此部份也可称为操作部分。

方案的第二方面包括研究必要性的背景、目的、要检验的假设以及所选择的试验设计的理由等,这一部分是研究方案的科学设计部分。

研究方案的内容

研究方案应当包括(1)背景;(2)总目标;(3)具体目标;(4)选择病人的标准;(5)排除病人的标准;(6)处理的安排;(7)病人反应的评定指标及评定方法;(8)试验设计;(9)病人的登记和随机化;(10)征得病人同意参加试验;(11)研究样本大小的确定;(12)试验过程的监控;(13)表格和数据处理方法;(14)与方案发生偏离时的处理方案;(15)统计分析计划;(16)管理与经费等内容。这些仅为总的原则,各个不同的试验可以有所不同。

背景

背景是方案的第一部分,主要说明了为何要进行这项试验,以及提出过去与这项试验有关的研究成果。一份比较详细的试验方案应当对于过去及目前的研究情况有一个总结,类似文献综述,但应围绕本试验所需要的内容。这有利于强调本试验的必要性,也可避免不必要地重复别人的工作和失败。

例如,上海医科大学临床流行病学室刘天舒医师对单硝酸-5-异山梨醇酯(Isosorbide-5-Mononitrate—IS-5-MN)在食管静脉曲张出血(EVB)及再出血的发生率、发病原因和病死率以及国内引起EVB主要原因,如肝硬化及肝炎的发病率、患病率、转成慢性及由此转为肝硬化的情况;国外酒精性肝硬化的情况、由此造成的经济损失、在不同年龄和性别、城乡不同地区的情况等,同时分析了对EVB发生的各种影响因素,以及治疗再出血的方法,包括内镜、普洛奈尔等治疗的优缺点的讨论。在讨论中,研究者提出了IS-5-MN比普洛奈尔具有副反应少而有效的优点,而国外虽已有文献报道,但样本太小尚不能作出结论,从而得出有必要进行本次试验的结论。

在背景中提出的许多资料应当有文献的支持,而更重要的是提出进行研究理由的有力说明。

总目标

总目标是对本研究课题所要达到的总的说明。如在IS-5-MN的研究中,总目标为:确定IS-5-MN比同时接受标准治疗方案的食管静脉曲张出血病人是否有减少再出血的作用。

在这一总目标中所提到的减少再出血,在具体研究目的中将会有更清晰的说明。