

345332

YIXUEKEYAN SHEJI

医学科研设计与论文写作

陈修编



YU LUNWENXIEZUO

湖南医学院科研处



编写说明

1978年我院恢复了研究生培养工作，并开设了实验设计课程和查阅积累医学文献的讲座。我有机会担任这些教学任务，边教边学。根据研究生培养需要，结合自己的工作体会，把1963年在我院科研基本功训练班我担任的部分讲授及以后有关讲座的讲稿和发表过的有关论文整理成这本教材。在编写中参考了已故的陈祐鑫教授为1978年级研究生教学编的《实验设计》提要，和国内外有关文献资料。内容基本包括了科研的主要环节，从选题、查阅文献作文献综述，到科研设计，写出论文。为适应研究生教学需要，每讲之后有习题，对一般读者可供自我检查学习情况用。每讲后有主要文献资料，可供进一步钻研时查阅。

医学科研设计与论文写作是一门实用性和理论性都很强的课程。但归根到底学这门课是为了应用搞好科研。如果奢谈理论，脱离实际，难免不成为“纸上谈兵”。本教材力求理论联系实际，多用实例说明理论问题和设计方法。书中举例除个别虚拟者（为了便于讲清问题。）外，均为实例。但也不可能解决各个专业的应用问题。为了缩短学与用之间的距离，希望读者通过实例加深对基本理论和原则的理解，进而结合自己专业的实际，思考解决问题，达到学以致用要求。

医学科研设计涉及数理统计学及各专业学科的理论和应用问题很多。由于作者经验和理论水平有限，举例多为药物治疗或药理问题，希读者举一反三，联系自己的专业。本书承科研处处长游孟高教授在百忙中审阅全文，承研究生科科长钟奉贤同志大力支持，作者深表感谢。作者水平所限和编写仓促，本书会有不少缺点或错误，希读者与有关专家不吝指正。

陈 修 1982年3月

目 录

第一讲	医学科研的基本程序	1
第二讲	实验设计有关的基本概念	8
第三讲	实验设计三大原则	14
第四讲	配对设计、随机设计与随机区组设计	26
第五讲	拉丁方设计与析因设计	32
第六讲	正交设计	41
第七讲	序贯设计	47
第八讲	动物实验的几个基本问题	56
第九讲	科研论文写作	63
第十讲	怎样查阅积累医学文献资料	68
第十一讲	怎样作医学文献综述	78
附表 1	随机数目表	83
2	t 值表	
3	对两个率的差别作显著性检验时每组所需样本大小 n ($\alpha = 0.05$)	
4	对两个率的差别作显著性检验时每组所需样本大小 n ($\alpha = 0.01$)	
5	对样本均数与总体均数差别作显著性检验时所需样本大小 n	
6	对两个样本均数的差别作显著性检验时所需样本大小 n	
7	F 值表	
8	常用正交表	
9	质反应单向序贯试验边界系数表	

第一讲 医学科研的基本程序

科学研究工作的任务是探索事物发生发展的客观规律性，从而改造自然和社会，造福人类。医学科研的目的是认识生命过程和疾病发生发展的规律性，创造防治疾病提高健康水平的方法，促进人类健康长寿。这些目的任务赋予医学科研与其他科研工作的共同特征是创造性。“科学的生命在于创新”，科研的结果应该为人类已有的知识宝库增添新内容，科学才能发展。

医学科研可分应用性和理论性科研两大类。两类科研都必须具有创造性。日常医疗工作，或学生为验证理论进行的实验，或是推广别人的科研成果，虽然都是不可缺少的有意义工作，但如无创新，都不属于科学研究。如果在医疗中或推广科研成果中有新的发现或改进创新，则属于科研工作。理论性研究可以是阳性结果，也可以是阴性结果，都应该增加新的知识。对有争论的理论问题增加肯定性资料或否定性资料都有价值，但对公认有定论的理论问题如非深入一步，则不必重复他人研究。

科研工作的创造性来源于在实践中认真观察和创造性地思维。恩格斯说过：“一个民族想要站在科学的最高峰，就一刻也不能没有理论思维”。要克服思想僵化，反对懦夫懒汉无所作为的思想。既要学习外国的先进理论和技术方法；又要树立民族自尊心和自信心，而不盲目迷信外国，也不被任何专家权威的思想所束缚；要发扬中华民族固有的理论思维能力和创造精神，为实现“四化”和振兴中华进行创造性的科研工作。

科研工作的创造性应贯穿在科研工作的全过程中。医学科研的基本程序大体可分为以下五步：

一、选 题

正确选定研究课题十分重要，关系到能否创新出成果，出多大的成果问题。选题不当常导致科研失败或总结不出有价值的成果。

(一) 选题要点

1. 创造性：必须熟悉国内外对本课题的情况和薄弱的环节，以及有争论的问题，从中选出“突破口”。要有创新精神，而不是照搬或重复别人的研究。
2. 目的性：必须明确所选课题要解决什么问题，及解决的深度和广度。题目太大，目标散乱，目的性就不明确，难抓住要害。如“冠心病的研究”和“冠心病病因学的研究”都太大，“低脂饮食对冠心病发病率的影响”较适当。
3. 先进性：创造性有大有小，要有较高起点。要认识国内和世界先进水平何在，首先靠精通专业，发挥创造性思维，同时也要力争方法手段和仪器设备具有先进性。
4. 科学性：先进性和创造性不能以幻想为基础，而要有充分的理论和实践依据。例如要研究“人可以不吃饭不消耗氧生存”看来很先进，但没有科学依据。必须进行调

查研究，查阅文献资料，积累文献资料，并锻炼实践操作能力。“根深才能叶茂”，要有广博牢固的基础才能建立“金字塔”。

5. 现实性：充分考虑主客观条件，扬长避短，发挥优势，使创造性、先进性、有适当的物质基础，有人力、物力、财力的保证，不至于“空中楼阁”或半途而废。

(二) 科研选题步骤

1. 了解“四化”需要结合可能条件，具体可从国家、省市、学校的科研计划中了解，如肿瘤、心血管病、传染病与地方病的防治，计划生育、针刺麻醉、中医基本理论等等，从需要结合本单位人力设备条件确定选题方向。进一步了解这些大的计划中急待解决什么问题。如能看准苗头，当然也可个人选题。

2. 查阅文献，调查研究：了解所选课题国内外研究现状和水平，查阅主要文献，特别是综述也可向从事这方面工作有经验的专家请教，向有实践经验的同志调查了解，确定选题的目标和范围。注意了解有争议的理论问题，空白点问题，或防治效果不佳的情况“填补空白”因系别人未作过的工作，一般有创造性，其创造性大小决定于下一步——专业设计的创造性思维。

3. 形成假说，确定题目：

例如：调查得知有些冠心病病人经过劳动锻炼心电图和心绞痛发作明显改善，根据文献和调查运动员心率一般比普通人慢些，心搏量大些，有利于减少心肌耗氧，运动又能降低胆固醇和低密度脂蛋白及减轻体重等事实，可以假设适当运动量能治疗冠心病，（见卢兴：生理科学进展 12 (1):40,1981）进而决定研究“适量运动治疗冠心病的作用”的研究课题。

科研题目一般包括研究对象，处理措施和效果三个因素。上述题目可分解为：

题目： 适量运动 治疗 冠心病的作用

因素： 处 理 效 果 对 象

以上三个因素的变化组合可以扩大研究题目，见下表：

	1	2	3
处理	人参皂甙	山楂黄酮	氯压定、硝酸甘油
对象	狗	大白鼠	兔
指标	减少心肌耗氧	降低血管阻力和 血 压	防治心肌梗塞

要根据已知理论组合：盲目地“在文献缝找题目”不妥。

1. 氯压定和硝酸甘油通过扩血管降血压，降低心脏负荷，能抗心肌梗塞，山楂黄酮也降血压，扩血管，可能也有抗心肌梗塞作用。

2. 人参皂甙如能节约心肌耗氧，也可能有抗心肌梗塞作用。

二、设计

科研设计是制定实践方案实现选题目标，关系到能否按选题愿望出成果，及成果的创造性大小，可靠性如何，经济与否。

医学科研设计范围如下：

(一) 按专业分：

1. 基础医学科研设计：设计的特点要体现它的主要工作场所在实验室，研究的主要对象是动物、微生物等。

2. 临床科研设计：设计的特点体现它的主要工作场所在临床、（门诊、病房）也有实验室、研究的主要对象是病人。

3. 流行病学科研设计（调查设计）设计的特点体现它的主要工作场所是社会和现场，也有实验室，对象是人群及有关生物等。

(二) 按设计时间分：

1. 前瞻性科研设计：在研究工作开始前进行设计，按设计积累数据资料。基础医学科研主要用前瞻性设计，临床和流行病学科研的发展趋势也是多用前瞻性设计。因为这种设计目的性明确，便于控制条件和影响因素，有针对性地解决实践或理论问题。

2. 回顾性科研设计：是在事情发生后，根据设计收集资料，找出规律性。如临床的病例分析，中医的医案，病理学的尸体解剖，流行病的疾病调查、死因调查等。

前瞻性科研设计与回顾性科研设计可以密切配合，相互促进，便于找出规律性。例如1979年长沙市某小学师生发生流行病，表现为皮疹、皮疹、发热关节肿痛、及耳廓红肿疼痛等，经省防疫站进行流行病学调查表明这批师生前不久游岳麓山在松树下作游戏受松毛虫毒素侵袭有关。这是回顾性调查所得出的结论。为了判明这一结论的正确性并取得直接证据，将从现场取回的死虫作成水浸液，直接涂于小白鼠剃毛的腹部，结果病理学检查证明发生了局部皮炎和关节炎。（见陈大保等：湖南医药杂志（3）：1，1980）后一部分实验室工作是前瞻性研究。在实验之先已作出研究设计方案。这样通过对已发生的流行病作回顾性调查，再加上实验室进行前瞻性研究，确定了流行的是松毛虫病，在此基础上可以进一步作前瞻性致病因素的分离研究，从而可使问题更加深入了解。

(三) 按技术内容分：

1. 专业设计：是利用专业知识解决要通过科研解决的问题。专业设计水平的高低决定科研创造性大小和学术水平的高低。要搞好专业设计，首先要深入钻研本课题的基本内容，熟悉有关进展情况。为此，必须查阅文献，向有经验的人调查研究和学习，在掌握专业知识和课题进展的前提下，发挥创造性思维，提出工作假设。

成功的专业设计举例：

(1) 弗采明 (Fleming) 1928年发现青霉素，是他检查废弃的培养皿时偶然的机

会发现的，但不是碰运气。因为他多年来一直在研究溶菌现象，1922年他已首先发现溶菌酶 (lysozyme)，他深知葡萄球菌不易发生溶菌现象，所以一旦碰到被霉菌污染的培养皿里，青霉菌周围的葡萄球菌被溶解，引起他非常敏锐的注意。如果是一个对溶菌问题无知的人，对此现象会“视而不见”。他于是进一步研究，发现了青霉素。但因他缺乏化学知识，未能解决青霉素不耐酸硷问题，未能提纯；又因他缺乏药理和实验治疗学知识，未作抗感染实验，故未能应用。1940年Chain和Florey提纯了青霉素并发现注射青霉素的小鼠尿中仍有青霉素的抗菌作用，说明它在体内稳定，进行了动物实验治疗成功，1941年才用于临床治疗。(见Chain E; Trends in Pharmacol Sci 1(1); 6, 1979)

(2) 二十年代陈克恢研究麻黄的经验。他决定研究中药的药理时，先调查了解哪些中药在临床应用中效果非常肯定，老中医推荐他研究麻黄，因麻黄发汗平喘疗效确切。陈克恢运用专业知识，并借鉴日人 Amatsu 1917 年的研究结果认为麻黄可能有拟交感作用，因为发汗及平喘所需要的支气管扩张作用都是交感神经兴奋的表现。随后于1923年开始用动物实验证明麻黄所含麻黄素具有拟交感作用，为世界药物宝库增添了新武器。

(3) Otto Loewi 1921年证明乙酰胆硷为迷走神经递质的设计。他用刺激甲蛙心的迷走神经使心脏抑制，心率减慢，将灌流液引出灌流乙蛙心，结果乙蛙心也随之抑制。这为神经化学传导的理论，提供了直接的证据 Dale 等人虽早在1906——1914年已发现乙酰胆硷的作用与刺激副交感神经相似，而去甲肾上腺素的作用与刺激交感神经相似，但未能证实。Loewi通过创造性思维运用他深博的专业知识设计出上述简单而有说服力的实验，证明神经递质、效应的关系，创立了神经化学传递理论。这一专业设计的成功，说明有了专业上的创造性思维，即使简单的方法也可创造出有重大理论和应用价值的成果。实验方法虽很简单，但创造性思维要付出巨大的脑力劳动，发挥创新精神。在我国缺乏先进仪器设备的现实条件下，更应强调消除思想僵化的影响，发挥我国人民的创造性思维能力，创造新成果。

(4) 现代医学科研的重大成果更多依赖先进的技术方法与创造性的思维并用。例如1977年 Sokoloff 报道的用 $[^{14}C]$ 标记的 2-脱氧葡萄糖给动物静脉注射作脑组织放射自显影，可以确定微小到大白鼠一根胡子在大脑皮层上的定位。这一研究方法连同其他的新技术方法的应用，使一个古老的“垂死的”学科——神经解剖学获得新生，使它从一门僵硬的学科一跃而成现代最活跃的学科之一。创造这一技术方法的工作假设是创造性用“不用则废”的道理。已知葡萄糖是大脑代谢的主要能源，也知外周组织在大脑皮层感觉区有相应的投射部位。这一部位的细胞团在接受感觉信息时要消耗能量(葡萄糖)，设想如果在大鼠发育期拔掉一根胡子，由于废用将导致相应的大脑皮层细胞团萎缩或消失。 $[^{14}C]$ 2-脱氧葡萄糖代替葡萄糖进入脑组织。由于它可以示踪定位，从而证实了“不用则废”的道理也适用于大脑皮层。(见张香桐：生理科学进展10(1)；11, 1979)

失败的专业设计举例：

(1) 针刺治聋哑的曇花一现：在诊断上未认真鉴别神经性耳聋与一些非完全耳聋。认为聋人完全无听力，也完全不会说话。据调查聋哑学校的学生完全无听力的只占

10—20%，个别只是患了中耳炎，不适当地被送入聋哑学校。在判断疗效标准方面缺乏客观性。后来用电听力计记录听力图，不能证明针刺能恢复失去的听力。（见邓元诚等：中华医学杂志61（11）：657,1981）

（2）复脉汤（灸甘草汤）使“心肌细胞再生”的作用，实验缺乏基本理论和技术。

（3）牛奶浆草杀血吸虫：把导泻消腹水混同杀虫治愈。把治疗前多次粪检偶尔阳性诊断的血吸虫病，治疗后一次粪检阴性即作为治愈的错误判断疗效。从以上专业设计失败事例说明要作出经得起考验的成果，必须在专业上站稳脚跟，扎扎实实加强专业学习，提高专业水平，才能作好专业设计。

专业设计应在提高科学性方面下功夫，以下几点对临床专业设计尤应注意：①疾病的诊断要有可靠的根据，设计中应订出诊断标准。中医的辨证应力求有客观指标，这样才能避免同一病人不同的医生辨证结论不同。在缺乏客观指标时，有人主张采取集体评定的方法。②评定疗效的指标必须具体、客观，最好采用国际或国内统一标准。如肿瘤缩小 $1/2$ 应是垂直方向两个最长径的乘积数字减半，单位应是长度的平方，应避免含糊不清各人理解不同的判断疗效标准。③药物品种和质量应严格控制，以黄芪为例，用作研究时对其产地、品种、采集季节、剂型以至保存时间都应统一，如用提取物，对提取方法、质量鉴定等都应有规定，且同一批研究对象最好用同一批药物，才有可比性。④中药方剂如必须按辨证加减，应统一几种辨证和几种复方加减，治疗时不同医生均按协定方剂开处方，否则一人一证一方，则无法总结经验。

2. 数理统计设计：通常称为实验设计。

用数理统计方法保证科研结果的可重复性（可靠性）和经济性，使抽样误差减到最小程度，使处理的布置经济合理。作好数理统计设计，能用较少人力物力时间作出经得起重复的结果。这方面的经验教训也是很多的。例如欧美曾用金制剂治疗结核病几十年，长期作不出有无疗效的结论；而五十年代初应用异烟肼（雷米封）不久，就对它与链霉素PAS合用和单用的疗效作出了科学的结论；即这三种药不仅单用有效，合用效果则更好，30年来实践证明这一结论是正确的。又如我国对流脑菌苗的效果观察，1975年以前由于缺乏严格对照，诊断标准不统一，观察几百万人份的菌体菌苗效果波动在10—90%之间，得不出结论，浪费了大量人力物力。1976年加强了实验设计，用双盲法分组，提纯菌苗的预防效果在60—70%之间，结果较为可靠。

实验设计在现代医学研究中占有不可缺少的地位，但它不能代替专业设计。如果专业设计有错误或实验与临床观察不严谨，运用统计学方法处理也决不能提高其科学性。反之，专业设计也不能代替统计学设计，没有好的统计学设计要浪费人力物力，而且在许多情况下作结论没有把握。所以专业设计与统计学设计不是互相排斥的，而是相辅相成的。

〔附〕科研设计的主要内容

1. 题目名称

2. 承担单位、主要负责人

3. 国内外现状、水平和发展趋势

1. 研究目的和提出的根据（包括根据现有资料找出的关键问题和“突破口”，及

工作假设

5. 主要研究内容和技术关键（包括研究的项目、指标、需建立的技术方法等）
6. 实验设计（包括选用什么设计方案；如何分对照组与实验组；如何抽样；每组样本数；实验结果数据处理方法；观察测定记录表格等）
7. 计划进度（包括分几个阶段进行，每一阶段完成的任务，何时完成并作出总结）
8. 题目参加人员、协作单位及其协作任务
9. 经费预算（包括实验材料，试剂药品、动物、设备、调查研究和资料费用等）
10. 存在问题和困难

三、实 践

根据不同专业特点，根据科研设计进行实验或观察病人或作流行病学调查。这是用科学方法收集资料的重要步骤，决定将来据此作出的结论是否科学可靠。

实践中必须注意：

1. 避免主观性：实验者（如医生）和研究对象（如果是病人）都可以有先入为主的主观性，认定处理有效或无效。“带着有色眼镜”观测，甚至有意“制造”有效。如食醋熏蒸预防感冒，领导要求“只许做好，不许做坏”结果把实验组中发病者也算作对照组的，错误作出食醋熏蒸预防感冒有效的结论。现已为实践所否定。医生将病情暗示病人，也可得出与事实不符的结果和结论。

2. 加强“三基”“三严”：缺少基本功训练，即使有避免主观性的愿望，但技术方法不可靠，或工作粗心大意，结果仍然不可靠。如天平未校正或不会正确使用，以致称重错误；如未掌握听诊技术或不熟悉诊断疾病的标准，必然误诊，收集的资料也必然不可靠。

3. 认真记录，加强技术档案管理：观测结果必须把原始数据随手记录，不可事后回忆补记，更不可根据主观愿望删改记录。观测中有些意外情况，影响结果，应注明情况，以便分析整理资料时参考，有些实验或临床观察记录应事先表格化，以便整理总结。

四、资料整理

分析整理实验或流行病学调查资料是对研究结果进行条理化数量化的过程，是透过现象发现规律的重要步骤。应根据不同的资料作不同的处理。原始资料一般是数据，或图形（如从心电图、血压记录，曲线等）或切片照片，原则上凡能数量化的都应变为数字（如从血压图上测出变化的量，从心电图测出，P-R间期S-T段等数据，然后进行统计学处理。有时照片切片的结果不能数量化，可用分级计分或阴性、阳性等方法表示。要正确运用数理统计知识，不可撇开实验真象作“数学游戏”。

五、总 结

用分析综合归纳演释等多种形式，把感性材料上升为理性概念。要注意在一定条件下，对少数样本所得的结果，有一定局限性，勿任意外展扩大。如阿司匹林能降低发热动物和病人体温的观察，不可一般化为“能降低体温”，因它对正常人体温不能降低。

总结的结果，如有交流的价值，应写成论文发表。如无交流价值，也应总结资料和经验教训存档备查。论文写作要求和注意事项见后。

参 考 文 献

1. 医学科研座谈会记要，中华医学杂志61（6）：321，1981
2. 陈修：医学科研设计的几个基本问题，湖南医学院学报6（4）：339，1981
3. 中国医科大学科研处：医学科学研究基本方法，47—81页，1980
4. 王志均、关于培养研究生的几个问题，生理科学进展13(1):20,1982

第二讲 实验设计有关的基本概念

一、总体、样本与抽样误差

1. 同质的个体构成“总体”。例如一个人体内有无数的白细胞，世界上有千百万高血压病人。总体是很大的，甚至是难以计数的。科研的任务是观察少数个体组成的“样本”去推测总体。例如取少量外周血计数，可推测病人全身每立方毫米含有的白细胞数；观察某药对几百例高血压病人的疗效可推测该药对其他千百万病人疗效。

2. 因抽样引起的样本与总体间的差异称抽样误差。抽样误差来源于个体差异，用标准误 (SE或 $S_{\bar{x}}$) 表示。均数的SE表示重复抽样时所得均值的变异范围，是从标准差 (SD) 计算来的：

$$S_{\bar{x}} = \frac{SD}{\sqrt{N}}$$

由上式可见，减少误差方法有：

(1) 增大样本：样本大到接近于总体，抽样误差即接近于0。但增大样本需要大量人力物力，如果实验条件控制不严，方法不精确，也不能达到目的，故现在不主张片面追求扩大样本。

(2) 减少误差 (SD)：精心设计，随机抽取均匀的小样本比粗糙的大样本好。均数 \pm SD表示数据的分散程度，如果样本不均匀，个体反应差别大，实验方法不精确，所得结果的SD必然大。应该用随机化、分层等方法使样本均匀、有代表性，缩小抽样误差并提高实验方法精确性减少实验误差。

(3) 抽样误差大小举例：

心得安的降高血压的作用

批 次	治疗方法	例数	舒张压 (mmHg)		P	值
			$\bar{x} \pm SD$			
I	安慰剂	20	111 \pm 7 (97~125)	109 \pm 6 (97~125)	>0.05	
	心得安	20	110 \pm 6 (98~125)	95 \pm 3 (89~101)	<0.01	<0.05
II	安慰剂	20	120 \pm 14 (92~148)	114 \pm 15 (84~144)	>0.05	
	心得安	20	120 \pm 15 (90~150)	105 \pm 15 (75~135)	<0.05	>0.05

注：() 内数字为 $\bar{x} \pm 2SD$ 的数值代表95%数据分布情况

由表可见第一批实验结果统计学显著的原因是抽样误差小 (N均为20, SE即决定于SD)。第二批实验抽样误差大, 如果增大样本也可达到组间比较显著。

二、逻辑推理

1. 归纳 (Induction)

是由样本 (个别) 推论总体 (一般) 的方法, 也就是认识客观事物规律性从特殊到一般的过程。例如, 测定10只狗枳实注射液0.5g/kg升高血压 $\bar{x} \pm SD$ 为 70 ± 11 mmHg, 计算出 $SE = \frac{11}{\sqrt{10}} = 3.5$ 。可以推测再作100次实验 (每次10只狗), 有95次的平均升压范围为 $70 \pm 2 \times 3.5$, 即63~77 mmHg, 这是由样本推论总体的例子。又如受体学说的建立和发展, 可以看作是由个别推论一般的认识过程。开始是证明乙酰胆碱作为神经递质作用于胆碱能受体; 随后认识到去甲肾上腺素作为神经递质作用于肾上腺素能受体; 现在认识到还有多巴胺受体, 组织胺受体, 5HT受体、GABA受体、阿片受体、腺苷受体等几十种受体。

2. 演绎 (Deduction)

是由总体 (一般) 推论到样本 (个体) 的方法, 例如已知升血压药可治疗休克, 枳实注射液能升高血压, 故可推测它也能治疗休克。又如已知受体在体内多有激动剂, 1973年证实阿片受体的存在, 推论这一受体不可能只是为接受外源性吗啡类物质的作用, 体内应有阿片受体的激动剂, 两年后果然发现了内啡肽。〔药理学进展, (1978) 12—25页上海科技出版社, 1980〕进一步了解到内啡肽的复杂生理功能。

归纳与演绎是两种基本的逻辑推理方法, 两者相辅相成, 在认识过程中两者互相渗透, 在唯物辩证法指导下交替应用, 导致认识的深化, 在医学科研中有巨大的作用。它与哲学上的两个认识过程密切联系, 即毛主席在矛盾论中指出的: “这是两个认识的过程, 一个是由特殊到一般, 一个是由一般到特殊, 人类的认识是这样循环往复地进行的, ……使人类的认识不断地深化。”

不言而喻, 应用归纳推理时, 样本必须有代表性, 所得结果必须准确可靠, 才能正确推论总体, 否则推论必然错误。同理应用演绎推理时, 所依据的一般规律必须正确, 否则推论的个别也会错误。同时, 必须认识: 人们对客观规律性的认识是无限的, 不应以现在还未认识到的规律去限制探讨认识新事物。所以归纳与演绎推理有时不免有失误。

三、变异 (variation)

1. 生物界固有的变异是普遍存在的, 个体差异和抽样误差是不可避免的。但是, 可以认识且有规可循。医学中常见的分布为正态分布, 即两头小中间大的钟形分布特点如正常人的身高、体重、血压、心率等正常值, 可用 $\bar{x} \pm SD$ 表示, $\bar{x} \pm 1SD$ 代表样取约68%的分布范围, $\bar{x} \pm 2SD$ 代表样本95%的分布范围。如果从同一总体中随机抽出许多样本, 每一样本可求得样本均数 (\bar{x}), 这些样本的均数绝大多数分布在总体均数

附近。例如反复测定一个人的白细胞数，每次测定的结果不会相同，但大多数测定的结果是接近的，可以反映全身血中白细胞的数目，通常可用样本的均数 $\bar{x} \pm SD (S)$ 表示该样本分布的离散程度，用SE即 $S_{\bar{x}}$ 表示重复测定样本均数的分布情况。

2. 统计量与参数：上述样本的 \bar{x} 、 S 、 $S_{\bar{x}}$ 等都是样本的统计量 (statistics) 指标，多次重复测定可以得到接近总体的指标，称为参数 (parameter)。总体的均数、标准差、标准误分别以 μ 、 σ 与 $\sigma_{\bar{x}}$ 代表。

四、误差 (Error)

误差是指测得数值与真值之差统计量与参数之差可区别如下：

1. 抽样误差：已于前述，由于个体差异存在，抽样误差是不可避免的，但可在抽样和分配时用适当方法缩小之 (方法见后)。

2. 分配误差：实验标本组成不均匀，缺乏可比性造成分配误差，如高血压有原发性与继发性的，原发性中又有一、二、三期之分，继发性又有肾性、血管性、嗜铬细胞瘤等多种原因。如果将轻症、重症和病因不同的高血压各自单独分组进行药物试验，结果缺乏可比性，而应该用先分层 (见后) 再随机分组的方法减少分配误差。广义说分配误差是抽样误差的一种。

3. 系统误差 (偏差)：是人为的带有一定倾向性的误差，是可以并且应该避免。造成系统误差的原因如下：

(1) 主观性：研究者希望所试方法或药物有效，在估计观测结果时带“有色眼镜”，造成“估计误差”。病人可因有主观性，对所试方法或药物作出过高或过低的评价，造成“应答误差”。

(2) 客观条件：仪器设备未校正，未标准化，如不同的体温表可相差零点几度；天平、吸管、血压计均有误差，结果或偏高或偏低是单向性一致的。

(3) 时间条件：实验观测过程中随时间变化实验条件可引起变化，如动物的麻醉由深变浅；实验操作人员技术方法逐步从生疏到熟练，可造成“顺序误差”，生物体内固有的“生物钟”现象 (如上下午人的体温、血压、甚至白细胞数、血中肾上腺皮质激素水平都有规律性周期性变化)，如不注意避免，也可造成系统误差。

(4) 实验误差：实验结果受实验对象的反应性差异，操作者操作熟练程度，和实验技术方法精确度 (accuracy) 的影响所致误差。如比色法测定药物通常不及紫外吸收光谱法精确，而药物如能显出荧光、用荧光光度计测定更敏感，可达毫微克 (ng)；用放射免疫法测定则最敏感，可达微微克 (pg)。为了减少实验误差，应设法改善实验条件，采用先进技术方法，并应提高操作水平，通过准备试验阶段。达到操作熟练，找出引起操作误差原因，减少操作误差。

五、概率 (机率, probability)

概率反映某一事件发生的可能性大小的量，简称为 P 。

1. 必然事件： $P = 1$

例如失血过多必然休克，休克时间过久，必然死亡。人要生存必然要消耗能量。

2. 不可能事件：一定条件 $P = 0$

例如人不吃饭不喝水，时间过久不能生存，人的视力、听力、耐寒力、耐热力、耐缺氧力等等都有一定限度，超过一定限度，就不可能看见，听见和生存。

3. 随机事件（P）：在一定条件下，可能发生，也可能不发生，P在0与1之间，即 $0 \leq P \leq 1$ 。

例如孕妇生男生女的概率大约50%（ $P \doteq 0.5$ ）。

概率论是研究随机现象规律的一门数学学科，应用于科研工作中帮助人们通过反复实验找出随机事件的固有规律性，判断根据实验结果作出结论的可靠性（可重复性），如显著性测验的P值。

六、事物间的联系

医学科研要发现规律就要研究事物间的联系，特别是病因学治疗学研究中判断因果联系。

1. 非统计学联系：即事件的发生缺乏统计学显著性。如服避孕药的与不服避孕药的妇女癌症发病率差异无统计学显著性，说明避孕药与致癌无关。

2. 统计学联系：统计学上有显著差异的事件，在因果关系上有两种可能。

①非因果联系：如轻度高血压病人住院后，只服用维生素B和C，血压恢复正常，不一定是维生素的疗效，而可能是住院休息的效果。

②因果联系：如前所述某小学儿童在公园松树下玩游戏回家后1—3天内50%以上发生皮炎、关节肿痛，耳廓红肿疼痛等症状，经流行病学调查及动物实验证明系松毛虫的毒液侵入皮肤所致的松毛虫病。松毛虫的毒液与发病有因果关系。

七、假设 (Hypothesis)

1. 无效假设 (Null hypothesis, H_0)

假设该样本均数与总体均数之间的差别来自抽样误差，即假定该样本是从同一个总体中抽出，即 $H_0: \mu_1 = \mu_2$ 。无效假设是统计学比较逻辑的基石之一。当计算出二者差别由于机遇概率 (Chance, Probability) 所致的可能性很少时（一般以0.05为显著界限）则可拒绝无效假设，而认为样本差异可能不是从同一总体抽样所得，即非抽样误差所致。

2. 备选假设 (Alternative hypothesis, H_A)

在拒绝无效假设后而接受另一种可能性假设，即认为实验结果可能非来自同一总体，即 $H_A: \mu_1 > \mu_2$ 。

3. 假设的的两类错误

(1) I类（假阳性）错误：即 H_0 符合实际，实验结果错误地否定了 H_0 ，用 α 代表其概率。

(2) II类(假阴性)错误:即 H_0 不符合实际,实验结果错误地肯定了 H_0 ,用 β 代表其概率。

4. α 和 β 对判断事物的影响举例:

心电图运动试验作冠心病诊断存在的问题:假设在无症状人群中4%有冠心病,即每1000人中40人有冠心病。假设运动试验敏感性为80%, $40 \times 80\% = 32$ 人运动试验阳性,假阴性 $\beta = 0.2$ 共8人。

假设运动试验特异性为90%,即假阳性 $\alpha = 0.1$ 则无冠心病的960人中假阳性者为96人。

真阳性32人加假阳性者96人=128人,假阳性占 $96/128 = 75\%$ 。

结果:查1000人,假阳性96人,假阴性8人,正确诊断者32人故不能以运动试验的心电图变化诊断冠心病(见陶寿洪中华心血管病杂志9(1):75,1981和中华内科杂志20(4):253,1981)。

八、假说形成方法

科研要创新,必须在事实的基础上发挥创造性思维,即恩格斯说的“理论思维”把感性认识提高到理性认识水平。理论思维要善于根据实验资料提出假说,进一步在实践中验证,加深认识,发展科学理论。假说形成的一般逻辑方法如下:

1. 差异法:根据观察到事物的差异,提出假说,找出规律性,Ahlquist 1948年根据肾上腺素对心脏兴奋,对骨骼肌血管和支气管扩张,而对内脏及皮肤血管收缩等差别,提出假说:心脏、骨骼肌血管和支气管上存在 β 受体;而内脏及皮肤血管上存在 α 受体。1967年Lands比较了肾上腺素,异丙肾上腺素对不同器管等作用差别后提出有两类 β 受体,兴奋心脏的为 β_1 受体,扩张血管和支气管的为 β_2 受体。1974年Langer根据实验差异把 α 受体分为 α_1 和 α_2 受体。

2. 一致法:又称求同法或类同法,是根据事物发生的一致性,提出假说。如1973年——1975年美国调查了266,287个男性,找出冠心病多发于有高血压、高血脂、吸烟和糖尿病的病人,从而把这些因素作为冠心病的危险因素(risk factor),单用此法不易分清主次,分辨因果关系,常与差异法并用。美国随后又选出12,000人追踪到1982年,一组维持原生活方式,另组消除危险因素,进行“多项危险因素干预试验”(MRFIS)以消除危险因素,假说进一步验证,确定冠心病的发病因素。这种把差异法与一致法并用的方法,称为“类同与差异并用法”。

3. 伴随变异法:又称共变法。事物的某一因素总是与事物某现象伴随发生,该因素与某现象可能为因果关系。例如湖北等吃棉子油地区,多发低血钾和“软病”,推测棉子油可能与“软病”有关。又观察到吃棉子油的一些男性缺乏生育能力,推测棉子油中可能含有影响男性生育的成分。后来果然发现棉酚能影响精子生成,用作男性避孕药为我国一大创举。又例如1959——1962年德国妇女怀孕期服用镇静药反应停(Thalidomide)者,发生畸胎有二千例之多,此药与畸胎伴随发生,后经动物实验证明此药有致畸作用,以后提高了药物致畸的警惕性,并增加了这一试验项目。

4. 类推法:是根据已知事实推论未知事物,提出假说,例如已知现有的杀血吸虫

药物对人有效者对动物也有效，牛奶浆草在实验动物无效，对人也可能无效。临床用以消晚期血吸虫腹水的机理可能与其致泻作用消水有关，因为已知的致泻药也能消腹水，进一步实验证明牛奶浆草消血吸虫病腹水不是特异性的，因为它也能消除手术后狭窄的下腔静脉引起的大量腹水，与巴豆等致泻药相似（见新医药杂志（3）：28，1973）。

5. 剩余法：逐一排除可能引起作用的各因素，剩余的因素就可能是原因，例如肾上腺素能兴奋 α 与 β 两种受体，用 α 受体阻滞剂酚妥拉明后其升压作用反转为降压，推测其降压作用可能是兴奋 β 受体扩张血管所致。

习 题

1. 何谓抽样误差？如何减少抽样误差？

比较以下两例抽样误差大小，并计算重复测定时均数95%的可信限的范围。

(1) 100例正常人血浆胆固醇 $\bar{x} \pm SD$ 为 $197 \pm 25\text{mg}\%$

(2) 200例冠心病病人血浆胆固醇 $\bar{x} \pm SD$ 为 $210 \pm 20\text{mg}\%$

2. 何谓 I 类错误和 II 类错误？两者有什么关连？

某诊断方法 $\alpha = 0.05$ ， $\beta = 0.1$ ，设某病阳性率为3%，试推算检查2,000人有多少误诊和漏诊者，并对此法的敏感和特异性作出估计和评价。

3. 各举一例说明形成假说，伴随变异法和类推法的意义。

参 考 文 献

1. 上海第一医学院卫生统计学教研组；医学统计方法2-3页，24-25页，29页，36页，上海科学技术出版社，1979

2. 金正均；医学试验设计，1-28页，上海科学技术出版社，1964

3. Feinstein AR; Clinical biostatistics XXXI. The other sides of "statistical significance": alpha, beta, delta, and the calculation of sample size, Clin Pharmacol Ther 18:491, 1975

第三讲 实验设计三大原则

一、对 照 (Control) [1]

对照是实验设计的三大基本原则之一。没有对照就没有比较和鉴别，也就会影响科学性。目前我国临床报道很少对照。许多中草药的临床疗效，长期不能定论，与此有关。此外，由于缺乏严格对照而得出假阳性的结论盲目推广应用的实例很多。“鸡血疗法”、“卤硷疗法”、针刺治疗聋哑等为什么会风行一时而又否定？我们应从中吸取教训。严格对照观察。提高医学研究的科学性。

对照的原则即“齐对比”的原则。即除了要观察研究的因素外，实验组与对照组一切条件应尽量相同，有完全的可比性，才能排除影响因素，作出科学结论。根据不同的实验和临床观察要求，有以下对照可供应用。

(一) 空白对照与实验对照

空白对照是指不给任何处理的对照。例如研究利血平或口服避孕药有无致癌作用，应以同年龄组自然发癌率为对照。我们在研究矮地茶对电刺激猫的喉返神经引起的咳嗽的止咳效果，先观察在不给任何措施条件下在几个小时内引咳电阈值的自发变化，因为随着麻醉的变浅，电阈值会降低；而在补加麻醉后，或神经牵拉损坏后，电阈值又会升高。只有掌握了实验条件下的波动范围才能正确判断给药后的变化是否有意义。

在许多情况下，空白对照不能包含影响实验结果的全部因素。而应采用与实验组操作条件一样对照措施，称为实验对照。如果确知对照操作对实验结果无影响时，可以省略。但许多情况下对照操作对实验结果是有影响的，不可省略。例如作毒性试验时，试验组给药，对照组应用同样给药途径给予溶媒。因为不论灌胃或注射，在技术不够熟练时，操作本身可能造成损伤甚至动物死亡。

在作亚急性毒性试验或筛选抗肿瘤药时，对照组也必须每天给予溶媒灌胃或注射。因为有些是意料不到的影响的例如有人观察大白鼠每天抱在手里抚摸几分钟，其体重、骨长度和脂肪组织的增加都比不抚摸的快些，前者甚至切除甲状腺和付甲状腺的手术死亡率和某些药物中毒的死亡率也比后者低些。有时，一个笼子放入动物的数目也显著影响某些实验结果。例如苯丙胺对群居（1笼多只）小白鼠的毒性比独居（1笼1只）者要大得多，查其原因可能与群居者给药后体温升高加剧有关。

动物实验如需进行外科手术，对照组不应该用不作手术的空白对照，而应该进行与实验组相似的假手术，以排除手术创伤和感染造成的影响。

药物的溶媒有些是有药理作用的，如静脉注射吐温能降压、溶血、升高血胆固醇，二甲亚砜有止痛、抗炎、退热、扩血管、抗菌等作用。（David NA: Ann. Rev. Pharmacol. Toxicol. 12:253, 1972）中草药难溶成分吐温和二甲亚砜助溶和溶解，实验时如不作溶媒对照，可造成假阳性结果。植物药粗提物常呈酸性反应，作离体哺乳动物心脏灌流时可扩血管而增加冠脉流量。应该用同一pH的灌流液作对照，否则也可得假