

临床科研方法——理论与实践

LINCHUANG KEYAN FANGFA —— LILUN YU SHIJIAN

主 编： 刘建平（第三军医大学流行病学教研室）

冷泰俊（第三军医大学流行病学教研室）

编 者：（按姓氏笔画排列）

刘建平（第三军医大学流行病学教研室）

李亚斐（第三军医大学流行病学教研室）

许汝福（第三军医大学卫生统计学教研室）

冷泰俊（第三军医大学流行病学教研室）

薛国文（第三军医大学流行病学教研室）

军事医学科学出版社

内容提要

临床科研方法学即临床流行病学，是有别于基础医学研究的应用学科。它将流行病学的原理和方法应用于临床医疗实践和科研工作，以解决医疗实践中所遇到的问题。本书从临床医学角度出发，用群体的观点研究和评价疾病的病因、预后、诊断、防治等问题，并介绍了近年来国外临床科研领域中出现的新分支——循证医学，对提高临床医疗质量和科研水平会有所裨益。

全书共十一章，主要包括临床科研设计的原则和要点，疾病病因、诊断、治疗、预后研究的设计与评价，循证医学与 Meta-分析，临床科研中常见的偏倚、混杂和交互作用以及有关统计学的方法，并附有大量实例，便于临床医师学习、理解和应用。本书的主要对象为各科临床医师、从事临床和药物研究的专业人员、临床科研管理决策者、高等医学院校临床医学专业硕士研究生、本科生，可作为教学和科研的参考书使用。

* * *

图书在版编目 (CIP) 数据

临床科研方法：理论与实践/刘建平主编.—北京：

军事医学科学出版社.1999.11

ISBN 7-80121-192-8

I. 临… II. 刘… III. 临床医学-研究方法 IV. R4-3

中国版本图书馆 CIP 数据核字 (1999) 第 60092 号

* *

军事医学科学出版社

(北京市太平路 27 号 邮政编码 100850)

新华书店总店北京发行所发行

潮河印刷厂印刷

*

开本：787mm×1092mm 1/16 印张：12.5 字数：307.3 千字

2000 年 1 月第 1 版 2000 年 1 月第 1 次印刷

印数：1-3000 册 定价：18.00 元

(购买本社图书，凡有缺、损、倒脱页者，本社发行部负责调换)

前 言

随着临床医学和流行病学学科的发展和需要,于本世纪七十年代诞生了“临床流行病学”这一新兴的交叉学科。它是将流行病学和医学统计学的原理和方法应用于临床科研的设计、测量与评价(DME)之中,解决临床医学实践中的诊断、治疗、预防、病因、预后等问题。作为一门临床科研的方法学,它建立了自己的学科体系,并在临床医学研究和教学中产生着日益重要的作用,被誉为临床医学的“建筑学”。在国外已将其列入医学院校临床各专业必修的基础课,也是临床医师执业考试的必备内容。该学科自80年代初期引进我国以来,已经有了很大的发展和进步。国内已相继建立了中国临床流行病学网络(ChinaCLEN),成立了中华医学会临床流行病学学会,也即将诞生一本专业杂志——中国临床流行病学与循证医学杂志。

本书主要编著者均有多年临床医疗实践和科研经验,并在国内较早地接受“临床流行病学”的学习和训练,通过对国内医学期刊中发表的大量临床研究文献进行系统收集和评价发现,我国的临床研究整体水平不高,研究的设计存在不少问题。结合编著者对各级临床医师、临床医学专业硕士研究生十余年的“临床流行病学”培训及教学工作体会和经验,我们感觉到普及和推广临床研究的理论知识对于提高医疗质量和科研水平十分必要和迫切,临床医师的继续教育也急需这方面知识的学习和深造。鉴于目前国内针对临床医师的实用型科研方法参考书较少,因此,在以往多次修订的供临床医师、临床硕士研究生使用的《临床科研方法》和《临床流行病学》自编教材基础上,结合临床医疗实际和科研工作的需要,收集补充了大量实例,并参考国外该领域近年来的新进展如循证医学、Meta-分析介绍给读者,而编著形成了这本临床科研方法的参考书。其旨在为临床医师和从事临床科学研究者以及临床医学专业硕士生、本科生学习和掌握临床科研方法的基本理论和方法,以提高进行临床研究的能力和水平。

本书具有以下特点:

- 从理论到实践系统介绍临床试验的设计、测量与评价。
- 针对性强,专为临床医师编写,具有实用、易懂和可操作性。
- 主要编著者为临床医师出身,长期从事临床流行病学教学与科研,能将抽象难懂的流行病学与临床医疗实际较好地结合,解决临床实践中的问题。
- 核心思想贯穿了设计、测量与评价的临床科研精髓。

本书编撰过程中得到了“中国循证医学中心”的支持,也得到了第三军医大学科研处、预防医学系领导及流行病学和卫生统计学教研室全体老师的大力支持。经北京军事医学科学出版社领导的亲切关怀和促成,使本书得以顺利出版。在此一并表示衷心的感谢。

由于编者水平有限,虽经努力,书中缺点错误在所难免,请读者指正。

编者
1999年7月

目 录

第一章	绪 论	1
	一、定义和发展	1
	二、研究对象与任务	2
	三、研究内容	3
第二章	临床科研设计的基本原则与设计要点	6
	一、临床研究的基本原则	6
	二、临床科研的设计要点	12
第三章	疾病的筛检及其评价	20
	一、筛检的概念	20
	二、正常值及其范围的确定	21
	三、筛检试验的评价	22
	四、筛检过程中的效果评价	26
第四章	诊断试验的设计与评价	30
	一、诊断试验评价的基本方法与步骤	30
	二、诊断试验常用的评价指标	31
第五章	临床防治研究的设计与评价	41
	一、临床防治研究概述	41
	二、临床防治研究设计方案及其抉择	42
	三、常用防治研究的设计	43
	四、防治研究设计中的注意事项	58
	五、临床防治研究的评价	61
第六章	预后研究的设计与评价	66
	一、基本概念	66
	二、基本方法	68
	三、预后估计研究的评价	75
第七章	病因研究的设计与评价	77
	一、病因的概念	77
	二、病因研究的步骤与方法	79
	三、病因研究的评价	84
第八章	医学科研中的机遇、偏倚、混杂与交互作用	88
	一、机遇(chance)	88
	二、偏倚(bias)	89
	三、混杂(confounding)	95
	四、交互作用(interaction)	99
第九章	Meta-分析	104
	一、Meta-分析的定义	104
	二、Meta-分析的目的和适用对象	104
	三、Meta-分析的方法与步骤	104

四、Meta-分析应用实例	107
五、结语	112
第十章 循证医学概论	114
一、循证医学的概念和起源	114
二、EBM 与临床医疗实践	115
三、世界 EBM 协作网(The Cochrane Collaboration)与中国 EBM 中心	116
四、临床疗效资料库及其系统评价	118
五、实践 EBM 实例	122
六、实践 EBM 的优缺点	124
第十一章 临床科研中常用的统计学方法	127
第一节 随机化及其常用统计分组方法	127
一、完全随机化	127
二、分段随机化	128
三、配对设计	129
四、单位组设计分组	129
五、分层随机化	130
六、随机应从设计	130
七、随机抽样	130
第二节 样本含量的估计	131
一、样本率比较所需样本含量的估计	131
二、样本均数比较所需样本含量的估计	132
三、直线相关所需样本含量的估计	134
四、成组比较病例对照研究所需样本含量的估计	134
五、匹配比较病例对照研究所需样本含量的估计	135
六、整群抽样所需样本含量的估计	136
七、单纯随机抽样所需样本含量的估计	136
八、分层抽样所需样本含量的估计	137
第三节 常用统计分析方法	138
一、数据的预处理	138
二、分类变量资料的统计分析	140
三、数值变量资料的统计分析	145
第四节 生存分析	160
一、常用统计指标	160
二、生存期比较	163
三、比例危险率回归模型—Cox 模型	164
四、Logistic 模型	165

第一章 绪 论

流行病学方法用于临床由来已久。近十多年来由于临床与流行病学两方面发展和需要,已形成一门独立的学科——临床流行病学(Clinical Epidemiology)。它是临床、流行病学、社会医学、卫生经济学与统计学五门学科的边缘科学,以研究临床科研的设计、测量与评价(design, measurement, evaluation, DME)为主要内容。通俗的名称为临床科研方法,正越来越多地受到医学各学科的重视。

一、定义和发展

临床流行病学这一名词最早由 John R. Poul 首先提出(1938)。他认为公共卫生学的流行病学是社会人群的流行病学,而临床流行病学则是以临床病人为主要对象,他强调临床医师与流行病学家要紧密联系,协作共事,但当时尚未引起广泛的重视。60 年代后期 Alvan Feinstein, David L. Sackett 等提出临床流行病学是研究如何将流行病学与卫生统计学的原理和方法应用于临床,并用以指导临床日常工作与科研工作,提高临床医疗质量与科研工作的水平,其核心是临床医师要树立人群的观点。由于疾病问题千变万化,同一种致病因素或同一种防治措施,作用于不同的人可能产生千差万别,甚至完全不同的效应,而临床医师所接触的只是医院中收治的一小部分病例,这些病例并不能代表整个患病人群的全貌。又由于不同的病人临床表现不同,要求的治疗措施不能千篇一律,必须因人而异,对症下药,故日常诊疗活动强调的是个体特征。这种来源于个别病人的实践经验是否具有普遍意义必须经过人群的验证才能回答。这又从另一方面说明群体观念的重要性。

临床流行病学研究的主要内容是临床科研的设计、测量与评价,简单地说就是临床科研方法。由于临床日常工作主要是诊断与治疗,看书、学习和注意力也主要集中于新的诊疗方法,除少数对科研有兴趣的医师,一般不太注意科研方法与统计学处理,偶尔做一点科研,常因方法欠妥或统计学运用不当,得不到应有的结果,甚至劳而无功,直接影响日常工作与科研的提高。因而此门学科出现后很受临床医师的欢迎,反过来又促进了本学科的发展。

在洛氏基金会(Rockefeller Foundation)、世界银行和世界卫生组织的支持下,1982 年建立了国际临床流行病学网络(International Clinical Epidemiology Network, 简称 INCLEN)。现有 5 个国际级的临床流行病学资源与培训中心(CERTC),18 个国家的有关医学教育研究机构加入了该网络。作为该网络的成员,我国华西医科大学和上海医科大学已从临床流行病学单位(Clinical Epidemiology Unit, 简称 CEU)升为 CERTC。目前在 5 所医科大学成立了 CEU。我国自 80 年代初引入该学科以来,于 1983 年在华西医大、上海医大与广州中医学院建立了 3 个培训中心,并在有关医学院校中相继建立了有关教研室或咨询委员会,开设临床流行病学课程,培养专业人员,在国内外广泛进行了学术交流。于 1989 年建立了中国临床流行病学网络(China Clinical Epidemiology Network, 简称 ChinaCLEN),将此工作又向前推进了一大步。军队系统正在酝酿成立 PLACLEN。综上所述,临床流行病学是临床、流行病学和卫生统计学等三学科相结合的边缘学科。它将流行病学的群体观念引入

临床,用以探讨疾病的病因、诊断、治疗、预防和预后等临床问题,属临床基础学科之一。

二、研究对象与任务

临床流行病学的研究是以临床病人作为主要的研究对象,辅之以社会人群。临床医师比较重视个体病人。医院病人大多数是以个体为单位来就诊,医师接触的也就是一个个具体的病人。通过对某一具体病人的询问和检查来确定他罹患的是什么病,又根据病人的具体情况对症下药进行治疗。一个高明的医师决不会千篇一律地用同一种方法治疗一群病人,总希望搞清该具体病人个体的特点而进行诊断与治疗。这是临床工作的特点与需要,本来无可非议。但这种仅从个体出发来考虑的方法,对有些问题就难以回答。如某一症状或体征或某种处理方法所取得的效果是偶然的还是普遍的?其出现的概率是多少?不经一定数量病人的研究,就无法回答。作者曾经治过一名经病理确诊的原发性肝癌病人,经某中医师治疗存活3年多,一般情况明显好转,取得意想不到的效果。当时认为有可能是肝癌治疗上的重大突破,但经更多病例验证却均无效。这只能说该病人是一特殊病例,而不能确定该药对肝癌有效,类似的实例每位工作时间较长的医师,都可能有所体会。

另一个问题是能够收容住院的病人大多为临床表现明显的或比较重的病例。临床医师一般不太重视那些不显性病人或轻病人。这对了解疾病的自然史、认识疾病的全貌是不够的,由于不同医院或同一医院对不同病种收治的条件不同,入院率也就有所不同,用作研究就可能产生入院偏倚(详述于第八章),会使结果发生误差。因此必须辅之以社会人群来弥补其不足。流行病学是研究人群中疾病的分布及其影响分布因素的学科,与临床医师相比,更重视的是自然人群。包括病人、无病的健康人与未患该种疾病的其他病人,有其可取之处,但由于其研究的对象是人群,而非单个的具体病人,只能从群体的观念来回答某种现象出现的概率,而不能解决这一病人的具体问题,这又是其不足之处。一般来说,流行病学专家重视的是疾病是否发生,发生概率有多高,而不太深究单个病人的具体问题,也不深入了解其病理机理与发病机理。例如通过人群调查吸烟与肺癌可能具有因果联系,通过戒烟可减少肺癌的发生,从流行病学研究来说,这一研究就结束了。至于纸烟中的什么具体成分有致癌作用,该成分又是通过什么机理引起癌变的则就不太关心,认为是基础医学的工作,从这一角度来看,流行病学比较粗糙与表浅,不能满足临床医师的要求。当然上述问题都是相对而言,流行病学调查也可用分子生物学的方法,这就是分子流行病学。基础与临床医学也要运用流行病学原理和方法。由此可见,基础医学、临床医学与流行病学各有所侧重,必须相辅相成,紧密配合才能对疾病的本质有全面的了解。

临床流行病学可作为一门桥梁课程介于临床医学与流行病学或基础医学与流行病学之间。临床流行病学研究的主要任务有下列几方面:

- 引起某种疾病的危险因素(含致病因素)的探索;
- 新诊断方法的临床评价;
- 新防治措施的研究与疗效评价;
- 预后的预测与评价;
- 某种疾病的自然史;
- 临床诊断治疗措施的决策分析与决策分析。

上述任务实际上也是临床科研工作的主要内容。因此临床是临床流行病学研究的基

地，临床流行病学工作人员，一定不能脱离临床，否则将成空中楼阁，形同虚设。临床医师与临床流行病学工作者也各有其工作重点。前者主要就其专科，进行病人的诊断工作，后者则不受专科与病种的限制，重点在研究带有共性的科研方法，故本学科的特色是专门研究适合临床医师使用的流行病学。

三、研究内容

临床流行病学研究的主要内容，可概括为三个方面：

- 如何制定一份质量优秀的科研设计方案；
- 如何准确地收集、整理分析实施过程中的数据；
- 对所得结论如何进行科学的评价。

以上三点可归纳为设计、测量与评价三个词，简称 DME，也可以说，临床流行病学就是研究如何对临床科研进行 DME 的方法学，分述如下。

(一)设计

设计好比一项工程的蓝图，没有优良的工程蓝图，绝不可能建设质量良好的工程。不言而喻，没有良好的科研设计，也不可能完成质量良好的科研。有人说：“制定好一份质量优良的设计方案，科研工作就已完成了一半。”这句话并非危言耸听，言过其实。有关临床科研设计的基本要点将详述于第二章第二节，主要有下列 9 条：

1. 根据研究目的，选择恰当的受试对象。纳入研究的受试者诊断必须确定，要具有代表性。
2. 设置合理的对照，有比较才能鉴别，临床研究大部分均通过对比研究来进行，选定能说明问题的对照组，十分重要。
3. 分组与抽样均应尽可能采用随机化的方法以保证其均衡性。
4. 试验因素要明确、标准化与量化，并尽可能简单化以避免发生污染与干扰。
5. 选定适当的设计方案。原则是既要力求具有较高的论证强度，又要切实可行，要结合具体情况而定。
6. 评定指标与标准要求客观、可靠、量化。一般尽可能用不受主观因素影响的硬指标，并制定措施，保证从始至终，不管任何人执行均统一不变。
7. 科学估计合适的样本数量，以能达到研究目的为准，防止因样本太少，得不到应有结论，或者样本过大造成人力物力的浪费与拉长研究周期。
8. 选择正确的收集、整理与分析数据的方法，制定必要的统计表格。
9. 注意预防机遇、偏倚、混杂与交互作用所造成的误差。力争用盲法处理并争取受试者有良好的依从性。

(二)测量

所谓测量是指用定量的方法来衡量临床工作科研中所发生各种问题与现象，主要包括：

- (1) 疾病发生频数与分布的测量即各种率与比；
- (2) 症状与体征、分布规律及其变化；
- (3) 疾病带来的肉体、精神、经济及社会的影响；

(4) 疾病带来的费用消耗及如何提高临床工作的经济效能(efficacy)、效果(effectiveness)和效率(efficiency)等问题。

测量的主要问题是指标的选择,即选择适当的指标以科学准确地反映出其研究结果的问题。除上述设计中所论及的问题外,还要注意:

(1) 所选指标的目的性:所选用的指标必须与研究的目的有本质的联系,如对一名肝脏病病人为了反映其肝细胞损害的情况应选用 ALT(GPT),反映在肝细胞合成蛋白的功能应选蛋白电泳或白蛋白/球蛋白,反映肝脏的解毒分泌功能则以 BSP 与 ICG 为宜,可根据专业知识确定。指标的高、新、尖固然重要,因能将效应表达得更深刻和精密,但有时与研究目的并不一致,就没有什么意义,反而造成浪费,要注意选择。

(2) 指标的客观性:临床数据有硬、软两种。描述人群中疾病的发生与分布的指标如死亡率、病死率、患病率及标化死亡比等,数据比较明确,属硬数据,其他如体温、各种皮疹、实验室检查的数据以及因疾病引起的费用消耗数据等,也较明确属硬数据,但有些问题,如疼痛、乏力等症状及肿瘤、烧伤病人治疗后的生存质量等则不容易用明确的数字来测量,属软数据。

(3) 敏感度(sensitivity)与特异度(specificity)详见于第四章。

(4) 准确性与精确性:准确性表示测得的结果与真实结果接近的程度,主要受系统误差的影响。精确性或可靠性则表示重复测定时,其多次结果彼此接近的程度。准确性与精确性并不一定并行不悖,可以准确性好但精确性差,反之亦然。所选指标应兼顾这两个方面,详见第三章。

(三)评价

所谓评价指运用科学的方法,制定出某些标准,并运用这些标准来评价各种临床问题。有关临床的文献多如繁星,许多问题又常是褒贬不一,众说纷纭。如何在这浩如烟海的文海中去粗取精,弃伪存真,必须有一个科学的标准与方法,这也是 DME 工作中非常重要的一部分。McMaster 大学的临床流行病学家系统地介绍了各种评价的标准,转录如下以供参考。

1. 诊断试验的评价标准

- (1) 是否和公认的标准方法进行对照研究;
- (2) 所观察的病例是否包括了多种不同的临床情况,轻的、重的,治疗过的和未治疗过的,以及包括那些容易混淆的其他疾病;
- (3) 是否介绍了观察组和对照组病例的来源,是如何选择的;
- (4) 该试验测定的精确性、准确性如何,以及其观察误差的大小;
- (5) 正常值的确定是否合理;
- (6) 如该试验是作为一组试验或作为序列试验之一应用于临床者,应当检验该试验在该组试验总的效力中的作用;
- (7) 该试验的操作方法及注意事项是否作了详细介绍;
- (8) 该试验的实用性。

2. 病因和因果关系研究的评价标准

- (1) 该论断是否从真正的临床人体试验研究中获得;
- (2) 暴露于该因素和发生疾病之间的联系强度如何;

- (3) 不同作者、不同地区,用同样的研究方法,所获得的结论是否一致;
- (4) 先后关系是否正确,即是否先暴露于该因素而后发病;
- (5) 是否有剂量—效应梯度,其联系是否有流行病学意义;
- (6) 其联系是否有医学上的意义;
- (7) 其联系是否有特异性;
- (8) 这种联系是否和以前已被证实的病因联系相类同。

3. 防治措施效果的评价

- (1) 是否用真正的随机方法将观察对象分配到治疗组和对照组;
- (2) 是否报告了所有临床上有关的结局;
- (3) 文中是否介绍了研究对象的临床情况,如诊断标准、病例来源等;
- (4) 是否同时考虑了统计学和临床的意义;
- (5) 所介绍的治疗措施是否具有实用性;
- (6) 在下结论时是否交待了所有的观察对象。

4. 预后估计的评价

- (1) 所观察的病例是否是从最早出现症状时开始观察的,即观察病程的起点是否统一;
- (2) 是否介绍了观察病例来源;
- (3) 是否所有的病例都已随访,失访率多少;
- (4) 是否有客观的预后指标;
- (5) 是否应用盲法来估计预后指标;
- (6) 对其他影响预后的因素是否经过统计学的校正。

评价尚包括经济评价,即运用卫生经济学的方法来评价各种防治措施,医疗保健措施的成本—效益核算,还能应用卫生行政管理,评价卫生事业投资的方向;卫生管理措施的质量,经济效益和社会效益等,此外还可应用于教学效果的评价。

综上所述,临床流行病学所研究的内容就是从临床科研的设计、测量与评价各个环节着手,尽量排除各种主观或客观的因素对研究的影响,以保证所得结果与结论更科学、更准确、更切合客观实际的情况,是一门研究科研方法的科学。

(冷泰俊)

第二章 临床科研设计的基本原则与设计要点

医学是关于人体正常结构功能及其疾病的科学,属生命科学范畴。医学科研的任务为揭示人体生命的本质,认识健康与疾病相互转化规律,并据此而达到救死扶伤,提高人民健康水平的目的。因此它既有一般科研的共同特点和要求,更有其本学科不同的特点与要求。

临床医学研究是以人为研究对象,即使有些研究可借助动物模型来进行,但这些从动物实验中所得到的结果,最终必不可少地要经过人体试验后,其理论假设才能够得到肯定。临床医学研究最基本的出发点,在于阐明疾病的病因、诊断、治疗、预防、自然病程及其预后等方面的重要问题,从而认识疾病的本质,并进行有效的防治,达到保障人类健康和促进社会进步的目的。因此,临床科研具有以下共同特点。

1. 个体差异大、试验条件不易控制

众所周知,人是最复杂的生命体,人体的生命现象是最高级的物质运动形式,不但有生理、病理活动还有心理思想活动;不但有生物性活动,还有社会活动。人体之间的差异十分显著,试验条件难以标准化是其一大特点,不同于一般生物实验,更不同于理化实验。

2. 临床研究涉及到医德与伦理学问题

一切研究都必须在保证不危害受试者的生命与健康及伦理准则的前提下才能进行。如病因及有关致病因素研究,就不允许用有可能致病或使病情加重的因素作为实验因素,用人来做试验,对一些疗效尚不确实或是否有可能引起严重的毒副作用尚未弄清楚之前,绝不允许冒然进行临床试验。也就是说有许多研究因医德问题,不允许用人来做试验研究,因而只能用论证强度较低的观察或调查分析的方法来进行研究。

3. 临床研究的内容广泛,涉及的学科众多

由于疾病发生的模式已从生物医学向生物—心理—社会医学模式转变,因此,临床研究的范围涉及病理学、病理生理学、药理学、流行病学、心理学、社会学、卫生经济学、统计学等。从学科的联系和内部结构来看包括:①研究健康与疾病相互转化的机理与规律的基础医学;②防止由健康向疾病转化的预防医学;③患病后促使病人由患病向健康转化的临床医学等三大部分。临床科研方法主要是运用流行病学和卫生统计学的原理和方法解决临床实践中所遇到的疾病病因学、诊治、预后等问题。下面就一些临床研究的共性问题,即影响其研究结果的科学性、准确性、可信性及实用性等方面的问题,作一概括的介绍。

一、临床研究的基本原则

临床研究必须遵守以下共同的原则。

(一)医德与伦理学原则

前已述及,医学研究是以人为研究对象,涉及人道主义与伦理学的要求。原则上讲,不允许用人来做试验研究,而只能用观察或分析的设计方案进行研究。如吸烟与肺癌的因果关

系研究，就不可能人为地规定一组人吸烟、一组人不吸烟，前瞻性观察若干年，比较二组肺癌的发病率以证实吸烟与肺癌的联系。而只能用病例对照或队列研究等论证强度较低的设计方案来进行。防治方法的研究也必须先对该种治疗药物的药理、毒理作用有比较全面的了解后才能进行临床试验。一切试验措施均需要得到受试者的同意后才能试用。1996 年在南非召开的第 48 届世界医学大会所修订的赫尔辛基宣言 II，就是有关此问题的指导性文件。转录如下。

【附录】 赫尔辛基宣言 II

指导医师进行人体生物医学研究的建议

前 言

保护人民的健康是医师的天职。医师的知识和良知应贡献于完成这一使命。世界医学会日内瓦宣言对医师的要求是“首先考虑的应是其病人的健康”。医学伦理学国际准则规定，“当提供的医疗保健措施有可能损害病人的身体或精神状况时，医师必须按病人的利益行事”。

以人类为对象的生物医学研究，须以改进诊断、治疗和预防措施以及明确疾病的病因和发病为目的。当前，绝大多数诊断、治疗或预防措施均有一定风险；生物医学研究也如此。

医学的进步是以研究为基础的，这些研究最终在一定程度上均有赖于以人类为对象的试验。在生物医学研究领域，必须明确以诊断、治疗为主要目的的医学研究与纯学术的对病人没有直接诊断、治疗价值的医学研究的基本界限。在实施可能对环境有影响的研究项目时须特别谨慎；必须重视实验动物的权益。

把实验结果应用于人类以促进科学发展和病人的救治是至关重要的，因此世界医学会制订了以人为对象的生物医学研究应遵循的指导原则，今后还要随时加以审核。必须指出，此规定只是对全世界医学的一种指导方针，并不能以此免除医师们根据其本国法律所应承担的刑事、民事及伦理责任。

1. 基本原则

1. 涉及人的生物医学研究必须符合普遍接受的科学原则，应有充分的实验室及动物实验依据和对科学文献的全面知识。

2. 涉及人的每一实验步骤的设计与执行方案必须详细写出，送交独立的专门机构审查。

3. 涉及人的生物医学研究只能由合格的科学人员在一位合格的临床医务人员监督下进行，必须由合格的医务人员对受试者负责，而决不能由受试者本人承担责任，即使他本人同意参加研究也不行。

4. 涉及人的生物医学研究，除非其目的之重要性与其固有的危险性相比是值得的，否则不应进行。

5. 在进行任一项涉及人的生物研究之前，必须先对可预见的危险性和受试者或他人所得利益仔细权衡，必须把受试者的利益置于科学和社会的利益之上。

6. 受试者捍卫其自身完整性和健康的权利必须受到尊重！对个人隐私尤应谨慎，应将研究对受试者身体及精神和个性的影响减至最低限度。

7. 医师除非能预测潜在的伤害，否则不应从事该研究项目。任何研究，一旦伤害超过了

潜在的利益, 医师必须中止其进行。

8. 在发表研究结果时, 医师必须保持其结果的准确性。试验报告如果不符合本宣言的原则, 不应被同意发表。

9. 在任何对人的研究中, 应向每一个可能的受试者说明研究的目的、方法、预期的利益、潜在的伤害及可能引起的不适。必须向受试者讲明他们有决定是否参与的自由, 并且在任何时候均可退出试验, 医师必须得到受试者的知情同意, 最好是书面形式的。

10. 征求研究对象同意时, 医师要特别注重受试者和医师是否处于一种从属关系或是否被迫同意。如果有这种情况, 则应由不参与这一研究因而与这种关系无关的医师来取得同意。

11. 如受试者尚未符合法律资格, 则应按照国家立法的从合法的监护人处取得同意。如果由于受试者身体或精神的原因无法取得其同意时, 则应按照国家法律, 从负责的亲属处获得同意。

12. 研究计划中必须包括有关伦理问题的说明, 并且要与本宣言的原则一致。

II. 与专业性医疗保健结合的医学研究(临床研究)

1. 在治疗病人时, 医师应能自由应用新的诊断及治疗措施, 如果他认为这些措施有希望挽救生命、恢复健康或减轻病情。

2. 新方法的潜在利益、危害和不适, 必须与现有最好的诊断、治疗方法比较并权衡之。

3. 在任何医学研究中, 包括对照组(如果有的话)在内的每一名病人, 都应保证得到已知的最好的诊断与治疗。这并不排除在研究中使用一些尚未证实有诊断或治疗作用的安慰剂。

4. 病人拒绝参与研究绝不应影响医患关系。

5. 如果医师认为根本不需要取得病人的知情同意, 则应将这一提议的特殊理由在研究方案中说明并递交给独立的委员会。

6. 只有在医学研究有潜在的诊断或治疗价值时, 医师才可以结合医疗保健进行医学研究, 其目的在于获得新的医学知识。

III. 以人类为对象的非治疗性生物医学研究(非临床生物医学研究)

1. 对人进行纯科学的应用性医学研究时, 保护受试者的生命与健康是医师的责任。

2. 试验对象应当是自愿者——健康人或实验设计与其疾病无关的病人。

3. 如果研究者判定继续试验将损害受试者时, 必须中止研究。

4. 在以人为对象的研究中, 绝不能把科学及社会的利益凌驾于受试者的利益之上。

[引自: 美国医学会杂志(中文版), 1997; 277: 925-926, 文字经审者稍作改动]

(二)对比研究的原则

有比较才能鉴别, 许多问题就是在对比中得到澄清。临床研究由于个体差异大, 影响因素多, 试验条件难以标准化。没有严格的对照, 许多问题很难得到肯定的结论。例如感冒、肺结核、病毒性肝炎等许多疾病都有明显的自愈倾向。即使是恶性肿瘤等病死率极高的疾病也不一定 100% 死亡, 存活时间也会长短不一, 参差不齐。至于像高血压、类风湿性关节炎、慢性肾炎等慢性病, 整个病程中常有自行缓解与反复发作的情况。因此在判定某些药物的疗效、某种因素对预后的影响等研究中, 没有严格的对照就难以说明问题。即使是叙述性研究, 如某种疾病的自然史研究, 疾病的临床表现分析以及对无对照的一组病例防治效

果分析或流行情况调查等。虽然是无对照的叙述性研究，但在分析结果时仍然要与过去的经验或文献报告相比较。又如疾病的鉴别诊断，流行病学的三间分布，在分析结果时还是要采用对比的方法。只不过这种对照是历史对照或称潜在对照而已，常用对照类型有以下几种。

1. 空白(blank)或安慰剂(placebo)对照

“空白”是指对照组不施加任何处理。“安慰剂”是指用一种对自然病程不产生任何影响的制剂作为对照。常用者有淀粉、乳糖等，制成与试验组外形完全一样、气味相同的剂型以便于盲法的实施。二者共同的特点是保证对照组能够保持其固有的自然特征，可清楚地看出处理因素的作用。即凡能改变自然过程者，均可认为是有作用者，是一种论证强度很高的对照方法。

【例 2-1】 葡醛内酯（肝泰乐）、辅酶 Q₁₀ 对急性肝炎的疗效分析。结果见表 2-1。

表 2-1 葡醛内酯、辅酶 Q₁₀ 治疗急性肝炎的疗效分析

药 物	总例数	临床痊愈		好 转		无 效	
		例数	%	例数	%	例数	%
葡醛内酯	40	33	82.5	4	10.0	3	7.5
辅酶 Q ₁₀	40	32	80.0	6	15.0	2	5.0
对照	40	35	87.5	2	5.0	3	7.5

Kruskal-Wallis 秩和检验，Hc=0.650，P=0.7226

结论：不能肯定葡醛内酯与辅酶 Q₁₀ 对急性肝炎的治疗有效。

因本研究的对照是安慰剂，原料为淀粉，做成与葡醛内酯、辅酶 Q₁₀ 外型完全一致的糖衣片以供研究用。微量淀粉肯定不致影响肝炎病例的自然过程，外观完全一致又便于盲法的实施，因而此种对照方法论证强度很高。临床研究多采用安慰剂对照，空白对照很少使用，动物实验则以空白对照为常用。本对照方法的缺点为临床上容易产生医德纠纷。一般人常将其理解为“假药”，误认为欺骗病人而拒绝接受。因此安慰剂的使用应严格掌握适应证。安慰剂对照只适用于：①所研究的疾病是目前尚无特效治疗的疾病；②有明显自愈趋势的疾病；③自然病程复杂多样，个体差异很大或短时间不治疗对预后无明显影响者。凡不符合这三个条件者，不允许采用。

2. 标准(standard)对照

所谓“标准”是指肯定为有效的处理方法，如杀灭病原菌有特效的抗生素或其他抗菌药；退热、抗心衰、降血压、降血糖等有肯定效果的对症处理，是治疗研究中最常用的对照方法。因其既能得到肯定的结果，又符合人道主义的医德要求，是一种很好的对照。但对一些目前尚无特效治疗方法的疾病，则无标准对照可寻，是其不足之处。有人主张将目前惯用的或大家讨论协定的方法也列入标准对照，对此问题有不同意见。基础医学的研究中常采用此对照类型。

3. 实验(experimental)对照

指对照组给以除所研究的实验因素之外的其他伴随的因素以作对照。如用狗做胆总管结扎术，研究胆道梗阻后引起的胆道感染的感染来源与机理的问题。对照狗就必须给予同样的手术过程，只是不做结扎胆总管这一步骤，以排除手术过程的其他因素对结果的影响。这类研究很少能用人类来做研究，大多是在动物实验时采用。有些治疗措施要通过如静脉注射或静

脉滴法来处理, 对照组就应该采用同样的溶剂注射或滴注, 以避免该种溶剂对结果的影响, 此种方法也可算是实验对照。

4. 相互(mutual)对照

两种处理或同一种处理两种不同剂量或不同给药途径之间的相互比较, 可用以比较该两种实验措施的差异。因此种对照方法其用作比较的对照本身作用的效应就不能肯定, 因而用其比较出来的结果, 也就不能肯定。如表 2-1 所示, 如没有安慰剂对照, 就难以肯定葡醛内酯与辅酶 Q₁₀ 两种药物治疗效果是全有效, 还是全无效。有时还会发生以讹传讹的错误。

对照的种类还有多种, 历史对照或潜在对照已介绍于前, 自身前后左右对照将于设计方案类型中介绍。

(三)均衡原则

为了保证对比研究所得的结果准确、可靠, 除要选择合适的论证强度较高的对照外, 非常重要是对比组之间的可比性问题。所谓可比性就是试验组与对照组之间的比较的背景(background)相同或近似的程度。必须是两组差异无显著性, 即均衡性良好的情况下才能排除其他伴随因素的混杂, 保证对比结果的准确与可靠! 常说的齐同对比(control at the same background)就是这个意思。

理论上讲, 任何研究的效应(effect, E)都必须是该研究的处理因素(treatment, T)的处理结果, 即 $T \rightarrow E$, 才有实际价值, 但以人为对象的研究, 很难达到这个要求, 不可避免地会受到同时伴随的其他非处理因素(S)的影响, 而使结果发生偏倚, 变为下列模式:

$$T + S \rightarrow E + S$$

如果两种处理因素(T₁, T₂)或研究因素(T)与安慰剂(O)的对比, 则模式是:

$$T_1(T) + S_1 \rightarrow E_1 + S_1$$

$$T_2(O) + S_2 \rightarrow E_2 + S_2$$

由于个体差异的客观存在, 研究中无法完全避免。即使是动物实验可以用纯种纯系动物来进行, 以减少因动物种系间的差异造成偏差, 但其效果也不是绝对的。以人为研究对象者更无法用此种方法来进行研究。因而非处理因素的影响, 是难以完全避免的, 只能用两组非处理因素分配均等($S_1 = S_2$)的方法来抵销其影响, 此将详述于第八章偏倚与混杂的预防部分。齐同对比是保证结果准确可靠的非常重要的原则。达到均衡性良好的方法有以下几种。

1. 配对(matching)

以已知的对结果可能产生影响的非处理因素为配对条件, 选择与试验背景条件相同者为对照同时进入研究, 即称之为配对研究。本法适用于病例-对照设计方案的研究。多以病例组为准, 再从对照组中找出合适的对象进行比较分析。配对背景条件常用者包括性别、年龄、职业、文化水平、经济收入、营养状况等一般情况及同种疾病、同种手术及病型、病情、病程等疾病情况均相似者。需注意, 凡对结果有影响的非处理因素不能漏掉。同时也必须注意防止要求过严, 以免因过严而给挑选对照造成困难, 有时甚至难以找到符合条件的对照。所研究的处理因素绝不能列入配对条件, 因如果两组之间处理因素的构成无显著差异, 则此种处理因素的实际意义将被掩盖, 即称之为配对过度(over matching)。因此必须充分利用专业知识, 选择好必要的配对条件, 以利工作的顺利进行。

诊断试验研究中一份标本同时用两种方法进行检测, 也是配对处理的另一种方式。前瞻

性研究中的自身前后(左右)对照试验,交叉试验,也是配对的一种形式,可比性均较好。但以人为研究对象的研究,尤其是用病人来做前瞻性研究时,难以遇到有两个条件相似的病人同时入院接受研究,因而限制了配对设计的实施。

2. 分层(stratification)

配对比较是在单个病人中按一定的条件进行配比,分层则是先按对结果会有影响的因素进行分层,将一些条件近似的人群归入一层,再在此层中进行分组接受不同的处理,以求有较好的可比性。如按年龄分层,则老年人与老年人比;中青年与中青年比;儿童与儿童比,分别观察其结果,以防止年龄对结果的影响。其注意点与配对相同,关键是找准分层条件。分层不宜过细,以免实施困难。有人主张研究实施过程不分层,而在分析结果时进行分层分析,以排除混杂与交互作用对结果的影响。

3. 随机化(randomization)

随机化是指抽样调查或分组时,样本来自同一总体,按机会均等的原则而抽样或分组的方法。随机化分组或抽样是保证组间均衡的比较简便的方法。随机化方法详述于第十一章第一节。

(四)重复性(replication)原则

随机化分组与抽样可较好地达到“齐同对比”的目的,以消除非处理因素所造成的偏差。但这不是绝对的,因为齐同对比所能抵销的只是混杂、交互作用与一部份偏倚所造成的误差,并不能消除机遇所造成的误差。理论上讲只要是抽样研究,就一定会因抽样误差造成的机遇存在,机遇只能缩小而不能完全消除,此问题将详述于第八章。临床医师及基础研究人员在进行医学研究的过程中,常会遇到一些意想不到的成功或失败。如牛痘苗的发明就是 Jenner 在做实习医师时,偶然遇到一个挤牛奶女工,说她们挤牛奶女工很少患天花而得到的启发,经过他锲而不舍的研究,终于肯定了牛痘苗预防天花的作用。这是一个很成功的例证。但更多的偶然发现常是机遇的偶然结果,大多经不起实践的考验而被自然淘汰。实践是检验真理的唯一标准,重复就是多次实践。因此,任何研究必须是经多次重复中得到相同或相似的,即可以复制的结果,才是准确可信的,由此可见重复原则的重要性。一般地说,重复的次数越多,即样本数越大,越能反映机遇变异的客观真实情况。但这并不能说样本越大越好。因样本越大,试验条件越难控制,并且对每一个具体受试对象的观察就不可能做得很细。另一方面样本越大,参加研究工作的人数也必须增多,彼此之间在操作、观察、评价等方面都很难做到完全一致,又反而会带来许多误差。从经济上讲,样本越大,经费也一定会越大。因此,盲目追求大样本是不值得提倡的。那么,要多大的样本才算合适?就需要事先做出估计,也就是样本大小的估计问题,样本大小的估计方法根据设计方案而定,共同的要求是:

(1) 规定所允许的 I 型错误(α)与 II 型错误(β)标准,一般多定 α 为 0.05, 0.01; β 为 0.1, 0.2。要求越高,所需样本数越多,但可靠性就越强。

(2) 根据可能出现的结果,确定是单侧检验还是双侧检验。确定正常值研究时如高于或低于正常范围均为不正常者为双侧分布;如只有高于正常或只有低于正常才定为不正常者则为单侧分布。见图 2-1。

如定 $\alpha = 0.05$, 双侧检验 $2\alpha = 0.05$, 即每一侧的 $\alpha = 0.025$ 。单侧检验 $\alpha = 0.05$, β 值

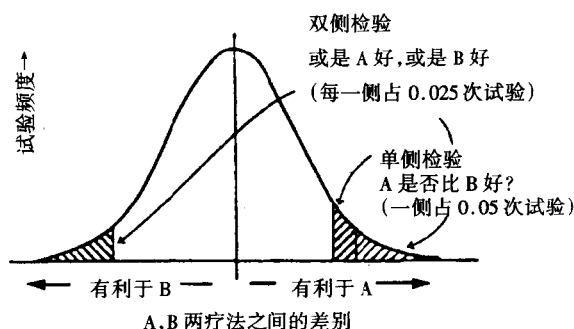


图 2-1 统计学意义的单侧和双侧检验的示意图

$P_0=0.05$, 如只需了解 A 是否优于 B, 而不需回答 B 是否优于 A 者则用单侧检验。对防治效果为 A 优于 B 或者 B 优于 A 均可能者用双侧检验

无单双之分。

(3) 根据资料性质规定必须的标准。

① 计数资料要求定出试验组与对照组总体率(π)。因实际中很难得到 π 值, 一般均用样本率 P_0 , P_1 代表。 P_0 为空白(安慰剂)对照的率, 可从文献中查到, P_1 为试验组的率, 一般也可从文献中查到, 如实在查不到时, 可以人为地规定要求比空白(安慰剂)高出多少方能认为处理是有实际意义者计算出 P_1 。

② 计量资料要求定出, 容许误差 δ 和标准差 α 。

常用的估计方法有查表法与公式计算法。查表法比较简便, 但条件是固定的, 不符合其条件者, 无表可查。后者则计算比较繁琐。常见医学研究的样本大小估计方法详见第十一章第二节。

二、临床科研的设计要点

不同研究类型设计要点有其不同的特点, 本节只能根据临床研究特点及应遵守原则, 就一些共性的问题简介如下。

(一)根据研究目的选择合适的研究对象

基础医学一般多以动物模型进行研究, 取得一定的结果后, 在不违背医德的前提下可进行人体的验证。临床医学中的治疗研究多以病人为研究对象, 观察治疗的效果。诊断性研究多在基础医学研究的基础上, 用一定数量的确诊为某病的病人与健康人或与该病无关的其他病种的病人对象进行对比, 以了解其敏感度、特异度、假阳性率, 假阴性率等为过渡到临床使用提供必要的资料, 此将详述于第四章第三节。病因与致病危险因素的研究则多以健康人为研究对象, 观察其在接受致病因素的暴露后发病的情况。预防医学的研究则多以易感人群为研究对象, 观察其对接受了致病因子的攻击后的预防效果。可见不同领域的研究对象差异很大, 但应该共同注意的事项如下。

1. 诊断要有确实的根据

无论是试验组或对照组诊断都必须确凿无疑。上文已经说明, 一个新的诊断方法进行