

# 医学统计基本知识

—《新医学》特刊—

中山医学院《新医学》编辑出版组

# 毛主席语录

对情况和问题一定要注意到它们的数量方面，要有基本的数量的分析。任何质量都表现为一定的数量，没有数量也就没有质量。我们有许多同志至今不懂得注意事物的数量方面，不懂得注意基本的统计、主要的百分比，不懂得注意决定事物质量的数量界限，一切都是胸中无“数”，结果就不能不犯错误。

应当从客观存在着的实际事物出发，从其中引出规律，作为我们行动的向导。为此目的，就要象马克思所说的详细地占有材料，加以科学地分析的研究。

# 医学统计基本知识

本院卫生学教研组 胡孟璇 黄小兰

在医学统计上长期以来，存在着两条路线的激烈斗争。资产阶级学术权威一面大力鼓吹“统计万能论”，把医学统计说得玄而又玄，利用统计学来扼杀广大工农兵群众的科学成果，垄断科研，把广大革命群众排斥在科学研究大门之外。在无产阶级文化大革命中，广大工农兵对此进行了严肃的批判。但由于刘少奇一类骗子修正主义路线的干扰，在批判了“统计万能论”之后，又出现了“统计无用论”的思潮，认为凡是要求对科研工作进行统计学处理，统统都是机械的和繁琐的框框。其实这种看法也是不全面的。毛主席教导我们说：

**“……不懂得注意基本的统计、主要的百分比，不懂得注意决定事物数量的数量界限，一切都是胸中无‘数’，结果就不能不犯错误。”**因此，正确运用医学统计的知识，可以帮助我们更有成效地进行科学实验，使科学工作得出的结论更符合客观事物的规律性，使科学成果更好地为工农兵服务。为此，我们试把一些医学统计的基本知识，特别是有关临床试验设计、常用统计指标的计算和应用以及抽样误差和显著性测验等问题，介绍给大家，作为开展科研工作的参考。

## 第一节 临床试验设计

一种新的药物或治疗方法是否有效以及疗效如何，这是临床工作者经常需要解决的问题。虽然经过长期的实践经验，其效果终究会被人们认识。但是凭经验来考察判断医疗效果，无论从时间、经济以及患者健康来看，或者从结论的准确性来看，都不是一个好方法。为了对某药物或疗法是否有效以及疗效如何作出比较肯定和精确的判断，必须进行经过科学方法周密设计的临床试验。由于临床试验的对象是人，是有生命有思想而且有个体差异性的生物体，因此，对同一种药物（或刺激），可以产生不同的反应，甚至同一个体在不同时间重复做同一个实验时，各次反应也不尽相同。由于生物个体差异性以及试验的时间、地点、条件的某些偶然因素的影响，便增加了评价疗效的困难和复杂性。因此，需要借助数理统计的指导来计划、安排试验，以便能够有效地获得完整、正确的资料，从而可以从中得出比较可靠的科学结论。从过去大量的临床试验中，我们可以看到不少通过良好设计的临床试验，很快就能对一些医疗方法的疗效做出正确结论的例子；同时也可以看到一些临床试验设计不好而使某些药物或疗法的效果长期得不出正确结论，或者是对一些实际有效的药物或疗法错误地做出无效的结论（假阴性错误），而对一些实际无效的药物或疗法却错误地做出有效的结论（假阳性错误）。由于这种错误造成了很大损失。临床试验设计的目的就是尽可能避免这些错误，并使试验能在更少的人力、物力和时间条件下得出比较肯定和正确的结论。

下面谈谈有关临床试验设计的一些主要方面：

## 对 照

**“有比较才能鉴别。”** 临床试验中设立对照的意义也是为了比较和鉴别。例如有人研究某些草药对慢性支气管炎的疗效，曾用该草药对部分慢性支气管炎患者进行了治疗，有效率为60%，从而得出了该草药对慢性支气管炎有效的结论。从表面来看，这样的结论似乎有道理，但若仔细分析，这样的结论还不是十分可靠的。因为我们知道，象慢性支气管炎这样的疾病，即使不用什么特殊治疗，随着天气的转暖也会有一部分患者的病情自然获得缓解。因此这60%的有效率到底是由于药物的作用，还是其他因素造成的自然缓解？若单独进行治疗组的观察，还不能做出肯定的结论，必须设对照试验。临床上很多疾病如感冒、溃疡病、早期高血压、风湿性关节炎以及肺结核等一类疾病，往往存在着自愈或自然缓解的倾向，即使没有用特殊疗法而只是休息、注意饮食和营养，或随着天气转暖等，也可能使一部分患者的症状逐渐减轻或消失。因此设对照试验是临床研究的重要问题。

在进行对照试验时，必须遵守在一致（齐同）条件下作对比的原则，也就是说，试验治疗组和对照组除疗法不同外，其他可能对结果有影响的因素要尽量一致（绝对一致是办不到的）。只有这样才可以从各组中看出所要研究的那一个因素是否发生了影响，以及影响的大小。有一些临床试验，虽也设有对照，却没有遵守这个原则，因而仍不能得

出正确的结论。例如有人对某种新疗法的试验事先没有作对照的计划，事后为了对照才从以往病历中无选择地收集一些病例或搜集别人的一些试验结果作对照。这种做法是有缺点的。因为两组患者除治疗方法不同外，其年龄、性别以及病情轻重等方面不一定是一致的，而且疗效鉴定标准、观察方法、病历记录等也不可能统一。因此，即使试验治疗组效果比对照组好，也还不能得出这种疗法有效或疗效较高的准确结论，因为效果好可能不是这种疗法的不同，而是疗法以外的其他不同因素影响的结果。

至于对照组病人给以什么处理，在大多数情况下可以给以平时认为最好的疗法或一般习惯用的有效疗法。还应当注意的是，试验组与对照组的人数应尽可能相等，这样在以后做统计学显著性测验（评价两组结果的差别是否明显，详第三节）时，效率才最高，而且计算起来也比较方便。

## 分 组

正确分组是保证在试验条件一致情况下作对比的关键之一，目的是减少试验的误差，增加试验结果的正确性。由于分组不正确以致试验得不出正确结论的例子并不少见。例如有人想比较手术疗法与非手术疗法对溃疡病合并穿孔的疗效，把患者分为两组：那些全身情况好、不伴有休克和穿孔范围小的用非手术疗法处理，而那些全身情况比较严重的、穿孔范围大的或是非手术疗法处理无效的患者，则用手术疗法处理。这样的分组从统计学看显然是不合理的，因为两组患者病情轻重不一致，所造成的误差，可以导致错误的结论，甚至

可能掩盖两种疗法的真正差异（如果存在这种差异的话）。又例如有人研究两种草药对治疗慢性支气管炎的疗效，用两组患者互作对照试验：第一组患者选自农村公社社员，用第一种草药治疗，试验在12~2月间进行；第二组患者选自矿区工人，用第二种草药治疗，试验在3~5月间进行。很明显这样的分组也是不合适的，因为两组患者所在地区、职业以及治疗季节都不相同，而这些因素又足以影响其治疗效果。例如矿区工人慢性支气管炎的发生除一般原因外，还可能职业因素（如粉尘等）的影响。此外，季节与该病的发生也有很大关系。

正确的分组还应遵守“随机化”的原则。“随机化”分组就是使每一个研究对象都有同等机会被分到试验组或对照组中去，而不是凭研究者的主观意愿，把某个患者列入试验组，另一个患者列入对照组。“随机化”分组的方法有许多种，其中用抽签法最简单。如上述第一个例子，我们可以根据研究目的和预先确定研究对象的标准，选出某一类型的溃疡病合并穿孔的患者为研究对象，然后用抽签（一签上写手术组，另一签写非手术组，每一个患者抽到什么组就列入什么组）或掷硬币、用随机数字表或其他“随机化”方法来决定谁是手术组，谁是非手术组。如用掷硬币法，则事先规定掷出正面的患者列入手术组，掷出反面的列入非手术组。如果要使两组患者的性别、年龄也相似的话，则可以把研究对象先按年龄、性别分好组，然后在每组中用抽签或掷硬币的方法确定谁是手术组，谁是非手术组。当然，如果病者的病情不适宜作为研究对象，不在此列。在我们考虑进行医学实验时，首先应以不影响病者的治疗和健康为前提。

在上述第二个例子中，为了排除地区、职业和季节对慢性支气管炎疗效的影响，在工矿区工人患者中用“随机化”方法分成两组，一组用第一种草药治疗，另一组用另一种草药治疗。同样在农村公社社员中也用同一方法分组处理。这样就可以使两组患者条件大致相同而减少误差。

## 样本的大小

在医学研究中如人口普查和一些慢性病普查是对研究对象全部（统计学中称总体）进行研究，这样最能反映事物的客观规律。但在临床试验中是不可能对总体进行研究的。例如研究某种中草药对慢性支气管炎的疗效时，就不可能给全部慢性支气管炎的患者都试验服该药，而只能对其中一部份患者（统计学中称为样本）进行研究，并通过样本推知总体情况。例如在100例慢性支气管炎的病人，证明某种中草药有效后，就可推知这种药对慢性支气管炎病人总体也应有效。但由于人具有生物差异性，因此样本所得结果可能与总体所得结果有差别，这是很容易理解的。一般来说，样本越大作出结论越接近总体的真实情况，也就是说可靠性越大；反之则可靠性就越小。例如研究某种中草药对慢性支气管炎的疗效，以10名患者进行试验，结果5人有效，有效率为50%。显然，若用这个结果代表总体情况是不大可靠的，因为偶然性的影响很大。但是如果该有效率是对100名甚至1,000名患者试验的结果，则上述结论的可靠性便大大增加，就更接近于事物的客观规律。从数理统计算出的结果也是这样。如仅用10例得出有效率为50%时，其总体真正有效率可



能在19~81%的范围内；而用100例得出有效率为50%时，则其真正有效率可能在40~60%的范围内；用1,000例得出有效率为50%时，其真正有效率可能在47~53%之间，波动范围大大缩小（计算方法以后详述）。从上述事实来看，在一定条件下，样本越大所得结论越接近于真实情况，样本过小作结论是不可靠的，甚至是危险的。

“事物都是一分为二的。”样本越大，固然作出结论的可靠性越大。但试验所花费的人力和物力也越多、时间也越长，而且，样本太大往往在选择研究对象时也不易严格加以控制。此外，样本大小还应考虑病人的来源、疗效指标的精确性等因素。因此，究竟样本以多少为宜则必须根据事物的各种矛盾关系全面考虑。在临床对照试验中每组人数以多少为合适，将在以后专门介绍。

## 记录及效果判断

在试验设计时还必须对试验的项目、方法、记录及效果鉴定标准等有详细而明确的规定和说明。特别当研究工作是由许多单位或许多人参加时，更要注意。务必使全体参与研究工作的人员统一思想、统一认识。对各项试验操作方法要准确熟练。在试验中所用的处理方法或剂量应事先规定，并尽可能在试验期间不随意改变。因为假如医生凭主观，随时改变处理，则所得资料将无法整理。但是当患者对处理产生较大副作用或在较长期的临床试验和比较严重的疾病中，患者反应严重或病情恶化时，则一切应从患者健康出发，适当调整剂量，或者应该及时中止试验而改用其他治疗方法。对

这些病例按情况作“无效”或“试验失败”处理。

在评价疗效时，应尽量选用一些能客观反映药物疗效的指标，如体征、X线征、细菌学及其它化验指标等。但在一些疾病主要靠患者自觉症状如疼痛等作评价指标时，则最好用多种指标来验证。同时对患者自觉症状的反映，不但要仔细听取，而且要深入细致的观察，并随时作好记录。应该创造一种气氛让病人不是按照试验者的“暗示”来表态，而是让病人毫无顾虑地准确反映自己的真实感觉。此外，不要等待试验结束才询问患者整个试验过程的感受，因为人的主观感受程度往往受到时间因素的影响。例如，当时感到十分剧烈难忍的疼痛，在疼痛结束后隔一定时间才向患者了解，患者的反映往往就不如当时那么强烈。

进行效果判断时，要忌带主观性和片面性。对“痊愈”、“进步”、“无效”和“恶化”等各项要有明确的指标，使任何人都能根据指标把被试者效果归入上述等级。为了避免某个人鉴定患者的主观偏见，最好在鉴定时组成三人以上的鉴定小组，对每个患者的治疗效果集体讨论，全面分析，共同鉴定。

## 第二节 常用统计指标的计算及其应用

毛主席教导我们：“任何质量都表现为一定的数量，没有数量也就没有质量。我们有许多同志至今不懂得注意事物的数量方面，不懂得注意基本的统计、主要的百分比，不懂得注意决定事物质量的数量界限，一切都是胸中无‘数’，结果就不能不犯错误。”因此，我们对于反映事物的各种指标，要有足够的重视。

在医学研究中搜集来的原始数字资料，一般可分为两大类：

一类是测量资料（即定量指标），以数值的大小来表示某事物的程度。例如：体温、脉搏、血压、血红蛋白、红细胞数、血糖浓度、时间、距离等。

另一类是计数资料（即定性指标或属性指标），是清点数目所得的记录。例如：疾病调查时记录患某种或某些疾病的有多少；治疗疾病时按痊愈、显效、无效、恶化等结果记录疗效；动物实验中记录动物生存或死亡的各有若干等。

从测量计数中得出的原始数据往往不容易看出它们的意义和规律性。通过统计学一定的处理，可以把这些原始数据加以提炼概括，使之成为比较简要易懂而且可以用来对实验结果某些规律性进行适当估价的数值。这些数值（如率、比、平均数、标准差、标准误……）就叫做统计指标。下面先讨论关于计数资料的一些统计指标。

## “率”与“比”的计算和应用

当收集来的原始资料经整理分组后，得到一系列绝对数。例如为了开展心血管防治研究工作，曾在某地区40岁以上的居民中进行了调查。结果在40岁以上的4,000名男子中发现患冠心病的有132人，在40岁以上的1,800名女子中发现患该病的有60人。从这些绝对数我们很难立即看出该地区男女居民的冠心病的患病水平有什么不同。但是如果把这些绝对数转化为相对数，则能更清楚地了解到事物相互间的联系，有助于分析事物的规律性。因此，相对数在医学中很常用。常用的相对数有两种：“率”和“比”。

**一、率：**“率”是一种表示在一定条件下某种现象实际发生的例数与可能发生该现象的总例数的比。“率”也叫频率指标，用来说明某种现象发生的频率。如从上述某地区的调查资料，可计算其频率指标：

$$\text{男子冠心病患病率} = \frac{132}{4,000} \times 100 = 3.3\%$$

$$\text{女子冠心病患病率} = \frac{60}{1,800} \times 100 = 3.3\%$$

根据这两个频率指标就可看出该地区40岁以上的男、女居民的冠心病患病率是一样的，即每100人各约有3.3人患该病。

频率指标的基数可用100、1,000、10,000或100,000表示，分别称为百分率、千分率、万分率或10万分率。医学上常用的频率指标有：

$$\text{某病患率} = \frac{\text{调查某病患人数}}{\text{调查人数}} \times 100$$

(常用于慢性病调查)

$$\text{某病发病率} = \frac{\text{某期间内某病新病例数}^*}{\text{某期间平均人口数}} \times 100$$

(常用于急性病)

$$\text{某病感染率} = \frac{\text{带有某病病原体人数}}{\text{检查人数}} \times 100$$

(阳性率)

$$\text{某病死亡率} = \frac{\text{某地某年某病死亡人数} \times 1000}{\text{该地同年平均人口数}}$$

$$\text{某病病死率} = \frac{\text{死于某病人数}}{\text{该病患人数}} \times 100$$

$$\text{某病治愈率} = \frac{\text{某病治愈人数}}{\text{该病患人数}} \times 100$$

\*新病例数：急性病每发病一次，则为一个新病例；慢性病每年第一次发病算1个新病例

二、比：“比”是一种用来表示事物或现象内部各构成情况的统计指标，又叫构成指标或结构指标。

表1 某地区各年龄组恶性肿瘤死亡率与百分比

年龄组	人口数 (1)	死亡总数 (2)	其中恶性肿瘤死亡数 (3)	恶性肿瘤死亡占总死亡之百分比(4)	年龄组恶性肿瘤死亡率(十万分率)(5)
0~	82,920	138	4	2.9	4.8
20~	46,639	63	12	19.0	25.7
40~	28,161	172	42	24.4	149.1
60及以上	9,370	342	32	9.3	341.5

如表1中第(4)栏,恶性肿瘤死亡占总死亡的百分比就是一种“比”。其计算方法为:  $\frac{\text{恶性肿瘤死亡数}}{\text{死亡总数}} \times 100$

如0~岁组恶性肿瘤死亡占总死亡之百分比

$$= \frac{4}{138} \times 100 = 2.9\%$$

从表1第(4)栏的百分比,可了解到各年龄组中恶性肿瘤死亡在全部死亡中所占的比重。例如0~岁组只占2.9%,而40~岁组却占24.4%。由此可见恶性肿瘤不是儿童青少年死亡的主要原因;而是40~岁组的死亡重要原因,几乎占全部死亡例数的1/4,因此在该年龄组防治恶性肿瘤就很重要。

三、“率”和“比”的应用:“率”和“比”的计算虽甚简单,但他们所表达的意义并不相同,使用时必须十分注意,否则将导致错误的结论。例如从上表第(4)栏中40~岁组之“比”为24.4%,而60岁及以上组之“比”为9.3%,这些数字只能说明恶性肿瘤死亡分别占该两年龄组死亡总人数中的比重,却不能由此得出结论认为60岁及以上的人因恶

性肿瘤死亡的机会反比40岁组的机会少，若要了解究竟那个年龄组死于恶性肿瘤的机会多，则必须计算各年龄组恶性肿瘤死亡率（表1第5栏），即用各年龄组人口数作分母，同年年龄组的恶性肿瘤死亡数为分子。从各年龄组恶性肿瘤死亡率中可看出，事实上因恶性肿瘤死亡的机会是随年龄而增加的，60岁及以上组的恶性肿瘤死亡率比40岁高一倍以上。为什么“比”和“率”这两个指标不一致呢？主要因为“比”是以总数作为100%，即各部份构成之和一定是100%，因此其中某一构成部份的比重增大了，另一部份的比重就会相应地缩小。60岁以上的老人，由于其它原因的死亡率比恶性肿瘤死亡率更高，因此分母中总的死亡人数多了，随之恶性肿瘤死亡所占总死亡数之比便相对地减少了。所以在用“比”来直接比较时，必须注意“比”的这一特点。

在临床研究的文献中，率和比错误应用的例子是比较常见的。如为了分析某病与年龄、职业等因素的关系，临床工作者常以住院或门诊患者的资料进行统计。

表2 沙眼与职业的关系

职 业	沙 眼 就 诊 人 数	百 分 比
工 人	1,792	42.5
学 生	920	21.8
农 民	754	17.9
干 部	320	7.6
其 他	432	10.2
合 计	4,218	100.0

表2是根据某门诊就诊者的资料，分析沙眼与职业的关系。显然表中的相对数都是百分比，只能说明来该门诊看沙眼的以工人为最多，学生次之，而不一定能说明工人患沙眼机会多。因为在总人口中这几个阶层的人数是不同的，来就诊的机会也受许多因素影响，如门诊部所处地区的各阶层人口组成、医疗关系等等。因此这个表的统计数字实际上不能说明沙眼与职业的关系。若要知道二者的关系，最好通过对各职业人群进行普查或抽样调查，计算各职业人群的沙眼患病率。

值得指出，虽然“率”一般可作比较，但在对比时，也要注意在第一节中所谈的资料是否在一致（齐同）条件下进行对比，以及样本例数是否足够等问题。关于两个“率”相比较（如某两种疗法治某病时，两种治愈率的比较）是否有显著差别的问题，将在第三节介绍。

## 平均数和标准差的计算与应用

前面谈过，在医学研究中还有一类指标是用数值大小来表示的，称为测量资料。例如对10名正常男子空腹血糖浓度进行测定，得出10个测量数值为93, 102, 110, 102, 98, 109, 92, 97, 100, 103毫克%（这些数值因人而异，是一种变数，称为变量值）。要从这10个变量值得出一个概括的印象，就必须用一统计指标来表明血糖浓度（或其他现象）的综合特征，最常用的指标就是平均数和标准差。

（一）平均数的计算方法：平均数的计算方法很简单。如欲求上述10个绝对数的平均数，只要把它们相加，求得总和，再除以变量值的个数10，即可得出平均数。



10名正常男子空腹血糖平均浓度

$$\begin{aligned} &= \frac{93 + 102 + 110 + \cdots + 103}{10} \\ &= \frac{1,006}{10} = 100.6 \text{毫克\%} \end{aligned}$$

平均数的算法可用公式表示如下：

$$\begin{aligned} \bar{X} &= \frac{X_1 + X_2 + X_3 + \cdots + X_n}{N} \\ &= \frac{\sum X}{N} \quad (\text{公式 1}) \end{aligned}$$

式中 $\bar{X}$ 为平均数， $X_1, X_2, X_3, \cdots, X_n$ 分别代表各个变量值， $\Sigma$ （读 Sigma）是总和的意思， $\Sigma X$ 即全部变量值的总和， $N$ 为变量值的个数。

当变量值的个数较少时，例如30或50个以下（统计学上称为小样本），可用上面公式1计算平均数。而当变量值的个数较多（即大样本）时，用公式1虽也同样可算出平均数，但较繁琐、费时、而且容易出错，错了也不易核对。有必要在这里介绍一种大样本平均数的简捷算法。兹举例说明：给250名男子注射组织胺后测定其胃游离酸总量，共得250个变量值（原始测定记录资料从略），试求这250个变量值的平均数。第一步先将它们分组并制成频数分配表（见表3第（1）、（2）两栏）。