

# 流行病学常用的统计方法

## Ⅲ. 实验流行病学研究的设计及统计分析

辽宁省卫生防疫站 章扬熙

实验流行病学(Experimental Epidemiology)又称现场实验,是流行病学研究的重要方法之一。实验研究与调查研究不同。其区别在于:处理因素(如疫苗、膳食、药物等)是人为给予的,而且是受控制的;实验对象是选定的,且其分组是随机的。由于具有这些特点,实验研究比调查研究更为严谨,组内误差小,组间可比性强,均衡性好,从而可导出更可靠、更科学的验证推断。实验流行病学研究通常是按随机分配的原则,把实验对象分为两组,一组给处理因素,另一组不给处理因素,然后观察一段时间,比较两组的发病率等效应的差异。这种方法常用于病因研究、预防措施效果以及药效、疗效研究,以验证疾病的病因或因素,评价预防措施、药物及疗法的效应有无或优劣以及效应的大小。

### 一、实验设计要点:

1. 确定实验的目的:因为设计是围绕目的进行的,所以应明确该实验要解决什么问题,是验证病因,还是考核效应,且要具体化。由于现场实验的对象是人群(或动物群),生物的现象千差万别、千变万化,从中验证规律不易,所以通常只有一个“目的”,即单靶。

2. 确定实验现场:实验现场依目的而定,比如,为了评价某种疫苗的预防效果,现场应具备以下条件:

①实验地区或单位的人口稳定,迁出、迁入少,有足够数量,实验人群能代表研究的总体人群。

②预期有较高且稳定的发病率。希望在观察期间内对照组有足够的病人发生,以评价疫苗的效果。不要选择刚流行过的地区,因为这种地区易感者相对少,以后的发病率往往很低。

③现场的卫生防疫、医疗保健的条件较好,有良好的疫情报告制度及诊断水平,以利研究。

④当地行政领导重视,群众乐于接受实验,这也是保证实验研究的重要条件。

3. 确定实验对象及样本大小:

①实验对象:依实验目的而定。比如,为了考核某种疫苗的预防效果,实验对象宜选择健康的且为该疫苗所预防疾病的易感者,同时近期未接受过与本病有关的生物制品或药物者。

②样本大小:以比较实验组与对照组的发病率为研究目的的实验研究,其样本大小当作单侧检验时用公式1计算,当作双侧检验时用公式2计算。若研究目的为比较两组的均数(比如:血压、胆固醇等),其相应的样本大小计算公式为公式3及公式4。如果研究目的为考察疫苗接种前后接种对象血清抗体水平有无上升的变化,其相应的样本大小计算公式为公式5及公式6(表1):

式中 $n$ 为每组所需的样本大小; $\alpha$ 及 $\beta$ 分别为第I类和第II类误差的概率,因为有些统计书中只有双侧P的 $u$ 值表和 $t$ 值表,故以下 $u$ 、 $t$ 值均为双侧, $u_{\alpha}$ 、 $u_{2\alpha}$ 、 $u_{2\beta}$ 、分别为 $\alpha$ 、 $2\alpha$ 、 $2\beta$ 显著性水准的 $u$ 值, $t_{\alpha}$ 、 $t_{2\alpha}$ 、 $t_{2\beta}$ 分别为 $\alpha$ 、 $2\alpha$ 、 $2\beta$ 显著性水准的 $t$ 值(在公式1~4中自由度为 $2n-1$ ,在公式5~6中自由度为 $n-1$ ); $P_1$ 及 $P_2$ 为对两组总体率的估计值; $\delta$ 为研究者提出要求的一个差值,即:当两组均数的差数达到 $\delta$ 时,经检验假设可查出差别; $S$ 为标准差; $S_d$ 为差数的标准差,若用于接种研究则为接种前后差数(滴度倒数的2底对数的差数)的标准差。

例1:为了考核某疫苗的流行病学效果进行现场实验,估计对照组发病率为2%,接种组发病率为0.5%。设 $\alpha=0.05$ , $\beta=0.10$ ,试问两组各需多少人?

按理即使疫苗无效,接种组的发病率也不应明显高于对照组,由于接种组的发病率只低不高,所以通常用单侧的公式1来求样本大小。查 $u$ 值表(或自由度为 $\infty$ 时的 $t$ 值表)知, $u_{0.1}=1.645$ , $u_{0.2}=1.282$ ,原题给出 $P_1=0.002$ , $P_2=0.0005$ ,对率作反正弦平方根的变换,所得数为以度(degree)为单位(在应用电子计算器计算时,用DEG方式),得:

$$n = 1641.6 \left( \frac{1.645 + 1.282}{\sin^{-1}\sqrt{0.002} - \sin^{-1}\sqrt{0.005}} \right)^2 = 6321$$

表1

现场实验样本大小的计算公式

	单侧检验	双侧检验
两组率的比较	$n = 1641.6 \left( \frac{u_{2\alpha} + u_{2\beta}}{\sin^{-1} \sqrt{P_1} - \sin^{-1} \sqrt{P_2}} \right)^2$ (公式1)	$n = 1641.6 \left( \frac{u_{\alpha} + u_{2\beta}}{\sin^{-1} \sqrt{P_1} - \sin^{-1} \sqrt{P_2}} \right)^2$ (公式2)
两组均数的比较	$n = 2 \left[ \frac{(t_{2\alpha} + t_{2\beta}) S}{\delta} \right]^2$ (公式3)	$n = 2 \left[ \frac{(t_{\alpha} + t_{2\beta}) S}{\delta} \right]^2$ (公式4)
配对计量资料的比较	$n = \left[ \frac{(t_{2\alpha} + t_{2\beta}) S_d}{\delta} \right]^2$ (公式5)	$n = \left[ \frac{(t_{\alpha} + t_{2\beta}) S_d}{\delta} \right]^2$ (公式6)

所以，每组需6 321人，两组共12 642人。

**例2：**在进行例1现场实验的同时，为了考察接种疫苗前后血清抗体水平的变化，据既往经验 $S_d=2$ ，设 $\alpha=0.01$ ， $\beta=0.01$ ， $\delta=1$ ，即希望接种后血清抗体比接种前升高1个滴度（即滴度倒数的2底对数的差数），就能查出其效果。问需对多少人进行免疫前、免疫后的血清抗体检测？

从题意知为单侧检验，查自由度为 $\infty$ 时， $t_{0.02} = 2.33$ ，已知 $\delta=1$ ， $S_d=2$ ，把有关数据代入公式5，得：

$$n = \left[ \frac{(2.33 + 2.33) \times 2}{1} \right]^2 = 87$$

所以，约对90人进行接种前、后血清抗体水平的检测即可〔因为87人的样本较大，当自由度为 $n-1=87-1=86$ 时， $t=2.378$ （与自由度为 $\infty$ 时， $t=2.33$ 很接近），再代入公式5，得 $n=90$ ，与所得的87的约数90相同，故无需再尝试用自由度 $n-1=90-1=89$ 的 $t$ 值复算〕。如果考虑有个别人第二次拒绝采血，则可计划检测100人。

4.对照组的设立：有比较，才能鉴别。处理因素对发病有无影响，如果有影响，其影响程度多大，都要通过实验组与对照组的结果比较才能得出。因为，实验组在人为地施加处理因素之后，所产生的效应不一定都归功于处理因素，这里还可能有其他因素的作用，怎样把处理因素与其他因素的作用分离，就需要设立除没有处理因素外，其他因素都与实验组相同的对照，用实验组的效应减去对照组的效应才得到处理因素的效应。可见，设立严格的对照组十分必要。

5.分组的方法：当实验对象选定后，怎样把实验对象分配到实验组与对照组中去，这要按随机的原则进行分配。所谓随机，就是每个实验对象都有同等的机会被分配到实验组或对照组。常用的方法有单纯随

机抽样与系统（机械）抽样。有时对重要的其他因素（如性别、年龄），可先把相同或相似的（如同年龄、同性别）实验对象配成对子后再随机分配。至于两组除处理因素外在其他因素方面分配的结果是否均衡，可进行两组的均衡性比较。

6.双盲法与安慰剂：在实验流行病学研究中，为了防止实验者的主观诱导及使受试者在精神因素方面齐同，常采用双盲法。盲的含意为盲无所知，双盲即实验者与受试者双方都不知道是用什么处理的，只有主持者知道。对照组给予安慰剂，其外形、色味等均与实验组相同，只是不含有效的处理因素成分而已。当实验结束后，主持者对两组的效应进行统计分析。这是一种有效地避免实验过程中产生偏性的方法。但也要指出其应用的局限性及注意事项，即：对于探索性的研究，往往要在实验进程中不断分析、不断处理，盲无所知是不行的；另外，从必须保证受试者的健康与安全考虑，安慰剂有时需在常规法的基础上实施。

7.实验的期限：依实验的目的及处理效应的长短而定。若作近期效果与远期效果观察，可进行阶段性总结。

8.实验的观察指标、方法、表格和设备：对效应的观察指标应简单明确，抓主要矛盾。宜选用客观、特异，最好是定量的指标。判断与诊断要有统一的标准。对疫苗接种效果的考核，不仅要观察血清抗体的免疫学效果，也要观察流行病学现场预防的效果（发病少、病情轻、病程短、合并症或后遗症少、病死率低等）。抗体测定的方法宜选用敏感度及特异度高的先进方法，并进行质量控制。要提出具体的实验观察方法，使处理与观察全面、不漏，准备所需表格及设备备用。

9.拟定整理与分析的计划：如果整理分析应用电

子计算机来完成，观察表格要有电子计算机的编码，同时准备好供整理分析用的程序备用。定群研究所用的整理分析方法，可用于实验流行病学的研究，当然后者还有些其他的方法。

**例3:** 某伤寒防治委员会于某年组织了一个有严格对照的伤寒疫苗现场实验。计划纲要如下：

实验目的为鉴定伤寒疫苗在人群中应用后，对预防伤寒是否有效？实验现场选择在伤寒发病率较高、卫生情况不良、但卫生保健组织较强的地区。实验者按完全随机抽样的方法分为两组，一组接种伤寒疫苗，一组接种痢疾疫苗作为对照。实验疫苗接种工作始于某年3月（伤寒发病最低月份），由经过培训的专业人员进行。接种结束后，将接种组与对照组的有关情况进行均衡性比较，结果如表2。从表2可见，由于完全随机抽样的结果，两组情况基本一致。关于发现病例的方法，除了由当地医疗单位医生发现疑似伤寒患者送传染病院确诊外，每3个月由护士对接种者家庭走访一次，以搜索被遗漏的病例。凡经血培养阳性的伤寒病例，定为伤寒确诊病例；凡血培养阴性但有典型伤寒症状和体征者，定为疑似病例。观察期限从该年6月起至该年末。比较两组的伤寒罹患率，以考核该疫苗的效果。同时比较两组的其他传染病罹患率，以进一步考核两组在接触疾病的机会与易感性方面有无可比性。该实验在正式实验前，于另一地区先作小规模预试验以取得经验，为修订设计提供依据。

表2 伤寒疫苗接种组与对照组的均衡性

	接种组	对照组
应接种人数	12412	12627
完成接种人数	11503	11988
男：女	6204：6208	6274：6363
15岁以下	3911	3877
16~20岁	2520	2463
21~30岁	2801	3028
30岁及以上	3180	3269
工人	3552	3743
职员	946	1017
家庭妇女	1782	1834
居住城市：农村	5498：6554	6004：6633
曾患副伤寒病者	14	17
曾患白喉病者	251	297
在该区居住不满3年者	1318	1383
在该区居住3年以上者	11054	11254

**二、现场实验计数资料的统计分析：**定群研究中所用的统计分析方法，均可用于现场实验。当现场实验用于研究保护因素时，可计算保护率和效果指数，通常先计算实验组与对照组的罹患率并进行 $\chi^2$ 检验。当经 $\chi^2$ 检验证明两组率有差别后，可进一步计算效果指数、保护率及其可信限，所用公式见表3：

式中 $P_1$ 为对照组的罹患率； $P_2$ 为实验组的罹患率；a及b分别为对照组发病数及未发病数；c及d分别为实验组发病数及未发病数；ln为自然对数。

表3 现场实验计数资料的统计指标及可信限的计算公式

统计指标	标准误	95%可信限
保护率 $PR = \frac{P_1 - P_2}{P_1} \times 100\%$ (公式7)	$S(PR) = 100 \sqrt{\frac{1}{P_1^2} \cdot \frac{P_2 Q_2}{n_2} + \frac{P_2^2}{P_1^4} \cdot \frac{P_1 Q_1}{n_1}}$ (%) (公式8)	$PR \pm 1.96S(PR)$ (公式9)
效果指数 $EI = \frac{P_1}{P_2}$ (公式10)	$S(\ln EI) = \sqrt{\frac{1 - P_1^2}{n_1 P_1} + \frac{1 - P_2^2}{n_2 P_2}}$ (公式11)	$\ln^{-1}[\ln EI \pm 1.96S(\ln EI)]$ (公式12)

**例4:** 例3现场实验的结果如表4，试进行统计分析。

先作 $\chi^2$ 检验。检验假设为两组总体率相等。

$$\chi^2 = \frac{(23 \times 11496 - 7 \times 11965)^2 \times 23491}{30 \times 23461 \times 11988 \times 11503} = 7.90$$

自由度 $df = (行 - 1)(列 - 1) = (2 - 1)(2 - 1) = 1$ ，查 $\chi^2$ 值表， $df = 1$ 时， $\chi^2_{0.05} = 3.84$ ， $\chi^2_{0.01} = 6.63$ ，今 $\chi^2 > \chi^2_{0.01}$ ，所以 $P < 0.01$ ，拒绝检验假设，

表4 例3各组伤寒的发病情况

组别	发病情况			罹患率 (/10万)
	发病	未发病	合计	
对照组	23	11965	11988	191.86
实验组	7	11496	11503	60.85

说明两组罹患率有差别，疫苗有效。可进一步求疫苗接种的保护率，用公式7计算，得：

$$PR = \frac{191.86 - 60.85}{191.86} \times 100\% = 68.28\%$$

$$S(PR) (\%) = 100 \sqrt{\frac{1}{0.0019186^2} \times \frac{0.0006085 \times 0.9993915}{11496} + \frac{0.0006085^2}{0.0019186^4} \times \frac{0.0019186 \times 0.9980814}{11988}} = 13.69\%$$

总体保护率的95%可信限为：

$$0.6828 \pm 1.96 \times 0.1369 = 0.4145 \sim 0.9511 = 41.45\% \sim 95.11\%$$

还可用公式10求效果指数，得：

$$EI = \frac{P_1}{P_2} = \frac{191.86}{60.85} = 3.157$$

这说明对照组的伤寒罹患率为接种组的3.157倍。再用公式11与公式12计算效果指数的标准误及总体效果指数的95%可信限。

$$S(1_n EI) = \sqrt{\frac{1 - 0.0019186^2}{11988 \times 0.0019186} + \frac{1 - 0.0006085^2}{11503 \times 0.0006085}} = 0.4317$$

总体效果指数的95%可信限为：

$$\ln^{-1}[\ln 3.157 \pm 1.96 \times 0.4317] = 1.35 \sim 7.36$$

这说明总体效果指数在1.35~7.36范围内的概率为95%。

例5：例3的现场实验，在观察期的6月1日至9月3日（当时伤寒大量发生），对97例诊断为伤寒及其他疾病的住传染病院的病人进行了研究，统计如表5。

表5 实验对象患各种传染病的患者数

病名	实验组	对照组
伤寒	2	13
疑似伤寒	2	3
细菌性痢疾	13	10
阿米巴痢疾	—	1
胃肠炎	2	4
副伤寒B	2	2
钩端螺旋体病	9	13
病毒性肝炎	5	4
其他	6	6

从表中可以看出，两组中伤寒发病数乃是唯一的显著区别。这一方面说明了疫苗有效，同时也说明两组在接触疾病机会与易感性方面是可比的。

三、现场实验计量资料的统计分析：计量资料的统计分析方法取决于分布。常见的分布有正态分布与对数正态分布。关于正态分布资料的统计分析方法，于统计学书中均可查到，兹不赘述。对数正态分布资

再用公式8与公式9计算保护率的标准误及总体保护率的95%可信限。

料的统计分析方法，原则上先把个体值X取对数，这时logx的分布则呈正态分布，就可以按正态分布来进行统计分析。

例6：在考核流脑提纯菌苗的预防效果现场实验中，对100名接种菌苗者检测免疫前及免疫后一个月的抗体滴度结果如表6，试进行统计分析。

首先说明d值的求法，比如，某受试者免疫接种前测定抗体滴度为1:2，接种后一个月测定为1:32

表6 免疫前后抗体滴度资料的统计分析

免疫前后抗体滴度 倒数对数值的差数	频数	频数 × 差数	频数 × 差数平方
d	f	fd	fd <sup>2</sup>
①	②	③ = ①②	④ = ①③
0	3	0	0
1	9	9	9
2	14	28	56
3	21	63	189
4	24	96	384
5	15	75	375
6	8	48	288
7	6	42	294
合计	100	361	1595

取其倒数的2底对数，分别为log<sub>2</sub>2=1，log<sub>2</sub>32=5，差数为5-1=4，即免疫前后抗体滴度倒数对数值的差数d值为4，余类推。

统计各差数的频数于第②栏，将频数乘以差数得第③栏，再求频数乘以差数的平方于第④栏。从第②栏可以看出，d的频数分布近似正态，故可按正态分布进行统计推断，先作t检验。检验假设为总体差数均数为零，求样本差数均数用下式（公式13），即： $\bar{d} = \frac{\sum fd}{n} = \frac{361}{100} = 3.61$ 。再求差数的标准差，用下式（公式14）：

$$S = \sqrt{\frac{\sum d^2 - \frac{(\sum d)^2}{n}}{n-1}} = \sqrt{\frac{1595 - \frac{(361)^2}{100}}{100-1}} = 1.72$$

然后求差数均数的标准误，用下式（公式15）：

$$S_{\bar{d}} = \frac{S}{\sqrt{n}} = \frac{1.72}{\sqrt{100}} = 0.172$$

求t值用下式(公式16):

$$t = \frac{|\bar{d} - 0|}{S_d} = \frac{3.61}{0.172} = 21.02$$

查t值表,当自由度 $df = n - 1 = 100 - 1 = 99$ 时, $t_{0.05} = 1.984$ ,  $t_{0.01} = 2.626$ ,本例 $t = 21.02$ ,  $P < 0.01$ ,拒绝检验假设,说明接种菌苗后抗体水平有提高,平均提高了3.61个滴度(注意,这是滴度倒数对数值的差数均值,差数的几何均数为 $2^{3.61} = 12.21$ )但是,3.61是样本的差数均数,总体的差数均数95%可信限用下式(公式17)求得:

$\bar{d} \pm t_{0.05} \cdot S_d = 3.61 \pm 1.984 \times 0.172 = 3.27 \sim 3.95$ ,这是滴度倒数的对数值结果,还原回来则为:

$$2^{3.27} \sim 2^{3.95} = 9.65 \sim 15.45$$

此值求法应用电子计算器十分方便。比如求 $2^{3.27}$ 时,在计算器上按 $2Y^x 3.27 =$ 即得9.65的结果。这个结果总的说明,接种流脑菌苗一个月后比接种前抗体水平提高了12.21倍;从样本推断总体,总体的流脑菌苗接种一个月后比接种前抗体水平提高的倍数在9.65~15.45范围内的概率为95%。从表6还可看出,如果接种后一个月比接种前抗体有一倍以上增长即为成功,则成功率为 $P = (100 - 3) / 100 \times 100\% = 97\%$ ,总体成功率的95%可信限查百分率可信限表可知为92~99%。

至于有关非正态或非对数正态分布资料以及分布不详资料的统计分析,可采用非参数统计分析的方法。

## 四川省A群脑膜炎奈瑟氏菌脂多糖抗原血清型分布及其流行病学意义

四川省卫生防疫站 许才华 兰纪康 罗隆泽 黎桂兰 祝小平

1984~1985年对我省十地市A群脑膜炎奈瑟氏菌(NM)95株(其中病人71株,带菌者24株),进行了脂多糖(LPS)抗原血清型鉴定及其流行病学意义的调查,分析结果如下:

1. A群NM95株中可分型85株;分为 $L_9$ 、 $L_{10}$ 、 $L_{11}$ 三个型,分型率89.47%,其中 $L_9$ 型9株(9.47%)、 $L_{10}$ 型42株(44.21%)、 $L_{11}$ 型34株(35.79%)。

2. 分析发病率10/10万以上的七个流行县的25株LPS抗原血清型分布表现,不同爆发点的菌株LPS型不同,但同一个爆发点内的菌株则只具有一个血清型。

3. 我省A群NM的LPS抗原血清型的特点是 $L_{10}$ 和 $L_{11}$ 两型占优势(80%),特别是 $L_{11}$ 型显著的高于其他省市。随着发病强度增大, $L_{11}$ 与 $L_{10}$ 型菌株数的比值增大,说明 $L_{11}$ 型菌株对我省流脑发病与流行起着重

要作用。

4. 病人与带菌者的菌型构成分析: $L_9$ 型带菌者(16.67%)高于病人(7.04%), $L_{10}$ 型两者相似(病人43.66%,带菌45.83%), $L_{11}$ 型则病人(42.25%)高于带菌者(11.67%)。

5. 58例流脑病例分析表明,不同LPS菌型其临床表现不同,尤以 $L_{11}$ 病情较重,爆发型9例有7例属 $L_{11}$ 型菌株引起。病人的体温、血象改变亦以 $L_{11}$ 型较为明显。

综上所述A群NM LPS分型,对我省流脑流行病学监测有重要价值。菌型的变迁可为流脑预测提供重要依据。由于LPS分型具有地区分布特点,它可区分不同爆发点,有助于正确地追溯传染源,是流脑流行病学调查的一个有效手段和方法。

## 《流行病学进展》第五卷出版在即

由于发行渠道的不畅通,《流行病学进展》自第一卷发行以来,读者往往购买不到该书。其中重要原因之一是控制印数较紧,不预订就不能保证购得,该书第五卷内容包括分子流行病学研究进展、围产流行病学、计划生育评价方法、危险状态分析法的应用、病毒性肝炎的进展、输血后肝炎流行病学、乙型肝炎病毒消毒进展、人类慢病毒感染、艾滋病、大肠杆菌性腹泻研究及地理流行病学等多篇文章。读者如尚未预订者,请速向北京医科大学流行病学教研组徐宝元同志追订。