

• 系列讲座 •

临床流行病学

第六讲 诊断性试验研究

章扬熙

一、研究设计:进行诊断试验研究的基本方法是与诊断该病的金标准作盲法和同步的比较。

1. 确立黄金标准(gold standard):是指公认的标准诊断方法,它包括病理学检查(组织活检和尸体解剖)、外科手术发现、特殊的影像诊断(如用冠状动脉造影诊断冠心病等)以及长期随访临床观察所获得的结论。

2. 选择研究对象:包括两组,一组是被金标准确诊的病例组,另一组是经金标准证实无该病的人群,即对照组。若研究目的是评价临床诊断价值,对照可来自其他病例,尤其是易与该病混淆的其他病例;若研究目的是评价筛检试验价值,对照应选自筛检人群。病例应包括各种不同临床类型和病期的病例,使之能代表该病患者的总体。

3. 样本大小:评价诊断试验的四格表用表 1 来表示。这里讨论以求得的样本敏感性和特异性能与其总体值相差小于允许误差的较好估计值为目的的研究。

表 1 评价诊断试验的四格表

		金 标 准		
		+ 有病	- 无病	合计
诊断试验	+ 阳性	a 真阳性	b 假阳性	a + b
	- 阴性	c 假阴性	d 真阴性	c + d
合计		a + c	b + d	N

样本大小用下式计算

$$N = \frac{u_a^2 P(1 - P)}{\delta^2} \quad (1)$$

式中 P 为敏感性或特异性, δ 为允许误差, α 为第一类错误的概率, u 值由 u 界表查得。

4. 用被研究的诊断试验同步地测试这些研究对象,为了避免偏性,在试验过程中应实行双盲的原则,即试验操作者和打报告者皆应不知道谁是病人,谁是对照。

5. 应用以下指标评价该诊断试验的价值。

$$\text{敏感性} = a / (a + c) \quad (2)$$

$$\text{特异性} = d / (b + d) \quad (3)$$

$$\text{阳性预测值} = a / (a + b) \quad (4)$$

$$\text{阴性预测值} = d / (c + d) \quad (5)$$

$$\text{总一致性} = (a + d) / N \quad (6)$$

$$\text{调整一致性} = \frac{1}{4} \left(\frac{a}{a+b} + \frac{a}{a+c} + \frac{d}{c+d} + \frac{d}{b+d} \right) \times 100\% \quad (7)$$

$$\text{Youden 指数} = \frac{a}{a+c} + \frac{d}{b+d} - 1 \quad (8)$$

$$\text{患病率} = (a + c) / N \quad (9)$$

$$\text{阳性似然比} = [a / (a + c)] / [b / (b + d)] \quad (10)$$

$$\text{阴性似然比} = [c / (a + c)] / [d / (b + d)] \quad (11)$$

敏感性和特异性是反映一个诊断试验的两个基本指标。敏感性是指真正病人组中经诊断试验查出阳性人数的比例,特异性是指对照组中经诊断试验查出阴性人数的比例。这两个比例自然越大越好。实际上,临床医生还关心用某诊断试验查出的阳性者中有多少是真正的该病患者,在查出的阴性者中有多少不是该病患者,这就是阳性预测值和阴性预测值。阳性预测值受患病率的影响大,当患病率很低时,阳性预测值会变得很小,也就是说在查出的阳性者中,真正的该病患者很少,效能不高,这也就是为什么在筛检时多在高危人群中进行的缘由。

阳性似然比是真阳性率与假阳性率之比,阴性似然比是假阴性率与真阴性率之比。似然比的优点是非常稳定,不受患病率的影响,在选择诊断试验时,选择阳性似然比高的方法。

对于一种诊断方法进行综合评价,常采用总一

致性、调整一致性或 Youden 指数。

例 1:应用测定血清肌酸磷酸激酶(CPK)的阳性结果($\geq 80u$)诊断心肌梗塞,据初步试验资料知其敏感性为 0.90,特异性为 0.85,设 $\alpha = 0.05$, $\delta = 0.08$,求样本大小。

查 u 值表, $u_{0.05} = 1.96$,将有关数值代入公式 1,得:

$$N_1 = \frac{(1.96)^2(0.90)(1-0.90)}{(0.08)^2} = 54$$

$$N_2 = \frac{(1.96)^2(0.85)(1-0.85)}{(0.08)^2} = 77$$

所以,作该诊断性试验需有心肌梗塞病组 54 例,无心肌梗塞病组 77 例。

例 2:研究胎甲球筛检诊断肝癌的价值,以肝穿刺作为金标准,结果如表 2,计算该诊断试验的评价指标。

表 2 胎甲球诊断肝癌的评价

		肝癌(肝穿刺)		合计
		有病	无病	
胎甲球	异常	205(a)	31(b)	236
	正常	29(c)	37(d)	66
合计		234	68	302(N)

$$\text{敏感性} = \frac{205}{234} = 0.8761$$

$$\text{特异性} = \frac{37}{68} = 0.5441$$

$$\text{阳性预测值} = \frac{205}{236} = 0.8686$$

$$\text{阴性预测值} = \frac{37}{66} = 0.5606$$

$$\text{总一致性} = \frac{205 + 37}{302} = 0.8013$$

$$\text{调整一致性} = \frac{1}{4} \left(\frac{205}{236} + \frac{205}{234} + \frac{37}{66} + \frac{37}{68} \right) \times 100\% = 71.24\%$$

$$\text{Youden 指数} = 0.8761 + 0.5441 - 1 = 0.4202$$

$$\text{患病率} = \frac{234}{302} = 0.7748$$

$$\text{阳性似然比} = (205/234)/(31/68) = 1.92$$

$$\text{阴性似然比} = (29/234)/(37/68) = 0.23$$

二、提高诊断试验效率的方法:

1. 若作筛检,宜在高危人群中进行,以提高效率。这个道理是显见的,比如,某诊断试验的敏感性为 0.90,特异性为 0.90,若在 200 人中,患病率为 50%,则阳性预测值为 0.90,效率很好,在查出的 100 个阳性者中,有 90 个是真正的该病患者。若患病率为 1%,则在 1 万人中阳性预测值仅为 0.009,在查出的 1 000 个阳性者中,只有 9 个是真阳性,991 个是假阳性,还要对 991 假阳性者作进一步的确认试验,结果是一场虚惊,可见效率是很低的。

2. 采用联合试验:由于同时具有很高的敏感性和特异性的诊断试验不多,所以往往需要采用联合试验方法提高敏感度和特异度。联合试验的方法有两种:平行试验和系列试验。平行试验是同时做几个试验,只要有一个阳性,即可认定患某病;系列试验是相继做几次试验,当几个试验皆为阳性才认定患某病;前者提高了敏感性,后者提高了特异性。若作几次独立试验,敏感性与特异性分别提高为原值的几次方。有时为了同时获得高敏感性与高特异性,比如,艾滋病 HIV 抗体的检测,可先用酶联免疫吸附试验(ELISA)法作平行试验,再用蛋白印迹法(WB)做系列试验。

三、评价诊断试验的方法学标准:(1)是否与金标准进行同步的盲法对比研究?(2)所研究的病例是否包括了各型病例(轻的、重的、已治和未治的)以及常易与其他病混淆的病例?(3)是否在论文中充分叙述了研究的安排和选择病例的方法?(4)是否在论文中详细地叙述了诊断性试验的方法、步骤,使之能据此重复该研究?(5)是否从样本研究对总体进行了统计推断,已确定了该试验结果的重复性及其临床意义,论文包括研究与应用两部分?(6)试验中规定的正常值是否合理(依据数据分布特征应用适宜的统计方法,第 1 类与第 2 类错误的概率 α 、 β 值合理且符合临床实际)?(7)如将该试验作为同类论文之一来对比,能否判定它是同类试验论文中正确性最高的?(8)试验是否具有效用性,具体应用该试验时病人能够接受,试验无副作用,不影响治疗的效果。

(待续)

(收稿:1996-12-15)