

a clinically useful nomogram in its proper context. Br Med J, 1996, 312:426-429.

7 Osiri M, Suarez-Almazor ME, Wells GA, et al. Number needed to treat (NNT); implication in rheumatology clinical practice. the Annals of the Rheumatic Disease, 2003, 62:316-321.

8 Heller D. Disease impact number and population impact number; population perspectives to measures of risk and benefit. Br Med J, 2000, 321:950-952.

9 康德英,王家良,洪旗,等.证据分级评价系统及其在临床决策分析中的应用.中华流行病学杂志,2004,25:351-354.

10 詹思延,秦颖,李立明. NNT 在常见慢性病防治效果评价中的应用.中国慢性病预防与控制,2000,8:99-102.

11 詹思延,秦颖. NNT:综合评价药物疗效及不良反应新指标.药物不良反应杂志,2001,1:62-65.

(收稿日期:2005-03-07)

(本文编辑:张林东)

优势比在诊断试验评价中的应用

童峰 陈坤 何寒青

诊断试验是辅助临床医师进行疾病诊断的试验方法,评价诊断试验准确度的常用指标,包括灵敏度(真阳性率, TPR)、特异度(真阴性率, TNR)、假阴性率(FNR)、假阳性率(FPR)、约登指数、似然比(LR)、阳性预测值(PPV)、阴性预测值(NPV)等。这些指标各有优缺点。其实,可以利用流行病学中的优势比(OR)概念,将其应用到诊断试验的评价中,与传统的指标相比,OR 在某些方面有其独特的优势和用途,为此拟对诊断试验 OR 的含义及应用做一介绍。

1. 基本原理:评价一项诊断试验的基本方法是对金标准已诊断为患者和非患者的研究对象,采用某种试验方法进行检测,根据测定结果,整理成四格表(表 1),诊断试验 OR 即为病例组中试验阳性的比值与对照组中试验阳性的比值之比^[1]。

$$OR = \frac{a/c}{b/d} = \frac{ad}{bc}$$

表1 单个诊断试验的四格表

诊断试验	金标准		合计
	有病	无病	
阳性	a	b	a + b
阴性	c	d	c + d
合计	a + c	b + d	N

当有某一单元格的频数为 0 时,可以在每格的数值上各加 0.5,再进行计算。OR 表示诊断试验的结果与疾病的联系强度。OR 数值越大表示该诊断试验的判别效果越好;OR = 1,表示该试验无法判别患者与非患者;若 OR < 1,则意味着对照组比病

例组更有可能得到诊断试验的阳性结果。

OR 值的 95% 可信区间(CI)估计,可以采用 Woolf 自然对数转换法。

OR 值的概念还可以应用于诊断试验的 Meta 分析。对诊断试验进行 Meta 分析时,如果只是简单地合并各个研究的灵敏度和特异度作为衡量某一诊断试验准确性的合并效应值并不恰当。因为各研究在报道某项试验的阳性和阴性结果时,采用的截断阈值可能不尽相同,所以研究间灵敏度和特异度的变异,部分可能是由于阈值的变动所致^[2]。而采用 OR 作为合并效应值,就可以克服这个问题。要计算合并 OR 采用 SROC 曲线的方法^[3],首先建立一个线性回归方程: $D = \alpha + \beta S$, 方程中 $D = \log(OR)$, $S = \logit(TPR) + \logit(FPR)$, S 与诊断界点的选择有关,可视为诊断阈值;β 为回归系数,反映 $\log(OR)$ 依赖于阈值(S)的程度;α 为方程的截距项,是 $S = 0$ 时的 $\log(OR)$ 值。如果 β 与 0 的差异无统计学意义,则说明 OR 值不随着诊断阈值的变动而改变,那么各个原始研究的合并 $OR = e^\alpha$ 。要比较两种诊断试验的准确性有无差异,可以在线性回归模型中加入一个哑变量(x),表示不同的诊断试验进行回归分析,那么对自变量 x 的回归系数进行反自然对数转换后,就得到了诊断试验的相对优势比(ROR)。此外,也可以将其他可能导致研究间异质性的协变量(如研究设计类型、患者临床特征、文献发表年份等)加入到模型中,最终得到调整的合并 OR 值。

2. 实例分析:为了系统地评价核磁共振血管造影术(MRA)中的二维模式(2-D MRA)和三维模式(3-D MRA)诊断下肢动脉闭塞的准确性,以传统动脉造影术和动脉内数字减影血管造影术作为参照金标准,利用 Koelmay 等^[4]报道的其中 40 个研究资

作者单位:310031 杭州,浙江大学医学院公共卫生系流行病学与卫生统计学教研室

通讯作者:陈坤, E-mail:ck@zju.edu.cn

料重新进行了 Meta 分析(表 2)。首先,对这两种诊断方式分别建立 SROC 线性回归模型,考虑各研究间样本含量的差异,以各研究方差的倒数 $W_i = (\frac{1}{a_i} + \frac{1}{b_i} + \frac{1}{c_i} + \frac{1}{d_i})^{-1}$ 作为权重,采用加权最小二乘法进行模型参数估计,结果显示:2-D MRA 法的模型为 $D = 4.557 - 0.083S$,3-D MRA 法的模型为 $D = 5.473 - 0.201S$,两个模型中 S 的回归系数均与 0 的差异无统计学意义,所以 2-D MRA 法和 3-D MRA 法的合并 OR 值分别为 95.3(95% CI:53.6~169.4)和 238.2(95% CI:107.7~526.9)。为了进一步比较这两种诊断方式的准确度有无区别,综合各研究数据,在模型中保留变量 S 的基础上,增加诊断方式作为自变量(2-D MRA 法编码为 0,3-D MRA 法编码 1),同时将动脉分支数目(<8 支编码为 0,≥8 支编码 1)、发表年份(<1997 年编码为 0,≥1997 年编码为 1)、疾病患病率(<27% 编码为 0,≥27% 编码为 1)等可能导致研究间异质性的协变量逐个纳入模型进行拟合,结果诊断方式 x_1 ($P=0.02$)和动脉分支数目 x_2 ($P=0.05$)这两个因素进入最终的模型: $D = 4.136 + 1.032x_1 + 0.847x_2 - 0.067S$,而 S 的回归系数仍然与 0 的差异无统计学意义,所以经调整后诊断方式的 $ROR = 2.8$ (95% CI:1.2~6.8),说明 3-D MRA 法诊断下肢动脉闭塞的准确度优于 2-D MRA 法。

3. 讨论:在诊断试验中,由于患者和非患者某种指标的分布在多数情况下并不完全分离,所以同一试验的不同诊断标准有不同的灵敏度和特异度,且两者通常是矛盾的,提高了试验的灵敏度,必然会降低部分的特异度,反之亦然。而 PPV 和 NPV 的高低在很大程度上与受检人群中目标疾病的患病率密切相关。在比较两个诊断试验的准确度时,会出现 A 试验的灵敏度高于 B 试验,但其特异度又低于 B 试验的情形,而利用其他的评价指标,如 PPV、NPV、LR+、LR- 等有时也难以对两个诊断试验做出孰优孰劣的最终结论。这时,利用 OR 常可以对此做出判断。它表示试验结果与疾病之间的联系程度,并且不依赖于受检人群中目标疾病患病率的高低,将灵敏度和特异度综合起来测量诊断试验的准确度,这在进行诊断试验的 Meta 分析时非常有用。但使用 OR 也存在局限性,它无法反映试验的错误率,对 FNR 和 FPR 可能造成的危害后果视为相同。

综上所述,将 OR 的概念拓展到诊断试验中,可以为评价诊断试验提供更多的信息。

表2 MRA 法诊断下肢动脉闭塞的研究数据

编号	文献发表年份	诊断方法(MRA 法)	动脉分支数	真阳性例数	假阳性例数	假阴性例数	真阴性例数	OR 值	观患率(%)
1	1993	2-D	1	9	0	2	4	34.2	73
2	1995	2-D	2	48	7	10	15	10.3	73
3	1996	2-D	7	16	3	2	140	373.3	11
4	1996	3-D	4	24	6	0	90	682.2	20
5	1997	2-D	4	29	8	2	79	143.2	26
6	1997	3-D	4	31	2	0	85	2154.6	26
7	1998	3-D	2	16	1	2	49	392.0	26
8	1995	2-D	2	20	37	0	11	12.6	29
9	1993	2-D	4	25	11	2	55	62.5	29
10	1997	3-D	4	36	5	1	114	820.8	24
11	1997	2-D	6	12	11	0	67	146.7	13
12	1997	3-D	6	12	0	0	78	3925.0	13
13	1998	2-D	1	18	2	1	26	234.0	40
14	1999	3-D	4	23	2	0	69	1306.6	24
15	1999	3-D	3	47	5	0	197	3411.4	19
16	1999	3-D	3	12	7	1	39	66.9	22
17	2000	2-D	7	35	20	8	197	43.1	17
18	2000	3-D	7	35	18	8	204	49.6	17
19	2000	3-D	8	22	3	0	82	1060.7	23
20	1993	2-D	1	15	0	2	9	117.8	65
21	1996	2-D	14	61	4	8	251	478.5	21
22	1998	3-D	6	40	2	4	110	550.0	28
23	1995	2-D	4	36	24	3	107	53.5	23
24	1993	2-D	3	40	3	3	36	160.0	52
25	1998	2-D	12	66	7	0	167	2970.3	28
26	1998	2-D	4	39	4	0	118	2080.3	24
27	1999	3-D	4	12	1	0	80	1341.7	13
28	1999	3-D	11	140	16	0	682	11623.2	17
29	1999	3-D	5	40	1	1	45	1800.0	47
30	1996	2-D	17	132	21	22	209	59.7	40
31	1998	3-D	4	32	0	3	69	1290.7	34
32	1995	2-D	2	22	6	2	58	106.3	27
33	1995	2-D	9	124	7	15	70	82.7	64
34	1996	2-D	13	172	10	3	208	1192.5	45
35	1997	2-D	11	100	3	6	45	250.0	69
36	1998	2-D	14	47	7	0	226	2869.0	17
37	1998	2-D	4	59	2	14	31	65.3	69
38	1999	3-D	10	30	0	2	200	4892.2	14
39	1999	3-D	9	183	22	14	482	286.4	28
40	1999	3-D	5	49	10	3	18	29.4	65

参 考 文 献

- Glas AS, Lijmer JG, Prins MH, et al. The diagnostic odds ratio: a single indicator of test performance. J Clin Epidemiol, 2003, 56: 1129-1135.
- Irwig L, Tosteson AN, Gatsonis C, et al. Guidelines for meta-analyses evaluating diagnostic tests. Ann Intern Med, 1994, 120: 667-676.
- Moses LE, Shapiro D, Littenberg B. Combining independent studies of a diagnostic test into a summary ROC curve: data-analytic approaches and some additional considerations. Stat Med, 1993, 12: 1293-1316.
- Koelmay MJ, Lijmer JG, Stoker J, et al. Magnetic resonance angiography for the evaluation of lower extremity arterial disease: a meta-analysis. JAMA, 2001, 285: 1338-1345.

(收稿日期:2005-04-28)
(本文编辑:张林东)