

诊断试验评价的 Meta 分析方法

刘云霞 王洁贞 庞春坤 刘建平

【摘要】 目的 介绍诊断试验评价的 Meta 分析方法。方法 以 7 篇 B 超诊断脂肪肝的研究为例,运用调整 SROC 法对其进行综合评价。结果 用调整 SROC 法求得 B 超的灵敏度为 0.89(95% CI:0.87~0.92),特异度为 0.94(95% CI:0.92~0.96),Q 值为 0.90。结果表明,B 超的诊断价值较高,可作为脂肪肝诊断的有效方法。结论 运用诊断试验评价的 Meta 分析方法,可对诊断试验的诊断价值做出综合评价,增加研究结论的统计功效,提高结论的可靠性。

【关键词】 Meta 分析; 调整 SROC 法; 诊断试验

The use of Meta-analysis in the evaluation on diagnostic tests LIU Yun-xia*, WANG Jie-zhen, PANG Chun-kun, LIU Jian-ping. *Institute of Evidence Based Medicine, The Second Hospital, School of Medicine, Shandong University, Jinan 250033, China

【Abstract】 Objective To introduce Meta-analysis in evaluating diagnostic tests. **Method** Adjusted SROC method was used in assessing 7 diagnostic tests on fatty liver. **Results** The pooled sensitivity and specificity of type B ultrasonography were 0.89[95% confidence interval(CI):0.87-0.92] and 0.94(95% CI:0.92-0.96) respectively while the Q value was 0.90 by adjusted SROC method. The results indicated that the diagnostic value of type B ultrasonography were high, thus could be regarded as an effective method for fatty liver diagnosis. **Conclusion** Meta-analysis on evaluating diagnostic tests could be used to assess the diagnostic test to increase the power of conclusion, and to improve its reliability.

【Key words】 Meta-analysis; Adjusted SROC method; Diagnostic tests

目前,诊断性试验研究的综合评价方法中应用最广泛的是 Moses 等^[1]和 Kardaun 等^[2]提出的固定效应模型(fixed-effects model):估计 SROC 曲线(summary ROC curve)的方法(SROC 法);而适用于该类研究评价的随机效应模型(random-effects model)仍在研究之中。Walter 等^[3]于 1999 年提出估计调整的 SROC 曲线(adjusting summary ROC curve)的方法(调整 SROC 法),是在 SROC 法的基础上对参考试验误差进行调整后估计 SROC 曲线。以下介绍调整 SROC 法,并应用该方法对 B 超诊断脂肪肝的价值作出科学评价。

基本原理

调整 SROC 法是应用 EM 算法对参考试验误差进行调整^[4],通过迭代估计出病例与非病例数后,再应用 SROC 法拟合得到调整的 SROC 曲线^[1,2,4-7]。

作者单位:250033 济南,山东大学医学院第二医院 山东大学循证医学研究所(刘云霞、王洁贞);山东省皮肤病性病防治研究所(庞春坤);International Health Research Group, Liverpool School of Tropical Medicine, UK(刘建平)

具体步骤:

1. 模型推导的前提是假设参考试验存在误差。将每个诊断试验与参考试验比较的基本数据整理成四格表形式(表 1)。考虑各研究可能存在误判,以 α 表示试验的假阳性率, β 表示试验的假阴性率, θ 表示某病的患病率,这样可把表 1 中的实际数转化成率(期望概率)的形式(表 2)。若以 α_1 和 β_1 分别代表诊断试验的假阳性率和假阴性率, α_2 和 β_2 分别代表参考试验的假阳性率和假阴性率, $\theta_k (k = 1, 2, \dots, k)$ 表示第 k 个诊断试验研究的患病率,则 Meta 分析中第 k 个研究的 4 种结果的期望频数分布如表 3 所示。由表 3 可见,每格均包括两部分,如第 1 格为 $N_k \theta_k (1 - \beta_1)(1 - \beta_2)$ 和 $N_k (1 - \theta_k) \alpha_1 \alpha_2$, 其中 $N_k \theta_k (1 - \beta_1)(1 - \beta_2)$ 表示病例中被诊断试验和参考试验均判为阳性的例数(两试验均正确); $N_k (1 - \theta_k) \alpha_1 \alpha_2$ 表示非病例中被诊断试验和参考试验均判为阳性的例数(两试验均错误)。表 3 中参数 $\theta_k, \beta_1, \beta_2, \alpha_1$ 和 α_2 均未知,运用 EM 算法可求得各参数的最大似然估计。

2. 利用模型的参数估计值推算第 k 个诊断试

验研究的结果,从而校正参考试验误差导致的错判。例如,设 $k=1$,即第一个研究中校正后的第 1 格中的期望阳性数应包括两部分,即校正前的第 1 格的

$$f_1 = \frac{\theta_1(1-\beta_1)(1-\beta_2)}{\theta_1(1-\beta_1)(1-\beta_2)+(1-\theta_1)\alpha_1\alpha_2}$$

和第 2 格的 $f_2 = \frac{\theta_1(1-\beta_1)\beta_2}{\theta_1(1-\beta_1)\beta_2+(1-\theta_1)\alpha_1(1-\alpha_2)}$ 。那么,校正后第 1

格的例数 $n_{111}^* = n_{111} f_1 + n_{121} f_2$,第 2 格的例数 $n_{121}^* = n_{111} + n_{121} - n_{111}^*$ 。同理,可求得校正后的第 3 格和第 4 格的例数 n_{211}^* 和 n_{221}^* 。依此类推,可分别求出 k 个诊断试验研究的 4 种结果的估计值,进而可求得诊断试验的灵敏度和特异度的调整估计值。

3. 应用调整后的灵敏度和特异度,按照 SROC 法绘制调整 SROC 曲线,并求得合并的灵敏度和特异度以及诊断试验效能综合测量的评价指标 Q 值, Q 值用以描述 SROC 曲线接近图形左上角的程度。

表1 诊断试验研究的观察频数

		参考试验	
		+	-
诊断 试验	+	n_{11k}	n_{12k}
	-	n_{21k}	n_{22k}

表2 受试对象的期望概率分布

		参考试验	
		+	-
诊断 试验	+	$\theta_k(1-\beta)$	$(1-\theta_k)\alpha$
	-	$\theta_k\beta$	$(1-\theta_k)(1-\alpha)$

注: θ_k 为第 k 个研究的患病率; $(1-\beta)$:灵敏度; $(1-\alpha)$:特异度

实例分析

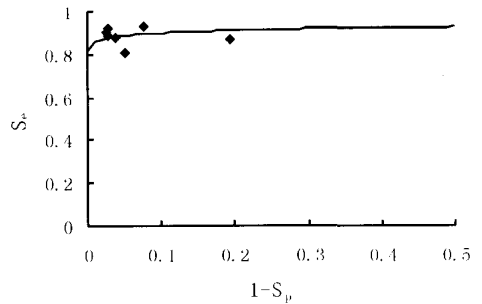
以笔者等^[8]所作“脂肪肝影像学诊断性试验的系统评价”一文中纳入的 7 篇以金标准肝活检为参考试验评价 B 超诊断准确性的研究为例。7 篇资料整理见表 4,为调整 0 观察值对研究分析的影响,对每一观察值均增加 0.5^[1,5,9]。

1. 通过 EM 算法估计参数。根据表 4 中各研究的基本数据求得 B 超的综合灵敏度为 82.90% ($\beta_1=0.171$),特异度为 89.63% ($\alpha_1=0.104$);参考试验(肝活检)的灵敏度和特异度分别为 87.99%

($\beta_2=0.149$)和 85.11% ($\alpha_2=0.120$)。以本例第 1 个研究为例,应用 EM 算法求得 $\theta_1=0.2545$ 。

2. 利用模型的参数估计值推算各诊断试验研究的结果,校正参考试验误差导致的错判。以第 1 个研究为例:将参数估计值代入公式可求得 $f_1=0.951, f_2=0.316$,故 $n_{111}^*=12.5 \times 0.951 + 2.5 \times 0.316=12.7, n_{121}^*=12.5 + 2.5 - 12.7=2.3$ 。同理求得 $n_{211}^*=3.1, n_{221}^*=43.9$ 。如此可求得其他各研究的诊断结果(表 5),进而计算各研究 B 超的灵敏度和特异度。

3. 根据表 5 中的调整参考试验误差后的诊断结果,应用 SROC 法估计调整 SROC 曲线(图 1)。同时应用 SROC 法求得:合并的 B 超灵敏度和特异度及其 95% CI 分别为 0.89 (0.87, 0.92) 和 0.94 (0.92, 0.96), Q 值为 0.90。表明 B 超诊断脂肪肝的准确性较高。



S_e :灵敏度; S_p :特异度; 图中的点代表纳入的研究

图1 B 超诊断脂肪肝的调整 S_e 和调整 $1-S_p$ (与肝活检比较)拟合的 SROC 曲线

讨论

系统评价(systematic review)作为循证医学的主要证据来源之一,亦逐渐引起临床医学界的关注。而 Meta 分析作为系统评价最常用的一种定量分析方法,其方法学也在不断完善。目前,Meta 分析在临床疗效评价、病因学评价方面的应用最多,方法亦最为成熟,应用 Cochrane 中心提供的 Review

表3 第 k 个研究的 4 种结果的期望频数分布

		参 考 试 验	
		+	-
诊断 试验	+	$N_k[\theta_k(1-\beta_1)(1-\beta_2)+(1-\theta_k)\alpha_1\alpha_2]$	$N_k[\theta_k(1-\beta_1)\beta_2+(1-\theta_k)\alpha_1(1-\alpha_2)]$
	-	$N_k[\theta_k\beta_1(1-\beta_2)+(1-\theta_k)(1-\alpha_1)\alpha_2]$	$N_k[\theta_k\beta_1\beta_2+(1-\theta_k)(1-\alpha_1)(1-\alpha_2)]$

注: N_k 表示第 k 个研究的样本量; θ_k 表示第 k 个研究的患病率; $(1-\beta_1)$ 、 $(1-\beta_2)$ 分别为诊断试验和参考试验的灵敏度; $(1-\alpha_1)$ 、 $(1-\alpha_2)$ 分别为诊断试验和参考试验的特异度

表4 B超与肝活检金标准诊断比较

作者	n_{11k}	n_{12k}	n_{21k}	n_{22k}
Foster, et al	12.5	2.5	8.5	38.5
Debondnie, et al	28.5	1.5	7.5	68.5
Yajima, et al	10.5	0.5	2.5	33.5
谭健民, 等	24.5	1.5	4.5	84.5
Savermuttu, et al	45.5	6.5	3.5	31.5
Joseph, et al	32.5	1.5	4.5	13.5
Hulterantz, et al	37.5	19.5	8.5	19.5

表5 参考试验误差调整后的 B 超诊断结果

作者	n_{11k}^*	n_{12k}^*	n_{21k}^*	n_{22k}^*
Foster, et al	12.68	2.32	3.10	43.90
Debondnie, et al	27.91	2.09	3.66	72.34
Yajima, et al	10.08	0.92	1.07	34.93
谭健民, 等	23.49	2.51	2.02	86.98
Savermuttu, et al	49.43	2.57	3.92	31.08
Joseph, et al	33.47	0.53	4.64	13.36
Hulterantz, et al	52.15	4.85	7.91	20.09

Manager(RevMan)软件计算非常方便。但对于诊断性试验研究的综合评价方法仍在不断的探索研究之中;迄今国内尚无诊断性试验的系统评价,国外该方面的评价也不是很多。

SROC法是诊断试验综合评价的常用方法,该方法考虑了诊断性试验研究在临界值不同时对诊断试验效能评价的影响,通过拟合 SROC 曲线综合评价诊断试验的准确性。在诊断试验的 Meta 分析中(包括 SROC 法),通常是假设诊断试验与无误差(完全正确)的参考试验相比较。然而,有研究表明参考试验往往有误差^[10,11],且通常是未知的、不可估计的,其存在会(甚至严重)影响诊断试验准确性的估计。而本文所介绍的调整 SROC 法是在 SROC 法的基础上对参考试验误差进行了调整,可更准确地估计试验的诊断准确性。

调整 SROC 法(SROC 法)是根据最小二乘法(least square method)原理拟合直线回归模型,可进行加权分析和不加权分析。不加权分析即赋予各研究以相同的权重;加权分析即以研究内和研究间方差和的倒数为权重(也可赋予其他权重)。但加权分析可能给予准确性较差的研究以较大的权重,从而可能影响对试验的真实诊断价值的估计,而并不能很好地调整研究间样本量的不同对研究结果的影

响^[12],故本文实例应用中仅给出非加权分析的结果。

参 考 文 献

- 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.
- Kardaun JW, Kardaun OJ. Comparative diagnostic performance of three radiological procedures for the detection of lumbar disk herniation. *Methods Inf Med*, 1990, 29: 12-22.
- Walter SD, Irwig L, Glasziou PP. Meta-analysis of diagnostic tests with imperfect reference standards. *J Clin Epidemiol*, 1999, 52: 943-951.
- Midgette AS, Stukel TA, Littenberg B. A Meta-analytic method for summarizing diagnostic test performances: receiver-operating-characteristic-summary point estimates. *Med Decis Making*, 1993, 13: 253-257.
- Irwig L, Tosteson ANA, Gatsonis C, et al. Guidelines for Meta-analysis evaluating diagnostic tests. *Ann Intern Med*, 1994, 120: 667-676.
- Vamvakas EC. Meta-analyses of studies of the diagnostic accuracy of laboratory tests: a review of the concepts and methods. *Arch Pathol Lab Med*, 1998, 122: 675-686.
- Nelemans PJ, Leiner T, de Vet HCW, et al. Peripheral arterial disease: Meta-analysis of the diagnostic performance of MR angiography. *Radiology*, 2000, 217: 105-114.
- 刘云霞, 王洁贞, 刘建平, 等. 脂肪肝影像学诊断性试验的系统评价. *中国医学科学院学报*, 2004, 26: 580-584.
- Irwig L, Macaskill P, Glasziou P, et al. Meta-analytic methods for diagnostic test accuracy. *J Clin Epidemiol*, 1995, 48: 119-130.
- Hadgu A. The discrepancy in discrepant analysis. *Lancet*, 1996, 348: 592-593.
- Kato I, Santamaria M, De Ruiz PA, et al. Inter-observer variation in cytological and histological diagnoses of cervical neoplasia and its epidemiologic implication. *J Clin Epidemiol*, 1995, 48: 1167-1174.
- Midgette AS, Stukel TA, Littenberg B. A Meta-analytic method for summarizing diagnostic test performances: receiver-operating-characteristic-summary point estimates. *Med Decis Making*, 1993, 13: 253-257.

(收稿日期: 2004-02-26)

(本文编辑: 张林东)