

• 方法评价 •

判断临床意见符合率的指标

赵晓梅¹

王建华²

自从 Kappa 值被作为两名观察者之间调整机会后的一致率的量度以来，人们发现了其解释中的几个难点。这是因为 Kappa 值不仅测量一致率，而且还受到观察者间存在的偏性，以及各范畴的数据的分布，即流行率的复杂的影响。笔者介绍了几个新的指标，这些指标提供了偏性、流行率以及观察一致率的独立的测量，并且得到了一个以这三个指标表达的 Kappa 值公式。在进行一致率研究之间的比较时，单独地报告 Kappa 值常导致错误，研究者还应同时报告偏性和流行率的定量指标。

判断临床意见符合率的指标 Kappa 值已有介绍^[1,2]。假设两位医生阅读相同的 N 张 X 光胸片，并分别做出有无肿瘤的判断，结果可归纳为下列 2×2 表（表 1）。

表 1 两名医生诊断同一批 X 光片结果归纳表

		观察者 A		合计
		是肿瘤	不是	
观察者 B	是肿瘤	a	b	g ₁
	不是	c	d	g ₂
合计		f ₁	f ₂	N

观察一致率 P_o 的比例为 (a+d)/N。该一致率部分可以单独因机会造成。比如，两位医生完全靠掷硬币的方法判断结果，其一致率大约 50%。此即机遇一致率，Cohen 提出的 Kappa 值就是评价校正机遇一致率之后的观察一致率的指标。Zwick 认为 Kappa 值可以表达为下式：

$$Kappa = (P_o - P_e) / (1 - P_e)$$

Pe 即机遇一致率。Pe = [(f₁+g₁)² + (f₂+g₂)²] / 4N² (Scott, 1955) 或 Pe = (f₁g₁ + f₂g₂) / N² (Cohen)。

Kappa 值还可利用以下简化公式计算：

$$Kappa = [N(a+d) - (f_1g_1 + f_2g_2)] / [N^2 - (f_1g_1 + f_2g_2)]$$

一、影响 Kappa 值的两种情况：尽管 Kappa 值作为衡量一致性的简单指标已得到广泛应用，但有

人指出与其结果有关的，使其解释发生困难的两种情况^[3]。

其一，如果 Pe 是大的，则机遇校正过程可以将相对高的 Po 值变为相对低的 Kappa 值，这样 Po 一样的情况下，不同的 Pe 可使 Kappa 值相差一倍多。例如表 2 左右两侧的 Po 均等于 0.85，但左侧 Pe = 0.50，Kappa = 0.7；右侧 Pe = 0.78，Kappa = 0.32。

表 2 Po=0.85, Kappa 值不同的情况

		观察者 A		合计	观察者 A		合计
		是	否		是	否	
观察者 B	是	40	9	49	80	10	90
	否	6	45	51	5	5	10
合计		46	54	100	85	15	100

其二，边缘值（指 f₁ 与 g₁ 和 f₂ 与 g₂）不平衡时比边缘值平衡时有较高的 Kappa 值。例如表 3 左右两侧的 Po 均等于 0.60，但左侧 Kappa = 0.13，右侧 Kappa = 0.26，而右侧边缘合计值一致性较差，

表 3 Po 相同边缘值不同时 Kappa 值

		观察者 A		合计	观察者 A		合计
		是	否		是	否	
观察者 B	是	45	15	60	25	35	60
	否	25	15	40	5	35	40
合计		70	30	100	30	70	100

二、几个新指标：

1. 偏性(bias)和偏性指数(bias index, BI): 如果观察者 A 和 B 评价某研究人群中某疾病发生的频率不同时，则观察者间有偏性。此时不同观察者的边缘值分布不等。偏性指数 BI 的定义是两个观察者评价“是”的比例的差别。BI = (a+b)/N - (a+c)/N = (b-c)/N。当 b=c 时，BI 绝对值最小 (BI=0)；

1 天津中医学院公卫教研室 300193

2 天津医科大学流行病学教研室

b=N 或 c=N 时, BI 绝对值最大 (BI=1)。如果, 并且只有边缘比例相等时, BI=0; 此时 $f_1=g_1, f_2=g_2$ 。比较表 4 左右两侧可见到偏性的影响。左侧的 Kappa=0.17, 两名观察者均判断 60% 的病人为“是”。右侧边缘值比例不同 (分别为 45% 和 75%), Kappa=0.24。左侧 BI=0, 右侧 BI=0.3。

表 4 偏性对 Kappa 值的影响

		观察者 A			观察者 B		
		是	否	合计	是	否	合计
观察者 B	是	40	20	60	40	35	75
	否	20	20	40	5	20	25
合计		60	40	100	45	55	100

因此 Byrt 等人引入了一个新的概念^[1], 调整偏性的 Kappa 值 (bias-adjusted Kappa, BAK)。BAK 定义为表 1 中的 b 和 c 均用其平均值 $m = (b+c) / 2$ 代替后得到的 Kappa 值。表 4 左右两侧的 BAK 均为 0.17。右侧的 Kappa 值高完全是因为观察者间的偏性造成的。这是因为 $Kappa = (Po - Pe) / (1 - Pe) = 1 - (1 - Po) / (1 - Pe)$, Po 固定时, 偏性增加, Pe 下降, (1 - Pe) 增加, $1 / (1 - Pe)$ 下降, Kappa 值上升。

2. 流行率 (prevalence) 和流行率指数 (prevalence index, PI): Kappa 值还受到“是”和“否”的相对概率的影响, 如果我们想估计整个人群的“是”的概率, 最好的估计是 $(a+b) / N$ 和 $(a+c) / N$ 的平均数。同样, “否”的概率的最好估计是 $(c+d) / N$ 和 $(b+d) / N$ 的平均数。“是”概率与“否”概率之间的差定义为流行率指数 PI, $PI = (a-d) / N$, 取值从 -1 (a=0, d=N 时) 到 +1 (a=N, d=0 时); 当“是”和“否”的概率相等时, 即“是”平均流行率为 50% 时, PI=0。

表 5 左右两侧的 Po 均为 0.8, 不一致均为 0.2, 但左侧 Kappa=0.6, PI=0; 右侧 Kappa=0.375, PI=0.6。Kappa 值的差别完全是因为流行率的影响, 因为 PI 值越大, Pe 越大, 则 1 - Pe 减小, $1 / (1 - Pe)$ 增大, Kappa 越小。

表 5 流行率对 Kappa 值的影响

		观察者 A			观察者 B		
		是	否	合计	是	否	合计
观察者 B	是	40	10	50	70	10	80
	否	10	40	50	10	10	20
合计		50	50	100	80	20	100

3. 调整流行率和偏差的 Kappa 值 PABAK

(prevalence adjusted bias-adjusted Kappa): PABAK 为调整了“是”和“否”流行率的差别, 同时也调整了观察者间偏性的观察者间一致率的指标。也即用 $m = (b+c) / 2$ 代替 b 和 c, 用 $n = (a+d) / 2$ 替代 a 和 d 而得到 K 值, 即根据下表计算 PABAK 做为 Kappa 值 (表 6)。

表 6 计算 PABAK 整理归纳表

		观察者 A	
		是	否
观察者 B	是	n	m
	否	m	n

很容易看出, 表 6 $Pe=0.5$, 因此 $PABAK = [(2n/N) - 0.5] / (1 - 0.5) = 2Po - 1$ PABAK 与观察一致率 Po 有线性关系, 其取值范围由 -1 (n=0) 到 +1 (m=0)。观察一致率为 50% 时, PABAK=0。

4. 偏性和流行率的联合影响: 根据任何一个评价一致率的 2x2 表均很容易推出以下几个式子:

$$a = N (Po + PI) / 2$$

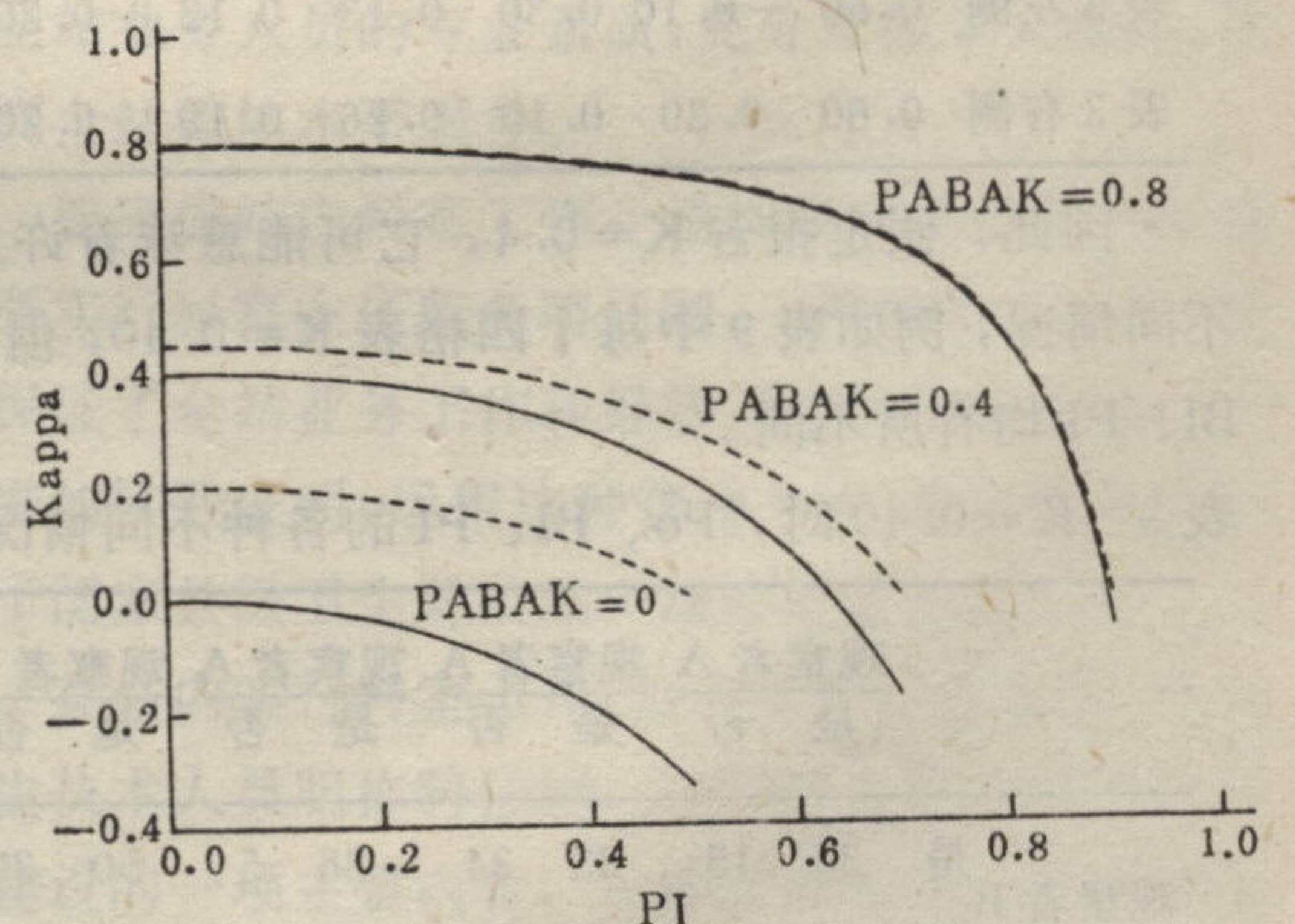
$$b = N (1 - Po + BI) / 2$$

$$c = N (1 - Po - BI) / 2$$

$$d = N (Po - PI) / 2$$

$$K = \frac{PABAK - PI^2 + BI^2}{1 - PI^2 + BI^2} \quad (PABAK = 2Po - 1)$$

由上式可知, 除非 PABAK=1, 否则如果 PI 不变, BI 的绝对值越大, K 越大; 如果 BI 不变, PI 的绝对值越大, K 越小。如果偏性和流行率的影响同时存在, 则根据 PI 和 BI 的相对大小, 所得 K 值可能比 PABAK 大或小。从附图可清楚地看出给定的 PABAK, K 对 PI 和 BI 的依赖。



附图 三种不同 PABAK 时, K 与 PI 的关系 (实线为偏性为零时, 虚线为偏性最大值)

BI 和 PI 可能互相独立地变化, 前者反映了不一致的格子 b 与 c 之间的差别, 后者反映的是一致的格

子 a 与 d 之间的差别。PABAK 或 P_o 的值决定了 BI 和 PI 的最大可能值, 并且对任何 P_o 值, 这些最大值之和为 1。

偏性对小的 K 值比对大的 K 值潜在影响大得多。当 PABAK=0.8 时 ($P_o=0.9$), BI 的最大可能值为 0.1, 所以偏性的影响必定小。但是当 PABAK 接近零时, BI 可高达 0.5, PABAK 和 K 之间的差高达 0.2。另一方面, 流行率对大的 PABAK 或 P_o 比对小的影响大得多。如表 7 的极端情况, $P_o=0.90$, PABAK=0.80, 但 K 是负的。

表 7 流行率极端情况对 Kappa 的影响

		观察者 A	
		是	否
观察者 B	是	90	5
	否	5	0

当既有偏性影响也有流行率的影响时, 往往难以区分开, 也正是这个困难产生了前述难以解释的两种情况。例如回到表 2 的情况, BI 分别为 0.03 和 0.05, 偏性影响很小; 但是流行率的影响 PI 分别为 -0.05 和 0.75, 差别很大; 这导致表 2 右侧的 K 比左侧小得多。表 3 有中等的流行率影响, $PI=0.3, 0.1$, 这被较强的偏性影响 ($BI=-0.1, 0.3$) 所压倒, 导致表 3 右侧 K 值比表 3 左侧大, 表 8 为表 2、表 3 各指数的概括。

表 8 表 2 和表 3 中各指数归纳

	P_o	BI	PI	K	BAK	PABAK
表 2 左侧	0.85	0.03	-0.05	0.70	0.70	0.70
表 2 右侧	0.85	0.05	0.75	0.32	0.31	0.70
表 3 左侧	0.60	-0.10	0.30	0.13	0.12	0.20
表 3 右侧	0.60	0.30	0.10	0.26	0.19	0.20

因此, 假定报告 $K=0.4$, 它可能意味着许多种不同情况, 例如表 9 中每个四格表 $K=0.40$, 但 P_o 、BI、PI 均有所不同。

表 9 $K=0.40$ 时, P_o 、BI、PI 的各种不同情况

		观察者 A		观察者 A		观察者 A		观察者 A	
		是	否	是	否	是	否	是	否
观察者 B	是	35	15	33	34	86	5	50	30
	否	15	35	0	33	5	4	0	20
P_o		0.7		0.66		0.9		0.7	
BI		0		0.34		0		0.3	
PI		0		0		0.82		0.3	

三、观察的阳性一致率和阴性一致率的比例: Cicchetti 和 Feinstein^[5] 提出, 前述两种难以解释的情况可通过计算观察的阳性一致率和阴性一致率的比例来解决。它们分别为

$$P_{pos} = 2a / (N + a - d)$$

$$P_{neg} = 2d / (N - a + d)$$

这两个指标与流行率指标密切相关:

$$P_{pos} = (P_o + PI) / (1 + PI)$$

$$P_{neg} = (P_o - PI) / (1 - PI)$$

但这两个指标不反映偏性的影响, 例如表 4 的 BI 分别为 0 和 0.3, 两侧的 P_{pos} 均为 2/3, P_{neg} 均为 1/2。因此, 独立地考虑偏性和流行率的影响是重要的。

四、报告 2×2 表一致率的一种策略: Byrt 等人^[4] 认为目前尚没有一个综合性的能够满足各种目的的需要的一致率指标, 作为解释 2×2 一致率表格和报告结果的一种策略, 他们建议:

1. 应该利用偏性指数 BI 估计偏性的存在, 同时独立地解释并根据其对 Kappa 的影响来解释。如果确实有较大的偏性存在, 应进一步研究揭示其原因, 此时引用一致率的指标可能是不适当和不必要的。

2. 应当利用 PI 的估计说明流行率的影响。多数情况下偏性影响相对小, 通过流行率影响的分析, 常可对 Kappa 值做出明确的解释。在实际分析中, 利用阳性一致率和阴性一致率指数 P_{pos} 和 P_{neg} 可能会有助于解释流行率的影响。

总之, Cicchetti 和 Feinstein 认为^[5], 只报告一个简单的 Kappa 系数会使解释和比较发生困难。Kappa 值可分解成观察一致率、偏性和流行率影响的组成成分。在一致率研究之间进行比较时, 不宜单独报告 Kappa 值, 而应同时讨论偏性和流行率的影响。

参 考 文 献

- 1 詹绍康. 一致度分析. 中华流行病学杂志, 1991, 12 (3): 182.
- 2 刘正乐. 临床意见分歧的分析和处理. 见: 王家良, 主编. 临床流行病学. 上海: 上海科技出版社, 1990: 135.
- 3 Feinstein AR, Cicchetti DV. High agreement but low Kappa: I. The problems of two paradoxes. J Clin Epidemiol, 1990, 43: 543.
- 4 Byrt E, Bishop J and Carlin JB. Bias, prevalence and Kappa. J Clin Epidemiol, 1993, 46: 423.
- 5 Cicchetti DV, Feinstein AR. High agreement but low Kappa: II. Resolving the paradoxes. J Clin Epidemiol, 1990, 43: 551.

(收稿: 1994-01-18 修回: 1994-04-02)