

· 系列讲座 ·

流行病学与计算机应用

第三讲 率差与归因比例

俞顺章

【导读】文中介绍率差、归因比例、归因危险度的意义及其计算方法。介绍了单一危险因素和多个危险因素,以及分层分析时如何去除混杂因素计算归因危险度的一些计算方法。最后讲解如何以归因危险度图示评估疾病防治的效果。

【关键词】率差; 归因比例; 归因危险度

1. 率差: 未暴露组发病(患病、死亡)率(A 率)减去暴露组发病(患病、死亡)率(B 率)即得率差(rate difference, RD), 见表 1^[1]。

表1 RD 计算模式

项目	暴露组	未暴露组	合计
发病	a_i	b_i	$a_i + b_i$
未发病	c_i	d_i	$c_i + d_i$
合计	A_i	B_i	T_i

当只有一层($i=1$)时: A 率 = a/A ; B 率 = b/B 。计算机程序可用 CPEPI 程序 ATFRAC(attributable and prevented fractions: case-control studies) 计算^[1]。计算举例见表 2。

表2 口服甲糖宁(甲苯磺丁脲)和安慰剂
年龄分层临床试验结果

项目	<55 岁(A 组, 第一层)		≥55 岁(B 组, 第二层)	
	口服甲糖宁	口服安慰剂	口服甲糖宁	口服安慰剂
死亡例数	8	5	22	16
参加人数	106	120	98	85

第一层: A 率 = 7.547%, B 率 = 4.167%
RD = 3.381%, $s_{xRD} = 3.148\%$

95% CI = -3.678% ~ 10.439%

$\chi^2_{RR} = 1.186, P = 0.276$

$\chi^2_{RR-校正} = 0.645, P = 0.422$

$RR_{第一层} = 1.811, s_{x \ln RR} = 0.554$

95% CI = 0.61 ~ 5.37

第二层: A 率 = 22.449%, B 率 = 18.82%

RD = 3.625%, $s_{xRD} = 5.978\%$

95% CI = -9.191% ~ 16.442%

$\chi^2_{RR} = 0.364, P = 0.546$

$\chi^2_{RR-校正} = 0.177, P = 0.674$

$RR_{第二层} = 1.193, s_{x \ln RR} = 0.293$

95% CI = 0.67 ~ 2.12

2 层合计: RD_{合计-MH} = 3.49%, $s_{xRD} = 3.19\%$

95% CI = -2.76% ~ 9.74%

$\chi^2_{RR} = 1.191, P = 0.275$

$\chi^2_{RR-校正} = 0.880, P = 0.348$

$RR_{合计-MH} = 1.326, s_{x \ln RR} = 0.259$

95% CI = 0.80 ~ 2.20

表 2 中两个年龄组口服甲糖宁临床试验结果显示: 服药组与安慰剂对照组在控制病死率率差方面的差异无统计学意义。

2. 人年发病率差: 计算人年发病率差时, 如单独一层, 四格表中各数字都应足够大, 才可作统计分析。如多层时病例数可以是零或小数字(表 3)。计算可用 CPEPI 中 RATE2 (comparison of two rates or proportions) 计算^[1]。

表3 人年发病率差计算模式

项目	暴露组	未暴露组	合计
病例	a_i	b_i	M_i
人年	A_i	B_i	T_i

注: i 指分层数据, 如等于 1, 即不分层四格表数据

暴露组人年发病率 = $\frac{a}{A}$, 非暴露组人年发病率 = $\frac{b}{B}$; 人

年发病率率差: $RD = \frac{a}{A} - \frac{b}{B}$, 如只有一层时: $\chi^2 =$

$\frac{\left(a - \frac{MA}{T}\right)^2}{\frac{MAB}{T^2}}$, 可信限: $RD \pm Z_{\alpha/2} \sqrt{\frac{a}{A^2} + \frac{b}{B^2}}$; 如有多层时: 先求

权重率差, 再作联系卡方和趋势卡方, 亦可求出总卡方减去联系卡方得趋势卡方, 或减去趋势卡方得联系卡方。

权重率差: 各层权重率差平均数 $RD_p = \frac{\sum w_i RD_i}{\sum w_i}$, 权重

系数 $w_i = \frac{A_i^2 B_i^2}{bA_i^2 + aB_i^2}$; $s_{xRD_i} = \frac{1}{\sqrt{\sum (w_i)}}$, 方差 $Var(RD_i) =$

$(s_{xRD_i})^2$; 趋势卡方 $\chi^2_{het} = \sum \frac{(RD_i - RD_{ML})^2}{Var(RD_i)}$, 其中 RD_{ML} 由最大似然法求得, 比 RD_{MH} 略大。95% CI: $RD \pm 1.96(s_{xRD_i})$ 。

3. 归因比例(AF)和归因危险度(AR): AF 可分为暴露归因比和人群归因比。暴露归因比指暴露组人群如果不暴露于该危险因素, 将不会发生该病的比例; 人群归因比指总人口中(暴露和非暴露者)如果不暴露于该因素所减少的病例或死亡数占总人数的比例。暴露归因比: $AF_{exp} = (I_e - I_u)/I_e = (RR - 1)/RR = 1 - 1/RR$ (其中相对危险度 $RR = I_e/I_u$, 即暴露组发病率 I_e 被非暴露组发病率 I_u 除), $I_e - I_u$ 即率差, 率差被暴露组发病率相除, 即暴露归因危险度。如某一食物中毒暴发调查中, 吃盒饭者腹泻发病率为 48.3%,

未吃盒饭者腹泻发病率为 5%，吃盒饭暴露归因比为 (48.3 - 5)/48.3 = 0.8965，即 89.65%。

人群归因危险比 (population attributable proportion, PARP) 即人群中发病率减去非暴露组发病率 (一般称超额发病率)，然后被人群中发病率除。AR_P = (I_p - I_u)/I_p = [P_e(RR - 1)]/[1 + P_e(RR - 1)]。I_p - I_u 即超额发病率。具体计算时先算分子，人群中人数 (N) 乘上暴露于该因素比例 (P_e) 得暴露人群，再乘上暴露组发病率 (I_e)，得到人群中暴露于该因素的总病例数 NPI_e。在 NPI_e 中，有多少人由暴露于该因素得病，即乘以暴露归因比 [(RR - 1)/RR] 得人群归因暴露人数 = NPI_e × [(RR - 1)/RR]。再计算分母，即人群中暴露和非暴露总人数。上述两者相除，亦可得到 PARP，如以百分数表示，即人群归因危险度 (population attributable risk, PAR, 简称 AR)。

$$PARP = AR = AF = \frac{NPI_e \left(\frac{RR - 1}{RR} \right)}{NPI_e + N(1 - P)I_u}$$

$$= \frac{\frac{NPI_e \left(\frac{RR - 1}{RR} \right)}{NI_u}}{\frac{NPI_e + N(1 - P)I_u}{NI_u}} = \frac{P(RR) \left(\frac{RR - 1}{RR} \right)}{P(RR) + (1 - P)}$$

$$= \frac{P(RR - 1)}{1 + P(RR - 1)}$$

人群归因度另一种算法：AF = $\frac{\text{合计率} - \text{B组率}}{\text{合计率}}$ ，合计率 = $\frac{\sum a_i + \sum b_i}{\sum A_i + \sum B_i}$ ，B组率 = $\frac{\sum b_i}{\sum B_i}$ ，如 AF 为负值，可计算其可预防比 (prevented fraction)：PF = $1 - \frac{1}{1 - AF}$ 。

(1) 单一危险因素时^[1,2]：

在病例对照调查中计算 AF：见表 4。

表 4 病例对照调查归因比计算模式

分组	是否暴露于因素	
	是 +	否 -
病例组	a	b
对照组	c	d

$$\text{人群归因危险度 } AF = 1 - \frac{b(c + d)}{d(a + b)}$$

$$\text{暴露归因比 } AF_{\text{exp}} = AF \left(\frac{a + b}{a} \right)$$

$$s_x(AF) = \frac{d(a + b)}{b(c + d)} (1 - AF)^2 \sqrt{\frac{a}{b(a + b)} + \frac{c}{d(c + d)}}$$

如人群中发病率 (P) 已知，计算更精确：

$$AF = 1 - \frac{1}{P + (1 - P) \frac{d(a + b)}{b(c + d)}}$$

$$s_x(AF) =$$

$$\frac{d(a + b)}{b(c + d)} (1 - P)(1 - AF)^2 \sqrt{\frac{a}{b(a + b)} + \frac{c}{d(c + d)}}$$

在队列调查中计算 AR：举例见表 5。

表 5 美国 Framingham 40~59 岁组人群胆固醇与心脏病队列调查结果

血清胆固醇 (mmol/L)	6 年后发生冠心病例数		
	发生	未发生	合计
>5.72	72(a)	684(b)	756
≤5.72	20(c)	553(d)	573
合计	92	1237	1329(T)

$$RR = \frac{(72/756)}{(20/573)} = 2.73, P = \frac{756}{1329} = 0.569, 95\% CI: 1.72 \sim 4.73$$

$$PARP = AR = \frac{P(RR - 1)}{1 + P(RR - 1)} = \frac{0.569(2.73 - 1)}{1 + 0.569(2.73 - 1)} = 0.496$$

$$\text{方差 } Var(AR) = \frac{cT[ad(T - c) + bc^2]}{(a + c)^3(c + d)^3} = \frac{20 \times 1329[72 \times 553 \times (1329 - 20) + (684 \times 20^2)]}{(72 + 20)^3(20 + 553)^3} = 9.506 \times 10^{-3}$$

$$s_x(AR) = \sqrt{Var(PARP)} = 0.0975, 95\% CI: 0.496 \pm 1.96 s_x(AR) = 0.496 \pm 1.96 \times 0.0975 = 0.305 \sim 0.678$$

(2) 多危险因素时 AF 的计算方法^[3,4]：如有多个危险因素可用下式计算或在 logistic 回归计算后调整到百分位数内。

如 OR₁, OR₂, ..., OR_n 时可用 AF 计算公式如下：

$$AF = \frac{[P_0 + P_1(OR_1) + P_2(OR_2)] - 1}{P_0 + P_1(OR_1) + P_2(OR_2)}$$

式中，P₀：人群中无危险因素占总人群百分比；P₁：对照中有第一个危险因素人数百分比；P₂：对照中有第二个危险因素人数百分比，依此类推。

上述公式可以改写为：

$$AF = \frac{\sum_{i=1}^n P_i RR_i - \sum_{i=1}^n P'_i RR_i}{\sum_{i=1}^n P_i RR_i}$$

当危险因素为连续数时 (血压值、大气颗粒浓度) 则变为

$$AF = \frac{\int_{x=0}^m RR(x)P(x) - \int_{x=0}^m RR[x_0 P'(x)]}{\int_{x=0}^m RR(x)P'(x)}$$

式中，RR(x) 为暴露于 x 因素发病或死亡的相对危险度；P(x) 为暴露于 x 因素下人群暴露率；P'(x) 为人群中暴露 x 因素下虚拟分布。对疾病或伤害负担比例，亦即某因素 AR 乘以总负担 (B)，即 AB = AR × B。

进行 AR 评估时应采取的主要步骤：① 选择危险因素，应该选择影响较大，比较接近病因，采取一定政策后可以改变的，其分布也可以获得的；② 选择的因素可以测量，有单位的暴露变量 (如每天饮酒几两或几毫升)，变量常是连续或分类变量；③ 应能收集暴露人群分布的数据；④ 危险因素具有特异性要系统收集流行病学的有关文献；⑤ 要进行不确定性分析。

4. 多层暴露时 AF 的计算^[2,3]：

(1) 多层暴露时要用权重平均：当暴露人群不均匀的时候，应该对每一层暴露归因危险度乘以一个权数 CF (即总病例数中的暴露病例数) 进行调整。

分层数据时:

$$AF_p = \frac{\text{超额发病数}}{\text{总病例数}} = \frac{\text{超额发病数}}{\text{暴露病例数}} \times \frac{\text{暴露病例数}}{\text{总病例数}}$$

$$= \sum (AF_i \times CF_i)$$

$$AF_p = 1 - \left(CF_0 \times \frac{1}{RR_0} + CF_1 \times \frac{1}{RR_1} + CF_2 \times \frac{1}{RR_2} \right)$$

PAR 计算举例见表 6。

表6 126 人调查三种不同暴露水平人群总归因危险度

暴露水平	人数	人群暴露构成比 (%)	RR 值	PAR (%)	
				不正确计算 ^a	正确计算 ^b
低	50	50	1.0	-	-
中	42	30	1.4	10.7	9.5
高	34	20	1.7	12.3	11.1
合计	126	100	-	23.0	20.6

注:^a 不正确计算:暴露水平中:(0.4×0.3)/(1+0.4×0.3)=0.107,暴露水平高:(0.7×0.2)/(1+0.7×0.2)=0.123; ^b 正确计算:暴露水平中:(0.4×0.3)/(1+0.4×0.3+0.7×0.2)=0.095,暴露水平高:(0.7×0.2)/(1+0.4×0.3+0.7×0.2)=0.111

超额发病数 = 占人群中% × RR, 计算式合计:(CF₀ = 50/126, CF₁ = 42/126, CF₃ = 34/126, 1/RR₁ = 10/14, 1/RR₂ = 10/17); AF_p = 1 - [(50/126) × 10 + (42/126) × 10/14 + (34/126) × 10/17] = 1 - 100/126 = 26/126 = 0.06。

(2) 多层数据时还应调整混杂因素: 分层数据中如有混杂因素, 计算 PAR 要予以调整(表 7)。

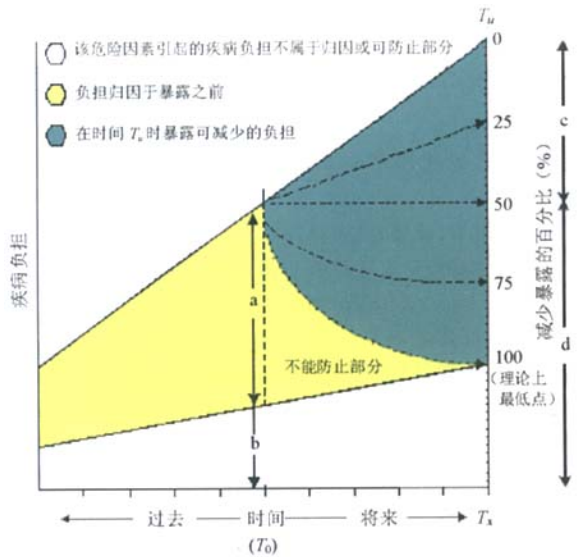
表7 有混杂因素的分层数据时计算 PAR 的正确与不正确方法(数据假设)

混杂因素	暴露因素	人群中构成比 (%)	RR 值	病例数(超额例数)	PAR (%)
-	-	40	1	40	-
-	+	10	1.5 ^a	15(5)	-
+	-	10	2	20	-
+	+	40	3	120(40)	-
合计		100		195(45)	45/195 = 23
不正确计算:					
-		50	1	60	-
+		50	1.5 ^b	135	$\frac{0.5 \times (1.5 - 1)}{1 + 0.5 \times (1.5 - 1)} = 20$
正确计算:					
-			1.5 ^b	55(5)	5/55 = 9.1
+			1.5 ^b	140(40)	40/140 = 28.6
合计 ^c				$55 \times \frac{5}{55} + 140 \times \frac{40}{140} = 23$	

注:如混杂因素(-)和暴露因素(-)时的 RR 值为 1.0;^a 混杂和暴露因素皆(+)比混杂(+)暴露(-):3:2 = 1.5/1 = 1.5;^b 调整 RR 值;^c 正确计算当 RR = 1.5 时 AF_p = AF₁ × CF₁ = (135/195) × (0.5/1.5) = 0.2308

5. AR 的图示: 使用 AR 评估疾病防治效果要慎重。首先, 调查的对象应该有一定的代表性, 调查方法要求无偏倚、正确。这样调查出来的 AR 才能正确估计防治的效果。在分析疾病防治效果时, 还应该考虑原先其他因素对疾病负担所起的作用, 在防治措施前疾病负担的现状, 以及措施后经过一定时间, 去除危险因素百分比后所发挥的作用。具体图

形如图 1^[5]。



注:a = T₀ 时归因于暴露前的疾病负担; b = T₀ 时不属于归因危险因素的疾病负担(由其他危险因素引起); c = T₀ 时暴露减少 50% 后, T_x 时疾病负担; d = T_x 时危险因素减少 50% 后疾病负担

图1 AR 在疾病负担分析时的图解

图 1 中 T₀ 时归因于暴露前的疾病负担 $AF = \frac{a}{a+b}$; T_x 时由于危险因素暴露减少 50% 后疾病负担可防止部分 T₀ = $\frac{c}{c+d}$; 一般由于 T₀ 时减少暴露 50%, T_x 可以防止部分为绿色区域, 点线箭头部分为 T₀ 减少后部分。根据暴露降低的可行性、合理性和费用效益, 选择不同的分布。

6. 结语:

(1) RD 是流行病学调查时常用的指标。本文介绍该指标的统计方法, 显著性检验和可信限计算。

(2) AF、AR 和超额发病(死亡)率在危险度分析和慢性病控制中亦经常使用, 本文阐明了其中的关系。

(3) 介绍了单因素和多因素归因度权重平均数和混杂因素调整的计算方法。

参 考 文 献

- [1] Gahlinger PM, Abranson JH. Computer programs for epidemiologic analysis. Makapuu Medical Press, Hnolulu, Hawaii, 1993: 7-14, 119-125.
- [2] Rothman KJ, Greenland S. Modern epidemiology. 2nd ed. Philadelphia: Lippincott-Raven, 1998.
- [3] Hanley JA. A heuristic approach to the formulas for population attributable fraction. J Epidemiol Community Health, 2001, 55: 508-514.
- [4] Bruzzi P, Green SB, Byar DP, et al. Estimating the population attributable risk for multiple factors using case control data. Am J Epidemiol, 1985, 122: 904-914.
- [5] Mathers CD, Vos T, Hopez AD, et al. National burden of disease studies: a practical guide. Ed 2. 00, World Health Organization Global Program of Evidence for Health Policy. Geneva, 2001: 126-135.

(收稿日期: 2007-04-26)
(本文编辑: 张林东)