

多种危险因素疾病中各因素发病贡献值的测量方法研究

何寒青 陈坤 周曼 童峰

【摘要】 目的 为了评价某一危险因素对个体发病所起的作用,并在多因素暴露情况下比较各个因素对发病的影响大小。方法 运用贝叶斯公式,依据相关流行病学理论,建立由相对危险度数值等来测算单个危险因素对个体发病贡献权重的方法。结果 建立的方法可以用来比较各个危险因素在个体发病中的贡献大小。结论 该测算方法简单方便,结合人群研究及 Meta 分析结果,能够为疾病的病因学研究和临床决策提供一定依据。

【关键词】 危险因素;贝叶斯公式;危险评估;流行病学

A method for measuring the contribution of individual factor to disease caused by multiple risk factors
HE Han-qing*, CHEN Kun, ZHOU Man, TONG Feng. *Department of Epidemic and Health Statistics, School of Medicine, Zhejiang University, Hangzhou 310006, China
Corresponding author: CHEN Kun, Email: ck@zju.edu.cn

【Abstract】 Objective To evaluate the contribution of individual risk factor to a disease on someone with several risk factors. **Methods** A method based on epidemiological theory and Bayes' theorem was established to measure the contribution of individual risk factor, using the relative risk (RR) or odds ratio (OR) value obtained from population-based cohort studies or meta-analysis. **Results** The proportional contribution for individual risk factor to disease in one person can be measured or estimated. **Conclusion** This method can be applied to risk assessment in a patient with more than one risk factor, and the results also contribute to our etiological study and clinical decision-making strategy.

【Key words】 Risk factors; Bayes' theorem; Risk assessment; Epidemiology

在人类疾病的流行病学病因研究中,常常存在多个危险因素同时暴露的情况,这时怎样才能比较各个危险因素对发病的影响大小?在个体发病时,又如何应用人群研究的结果来评价所暴露的每一危险因素对其发病的贡献比重?这一系列问题是疾病多因素理论建立后,在医学活动过程中亟需解决的问题。目前,对于多数肿瘤和慢性病的发病都认为多种危险因素作用的结果,包括遗传因素、身体状况、饮食习惯、生活和行为特征、体育锻炼等因素^[1-5]。然而,当前有关疾病危险因素的研究还只是对于人群发病相对危险度(RR)或归因危险度(AR)的评价,没能对各个因素的作用进行合理比较。尤其在比较单个危险因素对于个体多因素疾病的发病贡献大小时,以往的研究就很难找出有效的

方法。我们根据数理流行病学研究理论,应用概率论中的贝叶斯计算公式,直接利用人群 Meta 分析和队列研究的结果,建立评价单个因素对个体多因素疾病发病贡献大小的权重测算方法。

基本原理

1. 贝叶斯公式和相关参数^[6]:假定 A 指某一事件的发生,它的样本空间(发生条件)B 可以分成独立的事件 B_1, B_2, \dots, B_n , 其中 B 事件存在的概率 $P(B_k) \geq 0$, 则 A 事件发生是条件 B_i 作用的条件概率 $P(B_i/A)$ 为:

$$P(B_i/A) = \frac{P(B_i) \times P(A/B_i)}{\sum_{k=1}^n P(B_k) \times P(A/B_k)}$$

式中 $P(A/B_k)$ 是指存在 B_k 条件下, A 发生的概率。

假设 F_1, F_2, \dots, F_i 为某一疾病的各种独立的危险因素, P_1, P_2, \dots, P_i 为人群中各危险因素所对

作者单位:310006 杭州,浙江大学医学院流行病与卫生统计学教研室(何寒青、陈坤、童峰),理学院数学系(周曼)

通讯作者:陈坤, Email: ck@zju.edu.cn

应的暴露率(P), Ie_i , RD_i , RR_i 和 AR_i 分别为某一危险因素 F_i 暴露下, 人群的发病率(Ie)、超额危险度(RD)、 RR 和 AR (表 1)。

表1 各种相关参数的说明

因素	F_1	F_2	...	F_i
P	P_1	P_2	...	P_i
Ie	Ie_1	Ie_2	...	Ie_i
RD	RD_1	RD_2	...	RD_i
RR	RR_1	RR_2	...	RR_i
AR	AR_1	AR_2	...	AR_i

注: 设 I_0 为非暴露组发病率, $RD = Ie - I_0$, $RR = Ie/I_0$, $AR = (Ie - I_0)/Ie$

2. 计算方法:

单个因素对人群发病的影响权重计算: 假设某一人群中存在两个独立的危险因素 F_1 和 F_2 , P 分别为 P_1 和 P_2 , 该人群相对于普通人群(非暴露人群)患某一疾病的 RD 为 RD_{12} 。对应于上述贝叶斯公式, 这里的危险因素 F 为条件事件 B , P 为 B 事件存在的概率 $P(B)$, RD 就是条件事件作用下事件发生概率项 $P(A/B_k)$, 根据以上计算公式可得:

RD_{12} 中危险因素 F_1 作用的权重(即条件概率):

$$W_1 = \frac{P_1 \times RD_1}{P_1 \times RD_1 + P_2 \times RD_2} \%$$

RD_{12} 中危险因素 F_2 作用的权重(即条件概率):

$$W_2 = \frac{P_2 \times RD_2}{P_1 \times RD_1 + P_2 \times RD_2} \%$$

同样可得, 同时存在多个独立因素 F_1, F_2, \dots, F_n 时, 在该人群发病的 $RD_{12, \dots, n}$ 中暴露因素 F_i 作用的权重为:

$$W_i = \frac{P_i \times RD_i}{\sum_{k=1}^n P_k \times RD_k} \%$$

3. 计算方法在个体多因素疾病发病评价中的应用:

(1) 假设有一个体同时暴露于独立的危险因素 F_1 和 F_2 , 其相对于普通个体(无危险因素暴露个体)发病的 RD 为 RD_{12} 。对个体而言, 暴露于某危险因素的 P 为一固定值, 若暴露于该因素, 则 $P=1$, 未暴露则 $P=0$ 。因此, 上述计算公式中的 P 就以具体数值 1 或 0 来取代。这时相应计算公式为:

该个体发病的 RD_{12} 中因素 F_1 作用所占权重:

$$W_1 = \frac{RD_1}{RD_1 + RD_2} \%$$

该个体发病的 RD_{12} 中因素 F_2 作用所占权重:

$$W_2 = \frac{RD_2}{RD_1 + RD_2} \%$$

同样可得, 个体同时暴露于多个独立因素 F_1, F_2, \dots, F_n 时, 其发病的 $RD_{12, \dots, n}$ 中因素 F_i 作用在所占比重:

$$W_i = \frac{RD_i}{\sum_{k=1}^n RD_k} \%$$

(2) 由于人群的队列研究结果中, 习惯用 RR 来表示某一危险因素引起发病的危险性, 因而我们根据流行病学理论进行换算, 将权重计算公式中的 RD 转化为 RR , 计算过程:

$$\begin{aligned} W_1 &= \frac{RD_1}{RD_1 + RD_2} \% = \frac{Ie_1 - I_0}{(Ie_1 - I_0) + (Ie_2 - I_0)} \% \\ &= \frac{(Ie_1 - I_0)/I_0}{[(Ie_1 - I_0) + (Ie_2 - I_0)]/I_0} \% \\ &= \frac{(RR_1 - 1)}{(RR_1 - 1) + (RR_2 - 1)} \% \end{aligned}$$

$$\begin{aligned} W_2 &= \frac{RD_2}{RD_1 + RD_2} \% = \frac{Ie_2 - I_0}{(Ie_1 - I_0) + (Ie_2 - I_0)} \% \\ &= \frac{(Ie_2 - I_0)/I_0}{[(Ie_1 - I_0) + (Ie_2 - I_0)]/I_0} \% \\ &= \frac{(RR_2 - 1)}{(RR_1 - 1) + (RR_2 - 1)} \% \end{aligned}$$

同样, 单因素 F_i 在多因素暴露下的发病贡献权重:

$$W_i = \frac{(RR_i - 1)}{\sum_{k=1}^n (RR_k - 1)} \%$$

实例分析

1. 多个危险因素的独立作用时各因素发病贡献权重计算: 如某一例肺癌患者, 已知有被动吸烟、肥胖、精神因素、肺癌家族史和室内煤气污染等肺癌发生的危险因素。假定其还暴露与可能的未知 n 危险因素, 记为 F_1, F_2, \dots, F_n 。查阅文献和研究证据, 得到当前最佳的人群肺癌危险因素的 Meta 分析文献^[7], 获得这些危险因素多因素 Meta 中的合并 OR 值(表 2)。对于发病率不高的疾病, 可以把 OR 值作为 RR 值的估计^[8]。按照上述方法, 计算这些危险因素对该患者肺癌发病的贡献权重:

$$\begin{aligned} \sum_{k=1}^n (RR_k - 1) &\approx \sum_{k=1}^n (OR_k - 1) = (1.53 - 1) + (1.30 - 1) \\ &\quad + \dots + (3.41 - 1) + (\sum OR_{未知} - 1) \\ &= 6.37 + (\sum OR_{未知} - 1), \text{ 即} \end{aligned}$$

$$W_{\text{被动吸烟}} = \frac{(RR_{\text{被动吸烟}} - 1)}{\sum_{k=1}^n (RR_k - 1)} \% \approx \frac{(OR_{\text{被动吸烟}} - 1)}{\sum_{k=1}^n (OR_k - 1)} \%$$

$$= \frac{(1.53 - 1)}{6.37 + (\sum OR_{未知} - 1)} \%$$

$$W_{肥胖} = \frac{0.30}{6.37 + (\sum OR_{未知} - 1)} \%$$

$$W_{精神} = \frac{1.57}{6.37 + (\sum OR_{未知} - 1)} \%$$

$$W_{家族史} = \frac{1.56}{6.37 + (\sum OR_{未知} - 1)} \%$$

$$W_{煤烟污染} = \frac{2.41}{6.37 + (\sum OR_{未知} - 1)} \%$$

在未知因素不确定情况下,由各因素发病贡献权重的分母相同,可以计算因素间贡献比,如煤烟污染与肥胖的发病权重贡献比为: $W_{煤烟污染} / W_{肥胖} = 8.03$, $W_{家族史} / W_{肥胖} = 5.20$, 这里各个因素都与肥胖计算发病贡献比,结果见表 2。若假定患者未知危险因素很少或未知危险因素的 OR 值接近于 1,即 $(\sum OR_{未知} - 1) \approx 0$, 可得各因素权重估计值: $\hat{W}_{被动吸烟} = 8.32\%$, $\hat{W}_{肥胖} = 4.71\%$, $\hat{W}_{精神因素} = 24.65\%$, $\hat{W}_{家族史} = 24.49\%$, $\hat{W}_{煤烟污染} = 37.83\%$ 。这里还要计算 AR(%), 结果见表 2。

表2 个体多种因素暴露时各个危险因素对肺癌发生贡献权重测算结果

危险因素	OR 值*	发病贡献比#	权重估计 $\hat{W}(\%)$	AR (%)
被动吸烟	1.53	1.77	8.32	34.64
肥胖因素	1.30	1.00	4.71	23.08
精神因素	2.57	5.23	24.65	61.09
肺癌家族史	2.56	5.20	24.49	60.94
室内煤烟污染	3.41	8.03	37.83	70.67

* 数据引自文献[7]; # 各因素都与肥胖因素发病贡献权重比

2. 因素间存在协同作用时各因素的发病贡献权重计算:为了研究需要,这里就假定某一冠心病事件发生者只是具有冠心病事件的三个危险因素(高血压、吸烟和超重),由相关人群队列研究得知这三个危险因素相应的 RR 值分别为 5.0、1.9 和 5.7,以及这三个因素联合作用时的 RR 值 ($RR_{协同} = 25.8$)^[9]。按照本研究的上述计算公式可得各个因素对疾病发生的贡献权重分别为:

$$W_{高血压} = \frac{(RR_{高} - 1)}{(RR_{高} - 1) + (RR_{吸} - 1) + (RR_{超} - 1)} \%$$

$$= \frac{5.0 - 1}{(5.0 - 1) + (1.9 - 1) + (5.7 - 1)} = 41.7\%$$

$$W_{吸烟} = \frac{(RR_{吸} - 1)}{(RR_{高} - 1) + (RR_{吸} - 1) + (RR_{超} - 1)} \%$$

$$= \frac{1.9 - 1}{(5.0 - 1) + (1.9 - 1) + (5.7 - 1)} = 9.4\%$$

$$W_{超重} = \frac{(RR_{超} - 1)}{(RR_{高} - 1) + (RR_{吸} - 1) + (RR_{超} - 1)} \%$$

$$= \frac{5.7 - 1}{(5.0 - 1) + (1.9 - 1) + (5.7 - 1)} = 48.9\%$$

结果表明高血压、吸烟和超重对于该患者冠心病事件发生的贡献权重分别为 41.7%、9.4% 和 48.9%。在三因素对冠心病事件发生协同作用中,它们所占贡献比例分别为 40.1%、9.0% 和 47.0%, 合计为 96.1% (表 3)。

表3 多种危险因素的个体冠心病事件发生的各因素贡献权重测算结果

危险因素	RR 值*	AR 值 (%)	贡献权重 W (%)	发病危险性的贡献率 (%)
高血压	5.0	80.0	41.7	40.1(41.7×96.1)
吸烟	1.9	47.4	9.4	9.0(9.4×96.1)
超重	5.7	82.5	48.9	47.0(48.9×96.1)
高血压+吸烟+超重	25.8	96.1	-	-

* 数据引自文献[9]

讨 论

1. 多种危险因素疾病中各因素发病贡献值测算方法的应用:

(1) 比较各因素对个体发病贡献权重大小:对于多危险因素的疾病,该方法可以利用人群研究和 Meta 分析的结果,计算各危险因素对个体发病的贡献权重,结果能用来对各个危险因素的贡献权重进行比较。以往常用的 AR 评价仅在人群研究中的应用,只能表明暴露后人群所增加的发病超额危险比例^[10,11]。如本文实例分析 1 中,被动吸烟、肥胖……室内煤烟污染等因素对于人群肺癌发生的 AR 分别为 34.64%、23.08%……70.67%, 合计值超过 100%, 无法合理解释危险因素的发病中贡献大小,也不能直接用于个人中比较。而根据上述计算方法,在无法确定其他未知危险因素情况下,通过计算因素间的发病贡献比,可对因素间比较并排出顺位为室内煤烟污染 > 精神因素 > 肺癌家族史 > 被动吸烟 > 肥胖,即室内煤烟污染对于该患者发病贡献权重最高,是肥胖因素的 8.03 倍。假定无其他未知因素或未知因素影响不大,即认为个体只暴露于这五个危险因素情况下,可以得出它们对于肺癌发生的贡献权重的具体数值的估计,如室内煤烟污染的肺癌发病贡献权重为 37.83%, 而肥胖为 4.71%, 便能更加直观地进行比较。

(2) 测算单个因素在多因素协同作用疾病的发病中贡献权重大小:多因素疾病的发生过程中,危险因素间一般都不是独立作用,往往是相互影响或存

在协同作用^[12]。运用该方法进行测算,还能得出存在因素间协同作用的疾病发生中各因素贡献权重值。如果只是依据 RR 值大小来比较,结果一般都不太合理,因为各因素暴露的 RR 累加值往往与它们协同作用的 RR 值不等。如本文实例分析 2 中,高血压、吸烟和超重三个因素单独作用的 RR 值合计为 12.6,与实际协同作用时的 RR 值不等($RR = 25.8$),表明存在因素间协同作用。而应用该方法对各因素的发病贡献权重计算,这三个因素在它们协同作用冠心病事件发生中的贡献权重分别为 40.1%、9.0% 和 47.0%,合计为 96.1%,与协同作用后的 AR 值($AR = 96.1\%$)相等。即存在协同作用下冠心病发生时,高血压和超重因素所占发病贡献权重较大,而吸烟因素最低。因而,存在多种危险因素协同作用时,可以用该计算方法所得权重来比较和分析各个危险因素各自对发病的贡献权重大小。

2. 多种危险因素疾病中各因素发病贡献值测算法的特点:计算方法通过对危险因素的各自发病贡献权重系数的合理分配,克服了以往方法的不足,结果能够用来比较各危险因素对疾病发生的贡献值大小。对于个体暴露于多种危险因素时单个因素权重的计算,存在未知因素情况下也可以比较各个因素在发病中的贡献大小,用来指导临床实践决策。同时,该测算方法简单方便,还能适合用于临床流行病学研究,有助于病因的寻找和预防措施的有效开展。

由于生物医学模式的特殊性,这一计算方法应用还有一定的条件的限制。获得当前最佳证据也即人群研究和 Meta 分析结果,是该计算方法应用的关键,这样才能将群体流行病学研究的结论很好地用于指导个体临床实践。疾病是一个复杂的过程,某些危险因素当前还未知,因而在发病贡献权重大小计算和比较时还需慎重,并结合专业知识进行。

参 考 文 献

- 1 Li X, Hemminki K. Parental cancer as a risk factor for bone cancer: a nation-wide study from Sweden. *J Clin Epidemiol*, 2002, 55:111-114.
- 2 Zhu BT. Medical hypothesis: hyperhomocysteinemia is a risk factor for estrogen-induced hormonal cancer. *Int J Oncol*, 2003, 22:499-508.
- 3 Goldenberg D, Golz A, Joachims HZ. The beverage mate: a risk factor for cancer of the head and neck. *Head-Neck*, 2003, 25:595-601.
- 4 Ahluwalia IB, Mack KA, Murphy W, et al. State-specific prevalence of selected chronic disease-related characteristics-Behavioral Risk Factor Surveillance System, 2001. *MMWR Surveill Summ*, 2003, 52:1-80.
- 5 Nguyen V, Sadler GR, Shi M, et al. Cancer risk factor assessment among Chamorro women. *J Cancer Educ*, 2003, 18:100-106.
- 6 范大茵,陈永华,主编.概率论与数理统计.第 2 版.杭州:浙江大学出版社,2003.25.
- 7 么鸿雁,施倡元.中国人群肺癌发病危险因素的 Meta 分析.中华流行病学杂志,2003,24:45-49.
- 8 李立明,主编.流行病学.第 5 版,北京:人民卫生出版社,2004.94-95.
- 9 吴兆苏,姚崇华,赵冬,等.11 省市队列人群心血管病发病前瞻性研究:个体危险因素聚集与心血管病发病的关系.中华心血管病杂志,2001,29:246-250.
- 10 李立明,主编.流行病学.第 4 版.北京:人民卫生出版社,1999.90-91.
- 11 Fioretti F, Tavani A, Gallus S, et al. Case-control study of thyroid cancer in Northern Italy: attributable risk. *Int J Epidemiol*, 1999, 28:626-630.
- 12 Au WW. Life style factors and acquired susceptibility to environmental disease. *Int J Hyg Environ Health*, 2001, 204:17-22.

(收稿日期:2004-10-28)

(本文编辑:张林东)