

叉生分析在基因-环境交互作用研究中的应用与意义

王培桦 沈洪兵 陈峰 赵金扣

【摘要】 目的 探讨叉生分析在基因-环境交互作用研究中的作用和意义。方法 以口服避孕药和 Leiden 因子 V 基因突变与静脉血栓栓塞的病例对照研究为例,分析 2×4 叉生分析表核心信息,并与分层分析、单纯病例研究相比较。结果 对基因-环境交互作用的分析依赖于相加或相乘模型的选择。基于相乘模型交互作用的叉生分析 OR 值为 1.35 ($P > 0.05$),其结果与分层分析和单纯病例研究一致;叉生分析还可计算相加模型的各项指标:交互作用指数为 3.90、交互作用归因比为 72.24%、交互作用超额相对危险度为 25.08 ($P > 0.05$)。结论 叉生分析在基因-环境交互作用分析中可进一步应用。

【关键词】 叉生分析;交互作用;分层分析;单纯病例研究

Study on the significance and application of crossover analysis in assessing gene-environmental interaction

WANG Pei-hua*, SHEN Hong-bing, CHEN Feng, ZHAO Jin-kou. *Department of Epidemiology and Statistics, School of Public Health, Nanjing Medical University, Nanjing 210029, China

Corresponding author: ZHAO Jin-kou, Jiangsu Provincial Center for Disease Control and Prevention, Nanjing 210009, China. Email: jkzhao@hotmail.com

【Abstract】 **Objective** To examine the significance of crossover analysis in gene-environmental interaction studies. **Methods** Through elaboration of a case-control study on the increased risk of venous thrombosis in oral-conceptive users who were carriers of factor V Leiden mutation, core information from 2×4 crossover table were analyzed and compared with stratified analysis and 'case only' study. **Results** Different models (additive or multiplicative) in analyzing gene-environmental interaction yielded different results. The result of interaction based on multiplicative model was 1.35 ($P > 0.05$), compatible with that of stratified analysis and case only study. Calculated by crossover analysis based on additive model, synergy index S(S), attributable proportion of interaction (AP) and relative excess risk of interaction (RERI) appeared to be 3.90, 72.24%, 25.08 ($P > 0.05$) respectively. **Conclusion** Crossover analysis should further be applied in gene-environmental interaction studies.

【Key words】 Crossover analysis; Interaction; Stratified analysis; Case only study

除很少一些疾病纯粹由遗传(如先天性遗传病)或环境因素(如车祸)引起外,绝大多数疾病是遗传和环境因素共同作用的结果,尤其是对于复杂性状的慢性病,基因和环境因素的交互作用在其中扮演着十分重要的“角色”。近年来,在研究基因-环境交互作用的方法学上进展很快,有分层分析方法(病例对照研究或单纯病例研究),有采用多因素 logistic

回归模型分析方法以及广义相对危险度模型方法(generalized relative risk model)^[1]。我们从病例对照研究方法中的基本流行病学单元—— 2×4 叉生分析表着手,阐明其在分析基因-环境交互作用中的作用,并试图从深层次挖掘其与分层分析、单纯病例研究的联系,同时给予例证。

基本原理

1. 叉生分析表及其核心信息: 2×4 叉生分析表是基因与环境因素相互作用研究中的基本研究单元,它表示遗传(G)、环境因素(E)均为二分类变量而组成的四种可能的组合表。同时暴露于两因素相对于同时不暴露于两因素的危险性(比值比, OR)

基金项目:江苏省医学“135 工程”重点学科资助项目(流行病学 02-02)、江苏省医学重点人才培养资助项目(RC2003090)

作者单位:210029 南京医科大学公共卫生学院流行病与卫生统计学系(王培桦、沈洪兵、陈峰);江苏省疾病预防控制中心慢性非传染病学科(赵金扣)

通讯作者:赵金扣 Email: jkzhao@hotmail.com

记为 OR_{eg} (简记为 A);单独暴露于遗传或环境因素的危险性分别记为 OR_g 、 OR_e (分别简记为 B、C);两因素均未暴露的病例和对照组作为共同参比组, $OR = 1$ 。参见表 1。

表1 遗传(G)与环境(E)因素作用 2×4 叉生分析

G	E	病例数	对照人数	OR 值	反映信息
+	+	a	b	$ah/bg(OR_{eg}, A)$	G、E 联合作用效应
+	-	c	d	$ch/dg(OR_g, B)$	G 单独作用效应
-	+	e	f	$eh/fg(OR_e, C)$	E 单独作用效应
-	-	g	h	1	共同对照

这里,基因与环境联合作用的效应不仅包括两者分别作用的效应,还可能包括基因与环境作用的叠加,也可能呈现基因与环境作用的相乘效应。通过不同的模型,我们可以判别基于不同模型的两因素间交互作用是否存在及其大小。

通过表 1 提供的数值,可以计算基于不同模型的交互作用值,进行不同组合的分层分析,并且可以计算单纯病例或单纯对照中的 OR 估计值,其反映的主要信息通过表 2 表示^[2,3]。交互作用的存在与否,与所选择的模型密切相关^[4],对其相应的结果应进行统计学假设检验。

根据 Rothman 等提出的基于相加模型的计算交互作用的指标^[4-8],有以下几种:

(1)交互作用指数(the synergy index S, S): $S = (A - 1) / [(B - 1) + (C - 1)]$ 当 $S = 1$ 时,无交互作用; $S > 1$ 时,两因子间有正交互作用; $S < 1$ 时,两因子间有负交互作用。

(2)交互作用归因比(attributable proportion of interaction, AP):表示总效应中有多大比例归因于两因子交互作用 $AP = [A - (B + C - 1)] / A$ 。

(3)纯交互作用归因比: $AP^* = [A - (B + C - 1)] / (A - 1)$ 表示由基因和环境两因素引起的效应中归因于两因子交互作用的比例。

(4)交互作用超额相对危险度(relative excess

risk of interaction, RERI):表示两因素联合作用与其单独作用之和的差值, $RERI = A - (B + C - 1)$,同时也表示交互作用与两因素以外的因素作用之间的关系(这里公式是假定后者 $OR = 1$),如果未知因子作用很大,则所研究的交互作用就变得十分次要而没有意义。这里, RERI 即为基于相加模型的两因素交互作用值^[9]。

从上述四项指标可以看出,公式均是以两因素的相加模型的交互作用为前提的。如果是基于相加模型的交互作用,就可以进一步计算指标(1)至(4)的值。

2. 与分层分析、单纯病例研究之间的关系:

(1)在分层分析中,如果分层后各层 OR 值不一致,并经假设检验差异有统计学意义,可认为分层因素与研究因素之间存在效应修饰(effect modification)(即交互作用)^[4]。它不同于混杂效应,是观察者希望识别并予以报告的。可以证明,分层分析得出的两层 OR 之比与基于相乘模型的交互作用相等。

以表 2 中分层分析 1、2 为例, G 与 E 的交互作用为 $(ad/bc) / (eh/gf)$ 将分子分母同时乘以 h/g , 就会推导出其值即为基于相乘模型的交互作用 $A / (B \times C)$ 。通过分层分析 3、4 之比可以得出同样的结果。

(2)单纯病例研究由于在研究中不需要对照,花费小,研究快,倚小而逐渐受到重视,但其应用需要满足两个条件:①环境、基因两因素相互独立;②交互作用的模式是基于相对危险度(RR)的乘法模型,病例需要有代表性。这两个条件任何一个得不到满足,此研究就不能检验交互作用的存在^[9,12]。在患病率较低的前提下,以病例对照研究的 OR 值近似估计 RR 值。应用病例对照研究叉生分析表,我们可以充分利用单纯病例和单纯对照的数据。利用对照的各项数值计算 OR 值,可以反映或检验基因和环境两因素在人群中的独立性,理论上其值应

表2 基于不同模型计算交互作用

方法	OR 值	反映信息
相乘模型	$A / (B \times C)$	值为 1 表示两因素作用符合相乘模型,不存在交互作用; > 1 示正交互, < 1 示负交互(需经统计学假设检验)
相加模型	$A - (B + C - 1)$	值为 1 表示两因素作用符合相加模型,不存在交互作用; > 1 示正交互, < 1 示负交互(需经统计学假设检验)
单纯病例研究	ag/ce	基于相乘模型的交互作用
单纯对照研究	bh/df	人群中基因与环境因素的独立性(与无效假设 $OR = 1$ 进行假设检验如无统计学意义,示基因与环境因素独立)
分层分析 1	ad/bc	G(+)时, E 的作用
分层分析 2	eh/gf	G(-)时, E 的作用
分层分析 3	af/eb	E(+)时, G 的作用
分层分析 4	ch/dg	E(-)时, G 的作用

为 1;但由于抽样误差,单纯对照的两因素作用 OR 值并不一定恰好等于 1,此时应进行假设检验。这也是直接利用单纯病例数据计算出的两因素交互作用值,并不正好等于基于相乘模型交互作用 OR 值的原因所在。可以证明,利用叉生分析表计算的单纯病例研究 OR 值和单纯对照两因素独立性检验 OR 值之比恰等于基于相乘模型交互作用 OR 值。两 OR 之比: $(ag/ce)/(bh/df) = (a/b)/(che/dgf) = (ah/bg)/[(ch/dg) \times (eh/fg)] = A/(B \times C)$ 。

实例分析

1. 研究对象:一项关于口服避孕药(环境因素 E)和 Leiden 因子 V 基因突变(遗传因素 G)与静脉血栓栓塞的病例对照研究,静脉血栓栓塞患者 155 例,健康对照 169 人,其叉生分析数据表如表 3 左四列所示,资料描述具体见参考文献[13]。

表3 口服避孕药(E)和 Leiden 因子 V 基因突变(G)与静脉血栓栓塞的病例对照研究

G*	E#	病例数	对照人数	OR 值	反映信息 [△]
+	+	25	2	OR _{eg,A}	34.72(7.83~310.00)
+	-	10	4	OR _{g,B}	6.94(1.83~31.80)
-	+	84	63	OR _{e,C}	3.70(2.18~6.32)
-	-	36	100	1	
合计		155	169		

* G(+):Leiden 因子 V 基因突变(包括突变杂合子和纯合子); G(-):无 Leiden 因子 V 基因突变; # E(+):最近服用口服避孕药; E(-):最近不服用口服避孕药; △OR 值 95% CI

2. 环境与遗传因素单独作用与联合作用效应计算:以表 3 中两因素均为阴性的病例与对照组作参比组,其他各行病例与对照数据分别与其组成四格表,从中可以计算出两因素单独作用和联合作用的 OR 值和 95% CI(表 3)。

3. 基于相乘模型的交互作用计算:基于相乘模型的交互作用值为 $OR_{eg}/(OR_g \times OR_e) = 34.72/(6.94 \times 3.70) = 1.35, P = 0.759$ (P 值由笔者根据不控制混杂因素的 logistic 模型计算而得),显示不存在相乘正交互作用(原文中笔者在控制年龄的情况下,运用 logistic 模型计算 Leiden 因子 V 基因突变和使用口服避孕药的基于相乘模型的交互作用值为 1.2,95% CI:0.2~9.7,显示不存在两因素的相乘交互作用。笔者直接利用表 3 数据计算,显示也不存在相乘交互作用)。

对上例进行单纯病例 OR 分析,运用表 3 原始数据计算,其值: $(25 \times 36)/(10 \times 84) = 1.07$ (注意此

处不等于 1.35),同时计算单纯对照的 OR 值 = $(2 \times 100)/(4 \times 63) = 0.79$,计算单纯病例和单纯对照 OR 值之比为 $1.07/0.79 = 1.35$,即与基于相乘模型的交互作用其值是相等的。

对上例进行分层分析见表 4。

表4 基于表 3 的分层分析

E	G(+)		G(-)		OR 值之比
	病例数	对照人数	病例数	对照人数	
+	25	2	84	63	
-	10	4	36	100	
OR 值(95% CI)					5.00(0.79~31.77) 3.70(2.24~6.12) 1.35

分层后,G(+)与 G(-)两层 OR 值不等,其 OR 值之比为 $5.00/3.70 = 1.35$,即等于基于相乘模型的交互作用。层间 OR 值一致性及交互作用的假设检验不赘述。

4. 基于相加模型的交互作用计算:这里,我们进一步对本例计算基于相加模型的交互作用值进行分析,根据文中前述的计算交互作用的指标,可以计算出:

(1)交互作用指数: $S = (A - 1)/[(B - 1) + (C - 1)] = 3.90, A - 1 > (B - 1) + (C - 1)$ 表明两者之间存在基于相加模型的正交互作用,有 Leiden 因子 V 基因突变者再口服避孕药可能增加静脉血栓栓塞的危险。该指数是否具有统计学意义需要进行假设检验。

(2)交互作用归因比: $AP = [A - (B + C - 1)]/A = 0.7224 = 72.24\%$,表明有上述基因突变的口服避孕药者,其产生的静脉血栓栓塞患者中,有 72.24% 是由该两因子的交互作用所致。

(3)纯交互作用归因比: $AP^* = [A - (B + C - 1)]/(A - 1) = 0.7438 = 74.38\%$,在有上述基因突变的口服避孕药者所产生的全部静脉血栓栓塞患者中,当除去其他或未知因子产生的病例后,因这两因子交互作用占有的与两因子有关的病例总数比例为 74.38%。

(4)交互作用超额相对危险度: $RERI = A - (B + C - 1) = 25.08$,表明归因于有 Leiden 因子 V 基因突变者口服避孕药两者交互作用所致的静脉血栓栓塞危险性是其他或未知因子(其 OR = 1)的 25.08 倍。

对相加模型的交互作用的假设检验较为复杂,不再详述。有兴趣可以参看 Mitchell 与 Jacques 合著的 Encyclopedia of epidemiologic methods (John

Wiley & Sons Ltd, 2000) 相关章节。笔者尝试对相加效应模型交互作用值即 RERI 用 Stata 软件编程进行假设检验, 得到其值为 25.073 95 (约为 25.08), $P=0.177 (P>0.05)$, 表明两因素虽然对静脉血栓栓塞作用的相加交互作用值很大, 达到 25.08, 但仍然未呈现统计学意义, 即两因素相加交互作用不具有统计学意义。究其原因, 笔者认为同时具有 Leiden 因子 V 基因突变者再口服避孕药的对照数极少, 仅为 2 例, 造成可信区间过宽, 两因素联合作用效应不稳定所致。要克服此问题, 需要进一步加大样本量。

讨 论

又生分析表作为病例对照研究中基本的流行病学分析单元, 具有信息量丰富, 计算直观等优点。本文以一项关于口服避孕药 (环境因素 E) 和 Leiden 因子 V 基因突变 (遗传因素 G) 与静脉血栓栓塞的病例对照研究为例, 通过又生分析表, 不仅分别获得基因和环境因素各自的主效应, 还获得基于不同模型 (相加或相乘模型) 的两因素对疾病的交互作用值。而且相加模型的交互作用比相乘模型的交互作用更具有公共卫生学意义^[4]。尽管从本例的计算表明两因素对静脉血栓栓塞作用既不具有相乘交互作用, 也不具有相加交互作用, 但通过本例的计算, 我们可以十分直观地计算基于不同模型的交互作用值, 以后再借助于 SAS、Stata 等统计软件, 对其进行假设检验。可以说, 又生分析表直观地呈现了流行病学中基本分析单元的绝大部分信息, 为我们提供了更广博的判断力。

从例证中还可可见若基于相乘模型计算交互作用, 其交互作用值与分层分析和单纯病例与对照的 OR 比值是一致的, 我们不仅从例中得证, 而且从普遍性的公式成功进行了推导, 这就使我们更加清楚又生分析表与后两者分析方法的关系。若交互作用是基于相加模型时, 我们还可具体计算其各项作用指标。而这是分层分析和单纯病例研究所不能给予的。

进一步挖掘又生表的信息, 发现还可以得到因素单独作用和联合作用下针对暴露组和针对人群的归因危险百分比; 如果对照是人群的代表性样本, 还可估计在对照人群中暴露的频率; 对于不同分层情

况下的优势比计算亦十分方便; 对单纯对照的 OR 值估计可帮助我们了解两因素在研究人群中作用是否独立 (本文未列出)。总之, 又生分析表提供的信息十分丰富, 在统计分析中应充分挖掘, 予以利用。但是, 由于 2×4 又生分析表只能分析两个二分类因素的作用, 在分析交互作用时, 未考虑其他不参加又生的危险因素的作用 (其他类型的又生分析表也存在同样的问题), 应将其与多因素回归分析 (如基于相乘模型的 logistic 回归分析, 或基于广义相对危险度模型的分析) 相结合, 才能得到更为合理的信息。尽管如此, 又生分析表在流行病学分析中, 尤其是交互作用的分析中其作用依然显而易见。

参 考 文 献

- 1 周利锋, 高尔生, 金丕焕, 等. 胃癌病因学研究中交互作用研究方法的比较. 中国卫生统计, 1997, 14: 9-11.
- 2 李辉. 临床流行病学研究中交互作用的判别及测量 (一). 中华流行病学杂志, 1996, 17: 243-247.
- 3 李辉. 临床流行病学研究中交互作用的判别及测量 (二). 中华流行病学杂志, 1996, 17: 305-307.
- 4 黄悦勤, 主编. 临床流行病学. 北京: 人民卫生出版社, 2002. 143-146.
- 5 Rothman KJ, Greenland S. Modern Epidemiology. 2nd ed. Lippincott-Raven Publishers, 1998.
- 6 向惠云, 余松林, 孙奕, 等. 疾病资料多元分析中交互作用指标及可信区间的估计. 中国卫生统计, 1999, 16: 130-132.
- 7 贺晓新, 金水高. 个体匹配病例对照研究中的交互作用及其处理方法. 中国卫生统计, 2002, 19: 43-45.
- 8 Mitchell HG, Jacques B. Encyclopedia of epidemiologic methods. John Wiley & Sons Ltd, 2000.
- 9 俞顺章, 蔡琳, 穆丽娜, 等. 单纯病例研究方法在流行病学研究中的应用. 中华流行病学杂志, 2003, 24: 406-409.
- 10 李大林, 李立明. 基因-环境交互作用研究方法: 无对照病例研究. 中华流行病学杂志, 2002, 23: 304-307.
- 11 Muin JK, Flanders WD. Nontraditional epidemiology approaches in the analysis of gene-environment interaction: case-control studies with no controls! A J Epidemiol, 1996, 144: 207-213.
- 12 David C, Paul MM. Epidemiological methods for studying genes and environmental factors in complex disease. Lancet, 2001, 358: 1356-1360.
- 13 Vandenbroucke JP, Koster T, Briet E, et al. Increased risk of venous thrombosis in oral-contraceptive users who are carriers of factor V Leiden mutation. Lancet, 1994, 344: 1453-1457.

(收稿日期: 2004-05-26)

(本文编辑: 张林东)