

基因-环境交互作用研究方法:无对照病例研究

李大林 李立明

从目前的观点来看,可以认为很少有疾病是纯粹由遗传或者环境因素决定的,基因-环境交互作用在许多疾病,特别是常见的慢性疾病或者是所谓的“复杂性状疾病”(complex-trait diseases)的发病中,具有重要的意义^[1]。基因-环境交互作用可以理解为遗传因素对环境因素易感性的影响^[2]。对于基因-环境交互作用的深入研究,有助于了解人群易感性差异的原因,进而可以对环境与疾病、基因与疾病的关系有更深入的认识。目前许多对复杂性状疾病病因的遗传流行病学研究得到非常不一致的结论^[3],没有考虑到基因-环境交互作用应该是重要原因之一。从另一个角度来看,如果能够对基因-环境交互作用有深入了解的话,我们就可以根据个人的基因型,制定更有效的干预措施以及个人保健措施^[4]。传统研究交互作用应用的方法是病例对照研究,但应用病例对照研究进行基因-环境交互作用的研究需要很大的样本量,同时一些潜在的因素可能会对研究结果有较大的影响。近几年来,研究者提出,可以用其他的研究方法进行基因-环境交互作用的研究,例如无对照病例研究(case-only study)、病例-亲本研究(case-parental study)等^[5-8]。我们将对无对照病例研究作一简单介绍。

一、概述

1994年Walter等^[7]提出无对照病例研究可以用于研究交互作用。在无对照病例研究中不采用对照,只对病例中已知遗传因素(G)与环境因素(E)的分布情况进行分析,判断在疾病的发病中是否存在已知基因与已知环境因素的交互作用,若交互作用存在,这种方法还可以对交互作用的效应进行估计。此方法最初是基于logistic模型提出的,在此我们利用四格表对其原理加以说明。

无对照病例研究是以交互作用的RR值(risk ratio)乘法模型为基础的。在RR值乘法模型成立时,若不存在遗传因素与环境因素的交互作用,则不同环境暴露水平下,相同水平的遗传因素的RR值是相同的。即设 RR_{ge} 为 $G=g, E=e$ 时相对于 $G=0, E=0$ 时的RR值,则交互作用符合RR值的乘法模型时,有:

$$\frac{RR_{g1}}{RR_{00}} = \frac{RR_{g1}}{RR_{01}} = \frac{RR_{g2}}{RR_{02}} = \dots = \frac{RR_{ge}}{RR_{0e}} \quad (1)$$

在此为简化起见,我们只考虑G与E都为二分变量的情况。此时在无对照病例研究中可以得到表1的分布。

此时若考虑一个以整个人群的情况,则可得到表2。

表1 无对照病例研究中遗传因素与环境因素的分布

		G	
		+	-
E	+	a	b
	-	c	d

表2 在无对照病例研究中考虑虚拟对照时的分布情况

		E(+) G		E(-) G	
		+	-	+	-
D	+	a	b	c	d
	-	a'	b'	c'	d'

此时根据式(1)可得:

$$\frac{ab'}{(a+a')(b+b')} = \frac{cd'}{(c+c')(d+d')} \quad (2)$$

变换可得:

$$\frac{ad}{bc} = \frac{(a+a')(d+d')}{(b+b')(c+c')} \quad (3)$$

式(3)的左侧部分被一些学者称为cross-product OR(COR),其右侧为一般人群中不同环境暴露水平下遗传因素各水平的比值之比,若在正常人群中遗传因素与环境因素相互独立,即E的分布对G的分布无影响,则式(4)的右侧为1。因此,在无交互作用时,COR=1,检验交互作用的存在与否,只需检验COR是否为1即可。同时COR还可以作为交互作用效应大小的指标,其值愈偏离1,则交互作用的效应愈强。以上所考虑的只是G与E都为二分变量的情况,在二者为分类变量的情况下,此结论依然能够成立。

由以上的介绍可以看出,无对照病例研究要求有两个基本假设必须成立:①交互作用的模式是RR值的乘法模型;②基因型与环境暴露 in 目标人群中的分布彼此独立。如果这两个条件任一个不能得到满足,那么无对照病例研究就无法检验基因-环境交互作用的存在。同时,由于没有对照,无对照病例研究要求选择的病例尽量能够代表整个人群的情况,以保证研究结果的真实性和可推广性。

二、统计分析

无对照病例研究的资料通常可以有两种分析方法:

1. 遗传因素与环境因素都是二分变量时,可以用四格表进行统计分析。此时无对照病例研究的结果可以用如表1所示的四格表表示,其统计分析包括三步:

(1)计算COR

$$COR = \frac{ad}{bc} \quad (4)$$

(2)计算COR的可信区间。

(3) 并利用 χ^2 检验无效假设 检验的自由度为 1。

2. 利用 logistic 回归进行分析^[5]。以基因型为因变量, 环境暴露情况为自变量, 进行 logistic 回归分析, 可得到 COR 的估计值及其可信区间, 并对无效假设进行检验。这种分析方法在遗传因素以及环境因素是分类变量或者考虑协变量的情况下, 更为适用。Walter 等^[7]认为, 这种分析方法要比利用病例和对照进行 logistic 回归分析具有更大的检验把握度。

计算得到的 COR 估计值若为 1, 则说明基因-环境交互作用符合交互作用的 RR 值乘法模型, COR > 1 说明遗传因素可加强环境因素的致病作用, 而 COR < 1 则说明遗传因素对环境因素的致病能力有保护作用。

值得注意的是, 在传统病例对照研究中, 计算交互作用效应的指标等价于易感基因携带者暴露于环境危险因素时的 OR 值与其未暴露于环境危险因素时 OR 值的比值(OR_i), 在以前的文献中, 有研究者将 COR 等同于采用传统病例对照研究交互作用所得到的交互作用 OR 值(OR_i)。Walter 等^[7]最初提出无对照病例研究时, 是采用 logistic 模型进行证明的, 他们在假设人群中患病率很低的情况下, 提出可以利用无对照病例研究方法研究遗传与环境因素之间的关系是否偏离交互作用的 OR 值乘法模型, 并认为此时得到的交互作用效应指标等同于传统病例对照研究得到的交互作用 OR 值(OR_i)。这实际上是在人群患病率很低的情况下得到的近似结果。根据式(3)可以看到, 当人群患病率很低时, 其可以做如下近似:

$$\frac{ad}{bc} = \frac{a'd'}{b'c'} \quad (5)$$

进一步变化可以得到:

$$\frac{ab'}{a'b} \times \frac{ac'}{a'c} = \frac{ad'}{a'd} \quad (6)$$

$$\text{即: } OR_{11} = OR_{01} \times OR_{10} \quad (7)$$

此即为交互作用的 OR 值乘法模型, 即在人群患病率较低的情况下, 可以近似认为符合交互作用的 RR 值乘法模型等价于符合交互作用的 OR 值乘法模型。

但若人群患病率较高, 式(3)至式(4)的近似变换是不能成立的。此时符合交互作用的 RR 值乘法模型并不能等价于符合交互作用的 OR 值乘法模型。Silke 和 Danief^[9]也于 1999 年用 Bayes 理论证明, COR 实际上等价于交互作用的 RR 值(RR_i), 即在无对照病例研究中, 计算交互作用效应的指标 COR 等价于二者的 RR 值之比(RR_i)。如果疾病患病率较低时, RR_i 与 OR_i 二者可近似相等, 但在其他情况下, 例如疾病的患病率较高时, 那么 RR_i 将明显小于 OR_i 。因此, 在一些常见疾病的基因-环境交互作用研究中, 无对照病例研究对于交互作用效应的研究结果不能与传统病例对照研究直接比较。

三、优势与存在的问题

1. 优势: 无对照病例研究实际上是采用虚拟对照的理想分布情况作为对照进行研究的, 有研究者因此将其与 case-crossover study 等研究方法归为一类, 称为 case-distribution

study^[10]。此种研究方法提供了一种简便而有效的分析基因-环境交互作用的方法, 它对于交互效应的估计要比传统病例对照研究更精确^[7], 而且在假设检验时, 由于只采用病例时统计检验的方差比采用病例和对照时小, 因此在采用无病例对照研究时将有更大的把握度检验交互作用的存在^[5, 11]。

同时, 采用无对照病例研究探索基因-环境交互作用, 可以使研究的样本量与成本显著降低。Yang 等^[11]曾根据他们的计算方法对无对照病例研究所需的样本量进行计算, 发现这种方法可以极大的减少所需样本量。例如在 $\alpha = 0.05$, 人群中 $e = 0.3$, $g = 0.1$, $R_e = 2$, $R_g = 1$, $R_i = 5$ 的前提下, 达到 0.80 的把握度, 在病例对照研究中即使采用 1:2 配比, 仅病例也需要 269 例。而采用无对照病例研究, 则只需要 123 例病例即可达到要求。

另外, 由于无对照病例研究没有采用对照, 因此由于采用对照所可能引入的各种偏倚^[6], 例如回忆偏倚、调查偏倚以及病例选择不当所可能导致的混杂偏倚也可以得到避免。

2. 不足: 无对照病例研究也存在其自身的不足, 首先, 无对照病例研究是基于交互作用的 RR 乘法模型的, 它只能对交互作用是否偏离 RR 值乘法模型进行检验。目前对于交互作用的定义仍无统一的结论, 但可以肯定, 基因-环境交互作用的方式决不仅 RR 值的乘法模型一种, 因此应用无对照病例研究有可能会遗漏其他的交互作用的情况^[5]。而且, 由于传统病例对照研究的结果判断的是交互作用是否偏离交互作用的 OR 值相乘或相加模型, 因此, 无对照病例研究的结果可能与传统病例对照研究的结果存在一定的差异, 二者的效应指标一般也不能直接比较。

其次, 无对照病例研究假设遗传与环境之间相互独立, 研究者也常常将这种假设作为默认前提。但在有些情况下, 这个假设不能成立。在不同的亚人群中, 遗传和环境的暴露情况都可能不同, 二者之间的共变可能导致在整个人群之中, 遗传和环境因素之间存在相关。此被称为“人群分层”(population stratification)。另外, 某些基因可能会通过一些生化机制影响个体的行为, 进而影响个体的环境暴露情况, 例如酒精的代谢酶——乙醛脱氢酶的突变可以导致酒精代谢延迟, 使个体摄入酒精后恶心呕吐的反应加重, 从而使个体不愿意饮酒, 进而可能产生该基因型与饮酒之间的负相关关系^[12]。此时若用无对照病例研究方法研究某些疾病与酒精摄入, 以及酒精代谢酶之间关系的时候就有可能得到虚假的结论。这些情况都不能满足无对照病例研究的基本假设, 从而使其无法正确的作出检验^[6, 11]。

另外, 从其统计分析的过程可以看出, 无对照病例研究只能检验基因-环境交互作用是否存在, 并估计其交互效应的大小。因为没有对照, 无法估计遗传和环境因素的主效应。

四、应用举例

Claire 等于 2000 年进行了一次双亲吸烟, CYP1A1 基因多态性以及儿童白血病之间关系的研究。研究采用病例对照设计, 但在分析过程中将病例单独作为一个亚研究, 利用

logistic 回归进行了无对照病例分析,表3是部分分析结果。

表3 Claire 等利用无对照病例分析研究母亲吸烟, CYP1A1 多态性的交互作用与儿童白血病关系的结果 (调整儿童的年龄与性别后)

	CYP1A1 * 4 多态性		COR 值(95% CI)
	野生型	突变型	
妊娠前期			
不吸烟	82	4	1.0
吸烟(支)			
1~20	43	2	1.1(0.2~6.4)
>20	21	2	1.0(0.3~11.7)
妊娠中期			
不吸烟	89	3	1.0
吸烟(支)			
1~20	45	3	2.3(0.4~11.8)
>20	12	2	5.3(0.8~36.8)
妊娠后期			
不吸烟	90	3	1.0
吸烟(支)			
1~20	44	3	2.3(0.4~12.2)
>20	12	2	5.4(0.8~37.3)

在此研究中,尽管由于基因突变型在病例中出现的频率很低而无法得到有显著意义的结果,但此种情况下若采用传统病例对照研究分析其交互作用,为达到足够的把握度,需要的样本量远大于采用无对照病例分析其交互作用所需要的样本量。

此研究中有一点很值得我们借鉴,即采用病例对照研究的方法估计遗传与环境因素的主效应,用无对照病例研究分析其交互作用。这样的设计同时具有传统病例对照研究与无对照病例研究的优点,所需样本量较少,而且可以同时估计遗传因素与环境因素的主效应与交互效应,具有很高的研究效率。

五、研究进展

1. 样本量估计:其样本量的估计目前尚无成熟的方法。Yang 等^[1]于1997年提出一种无对照病例研究的样本量计算方法。若所需的样本量为 n ,目标人群中环境因素的暴露率为 e ,其危险度为 R_e ,易感基因的暴露率为 g ,其危险度为 R_g ,基因-环境交互作用的效应为 R_i ,那么,在研究中,可以得到表4的预期分布。

研究所需的样本量从式(8)得到:

$$n = \frac{Z_{\alpha/2} \sqrt{V_N} + Z_{\beta} \sqrt{V_A}}{[\log(R_i)]^2} \quad (8)$$

其中

$$V_A = \frac{1}{n} \left(\frac{1}{a} + \frac{1}{b} + \frac{1}{c} + \frac{1}{d} \right)$$

$$V_N = \frac{1}{n} \left[\frac{1}{MN} + \frac{1}{(T-N)M} + \frac{1}{(T-M)N} + \frac{1}{(T-M)(T-N)} \right]$$

2. 应用范围:Yang 等^[1]提出,无对照病例研究除了可以用于研究基因环境交互作用外,还可以用于研究基因-基因交

互作用。这一点是比较容易理解的,但根据其原理可以看出,无对照病例研究在研究基因-基因交互作用时,也同样要求感兴趣的两个基因在人群中分布是独立的,即要求二者之间彼此不存在连锁不平衡。因此,应用无对照病例研究基因-基因的交互作用时,研究者应该清楚感兴趣的两个基因的位置,二者最好在不同的染色体上,如果在同一染色体上,也要求二者相距很远,否则,连锁不平衡将极大的影响研究结果。

表4 无对照病例研究的预期分布

		G		M
		+	-	
E	+	a	b	T-M
	-	c	d	
		N	T-N	T

$$a = n(g e R_e R_g R_i) / \Sigma; b = n((1-g)e R_e) / \Sigma$$

$$c = n((1-e)g R_g) / \Sigma; d = n((1-g)(1-e)) / \Sigma$$

$$M = a + b; N = a + c;$$

$$T = a + b + c + d$$

$$\Sigma = (1-g)(1-e) + g(1-e)R_g + (1-g)R_e + g e R_e R_g R_i$$

总之,无对照病例研究是近年来提出的一种新型的研究基因-环境交互作用的设计方法。其最大的特点在于简单高效,所需样本量少。但同时也存在一些问题。我们应根据研究目的和实际情况决定是否使用此种研究方法。如果研究的目的是初步筛选可能的基因-环境交互作用,并且在研究人群中,感兴趣的遗传与环境因素彼此独立,那么此时应用无对照病例研究将极大地减少研究的花费,同时可以提高检验效率。如果研究的易感基因在人群中比较罕见,采用传统病例对照研究时,由于对照中易感基因暴露率很低,将导致对交互效应估计的不稳定,在这种情况下,无对照病例研究是比较适宜的研究方法。

另一方面,由于有时候人群中感兴趣的遗传因素与环境因素之间的彼此独立不能得到满足,而且此种研究方法只能检验遗传因素与环境因素之间的作用是否偏离OR值的相乘模型,因此无对照病例研究的结果必须经其他研究方法进一步验证。另外,如果需要同时检验基因与环境因素的主效应与交互效应,也应采用其他研究方法。

参 考 文 献

- 1 Lander ES, Schork WS. Genetic dissection of complex traits. Science, 1994, 265:2037-2048.
- 2 Evaves LJ. The resolution of genotype environment interaction in segregation analysis of nuclear family. Genet Epidemiol, 1984, 1: 215-228.
- 3 Phillips PH, Linet MS, Harris EL. Assessment of family history information in case-control cancer studies. Am J Epidemiol, 1991, 133: 757-765.
- 4 Campbell H. Gene-environment interaction. J Epidemiol Community Health, 1996, 20: 397-400.
- 5 Khoury MJ, Flanders WD. Nontraditional epidemiology approaches in the analysis of gene-environment interaction: Case-control studies with no controls! Am J Epidemiol, 1996, 144: 207-213.
- 6 Weiberg CR, David MU. Choosing a retrospective design to access joint

genetic and environment contributions to risk. *Am J Epidemiol*, 2000, 152: 197-203.

7 Walter WP, Clarice RW, Jack AT. Non-hierarchical logistic models and case-only designs for assessing susceptibility in population-based case-control studies. *Statistics in Medicine*, 1994, 13: 153-162.

8 Khoury MJ. Case-parental control method in the search for disease-susceptibility genes. *Am J Hum Genet*, 1994, 55: 414-415.

9 Silke Schmidt, Daniel JS. Potential misinterpretation of the case-only study to access gene-environment interaction. *Am J Epidemiol*, 1999, 150: 878-885.

10 Sanger Greenland. A unified approach to the analysis of case-distribution

studies. *Statistics in Medicine*, 1999, 18: 1-15.

11 Yang QH, Khoury MJ, Flanders WD. Sample size requirements in case-only designs to detect gene-environment interaction. *Am J Epidemiol*, 1997, 146: 713-720.

12 Sherman DI, Ward RJ, Yoshida A, et al. Alcohol and aldehyde dehydrogenase gene polymorphism and alcoholism. *EXS*, 1994, 71: 291-300.

13 Yang QH, Muin JK, Sun FZ, et al. Case-only design to detect gene-gene interaction. *Epidemiology*, 1999, 10: 167-170.

(收稿日期 2001-12-04)

(本文编辑 张林东)

- 短篇报道 -

宁波市大肠埃希菌 O157:H7 的发现与监测

许国章 徐景野 王仁元 何国华 王建辉 徐奋奋 周爱明

为了了解和掌握宁波市腹泻病人、动物携带大肠埃希菌 O157:H7 状况,于 1998~2001 年在全市范围内开展了系统的监测工作,结果报告如下。

1. 材料与与方法:

(1) 标本采集:各县(市、区)根据监测方案的要求,采集腹泻病人(脓血便为主)、动物粪便及各类食品标本等,并填写统一的流行病学调查记录表。

(2) 培养基及诊断血清:肠道增菌肉汤(EC 肉汤)、山梨醇麦康凯琼脂(SMAC)、营养琼脂、麦康凯琼脂(MAC)由江苏省宜兴万石培养基厂生产,微量生化管由杭州市天和微生物试剂有限公司提供,其他培养基为宁波市卫生防疫站制备。O157 和 H7 诊断血清由中国疾病预防控制中心传染病预防控制所和浙江省疾病预防控制中心提供。

(3) 分离培养及鉴定:标本接种于 EC 肉汤,36℃ 增菌 16 h,划种 SMAC 平板,36℃ 培养 24 h,挑取无色菌落(不发酵山梨醇),转种克氏双糖铁斜面,36℃ 培养 24 h,结果符合大肠埃希菌反应特点者,在 MAC 平板上分纯,然后再进行形态学、初步生化试验和大肠埃希菌 O157:H7 血清学鉴定。

(4) 系统生化及分子生物学试验:生化测试由 VITEK32 细菌自动鉴定仪(由法国梅里埃公司生产)完成,同时将菌株送浙江省疾病预防控制中心,用 PCR 技术对菌株进行分子生物学方面的鉴定。

2. 结果:

(1) 人群监测:从 1998 年开始设点开展了大肠埃希菌 O157:H7 感染监测,至 2001 年底全市累计监测腹泻病人标本 3 120 份,分离到大肠埃希菌 O157:H7 1 株,检出率为 0.32‰。

作者单位 315010 宁波市卫生防疫站防疫科(许国章、徐景野、王仁元、何国华、周爱明),奉化市卫生防疫站微生物检验科(王建辉、徐奋奋)

(2) 动物粪便及各类食品标本监测:全市累计监测各类标本 3 865 份,培养检查大肠埃希菌 O157:H7 全部阴性。

(3) 首例病人的临床及流行病学特征:患者男性 65 岁,退休工人,居住在本市奉化城关,于 2001 年 9 月 11 日因发热 37.7℃,腹泻 2 次(系溏便)就诊于奉化市人民医院肠道科,即采集大便进行大肠埃希菌 O157:H7 检测,9 月 14 日分离到 1 株大肠埃希菌 O157:H7,后经省、市卫生防疫部门共同鉴定确系大肠埃希菌 O157:H7。据调查患者家庭卫生状况良好,无禽、畜圈养,近期也无外出史,自述近 2 年来一直有溏便史,每日 2~3 次,9 月 10 日在其女儿家吃晚餐,主要食用了蟹、虾、猪蹄、鸭肉(系外购熟食)及水果(哈密瓜、梨)等,11 日上午自感不适、腹泻、发热,即到医院求诊。疫情发生后,对病人立即采取了隔离治疗及病家的疫点消毒、密切接触者与环 境等采样检测。

(4) 大肠埃希菌 O157:H7 细菌学特征:本次分离到的大肠埃希菌 O157:H7 为革兰染色阴性,无芽孢短小杆菌,在营养琼脂平板上生长良好。生化试验符合大肠埃希菌特征。血清学鉴定与 O157 和 H7 诊断血清作玻片凝聚,其凝聚价均达 3 以上。分子生物学鉴定未检出 *SLT*、*hly*、*eae* 等毒力基因,为大肠埃希菌 O157:H7 不产毒株。

3. 讨论:目前我国大肠埃希菌 O157:H7 的分布已经比较广泛,近几年来全国各地已陆续从动物粪便(牛、猪、鸽)、食品(猪肉及其制品、鸡肉)及病人中分离到该菌。我们连续 4 年的系统监测,仅在腹泻病人标本中检出 1 株大肠埃希菌 O157:H7,检出率为 0.32‰,而动物粪便及各类食品标本中均未检出,说明本市腹泻病人中已有大肠埃希菌 O157:H7 的存在。因此,有必要进一步加强流行病学监测工作,防止肠出血性大肠埃希菌 O157:H7 爆发性流行。

(收稿日期 2001-11-19)

(本文编辑:张林东)