

病例同胞对照设计

易洪刚 陈峰 于浩 赵杨 娄东华

【摘要】 目的 介绍病例同胞对照设计的研究设计及统计分析方法,并对实际资料进行统计分析。方法 以亚甲基四氢叶酸还原酶基因 C677T 突变与冠心病连锁研究为实例,采用同胞-传递不平衡检验(s-TDT)和同胞组不平衡检验(SDT)等方法进行统计分析。结果 采用 s-TDT 分析,统计量 $Z=0.27, P>0.05$;采用 SDT 分析, $\chi^2=0.31, P>0.05$ 。提示受累子代与非受累子代 T 等位基因分布差异无统计学意义。结论 病例同胞对照设计使用患者同胞作为对照,避免了人群分层现象,可以用来检测基因与疾病之间的关联或连锁。

【关键词】 病例同胞对照设计; 关联; 连锁; 基因

Introduction on "case-sibling control design" YI Hong-gang, CHEN Feng, YU Hao, ZHAO Yang, LOU Dong-hua. Department of Epidemiology and Biostatistics, School of Public Health, Nanjing Medical University, Nanjing 210029, China

Corresponding author: CHEN Feng, Email: ohceptf@163.com

【Abstract】 Objective To introduce the design and statistical methods of case-sibling control design and to analyze the published data. **Methods** Data from an association study between the coronary heart disease and methylenetetrahydrofolate reductase gene C677T polymorphism was analyzed by the sib transmission/disequilibrium test (s-TDT) and the sibship disequilibrium (SDT) methods. **Results** Using s-TDT method, Z value was 0.27 with $P>0.05$. The result of SDT method showed that chi-square was 0.31 with 1 $df, P>0.05$. All results suggested that neither s-TDT nor SDT showed significant difference between the transmitted and untransmitted methylenetetrahydrofolate reductase gene C677T allele distributions. **Conclusion** Case-sibling control design might avoid population stratification by using siblings as controls thus might be used to test association and linkage between genes and disease.

【Key words】 Case-sibling control design; Association; Linkage; Gene

致病基因的定位与克隆在阐明疾病发病机制、治疗与预防等方面都具有重要意义。连锁分析(linkage analysis)是常用的定位致病基因的方法,对于许多符合孟德尔遗传规律的单基因疾病这种方法非常有效。但是对于复杂性疾病(complex disorder),由于疾病易感性基因的外显率较低,并且可能是多个基因联合作用而致病,所以连锁分析在多基因遗传疾病易感基因定位研究中的作用非常有限^[1]。以人群为基础的关联研究(population-based association study)是连锁分析之外的另一个用来研究特定遗传标记(genetic marker)与疾病关联的一种方法。虽然该方法简单易于实施,但是由于人群

分层现象的存在,有可能会出虚假的关联结果。近年来,以家系为基础的关联研究(family-based association study)逐渐得到了广泛的应用^[2]。由于使用家系内的成员作为对照,从而避免了人群分层(population stratification)等因素带来的混杂。病例同胞对照设计就是其中一种较优的方法。近年来该设计出现了许多统计分析方法,在研究遗传标记与疾病关联方面得到广泛的应用。现将这些方法简述如下。

基本原理

虽然早在 1958 年 Manuila^[3]认为使用同胞作为对照能够有效的避免人群分层导致的偏倚,但是直到最近才有学者开始探讨一系列的使用同胞作为对照的分析方法。由于患者及其同胞具有相似的遗传背景, Curtis^[4]、Spielman, Ewens^[5]、Horvath, Laird^[6]提出了使用同胞作为对照来研究疾病关联的一种新设计,称为病例同胞对照设计。以患者及患者未患

基金项目:国家重点基础研究发展计划(973)资助项目(2002CB512910);国家自然科学基金资助项目(30571619);江苏省高校自然科学基金重点项目资助(04KJB310081)

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

通讯作者:陈峰, Email: ohceptf@163.com

病的同胞作为研究对象,对其进行基因分型,通过比较同胞的等位基因或者基因型,来检测遗传标志是否与疾病位点基因关联或者连锁。

1. 同胞-传递不平衡检验^[5] (s-TDT): 将每个家系的资料整理为表 1。在每个同胞对中, a 表示患者数目, u 表示未患病同胞数目, 则所有的同胞数 $t = a + u$ 。 r 表示基因型为 1/1 的数目, s 表示基因型为 1/2 的数目。

s-TDT 检验统计量:

$$Z = (Y - A) / \sqrt{Var} \quad \text{或} \\ Z' = (|Y - A| - 1/2) / \sqrt{Var} \quad (1)$$

近似服从标准正态分布。式中

$$Y = \sum(n_{1i} + n_{2i}) = \sum a_i; \quad A = \sum(2r_i + s_i) a_i / t_i \\ Var = \sum au [4r(t - r - s) + s(t - s)] / [t^2(t - 1)]$$

表 1 同胞-传递不平衡检验分析

患病状态	同胞数			合计
	1/1	1/2	2/2	
患病	n_{1i}	n_{2i}	-	a_i
未患病	n_{3i}	n_{4i}	-	u_i
合计	r_i	s_i	-	t_i

2. 同胞组不平衡检验 (SDT)^[6]: 在每一个同胞对中, m_A^1 和 m_U^1 分别表示每个同胞对中患者和未发病同胞携带变异等位基因 1 的均数, 即: $m_A^1 = \{\text{患者携带等位基因 1 的个数}\} / \{\text{同胞对中发病人数的}\}$; $m_U^1 = \{\text{对照携带等位基因 1 的个数}\} / \{\text{同胞对中未发病人数的}\}$; $d^1 = m_A^1 - m_U^1$ 表示患者和对照携带变异等位基因 1 的均数之差。当 $d^1 = 0$ 时, 这样的同胞对资料就剔除; 用 b 表示 $d^1 > 0$ 的同胞对数目; c 表示 $d^1 < 0$ 的同胞对数目。

SDT 检验统计量:

$$T = (b - c)^2 / (b + c) \quad (2)$$

近似服从自由度为 1 的 χ^2 分布。

3. 同胞对关联检验法 (the sib association test, SAT)^[4]: 假如患者有多个未发病同胞时, 则选择基因型与患者有最大差别的一个同胞作为对照。每个同胞对中, 包括了一位患者和一位未发病的同胞。考虑双等位基因时, 双亲和患者基因型有如表 2 所示的三种有效组合。

检验统计量:

$$Z = (T_{AB} - N_1/2 - N_2) / \text{sqrt}(N_1/4 + N_2) \quad (3)$$

近似服从标准正态分布。其中, T_{AB} 表示标记为 t_{AB}

的次数, N_1 表示标记为 t_{AB} 或者 t_{BA} 的次数, N_2 表示标记为 $2t_{AB}$ 或者 $2t_{BA}$ 的次数。

表 2 同胞对关联检验法分析

双亲基因型	患者基因型	患者基因型概率	对照基因型	对照基因型概率	标记
AA, AB	AA	0.5	AB	$1 - 0.5^S$	t_{AB}
	AB	0.5	AA	$1 - 0.5^S$	t_{BA}
BB, AB	BB	0.5	AB	$1 - 0.5^S$	t_{BA}
	AB	0.5	BB	$1 - 0.5^S$	t_{AB}
AB, AB	AA	0.25	BB	$1 - 0.75^S$	$2t_{AB}$
			AB	$0.75^S - 0.25^S$	t_{AB}
	AB	0.5	AA	$0.5(1 - 0.5^S)$	t_{BA}
			BB	$0.5(1 - 0.5^S)$	t_{AB}
BB	0.25	AA	$1 - 0.75^S$	$2t_{BA}$	
		AB	$0.75^S - 0.25^S$	t_{BA}	

s_i : 作为对照的同胞数目; t_{AB} : 患者双亲传递等位基因 A 给患者、传递等位基因 B 给对照; t_{BA} : 患者双亲传递等位基因 B 给患者、传递等位基因 A 给对照

4. 表型不一致同胞对分析 (discordant sib pair analysis, DSP): Boehuke, Langefeld^[7] 提出了使用病例-同胞对照设计时的一些检测关联的方法。如等位基因 (基因型) 计数法 (allele-and genotype-counting statistics)、排列检验 (permutation test)、对称性检验 (symmetry statistics) 等。

在等位基因 (基因型) 计数法中, 假设遗传标志位点有 $m (m \geq 2)$ 个等位基因。当遗传标志与疾病没有关联时, 等位基因在患者和同胞中的计数应该是相等的。因此, 该方法通过比较患者及其同胞易感等位基因的计数来检测遗传标志与疾病的关联。

采用两种等位基因计数方法, 计数法 1 计算所有的等位基因的数目, 计数法 2 只计算患者和同胞遗传标志不同时的等位基因数目。 n_{1j} 表示 N 个患者中等位基因 $j (1 \leq j \leq m)$ 的计数, n_{2j} 表示在未发病的同胞中该等位基因的计数。将资料整理为表 3。

检验统计量:

$$AC_i = \sum_{j=1}^m \frac{(n_{1j} - n_{2j})^2}{n_{1j} + n_{2j}} \quad (4)$$

服从自由度为 $m - 1$ 的 χ^2 分布 ($i = 1, 2$ 分别表示计数法 1 和 2)。

类似可以计算基因型计数时的检验统计量。该方法基于等位基因或基因型的计数或者检测数据的对称性, 使用排列过程来计算统计量的检验水准。

5. 同胞对内重复抽样法 (within sibship paried resampling method, WSPR): 为了能够充分利用多个同胞的信息同时又能避免同胞对之间的相关,

Rieger 等^[8]提出了 WSPR 方法。从每个家系中随机抽取一个患者和一个未发病同胞组成一个配对资料,这些配对资料构成这一同胞对的子集,然后采用条件 logistic 回归进行分析。重复抽样后得到若干个子集,将所有子集的结果求平均后进行参数估计和假设检验。

表3 表型不一致同胞对分析

同胞对编号	患者基因型	同胞基因型	等位基因计数			
			计数法 1		计数法 2	
1	1/1	1/1	1,1	1,1	-	-
2	1/1	1/2	1,1	1,2	1	2
3	1/1	2/2	1,1	2,2	1,1	2,2
4	1/2	1/1	1,2	1,1	2	1
5	1/2	1/2	1,2	1,2	-	-
6	1/2	2/2	1,2	2,2	1	2

WSPR 利用了同胞对当中所有对象的信息,而且可以适用于患病风险存在差别的各个家系资料。与其他方法不同,WSPR 可以进行参数估计和构建可信区间。

实例分析

王晓玲等^[9]在研究亚甲基四氢叶酸还原酶(MTHFR)基因 C677T 突变是否与冠心病连锁时,选择 40 个符合要求的同胞组资料进行同胞传递不平衡检验和同胞组不平衡检验(表 4)。

表4 MTHFR 基因 C677T 基因型在同胞组中的分布

家系序号	同胞数	患者			未受累同胞		
		CC	CT	TT	CC	CT	TT
001	2	0	1	0	0	0	1
002	5	0	1	0	3	1	0
003	3	0	1	0	0	0	2
...
040	5	0	2	0	1	1	1

对满足一组同胞中至少有 1 个患者及 1 个非患者并且各成员的基因型不完全相同的 40 个家系进行 s-TDT 和 SDT 分析,比较患者与非患者标记位点基因型分布,检验患者标记等位基因频率是否显著异于未受累同胞。s-TDT:由表 4,依据公式(1)、(2)和(3)计算后可得 $A = 54.38$ 、 $V = 8.20$ 、 $Y = 53$ 、 $Z = 0.27$ 、 $P > 0.05$,提示患者与非受累同胞 T 等位基因分布的差异无统计学意义。SDT:由表 4 可得 $b = 16$ 、 $c = 13$,计算 $\chi^2 = 0.31$ 、 $P > 0.05$,结果表明患者与非受累同胞 T 等位基因分布差异无统计学意义,提示 MTHFR 基因 677T 等位基因与冠心病不存在连锁。

讨 论

近年来基于家系以患者双亲、同胞等为对照的关联研究得到了普遍地应用。这是因为:①在遗传标志与疾病位点紧密连锁的时候,检测连锁很困难,但是用家系为基础的关联研究能够检测到关联;②能够避免由于人群分层混杂所引起的虚假关联。病例-同胞对照设计选择患者未患病的同胞作为对照,由于患者及对照具有相同的遗传背景,就避免了由于对照不具有相同的遗传背景而引起的混杂,即有效地避免了人群分层混杂现象。对于一些晚期发作的疾病,诸如 2 型糖尿病、阿尔茨海默病、心血管疾病、肿瘤等,要测定双亲的基因型是非常困难的,有时甚至是不可能的。

相对而言,收集患者同胞的资料较为容易。为此,Curtis^[4]提出了病例-同胞对照设计的一系列统计分析方法。病例-同胞对照设计中,患者是随机抽样得到的,对照是患者未患病的同胞。这样使用同胞作为对照一方面可以避免人群分层的问题,另一方面在实施过程中相对来说又易于实施。

同胞对关联检验法选择使用基因型与患者相差最大的一个同胞作为对照,且分析时类似于配对的方式,消除了多个同胞时的相关问题,能够避免由于人群分层混杂带来的假阳性结果。在统计分析上,该方法并不十分精确,所以效能并不是很高,因此在实际工作当中专门按照该方法进行设计和统计分析并不可取。但是,考虑到许多用于连锁分析的资料已经获得了患者及其同胞的基因型数据,此时可以结合使用同胞对关联检验法来检测连锁不平衡。

SDT 方法可以用来在有关联时检测连锁,也可以用来在有连锁时检测是否关联。该方法对于资料中每一个家系,比较患者和未患病同胞携带致病等位基因数的差别,可以视为一种非参数的秩和检验,分析时会损失一部分信息。但是该方法非常简单方便,而且在同一个家系中有多个患者及多个未患病同胞时也可有效地检测关联,而其他方法此时只能检测连锁而不能检测关联。

s-TDT 对于基因型资料有如下的要求:①在每个同胞对中至少有一个患者和一个未患病的同胞;②同胞对中所有个体的基因型不能完全相同。s-TDT 可以用来检测遗传标志是否与疾病连锁,也可以检测关联。在检测连锁时,s-TDT 与 SDT 方法一样,可以分析含有多个患者和多个同胞的同胞对;

在检测关联时, s-TDT 需要同胞对是最小配置的 (minimal configuration), 即每个同胞对中只含有一个患者及一个未患病同胞。Monks 等^[10] 研究发现当遗传标志位点有多个等位基因时, s-TDT 与上述两种方法相比具有较高的效能。

病例-同胞对照设计的优点: 首先, 对照的资料容易获得。对于许多晚期发作的疾病, 选择未发病的同胞比选择患者双亲容易。因此, 病例-同胞对照设计能够解决其他设计选择对照的困难。其次, 与传统的病例-对照研究选择无关联的对照不同, 病例-同胞对照设计选择患者的同胞作为对照, 因而可以避免人群分层所导致的偏倚。最后, 该设计可以增加资料的质量。由于同一家系里有多成员填写调查表, 因此研究者可以交叉验证患者和同胞所填调查表的某些信息是否一致。

病例-同胞对照设计的缺点: 没有同胞的患者不能入组; 同胞年龄相差较大时, 需要考虑对年龄协变量的控制; 容易匹配过度, 由于同胞对的基因来源于相同的亲代, 因而患者和未发病的同胞之间等位基因频率的差别要小于随机人群中抽取的发病和未发

病者之间的差别。

参 考 文 献

- Schaid DJ, Sommer SS. Genotype relative risks: methods for design and analysis of candidate-gene association studies. *Am J Hum Genet*, 1993, 53:1114-1126.
- Risch N, Merikangas K. The future of genetic studies of complex human diseases. *Science*, 1996, 273:1516-1517.
- Manuila A. Blood groups and disease-hard facts and delusions. *J Am Med Assoc*, 1958, 167:2047-2053.
- Curtis D. Use of siblings as controls in case-control association studies. *Ann Hum Genet*, 1997, 61:319-333.
- Spielman RS, Ewens WJ. A sibship test for linkage in the presence of association: the sib transmission/disequilibrium test. *Am J Hum Genet*, 1998, 62:450-458.
- Horvath S, Laird NM. A discordant-sibship test for disequilibrium and linkage: no need for parental data. *Am J Hum Genet*, 1998, 63:1886-1897.
- Boehuke M, Langefeldt CD. Genetic association mapping based on discordant sib pairs: the discordant-alleles test. *Am J Hum Genet*, 1998, 62:950-961.
- Rieger RH, Kaplan NL, Weinberg CR. Efficient use of siblings in testing for linkage and association. *Genet Epidemiol*, 2001, 20:175-191.
- 王晓玲, 顾东风, 孙峰, 等. 亚甲基四氢叶酸还原酶基因 C677T 突变与冠心病的连锁分析. *中华医学遗传学杂志*, 2001, 18:206-208.
- Monks SA, Kaplan NL, Weir BS. A comparative study of sibship tests of linkage and/or association. *Am J Hum Genet*, 1998, 63:1507-1516.

(收稿日期: 2005-09-08)

(本文编辑: 张林东)

· 疾病控制 ·

260 例流行性腮腺炎患者流行特征及临床分析

张敏 彭怡纯

为了解近年来流行性腮腺炎(流腮)流行特征及临床特点, 加强流腮的控制工作, 对 2004 年 1-12 月在武汉市两所医院住院的 260 例流腮患者资料进行分析。诊断标准主要根据流行病学史及临床表现, 部分患者检测血清中腮腺炎病毒 IgM 抗体。

结果与分析: 性别与年龄分布显示, 260 例患者中男性 215 例, 女性 45 例, 男: 女 = 4.78:1, 男性明显高于女性。年龄范围在 1~65 岁之间, 平均年龄 (9.89 ± 2.13) 岁。其中 6~12 岁小学生居多 (176 例, 67.69%), 1 岁以内婴儿最少 (2 例, 0.70%)。发病高峰季节主要集中在 4-8 月, 占 63.4% (165/260 例), 但全年均有散发性发病, 11-12 月有一个小高峰 59 例 (22.69%)。12 月中旬在汉阳某社区同一段时间, 12 名小学生发病。发病地区主要分布在武汉市七城区及周边郊县, 计 233 例 (82.69%), 散居儿童次之 35 例 (12.85%)。上述结果表明流腮为腮腺炎病毒引起的急性呼吸道传染病, 多发生于儿童及青少年。近年流行病学特征呈

高度散发与局部暴发并存的方式, 散发以幼儿和成人为主, 暴发以中小學生发病为主。临床表现病程最短 4 天, 最长 25 天, 平均 (11.23 ± 8.46) 天。主要症状以腮腺非化脓性炎症性肿大、发热、头痛明显的为 253 例 (97.31%)。并发症发生率较高 68.08% (177/260), 其中最多为脑膜脑炎或脑膜炎 25.77% (67/177), 其次为胰腺炎 20.77% (54/177), 颌下腺炎 12.69% (33/177), 睾丸炎 7.31% (19/177)。而且在发生并发症的 177 例患者中, 有 123 例 (69.49%) 为 1 种并发症, 45 例 (25.42%) 为 2 种并发症, 8 例 (3.08%) 为 3 种并发症, 1 例 (0.56%) 为 4 种并发症。260 例患者中发生 2 种以上并发症者占 20.77% (54/260), 说明腮腺炎病毒可侵犯腮腺以外的其他腺体及多种器官, 导致各种并发症发生, 多在腮腺炎同时或其后出现, 临床应密切观察, 及时处理, 避免漏诊及延误治疗。以上结果提示, 无论在流腮流行期间还是散发期, 各级医疗机构均应加强疫情监测, 尽早诊断, 及时隔离治疗, 是有效控制流腮暴发的关键。

(收稿日期: 2005-10-27)

(本文编辑: 尹廉)