

病例家属提供流行病学研究资料可靠性研究

陈 坤¹ 刘希永¹ 沈永洲² 沈高飞²

摘要 流行病学调查中不同的资料收集途径其在真实性和可靠性方面可能不尽一致,为评价病人及其代理人(如配偶)对问卷问题的回答一致性(可靠性),作者分别对 50 例肺癌病人及其代理人进行了面对面访问调查。结果显示,对计量资料两者回答的基本一致率为 84.1%,经配对 t 检验除吸烟开始年龄外,余均差异无显著性($P > 0.05$);对计数资料比较表明,平均一致率为 87.8%,各项目经 Kappa 一致性检验, P 均小于 0.001。结果与国外文献报道基本一致,并提示在实际工作中对死亡病例作为研究对象,则可以从其代理人那里采集所需的资料。

关键词 病例代理人 流行病学

A Study on the Reliability of Data Provided by Patients' Family Surrogates Chen Kun*, Liu Xi-yong, Sheng Yong-zhou, et al. * Zhejiang Medical University, Hangzhou 310013

Abstract In epidemiological investigations, the validity and reliability of data under different ways of collection might not be consistent. In order to evaluate the consistency of data collected from cases themselves or from their surrogates (for example, the spouse) through answering the same questionnaire, at personal interview, was carried out among 50 cases of lung cancer and their surrogates respectively. Other than the item starting age of smoking by a paired- t test, the results showed that between the two groups of quantitative data sources, the overall consistent rate was 84.1% and the differences were not significant ($P > 0.05$). When Kappa consistency test was used to the count variables, an average consistent rate of 87.8% was noticed, but P values were less than 0.001 for all items. Results from this study were quite similar to those with have been published abroad and suggested that it is credible to collect data from surrogates when dead cases were selected as research subjects.

Key words Surrogate Epidemiology

在恶性肿瘤及其它病死率较高疾病的流行病学、病因学等研究中常常需要从已死亡病例的代理人(surrogate),如配偶、兄弟姐妹、子女等那里获取资料,而存活病人则通常由病例自己提供研究所需的资料。这实际上是反映同一研究对象情况的两种资料收集途径。众所周知,流行病学调查中,不同的资料收集途径其在真实性、可靠性方面可能不尽一致^[1],特别在同一个调查研究课题中,由研究对象本人和其代理人提供的资料同时存在时,更应对这两种方法的一致性作出评价,这是对这类研究结果能够正确解释的前提。

我们通过 50 例肺癌病人及其代理人(家属)对问卷问题的回答一致性分析,以评价病例代理人(家属)提供资料的可用性,即可靠性。

材料和方法

一、研究对象的确定:50 例肺癌病人均为海宁市恶性肿瘤登记系统中确诊的现患病人,年龄在 46~78 岁,平均年龄为 64.82 岁;男性 44 人,女性 6 人。代理人按照以下原则选择:①为病例的家属;②如配偶健在则由其为代理人;③如配偶已亡故,则选择一起生活时间最长的子女;④如既无配偶又无成年子女,则选择病例的兄弟姐妹。

二、调查内容和方法:调查表经过预试验

1 浙江医科大学公共卫生学院 杭州 310031
2 海宁市肿瘤防治研究所

后拟出。主要内容包括:年龄、性别、肺癌诊断等一般情况、吸烟史、被动吸烟史、家族肿瘤史和本人呼吸系统疾病史,以及参与烹调 and 燃料使用类型等情况。调查员经统一培训,由预防医学系学生 4 人及乡村医生若干人承担,备有调查员手册,其对调查员的要求及调查表上每一问题的定义均作了明确的规定。调查方法采用面对面的询问调查,在实施调查时,做到病例及代理人相互严格回避的设计要求,每对病例及其代理人的询问均由同一调查员完成,以减少不同调查员间的误差。

三、统计方法:所有资料均经统一编码后,进入 IBM-586 微机分析,对计量资料分别计算了病例和其代理人两者回答的完全一致率以及 ≤ 2 个单位及 ≥ 2 个单位的高估和

低估率,并作两者回答一致性的配对 t 检验;对计数资料计算了完全一致率及其 95% 可信区间 (0.95CI), 并作 Kappa 一致性分析^[2]。

结 果

对统计分析的 10 项计量资料,代理人回答与病例自己回答结果完全一致的在 44.2%~92.0% 之间,平均为 73.4% (表 1), 但如将高估或低于 ≤ 2 个单位的当作基本一致的,则平均一致率为 84.1%, 其中最低的是吸烟年数项为 70%, 最高的如年龄、诊断年份项可达 100%。配对 t 检验示,除了吸烟开始年龄达到显著性差异外,其余均未见有显著性差异 ($P > 0.05$) (表 1)。

表 1 若干计量资料的一致性分析

项 目	例数	一致率 (%)	高估率 (%)		低估率 (%)		P 值*
			≤ 2 个单位	> 2 个单位	≤ 2 个单位	> 2 个单位	
年龄(岁)	50	86.0(43)	6.0(3)	-	8.0(4)	-	0.785
出生年(年)	47	91.0(43)	4.3(2)	2.1(1)	2.1(1)	-	0.497
出生月(月)	50	76.0(38)		24.0(12)**			0.059
诊断年(年)	50	92.0(46)	4.0(2)	-	4.0(2)	-	0.990
诊断月(月)	50	72.0(36)		28.0(14)**			0.429
吸烟开始年(岁)	43	53.5(23)	11.6(5)	-	11.6(5)	23.3(10)	0.010
吸烟年数	43	44.2(19)	9.3(4)	23.3(10)	16.3(7)	7.0(3)	0.803
被动吸烟年数	6	50.0(3)	-	50.0(3)	-	-	0.584
参与烹调(次/周)	47	70.2(33)	6.4(3)	10.6(5)	-	12.8(6)	0.302
参与烹调年数	46	71.7(33)	6.5(3)	10.9(5)	-	10.9(5)	0.281
合 计	432	73.4(317)	5.1(22)	5.6(24)	4.4(19)	5.6(24)	

* 配对 t 检验; ** 代理人较病例回答结果 ≤ 2 单位或 ≥ 2 个单位

对进入分析的 10 项计数资料,则代理人与病例回答的一致率要比计量资料平均高出 15%, 其中一致率最低的为吸烟状态有否改变项(86.1%), 最高的为婚姻状态、是否吸烟等项(100%) (表 2)。Kappa 值在 0.35~1.0 之间, 对 Kappa 值进行统计学显著性检验 P 值均小于 0.001, 说明两者一致性极佳(表 2)。

从表 1 和表 2 统计分析的项目内容看, 对于出生年月、婚姻状态、疾病诊断年份、诊断依据、家族肿瘤史等项目, 其一致性很好, 均为 90% 以上。而年龄的回答一致率较低(86%) 则可能是民间有虚岁、实岁计算的习惯有关。对出生、诊断等的确切月份则两者误差较大, 一致率分别为 76.0% 和 72.0%。

在吸烟有关的项目上, 对是否吸烟, 病例

与代理人的一致率达到了 100%，但对具体的吸烟情况两者回答存在一些差别，如吸烟年数完全一致仅为 44.2%，但将误差 ≤ 2 年合并作为基本一致，则基本一致率可达到 70% 左右，进一步分析吸烟情况的有关项目(表 1)，可以发现：①代理人(家属)在吸烟年数上趋于高估，即家属回答 ≥ 2 年的达 23.3%；②而在吸烟年龄趋于低估，即 ≥ 2 年低估率的亦接近于 1/4(23.3%)。其原因可能来自于两个方面：一是病例对这类不良行为有可能趋于部分隐瞒，其次是家属估计的错误。

就疾病史而言，对家族中有无其它肿瘤病人则两者回答基本一致，而对病例自身过去疾病史，则一致性稍低[对 13 例不一致者作原因分析显示：主要由于病人对此问题未作回答(9 例)而引起的]；对两者均回答有呼吸系统疾病史的 21 例进一步分析表明：其中 19 例在具体病名上都是一致的(90.5%)。

表 2 若干计数资料的一致性

项 目	例数	一致率		Kappa 值*
		%	0.95CI	
文化程度	50	82.0(41)	69.0~91.0	0.348
婚姻状态	50	100.0(50)	93.0~100.0	1.000
诊断单位	50	80.0(40)	66.0~90.0	0.709
诊断依据	50	90.0(45)	78.0~97.0	0.718
是否吸烟	50	100.0(50)	93.0~100.0	1.000
吸烟改变	43	86.1(37)	76.0~97.0	0.787
家族肿瘤史	50	90.0(45)	78.0~97.0	0.608
呼吸病史	50	74.0(37)	59.0~85.0	0.490
烹调史	50	88.0(44)	76.0~95.0	0.758
燃料类型	50	88.0(44)	76.0~95.0	0.746
合 计	493	87.8(433)	84.5~90.9	-

* 计算方法见参考文献^[2]。

讨 论

在流行病学、病因学以及临床等研究工作中，由于病例死亡或其它原因而只能从其代理人那里获得所需资料。80 年代，国外一些学者对病例与其代理人那里获得所需资料的可靠性作了报道^[3~9]，而国内则未见有类

似专门研究报告。

Pickle 等^[3]通过对 2 606 个不同代理人类型提供资料的可靠性分析后指出：对有关病例早年生活事件的最佳代理人为兄弟姐妹，而成年生活中事件的最佳代理人为其配偶或子女。本研究中若干项目的一致率偏低，可能与设计中采用代理人的原则以其配偶为主有关。

Lerchen 等^[4]于 1983 年报道了 105 名妻子对其丈夫有关情况所提供资料的真实性研究，其结果表明：在吸烟、饮食等方面，两种来源途径的资料具有可比性。妻子能正确地报告其丈夫的吸烟状态，但对吸烟量，妻子趋于报告一个整数，如每日一包等。Mclaughlin 等^[5,6]的研究则发现代理人在吸烟、饮酒的量上比起病例自己的报告来趋于高估，以及比病例自己报告有较多的疾病经历，本研究结果与这些国外类似研究结果基本一致(表 1 和表 2)。

对于频数(计量)指标问题，Humble^[7]和 Marshall 等^[8]分别有研究指出：不管代理人的性别如何，其完全一致性在 60%~80% 之间。他们对饮食营养素摄入量的调查中发现，如将饮食指标改为频数(如每周几次等)，则一致率可达到 95% 或更高，比较表 1 和表 2 结果则有类似情况。

文献指出^[8]，在使用代理人为资料来源途径时，虽其可靠性(与病例的一致程度)，并非绝对的好，但对揭示病例和对照的较大差异方面仍是恰当的。菲律宾的一项研究指出^[9]，陪同病人来就诊的人员可作为病例对照研究中良好对照来源。Gordis^[10]认为在病例对照研究中，如若采用了死亡病例，其对照并不需要死亡者，即由其代理人来回答问卷，这是因为代理人提供的资料虽未完全与病例自己提供一致，但使用死亡对照本身就会使研究资料中介入偏倚。另外需要指出，至今的这类研究主要揭示了代理人与病例自己报告资料的一致程度(可能性)，而较少能评价其真实程度，故在解释这些不一致情况时较

难区分出其是完全由于代理人的错误还是由于病例本身隐瞒或夸张的情况存在。从本研究结果来看,总体而言一致性较好,与文献报道相似^[3~9],在实际工作中,死亡病例作为研究对象为唯一途径时,则可由其代理人来提供研究所需资料,但最好能对这两种途径来源的资料作出评价。

参 考 文 献

- 1 Harlow SD, Linet MS. Agreement between questionnaire data and medical records. *Am J Epidemiol*, 1989, 129: 233.
- 2 赵跃进,王洁贞. 一致性检验方法及应用. *中华预防医学杂志*, 1991, 25:238.
- 3 Pickle LW, Brown LM, Blot WJ, et al. Information available from surrogate respondents in case-control interview studies. *Am J Epidemiol*, 1983, 118:99.
- 4 Lerchen ML, Samet JM. An assessment of the validity of questionnaire responses provided by a surviving spouse. *Am*

- J Epidemiol*, 1986, 123:481.
- 5 Mclaughlin JK, Blot WJ, Mehl ES, et al. Problems in the use of dead controls in case-control studies. I. General results. *Am J Epidemiol*, 1985, 121:131.
- 6 Mclaughlin JK, Blot WJ, Mehl ES, et al. Problem in the use of dead control in case-control studies. II. Effect of excluding certain cause of death. *Am J Epidemiol*, 1985, 122:485.
- 7 Humble CG, Samet JM, Skipper BE, et al. Comparison of self- and surrogate-reported dietary information. *Am J Epidemiol*, 1984, 119:86.
- 8 Marshall J, Priore R, Haughey B, et al. Spouse-subject interviews and the reliability of diet studies. *Am J Epidemiol*, 1980, 112:675.
- 9 Ngelangel CA. Hospital visitor-companion as a source of controls for case control studies in the Philippines. *Int J Epidemiol*, 1989, 18(suppl 2):S50.
- 10 Gordis L. Should dead cases be matched to dead control? *Am J Epidemiol*, 1982, 115:1.

(收稿:1996-03-04 修回:1996-05-17)

北京市 1995 年儿童肺炎支原体感染的流行病学分析

徐文华

张美和

由于肺炎支原体(Mp)感染在呼吸道感染中占很大比例,所以检测 Mp 抗体对临床诊断和治疗具有非常重要的意义。我们对北京市 1995 年 1~14 岁的患儿血清用 ELISA 方法检测 Mp 特异 IgM 抗体,结果报告如下。

体阳性率(37.6%、27.0%、43.5%)明显高于年总阳性率(24.1%),差异有显著性($P < 0.01$)。

一、材料与方 法:

2. 支原体抗体阳性率与年龄的关系:

1. 血清来源:采集 1995 年住院和门诊部分患儿的静脉血共 1 989 份,其中临床诊断肺炎 986 例。

(1)986 例 1~14 岁肺炎患儿支原体抗体阳性率以 8 岁、10 岁、12 岁儿童为高,与总阳性率无显著性差异。

2. 血清学检查方法:血清标本做 1:50 倍稀释,用 ELISA 方法进行检测。酶标仪波长 492nm,以肺炎支原体抗体阳性血清为参照血清,样品吸光度 A 值等于或大于阳性参照值的为阳性。ELISA IgM 试剂由首都儿科研究所细菌室提供。

(2)送检的 1 989 份血清中,支原体抗体阳性率以 1 岁(15.6%)最低,但与总阳性率(25.0%)相比差异有显著性($P < 0.05$)。

二、结果与分析:

3. 支原体抗体阳性率与性别的关系:986 例肺炎患儿支原体抗体阳性率男性 32.74%,女性 39%,送检的 1989 份血清中支原体抗体阳性率男性为 30.27%,女性为 36.4%,均无显著性差异。

1. 支原体抗体阳性率与季节的关系:

(1)对 986 例肺炎患儿血清检测表明,支原体抗体阳性率以 1 月(42.7%)、12 月(47.1%)为高,与年总阳性率(26.2%)差异有显著性($P < 0.01$)。9 月为 11.5%,明显低于年总阳性率($P < 0.01$)。

以上结果表明,对 1995 年度 1 989 份患儿血清进行的支原体抗体检测发现,支原体感染一年四季均可发生,但以冬季发生率最高,1990~1995 年文献报道二起支原体感染流行亦均发生在冬季,本文再次证明冬季是支原体感染的易发季节。在 2~14 岁年龄中,支原体抗体阳性率无显著性差异,而 1 岁与 1 岁以下阳性率偏低可能与免疫系统发育不成熟有关。

(2)1 989 份血清中 1 月、11 月、12 月支原体抗

(收稿:1996-05-24 修回:1996-08-28)