

EPICURE 在 X 射线诊断医生肿瘤队列调查中的应用

余宁乐, 陈连生, 王 进, 许翠珍

中图分类号: R818 文献标识码: B 文章编号: 1004-714X(2005)03-0204-02

【摘要】目的 运用 Epicure 数据处理系统对江苏省 1950~1996 年间医用诊断 X 射线工作者肿瘤流行病学调查资料进行数据处理和分析。方法 采用 Epicure 中的 Datab 模块计算了人年数, 联合 AMFIT 最终得出了队列观察期间全部肿瘤和部位别肿瘤发生的相对危险。置信区间估计采用  $\chi^2$  似然比法计算。结果 照射组全部肿瘤相对危险 1.2, 95% 可信区间 0.95~1.59, 其中女性乳腺癌的相对危险达 3.3, 95% 可信区间为 1.39~8.07。全部死因也有增高的趋势。结论 EPICURE 在危险因素研究中具有广泛的前景。

【关键词】X 射线工作者; 队列研究; 肿瘤发生

Epicure 是由美国国立癌症研究所的生物统计学家和流行病学家 Preston DL, Lubin JH 和 Pierce DA 等人在对日本广岛、长崎原子弹爆炸幸存者群组观察资料分析实践中编写的一个 MS-DOS 交互式程序集, 由风险建模及人年计算的五个模块组成, 即 GMBO、PECAN、AMFIT、PEANUTS、DATAB 等<sup>[1]</sup>。在国内主要用于广东阳江高本底地区人群辐射效应观察<sup>[2]</sup>、云南锡业公司矿工氡与肺癌调查等<sup>[3]</sup>, 我们运用 Epicure 数据处理系统结合标准统计分析软件 SAS 对江苏省 1950~1996 年间医用诊断 X 射线工作者肿瘤流行病学调查资料进行数据处理和分析, 采用 Epicure 中的 Datab 模块计算了人年数, 联合 AMFIT 最终得出了队列观察期间全部肿瘤和部位别肿瘤发生的相对危险<sup>[4]</sup>。置信区间估计采用  $\chi^2$  似然比法计算。

1 对象与方法

1.1 研究队列的确定 采用回顾前瞻性队列研究方法, 1981 年确定了江苏省范围内 1950 年~1980 年间在医院放射科工作的医用诊断 X 射线工作者 3 975 人, 作为观察组; 同所医院未从事过放射工作的内科、外科、五官科、儿科等医务人员 3 726 人作为对照组。

1.2 调查方法 根据全国统一方案, 采用回顾前瞻性队列研究方法。1981 年首先从各医院人事部门获得所调科室 1950 年 1 月 1 日至 1980 年 12 月 31 日期间工作人员名单, 然后对所列人员进行面询填表调查, 已调离、离退休及死亡人员予以追访。随后每 5 年随访 1 次, 共 3 次, 调查对象均以 1981 年调查名单为准, 走访医院医务、人事及被调查者所在科室, 登记被调查者在所调期限内工作调动、离退休、恶性肿瘤和死亡情况, 新患肿瘤和死亡者需填写肿瘤死亡登记卡, 并对以往的调查结果进行核查和校正。

1.2.1 观察指标选择 调查截止日期 1996 年 12 月 31 日, 对于肿瘤患者, 以首次诊断恶性肿瘤为观察终点。对诊断为恶性肿瘤者要求明确肿瘤诊断、诊断医院、诊断日期、诊断依据。对死亡者要求确定死亡时间、死亡的直接原因。肿瘤分类按 ICD-9 编码。

1.2.2 数据确认 对上报的流行病学调查表和肿瘤、死亡卡逐张检查, 对缺项、逻辑错误进行核实和纠正。

1.2.3 建立数据库, 输入数据 数据库系统选择 Microsoft Access 97 关系型数据库。

1.2.4 数据的录入与核对 数据录入采用人工键盘输入, 通过多次的人工核对, 数据逻辑检查保证录入质量, 确保统计关键字段没有错误。

2 数据处理与分析

作者单位: 江苏省疾病预防控制中心, 江苏 南京 210009  
作者简介: 余宁乐(1960~), 女, 浙江乐清人, 主任医师, 从事放射医学研究。

2.1 数据核查 我们对观察队列各变量值及人数、性别进行了列表核查, 结果见表 1、2、3。

表 1 照射组和对照组各变量值核查列表

变 量	人数	最小值	最大值
出生年	7 701	1 892.0	1 965.0
开始工作年	7 701	1 913.0	1 980.0
开始工作年龄	7 701	15.0	68.0
诊断年	312	1 954.0	1 996.0
诊断年龄	312	22.0	82.0
ICD 编码	312	141.0	208.0
诊断级别	7 701	0.0	3.0
死亡年	448	1 954.0	1 996.0
工作变动年	2 906	1 957.0	1 996.0
工作变动原因	2 906	1.0	4.0
生死代码	7 701	0.0	1.0
失访	103	1 993.0	1 993.0
进队列年	7 701	1 950.0	1 980.0
出队列年	7 701	1 954.0	1 996.0

表 2 照射组和对照组人数

组别	调查例数	百分比(%)	累积例数	累积频率(%)
对照组	3 726	48.4	3 726	48.4
照射组	3 975	51.6	7 701	100.0

表 3 照射组和对照组性别

性别	照射组	对照组	频数	百分比	累积频数	累积百分比
男	3 317	2 598	5 915	76.8	5 915	76.8
女	658	1 128	1 786	23.2	7 701	100.0

2.2 数据分析 用下述模型估计照射组的致癌危险:

$$\lambda_{sex, age} \times e^{exposed}$$

式中:  $\lambda$  代表对照组的癌症基线率, sex, age, exposed 分别是性别、到达年龄和是否为照射组。我们在计算基线率时考虑了年龄与性别之间的交互作用, 以及在高龄组可能出现的癌症率的低估(age\*age)。利用上述基本模型, 采用对数似然比  $\chi^2$  近似法( $\chi^2$  approximation to likelihood ratio test)检验了随访年代、开始暴露时年龄和开始工作年代的效应修正作用, 认为随访年代、开始暴露时年龄是效应修正因素, 开始工作年代不是一个明显的效应修正因素(见表 4)。

表 4 肿瘤相关因素对数似然比检验

变 量	LR 检验 P 值
随访年代	0.009 2
开始暴露时年龄	0.004 8
开始工作年代	0.269 6

对随访年代和开始受照年龄进行了分层调整,按不同随访年代、不同受照年龄等因素估算了主要部位别癌症的危险。采用对数似然比 $\chi^2$ 近似法( $\chi^2$  approximation to likelihood ratio test)计算了危险估计的双侧显著性检验和 95%置信区间估计。

3 结果与分析

3.1 队列基本情况 建立队列 7 701 人,1996 年底止,死亡 448 人,其中照射组 194 人,对照组 254 人,两组共失访 103 人。照射组平均年龄 51.0 岁,对照组 54.7 岁。

3.2 所有癌症 1950 至 1996 年照射组恶性肿瘤 147 例,对照组 165 例。照射组和对照组各恶性肿瘤的发生数、照射组各主要肿瘤相对危险和 95%可信区间见表 5。照射组全部肿瘤相对危险 1.2,95%可信区间 0.95~1.59,其中女性乳腺癌的相对危险达 3.3,95%可信区间为 1.39~8.07,明显高于对照组。除食道癌、结肠癌、和肺癌以外,照射组其他部位别癌症都有增高的趋势。全部死因也有增高的趋势,照射组的非癌疾病危险也值得研究。

表 5 照射组肿瘤及死亡相对危险

肿瘤名称	例数	RR	95% CI
全癌	147(312)	1.23	0.95—1.59
肝癌	34(59)	1.28	0.70—2.33
胃癌	23(55)	1.09	0.58—2.02
食管癌	9(20)	0.80	0.27—2.24
结肠癌	4(13)	0.51	0.11—1.92
直肠癌	8(12)	3.10	0.86—12.76
膀胱癌	6(13)	1.50	0.44—4.98
肺癌	13(43)	0.50	0.23—1.03
女性乳腺癌	14(24)	3.33	1.39—8.07
白血病	8(14)	2.64	0.78—9.24
实体癌	139(298)	1.19	0.91—1.54
全死亡	194(448)	1.20	0.97—1.49

注:括号内为两组发病总数。

各随访阶段全癌、白血病、乳腺癌、实体癌相对危险值均大于 1.0(表 6)。随着随访时间的延长,照射组的危险趋于稳定,置信区间变小。白血病的危险大约为 2.6,实体癌的危险为 1.2,乳腺癌的危险为 3.3。乳腺癌的危险最大。

表 6 照射组不同随访阶段的相对危险

肿瘤名称	1950~1980		1950~1985		1950~1990		1950~1996	
	例数	RR	例数	RR	例数	RR	例数	RR
白血病	3	—	3	2.88	6	3.00	8	2.64
实体癌	30	1.61	53	1.15	84	1.02	139	1.19
乳腺癌	4	8.98	8	3.66	9	2.46	14	3.33
全癌	33	1.59	56	1.20	90	1.07	147	1.23

3.3 恶性肿瘤与其他因素的关系 参加放射工作时年龄与相对危险检验(表 7)可见,白血病最大危险出现在参加工作年龄最小的组。在乳腺癌中,没有观察到这一现象。参加放射工作时期与相对危险检验(表 8)可见,乳腺癌的危险随参加工作的年代延迟,而表现得更加明显,但无显著意义。

表 7 照射组开始工作年龄与相对危险关系

暴露年龄(岁)	白血病		乳腺癌		实体癌		全癌	
	例数	RR	例数	RR	例数	RR	例数	RR
0~19	2	9.00	0	—	4	0.52	6	0.80
20~24	1	1.06	5	2.85	21	0.92	22	0.93
25~29	1	1.13	5	3.33	40	1.58	41	1.57
30	4	—	4	36.42	74	1.10	78	1.17
P 值	0.32		0.16		0.18		0.38	

表 8 照射组开始工作年代与相对危险关系

参加工作年代	白血病		乳腺癌		实体癌		全癌	
	例数	RR	例数	RR	例数	RR	例数	RR
<1960	2	1.76	4	1.31	41	1.02	43	1.04
1960~1969	4	3.72	4	4.24	29	1.22	33	1.35
1970~1980	2	—	6	5.00	69	1.34	71	1.36
P 值	0.44		0.13		0.72		0.51	

4 讨论

通过以上分析发现,EPICURE 在危险因素研究中具有广泛的前景。它能运算大规模的队列研究数据,为队列资料的分析提供了更多的手段。由于放射科医师以男性为主,观察队列男性与女性人数比为 3.31,人数之比为 3.26。我国影像诊断在上世纪 70 年代后得以迅速发展,因此,放射组 73.0%为 70 年代后参加工作,而对照组同期只有 25.8%,造成放射组入年数(97 102)低于对照组入年数(120 269)。此外,80 年代以前放射工作人员多为普通医生转行,参加工作年龄与对照组相比偏大<sup>[4]</sup>,由此造成两组性别、年龄分布的差异 EPICURE 给予了修正。

在过去用于分组资料的标准模型有 Log—Linear 和 logistic 等回归模型<sup>[5,6]</sup>。但考虑到在大多数情况下,相乘性假设并不一定能够较好地描述暴露和危险度之间的关系<sup>[7]</sup>,公共卫生方面有许多问题属相加作用形式<sup>[8]</sup>,对于群组资料,Poisson 模型更合适<sup>[9]</sup>。只有根据数据特点,采用合适的模型,才能真正地揭示疾病和危险因素之间的关系。EPICURE 可以拟合 Poisson 模型,也可以拟合 Log—Linear 和 logistic 等回归模型,能够较好地满足上述要求。

危险度估计是流行病学资料分析的重点,旨在分析暴露剂量与癌症发生之间的数量变化关系,并分析其他因素对这一关系的影响。EPICURE 是专门为流行病学资料统计分析的需要编制的模型分析软件。江苏省医用诊断工作者 46 年(1950~1996)间肿瘤流行病学调查资料采用了 EPICURE 统计分析技术,深化了分析深度,增加了结果的可比性,提高了分析结果科学意义。

(中国疾病预防控制中心辐射防护与核安全医学所孙全富副研究员给予了指导和帮助,在此致谢!)

参考文献:

[1] Preston DL, Lubin JH, Pierce DA. Epicure user's guide Hirosft International Corp[ CP/CD ]. 1992. 43—49.  
[2] 孙全富, 邹剑明, 刘玉升, 等. 阳江高本底地区居民健康流行病学调查数据库与统计分析方法[ J ]. 中华放射医学与防护杂志, 1997, 17(6): 381.  
[3] 张颖, 倪宗瓚, 姚树祥, 等. 广义 Poisson 回归模型及其应用[ J ]. 中国卫生统计, 2000, 17(3): 130.  
[4] 余宁乐, 王进, 许翠珍, 等. 江苏省医用诊断 X 射线工作者肿瘤流行病学调查[ J ]. 中华放射医学与防护杂志, 2001, 21(4): 307.  
[5] Holford TR. The analysis of rates and of survivorship using log—linear models[ J ]. Biometrics 1980; 36: 229.  
[6] Mantel N, Haenszel W. Statistical aspects of the analysis of data from retrospective studies of diseases[ J ]. J Natl Cancer Inst. 1959; 22: 719.  
[7] Berry C. Dose—response in case—control studies. J Epidemiolommunity Health 1980; 34: 217.  
[8] Thomas DC. General relative risk models for survival time and matched case—control analysis[ J ]. Biometrics 1981; 37: 673.  
[9] 孙全富, 邹剑明. 放射流行病学群组研究资料 Poisson 回归分析及其进展[ J ]. 中华放射医学与防护杂志, 1998, 18(6): 445—447.

(收稿日期: 2004—11—14)