

胃癌个体危险度及人群筛检 数量化评价的研究

陈坤 俞维萍 朱益民

【摘要】 目的 建立胃癌个体危险度判定及人群筛检的数量化模型。方法 在人群为基础的胃癌危险因素病例对照研究的基础上筛选出有统计学意义变量,应用概率论、模糊数学等原理和方法对胃癌危险因素和保护因素数量化,建立用于个体胃癌危险度测量和人群筛检的数量化评价方法(数学模型)。采用两种不同的加权系数计算方法,分别确定该数量化方法的判别阈值为0.20和0.17,对63例胃癌患者和693名正常人进行实际判别。结果 灵敏度和特异度均分别达到69%和63%左右,加权系数计算方法不同并不影响实际判别效果。结论 建立的胃癌个体危险度及人群筛检的数量化模型灵敏度和特异度尚理想,方法简便、易行、经济,有利于获得满意的受检率;可用于个体和群体(社区)的胃癌危险度测定。

【关键词】 胃肿瘤;筛检;数学模型;数量化评价

Quantitative assessment of individual risk and population screening program on gastric carcinoma
CHEN Kun, YU Wei-ping, ZHU Yi-min. School of Public Health, Zhejiang University, Hangzhou 310031, China

【Abstract】 Objective To set up a mathematic model for determining individual risk and population risk of gastric carcinoma. **Methods** Based on principles and methods of probability and fuzzy mathematics, a case-control study was quantified and a mathematic model for the screening of gastric carcinoma was set up. Using two different calculation methods of weight coefficients, the thresholds were identified as 0.20 and 0.17, respectively. This mathematic model was then used to examine 63 cases and 693 normal persons. **Results** The sensitivities and specificities were 69% and 63%. The different calculative methods of weight coefficients did not affect the results of the identification. **Conclusion** The sensitivities and specificities were satisfactory indicating that it was convenient, feasible, economic and could meet the contented screening rate. It could be used to determinate the risk for gastric carcinoma both on individual and on population.

【Key words】 Gastric neoplasms; Screening; Mathematic model; Quantitative assessment

胃癌是常见恶性肿瘤之一,我国胃癌粗死亡率为25.16/10万,占全部恶性肿瘤死亡的23.24%,居第一位^[1]。降低胃癌死亡率关键在于早发现和早治疗。在人群中进行筛检是使胃癌患者得到早期诊断和治疗的有效方法,但以何种手段进行筛检值得探讨。任何一种筛检方法不仅要求漏诊(假阴性率)低,更重要的是要简便易行,易为受检者接受,获得满意的受检(应答)率^[2]。目前用于早发现和早诊断胃癌的方法尚较为繁复或灵敏度、特异度不够理想。自20世纪80年代起,国内、外开始有学者应用数量化评估方法进行肿瘤危险度评估^[3-5]。我们试图在人群病例对照研究的基础上,应用现代生物数学技术,建立胃癌个

体危险度及人群筛检的数量化方法(数学模型),以此作为第一线的筛检方法,并对所建立的数学模型的真实性和实用性分别作了实际应用评价。

资料与方法

1. 病例对照研究:1999年在浙江省洞头县居民中进行了基于全人群的,按同性别、同年龄组和同居住地等条件1:3配比设计的胃癌危险因素病例对照研究,共调查确诊胃癌患者66例,正常人对照198名。所有病例是1990~1997年间的胃癌病例,其中病理诊断36例,占55%;内窥镜、手术诊断25例,占38%;以影像、临床表现诊断9例,占8%。研究因素包括:一般情况、饮食、饮水史、个人习惯、医疗疾病史及家族史等。经过该项快速流行病学评价

方法的研究,采用 t 检验、 χ^2 检验及秩和检验单因素统计学分析后,确定了该县人群中胃癌的危险因素和保护因素。在 $\alpha=0.10$ 水平上,与胃癌有关的危险因素有:重体力劳动时间(>2 h/d)、常吃小黄鱼(干)、常吃虾皮(鲜)、常吃虾蛄、常吃螃蟹、生气后立即吃饭和母亲有胃病等,而配偶健在、冰箱使用和常吃辣味食品等则为保护因素(表 1)。

表1 胃癌危险因素和保护因素的 OR 值和 95% CI 及其加权系数

变 量	OR _i 值(95% CI)	α_i (方法 1)	α_i (方法 2)
重体力劳动时间 (>2 h/d)	2.00(1.14~3.54)	0.084	0.439
常吃小黄鱼(干)	1.5X(1.10~2.09)	0.064	0.490
常吃虾蛄(干)	6.1X(1.15~32.66)	0.257	0.782
常吃虾皮(鲜)	1.26(0.99~1.60)	0.053	0.462
常吃虾蛄(鲜)	1.70(0.92~3.13)	0.071	0.429
常吃螃蟹	1.76(1.00~3.09)	0.074	0.451
母亲有胃病	5.51(0.95~32.06)	0.231	0.774
生气后即吃饭	2.07(1.31~3.27)	0.087	0.522
配偶健在	0.89(0.81~0.99)	0.047	0.819
使用冰箱	0.47(0.19~1.15)	0.089	0.399
常吃辣味食品	0.52(0.28~0.97)	0.081	0.438

2. 数学表达式的建立:在上述病例对照研究的基础上,选择若干在统计学上有显著意义的因素(危险因素和保护因素),应用模糊数学和概率论的一些原理和方法^[6],提出以下胃癌的数量化评估方法^[7]将人群特征空间分成胃癌和非胃癌两个模糊子集,对每一个观察对象由隶属函数确定其所属子集。该决策的概率应最大或错误概率最小。设 \hat{A} 为罹患胃癌的一个模糊子集,为建立数学模型,首先须确定 \hat{A} 的“标准”因素群(U_i), U_i 具有的因素是胃癌(\hat{A})病例中最“典型”的,设 U_i 的加权和值($P \hat{A}$)为:

$$P \hat{A} = \sum_{j=1}^n \alpha_j C_j \quad (1)$$

式中, $j=1, 2, 3, \dots, n, m$ 为因素群中因素(F_i)的个数; C_j 为各因素的判别分数,具有因素 F_i 时,不管 F_i 为危险因素抑或保护因素, C_j 均取值 1; α_j 为因素 F_i 的加权系数,由此, $P \hat{A}$ 为当全部因素(F_i)具有时的加权和值。

以 $P \hat{A}$ 代表具有 F_i 中若干或全部因素的某个受检者的加权和值,即:

$$P \hat{A} = \sum_{j=1}^m \alpha_j C'_j (j=1, 2, 3, \dots, m) \quad (2)$$

式中的 α_j 同前,为各因素 F_i 的加权系数。当受检

者具有全部因素中某个因素 F_i 时,如 F_i 为危险因素, C'_j 取 C_j 的值(取值为 1);如 F_i 为保护因素,则 C'_j 取 0 值;反之,当受检者不具有因素 F_i 时,则 C'_j 相应分别取 0 或 1 值。

由上述 $P \hat{A}$ 和 $P \hat{A}$ 两个加权和值,可建立 \hat{A} (人群或个体)的隶属函数:

$$\mu \hat{A}(U_i) = \frac{P \hat{A}}{P \hat{A}} \quad (3)$$

式中, $\mu \hat{A}(U_i)$ 是 \hat{A} 隶属函数的表达式,每个受检者都可由(3)式求得一个具体的数值,即隶属度(AD),其数值在 0~1 之间波动。AD 值越大,其属于罹患胃癌模糊子集的可能性越大。

3. 各因素加权系数的确定:上述(1)(2)式中各因素(F_i)的加权系数(α_j)在本研究中采用两种计算方法。在方法 1 中,以各因素 F_i 的 OR 值(如为保护因素则取 OR 值的倒数,即 $1/OR$)与具有所有相关因素总 OR 值的比值作为各因素的加权系数,即

$$\alpha_i = \frac{OR_i \text{ (或 } 1/OR_i)}{\sum OR_i}$$

在方法 2 中以各因素的信息熵计算加权系数,计算公式为:

$$\alpha_i = \frac{H(F_i/D)}{H(F_i/D) + H(F_i/\bar{D})}$$

式中, $H(F_i/D)$ 为病例条件下 F_i 的信息熵, $H(F_i/\bar{D})$ 为正常人条件下 F_i 的信息熵。

$$H(F_i/D) = -P(F_i/D) \log_2 P(F_i/D)$$

$$H(F_i/\bar{D}) = -P(F_i/\bar{D}) \log_2 P(F_i/\bar{D})$$

$P(F_i/D)$:在患大肠癌(D)的条件下具有 F_i 的概率($i=1, 2, \dots, m$); $P(F_i/\bar{D})$:在不患大肠癌(\bar{D})的条件下具有 F_i 的概率($i=1, 2, \dots, m$)。

各因素按上述两种方法计算的加权系数如表 1 所示。

结 果

按上述方法建立的胃癌个体危险度及人群筛检数学模型,用于来自洞头县的 63 例胃癌患者和 693 名正常人,按(3)式对每一个对象进行 AD 计算,这些病例与前参加病例对照研究的 66 例病例无关,均经医院确诊。表 2 示加权系数 α_j 分别以方法 1 和方法 2 计算时,判别的灵敏度、特异度和 Youden 指数随 AD 阈值变化的变化趋势。虽然 α_j 的计算方法不同,但两种方法的灵敏度、特异度和 Youden 指数的

变化趋势相似,判别效果比较一致。由于 AD 值为一连续变量,在实际应用中,可根据实际需要,确定判别的阈值。如用于筛检患者时,可考虑将阈值定得低一些,以增加灵敏度,提高检出率;而用于鉴别诊断时,则可将阈值右移,增加特异度,减少误诊率。

综合考虑胃癌人群筛检时的灵敏度和特异度,为使两者达到最大(Youden 指数达到最大),将两种加权系数(α_j)计算方法所得的 AD 值分别设定为 0.20 和 0.17,作为判断阈值,则洞头县的 63 例胃癌患者和 693 名正常人的实际判别结果如表 2 注中所示。对上述 2 种方法作出的判别结果(Youden 指数)作显著性检验,差异并无统计学显著性意义($P>0.05$)。可见,以 Youden 指数达到最大为确定判别阈值的依据,由不同的 α_j 计算方法所得的诊断价值结果趋于一致。

讨 论

以往,数量化的方法多用于临床疾病的鉴别诊断^[6]。80 年代,国外出现了基于流行病学研究结果的健康损害/健康危险性(health hazard/health risk)的评价方法^[3]。至 90 年代初,美国 Arizona 大学发展了一个数量化记分的癌症危险度评价工具^[4],主要用于患者的健康教育。随后也有数量化方法用于老年患者的肿瘤死亡危险评价^[8]、干预试验效果评

价^[9]等,但少见用隶属函数和隶属度来判别的。国内在 80 年代末,陈坤^[7]报道了基于模糊数学隶属度判别用于人群大肠癌筛检的数量化方法,该方法随后在一个 6 万余人口的普通人群中应用被证实为一种简便、有效和非常经济的筛检方法^[10]。该数量化方法也被应用于肝癌危险度评价,并以三种不同的加权系数计算方法分别进行实际应用评价,提出加权系数计算方法不同并不影响实际判别效果^[11]。

目前胃癌的早发现、早诊断和早治疗的筛检方法,尚主要依赖于胃镜或胃液检查。丘新尧等^[12]曾应用胃癌高危因素,结合早期症状、体征及隐血珠试验等按 Bayes 分类器原理建立模式识别程序,用于现场普查取得了一定成效,但较为繁杂,费用高,作为筛查手段不易推广。我们介绍的胃癌个体危险度及人群筛检的数量化方法则具有简便、易行、经济和无创伤等优点。只要让每一位受检者填一张统一编制的简单询问调查表就可以进行判别,计算方法较之回归判别、Bayes 判别等传统数理方法简单易懂,计算的工作量也已被减轻到最低限度。而且数量化方法的诊断价值相对较高,实际适用能力较好^[10,11]。本研究对胃癌的判别灵敏度和特异度分别达到了 69% 和 63% 左右(表 2),虽不是特别理想,但反映出的实际适用能力尚可。

表2 63 例胃癌病例和 693 名正常人的实际判别结果

AD 阈值	方 法 1					方 法 2				
	真阳性例数	真阳性人数	灵敏度	特异度	Youden 指数	真阳性例数	真阳性人数	灵敏度	特异度	Youden 指数
0.10	61	76	0.968	0.110	0.078	54	231	0.857	0.333	0.190
0.12	54	232	0.857	0.335	0.192	51	311	0.810	0.449	0.259
0.14	53	245	0.841	0.354	0.195	46	369	0.730	0.561	0.291
0.16	51	294	0.810	0.424	0.234	45	406	0.714	0.586	0.300
0.17	48	368	0.762	0.531	0.293	43	435	0.683	0.628	0.311*
0.18	45	413	0.714	0.596	0.310	32	482	0.508	0.696	0.204
0.19	45	425	0.714	0.613	0.327	29	510	0.460	0.736	0.296
0.20	44	440	0.698	0.635	0.333*	29	517	0.460	0.746	0.206
0.21	39	491	0.619	0.709	0.328	29	531	0.460	0.766	0.226
0.22	36	509	0.571	0.734	0.305	27	563	0.429	0.812	0.241
0.23	29	535	0.460	0.772	0.232	24	586	0.381	0.846	0.227
0.24	28	544	0.444	0.785	0.229	21	600	0.333	0.866	0.183
0.25	25	584	0.397	0.843	0.240	20	611	0.317	0.882	0.199
0.26	25	593	0.397	0.856	0.253	20	617	0.317	0.890	0.207
0.27	21	611	0.333	0.882	0.210	20	619	0.317	0.893	0.210
0.28	20	617	0.317	0.890	0.207	15	648	0.238	0.935	0.173
0.30	19	631	0.302	0.911	0.213	13	660	0.206	0.952	0.158
0.32	14	661	0.222	0.954	0.176	12	665	0.190	0.960	0.150
0.34	14	663	0.222	0.957	0.179	8	676	0.127	0.975	0.102
0.36	10	680	0.159	0.981	0.140	7	682	0.111	0.984	0.095

* 两种方法最大 Youden 指数经 u 检验, $u = 0.048 < 1.96\alpha (u_{0.05})$, 故 $P > 0.05$, 差异无显著性

我们发现,虽然 α_j 的计算方法不尽相同,但判别效果则比较一致,随 AD 阈值取值不同,两种方法灵敏度、特异度和 Youden 指数均有相似的变化趋势(表 2)。由(1)(2)和(3)式可知,其 AD 值越大,则罹患胃癌的危险度越大,以 Youden 指数达到最大为确定判别阈值的依据,则由不同的 α_j 计算方法所得的诊断价值结果趋于一致。证实陈坤提出的在建立个体危险度及人群筛检的数量化评价模型时,加权系数的计算方法,并非为一个重要的影响判别结果的因子^[11]。由于 AD 值为一连续变量,在实际应用中,可根据实际需要,确定判别的阈值。

总之,本研究结果表明,在胃癌人群筛检中引入数量化方法和隶属判别是一有意义的尝试。本数量化方法可作为第一线(初筛)手段在居民中进行胃癌危险度评估。理论上说,本研究所建立的数量化模型可直接应用与人群和个人,也可用于以医院为基础的人群。不过考虑到胃癌发生是多种影响因素综合作用的结果,不同地区不同人群胃癌发生的影响因素可能不同,应用模型前,最好先进行基于本地区本人群的病例对照研究,以获得相应的危险因素和保护因素。随着对胃癌病因研究的发展及不断现场试验,这样的数量化方法将更趋完善。

参 考 文 献

1 周有尚,张思维.我国人口死亡和恶性肿瘤死亡情况分析.中国

- 肿瘤,1997,6:9-12.
- 2 王志瑾.筛检.见:谭红专,主编.现代流行病学.北京:人民卫生出版社,2001.160-170.
- 3 Wagner EH, Beery WL, Schoenbach VJ, et al. An assessment of health hazard/health risk appraisal. Am J Public Health, 1982, 72: 347-351.
- 4 Brown SL. Quantitative risk assessment of environmental hazards. Annual Review of Public Health, 1985. 6247-6267.
- 5 Lippman SM, Bassford IL, Meyskens FL. A quantitative scored cancer-risk assessment tool: its development and use. J Cancer Education, 1992, 7: 15-36.
- 6 临床模糊分类和诊断.见:周怀梧,孙维民,主编.临床计量医学.上海:上海医科大学出版社,1999. 51-72.
- 7 陈坤.人群大肠癌筛检的一个数量化方法.浙江医科大学学报, 1988, 17: 49-52.
- 8 Louise CW, Kenneth EC. Cancer screening in elderly patients: a framework for individualized decision making. JAMA, 2001, 285: 2750-2756.
- 9 Sara W, Deborah PM, Melva TR, et al. Nutrition and physical activity interventions to reduce cardiovascular disease risk in health care setting: a quantitative review with a focus on women. Nutrition Reviews, 2001, 59: 197-214.
- 10 陈坤,焦登鳌,余海,等.人群大肠癌筛检数量化方法的应用研究.中华肿瘤杂志,1993,15: 37-39.
- 11 陈坤,朱益民,张扬,等.肝癌个体危险度评价及人群筛检数量化的评价研究.生物数学学报,2002,17: 560-565.
- 12 丘新尧,史奎雄,施榕,等.胃癌的筛检——模式识别程序建立及现场应用.中国卫生统计,1994,13: 19-50.

(收稿日期:2002-06-28)

(本文编辑:张林东)

· 疾病控制 ·

173 例老年涂阳肺结核病例转归特征分析

杨国文

随着人口老龄化进程的加快,肺结核已成严重危害老年人群健康的疾病之一。为探讨其转归特征,分析评价防治效果,改进防治措施,提高老年人群的生活质量,分析了 173 例新登老年涂阳肺结核病例转归特征。

江川县 1983~2001 年老年涂阳结核患病率 67.5/10 万,占总涂阳结核病 533 例的 32.5%;男女患病比值 1.08:1;患病最大年龄 86 岁,76 岁以上组患病 8.7%(15 例),65 岁组患病 43.9%(76 例),说明该年龄段的老年结核患者仍是社会的主要成员。坝区乡镇人群患病 87.3%(151 例),山区、半山区患病 12.7%(22 例);工人患病 4.6%(8 例),干部患病 5.8%(10 例),农民患病 89.6%(155 例)。提示应加强对山区、半山区结核患者的发现及治疗管理工作,加大对结核

病防治知识宣传工作的力度,其次是结核病防治工作的重点仍在农村。在 173 例老年涂阳结核病例中,首诊村卫生所和乡镇医院的 65.9%(114 例),首诊县医院、中医院放射科的 12.7%(22 例),直接就诊县结核病防治专科的 21.4%(37 例);上述病例出现症状(咳嗽、乏力、消瘦等)到就诊平均延误时间(86±7)天。提示合并感染的机会增加。老年涂阳肺结核短程化疗具有“三高一低”的特征,即阳性率 8.1%(14 例),死亡率 15.6%(27 例),其中结核死亡 6.4%(11 例),肺癌死亡 4.1%(7 例),非结核死亡 5.2%(9 例);完成治疗率 13.9%(24 例),治愈率仅为 59.54%(103 例)。这说明交通和经费是影响老年涂阳结核患者落实 DOTS 督导化疗方案的关键因素。

(收稿日期:2002-12-25)

(本文编辑:张林东)