

2003~2007 年中国恶性肿瘤发病率水平的聚类分析

Clustering Analysis on Incidence Rates of Cancer in China from 2003 to 2007

ZHANG Xiao,ZHOU Qin,CHEN Qing,et al.

张 晓^{1,2},周 琴¹,陈 清²,林国桢¹

(1.广州市疾病预防控制中心,广东 广州 510080;

2.南方医科大学公共卫生与热带医学学院,广东 广州 510515)

摘 要: [目的] 利用聚类分析方法对中国 32 个肿瘤登记处的恶性肿瘤发病率资料进行综合评判分类,进一步了解全国恶性肿瘤的地区分布特征。[方法] 使用 SPSS13.0 对《中国癌症发病与死亡 2003—2007》中 32 个肿瘤登记处的恶性肿瘤发病率进行聚类分析。[结果] 当各肿瘤登记处按总发病率水平分成 4 类时,磁县、阳城县、涉县、扬中市和盐亭县聚为一类,其发病率为 226.02/10 万~267.03/10 万,明显高于其余三类;大连市、建湖县、林州市和广州市聚成一类,肥城市、中山市、海门市、大丰市、长乐市、嘉善县、哈尔滨市南岗区、武汉市、沈阳市、启东市、淮安市楚州区、杭州市、本溪市、上海市、嘉兴市和鞍山市这 16 个登记处聚成第三类,海宁市、马鞍山市、北京市、金坛市、临朐县、四会市和扶绥县恶性肿瘤发病水平低聚为一类,其发病率为 96.43/10 万~119.37/10 万。[结论] 中国癌症的发病存在明显的高发区和理论上的低发区,较多地区的发病率接近全国平均水平。

关键词: 恶性肿瘤;发病率;聚类分析;地区分布;中国

中图分类号: R73.31 **文献标识码:** A **文章编号:** 1004-0242(2013)01-0013-05

20 世纪 70 年代以来,我国癌症发病及死亡率一直呈上升趋势,癌症高发地区多在农村和西部地区,危害尤为严重,是当地农民因病致贫及因病返贫的重要原因^[1]。2003~2007 年全国 32 个肿瘤登记处主要分布在沿海地区,包括 7 个癌症高发地区,即启东市、嘉善县、长乐市、林州市、盐亭县、扶绥县和临朐县。本研究旨在利用聚类分析(clustering analysis)的统计学方法对 32 个登记处进行综合评判分类,验证和发现恶性肿瘤高发区,以了解我国恶性肿瘤发病的现状和分布规律,为进一步开展分析流行病学调查研究提供重要线索,对当地制定恶性肿瘤防治策略提供参考资料和决策依据^[2]。

聚类分析是将随机现象归类的统计学方法,已广泛应用于医学科学研究之中。聚类分析属于探索性统计分析方法,按照分类目的分为两大类,分别是 R 型聚类(又称指标聚类)和 Q 型聚类(又称样品聚类);也可分为系统聚类(层次聚类)、动态样品聚类(K-Means 聚类)和有序样品聚类^[3]。聚类分析具有探

索性研究的特性,能够快速识别潜在的空间区域,可以作为一种筛选工具,用于识别候选癌症的一种映射分析^[4]。

1 资料与方法

1.1 资 料

来源于全国 32 个肿瘤登记处癌症发病率资料,摘自《中国癌症发病与死亡 2003—2007》,32 个登记处分布在 16 个省(自治区、直辖市),其中地级以上城市 14 个,县和县级市 18 个,2003~2007 年登记覆盖人口 255 430 909 人年,其中男性 129 242 574 人年,女性 126 188 335 人年,性别比 1.02^[5]。对于同一监测点 5 年数据的处理,分析 5 年发病率的波动性,全国肿瘤登记中心利用数据库和分析软件对全国 44 个登记处上报的 2003~2007 年数据,分别计算病理学或组织学诊断比例、只有死亡医学证明书比例、死亡/发病比、其他或未指明部位及原发部位不明(继发)恶性肿瘤比例、发病(死亡)率、标化发病(死

收稿日期:2012-06-12;修回日期:2012-09-14

通讯作者:林国桢,E-mail: szk@gzcdc.org.cn

亡)率、构成比等指标,对登记数据的完整性及有效性进行了全面的审核和评价,根据审核结果,最终接受了其中质量较好的 32 个登记处的数据。

1.2 方 法

层次聚类分析,又称系统聚类,是将相似的样品或变量归类的最常用方法。聚类过程如下:开始将各个样品(或变量)独自视为一类,即各类只含一个样品(或变量),计算类间相似系数矩阵,其中的元素是样品(或变量)间的相似系数。相似系数矩阵是对阵矩阵;将相似系数最大(距离最小或相关系数最大)的两类合并成新类,计算新类与其余类间相似系数;重复第二步,直至全部样品(或变量)被并为一类。马氏距离的计算公式为 $d_{ij} = X'S^{-1}X$, 其中向量 $X = (X_{i1} - X_{j1}, X_{i2} - X_{j2}, \dots, X_{im} - X_{jm})'$ [4]。聚类完成后,用 χ^2 检验各类间是否具有统计学差异。

本研究将每个登记处视为一个样本,其 2003~2007 年恶性肿瘤的中国标准化发病率(以下称中标率)进行 Q 型聚类,本研究进行两次聚类分析,包括 2003~2007 年全国 32 个登记处癌症总发病率、12 种主要癌症(包括肺癌、胃癌、结直肠癌、肝癌、乳腺癌、食管癌、胰腺癌、膀胱癌、脑瘤、淋巴瘤、宫颈癌、鼻咽癌)的发病率。采用自由聚类分析法对 2003~2007 年全国 32 个肿瘤登记处的癌症总发病率进行聚类分析,类间距离采用组间平均距离法,不进行标化。对 12 种主要癌症发病率采用层次聚类法,类间距离采用离差平方和法(Ward's method),为了消除不同变量变异系数大幅波动对聚类结果的影响,有必要在聚类分析中对数据进行标准化处理,此处采用 Z scores 标准化。具体聚类过程采用 SPSS13.0 完成 [5]。

2 结 果

2.1 2003~2007 年全国 32 个肿瘤登记处癌症发病率情况

总体发病率按年龄构成调整后,最高的是阳城县,其次是磁县和扬中市。12 种主要癌症发病率最高的前三个地区:肺癌为哈尔滨市南岗区、广州市、磁县;胃癌为涉县、扬中市和阳城县;结直肠癌为广州市、上海市和沈阳市;肝癌为扶绥县、启东市、盐亭县;乳腺癌为广州市、大连市和上海市;食管癌为磁

县、阳城县、扬中市;胰腺癌为嘉善县、嘉兴市和海宁市;膀胱癌为大连市、鞍山市、本溪市;脑瘤为大连市、嘉善县、杭州市;淋巴瘤为广州市、杭州市、上海市;宫颈癌为阳城县、涉县、扬中市;鼻咽癌为四会市、中山市、广州市(Table 1)。

2.2 总发病率聚类分析结果

根据发病率水平高低进行分类,分为两类时,磁县、阳城县、涉县、扬中市和盐亭县首先聚成一类,其发病率 226.02/10 万~267.03/10 万,明显高于全国平均水平,其余地区分为一类。继续分类,分为四类时,第二类包括大连市、建湖县、林州市、广州市,发病率 165.10/10 万~191.82/10 万;肥城市、中山市、海门市、大丰市、长乐市、嘉善县、哈尔滨市南岗区、武汉市、沈阳市、启东市、淮安市楚州区、杭州市、本溪市、上海市、嘉兴市和鞍山市 16 个登记处聚成第三类,发病率 135.32/10 万~155.77/10 万;第四类包括海宁市、马鞍山市、北京市、金坛市、临朐县、四会市、扶绥县,发病率 96.43/10 万~119.37/10 万(Figure 1);差异有显著性($\chi^2=4911.673, P<0.05$)(Table 2)。

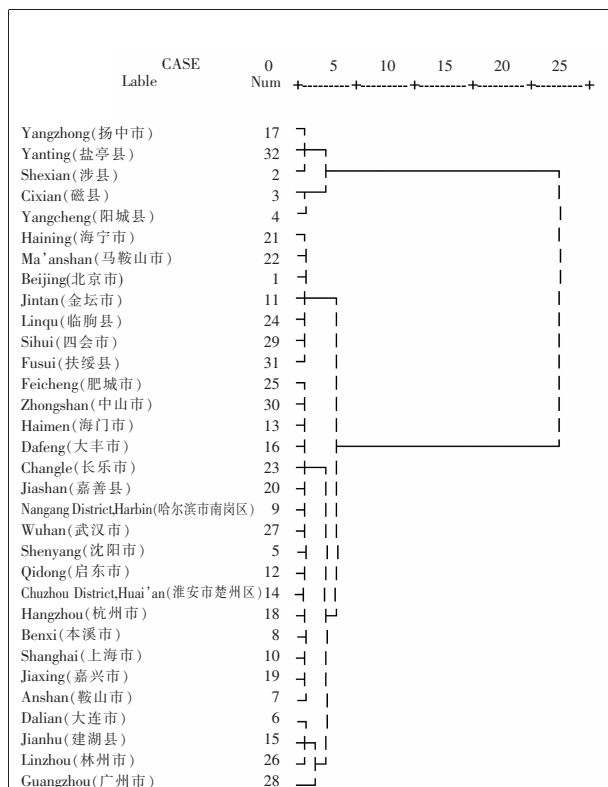


Figure 1 The results of cluster analysis of cancer incidence in China from 2003 to 2007

Table 1 The age-standardized incidence rate of by China population in cancer registration from 2003 to 2007(1/10⁵)

No.	Area	Incidence	Lung	Stomach	Colorectum	Liver	Breast	Esophagus	Pancreas	Bladder	Brain	Lymphoma	Cervix	Nasopharynx
1	Beijing(北京市)	119.37	21.07	7.96	12.92	8.25	28.26	3.65	3.05	3.94	2.77	4.70	3.95	0.67
2	Shexian(涉县)	226.02	15.74	97.67	7.32	16.07	6.75	56.95	0.61	1.43	2.93	1.91	19.20	0.79
3	Cixian(磁县)	262.12	33.44	51.74	8.37	20.50	10.31	109.72	1.80	1.23	4.38	2.06	4.84	0.61
4	Yangcheng(阳城县)	267.03	11.17	83.16	8.04	21.29	11.45	85.31	1.37	1.36	2.27	4.23	57.86	0.30
5	Shenyang(沈阳市)	138.70	31.74	12.29	18.16	10.66	27.99	4.37	3.51	4.16	3.50	1.89	8.59	0.67
6	Dalian(大连市)	165.10	32.23	17.91	16.36	14.95	31.20	3.58	4.76	5.41	6.37	4.72	6.32	0.92
7	Anshan(鞍山市)	145.58	33.26	10.38	18.12	11.40	30.08	4.71	3.16	4.24	4.02	3.19	10.60	1.20
8	Benxi(本溪市)	143.44	32.91	15.05	16.77	16.00	23.37	5.18	3.46	4.18	3.97	2.83	6.00	0.72
9	Nangang District, Harbin(哈尔滨市南岗区)	141.20	34.61	13.11	15.16	15.08	25.34	3.24	4.79	3.25	5.14	2.18	6.79	0.81
10	Shanghai(上海市)	143.24	20.28	14.61	18.49	10.63	30.28	2.87	4.38	3.46	5.23	4.73	4.05	2.00
11	Jintan(金坛市)	118.47	11.20	39.86	8.61	7.41	8.16	22.30	2.37	1.54	1.92	2.03	1.79	0.88
12	Qidong(启东市)	139.70	27.14	17.34	9.16	40.49	12.25	5.11	4.18	2.14	4.49	4.24	4.03	1.17
13	Haimen(海门市)	151.41	29.17	19.42	10.02	31.06	13.03	7.98	4.16	2.94	4.66	4.32	4.05	2.07
14	Chuzhou District, Huai'an(淮安市楚州区)	144.31	15.81	26.63	3.42	16.03	3.23	63.59	2.08	0.87	2.28	1.60	1.38	0.63
15	Jianhu(建湖县)	165.52	16.05	55.37	3.80	18.47	5.96	45.62	2.40	0.75	2.87	1.43	5.37	0.92
16	Dafeng(大丰市)	135.32	21.91	23.90	6.85	20.66	8.39	21.22	3.93	1.83	4.83	0.54	7.58	1.16
17	Yangzhong(扬中市)	234.38	14.61	84.20	13.75	14.94	19.78	67.64	2.74	1.37	1.07	0.90	12.46	1.56
18	Hangzhou(杭州市)	143.83	22.91	15.68	15.08	13.54	21.86	5.20	3.60	3.45	6.19	4.78	7.15	2.68
19	Jiaxing(嘉兴市)	142.86	26.19	13.69	16.93	14.99	20.66	6.54	5.39	2.69	6.11	4.17	3.65	2.25
20	Jiashan(嘉善县)	137.30	26.22	16.90	16.57	15.09	18.08	7.12	5.78	2.74	6.29	3.87	2.24	2.16
21	Haining(海宁市)	109.00	24.06	11.50	10.03	13.14	11.90	5.86	4.84	1.72	3.23	3.36	3.95	2.36
22	Ma'anshan(马鞍山市)	107.01	18.32	21.31	11.51	6.93	14.47	9.47	2.06	2.20	0.99	3.29	3.46	0.84
23	Changle(长乐市)	135.56	14.71	45.82	7.02	19.74	8.65	7.43	1.48	1.50	3.11	1.73	6.91	3.30
24	Linqu(临朐县)	114.93	29.85	28.48	5.76	17.45	7.36	8.11	2.97	1.68	2.58	1.51	1.08	0.61
25	Feicheng(肥城市)	154.31	21.64	27.04	5.99	15.48	11.00	53.50	1.45	1.01	3.71	0.30	1.28	0.43
26	Linzhou(林州市)	171.50	10.43	57.50	6.99	10.23	8.91	63.19	1.02	0.42	3.05	0.80	3.73	0.33
27	Wuhan(武汉市)	140.74	27.70	14.21	14.48	14.09	24.22	5.21	3.07	3.45	5.04	4.67	5.52	2.06
28	Guangzhou(广州市)	191.82	34.13	8.53	21.65	21.44	34.08	4.80	2.67	3.55	5.73	7.33	7.13	11.94
29	Sihui(四会市)	113.25	21.02	5.93	8.56	25.56	9.93	3.33	0.86	1.44	3.02	2.37	3.34	16.84
30	Zhongshan(中山市)	155.77	26.90	6.91	14.65	17.44	17.72	8.72	2.46	3.09	4.41	4.57	6.13	15.44
31	Fusui(扶绥县)	96.43	10.17	9.42	4.20	50.37	4.05	2.06	0.60	0.67	1.68	1.04	1.31	4.33
32	Yanting(盐亭县)	234.25	23.21	74.66	6.99	34.16	8.07	60.85	1.09	1.13	2.93	3.36	6.12	1.16

Table 2 The cancer incidence in China from 2003 to 2007

Category	New cases	Population	Incidence per year (1/10 ⁵)
No.1	35916	11302065	317.78
No.2	112839	38494142	293.13
No.3	409491	152858309	267.89
No.4	121017	52776393	229.30
Total	679263	255430909	265.93

2.3 12 种主要癌症发病率综合聚类分析结果

结果分 5 类:第一类包括嘉兴市、嘉善县,杭州市、武汉市、上海市、北京市、大连市、沈阳市、鞍山市、本溪市、哈尔滨市南岗区,结直肠癌、乳腺癌、胰腺癌、膀胱癌、脑瘤和淋巴瘤发病率均相对高发;第

二类为四会市和中山市首先聚成一类,而后与广州市再聚成一类,三地的鼻咽癌发病率居前三位,且明显高于全国其他地区,消化道肿瘤如胃癌、胰腺癌等发病率较低;第三类包括启东市、海门市,大丰市、临朐县、海宁市、扶绥县,均属于肝癌高发区,扶绥县、启东市和海门市的肝癌发病率均居于前列;第四类包括涉县、盐亭县,磁县、扬中市,淮安市楚州区、建湖县、肥城市、林州市,金坛市、马鞍山市、长乐市,此类的各市、县处于胶东半岛、江浙沿海和华北地区,属于食管癌、胃癌高发区,其发病率在 32 个登记处中均位居前列,胃癌以涉县最高,其次是扬中市、盐亭县,磁县的食管癌发病率远高于其他各地;第 5 类

阳城县单独聚成一类,其宫颈癌的发病率明显高于其他地区,胃癌、食管癌的发病率也较高,其他恶性肿瘤发病率均较低(Table 3、Figure 2)。

Table 3 The comprehensive clustering of 12 common cancers in cancer registration in 2003~2007

Category	New cases	Population	Main cancer
No.1	452409	164724508	Colorectum,breast,pancreas, bladder,brain cancer and lymphoma
No.2	74896	27895264	Nasopharyngeal cancer
No.3	58033	23729410	Liver cancer
No.4	87589	37143255	Esophageal cancer,stomach cancer
No.5	6336	1938472	Cervix ,stomach ,esophageal cancer

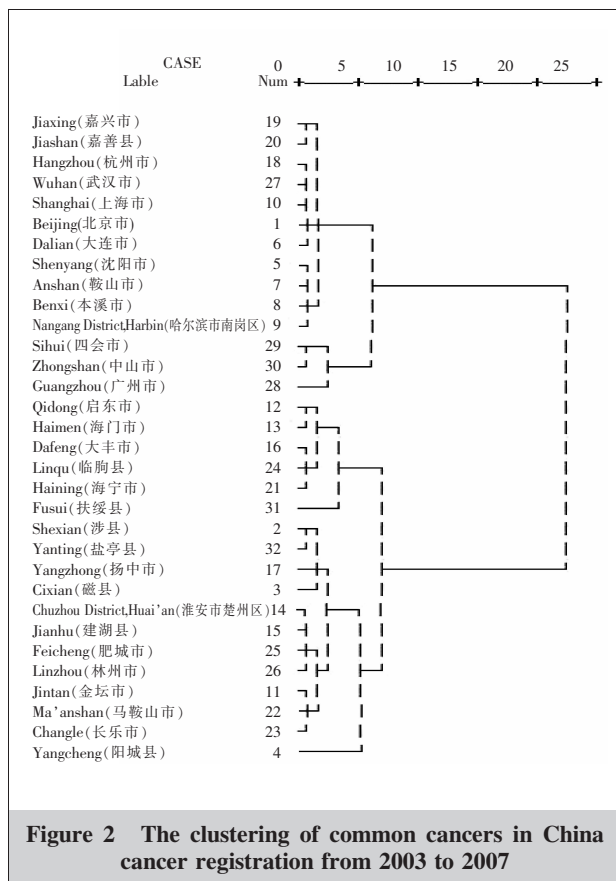


Figure 2 The clustering of common cancers in China cancer registration from 2003 to 2007

3 讨 论

从总发病率水平聚类分析的结果来看,我国癌症可分为高、中、低发地区,高发地区包括磁县、阳城县、涉县、扬中市、盐亭县、大连市、建湖县、林州市和广州市,其发病率高于全国平均水平的 144.25/10 万^[7],

这些登记处基本分布在华北地区的县级市,包括了我国多数的癌症高发现场,如林州、磁县和盐亭是我国食管癌高发现场,阳城县是宫颈癌、胃癌高发地区^[8];肥城市、中山市、海门市、大丰市、长乐市、嘉善县、哈尔滨市南岗区、武汉市、沈阳市、启东市、淮安市楚州区、杭州市、本溪市、上海市、嘉兴市和鞍山市这 16 个登记处的发病率水平与全国平均水平接近,划为中等发病地区;海宁市、马鞍山市、北京市、金坛市、临朐县、四会市和扶绥县的发病率低于全国平均水平,属于低发地区,可实际上,山东省临朐县是我国北方胃癌高发区之一^[9],四会市是我国鼻咽癌高发区之一,与国内其他地区相比较,扶绥县肝癌发病率处于较高水平,是国内高发水平地区之一^[10],这三个登记处因只是某种癌症的发病率很高,而其他癌症的发病率在低水平,故其总发病率较低。

12 种主要癌症发病率水平的层次聚类分析,基本上是某一种或几种高发癌症的登记处聚成一类,其统计分析结果与全国癌症高发种类有关。最明显的是我国鼻咽癌以广东省发病率最高^[11],其发病率水平远远高于全国其余各地,四会市、中山市和广州市聚成一类。我国肝癌的分布特点为东南地区高于西北地区,沿海地区高于内陆地区^[12],启东市、海门市、扶绥县、大丰市地处沿海地区,其肝癌发病率居前列^[13],聚成一类;我国甘肃河西走廊、胶东半岛、江浙沿海的胃癌发病率和死亡率较高,华北地区次之,中南、西南地区较低^[12],位于高发地区的涉县、磁县、扬中市,淮安市楚州区、建湖县、肥城市、林州市、金坛市、马鞍山市聚成一类。

32 个肿瘤登记处把我国主要癌症的部分高发地区纳入监测范围,这些登记处有一定的合理性,但是代表性和全面性不足,其多数位于沿海地区,内陆只有四川盐亭县、广西扶绥县、山西阳城县和河南林州市,而癌症高发地区多在农村和西部地区,危害尤为严重,是当地农民因病致贫及因病返贫的重要原因。对于全面掌握恶性肿瘤在全国的流行特征和趋势以及因地制宜地制定切实有效的防治措施,这些登记处的设置显然是不够的,扩大癌症监测范围势在必行,具有极大的公共卫生学意义,可喜的是,截至 2011 年,全国肿瘤登记处已经增至 195 个,肿瘤登记工作已经在我国 31 个省、自治区、直辖市(除港、澳、台外)开展,登记覆盖人口约 1.85 亿^[7]。

癌症登记可以为各种肿瘤病因学、流行病学研究、肿瘤的自然史和治疗、预防效果的评价提供重要的、甚至是惟一的信息来源,可以认为,癌症登记是我国肿瘤防治研究工作的一项基本建设^[14]。全世界癌症的发病率正以每年3%的速度递增,我国癌症的发病率与死亡率也呈上升趋势^[15],鉴于癌症对人类健康和社会可持续发展构成的极大威胁及肿瘤登记的重要性,应进一步完善全国肿瘤登记制度,增加登记点,扩大监测范围,尤其要重视对于西部和农村癌症高发地区登记工作的资源投入,做好癌症的早诊早治,努力降低癌症发病率和死亡率。

参考文献:

- [1] The Ministry of Health.Chinese cancer prevention and control program (2004–2010) [J]. China Cancer, 2004, 13 (2): 65–68.[卫生部. 中国癌症预防与控制规划纲要(2004–2010) [J]. 中国肿瘤, 2004, 13 (2): 65–68.]
- [2] Xiang YB.An analysis of data utilization of high risk area for cancer in China [J]. China Cancer, 2005, 14 (9):562–569.[项永兵. 中国恶性肿瘤高发发现场点资料的分析和利用 [J]. 中国肿瘤, 2005, 14 (9):562–569.]
- [3] Sun ZQ. Medical statistics[M].2nd edition.Beijing: People's Medical Publishing House, 2005.399–412.[孙振球. 医学统计学 [M]. 第2版. 北京:人民卫生出版社, 2005.399–412.]
- [4] Cassetti T, La Rosa F, Rossi L, et al. Cancer incidence in men: a cluster analysis of spatial patterns[J].BMC Cancer, 2008,8:344.
- [5] Zhao P,Chen WQ,Kong LZ.Cancer incidence and mortality in China 2003–2007[M]. Beijing:Military Medical Science Press,2012.[赵平,陈万青,孔灵芝.中国癌症发病与死亡2003–2007 [M]. 北京:军事医学科学出版社,2012.]
- [6] Yu CH. SPSS and statistical analysis[M]. Beijing: Electronic Industry Press, 2007.468–478.[宇传华. SPSS与统计分析 [M]. 北京:电子工业出版社, 2007. 468–478.]
- [7] Zhang SW,Chen WQ,Kong LZ,et al. An annual report: cancer mortality in 35 cancer registries in China, 2003[J]. China Cancer,2007, 16 (7) : 494–506.[张思维,陈万青,孔灵芝,等. 中国部分市县 2003 年恶性肿瘤发病年报报告 [J]. 中国肿瘤,2007, 16 (7) : 494–506.]
- [8] Zhang SW,Chen WQ,Kong LZ,et al. An analysis of cancer incidence and mortality from 30 cancer registries in China,1998~2002[J].China Cancer,2006, 15(7) :430–440.[张思维,陈万青,孔灵芝,等.中国部分市县 1998~2002 年恶性肿瘤的发病与死亡[J].中国肿瘤,2006, 15(7) :430–440.]
- [9] Ma JL,Liu WD,Zhang L,et al. Analysis of mortality trend of cancer from 1980 to 2002 in Linqu county Shandong province[J]. Chin J Prev Med,2006,40(6):405–408.[马峻岭,刘卫东,张联,等. 临朐县 1980 至 2002 年恶性肿瘤变化趋势分析[J]. 中华预防医学杂志,2006,40 (6) :405–408.]
- [10] You WC. The prevalence of common malignant tumors[J]. China Medical News,2005, 20(22):22.[游伟程.常见恶性肿瘤的流行情况[J]. 中华医学信息导报, 2005, 20(22):22.]
- [11] Wu F,Lin GZ,Zhang JX. An overview of cancer incidence and trend in China[J]. China Cancer, 2012, 21(2):81–85. [吴菲,林国桢,张晋昕.我国恶性肿瘤发病现状及趋势[J]. 中国肿瘤, 2012, 21(2):81–85.]
- [12] Li LM,Ye DQ,Zhan SY,et al. Epidemiology[M]. 6th edition. Beijing:People's Medical Publishing House, 2008. 376. [李立明,叶冬青,詹思延,等.流行病学[M].第6版. 北京:人民卫生出版社, 2008.376.]
- [13] Chen JG,Chen WQ,Zhang SW,et al.Incidence and mortality of liver cancer in China: an analysis on data from the National Registration System between 2003 and 2007 [J]. Chinese Journal of Epidemiology,2012, 33 (6) : 547–553. [陈建国,陈万青,张思维,等.中国 2003–2007 年肝癌发病率与死亡率分析 [J]. 中华流行病学杂志,2012, 33 (6): 547–553.]
- [14] Chen JG.The role of cancer registration in research on cancer[J]. China Cancer, 2003, 12(3):126–129.[陈建国. 癌症登记在肿瘤防治研究中的应用[J]. 中国肿瘤, 2003, 12(3):126–129.]
- [15] Yu PG. My view on treatment of malignant tumor[J]. Journal of Tongji University (Medical Science),2011, 32 (6) :125–128.[余培根.恶性肿瘤之我见[J]. 同济大学学报(医学类),2011, 32(6) :125–128.]