

# 食管癌理论流行病学研究

## I 食管癌年龄死亡率的Logistic曲线及寿命表的拟合检验

河北医学院流行病学教研组 黄朝南

河北省肿瘤研究所流行病学研究室 赵泽贞

随着统计学及计算技术的进展, 流行病学数学模型的应用, 已成为流行病学方法的重要研究内容, 并且形成了以数学方法描述疾病流行过程的理论流行病学<sup>[1]</sup>。在国内, 定量描述疾病流行特征的研究已被注意<sup>[2~4]</sup>, 至于食管癌理论流行病学研究尚乏报导。本文仅讨论食管癌年龄死亡率的Logistic曲线及寿命表的拟合检验, 为食管癌流行病学研究提供分析方法。

河北省于1977年8月在全省范围, 对从1974~1976年三年内常住户口的全部死亡人数及以肿瘤为主的各种死亡原因, 进行回顾性调查。调查的资料经人口统计, 恶性肿瘤死前就诊率、诊断级别的比例及诊断不明的比例等各项分析, 确认资料的完整性和可靠性<sup>[5]</sup>。本文所据的实际数据系以该资料的三年累计人口数和三年累计死亡数, 所求出的三年平均死亡率为基础。

### Logistic曲线的拟合检验

医学资料中两个变量间的关系往往是曲线关系。对于呈现曲线回归关系的资料, 要根据实测资料找到反映变量间关系的曲线回归方程。这种曲线回归分析法即称为曲线拟合。

当检验资料的曲线形状略呈“S”形, 即曲线的上下两端在接近某一限度时, 上升或下降均逐渐缓慢, 并形成两条相互平行的渐近线, 曲线的变动当在此两条渐近线之间的范围内。此种资料的检验值在半对数纸上可呈直线趋势, 便可进行Logistic曲线的拟合检验。

河北省食管癌的年龄死亡率曲线, 也表现出青少年组和老年组上升逐渐缓慢, 而在壮年组则急剧上升的特点。故先将本省食管癌年龄

死亡率的检验数据, 在方格纸上绘图, 各实测值点呈现“S”形的趋势, 籍作图用的曲线板, 目测曲线的上渐近线值与下渐近线值, 几次尝试得合适的d值和k值, d值和k值均系经验数值。尝试的步骤是以各年龄组(T)的死亡率(Y值)减去d值, 得各点与下渐近线的距离, 即Y-d。再以k值减去Y-d的差数, 是各点与上渐近线的距离, 即k-(Y-d)。前两项的比例, 即各点与上渐近线之距除以各点与下渐近线之距的商数为Z。尽量使 $[k-(Y-d)] / (Y-d) = Z$ 的对数与t呈直线关系, 以达到在算术对数纸上直线化。若在尝试过程中k值太小, 则各点所连的线将会凸起。反之, k值太大, 则各点连线将会凹下不能成一直线。此两种情况均应另换d值和k值, 直至近于直线为止。t为年龄变量, 从直线上可看到z=1时, 相应的T值为X, t是各T值与X的距离, 即 $t=T-X$ 。

从目测直线上, 取便于读数的两点:  $P_1(T_1 Z_1)$ ,  $P_2(T_n Z_n)$ 。以t值代替T值, 并将Z值化成对数, 求直线的坡度m。

$$m = \frac{\log Z_n - \log Z_1}{t_n - t_1}$$

求得m值, 即可求方程式中的r和c。

$$r = 2.30259m$$

此式的r值实为当直线化时的纵轴化为自然对数后, 所作回归直线的坡度。按普通对数, 以 $\log_e 10 = 2.30259$ 乘之, 则化为自然对数。

$$C = \text{Antilog} [\log Z_1 - (t_1) m]$$

将d、k、c、r等数值代入logistic曲线的方程式:

$$Y-d = \frac{k}{1 + ce^{rt}} \quad [6]$$



解开方程式即得食管癌各年龄组死亡率的理论值,以各理论值作图遂得“S”形曲线。曲线拟合是否合适,可用相关指数( $R^2$ )以表示拟合度。 $R^2$ 愈接近1,说明实测值与理论值愈接近,曲线拟合得愈好。其计算公式为:

$$R^2 = 1 - \frac{\sum(y - \hat{y})^2}{\sum y^2 - \frac{(\sum y)^2}{n}} \quad [7]$$

式中 $y$ 为食管癌各年龄组的调查死亡率, $\hat{y}$ 为相应年龄组的理论死亡率, $n$ 为年龄组数。

河北省1974~76年男性和女性食管癌各年龄组死亡率的Logistic曲线及其拟合度检验结果如表1。

表1 食管癌年龄死亡率的logistic曲线拟合检验(1/10万)

年龄组 T	男 性		女 性	
	实 测 死亡 率 $y$	理 论 死亡 率 $\hat{Y}$	实 测 死亡 率 $y$	理 论 死亡 率 $\hat{Y}$
10~	0.01	0.01	0.01	0.01
15~	0.05	0.03	0	0.03
20~	0.29	0.10	0.09	0.06
25~	0.47	0.30	0.38	0.16
30~	1.64	0.95	0.77	0.41
35~	4.83	2.95	2.81	1.05
40~	16.53	7.07	8.18	2.64
45~	35.60	27.12	17.01	6.60
50~	78.88	74.87	36.22	16.07
55~	140.77	171.72	59.19	37.03
60~	236.24	293.25	94.85	76.13
65~	316.78	379.21	131.25	130.39
70~	418.51	418.51	181.30	181.30
75~	427.29	432.88	199.26	214.27
80~	438.47	437.69	242.55	230.80

方程式

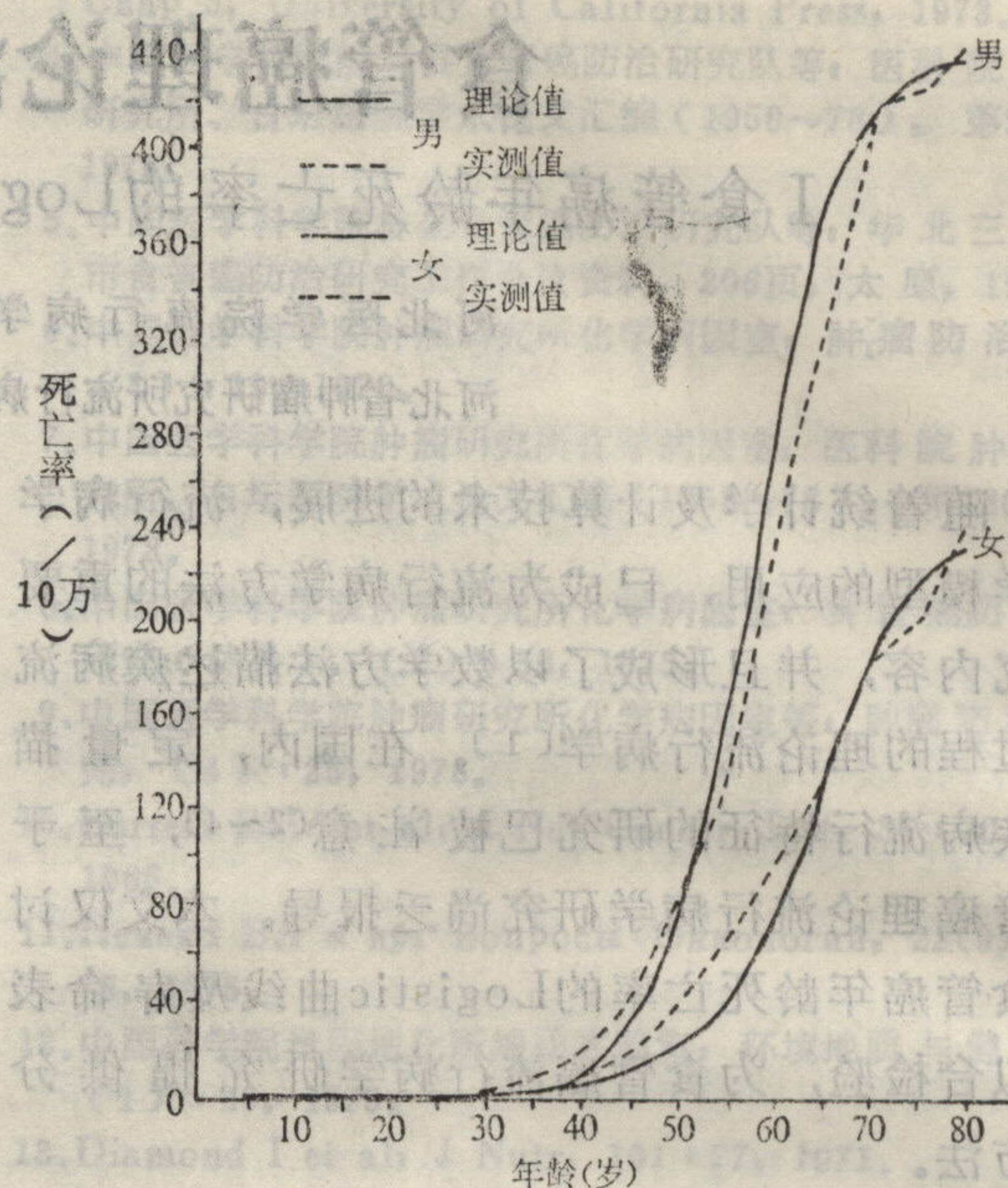
$$\hat{Y}_{男} = \frac{440}{1 + 0.56078e^{-1.1384t}}$$

$$\hat{Y}_{女} = \frac{243}{1 + 1.04054e^{-0.9313t}}$$

$R^2$  0.980731356      0.982386343

将表1的年龄组(T值)与食管癌的实测死亡率( $y$ )和理论死亡率( $\hat{Y}$ )作图:

Logistic曲线的优点在于明确显示食管癌死亡率的年龄分布趋势,便于肿瘤流行病学分析。目前肿瘤流行病学调查资料大多为横断面



附图 河北省1974~1976年食管癌死亡率的Logistic曲线调查,这种资料的比较分析往往需要采用“标化”和“队列”(Cohort)的方法。前者需要标化人口,后者需要积累数十年以上的资料,在实际工作中存在一定困难。本资料Logistic曲线拟合结果表明:河北省食管癌各年龄组死亡率,男性均高于女性。从理论曲线的上升趋势及速度,男女均以45~65岁组,呈直线峻陡上升,尤以男性更急剧。理论曲线的拟合度检验,男女两性的 $R^2$ 均为0.98,说明理论曲线对河北省食管癌各年龄组死亡率的分布状况,有一定代表性,为今后本省食管癌死亡率的年龄分布调查,建立流行病学分析的基础。理论曲线不仅可以比较各横断面调查资料的变化趋势,而且从建立的方程式,还可推算任一组年龄的理论死亡率。一般地,食管癌的死亡率是随年龄而升高,可是由于食管癌的分布特性,使其在某一次横断面调查可能出现抽样误差。表现在某些年龄组的死亡率,不一定都服从上升的趋势,反而出现下降的假象是屡有所见的。例如本资料中女性食管癌死亡率在10岁组为0.01/10万,而15~19岁组的死亡率反降为0,其后的20~24岁组又重升为0.09/10万。这提示实际调查资料,尚未能概括总体。存在抽样误差



的事实，说明样本与总体之间既存在着普遍差别，又有明显的一致趋势。如本文将调查的资料拟合成相应的理论趋势线，用于估计总体的特征。无疑地要比一般频数曲线描述的方法更为合理。

### 寿命表的配合

死亡率的年龄分组资料可用以编制寿命表〔6〕。寿命表中各项指标是综合性，不受人口年

龄构成或出生率的影响。因此，不同人群的寿命表指标，可以直接相比，这对比较不同地区、时间及人群的疾病分布特点的流行病学，提供另一种概括死亡率的方法。

本文按表1的Logistic曲线方程式所求得的食管癌理论死亡率与实际调查的食管癌年龄组死亡率编制简略寿命表及去食管癌寿命表〔7〕(表2、表3)。

寿命表运用各年龄组的死亡率并演算出表

表2 河北省1975年男性食管癌logistic曲线的寿命表检验

年 龄 组	去食管癌平均 期望寿命 $ex^{-i}$		全死因平均期 望寿命 $ex$		因食管癌减少 寿命数		百分比	
	实测值	理论值	实测值	理论值	实测值	理论值	实测值	理论值
10~	62.09	62.15	60.78	60.78	1.31	1.37	2.16	2.25
15~	57.27	57.33	55.96	55.96	1.31	1.37	2.35	2.45
20~	52.49	52.55	51.17	51.17	1.32	1.38	2.58	2.70
25~	47.76	47.82	46.44	46.44	1.32	1.38	2.85	2.97
30~	43.04	43.10	41.71	41.71	1.33	1.39	3.19	3.33
35~	38.35	38.41	37.01	37.01	1.34	1.40	3.62	3.78
40~	33.77	33.84	32.42	32.42	1.35	1.42	4.17	4.38
45~	29.30	29.39	27.96	27.96	1.34	1.43	4.80	5.11
50~	24.98	25.08	23.64	23.64	1.34	1.44	5.66	6.09
55~	20.91	21.02	19.61	19.61	1.30	1.41	6.64	7.19
60~	17.12	17.23	15.88	15.88	1.24	1.35	7.80	8.50
65~	13.84	13.92	12.64	12.64	1.20	1.28	9.47	10.13
70~	10.91	10.96	9.69	9.69	1.22	1.27	12.59	13.11
75~	8.85	8.85	7.43	7.43	1.42	1.42	19.05	19.11
80~	7.39	7.39	5.42	5.42	1.97	1.97	36.35	36.35

中的各项指标，本资料采用简略寿命表，不仅编制方便，而且因为各年龄组距加大，对寿命表各项指标起到修匀作用，更有利于对Logistic曲线的比较分析。从表2及表3的检验结果说明，用寿命表统计分析检验Logistic曲线是理想的。寿命表中各项指标的实测值和理论值都很接近，并且在“因食管癌减少寿命数”两个值的t测验，揭示差异无显著性。

去食管癌寿命表的编制，则是用寿命表法研究食管癌对人群生命影响程度的方法。从表2及表3的末两项纵列的数据，首先揭示食管癌使河北省男女各年龄组的人群寿命缩短1~2岁。除80岁以上的年龄组因食管癌减少的寿命数较明显的升高外，其他各年龄组的减少寿

命数基本相近。但是，因食管癌减少寿命数在全死因平均期望寿命所占的比重，以百分比表示，则表现随年龄而升高的特点，呈现为加速度的指数曲线型，揭示食管癌对人群生命的危害因年龄而加重并显出老年性疾病的特点。可见寿命表又从“寿命”的角度，配合Logistic曲线描述疾病的流行病学特征。

### 结 语

Logistic曲线和寿命表都是医学研究中常用的统计方法。本文讨论其在食管癌理论流行病学研究中的价值。

以河北省1974~76年的食管癌年龄死亡率为基本资料，先在方格纸观察趋势，又在半对



表3 河北省1975年女性食管癌logistic曲线的寿命表检验

年 龄 组	去食管癌平均 期望寿命 $ex^{-i}$		全死因平均期 望寿命 $ex$		因食管癌减少 寿命数*		百分比	
	实测值	理论值	实测值	理论值	实测值	理论值	实测值	理论值
10~	64.14	63.88	62.90	62.90	1.24	0.98	1.97	1.56
15~	59.28	59.03	58.05	58.05	1.23	0.98	2.12	1.69
20~	54.28	54.21	53.22	53.22	1.06	0.99	1.99	1.85
25~	49.53	49.45	48.46	48.46	1.07	0.99	2.00	2.04
30~	44.82	44.74	43.74	43.74	1.09	1.00	2.00	2.29
35~	40.14	40.06	39.05	39.05	1.09	1.01	3.00	2.58
40~	35.54	35.46	34.45	34.45	1.09	1.01	3.00	2.94
45~	31.03	30.96	29.94	29.94	1.09	1.02	4.00	3.42
50~	26.64	26.59	25.55	25.55	1.09	1.04	4.00	4.06
55~	22.44	22.41	21.35	21.35	1.09	1.06	5.00	4.95
60~	18.46	18.45	17.37	17.37	1.09	1.08	6.00	6.21
65~	14.88	14.88	13.77	13.77	1.11	1.11	8.00	8.06
70~	11.66	11.66	10.49	10.49	1.17	1.17	11.00	11.20
75~	9.22	9.23	7.83	7.83	1.39	1.40	18.00	17.88
80~	7.41	7.41	5.48	5.58	1.93	1.93	35.00	35.00

数纸直线化,遂进行Logistic曲线的拟合检验,并获得相应的数学方程式。解开方程式即得食管癌各年龄组死亡率的理论值。理论值曲线拟合度 $R^2$ 都接近于1,说明理论曲线拟合满意。明确显示了河北省食管癌死亡率的年龄分布趋势,并为今后本省的食管癌流行病学横断面调查,建立比较分析的基础。Logistic曲线不仅比一般频数曲线的描述方法更为合理,而且也具有比“标化”或“队列”的方法更为方便的优点。

本文进一步用寿命表法对Logistic曲线加以检验。证明寿命表中各项指标的实测值和理论值都很接近,t测验揭示两者差异不显著,Logistic曲线拟合理想。而且,去食管癌寿命表的编制,又从“寿命”的角度,配合Logistic曲线显示食管癌的老年性特征。这些表明Log-

istic曲线及寿命表拟合检验,为食管癌理论流行病学研究提供分析方法。

### 参 考 文 献

1. WHO: Dynamics of Acute Bacter Dis, 1978, Geneva.
2. 苏德隆: 流行病学调查分析(中), 内部资料, 1979.
3. 黄朝南: 河北省参加全国流行病学学术会议论文选编, 154页, 中华医学会河北分会, 1980.
4. 叶本法: 流行病学杂志, (1): 34, 1981.
5. 河北省卫生局肿瘤防治办公室: 肿瘤防治研究资料, (4): 3, 1980.
6. 郭祖超等编: 医用数理统计方法, 452~458页, 人民卫生出版社, 北京, 1963.
7. 上海第一医学院卫生统计学教研组: 医学统计方法, 98~99页, 220~223页, 上海科学技术出版社, 上海, 1979.
8. England JM: Med Resear, A Statistical and Epidemiological Approach, p 103-106, Churchill Livingstone, Edinburgh London & New York, 1975.