

细胞存活率—辐射剂量曲线拟合初探*

林万敏 王兆霖 马承武 罗祖玉

(复旦大学)

摘要 根据实验数据来拟合存活率——辐射剂量曲线是放射生物学的基础工作之一。这一工作常常在计算机上用最小二乘法完成。但是,由于所用的目标函数不同,从同样实验数据中得到的曲线参数也不同,本文通过数理推理证明,目标函数 $Q = \sum_{k=1}^m \left[\frac{Y_k - S(X_k, B)}{SE_k} \right]^2$ 比其它一些目标函数更适用于存活曲线拟合。本研究用三个存活模型对理论上的推论进行了验证,得到了较好的结果。这一目标函数看来也适用于其它一些分布跨越几个数量级的实验数据的拟合。

关键词: 存活率—辐射剂量曲线; 曲线拟合; 最小二乘法; 目标函数。

前言

制作细胞存活率——辐射剂量曲线(以下简称存活曲线)已成为放射生物学中重要的基础工作。它能定量地提供具有生物物理学意义的参数,是迄今在这一领域中把微剂量学、统计学与细胞学圆满结合起来的例证之一。

在数理统计中,一般用最小二乘法来拟合已知的数学模型。存活曲线拟合所依赖的数学模型,目前较为常用的有三个,即多靶单击模型 $S = 1 - (1 - e^{-\lambda_1 D})^N$ ^[1] 和带初斜率的多靶单击模型 $S = e^{-\lambda_1 D} [1 - (1 - e^{-\lambda_2 D})^N]$ ^[2], 以及一次——二次方模型 $S = e^{-(\alpha D + \beta D^2)}$ ^[3]。上述三式中, S 表示存活率, D 表示辐射剂量。这些模型分别带有二个或三个待定的参数,除了第三个模型能够变换成多项式的表达形式,从而用手算进行回归计算外,其余两个模型很难用人工计算来达到具有一定精度的拟合。但是,应用计算机这一问题就变得很容易了。存活曲线的计算机拟合已为越来越多的人所应用,近几年,国内也出现了这一方面的研究报道^[4]。

然而,从一些作者处理数据的方法来看,在用计算机作最小二乘法的曲线拟合时,回归的指标即目标函数不尽相同。本文依据数理统计的原理,并结合实验,分析、比较了不同的拟合方法,提出了较为合理的目标函数。

问题的提出和目标函数的推导

一般的非线性函数的最小二乘法曲线拟合,采用的目标函数为

$$Q = \sum_{k=1}^m [S_k - Y(X_k, B)]^2 \quad (1)$$

式中, X_k 、 S_k ($K = 1, 2, 3, \dots, m$) 为——对应的自变量和因变量(在存活曲线实验中即是

* 本项研究得到中国科学院科学基金会资助。

辐射剂量和存活率观察值); $Y(X, B)$ 为拟合时所用的数学模型, B 为模型中待求的参数, 可以是一个或几个。目标函数 Q 表示存活率的实验观察值与拟合估算值的残差的平方和。能满足 Q 取得最小值的 B 即为所求之参数。

在用式(1)进行存活曲线拟合时, 有两个不足难以克服, 首先, 存活率观察值变化范围很大。当剂量(X_k)为 0 时, 存活率(S_k)等于 1; 但当剂量增大到一定值时(对一般哺乳动物细胞而言, 大致在几个 Gy 以上), 存活率可以下降到 10^{-3} 以下, 而 $Y(X_k, B)$ 总是一个在数量级上与之相当的值。这样在函数 Q 中, 低剂量实验点的残差平方和比高剂量实验点的残差平方和大得多, 即主要是低剂量实验点的贡献。剂量越高, 其相应实验点的贡献越小。与高剂量点相比, 低剂量点的残差平方和可以高出几个数量级。观察表明, 当实验包含八、九剂量点时, 最后二、三个高剂量点的存活率的值(往往在 10^{-3} 左右或更小)取与不取竟然对拟合结果没有什么影响。这显然是很不合理的。其次, 假如不考虑数据本身的差异, 那么各实验点将对 Q 值作出同样的贡献。这也不符合存活曲线实验的实际情况。事实上, 由于各种因素的影响, 通过细胞培养取得重复的定量结果是相当困难的, 细胞的存活率数据都是由多批、多样品的实验值平均得到。因此, 各实验点的平均值的离中性是不同的。这在统计学上用标准误(SE)表示。在各实验点平均值相差不大的情况下, 某一实验点的标准误越小, 意味着该点数据越集中, 自然其可信度也越高。但在存活曲线实验中, 从各实验点得到的存活率平均值相差很大, 应以标准误和平均值的比, 即 SE_k/S_k , 来判定数据的集中程度。 SE_k/S_k 越小的实验点, 其数据变异程度越小。从拟合的角度看, 应该让变异程度小的点占有更大的比重。

国外已经有人注意到对上述两方面偏差的纠正。Millar^[6]等提出, 为了使各实验点具有同样的权重, 拟合存活曲线应用如下的目标函数:

$$Q = \sum_{k=1}^m \left[\frac{S_k - Y(X_k, B)}{S_k} \right]^2 \quad (2)$$

式(2)实际上解决了上述第一个问题。目标函数由各实验点的绝对残差平方和换成了相对残差平方和, 从而剔除了各实验点数据本身差异带来的不合理成份。

考虑到相对标准误 SE_k/S_k 在结果中应有所体现, 我们在式(2)的基础上进一步加权处理。在一般的最小二乘法曲线拟合中, 权重 P_k 与实验数据的标准误的平方成反比^[8]。在存活曲线的拟合中, 权重应与 SE_k/S_k 的平方成反比。

综上所述, 目标函数可表达为:

$$Q = \sum_{k=1}^m P_k \left[\frac{S_k - Y(X_k, B)}{S_k} \right]^2$$

而

$$P_k = \left(\frac{1}{SE_k/S_k} \right)^2$$

经代入整理, 即

$$Q = \sum_{k=1}^m \left[\frac{S_k - Y(X_k, B)}{SE_k} \right]^2 \quad (3)$$

实验验证与讨论

为了证实上述设想, 我们用实验数进行了验证。实验所用的材料为人肺腺癌细胞 SPC-A-1^[7]。

辐照用 Philips 深部治疗机产生的 220 KV X 线进行。详细操作步骤见前文^[3]。实验数据的最后处理是在 Commodore-64 微机上用 BASIC 语言编写的程序完成的。

表 1 中 A 行列了在空气中经 0-5.0 GyX 线照射的 SPC-A-1 细胞的存活率(由三批实验得到的平均值)。为了观察在不同的拟合方法中高剂量及低剂量区域实验数据的变化对最后结果的影响,在 B 行和 C 行假定高剂量区域的最后两点的存活率分别比 A 行提高一倍和降低一半;在 D 行

表 1 X 线照射后 SPC-A-1 细胞的存活率

Table 1 Survival fraction of SPC-A-1 cells irradiated with x-rays

剂量 (Gy)		0.0	0.5	1.0	1.5	2.0	2.5	3.0	4.0	5.0
存活率 Survival fraction	A	1.0	0.69	0.41	0.24	0.125	0.075	0.033	0.0074	0.0014
	B	1.0	同样本 A The same as sample A						0.0148	0.0028
	C	1.0	同样本 A The same as sample A						0.0037	0.0007
	D	1.0	0.80	0.50	同样本 A The same as the sample A					
	E	1.0	0.60	0.35						

表 2 两种不同的拟合方法对参数的影响

Table 2 Effect of two different fitting methods on the parameters

样本 Samples	参数 Parameters	方 法 Methods	$Q = \sum_{k=1}^m [S_k - Y(X_k, B)]^2$ →min.	$Q = \sum_{k=1}^m \left[\frac{S_k - Y(X_k, B)}{S_k} \right]^2$ →min.
A	D_0		0.8248	0.6866
	N		1.4864	2.3302
	Q		1.5188×10^{-4}	0.1139
B	D_0		0.8295	0.7989
	N		1.4838	1.6002
	Q		1.2828×10^{-4}	8.8870×10^{-5}
C	D_0		0.8315	0.5993
	N		1.4793	3.2677
	Q		1.3526×10^{-4}	0.3618
D	D_0		0.6838	0.6682
	N		2.5036	2.7351
	Q		7.4969×10^{-4}	5.8058×10^{-2}
E	D_0		0.9727	0.7076
	N		1.0082	1.9495
	Q		9.9838×10^{-4}	0.2346

表 3 用目标函数为 $Q = \sum_{k=1}^m \left[\frac{S_k - Y(X_k, B)}{SE_k} \right]^2$ 的最小二乘法处理样本A的结果

Table 3. Results from sample A using the least squares method with the objective

$$\text{function } Q = \sum_{k=1}^m \left[\frac{S_k - Y(X_k, B)}{SE_k} \right]^2$$

剂 量 (Gy)	0.0	0.5	1.0	1.5	2.0	2.5	3.0	4.0	5.0
存活率 (S_k)	1.0	0.69	0.41	0.24	0.125	0.075	0.033	0.0074	0.0014
相对标准误 ($\frac{SE_k}{S_k}$)	0.0	0.226	0.164	0.100	0.020	0.147	0.009	0.234	0.540
存活率计算值 (Y_k)	1.0	0.763	0.449	0.242	0.125	0.064	0.032	0.008	0.002
参数 Parameters	$D_0 = 0.721$			$N = 2.0797$			$Q = 2.5327$		

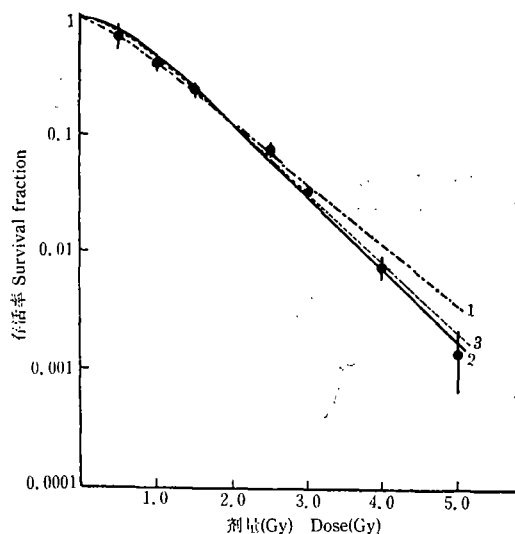


图 1 根据三种不同拟合方法处理样本A得到存活率—剂量曲线, 存活模型是 $S = 1 - (1 - e^{-D/D_0})^N$, 竖线表示 $S_k \pm SE_k$

Fig. 1 Survival-dose curves based on the treatment of sample A using three different fitting methods. The survival model is $S = 1 - (1 - e^{-D/D_0})^N$. The bar is $S_k \pm SE_k$.

$$\textcircled{1} Q = \sum_{k=1}^m \left[S_k - Y(X_k, B) \right]^2$$

$$\textcircled{2} Q = \sum_{k=1}^m \left[\frac{S_k - Y(X_k, B)}{S_k} \right]^2$$

$$\textcircled{3} Q = \sum_{k=1}^m \left[\frac{S_k Y - (X_k, B)}{SE_k} \right]^2$$

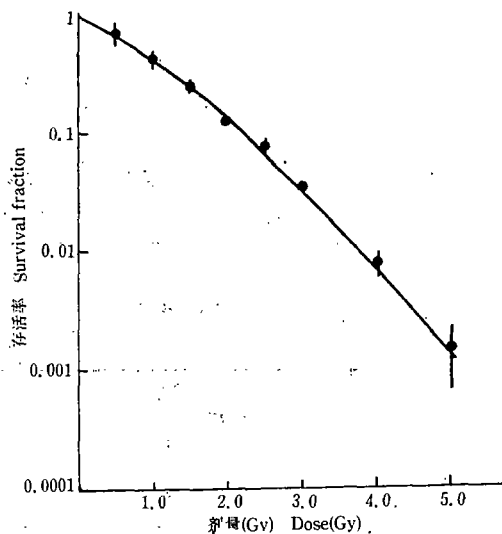


图 2 用目标函数为 $Q = \sum_{k=1}^m \left[\frac{S_k - Y(X_k, B)}{SE_k} \right]^2$

的最小二乘法处理样本A得到存活率—剂量曲线。存活模型为 $S = e^{-D/D_1} [1 - (1 - e^{-D/D_2})^N]$

Fig. 2 Survival-dose curve based on the treatment of sample A using the least squares method with the objective

$$\text{function } Q = \sum_{k=1}^m \left[\frac{S_k - Y(X_k, B)}{SE_k} \right]^2$$

The survival model is $S = e^{-D/D_1} [1 - (1 - e^{-D/D_2})^N]$

和E行则假定在低剂量区域的存活率在A行的基础上分别提高和降低15%左右。以 $S = 1 - (1 - e^{-\alpha D})^N$ 为模型进行处理。

表2是用 $Q = \sum_{k=1}^m (S_k - Y(X_k, B))^2$, 和 $Q = \sum_{k=1}^m \left[\frac{S_k - Y(X_k, B)}{S_k} \right]^2$ 进行拟合的结果。由表中可见,

用式①进行拟合, 高剂量点的存活率变化(B、C), 对参数 D_0 和 N 的影响甚微, 而低剂量区域存活率的变化(D、E)则对 D_0 和 N 有较大的影响。当用式②进行拟合, 上述弊病则有所克服, 结果显得较为合理。图1中的①、②分别表示这种方法对表1中A行数据进行拟合得到的图象。

表3是用 $Q = \sum_{k=1}^m \left[\frac{S_k - Y(X_k, B)}{SE_k} \right]^2$ 对表1中A行数据进行拟合的结果。得到的 N 和 D_0 值和

以式 $Q = \sum_{k=1}^m \left[\frac{(S_k - Y(X_k, B))}{S_k} \right]^2$ 进行拟合的相近。从表中可见, 凡是 SE_k/S_k 较小的点, 其 S_k 的

值与 $Y(X_k, B)$ 的值也比较接近, 说明这些点在拟合中所作的贡献较大, 反之亦然, 从而验证了所用方法的合理性。图1中的③是这一拟合的图象。

为了进一步对 $Q = \sum_{k=1}^m \left[\frac{S_k - Y(S_k, B)}{SE_k} \right]^2$ 进行

检验, 我们还用它配合另二个存活曲线模型对表1中的A行数据进行了拟合。图2、3是这二条曲线的图象。

本工作表明, 在用最小二乘法进行存活曲线拟合时, 采用不同的目标函数, 其结果是不一样的。如果选择的拟合方法不尽合理, 则得到的存活曲线参数未必能反映细胞对射线的反应。实验表明由于方法选择不当而引起的参数变异, 有时可能会超过某些剂量修饰因子的影响。这一点应予足够重视。看来适用于一般非线性曲线拟合的目标函数①对存活曲线拟合并不适用, 而代之以③较为合理。这一方与 Nilson等^[9]所采用的方法基本一致, 但表达更为简单。当每一剂量的存活率的相对标准误相等时, 式(2)和式(3)等价。本方法对那些实验数据分布与存活曲线实验相似的工作, 如细胞的药物毒性试验等, 也有一定的参考价值。在程序语言的编写上, 只要对一般非线性

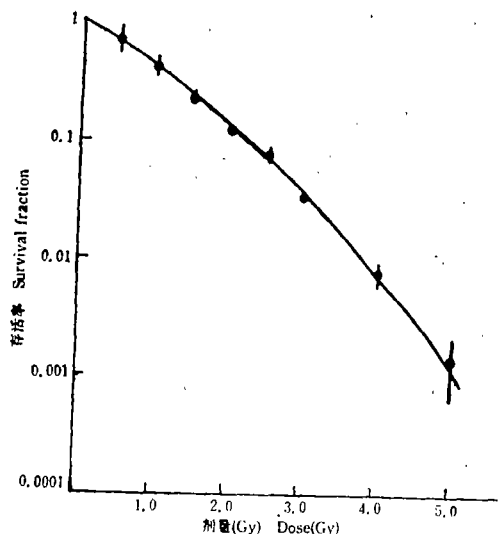


图3 用目标函数为 $Q = \sum_{k=1}^m \left[\frac{S_k - Y(X_k, B)}{SE_k} \right]^2$

的最小二乘法处理样本A得到的存活率—剂量曲线。存活模型为 $S = e^{-(\alpha D + \beta D^2)}$

Fig. 3 Survival-dose curve based on the treatment of sample A using the least squares method with the objective

$$\text{function } Q = \sum_{k=1}^m \left[\frac{S_k - Y(X_k, B)}{SE_k} \right]^2$$

The survival model is $S = e^{-(\alpha D + \beta D^2)}$

性最小二乘法的语言稍加修改, 就可适用本方法。而最小二乘法的计算机语言程序, 在一些程序汇编上可以找到。这对学过计算机语言的基础知识的放射生物学工作者并不太困难。

参 考 文 献

- [1] M. M. Elkind and G. F. Whitmore, The radiobiology of cultured mammalian cells, (Chapter 2) Gordon and Breach, New York, 1967.
- [2] M. A. Bender and P. C. Gooch, Int. J. Radiat. Biol., 5, 133 (1962).
- [3] K. H. Chadwick and H. P. Leenhouts, Phys. Med. Biol., 18, 78 (1973)
- [4] 李冬华等, 中华放射医学与防护杂志, 第2卷, 第5页, 1982年。
- [5] B. C. Millar, E. M. Fielden and J. L. Millar, Int. J. Radiat. Biol., 33, 599 (1978).
- [6] 北京农业机械化学学院主编: 应用数学, 农业出版社, 1979年。
- [7] 吴善芳等: 中国科学(B辑)第10期, 第953页, 1982年
- [8] Lin Wanmin, Ma Chengwu, Wang Zhaolin and Luo Zuyu Radiosensitization Newsletter, 4, 7 (1985).
- [9] S. Nilsson, J. Carlsson and B. Larsson, Int. J. Radiat. Biol., 37, 267 (1980).

(收到日期: 1985年10月7日)

PRELIMINARY STUDY ON THE COMPUTER FITTING OF SURVIVAL-RADIATION DOSE CURVES

Lin Wanmin, Wang Zaolin, Ma Chengwu, Luo Zuyu

(Department of Biology, Fudan University)

ABSTRACT It is essential in radiobiology to fit a survival-radiation dose curve from experimental data, which was usually done on a computer by the least square method. However, using different objective functions, the curve parameters from the same data may be different. The statistics inference in this paper shows that the objective function $Q = \sum_{k=1}^m \left[\frac{S_k - Y(X_k, B)}{SE_k} \right]^2$ is more suitable for survival curve fitting than those used by previous authors.

Three Survival models are used to verify the theoretical inference and a good fitting is achieved. Moreover, this objective function will be applicable to other kinds of fitting work for treating experimental data ranging over several orders.

KEY WORDS, Survival-radiation dose curve; Curve fitting; Least squares method; Objective function.