Survival distribution model and multistage model

参考文献: 柳川(2003): データサイエンスシリーズ, 環境と健康データ, 共立出版

It is relatively easy to obtain dose – response relationship in the case of acute toxicity evaluation. As we learned in the last week, it is relatively difficult to obtain dose – response relationship in the case of carcinogenic toxicity evaluation.

[Exercise 1] Why is it difficult?

In the last lecture, all the experimental results are evaluated as either positively affected or negatively affected. However, in most cases of exposure of chemical compounds, it will take long time before adverse effects can be observed. An example is shown below. Three doses of certain chemical compound were fed to rats and following results were obtained from 13 weeks observation. In the table, “X” means death and “O” means “Observation stopped due to certain reason”. Stopping observation is a usual matter in the experiments. For example, rats can die from several causes other than the chemical compound under examination.

Table 1. The survival of rats in the exposure to certain chemical in 13 weeks.

No.	Dose	Weeks of observation										
	(mg/kg)	3	4	5	6	7	8	9	10	11	12	13
1	0	-	-	-	-	-	X					
2	0	-	-	-	-	-	-	-	-	-	-	O
3	0	-	-	-	-	-	-	-	-	-	-	O
4	0	-	-	-	-	-	-	-	-	-	-	O
5	0	-	-	-	-	-	-	-	-	-	-	O
6	0.1	-	-	-	X							
7	0.1	-	-	-	-	-	X					
8	0.1	-	-	-	-	-	-	-	-	-	-	O
9	0.1	-	-	O								
10	0.1	-	-	-	-	-	-	O				
11	0.5	-	X									
12	0.5	-	-	-	-	X						
13	0.5	-	-	-	-	X						
14	0.5	-	-	-	-	-	-	-	-	-	-	O
15	0.5	-	-	O								

From the table, we can estimate

1. survival ratio
2. difference in survival ratio between zero doses cases and other cases
3. dose – response relationship

In this type of analysis, it is not appropriate to omit the data of stopping observation, because the

data of stopping observation imply that the rats survives until then under the exposure to the chemical compound. If we ignore these results of stopping observation, we select data with bias.
By plotting the survival ratio against survival time, the following figure will be obtained.

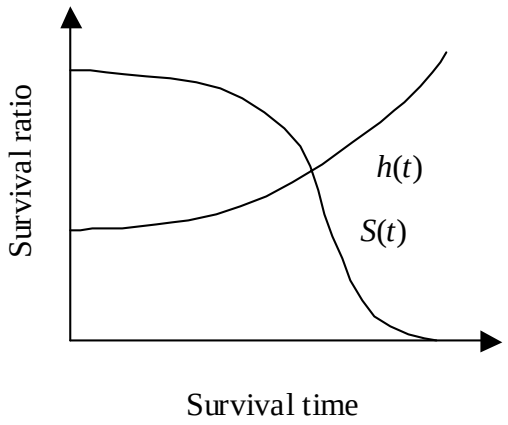

Figure 1. Hazard function and survival function

Here we define a hazard function $h(t)$ from the survival function $S(t)$. The hazard function means death rate at certain time.

$$h(t) = \frac{1}{S(t)}\left(-\frac{dS(t)}{dt}\right)$$

From this definition, we will obtain following equations.

1) $h(t) = -\dfrac{d\log S(t)}{dt}$

2) $S(t) = \exp\left(-\int_0^t h(x)dx\right)$

If the hazard function is a constant $h(t) = c$, then we obtain $S(t) = \exp(-ct)$. This gives a curve as we obtained in the first order reaction. If the hazard function is given as $h(t) = ct^{p-1}$, then we obtain $S(t) = \exp\left(\dfrac{-ct^p}{p}\right)$. This curve is called Weibull survival function and widely used in reliability engineering field.

Table 2. The correspondence of functions in environmental risk assessment and those in reliability engineering.

Environmental Risk Assessment	Reliability Engineering
$S(t)$: Survival function	$R(t)$: Reliability function
$h(t)$: Hazard function	$\lambda(t)$: Failure rate

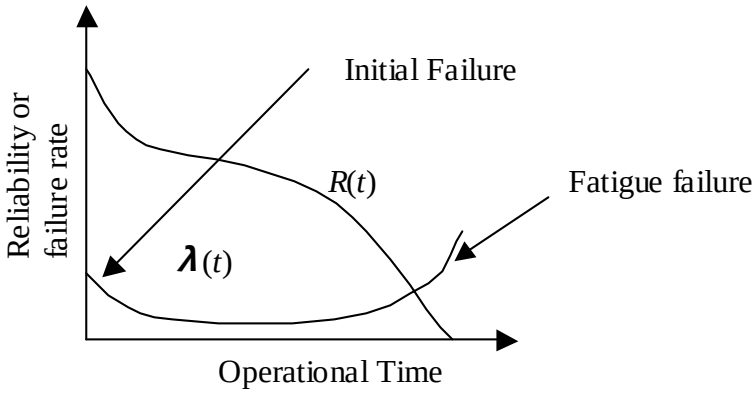

Figure 2. Bath-tub curve of failure rate.

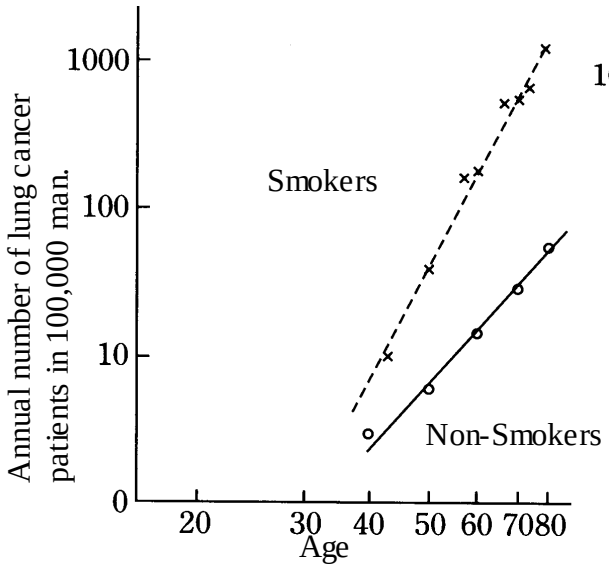

Figure 3. Death ratio of smokers and non smokers.

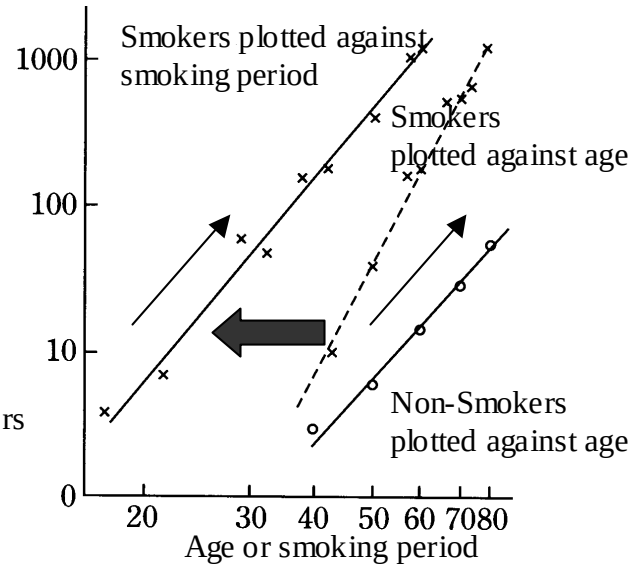

Figure 4. The plots of smokers were shifted from age to smoking experiences.

If we make log – log plots between annual number of long cancer patients in 100,000 men (male) and their age, straight line can be observed as shown in figure 3. This is quite natural because most of the cancer data can be explained by the $h(t) = ct^{p-1}$ model. If we take logarithm of this equation, we obtain

$$\log h(t) = \log c + (p-1)\log t$$

which appears as a straight line in figure 3. The figure 4 shows that the plots can be shifted, if hazard function is plotted not against age but against smoking period. It is of surprising that in figure 4, two straight lines are parallel. It means that the same mechanism is involved in the cancer between smokers and non-smokers.

Many researches were conducted to clarify the mechanism of cancer. An earlier research in 1940s already implied that becoming cancer requires two stages: initiation and promotion. If we feed rats with substance A and after that substance B, then the rats become cancer. If we feed rats with substance A and after long interval we feed them with substance B, the rats become cancer in the

same way. If we feed rats with substance B first and substance A later, the rats don't become cancer. This mean that A is the initiator and B is the promoter.

Due to these consideration on initiators and promoters, multistage model is widely accepted by the researchers.

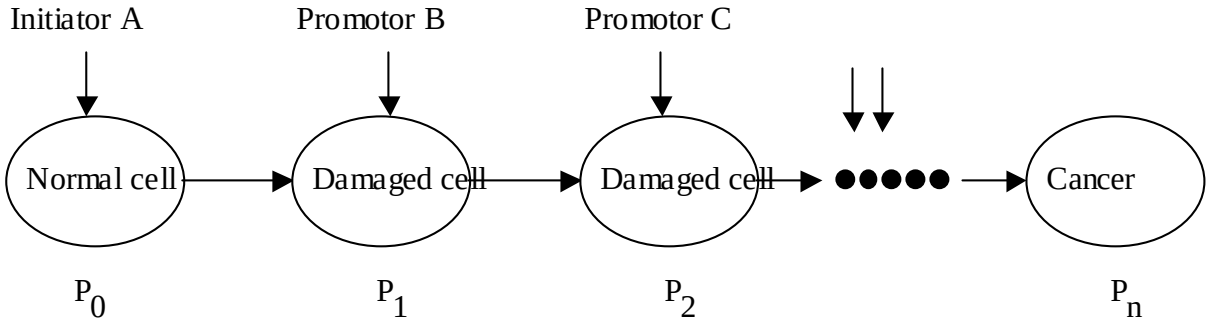

Figure 5. The concept of multi stage model

In the initial stage (in the case of completely healthy person), $P_0 = 1$, $P_1 = P_2 = \bullet\bullet\bullet = P_n = 0$ where P is the probability of the cells in the specific stage. Assuming that substance A converts a healthy cell to a damaged one with the probability of λ_1 during unit time length. In this case, the following equation can be drawn.

$$\frac{d}{dt}P_0(t) = -\lambda_1 P_0(t)$$

This equation can easily be solved and gives

$$P_0(t) = \exp(-\lambda_1 t)$$

In the low level exposure as is observed in environment, this equation can be expanded and by taking the first term of the expansion, we obtain

$$P_0(t) = 1 - \lambda_1 t$$

[Exercise 2] Show that

$$P_1(t) = \lambda_1 t\,,\quad P_2(t) = \frac{1}{2}\lambda_1\lambda_2 t^2\,,\quad P_n(t) = \frac{1}{n!}\lambda_1\lambda_2 \bullet\bullet \lambda_n t^n$$

Assuming that we will become certain cancer with the probability of $P_n(t) = \frac{1}{n!}\lambda_1\lambda_2 \bullet\bullet \lambda_n t^n$, the hazard function is calculated to be

$$h(t) = \frac{1}{1-P_n(t)}\left(\frac{d}{dt}P_n(t)\right) \approx \frac{1}{(n-1)!}\lambda_1\lambda_2 \bullet\bullet \lambda_n t^{n-1}$$

By taking logarithm of this equation, $\log h(t) = const + (n-1)\log t$

This means the slope in figure 4 tells the number of stages in the mechanism of cancer.