

# 医薬安全性研究会

*Bulletin of Japanese Society for Biopharmaceutical Statistics.*

会報 NO. 38

Dec. 1993

## 目次

〈前 付〉

\* 医薬安全性研究会スケジュール

---

Statistical Procedures for Carcinogenicity Testing Ajit K. Thakur, Ph.D. (Hazleton Washington) .....	1
癌原性試験の統計解析の手順 浜田知久馬 (東京大学) .....	23
発癌性の指標 $TD_{50}$ (Tumor Dose <sub>50</sub> ) の推定について 長谷川良平 (名古屋市立大学)・浜田知久馬 (東京大学) .....	25
3 B I S O の質疑応答より 佐藤俊哉 (統計数理研究所) .....	30
薬物の複合投与効果の考え方—マウス小核試験における理論と実際— 鈴木孝昌 (国立衛生試験所) .....	33
毒性試験における多重比較法—L. Hothorn の研究の解説と検討— 松田眞一 (南山大学)・滝沢毅 (日本ロシュ) .....	42

## 1994年 医薬安全性研究会 これからのスケジュール

- ☆1994年1月29日(土) ..... 第57回定例会(総評会館)
- ☆1994年4月16日(土) ..... 第58回定例会(総評会館)
- ☆1994年6月3日(金)～6月4日(土) ..... データ解析講習会(総評会館)
- ☆1994年7月2日(土) ..... 第59回定例会(総評会館)
- ☆1994年10月22日(土) ..... 第60回定例会(総評会館)

---

★6月のデータの解析講習会の日程は予定です。確定次第、改めてお知らせ致します。

★10月の第60回定例会で、当研究会は15周年を迎えます。記念パーティを只今思案中です。計画が決まりましたらお知らせします。

# BIostatISTICS IN CANCER RISK ASSESSMENT

Edited by D. G. Hoel (NIEHS) and T. Yanagawa (Kyushu Univ.)

A4 Size 176pp. ¥ 4,000 (¥ 310YEN)

---

This monograph is a set of selected papers from the two proceedings of the conferences that we organized in 1984 and 1989 in Japan. These proceedings are:

Environmental Risk Assessment and Statistical Methods,

*Environmental Health Perspectives*, Vol 63, 1985

The Third Japan-US Conference on Biostatistics in the Study of Human Cancer

*Environmental Health Perspectives*, Vol 87, 1989

The papers selected in this monograph cover important and essential topics relevant to a range of human health studies. After the publication of these proceedings, we have been asked by many mathematical scientists to make them more accessible to the statistical community so that they may be used for the purpose of training and education of Biostatistics.

<Part of Contents>

"Statistical Studies in Genetic Toxicology : A Perspective from the U.S. National Toxicology Program" by B. H. Margolin

"Multistage Models of Carcinogenesis" by P. Armitage

"Statistical Analysis of Disease Onset and Lifetime Data from Tumorigenicity Experiments" by S. W. Lagakos and L. M. Ryan

"Use of Historical Controls in Animal Experiments" by T. Yanagawa and D. G. Hoel

"Confidence Intervals for Effect Parameters Common in Cancer Epidemiology" by T. Sato

"Statistical Analysis of  $K_2 \times 2$  Tables : A Comparative Study of Estimators/Test Statistics for Association and Homogeneity" by T. W. O'Gorman, R. F. Woolson, M. P. Jones, and Jon H. Lemke

"Relative Risk Regression Analysis of Epidemiologic Data" by R. L. Prentice

"Empirical Bayes Methods for Smoothing Data and for Simultaneous Estimation of Many Parameters" by T. Yanagimoto and N. Kashiwagi

"A Nonparametric Method for Estimating Interaction Effect of Age and Period on Mortality" by M. Ohtaki, D. Kim, and M. Munaka

"An Index for Cancer Clustering" by T. Tango

.....

---

Orders should be directed to :Scientist Inc.

Yamazaki - bldg. 3 - 2

Kanda Surugadai, Chiyodaku,

Tokyo, 101 Japan

# INTERNATIONAL BIOSTATISTICS CONFERENCE IN THE STUDY OF TOXICOLOGY

B5size 135pp. ¥ 3000.

---

## CONTENTS

BIOSUTATISTICAL ISSUES IN THE DESIGN AND ANALYSIS OF ANIMAL CARCINOGENICITY EXPERIMENTS .....	Christopher J. Portier
CARCINOGENICITY TESTING : STATISTICAL ISSUES .....	Yasuo Ohashi
THREE STATE MODELS FOR RODENT CARCINOGENICITY EXPERIMENTS	
	Louise Lyan and Jane Lindsey
A NOTE ON THE USE OF HISTORICAL CONTROLS ..IsaoYoshimura and Kazuhiko Matsumoto	
DATA-BASED SCORES IN EXACT TREND TESTS FOR FETAL SURVIVAL .....	
	Keith A. Soper and Robert L. Clark
STATISTICAL METHODS FOR THE BETA-BINOMIAL MODEL IN TERATOLOGY	
	Eiji Yamamoto and Takemi Yanagimoto
MULTIPLE COMPARISONS IN LONG-TERM TOXICOLOGICAL STUDIES .....	Ludwig Hothorn
MULTIPLE COMPARISON AMONG GROUPS OF GROWTH CURVES.....	Toshinari Kamakura
STATISTICAL EVALUATION OF MUTAGENICITY TEST DATA : RECOMMENDATIONS OF THE U. K. ENVIRONMENTAL MUTAGEN SOCIETY.(UKEMS).....	David J. Kirkland
STATISTICAL ANALYSIS OF DATA IN MUTAGENICITY ASSAYS : IN THE CASE OF RODENT MICRONUCLEUS ASSAY.....	
	M. Hayashi, S.Hashimoto, Y. Sakamoto, C. Hamada, T. Sofuni and I. Yoshimura
BIOSTATISTICAL ISSUES IN THE DESIGN AND ANALYSIS OF MULTIPLE OR REPEATED GENOTOXICITY ASSAYS .....	Lutz Edler
NONPARAMETRIC REGRESSION ANALYSIS OF THE AMES MUTAGENICITY ASSAY	
	John B. Cologne and Norman E. Breslow
DATA MANAGEMENT FOR TOXICOLOGICAL STUDY .....	Ikuro Horii
STATISTICAL APPROACHES FOR STUDY OF THE INDUCTION KINETICS FOR SPECIFIC-LOCUS MUTATION IN <i>Neurospora crassa</i> .....	Frederick J. de Serres
POTENCY MEASURES IN MUTAGENICITY RESEARCH .....	Barry H.Margolin
STATISTICAL ISSUES ON THE NO-OBSERVED-ADVERSE-EFFECT LEVEL IN CATEGORICAL RESPONSE.....	Takashi Yanagawa, Yasuhiro Kikuchi and Kenneth G.Brown
PHYSIOLOGICALLY-BASED PHARMACOKINETICS IN CANCER RISK ASSESSMENT .....	
	Melvin E. Andersen
.....	

**こうすれば内容豊かな質問になる！**

## 医薬安全性研究会 質問募集

医薬安全性研究会では現場での疑問をもとに、より興味深い議論を行いたいと考えています。しかし、質問の仕方がうまくないために答えが面白くなくなっている場合があります。そこで次のように質問を募集します。

### 〔要 項〕

#### 〈答えにくい質問が多い〉

- 1) 抽象的な質問……………「Q：分散分析の位置づけは？」  
「A：野球におけるキャッチボールのようなものだ。」
- 2) 状況説明が一般的……………「Q：併用効果を調べるために、1.対照群 2. Aを投与 3. Bを投与 4. AとBを投与という4群の実験をしたときの検定方法は？」  
「A：単純な2元配置実験なので2元配置実験のデータ解析をするとよい。」

#### 〈質問は具体的に〉

- 1) 質問の状況をできるだけ明らかに
  - 群の大きさ、群の数、動物種／● 測定変数は何か(血圧、時間、白血球数……)／
  - 測定値の例／● 測定条件／● 実験の目的 etc.
- 2) 具体的な説明が必要な理由
  - 質問者が気がついてないことで大切なことがある。  
「イヌの実験で実験者は一元配置分散分析だと考えていたのが二元配置型の実験であった。」
  - 前提とする分布等を知ることにより、より適切な解答となる。

#### 〈公開の原則〉

- 質問は医薬安全性研究会で公表し、討論することを原則とする。

#### 〈公表してほしい部分への配慮〉

- 公表してほしい部分は、質問の際に付記してあれば尊重する。  
質問者、対象動物種、疾患名、具体的な名前(測定変数、値の大きさ)

#### 〔質問送付方法〕

- \* 質問は随時受付けています。
- \* 裏面の質問用紙をコピーして御記入の上、  
右記宛お送り下さい。

#### 医薬安全性研究会事務局

〒101 東京都千代田区神田駿河台3-2 山崎ビル  
TEL 03<253>8992/FAX 03<255>6847  
/振替 東京8-71335

株式会社 サイエコティスト社 内

# 質問用紙

No. \_\_\_\_\_

御氏名		会員No	非会員	年 月 日
〈具体的状況〉				
〈質問内容〉				

(コピーしてお使い下さい)

# Statistical Procedures for Carcinogenicity Testing

Ajit K. Thakur, Ph.D. (Hazleton Washington)

## PURPOSE

The purpose of our discussion is to present the statistical methodology currently used for evaluating incidence tables associated with lifetime carcinogenicity studies. The methods to be described are universally applied by statisticians practicing in this field. There will be examples demonstrating the usefulness and need for these methods under various experimental conditions.

## INTRODUCTION

In a typical carcinogenicity study one employs 50-100 rodents (generally various strains of rats and mice) per group in 4-5 groups. Because of regulatory requirements, in a typical study, one must demonstrate that the material under investigation has been applied at a high enough level to induce some level of toxicity ("MTD") without inducing any carcinogenicity. At the same time, there should be a level (generally higher than the therapeutic or effective level) at which there should not be any adverse toxic effect ("NOAEL"). These levels often are not determined by pharmacological, physiological, and pharmacokinetic investigations. Instead, toxicologists generally try to rely on results from acute and subacute studies. As a consequence, one generally encounters competing toxicity producing differential mortality during such a study. Also, often, particularly with certain agricultural chemicals, latency of spontaneous tumors in rodents may shorten. Because of such happenings, simple tests on proportions, such as  $\chi^2$  and Fisher-Irwin exact tests on contingency tables, may not produce optimal evaluation of the incidence data. In many cases however, statisticians still use some of these tests as methods of preliminary evaluation. These are unadjusted methods without regard for the mortality patterns in a study. Failure to take into account mortality patterns in a study sometimes causes serious flaws in interpretation of the results. We will discuss that with examples later. Hoel and Walburg (1972) addressed the survival issue in their classic work. Since then, a lot of works have been done in the field some of which we will encounter in this discussion. The mortality adjusted methods are simply termed "adjusted methods". Here, instead of a single 2 x k contingency table for the entire life span of the animals, one has a series of 2 x k contingency tables over the life span. By appropriate method of pooling evidence from these tables one provides overall statistical evidence of carcinogenicity. We will examine some of these methods in this presentation.

Based on Hoel and Walburg's original work, Peto et al (1981) examined these issues in detail in the IARC (International Agency for Research on Cancer). Many of the world regulatory agencies now follow some form

## SUMMARY

The purpose of our discussion is to present the statistical methodology currently used for evaluating incidence tables associated with lifetime carcinogenicity studies. The methods to be described are universally applied by statisticians practicing in this field. There will be examples demonstrating the usefulness and need for these methods under various experimental conditions.

In a typical carcinogenicity study one employs 50-100 rodents (generally various strains of rats and mice) per group in 4-5 groups. Because of regulatory requirements, in a typical study, one must demonstrate that the material under investigation has been applied at a high enough level to induce some level of toxicity (Maximum Tolerated Dose or "MTD") without inducing any carcinogenicity. At the same time, there should be a level (generally higher than the therapeutic or effective level) at which there should not be any adverse toxic effect (No Adverse Effect Level or "NOAEL"). These levels often are not determined by pharmacological, physiological, and pharmacokinetic investigations. Instead, toxicologists generally try to rely on results from acute and subacute studies. As a consequence, one generally encounters competing toxicity producing differential mortality during such a study. Also, often, particularly with certain agricultural chemicals, latency of spontaneous tumors in rodents may shorten. Because of such happenings, simple tests on proportions, cannot produce optimal evaluation of the incidence data. In many cases however, statisticians still use some of these tests as methods of preliminary evaluation. These are unadjusted methods without regard for the mortality patterns in a study. Failure to take into account mortality patterns in a study sometimes causes serious flaws in interpretation of the results.

Through several examples we will demonstrate the usefulness of the methods to be discussed in this presentation. As a morale to our story, hopefully we will remember that statistical evidence constitutes only one single piece of chronic oncogenicity testing. It is an important piece but not necessarily the finale of the saga. These tests are not designed for statistical hypothesis testing. They are designed to meet regulatory requirements and toxicological curiosity. The toxicologists, the pathologists, and the regulators will have the final decision regarding the toxicity or carcinogenicity of a chemical. Statisticians' job is to help them in that process.



of the methodology discussed in the IARC document. Based on the IARC discussion, all tumors are initially separated into two categories (Slide 1):

1. Observable or Palpable Tissue Masses: Generally skin and mammary lesions
2. Occult or Nonobservable or Internal Organ Lesions

Based on the above categorizations, different statistical methods have been proposed for their analyses. In the case of Category 1, one has information about the latency or onset time (first observation or palpation time). In the case of Category 2, one generally does not have that information available.

There are many other topics of both statistical as well as biological interests in carcinogenicity testing. We will address them in our discussion. We should appreciate the fact that cancer formation is a very complex and often unexplored mechanism. We must not expect statistical methods to answer all the issues involved in carcinogenicity. The best we can do at this point in time is to address the issues involving the alternative hypotheses of treatment effect and empirical examination of any plausible mechanisms.

#### METHOD

Let us briefly discuss the statistical methods without going into their arithmetic details which are provided in Gart et al (1986) and Thakur et al (1991).

(1) Unadjusted Tests on Proportions: These tests are based on the generalized logistic function (Cox, 1970). Specifically, one can use Cochran-Armitage test (or its parallel, Mantel-Haenszel version) for monotonic trend and Fisher-Irwin exact probability test on 2 x k contingency tables (Thakur et al, 1985) as heterogeneity test. The numbers at risk are generally the numbers of animals histopathologically examined for specific tissues. The methods are sketched out in Slides 2-5:

2 x k Contingency Table with Fixed Marginals

<u>Dose</u>	<u>Number with Tumor</u>	<u>Number without</u>	<u>Number at Risk</u>
L <sub>1</sub>	r <sub>11</sub>	c <sub>1</sub> - r <sub>11</sub>	c <sub>1</sub>
L <sub>2</sub>	r <sub>21</sub>	c <sub>2</sub> - r <sub>21</sub>	c <sub>2</sub>
.	.	.	.
.	.	.	.
L <sub>k</sub>	r <sub>k1</sub>	c <sub>k</sub> - r <sub>k1</sub>	c <sub>k</sub>
Total	$r_1 = \sum r_{i1}$	$r_2 = \sum (c_i - r_{i1})$	$c = \sum c_i$

Homogeneity Statistic: Pearson's  $\chi_h^2$ :

$$\chi_h^2 = ( \sum c_i \bar{p}_i^2 - c\bar{p}^2 ) / [\bar{p}(1-\bar{p})] \sim \chi^2 (k-1)$$

where

$$p_i \equiv r_{i1}/c_i \text{ and } \bar{p} = r_1/c$$

Trend Statistic: Cochran-Armitage  $\chi_t^2$ :

$$\chi_t^2 = c(c\sum r_{11}L_1 - r_1 \sum c_1L_1)^2 / (r_1r_2[c \sum c_1L_1^2 - (\sum c_1L_1)^2])$$

~  $\chi^2(1)$  (Two-tailed Test)

$$Z = \sqrt{\chi^2(1)} - z(0,1) \quad \text{(One-tailed Test)}$$

Departure from Linearity Statistic:  $\chi_d^2$ :

$$\chi_d^2 = \chi_h^2 - \chi_t^2 - \chi_d^2(k-2)$$

The departure statistic is not sensitive to departures from linearity as will be shown later.

Fisher-Irwin Exact Test: This is a hypergeometric exact test for 2 x 2 tables which is widely used as a multiple comparison test and is derived from the generalized logistic function. Under fixed marginals, the probability associated with the observed table is given by:

$$P_o = (r_1!r_2!c_1!c_2!) / [c!r_{11}!r_{21}!(c_1-r_{11})!(c_2-r_{21})!]$$

The tail is initially identified where the observed table appears. The cumulative one-tailed probability is the sum of all probabilities  $\leq$  the probability associated with the observed table in the identified tail. The two-tailed probability is the unconditional sum of all probabilities  $\leq$  the probabilities associated with the observed table. Network or recursion algorithms (Thakur et al, 1985) can generate these tail probabilities even for very large sample sizes swiftly eliminating the need for asymptotic  $\chi^2$  approximation.

Some gross adjustments on the numbers at risk can be made by eliminating early deaths or sacrifices by justifying that those animals were not at risk to have developed the particular tumor in question. Unless there is dramatic change in tumor prevalence distribution over time, the gross adjusted method provides fairly reliable evidence of treatment effect, at least for nonpalpable tissue masses.

(2) Survival Adjusted Tests on Proportions: As mentioned earlier, in the case of survival adjusted analyses, instead of having a single 2 x k table as in Slide 2, one has a series of such 2 x k tables across the entire lifetime of the study. The numbers at risk for such analyses will depend on the type of tumor one is dealing with. They are shown in Slide 6:

(a) Palpable or Lethal Tumors

The number at risk at time t = number of animals surviving at the end of time t-1.

(b) Incidental Tumors

The number at risk at time t = number of animals that either died or sacrificed whose particular tissue was examined histopathologically.

The methods of analyzing the incidences, once the appropriate numbers at risk are assigned for these tumors are rather similar, either binary regression based or by pooling evidence from individual tables (Gart et al, 1986). We will examine some of them here.

(a) Observable or Palpable Tumors: Life table techniques are used for these data. Specifically, one should use Kaplan-Meier product limit estimates from censored data graphically, Cox-Tarone binary regression, and Gehan-Breslow modification of Kruskal-Wallis tests (Thomas et al, 1977) on censored data.

Kaplan-Meier Product Limit Estimate,  $\hat{P}_i(t)$  (Slide 7):

$$\hat{P}_i(t) = \prod (1 - x_{ik}/n_{ik})$$

$$\sigma^2[\hat{P}_i(t)] = [\hat{P}_i(t)]^2 \sum x_{ik}/[n_{ik}(n - x_{ik})]$$

where

$x_{ik}$  = Number of animals with tumor at time  $t_k$  in group  $i$

$n_{ik}$  = Number of animals at risk at time  $t_k$  in group  $i$

The sums and the products are formed over the set of all  $k$  with times of deaths at or before  $t$ .

These estimates produce a step function for each group and are plotted over the lifetime of the animals. Planned, accidentally killed, and lost animals are censored. Moribund deaths are considered to be treatment related. The graphical representation of Kaplan-Meier estimates provide excellent interpretation of survival adjusted data except in the case where the curves cross between two or more groups. When the curves cross and change direction, no meaningful interpretation of the data can be made by any statistical method because proportional odds characteristic is totally lost over time. This would be a rare case where treatment initially produces more tumor or death and then, due to repair or other mechanisms, becomes beneficial.

Cox-Tarone Binary Regression (Tarone, 1975 ; Slides 8-10):

Define:  $V_{hi} = \sum \alpha_k A_{hk} (\delta_{hi} - A_{ik})$

where  $\alpha_k = x_{.k}(n_{.k} - \alpha_{.k}) / (n_{.k} - 1)$ ;  $A_{ik} = n_{ik} / n_{.k}$

and  $\delta_{hi} = 1$  if  $h=i$ , =0 otherwise;  $h, i=0, 1, \dots, I$ .

Defining  $V$  as the  $(I+1) \times (I+1)$  matrix of  $(h+1, i+1)$  entry  $V_{hi}$ , the homogeneity statistic  $\chi_h^2$  is derived:

$$\chi_h^2 = (O - E) V^{-1} (O - E) - \chi_h^2 (k-1)$$

where:  $O = \sum x_k$  (a vector of observed numbers)  
 $E = \sum \hat{e}_k$  (a vector of expected numbers)

The test statistic for linear trend  $\chi_t^2$  is derived as (Slide 9):

$$\chi_t^2 = [\sum L_i(O_i - E_i)]^2 / \sum V_{ij} L_i L_j - \chi_t^2 (1), \text{ for two-tailed test}$$

$$Z = \sqrt{\chi_t^2} - z(0,1), \text{ for one-tailed test}$$

The departure from linearity (not very sensitive) is given by  $\chi_d^2$ :

$$\chi_d^2 = \chi_h^2 - \chi_t^2 - \chi_d^2 (k-2)$$

Peto's log-rank test (Peto, 1974), tests based on the Weibull distribution, and Mantel-Haenszel (Mantel and Haenszel, 1952) test are very similar to this test when there are no covariates or stratifying variables in the design. The logistic regression based Cox-Tarone test is preferable because one can easily incorporate covariates and stratifying variables which one cannot in the IARC methods.

Gehan-Breslow Modification of Kruskal-Wallis Test (Slides 11-12):

The score associated with group  $i$  at time  $t$  ( $W_{ik}$ ) is given by

$$W_{ik} = n_{.k} x_{ik} - m_k n_{ik} \text{ with } W_i = \sum W_{ik} \text{ and } \underline{W} = (W_1, W_2, \dots, W_m)$$

Under the null hypothesis of no heterogeneity, the mean of  $W=0$  and

$$\text{Cov}(\underline{W}) = \underline{V}_W = \sum n_{.k}^2 \underline{V}_k = |V_{ij}^{(u)}|$$

The homogeneity statistic  $\chi_h^2 = \underline{W}' \underline{V}^{-1} \underline{W} \sim \chi_h^2 (k-1)$

The linear trend component of  $\chi_h^2$  is  $\chi_t^2$  (Slide 11):

$$\chi_t^2 = (\sum L_i W_i)^2 / [\sum \sum V_{ij}^{(u)} L_i L_j] - \chi_h^2 (1), \text{ for two-tailed test}$$

$$Z = \sqrt{\chi_t^2} - z(0,1), \text{ for one-tailed test}$$

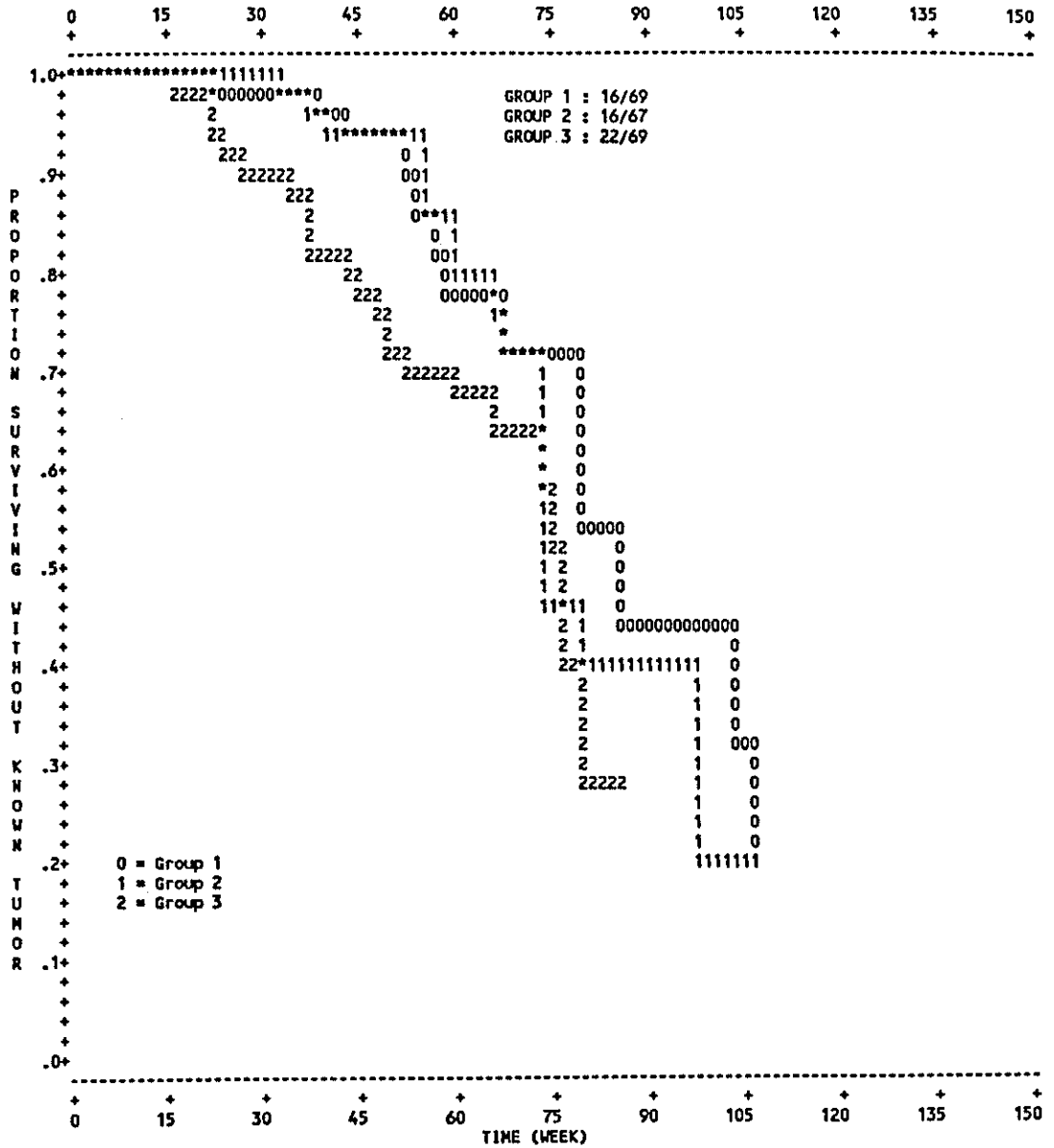
The departure from linearity (not very sensitive) is given by  $\chi_d^2$ :

$$\chi_d^2 = \chi_h^2 - \chi_t^2 - \chi_d^2 (k-2)$$

Gehan-Breslow test assigns more weight to early incidences compared to Cox-Tarone test.

The same methods apply to survival analyses where instead of palpation time one takes the time of death in the analyses. Animals dying or being sacrificed pre-planned are censored here. Slides 13-14 show an example.

KAPLAN-MEIER PRODUCT LIMIT ESTIMATES FOR MAMMARY FIBROADENOMA/CARCINOMA



Results of both unadjusted and life table analyses are shown below:

<u>Test</u>	<u>Trend p</u>	<u>Departure p</u>	<u>Heterogeneity p</u>	<u>0 vs 1 p</u>	<u>0 vs 2 p</u>
Unadjusted	0.1445	0.5782	0.4389	0.5423	0.1704
Cox-Tarone	0.0246 *	0.5002	0.0966	0.3698	0.0327 *
Gehan-Breslow	0.0097 **	0.2085	0.0293 *	0.4663	0.0146 *

As the above slide shows, because of earlier onset of these tumors, life table methods picked up treatment effect whereas the unadjusted tests on proportions could not.

(b) Occult or Incidental Tumors: The word "incidental" in carcinogenicity testing means tumors that are not cause of death of the animals; the animals died or were sacrificed for some other reasons. The tumors just happen to be there when necropsies are performed. Since the tumors are not observable during in-life phase of the study, there is no onset or palpation time for them. There are several types of occult tumors which require different types of analyses.

(i) Rapidly Lethal: If the pathologist, with some level of certainty, can determine that the occult tumor is rapidly lethal, the method of analysis is the same as in the case of palpable tumors. Instead of onset times for these tumors, one takes the time of death with such tumors as surrogate of onset time. This approximation is based on the fact that such a tumor generally kills the animals very shortly after induction. There are problems associated with such categorization as will be shown later.

(ii) Purely Incidental: The two most popular survival adjusted methods are the Dinse and Lagakos (1983) logistic regression based and Peto et al (1980) interval selection based prevalence techniques.

(a) Logistic Prevalence Method (Dinse and Lagakos, 1983): Tumor prevalence is empirically modeled as logistic function in dose, a polynomial in time, and a number of covariates (if needed) such as age of the animals, body weight, cage location, etc. It is briefly shown in Slide 15-17:

$$\theta(x,z,t) = \text{pr}(Y=1 | X=x, Z=z, T=t) / \text{pr}(Y=0 | X=x, Z=z, T=t)$$

$$= \exp(\alpha x w(t) + \beta z + \gamma(t))$$

where

$\alpha$  = A scalar of unknown parameters;

$\beta$  = A p-vector of unknown parameters

$\theta(x,z,t)$  = Prevalence odds

$w(t)$  = A pre-specified function of time

$\gamma(t)$  = A function of unknown parameters =  $\sum C^i t_i$ ,  $i$  is generally taken to be 2 (quadratic in time)

$Y = 0, 1$  (Binary tumor indicator)

$X$  = Dose (Treatment independent variable)

$Z = (Z^1, Z^2, \dots, Z^p)'$  = A p-vector of covariates

$T$  = Survival time

Test Statistic for Monotone Trend (Q):

$$\text{Score Function } U = \frac{\partial}{\partial \alpha} [\log(\text{likelihood})] = \sum W_j (n_j - E_j[Z])$$

where

$$E_j(Z) = \text{Expected number of tumor bearing animals} \\ = \sum [x_i (\exp(\beta z_i + \gamma_j) / (1 + \exp(\beta z_i + \gamma_j)))]$$

$$n_j = \text{Observed number of tumor bearing animals}$$

$$\text{Variance } \sigma^2(U) = \frac{\partial^2}{\partial \alpha^2} [\log(\text{likelihood})]$$

Finally:

$$Q = \hat{U}^2 / \hat{V} - \chi^2(1) : \text{For Two-tailed test for trend}$$

where

$$\hat{U} = U|_{\alpha=0}; \hat{V} = V|_{\alpha=0}$$

$$\text{For one-tailed test: } Z = \sqrt{Q} - z(0,1)$$

One can similarly derive two-sample test for 2 x 2 comparisons. Further, dose x time interactions are easily derived.

(b) IARC Interval Based Test (Peto et al, 1981) (Slide 18):

$$T = \sum D_i (O_i - E_i); \sigma^2(T) = \sum \sum D_i D_j \sigma_{ij}^2$$

$$\chi^2 = T^2 / \sigma^2(T) - \chi^2(1) : \text{For two-tailed test}$$

For one-tailed test,

$$Z = \sqrt{\chi^2} - z(0,1)$$

There are several features in the logistic prevalence method which makes it a better procedure (Slide 19):

- (i) Unlike the IARC method, it does not require interval selection
- (ii) Non-isotonic prevalences do not have any effect on the analysis
- (iii) Covariates and stratifying variables can be used to fully explore the nature of treatment effect

The IARC method, on the other hand, not only do not have the benefits of the prevalence method, it could introduce bias in the analysis because of the interval selection. With a different interval selection scheme, one could get different significance. The only advantage of this method is that it does not require maximum likelihood estimation which may be time consuming. In any case, for many experimental data sets, where covariates etc. are not needed to explore the nature of treatment effect, both methods provide reasonable approximation of statistical significance. We will examine the outcomes of both methods with examples.

For all the unadjusted and adjusted analyses we have discussed, it is possible to develop permutation based exact hypothesis testing. For large samples the methods could be extremely time consuming; however, with sparse tables, they should be utilized.

### DISCUSSION

In discussion, we will deal with some special issues arising in statistical analyses of tumor data with examples whenever appropriate. For the sake of confidentiality, the names of the chemicals cannot be mentioned in this discussion.

(a) Trend vs Heterogeneity: Group comparison tests for proportions notoriously lack power. Trend tests, because of their use of prior information (dose levels) are much more powerful. Also, it is generally believed that the nature of true carcinogenicity (or toxicity for that matter), manifests itself as dose-response. Because of the above facts, evaluation of trend takes precedence over group comparisons. In fact, in the absence of a trend, scattered significant group comparisons may be overwritten as being toxicologically irrelevant in an oncogenicity study. Also, since hypothesis testing (qualitative risk assessment) is a preliminary step of risk assessment (quantitative risk assessment), it is only natural that the investigator looks for to see whether there is a dose-response in the data. In this context, as will be shown later, one should examine the validity of the trend test carefully because it is weighted heavily by the magnitude of the high dose effect level. Often these studies have unequally spaced dose levels both on the arithmetic as well as on the logarithmic scale. In order to achieve optimal test statistics, many people use ordinal dose levels (0,1,2...etc.) instead of the true arithmetic dose levels to test for trend. However, such a decision should be made a priori. Slides 20-22 show a hypothetical case of tumor incidences with weakness of heterogeneity tests:

<u>Number at Risk</u>	<u>Number with Tumor</u>	<u>Dose Level</u>
50	2	0
50	4	1
50	6	2
50	7	3

#### COCHRAN-ARMITAGE TEST FOR TREND

	CALC. CHI2	DF	ALPHA	2-TAIL P
TREND	3.3446	1	.0500	.0674
DEPARTURE	.0694	2	.0500	.9659
HOMOGENEITY	3.4141	3	.0500	.3321



ONE-TAIL TESTS FOR TREND

TYPE	PROBABILITY
UNCORRECTED	.0337 *
CONTINUITY CORRECTED	.0426 *
EXACT	.0423 *

DIRECTION = +

MULTIPLE PAIRWISE GROUP COMPARISONS  
BY FISHER-IRWIN EXACT TEST

GROUPS COMPARED	ALPHA	ONE-TAIL PROB.
1 VS. 2	.0500	.33887
1 VS. 3	.0500	.13433
1 VS. 4	.0500	.07975

As is evident from this example, often group comparison tests will fail to identify significant treatment but trend test will. Same arguments apply to survival adjusted tests on proportions as well.

(b) One-tailed vs Two-tailed Hypothesis: In carcinogenicity testing, both trend and heterogeneity of groups are tested using one-tailed significance levels because of the nature of the alternative hypothesis (increased tumors with dose).

(c) False Positive Error Rate: Most group comparison tests in carcinogenicity testing are two-sample tests use as multiple comparisons. This may open the possibility of high false positive error rate. Gart et al (1979) and Haseman (1983) address this issue in detail. The U.S. regulatory agencies such as the Food and Drug Administration (FDA) and the Environmental Protection Agency (EPA) generally rely on tests for trend in tumor data. In the presence of a trend, they do not accept adjustments for multiple comparisons. In the absence of a trend, adjustments can be made in multiple comparisons. The thought behind such an approach is that a trend or dose-response generally does not happen by chance alone whereas one or two treated groups may show significant heterogeneity by chance alone. In any case, our approach is to make such adjustments in the absence of a trend through the application of Bonferroni inequality (Miller, 1981) (Slide 23):

$$\text{pr}[U(P_i \leq \alpha/n)] \leq \alpha (0 \leq \alpha \leq 1)$$

where

n = number of comparisons to be made  
 $\alpha$  = experimentwise error rate

The null hypothesis is rejected if  $P \leq \alpha/n$  rate. When n is too large, the adjustment provides very conservative tests.

Some statisticians believe that because of the number of tumor types analyzed in a carcinogenicity study, such adjustments should be made based on the number of such tumor types as well. As can be readily seen, such an approach will effectively make every comparison nonsignificant. Gart et al (1979) and Haseman (1983) provide discussion and procedures to approach such a scenario. Briefly,

they propose that for common spontaneous tumors which arise in large numbers, such as mammary tumors in female Sprague-Dawley rats, the comparison be made at  $p \leq 0.01$  level and for tumors that are not so common, such as mammary tumors in male Sprague-Dawley rats, the comparisons are made at  $p \leq 0.05$  level. Some preliminary works are underway using multivariate contingency techniques to minimize such errors. Generally, regulatory agencies do not subscribe to this concept.

(d) Continuity Correction: The test statistics used for both unadjusted and adjusted analyses come in two different forms: two-tailed and one-tailed. As we mentioned earlier, the one-tailed forms are the ones that are used by the regulatory agencies. The one-tailed versions are derived from the two-tailed ones asymptotically using the standardized normal deviate. Strictly speaking, such an asymptoticity requires a continuity correction. It can be shown that a continuity corrected one-tailed probability approximates the exact significant tail probability under permutation. Such a correction is done as follows with any of the trend statistics we have described so far (Slide 24):

$$Z_{\text{corr}} = \frac{\sqrt{(\text{Numerator of two-tailed statistic}) + \text{sgn}(\Delta)/2}}{\sqrt{(\text{Denominator of the two-tailed statistic})}} - z(0,1)$$

where

$\Delta$  = the difference between two adjacent dose levels  
 $\text{sgn} = -$  when the effect is direct (positive trend) and  $+$  when the effect is inverse (negative trend)

Unfortunately, because of unequal spacing of the dose levels, there may not be any unique  $\Delta$  in many studies. As a result, there is no unique continuity correction there either. Gart et al (1986) proposed bounds for such cases with  $\Delta_{\text{min}}$  and  $\Delta_{\text{max}}$ . A unique correction can be made if one uses rank or ordinal dose levels in computation of trend. As we saw in the previous example, there may be significant disagreement between uncorrected and continuity corrected p-values in many cases. With sparse tables, the problem becomes worse as we will see later. Whenever possible, one should try to exact permutation probabilities for these cases.

(e) Sparse Tables: As mentioned earlier, sparse tables cause a lot of problems in interpreting carcinogenicity data. These tables generally arise when there are uncommon or rare tumors such as brain astrocytoma or liver hepatoblastoma seen in a study. Asymptotic unadjusted or adjusted methods provide either no statistical significance or incorrect p-values even when there is indication of significance. For the unadjusted case, the classical "expected frequency = 5" practice may be emphasized and one should then use exact permutation tests for such tables. For survival adjusted cases this may be difficult or time consuming except in special circumstances. For example, if the life span of the animals can be subdivided into 5-10 intervals, exact probability tests may be feasible as a general practice. Computer programs are available for such special cases.

Slides 25-27 deal with an example involving both continuity correction and sparse table.

<u>Number at Risk</u>	<u>Number with Tumor</u>	<u>Dose Level</u>
50	0	0
50	0	1
50	0	2
50	2	3

COCHRAN-ARMITAGE TEST FOR TREND

	-----	-----	-----	-----
	CALC. CHI2	DF	ALPHA	2-TAIL P
TREND	3.6182	1	.0500	.0572
DEPARTURE	2.4121	2	.0500	.2994
HOMOGENEITY	6.0303	3	.0500	.1101

ONE-TAIL TESTS FOR TREND

-----	-----
TYPE	PROBABILITY
UNCORRECTED	.0286 *
CONTINUITY CORRECTED	.0565
EXACT	.0616

DIRECTION = +

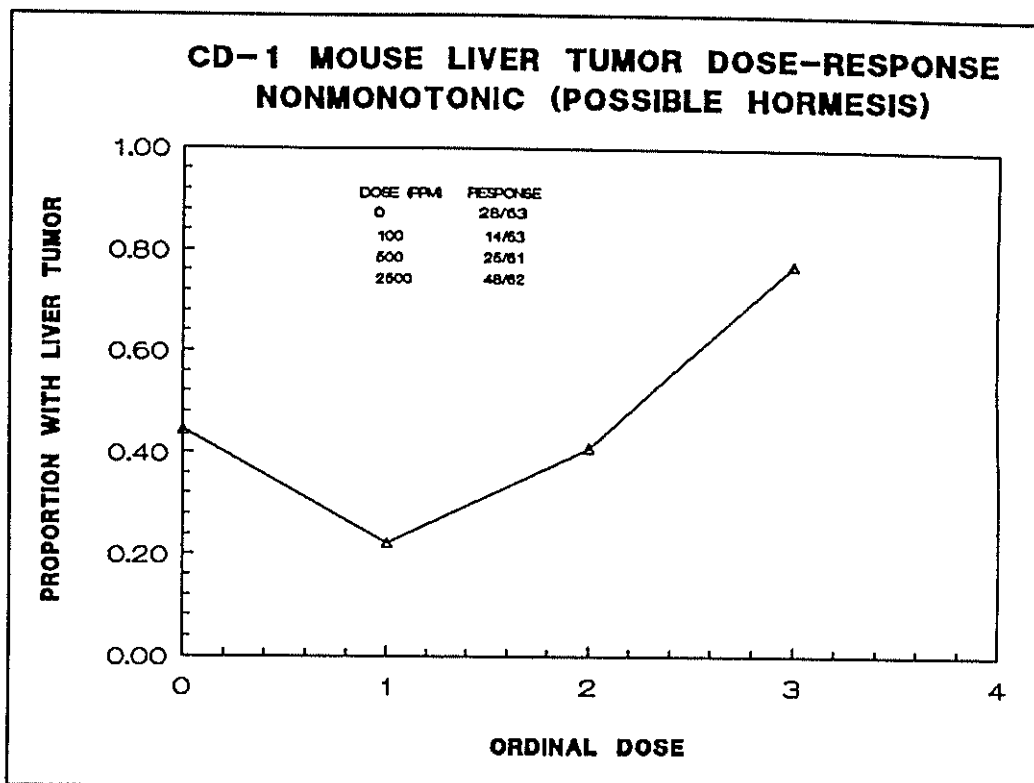
MULTIPLE PAIRWISE GROUP COMPARISONS  
BY FISHER-IRWIN EXACT TEST

-----	-----	-----	-----
GROUPS COMPARED	ALPHA	ONE-TAIL PROB.	
1 VS. 2	.0500	1.0000	
1 VS. 3	.0500	1.0000	
1 VS. 4	.0500	.2475	

The results from this example clearly shows that one would make misinterpretation of the data from this sparse table if one used uncorrected probability. The continuity corrected p-value comes closer to the exact p-value but there may be occasions where even that closeness is not enough.

(f) Problem of Threshold: Many biological functions indicate the presence of threshold where dose levels have to exceed a certain level predetermined by the system before one sees any treatment effect. Tumor formation is no exception. Many chemicals exhibit such thresholds. Unfortunately, tests for trend, even though they are strictly for linear to monotone increasing or decreasing functions, unduly overestimate the test statistics or underestimates the tail probabilities for such cases. This is because all these test statistics are heavily weighted by the magnitude of the high dose which may be sometimes 100-fold higher than the preceding dose level. An exact test under these conditions may be slightly better than the asymptotic ones although interpretation of such data still may be problematic. The previous case can also be taken as a case of threshold.

(g) Hormesis: Hormesis (Boxenbaum et al, 1988) is a well recognized toxicological mechanism by which the low dose may actually be beneficial for the animals. In other words, one may find a biphasic dose-response curve. Unfortunately, the tests available for us are unable to interpret such data. The high dose group will once again dominate the component for trend. Slides 28-30 show such an example.



COCHRAN-ARMITAGE TEST FOR TREND

	CALC. CHI2	DF	ALPHA	2-TAIL P
TREND	17.2426	1	.0500	.00000 **
DEPARTURE	22.2302	2	.0500	.00000 **
HOMOGENEITY	6.0303	3	.0500	.00000

ONE-TAIL TESTS FOR TREND

TYPE	PROBABILITY
UNCORRECTED	.00000 **
CONTINUITY CORRECTED	.00000 **
EXACT	.00002 **

DIRECTION = +

MULTIPLE PAIRWISE GROUP COMPARISONS  
BY FISHER-IRWIN EXACT TEST

GROUPS COMPARED	ALPHA	ONE-TAIL PROB.
1 VS. 2	.0500	.0068 (-) **
1 VS. 3	.0500	.4178 (-)
1 VS. 4	.0500	.0000 (+) **

Conclusion from this example is trivial. Fortunately for this example, since the departure from monotonicity is so severe that the departure statistic was able to pick it up.

What is happening in the examples for threshold and hormesis is that we are basically committing Type III error, i.e. testing the wrong hypothesis (linear trend or monotone trend) with the right data. Since these mechanisms are very difficult to prove experimentally, regulatory scientists generally ignore them indicating that they are noise. We as statisticians must explain any such irregularities in the data and point out the possibility of misinterpretation. As a statistician who has been practicing in this trade for over 15 years, I do not have any qualms about admitting the weakness of many statistical tests under unusual circumstances. Not all biological problems can be solved statistically given the present state of knowledge.

(h) Misclassification Problems in Tumor Data Evaluation: As mentioned earlier, the numbers at risk for lethal or palpable tumors are different from the numbers at risk for incidental but nonlethal tumors. Let us see what happens if a pathologist misclassifies an occult tumor type in his review and the statistician is asked to follow it through (Slide 31).

	<u>WEEK 50</u>		<u>WEEK 75</u>	
	<u>GP 1</u>	<u>GP 2</u>	<u>GP 1</u>	<u>GP 2</u>
TOTAL NUMBER	100	100		
NUMBER LIVE	95	80	75	75
NUMBER DEAD	5	20	20	5
NUMBER WITH TUMOR	2	8	8	2
INCIDENTAL TUMOR RATE	2/5	8/20	8/20	2/5
	(0.40)	(0.40)	(0.40)	(0.40)
P-VALUE	1.0000		1.0000	
LETHAL TUMOR RATE	2/95	8/80	8/75	2/75
	(0.021)	(0.100)	(0.110)	(0.027)
P-VALUE	0.0268 (+) *		0.0488 (-) *	
UNADJUSTED TUMOR RATE (LIFE-TIME)	10/100 VS 10/100			
	(0.100)		(0.100)	
P-VALUE	1.0000			

In this hypothetical case, if the tumors were really lethal, one would get completely wrong conclusion by incidental tumor methods. On the other hand, by the lethal tumor method one has the interesting scenario whereby treatment is causing detrimental effect early but beneficial effect late during the life span. This is a very unusual happening indeed! If the tumors were truly incidental then of course lethal tumor method misrepresents treatment effect.

(i) Mixtures of Occult Tumors: In the IARC classification scheme, the occult tumors are actually subdivided as follows (Slide 32):

- (1) Definitely Incidental
- (2) Probably Incidental
- (3) Probably Lethal
- (4) Definitely Lethal

The original proposal calls for pooling (1) and (2) together, and (3) and (4) together. One then proceeds as follows (Slide 33):

Pooled (1) and (2) : Incidental Tumor Analysis  
Pooled (3) and (4) : Lethal Tumor Analysis  
Combine evidence from both

For example (Slides 34-35):

$n_i$  = Total number of animals in i-th group

$t_{i1}$  = Number of incidental tumor bearing animals in i-th group

$t_{i2}$  = Number of lethal tumor bearing animals in i-th group

Perform incidental tumor analysis with  $n_i - t_{i1}$  animals at risk and perform lethal tumor analysis with  $n_i$  animals at risk out of which  $t_{i1}$  are censored.

Let

$T_i$  = Score statistic from incidental tumor analysis

$\sigma_i^2$  = Variance of  $T_i$

$T_i$  = Score statistic from lethal tumor analysis

$\sigma_i^2$  = Variance of  $T_i$

Combined evidence is given by:

$$Z_c = [T_i + T_i + \text{sgn}(\Delta/2)] / (\sigma_i^2 + \sigma_i^2) - z(0,1)$$

The above is a continuity corrected one-tailed test with the sign being assigned as before.

There is one important concern about such splitting of tumor bearing animals in categories. Suppose in a study the two tumor types are in different direction due to treatment. The lethal type shows decrease due to treatment and the incidental type shows an increase. No meaningful conclusion can be made then. This is a hypothetical rare case but could happen in reality.

(j) Extreme Mortality Problems: Often a chemical being tested is both toxic as well as potentially carcinogenic. In many cases, e.g. CD-1 mice, amyloidosis appears in large numbers of animals in each group. If the amyloidosis is severe or extends to multiple vital organs, many of the animals will die prematurely. Added to that is the burden of toxicity. As a result, very few animals may remain at the end of the study to satisfy the FDA requirement of at least 50% or 25 animals surviving at the end of the study. In Sprague-Dawley rats there may

have been a genetic drift causing in severely decreased survival, as little as 20-25%, in recent years in some studies. As a result, one may face the situation where one may have to repeat the study. This results in killing a lot of animals, spending a lot of money, and wasting at least 2-3 years of time. If the question regarding the compound under investigation is carcinogenicity and there is no indication of delayed carcinogenicity (geriatric effect), and if the study indicates both a NOAEL and MTD, one may argue and possibly win regarding the validity of the study. To be on the safe side, one can easily prevent such a scenario by assigning a few more animals, 10-15, in each group. Also, it is prudent to start a chronic study with at least five groups of animals per sex. Scientifically it makes more sense to design these studies such. Even though it may seem initially that it is costing more money, in the long run one gets more accurate information from the study at the same time avoiding a much larger expense. What may sound like an optimal design statistically, may turn out to have flaws because of toxicological reasons. After all, the dose levels and their spacings are often based on some crude short time studies or results from previous studies. Because of genetic drift, results from previous studies may not be as reliable as one would like to think they are.

When competing toxicity causes extreme differences in mortality or there is clustering effect in tumor prevalence in a very short interval of time, none of the adjusted methods works. One then must use biological intuition to explain the tumor data.

(k) Double Negative Controls: Many European and a few American pharmaceutical companies sometimes put two negative controls in a chronic study in the hope of investigating biological and statistical variations in the data. This creates statistician's nightmare particularly when the two controls seem to be different. This may lead to opening Pandora's Box in some cases. Questions arise: "why are the two controls different from each other? Is it because they may have been treated differently? Were there cage allocation problems as in some of the dye studies? etc.". Since the control group is used multiplicatively in hypothesis testing, it makes sense to have a larger such as double the size control to achieve adequate power. In any case, when such a situation arises, it is desirable that one first compares the two controls by the appropriate method described earlier. If the two controls are very similar, one then combines them for a single analysis. If they are different but the effect against each control is in the same direction, one can combine the evidence from each comparison by Fisher's method (1958) (Slide 36) or any other method. In the case the two controls are different and the effects against each control are in different directions, then the analyses should not be combined and the toxicologist and the pathologist should investigate any possible design flaws or provide justification of the treatment outcome.

$$\chi^2 = -2 \sum \ln p_i \sim \chi^2 (2k)$$

where

k = Number of tables

$p_i$  = Probability associated with each table

(l) Use of Historical Incidence Data: Use of these data in statistical evaluation is controversial. There are too many sources of variation in these data. For example, different pathologist uses different criteria for categorizing tumors (in fact, the same pathologist changes opinion over time); there is laboratory-to-laboratory variation; there may be genetic drift over time;

location of suppliers may make a difference; and finally, these data are not part of the randomized current control. Methods have been described by Tarone (1982) and Smythe et al (1986). It seems that inclusion of these data in analysis may produce stronger evidence of carcinogenicity (because of increased power) for the cases of uncommon and rare tumors. Regulatory agencies and pathologists generally use these data for qualitative evaluation. My personal view is that that is where they belong.

(m) Pooling of Neoplasms: In many instances benign tumors (adenoma, hyperplasia) of certain origin may progress to be malignant tumors (carcinoma). Under those circumstances, it is advisable to pool these lesions for statistical analysis. McConnell et (1986) provide an extensive list for such tumors. One reason for such pooling is the fact that often what one pathologist may call an adenoma may be called a carcinoma by a second pathologist. In any case, even when it advised to pool such lesions, one should still analyze the carcinoma incidences alone. It may so happen that for certain spontaneous tumors, the total number of tumor bearing animals may be approximately be the same, but because of treatment, many of the adenomas became carcinomas. Promotion mechanisms are notorious for producing such effects.

(n) Method to be Avoided: Some people propose to use only the animals surviving after one year in the study or use a two-strata approach where they do separate analyses for the tumors in animals killed during the first year of the study and the ones there after. Or they sometimes exclude all animals in the study that died on test and analyze only the animals that are sacrificed at the end of the study. Many tumors (e.g. mammary, liver, lung etc.) may appear within the first year of a study. In fact, it is not unusual to see mammary tumors by week 26 to 30 in Sprague-Dawley rats. As a result, the assumption of first year kill animals not being at risk is not correct in many cases. As a result, exclusion of those animals, as a general rule, is inappropriate. On the other hand, supposing the purpose is to perform two sets of analyses, one at the end of the first and the second (after exclusion of the first year kills) at the end of study. It is quite conceivable that the conclusions from these two segments may be contradictory and as a result, combining the evidence would be inappropriate. Similar problem arises when the Kaplan-Meier curves from two groups intersect each other and change the overall direction. No meaningful conclusion can be drawn under these circumstances because of contradictory evidence from the segments. In cases like this, the prevalence function is strictly nonisotonic. A second major problem arises during combining evidence from segments of a study. This is due to the reduction in sample sizes and numbers observed with tumors. Let me illustrate this situation with a hypothetical but conceivable example (Slide 37):

Week 52 Rates		Week 104 Rates		Lifetime Rates	
Group 1	Group 2	Group 1	Group 2	Group 1	Group 2
0/25	2/25	0/25	3/25	0/50	5/50

Fisher-Irwin Exact Test at Week 52 : p-value = 0.2449  
 Fisher-Irwin Exact Test at Week 104 : p-value = 0.1174  
 Combined Evidence by Fisher's Method : p-value = 0.1307



Fisher-Irwin Exact Test on Lifetime Rates: p-value = 0.0281 \*  
 Logistic Prevalence Test : p-value = 0.0331 \*  
 Peto et al (IARC) Interval Based Method : p-value = 0.0115 \*  
 Exact Interval Based Method : p-value = 0.0287 \*

As is evident from this example and the associated statistical analyses, such individuals fail to find treatment effect either individually by intervals or by combination of evidence. All other methods, on the other hand, find significant evidence of higher incidence rate in Group 2. One reason the Peto interval based method shows a smaller p-value is that the method does not perform continuity correction before applying the normal approximation to the  $\chi^2$  statistic. Since the mortality rates in both groups are similar in this study, Fisher-Irwin exact test on unadjusted data for the lifetime incidence rates, logistic prevalence analysis, and the exact interval based method provide very similar p-values for this example.

(o) Miscellaneous Issues: Tumor multiplicity is a rather difficult problem because of dependency as well as nonrandomization. Drinkwater and Klotz (1981) Gart et al (1986) address this in detail.

Gradation of tumors such as slight, moderate, severe, etc. or by numbers can be handled by Bhapkar's method (1965) or by RIDIT method (Bross, 1958; Selvin, 1972) for R x C contingency tables. One should be warned by the fact that such gradations may be rather subjective which might introduce bias in the analysis.

The final issue is subjectivity in slide reading by most pathologists who do not want to read them in a coded fashion whereby they will not know the group an animal is coming from. This is not under statisticians' control but they should be aware of that in any case.

## BIBLIOGRAPHY

### General Topic

D.G. Hoel and H.E. Walburg, Statistical analysis of survival experiments, J. Nat. Cancer Inst., 49:361-372, 1972.

R. Peto, M.C. Pike, N.E. Day, R.G. Gray, P.N. Lee, S. Parish, J. Peto, S. Richards, and J. Wahrendorf, Guidelines for Simple, Sensitive Tests for Carcinogenic Effects in Long-term Animal Experiments, In Long-term and Short-term Screening Assays for Carcinogens: A Critical Appraisal, IARC Monographs Suppl. 2, pp 311-426, Lyon, France, 1980.

J.J. Gart, D.Krewski, P.N. Lee, R.E. Tarone, and J. Wahrendorf, The Design and Analysis of Long-term Animal Experiments, IARC Scientific Publications No. 79, Lyon, France, 1986.

A.K. Thakur, L.T. Wetzel, and J.T. Stevens, Statistical issues in lifetime oncogenicity studies in rodents, Proc. Biopharmaceutical Section, American Statistical Association, 275-281, 1990.

### Logistic Regression

D.R. Cox, The Analysis of Binary Data, Chapman and Hall, London, 1970.

### Unadjusted Tests

A.K. Thakur, K.J. Berry, and P.W. Mielke, Jr., A FORTRAN program for testing trend and heterogeneity in proportions, Comput. Progr. Biomed., 19:229-233, 1985.

### Life Table Analyses

D.G. Thomas, N. Breslow, and J.J. Gart, Trend and homogeneity analyses of proportions and life table data, Comput. Biomed. Res., 10: 10:373-381, 1977.

R. Peto, Guidelines on the analysis of tumor rates and death rates in experimental animals, Br. J. Cancer, 29:101-105, 1974.

R. Peto and P. Lee, Weibull distributions for continuous carcinogenesis experiments, Biometrics, 29:457-470, 1973.

N. Mantel and W. Haenszel, Statistical aspects of the analysis of data from the retrospective studies of disease, J. Nat. Cancer Inst., 22:719-748, 1952.

R.E. Tarone, Tests for trend in life table analysis, Biometrika, 62:679-682, 1975.

### Prevalence Analysis

G.E. Dinse and S.W. Lagakos, Regression analysis of tumour prevalence data, J. Roy. Stat. Soc. Series B (Appl. Stat.), 32:236-248, 1948.

### False Positive Error Issues

J.J. Gart, K.C. Chu, and R.E. Tarone, Statistical issues in interpretation of chronic bioassay tests for carcinogenicity, J. Nat. Cancer Inst., 62:957-974, 1979.

J.K. Haseman, A re-examination of false positive rates for carcinogenesis studies, Fund. Appl. Toxicol., 3:334-339, 1983.

### Bonferroni Adjustment

R.G. Miller, Jr., Simultaneous Statistical Inference, Springer-Verlag, Heidelberg, 2nd Ed., 1981.

### Pooling of Neoplasms

E.F. McConnell, H.A. Solleveld, J.A. Swenberg, and G.A. Boorman, Guidelines for combining neoplasms for evaluation of rodent carcinogenesis studies, J. Nat. Cancer Inst., 76:283-289, 1986.

### Use of Historic Control

R.T. Smythe, D. Krewski, and D. Murdoch, The use of historical control information in modelling dose-response relationships in carcinogenesis, Stat. Prob. Letters, 4:87-93, 1986.

R.E. Tarone, The use of historical control information in testing for a trend in proportions, Biometrics, 38:215-220, 1982.

J.K. Haseman, J. Huff, and G.A. Boorman, Use of historical control data in carcinogenicity studies in rodents, Toxicol. Path., 12:126-135, 1984.

### Tumor Multiplicity

N.R. Drinkwater and J.H. Klotz, Statistical methods for the analysis of tumor multiplicity data, Cancer Research, 41:113-119.

### Categorized Neoplastic Tables

I.D.J. Bross, How to use riddit analysis, Biometrics, 14:18-38, 1958.

S. Selvin, A further note on the interpretation of riddit analysis, Am. J. Epidemiol., 105:16-20, 1977.

V.P. Bhapkar, On the analysis of contingency tables with a quantitative response, Biometrics, 24:329-338, 1968. I.D.J. Bross, How to use riddit analysis, Biometrics, 14:18-38, 1958.

### Hormesis

H. Boxenbaum, H., P.J. Neafsky, and D.J. Fournier, Hormesis, Gompertz functions and risk assessment, Drug Metab. Rev., 19:195-210, 1988.

### Computer Software

1. Unadjusted Analyses: Thakur et al (loc cit)- FORTRAN
2. Life Table Analyses: Thomas et (loc cit)- FORTRAN; Peto et al (loc cit)- FORTRAN
3. Prevalence Analyses: Dinse and Lagakos (loc cit)- FORTRAN; Peto et al (loc cit)- FORTRAN
4. Exact Interval Based Tests: M. Ali, U.S. FDA, Rockville, Maryland, .U.S.A.- BASIC.
5. Exact Logistic Prevalence Test: LogXact: CYTEL Software Corp., 675 Massachusetts Avenue, Cambridge, MA02139, U.S.A. (Executable PC DOS Version; \$495.00 Commercial Version). The same group also sells StatXact, which provides permutation exact tests for many asymptotic procedures. StatXact sells for \$495.00 for regular DOS and \$895.00 for extended memory DOS versions commercially.

---

第54回定例会（93年4月10日）講演より

# 癌原性試験の統計解析の手順

浜田 知久馬 (東京大学)

## 1. 統計解析手法

### (1) 生存時間について調整を行わない腫瘍発生割合の検定

群間の等質性の検定：カイ2乗検定

傾向性検定 : コ克蘭・アーミテッジ検定

群間の対比較 : フィッシャーの正確検定

### (2) 生存時間について調整を行う腫瘍発生割合の検定

#### (a) 触診によって観測可能な腫瘍, 表在性の腫瘍, および致死的な腫瘍

累積腫瘍発生割合の推定: Kaplan-Meier法

検定: Cox-Tarone test (ログランク検定)

Gehan-Breslow のクラスカル・ワリス検定

(一般化Wilcoxon検定)

それぞれについて群間の等質性の検定, 傾向性検定が可能

#### (b) 偶発的に発見される(非致死的な)腫瘍

(i) 腫瘍発生割合についてのロジスティック回帰

(ii) IARCの方法(Petoの方法)

## 2. デスカッション

### (a) 傾向性検定と群間の等質性の検定

群間の対比較より傾向性検定の方が検出力が高い。ただし最高用量の重みが高くなる傾向があると、用量のスケール(生の用量, 対数用量, 順序スコア等)をどうとるかによって結果が異なってくる可能性があるので注意が必要である。

### (b) 両側検定と片側検定

癌原性試験では想定する対立仮設の構造から片側検定を用いるべきである。

### (c) 第1種の過誤の大きさ

群間の対比較を多数行う場合で、第1種の過誤を名義有意水準以下に抑えたいときはBonferroni法による調整等を行う。

### (d) 連続修正

等間隔でない用量との傾向性検定については連続修正の式が1つに決まらない。

腫瘍発生割合が小さい場合には、連続修正を行う場合と行わない場合で結果がかなり異なる場合があるが、このような場合には正確な検定を行った方がよい。

### (e) スパースな分割表

まれにしか発生しない腫瘍について解析する場合、度数0のセルが多いまばらな分割表が生じる。このようなデータを解析する場合には(d)で述べたことを含めて様々な統計解析上の問題点が生じる。

(f) シキイ値の問題

(g) 用量反応に単調関係が成り立たない場合がある。

(h) 腫瘍の致死性についての誤分類

非致死的な腫瘍を誤って致死的な腫瘍と分類することによって、全く反対の結論を導く可能性がある。

(i) 腫瘍の種類によっては個体に対して、致死的な影響を与えたり、与えなかったりする場合がある。IARCの方法では致死的な腫瘍の解析結果と、致死的でない腫瘍の解析結果をまとめて評価を行うが、処置によって致死的な腫瘍は増加し、非致死的な腫瘍は減少するような場合、発癌性が隠される。

(j) 処置によって極端に死亡率が高くなる場合

薬物の毒性によって死亡数が極端に増えるような場合、結果の解釈が極端に難しくなる場合がある。

(k) 2つの対照群(dual control)をおいた場合の結果の評価

欧州系の製薬企業では、全く同一の対照群を2つにおいて試験を行うことが多いが、2つの対照群の間で結果が大きく異なるような場合には結果の解釈は難しくなる。

(l) 背景データの利用

(m) 悪性腫瘍と良性腫瘍をプールした解析

良性腫瘍が進行して悪性腫瘍になるような場合には、悪性腫瘍と良性腫瘍をプールして解析を行う場合がある。このような解析が行われる理由の1つは悪性腫瘍と良性腫瘍の区別が病理学者によって異なる可能性があるからである。

(n) 推薦できない方法

・試験開始初期に死亡した動物を除いて解析を行う。

・1年半までとそれ以降というように区間ごとに解析を行う。

区間ごとに相反する結果がでる場合があり、またサンプル・サイズが減るため検出力が下がる

(o) その他の問題

・腫瘍の種類の多重性

・腫瘍のグレードを反映させた解析

・病理組織を観測する際のブラインド化

# 発癌性の指標 TD<sub>50</sub> (Tumor Dose<sub>50</sub>) の推定について

長谷川 良平 (名古屋市立大学)

浜田 知久馬 (東京大学)

## 1. はじめに

ある薬物の薬効あるいは毒性の強さを表す指標として ED<sub>50</sub> (effective dose 50), LD<sub>50</sub> (lethal dose 50) 等が知られている。このうち LD<sub>50</sub> は急性毒性を表す指標で、薬物を実験動物に単回投与したときに 50% の動物を死に至らしめる用量として定義される。これに対し薬物を反復投与したときの発癌性を表す指標が TD<sub>50</sub> (Tumor Dose 50) である。急性毒性を表す指標と慢性毒性を表す指標で決定的に異なるのは、時間という要素が加わる点である。用量を増やすことに加えて、投与期間を長くすることによっても毒性は増加することが予想される。したがって投与期間を固定した上で毒性の強さを評価する必要がある。TD<sub>50</sub> はその薬物を実験動物に一生涯投与したとき 50% の動物に腫瘍を発生させる体重キログラム当たりの量として定義される。一生涯の期間の長さは動物の寿命の長さによって異なってくる。対照群と薬物のある用量を与えた用量群の 2 群から構成される実験データから TD<sub>50</sub> を求めるために (1) 式が知られている。

$$TD_{50} = \frac{\log 2 \times d \times t^*}{\log(1 - n_c / N_c) - \log(1 - n_d / N_d)} \quad (1)$$

- d : 実験用量
- t\* : 実際の実験期間 (t) と基準化する期間 (t<sub>0</sub>) の比  
t<sub>0</sub> : rat 756 日, mouse 630 日, hamster 630 日
- n<sub>c</sub> : 担癌動物数 (対照群)
- N<sub>c</sub> : 有効匹数 (対照群)
- n<sub>d</sub> : 担癌動物数 (薬物投与群)
- N<sub>d</sub> : 有効匹数 (薬物投与群)

本発表では、この式が導出された背景にあるモデルについて説明し、次に示すような多群の実験データについて、TD<sub>50</sub> を推定する方法を示す。

### データの例

実験期間	用量	大腸癌 n/N (%)	乳癌 n/N (%)	
52W	400ppm	16/29 (55)	14/30 (47)	①
52W	0ppm	0/40 (0)	0/40 (0)	
104W	100ppm	13/30 (43)	14/30 (47)	②
104W	25ppm	0/30 (0)	2/30 (7)	
104W	0ppm	0/30 (0)	0/30 (0)	

## 2. 計算例と問題点

(1) 式を適用して、TD50を求める場合、3通りのTD50が計算される。

: 大腸癌

$$\textcircled{1} \quad \text{TD50} = \frac{\log 2 \times 400 \times (52/104)^3}{\log(1 - 0/40) - \log(1 - 16/29)} = 43.21 \quad (t_0 = 104W)$$

$$\textcircled{2} \quad \text{TD50} = \frac{\log 2 \times 25 \times (104/104)^3}{\log(1 - 0/30) - \log(1 - 0/30)} = \text{不定}$$

$$\textcircled{3} \quad \text{TD50} = \frac{\log 2 \times 100 \times (104/104)^3}{\log(1 - 0/30) - \log(1 - 13/30)} = 122.03$$

: 乳癌

$$\textcircled{1} \quad \text{TD50} = \frac{\log 2 \times 400 \times (52/104)^3}{\log(1 - 0/40) - \log(1 - 14/30)} = 55.13$$

$$\textcircled{2} \quad \text{TD50} = \frac{\log 2 \times 25 \times (104/104)^3}{\log(1 - 0/30) - \log(1 - 2/30)} = 251.17$$

$$\textcircled{3} \quad \text{TD50} = \frac{\log 2 \times 100 \times (104/104)^3}{\log(1 - 0/30) - \log(1 - 14/30)} = 110.28$$

TD50を(1)式にしたがって求めるには、次の問題点がある。

1) TD50の推定値が①、②、③でかなり異なる。TD50の定義にしたがえば、大腸癌も乳癌も100ppmの104Wで担癌率がほぼ50%になっているので、TD50は100ppm近くの値になるはずである。

2) 個々の時点で、個々の用量群ごとにTD50の値を推定しているが、5群全部の情報を使って推定した方が推定精度がよくなると考えられる。

## 3. TD50の推定の背景にあるモデル

(1)式は、腫瘍の発生に関する時間と薬物の影響について次のようなモデルを想定することで導出される。

$$h(t, d) = 3 \times \lambda d t^2 \quad (2)$$

$h(t, d)$  : ハザード関数、ある瞬間に癌になる確率(リスク)は、用量  $d$  に比例し、時間  $t$  の2乗に比例する

$$F(t, d) = 1 - \exp(-\lambda d t^3) \quad (3)$$



$$F(t, d) = 1 - \exp(-\lambda \cdot (d^B t)^3) \quad (6)$$

$$h(t, d) = 3\lambda t^2 d^{3B} \quad (7)$$

このモデルはハザード関数は用量  $d^B$  に比例し、時間  $t$  の 2 乗に比例するモデルでパラメータが (3) 式と比べて 1 つ増えている。このモデルの下では

$$TD50 = (\log(2) / (\lambda \cdot t_0^3))^{1/3B} \quad (8)$$

となる。

やはり SAS の LIFEREG プロシジャを用いて  $\lambda$  と  $B$  と最尤法によって求める。

：大腸癌  $\lambda = 10^{-11} \times 6.136$   $B = 0.6401$   $TD50 = 121.364$  対数尤度 = -41.687

：乳癌  $\lambda = 10^{-10} \times 4.766$   $B = 0.5096$   $TD50 = 108.432$  対数尤度 = -48.832

対数尤度を (3) のモデルと比較すると、 $B$  をモデルに含めることによってモデルが有意に改善されることがわかる。また  $TD50$  は③によって求めた値にほぼ等しくなっている

#### 実測値と予測値

実験期間	用量	大腸癌		乳癌	
		実測値	予測値	実測値	予測値
52W	400ppm	0.5517	0.5751	0.4667	0.47146
52W	0ppm	0	0	0	0
104W	100ppm	0.4333	0.3800	0.4667	0.45805
104W	25ppm	0	0.0328	0.0667	0.07092
104W	0ppm	0	0	0.0000	0

実測値と予測値を比較した結果でも、モデルがよくあてはまっていることがわかる。

#### 5. 問題点

① すべての腫瘍について (3) 式  $F(t, d) = 1 - \exp(-\lambda d t^3)$  が成り立つ保証はない。実験期間と用量の組み合わせがもう少し多ければ

$$F(t, d) = 1 - \exp(-\lambda \cdot (d^{B1} t)^{B2}) \quad (9)$$

等のより複雑なモデルを仮定した方がよいかもしれない。ただしこのデータから  $B1$ 、 $B2$  を両方推定するには無理がある。

②  $TD50$  の信頼区間をどのようにして求めるか？

$LD50$  の信頼区間を求めるには、Fieller の公式を利用するのが一般的であるが、 $TD50$  の場合にもこの方式で信頼区間を求めればよいのか？

なお本発表で示した  $TD50$  の推定を SAS の LIFEREG プロシジャで実行する方法については、大橋 他に詳しく示されているので参照されたい。

F ( t , d ) : 累積分布関数

このモデルはワイブルモデルと呼ばれる。Portier 等はNTPの数千例の背景データに基づいて腫瘍の種類ごとにワイブル分布をあてはめ、ワイブル分布の適合度と形状母数の値を報告している。ワイブルモデルを仮定する背景には、このような研究の成果が利用されていると思われる。ちなみにプロビット法によるLD50の推定では、Fに正規分布の累積分布関数を仮定している。

(3) 式のモデルが正しいとするならば、最尤法によって5群全部の情報を使ってλを推定し、t<sub>0</sub>に固定してF ( t<sub>0</sub> , d ) = 0.50となるようなdを求めれば、これがTD50の推定値となる

$$F ( t_0 , d ) = 1 - \exp ( - \lambda \cdot TD50 \cdot t_0^3 ) = 0.5 \quad (4)$$

(4) 式をTD50について解くと

$$TD50 = \log(2) / ( \lambda \cdot t_0^3 ) \quad (5)$$

となる。以上よりデータからλを推定できればTD50を求めることができる。

ここで

$$t' = d t^3 \text{ とおくと}$$

$$F ( t , d ) = 1 - \exp ( - \lambda t' )$$

となり、この問題はt'について指数分布をあてはめる問題に他ならない。統計解析ソフトウェアSASのLIFEREGプロシジャを用いてλを最尤法によって推定すると

: 大腸癌 λ = 10<sup>-9</sup> × 6.3668      TD50 = 96.7835      対数尤度 = -50.131

: 乳癌 λ = 10<sup>-9</sup> × 6.5568      TD50 = 93.9789      対数尤度 = -52.004

となる。TD50は100ppmに近づいている。しかしこのモデルはあまりあてはまりがよくない。累積の腫瘍発生率の実測値とモデルから求めた予測値を示す

実測値と予測値		大腸癌		乳癌	
実験期間	用量	実測値	予測値	実測値	予測値
52W	400ppm	0.5517	0.3010	0.4667	0.30842
52W	0ppm	0	0	0	0
104W	100ppm	0.4333	0.5114	0.4667	0.52172
104W	25ppm	0	0.1639	0.0667	0.16839
104W	0ppm	0	0	0.0000	0

#### 4. モデルの改善

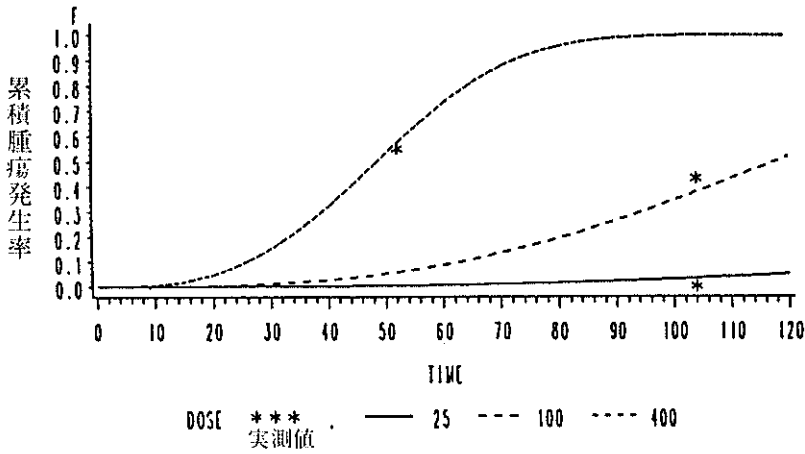
(3) 式のモデルはあてはまりがよくない、そこで改善したモデルを示す。

参考文献

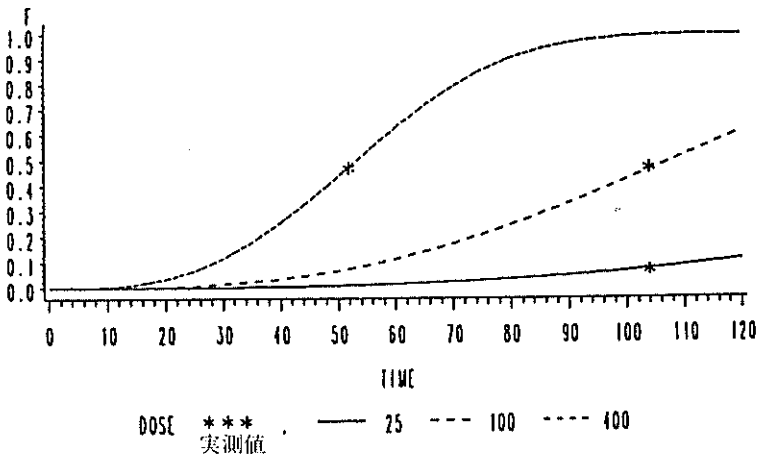
大橋 他(1993 刊行予定), SASによる生存時間解析, 東京大学出版会  
 C.J.Portier, J.C.Hedges, and D.G.Hoel (1986), Age-specific models of mortality and tumor onset for historical control animals in the national toxicology program's carcinogenicity experiments, CANCER RESEARCH 46, 4372-4378

改善したモデルのあてはめ

DISTRIBUTION FUNCTION  
 TUMOR=1 大腸ガン



DISTRIBUTION FUNCTION  
 TUMOR=2 乳ガン



### 3 BISO の質疑応答より

講師 佐藤俊哉 (統計数理研究所)

1  $2 \times 2$  の  $\chi^2$  検定および Wilcoxon 検定での連続修正を行うことについて、良いのか悪いのかの議論を整理して、現状でどちらの方が主流であるのか示してほしい。

この議論の本質は連続修正をする、しないではなく、conditional testかunconditional testかという点である (連続修正をするということはFisherの正確なp値を求めるということとほとんど同義である)。つまり、

conditional test = 連続修正は必要  
unconditional test = 連続修正は不必要

である。

ランダム化臨床試験の場合

表のようなランダム化臨床試験を考える。  
この100人の患者に対して、新治療は効果があるだろうか? 調べたいcausal parameterとして、population causal リスク差を考える。

	死亡	生存	標本数
新治療	8	42	50
標準治療	17	33	50
合計	25	75	100

population causal リスク差

= (100人の患者が全て新治療を受けた場合の死亡割合)

- (100人の患者が全て標準治療を受けた場合の死亡割合)

ここで、100人の有限母集団に対して推測を行うとする。causal 帰無仮説として「新治療は患者の死亡状況を変化させない」を考えると、この仮説が正しいときどちらの治療を割り付けられても患者の死亡状況は変化しないから、死亡した25人は必ず死亡する。つまり、

1. 有限母集団に対する推測
2. Randomization
3. causal 帰無仮説

という条件のもとではpermutation testとしてのFisherの正確な確率が正当化される。

次に上の100人の母集団の患者集団についてcausal inferenceをすることを考える。母集団でのcausal 帰無仮説を

$H_p$ : 全てのpopulation membersについて新治療でも標準治療でも死亡状況は変わらない

とする。この場合もやはり、ランダム化臨床試験を行って $H_p$ を調べようとするであろう。そのときも、死亡した25人を固定して考えるべきかどうか、100人を有限母集団として扱っていいかが問題となる。そこで、サンプルでのcausal 帰無仮説

$H_s$ : 全てのsample members について、新治療でも標準治療でも死亡状況は変わらない

を考えると、 $H_s$ はサンプルをとるまで定まらないが、一度サンプルが得られれば、 $H_s$ の検定は有限母集団の場合とまったく同じになる。 $H_p$ は $H_s$ を含むので、 $H_s$ が棄却されると $H_p$ も棄却される。しかし、 $H_s$ が棄却できなかったときは、 $H_s$ は $H_p$ を含まないので、「 $H_p$ を棄却できない」とはならない。かといって $H_p$ を棄却すると、サンプルでは観察されていない事実をもとにpopulationに対して推論を行っていることになる。したがって、 $H_p$ に関する推論は $H_s$ に基づいて行う。

#### 観察研究の場合

表のような結果が観察研究により得られたとすると、これは新治療を行った仮想的なsuper population、標準治療を行ったsuper populationからのランダムサンプルだと考えることができる。この場合は2つの独立な2項分布モデルを考えればよい。この場合

記述的帰無仮説: 2つのsuper populationsの死亡割合は等しい

となる。これは、unconditional testであるので、割合の差の検定、 $\chi^2$ 検定、などを行えばよい。

Greenland, S. (1991) On the logical justification of conditional tests for two-by-two contingency tables. The American Statistician 45, 248-251.

## 2 Fisherの直接確率計算は片側検定か?

片側検定であるとの説明があったように思われるが、両側とも計算、片側を計算して2倍するなどの方法がとられているように思われる。どの方法がよいのか?

表のような小さいランダム化臨床試験の結果が得られたとする。この場合、

	改善	変化無し	標本数
治療A	17	3	20
治療B	6	6	12
合計	23	9	32

causal 帰無仮説: 治療Aに変えて治療Bを行っても改善する割合は変わらない

これに対しFisherの正確な確率を計算すると、すべての周辺度数を固定した上で、治療Aで改善した患者の数が取り得るのは、11~20までであるから

x	超幾何確率	
11	0.0060	下側確率=0.0060
12	0.0539	
13	0.1824	
14	0.3040	
15	0.2736	
16	0.1368	
17	0.0376	
18	0.0054	上側確率=0.0433
19	0.0004	
20	0.0000	

となる。

今、両側5%検定を、1. 両側とも計算、2. 片側2.5%、で行うと、

1. 両側確率  $0.0060 + 0.0433 = 0.0493 < 0.05$  であるから、ギリギリではあるが有意であり、治療Aの効果あり

2. 上側確率  $0.0433$  は  $0.025$  に比べて結構大きいので、有意とはいえない

となるが、全ての統計的推測は、観察されたデータとconsistentでなければならない。1のようにして治療Aを効果ありとするのは、観察されていないデータに基づいた判断である。したがって、自由度1の両側検定は、有意水準を半分にした片側検定を2回行うべきである。

Cox, D.R. and Hinkley, D.V. (1974). Theoretical Statistics. pp.105-106. London: Chapman and Hall.

---

(佐藤俊哉先生ならびに日本科学技術連盟の許可を得て、92年3月26日に行われた3B10Sでの質疑の講義記録を転載いたしました。)

# 薬物の複合投与効果の考え方

## — マウス小核試験における理論と実際 —

鈴木 孝 昌 (国立衛生試験所・変異遺伝部)

これまでに数多くの化学物質について変異原性試験がなされてきたが、それらほとんどの場合においては、被験物質単独での効果についてのみ検討されている。しかし我々を取りまく環境中には数多くの化学物質が混然一体となって存在しており、実際にはこれら複数の物質に同時に暴露されているわけである。医薬品に目を向けてみても、複数の薬物を同時に服用するということは一般的に行われていることである。単独では現れなかった作用が複合投与時にはじめて現れることも起こりうるため、より正しい評価を下すためには、複数の物質に対して同時に暴露されたときの効果を調べるのが重要となってくる。変異原性試験の分野においても、将来的にこういった複合投与効果について検討を行うことが重要な課題となってくるだろう。

そして最終的には数多くの化学物質について複合効果を検討しなければならないが、そのための第一歩として、今回はまず2つの薬物の併用時の効果についてマウス小核試験を用いて検討し、複合投与効果に関する基礎的な考え方について検討を行った。

ここで、今回用いた小核試験について少し説明しておく。マウスを用いる小核試験は現在 *in vivo* での変異原性試験として最も汎用されており、被験物質投与により骨髓細胞中に生じる“小核”を指標としている。図1に示したように被験物質によるDNA傷害によって前赤芽球中に生じた染色体異常に起因する染色体断片は、細胞分裂の際に細胞質に取り残され主核とは別に小核と呼ばれる小さな核を

末梢血を用いる小核試験法

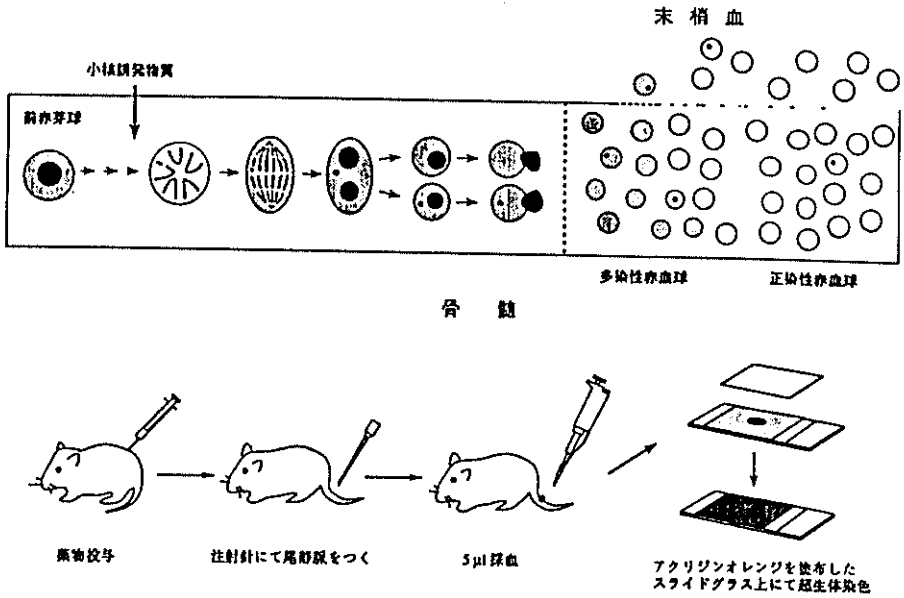
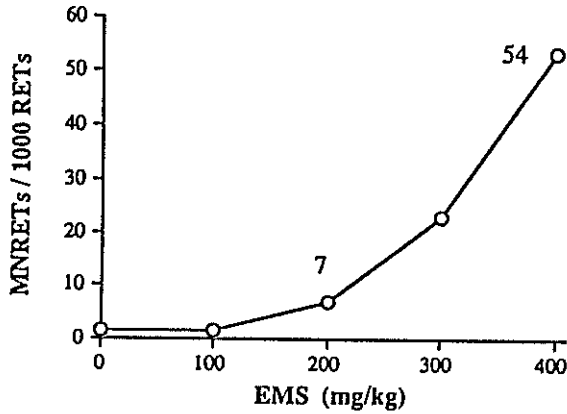


図 1

Micronucleus Induction by Ethyl methanesulfonate



MNx: 被験物質の投与量Xによって誘発されるMNRETs

EMS処理の場合、グラフより MN200 = 7, MN400 = 54

$$\text{MN400} = \text{MN200} + 200 \neq \text{MN200} + \text{MN200}$$

(54) (14)

Dose-response が直線的ではない場合、このように単純な加算は成立しない。

図 2



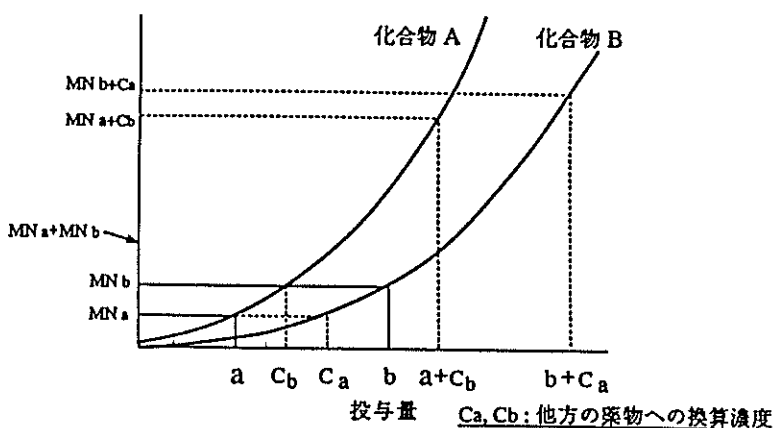
生じる。やがて赤血球の成熟の最終段階で脱核が起きる際に小核は取り残されて本来無核の赤血球中に小さな核を持つ血球が出現する。末梢血を観察して、小核を有する幼若な赤血球 (MNRET:micronucleated reticulocyte)の出現頻度を求めることにより被験物質の変異原性を評価する。

図2に代表的な変異原物質であるアルキル化剤のethyl methanesulfonate(EMS)の小核試験の結果を示した。縦軸はEMSによって誘発された小核を有する幼若赤血球”MNRET”数を1000個の幼若赤血球当たりの数として示した。これは単独物質による結果であるが、ここで仮にEMSという同じ薬物を併用した場合を考えてみる。例えばEMS 400mg/kg投与をEMS 200 mg/kgとEMS 200 mg/kgの併用と考える。この場合EMS 200mg/kg単独による小核誘発量(MN<sub>200</sub>)はグラフより7個であり、併用時に単純な加算効果として期待されるMNRET数は7+7で14個である。これは実測値すなわちEMS 400mg/kgによって誘発されるMNRET数(MN<sub>400</sub>)54個とは大きく異なっている。これはEMSによる小核誘発のdose-responseが直線的でないことより当然の結果であり、このような場合には上記のような単純な足し算は成立しない。

この場合は同じ薬物の併用であるため明かなことであるが、同じようなことが異なる薬物の併用時、特に類似した作用を示す薬物どうしの併用時にも考えられるのではないか。

我々は、バクテリアを用いた復帰突然変異試験において提唱された川添ら\*の考え方を小核試験へも取り入れ、図3に示した様な小核誘発のdose-responseをもつAとBという2つの薬物について、それらを複合投与したときの効果について以下のように考えてみた。まず、薬物A,Bをそれぞれ投与量a,bで併用した場合の小核の誘発量をMN<sub>a+b</sub>とし、それぞれ単独での小核誘発量をMN<sub>a</sub>,MN<sub>b</sub>とする。もし薬物AとBの作用が全く独立していれば併用時の効果は単純な加算値となると考えられ、 $MN_{a+b}=MN_a+MN_b$ となる。

” 2つの化合物の併用により期待される小核の誘発量”



2つの化合物を併用したときの効果(MN a+b)は化合物の作用機序より以下のように分類される

- ・ Independent :  $MN a+b = MN a + MN b$
- ・ Equivalent :  $MN a+b = p(MN a+Cb) + q(MN b+Ca)$  ( $p+q=1$ )
- ・ Synergistic :  $MN a+b > MN equivalent$
- ・ Antagonistic :  $MN a+b < MN independent$

\* (Kawazoe et al, Chem. Pharm. Bull., 34, 250-255 より改編)

図 3

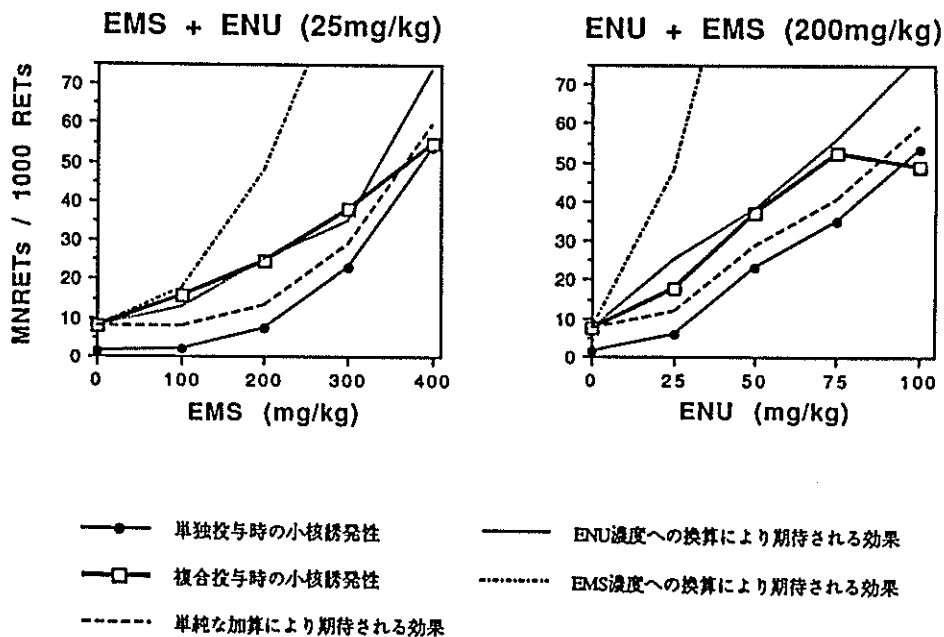


図 4

次に、併用する薬物の作用が小核誘発において同等である場合には、それぞれの薬物の投与量をひとつの薬物の投与量へと変換して複合効果を考えてみる。まずBの薬物の投与量bと同じ小核誘発量(MN<sub>b</sub>)を与える薬物Aの投与量はグラフよりC<sub>b</sub>と求まり、薬物Aへと換算した併用時の投与量としてa+C<sub>b</sub>が得られる。このとき期待される小核誘発量MN<sub>a+c<sub>b</sub></sub>は単純な加算値(MN<sub>a</sub>+MN<sub>b</sub>)よりも大きな値を与える。これは見かけ上相乗効果に見えるが、このような変換を行ったときには加算効果として説明される。同様にして薬物Bへと換算した時の複合効果は、薬物Aの薬物Bへの換算投与量C<sub>a</sub>を使ってMN<sub>C<sub>a</sub>+b</sub>として求められる。これらより、作用の同等な2つの薬物による複合投与効果はMN<sub>a+c<sub>b</sub></sub>とMN<sub>C<sub>a</sub>+b</sub>との間の値として得られることが期待され、これを式で表せば、MN<sub>a+b</sub>=p·MN<sub>a+c<sub>b</sub></sub>+(1-p)·MN<sub>C<sub>a</sub>+b</sub>となる。この値よりも併用時の効果が大きい場合には真の相乗効果(synergism)と考え、単純な加算値よりも小さい場合は拮抗効果(antagonism)とする。

この考え方をもとに、まず小核試験におけるアルキル化剤の複合投与効果について解析を行った。用いたのはethyl nitrosourea(ENU), ethyl methanesulfonate(EMS), ethyl ethanesulfonate(EES)の3種のエチル化剤であり、このうち2種の薬物の併用効果について検討を行った。小核試験にはddYマウス雄8週齢を1群あたり4匹用い、被験物質を単独および2種同時に投与した。そして48時間後に末梢血中の小核を有する網赤血球(MNRET)の出現を調べた。末梢血はマウス尾部より5μl採取し、acridine orange色素を塗布したスライドグラス上にて超生体染色後、蛍光顕微鏡で1000個の網赤血球(reticulocyte)あたりに出現するMNRET数をカウントした。

図4にENUとEMSとの組み合わせによる結果を示した。グラフにはそれぞれの薬物単独でのdose-response(—)および一定量他薬物を併用したときの複合投与効果(—)の実験値とともに、単純な加算により期待される効果(---)および、作用

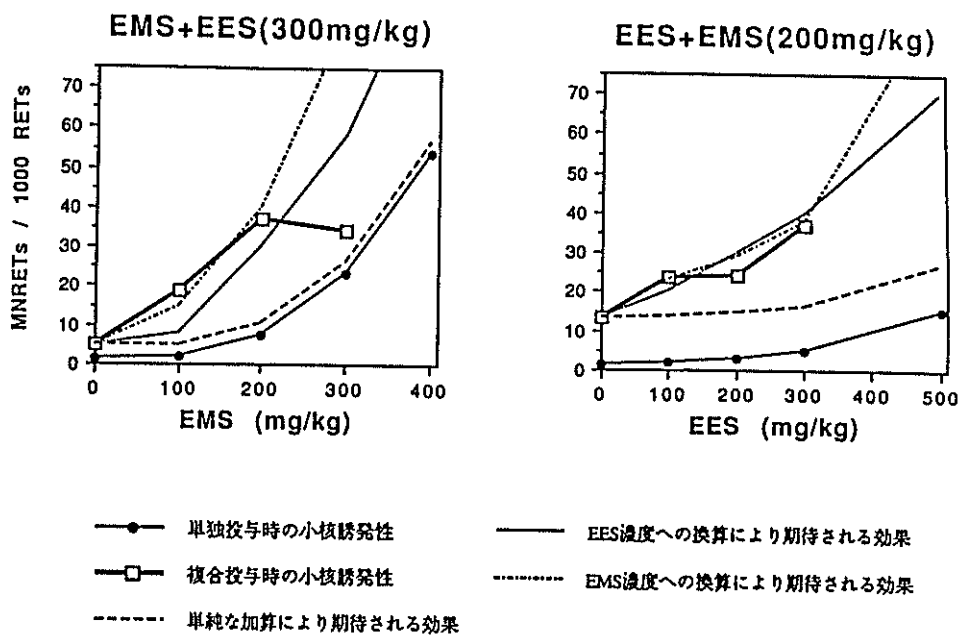


図 5

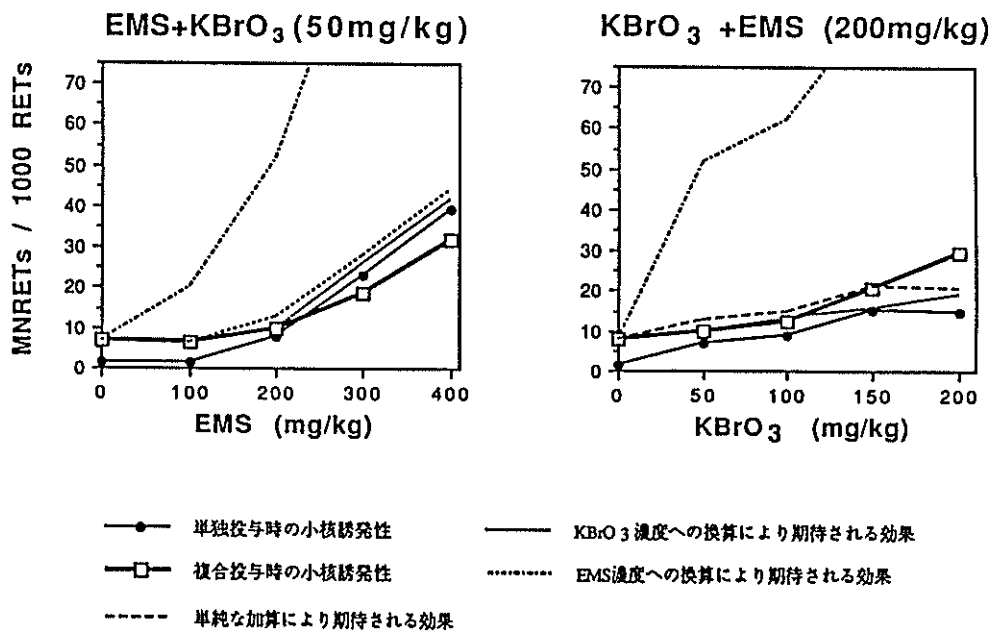


図 6

を同等と考えて図3のモデルのように片方の薬物投与量に変換して計算したときに期待される効果(—および----)を示した。どちらの組み合わせにおいても、併用時において単純な加算効果を上回る結果が得られた。これはENUへ換算したときに期待される効果に近い値となったが、EMSへと変換した値よりは低かった。

つぎにEMSと、同じスルホン酸エステルであるEESとの組み合わせによる結果を図5に示した。この場合にも単純な加算値を上回る見かけの相乗効果が得られたが、これは各薬物へ変換して考えたときに期待される効果と一致した(高用量で併用時に効果が落ちているのは細胞毒性のためと考えられる。このため一部ではデータが得られなかった。)

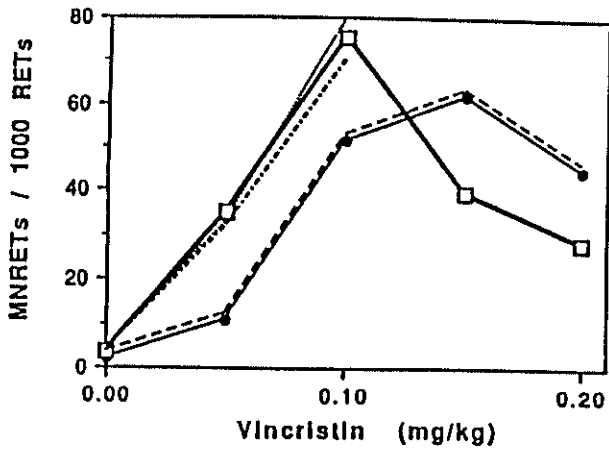
以上の結果より、アルキル化剤どうしの複合投与により得られた相乗効果は、“みかけ”上のものであり、図3に示したモデルに基づいて解析した場合には加算効果として説明できることが明かとなった。また、EMSとENUの組み合わせによる複合効果はEMSとEESによる複合効果に比べて小さかったが、これは投与量換算によって期待される複合効果が前二者において差があったことも含めて、ENUとEMSの作用が全く同等ではないことに起因するものと推測される。事実ENUとEMSは同じエチル化剤であるが、その化学的反応性はやや異なっている。

そこで次にアルキル化とは全く違った作用を持つ薬物として、活性酸素の関与により染色体異常を誘発すると考えられる臭素酸カリウム( $\text{KBrO}_3$ )を選び、EMSとの複合効果を検討した。その結果を図6に示す。

EMSと $\text{KBrO}_3$ の複合投与においてはこれまでとは違って、単純な加算効果と同じかむしろそれをやや下回る結果となってあらわれた。この結果は、両者の作用の違いを反映するものと考えられる。

さらに、同等の小核誘発機序性を持つと考えられる紡錘体阻害剤であるvincristinとcolchicineを用いて検討を行った。その結果図7のように毒性のあらわれないうち低濃度領域では複合投与時に“みかけ”の相乗効果が得られ、これは投与量換

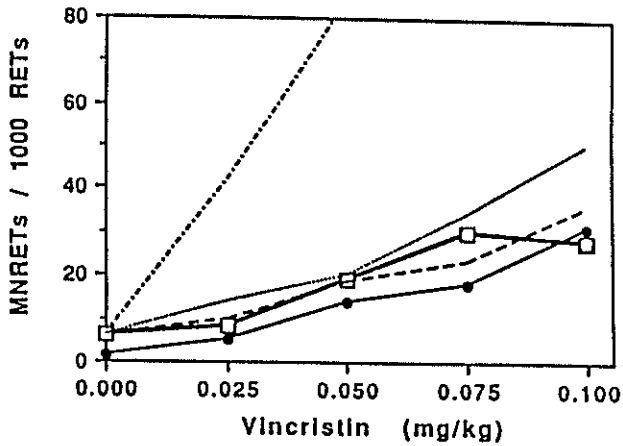
Vincristin + Colchicine (0.5mg/kg)



— Colchicine濃度への換算により期待される効果  
 ..... Vincristin濃度への換算により期待される効果

図 7

Vincristin + EMS (200mg/kg)



— Vincristin濃度への換算により期待される効果  
 ..... EMS濃度への換算により期待される効果

図 8

算により得られる加算効果に一致した。

また図8に示すようにvincristinとEMSという作用機序の異なる薬物どうしの組み合わせでは、単純な加算効果が得られた。

以上、マウス小核試験における薬物の複合投与効果について基礎的な検討を行ったが、作用機序が同じという前提のもとに、二つの薬物の投与量を一つの薬物の投与量として変換した値に基づく解析を行ったところ、見かけ上得られた相乗効果が実は加算効果として説明できることが明かとなった。一般に発がん性試験などの動物実験ではきちんとしたdose-responseを求められないことも多く、このような場合に単純な加算値よりも高い複合効果が得られれば単純に相乗効果が得られたと考えられていると思われるが、今回の結果は、その相乗効果が、それぞれの薬物のdose-responseを考慮した時、実は単純な加算効果として予想され得ることを示唆している。また視点を変えると、指数関数的なdose-responseを示すような作用の類似した化合物どうしの複合投与により、それぞれ単独の効果からは予想できないような強い複合効果があらわれることが予想され、このような場合には特に注意を要するといえよう。

今後さらにいくつかの薬物に関して小核試験を用いた同様の検討を進め、今回用いたモデルに基づく複合効果の解析を行ってゆく予定である。また、こういった検討は小核試験にとどまらず他の変異原性試験や毒性試験などへも適応可能であり、そういった試験系においても今回のような複合効果の解析が応用されていくことを期待したい。

---

第53回定例会（93年1月23日）講演より

# 毒性試験における多重比較法

— L.Hothorn の研究の解説と検討 —

松田 眞一 (南山大学)

滝沢 毅 (日本ロシュ)

1991年5月23日～25日、東大山上会館で行なわれた「毒性試験に関する国際計量生物会議」で、L.Hothorn が「毒性試験における多重比較法」と題して彼の研究を発表した。この問題は毒性試験関係者にとって最大関心事の一つであり、ここに解説と検討を行なう。

## 1、Hommel / Hochberg の手順

L.Hothorn は、毒性反応は必ずしも単調な用量反応を前提としていないので最もすすめられるとしている。

[手順]

- 1) 対照群と各用量群との2群比較を行なう。
- 2) 最も大きな p 値を有意水準  $\alpha$  と比較する。  
小さければ各用量群はいずれも有意水準  $\alpha$  で有意と判定する。  
大きければその p 値に対応する用量群は有意でないとし次の比較に移る。
- 3) 次に最も大きな p 値を有意水準  $\alpha / 2$  と比較する。  
小さければその p 値以下の用量群は有意水準  $\alpha$  で有意と判定する。  
大きければその p 値に対応する用量群は有意でないとし次の比較に移る。
- 4) 次に大きな p 値を有意水準  $\alpha / 3$  と比較する。  
小さければその p 値以下の用量群は有意水準  $\alpha$  で有意と判定する。  
大きければその p 値に対応する用量群は有意でないとし次の比較に移る。
- 5) 最も小さな p 値まで上記の手順を繰り返す。  
用量群の数を  $k$  としたとき 最も小さな p 値は  $\alpha / k$  と比較する。  
小さければその用量群だけ有意水準  $\alpha$  で有意とする。  
大きければその用量群も有意でないとする。



### [解説]

H/H 手順は Simes (1986) の手順を閉手順に拡張したもので、Simes の手順が有意水準を保つのなら問題がないことになる。しかし、Simes の手順が有意水準を保つことは一般的には証明されていない。ただし、各検定が独立の場合は証明があり、多変量正規分布の場合はシミュレーションによって保つことが分かっている。

## 2、CTP-T 手順

Closed Testing Procedure (閉手順) で t 検定を使う。L.Hothorn は、単調な用量反応を前提とできるとき、最も検出力が高いことを示した。

### [手順]

- 1) 対照群と各用量群との 2 群比較を行なう。
- 2) 最高用量群の p 値と有意水準  $\alpha$  の比較をする。  
小さければ、最高用量群は有意水準  $\alpha$  で有意と判定し次のステップに進む。  
大きければ、各用量群はいずれも有意でないと判定する。
- 3) 次の高用量群の p 値と有意水準  $\alpha$  を比較する。  
小さければ、その用量群も有意水準  $\alpha$  で有意と判定し次のステップに進む。  
大きければ、最高用量のみ有意で他は有意でないと判定する。
- 4) 最低用量群まで上記の手順を繰り返す。

### [解説]

CTP-T 手順は用量反応の単調性を前提としているものの、それを用いるのは検定の順序に関してのみである。順序付けて検定を行なうわけだから、各々の検定が有意水準を保つ限り全体としての有意水準も保たれる。

## 3、群の数の影響

毒性試験では、低、中、高用量と 3 用量設定することが多いが、4 用量、5 用量になることもある。これは低用量群に対しては有意差がなく、高用量群では有意差があるという結果にならないと、無影響量が求められないため、用量設計を慎重に行なうためである。

このとき 3 用量でも 4 用量でも各用量群と対照群の比較が同じ結果となるなら、問題ないが、L.Hothorn が示したごとく、違ってくる。

次の例で示すように3用量では有意差のある1、000mg/kg群が、全く同じ値であるにもかかわらず4用量では有意でなくなる。

	Dose 1	Dose 2	Dose 3	
Control	20 mg/kg	200 mg/kg	1,000 mg/kg	
2.087	2.064	2.151	2.649	$x_j$
.569	.304	.722	.831	$s_j$
	-.106	2.048	2.499	$s_j$
	2.14	2.14	2.14	$d_{k=3}$
	-	-	>	result

	low-dose	Dose 1	Dose 2	Dose 3	
Control	10 mg/kg	20 mg/kg	200 mg/kg	1,000 mg/kg	
2.087	2.072	2.064	2.515	2.649	$x_j$
.569	.491	.304	.722	.831	$s_j$
	-.071	-.106	2.048	2.499	$d_j$
	2.55	2.55	2.55	2.55	$d_{k=4}$
	-	-	-	-	result

[検討]

実験を計画する上でそれぞれの検定方法が群の数の多少や用量の設定にどのように影響されるかは知っておいてもよいと考えられる。

まず、群の数の多少については、以下のように考えられる。

少ない：個々の群の影響が大きくなる。

全体的に傾向は掴みにくくなる

多い：個々の群の影響は小さくなる。

用量反応関係が明確になり、用量の設定ミスに強い。

そのため、自ずと実験に応じた群の数というのが、決定されるはずである。

次に用量群の設定において高用量と低用量のどちらが手順に影響を与えるかということであるが、以下のように考えられる。

- (a) Dunnett の手順： 元々個々の群の間に順序を仮定しないのだからそういうことを考えることに意味がない。すなわち単調な用量反応関係があると推測されるときにこの方法を用いること自体に問題がある。
- (b) H/H 手順： p 値の大きいものから検出するので低用量群のようが影響が大きく、高用量群の影響は小さい。（Holm の手順では逆になる。）
- (c) CTP-T 手順：高用量群から検定するためその影響を受ける。低用量群の影響は受けない。
- (d) ウィリアムズの手順：どちらの群も影響があり、不利か有利かは自明ではない。

これらの結果を踏まえて事前に計画を練ることはいいのだが、後知恵に用いることがあってはならない。

#### 4、CTP（閉手順）と無影響量

「毒性試験の統計解析」p.110によれば「閉手順はどのような検定手法を使っても、一般に無影響量を過大評価する傾向がある」が、毒性試験では閉手順は適切ではないのだろうか？

##### [検討]

この記述では何に対して過大なのか 述べられていない。

まず、無影響量とは何かを考える。この本の前の方の記述にもあるようにガイドラインでは影響が認められない用量ということになっている。この曖昧な概念に基づく議論をするためによけいに話が難しくなっている。

例えば、「用量が0であってもある定常的な反応があり、影響とはある 'いき値' を超えた場合にそれより反応が大きくなることだ」と考えると、ホッケースティック型の折れ線の当てはめを考えていることになり、折れ曲がるところが無影響量となる。

そうすると離散的な用量に基づく検定を用いる限り、それが閉手順であるかどうかにかかわらず、そういった回帰による推定結果より過大評価となることが容易に分かる。

ただし、検定手法間の比較では閉手順だからといって過大評価ということにはならない。ある手法を閉手順に拡張した場合には明らかに検出力が向上するのだから逆に閉手順にすべきである。

一方、順序を考慮しないダネットの手順とウィリアムズの手順を比較するなら確かに過大評価になることがあり得るが、中、高用量で影響なしとされたものがあつた場合どうすればいいのかといった問題から逃れられなくなる。

## 5、リコメンデーション

「毒性試験データの統計解析」 p.109 では、単調な用量反応を前提とできるならウィリアムズの多重比較をすすめており、そうでないときはダネットをすすめているが、L.Hothornの研究によれば単調な用量反応を前提とできるなら CTP-T 手順、そうでないなら H/H 手順がよいとしている。

### [検討]

まず、単調な用量反応関係がある場合で考える。

ウィリアムズの手順は平均の補正を行なっているので、高用量群にとっては群の数が適当に多ければ偶然反応が低くて検出されないということを妨げる。一方、CTP-T 手順はそのような偶然性を排除できない。ただし、検出力の面で本当に有利かどうかは自明ではない。

次に、単調な用量反応関係がない場合について考える。

通常のダネットの手順と H/H 手順とを比較すると、最初の対の棄却に対してはダネットの手順の方が優れていてもっと小さな差に対しては H/H 手順が優れていると思われる。（H/H 手順はボンフェローニの方法の拡張なのだから大きな差に対してはそれと変わらない。ただ、閉手順としての拡張であるので小さな差には強い。）そういうことを念頭において（後知恵ではなく）使い分けることもできるが、ダネットの手順も閉手順に拡張すると小さな差に対しても検出力が上がり一番よいと思われる。しかし、手作業での手順は複雑である。

## 第56回定例会出席者名簿

日時：1993年10月30日（土）

11：00～12：00 基礎講座

場所：総評会館

（途中昼食）

13：00～17：00 定例会

### ご出席頂いた先生方

- |                  |                |
|------------------|----------------|
| *岩崎 学（成蹊大学）      | *橋本修二（国立公衆衛生院） |
| *佐久間 昭（東京医科歯科大学） | *浜田 知久馬（東京大学）  |
| *菅野 純（東京医科歯科大学）  | *松田 眞一（南山大学）   |
| *芳賀 敏郎（東京理科大学）   | *吉村 功（東京理科大学）  |

- |                           |                            |
|---------------------------|----------------------------|
| 1 北村 知宏（山之内製薬）            | 26 酒井 弘憲（三菱化成㈱）            |
| 2 相馬 義徳（日本ロシュ㈱）           | 27 大石 正平（科研製薬㈱）            |
| 3 清水 辰己（キリンビール㈱）          | 28 入野田 一彦（カビファルマシア㈱）       |
| 4 本川 裕（キリンビール㈱）           | 29 松本 道治（日本バイオアッセイ研究センター）  |
| 5 小原 直樹（大塚製薬）             | 30 奥田 裕計（日本バイオアッセイ研究センター）  |
| 6 矢田 武（協和発酵工業㈱）           | 31 加納 浩和（日本バイオアッセイ研究センター）  |
| 7 旦 喜夫（山之内製薬）             | 32 石川 寛明（日本バイオアッセイ研究センター）  |
| 8 鈴木 睦（キリンビール㈱）           | 33 池田 高志（帝国臓器製薬㈱）          |
| 9 村野 弘行（持田製薬㈱）            | 34 鈴木 孝昌（国立衛生試験所）          |
| 10 矢嶋 勉（持田製薬㈱）            | 35 西田 隆一（堺化学工業）            |
| 11 大橋 信之（(財)安評センター）       | 36 森 幸生（日本商事㈱医薬研究所）        |
| 12 岸林 伸行（協和発酵工業㈱）         | 37 豊島 恵一（ヤンセン協和㈱）          |
| 13 田村 幸資（雪印乳業㈱）           | 38 古川 敏仁（㈱エスアールエル）         |
| 14 日比 茂晴（雪印乳業㈱）           | 39 楠 正隆（㈱ミヤリサン生物医学研究所）     |
| 15 安藤 正一（山之内製薬㈱）          | 40 原 英彰（鐘紡㈱）               |
| 16 島村 義一（科研製薬㈱）           | 41 吉池 通晴（第一製薬㈱）            |
| 17 温井 一彦（住友製薬）            | 42 石原 信夫（東北労災病院）           |
| 18 西村 和夫（住友製薬）            | 43 長谷川 良平（名古屋市立大学医学部）      |
| 19 鈴木 隆一郎（大阪府立成人病センター）    | 44 田中 宏太佳（東京労災病院）          |
| 20 布能 英一郎（関東学院大学）         | 45 惣田 隆生（シオノギ製薬）           |
| 21 宮井 一浩（エスエス製薬㈱統計解析セクター） | 46 松林 哲夫（昭和薬科大学）           |
| 22 長岡 克彦（エスエス製薬㈱）         | 47 今井 節夫（安全性試験研究センター）      |
| 23 小林 和子（日本ロシュ）           | 48 麻坂 美智子（プリストル・マイーズ・スクイフ） |
| 24 滝沢 毅（日本ロシュ）            | 49 花岡 裕吉（わかもと製薬）           |
| 25 上原 秀昭（㈱ツムラ）            | 50 東島 秀樹（ロート製薬㈱）           |

- 51 神谷隆信(株薬効開発研究会)
- 52 田村博信(日本新薬)
- 53 寺尾哲(武田薬品)
- 54 山岡秀明(住友化学工業)
- 55 平河威(三井製薬工業)
- 56 稲葉太一(神戸大学)
- 57 西尾伸太郎(東レ株)
- 58 佐々木貴(ファルミタリア カルロエルバ株)
- 59 青柳誠(ファルミタリア カルロエルバ株)
- 60 田村友子(ファルミタリア カルロエルバ株)
- 61 山口達哉(東洋紡)
- 62 佐藤昇(ファイザー製薬)
- 63 中辻堅造(明治製菓株)
- 64 松本正人(明治製菓株)
- 65 鷹見勲(サンド薬品株)
- 66 光森達博(鐘紡株)
- 67 杉山公仁(昭和薬品化工株)
- 68 光丘元弘(アプゾソファーマシューティカルズリミテッド)
- 69 古賀正(アプゾソファーマシューティカルズリミテッド)
- 70 瀬川光久(スリーエム薬品株)
- 71 添田幸広(株サン・メディア)
- 72 小崎章夫(サントリー株)
- 73 日俣克一(山崎製パン株)
- 74 八木(キリンビール株)
- 75 長谷川要(日本レダグリー株)
- 76 正木文夫(株富士生物科学研究所)
- 77 深沢正樹(キッセイコムテック株)
- 78 伊藤博之(ノボノルディスクファーマ株)
- 79 内田浩伸(日本シェリング株)
- 80 鈴木富佑加(鈴木病院)
- 81 桑山典之(帝国臓器製薬株)
- 82 三橋弘明(帝人株)
- 83 清水明文(株科学飼料研究所)
- 84 中谷知弘(協和発酵工業株)
- 85 松永洋二(鳥居薬品株)
- 86 渡部則充(三井製薬工業)
- 87 荒川和仁(ヘキストジャパンK.K.)
- 88 加藤雅俊(サトウ製薬株)
- 89 鶴亀隆之(鐘紡株)
- 90 山田雅之(富士レビオ株)
- 91 坂本比呂志(吉富製薬)
- 92 松本一彦(日本たばこ産業株)
- 93 高橋行雄(日本ロシュ株)
- 94 下井信夫(株ユックムス)
- 95 半田淳(日本化薬株)
- 96 長谷文雄(日本ルセル株)
- 97 今溝裕(旭化成工業株)
- 98 内村直之(朝日新聞)
- 99 大塚芳正(日本ロシュ株)
- 100 佐々木秀雄(旭化成工業株)
- 101 工藤善隆(杏林製薬)
- 102 清水達也(株イナリサーチ)
- 103 林悦生(雪印乳業株)
- 104 山下和昭(森下ルセル株)
- 105 宮里盛幸(日本ルセル株)
- 106 杉浦直美(グレラン製薬株)
- 107 小沢義人(日本化薬株)
- 108 金子英志(帝人株)
- 109 菊地圭一(フリスト・マイヤーズ・スクイブ株)
- 110 森川敏彦(武田薬品工業株)
- 111 松井和浩(萬有製薬株)
- 112 神谷達郎(株保健科学研究会)
- 113 山田英樹(興和)
- 114 阿部いくみ(三菱化成総合研究所)
- 115 吉田和彦(ヘキストジャパン株)
- 116 戎五郎(株大塚製薬工場)
- 117 阿上(前臨床研究所)
- 118 干場敏男(生化学工業株)
- 119 川島英敏(エーザイ株)
- 120 渡辺理砂(旭化成工業株)
- 121 加藤義則(エーザイ株)
- 122 奥喜正(流通経済大学)
- 123 佐久間昭(東医歯大難研)
- 124 大島猛(日本化薬)
- 125 椿広計(慶応義塾大学)

〔事務局だより〕

10月の定例会では、当研究会としては初めての（恐らくは他の学会でも例のない）試みとしてディベートを行いました。会場の雰囲気からも、当日配ったジャッジシートに寄せられた感想・意見からも大好評のうちに実行できたと思います。一見奇抜なアイデアも積極的に実現してしまうのが当会のいいところです。合宿を含めた準備を重ね、各ディベーターも大いに個性を発揮して、本格的なサイエンス・ディベートを誕生させたのではないかと考えています。その結果、今まで当会で扱ってきた基本的なテーマに新しい光を当てることが出来た事、ゲーム性を前提にすることで対立点を明確にした論争を組めたことなど、新しい成果を当研究会にもたらしてくれたのではないかと思います。今後も定期的にこのディベートを実行して、基本的なテーマを掘り下げていく予定です。

ともあれ、何とか年内に38号を発刊出来て、ほっとしております。94年の会員各位の御活躍をお祈り申し上げます。(M.O.)

## 医薬安全性研究会 会報NO. 38

*Bulletin of Japanese Society for Biopharmaceutical Statistics.*

1993年12月31日 発行

定価1,030円（本体1,000円）

編集・発行 (株)サイエンティスト社

〒101 東京都千代田区神田駿河台3-2 山崎ビル

TEL. 03 (3253) 8992

FAX. 03 (3255) 6847

振替 東京8-71335

印刷・製本 (有)ナガノ印刷