European Federation of Statisticians in the Pharmaceutical Industry

President:
Paul Koopman
Kendle International Inc., Biometrics
Einsteindreef 123-131
NL-3562 GB Utrecht, The Netherlands
Tel: +31 30 258 4632
Fax: +31 30 258 4660
koopman.paul@kendle.com

Vice-President:
Bernhard Huitfeldt
AstraZeneca R&D
Global Clinical Science
S-151 85 Södertälje, Sweden
Tel: +46 8 553 27 385
Fax: +46 8 553 28 884
bernhard.huitfeldt@astrazeneca.com

Executive Office:
Jayne Fox
Resources for Business
3 St. James Close
Stretham, Cambridge, CB6 3ND, UK
Tel: +44 1353 649103
Fax: +44 1353 649103
efspi@resources.demon.co.uk

3 May 2000

Dear Paul

Enclosed please find a list of the publications in the DIA journal that refers to the Bruges meeting.

For the archive I have enclosed a copy of the papers as well.

Best regards,

Merete Jørgensen

List of literature

Presentations from DIA Workshop:
Statistical Methodology in Non-Clinical and Toxicological Studies,
March 25-27, 1996, Bruges, Belgium.

Hothorn, Ludwig A.
Guest Editor's Note: Biostatistics in Preclinical Studies.
Drug Information Journal, Vol. 31, pp. 321-322, 1997.

Hauschke, Dieter, Makoto Hayashi, Karl K. Lin, et.al.
Recommendations for Biostatistics of Mutagenicity Studies.
Drug Information Journal, Vol. 31, pp. 323-326, 1997.

Hauschke, Dieter.
Statistical Proof of Safety in Toxicological Studies.
Drug Information Journal, Vol. 31, pp. 357-361,1997.

Neuhäuser, Markus, Ludwig A. Hothorn.
The Control of the Consumer Risk in the Ames Assay.
Drug Information Journal, Vol. 31, pp. 363-367, 1997.

Hothorn, Ludwig A.
Modifications of the Closure Principle for Analyzing Toxicological Studies.
Drug Information Journal, Vol. 31, pp. 403-412,1997.

Kropf, Siegfried, Ludwig A. Hothorn, Jürgen Läuter.
Multivariate many-to-one Procedures with Applications to Preclinical Trials.
Drug Information Journal, Vol. 31, pp. 433-447, 1997.

Neuhäuser, Markus, Ludwig A. Hothorn.
Trends tests for Dichotomous Endpoints with Application to Carcinogenicity Studies.
Drug Information Journal, Vol. 31, pp. 463-469,1997.

Hothorn, Ludwig A., Ullrich Martin.
Application of Adaptive Interim Analysis in Pharmacology.
Drug Information Journal, Vol. 31, pp. 615-619, 1997.

Fairweather, William et al.
Biostatistical Methodology in Carcinogenicity Studies.
Drug Information Journal, Vol. 32, pp. 401-421, 1998

Front pages of all references papers are scanned in the remainder

Meeting
)PMENT:

ınada

a
or, Department
France
'hRMA, USA

area will provide
: are designed to
and Registration
ıeons, and social

SESSION
ɔ exhibit scientif-
istration Process.
Hall.

ing has been ten-

ıoon, June 26.

:erest and knowl-
liminary program

ı offers attendees
dessert reception,

Drug Information Journal, Vol. 31, pp. 321–322, 1997
Printed in the USA.

0092-8615/97

GUEST EDITOR'S NOTE: BIOSTATISTICS IN PRECLINICAL STUDIES

LUDWIG A. HOTHORN, PHD
University of Hannover, Hannover, Germany

PRECLINICAL studies are not as clearly defined as clinical studies with their Phase I-Phase IV classification. Moreover, the statistical methodology varies more in preclinical studies depending on the rather different nature of the studies. This may be one reason for the underrepresentation of statisticians working in this field as well as of the number of papers published on this topic, again, in comparison with clinical trials. Pharmacology, toxicology, substance screening, bioassay, stability testing, and so forth, however, play an important role in the drug development process.

It is difficult to find statistical publications in the nonclinical area, because they are widespread in biostatistical papers, for example, in Biometrics (1), in toxicological papers, for example, in Mutation Research (2), and in pharmacological papers, for example, in Regulatory Toxicology and Pharmacology (3). Several books were published in the last seven years, for example, Kirkland (4), Hothorn (5), Krewski and Franklin (6), Vollmar (7), and Morgan (8).

In toxicology, for example, two objectives of biostatistical work can be distinguished: academic—to find the most sophisticated approach—and practical—directed by the so-called "regulatory toxicology," that is, the toxicological studies defined by national and international guidelines or recommendations in the industrial drug development process. These studies, performed in either pharmaceutical companies or in contract research organizations, are routine in character. If, for example, hundreds of Ames assays a year are analyzed in a laboratory, using very different substances with different mechanisms in the same assay, this is done due to the need for cost minimization. Therefore, statistical approaches should be as robust as possible against real data configurations, as simple as possible for physicians to interpret, as clear as possible in relation to the false positive/false negative rate, and available as validated software. One dilemma is that the bandwidth of approaches for design and analysis of toxicological studies is too broad from the above viewpoint. Moreover, some of the International Conference on Harmonisation (ICH) documents and national guidelines on toxicological issues include facts relevant to statistics. Unfortunately, these seem to be largely written by nonstatisticians. Therefore, there is a need for recommendations on statistical analysis. Two such papers were developed in preparation for the DIA meetings in Brugge (March 1996) and Tokyo (August 1996): one on mutagenicity studies (9), the other on repeated toxicity studies (10). A third paper, on animal carcinogenicity studies, is still under preparation by an international team under the supervision of W. Fairweather.

The DIA workshop held in March 1996 in Brugge (Belgium), consisted of five sessions which included:

- Animal carcinogenicity studies,
- *In vivo* and *in vitro* mutagenicity studies,

Reprint address: Prof. Dr. L. A. Hothorn, Herrenhauser Str. 2, D-30419, Hannover, Germany.

Drug Information Journal, Vol. 31, pp. 323–326, 1997
Printed in the USA. 0092-8615/97

RECOMMENDATIONS FOR BIOSTATISTICS OF MUTAGENICITY STUDIES*

DIETER HAUSCHKE, PHD
Byk Gulden, Germany
MAKOTO HAYASHI, DSC
National Institute of Health Sciences, Japan
KARL K. LIN, PHD
Food and Drug Administration, United States
DAVID P. LOVELL, PHD
BIBRA International, United Kingdom
W. DAVID ROBINSON, MSC
Glaxo Wellcome, United Kingdom
ISAO YOSHIMURA, PHD
Science University of Tokyo, Japan

A Drug Information Association Workshop on "Statistical Methodology in Non-Clinical and Toxicological Studies" was held at the end of March 1996. The purpose of this meeting was to discuss and to obtain a consensus on the appropriateness of current and new biostatistical methods relevant in this field of drug development. The following summary represents a consensus of the Working Group on Biostatistics in Mutagenicity Studies. The recommendations outline the relevant principles of design and analysis rather than provide detailed specification of statistical methodology, thus providing the possibility of making reasonable choices between alternative approaches.

Key Words: Mutagenicity; Biostatistics; Recommendations

RECOMMENDATIONS

1. **How can statistics contribute to the evaluation of test data?** In the evaluation of mutagenicity experiments, statistical methods should be used to support the decision on whether a result is negative or positive by:
   - Estimating effects,
   - Investigating the dose-response relationship,
   - Identifying sources of variation,
   - Analyzing confounding factors,
   - Exploring the mechanisms of mutagenesis, and
   - Summarizing the data.

   Statistics should play a fundamental

*The views are those of the authors and not necessarily those of their employers.

Presented at the DIA Workshop "Statistical Methodology in Non-Clinical & Toxicological Studies," March 25–27, 1996, Bruges, Belgium.

Reprint address: Dr. Dieter Hauschke, Department of Preclinical Biometry, Byk Gulden Pharmaceuticals, P. O. Box 10 03 10, 78403 Konstanz, Germany.

Drug Information Journal, Vol. 31, pp. 357–361, 1997
Printed in the USA. 0092-8615/97

STATISTICAL PROOF OF SAFETY IN TOXICOLOGICAL STUDIES

DIETER HAUSCHKE, PHD
Department of Preclinical Biometry, Byk Gulden Pharmaceuticals, Konstanz, Germany

The statistical test of the traditional hypothesis of no treatment effect is commonly used in toxicological experiments. Failing to reject the hypothesis often leads to the conclusion of evidence in favor of safety. The major drawback of this indirect approach is the fact that what is controlled by a prespecified level is the probability of erroneously concluding hazard (producer risk). The primary concern of safety assessment, however, is the control of the consumer risk, that is, limiting the probability of erroneously concluding safety. In order to restrict this risk, safety has to be formulated as the alternative and hazard, that is, the opposite, has to be formulated as the hypothesis.

Key Words: Toxicology; Proof of hazard; Proof of safety; Biostatistics

INTRODUCTION

THE PURPOSE OF toxicology testing is the safety assessment of a new compound relative to a control. Based on this aim, statistical α-tests of the classical null hypothesis of no difference are usually performed. Failing to reject the hypothesis often leads to the conclusion that the compound has no deleterious effect in the corresponding biological model. The most important disadvantage of this indirect procedure is the fact that only the producer risk can be directly controlled by α. The primary concern, however, is control of the consumer risk. Thus, from a statistical point of view, the adequate test problem should be formulated by reversing the null hypothesis and the alternative and incorporating an a priori or a posteriori defined threshold. In this paper the direct approach will be compared with the indirect approach in the two-sample situation.

PROOF OF HAZARD

Let X_{ij} be independent random variables and suppose that the distribution functions $F_i(x) = F(x - \mu_i)$ are continuous ($i = 0, 1$ and $j = 1, \ldots, n_i$), with μ_0 denoting the location parameter for the control (vehicle or negative group) and μ_1 that for the compound (treatment group). Under the prior assumption of $\mu_0 \leq \mu_1$, that is, appropriateness of a one-sided test, the classical test problem (proof of hazard) is formulated as follows:

Presented at the DIA Workshop "Statistical Methodology in Non-Clinical & Toxicological Studies," March 25–27, 1996, Bruges, Belgium.

Reprint address: Dr. Dieter Hauschke, Department of Preclinical Biometry, Byk Gulden Pharmaceuticals, P.O. Box 10 03 10, 78403 Konstanz, Germany.

Drug Information Journal, Vol. 31, pp. 363–367, 1997
Printed in the USA. 0092-8615/97

THE CONTROL OF THE CONSUMER RISK IN THE AMES ASSAY

MARKUS NEUHÄUSER
Solvay Pharmaceuticals, Hannover, Germany
LUDWIG A. HOTHORN
Universität Hannover, LG Bioinformatik, Hannover, Germany

Preclinical tests in genetic toxicology represent safety studies. Therefore, the primary concern in the evaluation of Ames test data is the control of the consumer risk, that is, the risk of erroneously concluding safety. Hence, an equivalence test procedure is adequate. This approach is presented beside the two-fold rule and classical tests for differences with respect to an order restricted alternative.

Key Words: Ames assay; Consumer risk; Difference approach; Equivalence test procedures for ordered many-to-one designs

INTRODUCTION

THE AMES (1) ASSAY IS a short-term test in genetic toxicology (2) most often used worldwide. This assay is based on the detection of mutated histidine-dependent cell strains of *Salmonella typhimurium*. The mutated bacteria can mutate back to the wild-type, particularly if they are exposed to mutagenic compounds. The endpoint for evaluation is the number of revertant colonies on replicated plates. This endpoint is a count.

There are some difficulties in the statistical evaluation of the Ames assay:

- As the distribution of the discrete endpoint is not clear, different distributions are used: the Poisson distribution and two special cases of the generalized Poisson distribution according to Consul and Jain (3), the negative binomial distribution is an extension to the Poisson distribution as well (4). Furthermore, the normal distribution is used after a suitable transformation,
- Small and unbalanced sample sizes are used according to guidelines, for example, three per dose group and five for the negative control group are common sample sizes. With the small sample sizes, the variance may be reduced by ties. As a consequence, this can lead to a false positive result,
- The variances are frequently nonhomogeneous because of small sample sizes and ties. Moreover, the variance can increase with increasing effects, and

Presented at the DIA Workshop "Statistical Methodology in Non-Clinical and Toxicological Studies," March 25–27, 1996, Bruges, Belgium.

Reprint address: Markus Neuhäuser, Solvay Pharmaceuticals, Hans-Böckler-Allee 20, D-30173 Hannover, Germany.

Drug Information Journal, Vol. 31, pp. 403–412, 1997
Printed in the USA. 0092-8615/97

MODIFICATIONS OF THE CLOSURE PRINCIPLE FOR ANALYZING TOXICOLOGICAL STUDIES

LUDWIG A. HOTHORN, PHD
University of Hannover, Hannover, Germany

A Drug Information Association Workshop on "Statistical Methodology on Non-Clinical and Toxicological Studies" was held on March, 25–27, 1996. The purpose of this meeting was to discuss the appropriateness of current and new biostatistical methods in this field of drug development. This paper proposes a simple closure test for dose-response relationships under real data conditions. This approach takes into account deviation from variance homogeneity and monotonicity assumptions. Moreover, this approach can be easily adapted for "any" kind of two-sample tests, for example, nonparametric, dichotomous, censored, and so forth. Power was compared by a simulation study for selected conditions.

Key Words: MED; Closure principle; Trend test; Toxicological studies

INTRODUCTION

THIS ARTICLE DESCRIBES a simple closure testing procedure for dose-response relationships under real data conditions. Although this approach can be used for randomized clinical trials, pharmacological studies, and several toxicological studies, for example, reproductive toxicity, repeated toxicity studies (1) are considered here for analyzing the dose-response relationship. A one-way design: $\{C-, D_1, \ldots, D_k\}$ (with $C-, \ldots$ negative control; $D \ldots$ dose group; with k not too small $\in 2,3,4$) is assumed and multiple endpoints of a different scale, for example, approximate gaußian distributed (eg, body weight), a highly skewed distribution (eg, ASAT enzyme), dichotomous (eg, number of animals that died/number of animals at risk), or ordered categorical (eg, scored histopathological findings). Independent univariate analyses of the dose-response relationship should be performed, however, based on the same methodology. Only model-free approaches should be discussed here, because a priori in such "screening" studies a model for the profile cannot be assumed, even for a single endpoint.

The purpose of toxicological studies is safety assessment by comparing the doses to the concurrent negative control (many-to-one comparisons). Traditionally, the classical hypotheses are used: null-hypothesis for no difference and alternative hypothesis for a difference (or one-sided for an increase). In this setting, type I error α represents producer's risk and type II error β consumer's risk. A direct control of producer's risk (via a priori definition of α) is possible, but control of consumer's risk is of primary concern in risk assessment. Therefore, Hauschke (2,3) proposed the reversing of the hypotheses in the sense of a k-sample equivalence problem. For this approach a threshold δ must be de-

Presented at the DIA Workshop "Statistical Methodology in Non-Clinical and Toxicological Studies," March, 25–27, 1996, Bruges, Belgium.

Reprint address: Prof. Dr. L. A. Hothorn, Herrenhauser Str. 2, D-30419 Hannover, Germany.

Drug Information Journal, Vol. 31, pp. 433–447, 1997
Printed in the USA. 0092-8615/97

MULTIVARIATE MANY-TO-ONE PROCEDURES WITH APPLICATIONS TO PRECLINICAL TRIALS

SIEGFRIED KROPF, PHD
Research Assistant, Institute of Biometrics and Medical Informatics, Otto von Guericke University Magdeburg, Magdeburg, Germany
LUDWIG A. HOTHORN, PHD
Director, Institute of Bioinformatics, University of Hannover, Hannover, Germany
JÜRGEN LÄUTER, PHD
Director, Institute of Biometrics and Medical Informatics, Otto von Guericke University Magdeburg, Magdeburg, Germany

Comparisons of several treatments with a control represent a standard situation in preclinical trials. Usually, they are considered with a single variable, resulting in multiple test procedures such as the Dunnett test (1). Here, the multivariate many-to-one problem is considered, where several variables are observed on each individual of the control and treatment groups.

Classical MANOVA tests and their derivatives for the many-to-one problem require large sample sizes in order to be powerful if the dimension is high. In this paper, a new class of stabilized multivariate tests proposed by Läuter (2) and Läuter, Glimm, and Kropf (3) is extended to this special design. The new tests are based on linear scores which are derived in a certain way from the original variables. They utilize factorial relations among the variables.

It is shown here that the procedures keep the multiple level. In simulation experiments several versions of multivariate tests are compared with each other. Standard approaches are included as well as different score versions and a comparison of Dunnett-like procedures with Bonferroni-type procedures. Generally, an improved power of the new tests compared to standard procedures is demonstrated.

Key Words: Multivariate tests; Stabilized scores; Many-to-one procedures; Dunnett test; Principal component test

INTRODUCTION

MULTIPLE ENDPOINTS CAN occur in many different situations. A variable may be observed under different conditions or in the course of time, or there may also be different

Presented at the DIA Workshop "Statistical Methodology in Non-Clinical and Toxicological Studies," March 25–27, 1996, Bruges, Belgium.

Reprint address: Dr. S. Kropf, Institut für Biometrie und Medizinische Informatik, Otto-von-Guericke-Universität Magdeburg, Leipziger Straße 44, D-39120 Magdeburg, Germany.

Drug Information Journal, Vol. 31, pp. 463–469, 1997
Printed in the USA. 0092-8615/97

TREND TESTS FOR DICHOTOMOUS ENDPOINTS WITH APPLICATION TO CARCINOGENICITY STUDIES

MARKUS NEUHÄUSER
Solvay Pharmaceuticals, Hannover, Germany
LUDWIG A. HOTHORN
Universität Hannover, LG Bioinformatik, Hannover, Germany

A trend test for dichotomous endpoints analogous to the nonparametric Jonckheere test is developed. The power of this and all other single trend tests for dichotomous endpoints strongly depends on the shape of the dose response curve. Combined tests which have a stable power over a wide range of the ordered alternative are suggested. One can combine several contrast tests to a so-called adjustive test which is more powerful than a Cochran-Armitage test with equally-spaced scores. The latter was recommended by Armitage (1) in case there is no a priori knowledge of the type of the trend.

Key Words: Carcinogenicity studies; Trend tests; Dichotomous endpoints; Order restricted alternative; Unknown shape

INTRODUCTION

DICHOTOMOUS ENDPOINTS frequently occur in toxicological studies, especially in carcinogenicity studies. One observes whether the animal has developed a tumor or not. In this paper only crude tumor rates are considered, with no mortality-adjustment. An extension for stratified trend tests will be presented in a further paper.

Tumors are rare events, therefore, the sample size per group ranges up to 100 rodents on authorities' recommendations. Due to mortality and sacrifice, the sample sizes are often unbalanced. Table 1 displays lung tumor data from a study on 1,2-dichloroethane (2, p. 82). One can see unbalanced sample sizes, which are common in carcinogenicity studies.

There are dichotomous endpoints in other toxicological studies as well, for example, nonneoplastic histopathological lesions. Moreover, it is possible to dichotomize a continuous endpoint according to a cut-off value of clinical relevance (3).

A dichotomous endpoint can be represented by a random variable, which has a binomial distribution. In carcinogenicity as well as in other toxicological studies the many-to-one design with a negative control group and k different dose groups is commonly used. The

Presented at the DIA Workshop "Statistical Methodology in Non-Clinical and Toxicological Studies," March 25–27, 1996, Bruges, Belgium.

Reprint address: Markus Neuhäuser, Solvay Pharmaceuticals, Hans-Böckler-Allee 20, D-30173 Hannover, Germany.

Drug Information Journal, Vol. 31, pp. 615–619, 1997
Printed in the USA. All rights reserved.

0092-8615/97

APPLICATION OF ADAPTIVE INTERIM ANALYSIS IN PHARMACOLOGY

LUDWIG A. HOTHORN, PHD
University of Hannover, Hannover, Germany
ULLRICH MARTIN, PHD
Boehringer Mannheim GmbH, Penzberg, Germany

A Drug Information Association Workshop on "Statistical Methodology on Non-Clinical and Toxicological Studies" was held March 25–27, 1996. The purpose of this meeting was to discuss the appropriateness of current and new biostatistical methods in this field of drug development. This paper describes the application of adaptive interim analysis in pharmacological studies. Bauer and Köhne (1) published a two-step approach in the case of unknown a priori information. This approach is now widely used for clinical trails. Here, the advantages of use in some pharmacological studies will be discussed.

Key Words: Interim analysis; Pharmacological studies

THE PROBLEM

IN PHARMACOLOGY animal studies can be categorized as screening studies for a series of substances on a well-defined animal model in which the outcome is frequently dichotomous (eg, responder/nonresponder), standardized bioassay for estimation of biological activity, and studies to demonstrate a clinically relevant effect for a selected substance in a specific—sometimes not established—animal model, usually based on a design "treatment or dose groups versus negative control." The last type of study will be considered here. The paradigm of these studies is similar to clinical trials: randomized, placebo (negative control C–) controlled, based on experimental design with sample size estimation, definition of an (clinical) endpoint, and using a standardized protocol and analysis.

Frequently for these studies, a stepwise testing scheme can be observed: starting with a small sample size pilot study, continuing with a larger sample size extended pilot study, and decision making by a final per-protocol study. The reasons for this stepwise procedure are:

- The *clinically relevant endpoint* is unknown a priori. This primary endpoint should be selected from some candidates using this study,
- The *variance* σ is unknown a priori. The same problem occurs in clinical trials as described by Gould (2),
- The *relevant effect difference* δ (either to negative control or to a competitor's drug) is unknown a priori,
- The *effective dose* is unknown a priori in dose-finding studies, and
- The *relevant time* is unknown a priori within the usually repeated measurements.

Presented at the DIA Workshop "Statistical Methodology on Non-Clinical and Toxicological Studies," March 25–27, 1996, Bruges, Belgium.

Reprint address: Prof. Dr. Ludwig A. Hothorn, Herrenhauser Str. 2, D-30419 Hannover, Germany.

Drug Information Journal, Vol. 32, pp. 401–421, 1998
Printed in the USA. All rights reserved.

0092-8615/98

BIOSTATISTICAL METHODOLOGY IN CARCINOGENICITY STUDIES*

WILLIAM R. FAIRWEATHER, PHD
Associate Director, Office of Epidemiology & Biostatistics, Center for Drug Evaluation & Research, Food & Drug Administration, Rockville, Maryland
AMIT BHATTACHARYYA, PHD
Senior Statistician, SmithKline Beecham Pharmaceuticals R&D, Hertfordshire, United Kingdom
PETER P. CEUPPENS, PHD
Zeneca Pharmaceuticals, United Kingdom
GUENTER HEIMANN, DR. RER. NAT.
Corporate Biometry, Schering AG, Berlin, Germany
LUDWIG A. HOTHORN, PHD
Professor, University of Hannover, Hannover, Germany
RALPH L. KODELL, PHD
Division of Biometry & Risk Assessment, National Center for Toxicological Research, Food & Drug Administration, Jefferson, Arkansas
KARL K. LIN, PHD
Division of Biometrics II, Office of Epidemiology & Biostatistics, Center for Drug Evaluation & Research, Food & Drug Administration, Rockville, Maryland
HARRY MAGER, PHD
Head, Preclinical and Clinical Biostatistics, Center of Drug Research, Biometry, Bayer AG, Wuppertal, Germany
BRIAN J. MIDDLETON
Zeneca Pharmaceuticals, United Kingdom
WOUT SLOB, PHD
National Institute for Public Health and Environment (RIVM), Bilthoven, The Netherlands
KEITH A. SOPER, PHD
Senior Research Statistician, Merck Research Laboratories, West Point, Pennsylvania
NIGEL STALLARD, PHD
Medical and Pharmaceutical Statistics Research Unit, University of Reading, Reading, United Kingdom
JOHN VENTRE, MS
Nycomed, United States
JANE WRIGHT, PHD
Quintiles, United Kingdom

Reprint address: William R. Fairweather, Associate Director, Office of Epidemiology & Biostatistics, HFD-700, US Food and Drug Administration, 5600 Fishers Lane, Rockville, MD 20857.
*The views expressed herein are those of the authors and not necessarily those of their employers.

This paper addresses the design, conduct, and statistical analysis of carcinogenicity studies, especially in the context of drug products for human use. It contains suggestions concerning the choice of dose levels, number of animals, methods of slide reading, and the ensuing statistical analysis, focusing on the significance testing approach. The purpose of this document is to describe the current thinking of statisticians and others who work in the area of carcinogenicity studies. The authors represent experience gained in the pharmaceutical industry, regulatory agencies, and academia.

Key Words: Carcinogenicity; Study design; Data analysis; Test statistics; Literature