David Morgan <david.morgan@axess.co.uk> on 25-10-99 15:39:49

To: Bernhard.Huitfeldt@astrazeneca.com, f.aubin@fournier.fr, antonella.bacchieri@sigma-tau.it, Linda.Danielson@ucb-group.com, s.driessen@organon.oss.akzonobel.nl, afe4362@glaxowellcome.co.uk, bianca.francucci@pharma.novartis.com, marlis.herbold@hmrag.com, MerJ@novo.dk, meinhard.kieser@schwabe.de, Paul Koopman/EI/KENDLELD, mariepierre_malice@merck.com, willi.maurer@pharma.novartis.com, moccatti@netgrs.com, norbert.neumann@roche.com, ks@spadille.dk, "Huitfeldt, Bernhard" <Bernhard.Huitfeldt@arcus.se.astra.com>, msust@esteve.es, MARTINEZ Eduard <emartinez@almirallprodesfarma.com>

cc:

Subject: Report from ISI Conference in Helsinki 1999

---

With regard to item 11 on the agenda of EFSPI Council on 4 November I attach the following brief report.

The EFSPI session at the ISI Conference in Helsinki took place on August 12, 1999. The session was attended by 50-100 people - a small number in the pharmaceutical industry, some with relevant experience and some with no knowledge at all!

The session was chaired by Dr B Huitfeldt, who gave an introductory presentation about EFSPI. The following four papers were presented:

The Professional Background of Statisticians in the European Pharmaceutical Industry
David Morgan
AXESS Limited, UK

Harmonising the regulatory background to the practice of statistics in the pharmaceutical industry
Karsten Schmidt
Spadille ApS, Denmark

Proof of Safety in Toxicology
Dieter Hauschke
Byk Gulden Pharmaceuticals, Germany
and Meinhard Kieser
Dr. Willmar Schwabe Pharmaceuticals, Germany

Introduction and Development of Novel Statistical Methods for Clinical Development within Pharmaceutical Industry
Willi Maurer and Gerd Rosenkranz
Novartis Pharma AG, Switzerland.

There was invited discussion from Prof H-J Trampisch of Bochum University, Germany; Dr J Seldrup of Quintiles, France was not able to be present, but sent his invited contribution to the meeting. There was also good discussion from the floor.

I have attached Word documents containing the abstracts of the four presentations. The abstracts were published in the Proceedings of the Conference. They were also available on the web, but the ISI99 web-site appears to have closed down now.

Kind regards

David

**EFSPI SESSION AT ISI CONFERENCE, HELSINKI, 10-18 AUGUST 1999**

| | | | |
|---|---|---|---|
| **Session** | **Statistics in the pharmaceutical industry** | | |
| **Organiser/ Chair** | Dr Bernhard Huitfeldt<br>Astra Arcus Pharmaceuticals<br>Sodertalje<br>Sweden<br>[on behalf of the EFSPI Council] | | |
| **Paper 1** | **The professional background of statisticians in the European pharmaceutical industry** | | |
| **Speaker** | David Morgan<br>Hoechst Marion Roussel<br>Denham<br>UK | **Discussant** | Prof. Dr Hans-Joachim Trampisch<br>Ruhr-Universität Bochum<br>Germany |
| **Paper 2** | **Harmonising the regulatory background to the practice of statistics in the pharmaceutical industry** | | |
| **Speaker** | Dr Karsten Schmidt<br>Spadille Biostatistik<br>Fredensborg<br>Denmark | **Discussant** | Dr Hans Melander<br>Medical Products Agency<br>Stockholm<br>Sweden |
| **Paper 3** | **Technical developments in statistics in the pharmaceutical industry** | | |
| **Speaker** | Dr Dieter Hauschke<br>Byk Gulden<br>Konstanz<br>Germany<br><br>**and**<br><br>Dr Willi Maurer<br>Novartis Pharmaceuticals<br>Basel | **Discussant** | Dr Jorgen Seldrup<br>Quintiles<br>Strasbourg<br>France |

- **Statistics in pharmaceutical industry.**
  ***La statistique dans l'industrie pharmaceutique.***

Organiser:

European Federation of Statisticians in Pharmaceutical Industries (*Fédération européenne des statisticiens de l'industrie pharmaceutique*); Bernhard Huitfeldt (Sweden)

Papers:
*Communications:*
*The professional background of statisticians in the European pharmaceutical industry.* David Morgan (Great Britain)

*Harmonising the regulatory background to the practice of statistics in the pharmaceutical industry.* Karsten Schmidt (Denmark)

*Technical Developments in statistics in the pharmaceutical industry.* Dieter Hauschke (Germany) and Willi Maurer (Switzerland)

Discussants:
*Intervenants:*
Hans-Joachim Trampisch (Germany)

Hans Melander (Sweden)

Jorgen Seldrup (France)

# August 12

- 09:00-11:15 Scientific Meetings
- 11:15-13:00 Lunch Break and Administrative Meetings
- 13:00-15:15 Scientific Meetings
- 15:15-17:30 Scientific Meetings
- 17:30-19:00 Administrative Meetings

**09:00 11:15**

IPM 4. Statistics in telecommunication.
IPM 10. Organisation of national statistical institutes.
IPM 38. Lévy processes application and theory.
IPM 69. Issues related to use of censuses of population, agriculture and enterprises in subsequent surveys in developing countries.
IPM 82. Statistics in pharmaceutical industry.

**13:00 15:15**

IPM 3. History of ideas in statistics and probability in Scandinavian countries.
IPM 30. Coalescent modelling of evolutionary ancestry.
IPM 46. How to measure deregulation?
IPM 57. Computational aspects of graphical models and Bayesian inference networks.
IPM 61. Statistical education for life.

**15:15 17:30**

IPM 17. Advances in resampling methods.
IPM 48. Surveying and estimating for the informal sector.
IPM 53. Statistical computing in business and industry.
IPM 58. Statistical education and significance tests controversy.

---

Home Previous

# The Professional Background of Statisticians in the European Pharmaceutical Industry

David Morgan
*AXESS Limited, Department of Statistics and Data Management*
*277 London Road*
*Isleworth, Middlesex, UK*
*david.morgan@axess.co.uk*

## STATISTICIANS IN THE PHARMACEUTICAL INDUSTRY

The developing role of statisticians in the pharmaceutical industry has resulted in over 2000 statisticians being employed in Europe within companies and Contract Research Organisations, chiefly working on clinical trials. Lewis (1996) and Koepcke et al (1998) point out that national drug regulatory authorities in Europe are employing an increasing number of statisticians in order to facilitate statistical review of marketing authorisation applications.

The International Conference on Harmonisation (1998) emphasises the importance of statistical input to the design and analysis of clinical trials and states that "...the statistician should have a combination of education/training and experience sufficient to implement the principles articulated in this guidance" (Section 1.2). Such an individual could be expected to take responsibility for the design and analysis of a clinical trial and would be recognised as an appropriate signatory of a trial protocol and report.

This statement raises the question of what qualifications and experience make an individual appropriate for such work. The pharmaceutical industry operates on a global basis, with the objective that standards of research work should be harmonised in different parts of the world. One key aim of this harmonisation is that conclusions drawn from clinical trials in one region should be recognised as providing sufficient evidence for regulatory approval of the product worldwide. Hence there needs to be some means of assessing statistical competence on a common basis. However, this is problematic when different countries have such widely differing educational systems and history of applying statistics in the pharmaceutical industry.

In countries such as the USA, Canada and UK, degree programmes in statistics have existed for some time, leading to a clear concept of a statistician's skills and role. In Germany and the UK some attempt has been made to define the experience necessary to provide professional accreditation for statisticians. However there is no common European-wide understanding of what it means to be "an appropriately qualified and experienced statistician". There are still striking differences in the number and nature of statisticians working in the industry in different European countries. One source of diversity is the wide variety of university educational systems in Europe.

## CURRENT SITUATION

The European Federation of Statisticians in the Pharmaceutical Industry (EFSPI) reviewed the current situation and future developments in the qualifications and

experience of individuals regarded as 'qualified statisticians' for the pharmaceutical industry in the major European countries. Their findings are outlined in a recent publication by the EFSPI Working Group (1999).

In the UK, Denmark and Sweden most pharmaceutical statisticians have passed through the well-established MSc or other degree courses in statistics. In Belgium, Germany, the Netherlands and Switzerland, the majority of statisticians have a primary qualification in mathematics, with some subsequent specialisation in statistics. In France, Italy and Spain, many people working in pharmaceutical statistics have a primary qualification in medicine or biology rather than formal qualifications in statistics. In general there is little opportunity for specialisation in medical statistics in the European university systems, though such focussed courses have been appearing in several countries during the 1990s.

In addition there are schemes for the accreditation or certification of statisticians, in several countries in Europe. Around one third of pharmaceutical statisticians in the UK are Chartered Statisticians, which requires a degree in Statistics or equivalent, plus five years practical experience. In Germany some 50 statisticians in the pharmaceutical industry have qualified as “Certified Medical Biometrician”, which generally involves a degree in statistics or mathematics, plus three years practical experience with biometrics in medicine and at least five years of further scientific and medical education after the degree. The qualification is also open to those qualified in medicine, who would then need further mathematical and statistics training: this route of entry weakens this certification in some eyes. In Spain an accreditation scheme for medical statisticians in the pharmaceutical industry was initiated in 1996 for statisticians meeting the criteria outlined in the next section.

## OUTLINE DEFINITION OF AN APPROPRIATELY QUALIFIED AND EXPERIENCED STATISTICIAN

The skills and knowledge required by a statistician in the pharmaceutical industry include those defined by Lewis (1994). He suggests that effective statistical professionals need a strong technical foundation, knowledge of the pharmaceutical context, plus skills in communication and project management. These attributes would clearly be gained through a mixture of education, other technical training and relevant experience.

Following extensive consultation within the industry EFSPI have proposed the definition that a "qualified medical statistician" is expected to have a university degree in statistics or equivalent, plus more than 3 years experience in medical statistics.

An example of an equivalent qualification would be a degree in mathematics or related subject, involving more than one year’s (full time equivalent) courses in statistics, i.e. where statistical content formed at least one third of a three year course or one quarter of a four year programme. The definition requires experience in Medical Statistics, involving clinical trials and associated regulatory requirements, but not necessarily experience in the pharmaceutical industry itself.

While inevitably being somewhat vague and lacking detail on what might be included in university courses, this definition does make certain things clear. First, it is not necessary for an individual’s degree to be in statistics, provided that it has a suitable

statistical content. Second, a degree in medicine, with only a small number of courses in statistics, would not entitle someone to be regarded as a qualified medical statistician.

In addition the generality of this definition gives the advantage that it is not restricted in application to Europe, or to the clinical arena. Its principles could easily be adapted to use in North America or Asia, or to non-clinical statisticians.

## ISSUES NOT ADDRESSED BY THIS DEFINITION

In striving for simplicity, the definition above does not address the content of statistics courses, which may differ considerably in their coverage. However it is felt that if the mathematical background and basic principles are well covered at an appropriate level, additional material in a specialised area should be readily acquired by personal study.

In developing this guideline it was recognised that in the short term many good pharmaceutical industry statisticians would not meet the definition adopted, due to lack of formal statistical qualifications. In addition, it should not be taken to imply that statisticians rely purely on material covered in their degree programme. It is important to keep up to date with developments in their area of statistics by attendance at training courses, conferences and other forms of continuing professional development. In addition pharmaceutical statisticians need to broaden their capabilities, for example in terms of communication, management and consultancy skills, in addition to gaining knowledge in the areas of other sciences relevant to their work.

The definition is given to provide guidance for those who may find it useful. It is not intended to be the basis of a formal accreditation scheme. However it has been designed to be broadly consistent with requirements already in place. For example, the Chartered Statistician certification awarded by the Royal Statistical Society in the UK requires an appropriate degree in statistics "or in a subject containing a substantial coverage of statistical method and theory", plus five years practical experience in applying statistics, "of which three years should involve the candidate taking responsibility for the statistical content of their work". The requirements are also largely consistent with those for the German "Certified Biometrician" award, though they would not embrace the minority whose qualifications are in medicine.

The definition also does not address the difficult issue of when an appropriately qualified and experienced statistician should no longer be regarded as such, i.e. when should someone be 'de-certified'.

It must be recognised also that the quality of statistical work cannot be guaranteed purely by someone satisfying the criteria given here – the work itself must be open to scrutiny. However these criteria may provide a useful screening step for those making this type of assessment.

## THE WAY FORWARD

It is hoped that the outline definition presented here will give guidance to pharmaceutical companies in a variety of situations, for example when selecting

candidates for statistical roles with unfamiliar qualifications, or when assessing the competence of Contract Research Organisations. In addition it may be helpful to regulatory authorities and to universities in terms of defining an appropriate background to provide statistical expertise to support clinical trial and other pharmaceutical development activities and may provide a foundation for the future of the statistical profession in the global pharmaceutical industry.

## REFERENCES


EFSPI Working Group (1999). Qualified statisticians in the European pharmaceutical industry. Drug Information Journal 33, 407-415.

International Conference on Harmonisation (1998). E9: Statistical principles for clinical trials. International Conference on Harmonisation, Geneva.

Koepcke, W., Jones, D. R., Huitfeldt, B. and Schmidt, K. (1998). Statistics and statisticians in European drug regulatory agencies. Drug Information Journal 32, 243-251.

Lewis, J. A. (1996). Editorial: Statistics and statisticians in the regulation of medicines. Journal of the Royal Statistical Society A 159, 359-362.

Lewis, T. (1994) Statisticians in the pharmaceutical industry. In Discovering New Medicines (ed P. D. Stonier), 153-163. Wiley. Chichester.

# Harmonising the regulatory background to the practice of statistics in the pharmaceutical industry

Karsten Schmidt
Spadille ApS
Jernbanegade 34 C
DK-3480 Fredensborg, Denmark
E-mail: KS@spadille.dk
Home Page: www.spadille.com

The use of statistical methods in the medical and pharmaceutical areas has a long tradition. Since the introduction of the randomized controlled clinical trial(RCT), in the middle of this century, statistical methods have routinely been used in connection with such trials. Marketing of medicinal products is regulated by legal requirements and regulatory recommendations in the form of guidelines. To get an marketing approval of a drug an extensive documentation for the efficacy and safety of the drug is reviewed by the regulatory agencies. This documentation comprises information on all phases of the pharmaceutical and clinical development of the drug. Although the use of statistical methods is beneficial at almost all phases they are not used to the same extent However the cornerstone of a new drug application is one or more pivotal confirmatory phase III clinical trial, confirmative of all the pre-clinical and clinical work which has gone before. Some medicinal products may be approved on the basis of bioequivalence meaning that the blood levels of the active ingredient is demonstrated to be the about the same for the new drug as for a comparator drug having been approved previously by the above mentioned procedure.

The approval of drugs and the preparations of guidelines has been and is still performed by national regulatory agencies according to the national regulations. However attempts to harmonise the regulations between nations have been made for several years both within the European Union (EU) and between the EU, the United States (US) and Japan. Within the EU The Committee for Proprietary Medicinal Products (CPMP) act as a sort of common agency for EU by approving drugs in more than one EU country and by preparing common guidelines and Points to Consider papers. In US and Japan Food and Drug Administration (FDA) and the Japanese Ministry of Health and Welfare (MHW), respectively are the corresponding authorities. An initiative known as ICH (International Conference on Harmonisation) is an attempt to find common standards across all three regions (EU, US and Japan), and both industry and regulators take part.

Several guidelines contains material directly or indirectly addressing statistical methods and principles, the most important one being the ICH guidelines E3, E9, and E10 on Structure and content of Clinical Study reports, Statistical Principles for Clinical Trials, and Choice of Control Groups in Clinical Trials, respectively. The E10 is still under discussion, the E3 that was adopted in 1995 without taking due considerations to critical comments from the statistical community, the E9 was adopted in 1998 after a thorough discussion with the statistical community. These and other guidelines can be downloaded from the internet at URL: http://www.eudra.org/Document.htm, the homepage of the European Medicines Evaluation Agency (EMEA).

The development of common guidelines is the easy part of harmonisation. Harmonising the culture of regulatory agencies is much more important and much more difficult to achieve. The fact that the staffing at various regulatory agencies is very different inherently lead to differences. Thus the FDA employs a lot of statisticians and have done so for many years, only a few leading EU regulatory agencies employs a few statisticians and most of them have been employed for less than five years, the MHW employs no statisticians. The pharmaceutical industry in all three regions employs a lot of statisticians.

The European Federation of Statisticians in the Pharmaceutical Industry (EFSPI) consist of ten national groups representing more than 2000 statisticians working in or for the pharmaceutical industry. Only in three (Germany, Sweden, and UK) out of the ten nations the regulatory agencies employ statisticians the total number being less than ten statisticians. Even the harmonised ICH guidelines reflects disagreements with respect to statistical issues. As an example on the difficulties it is surprising that the CPMP less than half a year after the E9 has come into operation in EU has felt a need to prepare concept papers on key issues because the have experienced that the guideline either do not go far enough to answer practical difficulties or have not been interpreted consistently. These concept papers propose a paper be written for CPMP giving an EU interpretation of these issues to supplement the discussion in the E9. Having a specific EU interpretation without consultation with the other regions is not going in the direction of harmonisation. Thus it is obvious that regulatory statistics is not harmonised yet neither within the EU nor within the three ICH regions. Nonetheless there are aspects of statistics that apply in all regulatory agencies that employ statisticians

Focussing on the most important regulatory statistics guideline, the E9 is natural. This guideline is primarily based on the EU guideline on Biostatistical Methodology in Clinical Trials, CPMP (1994).Therefore most of the comments regarding this guideline and other issues in regulatory statistics still apply, Andreas Zipfel, Annette Robertson, Marco Girelli, and Karsten Schmidt (1994), John A. Lewis (1995, 1996), and John A. Lewis and Karen M. Facey (1998). The E9 is on statistical principles not on the use of specific statistical procedures or methods. A recent review of the E9 is given by David Morgan (1999). Regulatory guidelines is at the end the responsibility of the regulators and even though there have been considerable input from statisticians in the pharmaceutical industry and most of it has been taken into due consideration, consensus was not achieved on all issues. However the guideline reflects to a great extent the way pharmaceutical statisticians has worked for many years. Actually the E9 explicitly mention that the responsibility for all statistical work associated with clinical trials should lie with an appropriately qualified and experienced statistician, since this has not generally been the case within medical statistics.

The main special characteristics of statistics in clinical trial forming the basis for new drug applications as compared with medical statistics in general are with respect to (i) confirmatory versus exploratory analysis, (ii) analysis populations and intention to treat versus per protocol analysis, (iii) statistical analysis plan and updates based on blind review of data, (iv) superiority versus equivalence or non-inferiority trials, and (v) multiplicity corrections including interim analysis.

(i) An analysis can only be confirmatory if the hypothesis to be tested and the parameter to be estimated as well as the analysis to be used has been fully pre-specified in the protocol for only one primary variable or with appropriate multiplicity correction if more than one primary variable.

(ii) Two major analysis sets are recognised. The full-analysis set reflecting the so called intention to treat principle implying that in superiority trials, designed to give evidence that the new drug is superior to placebo and/or a comparator drug, the primary analysis should include all randomised patients. It is recognised that the full-analysis set might exclude some subject e.g. subjects failing to satisfy major entry criteria, to take at least one dose of trial medication and to have any data post randomisation. For trial designed to give evidence of equivalence or non-inferiority the primary analysis could be based on either the full or the per protocol analysis set dependent on which one is judged to me most conservative.

(iii) The principle features of the confirmatory analysis and the way anticipated problems for this analysis should be handled should be pre-specified in the trial protocol. A statistical analysis plan may be written as a separate document to be completed after

finalising the protocol. The plan may include detailed procedures for executing the statistical analysis of the primary and secondary variables and other data including how to handle analysis problems. The plan should be reviewed and possibly updated as a result of a blind review of the data and should be finalised before breaking the blind.

(iv) The use of placebo controls in clinical trials may be unethical. The analysis of equivalence studies or non-inferiority studies is certainly a hot topic closely connected to the E10 guideline on Choice of control groups. The E9 discusses some of the problems and pitfalls of these trials without giving many solutions.

(v) Ways to deal with multiplicity issues and to avoid data dredging are discussed. This includes use of multiple variables, of covariates, analysis of repeated measurements , and subgroups and interaction analysis. Precautions for use of interim analyses are emphasized not only regarding type 1 error adjustment but also regarding measures to ensure that all staff involved in the conduct and analysis should remain blinded to the results of such analyses.

## REFERENCES


Lewis, John A. 'Statistical Issues in the Regulation of Medicines', *Statistics in Medicine*, **14**, 127-136 (1995).

Lewis, John A. 'Statistics and Statisticians in the Regulation of Medicine', *J.R. Statist. Soc. A*, **159**, Part 3, 359-365 (1996).

Lewis, John A. and Facey, Karen M. 'Statistical Shortcomings in Licensing Applications', *Statistics in Medicine*, **17**, 1663-1673 (1998).

Morgan, David D.V.'The Impact of Harmonization', *Applied Clinical Trials*, 62-66 (March 1999).

Zipfel, Andreas, Robertson, Annette, Girelli, Marco, and Schmidt, Karsten 'Discussion Papers: A European Concept for Good Statistical Practice in Global Drug Development', *Drug Information Journal*, **29**, 473-510 (1995). A series of eight papers.

# Proof of Safety in Toxicology

Dieter Hauschke
*Byk Gulden Pharmaceuticals, Dept. of Biometry*
*Byk-Gulden-Str. 2*
*78467 Konstanz, Germany*
*dieter.hauschke@byk.de*

Meinhard Kieser
*Dr. Willmar Schwabe Pharmaceuticals, Dept. of Biometry*
*Willmar-Schwabe-Str. 4*
*76227 Karlsruhe, Germany*
*meinhard.kieser@schwabe.de*

## 1. Introduction

The purpose of toxicology testing is the safety assessment of a new compound relative to a control. Based on this aim, statistical $\alpha$-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 $\alpha$. However, the primary concern 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 defined threshold.

## 2. Proof of Safety

It is assumed that the random variables are mutually independent and normally distributed with unknown but common variance $\sigma^2$, that is, $X_{ij} \sim N(\mu_i, \sigma^2)$, $j = 1, ..., n_i$, $i = 0, 1$, with $\mu_0$ denoting the population mean for the control group and $\mu_1$ that for the compound (treatment group). The classical test problem (proof of hazard) is formulated as follows:

$$(1) \qquad H_0: \ \mu_1 - \mu_0 = 0 \quad \text{vs} \quad H_1: \ \mu_1 - \mu_0 \neq 0 .$$

Failing to reject the hypothesis by the *t*-test at level $\alpha$, traditionally leads to the conclusion that the compound has no harmful effect in the experiment. Of course, a *p*-value greater than $\alpha$ might indicate that there is no adverse effect of the treatment. However, due to inadequate sample sizes and/or large variances, a quite large difference can be present yet not be statistically significant. It is intuitively clear that the consumer risk should be of primary concern in toxicology (Hauschke and Hothorn, 1998). Therefore, the adequate test problem for toxicological studies should be formulated as follows:

$$(2) \qquad H_0^{\delta}: \ \mu_1 - \mu_0 \leq \delta_1 \ \vee \ \mu_1 - \mu_0 \geq \delta_2 \quad \text{vs} \quad H_1^{\delta}: \ \delta_1 < \mu_1 - \mu_0 < \delta_2 ,$$

where the interval $(\delta_1, \delta_2)$, $\delta_1 < 0 < \delta_2$, defines the safety range. However, a more common situation in practice is that the equivalence limits $\delta_1$ and $\delta_2$ are expressed as proportions of the unknown population mean $\mu_0 (\neq 0)$. Suppose that $\delta_1 = f_1 \mu_0$ and $\delta_2 = f_2 \mu_0$, where $-1 < f_1 < 0 < f_2$.

The foregoing test problem can be formulated as

$$(3) \qquad H_0^{\delta}:\ \mu_1-\mu_0 \le f_1\mu_0 \ \vee\ \mu_1-\mu_0 \ge f_2\mu_0 \quad \text{vs} \quad H_1^{\delta}:\ f_1\mu_0 < \mu_1-\mu_0 < f_2\mu_0$$

or as

$$(4) \qquad H_0^{\theta}:\ \frac{\mu_1}{\mu_0} \le \theta_1 \ \vee\ \frac{\mu_1}{\mu_0} \ge \theta_2 \quad \text{vs} \quad H_1^{\theta}:\ \theta_1 < \frac{\mu_1}{\mu_0} < \theta_2,$$

where $(\theta_1,\theta_2)$, $\theta_1 = 1+f_1, \theta_2 = 1+f_2$, $0<\theta_1<1<\theta_2$, is the corresponding equivalence interval for the ratio of $\mu_1$ and $\mu_0$. A split into two one-sided hypotheses results in:

$$(5) \qquad H_{01}^{\theta}:\ \frac{\mu_1}{\mu_0} \le \theta_1 \quad \text{vs} \quad H_{11}^{\theta}:\ \frac{\mu_1}{\mu_0} > \theta_1 \ \wedge\ H_{02}^{\theta}:\ \frac{\mu_1}{\mu_0} \ge \theta_2 \quad \text{vs} \quad H_{12}^{\theta}:\ \frac{\mu_1}{\mu_0} < \theta_2.$$

According to the intersection-union principle (Berger, 1982), $H_0^{\theta} = \bigcup_{r=1}^{2} H_{0r}^{\theta}$ is rejected in favor of $H_1^{\theta} = \bigcap_{r=1}^{2} H_{1r}^{\theta}$ (safety) at level $\alpha$, if and only if $H_{01}^{\theta}$ and $H_{02}^{\theta}$ are rejected at nominal $\alpha$. Sasabuchi (1988) demonstrated that the size-$\alpha$ likelihood ratio test rejects $H_{01}^{\theta}$ and $H_{02}^{\theta}$, if

$$(6) \qquad t_1^{\theta} = \frac{\overline{X}_1 - \theta_1\overline{X}_0}{\hat{\sigma}\sqrt{\frac{\theta_1^2}{n_0}+\frac{1}{n_1}}} \ge t_{\alpha,n_0+n_1-2} \ \wedge\ t_2^{\theta} = \frac{\overline{X}_1 - \theta_2\overline{X}_0}{\hat{\sigma}\sqrt{\frac{\theta_2^2}{n_0}+\frac{1}{n_1}}} \le -t_{\alpha,n_0+n_1-2},$$

$t_{\alpha,\nu}$ is the (1-$\alpha$)-percentile of the central *t*-distribution with $\nu$ degrees of freedom, $\overline{X}_0$ and $\overline{X}_1$ denote the correponding sample means and $\hat{\sigma}^2$ the pooled estimator of $\sigma^2$. Algebraic rearrangement (Hauschke et al., 1999) shows that condition (6) is equivalent to

$$(7) \qquad \theta_- \ge \theta_1 \ \wedge\ \theta_+ \le \theta_2 \ \wedge\ \overline{X}_0^2 > a_0,$$

where

$$(8) \qquad \theta_{\pm} = \frac{\overline{X}_1\overline{X}_0 \pm \sqrt{a_0\overline{X}_1^2 + a_1\overline{X}_0^2 - a_1a_0}}{\overline{X}_0^2 - a_0}, \quad a_0 = \frac{\hat{\sigma}^2}{n_0}t_{\alpha,n_0+n_1-2}^2 \ \text{ and } \ a_1 = \frac{\hat{\sigma}^2}{n_1}t_{\alpha,n_0+n_1-2}^2.$$

Condition (7) can be interpreted as follows: if the 100(1-2$\alpha$)% Fieller confidence set for $\frac{\mu_1}{\mu_0}$ has finite length, it is given by the interval $I_F = (\theta_-,\theta_+)$. Furthermore, Fieller´s confidence set is an interval if and only if $\overline{X}_0^2 > a_0$ holds true. Hence, the following two procedures lead to the same decisions: (i) conclude equivalence if and only if $H_0^{\theta}$ is rejected by the level-$\alpha$ Sasabuchi test based on $t_i^{\theta}$, $i$ = 1, 2; (ii) conclude equivalence if and only if $\overline{X}_0^2 > a_0$ and $I_F \subset (\theta_1,\theta_2)$, i.e. the $(1-2\alpha)$ 100% Fieller confidence interval is included in the safety range.

## 3. Power and Sample Size Determination

Assuming a balanced designs, i.e. $n_0 = n_1 = n$, the probability of correctly accepting $H_1^\theta$ is given by:

$$(9) \qquad P[t_1^\theta \geq t_{\alpha,2n-2} \ \wedge \ t_2^\theta \leq -t_{\alpha,2n-2} \mid \theta_1 < \frac{\mu_1}{\mu_0} < \theta_2, \sigma^2].$$

It can be shown that the vector $(t_1^\theta, t_2^\theta)$ has a bivariate noncentral *t*-distribution with noncentrality parameters

$$(10) \qquad \Theta_i = \frac{\mu_1 - \theta_i \mu_0}{\sigma \sqrt{\frac{1+\theta_i^2}{n}}}, \ i = 1, 2 \text{ and correlation coefficient } \rho = \frac{1+\theta_1\theta_2}{\sqrt{(1+\theta_1^2)(1+\theta_2^2)}}.$$

The above expression for the power can be calculated from the difference of two integrals (Hauschke et al., 1999). Expressing $\sigma$ as a proportion of the control mean, i.e. $\sigma = CV_0 \mu_0$, simplifies the presentation of the noncentrality parameters

$$(11) \qquad \Theta_i = \frac{\frac{\mu_1}{\mu_0} - \theta_i}{CV_0 \sqrt{\frac{1+\theta_i^2}{n}}}, \ i = 1, 2.$$

Figure 1 shows the power curves for an equivalence range of $(\theta_1, \theta_2)$ = (0.8, 1.25), a coefficient of variation $CV_0$ of 0.15, $n$ = 10, 20 and selected values $\frac{\mu_1}{\mu_0}$ from the alternative.

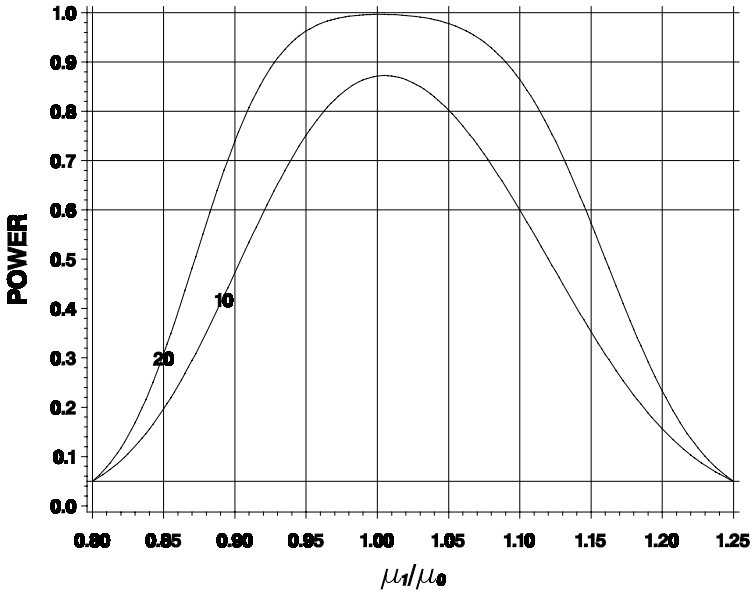


***Figure 1. Power curves refer to the equivalence range (0.8, 1.25), sample sizes (per group) of n = 10, 20, α = 0.05, $CV_0$ = 0.15***

The exact sample size to attain a pre-specified power of $1-\beta$ for a given level $\alpha$, a coefficient of variation $CV_0$ and a value from the alternative $\frac{\mu_1}{\mu_0}$, $\theta_1 < \frac{\mu_1}{\mu_0} < \theta_2$, can be directly obtained from the expression (9). However, the calculation of the integrals may not accessible to practioners, for that reason, the following approximation formula for sample size determination was provided by Kieser and Hauschke (1999). Assuming a safety range of the form $(\theta_1, \theta_2) = (1/\theta_2, \theta_2)$:

if $\frac{\mu_1}{\mu_0} = 1$ $\qquad n \geq \left(1+\theta_2^2\right)\left(\frac{CV_0}{\theta_2 - 1}\right)^2 \left(t_{\alpha,2n-2} + t_{\frac{\beta}{2},2n-2}\right)^2$,

$1 < \frac{\mu_1}{\mu_0} < \theta_2$ $\qquad n \geq \left(1+\theta_2^2\right)\left(\frac{CV_0}{\theta_2 - \frac{\mu_1}{\mu_0}}\right)^2 \left(t_{\alpha,2n-2} + t_{\beta,2n-2}\right)^2$

and

$\frac{1}{\theta_2} < \frac{\mu_1}{\mu_0} < 1$ $\qquad n \geq \left(1+\frac{1}{\theta_2^2}\right)\left(\frac{CV_0}{\frac{1}{\theta_2} - \frac{\mu_1}{\mu_0}}\right)^2 \left(t_{\alpha,2n-2} + t_{\beta,2n-2}\right)^2$.

Extensive comparisons between exact and approximate sample sizes over a wide range of different situations showed that the approximation formula nearly always gives the exact solution.

## REFERENCES


Berger, R. L. (1982). Multiparameter hypothesis testing and acceptance sampling. Technometrics, 24, 295-300.

Hauschke, D. and Hothorn, L. A. (1998). Safety assessment in toxicological studies: proof of safety versus proof of hazard. In Design and Analysis of Animals Studies in Pharmaceutical Development, (eds S.-C. Chow and J.-P. Liu) 197-225. Marcel Dekker, New York.

Hauschke, D., Kieser, M., Diletti, E.and Burke, M. (1999). Sample size determination for proving equivalence based on the ratio of two means for normally distributed data. Statistics in Medicine, 18, 93-105.

Kieser, M. and Hauschke, D. (1999). Approximate sample sizes for testing hypotheses about the ratio and difference of two means. Journal of Biopharmaceutical Statistics (accepted).

Sasabuchi, S. (1988). A multivariate one-sided test with composite hypotheses determined by linear inequalities when the covariance matrix has an unknown scale factor. Memoirs of the Faculty of Science, Kyushu University, Series A, Mathematics, 42, 9-19.

# Introduction and Development of Novel Statistical Methods for Clinical Development within Pharmaceutical Industry

Willi Maurer
*Novartis Pharma AG, Biostatistics*
*CH-4002 Basel/Switzerland*
*willi.maurer@pharma.novartis.com*

Gerd Rosenkranz
*Novartis Pharma AG, Biostatistics*
*CH-4002 Basel/Switzerland*
*gerd.rosenkranz@pharma.novartis.com*

## 1. Introduction

Biostatistics in pharmaceutical industry is operating in a highly regulated environment, in particular in those areas that are primarily related to the submission of new drug applications for approval by health authorities.

In order to stay competitive, however, also biostatistics departments in pharmaceutical companies are challenged to contribute to speeding up the development of new drugs. Requests usually include the reduction of the number of trials in a project or the number of patients within trials by means of better designs, or, if this is not possible, to make more effective use of the information contained in the data. Though regulatory bodies tend to be cautious with regard to the introduction of new statistical methodology that might help to achieve this goal, it is our experience that they are open to accept conclusions based on such methods if they are presented to them and discussed with them beforehand.

The present paper intends to highlight how careful introduction of new statistical methodology can help to meet this challenge. In most cases, this does not mean that these methods need to be developed from scratch. Biostatisticians in industry are more concerned with making theories available to (clinical) practice, to weigh the risks for unknown pitfalls pertaining to less known methods against their advantages, to get a feeling about how the results are affected by deviations from underlying assumptions by means of simulation studies and to make them work under real life conditions. There are however many cases where biostatisticians in the industry initiated and/or contributed to the further development of novel approaches. In the sequel we present a fairly subjective selection of examples from our own experience that to our opinion constitute such advancements. The paper closes with an equally subjective lists of unmet needs where industry statisticians would appreciate the support and cooperation in the development of solutions.

## 2. Continuous Reassessment Method

The accepted goal of dose-finding for oncology drugs is the determination of the maximum tolerated dose (MTD). Though this concept emerged from the development of cytotoxics it is still applied in the development of new drugs with different mechanisms of action. The standard method to find the MTD is the so-called 3+3 design where three patients are treated at a given dose which is going to be escalated for the next three patients if no dose limiting toxicity (DLT) was seen, kept constant if a DLT was seen in one patient and is reduced if two or more patients experienced a DLT. If the latter occurs, the resulting dose is declared the MTD and the trial is stopped after another 6 patients have been treated at the MTD.

This method has quite some drawbacks, e.g., that it cannot account for indication or patient population specific DLT rates, that it creates problems if one of the three patients is not evaluable and has to be replaced or if more than three patients have been treated at the same dose level. Also the precision of the estimator of the MTD is fairly imprecise and the method does not escalate the dose quickly enough if the starting dose has been selected too low and risks to underexpose too many patients.

The continuous reassessment method (CRM) introduced by O'Quigley, Pepe and Fisher (1990) is a Bayesian procedure that takes a prior distribution of a parameter of a one parameter dose-response model as a starting point. This distribution is 'updated' according to Bayes theorem as soon as new information about the occurrence of DLTs becomes available. The method can account for different numbers of patients per dose and can be targeted to any preselected DLT rate and even cope with the situation when investigators feel they have to overrule the proposal by CRM for the next dose to be used. The possibility to escalate by more than one dose level was seen as a disadvantage of CRM because it could lead to fairly high doses quite early and therefore to unacceptably severe toxicities. However, this drawback has been remedied by modifications proposed by several authors (see Piantadosi, Fisher and Grossman, 1998, for further references).

Despite some initial concerns about the acceptability of the results, the CRM has now replaced the old 3+3 design almost completely for dose-finding studies in oncology in our company. In addition, an alternative Bayesian method for dose escalation studies that does not assume a specific parametric curve for the dose-toxicity relationship was developed by Gasparini and Eisele (1998). Additional ongoing investigations include choice of cohort size and stopping rules with the goal of reducing trial duration and development of methods for which estimation of the MTD is not the ultimate goal.

## 3. Conditional/predictive Power

The determination of the probability of a trial to be "successful" at the design stage and during the trial bears some attractiveness. If these probabilities are used to modify other planned or running trials no adjustments of significance levels are necessary; if they are used to decide whether the trial should better be stopped for lack of efficacy only reductions in power have to be taken into consideration, since the probability of a type-I error can only decrease.

The conditional power is determined given the results of the assessments of a part of the patients and assumptions on the effect of the drug in the still outstanding patients (Lan, Simon and Halperin, 1982). The predictive power is basically the expectation of the conditional power where the a-posteriori distribution of the parameter of interest is used. This distribution is obtained from some a-priori distribution of the parameter of interest, updated according to Bayes theorem on the basis of the results obtained so far (Spiegelhalter, 1986). Statisticians in the industry developed these ideas further and derived the respective formulae for various types of variables and designs.

In an oncology trial with survival as the primary endpoint the results were presented summarised by treatment groups without unblinding treatment. The conditional power was calculated under different scenarios - the original alternative hypothesis and an estimate for the current trend and its confidence limits - with interchanged assignments of the two treatments. In most of the considered cases the conditional power was well below 0.2, therefore it was decided to stop recruitment into the trial for lack of evidence for superiority of the experimental drug.

In a second example the anticipated predictive power of a pivotal trial (to be conducted later) resulting from data of a pilot study was used to determine an optimal sample size for the pilot. In this example no data were available and predictive power was used as a concept for comparison of different scenarios a priori rather than during a trial. Under different assumptions regarding subject variability and the underlying parameter, a minimal pilot sample size could be found such that the predictive power was above 0.5.

## 4. Adaptive Designs

Standard methods for group sequential trials do not allow for changes of features of the design as for example the sample sizes after an interim analysis. For the case of normally and binomially distributed variables the effect of adapting the final sample size based on observation of the variability at an interim analysis have been evaluated and methods recommended. Gugerli, Maurer and Mellein (1993) showed that it is possible to take into account not only the observed variability but also the treatment difference observed at an interim analysis for the computation of final sample size while protecting the overall level of significance. A more general approach was proposed by Bauer and Koehne (1994). It even allows for changing the success criterion for the second phase of the trial. The principle works with any combination test that is based solely on p-values, e.g. $p_1$ and $p_2$ stemming from the independent phases of a two stage trial. Using Fisher's combination test, one of the proposed procedures allows for the choice of $\alpha$ (overall significance level) and $\alpha_0$, the probability of (correctly) accepting the null-hypothesis after the first phase if $p_1 \geq \alpha_0$. They show how to compute $\alpha_1$, the probability of erroneously rejecting the null hypothesis after the first stage if $p_1 \leq \alpha_1$, if in addition the respective criterion after the second stage (given the trial is not stopped at the first stage) is $p_1 p_2 \leq \exp\left[-0.5\chi^2_{4;1-\alpha}\right]$, where $\chi^2_{4;1-\alpha}$ is the (1-$\alpha$)–quantile of the central $\chi^2$ distribution with 4 degrees of freedom.

This design has been applied in a transplantation study designed to determine whether the plasma levels of a drug to prevent graft rejection can be more accurately monitored when trough levels are measured or when an AUC is estimated based on a sparse sampling scheme of drug levels at different time points. One monitoring method will be regarded superior over the other if one could determine a difference in clinical outcome, e.g., in the graft rejection rates or time to rejection. The study is currently ongoing and has not yet reached end of recruitment for the first stage after which it will be decided to enlarge the trial if it warrants to do so or to definitively stop recruitment.

## 5. Rank and Select

A different situation arises if one is not sure which one of several treatment options should be compared with a control in a phase III trial. Because of the huge amount of resources and costs involved one would be hesitant to conduct a big trial with several treatment arms. A possibility is to include initially all treatment options and to select the one that will be eventually compared with the standard in an interim analysis.

Designs for this situation have been described by Thall, Simon and Ellenberg (1988) for binomial (where software for sample size calculation has been developed) and Schaid, Wienand and Therneau (1990) for time-to-event data. Research in this area is ongoing at the MPS Research Unit, University of Reading, in cooperation with pharmaceutical industry.

## 6. Unmet Needs

The following constitutes a rather subjective and incomplete lists of topics that seem not or not sufficiently covered by statistical methodology for the time being:

- *Informative censoring:*

Many censoring mechanisms can hardly be treated as non-informative and the standard statistical methods for time-to-event data may not apply in these situation. The issue is similar to the missing data problem in longitudinal studies.

- *Dose-finding methods that account for both efficacy and safety*
- *Selection methods that allow for more than one selection criterion*
- *Methods that support decision making for a whole drug development project*

## ACKNOWLEDGEMENTS

The authors want to thank Jeff Eisele, Elisabeth Gruendl, William Mietlowski, Hans Prestele and Elisabeth Wehrle for implementing the methods described above and for helpful discussions.

**REFERENCES**

Bauer, P., Koehne, K. (1994). Evaluation of experiments with adaptive interim analyses. Biometrics 50, 1029-1042

Gasparini, M. and Eisele, J. (1998). A curve-free method for phase I clinical trials. Technical Report BS98-01, Department of Biostatistics, Novartis Pharma AG. Submitted for publication.

Gugerli, U., Maurer, W., Mellein, B. (1993). Internally Adaptive Designs for Parallel Group Trials. DIA Journal 27, 721-732.

Lan, K.K.G., Simon, R., Halperin, M.. (1982). Stochastically curtailed tests in long-term clinical trials. Communications in Statistics-Sequential Analysis 1, 207-219.

O'Quigley, J., Pepe, M., Fisher, L. (1990). Continual reassessment method: a practical design for phase 1 clinical trials in cancer. Biometrics 46, 33-48.

Piantadosi, S., Fisher, J.D., Grossman, S. (1998). Practical implementation of a modified continual reassessment method for dose-finding trials. Cancer Chemotherapy and Pharmacology 41, 429-436.

Schaid, D.J., Wienand, S., Therneau, T.M. (1990). Optimal two-stage screening designs for survival comparisons. Biometrika 77, 507-513.

Spiegelhalter, D.J. et al. (1986). Monitoring clinical trials: conditional or predictive power? Controlled Clinical Trials 7, 8-17.

Thall, P.J., Simon, R., Ellenberg, S.S. (1988). Two-stage selection and testing designs for comparative clinical trials. Biometrika 75, 303-310.

**RÉSUMÉ:**

*Les méthodes statistiques innovantes sont de plus en plus utilisées dans les essais cliniques afin de contribuer au développement clinique accéléré des médicaments. Le statisticien de l'industrie pharmaceutique a alors pour tache d'appliquer et d'adapter ces méthodes aux essais cliniques. Dans certain cas il contribue aussi au développement ultérieur de ces méthodes statistiques. Quelques méthodes statistiques innovantes sont présentées :*

*La méthode d'évaluation continue qui permet de déterminer la dose maximale tolérée, le calcul de la puissance prédictive conditionnelle et de la prédiction de la puissance finale qui permettent d'améliorer la prise de décision au cours d'un projet clinique, les plans adaptés qui permettent entre autre de recalculer le nombre nécessaire de sujets lors d'une analyse intermédiaire,les plans à deux étapes qui permettent de sélectionner lors d'une analyse intermédiaire les groupes de traitement les plus prometteurs pour l'analyse finale.*

*En conclusion une liste subjective de méthodes statistiques nécessitant un développement plus approfondi est présentée.*