MATHEMATISCHES FORSCHUNGSINSTITUT OBERWOLFACH

Tagungsbericht 06/2000

Medical Statistics:

Current Developments in Statistical Methodology for Clinical Trials and Statistical Challenges of Molecular Medicine.

06.02. – 12.02.2000

The meeting was organised by H. Schäfer (Marburg) and R. Simon (Bethseda). During the 5 days of the conference, 32 talks and a tutorial on Genetic Epidemiology were given, 38 scientists from Germany (# 19), USA (# 13), Great Britain (# 3) Denmark (# 1), Israel (# 1), Austria (# 1) participated. The intention of the conference was to touch different areas with new developments of statistical methods and applications in medicine and biomedical research. So, there was one focus on innovative clinical trial designs including adaptive designs and Bayesian methods in clinical trials, and there was another focus on statistical methods related to molecular medicine, including statistical genetics and statistical methods in bioinformatics. Bringing scientists from rather different fields together turned out to be a very successful concept due to the fruitful interaction between colleagues not focussed solely on their speciality.

The organisers and participants thank the “Mathematisches Forschungsinstitut Oberwolfach” to make the conference possible in the usual comfortable and inspiring setting. The abstracts follow in alphabetical order.

Innovations in the design of clinical trials using a bayesian approach

Donald Berry, Durham, USA

I describe the differences between the frequentist and Bayesian approach in developing drugs and medical devices. The Bayesian approach is more flexible and it is decision-oriented. I describe the status of the Bayesian approach in medical research in the United States. Bayesian design of experiments uses predictive probabilities to weigh consequences of any particular trial design. An instance of the flexibility possible when taking a Bayesian view is the ability to use adaptive designs in which accumulating data affect the future course of the trial, including affecting the therapy assigned to the next patient in the trial. I give two examples of adaptive clinical trials.

Perspectives of Genome Scans and Candidate Gene Strategies in the Light of Ascertainment Reality

Heike Bickeböller, Neuherberg, Germany

Recent focus in genetic epidemiology is the identification of genes where predisposing or protective alleles of the gene change the overall risk in the population by a moderate factor. If allele frequencies are high, the public health impact can be high. The two basic strategies, genome scan or candidate gene investigation, will be discussed.

The goal of a genome-wide scan is the localisation of a susceptibility gene, usually by linkage methods. It does not assume any knowledge about the underlying biological mechanism. Since statistical significance is extremely hard to reach with realistic sample sizes, comparisons across studies, populations and phenotypes can be very helpful for deciding on finemapping regions. We suggest a screening p-value limit of 0.01 for the first scan.

Candidate genes should be either strong candidates based on function or weaker candidates in suggestive regions for linkage. Difficulties with the power of realistic sample size are discussed in the context of multiple testing.

Prospects for Guiding Regimen Design by Biomathematical Disease-Process Modeling

Roger S. Day, Pittsburgh, USA

This talk describes the development of a cancer modeling workbench and ist application to clinical problems in treatment of breast cancer. In cancer, the development and treatment of the disease has been subject to many efforts at mathematical modeling as deterministic or stochastic processes. The most fervent hope is that modeling would yield insights leading to truly better treatments for patients. However, a huge spectrum of basic research shows that the actual complexity of the cancer process is extraordinary. Modelers who attempt to follow and utilize the research are quickly humbled. Nevertheless, several factors conspire to make highly elaborate mechanistic models of cancer attractive as a component of clinical study design. These are (1) accelerated discovery in the molecular biology of cancer, (2) major expansions in the capabilities of modeling technology to achieve comprehensiveness and flexibility, (3) greatly improved computing speed.

The Oncology Thinking Cap is a biomathematical cancer modeling computer program developed at the University of Pittsburgh. We have been utilizing this modeling workbench to tackle several high-profile controversies in the treatment of breast cancer including (1) duration of tamoxifen for ancillary treatment, (2) use of high-dose chemotherapy, (3) optimal strategic use of antiangiogenesis therapies. This type of endeavor places a heavy demand on the model-building and model-validation processes. The challenges of structuring and documenting these processes necessitate an ancillary "intelligent assistant" computer program, a "cancer information genie". A prototype has been developed using the Protege knowledge management system from Stanford University's Department of Medical Informatics.

Practical Designs for Phase I Studies and their Implementation

Lutz Edler, Heidelberg, Germany

Clinical phase I studies in oncology present the first instance where patients are treated experimentally with a new drug and they are therefore of pivotal importance for the development of anticancer treatment regimens. The primary goal is to define and to characterize new treatments in terms of the maximum tolerated dose (MTD) in humans for being investigated for efficacy in further clinical trials. There has been an ongoing debate on the preferred use of the Bayesian based continual reassessment method (CRM) for the determination of the MTD compared to the use of the traditional dose escalation rule (TER). With exception of few theoretical results, the operating characteristics of both methods and their comparison have been based on simulation studies. Comparing simulation results from various sources yields evidence of superiority of the CRM, eventually modified by restricting the dose escalation towards being more conservative. Still prevailing methodological gaps are the missing incorporation of the selection of the dose levels, the lack of the use of categorical toxicity and of pharmacokinetic information in Phase I dose finding and MTD estimation. Finally, standard methods of analyzing Phase I data are contrasted to new graphical methods allowing a comprehensive and multidimensional presentation of the trial's outcome.

The study of candidate genes in drug trials: sample size considerations

Robert C. Elston, Cleveland, USA

With the discovery of an increasing number of candidate genes that may affect inter-individual variability in response to drugs, the design of drug trials that incorporate their study has become relevant. We discuss the determination of sample size for such studies when the number of tests to perform is given, or alternatively, the number of tests that can be performed when the sample size is given. In many cases a uniformly most powerful test does not exist and normal approximations are not sufficiently accurate to determine sample size. We discuss briefly various tests of interest and give examples to illustrate some of the problems that arise.

The newly proposed guidance for population and individual bioequivalence studies

Paul D. Feigin, Haifa, Israel

The FDA has put out a second revision of a draft document as a Guidance for Industry on Average, Population, and Individual Approaches to Establishing Bioequivalence. An explanation of the difference between the three forms of bioequivalence is given. A description of the proposed ratio measures and testing procedures for Population and Individual bioequivalence are discussed. These procedures are based on considering whether a ratio of expected squared deviations is small enough, and is tested by computing an approximate 95% upper confidence bound for a linearized form. If that upper bound is less than zero, then bioequivalence is declared. Although sample sizes for achieving power properties given in the guidance were confirmed by simulation, it is not clear to what the relevant Type I error rates refer. In effect, for situations when the reference formulation has relatively high within subject variability, bioequivalence is virtually assured. Care has to be taken in applications to ensure that such within subject variability does not arise from measurement error in assays. Comparison to historical within subject variability estimates may therefore play a key role in such studies.

Search for genes which influence the development of malignant Schwannomas in the rat

Christine Fischer, Heidelberg, Germany
A. Kindler Röhrborn, Essen, Germany

Inbred rodent strains with differing sensitivity to experimental tumor induction provide models for the detection of genes either responsible for cancer predisposition or modifying the process of carcinogesis. Thus, rats of the BD strains are differentially susceptible to the induction of neural tumors by N-ethyl-N-nitrosourea (EtNU). While newborn BDIX-rats exposed to EtNU develop malignant schwannomas predominantly of the trigeminal nerves with an incidence > 85%, BDIV-rats are entirely resistant. Genetic crosses between BDIX and BDIV rats served (i) to investigate the inheritance of susceptibility, (ii) to obtain animals informative for the mapping of losses of heterozygosity (LOH) in schwannomas, and (iii) to localize, by linkage analysis, genes associated with schwannoma susceptibility. F_1 rats of both orientations show a markedly decreased incidence of schwannomas (~ 20%) as well as an extended latency time; however no genetic model is applicable. Different genes are likely to be involved in schwannoma development and variable latency time.
Schwannoma-free survival was the parameter determined. Animals deceased with competing risks either due to development of schwannomas or a malignancy different from schwannoma or a non-malignant disease.
Statistical methods to evaluate “competing risk data” in a genome screen are not commonly used. Therefore, to identify markers possibly linked to a gene involved in schwannoma development, an affecteds only analysis was performed first using allelic transmission disequilibrium tests. Chromsome wide type 1 error was estimated by simulation methods.
Currently established approaches for the analysis of optimized change points in prognostic factors are adapted and generalized to genome wide searches for genes responsible for a possibly complex disease in cooperation with Dr. B. Lausen, London. These methods can deal with traits measured on a quantitative or ordinal scale and with censored survival data as well and in addition control the chromosome wide type1 simultanously.

References

- Kindler-Röhrborn A, Koelsch B U Fischer Ch, Held S, Rajewsky M F Ethylnitrosourea-induced development of malignant trigenimal schwannomas in the rat: Two distinct loci on chromosome 10 involved in strain-specific susceptibility. Cancer research (1999) 59:1109-1114.
- Lausen, B., and Schumacher, M. (1996), Evaluating the effect of optimized cutoff values in the assessment of prognostic factors, Computational Statistics and Data Analysis 21, 307-326.

Blinded Sample Size Reestimation in Multi-armed Clinical Trials

Tim Friede, Heidelberg, Germany

When planning a clinical trial or other experiment the determination of the sample size is a key issue. However, there is usually uncertainty about the magnitude of the variance which may lead to inadequate sample sizes. To handle this problem, Wittes and Brittain (1990) introduced the internal pilot study design for the case of a normally distributed outcome variable. Thereby, the variance is estimated after recruitment of a certain number of patients. Hence, this variance estimate is used for sample size reestimation. It is a drawback of this design that the treatment code has to be broken for interim variance estimation. International guidelines (ICH, CPMP), however, stress the importance of blinding. To circumvent the disadvantage of unveiling, Gould and Shih (1992) proposed a procedure for blinded variance estimation which utilizes an EM algorithm (Dempster, Rubin, Laird 1977). Deficits of this procedure are described. Alternative variance estimators of a simple nature are proposed and their performance in clinical trials is demonstrated by a simulation study.

References

- Wittes J, Brittain E: The role of internal pilot studies in increasing the efficiency of clinical trials, Stat Med. 1990;9:65-72.
- International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use. ICH harmonized tripartite guideline: Statistical principles for clinical trials. Stat Med. 1999;18:1905-1942.
- CPMP Working Party on Efficacy of Medicinal Products. Biostatistical methodology in clinical trials in applications for marketing authorizations for medicinal products. Stat Med. 1995;14:1659-1682.
- Gould AL, Shih WJ. Sample size re-estimation without unblinding for normally distributed outcomes with unknown variance. Com Stat (A). 1992;21(10):2833-2853.
- Dempster AP, Laird NM, Rubin DB. Maximum Likelihood from incomplete data via the EM Algorithm. Journal of the Royal Statistical Society, Series B. 1977;39:1-38.

Meta-regression analysis to interpret randomised clinical cancer trials comparing polychemotherapies

Dirk Hasenclever, Leipzig, Germany

In part I of the talk, the generalised Skipper model was presented: Compared with well controlled experiments in animal systems, there are two major sources of variance in clinical patients' population that influence the outcome of chemotherapy: Variance in hemosensitivity (including differences in tumour burden) and variance in latency times (or growth rates). Simplistically assuming independent parametric distributions for these latent covariates and a Skipper type dose effect (linear on a log scale), a simple model was described that can be fitted to clinical data (progression free survival + dose actually received) estimating the latent distributions from the form of the PFS curves. This model was fitted to a large data set in advanced Hodgkin's disease. The model predicted that an increase of about 30% in total dose while simultaneously shortening the treatment duration would lead to a clinically relevant increase in cure rates from first line treatment (Hasenclever 1996, Annals of Oncology 7, Suppl. 4, S95-98).This has recently been confirmed in the large dose escalation HD9 trial of the German Hodgkin's Lymphoma Study group, the benefit being even more pronounced than anticipated (Diehl 1998 Blood 92,10 Suppl. 1 Abstract #2002).

Part II of the talk presented a new method to estimate the slope of the effective dose-response relationship and to estimate equipotency weights for individual cytostatic drugs from randomised clinical trials. The key idea is to define an appropriate crude measure of chemotherapy strength: A weighted sum of the total doses over all drugs used corrected for total treatment duration. The correction in line with the Skipper model assumes that the tumour regrows during treatment pauses and thus chemotherapy has to eradicate one tumour plus a regrown fraction which depends both on the treatment duration and the tumour-specific latency time (growth rate). This effective dose concept differs from the well known concept of summation dose intensity (Hryniuk) which is independent of the number of cycles given and thus of the total dose and does not take entities into the account differences in growth rates between tumour. In order to estimate the unknown drug weights and the parameter in the treatment duration correction, published results from all randomised clinical trials comparing chemotherapies are used. The observed log hazard ratio may be assumed to be proportional to the log effective dose ratio of the trial arms compared up to terms of third order. This defines a non-linear meta-regression. In Hodgkin's disease, analysis of 68 randomised comparisons demonstrates a clear clinically relevant effective dose-response relationship (as confirmed by the HD9 trial) and yields reasonable estimates for drug weights consistent with available single agent phase II data and clinical judgement. Non-surprisingly, confidence intervals are wide for less frequently used substances.

The proposed method is rather crude as several effects known or suspected to exist are not taken into the account. Nevertheless total dose and total treatment duration clearly are first order determinants of clinical outcome, and one has to get first order effects right before one can address second order effects (timing,, drug synergy, resistance induction etc.) Model-based interpretation of all the available (and reliable) clinical data does not intent to prove but to generate rough quantitative predictions to be confirmed in respectively focused clinical trials (evidence based trial design). The rationale of two new trials of the German high grade Non Hodgkin Lymphom a Study group is partially based on a model-based interpretation of NHL trials.

Adaptive weighting and variable selection in studies with multiple endpoints

Gerhard Hommel, Mainz, Germany

I consider studies with one interim analysis. The main topic is how one can modify the hypothesis/es after the interim analysis such that the type I error rate is controlled. If only a global statament is desired, a solution was given by Bauer (1989). If individual statements should be made, it is shown that a formal application of the closure test may lead to an excessive type I error rate; two proposals for a correction are given. For a general multiple testing problem, by Kieser, Bauer and Lehmacher (1999) and Bauer and Kieser (1999) solutions are given, by means of which the set of hypotheses can be reduced after the interim analysis. If weights for the tests within each of the two stages are chosen, the same ideas can be applied. Since it is allowed that a hypothesis has weight 0 in the first stage, but a weight >0 in the second stage, a formal way has been found to include additional hypotheses in the second stage.

Dealing with missing data problems in family-based allelic association studies

Steve Horvath, Bonn, Germany

Once a sufficiently dense map of genetic markers (SNPs) has been established, meiotic mapping of complex disease genes will enter the era of allelic association studies. Family-based association tests between a marker and a disease locus have become popular because they are powerful in the case of tight linkage between a marker and a disease locus and because they protect against detection of spurious associations that are due to population stratification. Since missing parental genotype data present a major challenge for diseases with late age of onset, several family-based association tests have been introduced that allow testing for linkage disequilibrium despite missing parental genotypes. I will discuss 3 approaches in the order of increasing complexity: First, the SDT (sibship disequilibrium test) which is a simple sign test. Second, the XRC-TDT which follows the logic of the reconstruction-combined TDT by Knapp 1999. Third, the FBAT tests which are based on the Rabinowitz-Laird (1999) algorithm. It turns out that the RC-TDT by Knapp 1999 is practically identical to a test that results from specializing FBAT to a dichotomous trait and an additive marker coding. This sets the stage for generalizing the RC-TDT to different inheritance models and different phenotypes.

Assessment of uncertainty inherent in genome wide mapping strategies

Berthold Lausen, London, Great Britain

An important challenge of genomics is the assessment of statistical significance in disease-gene discovery by genome scan data (ZHAO et al. 1999). TENG and SIEGMUND (1998) evaluate the effectiveness of such a multipoint linkage analysis using approximations via Gaussian processes. The presence or absence of marker alleles define the correlation structure of the process of two-sample type statistics along a chromosome. I adapt and generalise an established approach for the analysis of optimised change points in prognostic factors (LAUSEN and SCHUMACHER 1992) to genome wide searches for genes responsible for a possibly complex disease. The method can deal with traits measured on a quantitative or ordinal scale and with censored survival data as well. Finally, I discuss the possibilities to extend the proposal to more complex problems.

- LAUSEN, B., and SCHUMACHER, M. (1992): Maximally Selected Rank Statistics. Biometrics 48, 73–85.
- TENG, J., and SIEGMUND, D. (1998): Multipoint Linkage Analysis Using Affected Relative Pairs and Partially Informative Markers (with Discussion). Biometrics 54, 1247–1279.
- ZHAO, L.P., PRENTICE, R., SHEN, F., and HSU, L. (1999): On the Assessment of Statistical Significance in Disease-Gene Discovery. American Journal of Human Genetics 64, 1739–1753.

Sequential and Multiple Testing for Dose-Response Analysis

Walter Lehmacher, Köln, Germany

The analysis of a dose-response effect under the assumption of an ordered alternative is considered. Several global tests of trend are available for this problem. Further there are multiple comparison procedures identifying a minimum effective dose or effective dose steps. They control the experimentwise or multiple error rate. All these prodedures are based on the closed testing principle or shortened versions of it like the method of a priori ordered hypotheses or the method of Rom et al.. For a single test, several well established group sequential or recently proposed adaptive sequential prodedures are available. Here it is shown that multiple and interim analyses can be combined. For group sequential trials, shortcuts by early rejections are possible. For adaptive trials, further benefits are possible: A recalculation of the sample sizes, even unbalanced ones, a change of contrast test statistics, or an ommitting of dose groups by futility reasons. This gives a lot of flexibility for the further conduct planning after midtrial interim analyses.

Combining the advantages of group sequential designs and adaptive designs in statistical monitoring clinical trials

Hans-Helge Müller, Marburg, Germany

A general method for statistical testing in experiments with adaptive interim analyses was proposed by Bauer and Köhne. At the time of a planned interim analysis reasonable adjustments up to a flexible redesign of the remaining study may be performed. The paper focused on two-stage designs using Fisher's product criterion for combination of the p-values of independent stages. In the setting of normal means, Proschan and Hunsberger proposed a two-stage design by specifying a conditional error function. When performing the planned interim analysis, the method allows for a designed extension on the basis of conditional power calculations. However, planning of the sample size and critical values for an interim analysis to achieve an efficient procedure overall with respect to different perspectives of the design, as power and average sample size, was not treated so far in publications on adaptive design methods. The selection of a classical group sequential design for a clinical trial addresses this point because there exists a variety of designs with well studied characteristics. However, classical group sequential designs appear to be not as flexible as adaptive designs. In this contribution a general method will be presented to combine the advantages of the two sequential approaches. Moreover, the principle will allow a flexible change of every specified design at every time during the course of the trial (even if an interim analysis was not planned when the trial starts) without affecting the type I error level.

On the statistical analysis of allelic loss and comparative genomic hybridization

Michael Newton, Madison, USA

The genetic structure of cancer cells can be highly abnormal. Whole regions can be deleted or amplified, there can be rearrangements of material, and subtler changes at the base level. A range of molecular technologies are in use to identify these abnormalities, including the determination of allelic loss at molecular markers, cytogenetic methods, and comparative genomic hybridization. These data provide information about the location and effect of tumor suppressor genes and oncogenes, but inference is complicated by various factors including background genetic instability and statistical dependence. I review in this talk a framework for constructing statistical models for cancer genome abnormalites. Briefly, the idea is to postulate a stochastic model which creates instability randomly and homogeneously, and then to suppose that a cell having incurred damage at a putative cancer gene has an increased probability of transmitting descendants into an observable tumor. The probability of data is then obtained via Bayes rule by conditioning on tumorigenicity. I illustrate the method with both allelic loss and comparative genomic hybridization data.

Statistical methods used for evaluating chemical safety in the environment

Chris Portier, Research Triangle Park, USA

Much of modern environmental science concentrates on protecting people from potentially harmful chemicals such as pesticides and herbicides. Safeguarding humans from unhealthy exposures usually includes two problematic steps. First, even though humans experience low-level exposures to many compounds, human risk must be determined from toxicological experiments or occupational observations that generally have much higher exposures. Second, scientists usually must extrapolate from a chemical's effect in rodents—the traditional experimental models—to humans. The National Institute of Environmental Health Sciences' (NIEHS) Laboratory of Computational Biology and Risk Analysis (CoBRA) is working to solve the problems associated with these steps by designing biologically-based mechanistic models (BBMM) to guide the interpretion of the available data. The resulting knowledge can be applied to predict risks, for example, predicting a particular compound's carcinogenic potential. The eventual goal of this effort is to quantitatively link the likelihood of a disease to environmental causes.

Designing such a BBMM for a specific system involves multiple steps. The first step is to search the scientific literature for information on that particular system (e.g. physiological and biochemical characteristics). Then existing models (if any) of how a system works and additional information from the literature are combined to form a modified model. An attempt is made to quantify each step in the biological system associated with the onset of the disease. The models use any and all mathematics, including deterministic and stochastic models. The model's steps are arranged and transformed into mathematical equations. A computer program is written that uses the available data to estimate the model parameters. Results of the program address whether the available data is consistent with the model. If not, the model is rejected; if it is consistent, the model is fine-tuned and improved. Examples of this type of modeling exist for dioxins[1-6] and other agents.

1. Kohn, M.C., *et al.*, *A mechanistic model of effects of dioxin on thyroid hormones in the rat.* Toxicol Appl Pharmacol, 1996. **136**(1): p. 29-48.
2. Kohn, M.C. and C.J. Portier, *Effects of the mechanism of receptor-mediated gene expression on the shape of the dose-response curve.* Risk Anal, 1993. **13**(5): p. 565-72.
3. Kohn, M.C., *et al.*, *A mechanistic model of effects of dioxin on gene expression in the rat liver.* Toxicol Appl Pharmacol, 1993. **120**(1): p. 138-54.
4. Portier, C., *et al.*, *Ligand/receptor binding for 2,3,7,8-TCDD: implications for risk assessment.* Fundam Appl Toxicol, 1993. **20**(1): p. 48-56.
5. Portier, C.J., *et al.*, *Modeling the number and size of hepatic focal lesions following exposure to 2,3,7,8-TCDD.* Toxicol Appl Pharmacol, 1996. **138**(1): p. 20-30.
6. Portier, C.J. and M.C. Kohn, *A biologically-based model for the carcinogenic effects of 2,3,7,8-TCDD in female Sprague-Dawley rats.* Organohalogen Compounds, 1996. **29**: p. 222-227.

Recursive Combination Tests

Martin Posch, Wien, Austria
(together with Werner Brannath and Peter Bauer)

We present a method that extends the flexibility of adaptive designs to the number of interim analyses and to the choice of decision boundaries. It is based on a recursive application of a general class of two stage combination tests for p-values. The method covers as special cases the classical group sequential and adaptive two stage tests. Its recursive nature, under very general assumptions, gives a simple construction principle for an overall p-value, for confidence intervals and median unbiased point estimates. This method extends a procedure proposed by Müller and Schäfer (1999) based on the conditional error function for a replanned group sequential design.

Two-Stage Designs To Adaptively Modify Sample Size In Clinical Trials

Michael Proschan, Bethesda, USA

Sample size calculations are very important in clinical trials with a continuous outcome. Unfortunately, they depend on both the nuisance parameter (the variance) and the treatment effect. This talk covers two-stage designs in which the first stage is used to estimate parameters, and the final sample size is adjusted accordingly. The first part of the talk concerns sample size recalculation based only on the variance, while the second part deals with recalculation based on the observed treatment effect. The connection between two-stage tests, conditional error functions, and positive quadrant tests is shown.

Statistical Methods for HIV Genomics

Françoise Seillier-Moiseiwitsch, Chapel Hill, USA

The inefficiency of the replication process in HIV, like in any retrovirus, gives rise to many variants. The observed variability reflects both viability of the mutant and selection pressures from the immune system. This talk will review some recently developed methodology to study various anspects of the molecular evolution of HIV. One of the central themes is to quantify heterogeneity and to compare subgroups. Another focus is to detect correlated mutations and to incorporate them into phylogenetic reconstructions. Finally, it is becoming increasingly important to identify the link between the sequenceinformation and some specific biological characteristic.

Decision analysis and bioequivalence

Stephen Senn, London, Great Britain

The planning and analysis of bioequivalence studies is an example of an application of medical statistics that is unique to drug development[i]. Bioequivalence studies are trials in which the concentration time curves of a pharmaceutical in the blood are compared for different formulations of the same molecule. The planning and analysis of such studies has been accompanied by considerable controversy.

Bioequivalence studies can have at least two rather different *ultimate* purposes. Their immediate purpose is nearly always to show that two or more formulations are equivalent as regards therapeutic effects and side-effects. The argument is that equality of the formulations in terms of serum concentration necessarily (or at least very plausibly) implies equality in all other aspects. However, what has not always been clearly distinguished is that such equality can have two rather different ultimate purposes. It can be regarded as being a goal in itself or merely a means to an end. For the latter purpose a bioequivalence study is simply carried out as a shortcut for a full development: if the full development could be carried out cheaper and faster than the bioequivalence study it would be performed instead. It will be argued that this is the most important purpose of bioequivalence and that this is the only one with which the regulator should be concerned. Bioequivalence as an end in itself only arises if there is some desire to compare two or more formulations each of which is in any case perfectly acceptable for registration.

Failure to distinguish between these two purposes is at the root of much of the disagreement currently being aired in the subject of individual bioequivalence[ii]. This has centred around the notions of 'prescribability' and 'switchability' introduced by Anderson and Hauck[iii]. Two formulations are equally prescribable if a prescribing physician faced with a choice for a new patient yet to be treated with either could regard either formulation as being equally suitable. Two treatments are switchable if a physician could safely switch a patient from one formulation to the other. However, prescribability is all that is needed to register a drug. If two formulations are equally prescribable then we have no reason to choose either as being more likely to repeat the successes that led to one or other being registered in the first place. As such, to claim that one can be registered but the other cannot would be arbitrary and against the interests of patients and prescribers. On the other hand, the fact that two formulations are equally prescribable should not in itself entitle a health-care reimburser to force a physician to switch a patient from one to another: such a switch only involves no loss if switchability has

been demonstrated. Even here, however, the importance of switchability should not be over-emphasised. The risk in switching from one prescribable formulation to another for an existing patient cannot be greater than that to a naive patient in taking either for the first time. Yet if the drug is registered this risk must be deemed acceptable.

It will be argued that the notion of switchability, whilst of some theoretical interest, is of less practical relevance than has been supposed and that the subject would be better understood if some explicit modelling of losses were made.

This introduces a second theme. There has long been controversy over the appropriate analysis for bioequivalence studies even where simple mean bioequivalence is the object. The usual approach now used is to consider confidence intervals that are symmetric about the point estimate[i,iv]. In the past confidence intervals that were symmetric about the point of equivalence were proposed instead[v]. O'Quigley and Baudoin[vi] showed that these corresponded to answering two rather different fiducial questions: 1. Was the probability that the relative bioavailability lay within the region of equivalence as high as required? 2. Did the region of equivalence include the most probable region for the relative bioavailability?

Recently, however further proposals have been forthcoming with at first sight rather peculiar properties. For example an approach based on hypothesis testing[vii], given a large enough standard error, can permits the conclusion that the formulations are equivalent even where the point estimate is outside the limit of equivalence. Recently, Lindley has introduced an approach with an explicit consideration of losses[viii]. It will be considered to what extent, if any, these approaches improve on current practice.

References

[i] Senn, S.J. Statistical Issues in Drug Development, Wiley, Chichester, 1997

[ii] Senn, S.J. In the blood: proposed new requirements for registering generic drugs. The Lancet. 352, 85-86,1998

[iii] Anderson, S and Hauck, W.W. Consideration of individual bioequivalence. Journal of Pharmacokinetics and Biopharmaceuticals, 18, 259-273, 1990.

[iv] Schuirmann, D.J. A comparison of two one-sided tests procedure and the power approach for assessing the equivalence of average bioavailability. Journal of Pharmacokinetics and Biopharmaceutics, 15, 657-680, 1987.

[v] Westlake, WJ. Symmetrical confidence intervals for bioequivalence trials. Biometrics, 32, 741-744, 1976.

[vi] O'Quigley, J. and Baudoin, C. (1988) General approaches to the problem of bioequivalence, Statistician, 37, 51-58.

[vii] Brown, L.D., Hwang, J.T.G. and Munk, A. An unbiased test for the bioequivalence problem, Annals of Statisitics., 25, 2345-2367, 1997.

[viii] Lindley, D.V. Decision analysis and bioequivalence trials. Statistical Science, 13, 136-141, 1998.

Statistical Inference for Self-Designing Clinical Trials

Yu Shen, Houston, USA

In the process of monitoring clinical trials, it is appealing to use the interim findings to see if the sample size originally planned will provide adequate power when the alternative hypothesis is true, and to adjust the sample size if necessary. In Shen and Fisher (Biometrics, 1999), we propose a flexible sequential monitoring method for continuous outcomes with immediate responses, in which the maximum sample size does not have to be specified in advance. The final test statistic is constructed based on a weighted average of the sequentially collected data, where the weight function at each stage is determined by the observed data prior to that stage. Such a weight function is used to maintain the integrity of the variance of the final test statistic, so that the overall type I error rate is preserved. Moreover, the weight function plays an implicit role in termination of a trial, when a treatment difference exists. Furthermore, the design allows the trial to be stopped early when the efficacy result is sufficiently negative.

Challenges in the Development and Use of DNA Microarrays

Richard Simon, Bethesda, USA

DNA microarrays are a new technology for measuring the expression level of thousands of genes or for genotyping thousands of markers with a single hybridization. Microarrays have many potential applications for elucidating the pathogenesis of disease, revealing basic biological processes, and identifying molecular targets for disease prevention, early detection, classification and treatment. The utilization of microarrays in biomedicine presents many challenging statistical problems ranging from image analysis to pattern discovery and classification with thousands of variables. In my presentation I will describe microarray technology, biomedical questions being addressed with this technology and statistical research being conducted by my group and others to harness the power of this approach.

A sequential design for phase III clinical trials incorporating treatment selection

Nigel Stallard, Reading, Great Britain

Most methodology for phase III clinical trials focusses on the comparison of a single experimental treatment with a control. In practice, however, sufficient data to enable choice of a single experimental treatment may not be available prior to the trial. In this case, the phase III trial might start with several experimental treatments and a control, and include an interim analysis at which the most promising experimental treatment and the control treatment are retained, and all others eliminated. This talk describes an approach for the construction of sequential stopping rules that preserve error rates when the first interim analysis involves treatment selection.

Approximate Bayesian Evaluation of Multiple Treatment Effects

Peter F. Thall, Richard M. Simon and Yu Shen, Houston, USA

We propose an approximate Bayesian method for comparing an experimental treatment to a control based on a randomized clinical trial with multivariate patient outcome. Overall treatment effect is characterized by a vector of parameters corresponding to effects on the individual patient outcomes. We partition the parameter space into four sets where, respectively, the experimental treatment is superior to the control, the control is superior to the experimental, the two treatments are equivalent, and the treatment effects are discordant. We compute posterior probabilities of the parameter sets by treating an estimator of the parameter vector like a random variable in the Bayesian paradigm. The approximation may be used in any setting where a consistent, asymptotically normal estimator of the parameter vector is available. The method is illustrated by application to a breast cancer data set consisting of multiple time-to-event outcomes with covariates, and to count data arising from a cross-classification of response, infection and treatment in an acute leukemia trial.

Survival analysis and stochastic models of carcinogenesis

Alexander Tsodikov, Salt Lake City, USA

Our preliminary results strongly suggest that inadequacy of statistical methodology is largely responsible for many inconsistencies encounted in the currently available literature regarding possible role of covariates in cancer survival. Departures from proportional hazards encounted in the analysis of cancer data are intimately connected to biological processes underlying survival data. We propose to bring together the advantages of mechanistic models of cancer and statistical convenience model to enrich the existing tools of data analysis and its interpretation. Based on the theory of the PH model for improper survival functions a new class of biologically motivated statistical models is proposed. An algorithm is proposed to fit the PH model by maximizing the full likelihood. The concept is extended to apply to a broad class of semiparametric models including the proportional odds model and nested extentions of the cox model with nonproportional hazards. Various extentions of the PH model are studied. These include time-dependent covariates, combined additive/multiplicative effects, extended regression, limiting cure models. A class of mechanistic models of carcinogenesis is defined that matches the statistical models mentioned above. This allows one to biologically interpret the model assumptions (like the PH assumption) and the model parameters, thus providing a link to a deeper explanation of the data. Numerous examples of real data analysis are presented. These include clinical trials, animal carcinogenesis experiments, incidence of secondary tumors in cancer patients, analysis of cancer registry data.

'Complex models' in biostatistics - Computational and conceptual issues

Werner Vach, Odense, Denmark

In recent years we can observe more and more statistical approaches based on "complex models". With complex models I mean here models, which specify the joint distribution of several variables, where some of the variables are never or only partially observed, and where the joint specification is based on a composition of several, relatively simple, but different models. Typical examples are generalised linear mixed models, models for longitudinal data with informative dropouts, latent class models for analysing diagnostic procedures in the absence of a gold standard, models for incorporating information from validation substudies into the analysis of case control studies or frailty models for studying the genetic influence on longevity.

In the first part of the talk I report on the development of a software tool to assist computation of ML estimates in this very general class of models. Some general issues in using Monte Carlo integration methods for approximating the integrals of the likelihood are discussed. Besides these computational issues such complex models arise also the general difficulty of limited knowledge about sensitivity against violation of basic assumptions. We argue, that consideration of asymptotic bias and asymptotic variance can be a helpful tool in general and present some examples.

Power Assessment in Adaptive Designs

Gernot Wassmer, Köln, Germany

Several approaches for adaptive designs were proposed in the recent literature. The focus of my talk is on adaptive designs with more than one (adaptive) interim analysis. The methods are described in a unifying conceptual framework using a generalised conditional error function approach that is due to Proschan and Hunsberger (Biometrics, 1995) and the combination test priciple that is due to Bauer (Biom. und Inform. in Med. und Biol., 1989), respectively. Power comparisons in different designing strategies will be presented.

On the Use of Conditional Power in Clinical Trials

Janet Wittes, Washington, USA

The typical design of a clinical trial selects a sample size on the basis of the prespecified difference to be detected and on the desired power to detect that difference. Often the data accruing are inconsistent with an adequate power for alternative hypotheses of clinical interest. Many people, therefore, calculate an index of futility, where futility is defined as some level of conditional power so low that, conditional on the observed data and a "reasonable" trend for the remainder of the study, the trial is very unlikely to show a statistically significant benefit for treatment. The paper discusses the choice of indices of futility and presents examples of simulations for cases in which the drift in the data does not follow a Brownian motion. It points to the wide variability in estimated conditional power when the trend selected is based on the observed trend.

Analysis of SNP Data for Candidate Genes and Gene Finding

S. Stanley Young, Research Triangle Park, USA

The human genome project, gene chip technology and the desire of clinicians will lead to massive data sets where the question is Which genes are associated with disease phenotype, side effects, and drug efficacy? There is a need for statistical methods to address this question. The potential problems are formidable. SNPs in genes and random SNPs will need to be examined giving rise to multiple testing dilemmas. Phenotypes will be described as multivariate observations and can arise by divergent mechanisms so the statistical methods need to be able to handle mixture data. Interactions are expected to abound. Our idea is to borrow and modify data mining methods to address these problems. Multiple testing will be addressed. Genotype/environment interactions will be addressed. Real data and simulated data will be used to explicate methods. Analysis software will be demonstrated. If we are successfull, we can offer guidance to clinicians and help them prescribe “the right medicine for the right patient”.

New statistical approaches for model free linkage analyses of quantitative traits

Andreas Ziegler, Marburg, Germany

There has been a growing interest in model free linkage analysis for quantitative traits in the last decade. Three among them are the Haseman-Elston (H-E) regression approach, the new Haseman-Elston (new H-E) approach and the Weighted Pairwise Correlation (WPC) statistic. In this presentation, the relationship between the H-E approach and the WPC statistic for standardized phenotypic and genetic information for sib pairs is discussed first. Second, the equivalence between the new H-E approach and the WPC method is shown for the same situation. Third, the three methods are illustrated by using sib pair data from chromosome 5 on schistosoma mansoni.

Participants

Prof. Dr. Donald Berry
db@stat.duke.edu
Inst. of Statistics and Decision Sciences & Cancer Center Biostat.
Duke University, Box 90251
223 Old Chemistry Bldg.
Durham, NC 27708-0251
USA

Dr. Heike Bickeböller
bickeboeller@gsf.de
GSF-Inst. für Epidemiologie
Ingolstädter Landstr. 1
85764 Neuherberg

Inke Böddeker
boeddeke@mailer.uni-marburg.de
Institut für Medizinische Biometrie und Epidemiologie
der Philipps-Universität Marburg
Bunsenstr. 3
35037 Marburg

Dr. Jenny Chang-Claude
j.chang-claude@dkfz-heidelberg.de
DKFZ, Abt. Epidemiologie
Im Neuenheimer Feld 280
69120 Heidelberg

Prof. Dr. Roger S. Day
day@zydeco.pci.upmc.edu
U. Pittsburgh Cancer Institute
University of Pittsburgh
3600 Forbes Ave Suite 400A
Pittsburgh, PA 15213
USA

Dr. Peter Dirschedl
dir@ibe.med.uni-muenchen.de
Institut für Medizinische Informationsverarbeitung, Biometrie und Epidemiologie
Marchioninistraße 15
81377 München

Dr. Lutz Edler
edler@dkfz-heidelberg.de
Deutsches Krebsforschungszentrum
Biostatistik 0820
Postfach 101949
69009 Heidelberg

Prof. Dr. Robert C. Elston
rce@darwin.cwru.edu
Dept. of Epidemiology & Biostatist.
Case Western Reserve University
MetroHealth Medical Center
2500 MetroHealth Drive
Cleveland, OH 44109-1998
USA

Prof. Dr. Paul F. Feigin
iedean@ie.technion.ac.il
William Davidson Faculty of Industrial Engineering and Manage.
Technion Israel Institute of Technology
32000 Haifa
ISRAEL

Christine Fischer
christine_fischer@med.uni-heidelberg.de
Institut für Humangenetik
Universität Heidelberg
Im Neuenheimer Feld 328
69120 Heidelberg

Tim Friede
friede@imbi.uni-heidelberg.de
Abt. Medizinische Biometrie
Universität Heidelberg
Im Neuenheimer Feld 305
69120 Heidelberg

Prof. Dr. Olaf Gefeller
gefeller@imbe.med.uni-erlangen.de
Institut für Medizininformatik,
Biometrie and Epidemiologie
Universität Erlangen
Waldstraße 6
91054 Erlangen

Dr. Dirk Hasenclever
dirk.hasenclever@imise.uni-leipzig.de
Institut für Medizinische Inform.
Statistik u. Epidemiologie
Universität Leipzig
Liebigstr. 27
04103 Leipzig

Prof. Dr. Gerhard Hommel
hommel@imsd.uni-mainz.de
Institut für Medizinische Statistik
und Dokumentation
Universität Mainz
Langenbeckstr. 1
55131 Mainz

Stefan Horvath
horvath@imsdd.meb.uni-bonn.de
Institut für Medizinische Statistik
Dokumentation und Datenverarbeitung
Universität Bonn
Sigmund-Freud-Str. 25
53127 Bonn

Dr. Berthold Lausen
berthold.lausen@ic.ac.uk
Dept. of Medical Statistics and
Evaluation, Imperial College School
The Hammersmith Hospital
Du Cane Road
GB-London W12 0NN

Prof. Dr. Walter Lehmacher
walter.lehmacher@medizin.uni-koeln.de
Inst.für Medizinische Statistik
Medizinische Fakultät
Universität zu Köln
Joseph-Stelzmann-Str. 9
50931 Köln

Dr. Hans-Helge Müller
muellerh@mailer.uni-marburg.de
Institut für Medizinische Biometrie
und Epidemiologie
der Philipps-Universität Marburg
Bunsenstr. 3
35037 Marburg

Dr. Bertram Müller-Myhsok
bmm@bni.uni-hamburg.de
Bernhard-Nocht-Institut
Bernhard-Nocht-Str. 74
20359 Hamburg

Prof. Dr. Michael Newton
newton@biostat.wisc.edu
Department of Statistics
University of Wisconsin
1210 W. Dayton Street
Madison, WI 53706
USA

Christopher Portier
portier@niehs.nih.gov
Lab. of Comp. Biology & Risk Analys
National Ist. of Environ. Health
Sciences, MD A3-06
P.O.Box 12233
Research Triangle Park, Nc 27709
USA

Dr. Martin Posch
martin.posch@univie.ac.at
Institut für Medizinische Statistik
und Dokumentation
Universität Wien
Schwarzspanierstr. 17
A-1090 Wien

Michael A. Proschan
proscham@nih.gov
National Heart, Lung and Blood
Institute
Rockledge II Center
Rockledge Drive, MSC 7938
Bethesda, MD 20892-7938
USA

Prof. Dr. Helmut Schäfer
hsimbe@mailer.uni-marburg.de
Institut für Medizinische Biometrie
und Epidemiologie
der Philipps-Universität Marburg
Bunsenstr. 3
35037 Marburg

Prof. Dr. Martin Schumacher
ms@imbi.uni-freiburg.de
Institut für Medizinische Biometrie
und Medizinische Informatik
Klinikum der Universität
Stefan-Meier-Str. 26
79104 Freiburg

Prof. Dr. Francoise Seillier Moiseiwitsch
seillier@bios.unc.edu
Department of Biostatistics
School of Public Health,201 H
University of North Carolina
Chapel Hill, NC 27599-7400
USA

Prof. Dr. Stephen Senn
stephens@public-health.ucl.ac.uk
Dept. of Stat. Science & Epidemiol.
and Public Health
1-19 Torrington Place
GB-London WC1E 6BT

Yu Shen
yushen@odin.mdacc.tmc.edu
Dept. of Biomathematics
M.D. Anderson Cancer Center
Box 237
1515 Holcombe Boulevard
Houston, TX 77030
USA

Richard Simon
rich@brb.nci.nih.gov
NIH
6310 Executive Blvd.
EPN Room 739
Bethesda, MD 20892-0001
USA

Dr. Nigel Stallard
n.stallard@reading.ac.uk
Medical & Pharm. Statistics Res.
The University of Reading
P.O.Box 240
Earley Gate
GB-Reading RG6 6FN

Prof. Dr. Peter F. Thall
rex@mdanderson.org
Dept. of Biomathematics
M.D. Anderson Cancer Center
Box 237
1515 Holcombe Boulevard
Houston, TX 77030
USA

Prof. Dr. Alexander Tsodikov
atsodikov@hci.utah.edu
University of Utah
2000 E. N. Campus Dr
Room 461
Salt Lake City, UT 84112-5550
USA

Prof. Dr. Werner Vach
werner.v@statdem.ou.dk
Institut for Statistik oy
Demografi
Odense University
Campusveg 55
DK-5230 Odense 11

Dr. Gernot Wassmer
gernot.wassmer@medizin.uni-koeln.de
Institut für Medizinische
Dokumentation und Statistik
Universität zu Köln
Joseph-Stelzmann-Str. 9
50937 Köln

Dr. Thomas F. Wienker
wienker@uni-bonn.de
Institut für Medizinische Statistik
Dokumentation und Datenverarbeitung
Universität Bonn
Sigmund-Freud-Str. 25
53127 Bonn

Prof. Dr. Janet Wittes
janet@statcollab.com
Statistics Collaborative
1710 Rhode Island Ave, NW
Suite 200
Washington, DC 20036
USA

Dr. Stanley Young
ssy0487@glaxowellcome.com
Glaxo Welcome Inc.
5 Moore Drive
Research Triangle Park, NC 27709
USA

Dr. Andreas Ziegler
ziegler@mailer.uni-marburg.de
Institut für Medizinische Biometrie
und Epidemiologie
der Philipps-Universität Marburg
Bunsenstr. 3
35037 Marburg