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 $p_1 p_2 \leq \exp[-0.5 \chi^2_{2,1-\alpha}]$, 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


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


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.