ICSA Book Series in Statistics
Series Editors: Jiahua Chen · Ding-Geng (Din) Chen

Karl E. Peace
Ding-Geng Chen
Sandeep Menon *Editors*

Biopharmaceutical Applied Statistics Symposium

Volume 1 Design of Clinical Trials

Springer

Chapter 1
A Statistical Approach to Clinical Trial Simulations

Stephan Ogenstad

1.1 Introduction

Drug development is not for the fainthearted. We have heard repeatedly over the years regarding the process of bringing a new compound to the market, that every delay will add millions of dollars in added expenses and lost revenues. In order to address some of the concerns in this development process is to simulate the potential outcomes of the clinical study. Simulations of clinical trials go by different names, such as clinical trial simulations (CTS), modeling and simulation (M&S), computer-assisted trial design (CATD), model-based drug development (MBDD), and model-informed drug discovery and development (MID_3). CTS is being increasingly viewed as an integral part of clinical development programs and can be used to improve the understanding and decision making at every stage of drug development. These simulations help to develop better insight into the operating characteristic of a specific trial design. CTS provides the ability to test multiple scenarios, predict the potential study outcomes for each scenario and select the most advantageous study design. Hence, before conducting a study, examining various trial designs through computer simulations can help improve the likelihood of a successful study.

In the field of airplane development, already from the beginning of manned flight, there has been a symbiotic relationship between the airplane and simulation in all of its different forms. The role of simulation and flight simulators in airplane development, training and evaluation have evolved significantly over the past 80 years, often in response to technical innovations in both the airplane and ground support systems. In the same spirit, CTS ought to be an integral part of clinical development, in the design of the clinical protocol and in training of staff and investigators.

S. Ogenstad (✉)
Statogen Consulting LLC, 1600 Woodfield Creek Drive, #215, Wake Forest, NC 27587, USA
e-mail: sogenstad@statogen.com

K. E. Peace et al. (eds.), *Biopharmaceutical Applied Statistics Symposium*, ICSA Book Series in Statistics, https://doi.org/10.1007/978-981-10-7829-3_1

In clinical drug development, the process of development is classified into the four Phases I to IV. Phase I studies are frequently conducted in normal healthy subjects (except for the field of cancer where it is usually done in patients), where focus is on identifying tolerable doses, and on learning about what the body does to the drug (pharmacokinetics) and what the drug does to the body (pharmacodynamics), as well as examining if there are potential interactions with other classes of drugs. In Phase IIA the main objective is to evaluate whether or not the drug has initial encouraging efficacy in a small group of patients ('*proof of principle*' or '*proof of concept*'). The goal of Phase IIB is to learn how to use the drug in a larger group of patients for the indication under consideration. This is usually achieved by applying dose ranging, with or without simultaneous measurements of systemic exposure. In Phase III the efficacy and safety of the novel drug should be confirmed against an established treatment. Sometimes Phase IIIB outcome studies are conducted to learn if, for instance, a type-2 diabetes medication has cardiovascular benefits over other type-2 diabetes medications already on the market. In Phase IV the purpose is to accumulate more information on safety and efficacy from several thousands of volunteers who have the disease. Sheiner has viewed clinical development as two major learn-confirm cycles, the Phase I-IIA and the Phase IIB-III cycles (Sheiner 1997; Sheiner and Ludden 1992).

Nevertheless, even if the main objective of a clinical study is confirming, there are several opportunities to learn about variation in pharmacokinetics and pharmacodynamics in patient groups to increase the likelihood of identifying dosing strategies that will result in safe and effective treatment for the individual patient. Clinical trial simulations can be a valuable tool for decision making in drug development by applying diverse types of models. It then consists of three main components: a disease-placebo model, a drug model, and a trial design model. The disease-placebo model is concerned with the time course of the disease, relative risks with respect to morbidities and mortality. The drug model describes the relationship between therapeutic efficacy, toxicities, and doses. The clinical trial design model deals with components such as baseline characteristics (e.g. inclusion/exclusion, actual values the subjects have at baseline), compliance, missing values, endpoints, and statistical methods of analysis. The use of CTS for drug development has been shown to be a cost-effective approach, for instance, the exploration of multiple dosing regimens and their likely pharmacodynamic effects over diverse patient populations (Huang and Li 2007; Ette et al. 2003; Riggs et al. 2007; Holford and Ploeger 2010). Here, simulations provide a means to assess the effects of various loading and maintenance dosing parameters on steady-state concentrations; effects of dosing holidays (period when a patient is not taking the drug) on pharmacodynamics response; etc.

Without thorough planning, pretesting, and execution, the clinical trial implementation risks are high. Thus, optimization of the clinical trial design should be the main focus before starting the study. In the past, clinical trials were designed using ad hoc empirical approaches, where the '*organization*' impatiently desired the clinical trial to commence under the pretense not to lose any valuable time. Because data resulting from the clinical trial is often too complex to allow simple conclusions of what the outcome of the study is, the interest in CTS has been ongoing for the

past two decades (although there have been earlier success stories recounting the value of simulation for design of clinical trials), and has today become a frequently used tool in quantitative pharmacology investigations in academia, regulatory and the biopharmaceutical industry. Current trends within the pharmaceutical industry and within the offices of some regulatory agencies have suggested a reassuring future for clinical trial simulations (Chang 2010, 2014; Kimko and Peck 2010; Westfall et al. 2008; Duffull and Kimko 2002; Holford et al. 2000; Sheiner and Steimer 2000). If CTS is done thoughtfully, Peck et al. (2003) outline an ambitious but possible future that CTS might sometimes replace the second Phase III trial, and therefore only a single trial is needed.

CTS is the generation of biomarker or clinical responses in virtual subjects that take into account (a) the trial design and execution, (b) pathophysiological changes in subjects during the trial (disease-progress model), and (c) pharmacology (drug-intervention model), using mathematical, statistical and numerical methods and models. CTS can be applied in the design, analysis, and interpretation of human clinical drug trials in order to promote key decisions in drug development management and regulatory approval (Kimko and Peck 2010; Holford et al. 2000). The European Medicines Agency (EMA) and the Center for Drug Evaluation and Research (CDER) in the U.S. Food and Drug Administration (FDA) have each issued a number of guidances for drug developers that pertain to the role of CTS in development and regulation. The FDA's 2009 Guidance for Industry: End-of-Phase 2A Meetings urges sponsors to seek regulatory meetings to discuss quantitative modeling and trial simulations to improve dose selection and clinical trial design. Although not solely focused on CTS, these guidances describe standards and expectations concerning regulatory submission (Kimko and Peck 2010). CTS is included in the FDA's published strategic priorities and is expected to be incorporated in the 2017 PDUFA reauthorization.

Hence, CTS supports the project team to minimize risks and guide decision making by formalizing assumptions, quantifying and testing uncertainties. The simulations can be used for defining and testing analysis models, exploring study design properties, and performing analyses about precision and accuracy of potential endpoint estimates. The simulations can incorporate available scientific information to help the entire project team communicate and test ideas, and to plan significant, effective trials for every phase of clinical development. The CTS helps the team anticipate risks and preview the range of expected results before huge investments are allocated. Thus, CTS has the ability to transform drug development by making better use of prior data and information and to explore important clinical trial designs. As a result, the project team can receive swift feedback on the impact on trial outcomes that alternative designs and analysis methods could have presented in the future. CTS can gain credibility with the '*nonscientists*' as the trial design can be made understandable without technical terms and a different kind of reasoning, and can give clearness to otherwise difficult principles influencing opinion and behavior. The statistician has an imperative role to play within their organization and that by using professionally developed trial design software, such as EAST (Cytel Corporation), or if the organization has invested in the writing of their own computer

programs in, for instance, SAS (SAS Institute) or R. With the help of such software they can rapidly generate many alternative design scenarios that accurately address the questions at hand and the goals of the project team, freeing up time for vital discussions about the choice of endpoints, populations, and treatment regimens.

1.2 Protocol Deviations

Before undertaking any clinical research project, a fully developed and vetted study protocol is critical. In the field of clinical development, having a well written and thought out protocol means that we have a detailed plan that is available and consulted frequently during the conduct of the clinical research project and that the investigators and staff are well trained on at following the protocol. Before the clinical trial starts, it is critical that an efficient statistical methodology is selected and implemented in order to effectively analyze the data after database lock where no data is any longer allowed to be altered. The statistician is critical in conceptualizing the analytical methodology that should be used. Ideally, the statistician needs in a blinded fashion to continue to follow the study as the data is being collected and prior to final analysis of the data. It is not uncommon that the data that was planned to be collected, changes for pragmatic and to some unforeseen reasons. This means that the thoughts that go into the statistical analysis plan should if possible have considered the prospect of such changes could become a reality. Protocol deviations should be rare or unexpected if an intense effort has gone into writing the protocol, though unfortunately many times amendments need to modify the protocols. Consequences of protocol deviations on clinical trial outcomes depend on their qualitative and quantitative characteristics. Thus, while the consequence of one type of protocol deviation can be easily evaluated, some are more difficult to discern than others (e.g. noncompliance to treatment). It follows that the combination of several deviations of varying degrees may lead to unexpected consequences on study outcomes. Protocol deviations can result from many different circumstances, where the most critical deviations are noncompliance and missing data and dropped out subjects.

1.2.1 Noncompliance

Noncompliance or non-adherence to treatment protocol occurs when a patient does not carry out the clinical recommendations of a treating physician. In other words, it is the failure of the patient to follow the prescribed treatment regimen and procedures. Important questions are: What are the consequences if patients take fewer or extra doses of treatment medication than prescribed, but the remaining doses are taken on time, or if patients stop taking the treatment but remain on the study? Noncompliance is a significant problem in all patient populations, from children to the elderly. It applies to nearly all chronic disease states and settings and tends to worsen the

longer a patient continues on drug therapy (Spagnoli et al. 1989; Mardonde et al. 1989; Lacombe et al. 1996). Noncompliance rates with schizophrenia treatment could be as high as 40%, with partial noncompliance as high as 75% (Moore et al. 2000).

Noncompliance can result from a denial of the problem. Many diseases and conditions are easy to ignore, even when they have been diagnosed. This is particularly true for diseases that are asymptomatic and so does not bother the patient. For instance, patients with diabetes, or hypertension may not have symptoms that get in the way of everyday life. They may not even have known that they had the condition until it showed up on a routine examination, which can make it easy for patients to ignore the prescribed treatment regimens. The patients may have difficulty with the regimen and may have trouble following the directions. For instance, taking a pill in the middle of the night, or simply opening the '*child safe*' container may create a barrier to compliance for a patient with rheumatoid arthritis.

Bothersome previous experiences with medications prescribed by their physicians may lead the patients not to take their medication. As a consequence, some patients may not take the medication or may take another medication that they have at home for the same diagnosis. Whether the patients tell the investigators or not will cause difficulties interpreting the results and will bias the study results. Reasons for not disclosing to the investigator that the patient is not taking the medication could be that the patient does not want to affect their relationship with the investigator.

1.2.2 Dropouts and Missing Data

A common problem in clinical trials is the missing data that occurs when patients do not complete the study and drop out without further measurements are taken. Possible reasons for patients dropping out of the study could include death, adverse reactions, unpleasant study procedures, lack of improvement, early recovery, and other factors related or unrelated to trial procedure and treatments. Clinical trials that require adherence that is difficult to follow or have an extensive number of endpoints often suffer from missing data or even subject dropouts. The dropout and missing data mechanisms are often complex, and generally, cannot be assumed to be missing at random or missing completely at random (MCAR). More realistically, the missing values depend on patient experience in the trial. In some cases patient dropouts are infrequent with MCAR mechanism; in other cases, dropouts may be related to a lack of safety or efficacy of the patient's experience. There are several possible ways to model the dropout mechanism; some examples and further references are contained in O'Brien et al. (2005). Patient dropout is a real concern for clinical trials and one of the most problematic protocol deviations. Two types of dropouts exist, non-informative and informative dropouts. Non-informative dropouts simply mean that some patients may randomly stop to be reported in the study, this independently from the treatment they received, and thus independently of efficacy or side effects. Non-informative dropout will simply decrease the statistical study power which is easier to control. On the contrary, disease progress can be perceived by the patient in many

ways not measured in the study but, however, correlate with the endpoint that is being followed. In this case, the dropout is informative to the disease progress, and modeling the disease progress separately from the dropout process may be inefficient and may even produce biased estimates. The bias can be particularly notable if one wants to use the model to predict actually observed features, e.g., observed average disease progress. Imputing unobserved data, e.g., last value carried forward is commonly used as a conservative approach to demonstrate treatment differences, though last value carried forward is, however, inferior from a modeling standpoint as the pseudo data are treated as observed data, creating biases (Westfall et al. 2008).

1.3 Methods

Clinical trial simulations can produce a number of advantages that will help us predict likely outcomes for a range of assumptions about trial size, dose selection and operational considerations, such as:***Study specific aspects***

Study specific aspects

- Comparisons of different trial designs where we can evaluate what we might be losing in one aspect of one design in return for gaining another aspect with another design.
- Optimal dosing for each treatment arm to minimize overlap in exposures and subsequent responses.
- Anticipated patient exposures and responses for each treatment.

Improved specification of inclusion/exclusion criteria

- Optimizing inclusion/exclusion criteria to capture the desired subject population that is influencing the response.
- Potential effects of changes in recruitment rates and criteria on study timelines and results.

Safety and efficacy

- Effects of protocol deviations and treatment compliance on safety and efficacy.

Study results

- Placebo effects on patients over time.
- How the investigational treatment compares to the competitors' treatments.

Statistical analysis

- Whether planned study analysis can detect statistical significance.
- Since conventional statistical tests may be insensitive to a wide range of situations occurring commonly in practice, particularly when the effect of the factor under study is heterogeneous, an evaluation of the test can be made where approximations of the test statistic's distribution have been used in the past.

The computer models that simulate real scenarios are generally developed from previous datasets that may include preclinical data, as well as previous phases of real trials. As clearly stated in Burman et al. (2005) the CTS methodology can be summarized in four steps:

1. Utilizing relevant information.
2. Building a mathematical model (usually for the effect of a drug or device).
3. Predicting the outcome of potential clinical trials.
4. Optimizing the clinical trial program.

Before applying the four steps, the aims of the modeling effort must be defined. What is relevant information, what is a good statistical model, and what is an optimal clinical program depends on the aims we have with the model. The modeling is an interactive process between the formulation of the inputs to the model and the actual outcomes from the simulations. The models should include terms for covariate effects, as models used for simulation studies must deal with the variability from individual to individual. Covariate distribution models describe the relevant information that goes into the simulation, on the basis of preceding trials or clinical experience. The variability of patients' demographic and physiological characteristics in the population of interest that might affect the response. Data in clinical trials are naturally correlated and this should be considered. A number of things about the correlation structures can be learned from previous clinical trials. Baseline measurements are typically correlated with the response. Incorporation of them in the analysis will therefore often considerably improve the trial's effectiveness to show potential therapeutic effects. A baseline response model can help to select the target population or to interpret the trial data. Increasing the number of repeated measurements at baseline and at the end of the treatment period for each subject in a clinical trial will obviously increase the available information on treatment effects and could increase the statistical efficiency of the analysis (Frison and Pocock 1992; Ogenstad 1997). The most efficient way to allocate visits over time at the design stage (e.g., before or after randomization), and the best way to utilize the additional measurements from these visits at the analysis stage is not evident but could be explored via simulations. A model of the baseline response and the variability in the measurements can predict how much the gain would be in terms of efficiency, and could for instance influence the decision on whether the inclusion/exclusion criteria should be modified or not. The impact of the different covariate distributions on the expected outcome of a simulated trial can be assessed, which makes it possible to explore conditions that have been ruled out in the inclusion/exclusion procedures of the actual trial.

As Burman et al. (2005) point out, what information is relevant for the CTS largely depends on what the aims of the modeling are. It also depends on how much information is already available. The best information is perhaps hard endpoint data for the drug in question from a large, randomized, placebo-controlled clinical trial. Unfortunately, this kind of data is seldom available before the end of the clinical program, at the earliest. Hence, what we are concerned with is combining information from diverse sources and incorporating expert judgment in a nonbinding way, and remembering that not all experts are right all the time.

The goal of model building is to establish a model that is fit for the purpose, and not made too involved in order to fulfill the purpose of the design of the clinical study. We need to unify the thinking about the study design and inference. The CTS should make the design and future conduct of the study easier to understand. When a clinical trial is planned, it is supposed that the trial will be executed according to a specific protocol that defines all aspects of the study design, from its beginning to its completion. For instance, characteristics that should be precisely defined in any clinical protocol are whether the subjects are patients or healthy volunteers, inclusion/exclusion criteria, number of subjects to be accrued, treatments and allocation mechanism, blinding of investigators or subjects to the allocated treatment, dosage regimen (dose and timing of doses), endpoints, frequency of follow-up evaluations, and the length of the study.

Complete adherence to the study protocol will permit unbiased estimation of the treatment effects in terms of safety and efficacy with adequate statistical power if the assumptions at the planning stage were correct. Deviations from the protocol may lead to failure of the study to attain its declared purposes. It can be difficult at the planning stage to evaluate what the consequences are of a single protocol deviation, and almost impossible to do it for a combination of protocol deviations. One way to quantify the consequences of those deviations is by using models, describing individual behaviors and responses, combined with trial simulations that include these protocol deviations. When the results of the trial can be envisaged it is sometimes possible to choose, in a methodical and cogent way, between different possible trial designs. The features that are included in the model will unveil what design features can be compared using that model. Missing data cause the usual statistical analysis of complete or all available data to be subject to bias and will diminish the power of the study. Although there are a number of imputation methods, there are no universally applicable methods for handling missing data that will restore the dataset to what it could have been if no data had been missing. As has been noted in the ICH-E9 guideline, '*no universally applicable methods of handling missing values can be recommended*'. The issue of managing missing data is intrinsically difficult because it requires a large proportion of missing data to investigate a method. Moreover, a large proportion of absent data would make a clinical study less credible. The best suggestion is to minimize the chance of dropouts at the design stage and during trial monitoring. It should be reiterated that although an increase in a number of patients to the study will decrease the standard errors, but will not correct the bias that could have been caused due to the missingness of data.

Examples of other features that can be compared are study designs (e.g. sequential, adaptive, crossover, parallel), doses, dosing schedules, study duration, different endpoints, multiple endpoints and timing when the endpoints are measured. The data that is generated from the simulations, together with different statistical methods of analysis of the data may lead to an optimization of the whole study process.

When the purpose of the simulation is to estimate the powers of the statistical tests by the relative number of statistical significances it produces, it is important to use an adequate number of simulations. With 1000 simulations and a power around 90%, the estimation error is approximately 3.7% using a 95% confidence interval. Due to the propagation of uncertainty in the square of a quotient, the uncertainty in the

power estimates translates to an uncertainty in the relative efficiency of the tests in the order of 13–16%. With 10,000 simulations the estimation error is approximately 1.2% and the uncertainty in the relative efficiency of the order of 5%. In pursuance of getting reliable estimates of the true significance level, we recommend simulation sizes around 25,000.

1.4 The Clinical Trial Simulation System

We argue that the CTS system should be flexible, preserving the realism of the doubly-multivariate endpoint/timepoint correlation structures, the informative dropout mechanisms, non-normal distributions, non-monotonic hazard rate functions, survival endpoints, and noncompliance effects (Westfall et al. 2008). The assumptions should be a trade-off between ease of use of the system and realism and flexibility of its outputs. This type of framework for multivariate simulation is usually reasonably simple to program where a variety of software can be used, e.g. SAS (SAS Institute), R (R Foundation for Statistical Computing), SPLUS (TIBCO Software Inc.), Mathematica (Wolfram Research), and MATLAB (MathWorks, Inc.).

From literature, the goal with CTS is often to build a complete model. In Holford et al. (2000) their review on simulations in clinical trials, they state that the model should incorporate all scientific knowledge about the disease and drug. Burman et al. (2005) take a more modest view, where model components should be chosen according to the fit-for-purpose principle. We are convinced that simpler models can sometimes be very useful. Decisions where it may be useful cover a wide range of aspects, including choice of the drug candidate, stop/go for the further development of a compound, choice of patient population, and decisions regarding the positioning versus marketed competitor compounds.

In the PK/PD-phase of drug development, the introduction of population modeling has made it possible through the application of statistical non-linear mixed-effects models to data obtained from relatively few samples in many individuals to discern a genuine insight into the mechanistic aspects (Sheiner and Ludden 1992). More specifically, population models allow characterization of (a) mean pharmacokinetic/pharmacodynamic parameters, (b) extent of variability in these parameters and the sources thereof (e.g. gender, age, disease, comedication), and (c) relationships between pharmacokinetic (e.g. exposure) or pharmacodynamic (e.g. a biomarker) variables and clinical efficacy and safety endpoints. These models can then be used to simulate the outcomes of various trial designs under different assumptions. The usefulness of modeling and simulation in the PK/PD-phase of drug development and regulatory decision-making has been recognized (Holford 1990; Sheiner and Steimer 2000; Holford et al. 2000; Nestorov et al. 2001; Gobburu and Marroum 2001; Gobburu and Sekar 2002; Bhattaram et al. 2005; Burman et al. 2005). Exposure-response models may, for example, be used to support the use of a drug in new target populations through bridging, dose adjustment or no need for dose adjustment in subpopulations, new dose regimens, dosage forms and formulations, routes of administration,

and minor product changes (FDA Guidance for Industry 2003). A biological marker (*biomarker*) has been defined as a characteristic that is objectively measured and evaluated as an indicator of normal biological processes, pathogenic processes, or pharmacologic responses to a therapeutic intervention (Biomarkers Definition Working Group 2001). The most reliable way to assess the benefit and risk of a drug therapy is through its effect on well-defined clinical endpoints. However, this approach is sometimes impractical for the evaluation of long-term disease therapies and trials that require a large number of patients. A biomarker may then be substituted for clinical response, provided that it is reasonably likely to predict clinical benefit (FDA 1997). However, the single most important use of biomarkers is the selection of the dose range and doses for further investigation in the pivotal trials (Jadhav et al. 2004). To further facilitate the identification of optimal dosing regimens, the use of clinical utility functions has been proposed (Sheiner and Melmon 1978; Eriksen and Keller 1993; Graham et al. 2002; Jonsson and Karlsson 2005). Such functions serve to evaluate important desirable and undesirable effects of a drug on the same scale, under different assumptions of the relative severity of each outcome. In this way, the observed or predicted clinical outcome of different drug therapies, or different dosing regimens of the same drug, may be compared.

An appealing approach to building a statistical CTS system is found in (Westfall et al. 2008). Their approach starts with a model with a rich probabilistic structure to account for typical scenarios, using historical data where it is possible to validate the inputs and outputs, with specific emphasis on the economical yet flexible input of correlation structures. Here, patient responses are functions of underlying correlated $N(0, 1)$ clinical quantities; all distributional forms and dropout effects are determined from these underlying values. Evaluation of trial success then follows from the analysis of the simulated datasets. The goal is to generate realistic datasets having typical correlation structures for multiple endpoint/timepoint data with, say p, endpoints (safety, efficacy or both) indexed by $j = 1, \ldots, p$, and $T + 1$ timepoints indexed by $t = 0, \ldots, T$, where $t = 0$ can be the time of randomization of the patient. For patient i a $p(T + 1)$-vector of correlated $N(0, 1)$ variates Z_{ijt}, each of which may be thought of as a latent indicator of the patient's health relative to a population of similar patients, for endpoint j and timepoint t. Observations will be considered to be independent for different patients. Though, it is possible to include correlations, for instance, for random center effects. Obviously, for each specific patient the timepoint data $Z_{ijo}, \ldots, Z_{ijT}$ are correlated. For instance, the compound symmetry covariance structure model can be expanded easily to accommodate time-series carryover effects in addition to patient effects as $Z_{ijt} = \sqrt{\theta}S + \sqrt{1-\theta}\varepsilon_{ijt}$, where $S \sim N(0, 1)$ is the patient effect and $\varepsilon_{ij0}, \ldots, \varepsilon_{ijT}$ is a realization of a unit variance AR(1) process with parameter p. For simulation purposes, the parameters θ and p must be specified. For multiple endpoint data for patient and timepoint, it is suggested that the correlation between endpoints is best left as unstructured. For each patient, the observations between endpoints at different timepoints are correlated. There are a number of possibilities for defining this structure, the most convenient and well-known is the Kronecker product model used in multivariate longitudinal models (Westfall et al. 2008).

Most commercially available clinical trial software systems use parametric input into the systems. For instance, the exponential survival model is often used as input model. Though, the exponential survival model is a rather unrealistic model since it is assumed that the hazard rate function is constant over the entire observational study period. The Weibull model is many times a better choice than the exponential, but this model still has a monotonic hazard rate function, which might not be realistic either. A more flexible approach is to use Royston-Parmar models (Royston and Parmar 2002) that have great flexibility. Even better at times is to use mean structures as input for the different endpoint*timepoint*treatment combinations. Such structures can be determined purely a priori from earlier phase data, suggested by PK/PD models, or from studies on similar interventions. Survival analyses pose additional questions. Standard methods such as the log-rank test and Cox models are efficient when the hazards are proportional. This assumption is not always reasonable. The non-proportional hazards assumption that is a potential difficulty with the Cox model, could sometimes be handled in a simpler way, and the visualization of the hazard rate function could be made easier, using the Royston–Parmar framework. In Westfall et al. (2008), any types of distributions could be applied to the mean structures, and there they effectively made use of a missing value, dropout and noncompliance mechanism to generate '*real world datasets*'. Girard et al. (1998) developed a hierarchical Markov model for patient compliance with oral medications conditional upon a set of individual-specific nominal daily dose times and individual random effects that are assumed to be multivariate normally distributed. This model also has great flexibility and allows descriptions of almost all possible compliance profiles.

1.5 Some Published Clinical Trial Simulations

Wathen and Thall (2008) presented a new approach to the problem of deriving an optimal design for a randomized group sequential clinical trial based on right-censored event times. They were motivated by the fact that, if the proportional hazards assumption is not met, then a conventional design's actual power can differ substantially from its nominal value, and combined Bayesian decision theory, Bayesian model selection, and simulation to obtain a group sequential procedure that maintains targeted false-positive rate and power, under a wide range of true event time distributions. At each interim analysis, the method adaptively chooses the most likely model and then applies the decision bounds that are optimal under the chosen model. A simulation study comparing this design with three conventional designs showed that, over a wide range of distributions, their proposed methods perform at least as well as each conventional designs, and in many cases, it provides a much smaller trial.

Dragalin et al. (2010) presented a simulation study to compare new adaptive dose-ranging design. The main goals in an adaptive dose-ranging study are to detect dose-response, to determine if any doses meet clinical relevance, to estimate the dose-response, and then to decide on the dose(s) (if any) to take into the confirmatory Phase III. Adaptive dose-ranging study designs may result in power gains to detect dose-

response and higher precision in estimating the target dose and the dose response curve.

Kimko et al. (2000) simulated the anticipated results of a Phase III clinical trial of the antischizophrenic drug, quetiapine, based on input-output and covariate distribution models developed using data collected in earlier Phase I and II trials. The model development was performed using the NONMEM program with first order conditional estimation (Beal and Sheiner 1992). The proposed trial design was a double-blind, placebo-controlled, randomized, parallel group study of fixed-dose of quetiapine in hospitalized schizophrenic patients, who received one of five doses of quetiapine or placebo for a period of four weeks. The treatment was initiated after a placebo run-in period followed by a two week step-wise dose titration period. The executed study design was replicated by excluding individuals wrongly included in the study, as they failed to meet the entry criteria. In addition, placebo responders identified during the placebo run-in period were replaced. A random dropout algorithm using a multiplicative congruential method (such that the random number generated is the remainder of a linear transformation of the previous number divided by an integer) was used to simulate the high dropout rate observed in the earlier Phase II study. Based on the Phase II study result, 70% of the patients assigned to the placebo group, 60% assigned to the lowest dose group and 50% assigned to all other dose groups were withdrawn from the study. Simulations were performed for 100 sets of 50 patients per treatment group. Adequacy of the model to describe the original data was tested using sensitivity analysis and by comparing posterior parameter distributions and posterior predictions from the simulated trial design to parameters of the prior distribution and observed data. Dropout rates in the simulation and in the Phase III trial were comparable. Comparison of the simulated results with actual results obtained in the Phase III trial showed that the model adequately predicted responses to quetiapine. However, it was found to be inadequate in predicting the placebo response.

Clinical trial simulation for docetaxel was performed using pharmacokinetic/pharmacodynamic models previously developed from data obtained in earlier open-label, non-randomized, Phase II clinical trials of docetaxel in subjects with small cell lung cancer. The purpose of the simulation was to predict the influence of dose on survival time and time to disease progression in a high-risk group in a planned Phase III trial comparing doses of docetaxel of 100–125 mg/m^2 every three weeks. Input-output and covariate distribution models were developed using the NONMEM program. Hazard models were used to simulate the primary and secondary clinical endpoints, death and disease progression, respectively. In addition, the execution model included a separate hazard model for patient dropout. Different models were tested and the Weibull distribution was selected based on the goodness of fit assessed in the model-building phase of the analysis. A dose titration algorithm allowed for a 25% dosage reduction in the event of severe toxicity for each treatment cycle. To maintain consistency with study implementation, after two dosage reductions or if disease progression occurred, the patient was withdrawn from the study. Simulations were performed for 100 sets of subjects and the results were analyzed using SAS. Adequacy of the model to describe the Phase II data was tested using a posterior

predictive check of the following test quantities: number of deaths and progressions, median survival time, I year survival, median time to progression, patient characteristics at baseline, number of side-effects at the end of the first cycle, number of treatment cycles per patient and total dose. Tabulated median and 95% confidence intervals of simulated test quantities agreed well with those obtained from the original data. In addition, 100 sets of 200 subjects per treatment group were simulated under the Phase III trial design and test quantities were calculated. The results of the Phase III trial simulation showed no clinical advantage of the higher docetaxel dose on survival or time to disease progression in high-risk subjects with small cell lung cancer. As a consequence of this analysis, it was determined that there would be no further clinical studies to evaluate the effect of dose intensification in subjects with small cell lung cancer.

1.6 Commercially Available Trial Design Software Packages

Performing simulations with most currently available simulation tools is an investment of time, requiring custom programming and at times moving between one software application to perform simulations and another application to visualize simulations. There is a great need for even more efficient simulation systems that facilitate interactive, real-time evaluation and iteration on simulation scenarios.

As indicated earlier, more adaptations give the investigator more flexibility in identifying best clinical benefits of the test treatment under investigation. However, multiple adaptive designs with more adaptations could be very complicated and consequently, appropriate statistical methods for assessment of the treatment effect may not be available and are difficult, if not impossible, to obtain. Thus, one of the major obstacles for implementing adaptive design methods in clinical trials is that the appropriate statistical methods are not well established with respect to various adaptations. Though, some practical methods in this field are emerging (Gao et al. 2013; Pong et al. 2010). Current software packages such as SAS cannot be applied directly and hence are not helpful here. Although there are some software available in the marketplace such as ExpDesign Studio (http://www.ctrisoft.net), EastSurvAdapt (Cytel Corporation), and ADDPLAN (http://www.addplan.com), which cover certain types of adaptive trial designs, new software packages for adaptive design methods in clinical trials are necessary to assist in implementing adaptive trial designs in clinical trials (Wassmer and Vandemeulebroecke 2006). An overview of software available for group sequential and adaptive designs can also be found in Herson (2009).

Some software (e.g., Certara, https://www.certara.com/software/; Lixoft, http://lixoft.com/) require PK/PD input as drivers for the simulation output. A well developed system is found with EAST 6 from Cytel Corporation that has a large variety of parametric design choices. Another software that produces data with '*flexible*'

statistical characteristics, which helps the decision making that statisticians typically must make is developed by Westfall et al. (2010).

Concerning the description of virtual patients, i.e. the distribution of covariates in a target population, general-purpose statistical packages can be employed. Note that, since IO models usually include terms for covariate effects, the choice of methodology for generating virtual subjects is often dependent on the software for IO modeling. Mouksassi et al. (2009) use the R package library GAMLSS, which facilitates the simulation of demographic covariates specific to the targeted patient populations. Other authors (Chabaud et al. 2002) prefer to resample patients from existing epidemiological databases rather than creating realistic virtual subjects.

The R software environment (by R Core Team 2014) has an excellent set of tools for analyzing and visualizing simulation results in real time. The new RxODE package facilitates quick and efficient simulations of ordinary differential equation (ODE) models in R. RxODE provides an elegant, efficient, and versatile way to specify dosing scenarios, including multiple routes of administrations within a single regimen, sampling schedules, etc. It also enables simulations with between-patient variability and minimizes the amount of custom coding required for pharmacometrics simulations (Wang et al. 2015).

A system specifically designed for IO-modeling of data in this context are the non-linear mixed-effect model program NONMEM (developed by Stuart L. Beal and Lewis B. Sheiner in the late 1970s at UCSF for population pharmacokinetic modeling). It is still widely used.

ADAPT (Biomedical Simulations Resource (BMSR) in the Department of Biomedical Engineering at the University of Southern California) is a computational modeling platform developed for PK/PD applications. It is intended for basic and clinical research scientists and is designed to facilitate the discovery, exploration and application of the underlying pharmacokinetic and pharmacodynamic properties of drugs, which includes an extensive library of models to choose from.

MATLAB (MathWorks) is a multi-paradigm numerical computing environment and fourth-generation programming language. MATLAB allows matrix manipulations, plotting of functions and data, implementation of algorithms, creation of user interfaces, and interfacing with programs written in other languages, including C, C++, C#, Java, Fortran and Python. MATLAB provides a software tool, the so-called SimBiology, for the complete PK/PD workflow. Since Sim-Biology is based on MATLAB, users can employ MATLAB in order to program their simulations.

Mathematica (Wolfram Research) is a quality symbolic computation system. For clinical trial simulations SystemModeler is excellent for modeling and analysis throughout drug discovery, development, clinical trials, and manufacturing. The flexible environment supports application areas such as systems biology, bioinformatics, and more.

Mathematica and MATLAB are very different products. Mathematica focuses on quality symbolic computation and features like unlimited precision arithmetic. MATLAB focuses on high speed algorithms for numerical computation.

1.7 Discussion

In the past few years, scientific journals covering clinical pharmacology and pharmacokinetics and trials in later phases have published a large number of papers related to CTS. The interest in CTS in statistical literature within statistical university departments has been much lower. Still, statistics and statisticians are needed in CTS activities. By writing this paper, we would like to stimulate more statisticians to take an active part in applied modeling work and research related to CTS.

Some of the examples given have hopefully shown that even quite simple modeling exercises can prove very useful. One task for the modeler is precisely that of finding those questions where a limited amount of work is likely to give significant benefits. It might be hard for some statistical scientists to accept that being too rigorous may be harmful. The model need not be perfect. What matters is that the work is good enough to help make the right decisions.

Even though practical modeling work may sometimes be "quick and dirty", rigorous statistical research is needed in the CTS area. We would especially like to point out the need to apply and integrate different areas within statistics and to integrate statistical results into other disciplines, such as pharmacometrics and pharmacoeconomics.

CTS integrates expert knowledge in the relevant fields (primarily pharmacology and medicine in the clinical phase) with new data in a structured process to create quantitative models. The cooperation between different skills is thus essential. Some modeling work can, of course, be done by a single individual. In many situations, however, the greatest benefits are likely to result from a joint collaboration with several skills working in concert (e.g., Biomarkers Definition Working Group). What skills to include in the modeling team is, of course, depending on the modeling questions. Good organization is critical both internally in the modeling team and for the team's relations with decision makers and experts from different parts of the research organization.

CTS aims at optimizing a clinical development program. This program, however, is not totally isolated from the rest of drug development and commercialization. What is 'optimal' in clinical development depends on factors such as the medical need for a new treatment, its commercial value, the regulatory requirements, and the ability to find patients and produce tablets in time for the clinical trials etc. CTS should therefore not be seen as separate from other modeling activities. Preclinical, epidemiological and commercial models could provide useful input to CTS. The results of CTS, on the other hand, may be of great value for predicting market penetration and sales.

Execution models are used to examine the influences of protocol deviations on study outcomes. When implemented as a part of a clinical trial simulation, they allow "virtual" clinical trials to the run under varying conditions, from simple errors in data gathering to complex combinations of protocol deviations that emulate real-world situations. Thus, execution models arc powerful tools for identifying weaknesses or limitations in a proposed study design, which may be anticipated, avoided or resolved

in order to increase the robustness of the study design prior to implementation of the actual clinical study. As such, they are an integral component of clinical trial simulation and essential tools for identifying weaknesses or limitations in a proposed study design, which may be anticipated, avoided or resolved in order to increase the robustness of the study design prior to implementation of the actual clinical study. As such, they are an integral component of clinical trial simulation and an essential tool in clinical trial design. Execution models for protocol departures do not necessarily require data to be identified, except for dropout. Many trials can be performed in simulators that are just too risky in real life and they can be repeated multiple times. Simulators tend to prevent trial failures or overpowered studies by their ability to point what part of the experiment is the most sensitive to protocol departures. Indeed clinical trial simulation provides an invaluable tool to prospectively force experimental study designs to the point of failure.

References

Beal, S. L., & Sheiner, L. B. (1992). *NONMEM user's guide*. San Francisco: NONMEM Project Group, University of California, San Francisco.

Bhattaram, V. A., Booth, B. P., Ramchandani, R. P., Beasley, B. N., Wang, Y., Tandon, V., et al. (2005). Impact of pharmacometrics on drug approval and labeling decisions: a survey of 42 new drug applications. *American Association of Pharmaceutical Scientists Journal, 7,* E503–E5122005.

Biomarkers Definition Working Group. (2001). Biomarkers and surrogate endpoints: Preferred definitions and conceptual framework. *Clinical Pharmacology and Therapeutics, 69*(3), 89–95.

Burman, C. F., Hamren, B., & Olsson, P. (2005). Modelling and simulation to improve decision-making in clinical development. *Pharmaceutical Statistics, 4*(1), 47–58.

Chabaud, S., Girard, P., Nony, P., & Boissel, J. P. (2002). Clinical trial simulation using therapeutic effect modeling: Application to ivabradine efficacy in patients with angina pectoris. *Journal of Pharmacokinetics and Pharmacodynamics, 29*, 339–363.

Chang, M. (2010). *Monte Carlo simulation for the pharmaceutical industry: Concepts, algorithms, and case studies*. Chapman & Hall/CRC Biostatistics Series.

Chang, M. (2014). *Adaptive design theory and implementation using SAS and R*. CRC Press.

Dragalin, V., Bornkamp, B., Bretz, F., Miller, F., Padmanabhan, S. K., Patel, N., et al. (2010). A simulation study to compare new adaptive dose-ranging designs. *Statistics in Biopharmaceutical Research, 2,* 487–512.

Eriksen, S., & Keller, L. R. (1993). A multiattribute-utility-function approach to weighing the risks and benefits of pharmaceutical agents. *Medical Decision Making, 13*(2), 118–125.

Ette, E., Godfrey, C., Ogenstad, S., & Williams, P. (2003). Analysis of simulated clinical trials. In H. Kimko, & S. Duffull (Eds.), *Simulation for designing clinical trials: A pharmacokinetic/pharmacodynamic modeling perspective*. New York: Marcel Dekker.

Food and Drug Administration. (1997). *The Food and Drug Modernization Act of 1997, in Code of Federal Regulations Title 21 Part 314 Subpart H*.

Food and Drug Administration, Guidance for Industry. (2003). *Exposure-response relationships—Study design, data analysis, and regulatory applications* (pp. 1–25). Rockville, MD.

Frison, L., & Pocock, S. J. (1992). Repeated measures in clinical trials: Analysis using mean summary statistics and its implications for design. *Statistics in Medicine, 11,* 1685–1704.

Gao, P., Liu, L., & Metha, C. (2013). Exact inference for adaptive group sequential designs. *Statistics in Medicine*.

Gobburu, J. V. S., & Marroum, P. J. (2001). Utilisation of pharmacokinetic/pharmacodynamic modeling and simulation in regulatory decision-making. *Clinical Pharmacokinetics, 40*(12), 883–892.
Gobburu, J. V. S., & Sekar, V. J. (2002). Application of modeling and simulation to integrate clinical pharmacology knowledge across a new drug application. *International Journal of Clinical Pharmacology and Therapeutics, 40*(7), 281–288.
Graham, G., Gupta, S., & Aarons, L. (2002). Determination of an optimal dosage regimen using a bayesian decision analysis of efficacy and adverse effect data. *Journal of Pharmacokinetics and Pharmacodynamics, 29*(1), 67–88.
Herson, J. (2009). *Data and safety monitoring committees in clinical trials*. Boca Raton: Chapman & Hall/CRC.
Holford, N. H. G. (1990). Concepts and usefulness of pharmacokinetic/pharmacodynamic modeling. *Fundamental & Clinical Pharmacology, 4*(Suppl. 2), 93S–101S.
Holford, N., & Ploeger, B. A. (2010). Clinical trial simulation: A review. *Clinical Pharmacology and Therapeutics, 88,* 166–182.
Holford, N. H. G., Kimko, H. C., Monteleone, J. P. R., & Peck, C. C. (2000). Simulation of clinical trials. *Annual Review of Pharmacology and Toxicology, 40,* 209–234.
Huang, X., & Li, J. (2007). Pharmacometrics: The science of quantitative pharmacology. *American Journal of Pharmaceutical Education, 71,* 75.
Jadhav, P. R., Mehta, M. U., & Gobburu, J. V. S. (2004). How biomarkers can improve clinical drug development. *American Pharmaceutical Review, 7*(3), 62–64.
Jonsson, S., & Karlsson, M. O. (2005). Estimation of dosing strategies aiming at maximizing utility or responder probability, using oxybutynin as an example drug. *European Journal of Pharmaceutical Sciences, 25*(1), 123–132.
Kimko, H., & Duffull, S. B. (2002). *Simulation for designing clinical trials: A pharmacokinetic-pharmacodynamic modeling perspective*. CRC Press.
Kimko, H., & Peck, C. C. (2010). *Clinical trial simulations: Applications and trends*. Springer.
Lacombe, P. A., Vicente, J. A. G., Pages, J. C., & Morselli, P. L. (1996). Causes and problems of non response or poor response to drugs. *Drugs, 51,* 552–570.
Maronde, R. F., Chan, L. S., Larsen, F. J., Strandberg, L. R., Laventurier, M. F., & Sullivan, S. R. (1989). Underutilization of antihypertensive drugs and associated hospitalisation. *Medical Care, 27,* 1159–1166.
Moore, A., Sellwood, W., & Stirling, J. (2000). Compliance and psychological reactance in schizophrenia. *British Journal of Clinical Psychology*.
Mouksassi, M. S., Marier, J. F., Cyran, J., & Vinks, A. A. (2009). Clinical trial simulations in pediatric patients using realistic covariates: Application to teduglutide, a glucagon-like peptide-2 analog in neonates and infants with short-bowel syndrome. *Clinical Pharmacology & Therapeutics*.
Nestorov, I., et al. (2001). Modeling and simulation for clinical trial design involving a categorical response: A phase II case study with naratriptan. *Pharmaceutical Research, 18*(8), 1210–1219.
O'Brien, P. C., Zhang, D., & Bailey, K. R. (2005). Semi-parametric and non-parametric methods for clinical trials with incomplete data. *Statistics in Medicine, 24,* 341–358.
Ogenstad, S. (1997). Analysis and design of repeated measures in clinical trials using summary statistics. *Journal of Biopharmaceutical Statistics, 7*(4), 593–604.
Peck, C. C., Rubin, D. B., & Sheiner, L. B. (2003). Hypothesis: A single clinical trial plus causal evidence of effectiveness is sufficient for drug approval. *Clinical Pharmacology & Therapeutics, 73,* 481–490.
Pong, A., & Chow, S. C. (2010). *Handbook of adaptive designs in pharmaceutical and clinical development*. Taylor & Francis, New York: Chapman and Hall/CRC Press.
R Core Team. (2014). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. [Software] (Version 3.1.1). http://www.R-project.org/.
Riggs, M. M., Godfrey, C. J., & Gastonguay, M. R. (2007). Clinical trial simulation: efficacy trial. In E. I. Ette, & P. J. Williams (Eds.), *Pharmacometrics: The science of quantitative pharmacology* (pp. 881–900). Hoboken, NJ: Wiley.

Royston, P., & Parmar, M. K. B. (2002). Flexible parametric proportional-hazards and proportional-odds models for censored survival data, with application to prognostic modelling and estimation of treatment effects. *Statistics in Medicine, 21,* 2175–2197.

Sheiner, L. B. (1997). Learning versus confirming in clinical drug development. *Clinical Pharmacology and Therapeutics, 61*(3), 275–291.

Sheiner, L. B., & Ludden, T. M. (1992). Population pharmacokinetics/dynamics. *Annual Review of Pharmacology and Toxicology, 32,* 185–209.

Sheiner, L., & Melmon, K. (1978). The utility function of antihypertensive therapy. *Annals of the New York Academy of Sciences, 304,* 112–127.

Sheiner, L. B., & Steimer, J.-L. (2000). Pharmacokinetic/pharmacodynamic modeling in drug development. *Annual Review of Pharmacology and Toxicology, 40*(1), 67–95.

Spagnoli, A., Ostino, G., & Borga, A. D. (1989). Drug compliance and unreported drugs in the elderly. *Journal of the American Geriatrics Society, 37,* 619–624.

The International Conference on Harmonization (ICH) http://www.ifpma.org/ich1.html.

Wang, W., Hallow, K. M., & James, D. A. (2015). A tutorial on RxODE: Simulating differential equation pharmacometric models in R. *American Society for Clinical Pharmacology and Therapeutics*. CPT: Pharmacometrics & Systems Pharmacology.

Wassmer, G., & Vandemeulebroecke, M. (2006). A brief review on software development for group sequential and adaptive designs. *Biometrical Journal, 48,* 732–737.

Wathen, J. K., & Thall, P. F. (2008). Bayesian adaptive model selection for optimizing group sequential clinical trials. *Statistics in Medicine, 27*(27), 5586–5604.

Westfall, P. H., Tsai, K., Ogenstad, S., Tomoiaga, A., Moseley, S., & Lu, Y. (2008). Clinical trials simulation: A statistical approach. *Journal of Biopharmaceutical Statistics, 18,* 611–630.

Westfall, P. H., Dmitrienko, A., DeSouza, C., & Tomoiaga, A. (2010). *Clinical trials simulation: A publicly available, grid-enabled, GUI-Driven SAS® System*. SAS Global Forum 2010, Paper 180-2010.