Alessandra Giovagnoli and Maroussa Zagoraiou

Simulated Clinical Trias: some design issues

Quaderni di Dipartimento

Serie Ricerche 2010, n.1
ISSN 1973-9346

ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

Dipartimento di Scienze Statistiche “Paolo Fortunati”

1. Introduction

Simulation is a formidable tool to aid and complement real experimentation. It presupposes the availability of a "simulator", i.e. a computer code that can be run to imitate the behaviour of the system of interest. Simulators make it possible to explore complex relationships between input and output variables and can be used in settings where physical experimentation is impossible, such as rare event risk assessment. They are also invaluable when only a few physical runs can be made due to their high cost. For these reasons the practice of complementing laboratory experiments or field observations with simulated ones has been steadily growing in recent years, and a large number of scientific problems in the aerospace industry and other engineering set-ups, as well as in finance, marketing and several other disciplines, nowadays are explored with the aid of computer simulators. The books by Santner, Williams and Notz (2003) or Fang, Li and Sudjianto (2005) provide a useful introduction. In a recent conference dedicated to computer experiments Steinberg (2009), starting from applications, reviews some of the main ideas that have been proposed for the statistical analysis and design of studies that use computer simulators, including a brief mention of validation of the simulator by means of real data.

In this paper we intend to take a look at computer simulation in the context of clinical trials, paying special attention to the design aspects. One of the characteristic features of clinical trials is the well-known "individual-versus-collective ethics" dilemma. Potential harm to the subjects must be minimized, especially when they are patients presently under care and at the same time the trial must maximize the experimental information for the sake of future patients. As well as the ethical considerations, time and costs are also important. According to the Phrma 2009 Annual Report (2009), the complex process of researching and developing new medicines takes an average of 10 to 15 years and can cost $1.2 to $1.3 billion. Besides, only an average of 1 in 5 new drugs gets approved for general use. The burden of paying for all necessary people and services is usually borne by the pharmaceutical or biotechnology company that wants to develop the agent under study. Over the last decade

the pharmaceutical industry has encountered a vast number of failures of trials for drug development due to inadequate trial design, wrong statistical analysis, mistakes in the chosen dosages etc. Thus, to bring down the costs, prevent possible failures in future trials, reduce the trial time frame and avoid possible side effects in humans, researchers and drug companies have started to perform virtual experiments. Clinical trial simulation is asserting itself as an emerging technique, thanks also to the advent of new powerful software tools. The U.S. Food and Drug Administration have recommended that the industry expand and accelerate development of simulated clinical trials (FDA, U.S. Department of Health and Human Services, 2004). An important contribution is the collective volume edited by Kimko and Duffull (2003) which gives a general overview of simulation for clinical trials presenting a large number of case studies (see also Taylor and Bosch, 1990 and Holford et al., 2000).

Despite understandable misgivings in non-experts, the idea that the functioning of the human body can be mathematically modelled and analyzed has been widely accepted in the scientific community at least since the second half of last century. Mathematical models and numerical methods are used to approximate physiological functioning, disease progress and drug behaviour in the human body, thus making computer simulation possible in the pharmaceutical/biomedical field too. However, simulation studies require proper protocols just like real trials, but at present a theory of simulated experimental design seems to be lacking. It is up to the medical statisticians to meet this challenge, and develop appropriate methodological tools.

This paper is mainly of a review character and is entirely based on recently published research. The statistical issues involved in a simulated trial that we present are

- planning the simulation,
- modelling,
- experimental design,
- validation of the simulator.

We begin in Section 2 with a selection of simulated trials from the medical literature. In Section 3 we examine the potential aims of a

simulation experiment. Simulation models used in the clinical context are examined in Section 4. In Section 5 we discuss the ensuing experimental design problems. Section 6 contains a brief introduction to the use of metamodels in medicine. Section 7 is dedicated to the issue of validating the simulator of a virtual trial, and the final section contains our comments.

2. Applications of Simulated Trials

The goal of this section is to show by means of examples the variety of purposes for which simulated trials have actually been employed in medical research in recent years with the aim of illustrating a broad spectrum of applications.

2.1 Dosage optimization

In a recent study (Ozawa et al., 2009) the authors perform trial simulations in order to evaluate the dose reduction strategy in patients with liver dysfunction of a clinically well established medication - called docetaxel - used to treat breast, ovarian, non-small cell lung and other types of cancer. Docetaxel clearance is decreasing in patients with liver dysfunction therefore it may be indispensable to reduce the dose for this kind of patients and a reduction strategy linked to the gravity of liver dysfunction has been proposed (Minami et al., 2009). Since it is difficult to have a sufficiently large number of these patients for a real clinical trial, because of the typical exclusion criteria, the authors of this paper use a number of dose-response models and a pharmacokinetic model of docetaxel in order to simulate drug exposure. “Survival” and “number of patients who had a particular side effect (febrile neutropenia)” were the two chosen endpoints. A Weibull model was used to express the time to drop-out and the patients’ characteristics were simulated according to previous Phase II studies. The model was validated with Phase II data provided by Kunitoh et al. (1996) by comparing the predicted trial results obtained by the medians of simulation with the real data. The results of the clinical trial simulations suggested that it is possible to decrease toxicity via a reduced amount of docetaxel without loss of efficacy.

2.2 Dosage scheduling

Albers et al. (2007) conducted a simulation study aimed at developing an age-suitable carvedilol dosing strategy for paediatric patients, since the well established dosing scheduling was uniform for all age groups of this kind of patients at the time of the study. Carvedilol is a non-selective beta blocker used for the treatment of hypertension and congestive heart failure. For this purpose, a pharmacokinetic model was developed based on data from a prospective, nonplacebo-controlled study. The model was used for simulations of different dosing strategies. Covariates were included via stepwise forward inclusion procedure. In order to evaluate the pharmacokinetic model the authors compared simulated and measured carvedilol concentrations. The findings of the paper suggest that, in general, higher doses of the carvedilol are probably required for younger patients with respect to body weight. However, it is worth underlining the authors' words: "Further randomized controlled trials are necessary to establish a safe and effective use of carvedilol in paediatric patients with congestive heart failure".

2.3 Understanding treatment effects in population studies

Lee et al. (2010) have tried to gain a better understanding of the possible effects of vaccinating employees with the new H1N1 influenza vaccine through the development of a simulation model. In particular, they develop an agent-based computer model "consisting of a virtual population of computer commuter agents, each having a set of sociodemographic characteristics and behaviours, and which, like virtual people, moved among virtual households, workplaces, schools, and other locations every day and interacted with each other through simulated social networks" (Lee et al., 2010). The model outcomes were daily disease incidence, prevalence, clinic visits, work absenteeism, hospitalizations and deaths. The study shows the way in which several actions regarding vaccination may have an important impact during an epidemic especially in terms of the labour force.

2.4 Determining an efficient screening protocol

Since a randomized controlled trial to assess the efficacy of screening for ovarian cancer is costly because ovarian cancer is a rare disease and its diagnosis requires surgery, Urban et al. (1997) simulated the effects of offering screening to a given population, i.e. a virtual cohort of women of age 50 (at the beginning of the 30-year screening period) in order to identify an efficient protocol. A stochastic model was developed with the aim of evaluating the cost-effectiveness, namely the ratio that measures the cost per year of life saved attributable to the screening strategy, of several alternative protocols involving transvaginal sonography and/or a cancer antigen/biomarker called CA 125. The model simulates the natural progression of disease, then considers a screening program and evaluates how the screening strategy used alters longevity and costs. Assumptions and inputs for the model (cohort characteristics, disease, detection, survival and cost components) were based on previous data and literature reports. The aim of the simulation model was to rank possible strategies of screening in terms of benefits regarding health and cost advantages. The study suggests that screening once a year by transvaginal sonography conditional on high (or rising) values of CA 125 is more efficient than screening once a year by transvaginal sonography without considering CA 125.

The use of simulation models can be very fruitful as regards identifying questions to be addressed by a screening trial, as well as for suggesting screening strategies.

2.5 Comparison of trial designs

Simulation can be used to compare the properties of various experimental designs. Lockwood et al. (2006) used clinical trial simulation to select a robust design in order to test the hypothesis that a novel treatment was effective for Alzheimer's disease and therefore the primary aim of the study was to compare the power of several experimental designs to detect a treatment effect. Basing themselves on literature reports and previous Phase I data, the authors developed a family of reasonable disease and drug models (as the true effect for the new treatment was not known at the time of the study) describing the time course of the Alzheimer's

disease assessment scale (i.e. a test that evaluates language, memory, attention, reasoning etc). The models included pharmacokinetic, pharmacodynamic, disease progression, and placebo components. An execution model expressing the expected percentage of patients remaining in the trial was used based on past experience of 1% weekly dropout. The simulation results allowed the research team to compare eight trial designs and one of those proved to be more efficient than the traditional one, leading to savings in time and costs. This design was implemented later in a real trial.

2.6 Sample size determination

Simulations are frequently used to explore the impact of sample size on the study results. For instance, Chabaud et al. (2002) have simulated several clinical trials to investigate the best compromise between safety, efficacy, drug regimen, and number of patients to include in a Phase III study of a bradycardic agent called ivabradine developed for the treatment of stable angina pectoris. The authors examined the use of a physiological model aimed at transforming a biomarker (heart rate) into a clinical binary outcome ("absent" or "at least one chest pain"). Moreover, they developed a therapeutic model which assumed that the reduction in heart rate decreased the risk of angina attacks in patients with coronary artery disease. They also developed a pharmacokinetic–pharmacodynamic model that governed the decrease in heart rate based on the data of a Phase I randomized study with twelve healthy volunteers and different doses of ivabradine. The peculiarity of this paper lies in the fact that the authors use an epidemiologic database in order to obtain real heart rate profiles instead of a simulated model, i.e. they resample patients from a patient database rather than creating virtual patients.
The findings of the paper suggest that it is necessary to include 200 patients per group (control placebo and treated group) under an alpha of 5% and a power of 90% in order to detect a reduction of the risk of at least one chest pain in 15% of the treated patients.

3. Purposes of a simulated trial

In the light of the applications, we overview some possible types of simulated clinical trials, according to their purpose. It goes without saying that similar remarks extend to simulation in areas other than clinical research.

A virtual experiment may either be a complement to or take the place of a physical one. Let us first look at the most common case, in which simulation is an aid to real-life experimentation. Simulations may be run for

- **pre-trial** purposes

When run before a trial, simulation usually involves

— testing several scenarios to evaluate the implications of the assumptions and / or testing various models for model selection.

For instance Abbas et al. (2006) develop five simulation models of a clinical trial for evaluating the changes in cholesterol as a surrogate marker for lipodystrophy in HIV patients treated with different drugs. The models are based on different assumptions on treatment variability and cholesterol reduction over time. The primary aim of the paper is to validate and select the "best" model. Selection of the best model is based on the principle of parsimony and specific validation criteria proposed by the authors.

— choosing the **experimental design**

This typically means running simulations to assess the power of the test that we intend to perform once we observe the data, in order to recommend a given sample size when analytical calculations are not feasible. The common assumption is that there are no dropouts leads to underestimating the number or patients who need to be recruited to achieve a desirable statistical power.

But simulations may also be of help in improving the very design of the experiment, for instance by exploring its convergence properties in the case of a sequential one, studying the impact of possible protocol violations that may occur in the actual trial, and also

changing the inclusion / exclusion criteria.
Improvement of the experimental design can also be viewed as a **post-trial** purpose in order to prospectively optimize the design of future physical trials. In Skolnik et al. (2008) the authors performed a trial simulation comparing the so-called 3+3 patients per cohort design to the novel "rolling six" design with the aim of reducing the execution time of paediatric Phase I oncology trials, which is long with respect to the completion time related to adult cancer. Clinical trial simulations and virtual patients characteristics were based on historical data from 14 completed Phase I oncology trials conducted by the Children's Oncology Group and Pilot Consortium between 2000 and 2006. Of the 14 above-mentioned studies, 12 were judged suitable for investigating characteristics and timeline data for analysis. The study suggests that the new proposed design might decrease the duration of pediatric phase I oncology trials without increasing the risk of toxicity and, at the end of the paper, the authors state that they plan to prospectively evaluate the "rolling six" design in the upcoming generation of paediatric Phase I trials.

- **assessing** what might happen in a trial yet to be conducted e.g. predicting the outcome of Phase (k+1) using data from Phase k.

De Ridder (2005) illustrated a case study where the aim was to predict the outcome of a Phase III trial through data from two Phase II trials with five different doses. In particular the real data were related to two placebo-controlled double-blind Phase II dose ranging trials with patients treated for 4 weeks. Simulations were used in order to:

- obtain the outcomes of the Phase III trial;
- assess the robustness of an ongoing Phase III trial in the same context (patient variability, dose-response, drug-response);
- assess the chance of achieving a clinically relevant response with a reduced dose as compared with those included in the trial.

- **extrapolation** purposes

As M. Sale states (in Bonate and Howard Eds, 2004), the dimensions across which one may extrapolate include: different species (e.g. mouse/rat/dog to human), time (from a trial to demonstrate clinical

efficacy with a small number of strictly selected patients to a full clinical study), endpoint (from a surrogate to a clinical endpoint, namely a characteristic that reflects how a patient feels, functions, or survives), population (e.g. healthy to patients, adults to paediatric), dose/dosing regimen. In all these cases, the domain of these simulated trials is outside what has been investigated so far.

An alternative scenario is when the virtual experiment is run instead of a physical one or interactively. Therefore, the simulated trial can be conducted:

- **in replacement of a real trial**, to provide direct knowledge about the drug(s) under investigation. (Most of the trials of Section 2 seem to belong to this category.)
- **interactively with a real trial**, to build knowledge about the "true" state of nature, while dynamically modifying the computer code to get closer and closer approximations to the real phenomenon under study. We shall discuss this case in the final Section of the paper.

4. Simulation models

Computer experiment models simulate scenarios that might arise in real situations. In a clinical situation, a simulation model will include at least three submodels:

- an input-output (IO) model
- a covariate distribution model
- an execution model

Input-output models These are the models that describe the patient's response to the treatment in mathematical terms. They include pharmacokinetic, pharmacodynamic, disease progression models or a combination of these, and should incorporate all the available information about the treatment and/or the disease derived from biological considerations, previous trials and other reliable sources. Pharmacokinetics models describe how the body processes the drug (absorption, distribution, metabolism and elimination), while

pharmacodynamics specifies how the drug works in the body. The time course of drug action in the body can be understood in terms of pharmacokinetics and pharmacodynamics (both in the absence or the presence of a disease). Disease progression models account for the time course of the disease in treated and untreated conditions. In particular, these models include a baseline disease status, a placebo response, a natural history component, namely in the absence of treatment the model should define the history of the disease, while in the presence of a particular treatment the disease status is modified and therefore the disease model should also take into account an active effect (for a thorough discussion see Chan and Holford, 2001).

Examples of input-output models can be found in Duffull et al. (2000), Pillai et al. (2004), Gruwez et al. (2007), Zierhut et al. (2008), Habtemariam et al. (2009). However, other types of simulation models can be used such as agent-based models, which are based on simulating the behaviour of individuals and the overall consequences of their local interactions (see for example Lee et al. 2010), physiology models (see Chabaud et al. 2002) etc.

A word of warning: features of a model that are not relevant to the questions that have been posed from the simulation team should not be considered. For instance, even though "weight" could be a covariate of primary importance for a real trial, if the virtual experiment we want to conduct regards the same weight group, we should not include "weight" in the model. This may seem a fairly obvious statement, but in real life it is frequently violated.

Covariate distribution models: Input-output models usually include terms for covariate effects (prognostic factors), as models used for simulation studies must deal with the variability from individual to individual. Covariate distribution models describe in a probabilistic way, on the basis of previous trials or clinical experience, the variability of patients' demographic and physiological characteristics in the population of interest that might affect the response. Given an input-output model, the covariate distribution may be changed to reflect different characteristics in another population. Therefore, the impact of the different covariate distribution(s) on the expected outcome of a simulated

trial can be assessed, thus making it possible to explore conditions that have been ruled out in the inclusion/exclusion procedures of the actual trial.

Execution or drop out models: Although the protocol of a clinical trial is a binding document, it is well-known that some deviations from protocol are inevitable, due to patients dropping out, patients' non-compliance, patients lost to follow-up etc. (but also due to acquiring subsequent information which was not available when the study protocol was written). In simulation, execution models describe uncontrollable factors leading to deviations from protocol and therefore can be extensively used as a tool for anticipating weaknesses and limitations in a proposed study. Indeed, consequences of protocol deviations such as insufficient statistical power and patients' discontinuation can be studied via modelling and simulation techniques. A simple example is a dropout model in Lockwood et al. (2003) describing a random 3% weekly dropout rate derived from previous studies. Girard et al. (1998) develop a Markov execution model for patients' non-compliance assuming that the probability of taking a wrong dose (or not taking any dose at all) at a given time depends on the number of doses taken at the previous dose timing. Wang, Husan and Chow (1996) propose statistical models in the case of multiple dose regimen trials aimed at studying the impact of two different non-compliance scenarios: patients who do not take the prescribe dosage or patients who do not adhere to the dosing schedule.

For a discussion of execution models see also Girard (2005).

5. Experimental designs for simulation

In the Western world and the major developing countries, guidelines for the correct conduct of a clinical study have been issued by authoritative regulatory agencies. In drug development studies, a joint regulatory-industry initiative is the Technical Requirements for Registration of Pharmaceuticals for Human Use by the International Conference on Harmonization (ICH). A protocol is demanded for every trial, aimed at assuring safety and health of the trial subjects, and also adherence to the same standards by all study investigators, since most trials are multicentre

ones. In particular, the most important design decisions of the protocol may involve:

- the choice of the treatments, which for most trials include one or more controls.
- the eligibility criteria (inclusion/esclusion of potential subjects) and the sampling rule.
- the sample size. When the design is carried out sequentially, this is replaced by the stopping rule.
- the allocation rule of the subjects to the treatment arms. Very often this rule has a randomization component in it.
- the use of blinding or double blinding i.e. masking the treatments to the subjects and often to the investigators as well.

One can safely assume that there are no ethical problems involved in simulated trials, and the costs are often a minute fraction of those of a real trial, so is a "protocol" for virtual trials still necessary? We maintain that it is, for instance in all the cases described in Section 3, a simulation plan approved by the research team would be needed, although it might be different from what a real experiment would require. The document should specify, for instance:

- questions that have to be answered via the simulation
- assumptions
- description of the virtual experiment
- statistical methods and analyses
- suitable data to support the simulation model
- techniques for model validation
- extrapolation questions.

The primary focus in the preparation of a simulation plan is to identify the question(s) that the project team wants to answer by the simulation experiment, but there are further advantages, as Kimko and Duffull (2003) state:

- A simulation plan may work as a communication tool, especially in the model-building procedures where many assumptions should be listed in the plan.
- It should convey a level of transparency that allows any or all of the work to be reproduced or continued by newly added persons

to the simulation team.

- In addition, the simulation plan can provide a pro forma for the development of similar drugs or similar types of trial designs.
- Approved simulation plan adds credibility and acceptance of the clinical trial simulation process.

In simulation too there is scope for designing the experiment efficiently so as to gather information in the best possible way, so we now move on to discuss the experimental plan itself. The design and analysis of deterministic computer experiments has a vast literature (see for instance Santner, Williams, and Notz, 2003 or Fang, Li and Sudjianto, 2006). The design consists in choosing the settings of the input variables, with the proviso that a deterministic simulator provides "observations" without error, so replication is pointless. Space-filling, Latin Hypercubes, Minimax and Maximin Distance criteria, Uniform designs are used in a non-model based approach, and special analysis procedures such as the Kriging methodology are employed (Santner, Williams, and Notz, 2003).

However, the simulator of a clinical trial – the input-output model, as well as the covariate model and the execution model – will very likely include a stochastic component and the rationale for using standard statistical tools, in particular, standard experimental design theory, is restored. This includes traditional design techniques going back to Fisher, based on replication, randomization and blocking, and/or design optimality criteria (see for instance Atkinson, Donev and Tobias, 2007). Here it is worth mentioning that several papers address the problem of determining optimal experimental designs for pharmacokinetic-pharmacodynamic models: (see for instance Duffull et al., 2005, Ogungbenro et al., 2007, McGree, Eccleston and Duffull, 2009).

In addition, if simulating is cheap, we can expand the range and possibilities of the trial design. For a start, the choice of the covariate levels is under the experimenter's control and this allows for exploring conditions that have been ruled out in the inclusion/exclusion procedures of the actual trial exploring in depth all possible levels of the concomitant variables. The full strength of simulation lies in being able to treat prognostic factors as random noise in the simulated experiment, and letting them vary according to a prescribed probability law, whereas

in an actual trial we would have to content ourselves with just a few set levels, either chosen by the experimenter or occurring by pure chance.

As regards factors of interest, we can experiment on a wider design space and/or increase the number of factors that are simultaneously tried and their levels. When simulating, we would normally not confine ourselves to fractional factorials but use full factorials instead. An important point is that the usual rules of factorial experiments apply, namely we should not vary the factor levels one-at-a-time, to avoid masking possible interactions, not just possible interactions among the experimental factors (e.g. dosage and dose timing of the drug) but also possible interactions between treatments and prognostic factors. In actual practice often only a subset of factors proves to be responsible for most of the output variation, but not much use is made by clinical triallists of the literature on screening experiments, i.e. experiments for choosing a few relevant factors out of a potentially very large number (see Dean and Lewis, 2006). It is also sensible to use simulation for detecting possible side effects, and for accounting for possible protocol deviations.

Sequential design deserves special attention. Most real clinical trials are extremely lengthy. Recruitment is typically sequential – patients join the trial one by one – and very slow. Results too become available serially. Thus in general trials are conducted sequentially on groups of patients and interim analyses of the data are performed. Adaptive designs have come into use: adaptation of the study protocol involves changes in sample size, changing doses, dropping treatment arms, changing the timing and number of interim analyses, etc. Clearly the crucial inferential problem is to assess the impact of such changes on the statistical analysis (Bauer and Köhne, 1994, Posch, Bauer and Brannath, 2003, Cui, Hung and Wang, 1999). Going from real to virtual, it makes sense to ask ourselves whether a simulated trial in clinical research should or should not be carried out sequentially, since the above mentioned issues (slow patient recruitment to the trial, side effects, ethical demand of early stopping, etc.) do not apply to computer experiments. One possible answer is that sometimes the sequential nature of the experiment is dictated by inferential aspects, e.g. recursive estimation of unknown parameters of the model in binary response comparative trials (see for

instance Hu and Rosenberger, 2006) or non-parametric convergence to the unknown MTD in the Up-And-Down experiments for Phase I (see Bortot and Giovagnoli, 2005). The severe handicap of the generally slow convergence of the algorithms is no longer a problem when the experiment is a simulated one.

As a final thought, we like to add that often the choice of the simulator itself is the output of a trial-and-error process that can be regarded as a virtual experiment. In other words, maybe we should also be looking into techniques of experimental design for choosing the simulator as well, and use, for example, designs for model-selection (as in Atkinson, Donev and Tobias, 2007).

6. Metamodels

The requirement for the input-output model to be accurate in describing the problem under investigation means that the simulator may be rather complex. In some instances the simulator consists of the simultaneous solution of a large number of linear or non-linear, ordinary and/or differential equations and, consequently, running it does take up an appreciable amount of computer time or other resources. A possible solution consists in employing so-called emulators or surrogates, i.e. simpler models which represent a valid approximation of the original simulator. Since emulators imitate the original simulator, which is itself a model of reality, they are often called metamodels. One of the fundamental characteristics of these surrogate models is computational speed.

Furthermore, the case where data cannot support estimating all of the parameters in a complicated simulation model is not rare. Therefore, models with fewer parameters should be fitted to the data. In a study by Pillai et al. (2004), the authors state that "Although the complex physiological PK-PD model described the data well, its major disadvantages were the long computer run-times […] and the numerical difficulties associated with solving a rather stiff problem". In order to reduce the computer run-times associated with the simulator, the authors

have constructed a ‘kinetics of drug action’ (K-PD model) and its performance was assessed by fitting data simulated with the PK-PD model under various scenarios. The authors observe that the simplified model was virtually indistinguishable from the complex one.

Another use of metamodels in clinical research is to be found in Kowalski and Hutmacher (2001), who decided to adopt a one-compartment model instead of a two-compartment one to face the problems arising from a sampling design that, due to logistic reasons and clinical convenience, was inadequate for the more complex model.

7. Validating simulated trials: some examples

Especially in the context of clinical trials we need to make sure that the simulators i.e. the models we use are “reasonable”. The key issue is whether a particular simulator is an adequate representation for the real system it is trying to represent, and consequently the question of its ability to accurately predict real situations. This concern is related to model verification and validation (e.g. see Sargent, 2008). Model verification is concerned with mistakes that may occur in the computer program of the model and its implementation, while model validation is usually defined to mean “substantiation that a computerized model within its domain of applicability possesses a satisfactory range of accuracy consistent with the intended application of the model” (Schlesinger et al., 1979). Thus, the primary aim of validation is to make the model useful in the sense that it addresses the right problem and provides accurate information about the trial of interest. It goes without saying that to a certain extent this question arises in real experiments too, since real data too are subject to error, but in most cases we are inclined to believe that a real experiment has “empirical validity”, whereas a simulated one is fictitious and therefore far away from reality. When real data provided by physical experiments are taken to be the “gold standard” of the true relationship between factors and outputs, they should be used to confirm the computer model and the results obtained by simulation. In some cases, experimental data may not be available and data derived from observational studies or surrogate data (e.g. derived from experiments on animals or prototypes) may be used.

We can distinguish between retrospective and prospective validation. The so-called prospective validation is the one that uses data from simultaneous or subsequent clinical trials in the same context (e.g. same disease). Retrospective external validation uses the data of earlier trials to validate the model and, if necessary, modify it in order to present higher degree of credibility and confidence. Sometimes it is even possible to collect a new dataset for validation. If not (e.g. studies of rare diseases), an internal validation is used, which is based on "cheap" methods such as data-splitting, where data utilized in order to build the simulator are compared with data generated by the model. The validation problem is tackled with the aid of a family of resampling methods, at the expense of further computations.

Concordance of simulated with real data under the same study design can be performed via:

- the use of graphs (or descriptive statistics), e.g. predicted versus observed dependent variable, residuals versus predicted values of the dependent variable;
- metrics (e.g. standardized distances);
- a posterior predictive check;
- statistical check methods (e.g. chi-squared or Kolmogorov–Smirnov tests).

In what follows we describe the validating methodology applied in some simulation projects taken from the literature.

In the carvedilol dosing strategy study described earlier (see §2.2), Albers et al. (2007) make use of a visual predictive check in order to evaluate the proposed simulation model: plasma concentrations (dependent variable) from 17 real patients were observed and compared with the simulation data. The authors observe that about 90% of the real data are within the 90th percentile of the simulated concentrations. The precision of the unknown parameter estimates of the pharmacokinetic model was assessed by establishing 95% confidence intervals using a bootstrap analysis.

Eddy and Schlessinger (2003) validate the so-called Archimedes diabetes model, namely a representation of the anatomy, treatments and outcomes related to diabetes, by comparing Kaplan-Meier curves of real

and virtual data. In particular, they examine whether the difference between the outcome of the actual trial and the model is statistically significant by using the corrected chi-squared and the correlation coefficient between the outcomes calculated by the model and the outcomes of the real trial.

Duffull et al. (2000) develop a pharmacokinetic model for ivabradine and they use two different kinds of datasets in order to test its ability to describe the real data: all the observations used for the construction of the model and data were collected from a different study regarding 12 subjects. The authors state "The posterior predictive performance is a test of the degree of similarity between the system model and the system itself. It is performed by simulating data from the model under the same experimental design that the original data was obtained. Ideally, the simulated data should exactly represent the observed data. […] We have assessed the predictive performance by inspection of the prediction plots visually and comparing the cumulative density functions of the simulated and observed using a Kolmogorov–Smirnov test for two samples".

Abbas et al. (2006) propose an innovative approach for the validation and selection of a simulation model based on the standardized distance, in mean and variance, between real and simulated data.

In engineering problems, several papers address the problem of validating the computer model via Bayesian techniques. For instance Bayarri et al. (2007) introduce a fully Bayesian approach for modelling the bias between the computer model and the physical data. See also Wang, Chen and Tsui (2009) and Kennedy and O'Hagan (2000) among others, but in the clinical context a Bayesian approach for validation does not appear to be as widely used.

There may also be alternative ways for validation that have never been explored so far, e.g. tests for agreement (2004).

8. Some challenges

Undoubtedly the successful execution of a simulation project requires a multi-disciplinary approach: interaction and cooperation are needed among scientists from various disciplines (clinicians, statisticians, computer scientists) and institutions (e.g. regulatory agencies and industry). The recent interest shown by pharmaceutical companies in clinical trial simulation poses several challenging questions:

- Scientificity: Is this new discipline rigorous enough? Can results obtained by computer experiments be trusted?
- Efficacy: Is it true that simulated clinical trials can speed the drug development process? After all, the model development procedure too is associated with time and high costs.
- Ethics: Is it safe for the patients? Is it to their best advantage? Or do these efforts only help the pharmaceutical companies to reduce costs without any benefit for the patient community?

We stress that this method of investigation is not aimed at replacing real life trials; rather, physical and computer experiments are two complementary sources of information with distinct roles and different degrees of cost, speed, and reliability. Simulation is usually cheaper and faster, and, what is more important, avoids the major ethical problems involved in clinical research, but in order to be of use, simulation must be fairly close to the physical set-up. What is the best way of integrating real and simulated trials to build actual knowledge while dynamically modifying the computer code to get closer and closer approximations to the reality? A virtual experiment may be part of a sequence in which simulations and physical observations play a part with alternating roles. The fundamental steps in designing such a mixed trial would consist of

- designing actual (small) trials that provide the physical data;
- designing the simulated ones, to be run in groups, one after another, to improve our knowledge of the process;
- choosing a "switching rule": when do we change over from a virtual experiment to a real one to acquire more data, and vice-versa?

- choosing a final stopping rule.

To the best of our knowledge, this type of strategy has not yet been the object of theoretical investigation in a clinical research context. We are convinced that much work lies ahead.

Last but not least, it is worth pointing out that although in the majority of cases the aim of a clinical trial is drug development, as shown in Section 2 there is a wide variety of additional areas of investigation that require trials on humans: in particular, new approaches to surgical and radiation therapies, physiotherapeutic treatments, new vaccines, new medical devices and test kits, new diagnostic tools and procedures, new methods of population screening, not to mention improving the quality of life: healthy eating, lifestyle changes, comfort for chronic illnesses, old age, etc. In all of them the practice of simulating experiments, wholly or partially, will sooner or later gather momentum.

Acknowledgements This research was partly supported by the Research Project of National Interest of Italian Ministry for Research (MIUR) PRIN 2007 Statistical Methods for Learning in Clinical Research. The second author was supported by a grant from the University of Bologna "Esperimenti simulati e loro applicazioni tecnologiche e bio-sanitarie".

References

I. Abbas, J. Rovira, J. Casanovas (2006), *Validation by simulation of a clinical trial model using the standardized mean and variance criteria*, Journal of Biomedical Informatics, **39**: 687–696.

S. Albers, B. Meibohm, T.S. Mir, S. Läer (2007), *Population pharmacokinetics and dose simulation of carvedilol in paediatric patients with congestive heart failure*, British Journal of Clinical Pharmacology **65**: 511-522.

A. C. Atkinson, A. N. Donev, R. D. Tobias (2007), *Optimum Experimental Designs, With Sas*. Oxford University Press.

P. Bauer, K. Köhne (1994), *Evaluation of Experiments with Adaptive Interim Analyses*, Biometrics, **50**: 1029-1041.

M. J. Bayarri, J.O. Berger, R. Paulo, J. Sacks, J. A. Cafeo, J. Cavendish, C. H. Lin, J. Tu (2007), *A Framework for Validation of Computer Models*, Technometrics, **49**: 138-154.

P. Bortot, A. Giovagnoli (2005), *Up-and-down experiments of first and second order*, Journal of Statistical Planning and Inference, **134**: 236-253.

S. Chabaud, P. Girard, P. Nony, J. P. Boissel (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.

P. L. Chan, N. H. Holford (2001), *Drug treatment effects on disease progression*, Annual Review of Pharmacology and Toxicology, **41**:625-659.

[9] L. Cui, H. M. J. Hung, S.-J. Wang (1999), *Modification of Sample Size in Group Sequential Clinical Trials*, Biometrics, **55**: 853-857.

F. De Ridder (2005). *Predicting the Outcome of Phase III Trials using Phase II Data: A Case Study of Clinical Trial Simulation in Late Stage Drug Development*, Basic & Clinical Pharmacology & Toxicology **96**: 235-41.

A. Dean, S. Lewis Eds, (2006) *Screening Methods for Experimentation in Industry, Drug Discovery, and Genetics* Springer New York.

S. Duffull, S. Chabaud, P. Nony, C. Laveille, P. Girard, L. Aarons (2000), *A pharmacokinetic simulation model for ivabradine in healthy volunteers*, European Journal of Pharmaceutical Sciences, **10**: 285–294.

S. Duffull, T. Waterhouse, J. Eccleston (2005), *Some considerations on the Design of Population Pharmacokinetic Studies*. Journal of Pharmacokinetics and Pharmacodynamics, **32**: 441-457.

D. M. Eddy, L. Schlessinger (2003), *Validation of the Archimedes Diabetes Model*, Diabetes Care **26**: 3102-3110

K.T. Fang, R. Li, A. Sudjianto (2006). *Design and Modeling for Computer Experiments*, Chapman & Hall/CRC New York.

FDA U.S. Department of Health and Human Services, Challenge and Opportunity on the Critical Path to New Medicinal Products, March 2004. Available at

www.fda.gov/downloads/ScienceResearch/SpecialTopics/CriticalPathInitiative/CriticalPathOpportunitiesReports/ucm113411.pdf

P. Girard, T. F. Blaschke, H. Kastrissios, L. B. Sheiner (1998), *A Markov mixed effect regression model for drug compliance*, Statistics in Medicine **17**: 2313-2334.

P. Girard (2005), *Clinical Trial Simulation: a tool for understanding study failures and preventing them*, Basic & Clinical Pharmacology & Toxicology **96**: 228–234.

B. Gruwez, Marie-France Poirier, A. Dauphin, Jean-Pierre Olié, M. Tod (2007), *A kinetic-pharmacodynamic model for clinical trial simulation of antidepressant action: Application to clomipramine–lithium interaction*, Contemporary Clinical Trials **28**: 276–287.

B. Habtemariam, W. Sallas, G. Sunkara, S. Kern, V. Jarugula, G. Pillai (2009), *Population pharmacokinetics of valsartan in pediatrics*, Drug Metabolism and Pharmacokinetics **24**: 145-152.

N. G. H. Holford, H. C. Kimko, J. P. R. Montelone, C. C. Peck (2000), *Simulation of clinical trials*. Annual Review of Pharmacology and Toxicology, **40**: 209-234.

F. Hu, W.F. Rosenberger (2006). *The Theory of Response-Adaptive Randomization in Clinical Trials*, New York: Wiley & Sons.

M. C. Kennedy, A. O'Hagan (2000), *Predicting the Output from a Complex Computer Code when Fast Approximations Are Available*. Biometrika **87**: 1-13.

H. C. Kimko, S. B. Duffull (eds.). *Simulation for designing clinical trials. A pharmacokinetic-pharmacodynamic modeling perspective*, Dekker, New York, 2003.

K. G. Kowalski, M. M. Hutmacher (2001), *Design evaluation for a population pharmacokinetic study using clinical trial simulations: a case study*, Statistics in Medicine, **20**: 75-91.

H. Kunitoh, K. Watanabe, T. Onoshi, K. Furuse, H. Niitani, T. Taguchi (1996), *Phase II trial of docetaxel in previously untreated advanced non-small-cell lung cancer: a Japanese cooperative study*, Journal of Clinical Oncology **14**: 1649-1655.

B. Y. Lee, S. T. Brown, P. C. Cooley, R. K. Zimmerman, W. D. Wheaton, S. M. Zimmer, J. J. Grefenstette, T. Assi, T. J. Furphy, D. K. Wagener, D. S. Burke (2010), *A Computer Simulation of Employee Vaccination to Mitigate an Influenza Epidemic*. American Journal of Preventive Medicine, **38**: 247–257.

P. Lockwood, J. A. Cook, W. E. Ewy, J. W. Mandema (2003), *The use of clinical trial simulation to support dose selection: Application to development of a new treatment for chronic neuropathic pain*, Pharmaceutical Research, **20**: 1752-1759.

P. Lockwood, W. Ewy, D. Hermann, N. Holford (2006), *Application of clinical trial simulation to compare proof-of-concept study designs for drugs with a slow onset of effect; An example in Alzheimer's disease*, Pharmaceutical research, **23**: 2050-2059.

J. M. McGree, J. A. Eccleston, S. B. Duffull (2009), *Simultaneous versus sequential optimal design for pharmacokinetic-pharmacodynamic models with FO and FOCE considerations*, Journal of Pharmacokinetics and Pharmacodynamics, **36**:101–123.

H. Minami, K. Kawada, Y. Sasaki, M. Tahara, T. Igarashi, K. Itoh, H. Fujii, T. Saeki, K. Ozawa, H. Sato (2009), *Population pharmacokinetics of docetaxel in patients with hepatic dysfunction treated in an oncology practice*, Cancer Science, **100**: 144-149.

K. Ogungbenro, I. Gueorguieva, O. Majid, G. Graham, L. Aarons (2007), *Optimal Design for Multiresponse Pharmacokinetic–Pharmacodynamic Models – Dealing with Unbalanced Designs*. Journal of Pharmacokinetics and Pharmacodynamics, **34**: 313-331.

K. Ozawa, H. Minami, H. Sato (2009), *Clinical trial simulations for dosage optimization of docetaxel in patients with liver dysfunction, based on a log-binomial regression for febrile neutropenia*. Yakugaku Zasshi, **129**: 749-757.

Phrma Annual Report 2009. Available at

www.phrma.org/files/attachments/09-103%20PhRMA_AR071409SP.pdf

G. Pillai, R. Gieschke, T. Goggin, P. Jacqmin, R.C. Schimmer, J.-L. Steimer (2004), *A semimechanistic and mechanistic population PK–PD model for biomarker response to ibandronate, a new bisphosphonate for the treatment of osteoporosis*, British Journal of Clinical Pharmacology, **58**: 618-631.

M. Posch, P. Bauer, W. Brannath (2003), *Issues in designing flexible trials*, Statistics in Medicine, **22**: 953-969.

M. Sale. Clinical Trial Simulation. In: *Pharmacokinetics in Drug Development: Clinical Study Design and Analysis* (Volume 1). Peter L. Bonate and Danny R. Howard Editors (2004). AAPS Press, Arlington.

T. J. Santner, B.J. Williams, W. I. Notz (2003). *The Design and Analysis of Computer Experiments*, Springer, New York.

R. G. Sargent (2008), *Verification and validation simulation models*. In: S. J. Mason, R. R. Hill, L. Mönch, O. Rose, T. Jefferson, J. W. Fowler, editors. Proceedings of the 2008 winter simulation conference.

S. Schlesinger, R.E. Crosbie, R.E. Gagné, G.S. Innis, C.S. Lalwani, J. Loch, R.J. Sylvester, R.D. Wright, N. Kheir and D. Bartos (1979), *Terminology for model credibility*, Simulation, **32**: 103–104.

M. M. Shoukri (2004), *Measures of Interobserver Agreement*. Chapman & Hall/CRC.

J. M. Skolnik, J. S. Barrett, B. Jayaraman, D. Patel, P. C. Adamson (2008), *Shortening the Timeline of Pediatric Phase I Trials: The Rolling Six Design*, Journal of Clinical Oncology, **26**: 190-195.

D. M. Steinberg. An overview of Computer Experiments. ENBIS-EMSE Saint-Etienne France, 2009. Available at http://www.emse.fr/enbis-emse2009/pdf/slides/D.%20Steinberg.pdf

D. W. Taylor, E.G. Bosch (1990). *CTS: a clinical trial simulator*. Statistics in Medicine **9**: 787-801.

N. Urban, C. Drescher, R. Etzioni, C. Colby (1997), *Use of a Stochastic Simulation Model to Identify an Efficient Protocol for Ovarian Cancer Screening*, Controlled Clinical Trials **18**:251-270.

W. Wang, F. Husan, S.C. Chow (1996), *The impact of patient compliance on drug concentration profile in multiple doses*. Statistics in Medicine, **15**: 659-669.

S.Wang, W. Chen, K.-L. Tsui (2009), *Bayesian Validation of Computer Models*, Technometrics **51**: 439-451.

M.L. Zierhut, M. R. Gastonguay, S. W. Martin, P. Vicini, P. J. Bekker, D. Holloway, P. T. Leese, M. C. Peterson (2008), *Population PK–PD model for Fc-osteoprotegerin in healthy postmenopausal women*, Journal of Pharmacokinetics and Pharmacodynamics, 35: 379–399.