STATISTICA, anno LXXII, n. 1, 2012

SIMULATION OF CLINICAL TRIALS: A REVIEW WITH EMPHASIS ON THE DESIGN ISSUES

A. Giovagnoli, M. Zagoraiou

1. INTRODUCTION

Simulation run on a computer is a formidable tool to aid and complement real life experiments. 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 few physical runs can be made due to their high cost. For these reasons the practice of complementing laboratory experiments or field observations by means of simulated ones has been steadily growing in recent years. The books by Santner, Williams and Notz (2003) and by Fang, Li and Sudjianto (2005) provide a useful introduction. In a recent conference dedicated to computer experiments, Levy and Steinberg (2010), starting from applications, have reviewed 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.

Despite the understandable misgivings of the 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 progression and drug behaviour in the human body, thus making computer simulation possible in the pharmaceutical/biomedical field too. 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. To bring down the costs, prevent possible failures in future trials, reduce the trial time frame and avoid possible side effects in humans, clinical trial simulation (CTS) is asserting itself as an emerging technique to im-

prove the efficiency of the drug development process, thanks also to the advent of new powerful software tools. The excellent set of guidelines (Holford *et al.*, 1999) for correct CTS suggested in 1999 covers the following topics: planning a simulation project, models for simulation, computational methods, execution, critical assessment of simulation results and reporting, but it is not clear whether they are used or not in actual practice. A very recent substantial review by Holford, Ma and Ploeger (2010) of relevant papers published during the period 2000-2010 discusses methodological developments and applications of CTS. An important contribution is also 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; Holford *et al.*, 2002). A very useful introductory article has appeared recently (Krause, 2010).

In this paper – which is mainly of a review character – we look at computer simulation in clinical trials paying special attention to the design aspects. We aim at making medical statisticians more aware of the statistical issues and problems arising in this field. Section 2 presents some remarks about protocols for simulation studies. All the potential aims of simulation in clinical research are overviewed in Section 3. Section 4 contains a short description of the models used in clinical contexts, which must be implemented in a simulator, and in Section 5, the central part of this paper, we discuss the ensuing experimental design problems: the design of a simulated experiment is not necessarily the same as for a real one, due also to a possible difference in the endpoints, the aims etc. Section 6 explores existing software for CTS and Section 7 contains a brief introduction to the use of metamodels in medicine. Section 8 is dedicated to the question of validating the simulator of a virtual trial, in which statistics should play a crucial role. All the above topics will be illustrated by studies recently published in biopharmaceutical or biomedical journals. In the final section we make some comments and express some criticisms. Given that the subject is vast, we have made no attempt at covering all the existing bibliography: we refer to Holford *et al.* (2010) and some of the other papers we cite for additional references.

2. THE SIMULATION PROTOCOL

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). As is well known, a protocol is demanded for every trial, namely a written document setting out the rules and the steps to follow in the study, aimed at assuring the safety and health of the trial subjects, and also adherence to the same standards by all the study investigators when the trial is a multicentre one. Among the statistical decision to be made in advance of the trial, there is the description of the experiment itself, which includes:

– the choice of the treatments, which often include one or more controls
– the eligibility criteria (inclusion/esclusion of potential subjects)
– stratification of the 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 strong 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

and so on. In simulated trials one can safely assume that there are no ethical problems involved, and the costs are often a minute fraction of those of a real trial, but even for a virtual trial a protocol is still necessary, as clearly explained in the 1999 Guidelines (Holford *et al.* 1999). The primary focus of the protocol is to identify the question(s) that the project team wants to answer by means of the simulation experiment, but the document should also specify:

– assumptions
– description of the virtual experiment
– statistical methods and analyses
– suitable data to support the simulation model
– techniques for model validation
– extrapolation questions

and many more issues. The added value of a simulation protocol is discussed by Kimko and Duffull (2003): among other things, an approved simulation plan increases the credibility and acceptance of the trial simulation process.

3. PURPOSES OF SIMULATION IN CLINICAL RESEARCH

In a drug development program, virtual experimentation may be resorted to for a variety of purposes, both as an aid to in vivo experimentation and in place of a physical trial. We illustrate them quoting specific studies.

- *Pre-trial purposes*

Simulation is often run before a trial with one or more of the following purposes:

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

2. choosing the sample size
   This typically means running simulations to assess the power of the test that we intend to perform once we observe the data, when analytical calculations are not feasible, keeping in mind, however, that the common assumption of no dropouts leads to underestimating the number of patients who need to be recruited to achieve a desirable statistical power.
   – Chabaud *et al.* (2002) have simulated several clinical trials to investigate the number of subjects to include in a Phase III study of a bradycardic agent called ivabradine developed for the treatment of stable angina pectoris. The findings of the paper suggest that in order to obtain a desired reduction of the outcome it is necessary to include 239 patients per group (control placebo and treated group) with a twice-a-day low dose or 196 patients with a higher everyday dosage assuming an alpha risk of 0.05 and a power of 0.9.
3. finding robust designs, namely designs not too sensitive to some particular experimental choice
   – Lockwood *et al.* (2006) use clinical trial simulations to find a robust design in order to test the hypothesis that a novel treatment was effective for Alzheimer. The primary aim of the study was to compare the power of several experimental designs to detect a treatment effect using several dose response models, since the true effect of the treatment taken into account was unknown. The simulation results allowed the research team to compare the trial designs and one of those proved to be more efficient than the traditional one, leading to savings in time and costs.
4. predicting the outcome of real trials (this issue can also be viewed as a post-trial purpose)
   – Chan *et al.* (2007) use CTS to predict the outcome of a failed real trial in order to improve the understanding of its failure. The trial had been performed to detect a difference between placebo and levodopa, a drug therapy for Parkinson's.

- *Extrapolation purposes*

As stated by Sale (in Bonate and Howard Eds, 2004), the dimensions across which one may extrapolate include:

1. Species (e.g. mouse/rat/dog to human)
   – Dickinson *et al.* (2007) make predictions of pharmacokinetic and pharmacodynamics based on in vitro to in vivo extrapolations via simulations.
2. Phases (from a small number of strictly selected patients to a full clinical study or from Phase (k-1) to Phase k trial)
   – De Ridder (2005) illustrates a case study where the aim was to predict the outcome of a Phase III trial through data from two Phase II trial. 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.

3. Endpoints (from a surrogate to a clinical endpoint, namely a characteristic that reflects how a patient feels, functions, or survives)
   – Chabaud *et al.* (2002) examine 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").
4. Populations (e.g. healthy to patients, adults to paediatric)
   – Albers *et al.* (2007) conducted a simulation study aimed at developing an age-suitable carvedilol dosing strategy for paediatric patients since the dose given to young subjects was generally derived via linear extrapolation on the basis of the dose for adults but with dubious results.
5. Dose/dosing regimens
   – Ozawa *et al.* (2009) 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. 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.

- *Learning about the effects of a new drug, or new dosage, new dose scheduling, etc.*

The virtual experiment is run instead of a physical trial, or interactively, to provide direct knowledge about the drug(s) under investigation. This is what is properly meant by a simulated trial

   – Lockwood *et al.* (2003) use simulations to determine how precisely the minimum effective dose of a new treatment for neuropathic pain could be estimated. Concerning this treatment, only limited preclinical information was available and therefore clinical data on a different drug, shown to be clinically effective in diabetic neuropathy, were used in the simulation study.

In the final Section of the paper we shall discuss running simulations interactively with a real trial.

We end this section with two more examples relative to population studies and not drug development, to illustrate the broad spectrum of applications in medical research.

– 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 simulation shows how several actions regarding vaccination may have an important impact during an epidemic, especially in terms of the labour force.
– The use of simulations can be very fruitful as regards identifying questions to be addressed by a screening trial, as well as for suggesting screening strategies. Indeed, Urban *et al.* (1997) simulated the effects of offering screening to a given population in order to identify an efficient protocol because a randomized controlled trial to assess the efficacy of screening for ovarian cancer is costly (ovarian cancer is a rare disease and its diagnosis requires surgery). A stochastic model was developed with the aim of evaluating the cost-effectiveness of several alternative protocols involving transvaginal sonography and/or a cancer antigen/biomarker called CA 125, and the study suggests the importance of considering CA 125.

4. SIMULATION MODELS FOR CLINICAL TRIALS

The computer models that simulate real scenarios are generally developed from previous data sets that may include preclinical data, as well as previous phases of real trials. As clearly stated in the 1999 Guidelines (Holford *et al.*, 1999), a model for fully simulating a trial in drug development will include at least three submodels:

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

Input-output models: They are the models that describe the patient's response to the treatment in mathematical terms and they would normally be used for an in vivo experiment as well. These models include pharmacokinetic, pharmacodynamic, disease progression models or a combination of these. Often IO models are defined implicitly by a set of differential equations, which makes the implementing code computationally much slower to run. This will be discussed in Section 7. However, other types of models can also be used, such as physiological models (Chabaud *et al.*, 1999) or agent-based models, for simulating the behaviour of in-

dividuals and the overall consequences of their local interactions (Lee *et al.*, 2010). For a rich collection of PK/PD model equations see Chapter 11 of ADAPT 5 User's Guide (D'Argenio *et al.*, 2009). Descriptions of the IO models actually used by the authors can be found in the papers of Pillai *et al.* (2004), Gruwez *et al.* (2007), Zierhut *et al.* (2008); the paper by Post *et al.* (2005) includes a family of disease progression models.

Covariate distribution models: IO 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. Correlation between covariates should be considered, where appropriate. Methods for simulating from a joint distribution (whether continuous, discrete or mixed) are well-known in statistics. Given an IO model, the distribution function of covariates may be altered in the what-if scenarios of simulation to reflect different characteristics in another population. Thus the impact of the different covariate distributions on the expected outcome of a simulated trial can be assessed, making it possible to explore conditions that have been ruled out in the inclusion/exclusion procedures of the actual trial.

Execution and/or dropout 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, non-compliance, 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.* (2006) describing a random 1% 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 further discussion of execution models see also Girard (2005).

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 concerns the same weight group, we should not include "weight" in the model. This may seem a fairly obvious statement, but it is frequently violated.

5. EXPERIMENTAL DESIGNS FOR SIMULATION

All the modern books on clinical trial methodologies, see for instance Piantadosi (2005), Senn (2007), Friedman *et al.* (2010), devote at least one chapter to the experimental design. Here we want to discuss the design of a virtual experiment, which will be different from planning a real trial. However the design still needs to be efficient so as to gather information in the best possible way.

The design and analysis of deterministic computer experiments has a vast literature (Santner *et al.*, 2003; Fang *et al.*, 2005). 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 *et al.*, 2003). However, the simulator of a clinical trial – the IO model, as well as the covariate model and the execution model – will very likely include a stochastic component and the rationale of 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 also the use of specific designs, for instance cross-over designs and play-the-winner. It must be borne in mind that the choice of the experimental design will depend on the statistical model, and a model-based theory of optimal experimental design for clinical trials, including dose-finding ones, has come to a mature development stage, as shown in statistical journals and conferences (see for instance Giovagnoli *et al.*, 2010). But how relevant is this literature to the simulated experiments?

In simulations, we would normally experiment on a wider design space and/or increase the number of factors of interest and their levels that are simultaneously tried. 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. When simulating, we would normally not confine ourselves to fractional factorials but instead use full factorials to evaluate all the interactions among the experimental factors (e.g. dosage and dose timing of the drug). Fractional factorials would still be required, however, when the number of combinations of factors and levels is too large, as pointed out in the 1999 Guidelines (Holford *et al.*, 1999). 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 (Dean and Lewis, 2006). Furthermore, since virtual experiments are often run for choosing among

possible models, the theory of designs for model-selection (to be found for instance in Atkinson *et al.*, 2007) may be useful.

It is important to note however that often the experiment is a comparative one for the choice between two treatments. One may wonder about the role of randomization in simulation: it is usually preserved for realistic purposes, at the expense of exact balancing. However, more sophisticated ways of trading randomization and balance exist, for instance, the Biased Coin Design (Efron, 1971), or the Adjustable Biased Coin Design (Baldi Antognini and Giovagnoli, 2004), where at each step the probability of selecting the under-represented treatment is a non-increasing function of the current difference between the two groups of allocations, so that the tendency towards balance is stronger the more we move away from it. These could be implemented in simulation too.

Sequential design deserves special attention. In general clinical 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 (Posch *et al.*, 2003; Cui *et al.*, 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 frequently recurring issues of slow patient recruitment to the trial, side effects, ethical demand of early stopping, etc. do not apply to computer experiments. One answer is, again, to achieve greater realism, but also sometimes the sequential nature of the experiment is dictated by inferential aspects, e.g. recursive estimation of unknown parameters of the model in response adaptive trials (Hu and Rosenberger, 2006) or parametric and non-parametric convergence to the unknown MTD in the Up-And-Down experiments for Phase I (Baldi Antognini *et al.*, 2008; 'O Quigley, 2002). The severe handicap of the generally slow convergence of the algorithms is no longer a problem when the experiment is a simulated one.

It is worth mentioning that the problem of determining optimal experimental designs for pharmacokinetic and pharmacodynamic models has been addressed by several authors in the statistical and biomedical literature (for instance Fedorov *et al.*, 2007; Ogungbenro *et al.*, 2007; McGree *et al.*, 2009). However, Holford, Ma and Ploeger (2010) regret that the statistical theory of optimal design of experiments deals mainly with parameter estimation rather than hypothesis testing, whereas the main purposes of a clinical trial is usually assessing superiority or equivalence/non-inferiority of one drug over another. It will be interesting to see if a combined approach of optimal design methods and simulation will bring useful results: optimal design theory deals more often than not with designs that are most efficient for asymptotical inference, but possibly not fully so for small sample sizes. So, to be able to simulate a large sample according to an optimal experimental design should prove to be a good choice for accurate inference from the virtual data.

As regards covariates, in simulations the choice of their levels is under the experimenter's control and this allows for exploring conditions that are ruled out in

the inclusion/exclusion procedures of the actual trial, exploring in depth all possible levels of the concomitant variables, looking for possible interactions also between the treatments and the prognostic factors, since in general one wishes to use simulation for detecting also the possible side effects of a therapy. More in general, the full strength of simulation lies in being able to treat prognostic factors as random noise in the virtual 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. The statistical literature on experimental design does not seem to have caught up with this novelty. An IO model including random covariates is a mixed effect one (linear or non-linear), and appropriate experimental designs for these models are present in the literature, but they are all non-stochastic.

Lastly, what are the appropriate designs that enable accounting for possible protocol deviations? Again, this aspect has not been the object of statistical investigation as yet.

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. This is, yet again, a different problem, since in this case the endpoint is a measure of the performance of our simulator. In other words, maybe we should apply experimental design for choosing the simulator as well. Different techniques and different computer codes should be compared by the expert members of the simulation team.

6. SOFTWARE

Simulation for clinical trials includes different types of models and involves several statistical issues. Therefore often researchers use more than just one software, each software being targeted for specific purposes. In particular, sophisticated software packages are employed for IO models, which are usually quite complicated. Programs specifically designed for IO modeling of data in this context are the non-linear mixed-effect model program NONMEM or the Pharmacokinetic/Pharmacodynamic Systems Analysis Software ADAPT, which includes an extensive library of models to choose from. MathWorks provides a software tool, the so-called SimBiology, for the complete PK/PD workflow. Since SimBiology is based on MATLAB, users can employ MATLAB in order to program their simulations.

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.

To our knowledge, there are no particular software specifically designed for simulating execution models, but often a random number generator suffices. However, there also exists multi-purpose software for full clinical trial simulation that incorporates specialized methodologies for patients dropouts or for the solution of awkward differential equations, such as the Pharsight Trial Simulator and another, originally developed for Vertex Pharmaceuticals, which has recently become publically available. The software documentation can be found in

http://www.biopharmnet.com/doc/2010_02_13_cts_documentation.pdf.

In general, however, existing prepackaged software is, by definition, not flexible and this may turn out to be an obstacle. Furthermore, without a reasonable understanding of the statistical methods behind a specific clinical trial simulation software it is difficult to interpret the results correctly. Thus, rather than accepting library models and their assumptions, some scientists create models according to their own needs using the free environment statistical package R, or some general modeling and simulation packages such as Sigma for Windows (see Abbas *et al.*, 2008).

7. METAMODELS

The requirement for the IO 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. Particular optimal design problems for metamodels can be found in the recent literature (see for instance Baldi Antognini and Zagoraiou, 2010) but, in the clinical context, this aspect has not been the object of statistical investigation.

– 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.

8. VALIDATING SIMULATED TRIALS

In the context of clinical trials there is special emphasis on the need for the simulators to be "reasonable". The key issue is whether a particular simulator is an adequate representation for the real system that 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 (Sargent, 2010). Model verification deals with errors that might have occured in the computer program and its implementation, while model validation is usually defined as "substantiation that a computerized model within its domain of applicability possesses a satisfactory range of accuracy consistent with the intended application" (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 as well, since real data too are subject to random or systematic errors, 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 obtained 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 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 checked via:

- statistical goodness-of-fit methods (e.g. chi-squared or Kolmogorov-Smirnov tests).

- the use of graphs (or descriptive statistics), e.g. visual comparison of predicted versus observed values of the response variable, or residuals versus predicted;
- metrics (e.g. standardized distances between observed and predicted values);
- other methods like for instance PPC (posterior predictive check).

8.1 *Examples*

– In the carvedilol dosing strategy study described earlier (see §4), 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.
– In Ozawa *et al.* (2009) 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.
– 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.
– 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. The authors "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.

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

9. SOME CHALLENGES

It is worth pointing out that although we have concentrated on research for drug development, which is the aim of the majority of clinical trials, there is a wide variety of additional areas of investigation that require trials on humans: in particular, new approaches to surgical and radiation therapies, to 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.

It goes without saying that clinical trial simulation poses several challenging problems. First of all, some burning questions need an answer that is convincing for the laymen too.

- Scientificity: Is this new discipline rigorous enough? Can results obtained by computer experiments really 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?

Much work lies ahead for statisticians. 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) and it is up to the statisticians to develop appropriate methodological tools including, among other things, a suitable theory of experimental design for simulation. We stress that simulations are 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. Thus 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, the best strategy 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, has not yet been the object of theoretical investigation in a clinical research context.

ACKNOWLEDGEMENTS

This research was partly supported by the Research Project of National Interest of the Italian Ministry for Research (MIUR) PRIN 2007 Protocol 2007AYHZWC

"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".

Department of Statistical Sciences ALESSANDRA GIOVAGNOLI
University of Bologna MAROUSSA ZAGORAIOU

REFERENCES

I. ABBAS, J. ROVINA, J. CASANOVAS (2006), *Validation by simulation of a clinical trial model using the standardized mean and variance criteria,* "Journal of Biomedical Informatics", 39, pp. 687-696.

I. ABBAS, J. ROVIRA, J. CASANOVAS, T. GREENFIELD (2008), *Optimal design of clinical trials with computer simulation based on results of earlier trials, illustrated with a lipodystrophy trial in HIV patients,* "Journal of Biomedical Informatics", 41, pp. 1053-1061.

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, pp. 511-522.

A.C. ATKINSON, A.N. DONEV, R.D. TOBIAS (2007), *Optimum Experimental Designs, With Sas,* Oxford University Press, Oxford; New York.

A. BALDI ANTOGNINI, P. BORTOT, A. GIOVAGNOLI (2008), *Randomized Group Up and Down Experiments,* "Annals of the Institute of Statistical Mathematics", 60, pp. 45-59.

A. BALDI ANTOGNINI, A. GIOVAGNOLI (2004), *A new 'biased coin design' for the sequential allocation of two treatments,* "Journal of the Royal Statistical Society", Series B, 53, pp. 651-664.

A. BALDI ANTOGNINI, M. ZAGORAIOU (2010), *Exact optimal designs for Computer Experiments via Kriging metamodeling,* "Journal of Statistical Planning and Inference", 140, pp. 2607-2617.

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, pp. 339-363.

P.L. CHAN, J.G. NUTT, N.H. HOLFORD (2007), *Levodopa slows progression of Parkinson's disease: external validation by clinical trial simulation,* "Pharmaceutical Research", 24, pp. 791-802.

L. CUI, H.M.J. HUNG, S-J. WANG (1999), *Modification of Sample Size in Group Sequential Clinical Trials,* "Biometrics", 55, pp. 853-857.

D.Z. D'ARGENIO, A. SCHUMITZKY, X. WANG (2009), *ADAPT 5 User's Guide: Pharmacokinetic/ Pharmacodynamic Systems Analysis Software.* Biomedical Simulations Resource, Los Angeles.

A. DEAN, S. LEWIS (eds.), (2006), *Screening Methods for Experimentation in Industry, Drug Discovery, and Genetics,* Springer, New York.

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, pp. 235-41.

G.L. DICKINSON, S. REZAEE, N.J. PROCTOR, M.S. LENNARD, G.T. TUCKER (2007), *Rostami-Hodjegan A. Incorporating in vitro information on drug metabolism into clinical trial simulations to assess the effect of CYP2D6 polymorphism on pharmacokinetics and pharmacodynamics: dextromethorphan as a model application,* "Journal of Clinical Pharmacology", 47, pp. 175-186.

S.B. 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, pp. 285-294.

D.M. EDDY, L. SCHLESSINGER (2003), *Validation of the Archimedes Diabetes Model,* "Diabetes Care", 26, pp. 3102-3110.

B. EFRON (1971), *Forcing sequential experiments to be balanced,* "Biometrika", 58, pp. 403-417.

K.T. FANG, R. LI, A. SUDJIANTO (2005), *Design and Modeling for Computer Experiments.* Chapman and Hall/CRC, London.

V. FEDOROV, R. GAGNON, S. LEONOV, Y. WU (2007), *Optimal design of experiments in pharmaceutical applications*, in A. Dmitrienko, C. Chuang-Stein, R. D'Agostino (eds), Pharmaceutical Statistics using SAS. A Practical Guide, SAS Press, Cary, NC, pp. 151-195.

L.M. FRIEDMAN, C.D. FURBERG, D.L. DEMETS (2010), *Fundamentals of Clinical Trials, 4th Edition,* Springer Verlag, New York.

A. GIOVAGNOLI, A.C. ATKINSON, B. TORSNEY (eds.), C. MAY (co-editor) (2010), *mODa 9 – Advances in Model-Oriented Design and Analysis,* Springer-Verlag, Berlin Heidelberg.

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

P. GIRARD, T.F. BLASCHKE, H. KASTRISSIOS, L.B. SHEINER (1998), *A Markov mixed effect regression model for drug compliance,* "Statistics in Medicine", 17, pp. 2313-2334.

B. GRUWEZ, M-F. POIRIER, A. DAUPHIN, J-P. OLIÉ, M. TOD (2007), *A kinetic-pharmacodynamic model for clinical trial simulation of antidepressant action: Application to clomipramine-lithium interaction,* "Contemporary Clinical Trials", 28, pp. 276-287.

N.H.G. HOLFORD, M. HALE, H.C. KO, J-L. STEIMER, L.B. SHEINER, C.C. PECK (eds.) (1999) *Simulation in Drug Development: Good Practices.* Available at http://bts.ucsf.edu/cdds/research/sddgpreport.php. Last accessed 29 February 2012.

N. HOLFORD, S.C. MA, B.A. PLOEGER (2010), *Clinical Trial Simulation: A Review,* "Clinical Pharmacology & Therapeutics", 88, pp. 166-182.

N.G.H. HOLFORD, H.C. KIMKO, J.P.R. MONTELONE, C.C. PECK (2002), *Simulation of clinical trials,* "Annual Review of Pharmacology and Toxicology", 40, pp. 209-234.

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

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

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, pp. 75-91.

A. KRAUSE (2010), *The Virtual Patient: Developing Drugs Using Modeling and Simulation,* "Chance", 23, pp. 48-52.

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, pp. 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, pp. 247-257.

S. LEVY, D.M. STEINBERG (2010). *Computer experiments: a review.* "Advances in Statistical Analysis", 94, pp. 311-324.

P.A. 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, pp. 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, pp. 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, pp. 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, pp. 144-149.

M.S. MOUKSASSI, J.F. MARIER, J. CYRAN, A.A. VINKS (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", 86, pp. 667-671.

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, pp. 313-331.

J. O' QUIGLEY (2002), *Curve-free and model-based continual reassessment method designs,* "Biometrics", 58, pp. 245-249.

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, pp. 749-757.

S. PIANTADOSI (2005), *Clinical trials: a methodologic perspective,* Wiley, Hoboken New Jersey.

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, pp. 618-631.

M. POSCH, P. BAUER, W. BRANNATH (2003), *Issues in designing flexible trials,* "Statistics in Medicine", 22, pp. 953-969.

T.M. POST, J.I. FREIJER, J. DEJONGH, M. DANHOF (2005), *Disease System Analysis: Basic Disease Progression Models in Degenerative Disease,* "Pharmaceutical Research", 22, pp. 1038-1049.

M. SALE (2004), *Clinical Trial Simulation,* in P.L. BONATE AND D.R. HOWARD (eds.), Pharmacokinetics in Drug Development: Clinical Study Design and Analysis (Volume 1), 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 (eds.), Proceedings of the 2010 Winter Simulation Conference, Baltimore, USA, 5-8 Dec. 2010, (B. Johansson, S. Jain, J. Montoya-Torres, J. Hugan, and E. Yücesan, eds.) pp. 166-183.

S. SCHLESINGER, R.E. CROSBIE, R.E. GAGNÉ, G.S. INNIS, C.S. LALWANI, J. LOCH, R.J. SYLVESTER, R.D. WRIGHT, N. KHEIR, D. BARTOS (1979), *Terminology for model credibility,* "Simulation", 32, pp. 103-104.

S. SENN (2007), *Statistical issues in drug development, 2nd Edition,* Wiley, Chichester, England; Hoboken, New Jersey.

MM. SHOUKRI (2004), *Measures of Interobserver Agreement,* Chapman & Hall/CRC, Boca Raton.

D.W. TAYLOR, E.G. BOSCH (1990), *CTS: a clinical trias simulator,* "Statistics in Medicine", 9, pp. 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, pp. 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, pp. 659-669.

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, pp. 379-399.

SUMMARY

Simulation of clinical trials: a review with emphasis on the design issues

Simulation is a widely used tool to investigate real-world systems in a large number of fields, including clinical trials for drug development, since real trials are costly, frequently fail and may lead to serious side effects. This paper is a survey of the statistical issues arising in these simulated trials, with particular emphasis on the design of such virtual experiments, stressing similarities and differences with the design of real trials. We discuss the aims and peculiarities of the simulation models used in this context, including a brief mention of metamodels, and different validating techniques. We illustrate each specific issue through one or more studies recently reported in the medical and/or pharmaceutical literature. We end the paper with some challenging questions on the scientific rigour, ethics and effectiveness of simulation in clinical research, and the interesting research problem of how to integrate virtual and physical experiments in a clinical context.