Citation: CPT: Pharmacometrics & Systems Pharmacology (2013) 2, e23; doi:10.1038/psp.2012.24

www.nature.com/psp

ORIGINAL ARTICLE

Comparisons of Analysis Methods for Proof-of-Concept Trials

KE Karlsson[1], C Vong[1], M Bergstrand[1], EN Jonsson[1,2] and MO Karlsson[1]

Drug development struggles with high costs and time consuming processes. Hence, a need for new strategies has been accentuated by many stakeholders in drug development. This study proposes the use of pharmacometric models to rationalize drug development. Two simulated examples, within the therapeutic areas of acute stroke and type 2 diabetes, are utilized to compare a pharmacometric model–based analysis to a *t*-test with respect to study power of proof-of-concept (POC) trials. In all investigated examples and scenarios, the conventional statistical analysis resulted in several fold larger study sizes to achieve 80% power. For a scenario with a parallel design of one placebo group and one active dose arm, the difference between the conventional and pharmacometric approach was 4.3- and 8.4-fold, for the stroke and diabetes example, respectively. Although the model-based power depend on the model assumptions, in these scenarios, the pharmacometric model–based approach was demonstrated to permit drastic streamlining of POC trials.
CPT: Pharmacometrics & Systems Pharmacology (2013) **2,** e23; doi:10.1038/psp.2012.24; advance online publication 16 January 2013

The need for new methodologies within drug development has been highlighted several times in the last decade by regulators,[1] the drug industry,[2] and external evaluators,[3] and various approaches such as modeling and simulation, adaptive trial designs, the use of surrogate end points, etc. have been suggested as new strategies. One of the proposed approaches is the use of mixed-effects modeling for the development of pharmacometric models,[4–8] as the methodology is well-suited for the use of all available data (e.g., repeated measurements over time and multiple end points) and mechanistic interpretations of the model parameters.

Jonsson and Sheiner[9] have previously shown that the use of scientific model–based statistical tests can improve the efficiency of clinical trials and the use of pharmacometric models as decision making tools within drug development is increasing,[6,7,10] but there are only a few examples of pharmacometric models being used in the primary analysis of clinical trials. Pharmacometric models also offer an improved possibility of information propagation between development phases, mechanistic interpretation of the model parameters, and exploration of different study designs by means of clinical trial simulations.[11]

Proof-of-concept (POC) studies are designed to give preliminary evidence of efficacy and safety, with the aim to inform a decision about proceeding into full development of the drug. In practice, the POC decision is often based on whether a required effect size can be detected in comparison to placebo or a comparator treatment; be able to answer the addressed question within a reasonable time frame and the allotted budget; and the studies should be as small as possible. The study size is usually determined by the primary objective of the trial[12] and the number of subjects should be large enough to be able to detect the defined drug effect but at the same time expose a minimum number of subjects to an experimental drug. It is also common that dose-ranging POC studies are performed to address a secondary objective of exploring a dose–response relation.

In this study, we present two motivating examples, within the areas of acute stroke and type 2 diabetes, in which the use of a pharmacometric models has the potential to reduce sample size in POC studies. The examples were addressed through clinical trial simulations using previously developed pharmacometric models.

Drug development within the area of acute stroke has for many years struggled to find a successful drug, and several late phase failures[13–15] have been experienced. One of the issues is the inefficiency of the clinical trials, and several suggestions have been made on how to improve the efficiency of these trials.[16,17] With the development of pharmacometric models for the NIH stroke scale, the Barthel index[18] and the Scandinavian stroke scale,[19] the possibility of using a model-based analysis on clinical trials within the stroke area have emerged.

Type 2 diabetes mellitus is a progressive disease with continuous worsening of glycemic control.[20] Fasting plasma glucose (FPG) and insulin levels are often used as short-term assessments of the disease; however, for long-term assessments of the disease state, these levels are not reliable. Instead, glycosylated hemoglobin (HbA1c) is commonly used as a biomarker for long-term glycemic control. Hamrén and colleagues[21] have developed a mixed-effects mechanistic model for the interplay between FPG, HbA1c, and red blood cells (RBC). Such a mechanism-based model may offer great advantages in the analysis of clinical trials because it allows for a mechanistic interpretation of the result as well as a simultaneous analysis of multiple end points.

The aim of this article was to compare a pharmacometric model–based analysis to a conventional statistical analysis

[1]Department of Pharmaceutical Biosciences, Uppsala University, Uppsala, Sweden; [2]Exprimo NV, Mechelen, Belgium. Correspondence: KE Karlsson (Kristin.Karlsson@farmbio.uu.se)
Received 31 July 2012; accepted 19 November 2012; advance online publication 16 January 2013. doi:10.1038/psp.2012.24

with respect to power of detecting a defined drug effect in stroke and diabetes POC studies. The pharmacometric models applied for trial simulation and likelihood ratio testing (LRT) are presented in **Figures 1** and **2** and detailedly explained in the Methods section.

RESULTS

Two study designs were explored: a pure POC design with a placebo and an active arm; a dose-ranging scenario. For simplicity, the comparison between conventional study power and pharmacometric model–based power were made at 80% study power in all examples and scenarios; however, the full power curves are presented in the graphs.

In the POC stroke example, using a two-sided *t*-test to detect a difference in the change from baseline and day 90 National Institutes of Health Stroke Scale (NIHSS) score (using last observation carried forward) between placebo and the active dose group resulted in a study size of 388 patients (194 patients/arm), visualized in **Figure 3a**. Using a pharmacometric model–based approach, the 80% study power was reached with a study size of 90 patients (45 patients/arm), resulting in a 4.3-fold difference in total study size between the two methods. In the diabetes example, the conventional power calculation resulted in a study size of 84 patients (42 patients/arm) and the pharmacometric approach resulted in a study size of 10 patients (5 patients/arm), presented in **Figure 3b**, corresponding to an 8.4-fold difference between the two methods. The pharmacometric model–based 80% study power assessed with Monte-Carlo Mapped Power (MCMP, further described in the Methods section) was verified by stochastic simulations and estimations (data not shown). Both the investigated examples show a several fold reduction in study sizes when employing a model-based analysis. The reasons that the diabetes trial benefits the most are as follows: (i) FPG inherently contain more information (i.e., more sensitive to a drug effect as compared with stroke scores) than stroke scores, and (ii) a more informative study design which included a run-in phase to separate the placebo effect (which contained most of the between patient variability) from the drug effect, and a total of 10 repeated measurements of FPG were obtained.

The dose-ranging POC study scenario also resulted in a several fold difference between the two analysis methods for both disease areas, as visualized in **Figure 4**. In the stroke example, the pharmacometric approach resulted in a total study size of 184 patients and the *t*-test based study size was 776 patients (i.e., a 4.3 factor difference), as displayed in **Figure 4a**. Both the investigated study designs resulted in several fold larger study size to ensure similar power between the two analysis methods. However, due to the linear drug effect, the factor difference in study sizes remains the same when adding low and median dose groups. In the diabetic example, using the *t*-test to detect a significant difference between the placebo and the active treatment resulted in a total study size of 168 patients (42 patients/arm) to reach an 80% power, as shown in **Figure 4b**. The sample size required to reach the same power using the pharmacometric model–based approach resulted in a study size of 12 patients (three patients/arm), resulting in a 14 factor difference in study size between the two methods. The reasons for the increased difference between the methods, as compared with the pure POC scenario, are the nonlinear exposure–response relation that is more informed by multiple dose groups, and the inclusion of a follow-up observation adding more support to the drug effect.

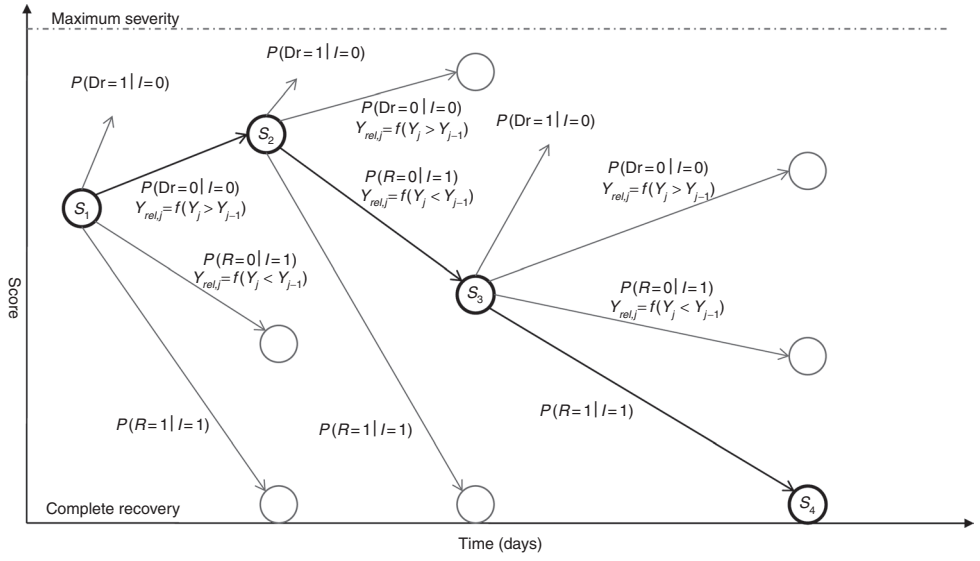

Figure 1 A schematic illustration of the concept of the National Institutes of Health Stroke Scale model, in which a zero score represent complete recovery and a score of 42 represent maximum severity. S_1, S_2, S_3, and S_4 are observed scores at observations 1, 2, 3, and 4, respectively. Gray circles indicate potential scores after each type of transition (which, in reality, could be any value between the score minimum and the last observation in the event of a score decline, between the last observation and one unit below the score maximum in the event of a score improvement, or the score maximum). Bold lines indicate actual score progression, whereas gray lines represent events that were possible, but did not take place, at every transition. $P(R=1|I=1)$, $P(R=0|I=1)$, $P(\text{Dr}=0|I=0)$, and $P(\text{Dr}=1|I=0)$ are the probabilities of reaching maximum score, improvement in score, decline in score, and dropout, respectively. $Y_{\text{rel},j} = f(Y_j > Y_{j-1})$ and $Y_{\text{rel},j} = f(Y_j < Y_{j-1})$ indicate the continuous functions describing the relative score change given a decrease, or an increase in disease state, respectively.

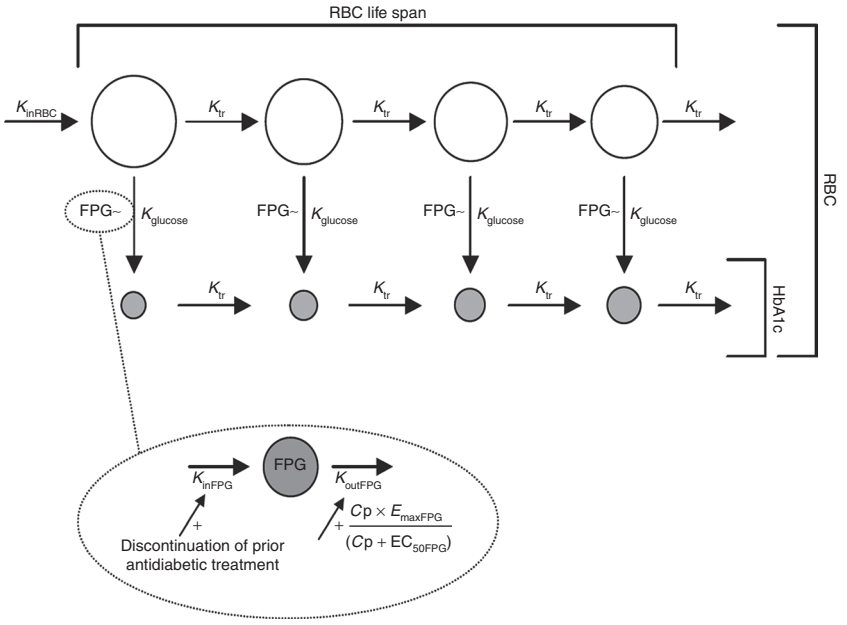

Figure 2 Schematic representation of the mechanism-based model for the FPG–HbA1c relationship, in which plasma concentration-*C*p vs. FPG effect is described by a sigmoidal E_{maxFPG} function including E_{maxFPG}, the maximum effect on K_{outFPG} the first-order degradation rate constant of FPG and EC_{50FPG}, the plasma concentration achieving half-maximal effect on E_{maxFPG}; red blood cell (RBC) maturation is described by K_{inRBC}, a zero-order rate constant of RBC release in blood circulation and K_{tr}, a first-order transit rate constant between each maturation stage; FPG mechanism is described by zero-order production rate constant of FPG and a glycosylation rate $K_{glucose}$ from RBCs to HbA1c. FPG, fasting plasma glucose; HbA1c, glycosylated hemoglobin. Reprinted with permission from Macmillan Publishers, ref. 21 copyright 2008.

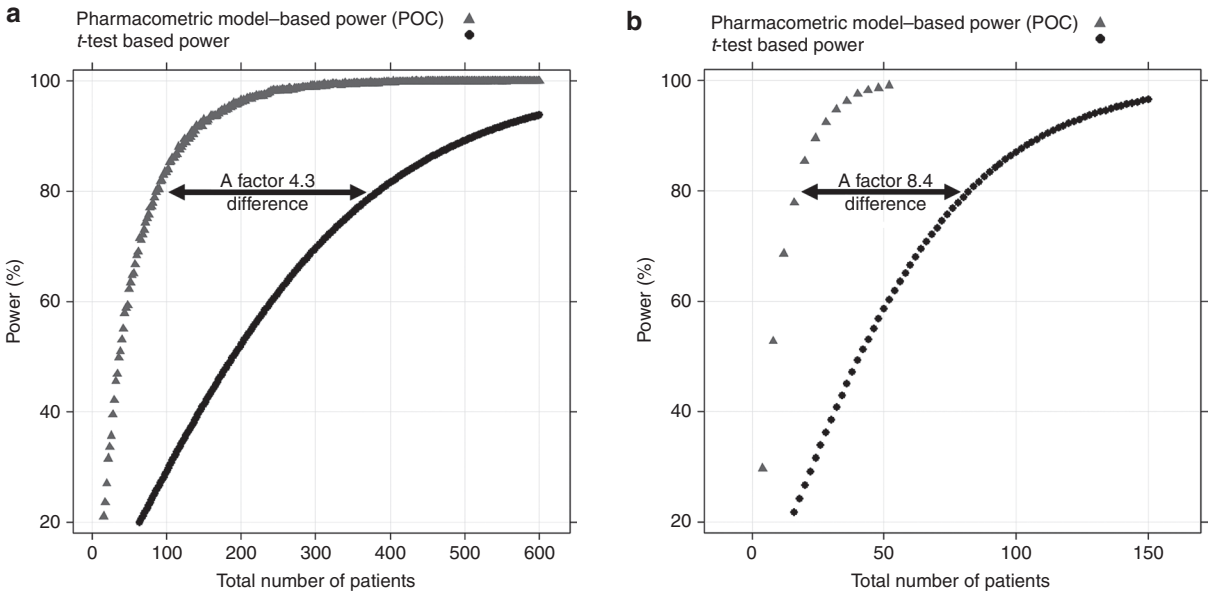

Figure 3 Power curve comparison between the pharmacometric model–based power (gray triangles) and the *t*-test based power (black diamonds), for the proof-of-concept scenario. (**a**) The power curves for the stroke example in which the difference in study size is a factor of 4.3 (90 vs. 388 total number of patients) is displayed. (**b**) In the diabetes example, the difference in study size was 8.4-fold (10 vs. 84 total number of patients) in favor of the pharmacometric approach.

DISCUSSION

POC studies (phase 2A) are often categorized as the first confirmatory trial in a drug development program,[22] and it is not uncommon that the primary analysis is similar to the analyses used in the phase 3 trials. This is unfortunate because the informativeness of the trial is diluted when, for example, using end of study observations only, discarding all other information. The first example, for each therapeutic area, represents a pure POC scenario in which the objective is to detect a defined drug effect between one active dose and placebo. However, because POC trials are often executed with multiple treatment

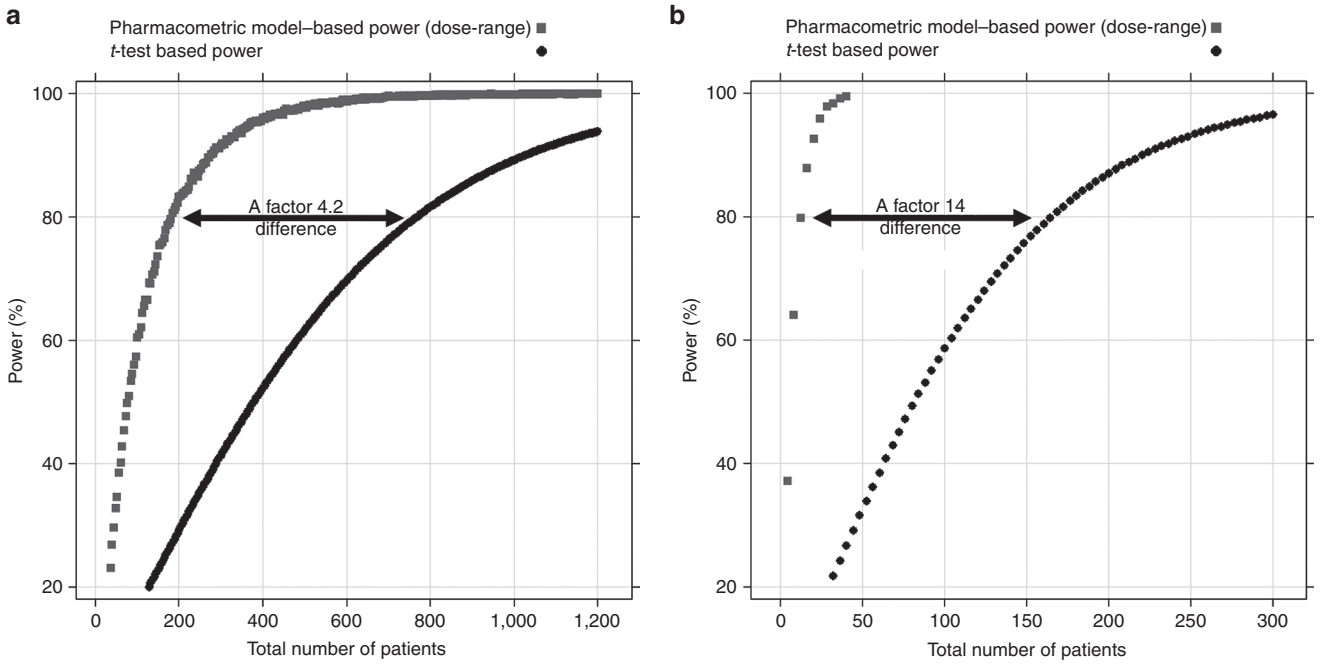

Figure 4 Power curve comparison between the pharmacometric model–based power (gray squares) and the *t*-test based power (black diamonds), for the dose-range scenario, with four parallel arms. (**a**) The power curves for the stroke example in which the difference in study size is a factor of 4.2 (184 vs. 776 total number of patients) is displayed. (**b**) In the diabetes example, the difference in study size was 14-fold (12 vs. 168 total number of patients) in favor of the pharmacometric approach. The *t*-test was based on the difference between placebo and the highest dose group and the total study size was calculated with addition of two equal sized treatment arms.

arms to fulfill secondary objectives such as exploring dose–response relations, a dose-ranging POC scenario was also investigated. Multiple active doses can contain valuable information to inform the POC decision, as well as support a dose/exposure–response relation in a pharmacometric model–based approach. In a conventional approach, *t*-tests are often applied individually on each treatment arm in comparison to placebo which make interpolations between treatment arms difficult and reduce the ability to propagate knowledge about dose/exposure–response to future studies.

As these results show, the use of a pharmacometric model–based approach within drug development has the potential to reduce study sizes of clinical trials. One of the main reasons for this is the use of longitudinal data as the pure POC results show. The POC example contains minimum information about the drug effect, involving only one active treatment arm and placebo, nevertheless by including all data available (i.e., repeated measurements) the pharmacometric approach results in a several fold reduction in study size when addressing the question of POC. Mixed-effects modeling is a powerful and flexible method when dealing with unbalanced repeated measurements,[23] which is often the case in clinical trials, and utilization of a pharmacometric model–based analysis does not necessarily mean that the design of the trial has to change, just that all available data are used in the primary analysis thereby increasing the information content of the trial, as these examples clearly illustrate.

We acknowledge that there are many methods available for the statistical analysis of clinical trials.[24–28] In this study, we have chosen to compare the *t*-test and pharmacometric modeling as these two methods can be viewed as the two extremes in terms of statistical power. The statistical power using other methods can be expected to fall somewhere in between the ones from the *t*-test and the pharmacometric modeling. There are few comparisons of statistical power between pharmacometric model–based analyses and other statistical methods; however, both Jonsson and Sheiner[9] and Hooker *et al*.[29] have presented results that indicate that model-based methods lead to a reduction in study sizes.

As already mentioned, POC trials and dose-ranging trials are often combined into one single phase 2 study to address both POC and dose-finding questions within the same trial. In a pharmacometric analysis, the aim is to detect a drug effect by establishing a model for the dose/exposure–response relation and naturally that will be more informed if multiple levels of doses are included. This is particularly true if the relation is nonlinear, as the results from the diabetes example indicate. In the stroke example, the drug effect was linear with respect to dose which resulted in the same factor difference in study sizes between the POC and dose-range scenarios. However, although not explored in the present investigation, the precision of the drug parameter will most likely increase with the addition of more dose levels.

Pharmacometric models have other advantages such as mechanistic interpretation of the model parameters and simultaneous analysis of multiple end points, as exemplified with the HbA1c model. Furthermore, a pharmacometric model–based power can be combined with a formal optimal design to pin down the most informative clinical trial design in terms of both study power and parameter precision.

The use of pharmacometric models when calculating the study power is of course dependent on the availability

Table 1 Examples of pharmacometric models available in the literature

Disease area	Clinical end point	Reference(s)
Acute and chronic pain	COX-2	Kowalski *et al*[41]
Alzheimer's disease	Adas-cog	Holford and Peace,[42] Ito *et al.*[43,44]
Hepatitis C	Sustained virologic response	Snoeck *et al.*[45]
Non-small cell lung cancer	Survival	FDA[46]
Obesity	% weight loss	FDA[46]
Parkinson's disease	Unified Parkinson's disease rating scale	Holford *et al.*,[47] FDA[39]
Rheumatoid arthritis	ACR20	Lacroix *et al.* [48]
Sleep disorders	Sleep stage	Karlsson *et al.*,[49] Bizzotto *et al.*[50]

FDA, Food and Drug Administration.

of a pharmacometric model. Several pharmacometric models exist in the literature for many clinical end points, and **Table 1** consists of a nonexhaustive list of pharmacometric model candidates in different therapeutic areas. Alternatively, a model for placebo treatment can often be generated from data from previous trials, if available to the investigator. If the compound is a follow-up compound, it may be possible to use a model developed for the predecessor, or if the compound is first in class, model developed in preclinical or early phase studies can be used or a selection of hypothesized models can be used to create a "best guess." However, if the information about a model is very limited, a model-based power calculation may not be sensible.

The model-based results in this study rely on the assumptions of no model misspecification and the detection of a drug effect different from zero. It is reasonable to believe that model misspecifications will lead to imprecision in the statistical power[30] and in the case in which the model is so uncertain that further model building needs to be done, the analysis will suffer from uncontrolled type I error rate. Although no clinical relevance criterion was applied in the power calculations, a clinically relevant drug effect was used in the simulation of data. In a real-life scenario, the clinical relevance should be evaluated by clinical trial simulations in which you have the option to investigate various outcome measurements, optimal doses, dose regimens, and the impact of possible uncertainties in the model.

The study power calculations based on pharmacometric models historically often rely on simulation and estimation exercises which can be very time consuming and, therefore, not extensively used. The newly developed MCMP method[31] for calculating the study power has the advantage of being a much faster method than the traditional simulation and estimation procedures, making a pharmacometric model–based power calculation more accessible. The increased speed also enables investigations of multiple pharmacometric models and parameter values to explore the assumptions made in the pharmacometric model–based approach.

The members of the Pharmaceutical Research and Manufacturers of America Proof of Concept Working Group recently recommended a more complex POC definition:[32] POC is the earliest point in the drug development process at which the weight of evidence suggests that it is "reasonably likely" that the key attributes for success are present and the key causes of failure are absent. To obey this definition, it is necessary to combine information about the drug from several sources, not only from a single efficacy study, and a pharmacometric analysis is well suited for including data from several studies and to combine models for both efficacy and safety into a single quantitative POC metric. Ideally, one could also include health economic aspects to further inform the POC decision.

Obvious benefits of reduced study sizes are the reduced costs and that fewer patients/volunteers will be exposed to an experimental drug before efficacy can be confirmed. This together with the increased potential for new drugs to reach the market faster provide strong incentives to consider a pharmacometric approach in the planning of a POC study.

METHODS

In the two example applications, a fixed study design was simulated and the results were analyzed using a pharmacometric model–based approach, as described below, and a conventional statistical analysis using a *t*-test. The pharmacometric based power, in both examples, was assessed by using the MCMP tool implemented in PsN version 3.2.7 (PsN, Uppsala University, Uppsala, Sweden)[33] and NONMEM version 7.1.2 (ICON Development Solutions, Ellicott City, MD),[34] run on a Linux cluster with a Red Hat 9 operating system using OpenMosix and a G77 Fortran compiler. The *t*-tests were performed in the statistical software R, version 2.11.1.

MCMP. The power/sample size calculation method proposed by Vong *et al.*[31] is based on the hypothesis testing principle of the LRT in nonlinear mixed-effect models. Several studies have investigated the performance of the LRT in nonlinear mixed-effects models, and the performance is generally good with type I error rates close to the nominal χ^2 distribution.[35–37] The MCMP method is a faster alternative to the multiple simulations and estimations of studies with different study size that constitutes the traditional power/sample size calculation method for LRT in mixed-effect models.[38] With the MCMP method, multiple random samples of individual objective function values (iOFV) are used as a substitute for multiple simulated and estimated studies. This substitution is based on the fact that the individual OFV sum up to the overall OFV of a model for a given data set as shown in Eq. 1 (where *iL* denotes the individual likelihood and *L* the total likelihood).

$$\mathrm{OFV} = -2\ln(L) = \sum_{j=1}^{n} -2\ln(iL_j) = \sum_{j=1}^{n} i\mathrm{OFV}_j \qquad (1)$$

The hypothesis of a possible drug effect can be tested with the LRT by assessing the difference in the OFV (ΔOFV) between two nested models (i.e., including or not including the hypothesized drug effect). The ΔOFV follows a χ^2 distribution with degrees of freedom corresponding to the difference in number of parameters between the two competing models. A model that corresponds to the null hypothesis of no drug effect will hereafter be referred to as a reduced model, and a model corresponding to the alternative hypothesis of an existing drug effect will be referred to as a full model. With the MCMP method, iOFV values estimated with a single full and single

reduced model are used to calculate ΔiOFV (Eq. 2). The sum of *n* randomly sampled ΔiOFV is used as a surrogate for the ΔOFV of a study with *n* number of subjects (Eq. 3). The single estimation step is performed with a large data set (typically ≥20 times the sample size needed for 80% power) simulated under the full model to form a large pool of ΔiOFVs.

$$\Delta iOFV = iOFV_{FULL} - iOFV_{REDUCED} \quad (2)$$

$$\Delta OFV = \sum_{j=1}^{n} \Delta iOFV_j \quad (3)$$

The ΔOFV calculation is repeated 10,000 times and the study power is computed as the percentage of ΔOFVs out of 10,000 scenarios that are greater than the significance level criterion defined by the LRT. The procedure is repeated with varying sample size (e.g., in increments of one patient) to map the power vs. sample size relationship up to a defined maximum power of interest.

Clinical trial design. In both investigated examples, two study scenarios were defined: a pure POC study in which a placebo arm was compared with an active dose group, and a dose-ranging scenario with placebo and three active treatment arms in which the objectives were to address both a POC definition and explore the dose–response relationship. Data for placebo and three active doses were simulated. In the POC study, only the placebo and the highest dose group were used whereas in the dose-ranging study, all four study arms were used in the pharmacometric approach. The conventional study sizes were based on *t*-test comparing placebo and the highest dose group, and the size of the dose-ranging study was calculated under the assumption of four equal sized groups, i.e., the conventional dose-ranging study was twice the size of the conventional POC study.

Stroke example. A nonlinear mixed-effects model has previously been developed for stroke disease progression after an acute ischemic stroke,[18] assessed by the 42 point NIH stroke scale[39] (NIHSS). The model consists of three submodels for conditional probabilities reflecting the likelihood of disease improvement or deterioration, reaching complete recovery (i.e., NIHSS = 0 as in no neurological disability) and dropout of the study, in combination with two linear submodels for of the relative magnitude of improvement or deterioration (visualized in **Figure 1** and model code available in **Supplementary Appendix S1a** online). The model also includes covariates such as age and baseline NIHSS score. This structure of the model enables several options on where to introduce a drug parameter, depending on the mechanism of the drug. However, in this simulation study, the drug effect was only added linearly on the magnitude of improvement (i.e., relative score change given an improvement in disease state).

Data were simulated using four arms: placebo and three active doses. Score assessments were made at day 0, 7, 30, and 90. The dose–effect relation was calibrated such that a low, medium, and high dose level would result in 25, 33, and 55% increase in the proportion of fully recovered patients at end of study as compared with placebo (resulting in a drug parameter value of 0.1 and dose levels of 2.5, 3.8, and 5.8). The definition of a fully recovered patient was a NIHSS score of 0 or 1.[40] Fifty-five percent relative proportion of fully recovered patients was a clinically relevant effect, assuming an equal treatment effect as the potential competitor tissue plasminogen activator treatment.[40]

The pharmacometric model–based study power was defined as the power to detect a drug effect, i.e., the possibility to estimate a drug parameter different from zero with a 5% significance level. For the purpose of the model-based power calculations, a large data set comprising of 2,500 patients/arm were simulated and estimated under the full model. In the POC scenario, placebo and the 5.8 dose arm were used, and in the dose-ranging study, all four treatment arms were kept in the data set.

To generate a conventional power curve, calculations were made using a two-sided *t*-test ($P < 0.05$) under the assumption that the difference in average change from baseline values at end of study, between the highest dose and placebo was 1.75 with SDs of 6.23 and 5.98, respectively. These values were based on the same distribution of responses (population) as used in the MCMP calculations. Last observation carried forwarded was applied and patients with only a baseline score were omitted from the analysis population.

Diabetes example. A nonlinear mixed-effects model in type 2 diabetes mellitus has previously been developed by Hamrén *et al.*[21] to describe the mechanistic relationship between tesaglitazar exposure, FPG, HbA1c, and aging RBC. The model as shown in **Figure 2** (NONMEM control stream available in **Supplementary Appendix S1b** online) consists of three submodels including an indirect response model on the effect of drug exposure on FPG over time, a transit compartment model to describe the RBC life span with a zero-order release of RBC into blood circulation, and a model that also includes at any stage of the RBC maturation, a function describing the glycosylation of RBC into HbA1c related to the FPG level. The structure of the model allows the possibility to evaluate different mechanisms for the drug effect. However, this simulation study has only investigated one plausible mechanism, which is a drug effect increasing the rate of elimination (K_{out}) of FPG.

The dose-ranging study investigated for the diabetes example was similar to the original study described in Hamrén *et al.*[21] The study design featured a 6-week run-in period, 12 weeks treatment, and a 3-week follow-up period. In total, it included 11 visits and as many samples of FPG, HbA1c was measured at the beginning of run-in, start of treatment, after 6 weeks of treatment, and at the end of treatment. Four trough pharmacokinetic samples were sampled during the treatment period. The study was assumed to include only male subjects with previous antidiabetic treatment withdrawn at the start of the run-in period. The investigated drug was assumed to have similar pharmacokinetic and pharmacodynamic properties as tesaglitazar, but lacking the hemodiluting effect of tesaglitazar. Doses 0.1, 0.5, and 1 mg were tested in parallel with placebo. Simulations were performed based on the final model and parameters in Hamrén *et al.*[21] with two small adjustments to the pharmacokinetic model. The final published model included a small dose and time-dependent effect on clearance, these two effects were excluded in the simulations. Estimation was performed with the same true/full model corresponding as the alternative hypothesis and a model without any drug dependent effect on K_{out} corresponding as the null hypothesis.

The POC study design was identical to the dose-ranging example but with only placebo and a 1 mg dose. Information about pharmacokinetic and treatment follow-up was ignored during parameter estimation based on the simulated POC study. The full model included a categorical drug effect parameter on K_{out}. Log-normal between subject variability was estimated for the drug effect parameter. The corresponding reduced model included no drug dependent effect on any glucose parameters.

The large data set utilized for the MCMP calculations comprised of 500 patients/arm (i.e., in total 2,000 patients for dose-ranging trial and 1,000 patients for the POC). The simulated mean (SD) baseline characteristics for FPG and HbA1c were 9.85 (2.20) mmol/l and 7.1 (1.1)%, respectively. The 1 mg dose resulted in a mean (SD) placebo corrected change from baseline of 0.63 (1.02)% for HbA1c at the end of treatment. Power calculations for a two-sided *t*-test ($P < 0.05$) was carried out assuming a group wise comparison between placebo and the 1 mg dose. For the likelihood ratio–based power calculations with the MCMP method, a two degree of freedom difference (one fixed effect and one random effect parameter) was assumed in the POC example (ΔOFV > 5.99) and a three degree of freedom difference (two fixed effect and one random effect parameter) in the dose ranging study example (ΔOFV > 7.81).

Author Contributions. K.E.K., C.V., and M.B. wrote the manuscript. E.N.J. and M.O.K. designed the research. K.E.K., C.V., and M.B. performed the research.

Conflict of interest. The authors declared no conflict of interest.

Study Highlights

WHAT IS THE CURRENT KNOWLEDGE ON THE TOPIC?
Pharmacometric mixed-effects models are used in the design, analysis, and simulation of clinical trials as they are well suited for the use of all available data and mechanistic interpretations of model parameters. However, in the planning of clinical trials, pharmacometric models are rarely used as the basis for power calculations to inform on the sample size.

WHAT QUESTION DID THIS STUDY ADDRESS?
Can study sizes be reduced with a pharmacometric analysis of a POC clinical trial, as compared with a conventional analysis?

WHAT THIS STUDY ADDS TO OUR KNOWLEDGE
The present investigations show that pharmacometric models can be used in power calculations to inform on sample sizes in clinical trials.

HOW THIS MIGHT CHANGE CLINICAL PHARMACOLOGY AND THERAPEUTICS
In applying a pharmacometric model–based analysis, there is a potential of drastically reducing the required study size in POC clinical trials and thereby, reducing costs, time spent, and the number of patients exposed to an investigational drug.

1. FDA. Challenge and Opportunity on the Critical Path to New Medical Products. <http://www.fda.gov/downloads/ScienceResearch/SpecialTopics/CriticalPathInitiative/CriticalPathOpportunitiesReports/ucm113411.pdf> (2004). Accessed 28 July 2012.
2. Orloff, J. *et al.* The future of drug development: advancing clinical trial design. *Nat. Rev. Drug Discov.* **8**, 949–957 (2009).
3. Price Waterhouse Coopers. Pharma 2020 Series. <http:// www.pwc.com/pharma> (2007–2011). Accessed 28 July 2012.
4. Goldberger, M.J. *et al.* ASCPT Task Force for advancing pharmacometrics and integration into drug development. *Clin. Pharmacol. Ther.* **88**, 158–161 (2010).
5. Barrett, J.S., Fossler, M.J., Cadieu, K.D. & Gastonguay, M.R. Pharmacometrics: a multidisciplinary field to facilitate critical thinking in drug development and translational research settings. *J. Clin. Pharmacol.* **48**, 632–649 (2008).
6. Lalonde, R.L. *et al.* Model-based drug development. *Clin. Pharmacol. Ther.* **82**, 21–32 (2007).
7. Wang, Y. *et al.* Leveraging prior quantitative knowledge to guide drug development decisions and regulatory science recommendations: Impact of FDA pharmacometrics during 2004–2006. *J Clin Pharmacol* **48**, 146–156 (2008).
8. Zhang, L. *et al.* Model-based drug development: the road to quantitative pharmacology. *J. Pharmacokinet. Pharmacodyn.* **33**, 369–393 (2006).
9. Jonsson, E.N. & Sheiner, L.B. More efficient clinical trials through use of scientific model-based statistical tests. *Clin. Pharmacol. Ther.* **72**, 603–614 (2002).
10. Powell, J.R. & Gobburu, J.V. Pharmacometrics at FDA: evolution and impact on decisions. *Clin. Pharmacol. Ther.* **82**, 97–102 (2007).
11. Lockwood, P., Ewy, W., Hermann, D. & Holford, N. 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. *Pharm. Res.* **23**, 2050–2059 (2006).
12. Industry., F.a.D.A.C.a.C.G.f. E9 Statistical Principles for Clinical Trials. (1998).
13. Lees, K.R. *et al.* NXY-059 for acute ischemic stroke. *N. Engl. J. Med.* **354**, 588–600 (2006).
14. Krams, M., Lees, K.R., Hacke, W., Grieve, A.P., Orgogozo, J.M. & Ford, G.A. Acute Stroke Therapy by Inhibition of Neutrophils (ASTIN): an adaptive dose-response study of UK-279,276 in acute ischemic stroke. *Stroke.* **34**, 2543–2548 (2003).
15. Hacke, W. *et al.* Intravenous desmoteplase in patients with acute ischaemic stroke selected by MRI perfusion-diffusion weighted imaging or perfusion CT (DIAS-2): a prospective, randomised, double-blind, placebo-controlled study. *Lancet Neurol.* **8**, 141–150 (2009).
16. Optimising Analysis of Stroke Trials Collaboration. Calculation of sample size for stroke trials assessing functional outcome: comparison of binary and ordinal approaches. *Int J Stroke* **3**, 78–84 (2008).
17. Young, F.B., Lees, K.R. & Weir, C.J. Improving trial power through use of prognosis-adjusted end points. *Stroke.* **36**, 597–601 (2005).
18. Karlsson, K.E., Wilkins, J.J., Jonsson, F., Zingmark, P.H., Karlsson, M.O. & Jonsson, E.N. Modeling disease progression in acute stroke using clinical assessment scales. *AAPS J.* **12**, 683–691 (2010).
19. Jonsson, F., Marshall, S., Krams, M. & Jonsson, E.N. A longitudinal model for non-monotonic clinical assessment scale data. *J. Pharmacokinet. Pharmacodyn.* **32**, 795–815 (2005).
20. Skelly, A.H. Type 2 diabetes mellitus. *Nurs. Clin. North Am.* **41**, 531–47, vi (2006).
21. Hamrén, B., Björk, E., Sunzel, M. & Karlsson, M. Models for plasma glucose, HbA1c, and hemoglobin interrelationships in patients with type 2 diabetes following tesaglitazar treatment. *Clin. Pharmacol. Ther.* **84**, 228–235 (2008).
22. Sheiner, L.B. Learning versus confirming in clinical drug development. *Clin. Pharmacol. Ther.* **61**, 275–291 (1997).
23. Marie Davidian, D.M.G. Nonlinear Models for Repeated Measurement Data (Chapman and Hall/CRC, Boca Raton, FL, 1995).
24. Statistical guide for clinical pharmacology & therapeutics. *Clin Pharmacol Ther* **88**, 150–152 (2010).
25. Mallinckrodt, C.H., Clark, S.W., Carroll, R.J. & Molenbergh, G. Assessing response profiles from incomplete longitudinal clinical trial data under regulatory considerations. *J. Biopharm. Stat.* **13**, 179–190 (2003).
26. Mallinckrodt, C.H., Kaiser, C.J., Watkin, J.G., Molenberghs, G. & Carroll, R.J. The effect of correlation structure on treatment contrasts estimated from incomplete clinical trial data with likelihood-based repeated measures compared with last observation carried forward ANOVA. *Clin. Trials* **1**, 477–489 (2004).
27. Siddiqui, O., Hung, H.M. & O'Neill, R. MMRM vs. LOCF: a comprehensive comparison based on simulation study and 25 NDA datasets. *J. Biopharm. Stat.* **19**, 227–246 (2009).
28. Wang, X. & Ting, N. A proof-of-concept clinical trial design combined with dose-ranging exploration. *Pharm. Stat.* **11**, 403–409 (2012).
29. Hooker, A.C., Ueckert, S., Andrew, M., Karlsson, M.O., Ito, K. & Corrigan, B. Model-based trial optimization for phase II and III designs in Alzheimer's disease. <http://www.go-acop.org/sites/default/files/webform/presentations/ACOP_HOOKER.pdf> (2011).
30. Plan, E.L., Karlsson, K.E. & Karlsson, M.O. Approaches to simultaneous analysis of frequency and severity of symptoms. *Clin. Pharmacol. Ther.* **88**, 255–259 (2010).
31. Vong, C., Bergstrand, M., Nyberg, J. & Karlsson, M.O. Rapid sample size calculations for a defined likelihood ratio test-based power in mixed-effects models. *AAPS J.* **14**, 176–186 (2012).
32. Cartwright, M.E. *et al.* Proof of concept: a PhRMA position paper with recommendations for best practice. *Clin. Pharmacol. Ther.* **87**, 278–285 (2010).

33. Lindbom, L., Ribbing, J. & Jonsson, E.N. Perl-speaks-NONMEM (PsN)–a Perl module for NONMEM related programming. *Comput. Methods Programs Biomed.* **75**, 85–94 (2004).
34. Beal, S.L., Sheiner, L.B., Boeckmann, A. & Bauer R.J. NONMEM User's Guide (1989–2009). (Icon Development Solutions, Ellicott City, MD, 2009).
35. Wählby, U., Bouw, M.R., Jonsson, E.N. & Karlsson, M.O. Assessment of type I error rates for the statistical sub-model in NONMEM. *J. Pharmacokinet. Pharmacodyn.* **29**, 251–269 (2002).
36. Wählby, U., Jonsson, E.N. & Karlsson, M.O. Assessment of actual significance levels for covariate effects in NONMEM. *J. Pharmacokinet. Pharmacodyn.* **28**, 231–252 (2001).
37. Wählby, U., Matolcsi, K., Karlsson, M.O. & Jonsson, E.N. Evaluation of type I error rates when modeling ordered categorical data in NONMEM. *J. Pharmacokinet. Pharmacodyn.* **31**, 61–74 (2004).
38. Lee, P.I. Design and power of a population pharmacokinetic study. *Pharm. Res.* **18**, 75–82 (2001).
39. Lyden, P. *et al.* Improved reliability of the NIH Stroke Scale using video training. NINDS TPA Stroke Study Group. *Stroke.* **25**, 2220–2226 (1994).
40. Tissue plasminogen activator for acute ischemic stroke. The National Institute of Neurological Disorders and Stroke rt-PA Stroke Study Group. *N Engl J Med* **333**, 1581–1587 (1995).
41. Kowalski, K.G., Olson, S., Remmers, A.E. & Hutmacher, M.M. Modeling and simulation to support dose selection and clinical development of SC-75416, a selective COX-2 inhibitor for the treatment of acute and chronic pain. *Clin. Pharmacol. Ther.* **83**, 857–866 (2008).
42. Holford, N.H. & Peace, K.E. Methodologic aspects of a population pharmacodynamic model for cognitive effects in Alzheimer patients treated with tacrine. *Proc. Natl. Acad. Sci. U.S.A.* **89**, 11466–11470 (1992).
43. Ito, K. *et al.* Disease progression model for cognitive deterioration from Alzheimer's Disease Neuroimaging Initiative database. *Alzheimers Dement.* **7**, 151–160 (2011).
44. Ito, K., Ahadieh, S., Corrigan, B., French, J., Fullerton, T. & Tensfeldt, T. Disease progression meta-analysis model in Alzheimer's disease. *Alzheimers. Dement.* **6**, 39–53 (2010).
45. Snoeck, E. *et al.* A comprehensive hepatitis C viral kinetic model explaining cure. *Clin. Pharmacol. Ther.* **87**, 706–713 (2010).
46. FDA. FDA Pharmacometrics. <http://www.fda.gov/ohrms/dockets/ac/06/briefing/2006-4248B1-04-FDA-topic-3-replacement.pdf> (2006).
47. Holford, N.H., Chan, P.L., Nutt, J.G., Kieburtz, K. & Shoulson, I. Disease progression and pharmacodynamics in Parkinson disease - evidence for functional protection with levodopa and other treatments. *J. Pharmacokinet. Pharmacodyn.* **33**, 281–311 (2006).
48. Lacroix, B.D., Lovern, M.R., Stockis, A., Sargentini-Maier, M.L., Karlsson, M.O. & Friberg, L.E. A pharmacodynamic Markov mixed-effects model for determining the effect of exposure to certolizumab pegol on the ACR20 score in patients with rheumatoid arthritis. *Clin. Pharmacol. Ther.* **86**, 387–395 (2009).
49. Karlsson, M.O. *et al.* A pharmacodynamic Markov mixed-effects model for the effect of temazepam on sleep. *Clin. Pharmacol. Ther.* **68**, 175–188 (2000).
50. Bizzotto, R., Zamuner, S., De Nicolao, G., Karlsson, M.O. & Gomeni, R. Multinomial logistic estimation of Markov-chain models for modeling sleep architecture in primary insomnia patients. *J. Pharmacokinet. Pharmacodyn.* **37**, 137–155 (2010).

Supplementary Information accompanies this paper on the *Pharmacometrics & Systems Pharmacology* website (http://www.nature.com/psp)