A Bayesian approach for dose-escalation in a phase I clinical trial incorporating pharmacodynamic endpoints

John Whitehead[1,*], Lisa Hampson[2], Yinghui Zhou[1], Edouard Ledent[3,4] and Alvaro Pereira[3,5]

[1]Medical and Pharmaceutical Statistics Research Unit, The University of Reading, UK
[2]Department of Mathematical Sciences, University of Bath, UK
[3] Lilly Research Laboratories, Eli Lilly and Company, Indianapolis, USA
[4]Present address: Biometrics, GSK Biologicals, Rixensart, Belgium
[5]Present address: Aepodia SA, Belgium

*j.r.whitehead@reading.ac.uk

ABSTRACT

Bayesian decision procedures have already been proposed for and implemented in phase I dose-escalation studies in healthy volunteers. The procedures have been based on pharmacokinetic responses reflecting the concentration of the drug in blood plasma and are conducted to learn about the dose-response relationship while avoiding excessive concentrations. However, in many dose-escalation studies, pharmacodynamic endpoints such as heart rate or blood pressure are observed, and it is these that should be used to control dose-escalation. These endpoints introduce additional complexity into the modelling of the problem relative to pharmacokinetic responses. Firstly, there are responses available following placebo administrations. Secondly, the pharmacodynamic responses are related directly to measurable plasma concentrations, which in turn are related to dose. Motivated by experience of data from a real study conducted in a conventional manner, this paper presents and evaluates a Bayesian procedure devised for the simultaneous monitoring of pharmacodynamic and pharmacokinetic responses. Account is also taken of the incidence of adverse events. Following logarithmic transformations, a linear model is used to relate dose to the pharmacokinetic endpoint and a quadratic model to relate the latter to the pharmacodynamic endpoint. A logistic model is used to relate the pharmacokinetic endpoint to the risk of an adverse event.

15 June 2007

1. Introduction

Phase I dose-escalation studies are generally conducted in small samples of healthy male volunteers in order to determine safe doses for later experimentation. The tolerability and safety of the experimental drug are closely monitored, with any adverse reactions being noted and pharmacokinetic responses such as AUC and C_{max} being recorded. Formal statistical designs for clinical trials in which binary adverse events are recorded have been studied by many authors, with the Continual Reassessment Method (O'Quigley, Pepe and Fisher, 1990) and its modifications being prominent in the literature. Recent summaries of that methodology can be found in Rosenberger and Haines (2002), Chevret (2006) and Ting (2006). Whitehead et al. (2001, 2006a) proposed Bayesian dose-escalation procedures for studies of pharmacokinetic responses from repeated dosing of healthy volunteers, Zhou et al. (2006) consider escalation based on bivariate data with one continuous and one binary component, and Zhou et al. (2007) describe a phase I study in which procedures for monitoring AUC and C_{max} values and incidence of adverse reactions were operated simultaneously. Piantadosi and Liu (1996) describe a dose-escalation procedure based on binary responses of toxicity in which pharmacokinetic measurements are taken into account. Their objective was to put observations of toxicity into the context of drug concentration rather than just dose administered, and as subjects were dosed only once in their design, the information was used in recommending doses for future subjects.

The methodology presented in this paper was motivated by a phase I dose-escalation trial that had been carried out according to a conventional design to investigate a compound aimed at treating Attention Deficit and Hyperactivity Disorder (ADHD). In addition to observing pharmacokinetic responses, investigators also recorded pharmacodynamic responses including pulse rate. The reason for interest in pulse rate was that the compound under study shares the same mechanism of action as

Atomoxetine, a highly specific presynaptic inhibitor of the noradrenaline (norepinephrine) transporter recently approved in the US for the treatment of patients with attention-deficit/hyperactivity disorder (ADHD). Adverse effects on the cardiovascular system, including abnormalities in heart rate, blood pressure, or cardiac rhythm have been associated with several noradrenergic medications. Under treatment, the mean heart rate and pulse rate increase as a consequence of the mechanism of action and its impact on the sympathetic and parasympathetic systems. The increases in blood pressure and pulse tends to occur early in therapy, stabilise, and then return toward baseline upon drug discontinuation (Wernicke et al., 2003). Because of the relationship between these effects and the mechnism of action of the drug on ADHD, elevation of pulse rate was used as a marker of pharmacological action.

2.   The motivating dataset

The data concern two cohorts of 9 healthy male volunteers. Doses investigated were 0, 20, 40, 80, 120, 180 and 200 mg. Cohort I participated in periods 1, 3 and 5 of the study, receiving doses 20, 80 and 180 mg in turn, while Cohort II participated in periods 2, 4 and 6, receiving doses 40, 120 and 200 mg. Within each period, six subjects received the active dose and three placebo, arranged so that each subject received two active doses over the course of the trial. Several measurements during each dosing period: here we restrict attention to the AUC, the pulse rate and the incidence of adverse events. A two week wash out interval separated consecutive dosing periods of the same cohort. Carry over of the drug from one treatment period to the next was neither anticipated nor observed. The study was conducted and interpreted in a conventional manner. In this section a model is fitted to the data to form the basis of a Bayesian approach developed in Section 3. In Section 4, illustrations of the Bayesian procedure are constructed by simulating realistic data from this model.

A model for AUC values was specified in terms of $y_{ij} = \log(AUC_{ij} + 1)$, where AUC_{ij} is the AUC for the ith subject during the jth period. The reason for adding 1s will be explained later, but other than that we follow Whitehead et al. (2001) and model y_{ij} according to $y_{ij} = \beta\ell_{ij} + s_i + \varepsilon_{ij}$, where $\ell_{ij} = \log(d_{ij} + 1)$ and d_{ij} is the dose administered to the ith subject during the jth period. The random subject effect s_i and the residual ε_{ij} are taken to be mutually independent normally distributed random variables with mean 0 and respective variances τ^2 and σ^2. The within subject correlation is $\rho = \tau^2/(\tau^2 + \sigma^2)$. Placebo data will not be used to fit the model, but it is constrained to pass through the origin since a placebo dose is certain to lead to an AUC of 0. Figure 1 shows the resulting fit of this model to the study data. Maximum likelihood estimates are $\hat{\beta} = 1.709$, $\hat{\tau}^2 = 0.340$ and $\hat{\sigma}^2 = 0.0554$, so that $\hat{\rho} = 0.860$.

FIGURE 1 NEAR HERE

The pulse rate of the ith subject during the jth period data was summarised as m_{ij}, the maximum increase in pulse rate over baseline during that period. A quadratic model in $\log(AUC + 1)$ was fitted to this endpoint: $m_{ij} = \theta_0 + \theta_1 y_{ij} + \theta_2 y_{ij}^2 + r_i + \delta_{ij}$, where the random subject effect r_i and the residual δ_{ij} are taken to be mutually independent normally distributed random variables with mean 0 and respective variances ζ^2 and ξ^2. The within subject correlation is $\eta = \zeta^2/(\zeta^2 + \xi^2)$. The model is fitted to data from all dosing periods, including placebo administrations, for which y_{ij} is taken to be zero. Figure 2 shows the resulting fit of this model. Maximum likelihood estimates are $\hat{\theta}_0 = 21.486$, $\hat{\theta}_1 = -2.169$, $\hat{\theta}_2 = 0.365$, $\hat{\zeta}^2 = 15.631$ and $\hat{\xi}^2 = 181.79$, so that $\hat{\eta} = 0.079$. The fit of the model appears to be satisfactory, with the relationship being close to horizontal between placebo and very low active doses, while at higher doses an increasing relationship is seen, as would be anticipated. The addition of 1 to the AUC before log transformation allows responses to placebo to be included in the model for pulse rate

against AUC, and the addition of 1 to dose values constrains the linear model relating AUC to dose to pass through the origin.

FIGURE 2 NEAR HERE

While it is accepted that the quadratic model does not reflect the precise relationship between pulse rate and AUC, it is proposed as a useful working model. One desirable property is that it will result in conservatism during dose-escalation, because the accelerating gradient at higher AUC values will give the appearance that high doses are particularly unacceptable. This would not be true, for example, of a split-line model.

Rather than counting all adverse events, only "dose limiting events" (DLEs) featured in analyses. These are loosely defined as events which would make investigators concerned about repeating that dose or administering higher doses, and more specifically taken to be any one of a list of events prepared before starting the study. The probability of a DLE will be modelled as a logistic regression on log(AUC+1). Denoting the probability that the ith subject suffers a DLE during the jth period by p_{ij}, the model logit(p_{ij}) = $\lambda_1 + \lambda_2 y_{ij}$ is fitted to data from active doses only. For the dataset available, maximum likelihood estimates were $\hat{\lambda}_1$ = −4.055 and $\hat{\lambda}_2$ = 0.347. The multivariate approach followed here allows us to relate the risk of a DLE to the amount of the drug *absorbed* (reflected by AUC), rather than the amount *administered* (reflected by dose). The former is thus closer to the underlying physiological mechanism, and makes some indirect allowance for the dependence of repeated observations on a subject.

The three models together describe the multivariate behaviour of AUC, pulse rate and DLE observations. It is being assumed that the occurrence of a DLE is conditionally independent of the pulse rate observation, given the AUC value. This is reasonable for the adverse events considered here, which took the form of vomiting,

headaches, urinary hesitation and dizziness. Conditions connected with heart rate, such as tachycardia, would not be accommodated in such a model. Tachycardia was on the list of events, but occurred only once during the course of the trial.

3. Bayesian modelling

For Bayesian modelling of the AUC data, a normal-gamma prior distribution is used for the parameters β and σ^2: $\beta \mid \sigma^2 \sim N(\mu_{A0}, \sigma^2 Q_{A0}^{-1})$; $\sigma^{-2} \sim Ga(a_{A0}, b_{A0})$, while a fixed value is set for ρ. A similar prior is imposed on the pulse rate data: $\boldsymbol{\theta} \mid \xi^2 \sim N(\boldsymbol{\mu}_{P0}, \xi^2 Q_{P0}^{-1})$; $\xi^{-2} \sim Ga(a_{P0}, b_{P0})$, with a fixed value set for η. The posterior distributions follow the pattern given in Whitehead et al. (2006a). For the DLE data, a Tsutukawa (1975) prior density $h_0(\lambda_1, \lambda_2)$ is used for λ_1 and λ_2, defined by

$$h_0(\lambda_1, \lambda_2) \propto \frac{\left\{\exp(\lambda_1 + \lambda_2 y_{-1})\right\}^{v_{-1}} \left\{\exp(\lambda_1 + \lambda_2 y_0)\right\}^{v_0}}{\left\{1 + \exp(\lambda_1 + \lambda_2 y_{-1})\right\}^{w_{-1}} \left\{1 + \exp(\lambda_1 + \lambda_2 y_0)\right\}^{w_0}},$$

where the parameters v_k and w_k reflect opinion about the risk of a DLE when $\log(AUC+1) = y_k$, $k = -1, 0$. This prior, and the corresponding posterior, is discussed by Whitehead and Williamson (1998).

The Bayesian decision procedure concerns the choice of dose to be administered to the ith subject in the jth treatment period. The pattern of administering placebo doses used in the real study is preserved in the new design. Thus, for the first cohort patients 1, 2 and 3 (randomly selected) receive placebo during the first treatment period, patients 4, 5 and 6 when they return for the third treatment period and patients 7, 8 and 9 during the fifth period. A similar system operates for the second cohort during treatment periods 2, 4 and 6. The Bayesian procedure is used to choose which active doses to administer to those due to receive them. By the jth treatment period, posterior models will have been developed from the data collected during the first $(j - 1)$ treatment periods. First, safety constraints will be set to avoid doses likely to lead to an excess

6

plasma concentration or a high risk of a DLE. Only doses for which the probability that $(y_{ij} \geq L)$ is less than 0.05 should be administered, and only doses for which the probability of a DLE is less than 0.2. The former probabilities are found from the predictive t-distribution of y_{ij}, which is described by Whitehead et al. (2006a), and estimated before each dosing period by replacing the unknown coefficients by their current posterior modal estimates. A posterior distribution is available for the subject effects relating to those who have previously been observed: for new subjects, the prior subject effect distribution is used. For the probability of a DLE to be less than 0.2, it is necessary that $\lambda_1 + \lambda_2 y_{ij} < \text{logit}(0.2) = -1.3863$. A simple approach would be to replace λ_1 and λ_2 by their posterior modal estimates, and y_{ij} by its expected value based on the dose d_{ij} administered and the posterior modal estimates of β and s_i, and determine whether the resulting inequality is valid. Here, we have used the predictive probability

$$P(\text{DLE}|\boldsymbol{\lambda}) = E\left\{\frac{\exp(\lambda_1 + \lambda_2 y)}{1 + \exp(\lambda_1 + \lambda_2 y)}\right\} = \int \frac{\exp(\lambda_1 + \lambda_2 y)}{1 + \exp(\lambda_1 + \lambda_2 y)} g(y) dy,$$

where g is the predictive t-density of y. Only doses for which this integral is less than 0.2 can be used. Once more λ_1 and λ_2 are replaced by their posterior modal estimates, but now variation in y and some of the dependency between the DLE risks of the same individual in different dosing periods is allowed for.

The set of available doses that satisfy the two safety constraints can be found for each subject in the next dosing period. These sets might differ from patient to patient according to current knowledge of their individual subject effect. The final recommendation from the subset of permissible available doses is determined according to the predictive distribution of the pulse rate. A desirable range of values (M_L, M_U) will be defined, M_L denoting the minimal maximum increase consistent with a beneficial effect and M_U the maximum safe increase. Ideally, the dose that maximises the probability that $m_{ij} \in (M_L, M_U)$ should be chosen. Here, we adopt the simpler but

more tractable criterion of choosing the dose that makes $E(m_{ij})$ closest to $\frac{1}{2}(M_L + M_U)$ amongst those in the available list of doses that satisfy the two safety constraints. The method for computing $E(m_{ij})$ is described in the Appendix.

4.   An illustration of the procedure

The settings chosen for this illustration were motivated by the real data collected in the completed trial. In practice, discussions with clinical investigators and consideration of data from previous, similar studies would be used to make these choices. The limiting value for $\log(AUC + 1)$ is to $L = 9.7$, corresponding to an AUC of approximately 16000 μg h mL^{-1}. The limits for maximum increase in pulse rate are set at $M_L = 10$ bpm and $M_U = 45$ bpm.

Fixed values for within subject correlations of the y_{ij} and the m_{ij} are taken as $\rho = 0.6$ and $\eta = 0.3$ will be imposed. Prior information concerning AUC values is expressed as two representative pseudo-observations on a single subject: AUC values of 550 and 6900 corresponding to doses 40 mg and 180 mg respectively are specified. These lead to a slope of a 1.7, consistently with the observed data. For the DLE model, following and Zhou and Whitehead (2003) prior opinion is expressed by supposing that 3 "pseudo-subjects" at dose 20 gave 0.6 of a DLE and 3 at dose 200 gave 1.5 DLEs. This leads to prior modal parameter estimates of $\lambda_1 = -3.255$ and $\lambda_2 = 0.361$.

Prior information on pulse rates is represented by three representative pseudo-observations on a single subject. These were maximum increases over baseline of 21, 25 and 50 bpm corresponding to AUC values of 0, 1000 and 22,000 respectively. The resulting prior modal parameter estimates are $\theta_1 = 21$, $\theta_2 = -4.611$ and $\theta_3 = 0.751$. This prior model predicts pulse rate values corresponding to doses in the region of 20mg to be similar to those expected under placebo.

To choose the gamma parameters for AUC, a_{A0} and b_{A0}, it was required that the lowest dose have a prior probability of an excessive AUC value equal to 0.05, while for the second dose this probability is 0.067 (see Whitehead et al.; 2001). This choice ensures that the only permissible dose for the first cohort will be the lowest dose. The parameter values $a_{A0} = 0.42529$ and $b_{A0} = 0.0355$ follow. It was not necessary to specify the gamma parameters for the prior distribution of ξ^{-2}, as the dose escalation procedure depends only on expectations of the m_{ij}. The data used in the following simulation examples were generated using the models described above, with AUC parameters set at $\beta = 1.709$, $\sigma^2 = 181.79$ and $\tau^2 = 15.631$, pulse rate parameters set at $\theta_0 = 21.5$, $\theta_1 = -2.17$, $\theta_2 = 0.36$, $\xi^2 = 0.0554$ and $\zeta^2 = 0.340$ and DLE parameters set at $\lambda_1 = -4.048$, and $\lambda_2 = 0.347$.

Each simulated run involved two cohorts of 9 healthy male volunteers and six dosing periods. Cohort I participated in periods 1, 3 and 5 and Cohort II in periods 2, 4 and 6. During each dosing period, three subjects received placebo, arranged so that each subject received placebo exactly once. The doses available were 0, 20, 40, 80, 120, 180 and 200 mg. Thus, the simulated studies resembled the real one, except that doses were chosen according to the recommendations of the Bayesian procedure. To enhance comparability of different design options, the investigation began with simulation of two subject responses to each of the available doses. Then in each run, subject responses were read from this complete dataset. This ensures that when two procedures coincide in dose allocation, the results of the simulated runs will be the same.

Before studying data from the simulated trials, the real data described in Section 2 will be revisited. Figure 3 summarises the responses to each active dose. Dose administered is plotted against dosing period. The symbol is heart-shaped if the maximum increase in pulse rate over baseline lies in the target interval (10, 45) bpm,

square if it lies below 10 and triangular if it lies above 45. The symbol is printed larger if the log(AUC + 1) value exceeded the limit of 9.7, (this occurs only in the actual dataset shown in Figure 3), and it is filled if a DLE was observed. The figure shows that DLEs were quite common, even at the lowest dose, and thus cannot have been considered to be a particularly vital limitation to escalation. At the highest dose, the AUC limit was exceeded twice. Suitable pulse rate increases were observed at all dose levels, including the lowest, indicating that raising the dose was not critical for the achievement of a potentially pharmacological effect.

FIGURE 3 NEAR HERE

Figure 4 shows the results from a simulated dose-escalation study, depicted in the same way as Figure 3. The data illustrate some important operating characteristics of the methodology. Excess AUC values are avoided completely. As satisfactory pulse rates are found from low doses, the higher range is not entered. Fewer DLEs are observed. One subject is given dose 80 during the last period: this subject exhibited individual effects indicating that a higher dose would be both safe and effective. Thus the procedure can tailor dosing to the individual subject.

FIGURE 4 NEAR HERE

The escalation shown in Figure 4 is conservative, with only one subject advancing beyond dose 40. Detailed scrutiny of the data suggests that this is because of the fear of DLEs. In the actual study, DLEs were observed, and yet escalation continued. This suggests that we need not be too concerned about their occurrence. Consequently, in the next run of the procedure, the tolerance relating to DLEs was altered to ensure that the risk was less that 0.3 rather than the stricter setting of 0.2. The result is shown in Figure 5. Doses up to 80mg are now administered to many subjects, and dose 120 appears twice. There are still no excessive AUC values, and a high proportion of the subjects have pulse rate increases in the target range.

FIGURE 5 NEAR HERE

Finally, a third refinement of the dose selection procedure was investigated. As subjects become used to taking the drug, and once it is established that no harmful rises in heart rate are experienced, the target values for the pulse rate measure can be progressively increased. Thus, for dosing periods 1, 2 and 3, M_L will set at 10, but it will be raised to 20 for periods 4 and 5 and finally to 30 for period 6. Within this scheme placebo subjects in Cohort II will start with $M_L = 20$, but the experience of Cohort I subjects dosed using the stricter criterion should provide sufficient assurance to allow this. Figure 6 shows the results from a simulated run using these criteria, and with the DLE tolerance set at 0.30. Now the top dose of 200 is reached, but still no excess AUC values are encountered. Most administrations result in an acceptable pulse rate. In this plot, the symbol is heart-shaped if the maximum increase in pulse rate over baseline lies in the current target interval, with lower limit 10, 20 or 30 according to dosing period: squares denote values that are too low and triangles values that are too high.

FIGURE 6 NEAR HERE

During all three simulation runs, very few dose allocations were constrained by the need to keep the AUC below its limiting value. This would occur in practice if the physiological target was reached well within the acceptable limit of plasma concentration. In conventional studies, there is often a desire to learn about the maximum tolerated dose of a drug. If therapeutic value can be found well within the tolerable range, knowledge of the maximum may not be pharmacologically important. It certainly may not be needed prior to the proof-of-concept studies that may lead to more extensive clinical development. Escalating to find the maximum tolerated dose is potentially one of the most hazardous parts of healthy volunteer testing, although it may

eventually be needed to determine the properties of or adverse reactions to the drug if overdosing occurs once the drug becomes widely available.

Table 1 presents maximum likelihood estimates from all three simulated runs. Maximum likelihood is used to avoid the influence of the prior information in this final analysis. The estimates from the run shown in Figure 6 are the most accurate and have the smallest standard errors, which follows because this run involved the widest range of doses. In the other runs, σ^2 is overestimated and τ^2 is underestimated, consistently with the findings of Whitehead et al. (2006a). In all runs ξ^2 is substantially underestimated. Table 1 also presents the maximum safe doses d^*_{AUC} and d^*_{DLE} under the AUC and DLE criteria, and the ideal dose d^*_{pulse} for achieving the target dose. These are underestimated in the runs shown in Figures 4 and 5 and overestimated in that shown in Figure 6.

TABLE 1 NEAR HERE

All runs were completed using the predictive probability $P(DLE \mid \lambda)$ described in Section 3. Earlier runs using the simpler approach in which y_{ij} is replaced by its expected value based on the dose d_{ij} administered gave the same dosing pattern. The exact values of the maximum safe doses d^*_{AUC} and d^*_{DLE} under the AUC and DLE criteria, and the ideal dose d^*_{pulse} for achieving the target dose were different, but the closest suitable doses amongst those available were unchanged.

5.   Discussion

This paper addresses issues of dose escalation in early phase studies in which a pharmacodynamic response that is predictive of the mechanism of action of the drug follows closely after administration of the drug and is quickly available. The rationale for anticipating that moderate elevation of pulse rate will indicate therapeutic value in

ADHD has been explained in Section 1. In other settings, different pharmacodynamic endpoints might be used. Whitehead et al. (2006b) discuss examples including the observation of anti-factor Xa activity as being indicative of thromboembolic disorders or inhibition of leukotriene B_4 being indicative of anti-inflammatory activity. In the current application, a quadratic model was used for the relationship between transformed versions of the pharmacodynamic and pharmacokinetic responses. This could be applied as a conservative alternative to a break point model or to the lower part of an E_{max} model: conservative in the sense that higher doses will appear more dangerous under the quadratic model and will thus be avoided until sufficient evidence of their safety has been collected. The authors recognise that each clinical study is unique, and that the details of other trials may be different from that studied here. Rather than presenting our method as a "one size fits all" approach, we hope that the considerations made in developing this design will serve as a useful model even when study details are different.

The objective of the methodology described is the identification of the dose that is most likely to have therapeutic value. More conventional approaches seek the maximum tolerated dose, but the importance and ethics of raising doses until harm is observed are doubtful if such activity occurs only well above the therapeutic range. Certainly, such questions become of interest only when the proof-of-concept stage has been passed, and it is clear that the drug has potential for widespread use. If necessary, the effects of higher doses can be addressed if and when that later stage has been reached.

The prior information for the approach will generally be derived from animal experimentation and from prior human experience with similar compound. The benefits of the methodology will depend on the extent and reliability of this prior information.

Sometimes there will be delays in processing pharmacokinetic data, so that observed AUC values are not available immediately after dosing. The Bayesian model can be updated using whatever data are available, and so can continue to be informative despite missing data. Nevertheless, every effort should be made to assemble most, if not all, of the necessary data in a timely fashion. Zhou et al. (2007) describe how this was successfully achieved in their study.

Some details of the scheme described could easily be varied. Summary measures for pulse rate other than the maximum could be used. For example, the weighted mean pulse rate (taken to be the area under the pulse rate-time curve) would penalise doses leading to a high pulse rate maintained over several hours, which might be of more concern than the occurrence of an isolated maximum value. Doses for the motivating study could have been made up from any combination of the available capsule sizes (10, 50 and 100 mg), and so more dose levels could have been created, and higher doses could have been made available. The dose-escalation scheme described in this paper does not require all available doses to be tried in turn, and so it can be used with a larger set of potential doses without extending its duration. The procedure homes in on the interesting doses, and it is advantageous to provide a finer grid of dose values than is conventionally done in order to identify the optimal choice more accurately.

The methodology described in this paper provides a framework from which the safety and pharmacodynamic effects of different doses can be foreseen, together with an expression of the reliability of such predictions. The method provides a formal recommendation of the doses to be used during the next dosing period, but a safety committee will usually weigh these against other information not captured by the formal scheme, and decide whether or not to accept the recommendation. The modelling provides more information than the final recommendation of doses, and this can be used

to help with decision making. In applying a similar dose-escalation scheme, Zhou et al. (2007) report that over 70% of formal dosing recommendations were accepted for administration, and that the fitted models were helpful in reaching the remaining dosing decisions.

Acknowledgement

References

Chevret, S. (Ed.) (2006). *Statistical Methods for Dose-finding Experiments*. Chichester: Wiley.

O'Quigley, J., Pepe, M. and Fisher, L. (1990). Continual reassessment method: a practical design for phase I clinical trials in cancer. *Biometrics* **46,** 33-48.

Piantadosi, S. and Liu, G. (1996). Improved designs for dose escalation studies using pharmacokinetic measurements. *Statistics in Medicine* **15**, 1605-1618.

Rosenberger, W.F. and Haines, L.M. (2002). Competing designs for phase I clinical trials: a review. *Statistics in Medicine* **21**, 2757-2770.

Ting, N. (Ed.) (2006). *Dose Finding in Drug Development*. New York: Springer.

Tsutakawa, R. K. (1975). Bayesian inference for bioassay. *Technical Report 52,* Mathematical Sciences, University of Missouri, Columbia.

Wernicke, J.F., Faries, D., Girod, D., Brown, J.W., Gao, H., Kelsey, D., Quintana, H., Lipetz, R., Michelson, D. and Heiligenstein, J. (2003). Cardiovascular effects of atomoxetine in children, adolescents, and adults. *Drug Safety* **26**, 729-740.

Whitehead, J. and Williamson, D. (1998). An evaluation of Bayesian decision procedures for dose-finding studies. *Journal of Biopharmaceutical Statistics* **8,** 445-467.

Whitehead, J., Zhou, Y., Mander, A., Ritchie, S., Sabin, A. and Wright, A. (2006a). An evaluation of Bayesian designs for dose-escalation studies in healthy volunteers. *Statistics in Medicine* **25,** 433-445.

Whitehead, J., Zhou, Y., Stevens, J., Blakey, G., Price, J. and Leadbetter, J. (2006b). Bayesian decision procedures for dose-escalation based on evidence of undesirable events and therapeutic benefit. *Statistics in Medicine* **25,** 37-53.

Whitehead, J., Zhou, Y., Patterson, S., Webber, D. and Francis, S. (2001). Easy-to-implement Bayesian methods for dose-escalation studies in healthy volunteers. *Biostatistics* **2,** 47-61.

Zhou, Y., Whitehead, J., Bonvini, E., Stevens, J. (2006). Bayesian decision procedures for binary and continuous bivariate dose-escalation studies. *Pharmaceutical Statistics* **5**, 125-133.

Zhou, Y. and Whitehead, J. (2003). Practical Implementation of Bayesian dose-escalation procedures. *Drug Information Journal* **37**, 45-59.

Zhou, Y., Whitehead, J., Korhonen, P. and Mustonen, M. (2007). Implementation of a Bayesian design in a dose-escalation study of an experimental agent in healthy volunteers. *Biometrics* (accepted).

Appendix: Derivation of the expected value of m_{ij}

According to the model of Section 2, $m_{ij} \mid R_1 \sim N(\theta_0 + \theta_1 y_{ij} + \theta_2 y_{ij}^2 + r_i, \xi^2)$ and

$y_{ij} \mid R_2 \sim N(\beta \ell_{ij} + s_i, \sigma^2)$, where $R_1 = \{y_{ij}, r_i, s_i, \theta, \beta, \xi^2, \sigma^2\}$ and $R_2 = \{r_i, s_i, \theta, \beta, \xi^2, \sigma^2\}$.

It follows that $E(m_{ij} \mid R_1) = \theta_0 + \theta_1 y_{ij} + \theta_1^2 y_{ij} + r_i$. We will now take expectations over

each of the conditioning variables in turn, in order to "unpeel" the expectation of m_{ij}.

Taking expectations over y_{ij}, remembering that $E(X^2) = \text{var}(X) + \{E(X)\}^2$, we obtain

$$E\left(m_{ij} \mid R_2\right) = \theta_0 + \theta_1\left(\beta \ell_{ij} + s_i\right) + \theta_2\left(\sigma^2 + \beta^2 \ell_{ij}^2 + 2s_i \beta \ell_{ij} + s_i^2\right) + r_i.$$

Now, in a manner similar to the derivation of equation (3.5) in Whitehead et al. (2001),

it can be shown that the posterior distribution of r_i, given θ, ξ^2 and η is normal, with

mean $w_{(1)i}\left\{\bar{m}_i - \theta_0 - \theta_1 \bar{y}_i - \theta_2\overline{\left(y_i^2\right)}\right\}$ and variance $\xi^2 w_{(1)i} / n_i$, and that the posterior

distribution of s_i, given θ, σ^2 and ρ is normal, with mean $w_{(2)i}\left\{\bar{y}_i - \beta\overline{\ell}_i\right\}$ and variance

$\sigma^2 w_{(1)i} / n_i$, where $w_{(1)i} = n_i \eta / \{1 + (n_i - 1)\eta\}$, $w_{(2)i} = n_i \rho / \{1 + (n_i - 1)\rho\}$, n_i is the

number of active doses that subject i has received and $\bar{m}_i$, $\overline{\ell}_i$, $\bar{y}_i$ and $\overline{\left(y_i^2\right)}$ are the means

of the m_{ij}, ℓ_{ij}, y_{ij} and y_{ij}^2 values recorded for this subject. Taking expectations over r_i

and s_i, we obtain

$$E\left(m_{ij} \mid R_3\right) = w_{(1)i}\bar{m}_i + \theta_0\left(1 - w_{(1)i}\right) + \theta_1\left[\beta \ell_{ij} - w_{(2)i}\beta \overline{\ell}_i - \left\{w_{(1)i} - w_{(2)i}\right\}\bar{y}_i\right]$$
$$+ \theta_2\left[\sigma^2\left\{1 + w_{(1)i}/n_i\right\} + \beta^2 \ell_{ij}^2 + 2w_{(2)i}\left\{\bar{y}_i - \beta\overline{\ell}_i\right\}\beta \ell_{ij} + \left\{\bar{y}_i - \beta\overline{\ell}_i\right\}^2 - w_{(1)i}\overline{\left(y_i^2\right)}\right]$$

where $R_3 = \{\theta, \beta, \xi^2, \sigma^2\}$. Finally, taking expectations with respect to θ and β, we

obtain $E(m_{ij})$ as the expression above with θ_0, θ_1, θ_2, and β replaced by their Bayes

estimates and β^2 replaced by the sum of its variance and the square of its estimate.

Captions for figures

Figure 1: *Scatter plot of log(AUC+1) data, with the fitted model superimposed. Dotted lines are the 95% confidence limits for the response of a new subject*

Figure 2: *Scatter plot of the pulse rate data against transformed AUC, with the fitted curve superimposed*
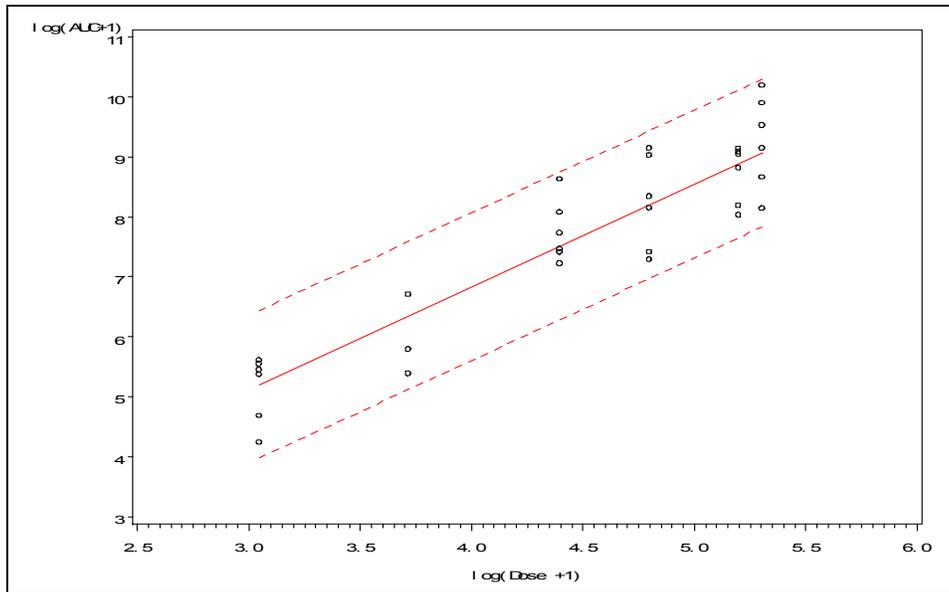
Figure 3: *Plot of the observed study data*

Figure 4: *A simulated dose-escalation study with a DLE tolerance of 0.2*

Figure 5: *A simulated dose-escalation study with a DLE tolerance of 0.3*

Figure 6: *A simulated dose-escalation study with a DLE tolerance of 0.3 and a variable pulse rate criterion*
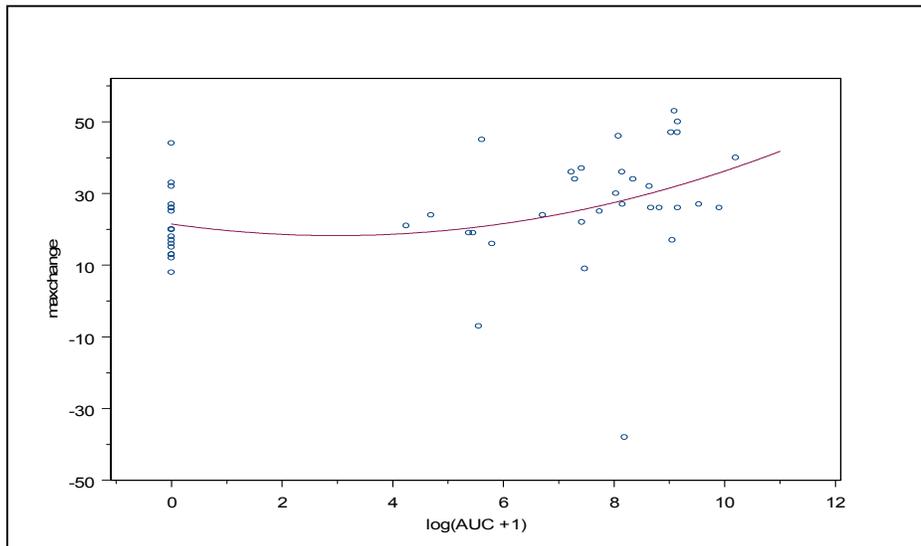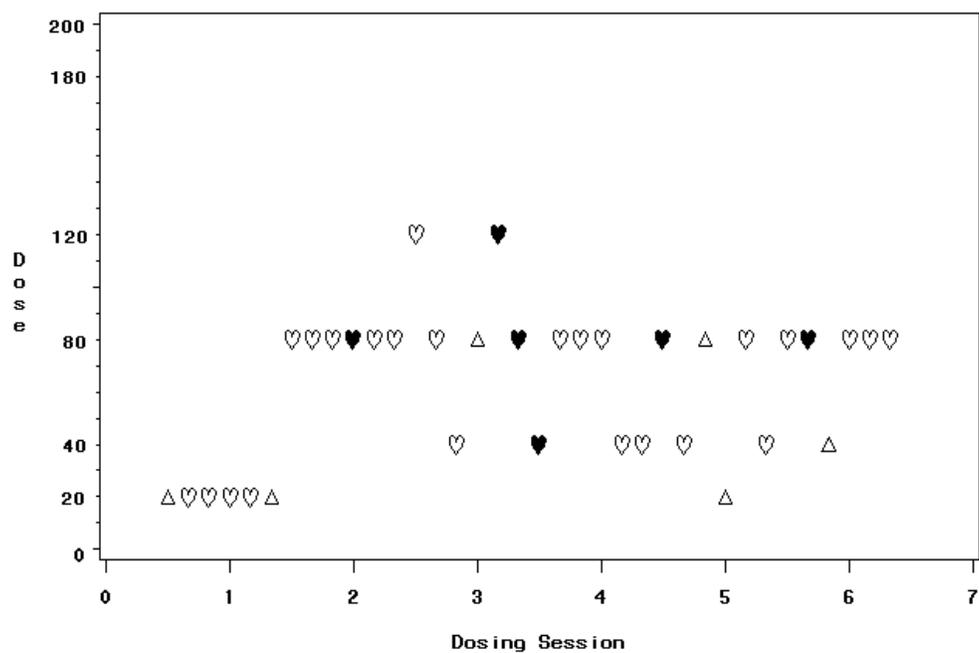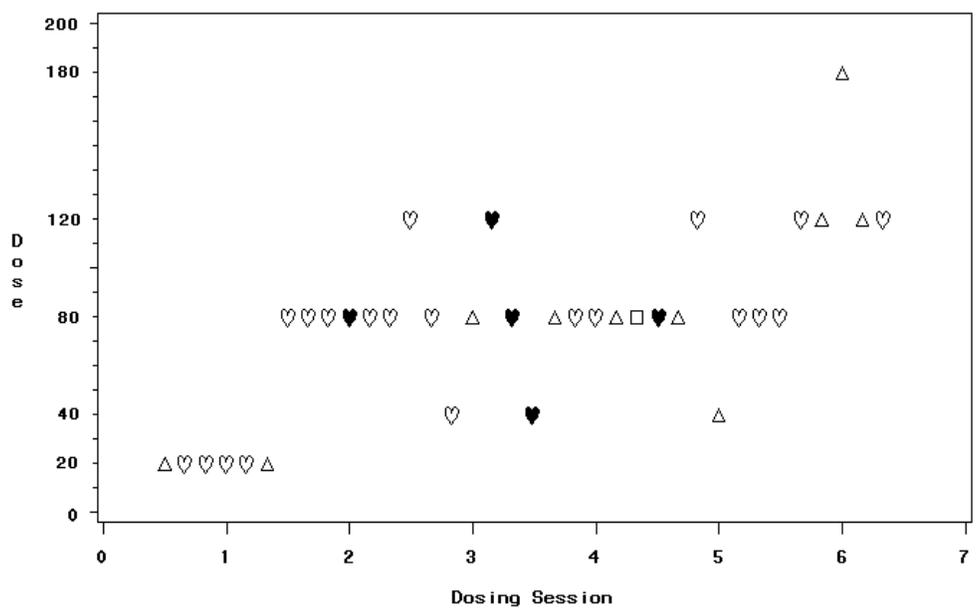
Figure 1

Figure 2

Figure 3

Figure 4

Figure 5

Figure 6

Table 1: *Maximum likelihood estimates of parameters following the three simulated runs*

Parameter	Truth	Figure 4	Figure 5	Figure 6
β	1.709	1.735	1.728	1.722
		(0.038)	(0.031)	(0.028)
σ^2	0.055	0.055	0.079	0.068
τ^2	0.340	0.311	0.264	0.253
θ_0	21.49	21.35	21.39	21.32
		(2.51)	(2.51)	(2.53)
θ_1	−2.169	−3.441	−5.153	−2.055
		(2.496)	(1.993)	(1.987)
θ_2	0.365	0.567	0.773	0.335
		(0.382)	(0.259)	(0.250)
ξ^2	182	106	103	94
ζ^2	15.6	7.9	10.0	21.2
λ_1	−4.055	−8.638	−9.210	−4.363
		(5.844)	(4.075)	(3.682)
λ_2	0.347	0.995	1.027	0.322
		(0.894)	(0.543)	(0.486)
d^*_{AUC}	158	150	156	162
d^*_{DLE}	90	64	76	151
d^*_{pulse}	104	73	83	124