Combining MCMC with 'sequential' PKPD modelling

**David Lunn · Nicky Best · David Spiegelhalter ·
Gordon Graham · Beat Neuenschwander**

Abstract We introduce a method for preventing unwanted feedback in Bayesian
PKPD link models. We illustrate the approach using a simple example on a sin-
gle individual, and subsequently demonstrate the ease with which it can be applied
to more general settings. In particular, we look at the three 'sequential' popula-
tion PKPD models examined by Zhang et al. (J Pharmacokinet Pharmacodyn
30:387–404, 2003; J Pharmacokinet Pharmacodyn 30:405–416, 2003), and provide
graphical representations of these models to elucidate their structure. An important
feature of our approach is that it allows uncertainty regarding the PK parameters to
propagate through to inferences on the PD parameters. This is in contrast to standard
two-stage approaches whereby 'plug-in' point estimates for either the population or
the individual-specific PK parameters are required.

D. Lunn (✉) · D. Spiegelhalter
Medical Research Council Biostatistics Unit, Institute of Public Health, University Forvie Site,
Robinson Way, Cambridge CB2 0SR, UK
e-mail: david.lunn@mrc-bsu.cam.ac.uk

N. Best
Department of Epidemiology and Public Health, Imperial College, London, UK

G. Graham
Pfizer Ltd., Sandwich, UK

B. Neuenschwander
Novartis Pharma AG, Basel, Switzerland

Introduction

It is natural to simultaneously model PKPD data through shared parameters, in particular when underlying concentrations estimated from a PK model are used as predictors in some PD model. This has become reasonably straightforward using either a likelihood [1] or a fully Bayesian [2, 3] approach. However, while theoretically sound, such 'full probability modelling' may have consequences that do not appear intuitively attractive. For example, [4] considered PKPD data from 27 rats, with 10–15 plasma midazolam concentrations and 32–50 EEG measurements per rat, taken over a five-hour time period. A plot of the PK data for one rat (given ~ 2.5 mg midazolam over a five-min intravenous infusion) is shown in Fig. 1, demonstrating the good fit provided by a two-compartment model (estimated using WinBUGS [5]).

A basic PD analysis might assume that clinical response depends on estimated concentrations, derived from the PK data, through the standard model:

$$E_j = E_0 + \frac{E_{\max} C_j}{EC_{50} + C_j}, \quad j = 1, \ldots, m,$$

where m is the number of PD observations, E_j denotes the jth PD observation, and C_j denotes the PK-model-predicted concentration corresponding to the time at which E_j was measured. When we simultaneously fit the PK and PD models for the rat whose data are depicted in Fig. 1, we obtain the fits shown in Fig. 2. The estimated concentrations now do not fit the PK data very well at all. This is due to the PK parameters now being influenced by the PD data as well as the PK data, and there existing some tension between what these two sources of information are saying (given the chosen PK and PD models). We emphasise that such *feedback* from the PD data is a consequence of the joint modelling, and so would happen within both a
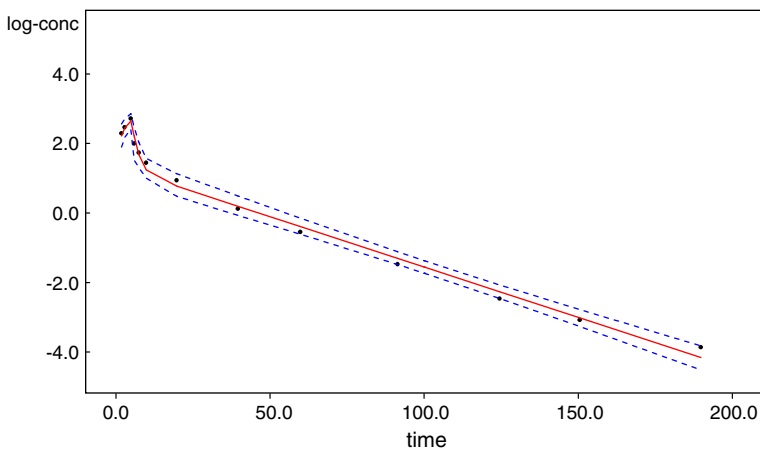

Fig. 1 Observed midazolam log-concentrations and predicted values (with 95% credible intervals) from fitted two-compartment model. Rat 7: •, observed values; —, posterior median; - - -, 95% credible interval
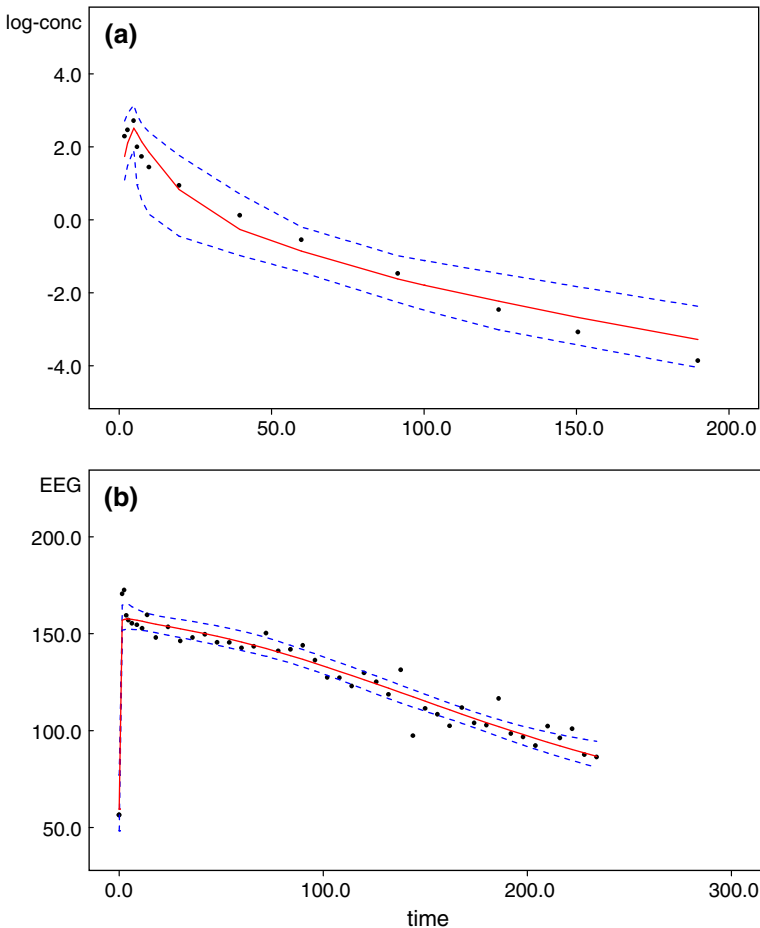
Fig. 2 Observed and predicted values (with 95% credible intervals) from simultaneous analysis of PK and PD data for rat 7: •, observed values; —, posterior median; - - -, 95% credible interval: **a** PK data, log-concentration versus time; **b** PD data, EEG versus time

full-likelihood and a full-Bayesian framework. We also emphasise that while the situation could possibly be improved by elaboration of the PD model, say, it is difficult, in general, to exercise any control over the extent to which this phenomenon occurs.

Part of the problem here is that there exists some inconsistency between the PK and PD models. However, it is typically impracticable to attempt to eradicate all such inconsistencies, especially in a population data set. Another aspect of the problem is that the two models are not being weighted according to their plausibility. In fact, in the above example the PK data are being swamped by the PD data, and the PK fit becomes unimportant as the PD fit is 'tweaked'. However, we typically have more confidence in the PK model and may feel it appropriate to weight the two data sets accordingly. In this paper we consider the approach of

discounting the likelihood contribution of the PD data to the estimation of the PK parameters, which has the flavour of a two-stage, or 'sequential', analysis.

Such sequential PKPD analyses have been considered by Zhang, Beal and Sheiner [6, 7] (henceforth referred to as ZBS), who contrast three different sequential strategies with a full-likelihood (simultaneous) PKPD analysis. For a population PKPD model with subject-specific PK parameters θ_i, drawn from a population distribution with parameters Θ, they consider:

1. Basing estimates $\hat{\Theta}$ on the PK data alone, and then estimating the θ_is using the PD data alone—termed the Population PK Parameters (PPP) approach by ZBS;
2. Basing estimates $\hat{\Theta}$ on the PK data alone, and then estimating the θ_i's using both the PK and PD data—termed the Population PK Parameters and Data (PPP&D) approach by ZBS;
3. Basing estimates $\hat{\theta}_i$ on the PK data alone—termed the Individual PK Parameters (IPP) approach by ZBS.

In a simulation study when the assumed models are correct, ZBS (paper I) showed that all three sequential methods were much faster to compute than the simultaneous procedure, and that the PPP&D sequential method in particular had similar performance to the simultaneous approach. However, if the assumed models are not correct, ZBS (paper II) show that sequential methods are far more robust, and in particular that the fitted PK model can be very sensitive to misspecification of the PD model. This behaviour is to be expected given the type of feedback from the PD model into the fitted PK model exemplified in Figs. 1 and 2.

A problem with the sequential approaches investigated by ZBS is that they condition on point estimates of PK parameters and hence do not propagate their associated uncertainty into the PD analysis. To do this a 'multiple imputation' [8] approach can be adopted, which for the third sequential strategy, say, would comprise three distinct stages:

1. From the PK analysis, estimate N plausible sets of possible values for the concentrations. This would generally be achieved through simulation;
2. Carry out N separate PD analyses, one for each of the estimated sets of concentrations;
3. Average over the N PD analyses with appropriate adjustment for estimates and intervals [8].

There has been a similar debate over simultaneous and sequential approaches to multiple imputation: [9] reviews approaches to missing covariates in regression of a response y on covariates X and points out that multiple imputation may condition only on X or on both X and y, while [10, 8, p. 217] discuss the potential advantages of adopting different models for imputation and analysis.

Little and Rubin [8] emphasise that multiple imputation is best considered as a Bayesian predictive procedure. Indeed, a Bayesian approach implemented within the MCMC framework [11–14] has already been proved to be of considerable value in population PKPD [15–21]. In the present context, it is natural to also consider a multiple imputation approach to *sequential* PKPD analysis. In this paper we demonstrate how such a multiple imputation approach can be implemented within a

full MCMC analysis via a simple adjustment to the sampling algorithm. This is achieved through the use of a "cut" function in the model description within WinBUGS [5], although the idea is easily transferable to any MCMC program. Section "Methods" describes the method and provides further insight into the modelling assumptions underlying each of ZBS's sequential methods, while sect. "Results" illustrates use of the method (and its impact) in a population PKPD setting. A concluding discussion is given in sect. "Discussion".

Methods

Graphical models

To help clarify the ideas discussed in this paper, it is convenient to start with a graphical representation of the structural assumptions relating the quantities in the PKPD model. Graphical models have become increasingly popular as 'building blocks' for constructing complex statistical models of biological and other phenomena [22]. These graphs consist of nodes representing the variables in the model, linked by directed or undirected 'edges' representing the dependence relationships between the variables. Here we focus on graphs where all the edges are directed and where there are no loops (i.e. it is not possible to follow a path of arrows and return to the starting node). Such graphs are known as Directed Acyclic Graphs (DAGs) and have been extensively used in modelling situations where the relationships between the variables are asymmetric, for example from cause to effect.

Figure 3a shows the generic situation we are interested in here. Unobserved variables are denoted by circular nodes while observed variables (i.e. the data) are denoted by square nodes. Arrows indicate directed dependencies between nodes. The model shown relates a response z to predictors x through parameters β, but where x is not directly observed. Instead we have observations y which depend on x through an assumed measurement error model. Note that, in order for the graph to represent a full joint probability distribution, we also assume that the unknown quantities at the top of the graph (i.e. β and x) are given appropriate prior probability distributions (usually chosen to be minimally informative); however, for clarity we suppress these dependencies in the graphical representation.

When considering the flow of information in a statistical model and the influence that one variable has on another, it is helpful to identify the conditional independence assumptions represented by the graph. In a directed acyclic graph, it is natural to draw analogies to family trees and refer to the 'parents' pa[v] of a node v as the set of nodes that have arrows pointing directly to v, 'children' ch[v] as nodes at the end of arrows emanating from v, and 'co-parents' as other parent nodes of a child. (The terms 'descendants' and 'ancestors', etc., then have obvious definitions.) A DAG expresses the assumption that any variable is conditionally independent of all its 'non-descendants', given its 'parents'. If we wish to define a joint distribution over all variables, V, say, in a given graph, such independence properties are equivalent [23] to assuming

$$p(V) = \prod_{v \in V} p(v|\text{pa}[v]), \tag{1}$$

that is, the joint distribution is the product of conditional distributions for each node given its parents.

The MCMC sampling-based algorithms used for Bayesian inference are able to exploit conditional independencies in the model in order to simplify computations. For example, the Gibbs sampler [11, 12] works by iteratively drawing samples for each unknown variable v from its full conditional distribution given (current) values for all the other variables in the model, which we denote $p(v|V\backslash v)$. Now $p(v|V\backslash v)$ ($\propto p(V)$) only depends on terms in $p(V)$ that contain v, which from Eq. 1 means that

$$p(v|V\backslash v) \propto p(v|\text{pa}[v]) \prod_{w \in \text{ch}[v]} p(w|\text{pa}[w]);$$

hence the full conditional distribution is proportional to the product of a 'prior' term $p(v|\text{pa}[v])$ and a likelihood term $p(w|\text{pa}[w])$ for each child w of v. In particular it is clear that the sampling of each variable depends only on the current values for its parents, children and co-parents in the graph. This simple result forms the foundation for the efficient algorithms implemented in the WinBUGS software [5], and in turn provides an easy rule for adapting MCMC algorithms to sample from the adjusted posterior distributions we describe below for sequential PKPD analysis.

Cutting the influence of children on their parents

Suppose we are interested in inference about β in the model represented in Fig. 3a. The correct posterior distribution $p(\beta|y,z)$ can be written
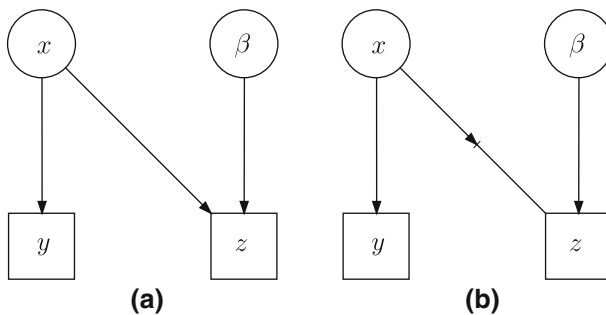
$$p(\beta|y,z) = \int p(\beta|x,z)p(x|y,z)\,dx$$

(a)   **(b)**

Fig. 3 Graphical models depicting the generic situation in which observed data z are related to predictors x through some parameters β, but where x is not directly observed—instead, we have observations y that depend on x through an assumed measurement error model: **a** full probability model; **b** model in which the influence of z on x has been cut. The 'diode'/'valve' symbol (▶|) between x and z in **b** represents the assumption that x is a parent of z but z is not a child of x, i.e. x influences z but not vice versa

since β is independent of y given z (its child) and x (z's co-parent). This means that the response data z are taken into account when making inferences about x, which subsequently influence the inferences about β. If the data z are more substantial than y, the estimation of x will tend to be dominated by the response model rather than the measurement error model. This may be particularly unattractive if we have greater confidence in the measurement error model, say through biological rationale for its functional form, while the response model is chosen more for convenience.

If we wanted to avoid the influence of z on the estimation of x, we would need to prevent or cut the feedback from z to x, allowing x to be estimated purely on the basis of y. This leads to the graphical model shown in Fig. 3b, which treats x as a parent of z but does not consider z to be a child of x. We denote this one-way flow of information by the 'valve' notation shown in the figure. When performing MCMC, all we have to do to prevent feedback is avoid including a likelihood term for z when sampling x. For example, if using Gibbs sampling, the conditional distribution used to generate values of x will not include any terms involving z.

Figure 4 shows the PD fit obtained when feedback is cut between the PD data (equivalent to z) and the PK parameters (equivalent to x) in the example discussed in sect. "Introduction". There is perhaps some underprediction towards the upper end of the time scale but the fit is by no means inadequate, nor is it necessarily inferior to that from the full model (Fig. 2b). Indeed, since the PK fit for this 'cut' model is as for the original PK-only model shown in Fig. 1, one could argue that the pair of fits in Figs. 1 and 4 is preferable to the pair in Fig. 2.

We emphasise that our conclusions based on this approach no longer arise from a full probability model, and as such disobey basic rules for both likelihood and Bayesian inference. However, as discussed further in sect. "Discussion", we may
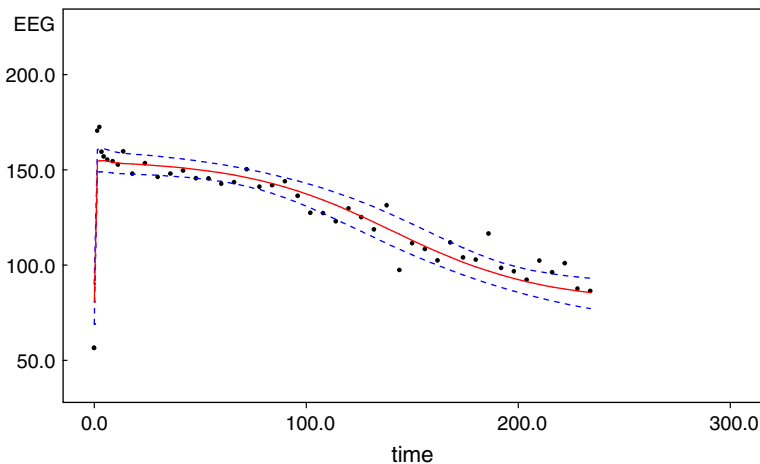

Fig. 4 Observed and predicted EEG values versus time (with 95% credible intervals) from simultaneous analysis of PK and PD data for rat 7, in the case where feedback from the observed EEGs to the estimation of the PK parameters is cut: •, observed values; —, posterior median; - - -, 95% credible interval
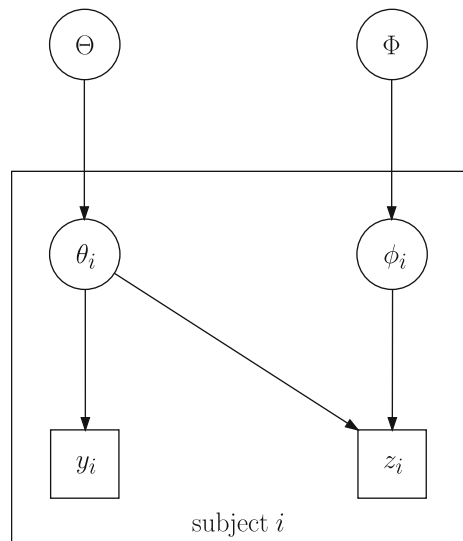
perhaps view this procedure as allowing a more 'robust' estimate of x that is not influenced by (possibly changing) assumptions about the form of the response model.

Population PKPD modelling

Following the notation in ZBS, we let y_i and z_i denote the vectors of observed concentrations and observed effects for individual $i = 1, \ldots, N$, with y and z denoting the collections of all PK and PD data across subjects. The (typically vector-valued) PK and PD parameters of the ith individual are denoted θ_i and ϕ_i, respectively, with θ and ϕ denoting the sets of all PK and PD parameters. Finally the PK and PD population parameters are denoted Θ and Φ respectively. Typically, the inter-individual distributions of the PK and PD parameters are assumed to be independently multivariate normal, so that $\Theta = (\mu_\theta, \Sigma_\theta)$ and $\Phi = (\mu_\phi, \Sigma_\phi)$ are the population means and covariances of the individual-level PK and PD parameters respectively, although the following discussion is general and applies to all distributional assumptions.

ZBS consider a variety of models for estimating the PKPD parameters in the above set-up. Their *simultaneous* approach (SIM) corresponds to specifying a full probability model for the PKPD data, and the graph corresponding to such a model is shown in Fig. 5. Unlike the previous graphs, Fig. 5 represents a *hierarchical* model. The large rectangle (known as a 'plate') denotes a repetitive structure—that is, the nodes enclosed within the plate are repeated for each subject $i = 1, \ldots, N$ in the study. Nodes outside the plate are common to all subjects, and represent the population PK and PD parameters in this context. The arrow linking θ_i to z_i in Fig. 5 represents the assumption that the PD responses, z_i, depend on the true PK responses (drug concentrations), where the latter are modelled as a deterministic function of

Fig. 5 DAG depicting full probability model underlying simultaneous population PKPD analysis

the PK parameters $f_{PK_i} = f(\theta_i)$. Note that dependence of the true and observed concentrations on quantities fixed by the study design, e.g. the dose and measurement times, is suppressed for notational clarity, as are error terms and other nuisance parameters.

From Eq. 1, the joint distribution of all the data and parameters of the model represented by Fig. 5 can be written

$$p(y, z, \theta, \phi, \Theta, \Phi) = \left[\prod_i p(z_i | \theta_i, \phi_i) p(y_i | \theta_i) p(\theta_i | \Theta) p(\phi_i | \Phi) \right] p(\Theta) p(\Phi).$$
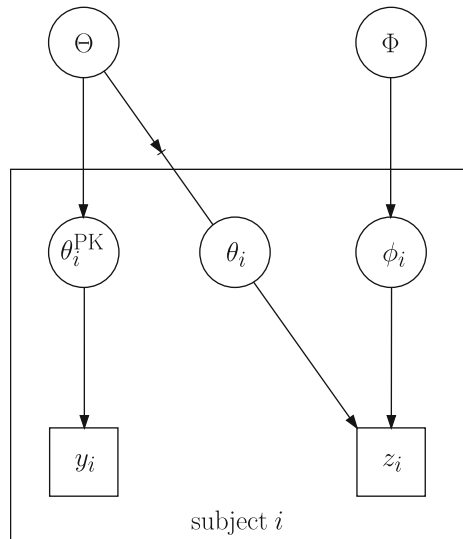
The key distinction between the inferential approaches we consider lies with the estimation of the θ_i's. Note that the posterior distribution for θ could be written

$$p(\theta | y, z) \propto p(z | \theta) p(\theta | y),$$

emphasising that the simultaneous posterior for θ conditional on both the PK and PD data is equivalent to having sequentially estimated the posterior for θ conditional only on the PK data, and then having used this posterior as the *prior* on θ for the second part of the analysis, conditioning on the PD data.

As noted previously, ZBS consider three alternative *sequential* approaches. The first of these, which they term PPP (Population PK Parameters) estimates the θ_i's on the basis of an estimate of Θ and the PD data alone: hence the θ_i's are influenced by the PK data only through estimation of Θ. Within our framework, this model may be represented by the graph shown in Fig. 6. The θ_i^{PK} nodes denote values of the PK parameters for subject i derived from the PK data alone, while the θ_i's are those parameters used to obtain the true concentrations $f_{PK_i} = f(\theta_i)$ for the PD model: the 'cut' ensures there is no influence of the PD data on the estimation of Θ. One way to interpret the cut is to imagine the PK data were analysed alone, giving rise to posterior distributions (given y) for both θ^{PK} and Θ. Then each subject is treated as a

Fig. 6 DAG depicting "PPP" sequential population PKPD analysis (see text for details). The 'cut' valve between Θ and θ allows θ to be influenced by Θ but prevents Θ (and hence θ^{PK}) being influenced by z

'new' individual and a posterior-predictive distribution is derived for his/her PK parameters. These predictive distributions then represent the priors for the subject-specific PK parameters in an analysis of the PD data alone. Our proposed framework allows all of this is to be performed simultaneously, however, as opposed to sequentially.

It is important to note that we do not know how to write down the resulting joint 'posterior' here! Our approach allows us to sample from it but we do not know its mathematical form (that is not to say that such a form doesn't exist). Hence we cannot compare it analytically with the correct posterior.

ZBS refer to the second of their sequential methods as PPP&D (Population PK Parameters and Data), which now estimates the individual PK parameters on the basis of both PK and PD data, but the PK population parameters using only the PK data. This method corresponds to the graphical model in Fig. 7. The node y^{copy} denotes a duplicate copy of the PK data, and emphasises the point made by ZBS that in the PPP&D method, the PK data are used twice (see below for further discussion of the implications of this). Interpretation of the cut is as before, with the posterior-predictive distribution (given y alone) forming the prior for each subject-specific PK parameter vector used in the PD analysis, except that now the PK parameters used in the PD analysis are influenced both by z and by the duplicated PK data. Re-use of the PK data might, at first glance, suggest spurious precision. However, this is not the case: by cutting the feedback between the θ_is and Θ, y^{copy} is used to inform about the subject-specific PK parameters (in the PD analysis) without providing further spurious information about the population PK parameters.

The final method considered by ZBS is called IPP (Individual PK Parameters), in which the subject-specific PK parameters do not depend on the PD data at all. The graphical model corresponding to this method is shown in Fig. 8. Unlike the

Fig. 7 DAG depicting "PPP&D" sequential population PKPD analysis (see text for details). The 'cut' valve between Θ and θ allows θ to be influenced by Θ but prevents Θ (and hence θ^{PK}) being influenced by y^{copy} and z
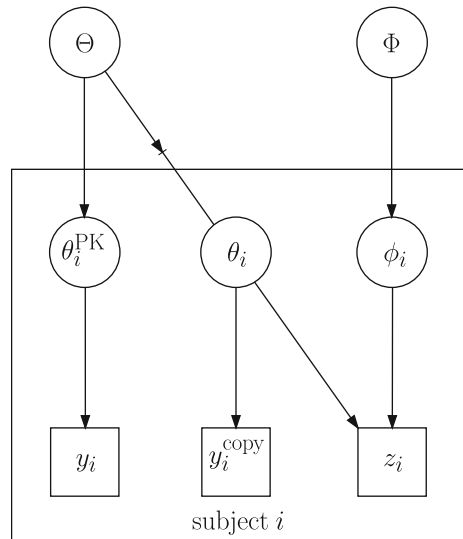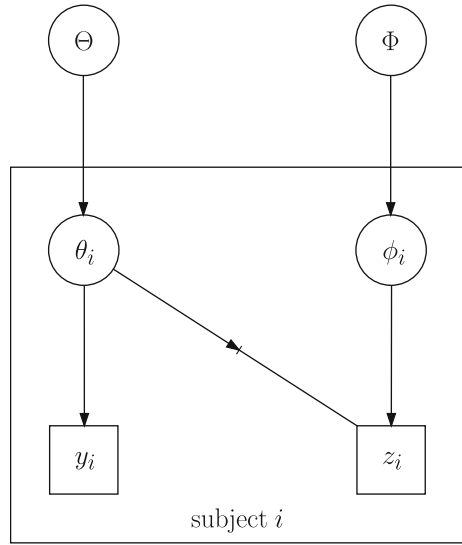
Fig. 8 DAG depicting "IPP" sequential population PKPD analysis (see text for details). The 'cut' valve between θ and z means that the subject-specific PK parameters θ_i do not depend on the PD data in any way

previous two sequential methods, the true concentrations needed for the PD model are derived from the original subject-specific PK parameters estimated from the PK model. In this case, the distribution of the individual θ_is used in the PD analysis cannot be thought of as a prior, since it is fixed and hence not updated in the PD analysis. Rather, we think of specifying θ as a 'distributional constant' in the PD analysis, which has the flavour of multiple imputation. Note that in the case of the PPP and PPP&D sequential models, it is Θ that is specified as a distributional constant. Table 1 summarizes the key differences between the four models in terms of how they condition on the 'data' in order to estimate the population and subject-specific parameters. As the nodes from which cut valves emanate can be interpreted as 'distributional constants' (i.e. fixed distributions that are not updated), we consider the corresponding distributions as 'data' here.

Table 1 Four different population PKPD models and the data on which their parameters depend

Parameter	Model			
	SIM	PPP	PPP&D	IPP
θ	z, y	$z, p(\Theta\|y)$	$z, y, p(\Theta\|y)$	y
θ^{PK}	–	y	y	–
Θ	z, y	y	y	y
ϕ	z, y	$z, p(\Theta\|y)$	$z, y, p(\Theta\|y)$	$z, p(\theta\|y)$
Φ	z, y	$z, p(\Theta\|y)$	$z, y, p(\Theta\|y)$	$z, p(\theta\|y)$

In cases where a cut valve is present, the node from which it emanates is interpreted as a 'distributional constant' and the associated distribution is considered 'data'
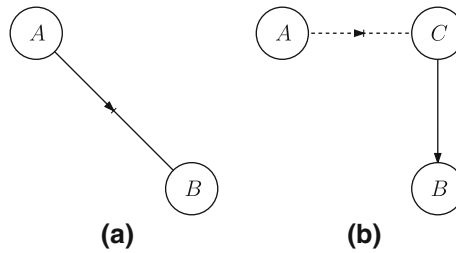
Fig. 9 Implementing cuts in the BUGS language: **a** graphical representation of the assumption that A is a parent of B but B is not a child of A; **b** to implement such an assumption we introduce a logical node $C = \text{cut}(A)$, which takes the same value as A but which severs the link between A and C as far as any children of C are concerned. A dashed edge represents a logical relationship

Implementing cuts in the BUGS language

Figure 9 depicts how cut valves are implemented in the BUGS language. Suppose we have two stochastic nodes A and B and we want to allow B to depend on A but prevent feedback from B to A (hence in observing B, our belief about A is unchanged). We introduce a *logical* (deterministic) node C, which is essentially equal to A but which severs the link between A and C as far as any children of C are concerned. Node B then becomes a child of C as opposed to A, and so no feedback can flow from B to A, while C provides the same information as A when it acts as a parent of B. In the BUGS language we might write, for example:

$$A \sim \texttt{dnorm}(\texttt{mu}, \texttt{tau})$$
$$C \gets \texttt{cut}(A)$$
$$B \sim \texttt{dnorm}(C, \texttt{gamma})$$

where $\texttt{cut}(.)$ is a logical function taking the same value as its argument (but which does not allow information to flow back towards its argument).

Results

We fit each of the four population PKPD models (SIM, PPP, PPP&D, IPP) to midazolam/EEG data [4] from 20 rats dosed intravenously, including the rat whose data are depicted in Figs. 1, 2 and 4 ("rat 7"). Each rat received ~ 2.5 mg midazolam over a 5-, 30- or 60-min infusion, and a maximum of 15 PK measurements were obtained from blood samples taken between 0 and 4 hours post-dose whereas a maximum of 49 PD measurements (EEG) were obtained over a similar time period. Annoated WinBUGS code for fitting the fully Bayesian model (SIM) is given in the appendix. The details of the model are somewhat superfluous here; what is important in the context of this paper are the modifications required for implementing each of the 'sequential' methods. These are outlined as follows.

- *IPP model:* We simply create a copy of the $\texttt{theta}[\,]$ variable using the $\texttt{cut}(.)$ function and use this copy, $\texttt{theta.cut}[\,]$, in the PD model (line 18) instead of

theta[ ]. The following BUGS code, inserted between lines 8 and 9, say, creates the new variable:

```
for (j in 1:4) {theta.cut[i, j] <- cut(theta[i, j])}
```

- *PPP model:* We first create copies of the population PK parameters, e.g.

```
for (i in 1:4) {
  mu.PK.cut[i] <- cut(mu.PK[i])
  for (j in 1:4) {Sigma.PK.Inv.cut[i, j] <- cut(Sigma.PK.Inv[i, j])}
}
```

We then change the name of the subject-specific PK parameters linked to the PK data (as in the graph of Fig. 6): we replace theta on lines 6 and 8 with theta.PK. Finally we specify a distributional assumption for the subject-specific PK parameters linked to the PD data (theta[ ]). We assume that they arise from a population distribution parameterised by the 'copied' population parameters, which is equivalent to assuming a prior equal to the posterior predictive distribution from the PK-data-only analysis. This is achieved by inserting the following, say, between lines 8 and 9:

```
theta[i, 1:4] ~ dmnorm(mu.PK.cut[ ], Sigma.PK.Inv.cut[, ])
```

- *PPP&D model:* We simply extend the PPP model by linking the subject-specific PK parameters used in the PD model (theta[ ]) to an exact copy of the PK data. For example, we could insert the following between lines 6 and 7:

```
y.copy[i, j] ~ dnorm(log.Cb.copy[i, j], tau.PK.copy)
log.Cb.copy[i, j] <- log(pkIVinf2(theta[i, ], PK.t[i, j], Dose[i], TI[i]))
```

where y[ ] is duplicated in the data set to form y.copy[ ].

For each model, two Markov chains with widely differing starting values were generated using WinBUGS. Seventy thousand iterations were performed and values from iterations 20001–70000 were retained for inference (giving a total sample size of 100000 for each parameter of interest). Note that this is a very conservative analysis, with run-lengths one tenth to one fifth as long typically being quite sufficient in practice. Run-times on a 1.2 GHz laptop machine were 172, 172, 184 and 95 min for the SIM, PPP, PPP&D and IPP models, respectively.

It is interesting to note that the apparent problem, highlighted earlier, with fitting a full probability model (SIM) to rat 7's data all but disappears when the population data are considered as a whole. This is presumably due to the additional information contained in the other rats' data regarding realistic values for the PK parameters. Note, however, that one cannot expect this to happen in general. Moreover, subtle differences between the model fits are still apparent on examination of the posterior deviance (minus twice the conditional log-likelihood). Table 2 shows the posterior mean deviance for each of the two sources of data when fitting each of the four models. This can be taken as a measure of model fit [24], with lower numbers indicating a better fit to the data. First note that we obtain the exactly the same fit

Table 2 Posterior mean deviances (as a measure of model fit) from each population PKPD model for each source of data

Data source	Model			
	SIM	PPP	PPP&D	IPP
PK (y)	-154.8	-162.9	-162.8	-162.9
PD (z)	6459	6392	6462	6491

(modulo Monte Carlo sampling variation) to the PK data with PPP, PPP&D and IPP. This is to be expected as none of these allows the PK model to be influenced by the PD data. Also note that each of the cut-models provides a better fit to the PK data than the fully Bayesian model (SIM), suggesting that feedback from the PD model actually damages the PK fit, at least in this example. In terms of fitting the PD data, the PPP model offers a substantial improvement over the fully Bayesian model whereas IPP performs considerably less well than SIM. PPP&D offers a similar goodness of fit to the SIM approach. For reasons that are discussed below we would expect our observations regarding the relative performance of the three cut-models to apply in general. In particular, in terms of goodness of fit of the PD model, PPP should always perform better than PPP&D, and PPP&D should always out-perform IPP. It is interesting to note that for these data, PPP offers both a better PK fit and a better PD fit than the SIM model.

Table 3 shows posterior summaries for each population parameter from each of the four models. SIM, PPP&D and IPP are all, generally, in good agreement. The PPP model, however, gives rise to substantially different estimates for $\log EC_{50}$ (both population mean and variance). In particular, the population mean corresponds to population median EC_{50} values around 30% higher than with the other models, and the population variance is inflated considerably. This is presumably due to the additional flexibility that PPP allows in terms of fitting the PD data (see below).

Discussion

We have constructed a general framework for preventing unwanted feedback in the simultaneous analysis of linked models. This has the flavour of a two-stage approach but unlike standard two-stage approaches our method allows uncertainty from the first stage to propagate fully through to the second stage. Thus the approach can be thought of as a form of multiple imputation. We have illustrated the use of our method for several PKPD models and shown how a graphical modelling perspective can elucidate the assumptions underlying established two-stage methods.

The four models considered can be thought of as representing varying degrees of confidence in the PK model relative to the PD model. With the full probability model (SIM), we are assuming equal confidence (in fact, total belief) in both models. If both PK and PD model specifications are optimal in some sense (we refrain from using the term 'correct' since all models are simplifications of reality)

Table 3 Posterior median point-estimates for population PKPD parameters from each of the four models; 95% credible intervals are given in parentheses

Parameter	Model			
	SIM	PPP	PPP&D	IPP
$\log CL$	−4.05	−4.05	−4.05	−4.05
	(−4.14, −3.95)	(−4.15, −3.96)	(−4.15, −3.96)	(−4.15, −3.96)
$\log Q$	−3.83	−3.77	−3.77	−3.77
	(−3.95, −3.72)	(−3.89, −3.65)	(−3.89, −3.65)	(−3.89, −3.65)
$\log V1$	−2.65	−2.70	−2.70	−2.70
	(−2.81, −2.49)	(−2.86, −2.54)	(−2.86, −2.54)	(−2.86, −2.54)
$\log V2$	−1.03	−1.03	−1.02	−1.03
	(−1.13, −0.923)	(−1.13, −0.919)	(−1.13, −0.918)	(−1.13, −0.919)
$\log E_0$	4.23	4.19	4.23	4.23
	(4.11, 4.34)	(4.09, 4.29)	(4.11, 4.34)	(4.11, 4.35)
$\log E_{max}$	4.35	4.41	4.35	4.35
	(4.28, 4.42)	(4.32, 4.49)	(4.28, 4.42)	(4.27, 4.42)
$\log EC_{50}$	−3.13	−2.82	−3.13	−3.15
	(−3.33, −2.93)	(−3.21, −2.46)	(−3.34, −2.93)	(−3.38, −2.91)
$Var[\log CL]$	0.039	0.040	0.040	0.040
	(0.021, 0.082)	(0.021, 0.084)	(0.022, 0.084)	(0.022, 0.084)
$Var[\log Q]$	0.014	0.014	0.014	0.014
	(0.004, 0.051)	(0.005, 0.049)	(0.005, 0.049)	(0.005, 0.047)
$Var[\log V1]$	0.056	0.045	0.045	0.045
	(0.013, 0.176)	(0.010, 0.154)	(0.010, 0.153)	(0.010, 0.151)
$Var[\log V2]$	0.037	0.040	0.040	0.040
	(0.016, 0.086)	(0.018, 0.092)	(0.018, 0.092)	(0.018, 0.091)
$Var[\log E_0]$	0.052	0.037	0.053	0.056
	(0.027, 0.112)	(0.018, 0.081)	(0.027, 0.113)	(0.028, 0.121)
$Var[\log E_{max}]$	0.018	0.028	0.018	0.019
	(0.008, 0.041)	(0.014, 0.062)	(0.008, 0.042)	(0.009, 0.044)
$Var[\log EC_{50}]$	0.114	0.411	0.117	0.173
	(0.020, 0.340)	(0.092, 1.113)	(0.021, 0.348)	(0.059, 0.447)
σ_{PK}	0.180	0.178	0.178	0.178
	(0.164, 0.200)	(0.162, 0.197)	(0.162, 0.196)	(0.162, 0.196)
σ_{PD}	9.11	8.77	9.13	9.25
	(8.68, 9.57)	(8.35, 9.22)	(8.70, 9.60)	(8.81, 9.73)

Q denotes distributional clearance; σ_{PK} and σ_{PD} denote the residual standard deviations for the PK and PD data respectively

then the PK parameter values supported by the PK model and data should be consistent with those supported by the PD model and data, and so allowing full feedback (i.e. borrowing strength) between the models using SIM is desirable. However, if the two parts of the PK-PD model specification lead to inconsistencies

between the parameter values supported by the different component models, the SIM approach can lead to fitting problems, especially if the PD data are more substantial than the PK. As discussed earlier, it is often realistic to have more confidence in the PK model specification than in the PD model. Hence, at the other extreme to SIM is the IPP model, which represents an uncompromising belief in the PK model, in the sense that the individual-specific parameters obtained from the PK model cannot be modified in any way to accommodate a better fitting PD model—they are input as 'distributional constants'. PPP and PPP&D lie somewhere in between these two extremes, with both preventing modification of the *population* PK parameters by the PD model, but allowing the *subject-specific* PK parameters to adapt to the PD model. This adaptation is stronger for PPP than for PPP&D since in the latter case it is tempered by the direct influence of the 'cloned' PK data.

In light of these observations we can see that the three 'sequential' methods can be ranked in terms of their expected ability, in general, to fit the PD data. (Recall that all three methods should fit the PK data equally well, since the PK model is identical in each case and no feedback from the PD model is permitted.) Fitting performance is (mostly) governed by the flexibility of the PK parameters used as input to the PD model, i.e. the extent to which they can be modified from the values that would be suggested by the PK data alone. The more flexible the input parameters, the better the PD fit achievable. As PPP, PPP&D and IPP represent increasing confidence in the PK parameters that would be obtained from the PK data alone, we would expect PPP, in general, to offer the best fitting performance and IPP the worst, with PPP&D lying somewhere in between (although they may, of course, all perform equally well). Hence if model fit is the main criterion for selecting between 'sequential' models, we might recommend PPP as the method of choice. However, perhaps it is preferable, in practice, to base such a decision on a careful consideration of one's relative confidence in the PK model instead. One might argue that PPP is particularly attractive because of the relatively weak assumption regarding the PK inputs to the PD model. However, this same assumption potentially increases the disparity between the PK parameters that best fit the PK data and those that are used as input to the PD model. We might question how meaningful a model that leads to two different, perhaps contradictory, sets of PK parameters may be—such an apparent internal inconsistency may, for some, be grounds for avoiding such an approach altogether. PPP&D also suffers from this problem, but to a lesser extent, since it allows less adaptation of the PK parameters to fit the PD model than does PPP; again, this may be viewed by some as a conceptual flaw in the approach, but others might think that allowing a degree of inconsistency is preferable to having to adopt an uncompromising belief in the PK model (by constraining the individual PK parameters to be distributional constants), as in IPP. Of the three sequential methods, we would expect the PD model fit for PPP&D to be most similar to that for SIM, since both allow the PK data to directly influence estimation of the PD model. However, unlike SIM, PPP&D yields a PK model fit that is not detrimentally influenced by feedback from the PD model in cases where relative confidence in the PD versus PK model specification is low.

We must emphasise, again, that models containing cuts do not correspond to an underlying full probability model (Bayesian or otherwise), in the same way that

sequential analyses neither correspond to some joint model. Hence, the 'joint distribution' from which we sample during our MCMC scheme is not a formal posterior; indeed it is possible that a joint distribution with the simulated properties does not even exist. This does not invalidate the approach, however; we simply think of cuts as representing the specification of 'distributional constants', an intuitive means of acknowledging a fixed degree of uncertainty regarding (otherwise fixed) input parameters, which is a natural objective in many contexts, for robustifying one's inferences. Even without the acknowledgement of uncertainty, cuts/sequential analyses can afford robustness by providing a mechanism whereby estimates relating to the measurement error model are not influenced by (possibly changing) assumptions about the response model. This is particularly important when there may be model misspecification. For example, we may have a well established (population) PK model, for which there is considerable biological rationale, and it is undesirable for our inferences regarding this model to change as we explore the PKPD relationship; the cut also ensures that the same inputs are used throughout the exploration process. ZBS have examined the performance of sequential methods under various model misspecification scenarios; while beyond the scope of the current paper, this is an area deserving of further investigation.

Appendix

Here we present annotated WinBUGS code for fitting the SIM model to the midazolam data. Much of the code is self-explanatory but some notes, pertaining to the line numbers shown in the right-hand margin, are provided for clarity. The

```
model {                                                               #1
  # PK model                                                          #2
  for (i in 1:K) {                                                    #3
    for (j in 1:PK.n) {                                               #4
      y[i, j] ~ dnorm(log.Cb[i, j], tau.PK)                           #5
      log.Cb[i, j] <- log(pkIVinf2(theta[i,], PK.t[i, j], Dose[i], TI[i]))  #6
    }                                                                 #7
    theta[i, 1:4] ~ dmnorm(mu.PK[], Sigma.PK.Inv[,])                  #8
  }                                                                   #9
  for (i in 1:4) {mu.PK[i] ~ dunif(-10, 10)}                         #10
  Sigma.PK.Inv[1:4, 1:4] ~ dwish(R.PK[,], 4)                         #11
  tau.PK ~ dgamma(0.001, 0.001)                                      #12
```

Appendix   continued

```
# PD model                                                              #13
for (i in 1:K) {                                                        #14
  for (j in 1:PD.n) {                                                   #15
    z[i, j] ~ dnorm(E.pd[i, j], tau.PD)                                 #16
    E.pd[i, j] <- E0[i] + Emax[i] * f.PK[i, j]/(EC50[i] + f.PK[i, j])   #17
    f.PK[i, j] <- pkIVinf2(theta[i,], PD.t[i, j], Dose[i], TI[i])       #18
  }                                                                     #19
  E0[i] <- exp(phi[i, 1])                                               #20
  Emax[i] <- exp(phi[i, 2])                                             #21
  EC50[i] <- exp(phi[i, 3])                                             #22
  phi[i, 1:3] ~ dmnorm(mu.PD[], Sigma.PD.Inv[,])                        #23
}                                                                       #24
for (i in 1:3) {mu.PD[i] ~ dunif(-10, 10)}                              #25
Sigma.PD.Inv[1:3, 1:3] ~ dwish(R.PD[,], 3)                             #26
tau.PD ~ dgamma(0.001, 0.001)                                           #27
Sigma.PK[1:4, 1:4] <- inverse(Sigma.PK.Inv[,])                          #28
Sigma.PD[1:3, 1:3] <- inverse(Sigma.PD.Inv[,])                          #29
sigma.PK <- 1 / sqrt(tau.PK)                                            #30
sigma.PD <- 1 / sqrt(tau.PD)                                            #31
}                                                                       #32
```

reader is also referred to [25] for a more detailed discussion on the general use of WinBUGS in PKPD contexts.

Line 3: K denotes the number of individuals.

Line 4: PK.n denotes the number of time-points for the PK data.

Line 5: log.Cb[i, j] denotes the natural logarithm of the 'true'/model-predicetd concentration for individual i at time-point j; tau.PK denotes the residual precision (1/variance) for the PK data (in WinBUGS, normal distributions are parameterised in terms of precisions rather than variances).

Line 6: With the *Pharmaco* interface (http://www.winbugs-development.org.uk/) installed in WinBUGS 1.4.x, pkIVinf2(.) is the syntax for a two-compartment, intravenous infusion model. This is a function of: (i) the appropriate four-dimensional parameter vector, theta[i,] in this case; (ii) the time at which the model is to be evaluated PK.t[i, j]; (iii) the dose D[i]; and (iv) the duration of infusion, in this case TI[i]. The model is parameterised in terms of $\log CL, \log Q, \log V1$ and $\log V2$ (in that order).

Line 8: Due to the log-parameterisation of the PK model, a multivariate normality assumption for the population distribution of the PK parameters is appropriate. The population mean and inter-individual inverse-covariance are denoted mu.PK[] and Sigma.PK.Inv[,], respectively.

Lines 10–12: Vague uniform, Wishart and gamma priors are specified for mu.PK[], Sigma.PK.Inv[,] and tau.PK, respectively. See [25] for more details regarding Wishart priors in WinBUGS.

Line 15: PD.n denotes the number of time-points for the PD data.

Line 16: E.pd[i,j] and tau.PD denote the model-predicted effect for individual i at time-point j, and the residual precision for the PD model, respectively.

Line 17: f.PK[i,j] denotes the model-predicted midazolam concentration for individual i at PD-time-point j (PD.t[i,j])—see line 18 for definition.

Lines 20–23: We assume that the logarithms of the PD parameters arise from a multivariate normal population distribution, with population mean and inter-individual inverse-covariance given by mu.PD[] and Sigma.PD.Inv[,], respectively.

Lines 25–27: As with the PK parameters, we assign vague uniform, Wishart and gamma priors to the population mean, inter-individual inverse-covariance and residual precision, respectively, of the PD model.

Lines 28–31: Here we define the inter-individual covariances and the residual standard deviations.

References

1. Beal SL, Sheiner LB (1992) NONMEM user's guide, parts I-VII. NONMEM Project Group, San Francisco
2. Wakefield JC, Aarons L, Racine-Poon A (1999) The Bayesian approach to population pharmaco-kinetic/pharmacodynamic modelling. In: Gatsonis C, Kass RE, Carlin B, Carriquiry A, Gelman A, Verdinelli I, West M, (eds) Case studies in Bayesian statistics. Springer-Verlag, New York, pp 205–265
3. Lunn DJ (2005) Bayesian analysis of population pharmacokinetic/pharmacodynamic models. In: Husmeier D, Dybowski R, Roberts S (eds) Probabilistic modeling in bioinformatics and medical informatics. Springer-Verlag, London, pp 351–370
4. Aarons L, Mandema JW, Danhof M (1991) A population analysis of the pharmacokinetics and pharmacodynamics of midazolam in the rat. J Pharmacokinet Biopharm 19:485–496
5. Spiegelhalter D, Thomas A, Best N, Lunn D (2003) WinBUGS user manual, Version 1.4. Medical Research Council Biostatistics Unit, Cambridge, 2003
6. Zhang L, Beal SL, Sheiner LB (2003) Simultaneous vs sequential analysis for population PK/PD data I: best-case performance. J Pharmacokinet Pharmacodyn 30:387–404
7. Zhang L, Beal SL, Sheiner LB (2003) Simultaneous vs. sequential analysis for population PK/PD data II: robustness of methods. J Pharmacokinet Pharmacodyn 30:405–416
8. Little RJA, Rubin DB (2002) Statistical analysis with missing data, 2nd edn. John Wiley & Sons, New York
9. Little RJA (1992) Regression with multiple x's: a review. J Am Statist Ass 87:1227–1237
10. Meng XL (1994) Multiple-imputation inferences with uncongenial sources of input. Stat Sci 9:538–558
11. Gelfand AE, Smith AFM (1990) Sampling-based approaches to calculating marginal densities. J Am Stat Assoc 85:398–409
12. Geman S, Geman D (1984) Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images. IEEE T Pattern Anal 6:721–741
13. Hastings WK (1970) Monte Carlo sampling-based methods using Markov chains and their applications. Biometrika 57:97–109
14. Metropolis N, Rosenbluth AW, Rosenbluth MN, Teller AH, Teller E (1953) Equations of state calculations by fast computing machines. J Chem Phys 21:1087–1091
15. Gueorguieva I, Aarons L, Rowland M (2006) Diazepam pharmacokinetics from preclinical to Phase I using a Bayesian population physiologically based pharmacokinetic model with informative prior distributions in WinBUGS. J Pharmacokinet Pharmacodyn 33:571–594
16. Mu S, Ludden TM (2003) Estimation of population pharmacokinetic parameters in the presence of non-compliance. J Pharmacokinet Pharmacodyn 30:53–81

17. Graham G, Gupta S, Aarons L (2002) Determination of an optimal dosage regimen using a Bayesian decision analysis of efficacy and adverse effect data. J Pharmacokinet Pharmacodyn 29:67–88
18. Lunn DJ, Aarons L (1998) The pharmacokinetics of saquinavir: a Markov chain Monte Carlo population analysis. J Pharmacokinet Biopharm 26:47–74
19. Lunn DJ, Aarons LJ (1997) Markov chain Monte Carlo techniques for studying interoccasion and intersubject variability: application to pharmacokinetic data. Appl Stat 46:73–91
20. Best NG, Tan KKC, Gilks WR, Spiegelhalter DJ (1995) Estimation of population pharmacokinetics using the Gibbs sampler. J Pharmacokinet Biopharm 23:407–435
21. Wakefield JC (1994) An expected loss approach to the design of dosage regimens via sampling-based methods. The Statistician 43:13–29
22. Spiegelhalter DJ (1998) Bayesian graphical modelling: a case-study in monitoring health outcomes. Appl Stat 47:115–133
23. Lauritzen SL, Dawid AP, Larsen BN, Leimer HG (1990) Independence properties of directed Markov fields. Networks 20:491–505
24. Dempster AP (1997) The direct use of likelihood for significance testing. Statist Comput 7:247–252
25. Lunn DJ, Best N, Thomas A, Wakefield J, Spiegelhalter D (2002) Bayesian analysis of population PK/PD models: general concepts and software. J Pharmacokinet Pharmacodyn 29:271–307