Rapid model exploration for complex hierarchical data: application to pharmacokinetics of insulin aspart

Robert J. B. Goudie,[a][*][†] Roman Hovorka,[b,c,d] Helen R. Murphy[b,c,d] and David Lunn[a]

We consider situations, which are common in medical statistics, where we have a number of sets of response data, from different individuals, say, potentially under different conditions. A parametric model is defined for each set of data, giving rise to a set of random effects. Our goal here is to efficiently explore a range of possible 'population' models for the random effects, to select the most appropriate model. The range of possible models is potentially vast, because the random effects may depend on observed covariates, and there may be multiple credible ways of partitioning their variability. Here, we consider pharmacokinetic (PK) data on insulin aspart, a fast acting insulin analogue used in the treatment of diabetes. PK models are typically nonlinear (in their parameters), often complex and sometimes only available as a set of differential equations, with no closed-form solution. Fitting such a model for just a single individual can be a challenging task. Fitting a joint model for all individuals can be even harder, even without the complication of an overarching model selection objective. We describe a two-stage approach that decouples the population model for the random effects from the PK model applied to the response data but nevertheless fits the full, joint, hierarchical model, accounting fully for uncertainty. This allows us to repeatedly reuse results from a single analysis of the response data to explore various population models for the random effects. This greatly expedites not only model exploration but also cross-validation for the purposes of model criticism. © 2015 The Authors. *Statistics in Medicine* published by John Wiley & Sons Ltd.

Keywords: Bayesian hierarchical models; variable selection; Markov chain Monte Carlo; pharmacokinetics; insulin

1. Introduction

Consider a hierarchical data set, where we have a number of sets of response data, from different patients perhaps. We wish to apply a parametric model to each individual's data set and then define a 'population' model relating all of the individual-level parameters (random effects) together. There may be a variety of credible models for the random effects, and it is important to fully explore a range of possibilities. For example, the parameters may depend on observed covariates, or there may be different ways of partitioning their variability across levels of the hierarchy. However, such exploration can be cumbersome and time-consuming, especially when individual-level models are complex. For example, pharmacokinetic models are typically nonlinear (in their parameters), often have complex functional forms and are sometimes only available as a set of differential equations, with no closed-form solution. Fitting such a

[a]*MRC Biostatistics Unit, Cambridge Institute of Public Health, Cambridge, U.K.*
[b]*Department of Paediatrics, University of Cambridge, Cambridge, U.K.*
[c]*University of Cambridge Wellcome Trust-MRC Institute of Metabolic Science, Level 4 Metabolic Research Laboratories, Cambridge, U.K.*
[d]*NIHR Cambridge Biomedical Research Centre, Cambridge, U.K.*
[*]*Correspondence to: Robert J. B. Goudie, MRC Biostatistics Unit, Cambridge Institute of Public Health, Cambridge, U.K.*
[†]*E-mail: robert.goudie@mrc-bsu.cam.ac.uk*

model for just a single individual can be a challenging task. Fitting a joint model for all individuals can be even harder, even without the complication of an overarching model selection objective.

We aim in this paper to develop a methodology for efficiently fitting a range of population models (including covariate selection) to the parameters (random effects) from individual-level models. We describe a *two-stage* approach that decouples inference on the population model from inference on each individual-level model but nevertheless fits the full, joint, hierarchical model, accounting fully for uncertainty. In the first stage, we estimate independent posterior distributions for the parameters in each individual-level model using Markov chain Monte Carlo (MCMC). We then store the resulting samples for use in the second stage, where they form 'proposal distributions' for the individual-level parameters in the full hierarchical model. The parameters are then updated via Metropolis–Hastings steps with an acceptance probability that is independent of the individual-level likelihood. This means that stage 2 is very efficient, allowing a wide range of models to be easily explored. In particular, our approach facilitates rapid exploration of covariate models using *reversible jump* MCMC, as well as exploration of different models for the variance components. It also facilitates criticism of the various population models via cross-validation techniques. Our approach is suited to situations in which the number of observations for each individual exceeds the number of parameters in the individual-level model and so is most likely to be useful in early clinical studies, in which detailed individual-level data are available.

We use this methodology to study four parameters relating to insulin kinetics in pregnant women with type 1 diabetes, using data from two clinical studies. We consider covariate selection models that enable identification of covariates that may be related to the kinetic parameters, as well as different structures for the hierarchical model, up to four levels. We believe that the complexity of our analysis would render it impracticable without the proposed two-stage methodology. Our two-stage approach is implemented in extensions to the OpenBUGS software [1, 2], which is freely available from www.openbugs.net.

Rapid-acting insulin analogues (such as insulin aspart and insulin lispro) can assist in safely optimising glucose control [3–5], but little is known about their pharmacokinetics and reproducibility in pregnancy. A significant gestational delay of approximately 30 min in time-to-peak plasma aspart concentration from early to late pregnancy has been previously described [6]. The aims of the study reported herein were to explore the relationship between aspart pharmacokinetics and clinical/demographic factors for subjects with type 1 diabetes undergoing continuous subcutaneous insulin infusion (CSII) during pregnancy and to assess reproducibility within and between subjects.

Basic clinical results from some of these analyses have been reported previously [7]. Here, we present and explore more thoroughly the statistical and methodological issues.

2. Data

Our data are from two 24-h trials of insulin aspart for pregnant women with type 1 diabetes conducted between March 2009 and April 2011 [8, 9]. Study protocols were approved by the Research Ethics Committee, and all participants provided written informed consent.

Table I. Summary of the 13 clinical and demographic factors.	
Factor	Summary
Maternal age (years)	32(4.5)
Body mass index (kg/m^2)	27(3.3)
Glycated haemoglobin (HbA1c) at booking (%)	7.1(1.0)
Duration of diabetes (years)	18(8.6)
Pregnancy gestation (weeks)	22(6.5)
Expected total daily dose (units/day)	55(18)
Peak bolus rate (units/hour)	9.3(5.0)
Recruited at Kings College [baseline Addenbrookes] (indicator)	5/22 women
Bolus delivered over longer than 1 min (indicator)	8/88 profiles
Multiple boluses (indicator)	17/88 profiles
Closed-loop basal insulin delivery (indicator)	24/88 profiles
Study 1 [baseline study 2] (indicator)	10/22 women
Breakfast [baseline dinner] (indicator)	44/88 profiles

Means (standard deviation) shown for continuous factors, and the number of profiles/women (out of total number) with that attribute are shown for dichotomous indicators.

During each study, women arrived at the clinical research facility (Addenbrooke's Hospital, Cambridge, UK) at midday and were monitored for 5 h after dinner on day 1 and breakfast on day 2. In study 1, 10 women were studied on two occasions: in early (12–16 weeks) and in later (28–32 weeks) gestation [8], with standardised dinner and breakfast, and under sedentary conditions. In study 2, 12 women were studied on two occasions in mid gestation (12–33 weeks) with standardised meals, snacks and exercise (3×20-min walks at 14:00, 19:30 and 09:00 h; and 2×50-min treadmill sessions at 15:00 and 09:30 h) [9].

A CSII delivering aspart maintained stable fasting and pre-meal glycaemic conditions during each visit using either closed-loop or conventional CSII. Under closed-loop CSII, the basal rate was adjusted every 15 min using continuous glucose measurements, whereas with conventional CSII, the women set temporary basal rates and used correction boluses according to capillary glucose measurements. The basal infusion rate was recorded from 14:00 h on day 1 in the first study and from 00:00 h on day 1, before the study started, in study 2. The median (interquartile range) infusion rate was 0.6(0.1–1.2) international units (U) per hour. All prandial insulin boluses were calculated according to capillary glucose levels and were initiated at 18:00 h on day 1 and 11:00 h on day 2. The median (interquartile range) prandial bolus dose was 8.9(6.5–12) U. We assume boluses infused steadily over a 1-min period, apart from eight boluses, which were delivered over longer periods. Insulin concentration readings were recorded from 16:30 h on day 1 in the first study and from 14:00 h on day 1 in the second study. Plasma insulin concentration was measured every 10 min for 90 min post-meal, every 15 min for 1.5–5 h post-meal and at 15- to 30-min intervals at other times, providing an average of 59 measurements per woman. Plasma insulin concentration was measured by an immunochemiluminometric assay (Invitron, Monmouth, UK; intra-assay coefficient of variation (CV) 4.7% and inter-assay CV 7.2–8.1%). Further details of study procedures are reported elsewhere [8, 9].

We consider each mealtime separately so that the basic unit of study is a time series of insulin concentrations (a 'profile') over a period around the evening (17:30–23:00 h on day 1) or morning (06:30–12:00 h on day 2) meal from a particular visit of a pregnant woman. A total of 22 women underwent two visits, each involving two meals; this gave rise to $22 \times 2 \times 2 = 88$ profiles to model.

Thirteen clinical and demographic factors were examined. These factors, which were considered time invariant at the profile-level, are summarised in Table I.

3. Methods

3.1. Profile-level model

3.1.1. Mechanistic model. We use a two-compartment model to represent insulin kinetics, as shown in Figure 1. At time t, $Q_1(t)$ and $Q_2(t)$ represent the insulin masses (in international units, U) in two subcutaneous tissue compartments, representing the delay in absorption into the blood. A controlled infusion of insulin enters compartment 1 at a rate $\mathrm{Inf}(t)$ (mU/min), supplemented by boluses of insulin, which we model as entering at a rate $\mathrm{Bol}(t)$ (mU/min). Insulin leaves both compartments according to a first-order process, with rate constant $(t^{\mathrm{max}})^{-1}$, where t^{max} is the time-to-peak insulin concentration in minutes. This gives rise to a pair of ordinary differential equations:

$$\frac{dQ_1(t)}{dt} = \mathrm{Inf}(t) + \mathrm{Bol}(t) - Q_1(t)/t^{\mathrm{max}}, \tag{1}$$

$$\frac{dQ_2(t)}{dt} = \frac{Q_1(t) - Q_2(t)}{t^{\mathrm{max}}}, \tag{2}$$

with $Q_1(0) = Q_2(0) = 0$, where $t = 0$ corresponds to midnight on day 1 before the trial starts. The run-in time between $t = 0$ and the time $t = t^{\mathrm{start}}$, when the trial starts, allows the states $Q_1(t^{\mathrm{start}})$ and $Q_2(t^{\mathrm{start}})$ to become independent of $Q_1(0)$ and $Q_2(0)$. Equations (1)–(2) do not, in general, have a closed-form solution. In this paper, we obtain a numerical solution, instead, via the Cash–Karp Runge–Kutta method [10].

We base our regression function on the observable quantity in these trials, which is plasma insulin concentration. This is assumed to equilibrate instantaneously with the efflux from compartment 2 and is given by $Q_2(t)/(t^{\mathrm{max}} \times wt \times \mathrm{MCR})$, where wt denotes the patient's body weight (kilogramme) and MCR is the *metabolic clearance rate* per unit of body weight: MCR is the volume of plasma (l) cleared of
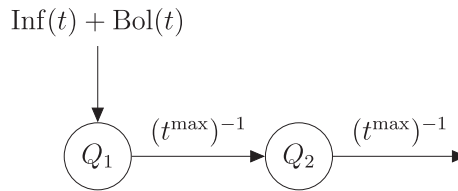
Figure 1. Two-compartment model for insulin kinetics, in which $Q_1(t)$ and $Q_2(t)$ represent the insulin masses in two subcutaneous tissue compartments, representing the delay in absorption into the blood. $\mathrm{Inf}(t)$ and $\mathrm{Bol}(t)$ represent insulin input via continuous infusion and supplemental boluses, respectively.

insulin per minute, per kilogramme. To account for any long-acting insulin taken before the start of the study (which continues to have an effect for around 24 h), we add a linear, 'residual insulin' term to our regression function. We assume that residual insulin concentration changes at a rate a (pmol/l/min) and that the post-prandial concentration at time t^{end} (5 h post-meal) is b (pmol/l). The regression function is then given by

$$\mu(\theta, z, t) = \frac{Q_2(t)}{t^{\mathrm{max}} \times wt \times \mathrm{MCR}} + a\left(t - t^{\mathrm{end}}\right) + b,$$

with unknown parameters $\theta = (t^{\mathrm{max}}, \mathrm{MCR}, a, b)'$ and observed data $z = (wt, t^{\mathrm{end}})'$.

3.1.2. Observation model. Denote by y_{ijkm} the mth measured plasma insulin concentration, taken at time t_{ijkm}, for individual i during and the following meal k of visit j ($i = 1, \ldots, N$, $j = 1, \ldots, J_i$, $k = 1, \ldots, K_{ij}$, $m = 1, \ldots, T_{ijk}$). We assume

$$y_{ijkm} \sim \mathrm{N}\left(\mu_{ijkm}, \sigma^2_{ijkm}\right), \quad \mu_{ijkm} = \mu\left(\theta_{ijk}, z_{ijk}, t_{ijkm}\right),$$

where θ_{ijk} and z_{ijk} are *profile-specific* vectors of unknown parameters and data, respectively. In addition, the residual variance is given by

$$\sigma^2_{ijkm} = \kappa^2_{ijk} + \lambda^2_{ijk} \times \mu^2_{ijkm},$$

which combines additive and multiplicative variance models, allowing for increases in measurement error above some baseline level as the modelled insulin concentration increases.

3.2. Population model

We wish to make inferences about the unknown parameters t^{max}, MCR, a and b. In particular, we are interested in establishing whether any relationships exist between the parameters and the observed covariates, and in quantifying their variabilities both within and between women. We begin by making separate distributional assumptions for each component (indexed by $l = 1, \ldots, 4$) of the parameter vector θ_{ijk}:

$$\theta_{ijkl} \sim \mathrm{LN}\left(\eta_{ijkl}, \sigma^2_{\theta_l}\right), \ l = 1, 2, \quad \theta_{ijkl} \sim \mathrm{N}\left(\eta_{ijkl}, \sigma^2_{\theta_l}\right), \ l = 3, 4,$$

where $\mathrm{LN}(., .)$ denotes a *log-normal* distribution with first and second parameters corresponding to mean and variance, respectively, on the log-scale. Note that we specifically avoid making a *multivariate* assumption for the whole θ_{ijk} vector, for reasons that we will discuss later (Section 5). By choosing different forms for η_{ijkl}, we can specify a variety of population models for the θ_{ijkl}s. In particular, we consider one-level, two-level and three-level population models as follows. In each case, we consider models both with and without covariates included.

3.2.1. One-level models. Here, we assume that the profile-specific parameters θ_{ijkl} are all conditionally independent, given a set of *global* parameters (or fixed effects). Hence

$$\eta_{ijkl} = \begin{cases} \phi_l + W_{ijkl}\beta_l & \text{with covariates} \\ \phi_l & \text{without} \end{cases}$$

where W_{ijkl} is a row-vector containing observed values (for individual i at meal k of visit j) of the covariates chosen as predictors for the lth kinetic parameter ($l = 1, \ldots, 4$); we discuss this further in Section 3.2.4. In this model, ϕ_l represents the global mean (or intercept if covariates are included in the model), and $\sigma^2_{\theta_l}$ represents the *total* variability (after any controlling for covariates) of parameter (or log-parameter) l.

3.2.2. Two-level models. Here, we acknowledge that profiles from the same woman may be correlated. It is tempting to achieve this by introducing patient-specific intercept and gradient parameters. However, with only four profiles from each woman, it is not very realistic to attempt to estimate patient-specific covariate effects, and so, we allow only the intercept to vary between women:

$$\eta_{ijkl} = \begin{cases} \psi_{il} + W_{ijkl}\beta_l & \text{with covariates} \\ \psi_{il} & \text{without} \end{cases}$$

with $\psi_{il} \sim \mathrm{N}(\phi_l, \sigma^2_{\psi_l})$, $i = 1, \ldots, N$. Here, $\sigma^2_{\psi_l}$ measures the between-patient variability (for parameter l), and $\sigma^2_{\theta_l}$ now represents the within-patient variability.
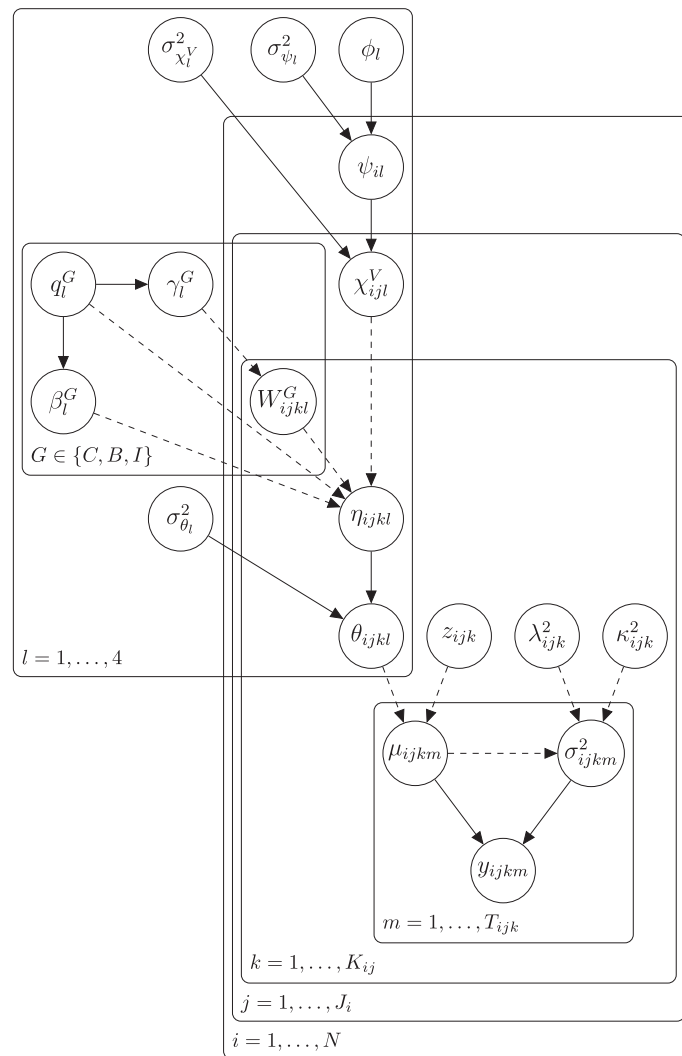

Figure 2. Directed acyclic graph representation of the three-level (visit) model with covariates. Each variable in the model is represented by a node, and links between the nodes denote direct dependence. Stochastic and deterministic (logical) dependence are represented by solid and dashed lines, respectively. Variables that are repeated are enclosed by 'plates' with the plate label denoting the range of repetition (e.g. $i = 1, \ldots, N$).

3.2.3. Three-level models.

Additionally, we might acknowledge that profiles from the same visit for a given woman may be correlated. Alternatively, profiles corresponding to the same mealtime for a given woman may be correlated. We might, therefore, allow visit-specific or mealtime-specific intercepts via

$$\eta_{ijkl} = \begin{cases} \chi^V_{ijl} + W_{ijkl}\beta_l & \text{with covariates} \\ \chi^V_{ijl} & \text{without} \end{cases} \quad \text{or} \quad \eta_{ijkl} = \begin{cases} \chi^M_{ikl} + W_{ijkl}\beta_l & \text{with covariates} \\ \chi^M_{ikl} & \text{without} \end{cases},$$

respectively. Here, $\chi^V_{ijl} \sim N\left(\psi_{il}, \sigma^2_{\chi^V_l}\right)$ or $\chi^M_{ikl} \sim N\left(\psi_{il}, \sigma^2_{\chi^M_l}\right)$, and again $\psi_{il} \sim N\left(\phi_l, \sigma^2_{\psi_l}\right)$, $i = 1, \dots, N$. In both cases, ψ_{il} represents the patient-specific mean intercept for parameter l, and $\sigma^2_{\psi_l}$ measures the between-patient variability. The terms $\sigma^2_{\chi^V_l}$ and $\sigma^2_{\chi^M_l}$, respectively, measure the variability between visits and between mealtimes (for a given individual) of the random intercepts for parameter l. In the case of visit-specific intercepts, $\sigma^2_{\theta_l}$ now measures the within-visit variability, whereas in the case of mealtime-specific intercepts, $\sigma^2_{\theta_l}$ measures the within-mealtime variability.

Figure 2 shows a graphical model representation of our statistical model in the case of three-level (visit) model with covariates.

3.2.4. Covariate selection.

Suppose we have c available covariates in total and we arrange their observed values in an $n \times c$ matrix X, where n is the total number of profiles. We expect the importance of each covariate may differ between parameters of the mechanistic model, and thus, η_{ijkl} for each $l = 1, \dots, 4$ may be a function of a different subset of the predictors. We represent the selected subset of covariates by a vector $\gamma_l = (\gamma_{l1}, \dots, \gamma_{lq_l})'$. This gives the column indices in X of the selected covariates. Let W_l denote the corresponding $n \times q_l$ design matrix and W_{ijkl} denote the row of W_l corresponding to meal k at visit j for individual i. Thus, the appropriate linear predictor term for inclusion in η_{ijkl} is $W_{ijkl}\beta_l$, where β_l is a vector of q_l regression coefficients. In this paper, we treat each q_l, γ_l and β_l as unknown parameters and estimate their values using *reversible jump* MCMC [11, 12]. Note that we standardise (and centre) the continuous covariates in X so that they are assessed in covariate selection on the same scale.

Interactions. Preliminary exploratory analyses suggested that parameters may differ between study and/or mealtime. However, when these covariates were included in X, neither was selected with high probability. We therefore decided to explore possible interactions. We allow for this by replacing study and mealtime in X with indicator variables for the following four combinations: study 1 breakfast, study 1 dinner, study 2 breakfast and study 2 dinner. We then choose a suitably weighted prior for the number of these indicators allowed in the model simultaneously, as discussed in Section 3.3.2.

3.3. Priors

3.3.1. Variance components.

The parameters of the residual variance model are treated as nuisance parameters and are assigned the following vague priors:

$$\kappa_{ijk} \sim U(0, 100), \quad \lambda_{ijk} \sim U(0, 1), \quad i = 1, \dots, N, \ j = 1, \dots, J_i, \ k = 1, \dots, K_{ij}. \tag{3}$$

The upper bound of 1 for the λ_{ijk}s is chosen as residuals larger in magnitude than the modelled concentration are implausible in this setting. The remaining variability parameters are all assigned vague uniform priors on the standard deviation scale:

$$\sigma_{\theta_l}, \sigma_{\psi_l}, \sigma_{\chi^V_l}, \sigma_{\chi^M_l} \sim U(0, 100), \quad l = 1, \dots, 4.$$

3.3.2. Fixed effects.

The global intercepts (or means if no covariates are included in the model), ϕ_l, $l = 1, \dots, 4$, are assigned vague normal priors with zero mean. The variance for the log-transformed parameters ($l = 1, 2$) is 100^2, whereas 1000^2 is chosen for $l = 3, 4$. The following prior distributions are assumed for the regression coefficients:

$$\beta_{lv} \sim N\left(0, \sigma^2_{\beta_{lv}}\right), \quad l = 1, \dots, 4, \ v = 1, \dots, q_l.$$

However, in reversible jump MCMC, the elements of each β_l will play different roles in the covariate model from one iteration to the next; that is, a given element may correspond to various different covariates during evolution of the simulated Markov chain. Hence, it seems appropriate to choose the same prior

variance for each element. In light of this, we can standardise the covariates so that the model treats them all equally. But we may still wish to allow for more or less diffuse priors for different types (or groups) of covariate. We thus decompose the linear predictor into separate terms for the continuous, binary and 'study–mealtime-interaction' covariates:

$$W_{ijkl}\beta_l = W_{ijkl}^C \beta_l^C + W_{ijkl}^B \beta_l^B + W_{ijkl}^I \beta_l^I,$$

where the *C*, *B* and *I* superscripts denote subvectors of W_{ijkl} and β_l, corresponding to continuous, binary and interaction covariates, respectively, and the same prior variance is specified for all elements of each subvector of β_l.

The prior standard deviation for coefficients corresponding to each covariate group is given by $\Delta\beta_l / 1.96\Delta x$, where $\Delta\beta_l$ is the width of the range of plausible values for the *l*th kinetic parameter (or log-parameter) and Δx is the width of the range of values for that covariate-type: $\Delta x = 1$ for binary covariates (including study–mealtime interactions) and $\Delta x \approx 2 \times 1.96$ for standardised continuous covariates. This is based on the assumption that the minimum and maximum plausible gradients define a 95% prior interval. Note that, lacking other prior information, we set $\Delta\beta_l$ to the range of the corresponding stage 1 posterior medians, and so, there is an element of using the data twice. While undesirable, this is necessary because the chosen prior variance influences the probability of inclusion of covariates in the model; hence, an informative prior is essential here.

The parameters associated with each covariate group $\left(q_l^G = \dim\left(\beta_l^G\right), \gamma_l^G, \beta_l^G, \text{ with } G \in \{C, B, I\}\right)$ are updated in the overall MCMC scheme as a separate reversible jump block. In each case, we wish to assume that all covariate models are equally likely a priori. We begin by assuming that all models of the same dimension are equally likely:

$$p\left(\gamma_l^G | q_l^G\right) = \binom{c^G}{q_l^G}^{-1}, \quad G = C, B, I; \; l = 1, \ldots, 4,$$

where c^C, c^B and c^I are the total numbers of available continuous, binary and interaction covariates, respectively. We then choose

$$q_l^C \sim \text{Bin}\left(c^C, 0.5\right), \quad q_l^B \sim \text{Bin}\left(c^B, 0.5\right), \quad p(q_l^I) = \begin{cases} \frac{1}{12} & q_l^I = 0 \\ \frac{4}{12} & q_l^I = 1 \\ \frac{6}{12} & q_l^I = 2 \\ \frac{1}{12} & q_l^I = 3 \\ 0 & q_l^I = 4 \end{cases},$$

where the latter prior excludes the possibility that $q_l^I = 4$ as this would lead to an unidentifiable model and acknowledges that all four possible models with $q_l^I = 3$ are essentially the same. Note that this specification renders equally probable a priori all distinct and identifiable models both within each covariate group *and* for the composite model defined by γ_l and q_l.

3.4. Inference

3.4.1. Hierarchical model. Let Ω denote the collection of all parameters in the population model for the θ_{ijkl}s. For example, for the simple one-level population model discussed in Section 3.2.1, $\Omega = \left(\phi, \beta, \sigma_\theta^2, \gamma, q\right)$, where, here and throughout, a quantity not indexed by *i, j, k, l* or *m* represents the collection of *all* quantities sharing the same variable name, for example, $\phi = (\phi_l)_{l=1,\ldots,4}$. The joint posterior distribution under the hierarchical model is then given by

$$p(\Omega, \theta, \kappa, \lambda \mid y) \propto p(y \mid \theta, \kappa, \lambda)p(\kappa)p(\lambda)p(\theta \mid \Omega)p(\Omega). \tag{4}$$

We make inference for the full hierarchical model (4) through a two-stage approach [13], which is outlined as follows.

3.4.2. Stage 1 analysis. In the first stage, we construct and estimate a posterior distribution for each profile independently, using the likelihood defined in Section 3.1.2, independent priors for the nuisance parameters given by (3) and independent, 'flat' priors for the kinetic parameters given by

$$t_{ijk}^{\max} \sim \mathrm{U}(5, 500), \quad \mathrm{MCR}_{ijk} \sim \mathrm{U}(0, 0.25), \quad a_{ijk} \sim \mathrm{N}\left(0, 100^2\right), \quad b_{ijk} \sim \mathrm{N}\left(0, 100^2\right), \tag{5}$$

where the lower and upper bounds for $t^{\max}$ and MCR represent the minimum and maximum physiologically plausible values, respectively. Using MCMC, we generate a sample of simulated values for the profile-specific parameters from each profile-specific posterior:

$$p_1\left(\theta_{ijk}, \kappa_{ijk}, \lambda_{ijk} \mid y_{ijk}\right) \propto p\left(y_{ijk} \mid \theta_{ijk}, \kappa_{ijk}, \lambda_{ijk}\right) p\left(\kappa_{ijk}\right) p\left(\lambda_{ijk}\right) p_1\left(\theta_{ijk}\right), \tag{6}$$

where y_{ijk} denotes the set of all measured concentrations for profile ijk and $p_1(\theta_{ijk})$ is the 'stage 1 prior' for θ_{ijk}, given by the product of terms in (5). We denote the samples by $\{\theta_{ijk}^{(h)}, \kappa_{ijk}^{(h)}, \lambda_{ijk}^{(h)}\}$, $h = 1, \dots, H_{ijk}$, and store them for use in stage 2, where they will be used to form proposal distributions for the profile-specific parameters in the full hierarchical model.

3.4.3. Stage 2 analysis. Stage 2 defines an MCMC scheme for updating all unknown parameters in the full hierarchical model. The parameters for each of the 12 covariate selection sub-models, $\beta_l^G, \gamma_l^G, q_l^G$, $G = C, B, I, l = 1, \dots, 4$, are jointly updated using reversible jump MCMC as described elsewhere [12]. The remaining parameters in Ω are updated by standard means. Each of the 'intercept' parameters, ϕ_l and, where appropriate, ψ_{il}, χ_{ijl}^V and χ_{ikl}^M, $i = 1, \dots, N, j = 1, \dots, J_i, k = 1, \dots, K_{ij}, l = 1, \dots, 4$, has a full conditional distribution that is available in closed form, and so a standard Gibbs step is appropriate. The variance components, $\sigma_{\theta_l}, \sigma_{\psi_l}, \sigma_{\chi_l^V}, \sigma_{\chi_l^M}, l = 1, \dots, 4$, can be updated by slice sampling [14], say, which is the default option in OpenBUGS for our model.

The profile-specific parameters $\theta_{ijk}, \kappa_{ijk}$ and λ_{ijk} are updated, jointly, as follows. From (4), their joint full conditional distribution is given by

$$p\left(\theta_{ijk}, \kappa_{ijk}, \lambda_{ijk} \mid \Omega, y\right) \propto p\left(y_{ijk} \mid \theta_{ijk}, \kappa_{ijk}, \lambda_{ijk}\right) p\left(\kappa_{ijk}\right) p\left(\lambda_{ijk}\right) p\left(\theta_{ijk} \mid \Omega\right). \tag{7}$$

We wish to make a Metropolis–Hastings update with this as the *target* distribution. A candidate update $(\theta_{ijk}, \kappa_{ijk}, \lambda_{ijk}) \rightarrow \left(\theta_{ijk}^\star, \kappa_{ijk}^\star, \lambda_{ijk}^\star\right)$ is drawn from the *proposal* distribution by choosing s uniformly from $\{1, \dots, H_{ijk}\}$ and setting $\left(\theta_{ijk}^\star, \kappa_{ijk}^\star, \lambda_{ijk}^\star\right) = \left(\theta_{ijk}^{(s)}, \kappa_{ijk}^{(s)}, \lambda_{ijk}^{(s)}\right)$, where the right-hand side is one of the samples stored in stage 1. From (6) and (7), the *target-to-proposal density ratio* is thus

$$R\left(\theta_{ijk}\right) \propto \frac{p\left(\theta_{ijk} \mid \Omega\right)}{p_1\left(\theta_{ijk}\right)}.$$

Note the cancellation of likelihood terms and nuisance priors. This makes for rapid computation in stage 2, facilitating exploration of a wide range of population models for the profile-specific parameters. That the ratio does not depend on κ_{ijk} or λ_{ijk} also means that these parameters need not actually be updated. The Metropolis–Hastings acceptance probability for the proposed update is $\min(1, \rho)$, where ρ is given by the target-to-proposal ratio at the proposed state divided by that at the current state:

$$\rho = \frac{p\left(\theta_{ijk}^\star \mid \Omega\right)}{p\left(\theta_{ijk} \mid \Omega\right)} \times \frac{p_1\left(\theta_{ijk}\right)}{p_1\left(\theta_{ijk}^\star\right)}.$$

If the stage 1 prior $p_1(.)$ is 'flat', as in our model, then the ratio on the right can be ignored as it is approximately equal to 1.

3.5. Model assessment via cross-validation

The two-stage method described earlier also expedites cross-validation to assess the various models considered. In leave-one-out cross-validation, a model is evaluated via predictions drawn from the model estimated with a single observation, or set of observations, excluded. Large disparities, measured by a discrepancy function, between these predictions and the excluded observations are indicative of model inadequacy. The procedure is repeated with each observation, or set of observations, omitted in turn. Such approaches have been widely discussed in the literature [15, 16].

Note that here we wish to assess the population model for the random effects θ_{ijk}, as opposed to the fit of the pharmacokinetic model to the observed response data. A discrepancy function defined in terms of the response data is not appropriate when the focus is on the random effects, because agreement or otherwise of random effect estimates is unnecessarily masked by the observation error. Hence we require a discrepancy function that measures differences between 'observed' and 'predicted' random effects, denoted $\theta_{ijk}^{\text{obs}}$ and $\theta_{ijk}^{\text{pred}}$, respectively. Clearly, the random effects cannot be observed, but the profile-specific posteriors defined by (6) can be used in lieu of observations [15]. The predicted random effects are obtained from the 'predictive prior':

$$p\left(\theta_{ijk}^{\text{pred}} \mid y_{\backslash ijk}\right) = \int p\left(\theta_{ijk}^{\text{pred}} \mid \Omega\right) p\left(\Omega \mid y_{\backslash ijk}\right) \mathrm{d}\Omega, \tag{8}$$

where $y_{\backslash ijk}$ denotes all observations except those from individual i during and the following meal k of visit j. This must be estimated with observations from each profile excluded in turn, which can be prohibitively time-consuming if the model is complex, as here. However, with the two-stage methodology presented here, only the second stage (which is computationally quick) needs to be repeated each time. Hence, this is a potentially important alternative to importance sampling [17], which can be unstable, or approximation [15, 18].

Let $D\left(\theta_{ijk}^{\text{obs}}, \theta_{ijk}^{\text{pred}}\right) = \theta_{ijk}^{\text{pred}} - \theta_{ijk}^{\text{obs}}$ be our discrepancy function. We define the Bayesian p-value $P = \Pr\left(D\left(\theta_{ijk}^{\text{obs}}, \theta_{ijk}^{\text{pred}}\right) \leq 0 \mid y\right)$ and estimate it, for each profile, by independent sampling from the predictive prior (8) and the stage 1 posterior (6), obtaining n p-values in total. Each p-value should be uniformly distributed under the 'true' sampling model [15], suggesting that model assessment guided by quantile–quantile plots of p-values is appropriate (note, however, that the p-values are not independent).

4. Results

4.1. Stage 1 analysis

To draw samples from the independent, profile-specific posterior distributions, we used the freely available BUGS software [2, 19]. Specifically, we used the OpenBUGS implementation (www.openbugs. net), which allows regression functions to be specified in terms of differential equations as standard.
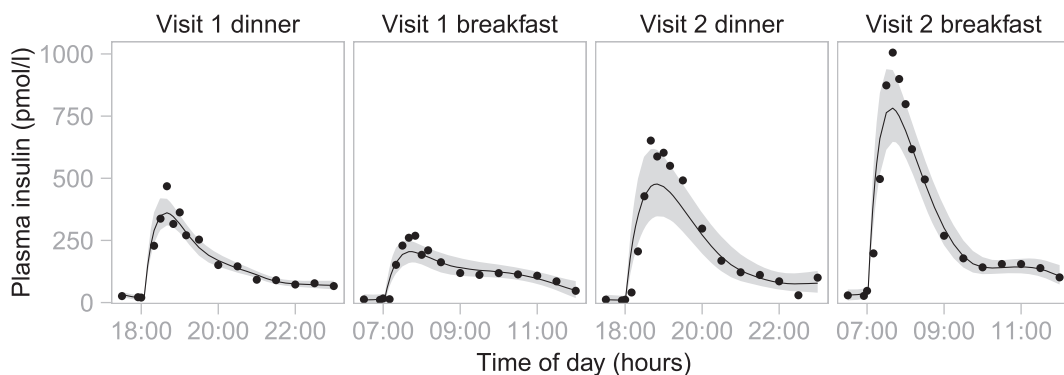

Figure 3. Four profiles from a typical individual: study 1, subject 1. The solid line is the (stage 1) posterior median model fit, and the shaded region is the 95% credible interval. The black circles (•) are the observed insulin concentrations. Note that some of the variabilities between these plots are due to differences in the input (bolus sizes were 8, 5, 20 and 20 international units, respectively).

Convergence of the generated Markov chains was assessed informally by visually examining chain-history plots and formally by applying the Brooks–Gelman–Rubin diagnostic [20, 21] to the output from two Markov chains starting from widely differing initial values. We found that a burn-in phase of 100 000 iterations was easily sufficient. We performed a further 1 000 000 iterations following burn-in for each profile, retaining every 100th value of each kinetic parameter (giving 10 000 approximately independent samples).
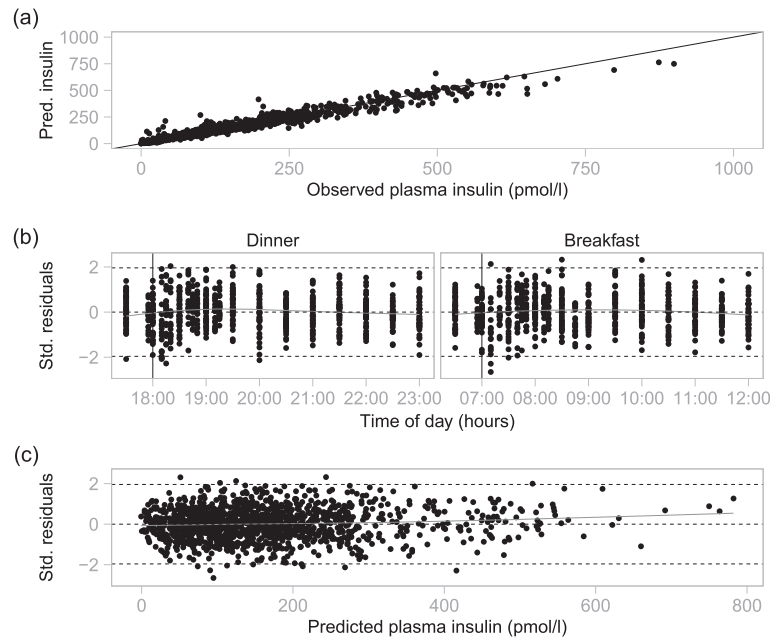
Figure 4. Model assessment following stage 1 analysis: (a) predicted (posterior median) versus observed insulin concentrations (•) with identity line $y = x$ (—); (b) posterior median standardised residuals versus time with mealtimes indicated by vertical lines (—); and (c) posterior median standardised residuals versus posterior median predicted concentration. The grey lines (—) are splines for the standardised residuals.

Figure 5. Marginal posterior inclusion probabilities for each of the 15 covariates considered under each of the four population models (including covariates) for each of the four kinetic parameters (t^{max}, MCR, a and b). The shading is proportional to the inclusion probability.

A total of 1302 plasma insulin concentrations were available for analysis. The model fits for a typical individual are shown in Figure 3. Our model was unable to fit six out of the 88 profiles available, and so these profiles were removed from our analysis. Figure 4(a) shows predicted (posterior median) versus observed insulin concentrations for all profiles fitted and indicates a good performance overall. To examine the model's performance in more detail, we also plot standardised residuals against time and against predicted concentration (Figure 4(b) and (c), respectively). There are no obvious trends, suggesting that the residual variance model is adequate and the residuals are generally within the expected range, although there is a hint of systematic bias about 10 min after each mealtime.
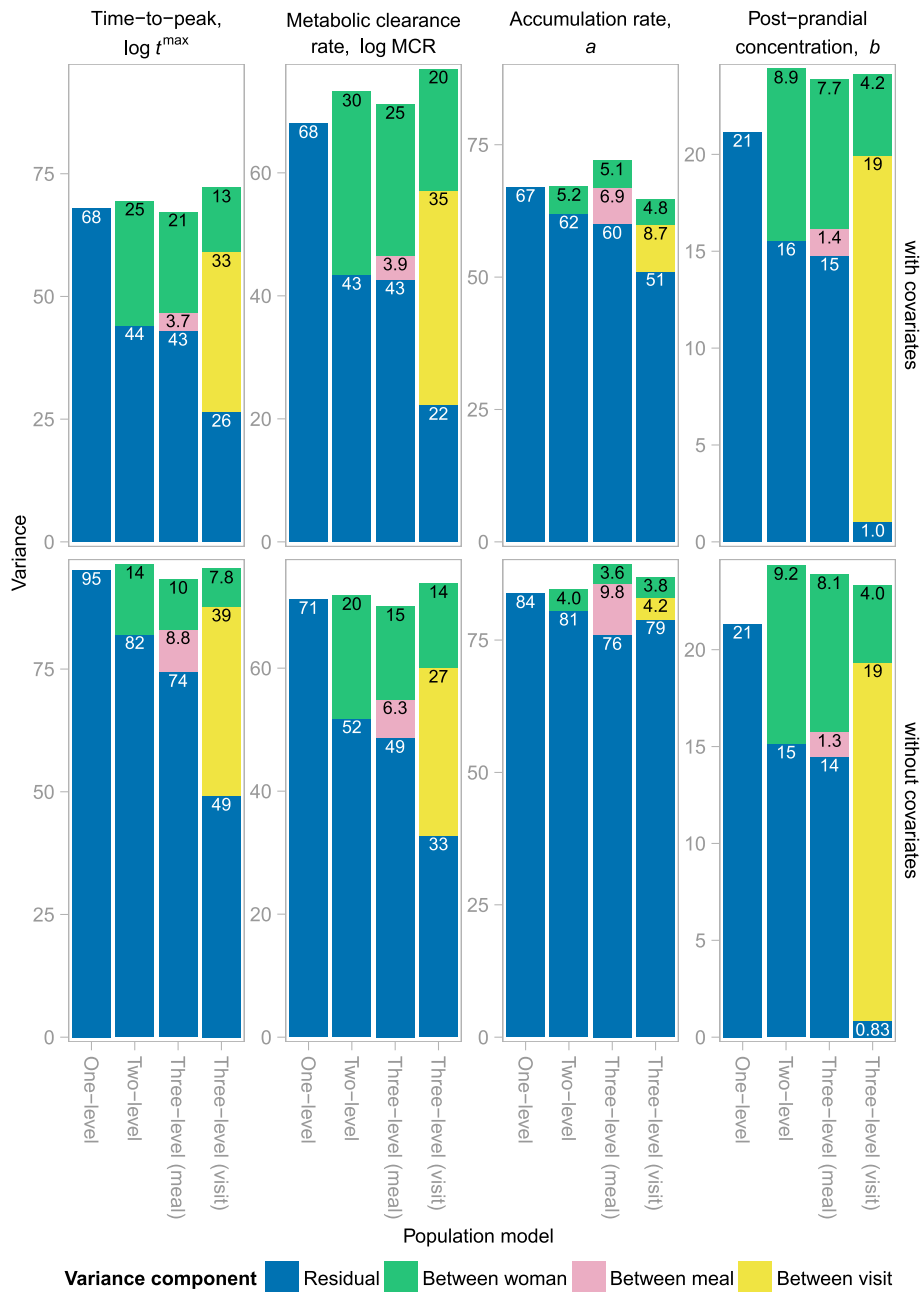

Figure 6. Decomposition of the variability in each pharmacokinetic (PK) parameter under each population model. Each panel represents the variance decomposition for a particular PK parameter for models either with (top row) or without (bottom row) covariates. Each coloured bar represents the proportion of total variance in the corresponding component for one of the four population models considered. Numbers shown within bars and on the *y*-axes are variances multiplied by the following factors, for $\log t^{max}$, $\log MCR$, a and b, respectively: $\times 10^3$, $\times 10^3$, $\times 10^4$ and $\times 10^{-2}$.

4.2. Stage 2 analyses

Figure 5 shows the posterior inclusion probabilities for each covariate being included in the linear predictor for each pharmacokinetic parameter, under each population model. We consider an inclusion probability greater than 0.5 to signify a notable association. The covariates identified as associated with each pharmacokinetic parameter were substantively consistent across the population models considered. All models strongly indicate that both t^{max} and a differ after breakfast in study 2 compared with the other meals and studies. We also found evidence for association of t^{max} with pregnancy gestation and diabetes duration and of a with peak bolus rate, expected total daily dose and delivery of multiple boluses. Evidence of association with MCR was less strong, although there was some evidence that MCR differs in study 2. No notable associations were observed for b.

Cross-validation analyses demonstrated that all of the population models we consider perform well: quantile–quantile plots (not shown) of the distribution of p-values under each model showed no obvious departures from uniformity. It was not possible to discern whether any specific model gave 'more uniform' plots than others. Differences amongst the models are more evident in the variance decomposition. These are shown in Figure 6. Two conclusions in common are apparent for t^{max} and MCR when comparing the various models. First, the models including covariates exhibit notably less residual variability, as might be expected given the evidence of parameter–covariate association described earlier. Second, the residual variability is reduced in the two-level and three-level models compared with the one-level model. The three-level, meal-specific model, however, offers little beyond the two-level model in terms of explaining variability of the random effects: the residual variability is almost identical. In contrast, the three-level, visit-specific model substantially reduces the residual variability beyond that in the two-level model. The two-level and three-level models are similarly an improvement for the post-prandial concentration b, but there is little difference between the models with and without covariates. Conversely, woman-specific effects are less apparent for the accumulation rate a, but the inclusion of covariates does make a notable difference to the residual variability.

Overall, the three-level, visit-specific model with covariates appears to offer the best explanation of the observed variation. Parameter estimates for this model are shown in Table II. The most notable effect is the faster absorption after breakfast in study 2 as implied by the substantially decreased time-to-peak t^{max} for these profiles. We also estimated that time-to-peak t^{max} increases by 1.7% per week of gestation in pregnancy but decreases by 1.1% per year of diabetes.

Table II. Insulin aspart pharmacokinetics in type 1 diabetes pregnancy.

Factor	Aspart PK parameter			
	Time-to-peak, t^{max} (min)	Metabolic clearance rate, MCR (l/kg/min)	Accumulation rate, a (pmol/l/min)	Post-prandial concentration, b (pmol/l)
Typical parameters for each study/mealtime combination				
Study 2 breakfast	**41(3.9)**	**0.028(0.0030)**	**0.056(0.044)**	33(12)
Study 2 dinner	57(4.7)	**0.022(0.0025)**	−0.047(0.033)	30(11)
Study 1 breakfast	55(3.8)	0.025(0.0021)	−0.029(0.027)	31(10)
Study 1 dinner	56(3.9)	0.025(0.0021)	−0.034(0.027)	31(10)
Effect sizes				
Peak bolus rate (U/h)	0.85(1.1)%	0.11(1.2)%	**0.012(0.0048)**	0.49(1.4)
Pregnancy gestation (week)	**1.7(0.73)%**	−0.33(0.87)%	−0.0039(0.0032)	0.47(1.4)
Multiple boluses	−11(14)%	17(15)%	**0.12(0.049)**	39(24)
Expected total dose (U/day)	0.37(0.36)%	−0.45(0.33)%	**−0.0027(0.0012)**	−0.19(0.62)
Duration of diabetes (year)	**−1.1(0.64)%**	0.66(0.76)%	0.0021(0.0024)	−0.12(1.2)

Top: posterior means (standard deviation (s.d.)) of each pharmacokinetic (PK) parameter for each study and mealtime combination for 'typical' covariate values, as defined by setting all continuous and binary covariates (except study–mealtime interactions) to zero. Bottom: posterior means (s.d.) of the effect sizes for covariates with a posterior inclusion probability > 0.5 in at least one model – estimates are conditional on inclusion of the corresponding covariate. The effect sizes shown for t^{max} and MCR are estimated *percentage* changes per unit change in the covariate; for a and b, we show the estimated *absolute* change. All notable effects are indicated in bold.

5. Discussion

We have presented a two-stage method for simplifying the analysis of complex hierarchical data. This can be thought of as a type of particle filtering (sequential Monte Carlo sampling; [22, 23]), where the resampling is carried out via Metropolis–Hastings. The method can be used whenever there is sufficient information in unit-specific data that units can be analysed independently (in stage 1). When unit-specific models are complex, as here, independent analyses may be a prerequisite for a full hierarchical analysis anyway, because the unit-specific data sets may require individual attention, for example, because of convergence issues or because we wish to assess whether a given unit-specific model can perform adequately over a wide range of units, say. However, our method is most likely to be useful when there is a range of 'population' models to consider for the unit-specific parameters. This is because the likelihood is dealt with in stage 1 and need not be computed in stage 2. Hence, the second stage can be performed repeatedly at little computational cost. This has allowed us to explore parameter–covariate relationships between four unknown parameters and 13 covariates of three different types, under a range of hierarchical structures (with up to four levels), for a model defined in terms of differential equations. The two-stage approach has also greatly facilitated cross-validation analyses to assess model performance. We chose to use cross-validation to explore models because we wished to focus our assessment of the models specifically on the choice of random effects model. Alternative approaches, such as the deviance information criterion (DIC) [24], would have undesirably focused on the pharmacokinetic model instead. It is our belief that the analyses presented herein would have been practically infeasible (or at least extremely cumbersome and time-consuming) without a two-stage approach.

We are mainly interested in the parameters t^{max} and MCR, because a and b relate to the 'residual' insulin model, which is somewhat speculative (although it may aid in compensating for model misspecification). Our analyses indicate that t^{max} is related to diabetes duration and gestational age in pregnant women using CSII. Time-to-peak increases by 1.7% per week of gestation and decreases by 1.1% per year of diabetes. In addition, there was a faster time-to-peak after breakfast in study 2, suggesting that moderately vigorous physical activity may counteract the gestational delay. The factors contributing to slower t^{max} in late pregnancy are unknown. The impact of diabetes duration may be related to loss of residual C-peptide activity [25]. While these clinical/demographic factors are not easily modified, the impact of exercise, most likely related to enhanced tissue perfusion and temperature, is potentially modifiable and may be a useful tool for speeding up insulin absorption as pregnancy advances. In contrast with Gagnon-Auger *et al.* [26], where substantially higher doses were used, our results do not indicate relationships between t^{max} and prandial dose, total daily dose or maternal body mass index.

There were no highly probable relationships identified for *MCR*, although in some models, there was a suggestion of decreased clearance following study 2 dinner and increased clearance following study 2 breakfast. This latter effect could be due to increased blood flow following physical exercise. There were also no relationships identified for b, but several were apparent for the drainage rate ($-a$). The drainage rate increased as the expected total daily dose increased but decreased with peak bolus rate and when multiple boluses were used.

A limitation of our analysis is that the data apply only to aspart delivered by CSII and are not necessarily applicable to lispro (another rapid-acting insulin analogue) or multiple daily injection therapy. However, as noted by Homko *et al.* [27], aspart and lispro have comparable pharmacokinetics, and CSII is increasingly recommended when glycaemic control targets are not achieved on multiple daily injections [28].

Although the two-stage method does not require independence between the four kinetic parameters, we chose in Section 3.2 to make separate distributional assumptions for each parameter. This is to avoid being too informative about the sizes (and relative sizes) of the variance components in our model, which characterise both between-patient and within-patient variabilities. The obvious alternative would be to assume multivariate normality for the vector $(\log t^{max}, \log \text{MCR}, a, b)'$, with a covariance matrix free to assume any symmetric–positive–semidefinite (SPSD) form, allowing for any correlations that may exist between the different kinetic parameters. The inverse-Wishart prior commonly chosen for such covariance matrices ensures the SPSD constraint automatically. However, it is more informative than might be expected and can have considerable influence on the posterior, particularly when, as here, there are multiple levels of variability and small sample sizes within levels (e.g. [29]). Other approaches proposed in the literature (e.g. [30, 31]) may be less informative, but we have chosen the relatively simple option of not allowing for correlations between kinetic parameters and assuming uniform priors for their variances.

We feel that this approach also reduces the possibility of confounding between covariance parameters and the inclusion of covariate effects.

Acknowledgements

References

1. Lunn D, Jackson C, Best N, Thomas A, Spiegelhalter D. *The BUGS Book: A Practical Introduction to Bayesian Analysis*. CRC Press: Boca Raton, 2013.
2. Lunn D, Spiegelhalter DJ, Thomas A, Best N. The BUGS project: evolution, critique and future directions. *Statistics in Medicine* 2009; **28**(25):3049–3067.
3. Lambert K, Holt RIG. The use of insulin analogues in pregnancy. *Diabetes, Obesity and Metabolism* 2013; **15**(10):888–900.
4. Heller S, Damm P, Mersebach H, Skjøth TV, Kaaja R, Hod M, Durán-García S, McCance D, Mathiesen ER. Hypoglycemia in type 1 diabetic pregnancy: role of preconception insulin aspart treatment in a randomized study. *Diabetes Care* 2010; **33**(3):473–477.
5. Mathiesen ER, Kinsley B, Amiel SA, Heller S, McCance D, Duran S, Bellaire S, Raben A, Insulin Aspart Pregnancy Study Group. Maternal glycemic control and hypoglycemia in type 1 diabetic pregnancy: a randomized trial of insulin aspart versus human insulin in 322 pregnant women. *Diabetes Care* 2007; **30**(4):771–776.
6. Murphy HR, Elleri D, Allen JM, Harris J, Simmons D, Rayman G, Temple RC, Umpleby AM, Dunger DB, Haidar A, Nodale M, Wilinska ME, Hovorka R. Pathophysiology of postprandial hyperglycaemia in women with type 1 diabetes during pregnancy. *Diabetologia* 2012; **55**(2):282–293.
7. Goudie RJB, Lunn D, Hovorka R, Murphy HR. Pharmacokinetics of insulin aspart in pregnant women with type 1 diabetes: every day is different. *Diabetes Care* 2014; **37**(6):e121–e122.
8. Murphy HR, Elleri D, Allen JM, Harris J, Simmons D, Rayman G, Temple R, Dunger DB, Haidar A, Nodale M, Wilinska ME, Hovorka R. Closed-loop insulin delivery during pregnancy complicated by type 1 diabetes. *Diabetes Care* 2011; **34**(2):406–411.
9. Murphy HR, Kumareswaran K, Elleri D, Allen JM, Caldwell K, Biagioni M, Simmons D, Dunger DB, Nodale M, Wilinska ME, Amiel SA, Hovorka R. Safety and efficacy of 24-h closed-loop insulin delivery in well-controlled pregnant women with type 1 diabetes: a randomized crossover case series. *Diabetes Care* 2011; **34**(12):2527–2529.
10. Cash JR, Karp AH. A variable order Runge–Kutta method for initial value problems with rapidly varying right-hand sides. *ACM Transactions on Mathematical Software* 1990; **16**(3):201–222.
11. Green P. Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika* 1995; **82**(4):711–732.
12. Lunn D, Best N, Whittaker JC. Generic reversible jump MCMC using graphical models. *Statistics and Computing* 2009; **19**(4):395–408.
13. Lunn D, Barrett J, Sweeting M, Thompson S. Fully Bayesian hierarchical modelling in two stages, with application to meta-analysis. *Journal of the Royal Statistical Society: Series C (Applied Statistics)* 2013; **62**(4):551–572.
14. Neal R. Slice sampling. *Annals of Statistics* 2003; **31**(3):705–741.
15. Marshall EC, Spiegelhalter DJ. Identifying outliers in Bayesian hierarchical models: a simulation-based approach. *Bayesian Analysis* 2007; **2**(2):409–444.
16. Gelfand AE, Dey DK, Chang H. Model determination using predictive distributions with implementation via sampling-based methods. In *Bayesian Statistics 4*, Bernardo JM, Berger JO, Dawid AP, Smith AFM (eds). Oxford University Press: New York, 1992; 147–167.
17. Stern HS, Cressie N. Posterior predictive model checks for disease mapping models. *Statistics in Medicine* 2000; **19**(17–18):2377–2397.
18. Marshall EC, Spiegelhalter DJ. Approximate cross-validatory predictive checks in disease mapping models. *Statistics in Medicine* 2003; **22**(10):1649–1660.

19. Lunn D, Thomas A, Best N, Spiegelhalter DJ. WinBUGS—a Bayesian modelling framework: concepts, structure and extensibility. *Statistics and Computing* 2000; **10**(4):325–337.

20. Gelman A, Rubin DB. Inference from iterative simulation using multiple sequences. *Statistical Science* 1992; **7**(4): 457–472.

21. Brooks SP, Gelman A. General methods for monitoring convergence of iterative simulations. *Journal of Computational and Graphical Statistics* 1998; **7**(4):434–455.

22. Doucet A, de Freitas N, Gordon N. Sequential Monte Carlo Methods in Practice. Springer: New York, 2001.

23. Andrieu C, Doucet A, Holenstein R. Particle Markov chain Monte Carlo methods. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 2010; **72**(3):269–342.

24. Spiegelhalter DJ, Best NG, Carlin BP, van der Linde A. Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 2002; **64**(4):583–639.

25. Wang L, Lovejoy NF, Faustman DL. Persistence of prolonged C-peptide production in type 1 diabetes as measured with an ultrasensitive C-peptide assay. *Diabetes Care* 2012; **35**(3):465–470.

26. Gagnon-Auger M, du Souich P, Baillargeon JP, Martin E, Brassard P, Ménard J, Ardilouze JL. Dose-dependent delay of the hypoglycemic effect of short-acting insulin analogs in obese subjects with type 2 diabetes: a pharmacokinetic and pharmacodynamic study. *Diabetes Care* 2010; **33**(12):2502–2507.

27. Homko C, Deluzio A, Jimenez C, Kolaczynski JW, Boden G. Comparison of insulin aspart and lispro: pharmacokinetic and metabolic effects. *Diabetes Care* 2003; **26**(7):2027–2031.

28. National Institute for Health and Care Excellence. *Diabetes in pregnancy. Management of diabetes and its complications in pregnancy from the pre-conception to the postnatal period.* NICE Guideline 63. National Institute for Health and Care Excellence, London, 2008.

29. Lunn D, Aarons LJ. Markov chain Monte Carlo techniques for studying interoccasion and intersubject variability: application to pharmacokinetic data. *Journal of the Royal Statistical Society: Series C (Applied Statistics)* 1997; **46**(1):73–91.

30. O'Malley AJ, Zaslavsky AM. Variance–covariance functions for domain means of ordinal survey items. *Survey Methodology* 2005; **31**(2):169–182.

31. Barnard J, McCulloch R, Meng XL. Modeling covariance matrices in terms of standard deviations and correlations, with application to shrinkage. *Statistica Sinica* 2000; **10**(4):1281–1311.