

Bayesian sequential integration within a preclinical PK-PD modeling framework: Potential challenges and opportunities

Fabiola La Gamba^{1,2} | Tom Jacobs¹ | Helena Geys^{1,2} | Thomas Jaki³ | Jan Serroyen¹ | Moreno Ursino⁴ | Alberto Russu¹ | Christel Faes²

¹Janssen Research & Development, Beerse, Belgium

²Interuniversity Institute for Biostatistics and Statistical Bioinformatics, Hasselt University, Diepenbeek, Belgium

³Department of Mathematics and Statistics, Lancaster University, Lancaster, UK

⁴INSERM, UMRS 1138, team 22, CRC, University Paris 5, University Paris 6, Paris, France

Correspondence

Fabiola La Gamba, Janssen Research & Development, Turnhoutseweg 30, 2340 Beerse, Belgium
Tel.: +32-485-999223
Email: flagamba@its.jnj.com

Funding information

This project has received funding from the European Unions Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 633567. Author Christel Faes also acknowledges support from the IAP Research Network P7/06 of the Belgian State (Belgian Science Policy). Author Moreno Ursino was funded by French National Cancer Institut (INCa) grant INCA 9539.

The present manuscript aims to discuss the implications of sequential knowledge integration of [small preclinical trials](#) in a Bayesian PK-PD framework. Whereas, at first sight, a Bayesian PK-PD framework seems to be a natural framework to allow for sequential knowledge integration, [the scope of this paper is](#) to highlight some often-overlooked challenges while at the same time providing some guidances in the many and overwhelming choices that need to be made. Challenges as well as opportunities will be discussed that are related to the impact of (i) the prior specification, (ii) the choice of random effects, (iii) the type of sequential integration method. In addition, it will be shown how the success of a sequential integration strategy is highly dependent on a carefully chosen experimental design [when small trials are analyzed](#).

KEYWORDS

Bayesian inference, Nonlinear Hierarchical Models, Pharmacodynamics, Pharmacokinetics, Recursive, Sequential

1 | INTRODUCTION

Research on pharmacokinetic and pharmacodynamic (PK-PD) models has evolved a lot over the last 50 years, allowing to better quantify the pharmacodynamic behavior of compounds, thereby facilitating drug development. We refer to Felmler et al., Mould et al. and Upton et al. [1, 2, 3] for a brief overview. Although the PK-PD literature has been originally developed in a frequentist framework using the maximum likelihood (ML) approach, only recently it has been recognized that Bayesian methods are a powerful tool for drug development [4, 5, 6, 7, 8, 9, 10]. Bayesian methods have several applications with respect to incorporating the knowledge from previous studies into the current study [11], such as bridging animal studies and first-in-man studies [12, 13], or using information from clinical studies on adults to analyze studies on neonates [14].

Previous work [15] illustrates the successful development and application of a Bayesian K-PD model for synergy on in-vivo data where the results from a historical dose-response trial were incorporated through a prior distribution. The data of all combination trials were collected at the same laboratory and the trials were conducted by the same scientist. The authors therefore argued that all trials could be pooled and that data could be analyzed together. In doing so, however, the sequential nature of the data collection, where each trial is conducted at a different point of time and the results from a trial are used to design the next trial, is completely ignored. The question arises to what extent a pooled analysis would differ from a (Bayesian) sequential modeling approach. In this paper such Bayesian sequential approach is first implemented. Whereas the integration of incoming information recursively over time is well established in the field of computer science under the name of recursive Bayesian estimation [16, 17], the technique has -to our knowledge- hardly ever been applied in the framework of a complex PK-PD modeling approach [18]. Micallef et al. [19] proposed a method to sequentially update the parameters of a PK model for caffeine in premature neonates using a stochastic particle algorithm. In our work, the Bayesian integration is performed by setting the hyperparameters of the prior distributions of a trial based on the posterior distributions resulting from the previous trial. Challenges and opportunities will be outlined in that respect as it is extremely important to understand the behavior and properties of this methodology on the intermediate results at different integration steps, particularly when small trials are analyzed with complex nonlinear models. In addition, it is shown that (under certain conditions) both approaches (pooling all data versus sequential integration) are expected to give similar results. As an important consequence, intermediate results allow for an adaptive design approach in the sense that they can guide the scientists towards a sensible dosing choice for the next trial.

The manuscript has the following structure: Section 2 describes the motivational case study and briefly revisits the model proposed by La Gamba et al. [15] to analyze these data; Section 3 describes different modeling aspects and compares various types of Bayesian sequential pooling. The results from these comparisons are shown in Section 4. A final discussion is given in Section 5.

2 | MOTIVATING CASE STUDY

This section provides a brief description of the case study (Section 2.1) and the proposed K-PD model for synergy by La Gamba et al. [15] to analyze these data (Section 2.2).

2.1 | Case study

This study was part of the pre-clinical safety evaluation of a novel compound meant to be co-administered with a compound already available on the market [20]. For the marketed compound, a historical trial was available to assess its safety [21]. An extensive dose range, from 0.04 to 10 mg/kg, was investigated. Placebo administration was included as well. A total of 55 rats were randomly allocated to receive a single dose of the marketed compound. For each of those rats body temperature, the biomarker of interest, was assessed up to 24 hours after single oral administration. No plasma concentration-time profiles were measured in this trial due to the possible impact of blood sampling on body temperature. [Dose related decrease in body temperature was observed for extremely high doses, indicating potential side effects beyond the clinical therapeutic window.](#)

With the intent to develop the novel compound for co-administration, 11 new [combination](#) trials were conducted [in different time periods](#) to evaluate whether a PD interaction between compounds occurs [at different combinations of dose levels](#) (the absence of a PK interaction had already been confirmed). [A new set of 20 rats \(from the same vendor and having the same strain\) was used in each trial](#), with 5 animals randomized to each of four treatment groups. Group 1 received a single dose of the vehicle, group 2 and group 3 received a single dose of the marketed and novel compound respectively. Finally, group 4 received a combination of both compounds. Body temperature was assessed every hour after administration for 4 hours. [As mentioned above, a different combination of dose levels was assessed in each trial, to understand this synergistic behavior. Therefore, the 11 combination trials only differ by dose combination assessed and conduction time period.](#) Table 1 shows the compound doses in each trial, whereas Figure 1 illustrates the time profiles for the rats belonging to trial 1 as an example. Animals in the vehicle and novel treatment groups did not show any effect on body temperature, whereas rats attributed to the marketed compound showed an effect in line with the historical trial. It was, however, unexpected that the combination group would have shown a change in body temperature which was more pronounced than the marketed treatment group. Also, the decrease in body temperature was more pronounced in trials assessing higher doses. Therefore, the presence of a PD synergy was hypothesized. We refer to [15] for a more detailed description of the case study.

2.2 | K-PD model for synergy

Let us now describe the model proposed by La Gamba et al. [15]. This is an example of a Type I indirect PD response model [22] with an inhibition process operating on the apparent production rate. So, if R_{it} denotes the observed body temperature for animal i at time t , then:

$$R_{it} \sim N(\bar{R}_{it}, \sigma_R^2), \quad (1)$$

$$\frac{d\bar{R}_{it}}{dt} = k_{in} \left(1 - \frac{I_{max} C_{it}}{IC_{50} + C_{it}} \right) - k_{out} \bar{R}_{it}. \quad (2)$$

where \bar{R}_{it} is the expected body temperature for animal i at time t . The parameters k_{in} and k_{out} are the rate for apparent production and elimination of body heat, respectively. C_{it} represents the (hypothetical) plasma concentration of the marketed compound and I_{max} is the maximal inhibition of the production rate. Finally, IC_{50} represents the concentration of the marketed treatment resulting in 50% of the maximal inhibition. The model takes into account residual error via equation (1), with σ_R^2 the measurement variance.

As the plasma concentration was not measured, a virtual one-compartment first order absorption PK profile [23]

was assumed to drive the inhibition of body heat (hence the use of the term "K-PD" in contrast to "PK-PD", to emphasize the absence of PK data):

$$\frac{dA_{e,it}}{dt} = -k_a A_{e,it}, \quad (3)$$

$$\frac{dC_{it}}{dt} = k_a A_{e,it} - k_e C_{it}, \quad (4)$$

where $A_{e,it}$ represents the hypothetical amount of the marketed compound in the absorption depot; the parameter k_a is the first-order apparent absorption rate constant, whereas the parameter k_e is the apparent elimination rate constant.

At time $t = 0$, $A_{e,i0}$ corresponds to the dose of the marketed compound, $C_{i0} = 0$, and $R_{i0} = k_{in}/k_{out}$ corresponds to the fact that body temperature is in a steady state condition prior to administration of the compound. Since the absorption constant is confounded with k_{out} , hence not identifiable, k_a was set to a very high value, i.e. to $\exp(5)$, which is consistent to the approach in Jacqmin et al. [23].

It is assumed that the co-administration of the novel compound increases the potency of the marketed compound for body temperature:

$$IC_{50} = \exp(\alpha A_{n,i0} + \beta A_{e,i0} A_{n,i0}), \quad (5)$$

where $A_{n,i0}$ is the dose of the novel compound for the i^{th} individual, α represents the main effect attributed to the novel compound and β is the interaction coefficient, representing the extent at which the effect of the novel compound on the potency of the marketed compound depends on the dose of the marketed compound itself.

To allow for heterogeneity among animals, a random effect was assumed for body temperature at baseline

$$R_{0i} \sim N(\bar{R}_0, \sigma_{R_0}^2).$$

3 | METHODS

In this section, a sequential integration of the information included in the different trials is discussed. The general idea is to use the posterior distributions resulting from one trial in order to determine the hyperparameters of the prior distributions of the next trial.

Compared to the traditional Bayesian pooling previously described by La Gamba et al. [15], the sequential methodology requires the application of the K-PD model on data from trial 1 as a first step, with the prior distributions based on the results from the historical trial. Performing a complex nonlinear hierarchical model such as the one described in Section 2.2 on little data may not be simple, and different aspects should be carefully taken into consideration during model building: Section 3.1 aims at describing how the model performance changes according to eliciting prior distributions with different precision, whereas the impact of assigning random effects on different parameters are described in Section 3.2. Finally, in Section 3.3 different sequential integration types are illustrated and compared.

The Bayesian model was estimated using the No-U-Turn Sampler (NUTS) algorithm in Stan (rstan version 2.14.1) [24] and the graphic representation was made using the statistical software package R (Foundation for Statistical Computing, Vienna, Austria) version 3.3.3. [The Stan code for the model is reported in Appendix A.](#)

3.1 | Prior specification

One of the aspects that differentiates Bayesian from frequentist inference is that the parameters are treated as random. Prior distributions represent the existing knowledge regarding the parameters before the data are collected, as well as the uncertainty of the knowledge itself; posterior distributions incorporate prior knowledge and the information coming from the data collected in the current experiment [25].

In case of low sample size, the use of informative prior distributions has proven to solve several performance issues, such as low power and biased parameter values [26]. In some areas of drug development, prior elicitation constitutes a crucial phase, since the specification of highly informative prior distributions compensates the small data, thus avoiding convergence issues in parameter estimation [27]. For example, in pediatric studies Bayesian methods are typically used to explicitly borrow information from other adult trials or previous pediatric studies [14, 28, 29]. See also Viele et al. [30], Ibrahim et al. [31], Takeda et al. [32] and Li et al. [33] for a more general overview of benefits resulting from borrowing historical information in the analysis of the current trial. On the other hand, the use of highly informative priors in case of low sample size may lead to a domination of the prior, rather than the data, on the posterior inferences. Therefore, prior elicitation should be cautiously performed in order to avoid the inappropriate introduction of artificial information which may bias the final results [34, 35, 36, 37].

In the manuscript of La Gamba et al. [15], the results from a frequentist analysis of the historical dose-response data on the marketed compound were incorporated in the prior distributions: expected values were set equal to the point estimates obtained from the analysis; the standard deviations were set equal to the double of standard errors to down-weight the information derived from the historical trial, which was conducted two decades before the combination trials. In particular, the priors for k_{θ} and k_{out} were assumed log-normally distributed, I_{max} was assumed to follow a beta distribution (as $0 \leq I_{max} \leq 1$), and a uniform distribution with parameters -10 and 10 was used as a prior for α and β as the historical data contain information on the effect of the marketed compound only, so there is no information regarding the PD synergy prior to the combination trials. An inverse gamma distribution was used as a prior for the measurement error variance $\sigma_{R_i}^2$, and lognormal and inverse gamma distributions were used as priors for the mean and variance of the random effect (\bar{R}_0 and $\sigma_{R_0}^2$). The prior distributions for each parameter are illustrated in Table 2.

In drug development, however, such a rich a priori information is not always available: weakly informative priors are often used to avoid that the results are overly influenced by potentially inaccurate prior distributions, as mentioned above. This section studies how prior informativeness, i.e. precision, influences parameter estimates and parameter correlations. The model described in Section 2.2 was run specifying three different prior distributions for the parameter I_{max} on data from trial 1, 2 and 3 pooled altogether, as these trials are analyzed during the first integration step (as specified in Section 3.3).

- (i) Prior 1 is the prior distribution that was used in La Gamba et al. [15], i.e., a beta distribution with expected value equal to the point estimate obtained from a frequentist analysis of the historical data, and standard deviation equal to the double of the standard error ($SD = 0.02$): $Prior1 \sim Beta(19.574, 174.532)$.
- (ii) Prior 2 has the same expected value as Prior 1, but doubled standard deviation ($SD = 0.04$): $Prior2 \sim Beta(3.557, 31.713)$.
- (iii) Prior 3 is a standard uniform distribution, i.e., $SD = 0.29$: $Prior3 \sim Beta(1, 1)$.

Parameter estimates and parameter correlations were consequently compared. The reason for the choice of parameter I_{max} for the prior specification evaluation is due to the fact that the estimate range of this parameter can be easily inferred from the historical trial, as it represents the maximal inhibition of body heat production. This eases the

assessment of the accuracy of the results from each analysis.

3.2 | Choice of random effects

As mentioned in Section 2.2, the K-PD model proposed by La Gamba et al. [15] was performed assuming a random body temperature at baseline. The choice of allocating random effects to certain parameters is not trivial in complex nonlinear ordinary differential equation (ODE) hierarchical models. In our case study, for example, the heterogeneity among animals belonging to the same treatment group could be observed, to some extent, in the temperature trend slopes asides from the baseline (Figure 1). The initial condition of indirect response models, $R_{i0} = k_{in}/k_{out}$ [22], implies that placing a random effect on R_{i0} leads to the implicit assumption of a random variation of k_{in} and k_{out} . However, it is not possible to detect which physiological parameter varies substantially between subjects solely through a graphical inspection. Therefore, a model assuming a random effect on k_{out} :

$$\log(k_{out_i}) \sim N(\mu_{\log(k_{out})}, \sigma_{\log(k_{out})}^2),$$

was compared with the random baseline model in terms of individual estimates and posterior predictions. Both models were performed on the eleven trials which were pooled together, keeping the same prior distributions as in the original analysis of the data [15], as described in the previous section. In particular, $\mu_{\log(k_{out})}$ was normally distributed: $\mu_{\log(k_{out})} \sim N(1.930, 6.840)$, whereas $\sigma_{\log(k_{out})}^2 \sim \text{Inv-Gamma}(2.022, 1.186)$. To ease the comparison, the random baseline model was also run assuming a log-normally distributed random effect:

$$\log(R_{0i}) \sim N(\mu_{\log(R_0)}, \sigma_{\log(R_0)}^2),$$

with priors $\mu_{\log(R_0)} \sim N(3.610, 0.002)$ and $\sigma_{\log(R_0)}^2 \sim \text{Inv-Gamma}(2.022, 1.186)$. Compared to the original model assuming a normally distributed random baseline, the greater a priori inter-individual variability of the latter model would allow to assess its performance under a more extreme scenario.

A model with random k_{in} was also evaluated, but it could not be fitted due to convergence issues.

3.3 | Sequential Bayesian integration

The aim of this section is to evaluate potentialities and challenges arising from integrating the trials in a Bayesian sequential framework and to assess the influence of factors such as study sample size or experimental design on the performance of this technique in comparison to pooling the trials together, originally explored by La Gamba et al. [15].

The methodological framework is as follows: let us suppose to have a number of trials conducted sequentially. At the first integration step, prior distributions for all parameters are chosen by setting the expected value to the point estimates obtained from the frequentist analysis of the historical data, whereas standard errors are doubled, consistently with the work of La Gamba et al. [15] (Table 2). At step s ($s = 2, \dots, S$) the priors are determined so that they have the same mean and variance as the posterior distributions obtained from step $s - 1$. From the second step onwards, the priors for α and β are assumed normally distributed, as a priori information is available from the previous integration steps. For all parameters, in case a posterior is found to have 2 or more modes, mixtures distributions are

considered.

In order to explore the impact of different factors on the performance of Bayesian sequential integration, the trials are sequentially pooled using four different methods which are then compared:

Method 1 This method respects the chronological order in which the trials were conducted (see Table 1). Thus, it is the most intuitive one and it is the type of sequential integration that would be chosen in practice, as each analysis would be performed each time new data become available. In order to allow the parameters α and β to be identifiable (such coefficients can only be estimated in presence of multiple dose combinations of both treatments), the first three trials are pooled together. Further, Bayesian integration was performed by adding one trial at a time.

Method 2 The eleven trials are integrated one at a time, but a random permutation of the trial order is used. It is therefore possible to check if the results of the sequential pooling are order-invariant. As before, the first three trials are pooled to guarantee identifiability of α and β .

Method 3 The trials are sequentially pooled three at a time, following the original trial order. [In this way, it is possible to assess if a higher sample size at each integration step has an impact on the final result.](#)

Method 4 New trials are sampled from the existing data such that each new trial contains one animal from each treatment group and from every single old trial (thus, for each dose combination tested). As such, a maximum exploration of the dose combination range can be guaranteed at each step of the sequential integration, [allowing to assess the impact of the experimental design on the final result.](#) Since each original trial contains five rats for each treatment group, five trials of 44 rats each are created.

Table 3 summarizes the sequential integration methods described above.

The rationale behind such comparison is that, if the elicited prior distributions for a trial are good approximations of the posterior distributions from the previous trial, it is in principle expected that each of the above listed sequential integration techniques would eventually lead to results which are comparable among each other, and comparable to the original Bayesian pooling described by La Gamba et al. [15].

To give a formal demonstration, let $\pi_1(\theta | X_1)$ be the posterior distribution of the parameter vector θ after observing data [at step 1](#), X_1 . Using Bayes' theorem:

$$\pi_1(\theta | X_1) \propto \mathcal{L}(X_1 | \theta) \pi_1(\theta), \quad (6)$$

where $\pi_1(\theta)$ is the prior distribution for θ , which was set based on the results from the analysis of the historical data, and $\mathcal{L}(X_1 | \theta)$ is the likelihood of the data [available during the first integration step](#) given θ . If the posterior in (6) is used to determine the prior for the analysis of data [at step 2](#):

$$\pi_1(\theta | X_1) = \pi_2(\theta), \quad (7)$$

it follows that the posterior distribution of θ after observing data [at step 2](#):

$$\pi_2(\theta | X_1, X_2) \propto \mathcal{L}(X_2 | \theta) \pi_1(\theta | X_1) \propto \mathcal{L}(X_2 | \theta) \mathcal{L}(X_1 | \theta) \pi_1(\theta). \quad (8)$$

where $\mathcal{L}(X_2 | \theta)$ is the likelihood of data [available during the second integration step](#).

Using the relationships in (7) and (8), it is possible to determine the posterior distribution at a generic step J :

$$\pi_J(\theta | \mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_J) \propto \prod_{j=1}^J \mathcal{L}(\mathbf{X}_j | \theta) \pi_1(\theta), \quad (9)$$

which, under the assumption of [trial exchangeability](#), is equal to:

$$\pi_J^P(\theta | \mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_J) \propto \mathcal{L}(\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_J | \theta) \pi_1(\theta), \quad (10)$$

that is, the posterior distribution resulting from the analysis of [all trials pooled together](#).

Although, as shown above, it is theoretically proven that the posterior of the sequential integration should coincide with the one resulting from pooling the trials together, in complex nonlinear ODE hierarchical models, where several numerical methods are used in the estimation process and in the approximation of the posterior distributions, [inconsistency between the two methods are to be expected](#). The sequential integration strategies mentioned above were compared among each other, and to the results of pooling the data altogether, in terms of parameter estimates and population predicted profiles.

4 | RESULTS

4.1 | Prior specification

The correlation matrices of the parameters of the K-PD model for synergy described in Section 2.2, performed on data from the first three trials using three different prior distributions for the parameter I_{max} (with a standard deviation equal to 0.02, 0.04 and 0.29) are shown in Figures 2, 3 and 4 respectively, whereas Table 4 compares the parameter estimates.

When a highly informative prior - incorporating the results of the historical trial - is used, the posterior parameter estimates are nearly uncorrelated: only a weak correlation can be observed between the parameters I_{max} and β , and between β and α . When the prior standard deviation is doubled, the correlations between the above mentioned parameters, as well as between k_e and k_{out} , and between I_{max} and k_{out} become stronger. These correlations increase even further when a non-informative (uniform) prior is chosen for I_{max} . This result shows that the use of highly informative priors reduces the correlation between parameters, that can compromise the validity of parameter estimation [33]: when weakly informative priors are chosen, the estimation method mainly relies on the current data and the chosen model; if the data are sparse, there is a lack of information which leads to a higher correlated parametric space, where parameters compensate each other. As such, the parameter estimates may be biased even though the model seems to fit the data well. For example, it is possible to observe (Table 4) that the increase of I_{max} (the maximal inhibition of body heat production) with increasing prior standard deviation is compensated by a decrease in the elimination rate of body heat k_{out} and an increase in β , which decreases the potency of the marketed treatment. However, it was known from the analysis of historical data, that I_{max} could not exceed a value of 0.18. Therefore, the estimates of I_{max} when $SD = 0.04$ and when $SD = 0.29$ cannot be considered reliable.

It should be noted that the low correlation among parameters following the specification of informative priors might be also due to the fact that the chosen prior distributions did not incorporate the correlations among parameters (as this would have increased the computational burden), thus setting a null a priori correlation. Such choice may be inappropriate in some contexts, but it prevents the above described parameter compensations arising from the use of

weakly or uninformative priors.

4.2 | Choice of random effects

In this section, the consequences of allocating a random effect on the parameter k_{out} are studied, and the performances of this random-effects model are compared with the model with a random baseline in terms of individual estimates and posterior predictions.

The individual body temperature estimates of data from trial 1 obtained from fitting the model with random k_{out} are shown in Figure 5 and compared with the individual estimates of the random baseline model (Figure 6). Both models show a good fit to the data, although some slight differences in the fitted trends were detected for the combination group.

When the posterior predictive profiles were computed (using the random effect mean rather than the subject-specific values and incorporating the inter-individual variability in the prediction intervals), substantial differences were observed between the two models: when the random effect was allocated to k_{out} (Figure 7), a systematic downward bias was observed in the prediction of the time profiles belonging to the combination group, whereas such bias is much reduced when the inter-individual heterogeneity was incorporated at a baseline level (Figure 8).

Through the inspection of the distribution of the posterior mean of subject-specific random effects of k_{out} (Figure 9), it was observed that the variable had approximately a bimodal distribution, where the values associated to the subjects belonging to the combination group were much lower compared to the other treatment groups. This is due to an overcompensation between the parameters k_{out} and β . In fact, only the combination group gives a contribution to the estimation of β : body temperature is at its steady state in the vehicle group and in the group that received the novel treatment only; in the group that received the marketed treatment only, IC_{50} is equal to 1. Thus, IC_{50} is expressed as a function of α and β only for the combination group. As mentioned in the work of La Gamba et al. [15] already, negative estimates of α and β imply that the presence of the novel treatment further decreases the body temperature, if co-administered with the marketed treatment. If the random effect is placed at k_{out} , the low values of the subject-specific random effects for the combination group (meaning a less pronounced dissipation of body heat) compensate the overly large decrease in body temperature induced by low values of α and β .

Nevertheless, when the posterior predictions are computed, the estimate of \bar{k}_{out} results on average particularly high in the combination group, thus producing an exaggerated and unrealistic reduction of body temperature.

Such behavior is very likely due to the correlation of the parameter space: specifically, the parameters β , k_{out} and T_{max} are highly correlated. The allocation of a random effect to a parameter which is part of the highly correlated parameter space can lead to a kind of escape route, where the subject-specific values of the random effect compensate other parameters. In this case, for example, both k_{out} and β compete to the estimation of the larger decrease in body temperature induced by the co-administration of marketed and novel compounds.

This result suggests that the allocation of the random effects to different parameters highly affects the performances of the model proposed.

It should be observed that the overcompensation among parameter estimates is not specific to the Bayesian approach, but it may also happen in a frequentist analysis. Such overcompensation, however, could sometimes be restrained in a Bayesian framework through the use of informative prior distributions, as shown in the previous Section.

4.3 | Sequential Bayesian integration

The posterior predictions of the resulting model at the end of each sequential Bayesian integration illustrated in section 3.3, on data from trial 1 are shown in Figures 10, 11, 12 and 13, respectively. Graphs displaying the parameter trends during the sequential integrations are shown in Appendix B.

Figures 10-13 illustrate that the models resulting from the first three types of sequential integration produce a substantial systematic underestimation in the predictions of the time profiles belonging to the combination group. Furthermore, the posterior predictive intervals are very large, meaning a large amount of uncertainty exists in predicting body temperature trends. Particularly, the models coming from the integration of one and three trials at a time keeping the original order show very similar predictions; the final model with the permuted trial order shows a lower underestimation, with narrower predictive intervals.

A noticeable improvement in prediction, however, can be observed for the model resulting from the sequential integration of five simulated trials containing all possible dose combinations: the underestimation is reduced, as well as the predictive intervals width. Moreover, the predictions resulting from such type of sequential integration are the most similar to the original Bayesian pooling described by La Gamba et al. [15] (Figures 6 and 8), although some differences between the posterior mean of parameters I_{max} , β and $\sigma_{R_0}^2$ can be noticed (see Appendix B).

Such results could be deemed counterintuitive: as mentioned in Section 3.3, assuming that the elicited prior distributions at step s ($s = 2, \dots, S$) are good approximations of the posterior distributions resulting from step $s - 1$, all different types of sequential Bayesian integration should lead to results which are equivalent to each other, and equivalent to the original Bayesian pooling. However, the set of relationships described in Formulas (6) - (10) may not be valid when the sequential integration is performed on small trials which do not provide enough information to estimate all parameters. In fact, during the first integration step using Methods 1 to 3, the information for the estimation of the eight model parameters originates from the assessment of three dose combinations only. Based on only few dose combinations, the MCMC algorithm may not be able to find a unique solution for the posterior estimation. This results in a highly correlated parameter space, and the posterior distributions arising from the first integration step might be biased.

Formally, let us suppose that the data from the first sequential integration step, \mathbf{X}_1^I , are scarce, or they are in general not informative enough for the estimation of the parameter vector θ . In this case, the posterior distribution $\pi_1^I(\theta | \mathbf{X}_1^I)$ may not be representative of the actual posterior distribution of θ :

$$\pi_1^I(\theta | \mathbf{X}_1^I) \neq \pi_1(\theta | \mathbf{X}_1), \quad (11)$$

which invalids the set of relationships in (6) - (10). In such scenario, the better performances of a sequential integration where the whole dose range is evaluated at each step would not be surprising: the time profiles corresponding to each treatment dose combination provide useful information to more accurately identify the real exposure-response relationship, which allows better predictions. As previously mentioned, although Method 4 shows better predictive performances, the different values observed in some of the final parameter estimates (see Appendix B) suggest that this last type of sequential integration cannot be considered equivalent to pooling the trials altogether. To guarantee the maximum accuracy level, an optimal sample size as well as sampling times should be determined, in addition to an optimal number of dose combinations. This is a topic of further research.

5 | DISCUSSION

The Bayesian K-PD model for synergy described by La Gamba et al. [15] was performed on several [preclinical](#) trials conducted in different time periods. These trials were pooled together, using prior knowledge from a historical dose-response trial of the marketed compound. In this work, the trials are integrated in a sequential way: the resulting posterior distributions from a trial are used to determine the priors for the analysis of the following trial. The advantage of adopting such approach in drug development is that it allows to analyze each trial immediately, instead of waiting for the end of data collection before running the analysis, or reanalyzing all available data up to the current trial, with a consequent computational burden. Moreover, the parameter estimates resulting from each sequential analysis step might be useful information for the design of the next trials. [Compared to the stochastic particle algorithm described by Micallef et al. \[19\], our method allows to incorporate multiple observations at each integration step: particle filters are only suitable when very limited data are added at each step; if a higher amount of data is included at once, the phenomenon called “particle depletion” can occur, i.e., one or few particles have a way higher probability, whereas all the others are not retained after the filtering step. Moreover, prior down-weight can be implemented in the case of incompatibility between previous information and newly observed data. Despite such benefits, performing Bayesian sequential integration of small preclinical trials using a nonlinear hierarchical K-PD model has several methodological implications, which were discussed in this manuscript.](#)

The theoretical foundation underlying the use of this technique is that the resultant posterior distributions would eventually be equivalent to the ones resulting from pooling the trials together. This is valid in case the elicited prior distributions for a trial are good approximations of the posterior distributions from the previous trial, and under parameter identifiability at each integration step. Parameter identifiability is challenging to assess in non-linear hierarchical models based on ordinary differential equations, and is still object of ongoing research [38]. Although a model can be proven to be identifiable when data is pooled together, identifiability issues may arise during the sequential integration, due to the fact that during the first integration steps the data do not provide enough information for the estimation of all parameters. [To illustrate this, one can consider the situation where only one of few dose combinations provide little information for the estimation of the interaction coefficient \$\beta\$. For this reason, different precautions should be taken before and during the sequential integration performance to avoid such identifiability issues.](#)

Firstly, [it was shown that](#) the use of weakly or non-informative priors [in the analysis of few small trials](#) increased the parameter correlation, which leads to a situation where some parameters compensate each other, compromising the accuracy of parameter estimates. This result suggests that, in case of little data during the first integration steps, the use of informative priors is recommended for identifiability of the model, as borrowing information from previous trials reduces the correlation of the parameter space and causes more stability and less convergence problems. On the other hand, the choice of priors which are too informative leads to a restrictive exploration of the parameter space during the MCMC sampling, which should be avoided. Thus, a trade-off in terms of prior informativeness should be carefully chosen by lowering, for example, the precision of the prior distributions representing the information coming from possible historical studies. A sensitivity analysis of the priors would enable to gain insights in the model performance in this respect.

A high parameter correlation is typical in nonlinear hierarchical PK-PD models, especially when sample size is small or the available dose range is narrow. One of its consequences concerns the fact that placing the random effect on a parameter that is part of a highly correlated parameter space leads to parameter overcompensation, where the population parameter on which the random effect is allocated incorporates part of the effect expressed by other parameters. A way to avoid such aberration is to limit the variability of the highly correlated parameters, by allocating the random effects on parameters that are easier to estimate, when it is feasible. In our case study, the random baseline

model had better performances because the temperature at baseline could easily be inferred from the data, and such parameter was weakly correlated with the others. Moreover, the functional relationship between baseline and k_{out} allowed the implicit assumption of random slope variability as well.

A careful prior elicitation and random effect allocation may not be enough to warrant an unbiased sequential integration, **especially when the trials are small**: such approach is successful only if adequate dose combinations, sampling times and replications are explored at each integration step. Our analysis showed that one or few fixed dose combinations per trial were producing biased results, and that such results were dependent on the order in which the trials were integrated. This highlights the need of thorough study planning, as the integration of a reduced number of well designed trials would have generated more accurate results in shorter time frames. In reality, it is not always possible to obtain ideally designed data for a number of reasons such as ethical or logistic considerations. In this case, a more pragmatic approach is required, such as prior elicitation from literature or from previous studies and the use of informative priors in the initial step of sequential data integration. However, it is crucial to be aware of the pitfalls of such approach, as stated in the current paper.

It should be noted that, although this manuscript focuses on the Bayesian framework, challenges such as parameter overcompensation following the choice of the random effect on a parameter that is difficult to estimate, or parameter identifiability issues arising from the analysis of small data using complex nonlinear models may occur in a frequentist framework as well. As such, the Bayesian approach allows to prevent some of these issues, if it is possible to borrow information from previous trials through the use of informative prior distributions.

The major limitation of our work is that parameter correlation was not incorporated through the use of multivariate prior distributions, as this would further increase the computational burden and convergence issues. This may imply a progressive information loss during the integration steps. However, specifying multivariate prior distributions is not trivial [39], especially in case of different marginal densities (in our case study, lognormal, beta and inverse-gamma distributions were used). In this respect, approaches such as Bayesian nonparametric techniques may be employed, and will be object of further research [40].

In conclusion, the sequential integration of PK-PD data in a Bayesian framework may be a promising approach, as it allows an early assessment of parameter estimates in reduced time frames. The assessment of a wide dose range is nevertheless advisable, especially in early studies. **Further work is being devoted to extending the findings of our case study to more general scenarios.**

REFERENCES

- [1] Felmler MA, Morris ME, and Mager DE. Mechanism-Based Pharmacodynamic Modeling. *Methods Mol Biol*, 929:583–600, 2012.
- [2] Mould DR and Upton RN. Basic Concepts in Population Modeling, Simulation, and Model-Based Drug Development—Part 2: Introduction to Pharmacokinetic Modeling Methods. *CPT Pharmacometrics Syst Pharmacol*, 2:e38, 2013.
- [3] Upton RN and Mould DR. Basic Concepts in Population Modeling, Simulation, and Model-Based Drug Development: Part 3—Introduction to Pharmacodynamic Modeling Methods. *CPT Pharmacometrics Syst Pharmacol*, 3:e88, 2014.
- [4] Wakefield JC, Smith AF, Racine-Poon A, and Gelfand AE. Bayesian analysis of linear and non-linear population models using the Gibbs sampler. *J R Stat Soc Ser C Appl Stat*, 43:201–221, 1994.
- [5] Wakefield JC. Bayesian analysis of population pharmacokinetic models. *J Am Stat Assoc*, 91:62–75, 1996.
- [6] Lunn DJ, Best N, Thomas A, Wakefield J, and Spiegelhalter D. Bayesian analysis of population PK/PD models: general concepts and software. *J Pharmacokinetic Pharmacodyn*, 29:271–307, 2002.
- [7] Jonsson F, Jonsson EN, Bois FY, and Marshall S. The application of a Bayesian approach to the analysis of a complex, mechanistically based model. *J Biopharm Stat*, 17:65–92, 2007.
- [8] Zhou H, Hartford A, and Tsai K. A Bayesian approach for PK/PD modeling with PD data below limit of quantification. *J Biopharm Stat*, 22:1220–43, 2012.
- [9] Lyons MA and Lenaerts AJ. Computational pharmacokinetics/pharmacodynamics of rifampin in a mouse tuberculosis infection model. *J Pharmacokinetic Pharmacodyn*, 42:375–389, 2015.
- [10] Erhardt EM, Ursino M, Biewenga J, Jacobs T, and Gasparini M. Bayesian knowledge integration for an in vitro-in vivo correlation model. *Biom J*, page 1–16, 2018.
- [11] Chang M and Boral A. ABC of Bayesian Approaches to Drug Development. *Pharmaceut Med*, 22:141–150, 2008.
- [12] Kathman SJ, Williams DH, Hodge JP, and Dar M. A Bayesian population PK-PD model of ispinesib-induced myelosuppression. *Clin Pharmacol Ther*, 81:88–94, 2007.
- [13] Gueorguieva I, Aarons L, and Rowland M. Diazepam pharmacokinetics from Preclinical to Phase I Using a Bayesian Population Physiologically Based Pharmacokinetic Model with Informative Prior Distributions in Winbugs. *J Pharmacokinetic Pharmacodyn*, 33:571–94, 2006.
- [14] Petit C, Samson A, Morita S, Ursino M, Guedj J, Jullien V, Comets E, and Zohar S. Unified approach for extrapolation and bridging of adult information in early-phase dose-finding paediatric studies. *Stat Methods Med Res*, 2016. 962280216671348.
- [15] La Gamba F, Jacobs T, Geys H, Ver Donck L, and Faes C. A Bayesian K-PD model for synergy: A case study. *Pharm Stat*, page 1–11, 2018.
- [16] Arulampalam MS, Maskell S, Gordon N, and Clapp T. A Tutorial on Particle Filters for Online Nonlinear/Non-Gaussian Bayesian Tracking. *IEEE Trans Signal Process*, 50:174–188, 2002.
- [17] Särkkä S. *Bayesian Filtering and Smoothing*. Cambridge University Press, Cambridge (UK), 2013.
- [18] Botsman K, Tickle K, and Smith JD. A Bayesian formulation of the Kalman filter applied to the estimation of individual pharmacokinetic parameters. *Comput Biomed Res*, 30:83–94, 1997.
- [19] Micallef S, Amzal B, Bach V, Chardon K, Tourneux P, and Bois FY. Sequential updating of a new dynamic pharmacokinetic model for caffeine in premature neonates. *Clin Pharmacokinetic*, 46:59–74, 2007.

- [20] Janssen PA, Niemegeers CJ, Awouters F, Schellekens KH, Megens AA, and Meert TF. Pharmacology of risperidone (R 64 766), a new antipsychotic with serotonin- S_2 and dopamine- D_2 antagonistic properties. *J Pharmacol Exp Ther*, 244:685–693, 1988.
- [21] Megens AA, Hens KA, and Vermeire JC (et al). *Neurological safety margin and behavioral side-effect liability of the fastdissociating D2 antagonist JNJ-37822681 in rats: an in vivo pharmacological comparison with 8 reference antipsychotics (Data available on file)*. Non-Clinical Pharmacology research report on JNJ-37822681-AAA, EDMS-PSDB-5345985:2.0; Beers (BE): Johnson & Johnson Pharmaceutical Research & Development, revised January 2010.
- [22] Dayneka NL, Garg V, and Jusko WJ. Comparison of four basic models of indirect pharmacodynamic responses. *J Pharmacokinet Biopharm*, 21:457–478, 1993.
- [23] Jacqmin P, Snoeck E, van Schaick EA, Gieschke R, Pillai P, Steimer JL, and Girard P. Modelling Response Time Profiles in the Absence of Drug Concentrations: Definition and Performance Evaluation of the K-PD Model. *J Pharmacokinet Pharmacodyn*, 34:57–85, 2007.
- [24] Carpenter B, Gelman A, Hoffman MD, Lee D, Goodrich B, Betancourt M, and Brubaker M (et al). Stan: A Probabilistic Programming Language. *J Stat Softw*, 76:1–32, 2017.
- [25] Gelman A, Carlin JB, Stern HS, Dunson DB, Vehtari A, and Rubin DB. *Bayesian Data Analysis, Third Edition*. Chapman and Hall/CRC Press, Boca Raton (FL), 2013.
- [26] van de Schoot R, Broere JJ, Perryck KH, Zondervan-Zwijnenburg M, and van Loey NE. Analyzing small data sets using Bayesian estimation: the case of posttraumatic stress symptoms following mechanical ventilation in burn survivors. *Eur J Psychotraumatol*, 6:25216, 2015.
- [27] Gelman A, Bois F, and Jiang J. Physiological pharmacokinetic analysis using population modeling and informative prior distributions. *J Am Stat Assoc*, 91:1400–1412, 1996.
- [28] Wadsworth I, Hampson LV, and Jaki T. Extrapolation of efficacy and other data to support the development of new medicines for children: A systematic review of methods. *Stat Methods Med Res*, 2016. 0962280216631359.
- [29] EMA. *Reflection paper on the use of extrapolation in the development of medicines for paediatrics (draft)*. 2017.
- [30] Viele K, Berry S, Neuenschwander B, Amzal B, Chen F, Enas N, Hobbs B, Ibrahim JG, Kinnersley N, Lindborg S, Micallef S, Roychoudhury S, and Thompson L. Use of historical control data for assessing treatment effects in clinical trials. *Pharm Stat*, 13:41–54, 2014.
- [31] Ibrahim JG, Chen MH, Gwon Y, and Chen F. The power prior: theory and applications. *Stat Med*, 34:3724–49, 2015.
- [32] Takeda K, Oba M, Kakizume T, Sakamaki K, Taguri M, and Morita S. Bayesian Approach to Utilize Historical Control Data in Clinical Trials. *Japanese Journal of Biometrics*, 36:25–50, 2015.
- [33] Li B, Sun Z, He Q, Zhu Y, and Qi ZS. Bayesian inference with historical data-based informative priors improves detection of differentially expressed genes. *Bioinformatics*, 32:682–689, 2016.
- [34] Van Dongen S. Prior specification in Bayesian statistics: three cautionary tales. *J Theor Biol*, 242:90–100, 2006.
- [35] Morita S, Thall PF, and Müller P. Determining the effective sample size of a parametric prior. *Biometrics*, 64:595–602, 2008.
- [36] Lehermeier C, Wimmer V, Albrecht T, Auinger HJ, Gianola D, Schmid VJ, and Schön CC. Sensitivity to prior specification in Bayesian genome-based prediction models. *Stat Appl Genet Mol Biol*, 12:375–91, 2013.
- [37] Thall PF, Herrick RC, Nguyen HQ, Venier JJ, and Norris JC. Effective sample size for computing prior hyperparameters in Bayesian phase I-II dose-finding. *Clin Trials*, 11:657–66, 2014.

-
- [38] Shivva V, Korell J, Tucker IG, and Duffull SB. An Approach for Identifiability of Population Pharmacokinetic-Pharmacodynamic Models. *CPT Pharmacometrics Syst Pharmacol*, 2:e49, 2013.
- [39] Garthwaite PH, Kadane JB, and O'Hagan A. Statistical methods for eliciting probability distributions. *JASA*, 100:680-701, 2005.
- [40] Moala FA and O'Hagan A. Elicitation of multivariate prior distributions: A nonparametric Bayesian approach. *Journal of Statistical Planning and Inference*, 140:1635-1655, 2010.

TABLE 1 Dose levels used in current trials of drug-drug interactions.

trial	1	2	3	4	5	6	7	8	9	10	11
Marketed compound (mg/kg)	10	2.5	10	0.63	10	0.16	2.5	0.63	0.16	0.04	0.04
Novel compound (mg/kg)	40	40	10	40	2.5	40	10	10	10	10	40

TABLE 2 Prior distributions and standard deviations relative to the parameters of the Bayesian K-PD model for synergy, first integration step.

Parameter	Prior distribution	SD
k_e	$\log N(-0.115, 0.514)$	0.559
k_{out}	$\log N(1.930, 6.840)$	Uninformative*
I_{max}	$Beta(19.574, 174.532)$	0.022
\bar{R}_0	$\log N(3.61, 0.002)$	0.074
α	$Unif(-10, 10)$	5.773
β	$Unif(-10, 10)$	5.773
$\sigma_{R_0}^2$	$Inv-Gamma(2.022, 1.186)$	3.910
σ_R^2	$Inv-Gamma(54.441, 24.476)$	0.063

* The standard error of the estimate of the parameter k_{out} resulting from the analysis of the historical trial on the marketed compound was high. Therefore, the prior distribution for k_{out} can be considered uninformative.

TABLE 3 Different Bayesian integration strategies evaluated. In Method 4, each new trial contains one animal from each treatment group and from every single old trial.

Method	Trial integration sequence
1: Original trial order	1, 2, 3 → 4 → 5 → 6 → 7 → 8 → 9 → 10 → 11
2: Random order permutation	5, 3, 8 → 11 → 6 → 1 → 2 → 9 → 7 → 4 → 10
3: Three trials at a time	1, 2, 3 → 4, 5, 6 → 7, 8, 9 → 10, 11
4: Five new trials	$1^{new} \rightarrow 2^{new} \rightarrow 3^{new} \rightarrow 4^{new} \rightarrow 5^{new}$

TABLE 4 Results of the Bayesian model performed on data from trials 1, 2 and 3 with three different priors for I_{max} : posterior mean and [standard deviation](#).

Parameter	$SD = 0.02, ESS = 194.106$	$SD = 0.04, ESS = 35.27$	$SD = 0.29, ESS = 2$
k_e	0.53 (0.08)	0.55 (0.08)	0.61 (0.10)
k_{out}	1.15 (0.17)	0.91 (0.17)	0.78 (0.16)
I_{max}	0.15 (0.02)	0.20 (0.03)	0.24 (0.05)
\bar{R}_0	37.12 (0.07)	37.15 (0.06)	37.16 (0.06)
α	-1.42 (0.40)	-1.58 (0.27)	-1.55 (0.24)
β	-2.85 (2.08)	-0.51 (0.89)	-0.13 (0.56)
$\sigma_{R_0}^2$	0.31 (0.08)	0.26 (0.07)	0.25 (0.07)
σ_R^2	0.41 (0.03)	0.41 (0.03)	0.41 (0.03)

SD: standard deviation; ESS: Prior's Effective Sample Size.

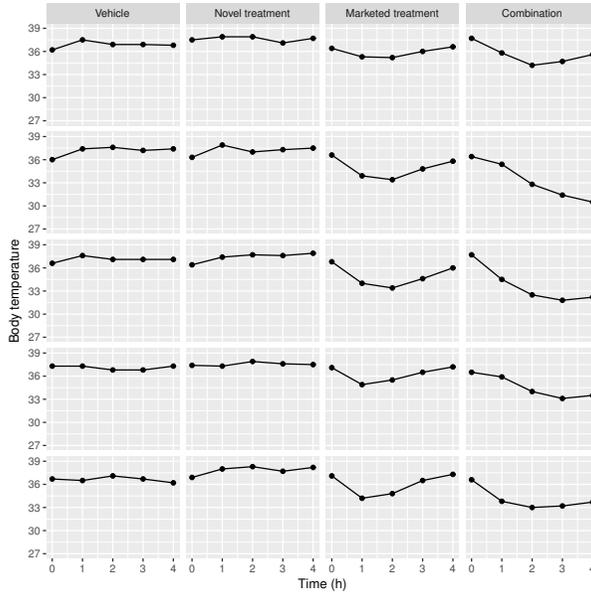


FIGURE 1 Time profiles of the individuals from trial 1: the first column represents the animals belonging to the vehicle group; the second and third columns show the animals assigned to the novel and marketed treatment, respectively; the last column illustrates the combination group.

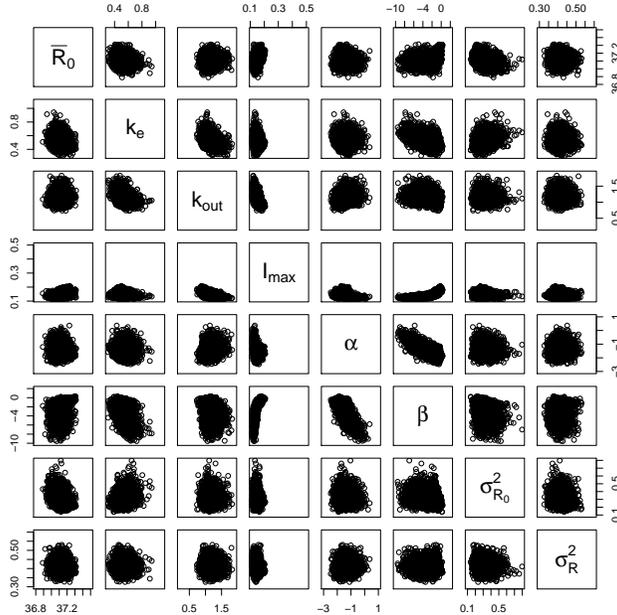


FIGURE 2 Impact of prior (prior 1): Correlation matrix relative to the parameters of the K-PD model for synergy.

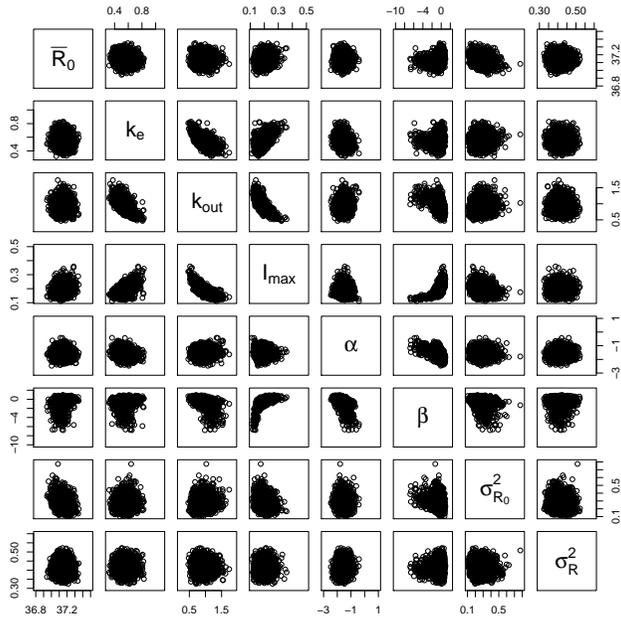


FIGURE 3 Impact of prior (prior 2): Correlation matrix relative to the parameters of the K-PD model for synergy.

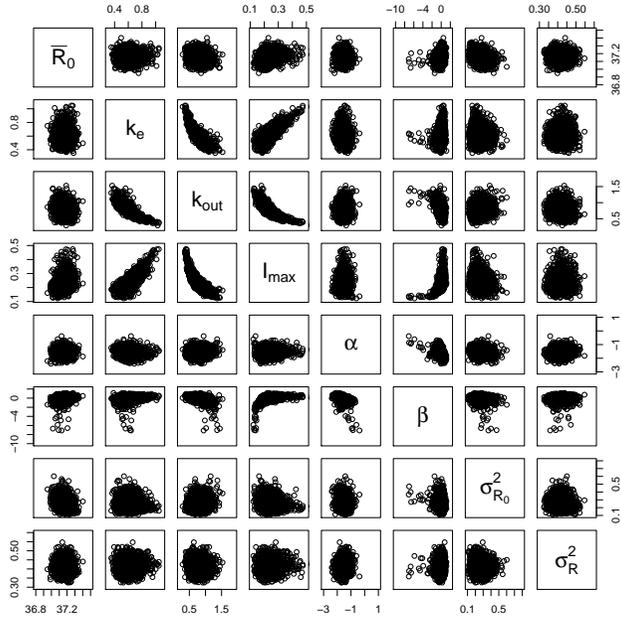


FIGURE 4 Impact of prior (prior 3): Correlation matrix relative to the parameters of the K-PD model for synergy.

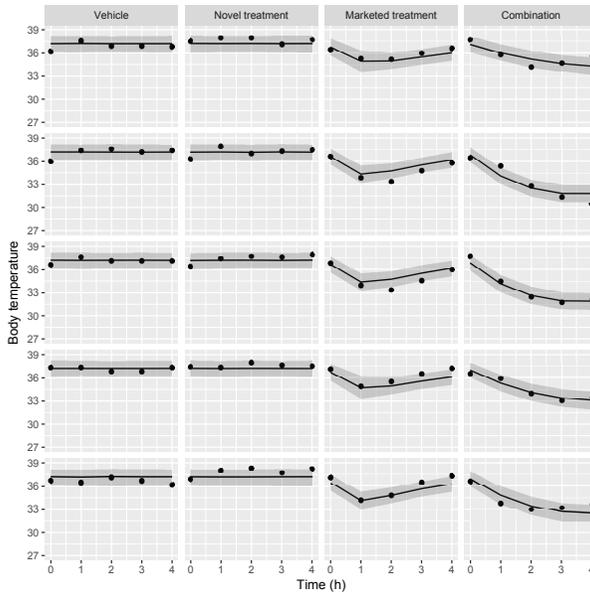


FIGURE 5 Impact of random effect (random k_{out}): Individual estimates (solid line) and credible intervals (grey bands), trial 1.

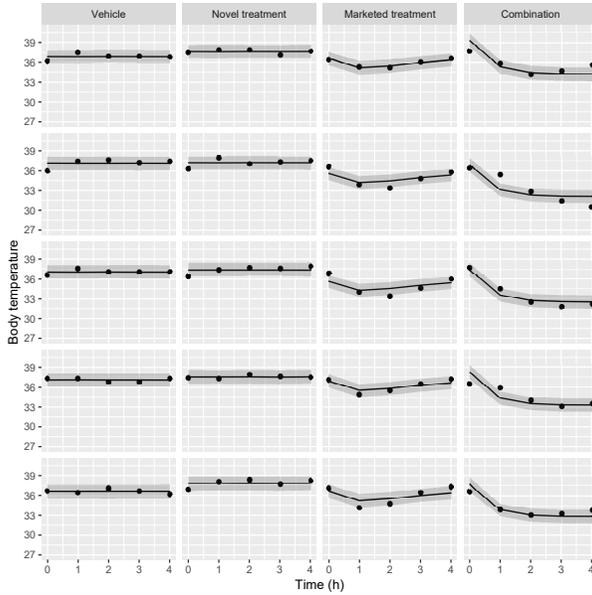


FIGURE 6 Impact of random effect (random R_0): Individual estimates (solid line) and credible intervals (grey bands), trial 1.

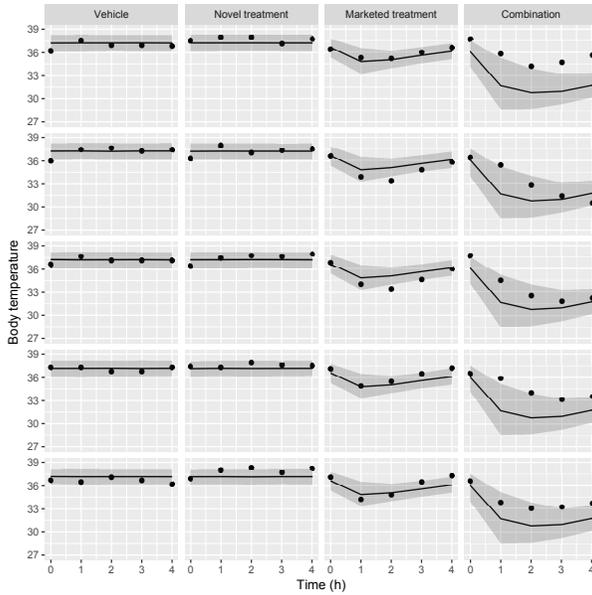


FIGURE 7 Impact of random effect (random k_{out}): Posterior predictions (solid line) and posterior predictive intervals (grey bands), trial 1.

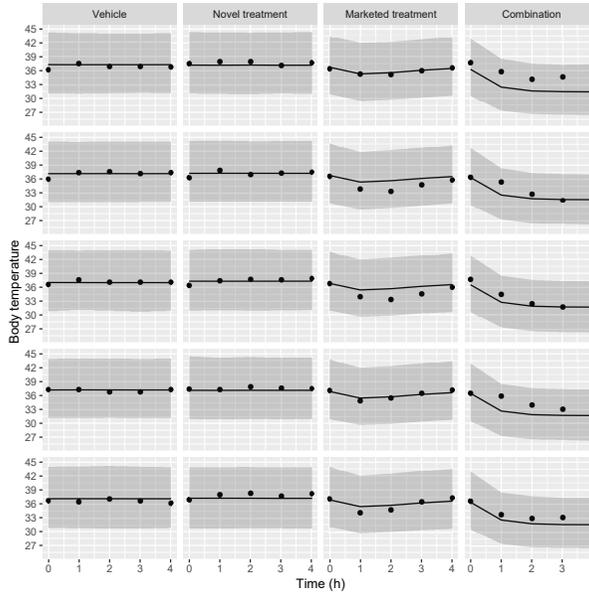


FIGURE 8 Impact of random effect (random R_0): Posterior predictions (solid line) and posterior predictive intervals (grey bands), trial 1

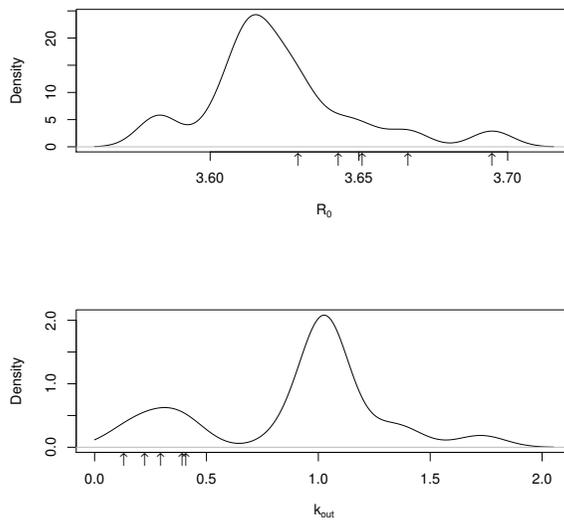


FIGURE 9 Kernel density estimation of the distribution of the posterior means of the subject-specific random effects values relative to trial 1, using the model with random R_0 (upper) and the model with random k_{out} (lower). The arrows represent the posterior estimates of the values for the subjects belonging to the combination group.

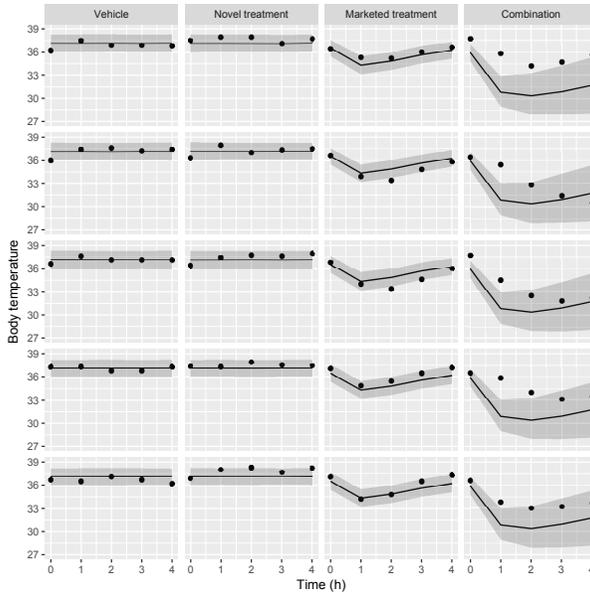


FIGURE 10 Impact of sequential integration (method 1): Posterior predictions (solid line) and posterior predictive intervals (grey bands), trial 1.

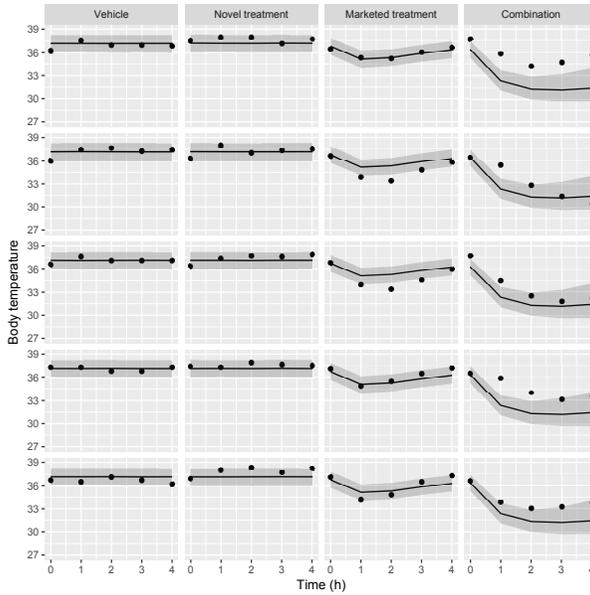


FIGURE 11 Impact of sequential integration (method 2): Posterior predictions (solid line) and posterior predictive intervals (grey bands), trial 1.

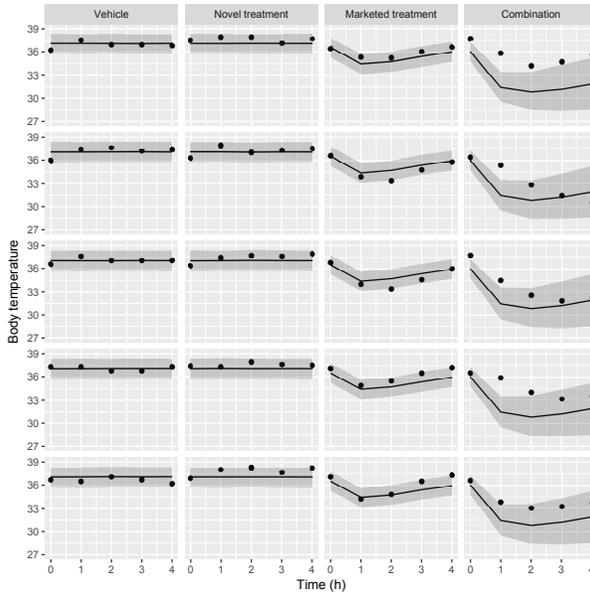


FIGURE 12 Impact of sequential integration (method 3): Posterior predictions (solid line) and posterior predictive intervals (grey bands), trial 1.

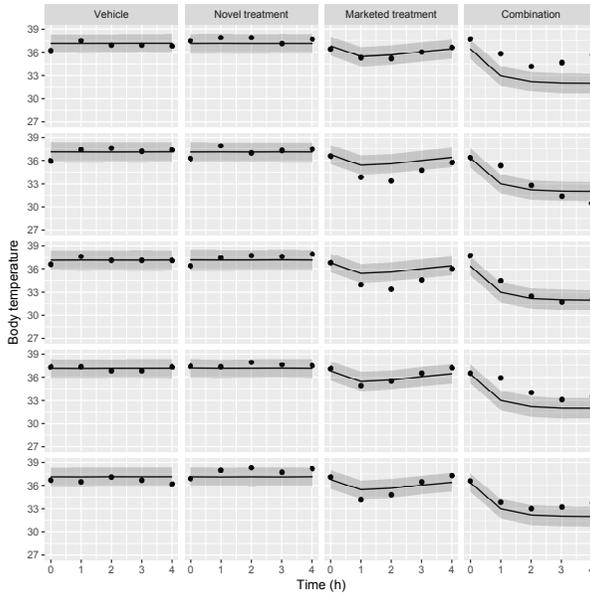


FIGURE 13 Impact of sequential integration (method 4): Posterior predictions (solid line) and posterior predictive intervals (grey bands), trial 1.