

TUTORIAL

Many Flavors of Model-Based Meta-Analysis: Part II – Modeling Summary Level Longitudinal Responses

Martin Boucher* and Meg Bennetts 

Meta-analyses typically assess comparative treatment response for an end point at specific timepoints across studies. However, during drug development, it is often of interest to understand the response time-course of competitor compounds for a variety of purposes. Examples of such application include informing study design and characterizing the onset, maintenance, and offset of action. This tutorial acts as a “points for consideration” document, reviews relevant literature, and fits a longitudinal model to an example dataset.

CPT Pharmacometrics Syst. Pharmacol. (2018) 7, 288–297; doi:10.1002/psp4.12299; published online 20 April 2018.

Part I of the model-based meta-analysis (MBMA) tutorial highlighted the critical importance for companies developing drugs to understand the key safety and efficacy attributes of other compounds, either on the market or in the pipeline.¹ The focus of many articles is to present study results based on primary and secondary end points in which more often than not, these end points will be landmark, for instance at a specific timepoint or an event, such as “end of study.” However, it is common for published study results to also include time-course information, usually in a graph or a table, thus providing a much more informative view of the data than a landmark end point.

Fitting models to the time-course of response can have many benefits especially in the learning phase of development (typically phase II) but it can also impact the confirmatory stage (typically phase III). Having clear research questions for a drug’s development will help decide whether longitudinal modeling will add value for a project team.

A benefit of modeling time-course data is to understand the full response profile for different compounds and/or placebo. This includes the onset of action, maintenance of effect, and any offset of response. Two competing drugs may have similar efficacy at week 6 but, with all other things being equal, a compound with a quicker onset of action is likely to be preferred by patients. A response that is maintained over a significant time-frame should be more meaningful than a response at a single point in time.

A second example is when previous proof of concept (POC) trials in a specific indication were typically 6 weeks in length but longitudinal data from historical trials of another mechanistically similar drug, to the one in development, demonstrate that a strong response can be shown as early as week 2. This could result in designing shorter POC trials in the future for this indication.

In addition, following the readout of a short-term POC study, it is useful to predict a response in a study of longer duration. Wang *et al.*² looked at how predictive a short-term result (3 months) was of a longer-term response (6 months) using a longitudinal model for responder rates. By using existing time-course information from similar types of

compounds reported in the literature, a scalar between two timepoints could be estimated and applied to the POC study result to generate the predictions for a future longer study design.

Depending on the types of models used, interpolation can be used to get estimates at timepoints that have been little studied (or not at all). This will be discussed further, later in the tutorial.

First, the aim of this tutorial is to review previous publications on longitudinal meta-analyses. Second, we wish to highlight important considerations in modeling time-course data. Finally, an maximum effect (E_{max}) model is fitted to an example osteoarthritis (OA) pain dataset using a selection of commonly used software in the pharmaceutical industry within (but not exclusively) Clinical Pharmacology. Landmark meta-analyses at selected timepoints will also be compared with the longitudinal model estimates. The dataset, model code, and outputs from the modeling will be provided in the **Supplementary Materials**. The primary focus here is on modeling aggregate data and, hence, MBMA of individual patient data (IPD) or combined aggregate data/IPD data will not be covered.

REVIEW OF LITERATURE COVERING LONGITUDINAL MBMA

The number of published articles on longitudinal meta-analyses, particularly for methodology development, is relatively small, as is the case for MBMA methodology in general. However, there has been a fairly consistent rate of such publications during the last 10 years, particularly with regard to the application of MBMA methods.

In terms of methodology, several articles have focused on the issue of accounting for residual correlations between timepoints, comparing the different approaches used to account for them, and the consequences of not doing so. Ishak *et al.*³ fitted a variety of models to deep-brain stimulation data from patients with Parkinson’s disease, all of which accounted for the correlations between timepoints within a study treatment arm. The article concluded that

accounting for these correlations could result in a better model fit and more precise parameter estimates. Similarly, Musekiwa *et al.*⁴ discussed and compared different covariance structures when fitting linear mixed effect models to an example dataset of 17 trials, which compared the combination of radiotherapy and chemotherapy with radiotherapy alone. With more of an MBMA focus, Ahn & French⁵ expanded on the work by Ishak *et al.*³ to nonlinear time-course and dose-response models with an emphasis on how compound symmetry correlations can be accounted for using the NONMEM software package and then, quantifying the resulting bias of not accounting for this correlation appropriately through an extensive simulation exercise.

Although E_{\max} and exponential models are common for fitting time-course in the Clinical Pharmacology arena, there are plenty of alternatives. Jansen *et al.*⁶ presented a network meta-analysis in which the time-course for OA pain was incorporated using fractional polynomials. These are flexible nonlinear models and are an extension of polynomial models.⁷ Whereas their flexibility is certainly an advantage, the parameter estimates themselves may not be so useful or intuitive in the same way that E_{\max} or ED_{50}/ET_{50} parameters tell us about relative maximal effects or potency/onset of action. Luu *et al.*⁸ used a cosine model to reflect circadian intraocular pressure of patients with glaucoma or ocular hypertension. There are also examples of authors transforming discrete measurements, such as responder rates or bounded mean pain scores, into logit space and then applying nonlinear models.^{9,10}

There are several other examples of applications of longitudinal MBMA that we do not discuss here but a table of these is provided in the **Supplementary Materials Table S1**, which includes the disease area, end points, and types of longitudinal models fitted. Additionally, further examples are referenced separately in the discussion section.

IMPORTANT CONSIDERATIONS WITH LONGITUDINAL MBMA

Time: Continuous variable or factor

Clinical pharmacologists conventionally consider “time” and “dose” to be continuous variables, as demonstrated by the routine use of E_{\max} or exponential models. These models readily support interpolation and, less commonly, extrapolation for future study prediction. Repeated measures analysis, in which time is treated as a factor, is more commonly used by statisticians to be used with good effect to describe the data at each timepoint with fewer assumptions. In summary, treating time as continuous is advisable when the goal of the analysis is broader than purely description, such as prediction, clinical trial simulation, or to reflect underlying pharmacology.

Different imputation methods

One of the major challenges with MBMA longitudinal data is that published summary level data are analyzed and reported in disparate ways across articles. Different articles may summaries different timepoints: some with a rich time-course of many results and others with just a baseline and end of study result. Access to original reports, from

company or regulatory websites, may be used to fill in unpublished timepoints, if available.

Identifying the method used by each publication to handle missing data at each timepoint and then determining how to account for any differences can be a major undertaking.¹¹ The end point time-course data in plots or tables may not be based on the same imputation method as the final end point analysis. For example, observed case (OC; no imputation) time-course data for the measure of interest can be plotted alongside the end of study statistical analysis result for which an imputation method has been applied, such as: last observation carried forward (LOCF), worst observation carried forward, baseline observation carried forward (BOCF), and multiple imputation, among others.¹²

End-point relationships are likely to vary over time depending on the imputation method used. However, this may not be important if we are only looking at early timepoints in which dropout is likely to be minimal. However, as the length of the study increases, differences in imputation method will become more of an issue. Access to IPD can reduce these differences by re-imputing summary level data using the most relevant method. However, for published data this is rarely possible and other methods accounting for these differences will need to be considered.

One straightforward approach would be to only include articles with an imputation method that matches the one planned for use in the specific drug development program the MBMA is aimed to inform. Alternatively, multiple independent MBMAs could be performed, one for each of the different imputation methods. However, this would be both time-consuming and not an efficient way to use the full dataset. A further approach might be to combine all the data from different imputation methods and then fit a covariate to estimate the effect for each method (e.g., LOCF vs. OC vs. BOCF). In this covariate approach, some studies may report summary data for more than one imputation method; in this case, correlation between endpoints/imputation methods, within trials, would also need to be considered.

Outcome reporting bias

The issue of outcome reporting bias needs careful consideration and arises when some or all of the included publications do not present the full time-course results of their underlying studies.¹³ It is important to understand which studies are contributing to the information at each timepoint and creating a table to show how this would be instructive.

Published plots will not always present the corresponding SEs of the means and, when they do, it may be difficult to digitize these accurately if the plots are of poor quality or the points/bars overlap. Digitization of these plots also introduces another source of transcription error. When OC data are presented on a plot without precision estimates and/or sample size information (which change over time when there is dropout), the analyst is left with the dilemma of how to weigh the residuals.

Correlations between timepoints within treatment arms

For time-course models, it is important to acknowledge that mean responses for a study arm will be correlated between

timepoints because these responses include the same individuals, subject to dropout and imputation method. Ignoring such correlation could lead to over precise estimates, bias, and more weight assigned to studies with many timepoints over those with few timepoints. It should be noted that correlations at the summary level may be different to those at the IPD level.

If correlation is accounted for using study-level and arm-level random effects (one of the approaches outlined in Ahn & French⁵), then arm-level effects should be nested within the corresponding study-level effects. The NONMEM version 7.3 software package has the option to include more levels of random effects with such nesting compared with previous versions so that setting up compound symmetry is now much easier.¹⁴

To implement in R, the nonlinear mixed effect (NLME) function for fitting NLME models has options for defining different correlation structures (e.g., compound symmetry, autoregressive (AR(1) etc)). The NONMEM version 7.3 also allows the use of AR residuals. In BUGS, compound symmetry can be set up either using the approach outlined in Ahn & French,⁵ which is discussed in the previous section, or by constructing the residuals in matrix form. The latter method is not straightforward when there are trials with different “dimensions” of timepoints.

As a recommendation, we suggest that, once the best available structural model has been chosen and fitted, the time-course of residuals are plotted by study and treatment. If runs of positive or negative values are observed within a study treatment arm, then this would suggest that there is still some residual correlation that needs to be accounted for. Compound symmetry would be a good starting point, only proceeding to other methods, such as AR, if necessary.

Residual weighting

Ideally for mean data, residuals should be weighted by the precision of the mean (the reciprocal of $SE^2 = 1/(SE^2) = 1/(SD^2/N)$). We recommend that available SDs over time are routinely plotted, as part of the exploratory analysis, to assess potential models for SD imputation, if required. This can also be a useful way of identifying unusual values (such as a SE being reported as an SD or vice-versa). There are many approaches to deal with missing SDs but Boucher used an NLME model to impute missing SDs over time for this type of data.¹⁵ Wiebe *et al.*¹⁶ provided a nice review of methods that have been used to impute missing variance data.

If, however, the number of missing SDs is high then it may be more appropriate to weight using the sample size. This makes an assumption that the within study SDs are the same across all studies and timepoints, which may not be a realistic assumption depending on the design and population characteristics of the included trials.

Covariate effects

In traditional meta-analyses, the term “meta-regression” is often used to describe the fitting of covariates, although “covariate analysis” is a more commonly used phrase in Clinical Pharmacology. Compared with landmark meta-analyses, there are potentially more parameters to which

covariates could be fitted when modeling longitudinal data. However, the limitations of covariates at the summary level remain the same, namely: small ranges of observed values at the mean level; inability to use the summary level covariate to make inferences at the patient level, due to ecological bias; and the often small numbers of studies involved in an MBMA.¹⁷ Plotting the relationship between endpoint and potential covariates is the first step to identifying those to take forward into a covariate analysis.

Between-study variability

The ability to get a good estimate of between-study variability will largely depend on the number of trials that are available for the analysis. If there are insufficient studies to get a good estimate, then one option would be to take the Bayesian approach and use a prior based on a similar analysis of different data. As will be seen in the example, longitudinal models provide scope for more than one random effect.

In the first tutorial, Q and I^2 were discussed as approaches to assess between-study variability along with reasons why they may not be particularly useful. They have not been adapted for MBMA, as far as we know, and we would not recommend their use for these longitudinal models.

Model diagnostics

There are a much greater number of potential diagnostics that can be carried out on a longitudinal model than for a landmark model. Broadly speaking there is residual-based and simulation-based diagnostics. Examples of residual-based diagnostics are weighted residuals over time, weighted residuals vs. predictions, and a histogram of residuals (or other plots that assess the distributional assumption for the residuals).

A common simulation-based diagnostic is the visual predictive check, which assesses how well the model describes the observed data. The mean and quartiles of the simulated data are compared with the mean and quartiles of the observed data. These can be produced using PsN.¹⁸ Normalized prediction distribution errors (NPDEs) are one of the newer simulation-based metrics used to evaluate NLME models and when the model is a good fit of the data, the NPDEs would be expected to be distributed $N(0,1)$.¹⁹ NPDEs can be produced in NONMEM version 7.3 by inserting “NPDE” in the table line of the command file. There is also a library in R (`npde`) that can produce these.

EXAMPLE DATASET: WESTERN ONTARIO AND MCMaster UNIVERSITIES PAIN IN OSTEOARTHRITIS

In order to understand the time-course characteristics of naproxen in OA, internal clinical study reports and publically available literature were searched to find relevant double-blind, randomized, placebo-controlled parallel-group studies. All trials included both naproxen and placebo treatment arms. The endpoint of interest was the Western Ontario and McMaster Universities (WOMAC) pain score, which is the same as the example used in part I of this tutorial, except that now, time-course information has been included. Study

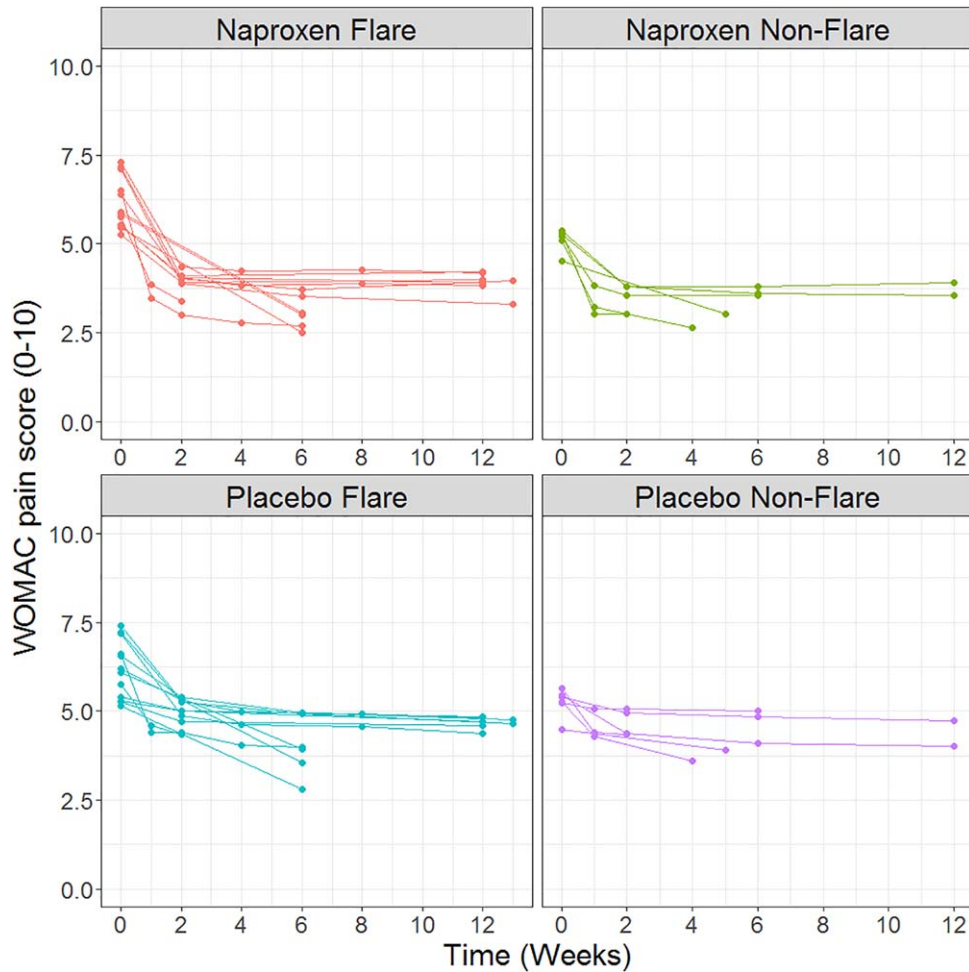


Figure 1 Mean Western Ontario and McMaster Universities (WOMAC) pain over time for naproxen and placebo split by study design.

characteristics are presented in **Supplementary Table S2**. For “flare” trials, subjects were washed out of their pain medications and were required to have a predefined increase in pain (a flare-up) to be eligible for randomization. Of the 18 trials included in this MBMA, 12 were flare designs and 6 were not.

Research questions

As emphasized in the first tutorial, formulating clearly defined research questions will lead to a more focused and efficient piece of work and avoid time-consuming “fishing” exercises. In addition to comparing the longitudinal methodology with landmark estimates, the research questions for this WOMAC pain example are: How does the onset of action and maximal effect compare between naproxen and placebo? Note: a quick onset of action could result in the possibility of running a shorter “first-in-patient” trial. 1) Does a flare design have any impact on treatment effect? Is there an advantage/disadvantage to using such a design in terms of the resulting treatment difference estimate? Note: flare designs are more selective and recruitment could take longer, therefore, if there is no advantage in this design then there is potential to complete the study sooner. Both of

these questions speak to specific model parameters that will be described in the next section.

METHODS

Figure 1 presents the mean WOMAC pain scores across time stratified by treatment (naproxen or placebo) and design (flare or nonflare). This plot shows a quick onset of action (<2 weeks) before reaching and maintaining a maximal effect over time.

There seems to be two key differences between the flare and nonflare trials. The baseline WOMAC score seems to be higher in flare designs, as does the maximal effect relative to baseline, for both the naproxen and placebo treatment groups.

This part of the tutorial will focus on the commonly used E_{\max} model, which will be applied using three different software platforms.²⁰ The results will be compared with the landmark equivalent for specified timepoints (weeks 2, 6, and 12). The E_{\max} models are normally applied to dose-response relationships (due to their pharmacological plausibility) but can also be used to characterize time-course response.

Table 1 Priors for longitudinal model parameters

Parameter	Prior	Comments
E_{onf} (nonflare)	Uniform (0, 10)	Covers the range of the bounded scale
ΔE_{of} (flare)	Uniform (-10, 10)	
$E_{\text{max,pnf}}$ (placebo nonflare)	Normal (0, 10,000)	
$\Delta E_{\text{max,pf}}$ (placebo flare)	Normal (0, 10,000)	
$\Delta E_{\text{max,n}}$ (naproxen)	Normal (0, 10,000)	
$\text{Ln}(ET_{50p})$	Uniform (-10, 10)	
$\text{Ln}(\Delta ET_{50n})$	Normal (0, 10,000)	
τ_1	Half normal (0, 1,000,000)	Can only take positive values
τ_2	Half normal (0, 1,000,000)	Can only take positive values

E_{max} , maximum effect.

Missing SDs were imputed with the approach used in Boucher.¹⁵

The three-parameter E_{max} model below was used to fit the WOMAC pain data (Y_{ijk}) for study i , and treatment arm j at time k .

$$Y_{ijk} = E_0 + \eta_{1i} + \frac{(E_{\text{max}} + \eta_{2i}) * t_{jk}}{ET_{50} + t_{jk}} + \varepsilon_{ijk}, \quad (1)$$

where E_0 is the effect at baseline (time = 0), E_{max} is the maximal effect over time, and ET_{50} is the time to get to 50% of E_{max} . Due to the observed differences between flare and non-flare designs, as discussed above, “flare” was fitted as a structural covariate to both E_0 and E_{max} such that:

$$E_0 = E_{0,\text{nf}} + I_f * E_{0,\text{f}} \quad (2)$$

and

$$E_{\text{max}} = E_{\text{max,pnf}} + I_f * E_{\text{max,pf}} + I_n * E_{\text{max,n}}, \quad (3)$$

where parameters with an additional suffix of “nf” related to non-flare designs and “f” to flare designs. I_f was an indicator variable for design (0 for “nonflare” and 1 for “flare”) and I_n was an indicator variable for treatment (1 for naproxen and 0 for placebo) ET_{50} was parameterized as follows:

$$ET_{50} = ET_{50p} + I_n * ET_{50n} \quad (4)$$

The ET_{50p} was the ET_{50} parameter for placebo with ET_{50n} being the additional ET_{50} for naproxen compared to placebo. ET_{50} was fitted in log space to ensure it was positive.

Research question 1 was addressed by comparing ET_{50p} with $(ET_{50p} + I_n * ET_{50n})$.

For research question 2, a covariate for “flare” design was added to $E_{\text{max,n}}$ such that:

$$E_{\text{max,n}} = E_{\text{max,nf}} + I_f * E_{\text{max,nf}} \quad (5)$$

Correlation was accounted for by fitting random effects to both E_0 and E_{max} , as shown in Eq. 1. These random effects η_1 (E_0) and η_2 (E_{max}) were assumed to be normally distributed both with mean 0 and variances τ_1^2 and τ_2^2 , respectively.

The residuals were assumed to be normally distributed with mean 0 and variance SD_{ijk}^2/n_{ijk} . As the weights were

based on the observed SEs, σ was fixed to 1 in the modeling. However, in R, using the NLME function, it was not possible to fix σ to 1.

Linear models for landmark comparison

A random effects linear model was used, as described in part I of the tutorial, to compare landmark estimates and precision with the corresponding longitudinal model estimates for selected timepoints.

$$\hat{\theta}_i = \theta_i + \varepsilon_i \quad (6)$$

where $\theta_i \sim N(\delta, \tau^2)$.

Bayesian considerations

The priors used for the Bayesian E_{max} model are summarized in **Table 1**. They were all set to be noninformative so that the resulting estimates should be in line with frequentist estimates. Generally, it is recommended that sensitivity analyses with different priors be considered. As this is an illustrative example, just a single set of priors has been used.

Three chains were used with a burn-in of 60,000 samples and then posterior distributions were summarized based on a further 20,000 samples.

Software

The models described above were fitted in NONMEM, BUGS (using R2OpenBUGS), and R (using function NLME).^{21,22} Landmark analyses were carried out in R using metafor.²³ Model diagnostics were produced in R. The results were compared but the main aim was to demonstrate the “how to” for each of these packages. The relative merits of each package will also be discussed in terms of their ease of use and limitations.

Model diagnostics

The following plots were produced, which are designed to be illustrative:

1. Observed versus individual predictions.
2. Weighted residuals (WRES in NONMEM) by time, stratified by treatment and flare design.
3. NPDEs by time, stratified by treatment and flare design.
4. Observed vs. predicted over time for each study and arm.

Table 2 Comparison of parameter estimates for BUGS, NONMEM, and R

Parameter	Estimate (SE)		
	BUGS ^a	NONMEM	R (NLME)
E_0 (nonflare)	5.22 (0.28)	5.20 (0.13)	5.20 (0.27)
ΔE_0 (flare)	0.93 (0.33)	0.96 (0.25)	0.96 (0.32)
$E_{\max p}$ (nonflare)	-1.14 (0.41)	-1.16 (0.24)	-1.15 (0.32)
$\Delta E_{\max p}$ (flare)	-0.86 (0.47)	-0.82 (0.09)	-0.82 (0.39)
$\Delta E_{\max n}$	-0.79 (0.06)	-0.79 (0.09)	-0.79 (0.07)
$\ln(ET_{50p})$	-0.40 (0.17)	-0.37 (0.20)	-0.40 (0.17)
ET_{50p} (weeks)	0.67	0.69	0.67
$\ln(\Delta ET_{50n})$	-1.24 (0.31)	-1.17 (0.20)	-1.20 (0.29)
ET_{50n} (weeks)	0.19	0.21	0.20
τ_1	0.86	0.62	0.62
τ_2	0.71	0.74	0.74

E_{\max} , maximum effect; NLME, nonlinear mixed effect; NONMEM, nonlinear mixed-effect modeling.

^aPosterior mean and SE.

5. Treatment difference (naproxen – placebo): observed vs. predicted over time for each study and arm.

All presented diagnostics were produced using NONMEM version 7.3 model output.

Publication bias

The authors are not aware of standard techniques to assess publication bias for longitudinal models but the following approach was taken for this example: a standard publication bias funnel plot for each timepoint (weeks 1, 2, 6, and 12 when there is more than one study).

RESULTS

Table 2 presents the parameter estimates for the E_{\max} model across the three packages.

The parameter estimates suggest that naproxen has a quicker onset of action than placebo but it should also be noted that the estimates are both below 1 week, which is earlier than any of the postdose observations across the studies.

The flare by treatment covariate was not found to be significant and the estimates in **Table 2** and **Table 3** are based on a model without it.

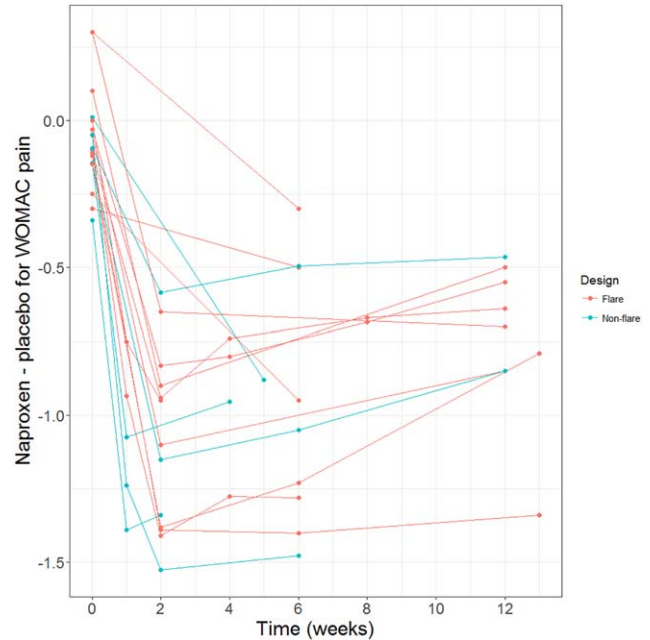


Figure 2 Difference in mean Western Ontario and McMaster Universities (WOMAC) pain over time between treatments (naproxen-placebo).

Table 3 presents longitudinal model estimates for timepoints 2, 6, and 12 weeks and compares them to the landmark estimates from a random effects model (all split by flare and nonflare). Note that the number of trials with 2, 6, and 12-week data reported were 13, 9, and 7, respectively, of 18 in total. The week 12 estimates are less comparable between the landmark and longitudinal models than the earlier timepoints. Both approaches demonstrate that treatment differences between naproxen and placebo decrease over time. This can also be seen in **Figure 2**, which shows the observed difference in means between naproxen and placebo over time, split by “flare.”

The publication bias plots shown in **Figure 3** for weeks 1, 2, 6, and 12, did not show any obvious evidence of publication bias.

Figure 4 presents the selected diagnostic plots that seem to demonstrate a good model fit for both naproxen and placebo, although there seems to be more positive

Table 3 Longitudinal and landmark model estimates of treatment difference at weeks 2, 6, and 12

Week	Estimates for naproxen – placebo (95% confidence/credible interval)			
	BUGS ^a	NONMEM	R (NLME)	Landmark
Nonflare				
2	-0.90 (-1.04, -0.76)	-0.90 (-1.04, -0.76)	-0.90 (-1.02, -0.78)	-1.08 (-1.52,-0.64) (4 trials)
6	-0.84 (-0.94, -0.74)	-0.84 (-0.98, -0.70)	-0.84 (-0.94, -0.74)	-0.95 (-1.51, -0.39) (3 trials)
12	-0.82 (-0.93, -0.71)	-0.82 (-0.98, -0.66)	-0.82 (-0.94, -0.70)	-0.62 (-1.00, -0.24) (2 trials)
Flare				
2	-1.04 (-1.18, -0.92)	-1.03 (-1.19, -0.87)	-1.03 (-1.15, -0.91)	-1.06 (-1.26, -0.86) (9 trials)
6	-0.90 (-0.99, -0.81)	-0.90 (-1.06, -0.74)	-0.89 (-0.97, -0.81)	-0.99 (-1.35, -0.63) (3 trials)
12	-0.85 (-0.95, -0.75)	-0.85 (-1.01, -0.69)	-0.84 (-0.94, -0.74)	-0.67 (-0.87, -0.47) (5 trials)

NLME, nonlinear mixed effect; NONMEM, nonlinear mixed-effect modeling.

^aPosterior mean and 95% credible interval.

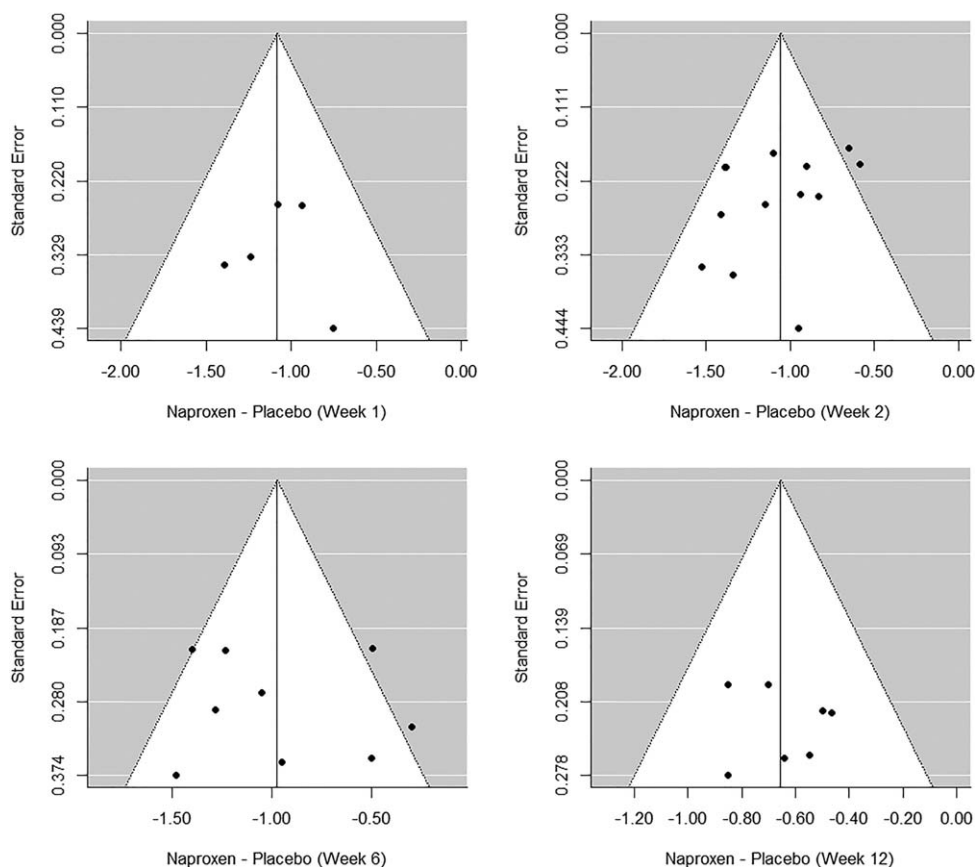


Figure 3 Funnel plots for weeks 1, 2, 6, and 12.

residuals and NPDEs for naproxen at weeks 12 and 13 with the exception of a single negative residual (week 13 of study 18). This observation does not seem to be an obvious outlier from the raw plots, although, where pain response has generally leveled off in most arms, here, it is still decreasing at week 13. The observed vs. predicted plots over time, split by study and arm, can be found in **Supplementary Figure S1**.

There seems to be more of a discrepancy between the predictions and observations for the treatment difference (naproxen – placebo) over time, which can be seen in **Supplementary Figure S2** and this will be further discussed in the next section. The treatment difference seems to be underestimated or overestimated by the model in roughly equal proportions across the 18 trials.

CLOSING REMARKS

This tutorial (part II) acts as both a “points for consideration” article when conducting longitudinal MBMA and also as an example of a common nonlinear model form being applied to such data across multiple platforms. There are many candidate models that could be used to fit such data, exponential being a common alternative, and it is not the intention of the authors to recommend a specific model or

to compare different models. Rather, this is intended as an illustrative example with code and dataset provided to allow readers to either reproduce results or try out alternative model constructs. Model building is an integral aspect of work carried out by clinical pharmacologists and, although one would not expect as many model steps as routinely seen for a patient level pharmacokinetic/pharmacodynamic analysis with potential multiple parameters in a longitudinal model, there is still scope for a variety.

One potential criticism of the models fitted in this tutorial, particularly outside of Clinical Pharmacology, is that baseline (E_0) is modeled with a random effect rather than fixed per study (when it is effectively treated as a nuisance parameter). Similar to modeling placebo response in a dose-response model, it may be desirable to make future predictions using certain criteria regarding baseline (the effect of covariates on E_0 , for example). The extent to which this might be a problem in terms of bias is unclear and it would be a useful avenue of future research.

The first tutorial used approaches that modeled the treatment difference in which the reference arm (e.g., placebo) was treated as a “nuisance parameter.” The modeling approach in this tutorial was taken at the arm level and, in the example, the fit seemed to be very good when assessing each arm within a study. However, diagnostic plots based on the treatment difference were less convincing. An

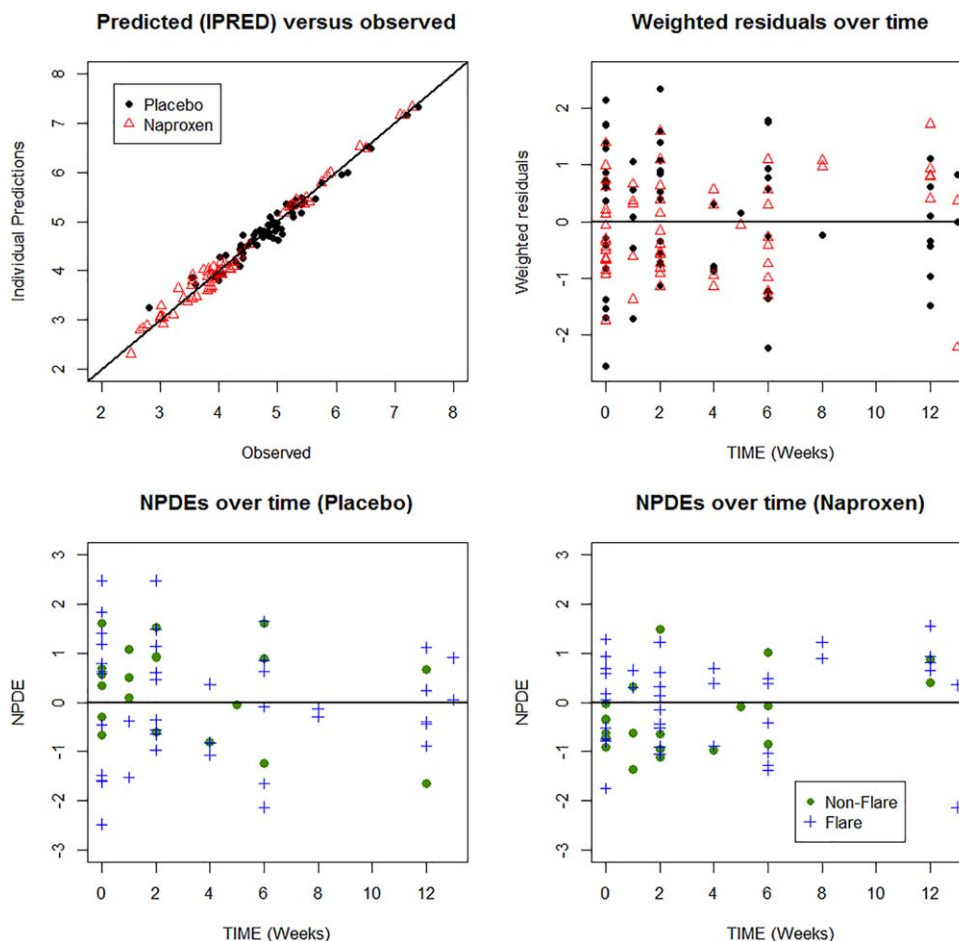


Figure 4 Diagnostic plots for the longitudinal model. IPRED, individual predicted; NPDE, normalized prediction distribution error.

issue with this arm-based approach is that the randomization is not being fully accounted for. This issue has been discussed in relation to traditional meta-analyses and network meta-analyses.^{24,25} An alternative would be to model the treatment differences but it is often of interest to characterize placebo response along-side active treatment response and, hence, it would be useful to look further into the consequences of the arm-based approach and the impact of biases.

For correlation, random effects were used to account for this in a simple way. This method is straightforward to implement in all three packages featured. It is not clear, generally, whether more complex forms of correlation, such as AR(1), would provide much advantage over the simpler forms when an appropriate structural model has been fitted. The AR(1) models can easily be fitted in the R NLME function and for NONMEM version 7.3 onward can also be easily fitted using a small amount of code that can be found in the user manual. To fit AR(1) models in BUGS would require more work as there are no “settings” as such. A comparison of different correlation structures for time-course modeling would be another useful area of research, together with worked examples in different software packages.

The number of potential diagnostics for these types of models, primarily developed for patient level data, can be overwhelming and this tutorial focusses on only a few of the most common ones. Other diagnostics commonly used are visual predictive checks and histograms of random effects (somewhat limited given the typical small number of studies in meta-analyses). Keizer *et al.*²⁶ provide a useful tutorial on a modeling and simulation workbench for NONMEM.

By including only trials with both naproxen and placebo arms, we have only concerned ourselves with direct information. An extension would be to look at all randomized controlled trials in several nonsteroidal anti-inflammatory drugs in OA (e.g., diclofenac and ibuprofen). There might be a need to make comparisons between the active compounds, which then may involve a combination of direct and indirect information. This could be an issue if there is inconsistency between direct and indirect information.²⁷

A future challenge for this type of MBMA is the move away from imputation methods, such as LOCF and BOCF in many new clinical trials. If a substantial proportion of the literature is of this form, it will be more of a challenge to relate historical data to new study readouts or to assess comparative effectiveness.

The model discussed here (E_{\max}) assumes a monotonically increasing/decreasing response. There are examples where an effect reaches a maximum and then starts to go back in the opposite direction (a rebound). This could be due to reasons such as resistance to drugs (e.g., human immunodeficiency virus viral load) or behavioral (weight loss in obesity trials).

This is evident in some published obesity trials, for example, and the models discussed here would need to be expanded to allow for this phenomenon.^{28,29} Pérez-Pitarch *et al.*³⁰ fitted a parameter to account for rebound of virologic response in patients with hepatitis C.

Another potential avenue of research could be multivariate time-course modeling. There are, for example, several correlated key endpoints in OA (WOMAC pain, WOMAC function, and weekly average pain scores). Published articles tend to report a mixture of these endpoints, with often only one of the three having a full time-course plot. Could a combined multivariate model borrow strength across the studies and endpoints to result in better estimates of treatment effect? Similarly, could multivariate models help with the issues of different imputation methods being reported across articles?

The first two parts of this tutorial have covered both time-course and dose-response separately but it may be useful to do a combined dose-response and time-course model. Mandema *et al.*³¹ did this comparing eletriptan and sumatriptan in the treatment of migraine pain. Similarly Checchio *et al.*³² also fitted a joint time-course and dose-response model to several endpoints related to the treatment of psoriasis.

Similarly, there are examples of models that incorporate a longitudinal MBMA as part of a wider pharmacokinetic/pharmacodynamic approach in which joint pharmacokinetic/MBMA models were fitted.^{8,33} Further examples are highlighted in the list of application examples in the **Supplementary Materials**.

The field of MBMA continues to expand and the number of published examples is growing but there is still very little published research around the methodology (Ahn & French⁵ and Mawdsley *et al.*²⁷ being two such examples) and there are many issues that need a deeper investigation. It is this future research that will make MBMA more acceptable to a wider audience, outside the Clinical Pharmacology sphere, such as statisticians, payers, and regulators.

Conflict of Interest. Martin Boucher and Meg Bennetts are employees and shareholders of Pfizer.

1. Boucher, M. & Bennetts, M. The many flavors of model-based meta-analysis: part I – introduction and landmark data. *CPT Pharmacometrics Syst. Pharmacol.* **5**, 54–64 (2016).
2. Wang, Y. *et al.* Short-term efficacy reliably predicts long-term clinical benefit in rheumatoid arthritis clinical trials as demonstrated by model-based meta-analysis. *J. Clin. Pharmacol.* **56**, 835–844 (2016).
3. Ishak, K.J., Platt, R.W., Joseph, L., Hanley, J.A. & Caro, J.J. Meta-analysis of longitudinal studies. *Clin. Trials* **4**, 525–539 (2007).
4. Musekiwa, A., Manda, S.O., Mwambi, H.G. & Chen, D.G. Meta-analysis of effect sizes reported at multiple time points using general linear mixed model. *PLoS One* **11**, e0164898 (2016).
5. Ahn, J.E. & French, J.L. Longitudinal aggregate data model-based meta-analysis with NONMEM: approaches to handling within treatment arm correlation. *J. Pharmacokinetic. Pharmacodyn.* **37**, 179–201 (2010).
6. Jansen, J.P., Vieira, M.C. & Cope, S. Network meta-analysis of longitudinal data using fractional polynomials. *Stat. Med.* **34**, 2294–2311 (2015).
7. Royston, P. & Altman, D.G. Regression using fractional polynomials of continuous covariates: parsimonious parametric modelling. *J. Appl. Stat.* **43**, 429–467 (1994).
8. Luu, K.T., Raber, S.R., Nickens, D.J. & Vicini, P. A model-based meta-analysis of the effect of latanoprost chronotherapy on the circadian intraocular pressure of patients with glaucoma or ocular hypertension. *Clin. Pharmacol. Ther.* **87**, 421–425 (2010).
9. Demin, I., Hamrén, B., Luttringer, O., Pillai, G. & Jung, T. Longitudinal model-based meta-analysis in rheumatoid arthritis: an application toward model-based drug development. *Clin. Pharmacol. Ther.* **92**, 352–359 (2012).
10. Mercier, F., Claret, L., Prins, K. & Bruno, R. A model-based meta-analysis to compare efficacy and tolerability of tramadol and tapentadol for the treatment of chronic non-malignant pain. *Pain Ther.* **3**, 31–44 (2014).
11. Jørgensen, A.W., Lundstrøm, L.H., Wetterlev, J., Astrup, A. & Gotzsche, P.C. Comparison of results from different imputation techniques for missing data from an anti-obesity drug trial. *PLoS One* **9**, e111964 (2014).
12. Rosenstock, J., Hollander, P., Chevalier, S. & Iranmanesh, A.; SERENADE Study Group. SERENADE: the study evaluating rimonabant efficacy in drug-naïve diabetic patients: effects of monotherapy with rimonabant, the first selective CB1 receptor antagonist, on glycemic control, body weight, and lipid profile in drug-naïve type 2 diabetes. *Diabetes Care* **31**, 2169–2176 (2008).
13. Chan, A.W. & Altman, D.G. Identifying outcome reporting bias in randomised trials on PubMed: review of publications and survey of authors. *BMJ* **330**, 753 (2005).
14. Beal, S., Sheiner, L.B., Boeckmann, A. & Bauer, R.J. *NONMEM User's Guides (1989–2009)* (Icon Development Solutions, Ellicott City, MD, 2009).
15. Boucher, M. Imputation of missing variance data using non-linear mixed effects modelling to enable an inverse variance weighted meta-analysis of summary-level longitudinal data: a case study. *Pharm. Stat.* **11**, 318–324 (2012).
16. Wiebe, N., Vandermeer, B., Platt, R.W., Klassen, T.P., Moher, D. & Barrowman, N.J. A systematic review identifies a lack of standardization in methods for handling missing variance data. *J. Clin. Epidemiol.* **59**, 342–353 (2006).
17. Wakefield, J. Ecologic studies revisited. *Annu. Rev. Public Health* **29**, 75–90 (2008).
18. Lindbom, L., Pihlgren, P. & Jonsson, N. PsN-Toolkit – a collection of computer intensive statistical methods for non-linear mixed effect modeling using NONMEM. *Comput. Methods Programs Biomed.* **79**, 241–257 (2005).
19. Brendel, K., Comets, E., Laffont, C., Laveille, C. & Mentré, F. Metrics for external model evaluation with an application to the population pharmacokinetics of gliclazide. *Pharm. Res.* **23**, 2036–2049 (2006).
20. Holford, N.H. & Sheiner, L.B. Understanding the dose-effect relationship: *clinical application of pharmacokinetic-pharmacodynamic models*. *Clin. Pharmacokinetic.* **6**, 429–453 (1981).
21. Lunn, D., Spiegelhalter, D., Thomas, A. & Best, N. The BUGS project: evolution, critique and future directions. *Stat. Med.* **28**, 3049–3067 (2009).
22. R Core Team. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. <<http://www.R-project.org/>> (2013).
23. Viechtbauer, W. Conducting meta-analyses in R with the metafor package. *J. Stat. Softw.* **36**, 1–48 (2010).
24. Senn, S. The many modes of meta. *Drug Inf. J.* **34**, 535–549 (2000).
25. Senn, S., Gavini, F., Magrez, D. & Scheen, A. Issues in performing a network meta-analysis. *Stat. Methods Med. Res.* **22**, 169–189 (2013).
26. Keizer, R.J., Karlsson, M.O. & Hooker, A. Modeling and simulation workbench for NONMEM: tutorial on Pirana, PsN, and Xpose. *CPT Pharmacometrics Syst. Pharmacol.* **2**, e50 (2013).
27. Mawdsley, D., Bennetts, M., Dias, S., Boucher, M. & Welton, N.J. Model-based network meta-analysis: a framework for evidence synthesis of clinical trial data. *CPT Pharmacometrics Syst. Pharmacol.* **5**, 393–401 (2016).
28. Davidson, M.H. *et al.* Weight control and risk factor reduction in obese subjects treated for 2 years with orlistat: a randomized controlled trial. *JAMA* **281**, 235–242 (1999).
29. Smith, S.R. *et al.* Multicenter, placebo-controlled trial of lorcaserin for weight management. *N. Engl. J. Med.* **363**, 245–256 (2010).
30. Pérez-Pitarch, A., Guglieri-López, B., Ferriols-Lisart, R. & Merino-Sanjuán, M. A model-based meta-analysis of sofosbuvir-based treatments in chronic hepatitis C patients. *Int. J. Antimicrob. Agents* **47**, 184–194 (2016).
31. Mandema, J., Cox, E. & Alderman, J. Therapeutic benefit of eletriptan compared to sumatriptan for the acute relief of migraine pain – results of a model-based meta-analysis that accounts for encapsulation. *Cephalalgia* **25**, 715–725 (2005).
32. Checchio, T. *et al.* Quantitative evaluations of time-course and treatment effects of systemic agents for psoriasis: a model-based meta-analysis. *Clin. Pharmacol. Ther.* **102**, 1006–1016 (2017).

33. Gibbs, J.P. *et al.* Quantitative model of the relationship between dipeptidyl peptidase-4 (DPP-4) inhibition and response: meta-analysis of alogliptin, saxagliptin, sitagliptin, and vildagliptin efficacy results. *J. Clin. Pharmacol.* **52**, 1494–1505 (2012).

© 2018 The Authors *CPT: Pharmacometrics & Systems Pharmacology* published by Wiley Periodicals, Inc. on

behalf of American Society for Clinical Pharmacology and Therapeutics. This is an open access article under the terms of the Creative Commons Attribution-NonCommercial License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

Supplementary information accompanies this paper on the *CPT: Pharmacometrics & Systems Pharmacology* website (<http://psp-journal.com>)