Shrinkage Estimation for Dose–Response Modeling in Phase II Trials With Multiple Schedules

Burak Kürsad Günhan, Paul Meyvisch, and Tim Friede

Department of Medical Statistics, University Medical Center Göttingen, Göttingen, Germany; Galapagos NV, Mechelen, Belgium

ABSTRACT

Recently, phase II trials with multiple schedules (frequency of administrations) have become more popular, for instance, in the development of treatments for atopic dermatitis. If the relationship of the dose and response is described by a parametric model, a simplistic approach is to scale doses from different schedules to a common unit and pool all rescaled doses. However, this approach ignores the potential heterogeneity in dose–response curves between schedules. A more reasonable approach is the partial pooling, that is, certain parameters of the dose–response curves are shared, while others are allowed to vary. Rather than using schedule-specific fixed-effects, we propose a Bayesian hierarchical model with random-effects to model the between-schedule heterogeneity with regard to certain parameters. Schedule-specific dose–response relationships can then be estimated using shrinkage estimation. Considering Emax models, the proposed method displayed desirable performance in terms of the mean absolute error and the coverage probabilities for the dose–response curve compared to the complete pooling. Furthermore, it outperformed the partial pooling with schedule-specific fixed-effects by producing lower mean absolute error and shorter credible intervals. The methods are illustrated using simulations and a phase II trial example in atopic dermatitis. A publicly available R package, ModStan, is developed to automate the implementation of the proposed method (https://github.com/gunhanb/ModStan).

1. Introduction

In phase II of any clinical development program, the investigations of the dose–response relationship of a compound is crucial. Usually, there are two main goals of these investigations: (a) establishing a dose–response signal and (b) estimating the dose–response function (Ruberg 1995). In addition to the dose, a treatment plan of a phase II trial includes the schedule (or dose regimen), that is, the frequency of the administration, for instance, a weekly or biweekly schedule. Recently, phase II trials with multiple schedules have become more popular, for instance, in the development of monoclonal antibodies as treatments for a variety of diseases including hypercholesterolaemia (Giugliano et al. 2012) and atopic dermatitis (Thaçi et al. 2016). Eichenfield and Stein Gold (2017) reviewed many therapies for atopic dermatitis which were in phase II or III of clinical development. Multiple schedules were investigated in phase II trials of almost half of the investigated therapies (Eichenfield and Stein Gold 2017). However, standard methods for dose–response estimation cannot account for multiple schedules.

Estimating separate dose–response curves for each schedule by a parametric model is a one way to tackle this problem, that is, full stratification of the dose–response curves. However, this method ignores the information shared between different schedules. Alternatively, one can scale doses from different schedules to a common unit using a reference schedule and then completely pool rescaled doses. The main problem with complete pooling is that it does not take into account the potential heterogeneity between different schedules. A more reasonable approach is the partial pooling, that is, certain parameters of the dose–response curves are shared, while others are allowed to vary. A placebo effect parameter can often be assumed shared between schedules, whereas this may not be true for the ED50 parameter, the dose at which half of the maximum effect is reached. Feller et al. (2017) proposed a partial pooling approach in which unshared parameters are treated as schedule-specific fixed-effects (Möllenhoff, Bretz, and Dette 2020).

We consider trials with (very) few schedules of small to moderate size. Here, borrowing available information is of great interest. Rather than using schedule-specific fixed-effects, we propose a Bayesian hierarchical model with random-effects to model the between-schedule heterogeneity with regard to certain parameters. Schedule-specific parameters can be estimated using shrinkage estimation. The basic idea of the shrinkage estimation is that stratified parameter estimates can be improved by shrinking toward the population mean. It has been shown that the shrinkage estimation improves the estimation accuracy in comparison to estimates obtained by pooling or stratification (Efron and Morris 1975). Shrinkage estimation in the context of clinical trials were investigated by Jones et al. (2011) and Freidlin and Korn (2013), among others. A popular application
is the estimation of the treatment effect in the presence of subgroups, for example, estimating response rate in a phase II trial with multiple patient populations (Neuenschwander et al. 2016). Here, we are interested in parametric dose–response models in the presence of multiple schedules, hence shrinkage estimators of the parameters of a dose–response model, for example, the ED<sub>50</sub> parameter of an Emax model. Shrinkage estimation allows dynamic borrowing (Viele et al. 2014), in which the weights for each schedule depend on the data instead of using fixed weights. Dynamic borrowing results in considerable gain in efficiency, while being a robust method against the heterogeneity between schedules. A theoretical justification for shrinkage can be established through the concept of exchangeability of the parameters between schedules. This means that finding no systematic reason to distinguish schedule-specific parameters, in other words, they are similar, but not identical (Greenland 2000). Usually, the assumption of exchangeability indicates the schedule-specific parameters come from a common distribution with an overall mean. For the ED<sub>50</sub> parameter, we assume the rescaled and log transformed ED<sub>50</sub> parameter estimates (using the corresponding frequency of each schedule) are exchangeable.

In this article, we propose a Bayesian hierarchical model which uses shrinkage estimation for certain parameters of the dose–response model to dynamically borrow strength across schedules in a phase II trial. Another contribution is the introduction of a publicly available R package, ModStan. In Section 2, two phase II trials with multiple schedules for the treatment of atopic dermatitis are described. We introduce the proposed method to analyze phase II trials with multiple schedules in Section 3. We also describe partial pooling with schedule-specific fixed-effects for certain parameters, discuss the choice of priors and implementation of the proposed method. We evaluated the repeated sampling properties of different methods in a simulation study in Section 4. One of the illustrative applications is revisited to display the proposed method and compare it to the alternatives in Section 5. We close with some conclusions and outlook.

### 2. Illustrative Applications

Atopic dermatitis, the most common form of eczema, is a chronic inflammatory disease that is characterized by skin rash and itching (Mayo Clinic 2018). Hypercholesterolemia is the presence of high levels of cholesterol in the blood and it can cause serious complications including heart attack and stroke (Mayo Clinic 2019). Recently, there is an increasing number of clinical trials investigating novel systemic agents for the treatment of atopic dermatitis and hypercholesterolemia (Giugliano et al. 2012; Alexander et al. 2019). We consider phase II trials of two human monoclonal antibodies, MOR106 and bococizumab, for the treatment of atopic dermatitis and hypercholesterolemia, respectively. Designs of two trials are listed in Table 1. We consider these two trials, since they were both designed to investigate multiple schedules. In MOR106 trial, the primary endpoint is the percentage change from baseline in Eczema Area and Severity Index (EASI) score at Day 85. The EASI scoring system is used to grade the severity of the signs of eczema. EASI scores take values between 0 and 72 and higher EASI score means higher severity. In the bococizumab trial, the primary outcome is the change from baseline in low-density lipoprotein cholesterol (LDL-C) at Day 85 (Ballantyne et al. 2014). These trials are used to motivate our simulation studies in Section 4. In October 2019, the MOR106 trial was terminated due to lack of efficacy in the interim analysis (MorphoSys AG 2019).

### 3. Statistical Methods

Assume that a response <i>y</i><sub><i>ijk</i></sub> (an efficacy or a safety outcome) is observed for schedule <i>i</i>, dose <i>j</i>, and patient <i>k</i>. Following Feller et al. (2017), we assume a normal likelihood for a continuous outcome:

\[
y_{ijk} \sim N(f(d_{ij}, \theta(i)), \sigma_i^2),
\]

where <i>θ</i><sub><i>i</i></sub> refers to the model parameters and <i>σ</i><sub><i>i</i></sub> to the error standard deviation. The <i>f(d_{ij}, θ(i))</i> represents the functional form of the dose–response relationship for schedule <i>i</i>. Other outcome types, for instance, dichotomous or count, can be modeled by specifying appropriate likelihood (e.g., Binomial or Poisson) and the link function (e.g., logit or log transformation).

There are a number of candidate models for the functional form including the popular Emax model (Thomas, Sweeney, and Somayaji 2014), that is

\[
f(d_{ij}, θ(i)) = E_0(i) + E_{max(i)} \frac{d_{ij}}{ED_{50(i)} + d_{ij}},
\]

where <i>E<sub>0</sub> (i)</i> is the placebo response and <i>E<sub>max</sub> (i)</i> is the maximum effect attributable to the drug. The ED<sub>50</sub> parameter represents the dose at which half of the maximum effect is reached. In the article, we exclusively use the Emax model, see Bretz, Pinheiro, and Branson (2005) for different candidate models.

As explained in the introduction, one way of modeling the dose–response curves is to treat all model parameters as schedule-specific fixed-effects. However, such an analysis is not the most efficient, when certain aspects of the dose–response curves in different schedules are similar. Alternatively, one can consider a complete pooled analysis in which all model parameters from different schedules are assumed to be the same. This approach is also problematic, since it ignores the potential heterogeneity between dose–response curves of different schedules. A more reasonable approach is the partial pooling (Feller et al. 2017), which strikes a balance between efficiency and robustness. It is often reasonable to assume that placebo

### Table 1. Designs of two phase II trials in atopic dermatitis (MOR106) and in hypercholesterolemia (bococizumab) involving different schedules.

| Arm | Schedule | Dose (mg/kg) | Arm | Schedule | Dose (mg) |
|-----|----------|--------------|-----|----------|-----------|
| 1   | Biweekly | 0            | 1   | Biweekly | 0         |
| 2   | Biweekly | 1            | 2   | Biweekly | 50        |
| 3   | Biweekly | 3            | 3   | Biweekly | 100       |
| 4   | Biweekly | 10           | 4   | Biweekly | 150       |
| 5   | Monthly  | 1            | 5   | Monthly  | 0         |
| 6   | Monthly  | 3            | 6   | Monthly  | 200       |
| 7   | –        | –            | 7   | –        | 300       |

Note: Clinicaltrial.gov identifiers are displayed for two trials.
Effect $E_i^{(1)}$ is the same for different schedules, that is, $E_i^{(1)} = E_i^{(2)} = \cdots$. This is especially the case, when there is only one placebo arm investigated in the trial as in the MOR106 trial described in Section 2. However, if different schedules were investigated in different studies, it may not be appropriate to assume a common $E_i^{(0)}$ parameter. In some situations, it might also make sense to assume that the maximum efficacy for high doses is same, $E_{\text{max}}^{(1)} = E_{\text{max}}^{(2)} = \cdots$. However, it might not be reasonable to assume the dose providing half of the maximum efficacy is the same for different schedules, that is, $ED_{50}^{(1)} \neq ED_{50}^{(2)} \neq \cdots$. Feller et al. (2017) suggested to treat the unshared parameters, for example, $E_{\text{max}}^{(1)}$ and/or $ED_{50}^{(1)}$ as schedule-specific fixed-effects in the partial pooling approach.

### 3.1. Proposed Method: Partial Pooling With Random-Effects

Rather than using schedule-specific fixed-effects, we propose a Bayesian hierarchical model with random-effects to model the between-schedule heterogeneity with regard to certain parameters in the partial pooling approach. In other words, we suggest partial pooling with schedule-specific random-effects for certain parameters of the dose–response model. To be concrete, assume that we want to obtain schedule-specific random-effects for $ED_{50}^{(i)}$. In the complete pooling, we scale doses to a common unit using a reference schedule $i_{\text{ref}}$. Hence, the rescaled doses are given by $d_i^{(1)} = d_i^{(2)} = \cdots$. The frequency of administration of the reference schedule $i_{\text{ref}}$ and the schedule $i$, respectively. The $ED_{50}^{(i)}$ is modeled on the log-scale, since a dose cannot be negative. We assume that the schedule-specific $ED_{50}^{(i)}$ estimates are exchangeable with

$$
\log(ED_{50}^{(i)}) \sim N(\mu_{\text{ED}_{50}}, \tau^2_{\text{ED}_{50}}),
$$

where $\mu_{\text{ED}_{50}}$ is the overall mean and $\tau_{\text{ED}_{50}}$ is the between-schedule heterogeneity in $\log(ED_{50}^{(i)})$. Our main interest is in the schedule-specific estimates, $ED_{50}^{(i)}$. If the heterogeneity $\tau_{ED_{50}}$ is zero, then the model reduces to a model assuming shared $ED_{50}^{(i)}$ parameters. Note that the results are invariant to the choice of the reference schedule. Furthermore, similar to the $ED_{50}^{(i)}$ parameter, shrinkage estimates of $E_{\text{max}}$ parameter can be obtained. There is no need to use the log transformation for the $E_{\text{max}}$ parameter. Treating $E_{\text{max}}$ and/or $ED_{50}^{(i)}$ parameters differently, assuming either one or both of them shared between schedules or assuming schedule-specific random-effects, results in a variety of alternative models.

Instead of rescaling doses to a common unit, one can rescale the $ED_{50}^{(i)}$ parameters, that is, $ED_{50}^{(i)} = ED_{50}^{(i)}/f_i^{(1)/f_i^{(ref)}}$. Then, the log transformation of the rescaled $ED_{50}^{(i)}$ parameters are assumed to be exchangeable as in (3). This parameterization is equivalent to the parameterization in which doses are rescaled. Hereafter, when we use the partial pooling with schedule-specific random-effects for the $ED_{50}^{(i)}$, we mean the parameterization in which $ED_{50}^{(i)}$ parameters are rescaled.

Complete pooling and partial pooling approaches can be fitted using likelihood estimation. For example, Möllhoff, Bretz, and Dette (2020) demonstrated the likelihood implementation of the partial pooling with schedule-specific fixed-effects for $ED_{50}^{(i)}$ using constrained nonlinear optimization via R package.
MOR106 trial, the primary outcome is the percentage change from baseline in EASI score, respectively. For instance, if we consider that a difference of approximately 40% between the 97.5% and the 2.5% point of the distribution of \( E_{\text{max}} \) values is large enough, we choose a half-normal prior with scale 10 (Röver et al. 2020).

The parameter \( ED_{50} \) is different from \( E_0 \) and \( E_{\text{max}} \) in the sense that it is the only parameter that enters the model non-linearly. In the frequentist framework, it is a common practice to impose bounds (lower and upper bounds) on the space for \( ED_{50} \), since the maximum likelihood estimator (MLE) often does not converge (Bornkamp 2014). However, the estimate will often exactly equal to the specified upper bound, which is unacceptable. In a Bayesian framework, simple prior choice for the \( ED_{50} \) are uniform distributions with finite bounds. However, uniform prior distributions on \( ED_{50} \) are problematic, since they strongly depend on the parameterization: One may end up with completely different implied prior distributions for the dose–response curve. A better prior for \( ED_{50} \) is the Jeffreys’ prior, which is invariant to parameterization. It is defined as \( p(\theta) \propto \sqrt{[\|\theta\|]}/\sqrt{[\|\theta\|]} \) where \( [\|\theta\|] \) is the Fisher information, and \( \theta \) is the vector of proportion of patients allocated at dose \( d \).

Hence, Jeffreys’ prior depends on the observed design \((x, w)\). One cannot state the Jeffreys’ prior before data collection, which is crucial in many applications, for example, in the presence of missing data or two stage designs.

Bornkamp (2012) introduced the functional uniform prior which is a modified version of the Jeffreys’ prior. Functional uniform priors are uniformly distributed on the potential different shapes of the underlying nonlinear model function. These priors are also invariant with respect to parameterization of the model function and typically result in rather nonuniform prior distributions on the parameter scale. Instead of the actual observed design, functional uniform priors are calculated using a grid of doses as \( x \) and equal weights for \( w \). More specifically, say, the gradient function of the Emax model is given by \( J_x(\theta) = (1/x(x + ED_{50}), -x/ED_{max}(x + ED_{50})^2) \). Let \( x \) be a grid of doses and \( F(\theta) \) be the matrix with \( J_x(\theta), x \) in the rows. Then, the functional uniform prior is proportional to \( \sqrt{[Z(\theta)]} \) where \( Z(\theta) = \mathcal{F}(\theta) F(\theta) \) (see Bornkamp 2014 for more detailed explanations). An approximation of the functional uniform prior for \( ED_{50} \) is given as the log-normal distribution with mean \(-2.5\) and standard deviation \( 1.8 \), when the \( ED_{50} \) is rescaled with the maximum available dose \( D \), that is, \( ED_{50}/D \) (Bornkamp 2014). For the simulations and the application, we used the approximation of the functional uniform prior, since it is computationally cheaper. In all models, we use the bounds \([0, 1.5 \times D]\) for the space of \( ED_{50} \) (or \( M_{ED_{50}} \)) parameter.

For the Bayesian implementation of the partial pooling assuming fixed-effects for the \( ED_{50} \), we specify vague priors for \( E_0, E_{\text{max}} \) and \( \sigma \), and functional uniform priors for \( ED_{50} \). The difference between partial pooling with fixed-effects and partial pooling with random-effects is that in the former, independent functional uniform priors are used for each \( ED_{50} \) parameter. However, in the latter, \( ED_{50}^{(i)} \) (rescaled \( ED_{50}^{(i)} \)) are assumed to come from a normal distribution with an overall mean \( \mu_{ED_{50}} \) and standard deviation \( \tau_{ED_{50}} \). A functional uniform prior is used for the \( \mu_{ED_{50}} \). Also, a half-normal prior with scale 1 is used for \( \tau_{ED_{50}} \). In other words, exchangeability is assumed to estimate schedule-specific \( ED_{50}^{(i)} \) parameters in the partial pooling with random-effects. In contrast, there is no underlying exchangeability assumption when estimating schedule-specific \( ED_{50}^{(i)} \) parameters in the partial pooling with fixed-effects.

### 3.3. Implementation of the Proposed Method

In a Bayesian framework, we fitted the described statistical models using the probabilistic programming language Stan which employs a modern Markov chain Monte Carlo (MCMC) algorithm, namely, Hamiltonian Monte Carlo with the No-U-Turn Sampler (Carpenter et al. 2017). The parameterization used for the statistical model influences the MCMC performance. A centered parameterization such as (3) may cause some computational difficulties such as difficulty in convergence in the presence of data sparsity such as meta-analysis of few studies (Betancourt and Girolami 2015) or dose–response modeling of phase II trials with few schedules. An alternative parameterization, that is, a noncentered parameterization, overcomes these computational difficulties. To be more precise, by the reparameterization of the location and scale parameters, (3) becomes \( \log(ED_{50}^{(i)}) = \mu_{ED_{50}} + \tau_{ED_{50}} \) where \( \tau_{ED_{50}} \sim \mathcal{N}(0,1) \) (Günhan, Röver, and Friede 2020). The Stan code defining the partial pooling with schedule-specific random-effects for \( ED_{50}^{(i)} \) is shown in Listing 1.

To facilitate the implementation of our proposed method for the practitioners, we have developed an R package, which is available as an online supplementary document. ModStan is a purpose-build package defined on top of the rstan, the R interface for Stan. We show how to install and use ModStan in Appendix A. The necessary code to reproduce the simulations and the bococizumab application is included to the online supplementary materials.

### 4. Simulation Study

To evaluate properties such as the mean absolute error of the dose–response function obtained by the proposed method and compare it with some alternative methods, a simulation study was conducted.

#### 4.1. Simulation Settings and Implementation

The scenarios considered are motivated by the MOR106 and bococizumab trials described in Section 2. Each generated trial consists of seven arms: one placebo arm and 1, 3, and 10 mg/kg for both biweekly and monthly schedules. The primary outcome is the percentage change from baseline in EASI score. Hence, the datasets are generated under the assumption of normally distributed outcomes, specifically (1). The underlying dose–response function is assumed to be an Emax model, that is, (2).

True values for \( E_0, E_{\text{max}} \), and \( \sigma \) are taken as \(-20\%\), \(-60\%\), and \(35\%\) for both schedules, respectively. Furthermore, \( ED_{50}^{\text{monthly}} \) is assumed to be 2 mg/kg. A total of 27 scenarios are obtained by varying the \( ED_{50}^{\text{monthly}} \) (\( ED_{50}^{\text{monthly}} \in \{1, 2, 3, 5, 6, 10\} \) (mg/kg))) and sample sizes of each arm \( (N \in \{30, 45, 60\}) \). \( ED_{50}^{\text{monthly}} \) values are chosen to investigate the influence of the
data {
  int<lower=1> N_obs; // num of observations
  int<lower=1> N_schedule; // num of schedules
  int<lower=1> N_pred; // num of predicted doses
  real resp[N_obs]; // responses
  real<lower=0> dose[N_obs]; // doses
  int schedule[N_obs]; // schedule indicator
  real<lower=0> freq[N_obs]; // frequency of administration (hrs)
}

parameters {
  real E0; // placebo effect (shared)
  real Emax; // Emax parameter (shared)
  real log_ED50_raw[N_schedule]; // rescaled log(ED50) parameters
  real<lower=0> sigma; // standard deviation for errors
  real<lower=0, upper=1.5> mu_ED50_raw; // mean of log(ED50) random-effects
  real<lower=0> tau_ED50; // between-schedule heterogeneity
}

transformed parameters{
  real mu_ED50;
  real log_ED50[N_schedule];
  real<lower=0> ED50[N_schedule];
  vector[N_obs] resp_hat;
  mu_ED50 = log(mu_ED50_raw * max(dose));
  for(i in 1:N_schedule)
    log_ED50[i] = mu_ED50 + log_ED50_raw[i] * tau_ED50;
  // Taking exponentials and rescaling ED50 parameters
  for(i in 1:N_schedule)
    ED50[i] = exp(log_ED50[i]) * (freq[i] / freq_ref);
  // Dose-response: Emax model
  for(i in 1:N_obs)
    resp_hat[i] = E0 + (Emax * dose[i]) / (ED50[schedule[i]] + dose[i]);
}

model {
  log_ED50_raw ~ normal(0, 1); // implies log(ED50) ~ normal(mu_ED50, tau_ED50)
  // likelihood
  resp ~ normal(resp_hat, sigma);
  // prior distributions
  sigma ~ normal(0, 100);
  E0 ~ normal(0, 100);
  Emax ~ normal(0, 100);
  // approximation to the functional uniform prior
  mu_ED50_raw ~ lognormal(-2.5, 1.8);
  tau_ED50 ~ normal(0, 1);
}

Listing 1. Stan code defining the partial pooling with schedule-specific random-effects for ED$_{50}$ parameter. The parameters $E_0$, $E_{\text{max}}$, and $\sigma$ are assumed to be shared between schedules.

difference between true values of ED$_{50}^{\text{biweekly}}$ and ED$_{50}^{\text{monthly}}$ on the performance. Note that the ED$_{50}^{\text{monthly}}$ values of 1 and 10 mg/kg may be regarded as extreme scenarios. However, we included them to show the trends. Figure 1 displays different dose–response curves for the monthly schedule investigated in the simulations. When ED$_{50}^{\text{monthly}}$ corresponds to 4 mg/kg, there is no heterogeneity in ED$_{50}$ parameters between schedules. This is because if we rescale ED$_{50}^{\text{monthly}}$ to transform on the biweekly scale (simply dividing by two), we obtain 2 mg/kg, which is the true value of ED$_{50}^{\text{biweekly}}$. Accordingly, when the true value of ED$_{50}^{\text{monthly}}$ deviates from 4 mg/kg, the heterogeneity between schedules in ED$_{50}$ increases. The simulations were carried out with 1000 replications per scenario.

In the proposed method, we assume that $E_0^{(i)}$, $E_{\text{max}}^{(i)}$, and $\sigma_i$ are shared between schedules, while ED$_{50}^{(i)}$ are assumed to be schedule-specific random-effects. In other words, the proposed method corresponds to the partial pooling with schedule-specific random-effects for ED$_{50}^{(i)}$ (“PP-RE”). As a comparator, we use the model in which ED$_{50}^{(i)}$ are assumed to be schedule-specific fixed-effects, while other parameters are shared (“PP-FE”). Both partial pooling approaches (PP-RE and PP-FE) are fitted via a Bayesian approach. We also consider the complete pooling method via a frequentist and a Bayesian approach (“CP (Frequentist)” and “CP (Bayesian)”). For the partial pooling with schedule-specific random-effects, we used the biweekly schedule as the reference schedule to rescale the ED$_{50}^{(i)}$ parameters. To implement the complete pooling approaches, all doses should be rescaled into the same schedule. For this purpose, we transform the doses from the monthly schedule into the biweekly schedule. Accordingly, the new set of doses becomes {0, 0.5, 1, 1.5, 3, 10 (mg/kg)} for complete pooling approaches.

The complete pooling (Frequentist) is fitted using fitMod function from the DoseFinding (Bornkamp, Pinheiro, and Bretz 2018) R package. All Bayesian methods are fitted using Stan and the prior distributions from Section 3.2 are used. More
B.K. GÜNHAN, P. MEYVISCH, AND T. FRIEDE

for the dose–response function (\( f \)) in the simulation study. Different curves are generated by varying \( ED_{50}^{\text{monthly}} \) parameter value.

specifically, we use vague priors, \( \mathcal{N}(0,100^2) \), for the parameters \( E_0 \) and \( E_{\text{max}} \), and a half-normal prior with scale 100 for \( \sigma \), \( \mathcal{HN}(100) \) in the complete pooling and both partial pooling approaches. The complete pooling include only one \( ED_{50} \) parameter, which is for the biweekly schedule. A functional uniform prior is used for the \( ED_{50} \) as described in Section 3.2. The partial pooling with fixed-effects for \( ED_{50}^{(i)} \) include two \( ED_{50}^{(i)} \) parameters to estimate, \( ED_{50}^{\text{biweekly}} \) and \( ED_{50}^{\text{monthly}} \). Here, also functional uniform priors are assumed for both \( ED_{50}^{(i)} \) parameters. In the partial pooling with random-effects for \( ED_{50}^{(i)} \), a functional uniform prior is specified for \( \mu_{ED_{50}} \) and a half-normal prior with scale of 1 is specified for the standard deviation \( \tau_{ED_{50}} \). The \( ED_{50} \) and \( \mu_{ED_{50}} \) parameters are assumed to be within the bounds of \([0,1.5 \cdot 10] \) to ensure identifiability for all methods.

Three MCMC chains were run in parallel for a total of 4000 iterations including 2000 iterations of burn-in. Convergence diagnostics are evaluated in some replications, these MCMC settings are chosen accordingly. Also, we calculated the Gelman-Rubin Rhat statistics and number of effective sample sizes as reported by rstan (Stan Development Team 2018). The Rhat values and number of effective sample sizes obtained by all three methods are displayed in histograms, which can be found in the online supplementary materials (Figure S1). Rhat values closer to 1 and higher numbers of effective sample sizes are more desirable. The 97.5% quantile of the Rhat values is 1.1 and 2.5% quantile of the number of effective sample sizes is 1359. Thus, we concluded that the MCMC settings used for simulations are adequate for reliable posterior calculations.

### 4.2. Simulation Results

For each simulation run, we calculated point estimates (\( \hat{f} \)) for the dose–response function (\( f \)) (the pointwise posterior median or the maximum likelihood estimate) at some prespecified dose levels. For this purpose, ten dose levels are chosen between 0 and 10 mg/kg equidistantly, namely \( \text{dose} = D = \{0.00,1.11,\ldots,10.00\} \). Additionally, interval estimates (95% confidence interval or 95% equi-tailed credible intervals) are derived across all doses (\( \text{dose} \)). These computations are done for the dose–response function of the biweekly schedule. The following three performance measures are calculated:

- **MAE**: Mean absolute error for the dose–response function, \( 1/10 \sum_{\text{dose} \in D} |f(\text{dose}) − \hat{f}(\text{dose})| \) across all doses where \( D = \{0.00,1.11,\ldots,10.00\} \).
- **Coverage probability**: Mean coverage probability of the interval estimates evaluated across all doses.
- **Mean length**: Mean length of the interval estimates across all doses.

Lower MAE for the point estimates, shorter interval estimates, and coverage probability of at least 95% for the interval estimates are desirable. The MAE obtained by the four methods is displayed in the first row of Figure 2. Different columns of Figure 2 correspond to different sample sizes \( N \) which are investigated in the simulations. Across different sample sizes, the relative performances of the four methods remain similar. The scenario of \( ED_{50}^{\text{monthly}} = 4 \) corresponds to the scenario without heterogeneity in the rescaled \( ED_{50}^{(i)} \) between biweekly and monthly schedules, which is shown by a vertical dashed line. The heterogeneity increases, when \( ED_{50}^{\text{monthly}} \) deviates from 4. Both complete pooling approaches display better performance than both partial pooling approaches in terms of the MAE, when the \( ED_{50}^{\text{monthly}} \) is 4. However, the partial pooling approaches result in more robust performance across \( ED_{50}^{\text{monthly}} \) values in comparison to the pooling approaches. The partial pooling with random-effects uses the prior \( \mathcal{HN}(1) \) for the heterogeneity parameter \( \tau_{ED_{50}} \). If we increase the value of the prior standard deviation (i.e., 1), then the performance of the partial pooling with random-effects will get closer to the partial pooling with fixed-effects. Similarly, if we assume that \( \tau_{ED_{50}} \) equals to zero, the partial pooling with random-effects reduces to, effectively, the complete pooling (Bayesian). The partial pooling with random-effects yields better performance than the partial pooling with fixed-effects in terms of the MAE across different \( ED_{50}^{\text{monthly}} \) values and sample sizes except the most extreme scenarios, namely \( ED_{50}^{\text{monthly}} = 1 \) or 10. Note that the main difference between the complete pooling (Bayesian) and complete pooling (Frequentist) is that in the former, functional uniform priors are used for \( ED_{50}^{(i)} \) parameters. The small discrepancy between the MAE obtained by the complete pooling (Bayesian) and the complete pooling (Frequentist) can be explained by this difference. Furthermore, when the sample sizes increase, the MAE decreases in the four methods as expected.

Figure 2 also shows coverage probabilities of the interval estimates obtained by the four methods. The complete pooling approaches result in a concave shape and display unacceptably low coverage when \( ED_{50}^{\text{monthly}} \) deviates from 4. This undesirable performance of the complete pooling approaches is more pronounced, when the sample size increases. As in the MAE,
Figure 2. Simulation results for different sample sizes \( N \) per arm. The mean absolute error (MAE) and coverage probabilities for the dose–response curve obtained by four methods with different sample sizes. Four methods include complete pooling approaches using frequentist, CP (Frequentist), and Bayesian methods, CP (Bayesian), and partial pooling approaches using schedule-specific fixed-effects (PP-FE) and schedule-specific random-effects (PP-RE) for \( ED_{50}^{(i)} \). The vertical dashed line indicates the scenario without heterogeneity in the rescaled \( ED_{50}^{(i)} \) between biweekly and monthly schedules. \( ED_{50}^{\text{biweekly}} \) is assumed to be 2 mg/kg.

Both partial pooling approaches show more robust performance in terms of the coverage probabilities in comparison to the complete pooling approaches. The partial pooling with random-effects yields superior performance in terms of the coverage probability compared to the partial pooling with fixed-effects across different \( ED_{50}^{\text{monthly}} \) values and sample sizes except when \( ED_{50}^{\text{monthly}} = 1 \). Figure 3 illustrates the ratios of lengths of credible intervals for the dose–response functions obtained by the partial pooling approaches. The denominator of the ratio is the length of the credible interval obtained by the partial pooling with random-effects. The partial pooling with random-effects results in slightly shorter credible intervals, while it produces slightly higher coverage probability compared to the partial pooling with fixed-effects in most of the scenarios.

We also examine the performance measures in different regions of the dose–response curve instead of taking the mean across all doses of the grid \( \{0.00, 1.11, \ldots, 10.00\} \). A boxplot showing the variability of mean absolute errors and mean coverage probabilities based on different regions of the dose–response curve are displayed in the online supplementary document (Plot S2). The boxplot shows that across all scenarios, there is less variability in the mean absolute errors and mean coverage probabilities from different dose regions compared to the complete pooling approaches. When the \( ED_{50}^{\text{monthly}} \) is 1, 2, 6, or 10 mg/kg, the complete pooling methods yield considerable variability in terms of mean absolute errors and mean coverage probabilities from different dose regions of the dose–response curve.

To examine the influence of the potential heterogeneity in \( E_{\text{max}}^{(i)} \) between schedules, we conducted additional simulations. True values for \( E_{\text{max}}^{(i)} \) and \( \sigma_i \) are taken as \(-20\%\) and \(35\%\) for
both schedules, respectively. The $ED_{50}^{\text{biweekly}}$ and $ED_{50}^{\text{monthly}}$ are assumed to be 2 and 4 mg/kg, respectively. This corresponds to assuming no heterogeneity in $ED_{50}^{i}$ parameters between schedules, since we focused on $E_{\text{max}}^{(i)}$ in these simulations. The $E_{\text{max}}^{(i)}$ is assumed to be $-60\%$. Sample size for each arm is 45. Three scenarios are generated by varying $E_{\text{max}}^{\text{monthly}}$ values ($E_{\text{max}}^{\text{monthly}} \in \{-70\%, -60\%, -50\%\}$). Notice that when $E_{\text{max}}^{\text{monthly}} = -60\%$, there is no heterogeneity in $E_{\text{max}}^{(i)}$ values. In the partial pooling with fixed-effects, both $E_{\text{max}}^{(i)}$ and $ED_{50}^{(i)}$ parameters are treated as schedule-specific fixed-effects. We assume $\mathcal{N}(0, 100^2)$ as prior for each $E_{\text{max}}^{(i)}$ parameter and functional uniform priors are used for each $ED_{50}^{(i)}$ parameter. In the partial pooling with random-effects, both parameters $E_{\text{max}}$ and $ED_{50}$ are assumed to be schedule-specific random-effects. To cover plausible $E_{\text{max}}$ values, we use half-normal priors with the scale 10, $\mathcal{H}N(10)$. We use $\mathcal{N}(0, 100^2)$ as prior for $\mu_{E_{\text{max}}}^{(i)}$ parameter. Moreover, a functional uniform prior is specified for $E_{\text{max}}$ and a half-normal prior with scale of 1 is specified for the standard deviation $r_{ED_{50}}$ parameter as in Section 4.2. The simulation results are listed in Table 3. In the scenario of $E_{\text{max}}^{\text{monthly}} = -60\%$, the complete pooling approaches result in lower MAE in comparison to the partial pooling approaches, while reaching the coverage probability of 95% for the confidence intervals. However, in other scenarios, complete pooling approaches yield worse performance in terms of the MAE and coverage probabilities compared to the partial pooling approaches. The partial pooling with random-effects results in smaller MAE and the shorter credible intervals compared to the partial pooling with fixed-effects in all three scenarios.

When we take into account all simulation results, the partial pooling approaches are more robust in terms of the MAE and the coverage probabilities across scenarios compared to the complete pooling approaches. The partial pooling with random-effects yields better performance than the partial pooling with fixed-effects in terms of the MAE and the mean length of the credible intervals with the exception of highly heterogeneous scenarios.

5. Revisiting the Bococizumab Trial

We return to the bococizumab trial which was described in Section 2. The least square means and standard errors for different arms of the trial are listed in Table 4 as reported in Pfizer (2017). The dataset is available from the clinvnDR R package (Thomas and Wu 2020). In total, 323 patients completed the trial. To compare dose–response curves of two schedules, the doses can be transformed into the biweekly scale, thus 200 and 300 mg monthly doses become 100 and 150 mg biweekly, respectively. The least square means and 95% confidence intervals of the rescaled doses are displayed in Figure 4(A). Least square means reflect some heterogeneity in dose–response curves between schedules.

We analyzed the dataset assuming normal distribution for least square means with the given standard errors as reported in ClinicalTrials.gov (Pfizer 2017). Note that this is different

### Table 3. Simulation results for varying $E_{\text{max}}^{\text{monthly}}$ scenarios.

| Schedule | Dose (mg/m$^2$) | Sample size | LS mean | Standard error |
|----------|----------------|-------------|---------|----------------|
| 1        | Biweekly       | 47          | -2.6    | 3.9            |
| 2        | Biweekly       | 44          | -36.9   | 4.7            |
| 3        | Biweekly       | 42          | -47.6   | 3.9            |
| 4        | Biweekly       | 46          | -56.0   | 4.0            |
| 5        | Monthly        | 46          | 4.6     | 4.6            |
| 6        | Monthly        | 48          | -23.0   | 3.9            |
| 7        | Monthly        | 50          | -40.3   | 4.5            |

### Table 4. The bococizumab trial: Sample sizes, least square (LS) means, and standard errors for each arm in the trial.

| Arm     | Schedule | Dose (mg/m$^2$) | Sample size | LS mean | Standard error |
|---------|----------|----------------|-------------|---------|----------------|
| 1       | Biweekly | 47            | -2.6        | 3.9     |                |
| 2       | Biweekly | 44            | -36.9       | 4.7     |                |
| 3       | Biweekly | 42            | -47.6       | 3.9     |                |
| 4       | Biweekly | 46            | -56.0       | 4.0     |                |
| 5       | Monthly  | 46            | 4.6         | 4.6     |                |
| 6       | Monthly  | 48            | -23.0       | 3.9     |                |
| 7       | Monthly  | 50            | -40.3       | 4.5     |                |
than assuming normality for the observations as described in (1). This will show that the proposed method also works with weaker assumption, as we only use an arm-level data instead of an observation-level data. Five different models were fitted in a Bayesian framework. We compare them via the approximate leave-one-out cross-validation information criteria (LOO-IC) (Vehtari, Gelman, and Gabry 2017). Note that LOO-IC has the same purpose as the Akaike information criteria (AIC) used in the frequentist framework and similar to the AIC, the lower value indicates the better model. All models assume an Emax model for the dose–response relationship. The model descriptions are listed in Table 5. These five models and the prior specifications were described in Section 4. For the prior specification of the $\tau_{\text{Emax}}$, we use a half-normal prior with the scale 20 $\mathcal{HN}(20)$ instead of $\mathcal{HN}(10)$ as in the simulations to be more conservative. This is because the outcome used in the bococizumab trial is not percentage change from baseline as in the MOR106 trial, but change from baseline in LDL-C. Model 1 corresponds to the complete pooling. In Models 2–5, the $E_{0}^{(i)}$ are assumed to be shared between schedules, while $ED_{50}^{(i)}$ and $E_{\text{max}}^{(i)}$ are treated differently in each model. Hence, Models 2–5 are partial pooling approaches. In Models 2 and 3, $E_{\text{max}}^{(i)}$ are assumed to be shared between schedules. Model 2 assumes schedule-specific fixed-effects for $ED_{50}^{(i)}$, while Model 3 uses schedule-specific random-effects for $ED_{50}^{(i)}$. Model 4 assumes schedule-specific fixed-effects both for $ED_{50}^{(i)}$ and $E_{\text{max}}^{(i)}$, whereas Model 5 uses schedule-specific random-effects both for $ED_{50}^{(i)}$ and $E_{\text{max}}^{(i)}$. For the complete pooling, the rescaled doses are used for the analysis. For Models 3 and 5, we use the biweekly schedule as the reference schedule.

Table 5 displays the LOO-IC values for the five models. The partial pooling with schedule-specific random-effects for $ED_{50}^{(i)}$ results in the best model in terms of the LOO-IC. The second, third, and fourth best models are the partial pooling with fixed-effects for $ED_{50}^{(i)}$, the partial pooling with random-effects for $ED_{50}^{(i)}$ and $E_{\text{max}}^{(i)}$, and the partial pooling with fixed-effects for $ED_{50}^{(i)}$ and $E_{\text{max}}^{(i)}$, respectively. The complete pooling yields the worst performance in terms of LOO-IC. Better performance of partial pooling approaches compared to the complete pooling was also demonstrated in simulations, when there is some heterogeneity in dose–response curves between schedules. Hereafter, we focus on Models 1–3 for convenience.

The posterior estimates obtained by Model 1 (Complete Pooling), Model 2 (PP-FE), and Model 3 (PP-RE) are shown in Table 6. Recall that for PP-FE and PP-RE, the $E_{\text{max}}$ parameters
are shared between schedules. The estimates $ED_{50}^{\text{monthly}}$ of the complete pooling are calculated by rescaling the estimate of $ED_{50}^{\text{biweekly}}$. Across three methods, estimates of $E_0$ are quite similar. For $E_{\max}$ and $ED_{50}^{(i)}$, however, three models yielded different results. Especially, the posterior mean of $E_{\max}^{\text{biweekly}}$ obtained by the complete pooling is quite low compared to partial pooling approaches. The heterogeneity parameter $\tau_{ED_{50}}$ (posterior mean 1.0 with standard deviation of 0.5) indicates that there is considerable heterogeneity in the dose–response curves between schedules.

The estimated dose–response functions $\hat{f}$ by the complete pooling, the partial pooling with fixed-effects (PP-FE), and the partial pooling with random-effects (PP-RE) are displayed in Figure 4. The $\hat{f}(t)$ are the posterior medians for the dose–response function $f(t)$ evaluated across all $i$ where $i \in \{0,5,2, \ldots, 150\}$ (mg-biweekly), equidistant sequence between 0 and 150 with 30 elements. Similarly, 95% equi-tailed credible intervals evaluated across all $i$ are displayed in Figure 4(B). The median dose–response curves obtained by the partial pooling with fixed-effects and the partial pooling with random-effects are very similar. The median dose–response curve estimated by the complete pooling is different from other methods, especially in the dose region between 50 and 150 mg. The bococizumab trial is similar to the scenarios when the rescaled $ED_{50}^{\text{monthly}}$ deviates from $ED_{50}^{\text{biweekly}}$, meaning that high heterogeneity in $ED_{50}^{(i)}$ between schedules. Thus, the bococizumab trial is similar to the simulation scenario with sample size for each arm of 45 and $ED_{50}^{\text{monthly}} = 6$ mg/kg. For this scenario, PP-FE and PP-RE performed better than complete pooling in terms of the mean absolute error in the simulations. Thus, the results from the bococizumab trial are in line with the simulation results.

Additionally, Figure B1 (Appendix B) shows the marginal posterior density estimates of $ED_{50}^{(i)}$ obtained by three methods alongside with the priors used for $ED_{50}^{(i)}$ in the partial pooling with fixed-effects. The marginal posterior distribution for the $ED_{50}^{(i)}$ of complete pooling is very different than the PP-FE and PP-RE approaches.

### 6. Conclusions and Outlook

An assumption of the homogeneity between schedules can be considered unrealistic in some clinical trials, hence a partial pooling is more reasonable than the complete pooling. Rather than using schedule-specific fixed-effects in a partial pooling approach, we have proposed to use schedule-specific fixed-effects for the certain parameters such as $ED_{50}$, allowing dynamically borrowing information in a fully Bayesian framework. In simulation studies, the proposed method displayed more robust performance in terms of the mean absolute error and coverage probabilities for the dose–response function $f(t)$ compared to the complete pooling. Furthermore, the proposed method produces lower mean absolute error and shorter interval estimates for $f(t)$ across most of the scenarios compared to using schedule-specific fixed-effects in a partial pooling approach.

In this article, we focused on the $E_{\max}$ model for the dose–response function. To account for the model uncertainty, it is important to consider alternative functions, such as log-linear or exponential. The shrinkage estimation can be applied to such alternative dose–response models, as well. One way of dealing with the model uncertainty is using a model selection criteria (e.g., AIC in the frequentist context) to decide the right functional form. Hence, by using a criterion such as LOO-IC, one can use the proposed approach to analyze data from a phase II trial with multiple schedules. Alternatively, a Bayesian model averaging approach (Schorning et al. 2016) can be used to deal with uncertainty of dose–response models. Here, we consider phase II trials with multiple schedules. Instead of multiple schedules, one may investigate phase II trials with multiple subgroups, for example, multiple patient populations. The proposed method is still applicable for such situations. We also did not consider to assume schedule specific $E_0$ parameters, which may be appropriate when information of different schedules come from different protocols. One could assume exchangeability to estimate schedule specific $E_0$ parameter in such circumstances. Another important point is that we only considered dose–response models, and not any exposure–response models, which also consider pharmacokinetic (PK) data such as drug concentration in the blood. To investigate the effect of the treatment schedule, an exposure-response modeling might be more informative (Ursino et al. 2017; Günhan, Weber, and Friede 2020). However, the necessity of PK data may result in logistical difficulties in the analysis of a phase II trial, which can be impractical for many trials.

The parameterization used in the proposed method (3) can be considered hard to motivate, since an overall mean of schedule-specific estimates does not have a meaningful interpretation. This can be overcome by adopting an asymmetric parameterization of schedule-specific estimates in terms of a reference schedule as follows

$$ED_{50}^{(k)} \sim N(\alpha_{ED_{50}}, 0) \quad (i.e., \quad ED_{50}^{(k)} = \alpha_{ED_{50}})$$

$$ED_{50}^{(k)} \sim N(\alpha_{ED_{50}}, \beta_{ED_{50}}^2),$$

where $\alpha_{ED_{50}}$ and $\beta_{ED_{50}}$ are the location and scale parameters, respectively (Röver and Friede 2020).

The bococizumab dataset and trials generated in simulations include two treatment schedules only. When analyzing a trial including three or more schedules, the implementation may be carried out slightly differently. Consider a trial including weekly, biweekly, and monthly schedules. When estimating schedule specific $ED_{50}^{(i)}$ in either partial pooling with fixed-effects or partial pooling with random-effects, the monotonic relationship
between ED$_{50}^{(i)}$ parameters of different schedules can be taken into consideration. This can simply be implemented in Stan code by declaring ED$_{50}[\text{weekly}] < \text{ED}_{50}[\text{biweekly}] < \text{ED}_{50}[\text{monthly}]$, which enforces the monotonic relationship between the parameters.

Although the partial pooling with random-effects is an improvement to complete pooling and the partial pooling with fixed-effects, the exchangeability assumption bears the risk of too much shrinkage. Perhaps, it is not very desirable to allow borrowing information for the extreme schedule. To overcome this, the exchangeability-nonexchangeability (EXNEX) models (Neuenschwander et al. 2016) can be considered. EXNEX models can be used to share information across similar schedules, while avoid too much borrowing for the extreme schedule. However, such complicated models should be calibrated well, due to sparse data available in a typical phase II trial.

Appendix A. How to Use the ModStan R Package?

The development version of ModStan is available as a supplementary materials. The bococizumab trial described in the text is available in the package, and it can be loaded as follows:

```r
library("ModStan")
data("dat.Bococizumab")
```

See ?dat.Bococizumab for the description of the dataset.

The `mod_stan` is the main fitting function of the package. The main computations are executed via the rstan package’s `sampling` function. We can fit the partial pooling method with schedule-specific random-effects for the ED$_{50}^{(i)}$ parameter as follows:

```r
PP.RE.Bococizumab.stan = mod_stan(dose = dose, resp = resp, sigma = sigma, schedule = schedule, freq = freq, freq_ref = 24 * 7 * 8, data = dat.Bococizumab, model = "PP-RE", tau_prior_dist = "half-normal", tau_prior = 1, chains = 3, stan_seed = 111, iter = 4000, warmup = 2000)
```

Convergence diagnostics and the results can be very conveniently obtained using the `shinystan` package as follows:

```r
library("shinystan")
launch_shinystan(as.shinystan(PP.RE.Bococizumab.stan$fit))
```

The posterior summary statistics can be obtained using the following command:

```
PP.RE.Bococizumab.stan
```

Appendix B. Marginal Posterior Density Estimates of ED$_{50}$ (Bococizumab Trial)

The marginal posterior density estimates of ED$_{50}^{\text{biweekly}}$ and ED$_{50}^{\text{monthly}}$ obtained by the three methods (CP, PP-FE, PP-RE) are demonstrated in Figure B1. Also, the prior distributions of ED$_{50}^{\text{biweekly}}$ and ED$_{50}^{\text{monthly}}$ used for the PP-FE are shown.

![Figure B1](image_url)

*Figure B1.* Marginal posterior density estimates of ED$_{50}$ for biweekly and monthly schedules obtained by the CP, the PP-FE, and the PP-RE. Also, the prior distributions of ED$_{50}^{\text{biweekly}}$ and ED$_{50}^{\text{monthly}}$ used for the PP-FE are shown.*
Supplementary Materials

The supplementary information includes R and Stan code to reproduce results in the article. There is also a Supplementary Material PDF document, which includes additional information to the main text.

Acknowledgments

We are grateful to Monika Jelizarow and Christian Röver who contributed valuable comments and pointed us to several important references. We thank the Associate Editor and two anonymous reviewers whose comments and suggestions helped to improve this article.

Conflict of Interests

P.M. is an employee of Galapagos NV. T.F. is a consultant to Galapagos NV. MOR106 was jointly discovered by Galapagos NV and MorphoSys, and an MOR106 trial was used to motivate and illustrate the investigations presented here.

References

Alexander, H., Patton, T., Jabbar-Lopez, Z., Manca, A., and Flohr, C. (2019), “Novel Systemic Therapies in Atopic Dermatitis: What Do We Need to Fulfill the Promise of a Treatment Revolution?,” *F1000Research*, 8, 132. [250]

Ballantyne, C., Neulet, J., Cropp, A., Duggan, W., Wang, E., Plochwalk, D., Sweeney, K., Kaila, N., Vincent, J., and Bays, H. (2014), “Efficacy and Safety of Bococizumab (RN316/PF-04950615), a Monoclonal Antibody Against Propionic Acid Subtilisin/Kexin Type 9 in Statin-Treated Hypercholesterolemic Subjects: Results From a Randomized, Placebo-Controlled, Dose-Ranging Study,” *Journal of the American College of Cardiology*, 63, A1374. [250]

Betancourt, M., and Girolami, M. (2015), *Current Trends in Bayesian Methodology With Applications* (Chapter 4), Boca Raton, FL: CRC Press, pp. 79–103. [252]

Bornkamp, B. (2012), “Functional Uniform Priors for Nonlinear Modeling,” *Biometrics*, 68, 893–901. [252]

——— (2014), “Practical Considerations for Using Functional Uniform Prior Distributions for Dose-Response Estimation in Clinical Trials,” *Biometrical Journal*, 56, 947–962. [252]

Bornkamp, B., Pinheiro, J., and Bretz, F. (2018), “Dosefinding: Planning and Analyzing Dose Finding Experiments,” R Package Version 0.9-16, available at [https://CRAN.R-project.org/package=DoseFinding](https://CRAN.R-project.org/package=DoseFinding). [253]

Bretz, F., Pinheiro, J., and Branson, M. (2005), “Combining Multiple Comparisons and Modeling Techniques in Dose-Response Studies,” *Biometrics*, 61, 738–748. [250]

Carpenter, B., Gelman, A., Hoffman, M., Lee, D., Goodrich, B., Betancourt, M., Brubaker, M., Guo, J., Li, P., and Riddell, A. (2017), “Stan: A Probabilistic Programming Language,” *Journal of Statistical Software*, 76, 1–32. [252]

Efron, B., and Morris, C. (1975), “Data Analysis Using Stein’s Estimator and Its Generalizations,” *Journal of the American Statistical Association*, 70, 311–319. [249]

Eichenfeld, L., and Stein Gold, L. (2017), “Systemic Therapy of Atopic Dermatitis: Welcome to the Revolution,” *Seminars in Cutaneous Medicine and Surgery*, 36, S103–S105. [249]

Feller, C., Schorning, K., Dette, H., Bermann, G., and Bornkamp, B. (2017), “Optimal Designs for Dose Response Curves With Common Parameters,” *The Annals of Statistics*, 45, 2102–2132. [249,250,251]

Freidlin, B., and Korn, E. (2013), “Borrowing Information Across Subgroups in Phase II Trials: Is It Useful?,” *Clinical Cancer Research*, 19, 1326–1334. [249]

Friede, T., Röver, C., Wandel, S., and Neuenschwander, B. (2017), “Meta-Analysis of Few Small Studies in Orphan Diseases,” *Research Synthesis Methods*, 8, 79–91. [251]

Gelman, A. (2006), “Prior Distributions for Variance Parameters in Hierarchical Models” (comment on article by Browne and Draper), *Bayesian Analysis*, 1, 515–534. [251]

Giuliano, R., Desai, N., Kohli, P., Rogers, W., Somaratne, R., Huang, F., Liu, T., Mohanavelu, S., Hoffman, E., McDonald, S., Abrahamsen, T., Wasserman, S., Scott, R., and Sabatine, M. (2012), “Efficacy, Safety, and Tolerability of a Monoclonal Antibody to Propionate Convertase Subtilisin/Kexin Type 9 in Combination With a Statin in Patients With Hypercholesterolaemia (LP troublesome-TIMI 57): A Randomised, Placebo-Controlled, Dose-Ranging, Phase 2 Study,” *The Lancet*, 380, 2007–2017. [249,250]

Greenland, S. (2000), “Principles of Multilevel Modelling,” *International Journal of Epidemiology*, 29, 158–167. [250]

Günhan, B. K., Röver, C., and Friede, T. (2020), “Random-Effects Meta-Analysis of Few Studies Involving Rare Events,” *Research Synthesis Methods*, 11, 74–90. [252]

Günhan, B. K., Weber, S., and Friede, T. (2020), “A Bayesian Time-to-Event Pharmacokinetic Model for Phase I Dose-Escalation Trials With Multiple Schedules,” *Statistics in Medicine*, 39, 3986–4000. [258]

Jones, H., Ohlssen, D., Neuenschwander, B., Racine, A., and Branson, M. (2011), “Bayesian Models for Subgroup Analysis in Clinical Trials,” *Clinical Trials*, 8, 129–143. [249]

Mayo Clinic (2018), “Atopic Dermatitis,” available at [https://www.mayoclinic.org/diseases-conditions/dermatitis/symptoms-causes/syc-20353273](https://www.mayoclinic.org/diseases-conditions/dermatitis/symptoms-causes/syc-20353273). [250]

Möllenhoff, K., Bretz, F., and Dette, H. (2020), “Equivalence of Regression Curves Sharing Common Parameters,” *Biometrics*, 76, 518–529. [249,251]

MorphoSys AG (2019), “MOR106 Clinical Development in Atopic Dermatitis Stopped,” available at [https://www.morphosys.com/media-investors/media-center/morphoag-mor106-clinical-development-atopic-dermatitis-stopped](https://www.morphosys.com/media-investors/media-center/morphoag-mor106-clinical-development-atopic-dermatitis-stopped). [250]

Neuenschwander, B., Wandel, S., Roychoudhury, S., and Bailey, S. (2016), “Robust Exchangeability Designs for Early Phase Clinical Trials With Multiple Strata,” *Pharmaceutical Statistics*, 15, 123–134. [250,259]

Pifer (2017), “Monthly and Twice Monthly Subcutaneous Dosing of PF-04950615 (RN316) in Hypercholesterolemic Subjects on a Statin,” Identification No. NCT01592240, available at [https://clinicaltrials.gov/ct2/show/results?NCT01592240?cond=NCT01592240&draw=2&rank=1&view=results.](https://clinicaltrials.gov/ct2/show/results?NCT01592240?cond=NCT01592240&draw=2&rank=1&view=results.) [256]

Röver, C., Bender, R., Dias, S., Schmid, C. H., Schmidli, H., Sturtz, S., Weber, S., and Friede, T. (2020), “On Weakly Informative Prior Distributions for the Heterogeneity Parameter in Bayesian Random-Effects Meta-Analysis,” arXiv No. 2007.08352. [252]

Röver, C., and Friede, T. (2020), “Dynamically Borrowing Strength From Another Study Through Shrinkage Estimation,” *Statistical Methods in Medical Research*, 29, 293–308. [258]

Ruberg, S. (1995), “Dose Response Studies I. Some Design Considerations,” *Journal of Biopharmaceutical Statistics*, 5, 1–14. [249]

Schorning, K., Bornkamp, B., Bretz, F., and Dette, H. (2016), “Model Selection Versus Model Averaging in Dose Finding Studies,” *Statistics in Medicine*, 35, 4021–4040. [258]

Spiegelhalter, D., Abrams, K., and Myles, J. P. (2004), *Bayesian Approaches to Clinical Trials and Health-Care Evaluation*, West Sussex: CRC Press. [251]

Stan Development Team (2018), “Stan Modeling Language Users Guide and Reference Manual, Version 2.18.0,” available at [http://mc-stan.org/](http://mc-stan.org/).

Thaci, D., Simpson, E., Beck, L., Bieber, T., Blauvelt, A., Papp, K., Soong, W., Worm, M., Szepietowski, J., Sofen, H., and Kawashima, M. (2016), “Efficacy and Safety of Dupilumab in Adults With Moderate-to-Severe Atopic Dermatitis Inadequately Controlled by Topical Treatments: A Randomised, Placebo-Controlled, Dose-Ranging Phase 2b Trial,” *The Lancet*, 387, 40–52. [249]

Thomas, N., Sweeney, K., and Somayaji, V. (2014), “Meta-Analysis of Clinical Dose-Response in a Large Drug Development Portfolio,” *Statistics in Biopharmaceutical Research*, 6, 302–317. [250]
Thomas, N., and Wu, J. (2020), “clindr: Simulation and Analysis Tools for Clinical Dose Response Modeling,” R Package Version 2.3, available at https://CRAN.R-project.org/package=clinDR. [256]

Ursino, M., Zohar, S., Lentz, É., Alberti, C., Friede, T., Stallard, N., and Comets, E. (2017), “Dose-Finding Methods for Phase I Clinical Trials Using Pharmacokinetics in Small Populations,” Biometrical Journal, 59, 804–825. [258]

Varadhan, R. (2015), “alabama: Constrained Nonlinear Optimization,” R Package Version 2015.3-1, available at https://CRAN.R-project.org/package=alabama. [251]

Vehtari, A., Gelman, A., and Gabry, J. (2017), “Practical Bayesian Model Evaluation Using Leave-One-Out Cross-Validation and WAIC,” Statistics and Computing, 27, 1413–1432. [257]

Viele, K., Berry, S., Neuenschwander, B., Amzal, B., Chen, F., Enas, N., Hobbs, B., Ibrahim, J. G., Kinnersley, N., Lindborg, S., Micallef, S., Roychoudhury, S., and Thompson, L. (2014), “Use of Historical Control Data for Assessing Treatment Effects in Clinical Trials,” Pharmaceutical Statistics, 13, 41–54. [250]