Meta-Analysis of Clinical Dose–Response in a Large Drug Development Portfolio

Neal THOMAS, Kevin SWEENEY, and Veena SOMAYAJI

This article reports the results of a meta-analysis based on dose–response studies conducted by a large pharmaceutical company between 1998–2009. Data collection targeted efficacy endpoints from all compounds with evidence of clinical efficacy during the time period. Safety data were not extracted. The goal of the meta-analysis was to identify consistent quantitative patterns in dose–response across different compounds and diseases. The article presents summaries of the study designs, including the number of studies conducted for each compound, dosing range, the number of doses evaluated, and the number of patients per dose. The E\text{max} model, ubiquitous in pharmacology research, was fit for each compound. It described the data well, except for a single compound, which had nonmonotone dose–response. Compound-specific estimates and Bayesian hierarchical modeling showed that dose–response curves for most compounds can be approximated by E\text{max} models with “Hill” parameters close to 1.0. Summaries of the potency estimates show pharmacometric predictions of potency made before the first dose ranging study within a (1/10, 10) multiple of the final estimates for 90% of compounds. The results of the meta-analysis, when combined with compound-specific information, provide an empirical basis for designing and analyzing new dose finding studies using parametric E\text{max} models and Bayesian estimation with empirically derived prior distributions.

Key Words: Bayesian pharmacometric models; Dosing design; E\text{max}.

1. Introduction

A large research portfolio of drugs was analyzed to answer the questions: (1) “What do clinical dose–response curves look like?,” (2) “How well were they determined by the studies conducted?,” and (3) “Are there consistent patterns in dose–response across diseases and compounds with different pharmacological targets?”

Data collection is described in Section 2.1. The study designs for each compound, and the data distributions within dose groups are described in Sections 2.2 and 2.3. The data collection and analyses were restricted to efficacy endpoints. The E\text{max} model is fit separately for each compound in Section 3 using maximum likelihood estimation (MLE). Analyses of the data from the only compound with a nonmonotone dose–response are in Section 3.4. The distribution of the parameters of the E\text{max} model are described in Section 4. The MLE’s of the parameters are summarized and then the parameter estimation is refined by the use of hierarchical Bayesian models representing the distributions of the parameters across compounds. The posterior distribution of the hierarchical parameters are used to derive the posterior predictive distribution for a new compound. This distribution can be used to evaluate the likely utility of different dosing designs for a future compound, and it provides an empirically derived prior distribution for future dosing studies. Supplementary summaries are provided for each compound. The summaries include graphs, tables, and model fits. The supporting documentation also includes
There is a large statistical literature on dose–response studies, including collected volumes such as Krishna (2006), and Ting (2006). Some researchers have focused on hypothesis testing approaches and corresponding confidence intervals as the basis for dose selection. Examples include methods in Bretz et al. (2003), and Hsu and Berger (1999), Tamhane et al. (1996), and Ruberg (1989), which build on the earlier work of Williams (1971) and Dunnett (1955). These methods invoke few assumptions regarding the nature of the dose–response curve beyond monotonicity. They tend to be sequential in nature and differ primarily in the family of hypotheses tested, and the extent of family-wide error control achieved. The MCP-Mod testing approach (Bretz, Pinheiro, and Branson 2005), in contrast, is derived using a set of parametric dose–response curves and seeks to test the global hypothesis of no drug effect. Many dose–finding approaches have also been developed within a curve estimation framework. These approaches vary widely regarding the assumptions made about the dose–response curve. Much recent work has focused on methods that make few assumptions about the shape of the dose–response curve, for example, the normal dynamic linear model, Berry et al. (2002), and other flexible models like Bornkamp and Lekstadt (2009). Such approaches may not assume monotonicity, as was done with earlier methods like isotonic regression (McDonald 2005). Methodological research based on specific parametric models has been less common because these methods are well developed. An exception is the estimation phase of the MCP-Mod approach (Bretz, Pinheiro, and Branson 2005), which emphasizes the selection or averaging of common parametric models. There has been recent work deriving optimal designs for parametric dose–finding methods (Bretz, Dette, and Pinheiro 2010; Dette et al. 2010, 2008). Bornkamp et al. (2011) is an example of adaptive designs for dose finding that incorporates many of the methods referenced here. It is not an objective of the current article to develop new methods, but rather to characterize the likely conditions (e.g., dose–response shapes, sampling designs) in clinical dose–response studies to guide the selection and development of dose–finding methods.

2. Data and Study Designs

2.1 Data Collection

A review of the study repository of a large sponsor (Pfizer) identified all dose–response studies initiated between 1998–2009. The portfolio represented roughly 10% of global pharmaceutical industry research and development spending during the time period (PhRMA 2008; Pfizer, Inc 2009; EP Vantage, Inc 2010). The repository included at least one compound in 13 of the 17 therapeutic classes containing three or more patented compounds in 2003 (Congressional Budget Office 2006). Two of the 13 classes included in the repository are not represented in our primary analyses because none of the compounds in these classes met the inclusion criteria. The sampling of compounds followed a prespecified plan with some modifications in response to unanticipated conditions.

Because the portfolio was focused on small molecules, and complex biological molecules may have different dose–response characteristics, the meta analysis was restricted to small-molecule compounds (i.e., < 25 atoms). Oncology compounds were also excluded because they have different dosing objectives and methods to achieve them. Our analyses consider only efficacy data, and they are restricted to a single timepoint. Trends in dose–response as a function of time are not reported. No safety data were extracted.

Compounds with evidence of clinical efficacy, which was based on assessments in clinical study reports, were selected for follow-up data collection. Dose–response studies for the selected compounds initiated before 1998 were also included when they could be located. Most of the studies were from phase 2 of development, but phase 3 studies with three or more doses were also included. Because of the time required to locate, extract, and verify the data from each study, it was not feasible to collect data on the large number of compounds that failed to demonstrate clinical efficacy. The implications of the compound sampling design are discussed in Section 4.

The initial data screening yielded 87 studies from 34 compounds. Patient-level efficacy data were extracted from 85% of the studies, and summarized results were obtained from study reports for the remaining studies. Upon reviewing the study-level data, four compounds from the original screening were excluded for lack of evidence of dose–response (three of the four were antiinfectives without placebo data). Two cross-over studies were identified, but both used different endpoints and visit schedules that could not be combined with the larger parallel group studies, so these studies were excluded. These exclusions yielded a total of 76 studies on 29 compounds. Additionally, the monotherapy data from a combination product were retained and merged with data from the compound’s monotherapy development, while the data from the combination product were excluded. The disposition of the compounds and studies are summarized in Figure 1. A list of the compounds is in Table 1, and a list of abbreviations is included in the supplemental materials. A numerical ID was assigned to each compound and used to reference compounds throughout. The missing numbers in the ID
sequence are the excluded compounds, which have already been described. Some compounds were evaluated in more than one distinct disease area (e.g., compound 25 for rheumatoid arthritis and psoriasis). The initial disease evaluated is denoted by appending “0.0” or “0.1” to the ID, and “0.2” or “0.3” for the second and third diseases studied. Except where noted, the same compound evaluated for more than one disease will be regarded as multiple distinct compounds. Another complication occurred when there were distinct subpopulations evaluated for the same disease so the dose–response curves may differ (e.g., treatment-naive and experienced patients in antiviral trials). When there was more than one subpopulation, the studies for the subpopulation with less data were identified by adding 100 to the compound ID. Separate model fits and other summaries were included in the supplemental materials for these secondary populations, but they were not included in the cross-compound summaries. Regarding different diseases treated by the same compound as separate “compounds” in cross-compound summaries increased the number of compounds from 29 to 33. The results of the hierarchical modeling in Section 4 were not substantially changed when the data from the additional diseases were excluded.

Data were collected on the primary efficacy endpoint(s) from each study at each planned study visit. When there were multiple endpoints and differing visit schedules, a visit and endpoint were selected that were available for all studies from a compound to permit pooling across studies. All of the analyses reported here are based on the single endpoint and timepoint. Our analyses are based on available cases without imputation; the missing data rates for most studies were between 5–20%, but one study had a missing data rate of 38%. As part of quality checking, dose groups summaries based on the available cases data were compared to corresponding primary analyses in the study reports. Exact agreement was not anticipated nor observed, but there were no changes to dose–response patterns apparent in graphical displays. Dosing data were also collected, including information...
Table 1. Compound characteristics. For data source, IRO = Investigator reported outcome, L = laboratory, PRO = patient reported outcome. CVMED is cardiovascular and metabolic diseases

| ID   | Therapeutic area | Data source | Endpoint                                      | Type          |
|------|------------------|-------------|-----------------------------------------------|---------------|
| 1.0  | CVMED            | L           | HbA1c (Diabetes)                              | Continuous    |
| 2.0  | CVMED            | L           | HbA1c (Diabetes)                              | Continuous    |
| 3.0  | CVMED            | L           | Lumbar Spine BMD                              | Continuous    |
| 4.0  | CVMED            | L           | Percent Change HDL                            | Continuous    |
| 5.0  | CVMED            | Vitals      | Body Weight                                   | Continuous    |
| 6.0  | CVMED            | L           | Percent Change LDL                            | Continuous    |
| 8.0  | CVMED            | PRO         | Freq Vasomotor Symptoms                       | Continuous    |
| 9.0  | CVMED            | IRO         | Venous Thromboembolism                        | Binary        |
| 10.0 | CVMED            | Vitals      | Systolic Blood Pressure                       | Continuous    |
| 11.0 | CVMED            | IRO         | Maintenance Sinus Rhythm                      | Binary        |
| 12.0 | CVMED            | Vitals      | Diastolic Blood Pressure                      | Continuous    |
| 13.0 | Neuroscience     | IRO         | PANSS (Schizophrenia)                         | Continuous    |
| 14.0 | Neuroscience     | PRO/L       | Continuous Quit (Smoking)                     | Binary        |
| 15.0 | Neuroscience     | PRO         | Headache Response                             | Binary        |
| 16.0 | Neuroscience     | L           | sWASO(Sleep)                                  | Continuous    |
| 17.1 | Neuroscience     | PRO         | Mean Pain Score                               | Continuous    |
| 17.2 | Neuroscience     | PRO         | Percent Change Seizures                       | Continuous    |
| 17.3 | Neuroscience     | IRO         | HAM-A (Depression)                            | Continuous    |
| 18.0 | Infectious Diseases | L     | HIV Log Viral Load                            | Continuous    |
| 22.0 | Infectious Diseases | L     | HIV Log Viral Load                            | Continuous    |
| 23.0 | Infectious Diseases | L     | HIV Log Viral Load                            | Continuous    |
| 24.0 | Inflammation     | PRO         | WOMAC (Osteoarthritis Pain)                   | Continuous    |
| 25.1 | Inflammation     | PRO         | ACR20 (Arthritis)                             | Binary        |
| 25.2 | Inflammation     | IRO         | PASI75(Psoriasis)                             | Binary        |
| 26.1 | Pain             | PRO/IRO     | Patient Global Assessment                     | Continuous    |
| 26.2 | Pain             | PRO         | Composite Pain Score                          | Continuous    |
| 27.0 | Pain             | PRO         | Composite Pain Score                          | Continuous    |
| 28.0 | Pain             | PRO         | Total Pain Relief                             | Continuous    |
| 29.0 | Genitourinary    | PRO         | Micturitions in 24 hours                      | Continuous    |
| 31.0 | Genitourinary    | PRO         | EDD(Erectile Dysfunction)                     | Continuous    |
| 32.0 | Genitourinary    | PRO         | Int Prostate Symptom Score                    | Continuous    |
| 33.0 | Genitourinary    | PRO         | EDD(Erectile Dysfunction)                     | Continuous    |
| 34.0 | CVMED            | L           | Percent Change HDL                            | Continuous    |
| 106.0| CVMED            | L           | Percent Change LDL                            | Continuous    |
| 117.1| Neuroscience     | PRO         | Mean Pain Score                               | Continuous    |
| 125.1| Inflammation     | PRO         | ACR20 (Arthritis)                             | Binary        |
| 126.1| Pain             | PRO/IRO     | Patient Global Assessment                     | Continuous    |

on dosing regimen (e.g., once or twice daily). A small number of compounds had a mix of dosing regimens; preliminary examination of the response data from the different regimens did not display marked differences, so data from different regimens were combined using total daily dosing, but this assumption was reexamined during model fitting. The response data were masked by linear rescaling so that the overall mean for each compound without regard to dose group or study was zero, and the overall standard deviation (SD) was 1.0. The dosing data were rescaled so the maximum dose tested for each compound was 1.0.

2.2 Designs for Estimating Dose–Response

All of the studies included in the final database had parallel group designs. Multiple dose–response studies per compound were common: 16 compounds were evaluated in a single study, 9 in two studies, and 8 compounds in three—six studies. Multiple studies were required to better characterize the dose–response as the first study was often inadequate, as will be described. Only one study used a substantially adaptive dosing design. Development of some of the compounds with a single study was terminated without completing dose selection for a variety of reasons. From the 33 unique compounds/diseases, only 20 advanced to phase three.

There were 9 compounds evaluated on 4 doses (including placebo), 4 compounds on 5 doses, 10 compounds on 6 doses, 7 compounds on 7 doses, and 3 compounds on 8–10 doses. The ratio of the highest dose to the lowest (nonplacebo) dose was computed for each compound across all studies. The 25th, 50th, and 75th quantiles of the dosing ratio were 8, 16, and 30, respectively. The common hyperbolic $E_{\text{max}}$ model, which is described in the Section 3.1, predicts a ratio of 81 between the doses yielding 90% and 10% of the maximum effect, so most of the compounds were evaluated over limited dosing ranges,
2.3 Data Distributions Within Dose Groups

There were six compounds with binary primary endpoints, and 27 compounds with continuous primary endpoints. There were 20 endpoints derived from patient or investigator assessment, and 13 endpoints based on laboratory measurements or vital signs (e.g., blood pressure, weight). Many of the clinical endpoints were collected from ordinal rating scales with numerous categories, which were regarded as continuous in the protocol-specified analyses. Some count data with high frequencies were also regarded as continuous by the researchers; one count endpoint with low frequency had been converted to a binary response. There were some compounds in the repository that had time-to-event primary endpoints, but none of these compounds met the clinical efficacy inclusion criteria. In the remainder of this section, the distributions of the continuous endpoints within dose groups are described for the 25 compounds with continuous patient-specific data available.

Boxplots of the data within dose groups (pooled across studies within compound) are included with the supplemental summaries. Most of the boxplots are consistent with symmetric bell-shaped data distributions, reflecting in part transformations selected by the data analysts (e.g., log transformation of viral load), and the use of dichotomous responder variables when continuous and polytomous variables had problematic distributions. Outliers are present in many of the boxplots. They were defined as any point exceeding the lower and upper quartiles by more than 1.5 times the interquartile range (Tukey 1977). The median (across compounds) proportion of outliers was 0.025 and its lower and upper quartiles were 0.014, and 0.047, respectively. To give context to these proportions, data were simulated in the dose groups of each compound from a homogeneous distribution with 7 degrees of freedom. One hundred portfolios were simulated, and the summarized proportions of outliers were averaged yielding a median proportion of outliers of 0.028, and averaged quartiles of 0.022 and 0.037, which were similar to those from the real data. In contrast, the median proportion of outliers from a normal distribution was 0.01. The rates of outliers in the dose–response data are consistent with broader experience reported from both laboratory (Posekany, Felsenstein, and Sykacek 2011), and less controlled data sources (Venables and Ripley 2002).

There was evidence of variance heterogeneity that increased with dose, but the increases were not as large as reported in some nonclinical dose–response settings. Because of the presence of outliers, robust measures of scale were computed from the interquartile ranges. To improve stability, the average of the interquartile ranges from the two highest dose groups were compared to the average from the two lowest dose groups (including placebo). Of the compounds with continuous endpoints, four had estimated ratios exceeding 2.0, with a maximum of 2.59. There were no ratios below 0.5. In the simple t-based simulations with homogeneous variability, no simulated sets of designs yielded more than two estimated ratios greater than 2.0.

The consistency of the placebo response when there were multiple studies for a compound was also evaluated. Likelihood ratio tests (binary data) or F-tests (continuous, normal model) were computed for each compound testing the homogeneity of the placebo response across studies. The proportion of compounds with placebo heterogeneity rejected at the 0.05 level was 0.19, and the p-values of the tests generally trended toward smaller values. The impact of the heterogeneity was apparent on some of the models fit in Section 3, so all models include a separate intercept (placebo) parameter for each study.

3. E_{\text{max}} Models for Dose–Response

3.1 E_{\text{max}} Models for Continuous and Binary Data

For continuous outcomes, denoted by Y, a normal model was used with homogeneous residual variance, \sigma^2, and a mean function, called the sigmoid E_{\text{max}} model, given by

\[ E(Y \mid D) = E_0 + \frac{E_{\text{max}} D^\lambda}{ED_{50}^\lambda + D^\lambda}, \tag{3.1} \]

where D is dose, and (E_0, ED_{50}, E_{\text{max}}, \lambda) are unknown parameters. The mean response is defined for D \geq 0, and changes monotonically from E_0 at D = 0 to a maximum difference of E_{\text{max}}. The parameter, \lambda, is called the Hill parameter, and the model in (3.1) is called a hyperbolic E_{\text{max}} model when \lambda = 1 (MacDougall 2006). For the hyperbolic E_{\text{max}} model, the ratio of the dose achieving 90% of the maximum difference with placebo, ED_{90}, and the dose achieving 10% of the effect, ED_{10}, is always 81. The model yields an increasingly steep sigmoidal shape as \lambda increases, and a more gradual approach to the maximum difference with \lambda < 1. When E_0 = 0 and E_{\text{max}} = 1, the model for the conditional mean is the distribution func-

-0.1
tion for the log-logistic distribution (Johnson and Kotz 1995). The model can be derived from receptor occupancy models in pharmacology when the \( \lambda \) parameter is a positive integer (Iliadis and Macheras 2006). Common parameters were assumed across different studies for the same compound, except for different placebo parameters, \( E_0 \), which were estimated separately for each study. The adequacy of the assumption of common treatment differences relative to placebo across studies, which is implied by the assumed common model parameters, was assessed through formal tests and graphical displays described in Section 3.3.

Corresponding \( E_{\text{max}} \) models were also used to fit data for compounds with binary outcomes, applied on the logit scale (e.g., Sheiner et al. 1997; Tan et al. 2011):

\[
\logit(P(Y | D)) = E_0 + \frac{E_{\text{max}}D^\lambda}{ED_{50} + D^\lambda}, \quad (3.2)
\]

dose–response curves based on this model were back-transformed to the response rate scale.

Parametric collections of dose–response curves based on other distribution functions could also be used to model the dose–response data. Because most of the dose–response curves were estimated with low precision, it is likely there are other parametric collections that would yield goodness-of-fit statistics similar to those produced by the \( E_{\text{max}} \) model.

### 3.2 Automated Fitting Algorithm

The \( E_{\text{max}} \) models for the continuous outcomes were fit using the nonlinear least squares R function, \( \text{nls} \), and the binary responses with MLE using the R function \( \text{nlm} \) (R Development Core Team 2011). An automated algorithm tried Newton-type maximization methods initially, and a partial least squares method if the Newton method failed. Two different starting values were tried if needed to achieve convergence, which was judged using the default criteria. The log of the \( ED_{50} \) parameter in (3.1) and (3.2) was estimated directly, and then back-transformed. The MLE’s of the parameters are denoted by \( (\hat{E}_0, \hat{ED}_{50}, \hat{E}_{\text{max}}, \hat{\lambda}) \).

Models were fit to all compounds including those in secondary subpopulations, which were included in the supplemental materials. The summaries here are for the primary populations. Convergence was achieved for the hyperbolic \( E_{\text{max}} \) models in 26 of the 27 compounds with continuous outcomes, while convergence was achieved for only 18 of the sigmoid models. The corresponding convergence for the 6 binary outcome models was 5 and 4, for the hyperbolic and sigmoid models, respectively. Summaries of the ML parameter estimates from converged fits are in Section 4. Model fits for the compounds where ML estimation failed were obtained using Bayesian methods, also described in Section 4.

Data from the compounds where the automated fitting failed were explored to understand the failed convergence. Most of these were settings with few doses (e.g., four doses including placebo), and the doses included were on the steeper linear portion of the \( E_{\text{max}} \) curve, or the doses were only near the plateau. Because of the lack of identification, further exploration of starting values and parameter transformations (Ross 1990) did not produce additional converged estimates. Convergence was sometimes sensitive to the convergence criteria.

### 3.3 Assessing Model Fit

The fitted sigmoid \( E_{\text{max}} \) curves for three compounds with well-studied pharmacological targets are displayed in Figures 2–4. The \( E_{\text{max}} \) curve of a statin compound (ID 6) for reducing low-density cholesterol (LDL) has \( \lambda = 0.35 \) (Figure 2). Mandema et al. (2005) found similar curves for four other statin compounds, which could be adequately described with a common \( E_{\text{max}} \) and a common \( \lambda \) parameter, while allowing for different potencies (\( ED_{50} \)) and placebo responses. An HIV C-C chemokine receptor 5 (CCR5) inhibitor (ID 22) for viral load reduction in Figure 3 had a common estimation challenge arising because doses near the plateau were not tested, so although convergence was achieved and the resulting \( E_{\text{max}} \) curve described the data well, the individual parameter estimates (e.g., \( \lambda = 0.51 \)) were unstable. A compound (ID 18) with the same target was tested over a broader dosing range, and produced \( \lambda = 1.08 \). The dose–response for a phosphodiesterase 5 (PDE5) inhibitor for treatment of erectile dysfunction, which has \( \lambda = 0.93 \) is displayed in Figure 4. An unpublished internal meta-analysis of two other marketed compounds, and another internal compound (ID 33) with the same target, were also well described with the same \( \lambda (= 0.8) \) and \( E_{\text{max}} \). The shared \( E_{\text{max}} \) and \( \lambda \) parameters of compounds with the same pharmacological target despite different potencies is predicted by common molecular binding models in pharmacology.

For compounds with more than one study, heterogeneity of the overlapping dose group differences from placebo was tested using a likelihood ratio (binary data) or F-test (continuous data) based on an unstructured model for the dose group differences. There was generally good consistency in the dose group differences despite heterogeneity in placebo response. Of the 17 compounds with multiple studies, only two had nominally significant heterogeneity at the 0.05 level. One compound (ID 15, \( p \)-value = 0.049) had moderately larger effects in the second study than the first, with no apparent explanation. The second compound (ID 17.1, \( p \)-value = 0.006), had different dosing regimens (once versus twice daily dosing).
between its two studies, which may have contributed to this discrepancy. Neither compound had dose–response curves that displayed nonmonotonicity or deviation in appearance from the usual E\(_{\text{max}}\) shapes.

Goodness-of-fit F-tests (Bates and Watts 1988) and likelihood ratio goodness-of-fit tests were computed for continuous and binary endpoints comparing the hyperbolic E\(_{\text{max}}\) model to an unstructured but homogeneous (across studies) model after adjustment for placebo amongst all compounds, except the two compounds with evidence of heterogeneous means. One compound (ID 34) displayed clear evidence of nonmonotone dose–response, and it is described in detail in Section 3.4. The compound in Figure 2 was well described by an E\(_{\text{max}}\) model with \(\lambda < 1\), but not \(\lambda = 1\). Two other compounds had a nominally significant test. One of these compounds (ID 4) had multiple protocols combining once and twice daily dosing, which accounted for the model discrepancies when the different dosing regimens were fitted separately. The other compound (ID 9) was the only one evaluated using an adaptive dosing design. The response to one dose was better than predicted by the model (approximately 3 SE), and better than observed for the highest dose. Mandema, Boyd, and DiCarlo (2011) performed a meta-analysis that included this compound, and two other compounds with the same target, both of which completed phase 3 testing. The dose–response for the other compounds was well described by an E\(_{\text{max}}\) model. Because of the lack of concurrent randomization, it is unclear whether the discrepant response was due to bias, or sampling variability consistent with the examination of numerous compounds.

The likelihood ratio test comparing the sigmoid and hyperbolic E\(_{\text{max}}\) model was significant for only one compound (ID 6, Figure 2). The power to reject the hyperbolic E\(_{\text{max}}\) model with \(\lambda = 1\) was low for most individual compounds, and the quality of the usual asymptotic approximations is suspect (Thomas 2006). The distribution of \(\lambda\) across compounds is quantified in Section 4.2.

### 3.4 Example of Nonmonotone Dose–Response

The dose–response of a compound (ID 34) to raise high-density cholesterol displayed evidence of
Figure 3. Dose–response curves for an HIV CCR5 inhibitor (ID 22) with viral load endpoint. The points are dose group means, the curves are sigmoid Emax models with a separate placebo response for each study, and the bars are 95% prediction intervals.

nonmonotone dose–response in Figure 5. The goodness-of-fit test for the hyperbolic Emax model was 0.002; the sigmoid Emax model did not converge for this data, which included only three dose groups and placebo. Prior to completion of the study, Nissen et al. (2007) published results from two studies that displayed a nonmonotone dose–response for a compound with the same target. A Bayesian Emax analysis had been proposed in the original analysis plan, but the plan was altered to a simple unstructured mean model after publication of results for the other compound. As part of the Bayesian Emax analysis, a posterior predictive check was planned to compare the best of the lower dose group sample means to the sample mean from the highest dose. The posterior predictive probability was 0.007, so even without the published information, it is likely that the analyses would have been appropriately amended. Both compounds were terminated, so the practical decision would have been unchanged if the incorrect assumptions about dose–response not been corrected.

4. Characterizing the Distribution of E_{max} Model Parameters

Bayesian hierarchical models with distributions summarizing the parameters from different compounds were fit to assess the variability in the parameters adjusting for estimation error. The data from all compounds with continuous outcomes were included, even if the data from a compound were inadequate to produce stable MLE, except the one compound (ID 34) with nonmonotone dose–response was excluded. The parameters from the compounds with binary endpoints were similar to those for continuous endpoints, but they were not included to simplify the modeling.

Let j index the compounds with continuous outcomes included in the analysis, and k = 1, ..., n_j index the n_j protocols for compound j. The model includes compound-specific E_{max,j}, ED_{50,j}, and \lambda_j parameters, and protocol-specific E_{0/k} parameters. The mean response for
Figure 4. Dose–response curves for a PDE5 inhibitor (ID 31) with erectile dysfunction endpoint. The points are dose group means, the curves are sigmoid E_max models with a separate placebo response for each study, and the bars are 95% prediction intervals.

A patient receiving dose $D$ of compound $j$ is

$$E_{0jk} + \frac{E_{\text{max},j} D^{\lambda_j}}{\text{ED}_{50j} + D^{\lambda_j}}.$$  

The residual variances for the (assumed) normally distributed errors are also compound-specific and denoted by $\sigma_j^2$.

Distributions are specified for the $\lambda_j$ and $\text{ED}_{50j}$ parameters across the compounds. Cross-compound distributions for the $E_{0jk}$ parameters are not specified. When fitting hierarchical models for the $\lambda_j$ and $\text{ED}_{50j}$, independent normal prior distributions with mean 0 and SD of 3 are used for the $E_{0jk}$ throughout. Recall that the data are normalized so the overall mean is 0 and the overall SD is 1 for each compound. Because the compounds were selected based on their differentiation from placebo, the distribution of effect sizes in the sample would not be an appropriate prior distribution for a new compound without demonstrated efficacy, so a cross-compound distribution for the $E_{\text{max},j}$ parameters was not estimated. For the hierarchical models, independent $t_3$ ($t$ with 3 degrees of freedom) prior distributions with mean 0 and scale parameter 1 are specified for $\log|E_{\text{max},j}|$ throughout, with the direction of the effect assumed to be known.

The magnitude of the ML estimates, $\hat{E}_{\text{max},j}$, are described in Section 4.1, and the prior distribution specified for each $E_{\text{max},j}$ is explained. Parameteric hierarchical distribution for the $\lambda_j$ are specified in Section 4.2.
Figure 5. Dose–response curve for a peroxisome proliferator-activated receptor α compound (PPAR-α, ID 34) to increase high-density cholesterol. The points are dose group means, the curve is the hyperbolic $E_{\text{max}}$ model, and the bars are 95% prediction intervals.

Preliminary estimates of these distributions are reported. Diffuse independent prior distributions for the $ED_{50j}$ are used for the models fit in Section 4.2. In Section 4.3, hierarchical parametric distributions are specified for the $ED_{50j}$, and preliminary estimates of these distributions are reported assuming hyperbolic $E_{\text{max}}$ models (i.e., the $\lambda_j = 1$). Section 4.4 summarizes the estimates of the hierarchical distributions from simultaneous estimation of different combinations of the hierarchical models for the $\lambda_j$ and $ED_{50j}$. Model sensitivity and other fitting issues are also assessed.

The distribution of the parameters across compounds requires special consideration because only compounds with demonstrated efficacy were included. When the $E_{\text{max}}$ parameter is 0, the $ED_{50}$ and $\lambda$ parameters are not identified. Compounds with nonzero $E_{\text{max}}$ parameters so small their effects cannot be demonstrated in clinical trials are excluded from our sample, but their dose–response shapes are of little practical interest, even if they differ from those included in our sample. An anomalous situation can arise when diffuse prior distributions are assumed for the $E_{\text{max}}$ and $ED_{50}$ parameters. A compound with $|E_{\text{max}}| > 0$ can fail to demonstrate effectiveness if the doses in the clinical trial design are much too low. The efficacy data from a null clinical trial cannot differentiate a small $|E_{\text{max}}|$ value from an $ED_{50}$ much larger than was anticipated when the trial was designed. The use of diffuse prior distributions assigns high probability to larger values of $|E_{\text{max}}|$ and $ED_{50}$, which can produce unintended high posterior probabilities for very large values of the $ED_{50}$ when the trial results are null. Our analyses included only efficacious compounds because it was not feasible to extract data from the numerous failed compounds, but this selection criteria also reduces sensitivity to the ambiguous implications for the model parameters that arise with null efficacy data.

To reduce computation time, the data were summarized by dose group means and within-dose group sums of squares. All of the models were evaluated using OpenBUGS software (Lunn et al. 2009), and confirmed using
often less than 1. the standardized treatment effects achieved were most areas where laboratory biomarkers (e.g., viral load, LDL) 1, and 3.77, while they were 0.45, 0.59, and 1.21 for the standardized effects on laboratory endpoints were 0.79, and 3.0, 3.77, and 1.21 for the portfolio of compounds, including those where the ML estimation failed. 

The second distribution for λ was a beta distribution scaled to the interval (0, 6). This range was specified because larger values of λ are practically indistinguishable, and a bounded distribution avoids potential numerical difficulties. Before scaling to (0, 6), the beta distribution was parameterized by the logit of the mean, denoted by α, and a size parameter, γ, as was done in Kahn and Raftery (1996) to support the assumption of prior independence of the parameters. The α parameter was assigned a normal prior distribution with mean 0, and an SD of 1.75. The prior distribution for γ is a Γ distribution with shape and rate parameters 1.0 and 0.1. When the inverse logit transform is applied to α, followed by scaling, the prior distribution for the mean of λ is centered in (0, 6) with support throughout. The prior distributions for α and γ were selected to yield an approximately uniform distribution for λ, with some additional probability near 0 and 6. The third prior distribution for λ was a discrete distribution with 12 possible values in (0.5, 1.0, 1.5, . . . , 6). The probabilities for the λ values were assigned a symmetric Dirichlet distribution with a concentration parameter of 1.0.

For the initial modeling of the λj distribution, the log(ED50) were assigned independent t2 distributions centered about initial projections of the ED50 with scale 2.0; the empirical basis for this prior distribution is explained in Section 4.3. The (5th, 10th, 25th, 50th, 75th, 90th, 95th) percentiles computed from the posterior predictive distribution for a new λ based on the prior log t2 distribution were (0.59, 0.67, 0.77, 0.86, 0.97, 1.09, 1.22). The corresponding posterior percentiles based on the scaled beta distribution were (0.33, 0.43, 0.62, 0.87, 1.16, 1.47, 1.7). The posterior distribution of λ based on the Dirichlet prior distribution had P(λ = 0.5) = 0.13, P(λ = 1) = 0.82, P(λ ≥ 1.5) = 0.05, with the small probability in the upper tail roughly uniformly distributed. The posterior distributions were concentrated at or below λ = 1, as was typically achieved in practice, and their estimation can be unstable. A more interpretable measure of the effect magnitude for the continuous endpoints was the ratio of the highest dose and placebo divided by the mean value for the continuous endpoints was computed by scaling, the prior distribution for the Emax parameters used throughout Section 4 has a conservatively selected scale parameter of 1.0.

4.1 The Emax Parameter

The estimated magnitude of the efficacy of the sampled compounds is summarized in this section, but a model for the Emax, across compounds is not developed. The 25th, 50th, and 75th percentiles of |Emax| estimated for the hyperbolic model from compounds with continuous endpoints were 0.86, 1.07, and 1.69, with minimum and maximum values of 0.2 and 3.76. The quantiles of log|Emax| were well approximated by the quantiles of a t distribution with mean 0 and scale parameter 0.44. The prior distribution for the Emax parameters used in the model fits that is described in Section 4.4. The mean was assigned a diffuse normal distribution with mean 0 and SD 10, and σ2 was assigned a diffuse gamma distribution with shape and rate values of 0.001. The t distribution was used for compound-specific parameters to reduce the shrinkage applied to potential outliers.

The second distribution for λ was a beta distribution scaled to the interval (0, 6). This range was specified because larger values of λ are practically indistinguishable, and a bounded distribution avoids potential numerical difficulties. Before scaling to (0, 6), the beta distribution was parameterized by the logit of the mean, denoted by α, and a size parameter, γ, as was done in Kahn and Raftery (1996) to support the assumption of prior independence of the parameters. The α parameter was assigned a normal prior distribution with mean 0, and an SD of 1.75. The prior distribution for γ is a Γ distribution with shape and rate parameters 1.0 and 0.1. When the inverse logit transform is applied to α, followed by scaling, the prior distribution for the mean of λ is centered in (0, 6) with support throughout. The prior distributions for α and γ were selected to yield an approximately uniform distribution for λ, with some additional probability near 0 and 6. The third prior distribution for λ was a discrete distribution with 12 possible values in (0.5, 1.0, 1.5, . . . , 6). The probabilities for the λ values were assigned a symmetric Dirichlet distribution with a concentration parameter of 1.0.

The 25th, 50th, and 75th percentiles of the converted λ amongst compounds with continuous endpoints were 0.85, 1.13, and 1.61, with minimum and maximum values of 0.35 and 4.6. Three different distributions were used to represent the variation of the underlying λj across the portfolio of compounds, including those where the ML estimation failed.

The first distribution was a tα distribution for log(λj) conditional on the mean, μj, and scale parameter, σj. The mean was assigned a diffuse normal distribution with mean 0 and SD 10, and σ−2 was assigned a diffuse gamma distribution with shape and rate values of 0.001. The tα distribution was used for compound-specific parameters to reduce the shrinkage applied to potential outliers.

The second distribution for λ was a beta distribution scaled to the interval (0, 6). This range was specified because larger values of λ are practically indistinguishable, and a bounded distribution avoids potential numerical difficulties. Before scaling to (0, 6), the beta distribution was parameterized by the logit of the mean, denoted by α, and a size parameter, γ, as was done in Kahn and Raftery (1996) to support the assumption of prior independence of the parameters. The α parameter was assigned a normal prior distribution with mean 0, and an SD of 1.75. The prior distribution for γ is a Γ distribution with shape and rate parameters 1.0 and 0.1. When the inverse logit transform is applied to α, followed by scaling, the prior distribution for the mean of λ is centered in (0, 6) with support throughout. The prior distributions for α and γ were selected to yield an approximately uniform distribution for λ, with some additional probability near 0 and 6. The third prior distribution for λ was a discrete distribution with 12 possible values in (0.5, 1.0, 1.5, . . . , 6). The probabilities for the λ values were assigned a symmetric Dirichlet distribution with a concentration parameter of 1.0.

For the initial modeling of the λj distribution, the log(ED50) were assigned independent tα distributions centered about initial projections of the ED50 with scale 2.0; the empirical basis for this prior distribution is explained in Section 4.3. The (5th, 10th, 25th, 50th, 75th, 90th, 95th) percentiles computed from the posterior predictive distribution for a new λ based on the prior log t2 distribution were (0.59, 0.67, 0.77, 0.86, 0.97, 1.09, 1.22). The corresponding posterior percentiles based on the scaled beta distribution were (0.33, 0.43, 0.62, 0.87, 1.16, 1.47, 1.7). The posterior distribution of λ based on the Dirichlet prior distribution had P(λ = 0.5) = 0.13, P(λ = 1) = 0.82, P(λ ≥ 1.5) = 0.05, with the small probability in the upper tail roughly uniformly distributed. The posterior distributions were concentrated at or below λ = 1,
with low probability assigned to $\lambda > 2.0$, even though there were MLE $\hat{\lambda} > 2$. The scaled beta prior distribution produced the most diffuse posterior predictive distribution for $\lambda$.

The Bayesian models fit the data well, as assessed by plotted dose–response curves (plots in the supplemental materials) and posterior predictive values for summaries of the MLE’s. Simulation of the data from the portfolio of compounds based on the posterior distribution was performed for each multivariate parameter set created using the MCMC methods by (1) simulating response data for each dose group in each protocol using the $E_{\text{max}}$ model evaluated with the simulated parameters, (2) applying the same automated ML estimation procedure to the simulated data that was applied to the real data, (3) storing summaries of the MLE’s for each compound. Note that the posterior predictive median (approximately 0.86) for $\lambda$ based on the continuous prior distributions was lower than the median of the $\hat{\lambda}$, (1.13). The posterior predictive distributions for the median $\hat{\lambda}$ across compounds, however, were in good agreement with the observed median $\hat{\lambda}$. The means of the posterior predictive distribution for the median of $\hat{\lambda}$ were 1.07, 1.14, and 1.13, for the log $t_3$, scaled beta, and Dirichlet prior distributions, respectively. Other percentiles of the $\hat{\lambda}$ were also reproduced by simulations from the posterior distributions. The relationship between the central tendency of the ML and Bayes estimates, and a more subtle multivariate sensitivity to the prior distribution are detailed in Section 4.4.

4.3 The ED$_{50}$ Parameter

The ED$_{50}$ from different compounds range from micrograms to decigrams, but much of this variation can be predicted from preclinical and early-stage clinical data before initiating clinical dose–finding studies. The variation in the ED$_{50}$ about these preliminary predictions is described in this section. The log transformation of the ED$_{50}$ is used throughout. The early ED$_{50}$ predictions were not consistently recorded, and confidence intervals for them were not typically computed. The predictions were approximated by the mid-point between the two lowest nonzero doses in the first dose–response study, because dose–response designs should have at least one dose below the ED$_{50}$, and clinicians are reluctant to include many doses that are predicted to have inadequate efficacy. The approximated preliminary predictions of the ED$_{50j}$ are denoted by $P_{50j}$. Figure 6 displays a $t_3$ quantile plot of the $\log(\hat{E}_{D50j}/P_{50j})$, where $\hat{E}_{D50j}$ are the MLE from the hyperbolic model. The $P_{50j}$ were well calibrated, as the line through the quantile plot has mid-point close to (0, 0). Because the $\lambda_j$ were concentrated near 1.0 in Section 4.2,
the MLE and posterior distribution of the ED_{50j} parameters are evaluated first with \( \lambda_j = 1 \). The consequences of joint estimation of all of the parameters are described in Section 4.4.

Despite estimation errors and potential differences between the \( P_{50} \) and the actual early predictions, the \( \log(ED_{50j}) \) were within a 2.3-fold multiple of the \( \log(P_{50j}) \) with only three exceptions. The implied interval is equivalent to \( (0.1 P_{50} < ED_{50} < 10 P_{50}) \). The compound (ID 3) whose \( \hat{ED}_{50j} \) was far below its preliminary prediction was confirmed as a real error. It was not due to an unstable MLE or a poor approximation of the preliminary prediction by the \( P_{50} \). The prior distribution for the \( ED_{50j} \) in Section 4.2 was a \( t_3 \) distribution for \( \log(ED_{50j}/P_{50j}) \) with mean 0 and conservatively selected scale parameter, 2.0.

The distribution of the \( \log(ED_{50j}/P_{50j}) \) across the portfolio of compounds was modeled by a \( t_3 \) distribution with mean, \( \mu_{50j} \), and scale, \( \sigma_{50j} \). A diffuse prior distribution was specified using a normal distribution for \( \mu_{50j} \) with mean 0 and SD of 10, along with a Gamma distribution with shape and rate parameters of 0.001 for \( \sigma_{50j} \). A second more informative prior distribution was specified with \( \mu_{50j} \) assigned a mean of 0 and SD of \( \log(2) \). A corresponding scaled chi prior distribution was specified for \( \sigma_{50j} \) with \( \log(2)/\sigma_{50j} \sim \chi_{df}/\sqrt{df} \), and \( df = 2 \). The more informative prior distribution is based on an internal pharmacometric objective to predict the \( ED_{50j} \) within a multiple of 2 using the preliminary data sources.

The (5th, 10th, 25th, 50th, 75th, 90th, 95th) percentiles from the posterior predictive distribution for the \( ED_{50j}/P_{50j} \) based on the less informative prior distribution were \( (0.29, 0.49, 0.93, 1.55, 2.57, 4.82, 7.98) \), and the percentiles based on the more informative prior distribution were \( (0.28, 0.46, 0.86, 1.49, 2.54, 4.69, 7.78) \). The percentiles of the \( \hat{ED}_{50j}/P_{50j} \) were \( (0.17, 0.38, 0.66, 1.11, 2.34, 5.82, 10) \). In the absence of an interval derived from compound-specific data, \( (P_{50j}/10, 10 P_{50j}) \) provides a conservative 90% interval for the initial \( ED_{50j} \).

### 4.4 Prior Sensitivity and Posterior Predictive Values

The three hierarchical distributions for \( \lambda_j \) in Section 4.2 were combined with the two for the \( ED_{50j} \) in Section 4.3 to jointly model the variation in these parameters. The prior distributions for the \( E_{0j/k} \) and \( E_{\text{max}} \) parameters were unchanged. The posterior predictive distributions of \( \lambda \) and \( ED_{50j} \) from the combinations based on the log \( t_j \) and scaled beta distributions for \( \lambda \) are summarized in Table 2. Trends for the posterior distributions based on the discrete prior distribution of \( \lambda_j \) were consistent with those from the continuous prior distributions and are not reported. The percentiles in Table 2 varied depending on the differing prior distributions, and when compared to the posterior distributions in Sections 4.2 and 4.3. Two consistent trends were (1) the highest posterior density region had \( \lambda < 1.0 \), and (2) the probability that \( \lambda > 1.5 \) was low, with a maximum value of 0.09.

The other pronounced trend was the increased central tendencies of the \( ED_{50j}/P_{50j} \) when compared to results in Section 4.3 based on the hyperbolic model. The increased \( ED_{50j} \) were associated with decreased values of the \( \lambda \). The central tendencies of the posterior distributions of the compound-specific \( |E_{\text{max}}| \), which are not displayed, were also increased. With small treatment effects and limited dosing ranges, there was a multivariate indeterminacy in the likelihood/posterior distribution, with decreasing values of \( \lambda \) compensated by increased values of the \( ED_{50j} \) and \( |E_{\text{max}}| \). This indeterminacy is also common when fitting sigmoid \( E_{\text{max}} \) models to data for a specific compound (Bornkamp 2012; Woodward 2012). All of the posterior distributions yielded similar dose–response curves over the observed dosing ranges, and each curve described the data well. Plots of the dose–response curves derived from the scaled beta prior distribution for \( \lambda \) and the more informative distribution for \( ED_{50j} \) are included in the supplemental summaries. Simulations from the posterior predictive distributions of the percentiles of the \( \hat{\lambda} \) and \( ED_{50j} \) reproduced the MLE’s as described at the end of Section 4.2, even though the central tendencies

---

Table 2. Posterior predictive distributions for \( \lambda \) and \( ED_{50j} \). The MI and LI refer to the more and less informative prior distributions for \( ED_{50j} \). The B and L refer to the scaled beta and log \( t_j \) distributions for \( \lambda \).

| Percentile | Prior ED_{50j} MI | \( \lambda \) | Prior ED_{50j} LI | \( \lambda \) | \( ED_{50j}/P_{50j} \) B | \( ED_{50j}/P_{50j} \) LI |
|------------|------------------|-------------|------------------|-------------|-----------------|-----------------|
| 0.05       | 0.26             | 0.23        | 0.48             | 0.35        | 0.52            | 1.11            |
| 0.10       | 0.35             | 0.31        | 0.58             | 0.45        | 0.81            | 1.52            |
| 0.25       | 0.54             | 0.48        | 0.73             | 0.59        | 1.37            | 2.24            |
| 0.50       | 0.79             | 0.72        | 0.86             | 0.76        | 2.20            | 3.10            |
| 0.75       | 1.11             | 1.02        | 0.98             | 0.93        | 3.44            | 3.97            |
| 0.90       | 1.45             | 1.35        | 1.16             | 1.17        | 5.65            | 5.07            |
| 0.95       | 1.68             | 1.58        | 1.36             | 1.44        | 8.41            | 6.28            |
of the distributions in Table 2 were further from those of the MLE's. Increasing the prior variance of the $|E_{\text{max}}|$ combined with the more diffuse prior distribution for the $ED_{50}$ yielded yet more pronounced trends toward lower values of $\lambda$ and higher values of $ED_{50}$.

The relationship between the compound-specific MLE's and the Bayesian estimation was also assessed by simulation of the portfolio. The simulation model parameters, $(E_{0,jk}, ED_{50,j}, E_{\text{max},j})$, were set equal to $(\hat{E}_{0,jk}, \hat{ED}_{50,j}, \hat{E}_{\text{max},j})$ from the hyperbolic $E_{\text{max}}$ models. When the data were simulated with $\lambda_j = 1$ for each compound, the means over 5000 simulated portfolios of the 25th, 50th, and 75th percentiles of the MLE's of the $\lambda_j$ were 0.87, 1.16, 1.71, which were in close agreement with observed $\hat{\lambda}_j$ percentiles, 0.85, 1.13, and 1.61, respectively. When the simulations were repeated with $\lambda_j = 1.5$ in the simulation model, the means of the percentiles from the simulated portfolios were 1.23, 1.6, 2.17. The proportion of simulated portfolios yielding percentiles above those from the actual $\hat{\lambda}_j$ was $> 0.99$ when $\lambda_j$ was fixed at 1.5. The converged ML estimates of $\lambda_j$ were biased upward.

The simulation of the portfolio was repeated and the posterior distributions of the parameters were computed based on the $t_3$ prior distribution for $\log(\lambda)$; computation time limited the number of replications to 200. When $\lambda$ was set to 1 in the simulation model, the mean across the 200 simulation replications of the posterior predictive median for a new $\lambda$ was 0.94 for the more informative prior distribution of the $ED_{50}$, and 0.92 for the less informative prior distribution. When $\lambda$ was set to 1.5, the means of the posterior median were 1.4 and 1.39. The portfolio of compounds was also simulated with the dosing designs for each compound unchanged, but with the sample sizes per dose increased proportionally. Consistent with asymptotic theory, the Bayes posterior means and medians for $\lambda$ approached the fixed population values with increasing sample sizes, but the approach to the population $\lambda$ was monotone from below.

5. Summary

The combination of low signal-to-noise, a small number of doses, and limited dosing range often obscured the dose–response shape in individual studies, but when aggregated across studies and compounds, a consistent pattern of dose–response was present, and it was well described by the hyperbolic $E_{\text{max}}$ model (Question 1, Section 1). In addition to displaying a consistent dose–response pattern, data from multiple protocols for the same compound were consistent, despite variation in placebo response. The consistent $E_{\text{max}}$ pattern may not generalize to other contexts. Gries et al. (1999), for example, described an example of a biologic compound with nonmonotone efficacy determined by its receptor binding. A meta-analysis of dose–response in toxicology studies by Calabrese and Baldwin (2001) suggested nonmonotone dose–response was more common.

The dosing range and the number of doses tested were inadequate to characterize the dose–response for most compounds (Question 2, Section 1). More than 1/3 of the compounds were tested on three or four nonplacebo doses, and more than 1/2 had a maximum dosing ratio less than 20. For some compounds, the limited dosing range was due to safety concerns, as noted for example, by Dutta, Matsumoto, and Ebling (1995), but even when there were safety concerns, investment of more resources in toxicology studies could sometimes expand the potential dosing range. Lower doses were also omitted from many studies. The proportion of compounds with the final $ED_{50}$ from the hyperbolic $E_{\text{max}}$ model below the lowest initial dose was 0.2. Although not captured in our data collection, the availability of different manufactured doses may be as important as safety and efficacy considerations when determining the dosing designs. Dosing ranges less than 20-fold are dubious given the uncertainty in the $ED_{50}$ at the initiation of dose–finding studies, and the nearly 100-fold range in doses between minimal and maximal response predicted by the hyperbolic $E_{\text{max}}$ model. The inadequacy of the designs is further supported by the frequent need to conduct multiple studies with additional doses.

Patterns in dose–response were summarized by the distributions of the $E_{\text{max}}$ model parameters (Question 3, Section 1). The modeling showed the Hill parameters were concentrated near 1.0. The assessment of this distribution between approximately (0.25, 1.5) is difficult because of the lack of identification within this range. Steeper dose–response curves represented by $\lambda > 2$ had low posterior predicted probability. This low probability depends on the assumption that our portfolio of compounds was a random sample from the collection of compounds that will be sampled by future drug developers. Dutta, Matsumoto, and Ebling (1995) reported without details on a meta-analysis yielding typical $\lambda$ parameters closer to 2.0. Holford and Sheiner (1981) summarized an example from Winkle et al. (1976) and Meffin et al. (1977) that reported an estimated $\lambda$ much greater than 2.0. All of these authors were evaluating concentration response rather than dose–response, so attenuation of the curves due to the use of dose rather than concentration might partially explain the steeper curves they observed.

These considerations suggest using the most diffuse of the posterior predictive distributions for $\lambda$. Alternatively, the hyperbolic model could be justified as the simplest model consistent with most past data, and it could be used unless the model is rejected by goodness of fit testing. The $\lambda$ and $ED_{50}$ parameters
for a new compound were approximately independent for each posterior predictive distribution in Table 2. The approximate independence was found even though the parameters for a compound with observed dosing data are sometimes highly dependent in their posterior distribution. The most diffuse posterior predictive distribution for λ (first column) can be approximated over a broad range of values by a beta distribution scaled to (0, 6) with parameters 3.03 and 18.15, computed by matching moments. The corresponding distribution for \( \log(ED_{50}/P_{50}) \) can be approximated by a \( t_3 \) distribution with mean 0.79 and scale parameter 0.6. In addition to their potential use in analyses, the distributions for \( (ED_{50}, \lambda) \) also provide an empirical quantitative basis for model-based designs such as those in Dette et al. (2008).

The accuracy of the preliminary \( ED_{50} \) estimates was conservatively assessed because the predicted \( ED_{50} \) had to be inferred from the initial dosing design. The methods and data used for the preliminary estimates vary between compounds and therapeutic areas, so there could be important differences in methodology between sponsors. When a compound-specific estimate of precision is available, this information will dominate. Even in these situations, the results in Section 4 provide a useful check on the claimed precision, which can be difficult to assess.

The results in Section 4.4 displayed the multivariate sensitivity of the \( E_{\text{max}} \) model parameter estimation to the specification of the prior distribution. The common practice of specifying diffuse prior distributions for the \( ED_{50} \) and \( E_{\text{max}} \) parameters when the \( \lambda \) parameter is also estimated favors model fits with smaller \( \lambda \) and larger \( ED_{50} \) and \( |E_{\text{max}}| \). Our experience has been that the preliminary estimates of the \( ED_{50} \) have utilized the hyperbolic \( E_{\text{max}} \) model, which might explain why the preliminary \( ED_{50} \) were best calibrated with the \( ED_{50} \) estimated from the hyperbolic \( E_{\text{max}} \) model.

There was one compound in the portfolio with nonmonotone dose–response. Despite demonstrated efficacy for the compound and another in same therapeutic class, both compounds were terminated. Replication of this and the other findings here using dose–response studies conducted by other large sponsors would be helpful. While graphical inspection and goodness of fit testing are warranted for all compounds, routine planning for nonmonotone dose–response was not indicated by the data in a large and diverse portfolio, or by the likely utility of compounds with nonmonotone efficacy.

### Supplementary Materials

Summaries and additional information on the compounds discussed in the article.

[Received October 2013. Revised March 2014]

### References

Bates, D., and Watts, D. (1988), *Nonlinear Regression Analysis and Its Applications*, New York: Wiley. [308]

Berry, D., Muller, P., Grieve, A., Smith, M., Parke, T., Blazek, R., Mitchard, N., and Krams, M. (2002), “Adaptive Bayesian Designs for Dose-Ranging Trials,” in *Case Studies in Bayesian Statistics V*, eds. B. Carlin, A. Carriquiry, C. Gatsonis, et al., New York: Springer, pp. 99–181. [303]

Bornkamp, B. (2012), “Functional Uniform Priors for Nonlinear Modeling,” *Biometrics*, 68, 893–901. [314]

Bornkamp, B., and Ickstadt, K. (2009), “Bayesian Nonparametric Estimation of Continuous Monotone Functions With Applications to Dose–Response Analysis,” *Biometrics*, 65, 198–205. [303]

Bornkamp, B., Bretz, F., Dette, H., and Pinheiro, J. (2011), “Response-Adaptive Dose-Finding Under Model Uncertainty,” *The Annals of Applied Statistics*, 5, 1611–1631. [303]

Bretz, F., Hothorn, L., and Hsu, J. (2003), “Identifying Effective and/or Safe Doses by Stepwise Confidence Intervals for Ratios,” *Statistics in Medicine*, 22, 847–858. [303]

Bretz, F., Pinheiro, J., and Branson, M. (2005), “Combining Multiple Comparisons and Modeling Techniques in Dose–Response Studies,” *Biometrics*, 61, 738–748. [303]

Bretz, F., Dette, H., and Pinheiro, J. (2010), “Practical Considerations for Optimal Designs in Clinical Dose–Finding Studies,” *Statistics in Medicine*, 29, 731–742. [303]

Calabrese, E., and Baldwin, L. (2001), “The Frequency of u-Shaped Dose Responses in the Toxicology Literature,” *Toxicological Sciences*, 62, 330–338. [315]

Congressional Budget Office (2006), *Research and Development in the Pharmaceutical Industry*, Congressional Budget Office Study, available at [http://www.cbo.gov/sites/default/files/cbofiles/fpd/docs/7615/10-02-drugr-d.pdf] [303]

Cook, S., Gelman, A., and Rubin, D. (2006), “Validation of Software for Bayesian Models Using Posterior Quantiles,” *Journal of Computational and Graphical Statistics*, 15, 675–690. [312]

Dette, H., Bretz, F., Pepelyshev, A., and Pinheiro, J. (2008), “Optimal Designs for Dose-Finding Studies,” *Journal of the American Statistical Association*, 103, 1226–1237. [303,316]

Dette, H., Kiss, C., Bevanda, M., and Bretz, F. (2010), “Optimal Designs for the Emax, Log-Linear and Exponential Models,” *Biometrika*, 97, 513–518. [303]

Dunnett, C. (1955), “A Multiple Comparison Procedure for Comparing Several Treatments With a Control,” *The Journal of the American Statistical Association*, 50, 1096–1121. [303]

Dutta, S., Matsumoto, Y., and Ebling, W. (1995), “Is It Possible to Estimate the Parameters of the Sigmoid E_{max} Model With the Truncated Data Typical of Clinical Studies?” *Journal of Pharmaceutical Sciences*, 85, 232–239. [315]

EP Vantage, Inc. (2010), “Pharma Industry Cutting Its R&D Cloth to More Sustainable Levels,” available at [http://www.evaluategroup.com/UniversalView.aspx?type=Story&id=217946&sectionID=&isEVPantage=yes] [303]

Gries, J., Munafo, A., Porchett, H., and Verotta, D. (1999), “Down-Regulation Models and Modeling of Testosterone Production Induced by Recombinant Human Choriogonadotropin,” *The Journal of Pharmacology and Experimental Therapeutics*, 289, 371–377. [315]

Holford, N., and Sheiner, L. (1981), “Understanding the Dose–Response Relationship: Clinical Application of Pharmacokinetic-Pharmacodynamic Models,” *Clinical Pharmacokinetics*, 6, 429–453. [315]
Meta-Analysis of Clinical Dose–Response in a Large Drug Development Portfolio

Hsu, J., and Berger, R. (1999), “Stepwise Confidence Intervals Without Multiplicity Adjustment for Dose–Response and Toxicity Studies,” Journal of the American Statistical Association, 94, 468–483. [303]

Iliadis, A., and Macheras, P. (2006), Modeling in Biopharmaceutics, and Pharmacodynamics: Homogeneous and Heterogeneous Approaches, New York: Springer. [307]

Johnson, N., and Kotz, S. (1995), Continuous Univariate Distributions (vol. 2), New York: Wiley. [307]

Kahn, M., and Raftery, A. (1996), “Discharge Rates of Medicare Stroke Patients to Skilled Nursing Facilities: Bayesian Logistic Regression With Unobserved Heterogeneity,” Journal of the American Statistical Association, 91, 29–41. [312]

Krishna, R. (2006), Dose Optimization in Drug Development, Boca Raton, FL: CRC Press. [303]

Kroon, J. (1997), “Application of the 𝜀 𝜖 𝜁 𝜆 𝜀 𝜅 𝜀 𝜁 𝜆 𝜉 𝜅 𝜆 𝜅 𝜁 𝜆 𝜅 𝜁 𝜆 𝜅 𝜁 𝜆 𝜅 𝜁 𝜆 𝜅 𝜁 𝜆 𝜅 𝜁 𝜆 𝜅 𝜁 𝜆 𝜅 𝜁 𝜆 𝜅 𝜁 𝜆 𝜅 𝜁 𝜆 𝜅 𝜁 𝜆 𝜅 𝜁 𝜆 𝜅 𝜁 𝜆 𝜅 𝜁 𝜆 𝜅 𝜁 𝜆 𝜅 𝜁 𝜆 𝜅 𝜁 𝜆 𝜅 𝜁 𝜆 𝜅 𝜁 𝜆 𝜅 𝜁 𝜆 𝜅 𝜁 𝜆 𝜅 𝜁 𝜆 𝜅 𝜁 𝜆 𝜅 𝜁 𝜆 𝜅 𝜁 𝜆 𝜅 𝜁 𝜆 𝜅 𝜁 𝜆 𝜅 𝜁 𝜆 𝜅 𝜁 𝜆 𝜅 𝜁 𝜆 𝜅 𝜁 𝜆 𝜅 𝜁 𝜆 𝜅 𝜁 𝜆 𝜅 𝜁 𝜆 𝜅 𝜁 𝜆 𝜅 𝜁 𝜆 𝜅 𝜁 𝜆 𝜅 𝜁 𝜆 𝜅 𝜁 𝜆 𝜅 𝜁 𝜆 𝜅 𝜁 𝜆 𝜅 𝜁 𝜆 𝜅 𝜁 𝜆 𝜅 𝜁 𝜆 𝜅 𝜁 𝜆 𝜅 𝜁 𝜆 𝜅 𝜁 𝜆 𝜅 𝜁 𝜆 𝜅 𝜁 𝜆 𝜅 𝜁 𝜆 𝜅 𝜁 𝜆 𝜅 𝜁 𝜆 𝜅 𝜁 𝜆 𝜅 𝜁 𝜆 𝜅 𝜁 𝜆 𝜅 𝜁 𝜆 𝜅 𝜁 𝜆 𝜅 𝜁 𝜆 𝜅 𝜁 𝜆 𝜅 𝜁 𝜆 𝜅 𝜁 𝜆 𝜅 𝜁 𝜆 𝜅 𝜁 𝜆 𝜅 𝜁 𝜆 𝜅 𝜁 𝜆 𝜅 𝜁 𝜆 𝜅 𝜁 𝜆 𝜅 𝜁 𝜆 𝜅 𝜁 𝜆 𝜅 𝜁 𝜆 𝜅 𝜁 𝜆 𝜅 𝜁 𝜆 𝜅 𝜁 𝜆 𝜅 𝜁 𝜆 𝜅 𝜁 𝜆 𝜅 𝜁 𝜆 𝜅 𝜁 𝜆 𝜅 𝜁 𝜆 𝜅 𝜁 𝜆 𝜅 𝜁 𝜆 𝜅 𝜁 𝜆 𝜅 𝜁 𝜆 𝜅 𝜁 𝜆 𝜅 𝜁 𝜆 𝜅 𝜁 𝜆 𝜅 𝜁 𝜆 𝜅 𝜁 𝜆 𝜅 𝜁 𝜆 𝜅 𝜁 𝜆 𝜅 𝜁 𝜆 𝜅 𝜁 𝜆 𝜅 𝜁 𝜆 𝜅 𝜁 𝜆 𝜅 𝜁 𝜆 𝜅 𝜁 𝜆 𝜅 𝜁 𝜆 𝜅 𝜁 𝜆 𝜅 𝜁 𝜆 𝜅 𝜁 𝜆 𝜅 𝜁 𝜆 𝜅 𝜁 𝜆 𝜅 𝜁 𝜆 𝜅 𝜁 𝜆 𝜅 𝜁 𝜆 𝜅 𝜁 𝜆 𝜅 𝜁 𝜆 𝜅 𝜁 𝜆 𝜅 𝜁 𝜆 𝜅 𝜁 𝜆 𝜅 𝜁 𝜆 𝜅 𝜁 𝜆 𝜅 𝜁 𝜆 𝜅 𝜁 𝜆 𝜅 𝜁 𝜆 𝜅 𝜁 𝜆 𝜅 𝜁 𝜆 𝜅 𝜁 𝜆 𝜅 𝜁 𝜆 𝜅 𝜁 𝜆 𝜅 𝜁 𝜆 𝜅 𝜁 𝜆 𝜅 𝜁 𝜆 𝜅 𝜁 𝜆 𝜅 𝜁 𝜆 𝜅 𝜁 𝜆 𝜅 𝜁 𝜆 𝜅 𝜁 𝜆 𝜅 𝜁 𝜆 𝜅 𝜁 𝜆 𝜅 𝜁 𝜆 𝜅 𝜁 𝜆 𝜅 𝜁 𝜆 𝜅 𝜁 𝜆 𝜅 𝜁 𝜆 𝜅 𝜁 𝜆 𝜅 𝜁 𝜆 𝜅 𝜁 𝜆 𝜅 𝜁 𝜆 𝜅 𝜁 𝜆 𝜅 𝜁 𝜆 𝜅 𝜁 𝜆 𝜅 𝜁 𝜆 𝜅 𝜁 𝜆 𝜅 𝜁 𝜆 𝜅 𝜁 𝜆 𝜅 𝜁 𝜆 𝜅 𝜁 𝜆 𝜅 𝜁 𝜆 𝜅 𝜁 𝜆 𝜅 𝜁 𝜆 𝜅 𝜁 𝜆 𝜅 𝜁 𝜆 𝜅 𝜁 𝜆 𝜅 𝜁 𝜆 𝜅 𝜁 𝜆 𝜅 𝜁 𝜆 𝜅 𝜁 𝜆 𝜅 𝜁 𝜆 𝜅 𝜁 𝜆 𝜅 𝜁 𝜆 𝜅 𝜁 𝜆 𝜅 𝜁 𝜆 𝜅 𝜁 𝜆 𝜅 𝜁 𝜆 𝜅 𝜁 𝜆 𝜅 𝜁 𝜆 𝜅 𝜁 𝜆 𝜅 𝜁 𝜆 𝜅 𝜁 𝜆 𝜅 𝜁 𝜆 𝜅 𝜁 𝜆 𝜅 𝜁 𝜆 𝜅 𝜁 𝜆 𝜅 𝜁 𝜆 𝜅 𝜁 𝜆 𝜅 𝜁 𝜆 𝜅 𝜁 𝜆 𝜅 𝜁 𝜆 𝜅 𝜁 𝜆 𝜅 𝜁 𝜆 𝜅 𝜁 𝜆 𝜅 𝜁 𝜆 𝜅 𝜁 𝜆 𝜅 𝜁 𝜆 𝜅 𝜁 𝜆 𝜅 𝜁 𝜆 𝜅 𝜁 𝜆 𝜅 𝜁 𝜆 𝜅 𝜁 𝜆 𝜅 𝜁 𝜆 𝜅 𝜁 𝜆 𝜅 𝜁 𝜆 𝜅 𝜁 𝜆 𝜅 𝜁 𝜆 𝜅 𝜁 𝜆 𝜅 𝜁 𝜆 𝜅 𝜁 𝜆 𝜅 𝜁 𝜆 𝜅 𝜁 𝜆 𝜅 𝜁 𝜆 𝜅 𝜁 𝜆 𝜅 𝜁 𝜆 𝜅 𝜁 𝜆 𝜅 𝜁 𝜆 𝜅 𝜁 𝜆 𝜅 𝜁 COVID-19 and Public Health in a Changing World瞒

R Development Core Team (2011), R: A Language and Environment for Statistical Computing, Vienna, Austria: R Foundation for Statistical Computing. [307]

Ross, G. (1990), Nonlinear Estimation, New York: Springer. [307]

Ruberg, S. (1989), “Dose-Reponse Studies II. Analysis and Interpretation,” Journal of Biopharmaceutical Statistics, 5, 15–42. [303]

Sheiner, L., Beal, S., and Dunne, A. (1997), “Analysis of Nonrandomly Censored Ordered Categorical Longitudinal Data From Analgesic Trials,” Journal of the American Statistical Association, 92, 1235–1244. [307]

Stan Development Team (2013), Stan: A C++ Library for Probability and Sampling, available at http://mc-stan.org/, 1.3 edition. [312]

Tamhane, A., Hochberg, Y., and Dunnett, C. (1996), “Multiple Test Procedures for Dose–Finding,” Biometrics, 52, 21–37. [303]

Tan, H., Gruben, D., French, J., and Thomas, N. (2011), “A Case Study of Model-Based Bayesian Dose–Response Estimation,” Statistics in Medicine, 30, 2622–2633. [307]

Thomas, N. (2006), “Hypothesis Testing and Bayesian Estimation Using a Sigmoid 𝐸max Model Applied to Sparse Dose–Response Designs,” Journal of Biopharmaceutical Statistics, 16, 657–677. [308]

Ting, N. (2006), Dose–Finding in Drug Development, New York: Springer. [303]

Tukey, J. (1977), Exploratory Data Analysis, Reading, MA: Addison-Wesley. [306]

Venables, W., and Ripley, B. (2002), Modern Applied Statistics With S (4th ed.), New York: Springer. [306]

Williams, D. (1971), “A Test of Differences Between Treatment Means When Several Doses are Compared With a Zero Dose,” Biometrics, 27, 103–117. [303]

Winkle, R., Meffin, P., Fitzgerald, J., and Harrison, D. (1976), “Clinical Pharmacokinetics and Tocainide,” Efficacy and Pharmacokinetics of a New Orally Effective Antiarrhythmic Agent,” Journal of the American Medical Association, 297, 1362–1373. [309]

About the Authors

Neal Thomas (E-mail: neal.thomas@pfizer.com) and Kevin Sweeney (E-mail: kevin.sweeney@pfizer.com), Pfizer, Inc., Eastern Point Road, Groton, CT 06340. Veena Somayaji, Pfizer, Inc., 620 Memorial Drive, Cambridge, MA 02139 (E-mail: veena.somayaji@pfizer.com).