Applied Modelling in Drug Development using `brms`

Joint Statistical Meetings, Portland, OR

David Ohlssen[1], Andrew Bean[1], Björn Holzhauer[2]

[1]Novartis Pharmaceuticals Corporation [2]Novartis Pharma AG

2024-08-05

NOVARTIS

Learning Goals

After this course, you should:

- Be familiar with `brms` syntax and workflow
- Recognize its versatility for statistical modelling in drug development
- Have hands-on experience with the package from two guided exercises

and of course:

- Feel empowered to use `brms` the future!

Housekeeping

- Q&A: you may raise your hand at any time, or hold for Q&A sessions at the end of each section
- Laptop charging: we recommend conserving battery by keeping your laptop powered down except during the hands-on exercises
- For hands-on exercises, we will use Posit Cloud: link to join our space
  - More instructions to come when we begin the first exercise
- Online case study library: http://opensource.nibr.com/bamdd
- Course materials: https://github.com/Novartis/bamdd/tree/main/workshops/jsm2024

Acknowledgements

- Paul Buerkner (Technical University of Dortmund)
- Sebastian Weber (Novartis)
- Lukas Widmer (Novartis)
- Cong Zhang (Novartis)
- Michael Mayer and Phil Bowsher (Posit)

Agenda

	item	time	start	end
Orientation (David)	Welcome	5	1:00 PM	1:05 PM
	Basics of Bayesian inference	15	1:05 PM	1:20 PM
	brms overview	15	1:20 PM	1:35 PM
Case study 1 (Andrew)	Historical control data + Q&A	20	1:35 PM	1:55 PM
	Hands-on exercises + wrap-up	25	1:55 PM	2:20 PM
Bayesian statistics in drug development (David)	Landscape: Bayes in drug development + Q&A	20	2:20 PM	2:40 PM
	Considerations for prior specification + Q&A	20	2:40 PM	3:00 PM
Break	Break	15	3:00 PM	3:15 PM
Case study 2 (Bjoern)	Bayesian Mixed Model for Repeated Measures (MMRM) + Q&A	25	3:15 PM	3:40 PM
Case study 3 (Bjoern)	Dose finding + Q&A	25	3:40 PM	4:05 PM
	Hands-on exercises	25	4:05 PM	4:30 PM
Case study 4 (Andrew)	Modelling time-to-event data + Q&A	25	4:30 PM	4:55 PM
Closing (Andrew)	Outlook	5	4:55 PM	5:00 PM

Bayesian inference basics

Review of Bayesian Inference

Probability distributions

Data Y, parameter(s) θ

1. Sampling distribution (statistical model, "likelihood")
   The distribution of Y conditional on θ
   $p(Y \mid \theta)$
2. Prior distribution of θ expresses knowledge about θ prior to observing data Y
   $p(\theta)$
3. Posterior distribution of θ expresses knowledge about θ after observing Y
   $p(\theta \mid Y)$

Bayes theorem: $p(\theta \mid Y) \propto p(Y \mid \theta) p(\theta)$

Posterior $\propto$ Likelihood × Prior

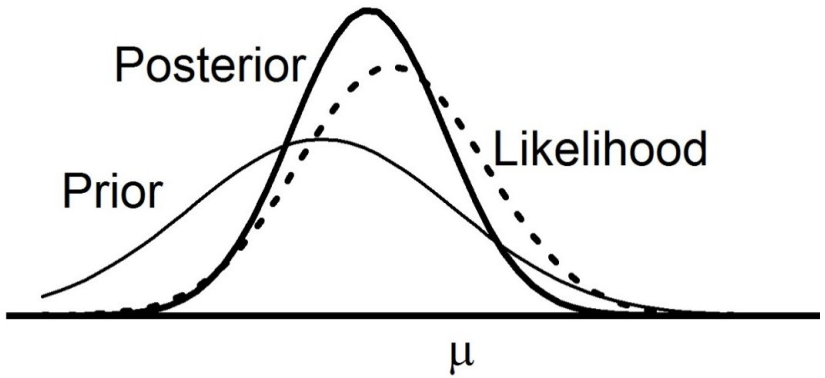
Review of Bayesian Inference

Normal data

Likelihood	$p(\text{data}\|\theta)$
Prior	$p(\theta)$
Posterior $\propto$ Likelihood $\times$ Prior	$p(\theta\|\text{data}) \propto p(\text{data}\|\theta) \times p(\theta)$

Example: Normal data with known σ

Likelihood	$p(\bar{Y}\|\mu) = N(\mu, \sigma^2/n)$
Prior	$p(\mu) = N(\mu_0, \sigma^2/n_0)$
Posterior $\propto$ Likelihood $\times$ Prior	$p(\mu\|\bar{Y}) = N(\{n_0\mu_0+n\bar{Y}\}/(n_0+n), \sigma^2/(n_0+n))$



As if n_0 additional patients with average response μ_0 had been included

n_0 = prior sample size

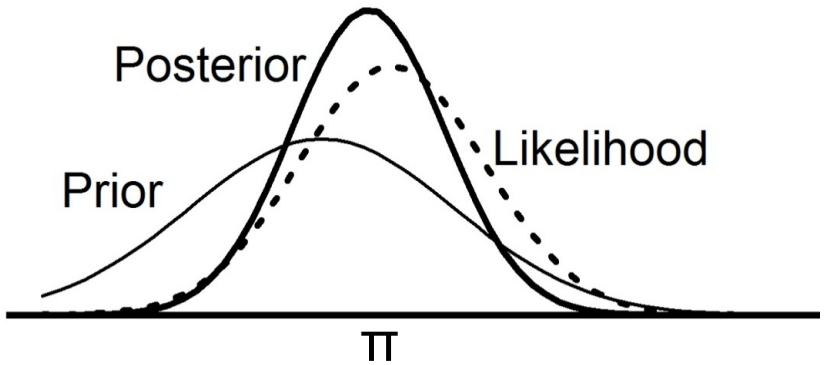
Review of Bayesian Inference

Binary data

Likelihood	$p(\text{data}\|\theta)$
Prior	$p(\theta)$
Posterior $\propto$ Likelihood $\times$ Prior	$p(\theta\|\text{data}) \propto p(\text{data}\|\theta) \times p(\theta)$

Example: Binary data

Likelihood	$p(Y\|\pi) = \text{Binomial}(\pi, n)$
Prior	$p(\pi) = \text{Beta}(a, b)$
Posterior $\propto$ Likelihood $\times$ Prior	$p(\pi\|Y) = \text{Beta}(a + Y, b + \{n-Y\})$



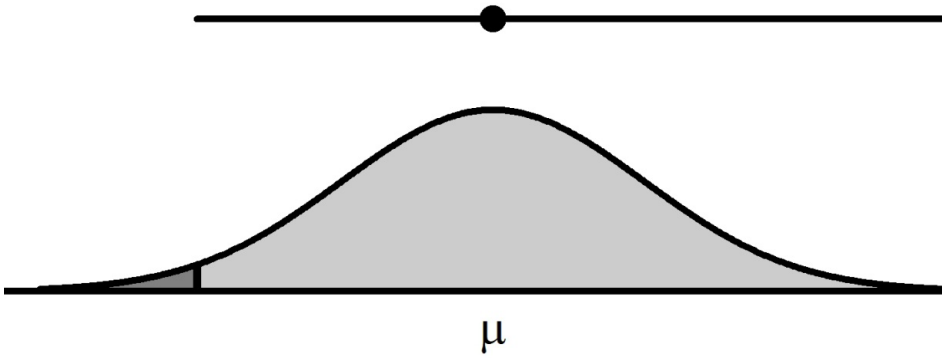
As if n_0=a+b additional patients with response rate a/n_0 had been included

n_0 = prior sample size

Review of Bayesian Inference

Bayesian inference vs classical inference

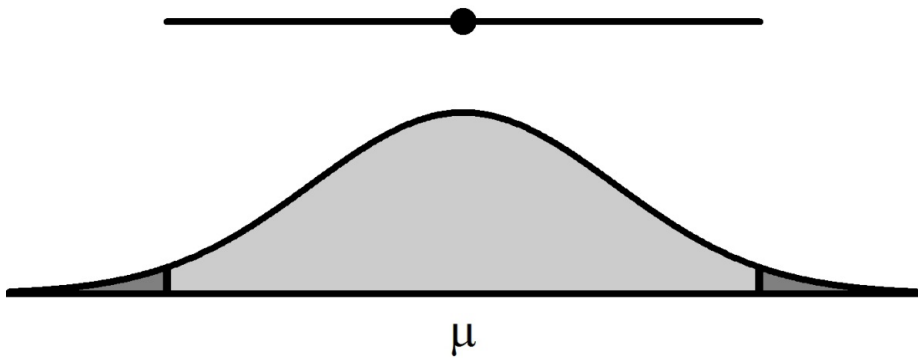
$\bar{Y}$=2.5, n=25 Normal distribution, σ=5 known, vague prior (n_0=0.001)



one-sided 97.5% CI

(0.54,∞)

$p(\mu>0.54 \mid \bar{Y})=97.5\%$



two-sided 95% CI

(0.54,4.46)

$p(0.54<\mu<4.46 \mid \bar{Y})=95\%$

NOVARTIS

Review of Bayesian Inference

Prediction

Data Y, parameter(s) θ , new data Y_* (planned)

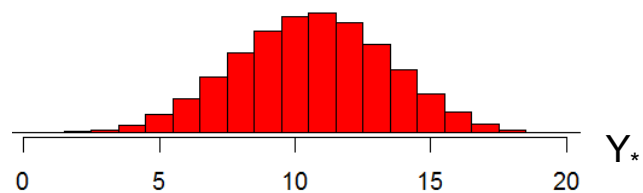
Predictive distribution of Y_* expresses knowledge about Y_* after observing Y, but before observing new data Y_*

$$p(Y_* \mid Y) = \int p(Y_* \mid \theta)\, p(\theta \mid Y)\, d\theta$$

Example – clinical trial in 50 cancer patients

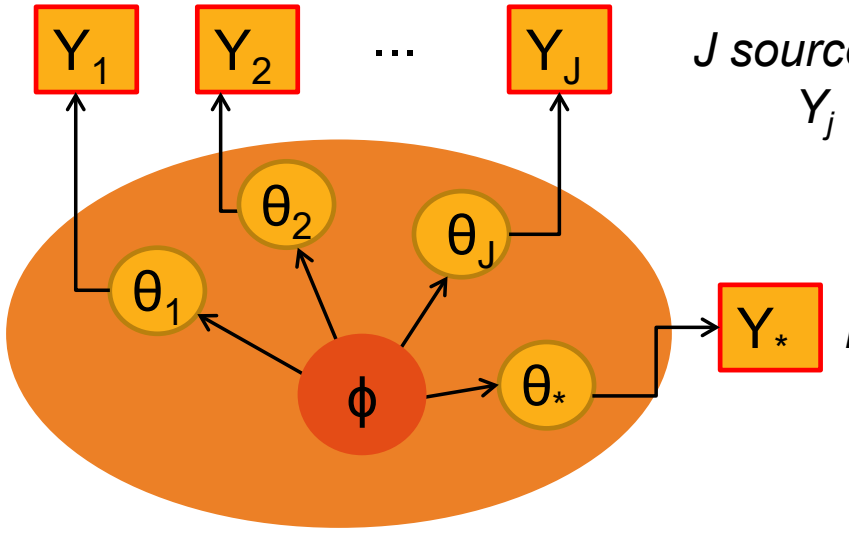
- All patients receive the test treatment, outcome=response yes/no
- Test treatment will be further investigated if at least 30/50 respond
- Y=16 of the first 30 patients responded. What is the probability that at least 14 of the next 20 patients respond?

Prior response rate Beta(1,1)
Posterior Beta(17,15)
Predictive distribution $p(Y^* \mid Y=16)$
$p(Y^* \geq 14 \mid Y=16) = 16\%$



Review of Bayesian Inference

Evidence synthesis – general statistical model



J sources of historical data

$Y_j \mid \theta_j \sim F_j(\theta_j)$

$p(Y_j \mid \theta_j)$

New source of data (plan)

$Y_* \mid \theta_* \sim F_*(\theta_*)$

$p(Y_* \mid \theta_*)$

Model to link parameters (hyper-parameter ϕ)

$(\theta_*, \theta_1, \ldots, \theta_J) \mid \phi \sim G(\phi)$

$p(\theta_*, \theta_1, \ldots, \theta_J \mid \phi)$

Bayesian inference on unknowns $Y_*, \theta_*, \theta_1, \ldots, \theta_J, \phi$

Review of Bayesian Inference

Bayesian computation

Many parameters $\theta = (\theta_1, \ldots, \theta_p)$ (p may be >> 100)

Posterior distribution:

$$p(\theta \mid Y) = p(Y \mid \theta)\, p(\theta) / \int p(Y \mid \theta)\, p(\theta)\, d\theta$$

Marginal posterior distribution:

$$p(\theta_1 \mid Y) = \int p(\theta \mid Y)\, d\theta_2 \ldots d\theta_p$$

Predictive distribution:

$$p(Y_* \mid Y) = \int p(Y_* \mid \theta)\, p(\theta \mid Y)\, d\theta$$

Requires high-dimensional integration

- Analytical evaluation for simple cases only
- Numerical integration for low dimensions only

Review of Bayesian Inference

Bayesian computation

Bayesian data analysis revolutionized by Markov Chain Monte Carlo - MCMC (Gelfand and Smith,1990)

Generate a very large sample from the posterior distribution, without need to know $\int p(Y \mid \theta) p(\theta) d \theta$

$\theta^{(1)}, \ldots, \theta^{(M)}$ (e.g. M=10‘000) $\theta^{(k)} = (\theta_1^{(k)}, \ldots, \theta_p^{(k)})$

Posterior distribution ≈ Empirical distribution of sample

e.g. $p(\theta_1 \mid Y) \approx$ empirical distribution of $\theta_1^{(1)}, \ldots, \theta_1^{(M)}$

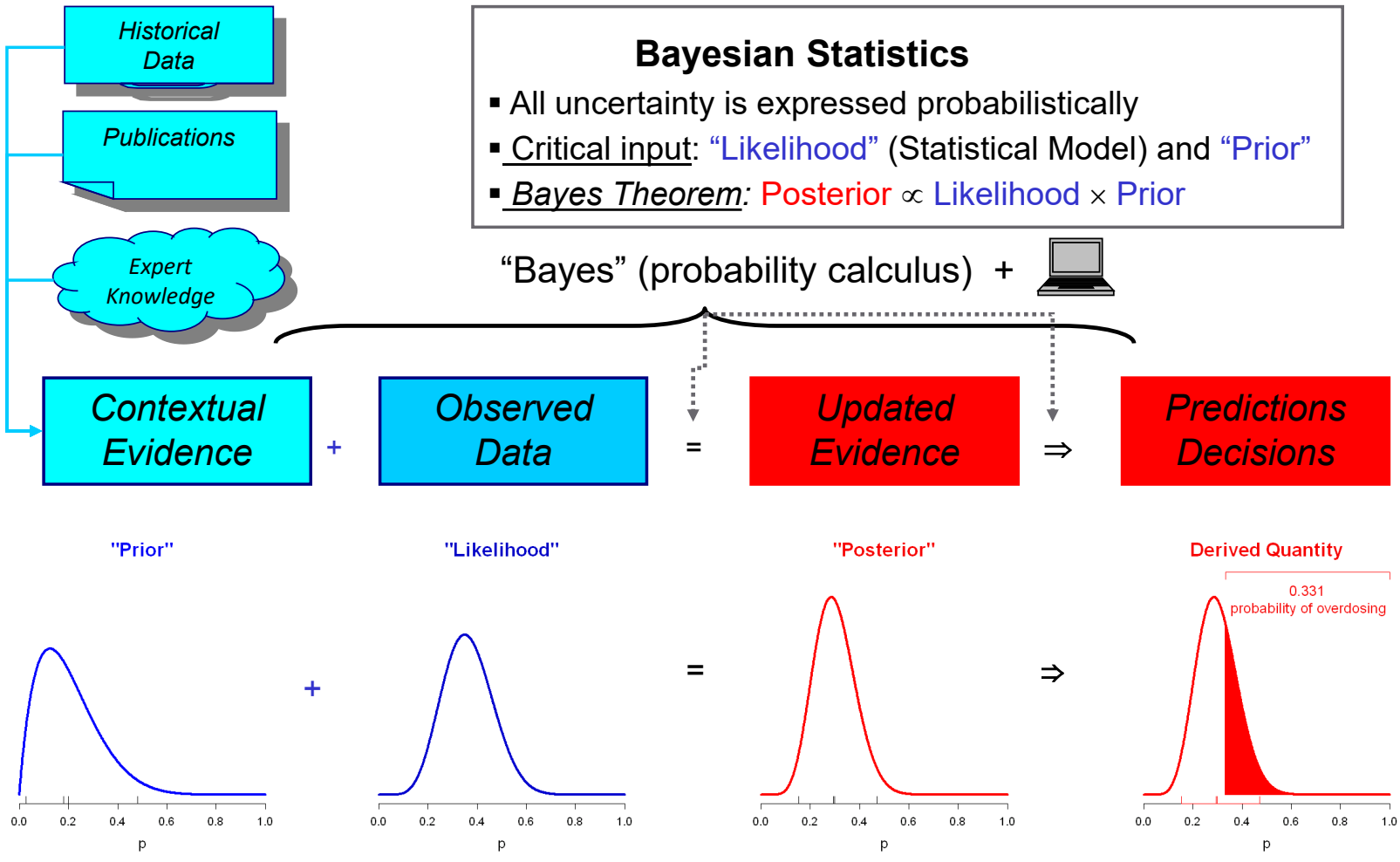
$p(g(\theta) \mid Y \approx$ empirical distribution of $g(\theta^{(1)}), \ldots, g(\theta^{(M)})$

Software

- WinBUGS, JAGS, Stan, **brms** ... Nimble, Turing, PyMC3, etc.
- SAS

Bayesian Statistics

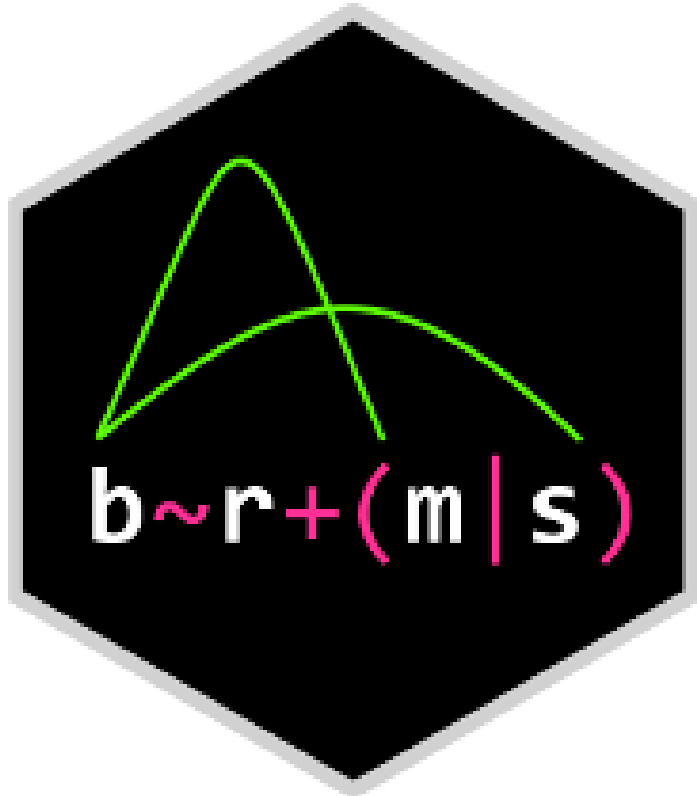
Summary



NOVARTIS

A `brms` modelling workflow

Bayesian Software: brms



- Specify models via extended R formula syntax
- Internally write Stan code that is readable yet fast
- Provide an easy interface for defining priors

NOVARTIS

Some Highlights of brms

- Flexible hierarchical (random effects) modeling
- Both built-in and user-defined likelihoods
- Explicit and implicit non-linear modeling
- Distributional regression
- Within-chain parallelization
- Posterior and prior predictions
- Highly dense feature matrix

Model specification in brms: `formula`

varying intercept model for a single grouping factor:

```
formula = y ~ 1 + x + (1 | g)
```

Varying intercept-slope model for a single grouping factor:

```
formula = y ~ 1 + x + (1 + x | g)
```

Advanced non-linear terms such as Gaussian processes:

```
formula = y ~ 1 + gp(x) + (1 + x | g)
```

Model specification in brms: `formula`

Linear formulas for multiple distributional parameters (e.g., predict mean and overdispersion of negative binomial):

```
formula = bf(
  y ~ 1 + x + (1 | g) + ...,
  par2 ~ 1 + x + (1 | g) + ...,
  par3 ~ 1 + x + (1 | g) + ...,
)
```

Non-linear formula for a single distributional parameter:

```
formula = bf(
  y ~ fun(x, nlpar1, nlpar2),
  nlpar1 ~ 1 + x + (1 | g) + ...,
  nlpar2 ~ 1 + (1 | g) + ...,
  nl = TRUE
)
```

Model specification in brms: `family` (likelihood)

General structure:

```
family = brmsfamily(
  family = "<family>", link = "<link>",
  more_link_arguments
)
```

Gaussian likelihood (default):

```
family = brmsfamily(family = "gaussian", link = "identity",
                    link_sigma = "log")
```

Poisson likelihood:

```
family = brmsfamily(family = "poisson", link = "log")
```

See also `vignette("brms_families")` for details on the families.

Global brms Settings

Some global options that are useful to set for your brms analysis

```
options(
  # how many processor cores would you like to use?
  mc.cores = 4,
  # how would you like to access Stan?
  brms.backend = "cmdstanr",
  # cache model binaries
  cmdstanr_write_stan_file_dir=here::here("_brms-cache"),
  # no need to normalize likelihoods
  brms.normalize = FALSE,
  # when you are storing your model to file,
  # how shall it be updated?
  brms.file_refit = "on_change"
  # alternatives: "never", "always"
  # use "never" for production
)
# create cache directory if not yet available
dir.create(here::here("_brms-cache"), FALSE)
```

Assign at least 8 GB of RAM to ensure that everything works smoothly

Case study 1: historical control data

Case study background

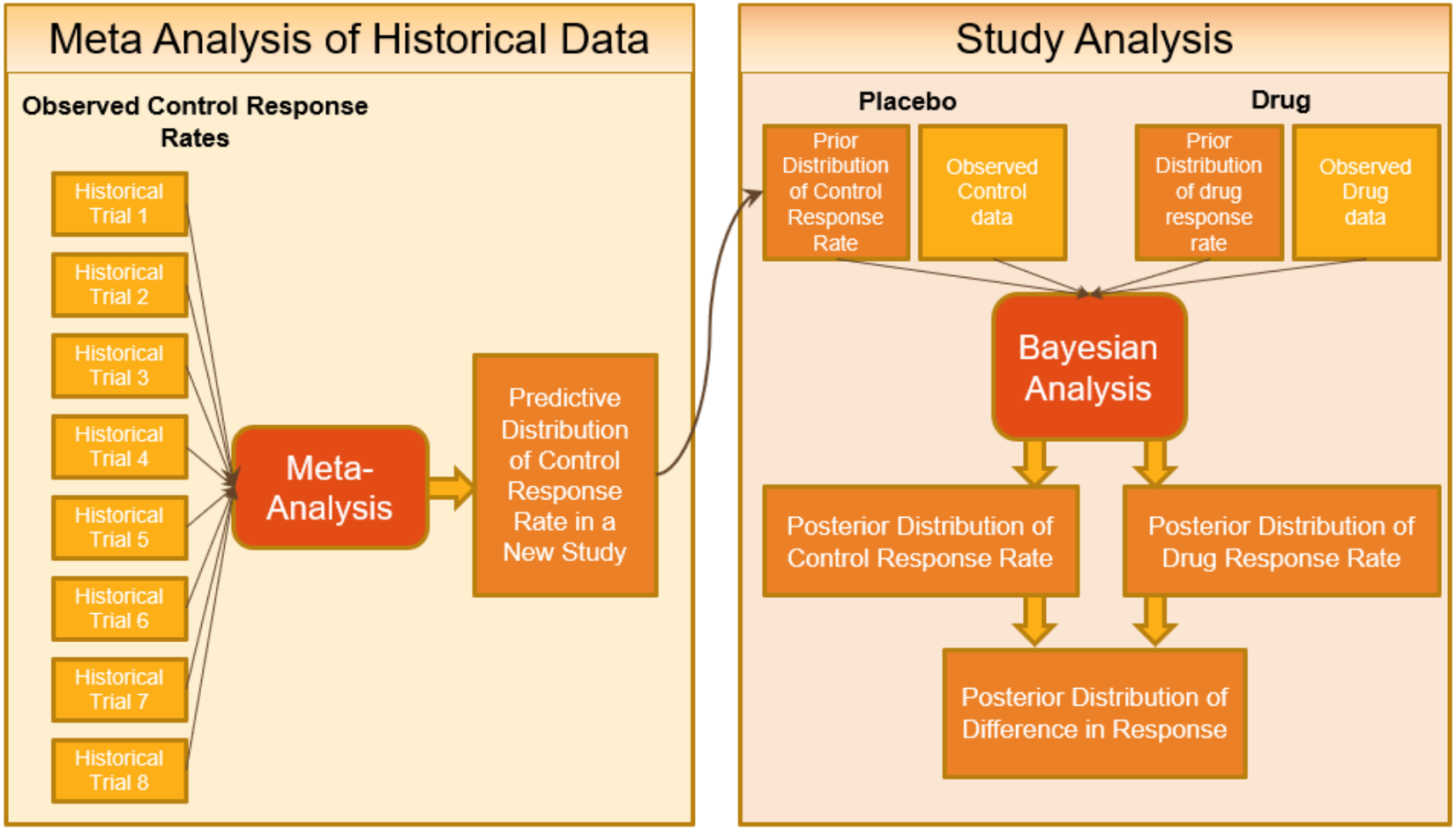
- Suppose we are planning a Phase-II study in ankylosing spondylitis
- The study will be randomized, comparing a test treatment with placebo
- Each patient will be followed, and recorded as a responder or non-responder
- Binary endpoint: percentage of responders
- Goal: minimize number of patients exposed to placebo, using data from past studies about placebo response rates

Historical data

Historical data for placebo:

study	n	r	region
Study 1	107	23	asia
Study 2	44	12	asia
Study 3	51	19	north_america
Study 4	39	9	north_america
Study 5	139	39	north_america
Study 6	20	6	europe
Study 7	78	9	north_america
Study 8	35	10	europe

Meta-Analytic-Predictive (MAP) approach



NOVARTIS

MAP priors

- Approach for augmenting the internal control arm: derive a Meta-Analytic-Predictive (MAP) prior, and use it in the analysis of the Phase-II study data
- The MAP prior is

$$p(\theta_{\text{new}} \,|\, x_{\text{hist}}) = \int p(\theta_{\text{new}} \,|\, \theta_{\text{hist}}) p(\theta_{\text{hist}} | x_{\text{hist}}) \, d\theta_{\text{hist}},$$

- θ_{new} is the probability of responding to placebo in the new study
- x_{hist} is the historical data (responder count), and the posterior distribution

$$p(\theta_{\text{hist}} \,|\, x_{\text{hist}}) \propto p(x_{\text{hist}} \,|\, \theta_{\text{hist}}) \cdot p(\theta_{\text{hist}})$$

is based on Bayesian random-effects meta-analysis

Deriving MAP Priors: Model Specification

The random-effects meta-analysis to derive the MAP prior can be specified as:

```
form_AS <- bf(r | trials(n) ~ 1 + (1|study),
              family = binomial("logit"))
```

```
get_prior(form_AS, data = AS)
```

```
bprior_AS <- prior(normal(0, 2), class = "Intercept") +
  prior(normal(0, 1), class = "sd", coef = "Intercept",
        group = "study")
```

```
fit_AS <- brm(
  form_AS, data = AS, prior = bprior_AS,
  seed = 2454
)
```

Deriving MAP Priors: Summary

```
summary(fit_AS)
```

```
 Family: binomial
  Links: mu = logit
Formula: r | trials(n) ~ 1 + (1 | study)
   Data: AS (Number of observations: 8)

Group-Level Effects:
~study (Number of levels: 8)
              Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
sd(Intercept)     0.38      0.21     0.04     0.86 1.01     1074     1195

Population-Level Effects:
          Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
Intercept    -1.10      0.19    -1.47    -0.70 1.00     1482     1142
```

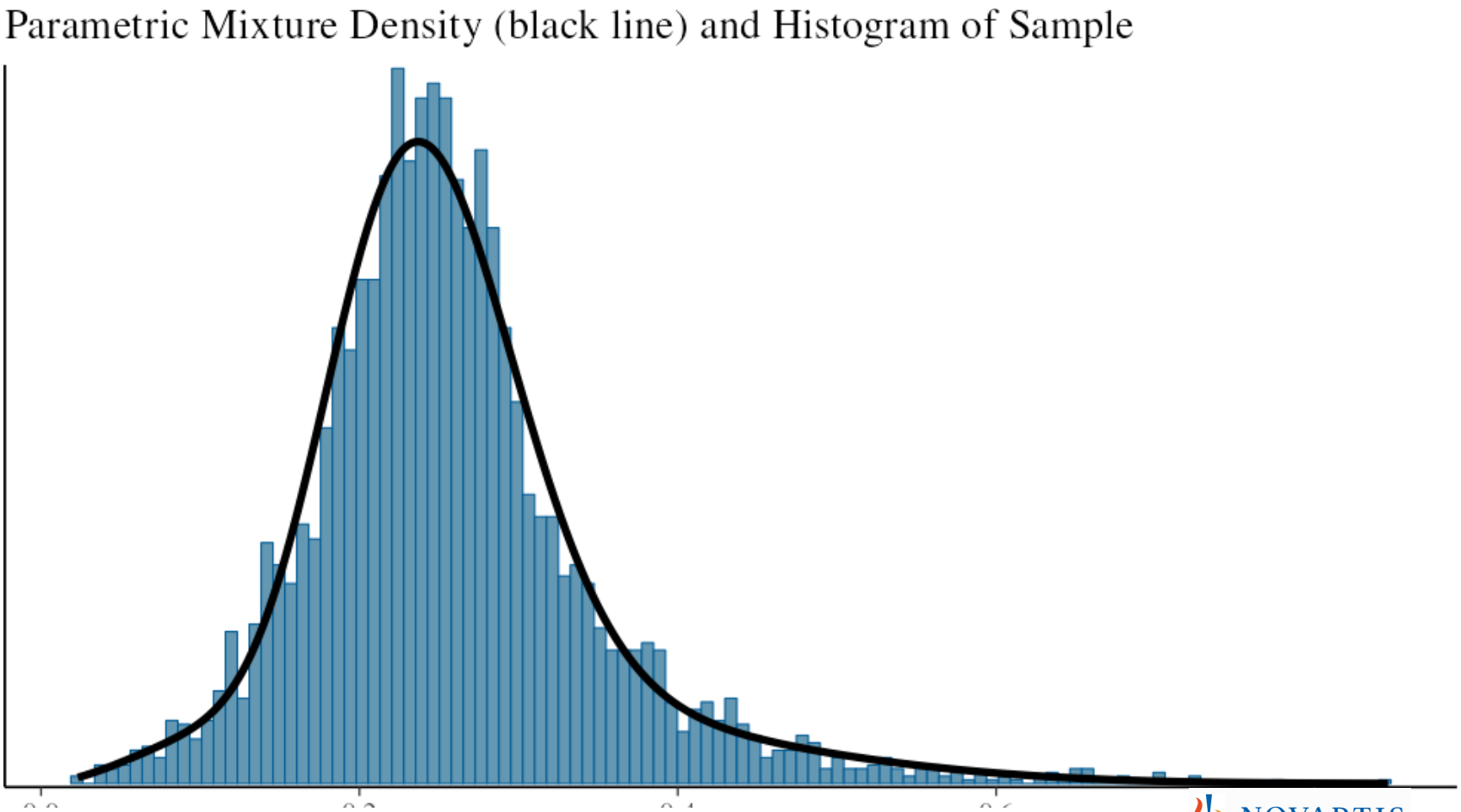
Predicting the placebo response rate in a new study

```
AS_new <- data.frame(study = "new_study", n = 1)
pe <- posterior_epred(
  fit_AS, newdata = AS_new, allow_new_levels = TRUE,
  sample_new_levels = "gaussian"
)
posterior_summary(pe)
```

```
      Estimate  Est.Error      Q2.5     Q97.5
[1,] 0.2582027 0.09086392 0.1075108 0.4811092
```

Approximation with a finite mixture

```
pe_mix <- RBesT::automixfit(pe[, 1], type = "beta")
plot(pe_mix)$mix
```



NOVARTIS

Deriving MAP Priors: Varying Regions Model

```
form_AS_region <- bf(r | trials(n) ~ 1 + (1 | region/study),
                     family = binomial("logit"))
```

```
bprior_AS_region <- prior(normal(0, 2), class="Intercept") +
  prior(normal(0, 0.5), class="sd", coef="Intercept",
        group="region") +
  prior(normal(0, 0.25), class="sd", coef="Intercept",
        group="region:study")
```

```
fit_AS_region <- brm(
  form_AS_region, data = AS_region,
  prior = bprior_AS_region, seed = 2341
)
```

Deriving MAP Priors: Summary

```
summary(fit_AS_region)
```

```
 Family: binomial
  Links: mu = logit
Formula: r | trials(n) ~ 1 + (1 | region/study)
   Data: AS_region (Number of observations: 8)

Group-Level Effects:
~region (Number of levels: 3)
              Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
sd(Intercept)     0.22      0.19     0.01     0.71 1.00     1373     1434

~region:study (Number of levels: 8)
              Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
sd(Intercept)     0.26      0.13     0.02     0.54 1.00     1382     1084

Population-Level Effects:
          Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
Intercept    -1.09      0.22    -1.55    -0.64 1.00     1231     1021
```

Deriving MAP Priors: Extract MAP MCMC samples

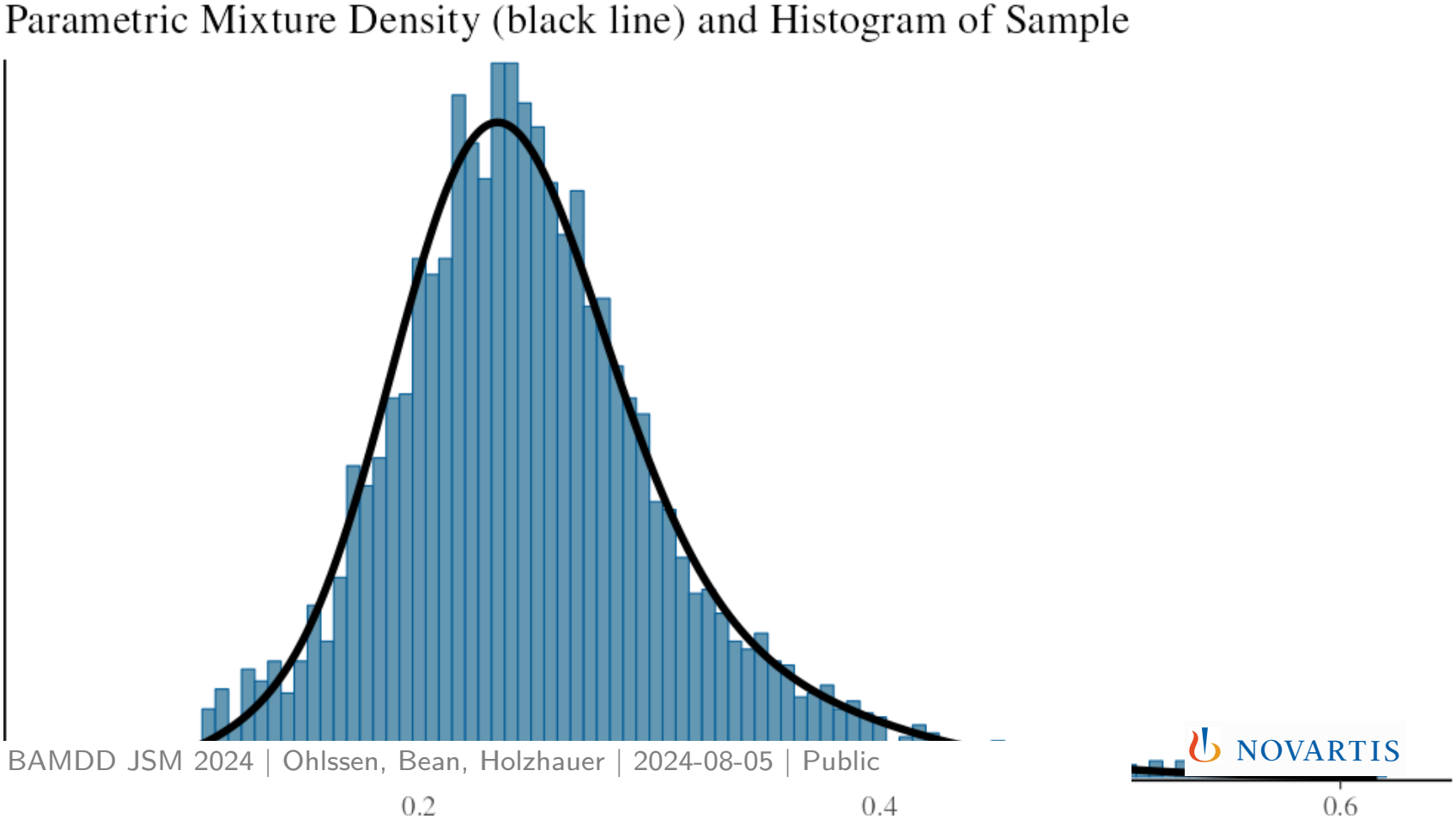
```r
AS_region_new <- data.frame(study = "new_study_asia",
                            n = 1, region = "asia")

pe_region <- posterior_epred(
  fit_AS_region, newdata = AS_region_new,
  allow_new_levels = TRUE,
  sample_new_levels = "gaussian"
)
posterior_summary(pe_region)
```

```
      Estimate Est.Error      Q2.5     Q97.5
[1,] 0.2504935 0.0680455 0.1284533 0.4141241
```

Deriving MAP Priors: Obtain Parametric MAP Prior

```
pe_mix_region <-
  RBesT::automixfit(pe_region[, 1], type = "beta")
plot(pe_mix_region)$mix
```



NOVARTIS

Leveraging historical control data: summary

- Bayesian random-effects meta-analysis models can be used to derive Meta-Analytic-Predictive (MAP) priors
  - Predictions for the mean in a new study inform the MAP prior
- Specification of and inference for these models is simple in `brms`
- Including new-study predictions

Hands-on exercises: historical control data

Posit Cloud link

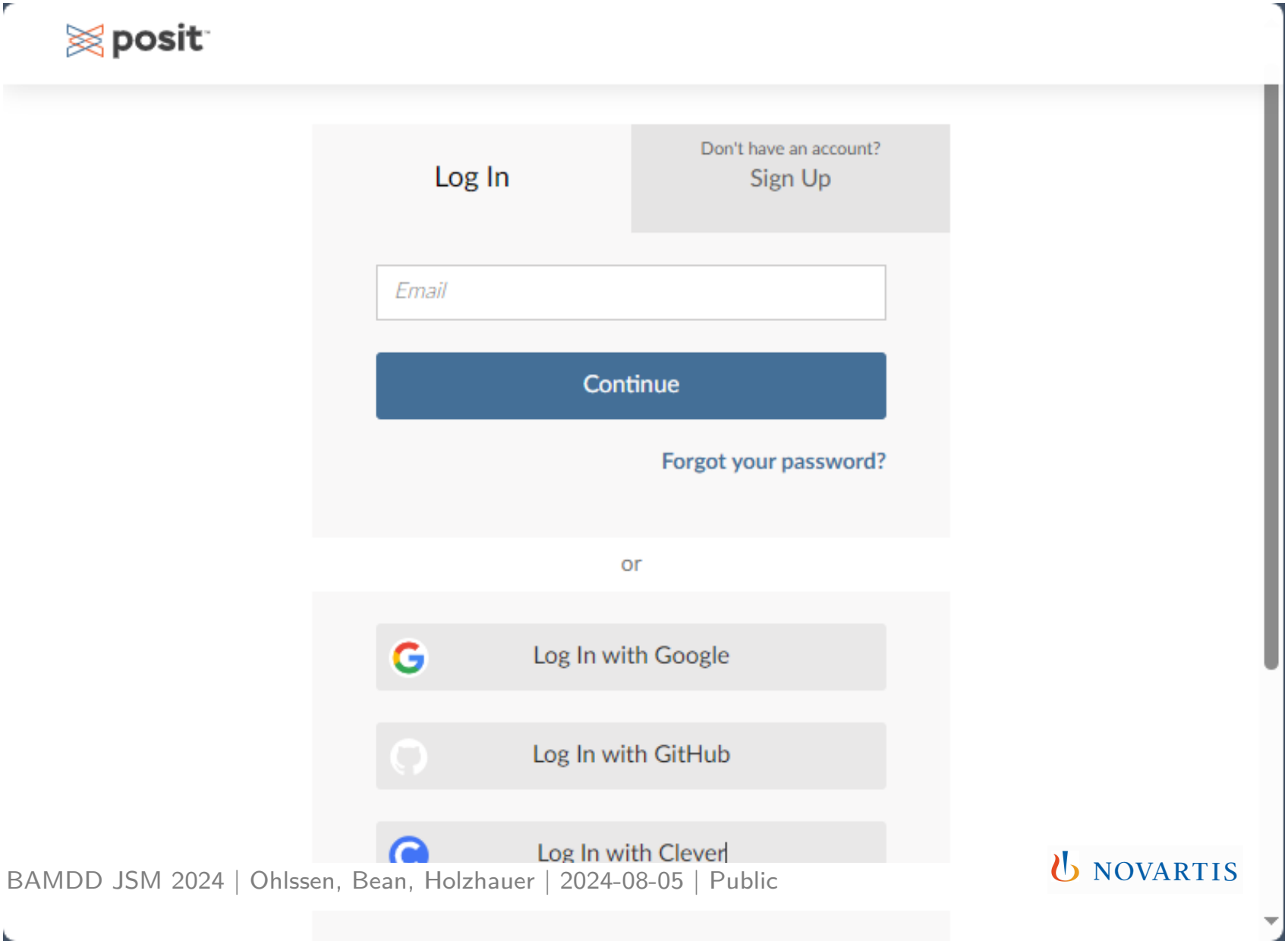
Link to join our Posit Cloud space (shared RStudio workspace):

Link

Step-by-step instructions for access

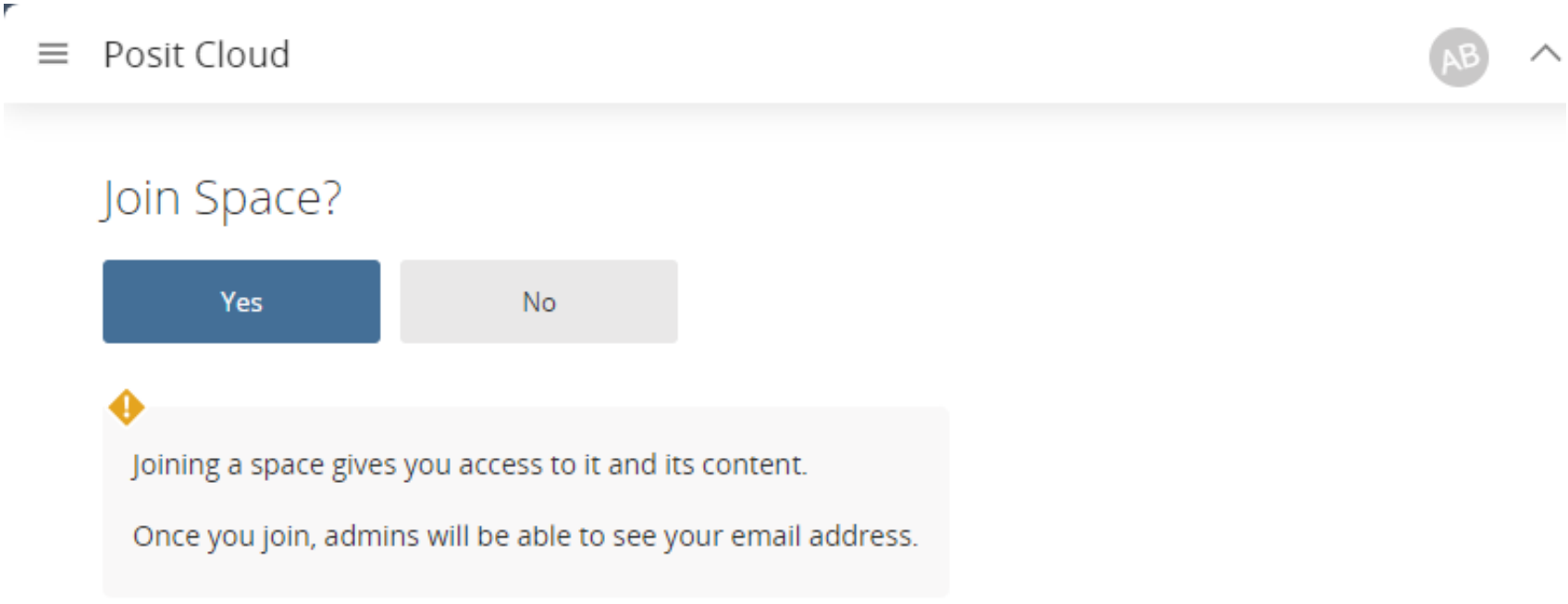
1. (Create an account and) log in to Posit Cloud at the link
2. Agree to join the space
3. Once in the space, go to “Content”
4. Open the “brms-jsm2024” workspace
5. From the “Files” tab in the bottom right, open “exercises/ex1_historical_controls.qmd”

Step 1: Log in



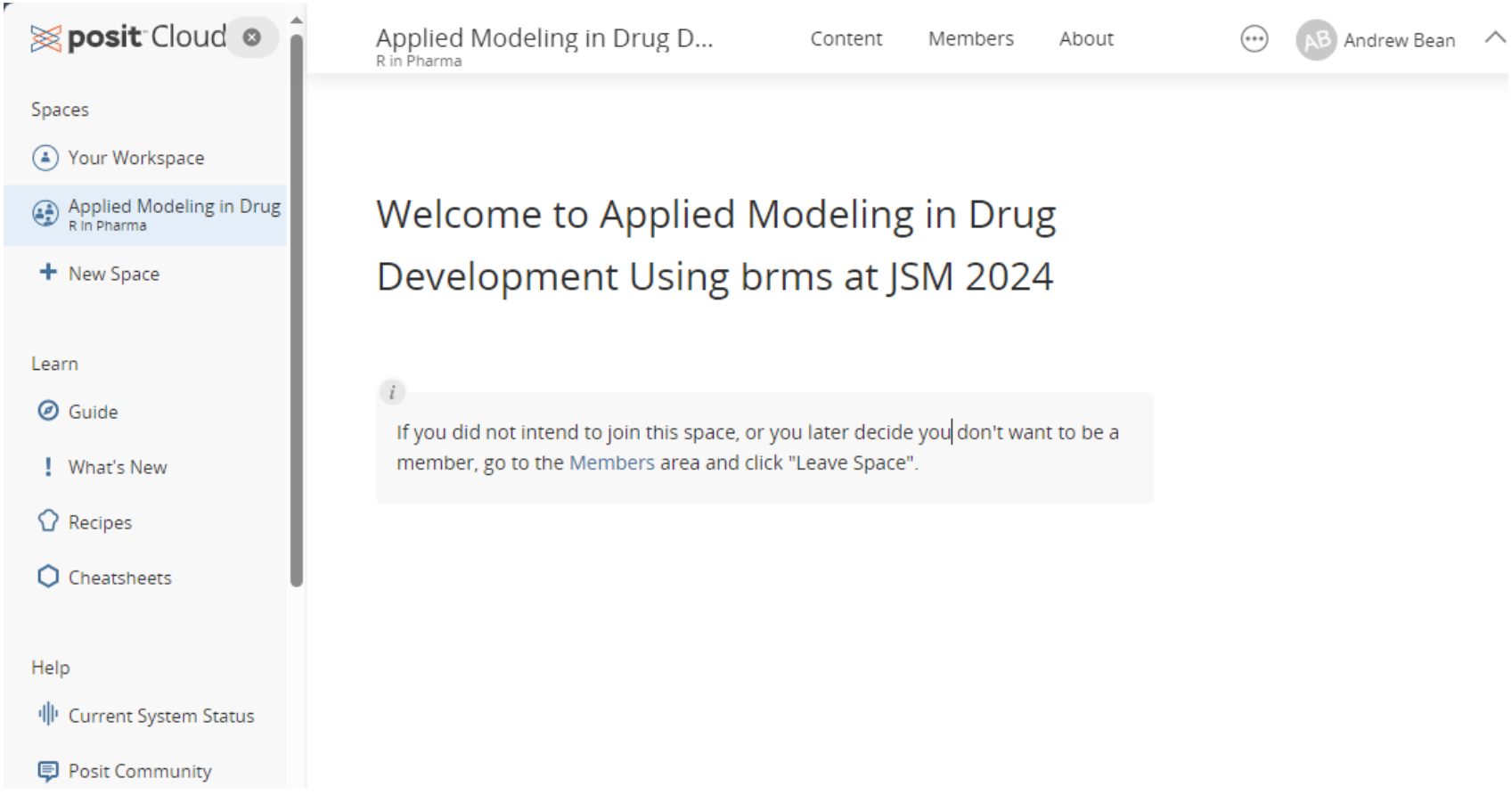
NOVARTIS

Step 2: Join space



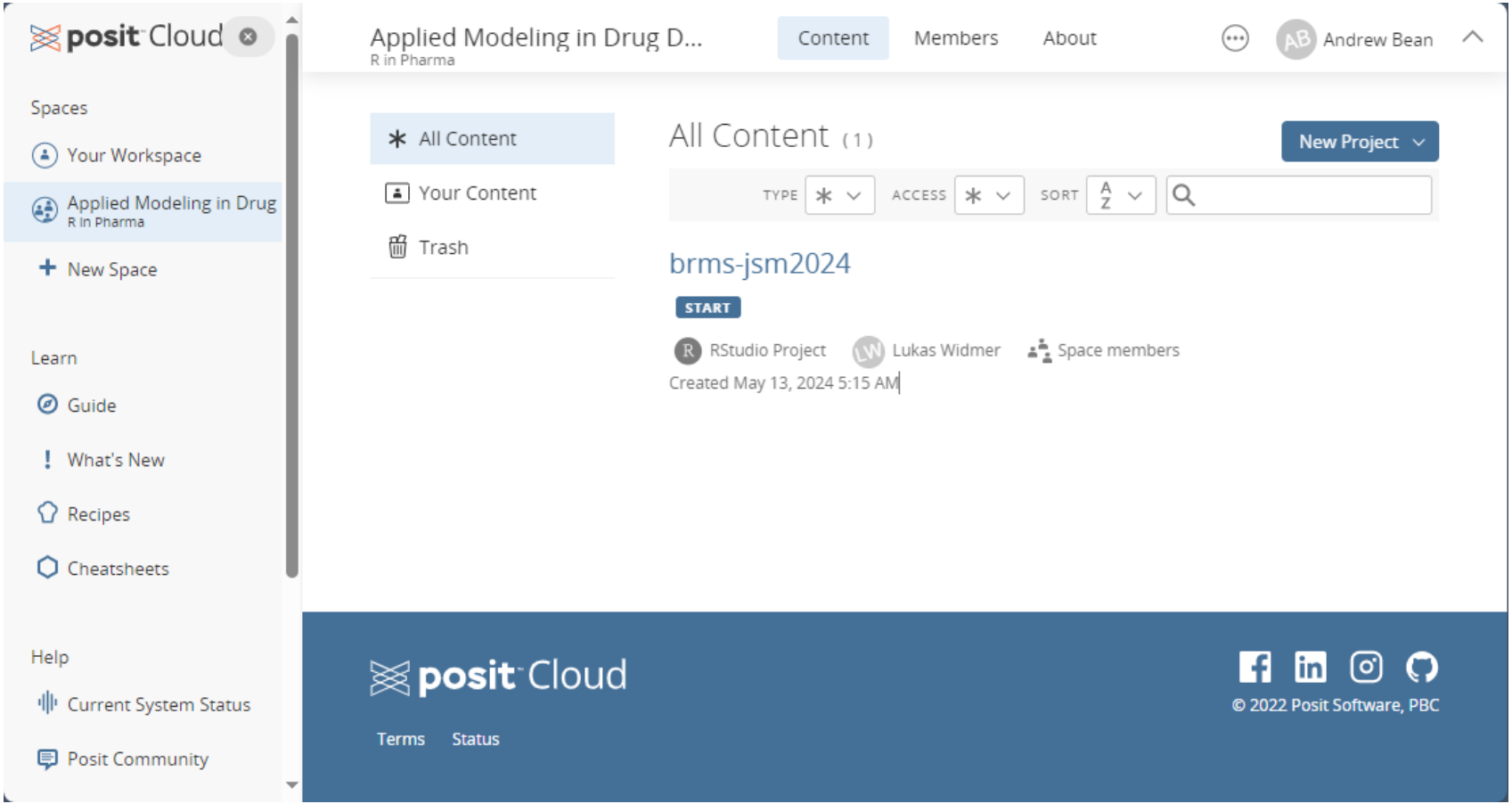
NOVARTIS

Step 3: Click content



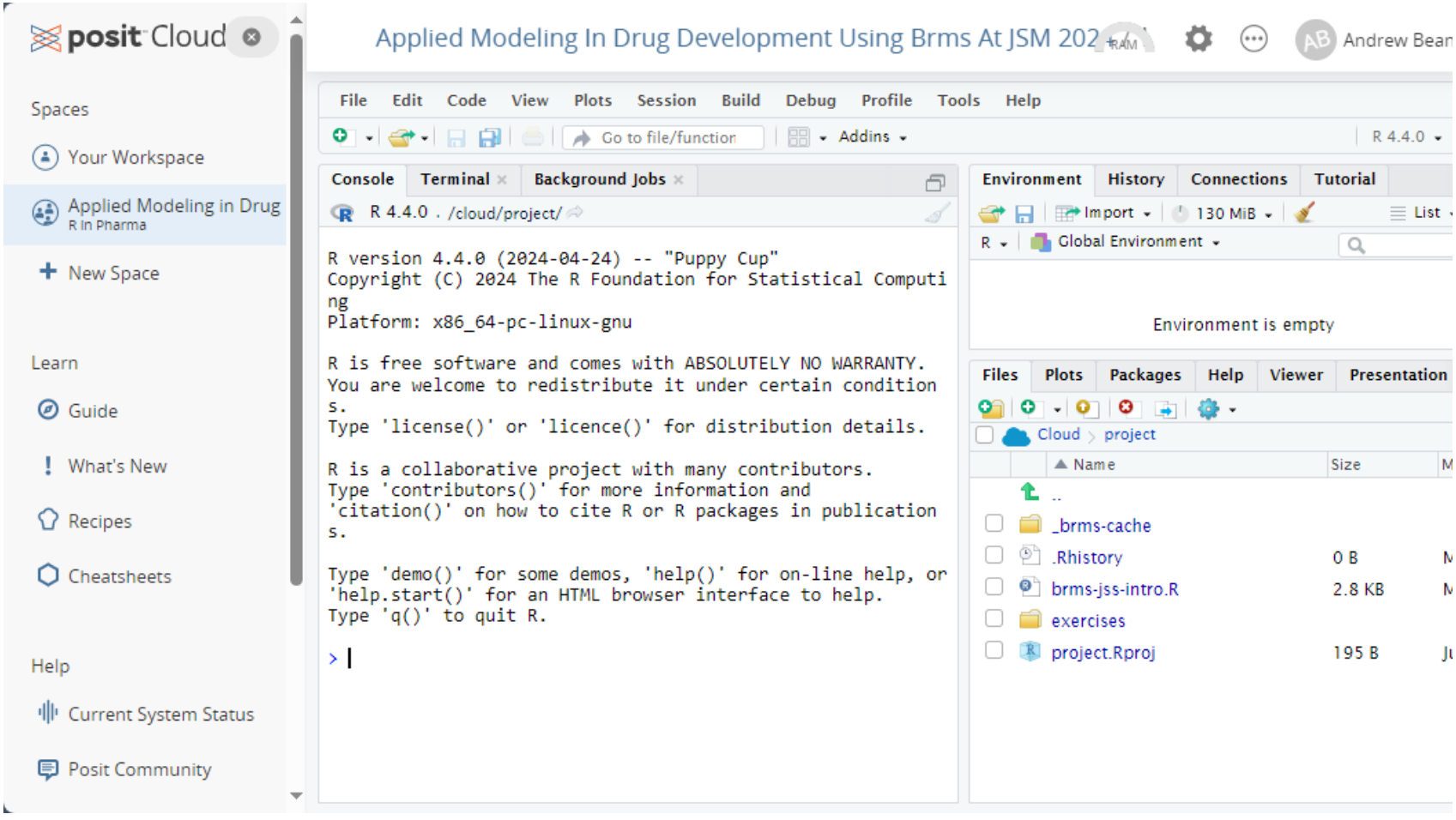
NOVARTIS

Step 4: Open the `brms-jsm2024` workspace



NOVARTIS

Step 5: Open `exercises/ex1_historical_controls.qmd`



NOVARTIS

Bayesian statistics in drug development

Bayesian Statistics in Drug Development

Current landscape

- Motivation
- Bayesian thinking
- Recent regulatory perspective
  - FDA Complex innovative designs
  - FDA Bayesian Supplementary Analysis
  - EMA Methodology Working Party Plan
- Industry applications
  - Proof of concept studies
  - Bayesian decision rules
  - Portfolio assessment via probability of success

Motivation

Important decisions in medicine

- Health authorities

  Should a drug be approved? Or marketing authorization withdrawn?

- Reimbursement agencies

  Should a drug be reimbursed – is it cost effective?

- Medical societies

  Should screening be done and how?

- Pharmaceutical companies

  Should drug development be continued?

- Health organizations

  Should specific food be declared cancer-causing?

Motivation

Important decisions in medicine

- Decisions in medicine may have far reaching consequences

  Patients, medical doctors, payers, pharmaceutical companies, society

- Decisions should be
  - Clear
  - Transparent
  - Evidence based

- Evidence from various sources have to be considered
  - Clinical studies
  - Observational studies
  - Preclinical experiments

Sources of Information

Clinical trials

- Clinical trials are often the key source of information
- A typical randomized clinical trial (RCT)
  - Participating patients either receive *test* or *control* treatment
  - At the end of the trial, the effects of *test* vs *control* are compared
  - Trials are often double-blind, i.e. neither the patient nor the medical doctor knows whether the patient received *test* or *control*
- Available information
  - Summary information on trial results are typically published in clinical journals, or elsewhere, e.g. at ClinicalTrials.gov
  - Individual patient data usually not publicly available

Bayesian Thinking in Healthcare Evaluation

- “The explicit quantitative use of external evidence in the design, monitoring, analysis, interpretation and reporting of a health-care evaluation” (Spiegelhalter et al.; 2004)
- “...The Bayesian view is well suited to this task because it provides a theoretical basis for learning from experience; that is, for updating prior beliefs in the light of new evidence.
- “I am using the term Bayesian here to describe a point of view, and not a particular statistical method involving use of a prior probability distribution when analysing data. ...”
- “...prior knowledge (i.e., validated scientific theory) is to be incorporated into the analysis of current data, and thereby be updated. Prior knowledge can be introduced, as I stress here, through the assumption of mechanistic scientific models for the data,...” (Adapted from Learn and Confirm Sheiner;1997)

Challenges to using Bayes in Drug Development

- Using Bayes in practice is easier said than done
  - Deciding on the relevance of different sources of information is subjective and requires scientific expertise
  - Bayesian thinking usually require a much greater level of engagement and resource
  - How to link together relevant evidence and form realistic complex Bayesian models (subjective, requires technical expertise)
- Traditionally strong emphasis placed on bias and (strict) type one error control leads to
  - Inference based on one or two pieces of evidence (e.g. confirmatory clinical trials) that are the most rigorous and relevant
  - Being more descriptive and qualitative when assessing other evidence
  - Use of simple methods that focus on population average effects try to avoid models and assumptions

Enabling Bayesian methods with a Structured Framework

- Bayesian statistics often requires a structured framework to be used in practice
- Without a structure it is difficult to convince people you are synthesizing evidence appropriately
  - In Europe, Bayesian methods have been widely used in health technology assessment. However, the backbone of this is a careful systematic review
  - CDRH/ CBER Bayesian guidance on the Use of Bayesian Statistics in Medical Device Clinical Trials has greatly helped to provide a structure

Prescription Drug User Fee Act VII(PDUFA VII)

Complex Innovative Designs

- Enhancing regulatory decision tools to support drug development and review
- Complex Innovative Trial Designs (CIDs)
- Includes designs involving complex adaptations, Bayesian methods, or other features requiring simulations to determine statistical properties
- Uses of CIDs
  - Leverage data
  - Rare diseases
  - Multiple body sites in anti-infective drug development
  - Assess multiple interventions, diseases, and/or subgroups under a master protocol

Pediatric Multiple Sclerosis: Bayesian CID

Example posted on FDA website

- FDA considers the following trial design features to be innovative:
  - Use of an active-controlled non-inferiority design that has not been previously used in this setting
  - Borrowing information from historical studies to increase the study power and increase the probability of stopping the trial at the interim analysis
  - Model-based extrapolation from adults to the pediatric population
- Potential Benefits of Design:
  - The proposed non-inferiority trial uses an FDA-approved effective comparator, which can be appealing to patients and their families and can improve recruitment and retention.
  - The Bayesian framework allows for the incorporation of prior knowledge and can make the trial more efficient.
  - Historical information is incorporated using RMAP priors with a 2-component normal distribution and a robust non-informative component which may mitigate the risk of borrowing patient data that is not compatible with that observed in the proposed trial.

Bayesian Supplementary Analysis

FDA Center for Drug Evaluation and Research

- FDA has a commitment under the PDUFA VII agreement to publish draft Guidance on the Use of Bayesian Methodology in Clinical Trials of Drugs and Biologics by September 30, 2025

CDER Center for Clinical Trial Innovation (C3TI)

- C3TI aims to increase experience in Bayesian statistical methods in simple trial settings across sponsors, CDER clinical reviewers, and CDER statisticians, including deepening an understanding of their applicability, opportunities, and challenges.
- With this demonstration project, C3TI will partner with sponsors to use Bayesian methods in supplementary analyses during their trial, providing an opportunity for both CDER and sponsors to learn new methods without impacting review criteria.

Example Bayesian Statistical

Plans Posted on FDA Website

Parallel-Group Trial with a Continuous Outcome	A double-blind trial to assess a drug's effectiveness in lowering acute hypertension in an emergency department setting, utilizing Bayesian analysis to leverage prior medical knowledge and focusing on 2-hour blood pressure reduction.
Supplemental Bayesian Analysis: Unification of Evidence	A double-blind trial with multiple endpoints. A Bayesian approach lets researchers clearly define the specific condition that would change clinical practice and then calculate the likelihood of that condition being met. This condition can be a combination of multiple factors.
Bayesian Subgroup Analysis: Sharing of Information Across Subgroups	This example illustrates how a Bayesian hierarchical model could be used to simultaneously determine estimated treatment effects (via hazard ratios) across four regions for a time-to-event endpoint. Data from all four regions are used in estimating each region-specific hazard ratio.

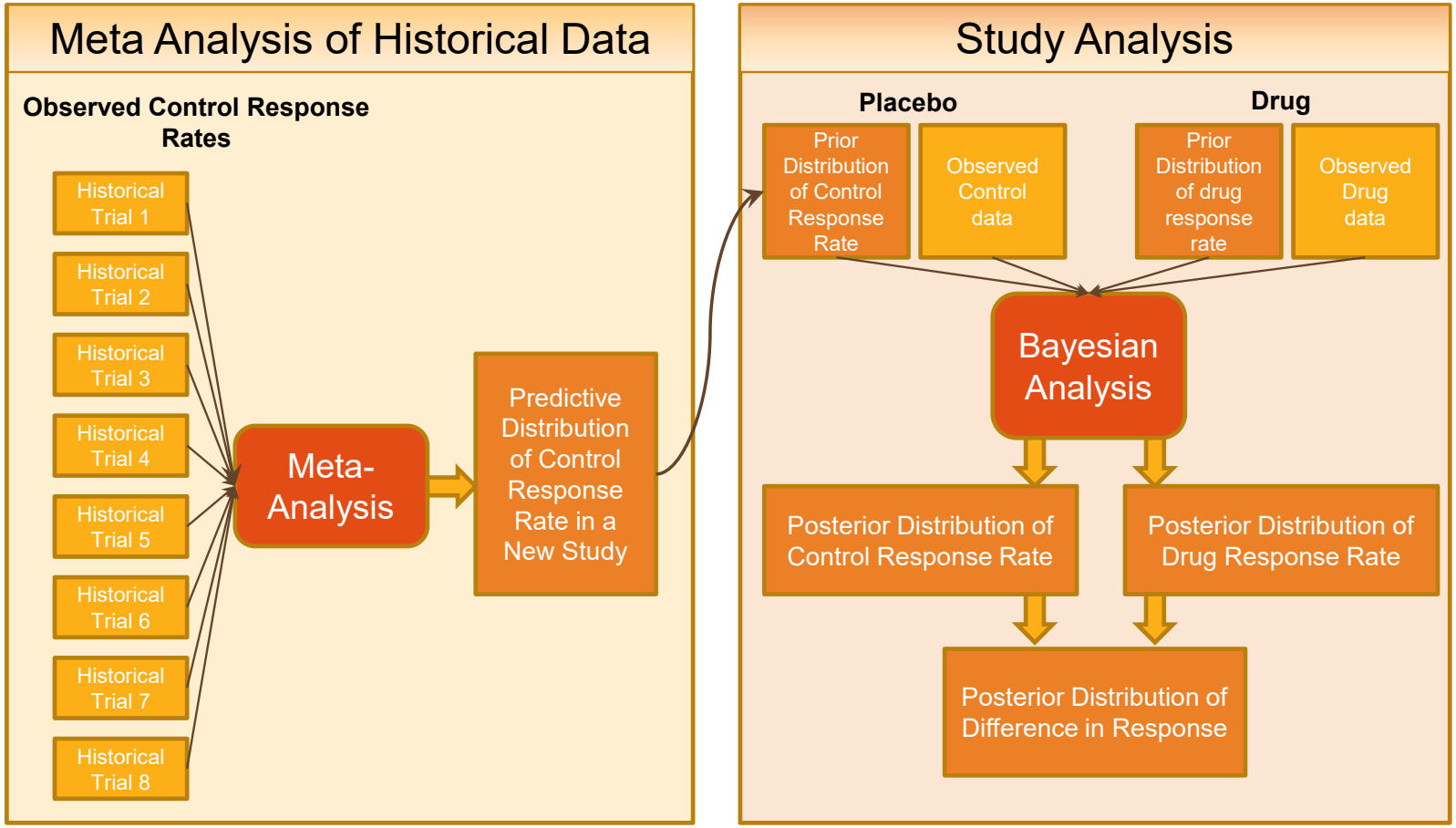
EMA: Methodology Working Party (MWP)

Clinical Trial Modernisation revised 3-year work plan

- Across the clinical research landscape, how trials are conducted is also changing with an increasing number of proposals utilising tools such as master protocols and Bayesian methods.
- There is a need for new guidance in these areas to ensure these novel approaches meet the required evidentiary standards and facilitate their evaluation.
- This will aid their integration into our established system for benefit-risk assessment, balancing innovation with stringent safety and efficacy criteria.

Bayesian proof of concept trial

Historical control prior



Example Ankylosing Spondylitis Study

Application in of using historical control data in a Proof-of-Concept Study

- Disease
  Ankylosing spondylitis
- Experimental treatment
  Monoclonal antibody
- Endpoint
  Binary: response at week 6
- Traditional clinical trial design
  - Experimental (n=24) vs. Placebo (n=24)
  - Fisher's exact test

However: 8 similar historical placebo-controlled clinical trials with different experimental treatments available

Could this historical placebo information be used?

Historical Controls

Motivating example: Trial design and analysis with historical controls

Historical placebo information

- Bayesian primary analysis
- *Prior Placebo* Derived from 8 historical trials (N=533), using a Meta-Analytic-Predictive (MAP) approach

  Beta(11,32) worth 43=11+32 patients
- *Prior Experimental* Weakly informative

  Beta(0.5,1) worth 1.5=0.5+1 patients
- Design:

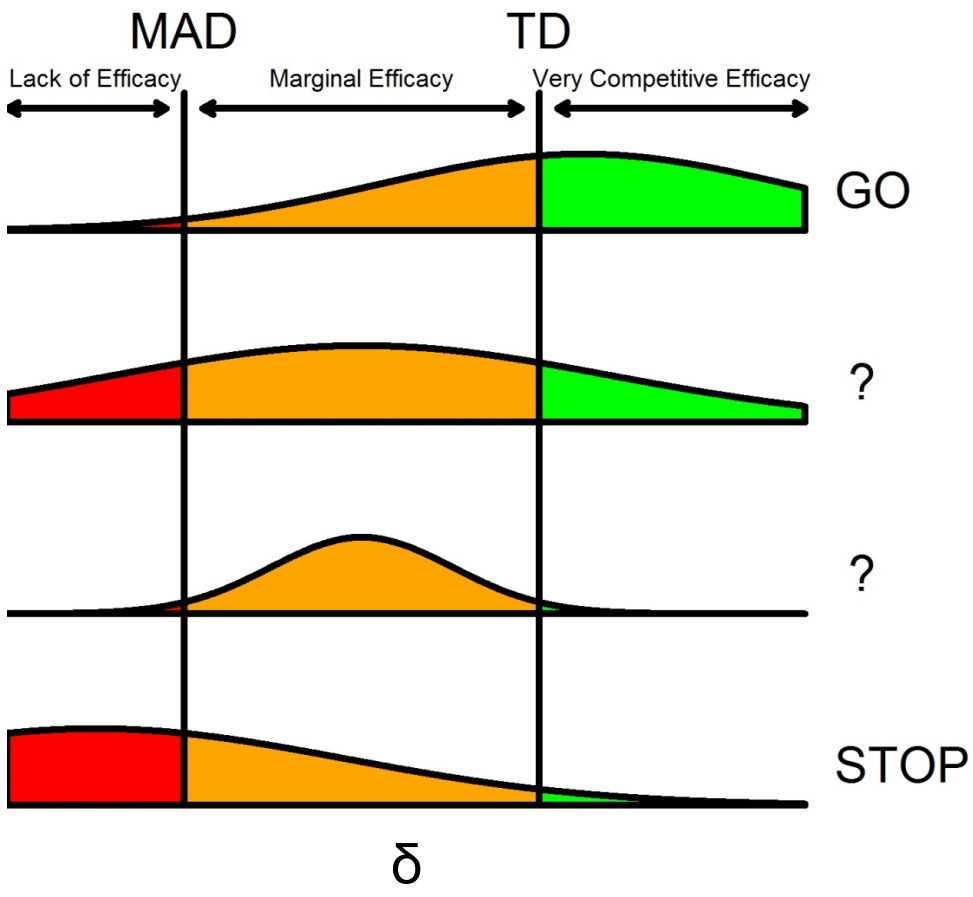
  Treatment (n=24) vs. Placebo (n=6)
- Results:

  14/24 Treatment vs. 1/6 Placebo, $p(\delta > 0 \mid \text{Data}) > 99.8\%$

  Baeten et al. (2013) Lancet 382(9906):1705-1713

Decision rules based on Posterior Probability

Double criterion - minimal acceptable difference target difference



Treatment vs. Control

p(δ > MAD | data) > 97.5%

p(δ > TD | data) > 50%

indeterminate:

neither STOP nor GO

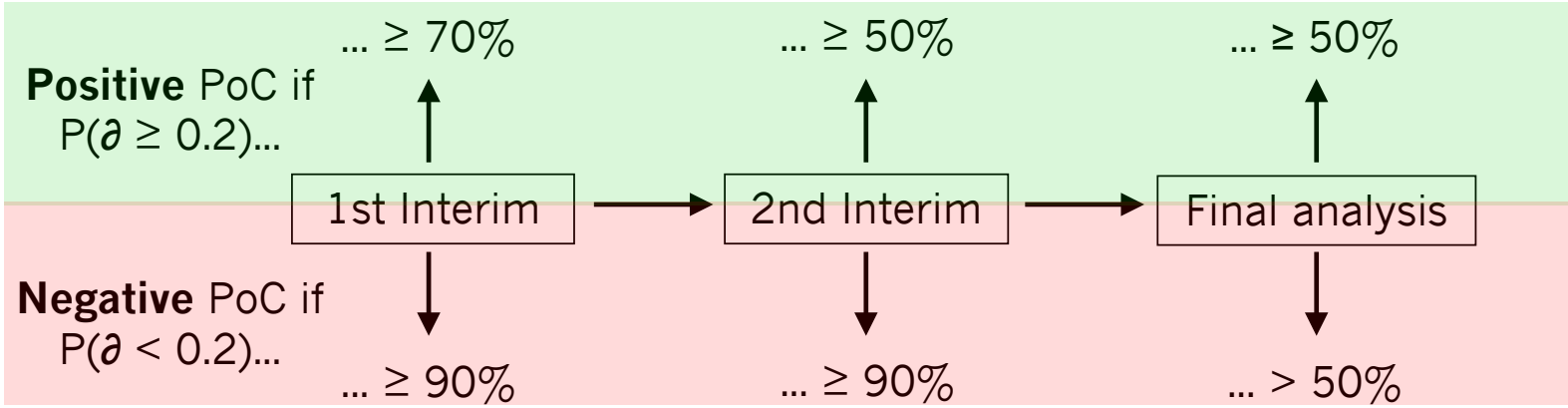
p(δ < MAD | data) > 50%

p(δ < TD | data) > 80%

NOVARTIS

Utilization in a Quick kill Quick win PoC Design

Assessing the design using Frequentist Operating Characteristics



With N=60, 2:1 Active:Placebo, IA's after 20 and 40 patients

	First interim		Second interim		Final		Overall power
Scenario	Stop for efficacy	Stop for futility	Stop for efficacy	Stop for futility	Claim efficacy	Fail	
∂ = 0	1.6%	49.0%	1.4%	26.0%	0.2%	21.9%	3.2%
∂ = 0.2	33.9%	5.1%	27.7%	3.0%	8.8%	21.6%	70.4%
∂ = 0.5	96.0%	0.0%	4.0%	0.0%	0.0%	0.0%	100.0%

With $p_{Placebo}$ = 0.15, 10000 runs

Drug Development: Probability of Success

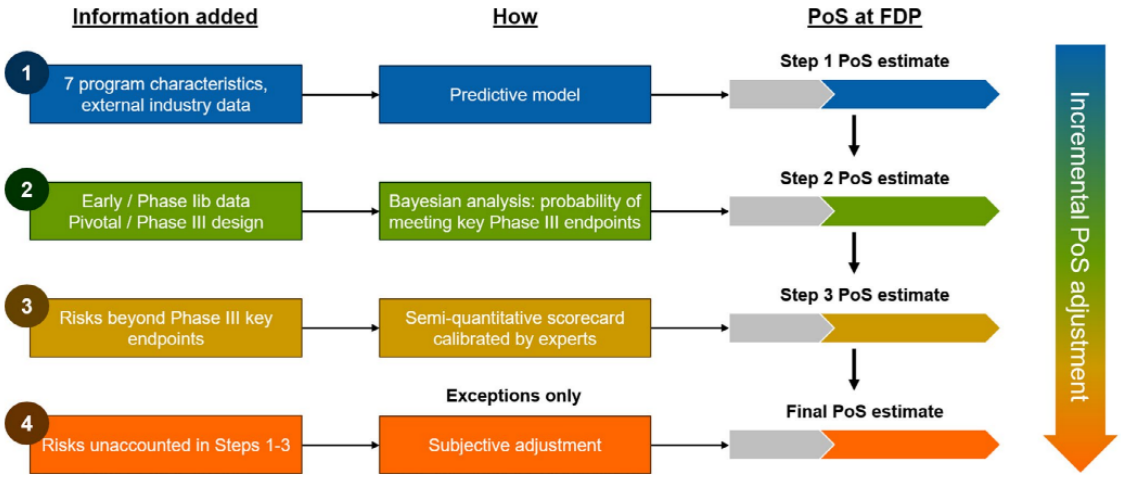
Received: 8 February 2021 | Revised: 30 August 2021 | Accepted: 31 October 2021
DOI: 10.1002/pst.2179

MAIN PAPER

WILEY

Improving the assessment of the probability of success in late stage drug development

Lisa V. Hampson[1] | Björn Bornkamp[1] | Björn Holzhauer[1] | Joseph Kahn[2] | Markus R. Lange[1] | Wen-Lin Luo[2] | Giovanni Della Cioppa[3] | Kelvin Stott[4] | Steffen Ballerstedt[1]



Incorporates background information on historical 'benchmark' rates of success amongst drugs sharing a similar mechanism of action

NOVARTIS

Outlook

- Within some companies Bayesian methods are widely used for internal decision making (e.g., PoS)
- Frameworks, such as CDRH guidance and UK NICE approach to HTA assessment, have helped move Bayesian methods into regulatory decision making
- Recent development by the FDA and EMA have encouraged greater use of Bayesian approaches in drug development
- Bayesian thinking is more important than Bayesian statistics

Priors

Priors

- Introduction
- Priors in brms
- Discussion on ‘weakly informative priors’
  - Dangers of failing to account for background/context
- Strategies for hierarchical models
- Making use of Empirical Evidence
- Summary
- Additional resources

Introduction

Concerns around priors

- Within the Bayesian regression modeling priors are required to perform inference
- This is often seen as a contentious aspect with concerns such as:
  - “Priors are inherently subjective”
  - “Priors bias your analysis”
  - “I have no idea how to set appropriate priors”

Introduction

Potential positive aspects of priors

- Priors can also be viewed as a strength providing flexibility to:
  - Make a-priori implausible values unlikely (weakly informative priors)
  - Incorporate specific expert information into the model (“subjective” priors)
  - Incorporate Empirical Evidence into the model
  - Mimic frequentist methods (uninformative/“objective” priors)
  - Regularize the model to avoid overfitting (shrinkage/sparsifying priors)
  - Facilitate convergence
  - ...
- In many cases the posterior is dominated by the data, which means that the likelihood term $p(y|\theta)$ is much larger than the prior term $p(\theta)$
- This is the default strategy used in brms

Priors brms

- To just get started with brms one may choose to not specify priors when calling brm.
- Doing so will let brms provide in most cases reasonable default priors.
- These default priors are intended to avoid any influence on the calculated posterior.
- The results are fully data driven and will be very close to the respective Frequentist maximum likelihood inference result
- However, the default prior is not guaranteed to stay stable between releases and can thus change whenever the brms version changes.
- Given that any Bayesian analysis requires a prior, we recommend to always explicitly define these - even if these just repeat the default prior from brms, which one can easily obtain.

Priors brms

Further tips

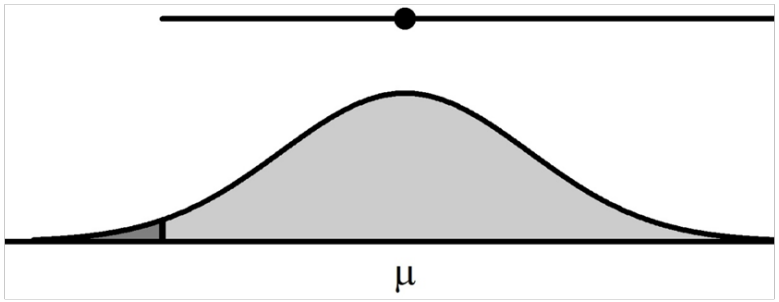
- Respecting boundaries
  - Use hard bounds in priors exactly where parameters have natural bounds
- Expressiveness
  - Use prior families flexible enough to express different plausible prior knowledge
- Scale awareness
  - Ensure that priors take the scales of parameters into account
- Data Informed
  - Use previous data to inform the current priors

Weakly informative/ default priors

Bayesian v Classical inference

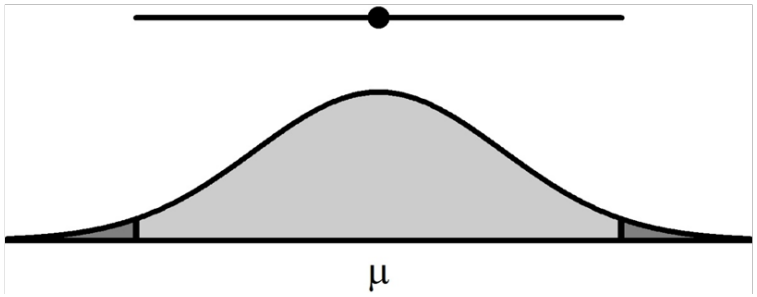
- In many cases the posterior is dominated by the data, which means that the likelihood term $p(y|\theta)$ is much larger than the prior term $p(\theta)$

$\bar{Y}$=2.5, n=25 Normal distribution, σ=5 known, vague prior (n_0=0.001)



one-sided 97.5% CI
(0.54,∞)

p(μ>0.54 | $\bar{Y}$)=97.5%



two-sided 95% CI
(0.54,4.46)

p(0.54<μ<4.46 | $\bar{Y}$)=95%

Bayesian v Classical inference

Critical differences in interpretation: Randomized control trial example

- Bayesian inference leads to a final model that is fully probabilistic
- Conclusions based on probability are potentially much stronger than conclusions based on hypothesis testing or confidence intervals
- For example, in the context of a randomized control trial
  - A low p-value would indicate that the result is unlikely if the true treatment effect is 0 or a treatment effect of 0 is incompatible with the trial data
  - A high posterior probability that the treatment effect is positive allows the conclusion there is a high chance the treatment works

Randomized control trial example

Dangers of weakly informative priors

- Can Bayesian inference with weak priors be used to reach stronger conclusion (e.g., the treatment is efficacious)
- One risk with such an approach is that a crucial piece of evidence or context is missed when assuming weak prior
- For example, if the disease area is known to have a very low drug development success rate
- Priors which incorporating this background could be developed to allow a full probabilistic interpretation and a level of consistency when assessing evidence.
- This approach has been used to support quantitative decision making in drug development (Hampson et al., 2022)

Hierarchical Models General Structure

E.g., Random effects logistic regression model

Specific model

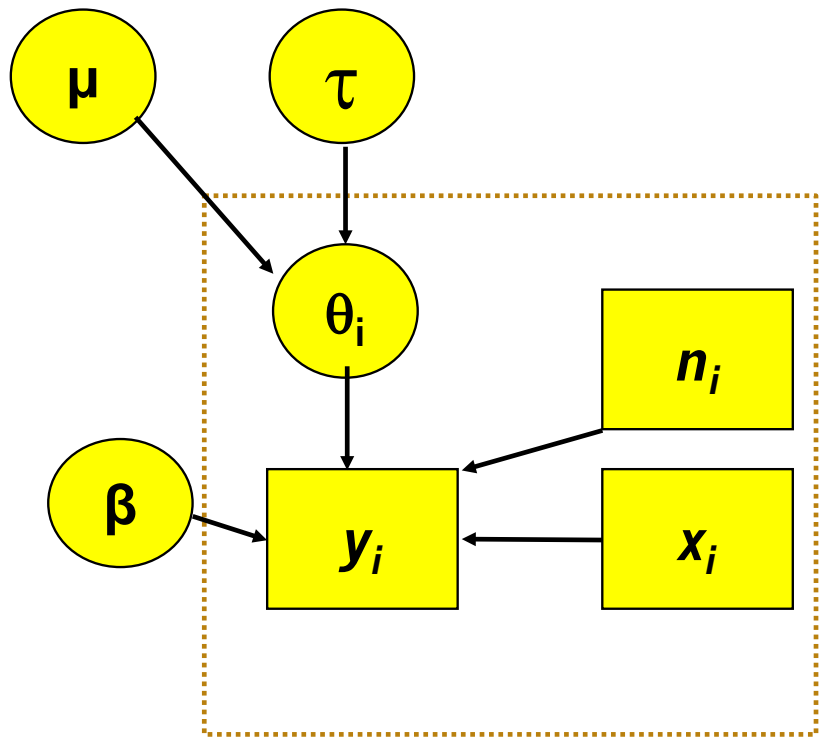
$Y_i \sim \text{Binomial}(N_i, \pi_i)$

$\text{logit}(\pi_i) = \theta_i + x_i \beta$

Study i, Y_i = number of events, N_i = number of patients, π_i = event rate

- $\theta_i \sim N(\mu, \tau^2)$: random study effect
- x_i : design matrix (Study level covariates)

Directed acyclic graphical model



Hierarchical Models

Example Historical trial data

```
kable(RBesT::AS)
```

study	n	r
Study 1	107	23
Study 2	44	12
Study 3	51	19
Study 4	39	9
Study 5	139	39
Study 6	20	6
Study 7	78	9
Study 8	35	10

To use this data as historical control a random intercept model is used, which casts this into a meta-analytic model

Priors for the Basic Hierarchical Model

- Basic hierarchical model
  - $\theta_*, \theta_1, ..., \theta_J \mid \mu, \sigma^2 \sim N(\mu, \tau^2)$
  - priors $p(\mu, \sigma) = p(\mu)\, p(\tau)$
- Population mean μ

  Well informed by data, hence very vague prior can be used, e.g. Normal distribution with very large variance
- Between-trial standard deviation τ
  - Inference sensitive to prior choice if few trials available
  - Use of weakly informative prior recommended, which place most probability mass on plausible values of τ
  - Spiegelhalter, Abrams, Myles (2004), Gelman (2006) Bayesian Analysis

Priors for τ

Priors for the between-trial standard deviation

- The key parameter in this model is the between-trial heterogeneity parameter
- If there are few studies/ groups (less than 5, say), many «default» priors for standard deviations/variances are not appropriate:

  e.g. Inverse-gamma(0.001,0.001), uniform(0,1000), ...
- Half-Normal
  - HN(s^2) is N(0, s^2) truncated at 0
  - Scale *s* should be chosen such that most probability mass placed on plausible values of τ
  - Plausible range for τ depends on endpoint and context
  - This distribution of a half-normal density has been studied extensively in the literature and found to be a robust choice in a wide range of problems.

Making use of Empirical Evidence: Hierarchical Model

Bayesian meta-analysis for Health technology assessment

- Empirical priors study for HTA treatment effect evaluation by the German IQWIG (Lilienthal et al. 2023)
- Empirical priors for meta-analyses organized in disease specific manner (Turner et al. 2015)
- Endpoint specific considerations for between-trial heterogeneity parameter priors in random effect meta-analyses (Röver et al. 2021)

Making use of Empirical Evidence: Non-Linear models

Bayesian model-based dose response

$$Y = E_0 + \frac{E_{max} dose^h}{ED_{50}^h + dose^h}$$

- dose is the treatment group dose
- E_0= response under placebo treatment
- E_{max} = maximum difference with PBO
- ED_{50} = dose producing half the maximum response
- The power parameter h determines the steepness of the curve

FDA Fit for purpose Integrated review

Submission	Empirically Based Bayesian Emax Models for Dose Response Design and Analysis
Submitter	Pfizer
Submission Date	April 21, 2021
OCP Reviewer(s)	Junshan Qiu, Jingyu Yu
OCP Concurring Reviewers	Rajanikanth Madabushi, Hao Zhu
OB Reviewer(s)	Junghi Kim, Qianyu Dang, Donald Schuirmann,
Final Signatory	Issam Zineh, OCP Director Sylva Collins, OB Director

- Empirically-based prior distribution combining dose response meta-data and compound-specific information
- Priors developed for key non-linear parameters ED_{50}, h

Summary

Priors

- Bayesian inference with priors provides inference leading to a fully probabilistic model
- This potentially leads to very clear interpretation based on probability statements
- While the default priors in brms provide a good starting point for many modeling problems:
  - Need to be aware of dangers of missing important background/ context that should be incorporated into a prior
  - More care and subtlety is often needed in more complex models (e.g., hierarchical models and non-linear models)
  - Think about scales, boundaries and plausible range
  - Where possible consider empirical evidence

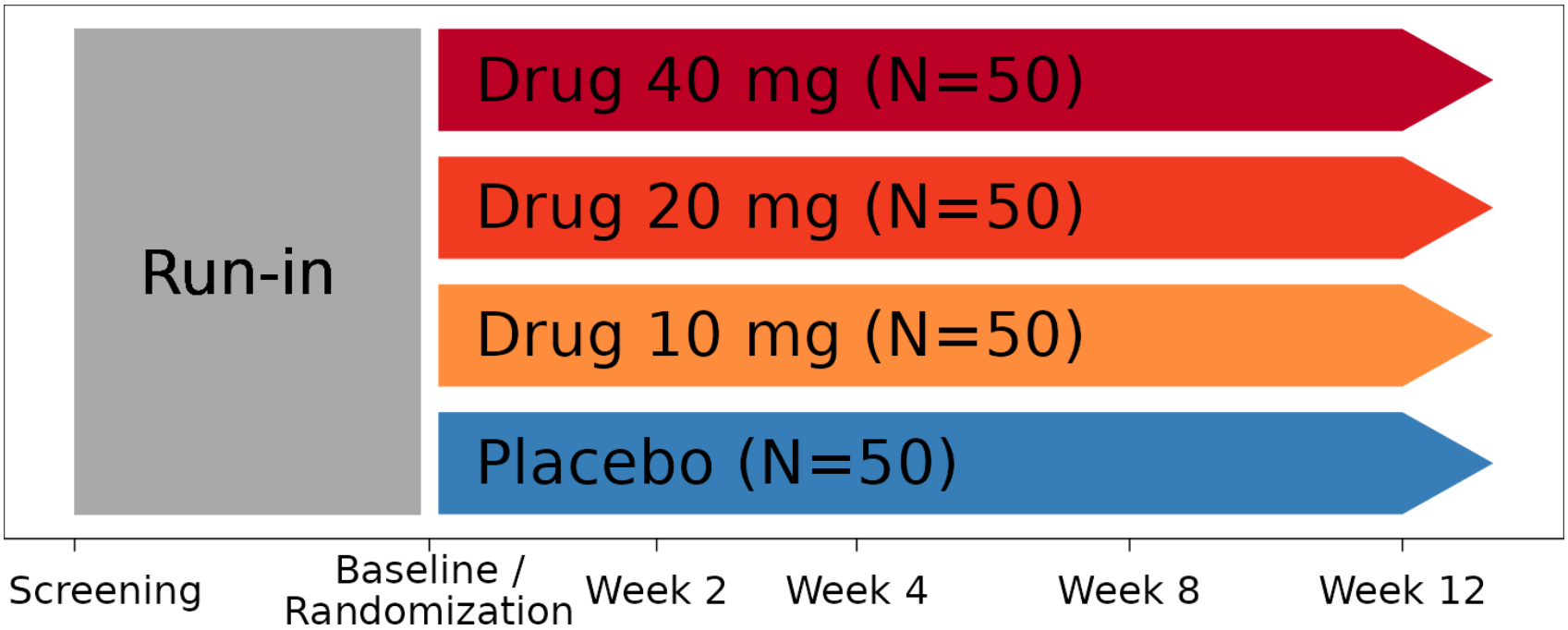
Additional literature for consideration

- Comprehensive introductory book to applied Bayesian data analysis with detailed discussion on many examples (Gelman et al. 2014)
- Live wiki document maintained by Stan user community (heavily influenced by Andrew Gelman & Aki Vehtari) (Stan 2024)
- Prior strategy based on nested modeling considerations (penalization of more complex models), (Simpson et al. 2014)
- Global model shrinkage regularized horseshoe prior (Piironen and Vehtari 2017) or R2D2 prior (overall $R2$) (Zhang et al. 2022)

Case study 2: Bayesian Mixed Models for Repeated Measures (MMRM)

Example study

- Hypothetical dose findng study (N=200 randomized patients)
- Continuous outcome measured for each patient at baseline, as well as weeks 2, 4, 8 and 12

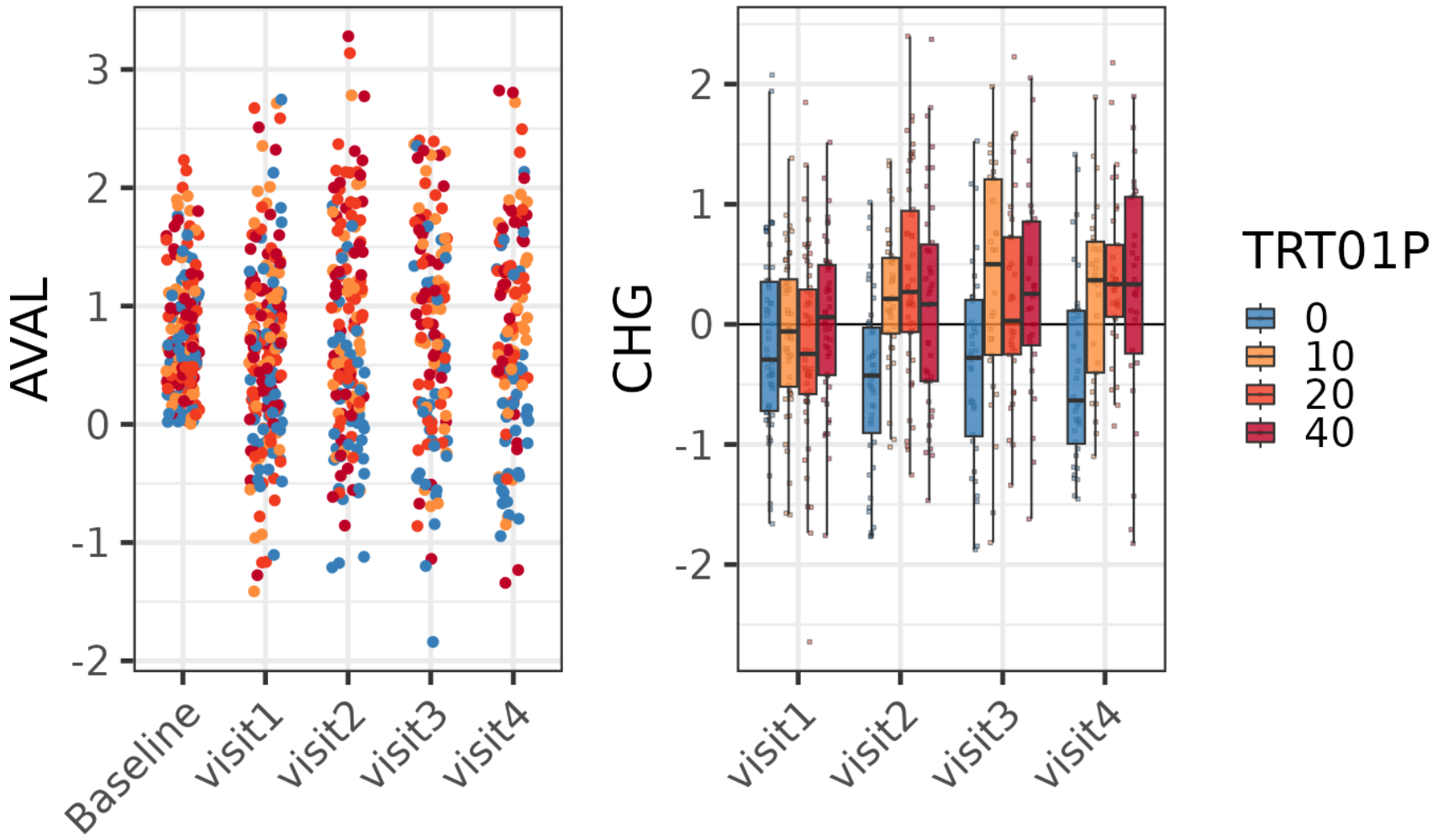


NOVARTIS

Overview and Analysis Goals

- Mixed Effects Model for Repeated Measures (MMRM) commonly used for longitudinal data (each patient measured at multiple visits)
- Direct likelihood analysis that can address hypothetical estimand of all patients completing the study on treatment without doing missing data imputation first
- Commonly no structure assumed for correlations between visits and different variance allowed for different visits (to avoid unnecessary assumptions)
- Convergence issues with REML fit common, especially for small sample sizes, which is alleviated by Bayesian inference with (weakly-)informative priors
- Bayes allows us to incorporate prior information and historical data, which is very interesting for Phase I studies
- `brms` lets us easily add more components & structure to the model

What do our data look like?



NOVARTIS

Analysis Data Model (ADaM) Basic Data Structure (BDS)

USUBJID	TRT01P	AVISIT	ADY	AVAL	CHG	BASE
3	10	visit1	14	1.32	0.54	0.78
3	10	visit2	28	1.30	0.52	0.78
3	10	visit3	56	−0.24	−1.02	0.78
3	10	visit4	84	1.40	0.63	0.78
9	20	visit1	14	0.98	0.87	0.11
9	20	visit2	28	1.54	1.44	0.11
9	20	visit3	56	−0.86	−0.97	0.11
9	20	visit4	84	1.33	1.22	0.11
13	10	visit1	14	2.35	0.45	1.90
13	10	visit2	28	2.04	0.14	1.90

Model Specification: Informal

A widely used default analysis is to have the following fixed effects:

- visit as a factor
- treatment as a factor
- treatment by visit interaction
- baseline (pre-treatment) value of the continuous endpoint as a continuous covariate
- visit by baseline value interaction

And the following random effects:

- Random subject effect on the visit main effect or equivalently correlated residual error terms within subjects

Model Specification: Formal

Formally, let us assume that there are V visits. We usually assume that the V-dimensional response Y_i for patient i satisfies

$$Y_i = X_i\beta + Z_i b_i + \epsilon_i$$

with $b_i \sim \text{MVN}(0, D)$ and $\epsilon_i \sim \text{MVN}(0, \Sigma)$, where Σ is a diagonal matrix. This implies

$$Y_i \sim \text{MVN}(X_i\beta, V_i),$$

where $V_i = Z_i D Z_i^T + \Sigma$. Model the correlated Y_{ij} either by

1. marginalizing out random effects & account for them with correlated residual errors (residual covariance matrix V_i), or
2. conditionally on (V-dimensional) random effects b_i with residual errors ϵ_i independent (once we condition on b_i).

MMRMs in SAS

Widely-used high-quality reference implementation

```
PROC MIXED DATA=simulated_data;
  CLASS TRT01P AVISIT USUBJID;
  MODEL CHG ~ TRT01P AVISIT BASE TRT01P*AVISIT AVISIT*BASE
    / SOLUTION DDFM=KR ALPHA = 0.05;
  REPEATED AVISIT / TYPE=UN SUBJECT = USUBJID R Rcorr GROUP=TRT01P;
  LSMEANS TRT01P*AVISIT / DIFFS PDIFF CL OM E;
RUN;
```

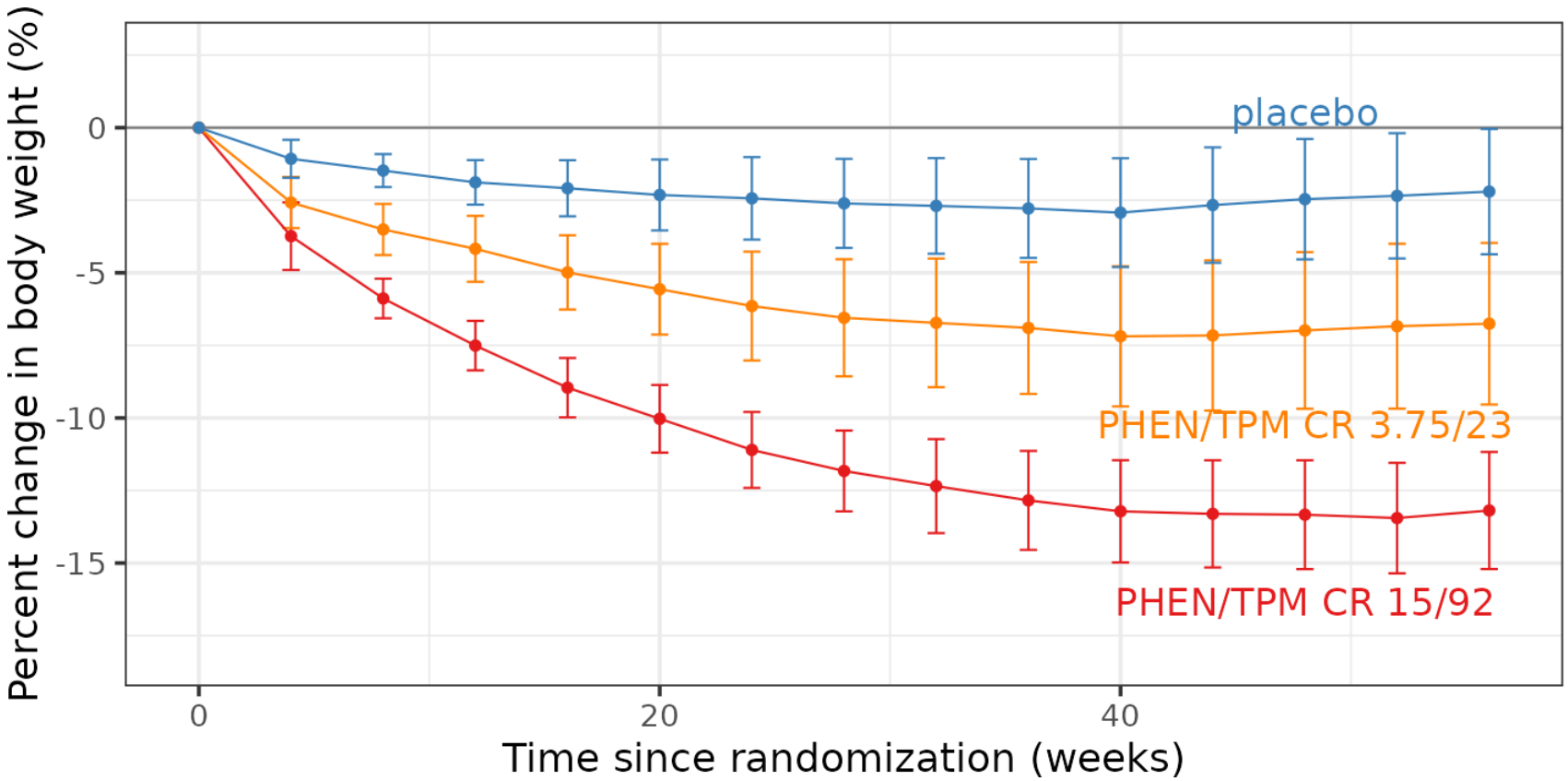
MMRMs with the `mmrm` R package

Fit the model in R in a frequentist framework

```
library(mmrm)
mmrm_fit <- mmrm(
  formula = CHG ~ TRT01P + AVISIT + BASE + AVISIT:TRT01P +
    AVISIT:BASE + us(AVISIT | TRT01P / USUBJID),
  method = "Kenward-Roger",
  vcov = "Kenward-Roger-Linear", # to match SAS
  data = simulated_data %>% mutate(USUBJID=factor(USUBJID))
)
```

Forward Difference Parametrization: Motivation

We would like to put priors on the differences from visit to visit



NOVARTIS

Difference contrasts

Setup forward difference contrasts for changes between visits:

```
contrasts(simulated_data$AVISIT) <- MASS::contr.sdif
```

Hard to interpret the contrast matrix directly:

```
       visit2-visit1 visit3-visit2 visit4-visit3
visit1 -3/4          -1/2          -1/4
visit2  1/4          -1/2          -1/4
visit3  1/4           1/2          -1/4
visit4  1/4           1/2           3/4
```

Learn more about contrasts:
https://bbolker.github.io/mixedmodels-misc/notes/contrasts.pdf

What do these contrasts mean?

Inverting the contrast matrix reveals the dummy variables' interpretation:

```
# add the intercept
cmat <- cbind("1" = 1, contrasts(simulated_data$AVISIT))
# compute the inverse matrix
solve(cmat) %>% MASS::fractions()
```

```
                visit1 visit2 visit3 visit4
1                 1/4    1/4    1/4    1/4
visit2-visit1      -1      1      0      0
visit3-visit2       0     -1      1      0
visit4-visit3       0      0     -1      1
```

MMRMs in brms

```
mmrm_model1 <- bf(
  CHG ~ 1 + AVISIT + BASE + BASE:AVISIT + TRT01P + TRT01P:AVISIT
    + unstr(time = AVISIT, gr = USUBJID),
  sigma ~ 1 + AVISIT + TRT01P + AVISIT:TRT01P
)
```

MMRMs in brms

```
mmrm_model1 <- bf(
  CHG ~ 1 + AVISIT + BASE + BASE:AVISIT + TRT01P + TRT01P:AVISIT
    + unstr(time = AVISIT, gr = USUBJID),
  sigma ~ 1 + AVISIT + TRT01P + AVISIT:TRT01P
)
```

```
mmrm_prior1 <- prior(normal(0, 2), class=Intercept) +
    prior(normal(0, 1), class=b) +
    prior(normal(0, log(10.0)/1.64), class=Intercept, dpar=sigma) +
    prior(normal(0, log(2.0)/1.64), class=b, dpar=sigma) +
    prior(lkj(1), class=cortime)
```

MMRMs in brms

```
mmrm_model1 <- bf(
  CHG ~ 1 + AVISIT + BASE + BASE:AVISIT + TRT01P + TRT01P:AVISIT
    + unstr(time = AVISIT, gr = USUBJID),
  sigma ~ 1 + AVISIT + TRT01P + AVISIT:TRT01P
)
```

```
mmrm_prior1 <- prior(normal(0, 2), class=Intercept) +
    prior(normal(0, 1), class=b) +
    prior(normal(0, log(10.0)/1.64), class=Intercept, dpar=sigma) +
    prior(normal(0, log(2.0)/1.64), class=b, dpar=sigma) +
    prior(lkj(1), class=cortime)
```

```
fit_mmrm1 <- brm(
  formula = mmrm_model1,
  data = simulated_data,
  prior = mmrm_prior1,
  ...
)
```

Expected marginal (least-squares) means

```
emm2 <- fit_mmrm1 %>%
  emmeans(~ TRT01P | AVISIT, weights="proportional")
```

```
AVISIT = visit1:
 TRT01P  emmean lower.HPD upper.HPD
 0      -0.1575   -0.3563    0.0414
 10     -0.0314   -0.2328    0.1611
 20     -0.1613   -0.3849    0.0460
 40      0.0543   -0.1353    0.2295

AVISIT = visit4:
 TRT01P  emmean lower.HPD upper.HPD
 0      -0.3686   -0.6177   -0.0983
 10      0.2235   -0.0548    0.4952
 20      0.4040    0.1764    0.6199
 40      0.3541    0.0173    0.6954

Point estimate displayed: median
HPD interval probability: 0.95
```

With `emm2 %>% as.mcmc() %>% summarize_draws()` we can work with MCMC samples of the expected marginal means & to summarize them exactly as we want (e.g. quantile credible intervals)

Expected Marginal Contrasts per Visit

```
contrast(emm2, adjust="none", method="trt.vs.ctrl", ref="TRT01P0")
```

```
AVISIT = visit1:
 contrast           estimate lower.HPD upper.HPD
 TRT01P10 - TRT01P0  0.12727   -0.1700     0.394
 TRT01P20 - TRT01P0 -0.00262   -0.3027     0.283
 TRT01P40 - TRT01P0  0.20851   -0.0692     0.465

AVISIT = visit4:
 contrast           estimate lower.HPD upper.HPD
 TRT01P10 - TRT01P0  0.59344    0.2103     0.971
 TRT01P20 - TRT01P0  0.77371    0.4205     1.104
 TRT01P40 - TRT01P0  0.72327    0.3095     1.149

Point estimate displayed: median
HPD interval probability: 0.95
```

`as.mcmc()` to work with MCMC samples of the difference

Monotonic Effects Across Ordered Factor Levels

```
mmrm_model2 <- bf(
  CHG ~ 1 + AVISIT + mo(TRT01P) + BASE + mo(TRT01P):AVISIT
    + BASE:AVISIT + unstr(time = AVISIT, gr = USUBJID),
  sigma ~1 + AVISIT + mo(TRT01P) + mo(TRT01P):AVISIT
)
```

Monotonic Effects Across Ordered Factor Levels

```
mmrm_model2 <- bf(
  CHG ~ 1 + AVISIT + mo(TRT01P) + BASE + mo(TRT01P):AVISIT
    + BASE:AVISIT + unstr(time = AVISIT, gr = USUBJID),
  sigma ~1 + AVISIT + mo(TRT01P) + mo(TRT01P):AVISIT
)
```

For category $c = 0, \ldots, (\text{categories} - 1)$, the monotonic term is

$$\text{coefficient} \times (\text{categories} - 1) \times \sum_{k=1}^{c} \zeta_k,$$

where $\zeta_k \in [0, 1]$ and $\sum_{k=1}^{\text{categories}-1} \zeta_k = 1$. For more details see this vignette: https://paul-buerkner.github.io/brms/articles/brms_monotonic.html

Monotonic Effects Across Ordered Factor Levels

```
mmrm_model2 <- bf(
  CHG ~ 1 + AVISIT + mo(TRT01P) + BASE + mo(TRT01P):AVISIT
    + BASE:AVISIT + unstr(time = AVISIT, gr = USUBJID),
  sigma ~1 + AVISIT + mo(TRT01P) + mo(TRT01P):AVISIT
)
```

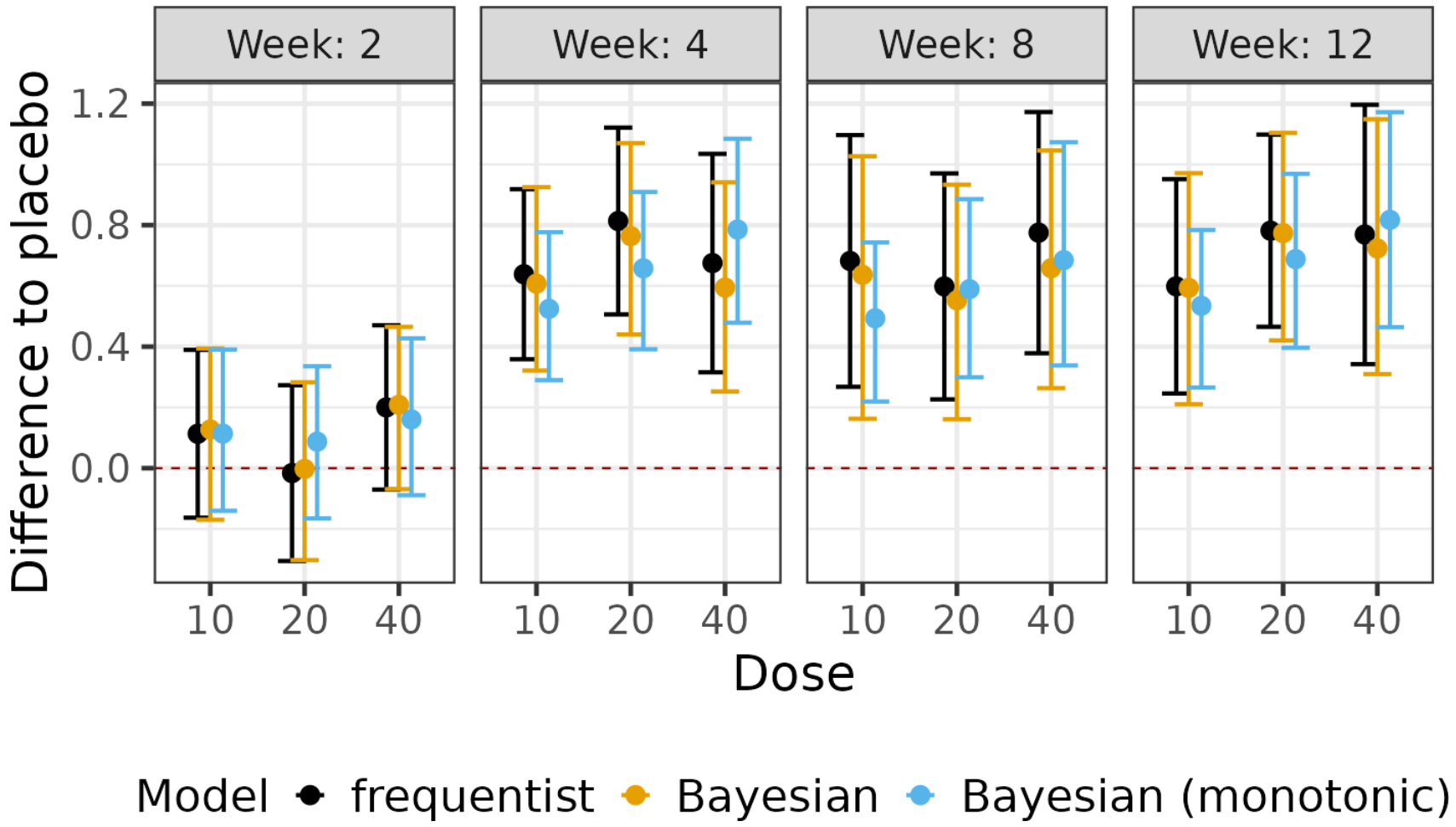
For category $c = 0, \ldots, (\text{categories} - 1)$, the monotonic term is

$$\text{coefficient} \times (\text{categories} - 1) \times \sum_{k=1}^{c} \zeta_k,$$

where $\zeta_k \in [0, 1]$ and $\sum_{k=1}^{\text{categories}-1} \zeta_k = 1$. For more details see this vignette: https://paul-buerkner.github.io/brms/articles/brms_monotonic.html

```
fit_mmrm2 <- brm(
  formula = mmrm_model2,
  data = simulated_data %>% mutate(TRT01P=ordered(TRT01P)),
  prior = mmrm_prior1,
  ...)
```

Results from different MMRMs



NOVARTIS

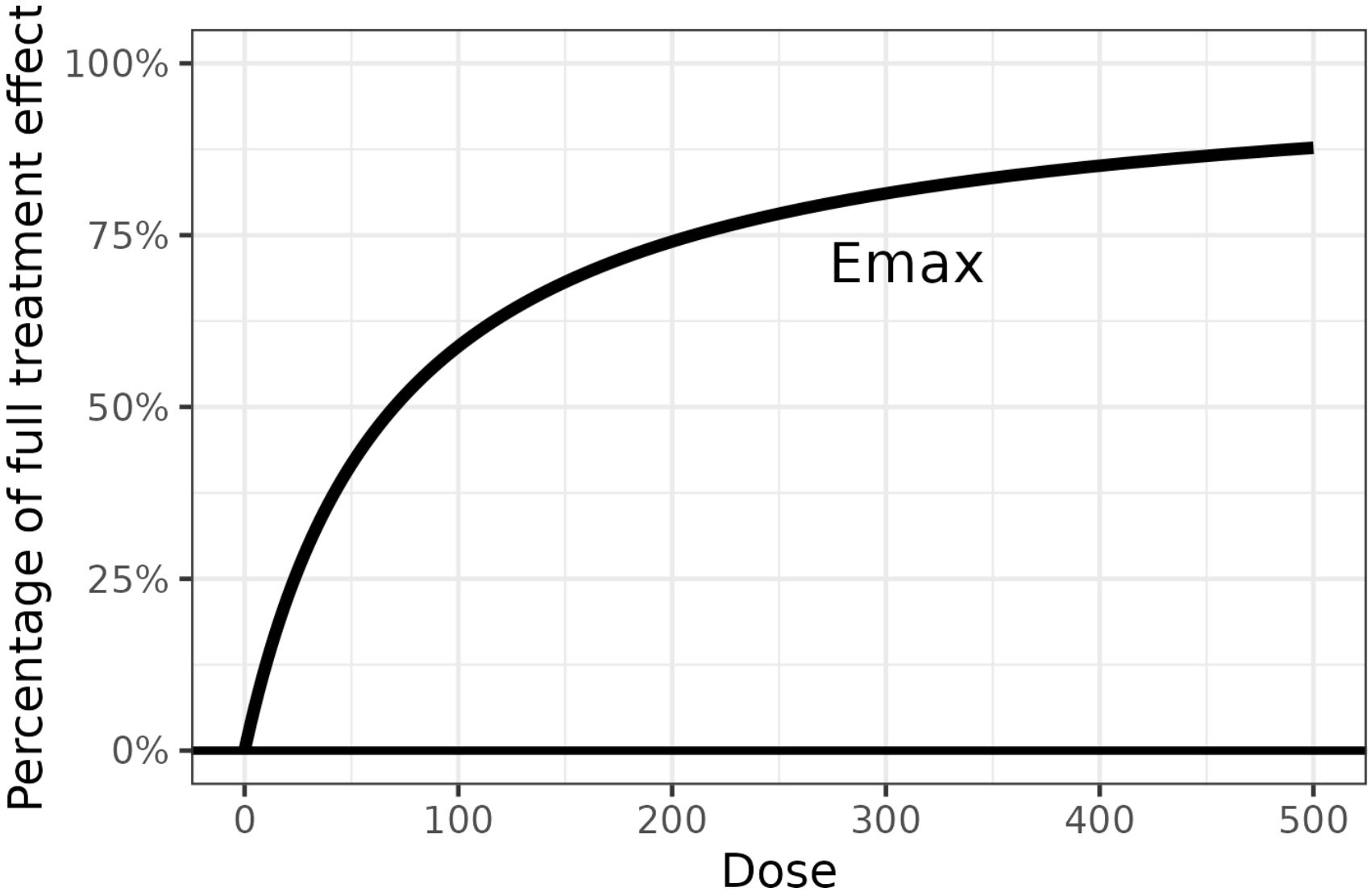
MMRMs: Outlook

In the case study on http://opensource.nibr.com/bamdd you additionally find:

- Data and full code
- More on estimands, parametrization, contrasts & setting priors
- Estimating average differences across visits
- Meta-analytic combined (MAC) approach using historical data
- Robustifying MAC via a "slab-and-spike"-type prior
- Non-linear functions over time & doses in MMRMs
- Excercises

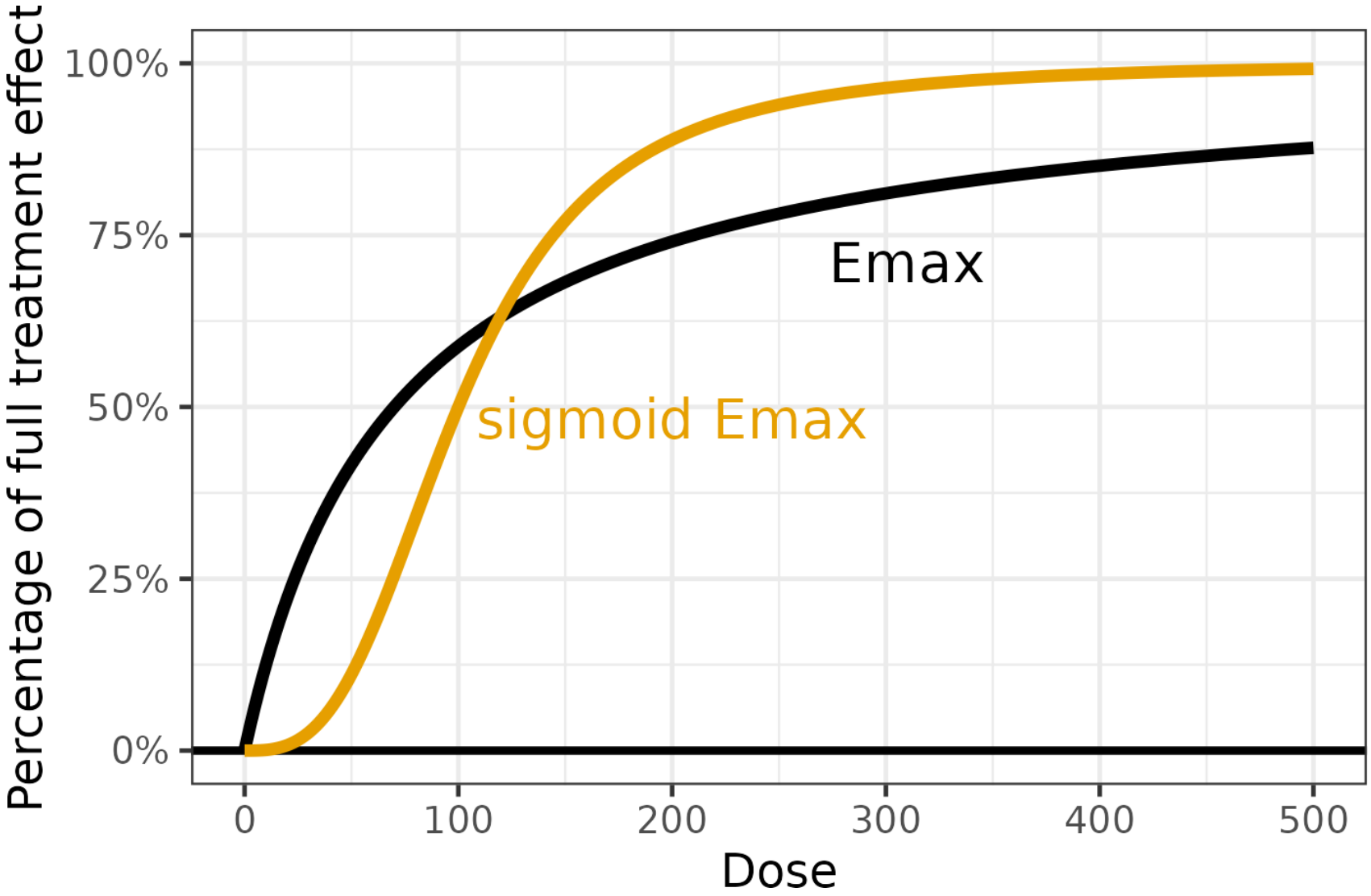
Case study 3: Dose finding

Typical dose response shapes



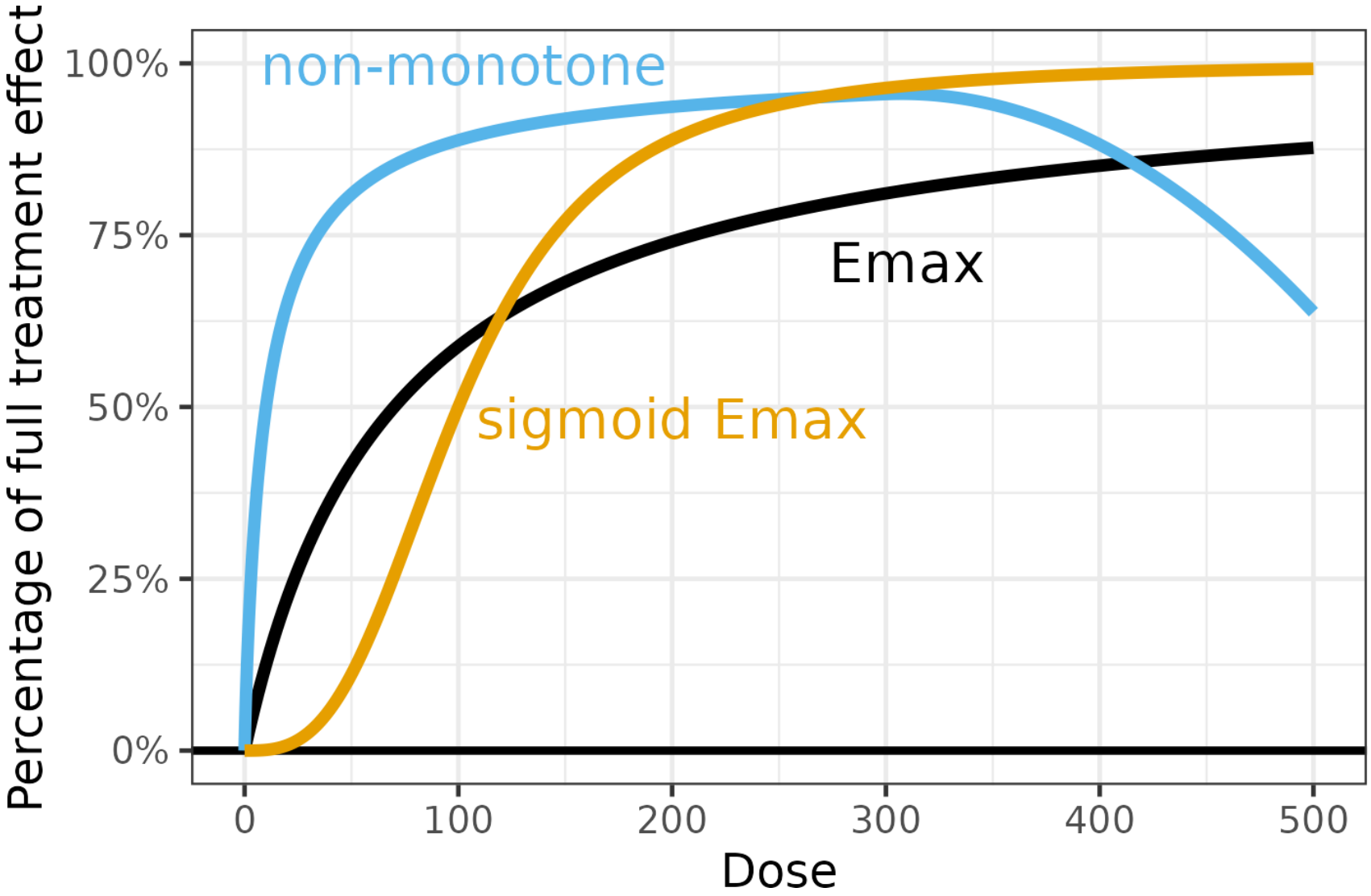
NOVARTIS

Typical dose response shapes



NOVARTIS

Typical dose response shapes



NOVARTIS

Data Set PATHWAY

- Placebo controlled trial
- Treatment of severe asthma with tezepelumab
- Three different doses + placebo
- Endpoint: annualized rate of asthma exacerbations
- Estimates per arm from negative binomial regression (like in "arm-based meta-analysis"), not individual patient data

dose	group	log_est	log_stderr
0	placebo	−0.400	0.103
70	tezepelumab 70 mg q4w	−1.347	0.177
210	tezepelumab 210 mg q4w	−1.661	0.222
560	tezepelumab 280 mg q2w	−1.514	0.191

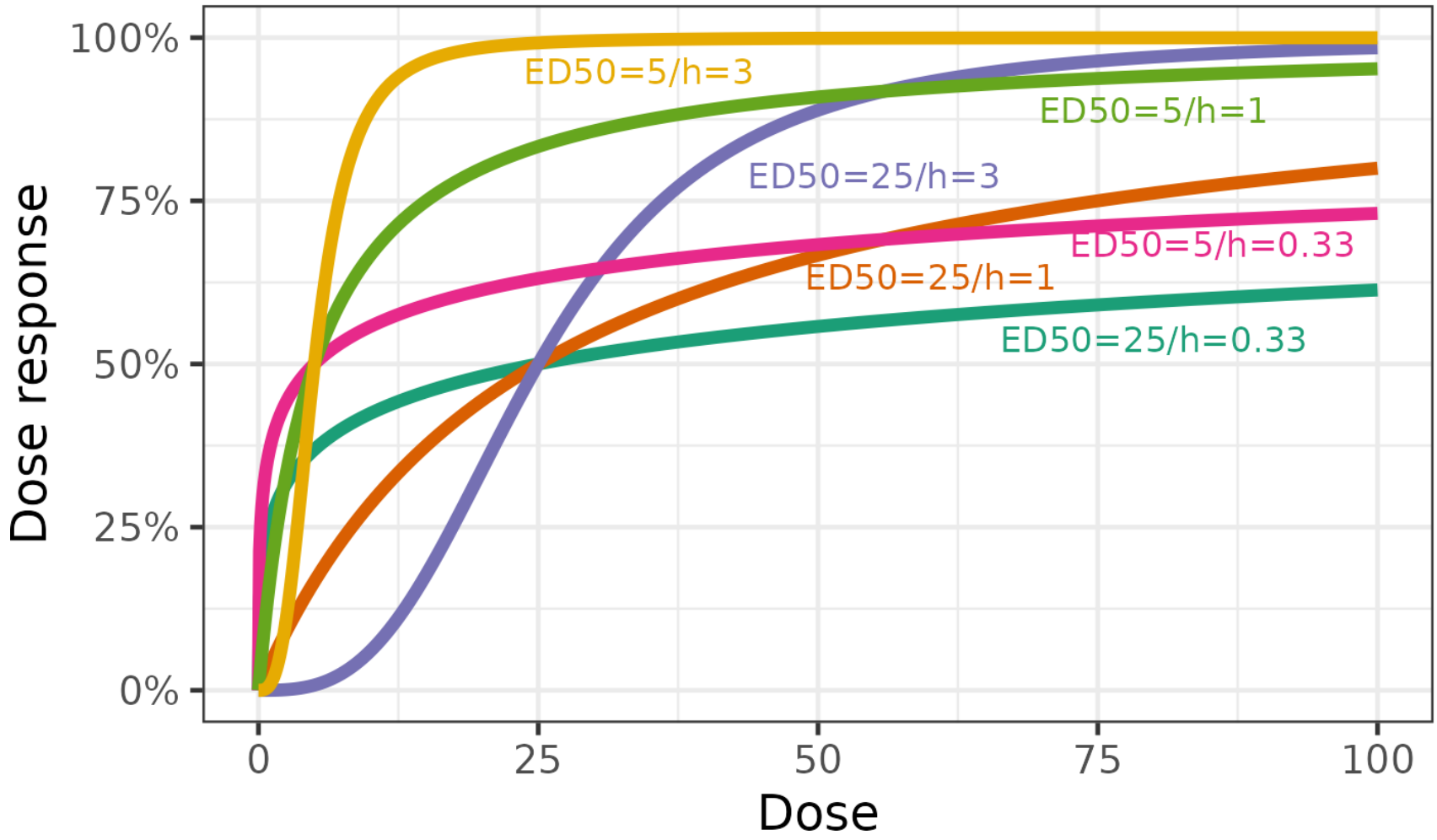
Sigmoid Emax Model

$$f(\text{dose}; \text{parameters}) = \text{E}_0 + \text{E}_{\text{max}} \times \frac{\text{dose}^h}{\text{dose}^h + \text{ED}_{50}^h}$$

Parameters:

- $\text{E}_0 \in \mathbb{R}$: Expected placebo outcome
- $\text{E}_{\text{max}} \in \mathbb{R}$: Maximum effect size
- $h \in \mathbb{R}_+$: Hill (steepness) parameter
- ED_{50}: Dose at which 50% of E_{max} is reached

Sigmoid Emax Model: Visualization



NOVARTIS

Specifying sigmoid Emax Model with `brms`

```
form_sig <- bf(
  log_est | se(log_stderr) ~ E0 + Emax * dose^h /
                             (dose^h + ED50^h),
  nlf(h ~ exp(logh)), nlf(ED50 ~ exp(logED50)),
  E0 ~ 1, Emax ~ 1, logh ~ 1, logED50 ~ 1,
  nl = TRUE,
  family = gaussian()
)

prior_sig <- prior(normal(0,1), nlpar="E0") +
  prior(normal(0,1), nlpar="logh") +
  prior(normal(0,1), nlpar="Emax") +
  prior(normal(4,2), nlpar="logED50")
```

Fitting the sigmoid Emax Model with `brms`

```
fit_sig = brm(
  formula = form_sig,
  data = pathway,
  prior = prior_sig,
  control = list(adapt_delta = 0.999)
)
```

Sigmoid Emax Model: Results Summary

```
summary(fit_sig)
```

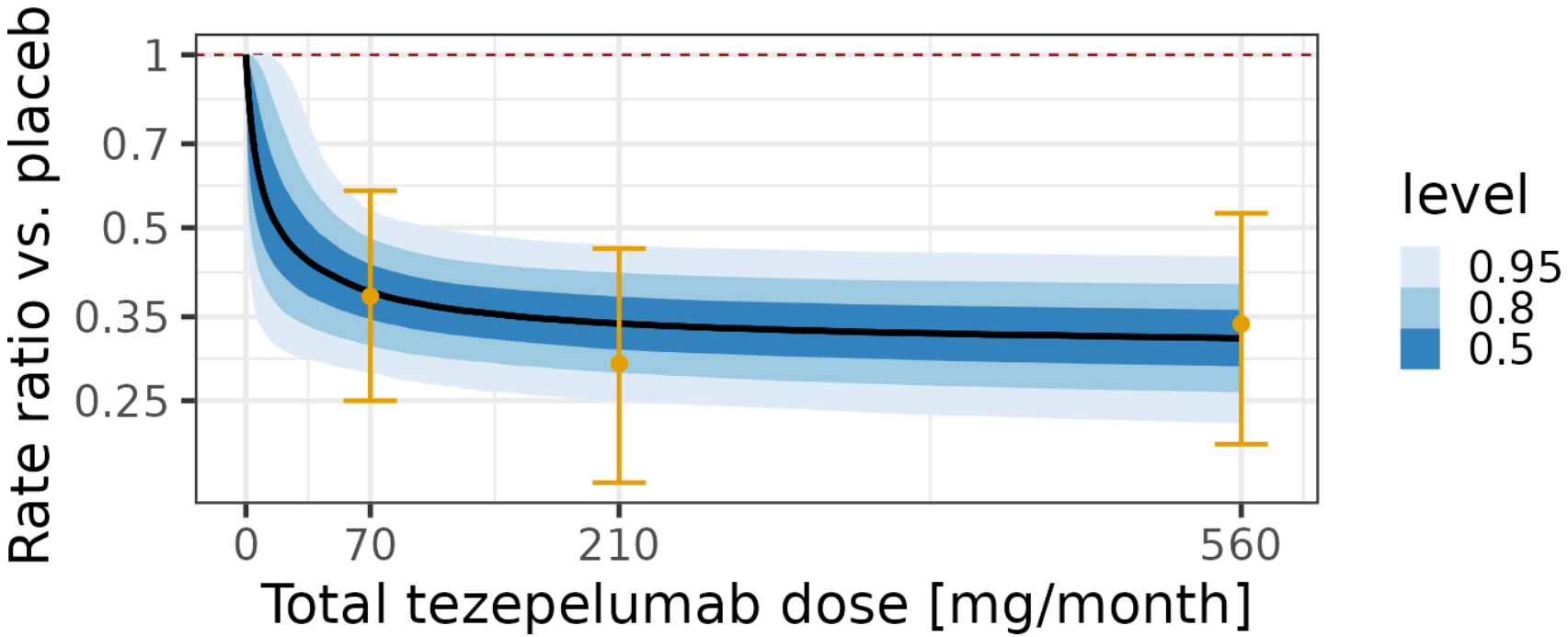
```
 Family: gaussian
  Links: mu = identity; sigma = identity
Formula: log_est | se(log_stderr) ~ E0 + Emax * dose^h/(dose^h + ED50^h)
         h ~ exp(logh)
         ED50 ~ exp(logED50)
         E0 ~ 1
         Emax ~ 1
         logh ~ 1
         logED50 ~ 1
   Data: pathway (Number of observations: 4)

Population-Level Effects:
                 Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
E0_Intercept        -0.42      0.10    -0.61    -0.21 1.00     2065     2179
Emax_Intercept      -1.30      0.32    -2.11    -0.84 1.00     1172     1199
logh_Intercept      -0.08      0.98    -1.90     1.92 1.00     1306     1914
logED50_Intercept    2.73      1.38    -0.27     5.32 1.00     1341     1273

Family Specific Parameters:
      Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
sigma     0.00      0.00     0.00     0.00   NA       NA       NA
```

Visualizing the Fitted Sigmoid Emax Model

```
tibble(dose = seq(0, 560, 1), log_stderr=1) %>%
  add_epred_rvars(object=fit_sig) %>%
  (\(x) x %>%
     left_join(x %>% filter(dose==0) %>% rename(pbo = .epred) %>% dplyr::select(-dose),
               by="log_stderr"))() %>%
  mutate(.delta = .epred - pbo) %>%
  ggplot(aes(x=dose, ydist=.delta)) +
  stat_lineribbon()
```



NOVARTIS

Modified Beta Model

$$f(\text{dose}; \text{parameters}) = \text{E}_0 + \text{E}_{\text{max}} \frac{(\delta_1 + \delta_2)^{(\delta_1+\delta_2)}}{\delta_1^{\delta_1}\,\delta_2^{\delta_2}} \left(\frac{\text{dose}}{S}\right)^{\delta_1} * \left(1 - \frac{\text{dose}}{S}\right)^{\delta_2}$$

Parameters:

- $\text{E}_0 \in \mathbb{R}$: Expected placebo response
- $\text{E}_{\text{max}} \in \mathbb{R}$: Maximum effect size
- $\delta_1, \delta_2 \in \mathbb{R}_+$: Shape parameters
- S: constant $>$ maximum dose, e.g. $1.5 \times \max(\text{dose})$, here we choose S=850

Specifying the Modified Beta Model with `brms`

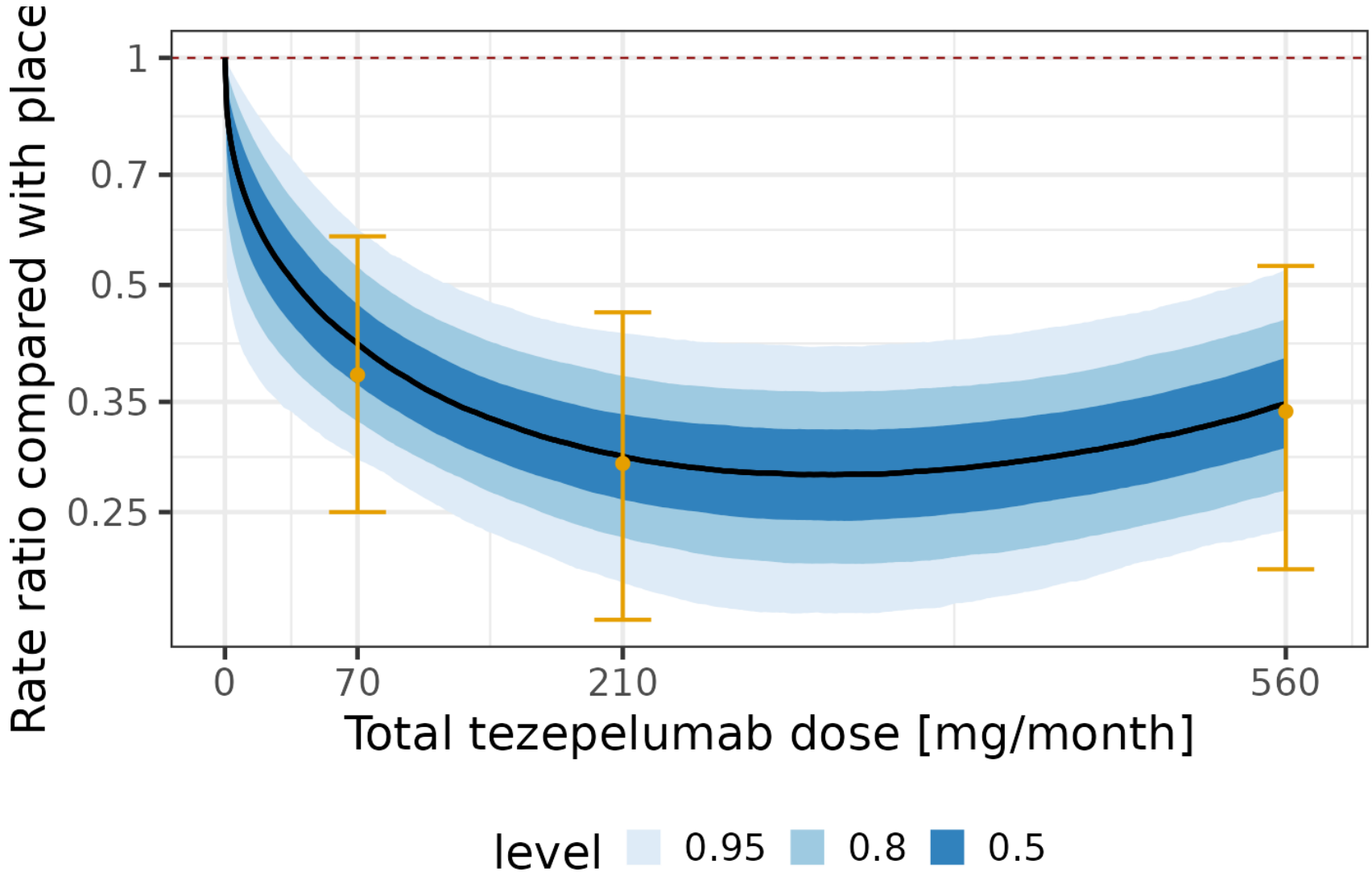
```
form_mbeta <- bf(
  log_est | se(log_stderr) ~ E0 +
    Emax * (delta1+delta2)^(delta1+delta2) /
    (delta1^delta1 * delta2^delta2) *
    (dose/850)^delta1 * (1-dose/850)^delta2,
  nlf(delta1 ~ exp(logdelta1)), nlf(delta2 ~ exp(logdelta2)),
  E0 ~ 1, Emax ~ 1, logdelta1 ~ 1, logdelta2 ~ 1,
  nl = TRUE,
  family = gaussian()
)

prior_mbeta <- prior(normal(0,1), nlpar="E0") +
  prior(normal(0,1), nlpar="Emax") +
  prior(normal(0,1), nlpar="logdelta1") +
  prior(normal(0,1), nlpar="logdelta2")
```

Fitting the Modified Beta Model with `brms`

```
fit_mbeta <- brm(
  form_mbeta,
  data = pathway,
  prior = prior_mbeta,
  control = list(adapt_delta = 0.999)
)
```

Visualizing the Fitted Modified Beta Model



NOVARTIS

Model Evaluation (failed attempt)

```
(loo_mbeta <- loo(fit_mbeta))
```

```
Computed from 4000 by 4 log-likelihood matrix

         Estimate  SE
elpd_loo      0.1 0.4
p_loo         2.2 0.6
looic        -0.2 0.7
------
Monte Carlo SE of elpd_loo is NA.

Pareto k diagnostic values:
                         Count Pct.    Min. n_eff
(-Inf, 0.5]   (good)     1     25.0%   1282
 (0.5, 0.7]   (ok)       0      0.0%   <NA>
   (0.7, 1]   (bad)      3     75.0%   25
   (1, Inf)   (very bad) 0      0.0%   <NA>
See help('pareto-k-diagnostic') for details.
```

Same problem for sigmoid Emax model (`loo(fit_sig)`)

Model Evaluation (failed attempt #2)

```
(loo_mm_mbeta <- loo_moment_match(fit_mbeta, loo_mbeta))
```

```
Computed from 4000 by 4 log-likelihood matrix

         Estimate  SE
elpd_loo     -0.1 0.5
p_loo         2.0 0.6
looic         0.1 1.0
------
Monte Carlo SE of elpd_loo is NA.

Pareto k diagnostic values:
                         Count Pct.    Min. n_eff
(-Inf, 0.5]   (good)     2     50.0%   418
 (0.5, 0.7]   (ok)       1     25.0%   200
   (0.7, 1]   (bad)      1     25.0%   26
   (1, Inf)   (very bad) 0      0.0%   <NA>
See help('pareto-k-diagnostic') for details.
```

Worked for sigmoid Emax model (`loo_moment_match(fit_sig, loo(fit_sig))`)

Model Evaluation (works)

```
(loo_exact_mbeta <- kfold(fit_mbeta, folds = "loo"))
```

```
Based on 4-fold cross-validation

           Estimate  SE
elpd_kfold     -2.0 1.6
p_kfold         4.3 2.1
kfoldic         4.0 3.1
```

Model Comparison

```
loo_compare(loo_mm_sig, loo_exact_mbeta)
```

```
           elpd_diff se_diff
fit_sig     0.0       0.0
fit_mbeta -3.0       1.8
```

```
fit_sig$criteria$loo <- loo_mm_sig
fit_mbeta$criteria$loo <- loo_exact_mbeta
(w_dose <- model_weights(fit_sig, fit_mbeta, weights = "loo"))
```

```
   fit_sig  fit_mbeta
0.95474875 0.04525125
```

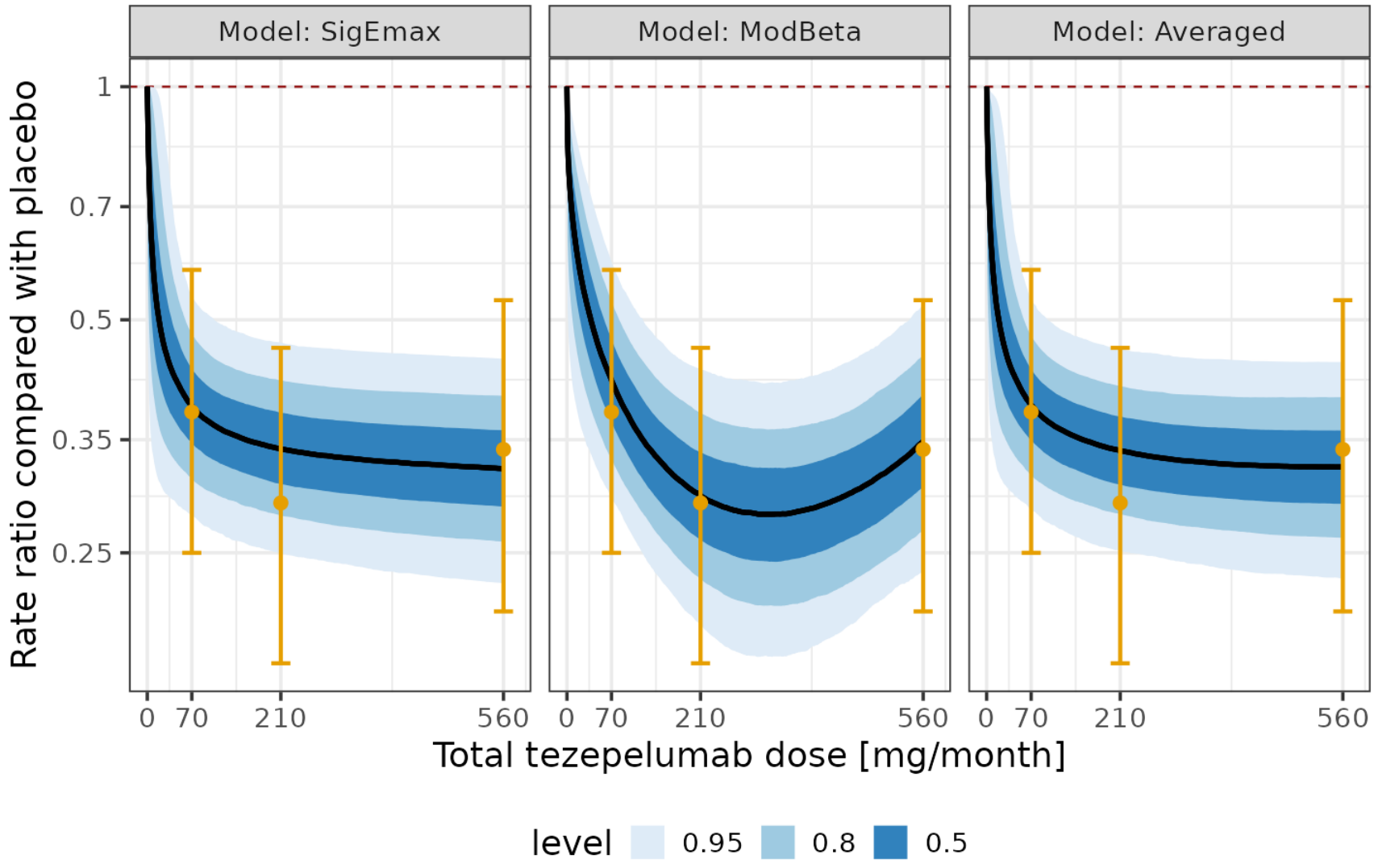
Bayesian Model Averaging

```r
pe_sig <- posterior_epred(fit_sig, newdata = dose_df)
pe_mbeta <- posterior_epred(fit_mbeta, newdata = dose_df)
pe_avg <- pe_sig * w_dose[1] + pe_mbeta * w_dose[2]
```

```r
pe_avg <- pe_avg %>%
  posterior_summary() %>%
  as.data.frame() %>%
  bind_cols(dose_df)
```

```
    Estimate  Est.Error       Q2.5      Q97.5      dose
1 -0.4155670 0.09671043 -0.5997071 -0.2250888  0.000000
2 -0.8377918 0.29297796 -1.4016708 -0.3511670  5.656566
3 -0.9668007 0.29058028 -1.4621438 -0.4051060 11.313131
4 -1.0534117 0.27577332 -1.4997970 -0.4525746 16.969697
```

Visualizing the Model Averaging

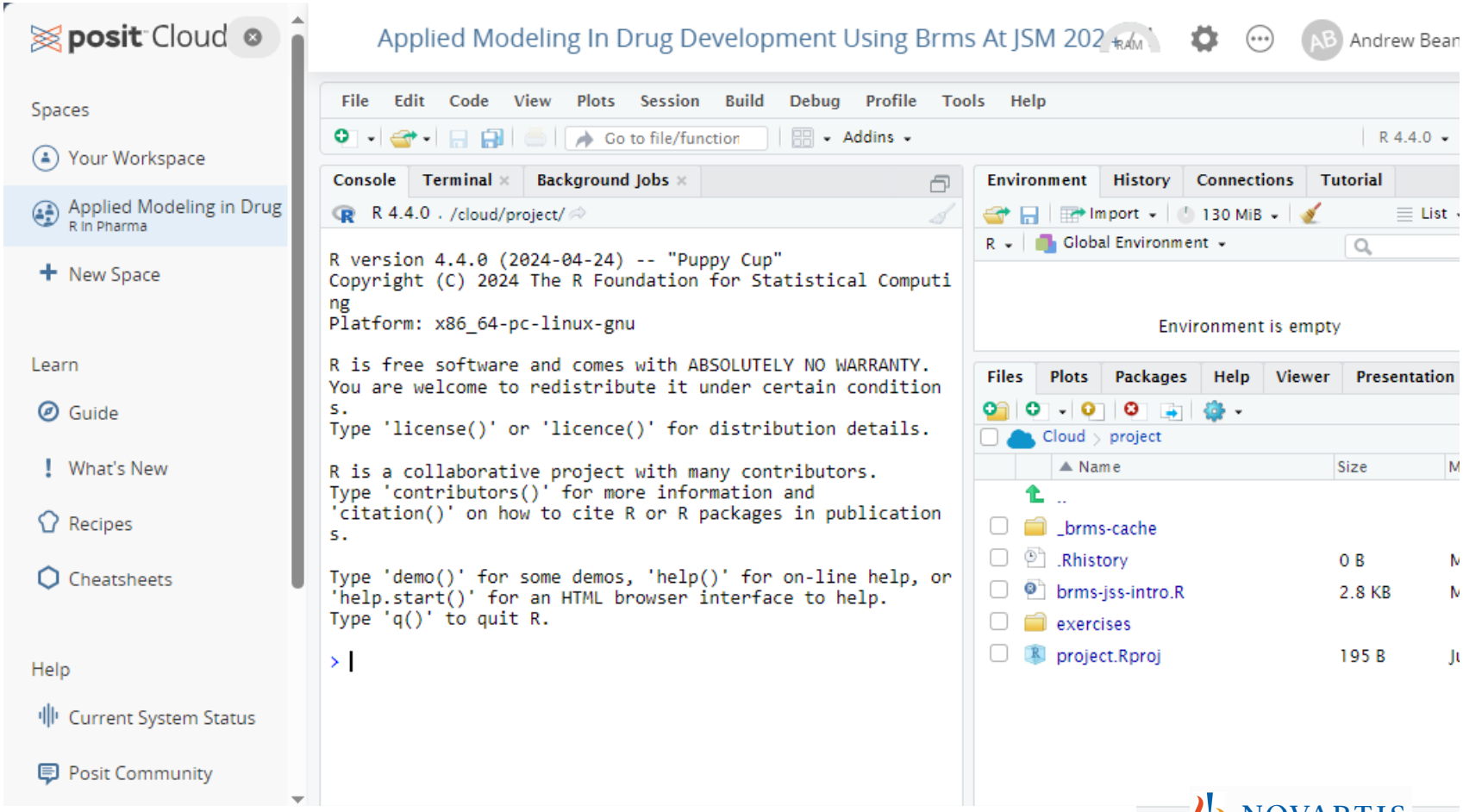


NOVARTIS

Hands-on exercises: dose finding

Guided exercise 2: Open `exercises/ex2_dose_finding.qmd`

Access steps are the same as in the first set of exercises, and then from here, open `exercises/ex2_dose_finding.qmd`



NOVARTIS

Case study 4: Time-to-event data

Overview and Analysis Goals

- Oncology late phase trial to evaluate efficacy of an active drug given in addition to two similar standard of care (SoC-A and Soc-B), which vary geographically
- A total of 4 trial arms active/control combined with SoC-A / SoC-B are studied
- Analysis needs to account for:
  - The efficacy of SoC-A and SoC-B are known to be similar
  - Active drug efficacy is expected to be consistent with SoC-A and SoC-B $\rightarrow$ interest in average treatment effect
- Key analysis goal: Need to control parametrization of model to reflect prior knowledge on similarity and increase efficiency in estimating average treatment effect

Simulated Data Set

First few rows of the simulated dataset:

y	event	trt	soc	arm
7.6954096	0	ctl	ChA	ctlChA
0.0950267	0	act	ChA	actChA
4.7481606	0	ctl	ChA	ctlChA
2.7468766	0	act	ChA	actChA
3.6137101	1	ctl	ChA	ctlChA
0.9358058	1	act	ChA	actChA
0.2591939	1	ctl	ChA	ctlChA
9.2119778	1	act	ChA	actChA

Contrasts: Math

Overall mean (intercept):

$$\mu = \frac{1}{4}(\mu_{actChA} + \mu_{ctlChA} + \mu_{actChB} + \mu_{ctlChB})$$

Average difference between the active and control arms:

$$\delta_{avg.diff} = \frac{1}{2}([\mu_{actChA} - \mu_{ctlChA}] + [\mu_{actChB} - \mu_{ctlChB}])$$

Half of the difference in treatment effect between the two SOC:

$$\delta_{effect} = \frac{1}{2}([\mu_{actChA} - \mu_{ctlChA}] - [\mu_{actChB} - \mu_{ctlChB}])$$

Difference between the two control arms:

$$\delta_{control} = -\mu_{ctlChA} + \mu_{ctlChB}$$

Contrasts: Inverse Matrix

First specify the groups as a function of the contrasts:

```
cc_inv
```

```
              arm
contrast       actChA ctlChA actChB ctlChB
  intercept      1/4    1/4    1/4    1/4
  effectAvg      1/2   -1/2    1/2   -1/2
  deltaEffect    1/2   -1/2   -1/2    1/2
  deltaControl     0     -1      0      1
```

Contrasts: Contrast Matrix

Then invert the matrix to get the actual contrast matrix:

```
cc <- solve(cc_inv)
```

```
        intercept effectAvg deltaEffect deltaControl
actChA          1       1/2           1        -1/2
ctlChA          1      -1/2           0        -1/2
actChB          1       1/2          -1         1/2
ctlChB          1      -1/2           0         1/2
```

The Weibull Family in brms

When using family `weibull` in brms, we are modeling the time until the event, **not** the hazard function!

Parameterize as mean μ and shape α such that, with scale $s = \mu / \Gamma(1 + \frac{1}{\alpha})$:

$$\text{Weibull}(t) = \frac{\alpha}{s} \left(\frac{t}{s} \right)^{\alpha - 1} \exp\left(- \left(\frac{t}{s} \right)^{\alpha} \right)$$

- This is an accelerated failure time model since the survivor function has the property of $S_i(t) = S_{\text{weibull}}(t / \mu_i)$.
- When using a log linear model on μ the regression coefficients are interpretable as relative speedup/slowdown of the process progression

Specify brms Weibull Model

```
model_weibull1 <- bf(y | cens(1-event) ~ 1 + arm,
                     family=weibull())
```

```
prior_weibull1 <-
  prior(normal(meanInter, log(4)/1.64), class="Intercept") +
  prior(normal(0, sdEffect), coef=armeffectAvg) +
  prior(normal(0, sdDeltaEffect), coef=armdeltaEffect) +
  prior(normal(0, sdDeltaControl), coef=armdeltaControl) +
  prior(gamma(0.1, 0.1), class=shape)
```

```
stanvars_weibull1 <-
  stanvar(-log(log(2)/8), name = "meanInter") +
  stanvar(log(2)/1.64, name = "sdEffect") +
  stanvar(log(1.25)/1.64, name = "sdDeltaEffect") +
  stanvar(log(1.25)/1.64, name = "sdDeltaControl")
```

Fit brms Weibull Model

```
fit_weibull1 <- brm(
  formula = model_weibull1,
  data = sim,
  prior = weibull_prior1,
  stanvars = stanvars_weibull1,
  ...
)
```

```r
summary(fit_weibull1)
```

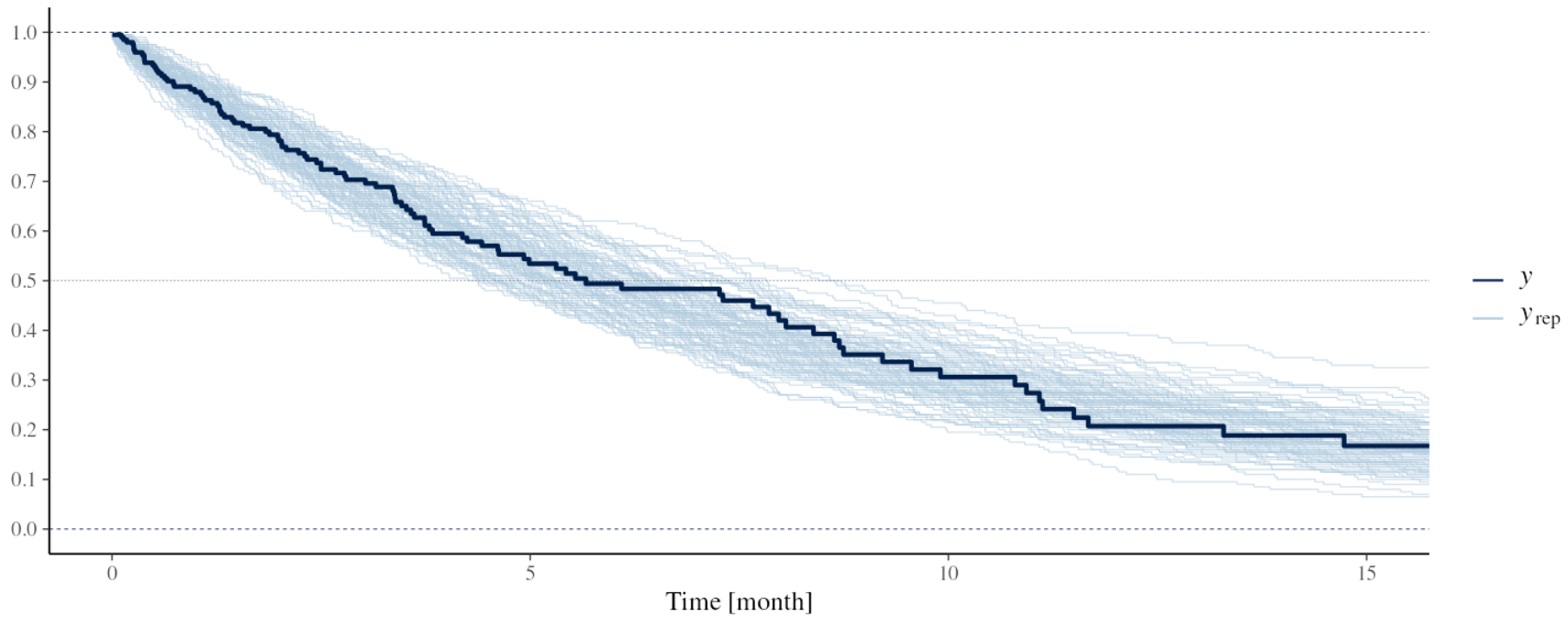
```
 Family: weibull
  Links: mu = log; shape = identity
Formula: y | cens(1 - event) ~ 1 + arm
   Data: sim (Number of observations: 200)

Population-Level Effects:
               Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
Intercept          2.16      0.12     1.94     2.42 1.00     3759     2829
armeffectAvg       0.27      0.19    -0.10     0.65 1.00     4386     2915
armdeltaEffect     0.00      0.11    -0.21     0.20 1.00     4763     3006
armdeltaControl    0.06      0.12    -0.17     0.29 1.00     4445     3215

Family Specific Parameters:
      Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
shape     0.97      0.08     0.83     1.12 1.00     3514     2838
```

Posterior Predictive Checks

```
p_full_fup <- pp_check(
  fit_weibull1, type = "km_overlay",
  status_y = sim$event, ndraws = 100
)
```



The model predictions assume no censoring

NOVARTIS

Time-to-Event Modeling: Outlook

In the case study on https://opensource.nibr.com/bamdd you additionally find:

- Additional details and model justification based on a real dataset
- Include historical data of average SoC
- Add custom coded contrasts to further improve flexibility of historical data analysis

Course wrap-up

Summary

- Diverse opportunities for applied modelling to inform good drug-development decisions
- Bayesian paradigm is well suited for many of these situations
  - Availability of meaningful prior information
  - Desire for probabilistically interpretable statements about unknowns and future observable quantities
- `brms` is a powerful and highly flexible engine for applied modelling , facilitating straightforward model specification and inference

Looking ahead

- We hope you have:
  - Become familiar with `brms` syntax and workflow
  - Seen its versatility for statistical modelling in drug development
  - Gained hands-on experience with the package from guided exercises
- And that you feel empowered to use `brms` the future!

Resources

- Our open-source book: Bayesian Applied Modelling in Drug Developoment (BAMDD)
- `brms` documentation: https://paul-buerkner.github.io/brms/
- Stan homepage: https://mc-stan.org
- Stan Forums: https://discourse.mc-stan.org/

Thank you

- Thank you for your interest and participation!
- Our contact information:
  - David: david.ohlssen [at] novartis [dot] com
  - Andrew: andrew.bean [at] novartis [dot] com
  - Björn: bjoern.holzhauer [at] novartis [dot] com