Bayesian methods and thinking in Pfizer's PharmaTherapeutics Division.

How we got here, what we've done and where we're headed.

Phil Woodward
VP Global Head of BioTherapeutics Statistics, Pfizer

Pfizer WORLDWIDE RESEARCH & DEVELOPMENT

Summary

- How did change come about?
- How it was actively managed
- Education, training and software
- What are we doing?
- The future?

How did change come about?

- Dissatisfaction with the current way of working
  - Usual issues with convoluted explanations of classical methods
  - Desire by all to interpret results as if they were Bayesian
  - Unsatisfactory approach to current use of Bayesian methods
- Vision of how things can be better
  - If we want Bayesian interpretation than let's be Bayesian!
  - Benefits of formally incorporating prior knowledge
  - Ability to directly tackle the questions of interest
- Pragmatic and strong focus on benefits to the Business
  - Why should others support and join us to make these changes?
  - What are the immediate tangible benefits? Gain momentum.
  - What will be the longer term benefits? Prepare groundwork.

How did change come about?

- Someone with the authority & passion to bring about change
  - Head of Statistics supporting projects up to Proof-of-Concept
  - Backed up by a manager with the same vision
  - Colleagues enthusiastic by the challenge to make these changes
  - Allies in partner lines, e.g. Clinical Pharmacology, Clinical
- Willingness to take risks
  - Make changes without having all potential issues worked out
  - Leader has the backs of the colleagues implementing changes
  - Belief that things will be better by making the changes
- Address the major obstacles to successful implementation
  - Education, training & software
  - Expertise and experience to “hold the hand” of “Bayesian virgins”
  - Pragmatic approximate methods to minimise delays in getting started

Include in Departmental Goals

A default position of utilising Bayesian methods in all PoM and PoC studies, at least, unless there is good reason not to do so.

... informative priors to reduce the number needed on placebo, and perhaps, standard-of-care

... interim analyses based either on predictive probability of success or failure at the end of the study, or the current posterior probability that the effect size criteria have been met

... additional exploratory use of credible informative priors for the treatment effect

Informative Priors Guidance

Guidance Document produced

1. Select relevant historical data to derive prior
2. Formulate model to relate historical studies to new study
3. Calculate “effective N” given by informative prior
4. Build checks for prior misspecification into analysis
5. Document prior derivation

Summarising Past Evidence.

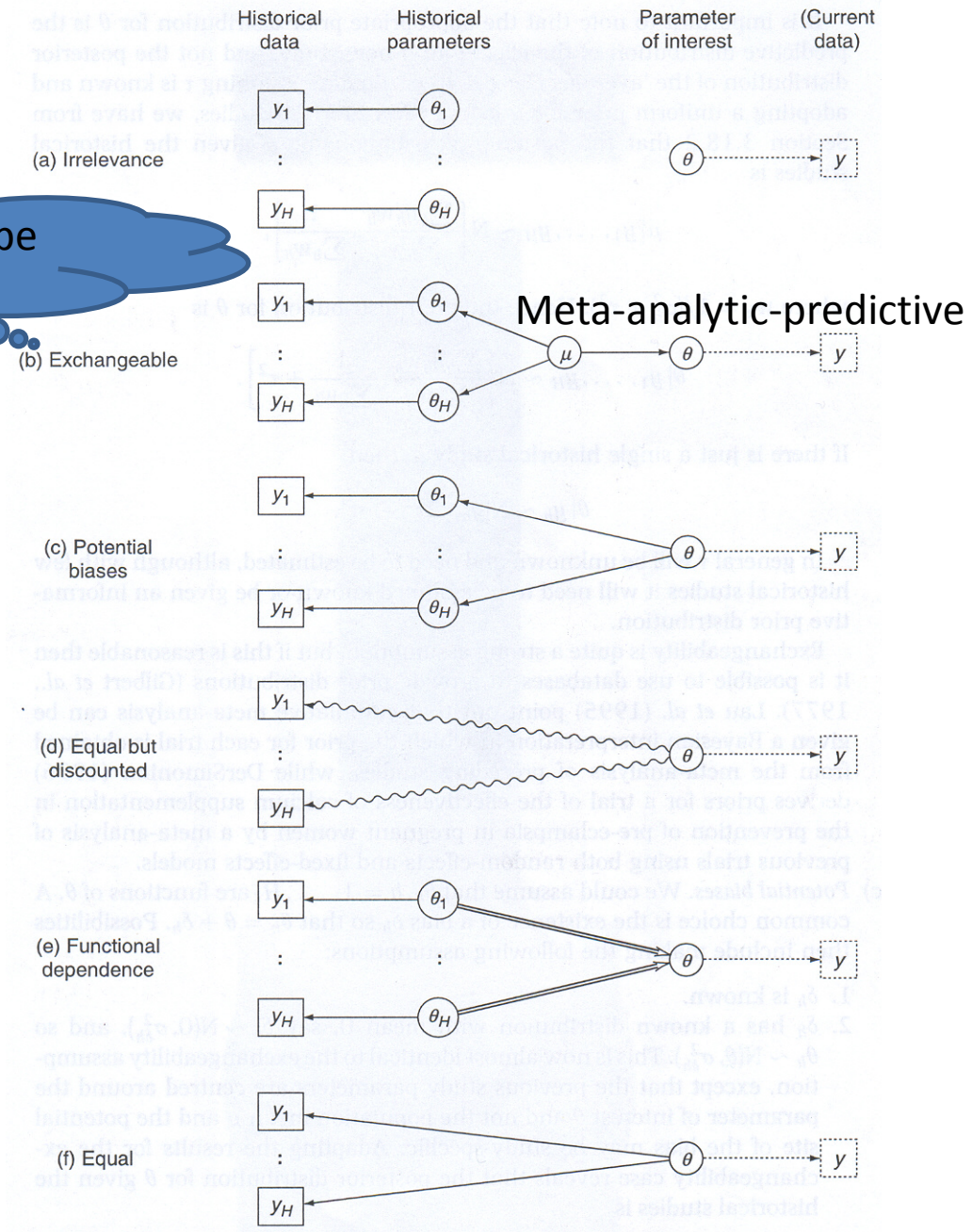
Typically (b) adequate, maybe with more complexity.

Typically,
$y_h \sim N(\theta_h, \sigma_h^2)$

Exchangeable
$\theta, \theta_h \sim N(\mu, \tau^2)$

(a) $\tau = \infty$, $\mu = K$
(b) $\tau \sim$ dist.
(f) $\tau = 0$, $\mu = \theta$

(c) $\theta_h = \theta + \delta_h$
$\delta_h \sim N(0, \sigma_{\delta h}^2)$
$\theta_h \sim N(\theta, \sigma_{\delta h}^2)$



Document prior derivation

- The derivation of the informative prior should be documented in a technical report that is stored along with the source data

- The report should contain the following information:
  - Study background and objectives
  - List of studies used to build prior
  - Criteria used to select these studies
  - Model used to relate historical studies to new study
  - Prior details (i.e. mean and standard deviation)
  - Description/model of how the prior will be used

Bayes Limited Duration Team Objectives

- Review and critique the current proposal
  - Ensure aims & benefits are clear
  - Identify issues with achieving these aims
  - Help identify what, if anything, should be mandatory or consistent
    - Consider trade-off between ease of communicating each project with need for flexibility of methods.
  - Help define and structure the goal to maximise wider engagement & impact

- Plan and assist with implementation
  - How should we address issues raised in 1b?
  - Who are the pioneers – PoC & large PoM studies?
  - Determine and help develop training, software, “self-help buddies”
    - Do we need external expertise to assist?

- Have our preferred way of incorporating Bayesian methods in operation within PTx

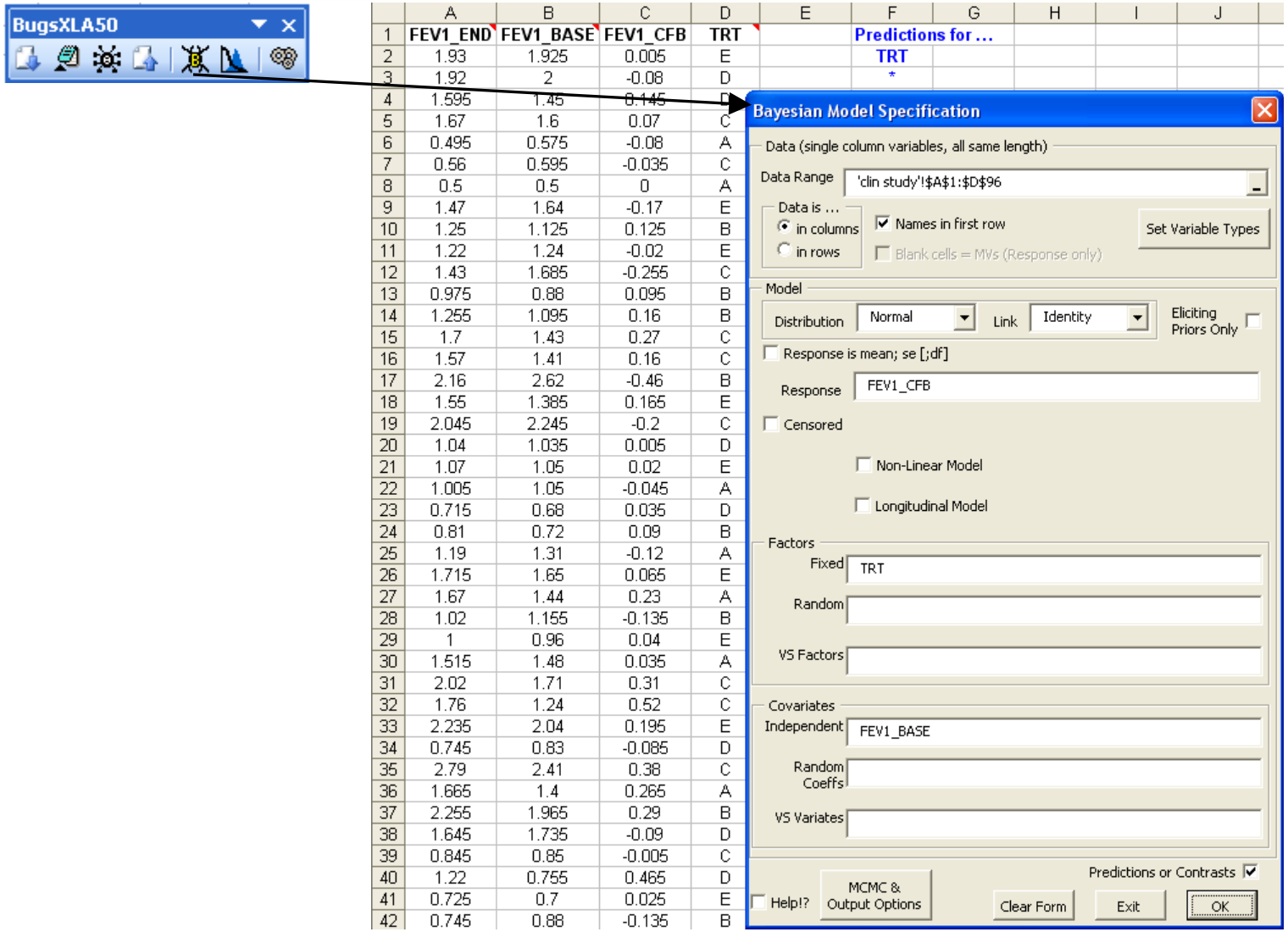
Education, Training & Software

- How is Bayes different? Why should we bother?
- Review theory, but with focus on implementation.
  - It is in our goals now!
- Win/OpenBUGS
  - BugsXLA
- SAS PROC MCMC
- R Scripts for study design operating characteristics
- Pragmatic approximations

How does Bayes add value? (statistical audience)

- Informative Prior
  - Natural approach for incorporating information already available
  - Smaller, cheaper, quicker and more ethical studies
  - More precise estimates and more reliable decisions
  - Sometimes weakly informative priors can overcome model fitting failure
- Probability as a "degree of belief"
  - Quantifies our uncertainty in any unknown quantity or event
  - Answers questions of direct scientific interest
    - P(state of world | data) rather than P(data* | state of world)
- Model building and making inferences
  - Nuisance parameters no longer a "nuisance"
  - Random effects, non-linear terms, complex models all handled better
  - Functions of parameters estimated with ease
  - Predictions and decision analysis follow naturally
  - Transparency in assumptions
- Beauty in its simplicity!
  - $p(\theta \mid x) = p(x \mid \theta)\, p(\theta) / p(x)$
  - Avoids issue of identifying "best" estimators and their sampling properties
  - More time spent addressing issues of direct scientific relevance

BugsXLA
Quick access to the power of Win/OpenBUGS



	A	B	C	D	E	F	G	H	I	J
1	FEV1_END	FEV1_BASE	FEV1_CFB	TRT		Predictions for ...				
2	1.93	1.925	0.005	E		TRT				
3	1.92	2	-0.08	D		*				
4	1.595	1.45	0.145	D						
5	1.67	1.6	0.07	C						
6	0.495	0.575	-0.08	A						
7	0.56	0.595	-0.035	C						
8	0.5	0.5	0	A						
9	1.47	1.64	-0.17	E						
10	1.25	1.125	0.125	B						
11	1.22	1.24	-0.02	E						
12	1.43	1.685	-0.255	C						
13	0.975	0.88	0.095	B						
14	1.255	1.095	0.16	B						
15	1.7	1.43	0.27	C						
16	1.57	1.41	0.16	C						
17	2.16	2.62	-0.46	B						
18	1.55	1.385	0.165	E						
19	2.045	2.245	-0.2	C						
20	1.04	1.035	0.005	D						
21	1.07	1.05	0.02	E						
22	1.005	1.05	-0.045	A						
23	0.715	0.68	0.035	D						
24	0.81	0.72	0.09	B						
25	1.19	1.31	-0.12	A						
26	1.715	1.65	0.065	E						
27	1.67	1.44	0.23	A						
28	1.02	1.155	-0.135	B						
29	1	0.96	0.04	E						
30	1.515	1.48	0.035	A						
31	2.02	1.71	0.31	C						
32	1.76	1.24	0.52	C						
33	2.235	2.04	0.195	E						
34	0.745	0.83	-0.085	D						
35	2.79	2.41	0.38	C						
36	1.665	1.4	0.265	A						
37	2.255	1.965	0.29	B						
38	1.645	1.735	-0.09	D						
39	0.845	0.85	-0.005	C						
40	1.22	0.755	0.465	D						
41	0.725	0.7	0.025	E						
42	0.745	0.88	-0.135	B						

Bayesian Study Design

Consider a generic decision criterion of the form

$$\text{GO decision if } \Pr(\delta \geq \Delta) > \pi$$

δ is the treatment effect

Δ is an effect size of interest

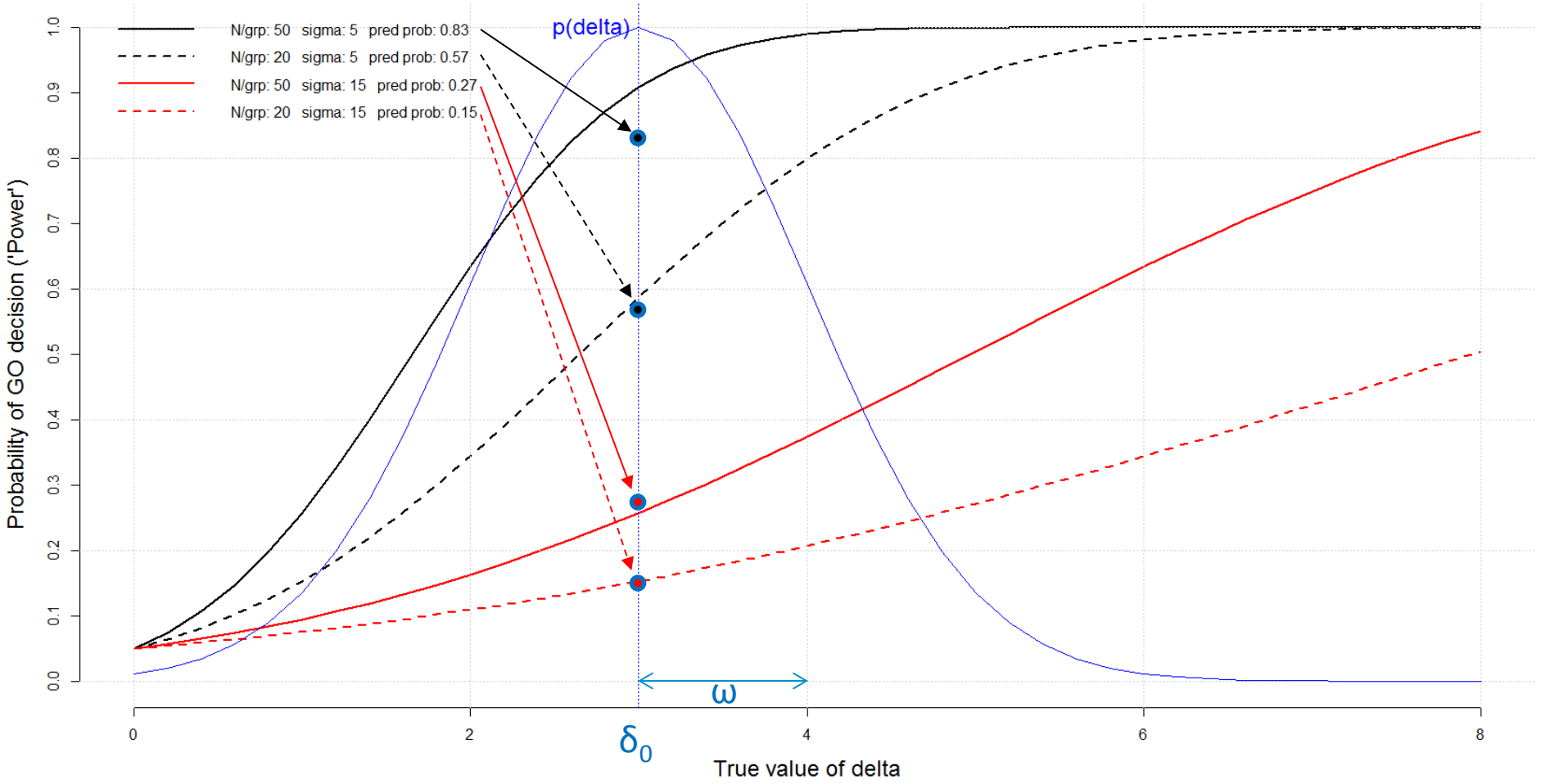
π is the probability required to make a *positive* decision

A Bayesian analogy to significance could be

$$\Pr(\delta > 0) > 0.95$$

Bayesian Study Design (Assurance)

Plot comparing classical ('conditional power') OC and assurance



Bayesian Study Design
(using informative prior to reduce sample size)

Worked example

Suppose predictive distribution (placebo prior)

$p(\gamma) \sim N(18, 12^2)$

Forecast residual standard deviation

(obtained in usual way, not shown here)

$\sigma = 70$

Effective N of placebo prior

$\text{Eff.N} = (70 / 12)^2 = 34$

Design study in usual way, ignoring informative prior. Then reduce placebo arm by 34 and have same power / precision.

Bayesian Study Design
(using informative prior to reduce sample size)

Unless no doubts at all, use Robust Prior

i.e. a mixture of informative and vague prior distributions

p(placebo mean) ~ 0.9 x $N(18, 12^2)$ + 0.1 x $N(18, 120^2)$

Represents 10% chance meta-data not exchangeable
in which case, will effectively revert to vague prior
(can also be thought of as heavy tailed distribution)

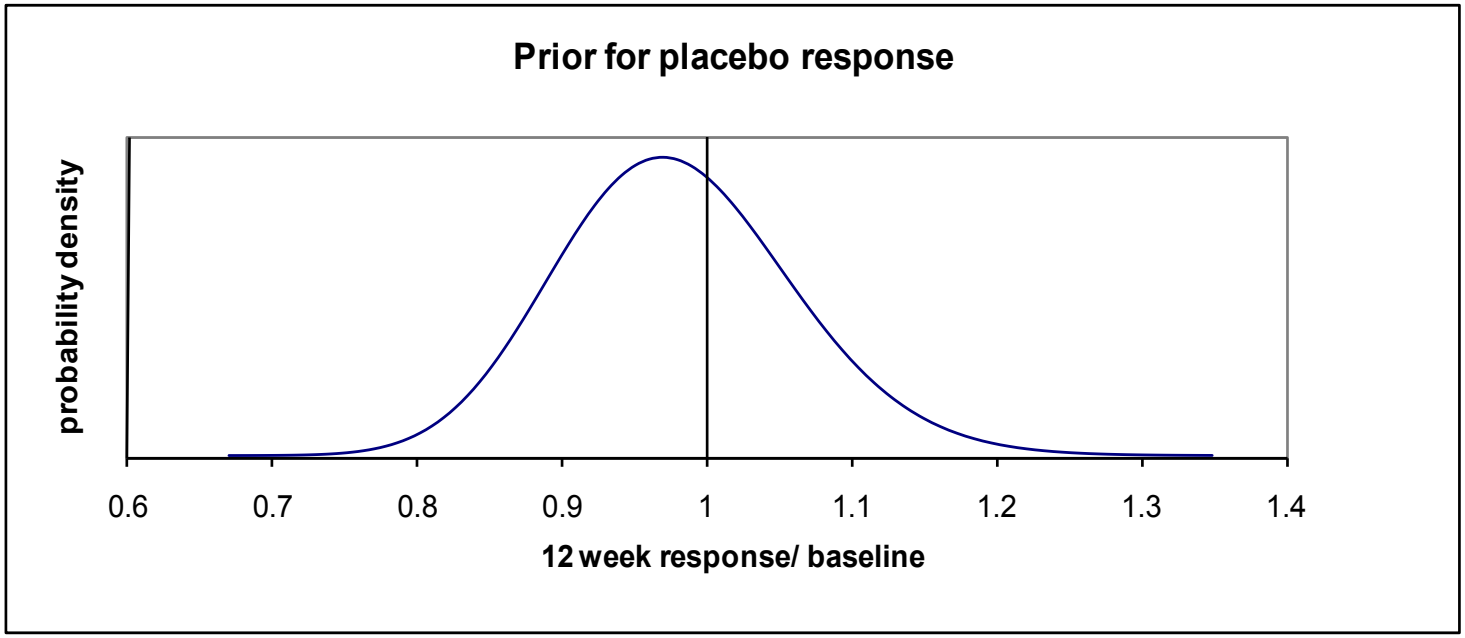
Also compute Bayesian p-value of data-prior compatibility

Pr("> observed mean" | prior ~ $N(18, 12^2)$)

Note: predictive dist. for obs. mean ~ $N(18, 12^2 + \sigma^2 / n_P)$

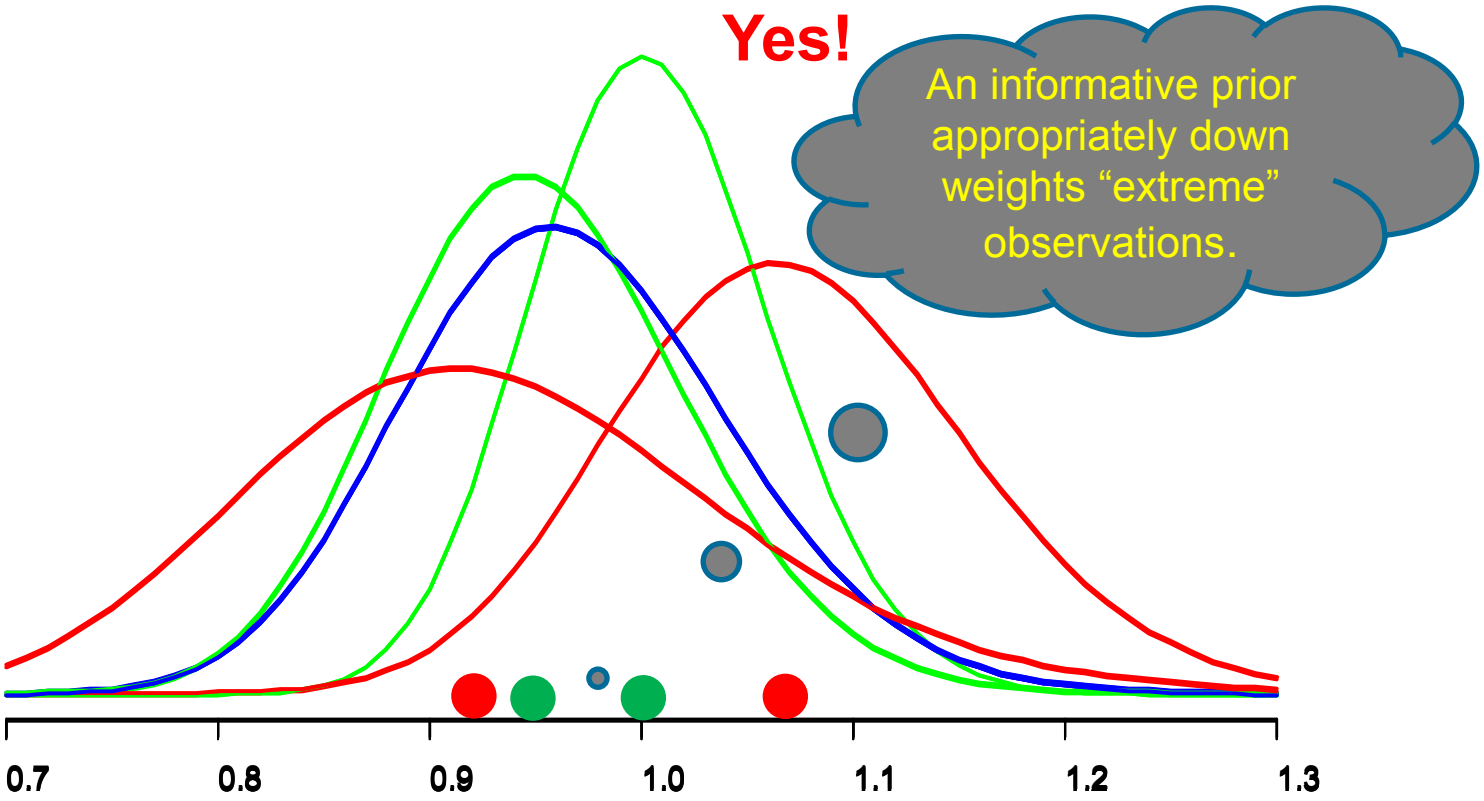
Diabetic Nephropathy PoC

- Expert elicitation and consistent with the literature
  - 632 placebo subjects in two large studies
  - Uncertainty in similarity with planned study accounted for
- Bayesian approach used this information in PoC study
  - Prior knowledge of placebo response equivalent to 100 subjects
  - Study completed 12 months sooner and with >$5M saving



Diabetic Nephropathy Biomarker

Were the observed placebo data consistent with the prior?

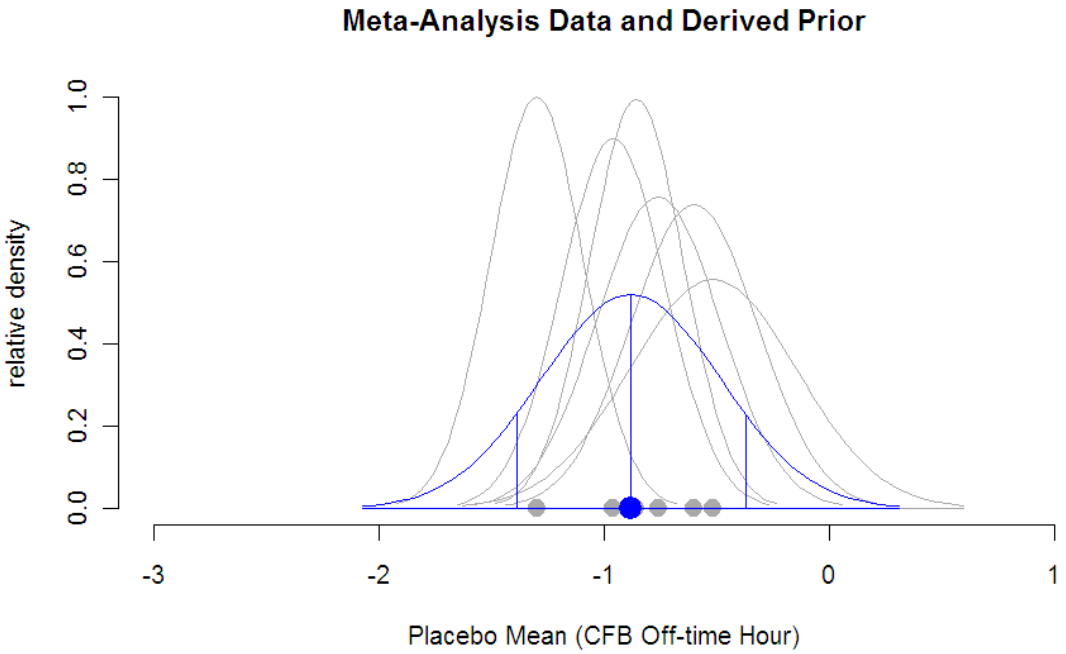


Also ran an identical study with a different compound.

Yes again!

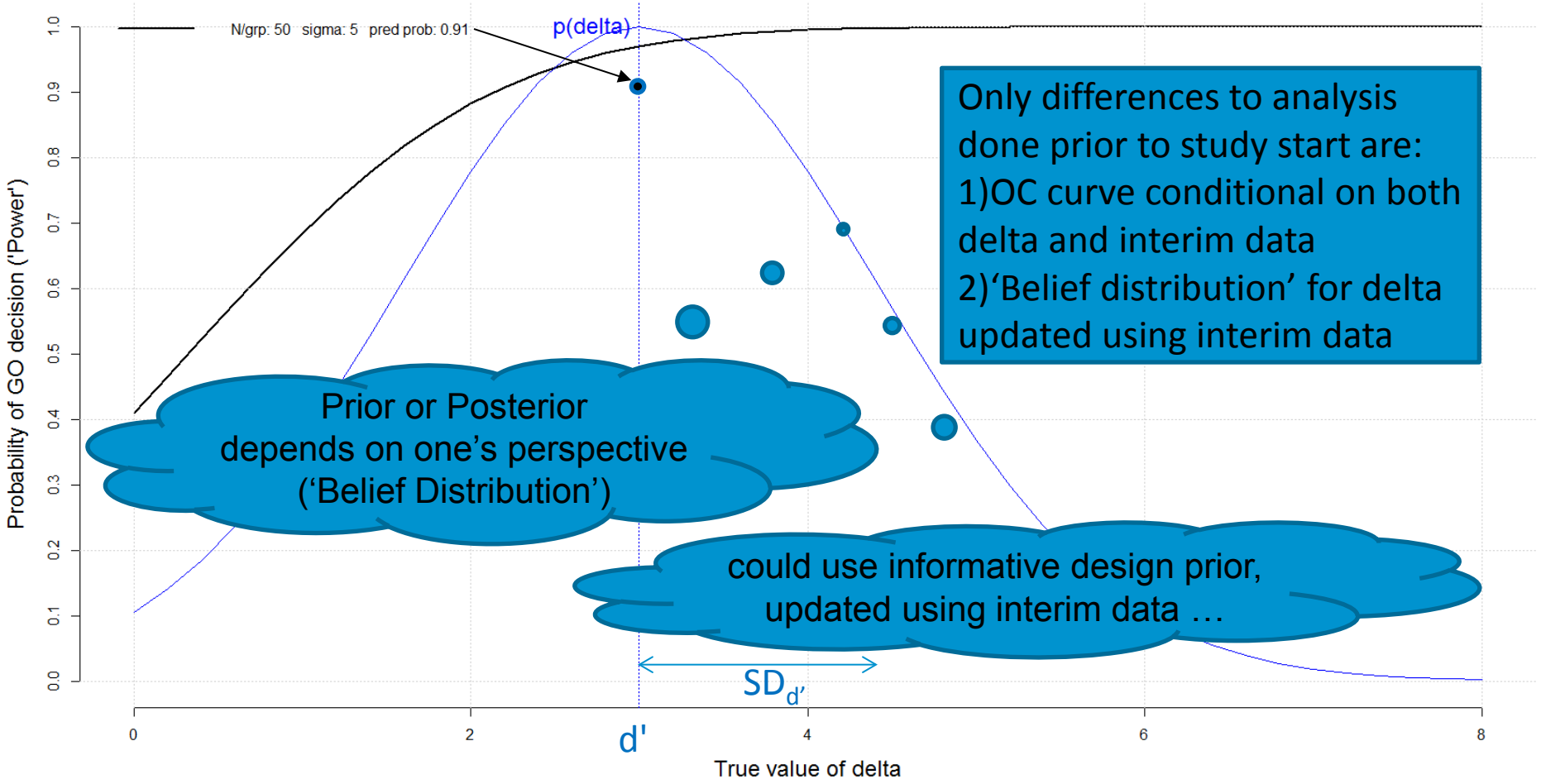
Parkinsons Disease Off-time

- Utilised 6 recent PD L-Dopa studies
  - 707 placebo subjects in total: 2008-2013
- Bayesian approach uses this information in planned study
  - Prior knowledge of placebo response equivalent to 53 subjects
  - Expected to save ~3 months and >$2M



Bayesian Interim Analysis
(Interim analysis predictive probability)

Plot comparing 'conditional power' and predictive probability following interim analysis (25/grp), vague prior distribution



Pragmatic Approximations
Predictive probability of end of study success at an interim.

End of study success criterion:

$$\Pr(\delta > \Delta) > \pi$$

Data at the Interim (vague priors):

$M_I = mean\ estimate\ of\ \delta\ at\ the\ interim\ \ (sometimes\ the\ ML\ estimate)$;
$approximates\ the\ posterior\ mean\ of\ \delta\ at\ the\ interim$
$V_I = variance\ of\ M_I\ \ (sometimes\ the\ squared\ standard\ error\ of\ M_I)$;
$approximates\ the\ posterior\ variance\ of\ \delta\ at\ the\ interim$

$$\mathbf{1 - \Phi\left(\sqrt{\frac{V_U}{V_I}}\left(z_\pi - \frac{(M_I - \Delta)}{\sqrt{V_E}}\right)\right)}$$

When we have exactly half the information at the interim, this simplifies:

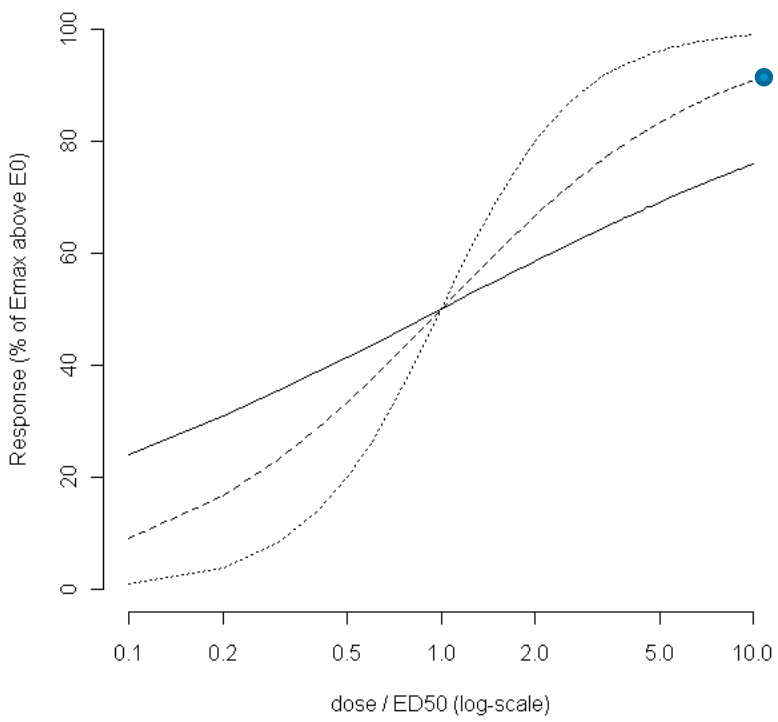
$$\mathbf{1 - \Phi\left(z_\pi - \frac{(M_I - \Delta)}{\sqrt{V_I/2}}\right)}$$

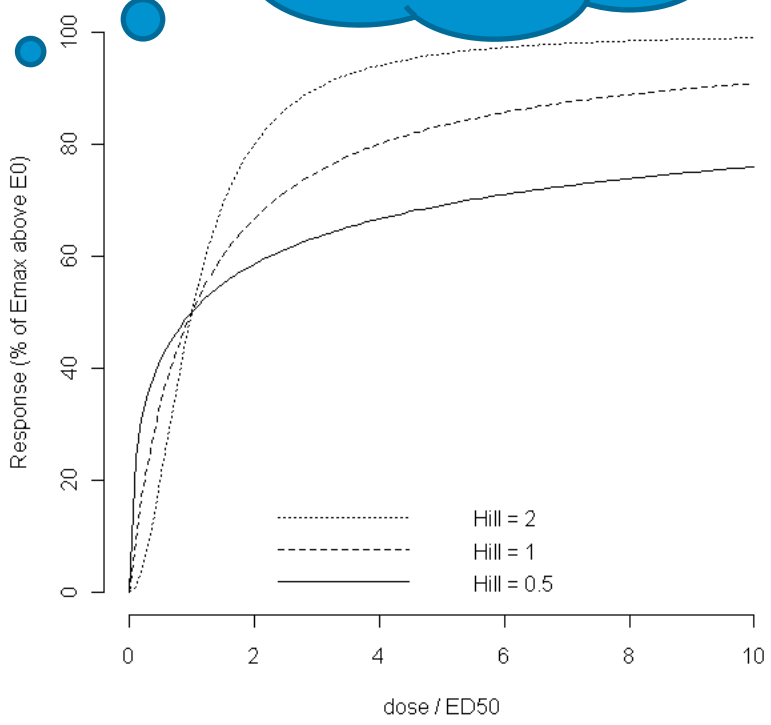
Bayesian Emax Model

$$E_0 + \frac{E_{max} X^{\lambda}}{ED_{50}^{\lambda} + X^{\lambda}}$$

λ sometimes referred to as 'Hill slope'

when λ = 1 need ~80 fold range to cover ED_{10} to ED_{90}

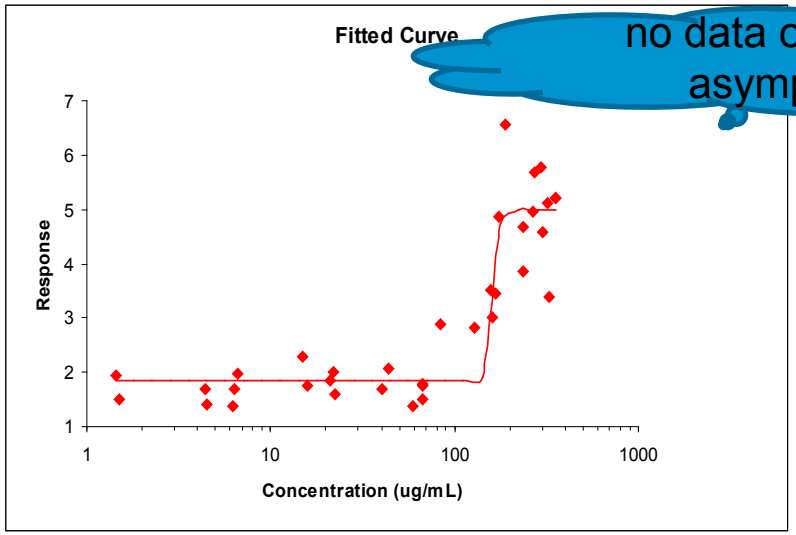




Bayesian Emax Model

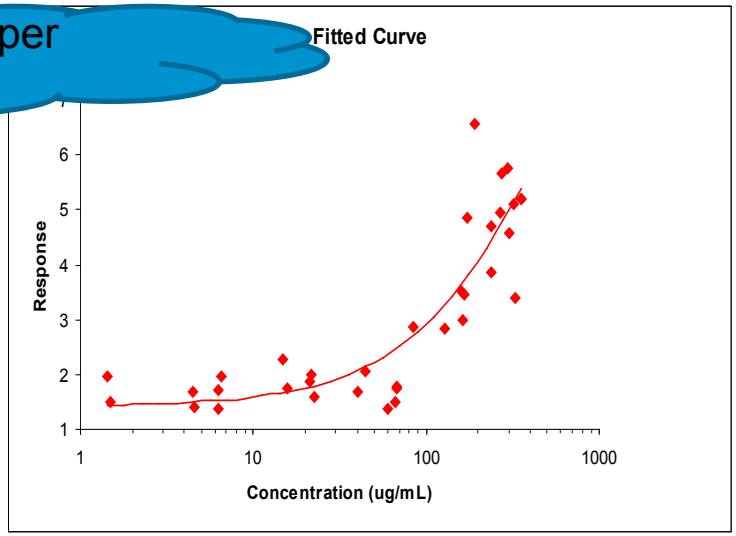
Convergence issues are common with MLE of Emax models

Hill slope not restrained



no data on upper asymptote

Hill slope = 1

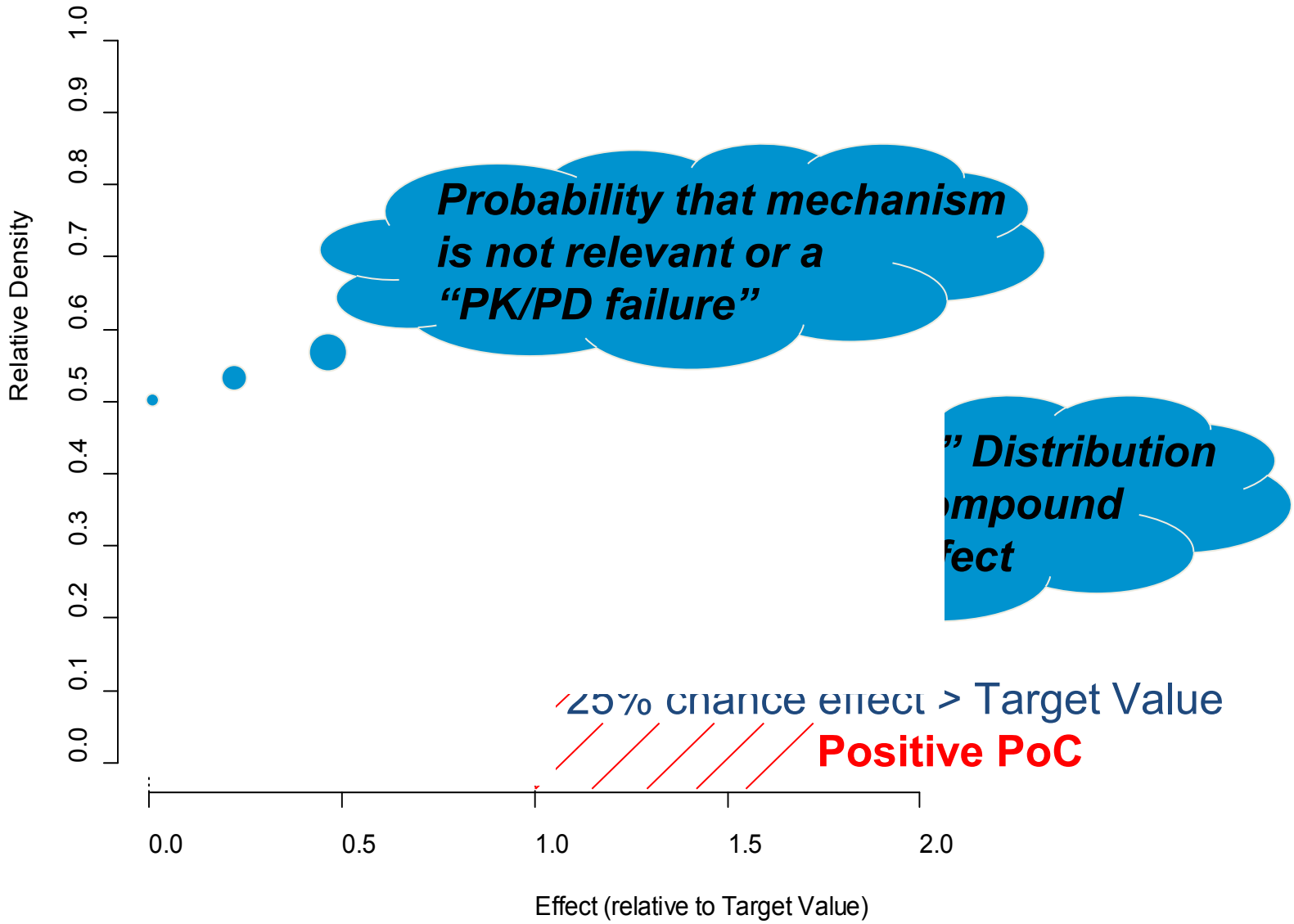


Most clinical data more variable than this and smaller dose range

Classical fitting algorithms can fail to provide any solution

Utilise prior from other relevant studies to avoid these two extremes

The Future?
A more realistic "Discovery/ED prior"?



Summary

- How did change come about?
  - Vision & determination by leader with authority to implement
  - Colleagues able and enthusiastic to make it happen
- How it was actively managed
  - Goals, documentation and strong advocate at Technical Reviews
- Education, training and software
  - Statisticians and scientific colleagues and other stakeholders
- What are we doing?
  - Bayesian:
    Study Design & Reporting, Interim Analyses, Complex Modelling
- The future?
  - Greater use of informative priors: “beyond MA of past studies”
  - Better decision making