General Bayesian approach to design clinical trials with time-to-event outcomes

James McGree
Professor of Statistics
School of Mathematical Sciences
Queensland University of Technology

james.mcgree@qut.edu.au | www.jamesmcgree.com | @j_mcgree

Collaborators: Overstall, Mahar, Jones, Snelling

March 1, 2023

Outline of talk

- Introduction
- Motivation: Re-design of ORVAC
- Background:
  - (Standard) Bayesian inference
  - General Bayesian methods for inference
  - (Standard) Bayesian design for clinical trials
- Extension: General Bayesian design for clinical trials
- Trial simulation for re-designing ORVAC
- Discussion

Adaptive trials

- COVID-19 brought an unprecedented push to undertake clinical trial assessments as quickly as possible
- Adaptive trials are particularly appealing in this regard
- Based on information from accruing data, e.g. trial can stop early due to success or futility, treatment can be dropped if ineffective or unsafe, etc
- Compared to non-adaptive trials, can complete sooner, cost less to run and reduce the number of participants on inferior treatments
- Potential benefits for participants and trialists

Adaptive trials

- Typically adaptive trials are designed via trial simulation
- Assumptions about the data generating process needed
- e.g. Size of treatment effect and distribution of data
- Misspecification can lead to a poorly designed trial e.g. insufficient power
- Could have consequences in terms of conclusions drawn from trial
- Propose to address this, at least partly, through general Bayesian methods

Motivation

- Introduction
- Motivation: Re-design of ORVAC
- Background
- Extension: General Bayesian design for clinical trials
- Trial simulation for re-designing ORVAC
- Discussion

ORVAC

- ORVAC: Optimising Rotavirus Vaccine in Aboriginal Children
- Double-blind, randomised, placebo-controlled Bayesian adaptive clinical trial
- Assess effectiveness of a third dose of Rotarix rotavirus vaccine in Australian Indigenous infants in providing improved protection (versus usual care) against gastroenteritis
- The third dose was active treatment, usual care was matched placebo.
- Randomly allocated to participants at a ratio of 1:1

ORVAC

- There were two primary outcomes:
  - Anti-rotavirus IgA seroconversion, defined as serum anti-rotavirus IgA $\geq$ 20 U/ml 28 to 55 days post Rotarix/placebo
  - Time from randomisation to presentation of acute gastroenteritis or acute diarrhoea illness (t2e outcome)
- Eligible for enrolment: Aged between 6 and 12 months
- Followed up to 36 months of age
- The maximum sample size: 1,000 participants
- Planned analyses occurred after 250 participants then every 50 thereafter.
- At each planned analysis, pre-specified decision rules: effectiveness and futility.

ORVAC

- To design trial, simulation was used
- Assumed proportional hazards where baseline hazard from exponential or Weibull
- Hazard function of the form:

$$h_i(t) = \psi_i h_0(t),$$

where t is time since enrolment, $\psi_i = \exp(x_i\beta)$, x_i is treatment, β is treatment effect and $h_0(t)$ is the baseline hazard function for participant i.

- Baseline hazard: Assumed constant under exp, and monotonically increasing or decreasing under Weibull

Background

- Introduction
- Motivation: Re-design of ORVAC
- Background
- Extension: General Bayesian design for clinical trials
- Trial simulation for re-designing ORVAC
- Discussion

Bayesian inference

- All inference based on the posterior distribution of $\boldsymbol{\theta}$ i.e.

$$p(\boldsymbol{\theta}|\boldsymbol{y}, \boldsymbol{x}) \propto p(\boldsymbol{\theta})p(\boldsymbol{y}|\boldsymbol{\theta}, \boldsymbol{x}),$$

  where $p(\boldsymbol{\theta})$ is the prior, $p(\boldsymbol{y}|\boldsymbol{\theta}, \boldsymbol{x})$ is the likelihood, $\boldsymbol{y} = (y_1, \ldots, y_N)^t$ and $y_i = (t_i, c_i)$.
- Make probability statements about treatment effect
- For t2e outcomes, likelihood can be formed based on hazard function
- E.g. Specify baseline hazard and assume proportional hazards between treatments

Bayesian inference

- Specifically, $f(t|\boldsymbol{\theta}) = h(t)S(t|\boldsymbol{\theta})$, where $S(t|\boldsymbol{\theta}) = \exp(-H(t))$ and $H(t) = \int_0^t h(t)\mathrm{d}t$
- To construct likelihood for ORVAC, participant i yield a t2e outcome from treatment allocation x_i.
- Assuming each observation is conditionally independent:

$$p(\boldsymbol{y}|\boldsymbol{\theta}, \boldsymbol{x}) = \prod_{i=1}^{N} f(t_i|\boldsymbol{\theta}, x_i)^{c_i} S(t_i|\boldsymbol{\theta}, x_i)^{1-c_i},$$

where c_i denotes censored or not for ith participant (uninformative right censoring).

General Bayesian inference

- Inference based on a general loss function i.e.

$$p(\boldsymbol{\theta}|\boldsymbol{y}, \boldsymbol{x}) \propto p(\boldsymbol{\theta}) \exp[-wl(\boldsymbol{\theta}, \boldsymbol{y}, \boldsymbol{x})],$$

  where $w \geq 0$ and $l(\boldsymbol{\theta}, \boldsymbol{y}, \boldsymbol{x})$ is the loss function.
- Standard Bayesian inference is a special case
- For continuous and distinct outcomes $t_{(1)}, \ldots, t_{(J)}$, negative partial log-likelihood as the loss function (with $w = 1$) i.e.

$$l(\boldsymbol{\theta}, \boldsymbol{y}, \boldsymbol{x}) = -\sum_{j=1}^{J} \log \frac{\exp(x_{(j)}\boldsymbol{\theta})}{\sum_{r \in R_j} \exp(x_r \boldsymbol{\theta})},$$

  R_j is risk set (set of participants not yet responded and not censored) at the jth ordered event time $t_{(j)}$ for a participant defined by $x_{(j)}$
- Note: Baseline hazard nuisance/undefined

Bayesian adaptive trials

- Typically trial design assessed via trial simulation
- Requires pre-specifying decision rules and what adjustments can be made
- When interim analyses will occur
- Used to provide insight into performance e.g. empirical estimates of power and type-1 error
- Adjust design to optimise summarise in some way

Bayesian adaptive trials

Simulate adaptive trial

1: Initialise $p(\theta)$, $h(t)$, treatment effect
2: **for** $k = 1 : K$ **do**
3: Enrol participants
4: Randomly assign treatments $\boldsymbol{x}$
5: Simulate observations based on $\boldsymbol{x}$ and $h(t)$
6: **if** planned interim analysis **then**
7: Evaluate decision rules
8: **if** any decision rule met **then**
9: Stop or adjust trial
10: **end if**
11: **end if**
12: **end for**
13: Summarise simulated trial

ORVAC - Effectiveness decision rule

- Expectation of declaring trial effectiveness based on:

$$P(\beta < 0 | \boldsymbol{y}, \boldsymbol{x}) > 0.97$$

- Expectation taken over data from participants who have not responded and are not censored i.e.

$$\left[\int_{\mathcal{Z}} \mathcal{I}(P(\beta < 0 | \boldsymbol{y}, \boldsymbol{z}, \boldsymbol{x}) > 0.97) p(\boldsymbol{z} | \boldsymbol{y}, \boldsymbol{x}) \mathrm{d}\boldsymbol{z} \right] > 0.9$$

- No analytic solution available but can use Monte Carlo:

$$\approx 1/B \sum_{b=1}^{B} \mathcal{I}(P(\beta < 0 | \boldsymbol{y}, \boldsymbol{z}_b, \boldsymbol{x}) > 0.97)$$

where $\boldsymbol{z}_b$ is simulated from posterior predictive distribution.

ORVAC - Effectiveness decision rule

Evaluate effectiveness decision rule

1: Initialise $\boldsymbol{y}, \boldsymbol{x}, p(\boldsymbol{\theta}|\boldsymbol{y}, \boldsymbol{x})$
2: **for** $b = 1 : B$ **do**
3: Simulate posterior predictive data $\boldsymbol{z}_b$ for enrolled participants where outcome not observed and not censored
4: Update posterior distribution
5: Evaluate trial effectiveness rule $\boldsymbol{\Lambda}^b = \mathcal{I}(P(\beta < 0|\boldsymbol{y}, \boldsymbol{z}_b, \boldsymbol{x}) > 0.97)$
6: **end for**
7: $\delta_e = \frac{1}{B} \sum_{b=1}^{B} \boldsymbol{\Lambda}^b$
8: **if** $\delta_e > 0.90$ **then**
9: Stop trial
10: **end if**

ORVAC - Futility decision rule

- Expectation of declaring trial effectiveness
- Expectation taken over:
  - Data from participants who have not responded and are not censored
  - Data from participants not yet enrolled, i.e.

$$\left[\int_{\mathcal{Z}} \sum_{\boldsymbol{v}\in\mathcal{V}} \mathcal{I}(P(\beta < 0|\boldsymbol{y}, \boldsymbol{z}, \boldsymbol{x}, \boldsymbol{v}) > 0.97) p(\boldsymbol{z}|\boldsymbol{y}, \boldsymbol{x}, \boldsymbol{v}) p(\boldsymbol{v}) \mathrm{d}\boldsymbol{z}\right] < 0.05$$

- No analytic solution available but can use Monte Carlo:

$$\approx 1/B \sum_{b=1}^{B} \mathcal{I}(P(\beta < 0|\boldsymbol{y}, \boldsymbol{z}_b, \boldsymbol{x}, \boldsymbol{v}_b) > 0.97)$$

where $\boldsymbol{z}_b$ is simulated from posterior predictive distribution, and $\boldsymbol{v}_b$ from randomisation scheme.

ORVAC - Futility decision rule

Evaluate futility decision rule

1: Initialise $\boldsymbol{y}, \boldsymbol{x}, p(\boldsymbol{\theta}|\boldsymbol{y}, \boldsymbol{x})$
2: **for** $b = 1 : B$ **do**
3: Simulate treatment allocation $\boldsymbol{v}_b$
4: Simulate posterior predictive data $\boldsymbol{z}_b$ for enrolled participants and for participants who are yet to enrol
5: Update posterior distribution
6: Evaluate trial effectiveness rule $\Lambda^b = \mathcal{I}(P(\beta < 0|\boldsymbol{y}, \boldsymbol{z}_b, \boldsymbol{x}, \boldsymbol{v}_b) > 0.97)$
7: **end for**
8: $\delta_f = \frac{1}{B} \sum_{b=1}^{B} \Lambda^b$
9: **if** $\delta_f < 0.05$ **then**
10: Stop trial
11: **end if**

Extension

- Introduction
- Motivation: Re-design of ORVAC
- Background
- Extension: General Bayesian design for clinical trials
- Trial simulation for re-designing ORVAC
- Discussion

General Bayesian adaptive trials

- Trial simulation: Need to make assumptions about the data generating process
- We seek robustness to the misspecification of these assumptions
- Our approach: Propose all inference conducted within a general Bayesian framework.
- Evaluate trial designs based on the partial likelihood i.e. Baseline hazard remains nuisance/undefined
- However, loss function need not be linked to the data generating process, so how to simulate data?
- i.e. Cannot generate data from partial likelihood (only)

General Bayesian adaptive trials

- To simulate data, propose to consider a *super model*
- Formulated such that it can describe a wide range of data sets
- Here, a wide variety of hazard functions such as constant, monotonic and non-monotonic function
- Could also extend to consider more complex censoring mechanisms
- Super model is out of the scope of inference
- So not desirable to estimate a treatment effect based on this model
- e.g. most likely overparameterised

General Bayesian adaptive trials

- Here, define super model based on a cubic spline representation of the baseline hazard function i.e.

$$h_0(t) = \sum_{q=1}^{Q} \xi_q g_q(t),$$

where ξ_q are parameters and $g_q(t) = t^q$ are the basis functions, for $q = 1, \ldots, Q$.

- Very flexible: Constant, monotonic and non-monotonic forms
- Use this model for simulation, and use partial likelihood for inference

Trial simulation

- Introduction
- Motivation: Re-design of ORVAC
- Background
- Extension: General Bayesian design for clinical trials
- Trial simulation for re-designing ORVAC
- Discussion

Set up

- Use trial simulation to explore new approach to re-design for ORVAC
- Models for data generation in trial simulation:
  - Exponential PHs model
  - Weibull PHs model
  - Super model
- Models for estimating treatment effect in trial simulation:
  - Exponential PHs model
  - Weibull PHs model
  - General Bayesian model
- Nine combinations
- Vaguely informative priors used throughout

Set up

- Initially, different values for treatment effect i.e. $\beta \in \{0, -0.075, -0.125, -0.175, -0.25, -0.5\}$
- Consider specific forms for baseline hazard

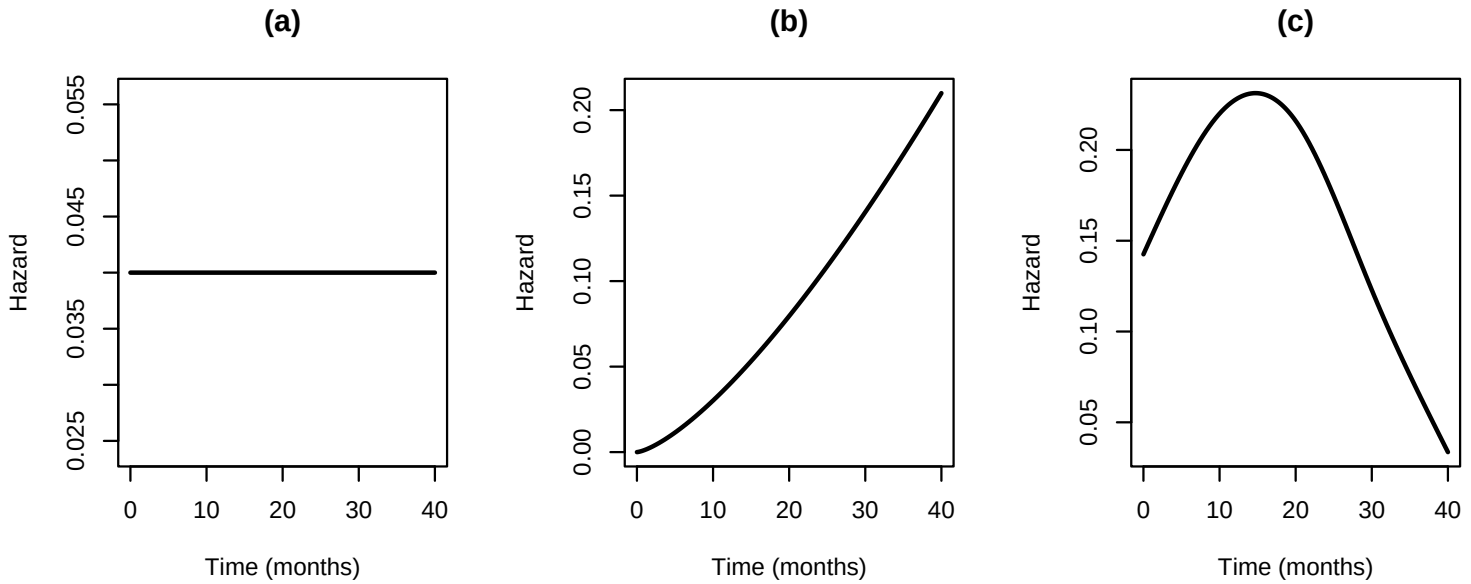

Figure: Assumed hazard functions for the (a) Exponential, (b) Weibull and (c) Super model.

Estimation results

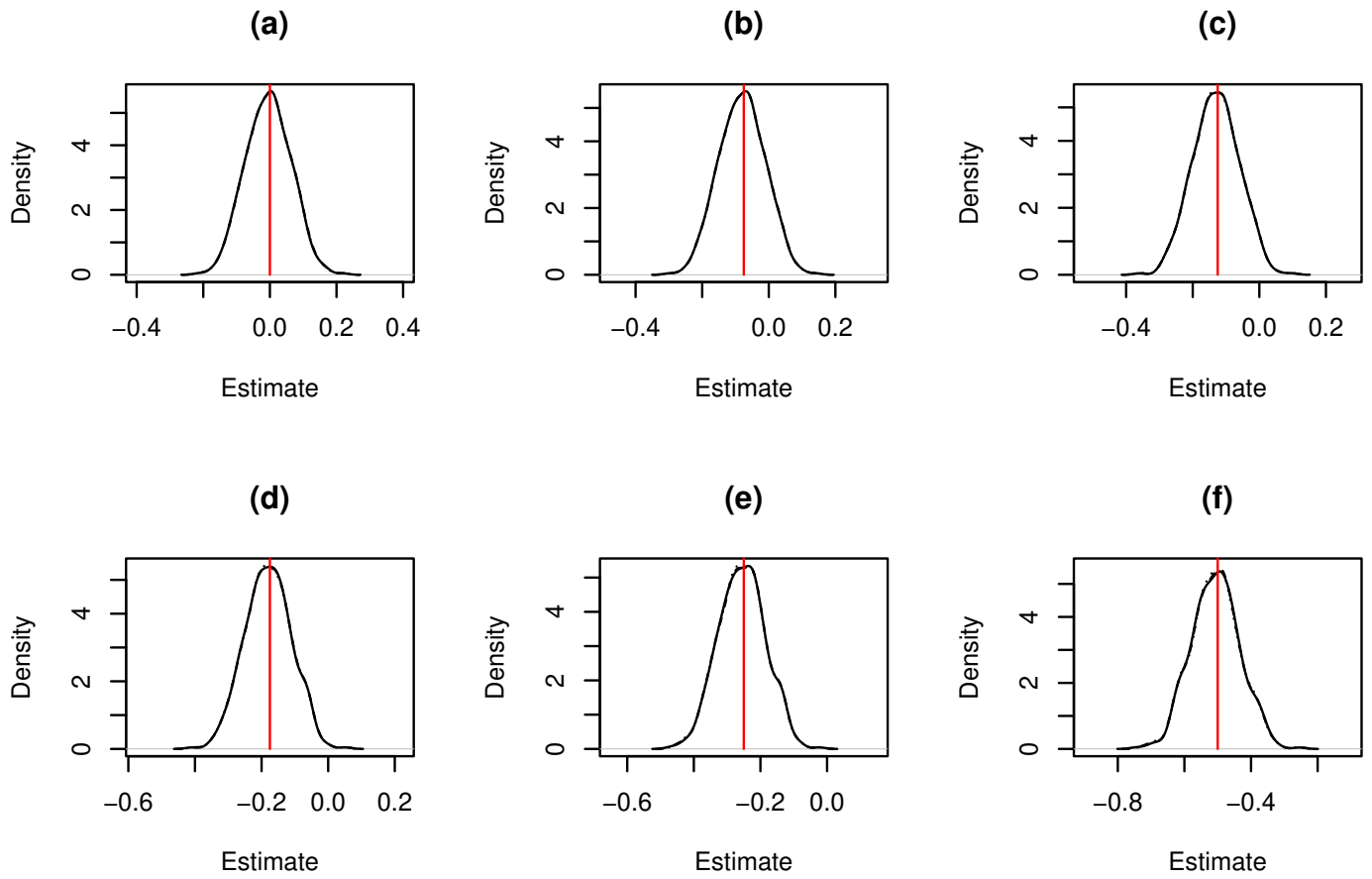

Figure: Data from exponential model, then fit Exponential $(\cdots)$, general Bayesian $(-)$ and Weibull $(- \ -)$.

Estimation results

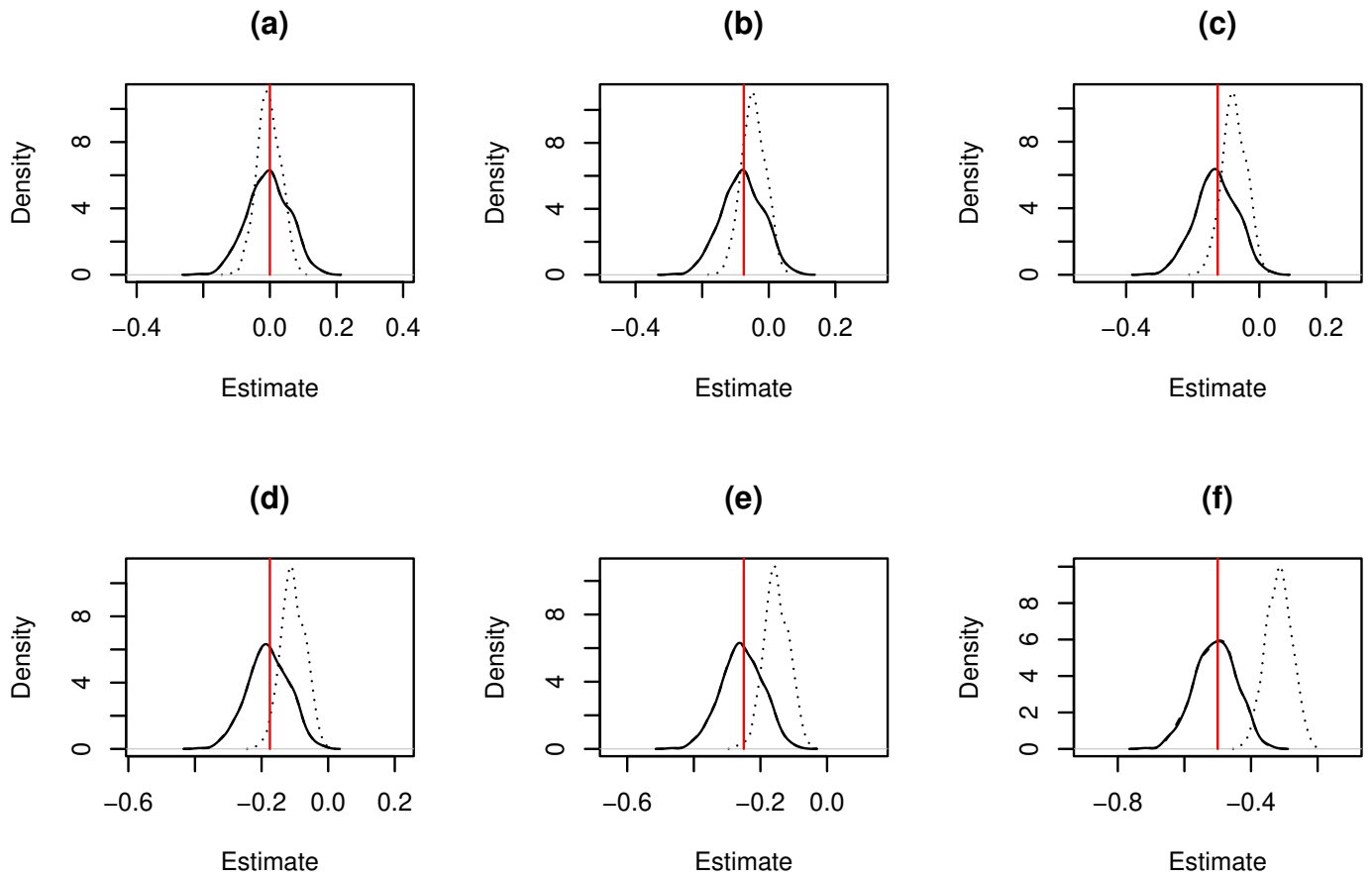

Figure: Data from Weibull model, then fit Exponential $(\cdots)$, general Bayesian $(-)$ and Weibull $(- \ -)$.

Estimation results

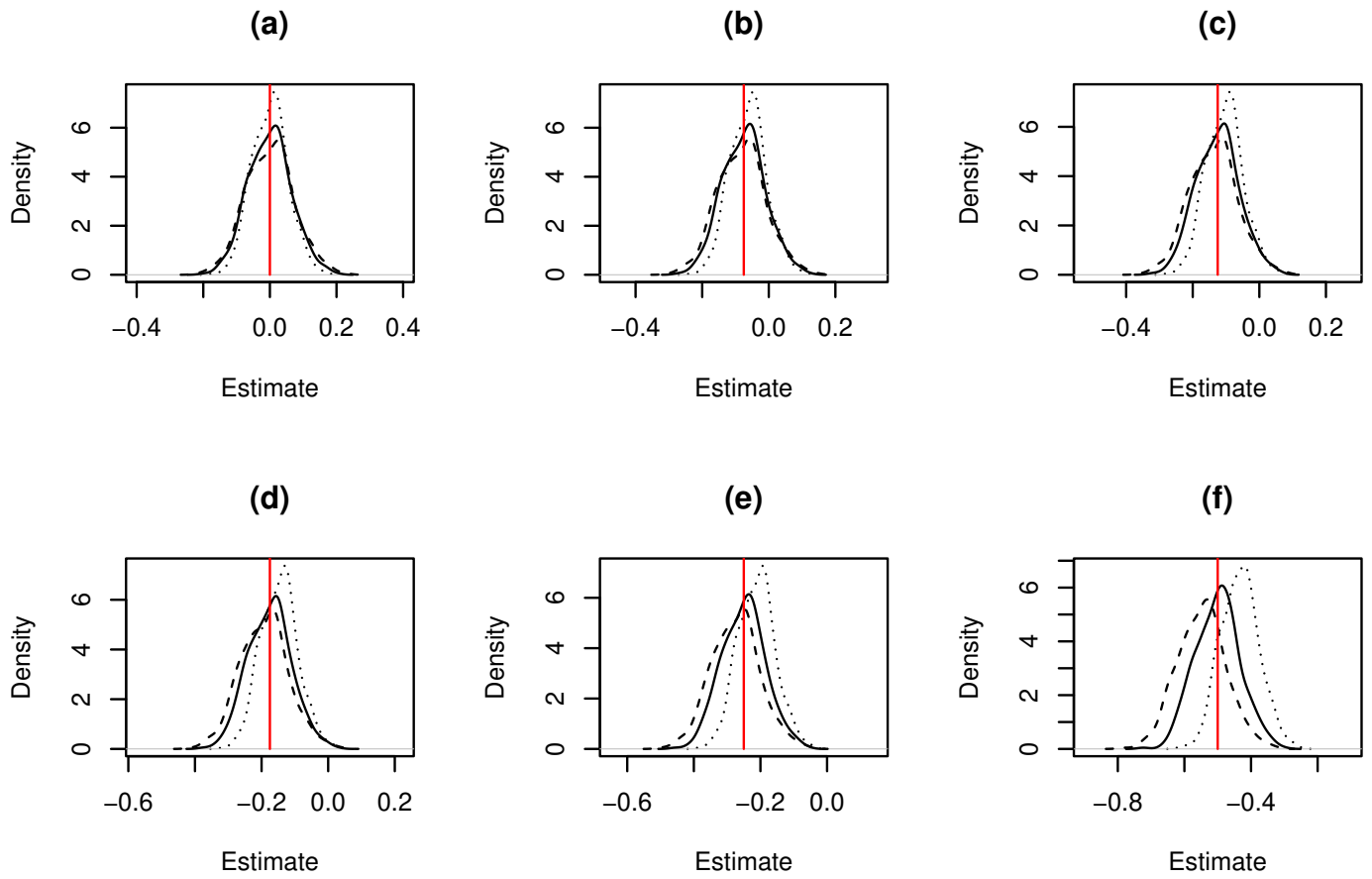

Figure: Data from super model, then fit Exponential $(\cdots)$, general Bayesian $(-)$ and Weibull $(- -)$.

Trial simulation results

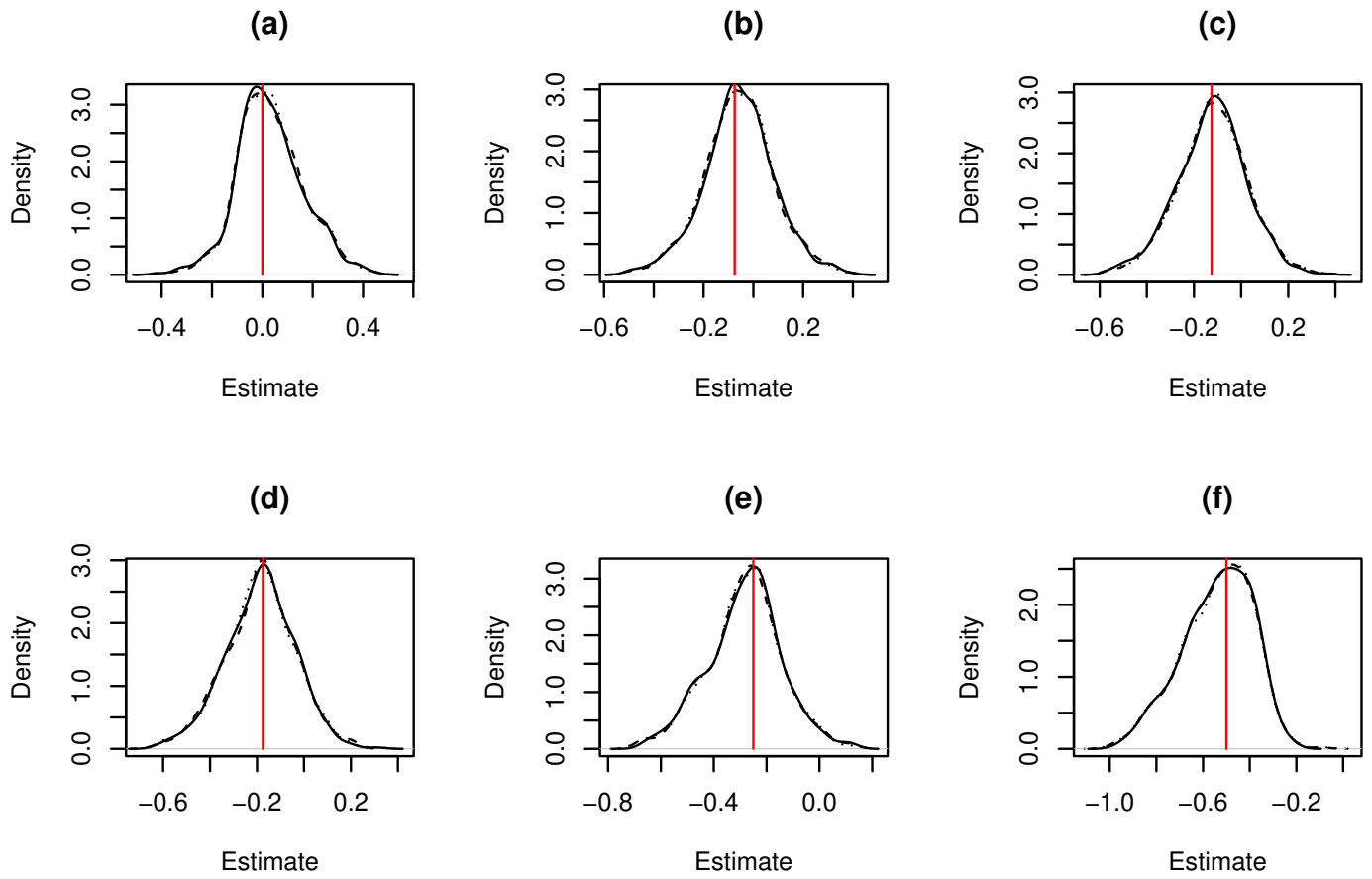

Figure: Data from exponential model, then fit exponential $(\cdots)$, general Bayesian $(-)$ and Weibull $(- -)$.

Trial simulation results

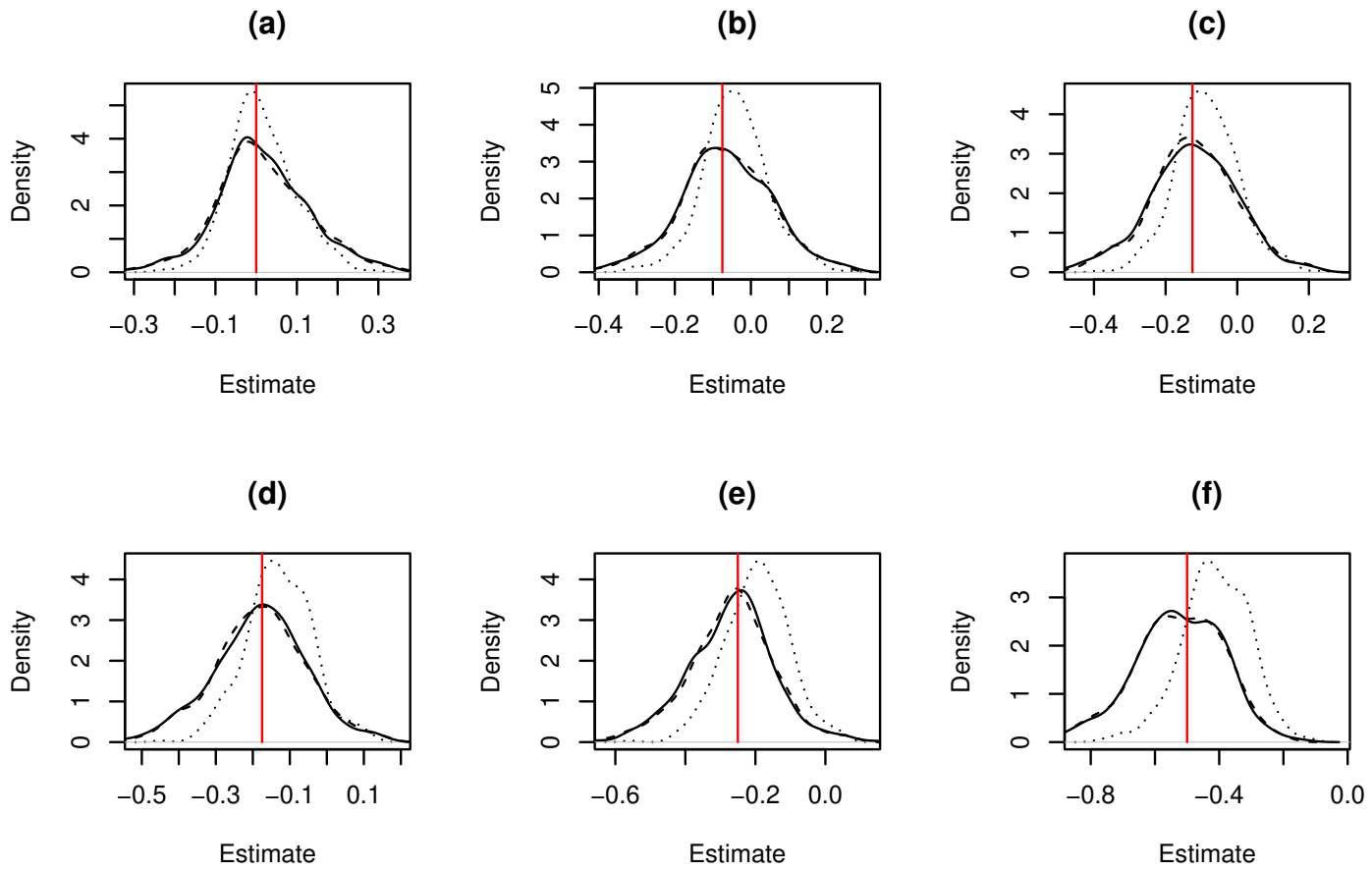

Figure: Data from Weibull model, then fit exponential $(\cdots)$, general Bayesian $(-)$ and Weibull $(- -)$.

Trial simulation results

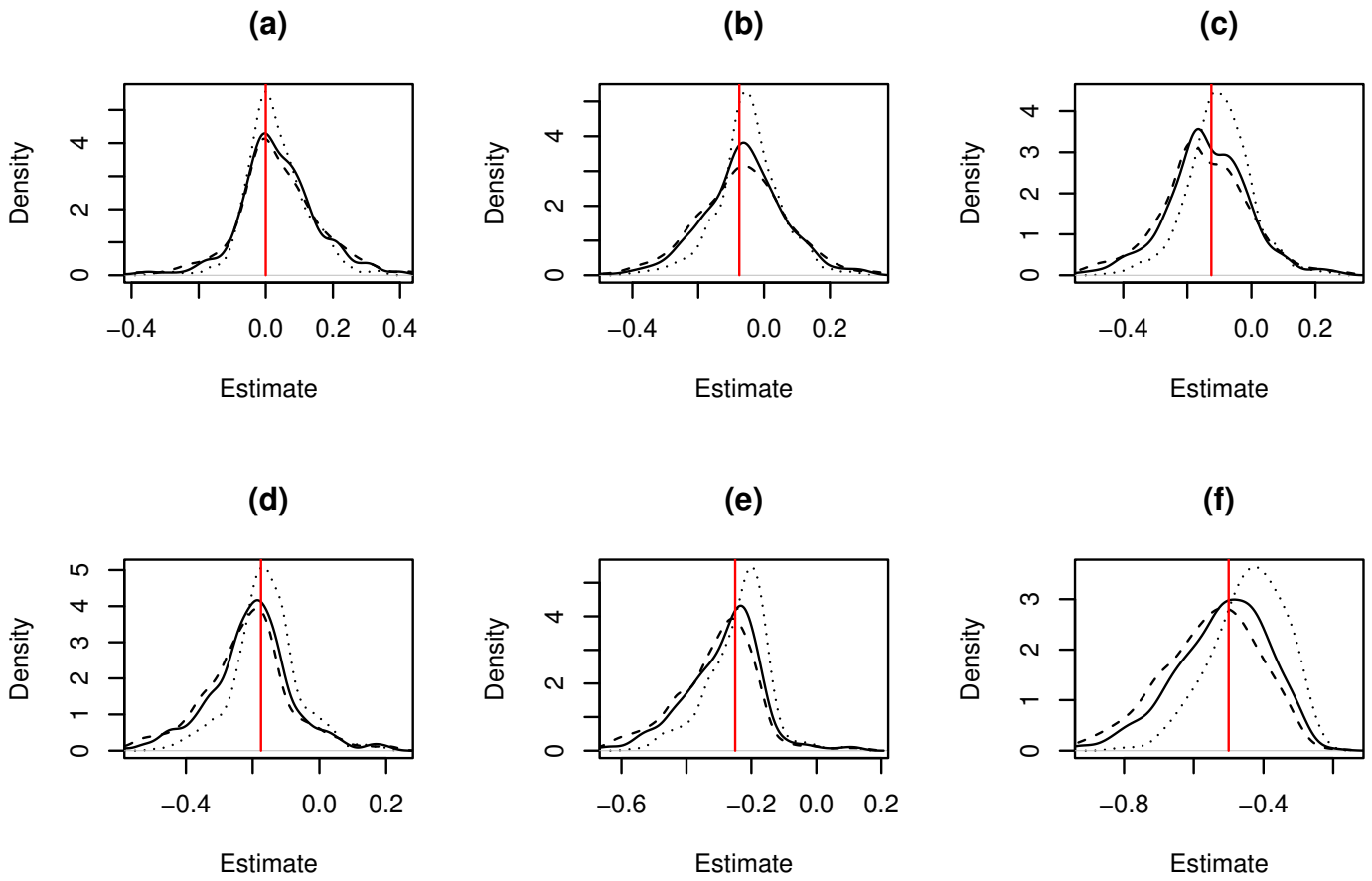

Figure: Data from super model, then fit exponential $(\cdots)$, general Bayesian $(-)$ and Weibull $(- -)$.

Trial simulation results

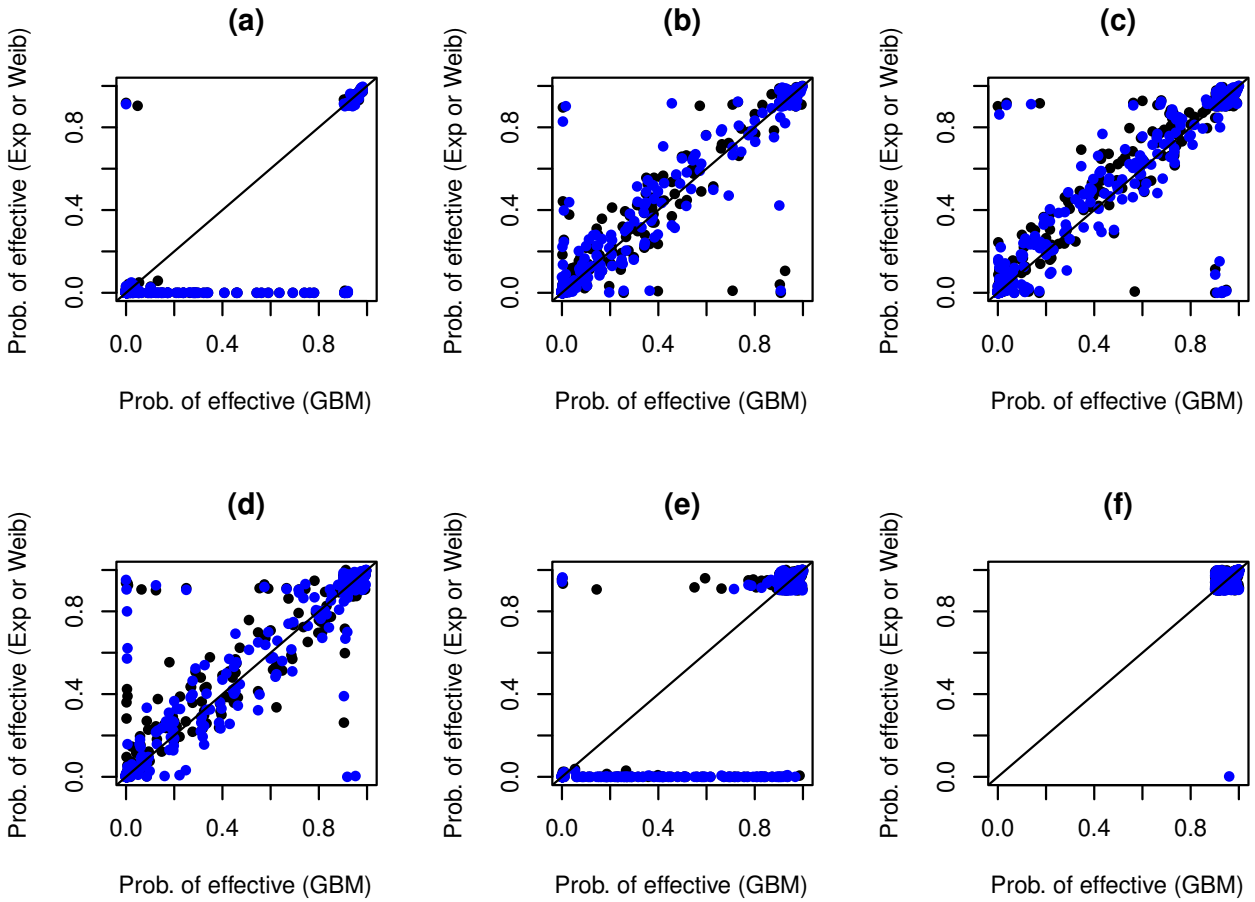

Figure: Data from exponential model, then fit general Bayesian, exponential (black) and Weibull (blue) models.

Trial simulation results

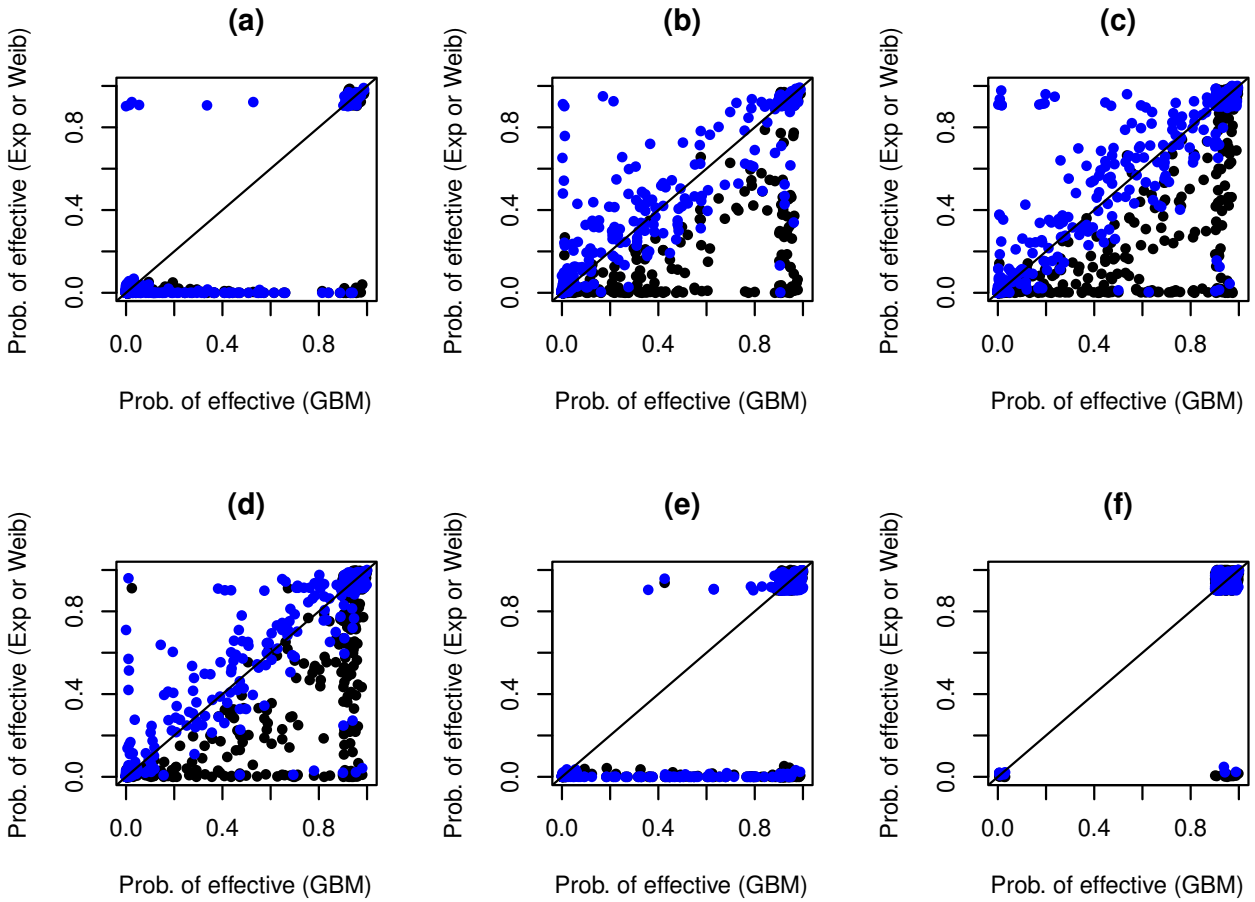

Figure: Data from Weibull model, then fit general Bayesian, exponential (black) and Weibull (blue) models.

Trial simulation results

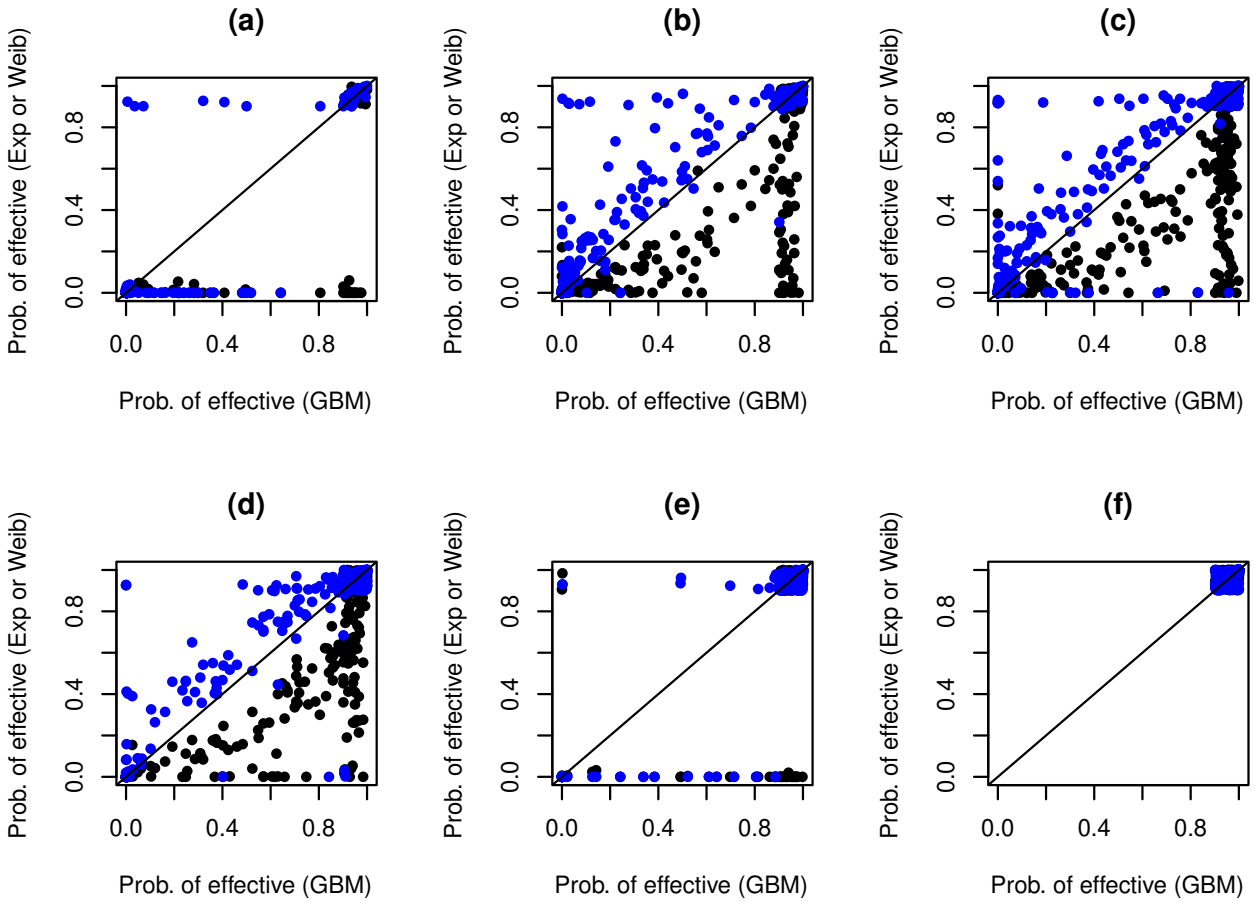

Figure: Data from super model, then fit general Bayesian, exponential (black) and Weibull (blue) models.

Trial simulation results

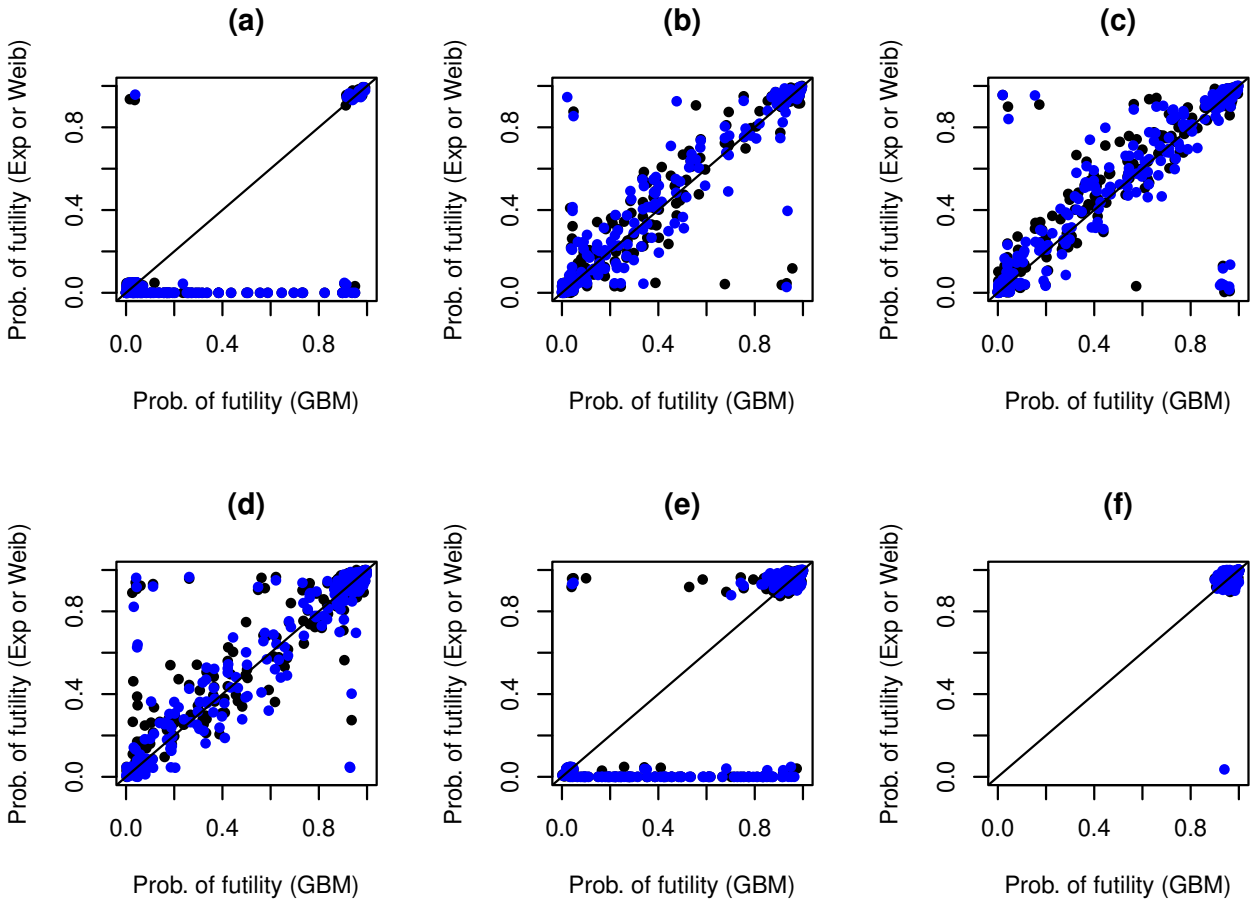

Figure: Data from exponential model, then fit general Bayesian, exponential (black) and Weibull (blue) models.

Trial simulation results

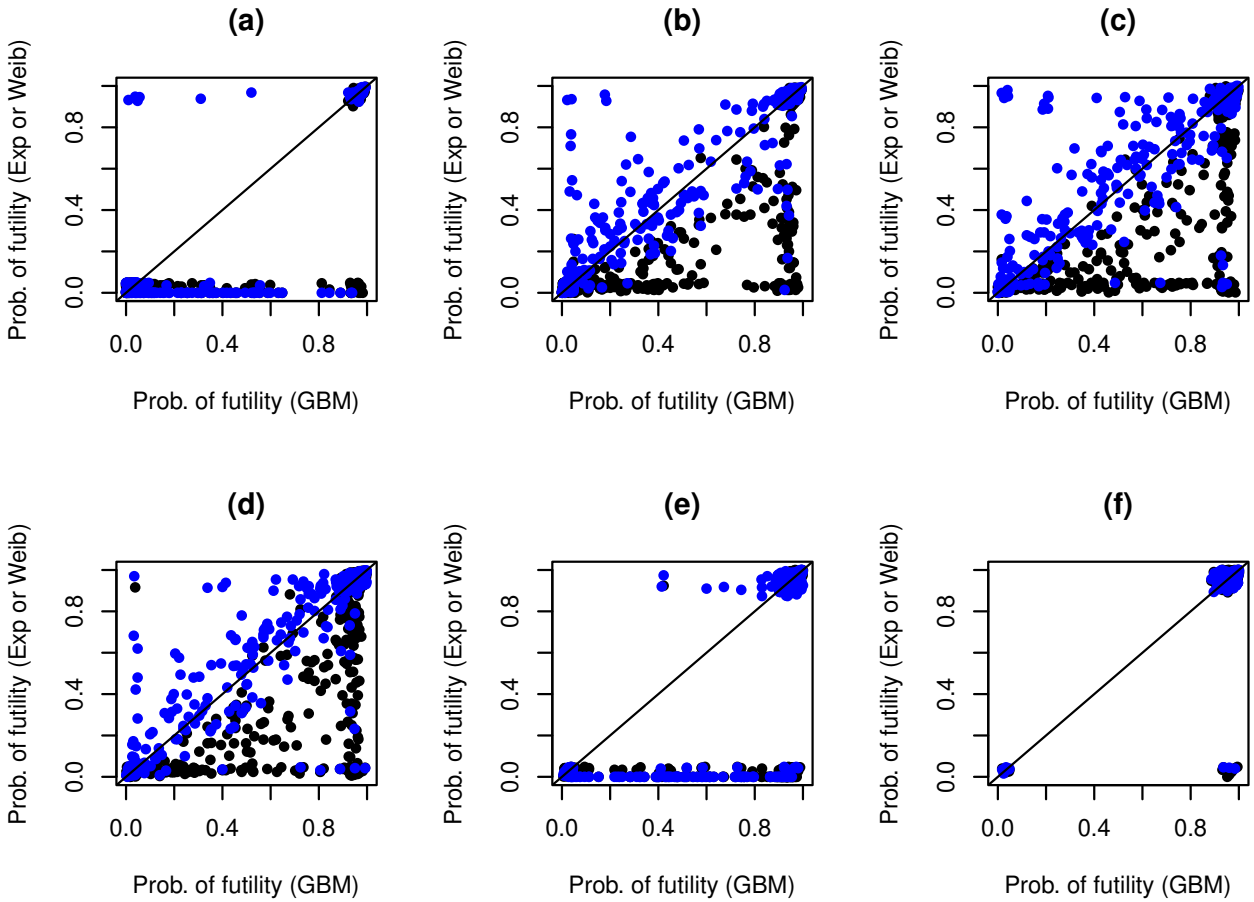

Figure: Data from Weibull model, then fit general Bayesian, exponential (black) and Weibull (blue) models.

Trial simulation results

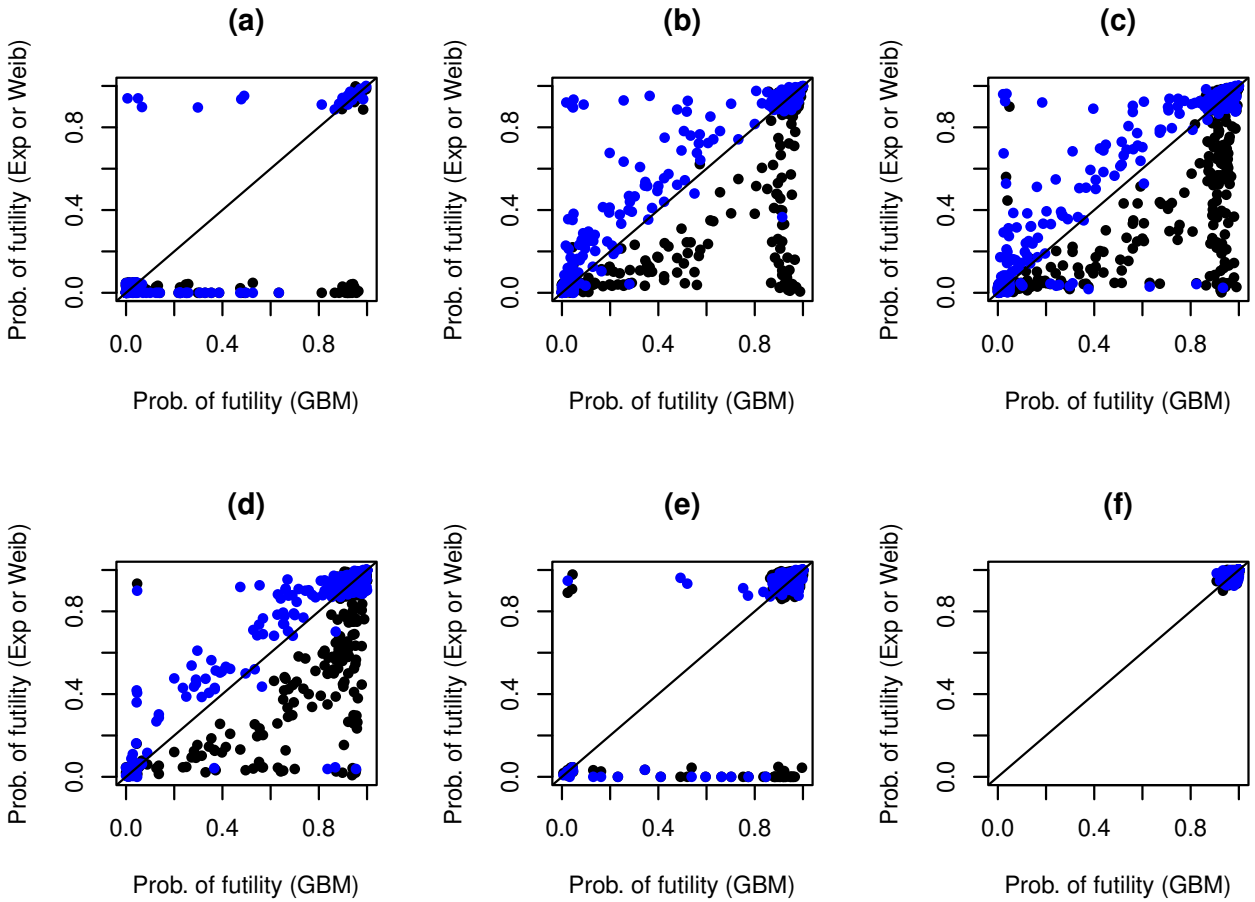

Figure: Data from super model, then fit general Bayesian, exponential (black) and Weibull (blue) models.

Trial simulation results

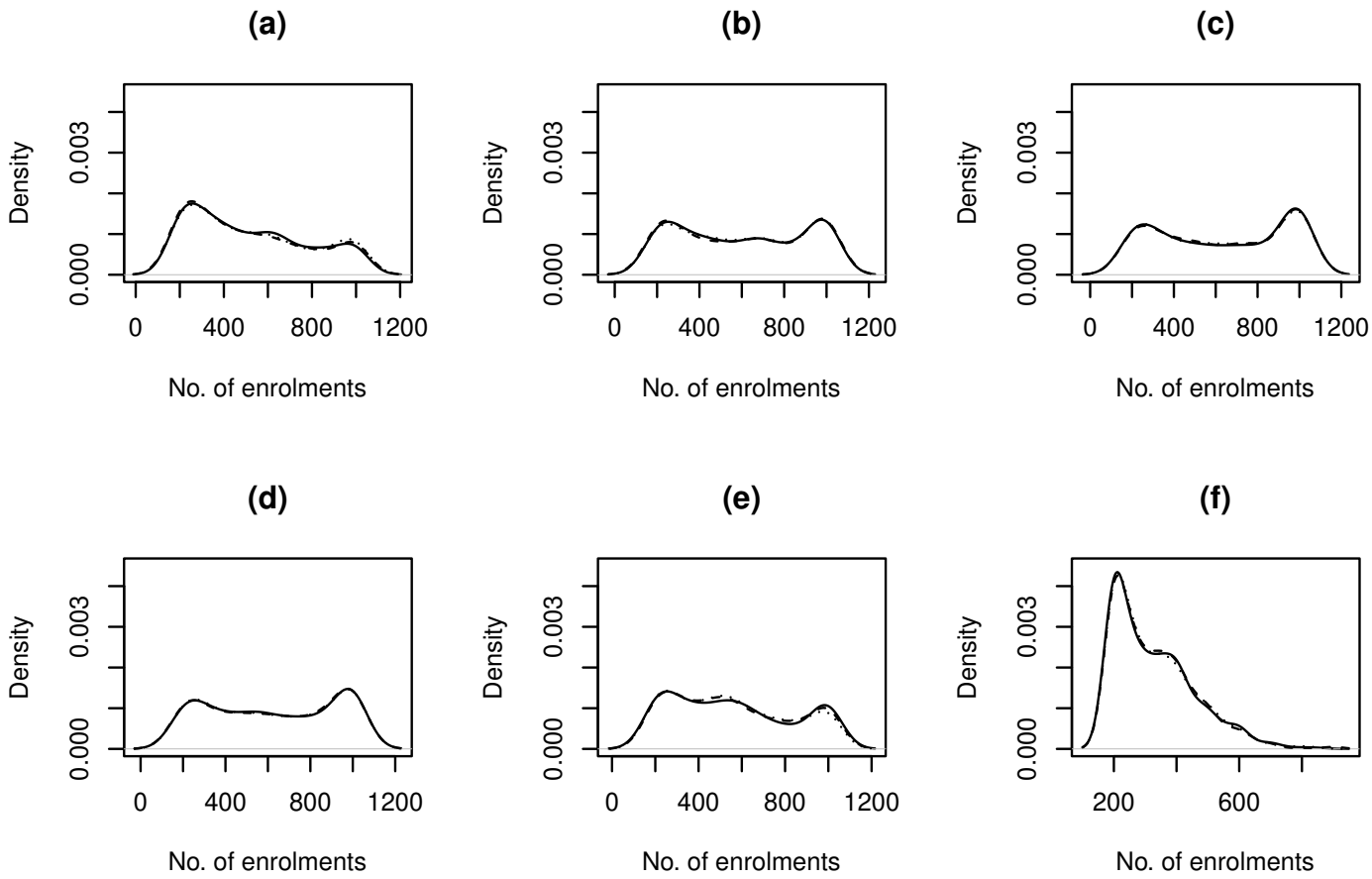

Figure: Data from exponential model, then fit exponential $(\cdots)$, general Bayesian $(-)$ and Weibull $(- \; -)$ models.

Trial simulation results

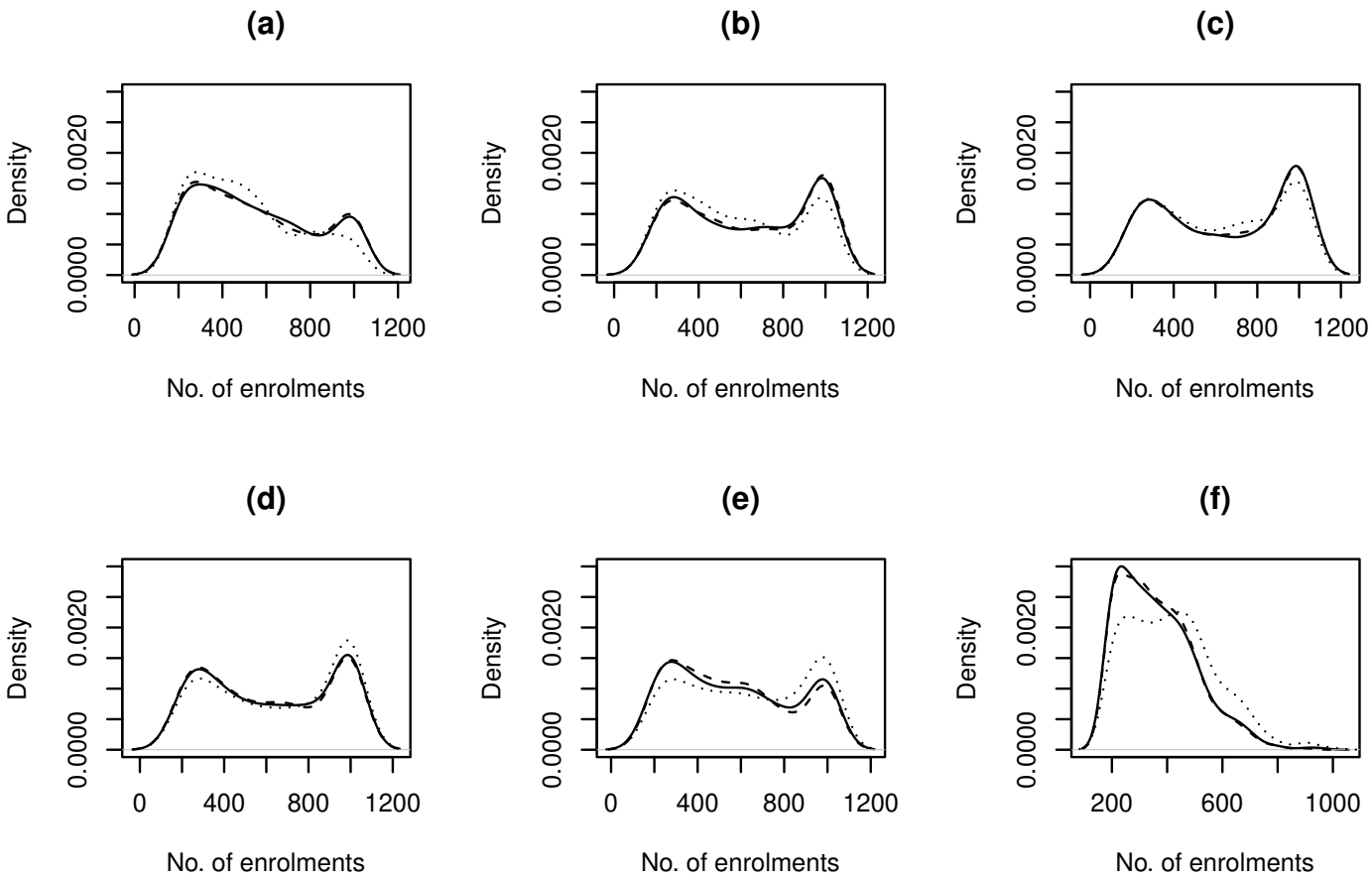

Figure: Data from Weibull model, then fit exponential $(\cdots)$, general Bayesian $(-)$ and Weibull $(- \ -)$ models.

Trial simulation results

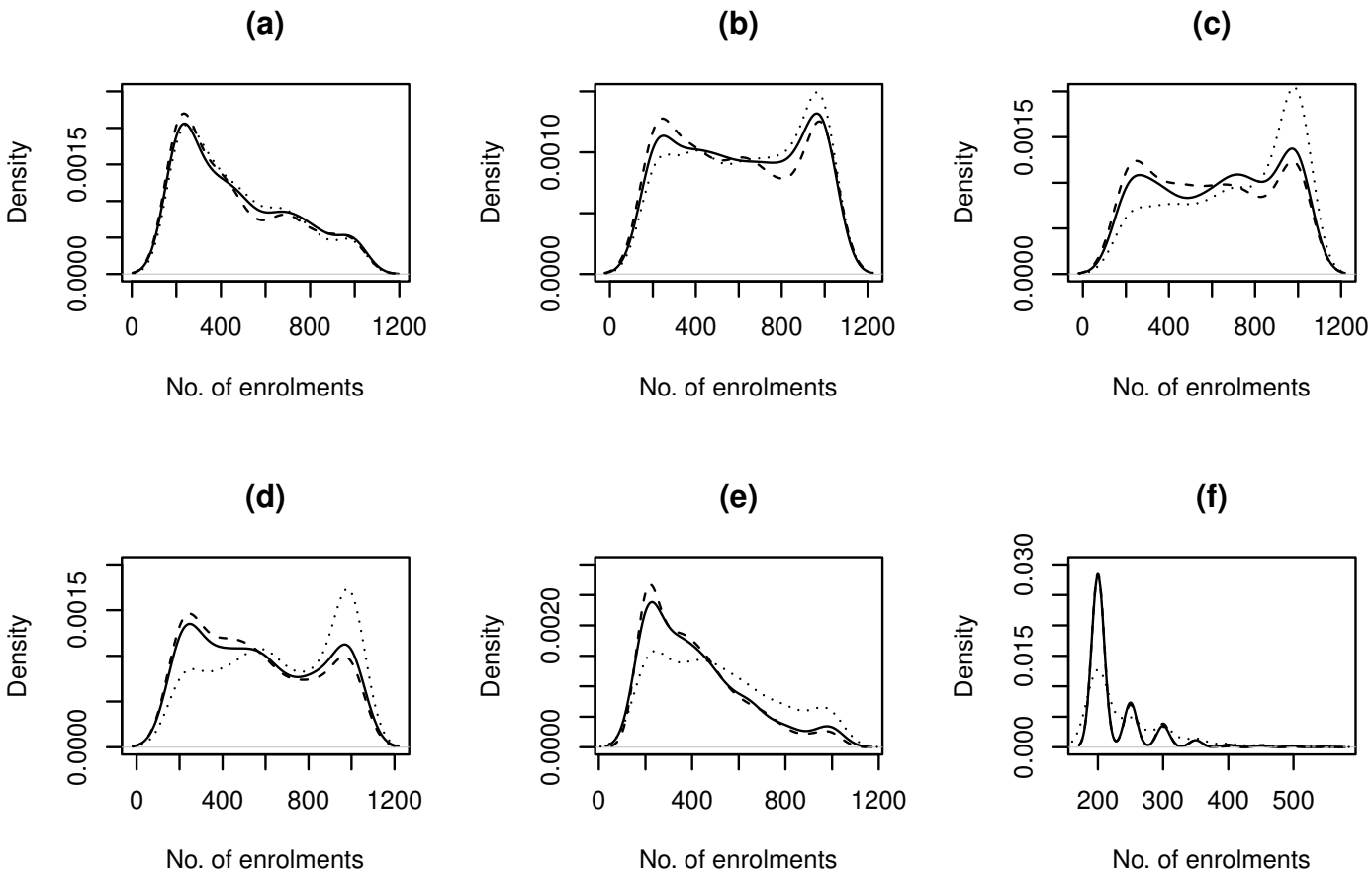

Figure: Data from super model, then fit exponential $(\cdots)$, general Bayesian $(-)$ and Weibull $(- \ -)$ models.

Extended adaptive trial results

- Next, explore the performance of models under a range of baseline hazard functions from the super model
- Form a prior for the super model
- i.e. Fix number of knots and knot position, then random generate corresponding values from U[0, 0.4] independently
- Prior on treatment effect: $\beta \sim U(-0.75, -0.25)$
- Assess performance as before

Extended adaptive trial results

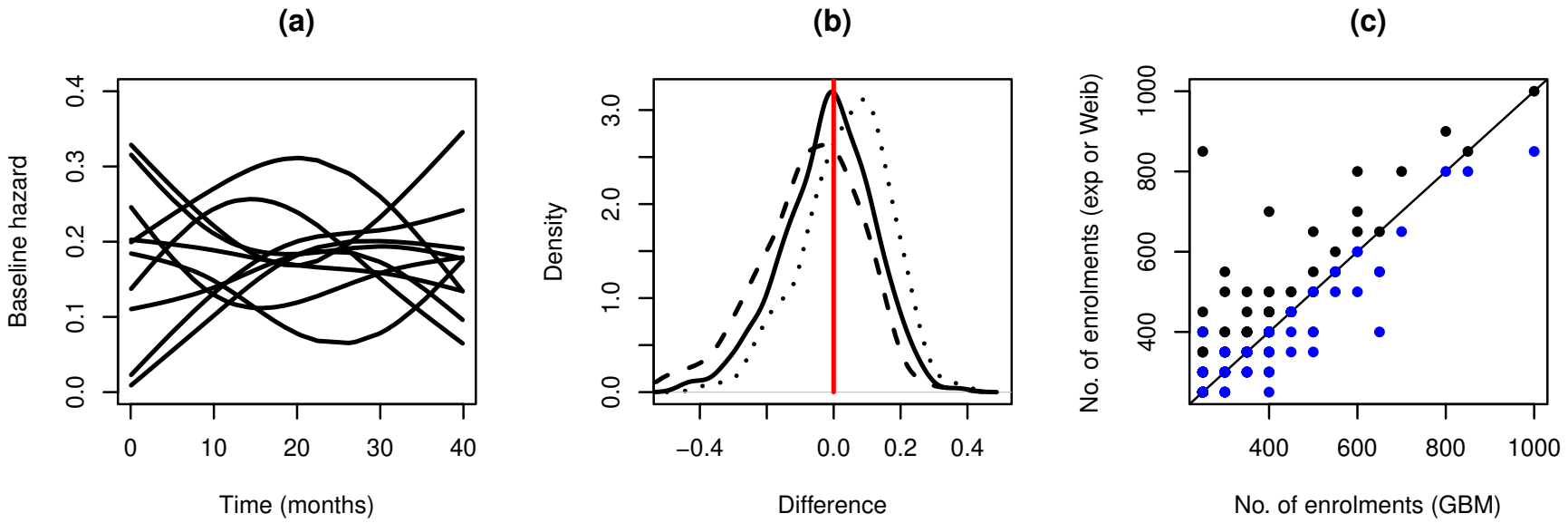

Figure: (a) 10 realisations of baseline hazard from super model; and (b) distribution of the difference between the true and estimated treatment effect under the exponential $(\cdots)$, general Bayesian $(-)$ and Weibull $(- -)$ models, and (c) number of enrolments under the general Bayesian model compared to the exponential (black) and Weibull (blue) models.

Extended adaptive trial results

- Exp and Weibull provide slightly biased estimates of treatment
- Median difference of 0.0503 and -0.0607, respectively.
- General Bayesian model relatively unbiased estimate of treatment with mean difference of -0.0087
- Exp/Weibull, on average, a larger/fewer number of enrolments compared to general Bayesian model
- Mean difference of $18/-5.2$ enrolments.
- No appreciable differences in probabilities of success and futility

Trial simulation

- Introduction
- Motivation: Re-design of ORVAC
- Background
- Extension: General Bayesian design for clinical trials
- Trial simulation for re-designing ORVAC
- Discussion

To conclude

- Proposed a general Bayesian method to design adaptive trials
- For time-to-event outcomes, not required to specify the specific data generating model
- Led to trial design that was robust to baseline hazard function
- Appears useful e.g. assuming the wrong baseline hazard function can lead to over/underestimation, and shorter/longer trials
- Seems preferable to base designs on flexible models

Future research

- Scope to extend to other outcomes e.g. Overdispersed counts, remove influence of outliers, etc
- More broadly, seek approaches to reduce reliance on assumptions e.g. PH.
- Need computationally efficient approaches to determine w
- Robust approach in other settings? e.g. GAMs, Gaussian Process?

Selected references

- Berry et al. (2010). Bayesian adaptive methods for clinical trials. Chapman and Hall.
- Bissiri, Holmes and Walker (2016). A general framework for updating belief distributions. JRSSB, 78, 1103-1130.
- Chaloner and Verdinelli (1995). Bayesian experimental design: A review. Statistical Science, 10, 273–304.
- Jones et al. (2020). The ORVAC trial: a statistical analysis plan. Trials, 21, 741.
- Overstall and McGree (2023). General Bayesian L2 calibration of mathematical models. Technometrics. Tentatively accepted.