

# Bayesian Approaches to Phase I Clinical Trials: Methodological and Practical Aspects

---

Design of Experiments in Healthcare  
Isaac Newton Institute of Mathematical Sciences  
August 2011, Cambridge UK

Beat Neuenschwander  
Biometrics Oncology, Novartis Pharma , Basel, Switzerland



# Outline

---

- Introduction
- Phase I Cancer Trials
  - overview
  - a wish list
  - a Bayesian implementation
- Case Study: Japanese Phase I Cancer Trial
- Conclusions

---

# Introduction

# Introduction

## *The Role for Statistics in Drug Development*

---

### ■ Phase III

- statistics well-established, sets confirmatory standards

### ■ Phase II

- proof-of-concept studies, dose-ranging studies
- a mix of approaches: exploratory, confirmatory, predictive

### ■ Phase I: statistics plays a minor role

- use of simplistic approaches prevails
- considerable potential for statistics, if seen as an approach/guide to understanding and quantifying uncertainty in order to arrive at better informed decisions

# Introduction

## *Statistical Thinking - Statistical Science – Statistical Engineering*

---

### ■ Statistical Thinking

- considerable uncertainty in many areas. Early clinical trials are one example.
- understand these uncertainties and act adequately

### ■ Statistical Science provides the know-how

### ■ Statistical Engineering

- how to make best use of statistical science and statistical thinking in order to successfully solve applied problems
- Statistical engineering provides the (missing) link!

# Introduction

## *Statistical Engineering*

---

### Statistical Engineering: a definition

*(Hoerl and Snee 2010)*

*Statistical engineering is the study of how to best use known statistical principles and tools to solve high impact problems for the benefit of mankind. It encompasses the tactical integration of statistical thinking (at the strategic level) with application of statistical methods and tools (at the operational level), and has the potential to provide the missing link that will drive proper application of statistical methods based on solid understanding of statistical thinking principles.*

*(see also: Roger Hoerl. *The Word is Calling: Should We Answer?* Deming Lecture, JSM Miami 2011)*

# Introduction

## *Objective of the Talk*

---

Objective of this talk is to

- show an application (phase I cancer trials) where statistical engineering is needed
- highlight different aspects of phase I trials
- present a concrete implementation of the Bayesian approach to phase I cancer trials
- show a case-study

---

# Phase I Cancer Trials



# Overview

## *Main Objective of Phase I Cancer Trials*

- Main objective: find maximum tolerable dose (MTD)
- MTD: dose  $d$ 
  - with acceptable rate of dose-limiting toxicity (DLT):  $\pi_d$
  - Target:  $\pi_d$  in the range 1/6 to 1/3 (20%, 25%, 30%)
- Phase I trials
  - Small dose-escalation trials
  - Adaptive: based on current data, decide on next dose
  - Substantial uncertainty: during and at the end of the trial
  - Statistics: guide to decision making under uncertainty

# Approaches

## *Algorithmic and Model-Based Designs*

---

- No consensus on how to do phase I trials
- Approaches
  - non-statistical, algorithmic designs
    - 3+3 design, a simple algorithm
    - more sophisticated algorithmic designs
  - statistical
    - on-study dose recommendations, and declaration of the MTD are based on statistical inference
    - CRM: continual reassessment method
    - BLR: Bayesian logistic regression
    - ...

# Approaches

## “3+3” Design

### ■ “3+3” design

- Rules: depend on how many toxicities are seen in 3 pts
  - 0/3  $\uparrow$ , 1/3  $\rightarrow$ , 2/3  $\downarrow$
  - plus stopping rule: 1/6 ok, 2/6 too toxic

|            | 1mg | 2mg | 4mg | 8mg |
|------------|-----|-----|-----|-----|
| • Cohort 1 | 0/3 |     |     |     |
| • Cohort 2 |     | 0/3 |     |     |
| • Cohort 3 |     |     | 1/3 |     |
| • Cohort 4 |     |     | 1/6 |     |
| • Cohort 5 |     |     |     | 1/3 |
| • Cohort 6 |     |     |     | 2/6 |

- stop: declare *MTD* = 4mg

- a very popular design
- 3+3 declares an incorrect MTD too often! (Thall&Lee 2003)

# Operating Characteristics

*Long-Run Properties: Model-based designs are far better than “3+3”*

| Scenario | Gemci Dose |                | 100   | 200  | 300  | 400  | 500  | 600 | None |
|----------|------------|----------------|-------|------|------|------|------|-----|------|
| C1       | BLR        | True $p_{TOX}$ | .15   | .25  | .40  | .60  | .75  | .85 | –    |
|          |            | % MTD          | 21    | 66   | 12   | 0    | 0    | 0   | 0    |
|          |            | #Pats          | 6.9   | 21.9 | 6.6  | 0.5  | 0    | 0   |      |
|          | CRM        | % MTD          | 22    | 64   | 14   | 0    | 0    | 0   | 0    |
|          |            | #Pats          | 9.7   | 18.8 | 7.0  | 0.5  | 0    | 0   |      |
|          |            | 3 + 3          | % MTD | 36   | 34   | 11   | 1    | 0   | 0    |
|          |            | #Pats          | 4.8   | 4.3  | 2.4  | 0.6  | 0    | 0   |      |
| C2       | BLR        | True $p_{TOX}$ | .00   | .02  | .08  | .24  | .45  | .63 | –    |
|          |            | % MTD          | 0     | 0    | 15   | 72   | 12   | 0   | 0    |
|          |            | #Pats          | 0     | 3.6  | 9.5  | 17.6 | 4.6  | 0.6 |      |
|          | CRM        | % MTD          | 0     | 0    | 20   | 74   | 6    | 0   | 0    |
|          |            | #Pats          | 0.2   | 3.6  | 11.0 | 17.8 | 3.4  | 0   |      |
|          |            | 3 + 3          | % MTD | 0    | 6    | 39   | 44   | 9   | 0    |
|          |            | #Pats          | 3.1   | 3.4  | 4.6  | 5.0  | 2.8  | 0.5 |      |
| C3       | BLR        | True $p_{TOX}$ | .00   | .00  | .02  | .12  | .30  | .50 | –    |
|          |            | % MTD          | 0     | 0    | 0    | 37   | 56   | 7   | 0    |
|          |            | #Pats          | 0     | 3.0  | 4.0  | 12.7 | 12.9 | 3.4 |      |
|          | CRM        | % MTD          | 0     | 0    | 0    | 46   | 53   | 1   | 0    |
|          |            | #Pats          | 0     | 3.0  | 4.3  | 15.3 | 12.8 | 0.6 |      |
|          |            | 3 + 3          | % MTD | 0    | 1    | 13   | 49   | 30  | 0    |
|          |            | #Pats          | 3.0   | 3.0  | 3.6  | 5.0  | 4.5  | 1.7 |      |

(from Thall&Lee 2003)

# The Bad News!

---

- Model-based designs are only rarely used
- Rogatko *et al.* (2007)
  - Investigated 1235 Phase I cancer trials (1991-2006)
  - Only 20 (**1.6%**) used innovative designs (other than 3+3)
  - (see also: Tourneau *et al.* 2009)
- This is disappointing
  - ... after 20 years of statistical research in Phase I
  - ... what are the reasons?
  - ... what is missing?
- A case for Statistical Engineering?

# Are model-Based designs too aggressive?

## *An Example*

- From Muler *et al.* (JCO 2004)

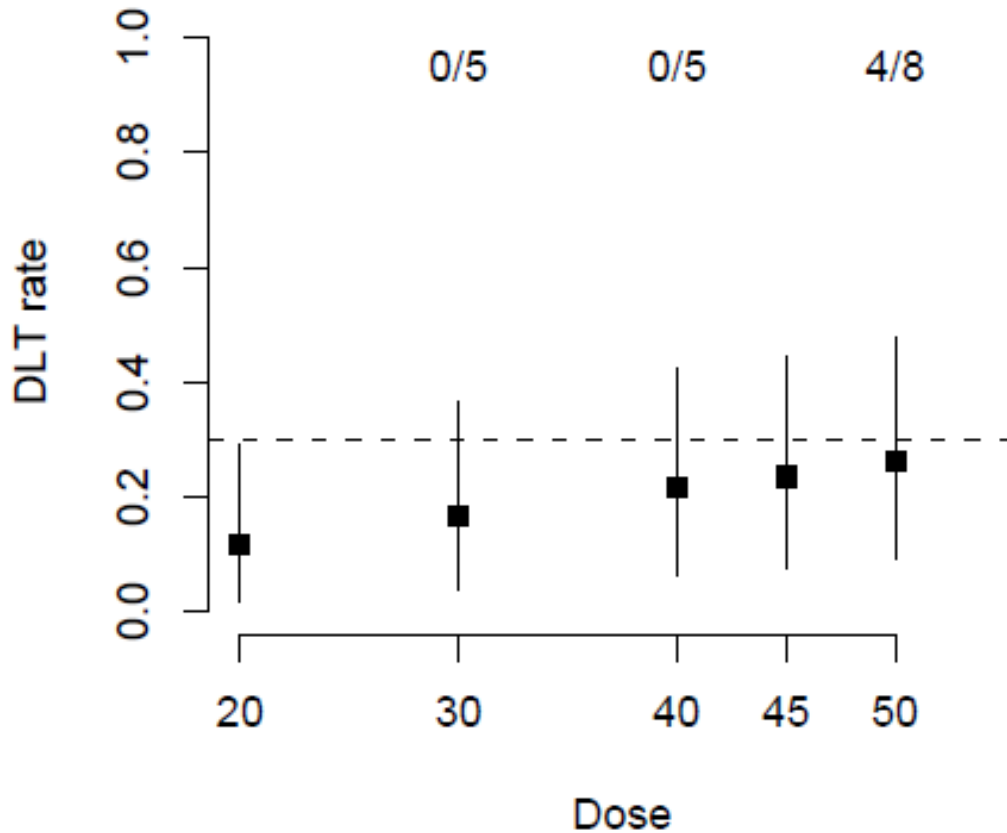
|           | 20mg | 30mg | 40mg | 50mg |
|-----------|------|------|------|------|
| #DLT/#Pts |      | 0/5  | 0/5  | 4/8  |

- Continual Reassessment Method (CRM)
- One-parameter CRM (CRM-1) was used
- MTD recommendation from CRM-1: 50mg
  - Indeed an aggressive recommendation.
  - Is it justified?
- This is one of several examples that raised concerns about the use of model-based designs

# Are model-Based designs too aggressive?

## An Example

CRM-1 posterior DLT rates



### Explanations

- Inappropriate prior?
- CRM algorithm violated?
- Inappropriate model?
- ...?

# A Wish List

*Multiple objectives...*

## ■ Multiple objectives

- Find the MTD: → operating characteristics
- Quantify dose-toxicity relationship: → model-based
- Derive and communicate appropriate summary metrics to clinical teams (tailored to the specifics of the trial)
  - measure for the risk of overdosing (safety!)
- Make use of contextual information: → prior information
- Ensure extendability of approach
  - e.g. including additional information (PK, dose regimens), combination trials, time-to-toxicity, multiple-cycle data, ordinal data...



# Implementation of a Bayesian Approach

... *model-based*

- The following implementation of a Bayesian approach is the Novartis company-standard. It has been used in > 60 phase I studies.
- Base model: logistic regression
  - $\text{logit}(\pi_d) = \log(\alpha) + \beta \log(d/d^*), \alpha, \beta > 0$
- Prior for  $\alpha, \beta$ 
  - $(\log(\alpha), \log(\beta)) \sim \text{BVN}(,)$
  - Alternative: gamma priors for  $\alpha$  and  $\beta$
- The base model is easily extendable

# Implementation of a Bayesian Approach

... uses problem-oriented metrics

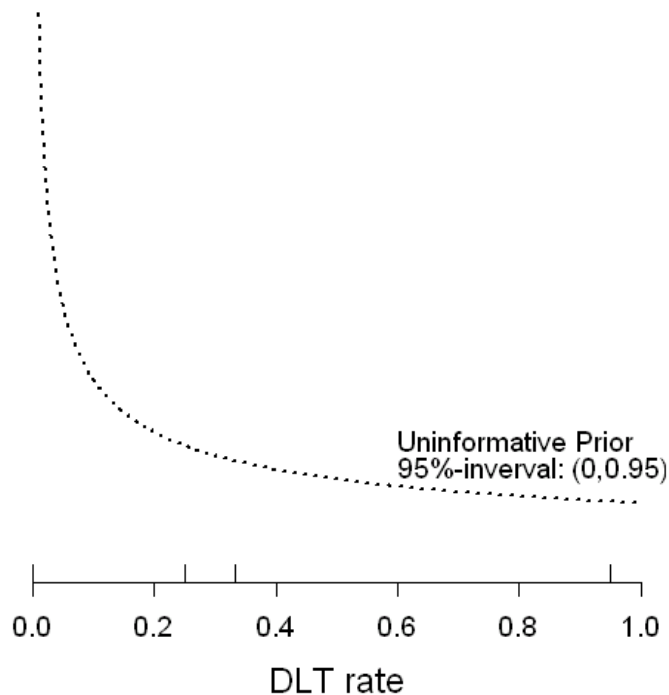
- Ideally, we would like to escalate to the dose with its estimate closest to the target (e.g. 25%)
- However, there are safety concerns
  - Uncertainty needs to be acknowledged
  - Toxicity intervals: true toxicity rate  $\pi_d$ 
    - <0.16 *underdosing*
    - between 0.16 and 0.33 *targeted toxicity*
    - >0.33 *overdosing*
  - Escalation with overdose control (Babb *et al.* 1998):  
Constraint:  $\Pr(\pi_d > 0.33 \mid \text{data}) < 0.25$
  - Alternative: fully Bayesian approach (utility-based)

# Implementation of a Bayesian Approach

... uses *problem-oriented metrics*

Example: 1 DLT in 6 patients.

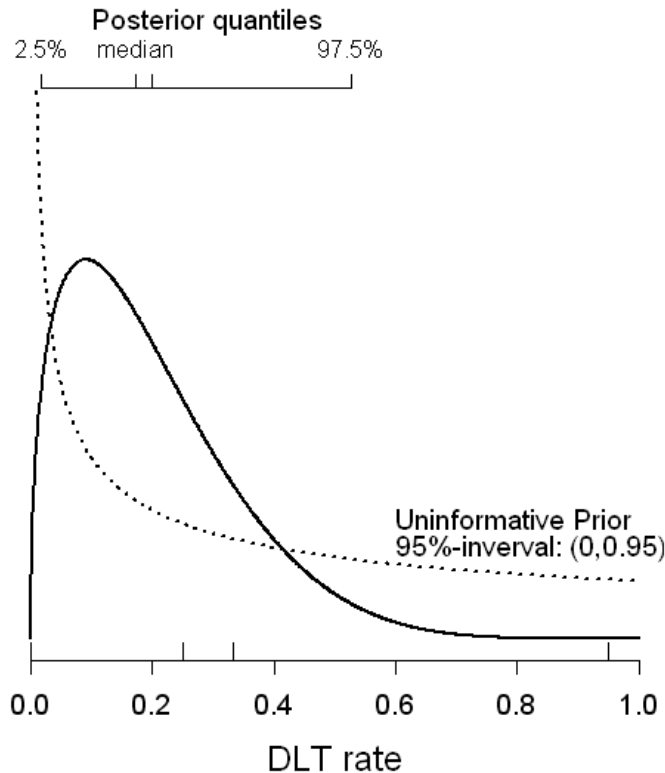
- Weakly-informative prior:  
 $0.25 (0.00, 0.95)_{95\%}$



# Implementation of a Bayesian Approach

... uses problem-oriented metrics

Example: 1 DLT in 6 patients.

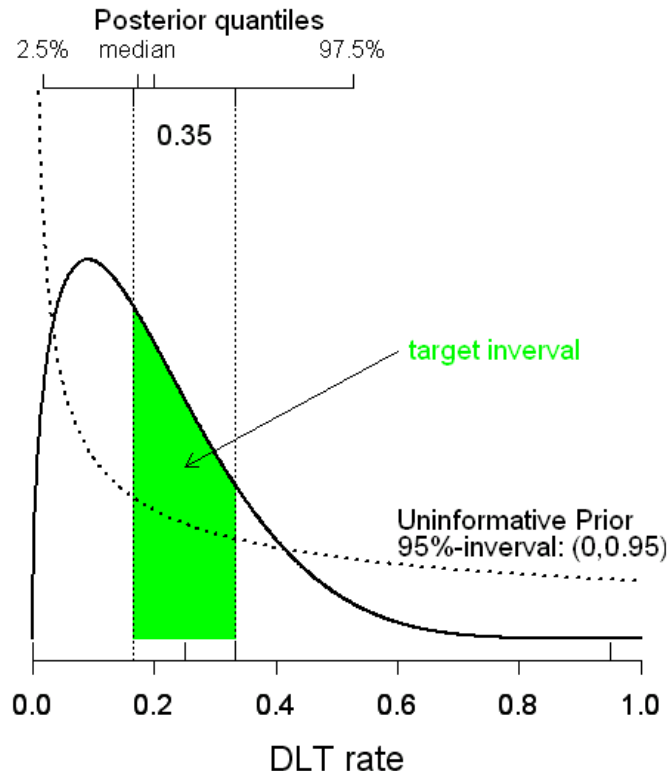


- Weakly-informative prior:  
 $0.25 (0.00, 0.95)_{95\%}$
- Data:  $1/6$
- Summary for  $p$ :  $0.17 (0.02, 0.53)_{95\%}$

# Implementation of a Bayesian Approach

... uses problem-oriented metrics

Example: 1 DLT in 6 patients.

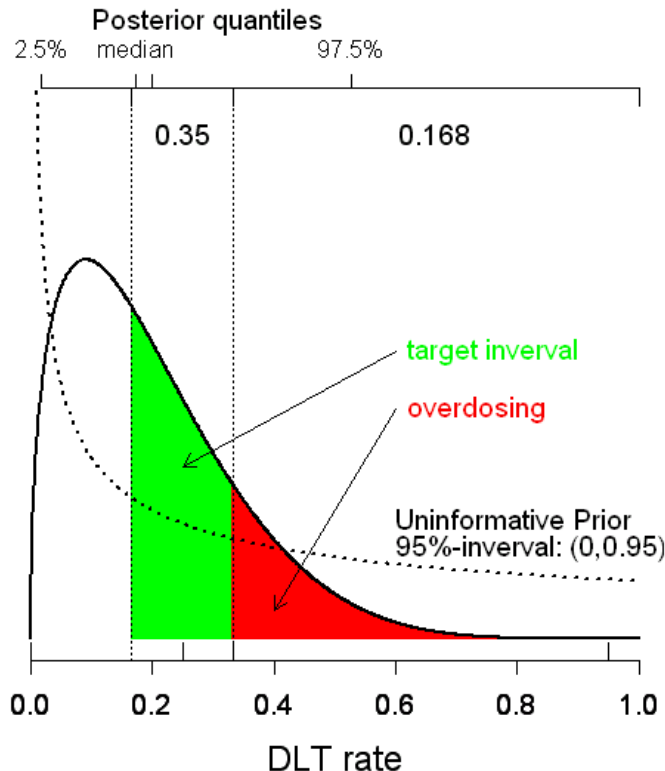


- Weakly-informative prior:  
 $0.25 (0.00, 0.95)_{95\%}$
- Data:  $1/6$
- Summary for  $p$ :  $0.17 (0.02, 0.53)_{95\%}$
- Additional information: there is a
  - **35%** probability for **targeted toxicity**:  
 $p$  in target interval  $(0.16, 0.33)$

# Implementation of a Bayesian Approach

... uses problem-oriented metrics

Example: 1 DLT in 6 patients.

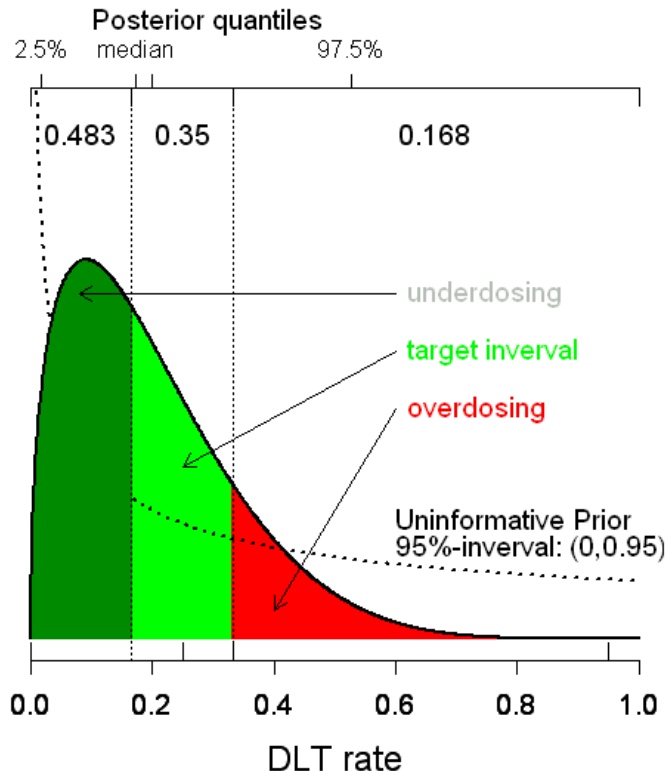


- Weakly-informative prior:  
 $0.25 (0.00, 0.95)_{95\%}$
- Data:  $1/6$
- Summary for  $p$ :  $0.17 (0.02, 0.53)_{95\%}$
- Additional information: there is a
  - **35%** probability for **targeted toxicity**:  
 $p$  in target interval  $(0.16, 0.33)$
  - **16.8%** probability for **overdosing**:  
DLT rate  $p > 1/3$

# Implementation of a Bayesian Approach

... uses problem-oriented metrics

Example: 1 DLT in 6 patients.



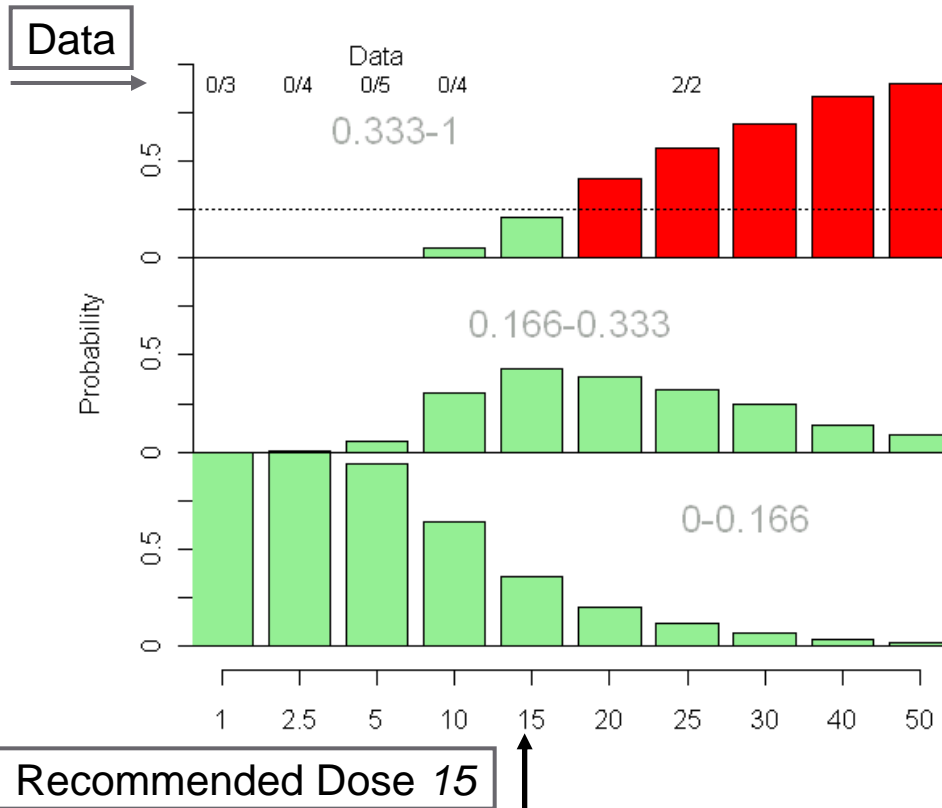
- Weakly-informative prior:  
 $0.25 (0.00, 0.95)_{95\%}$
- Data:  $1/6$
- Summary for  $p$ :  $0.17 (0.02, 0.53)_{95\%}$
- Additional information: there is a
  - **35%** probability for **targeted toxicity**:  
 $p$  in target interval  $(0.16, 0.33)$
  - **16.8%** probability for **overdosing**:  
DLT rate  $p > 1/3$
  - **48.3%** probability for **underdosing**:  
DLT rate  $p < 1/6$

# Implementation of a Bayesian Approach

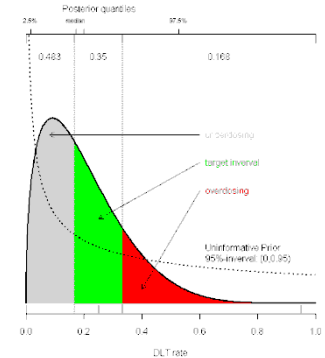
## Communication of Risk: the Risk Plot

### Risk Plot

Interval Probabilities by Dose



overdosing  
targeted toxicity  
underdosing



- ← probability of *overdosing*  
failed overdose criterion in red!  
 $\Pr(\text{true DLT rate} > 0.333 \mid \text{data}) > 0.25$
- ← probability of *targeted toxicity*
- ← probability of *underdosing*

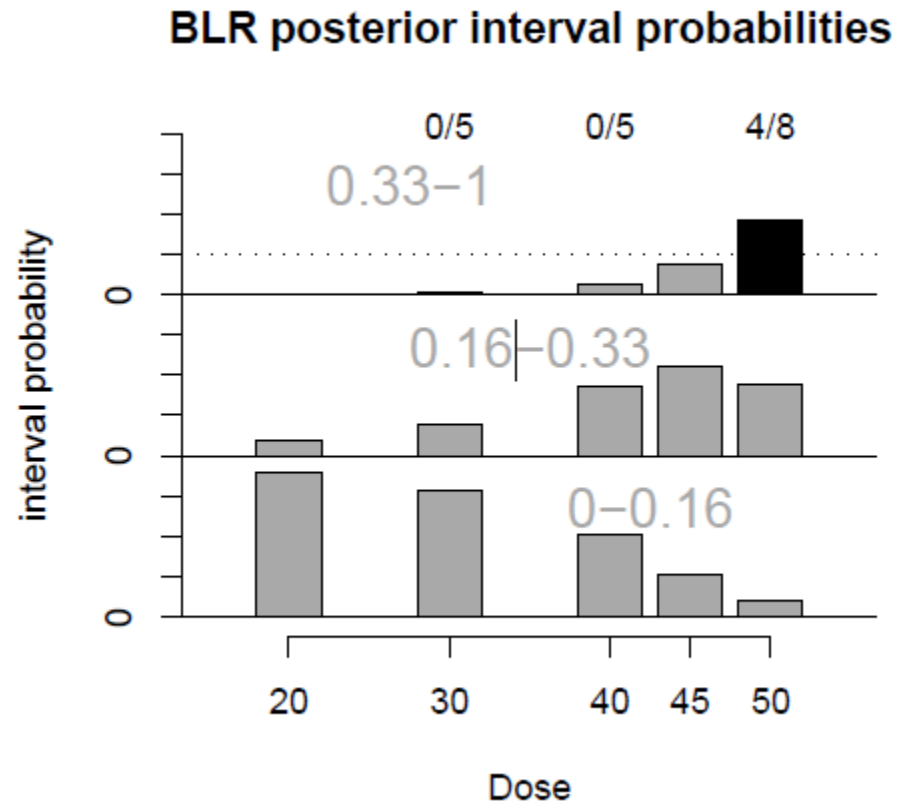
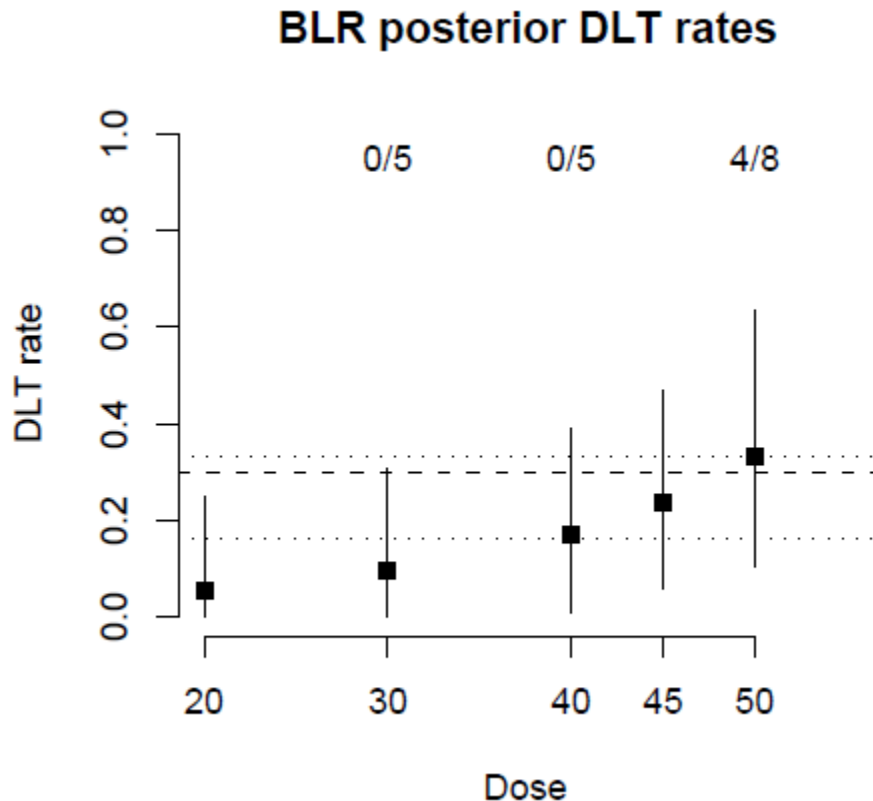


# Implementation of a Bayesian Approach

... Muler et al. JCO 2004 (continued)

## Bayesian logistic regression analysis

- recommended next dose (or MTD) is 40mg, or 45mg if feasible



# Implementation of a Bayesian Approach

*... allows for contextual information*

## ■ Other information

- incorporate (relevant) contextual evidence, not only for designing the study, but also for the analysis
- *“use of historical data”*

## ■ Challenges

- What is relevant?
- How can the evidence be formally incorporated?
- What if historical and actual data are in conflict?

## ■ Requirements: principled statistical approach combined with good judgment

---

# Case Study

## Japanese Phase I Cancer Trial

# Japanese Phase I Cancer Trial

## *Introduction*

---

- First-in-Human study is on-going (Western study)
  - Advanced solid tumors
  - Primary objective:  
to determine the maximum tolerated dose (MTD)
  - Primary endpoint: frequency of dose-limiting toxicity (DLT)
- Now we want to run a dose-escalation study in Japan to determine the Japanese MTD
  - Make use of Western data to accelerate trial but protect patients' safety...
  - Bayesian approach: prior based on Western data

# Japanese Phase I Cancer Trial

## Western First-In-Human Trial: Similarity Scenario

- Western dose limiting toxicity (DLT) data

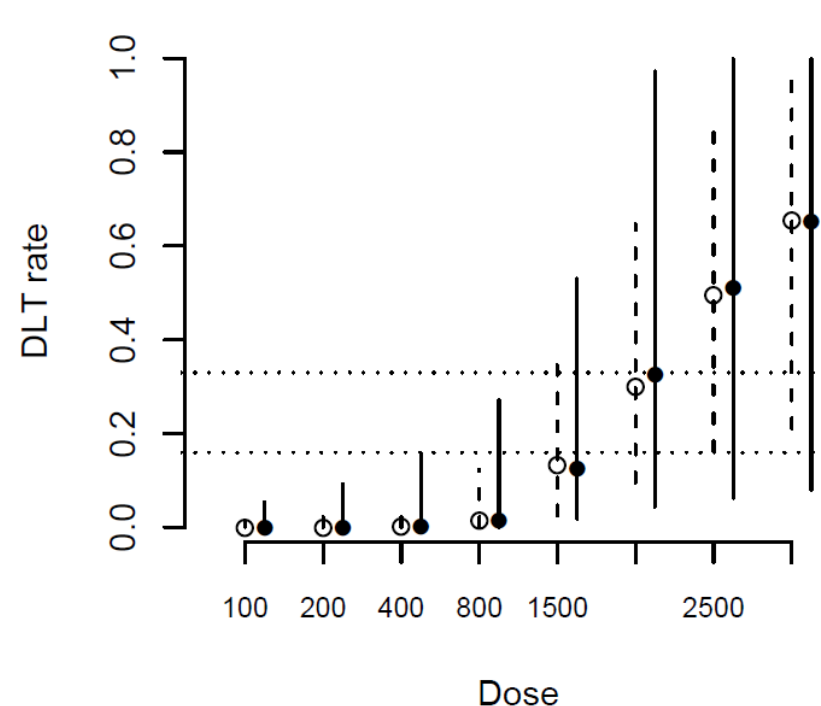
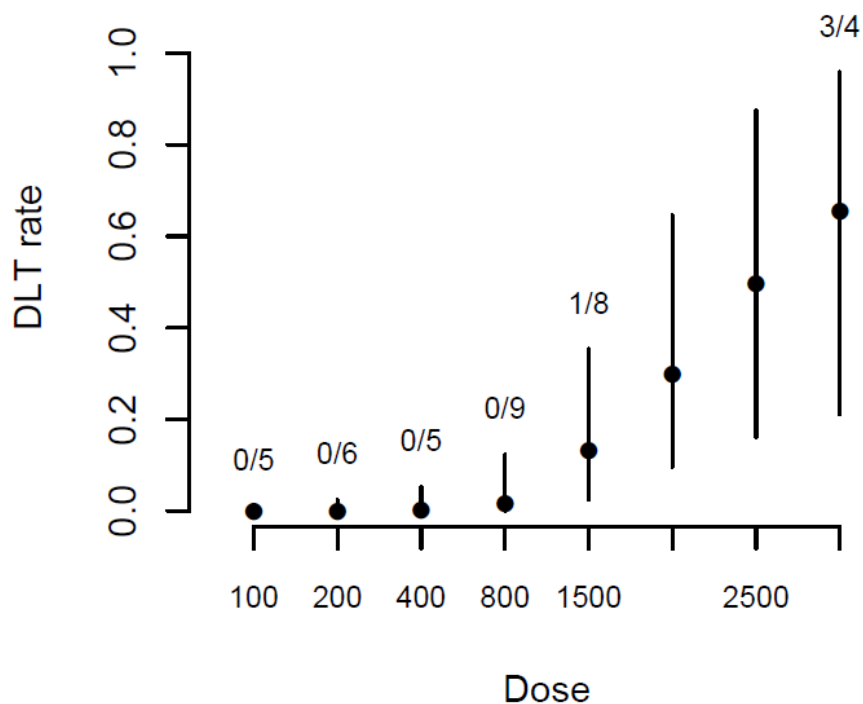
| Dose       | 100 | 200 | 400 | 800 | 1500 | 3000 | TOTAL |
|------------|-----|-----|-----|-----|------|------|-------|
| # Patients | 5   | 6   | 5   | 9   | 8    | 4    | 37    |
| # DLT      | 0   | 0   | 0   | 0   | 1    | 3    | 4     |

- Similarity scenario: Japan trial = “just another Western trial”
- Model for toxicity rate  $\pi_d$  at dose  $d$ 
  - Western:  $\text{logit}(\pi_d) = \log(\alpha) + \beta \log(d/1500)$
  - Japan:  $\text{logit}(\pi_d^*) = \log(\alpha^*) + \beta^* \log(d/1500)$
- Similarity of  $\alpha$  and  $\alpha^*$ , as well as  $\beta$  and  $\beta^*$  (exchangeability)
  - $\log(\alpha), \log(\alpha^*) \sim N(\mu_\alpha, \tau_\alpha^2)$ , and  $\log(\beta), \log(\beta^*) \sim N(\mu_\beta, \tau_\beta^2)$
  - $\tau$ : similarity parameter (between-trial sd)

# Japanese Phase I Cancer Trial

Prior for Japan for “Similarity Scenario”

- Left panel: posterior (Western data)
- Right panel: posterior (Western data, dotted line), and prior for Japan (solid line)  
(assuming substantial heterogeneity: between-trial sd  $\tau=0.5$ )

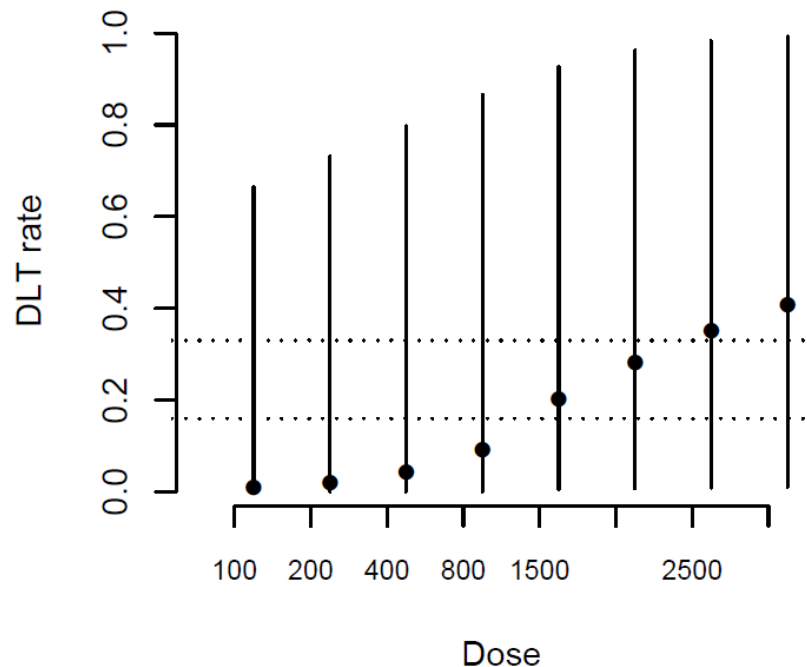


# Japanese Phase I Cancer Trial

## Prior for Japan for “Dissimilarity Scenario”

But what if ...

- Western data is irrelevant (due to relevant ethnicity differences)
- $\Rightarrow$  weakly-informative prior should be used

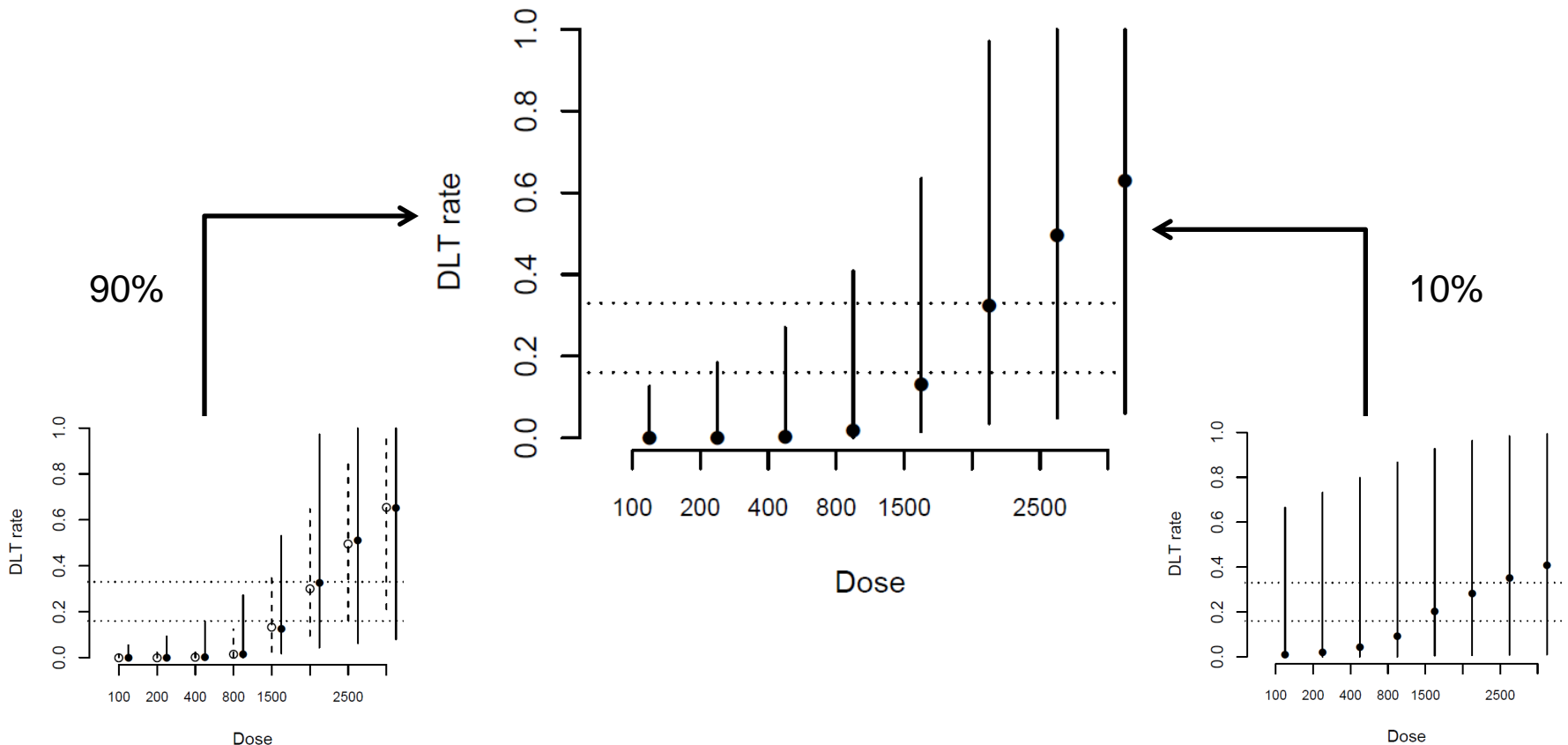


# Japanese Phase I Cancer Trial

*Prior for Japan: Mixture Prior*

Mixture prior (mixture of two BVN distributions) for the two scenarios

- 90% for similarity scenario, 10% for dissimilarity scenario





# Japanese Phase I Cancer Trial

## *Design Properties: Two Examples*

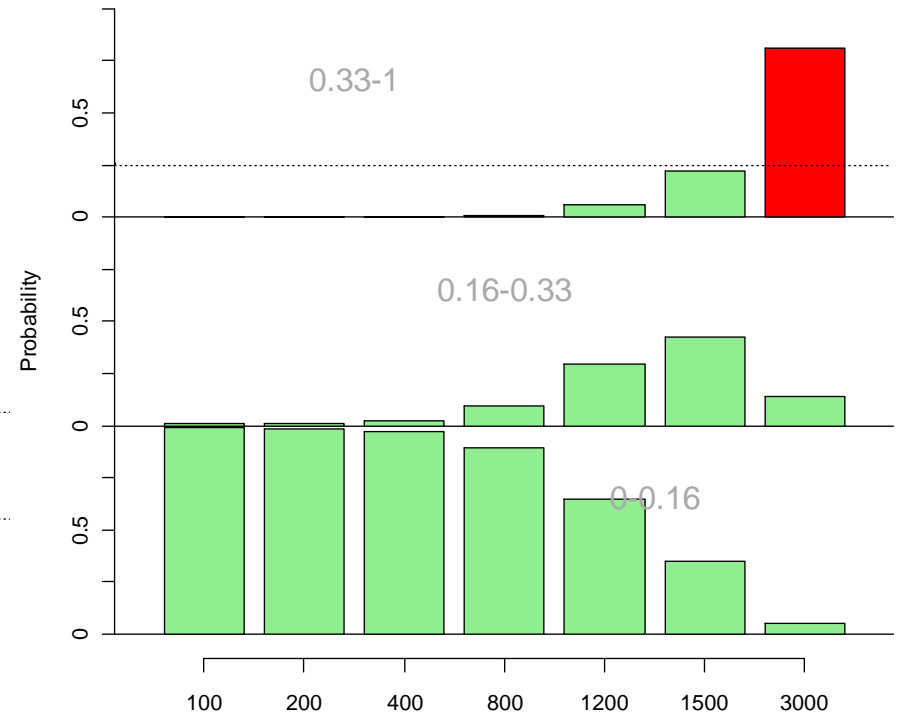
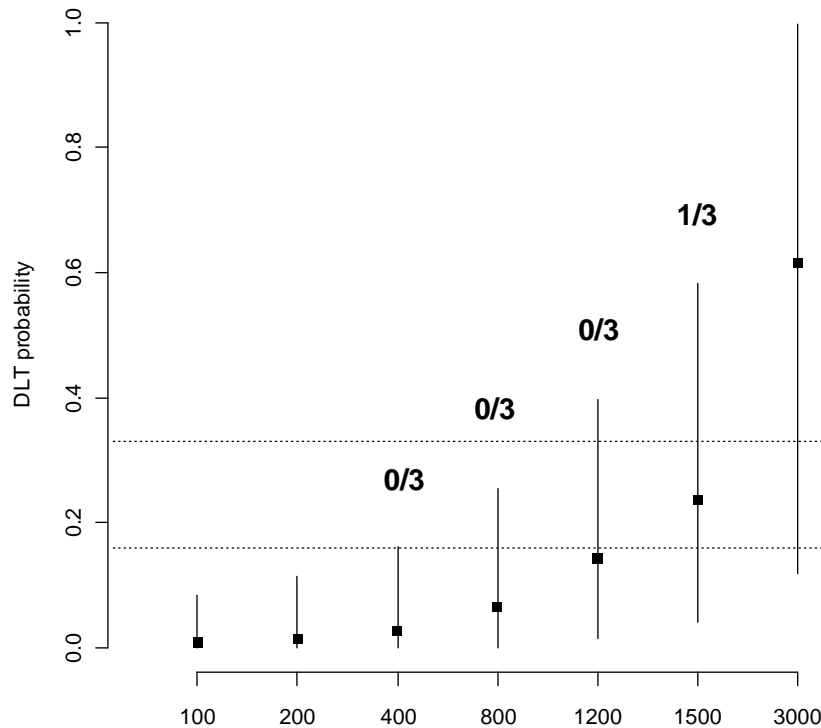
- Design properties
  - (Assess operating characteristics)
  - Assess data scenarios that may arise in the trial: how does the model react to these scenarios?

| Dose                              | 100   | 200   | 400   | 800   | 1200  | 1500  | 3000  |
|-----------------------------------|-------|-------|-------|-------|-------|-------|-------|
| Western Data                      |       |       |       |       |       |       |       |
| #DLT/#Pts                         | 0 / 5 | 0 / 6 | 0 / 5 | 0 / 9 |       | 1 / 8 | 3 / 4 |
| Japan: scenario 1 (similarity)    |       |       |       |       |       |       |       |
|                                   |       |       | 0 / 3 | 0 / 3 | 0 / 3 | 1 / 3 |       |
| Japan: scenario 2 (dissimilarity) |       |       |       |       |       |       |       |
|                                   |       |       | 0 / 3 | 2 / 3 |       |       |       |

# Japanese Phase I Cancer Trial

## 1. Similar data for Japanese patients

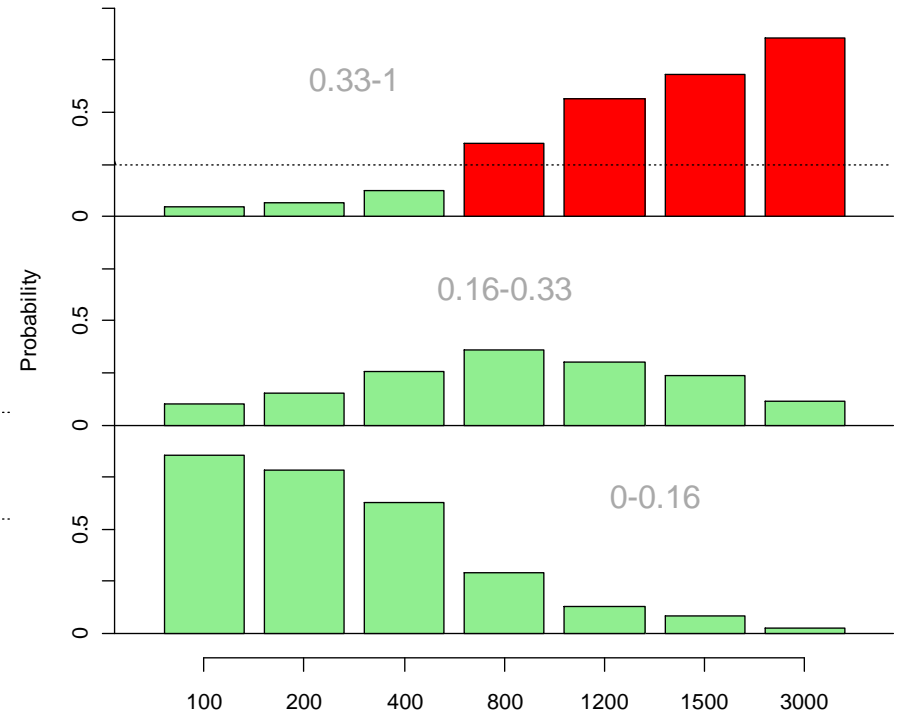
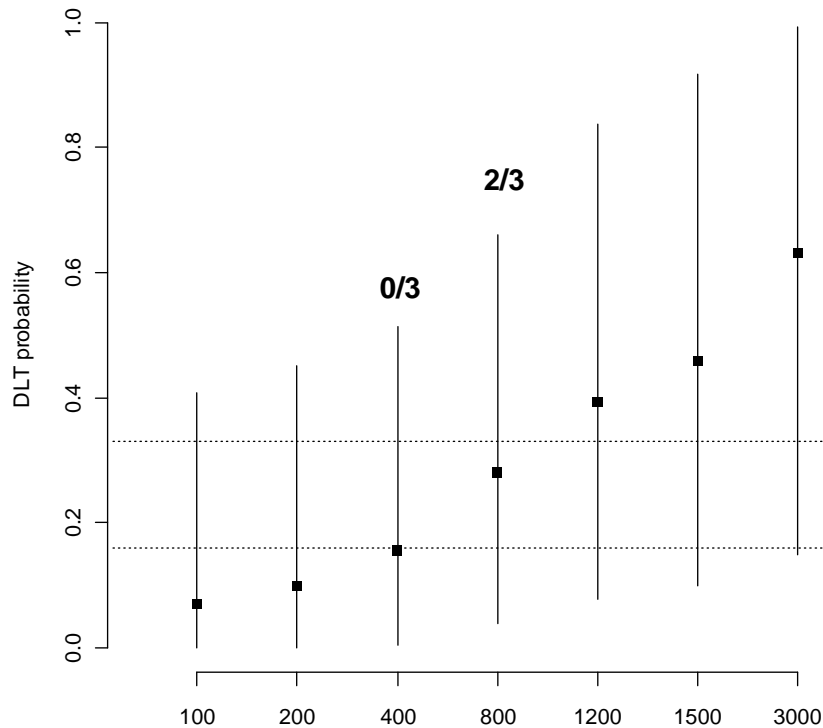
- Dose recommendation: retest at 1500
- Posterior mixture weights: 88% vs. 12%



# Japanese Phase I Cancer Trial

## 2. Dissimilar data for Japanese patients

- Dose recommendation: de-escalate to 400
- Posterior mixture weights: 63% vs. 37%



## Other Aspects

---

- Be aware of other challenges
- Initially, company-internal resistance due to belief that 3+3 design is the gold standard
  - Regular trainings for clinicians and statisticians have resolved these issues
- External review boards asking for 3+3 design (clinical) and operating characteristics (stats)
- Academia: KOL clearly prefer 1-parameter CRM
- Health Authorities:
  - Perception that only 3+3 protects patients' safety (clinical)
  - Lack of guidance with regard to Bayesian approach (stats)

---

# Conclusions

# Conclusions

---

- Massive statistical presence in drug development
- However, statistical innovation is slow, despite important contributions from statistical science
- Phase I is one of many examples
- How can we improve the situation?
- More emphasis on “statistical engineering”
- More comprehensive view, a more open-minded attitude, and better collaboration

# References

- Babb, Rogatko, Zacks. Cancer phase I clinical trials: efficient dose escalation with overdose control. *Statistics in Medicine* 1998
- Bailey, Neuenschwander, Laird, Branson. A Bayesian case study in oncology phase I combination dose-finding using logistic regression with covariates. *JBS* 2009
- Berry, Carlin, Lee, Müller. *Bayesian Adaptive Methods for Clinical Trials* 2010
- Cheung. *Dose Finding by the Continual Reassessment Method* 2011
- Hoerl. The Word is Calling: Should We Answer? *Deming Lecture*, JSM Miami 2011
- Le Tourneau, Lee, Siu. Dose escalation methods in Phase I cancer trials. *J Natl Cancer Inst* 2009
- Muler, McGinn, Normolle et al. Phase I trial using a time-to-event continual reassessment strategy for dose escalation of Cisplatin combined with Gemcitabine and radiation therapy in pancreatic cancer. *Journal of Clinical Oncology* 2004
- Neuenschwander, Branson, Gsponer. Critical aspects of the Bayesian approach to phase I cancer trials. *Statistics in Medicine* 2008
- Neuenschwander, Capkun-Niggli, Branson, Spiegelhalter. Summarizing Historical Information on Controls in Clinical Trials. *Clinical Trials* 2010
- Rogatko, Schoeneck, Jonas et al. Translation of innovative designs into phase I trials. *Journal of Clinical Oncology* 2007
- Spiegelhalter, Abrams, Myles. *Bayesian Approaches to Clinical Trials and Health-Care Evaluation* 2004
- Thall, Lee. Practical model-based dose-finding in phase I clinical trials: methods based on toxicity. *Int J Gynecol Cancer* 2003

---

John Maynard Keynes (1883-1946)

*It is better to be roughly right than precisely wrong.*

Isaac Newton (1642-1727)

*We build too many walls and not enough bridges.*