Designs for phase I oncology trials

Nolan A. Wages, PhD

Associate Professor
Division of Translational Research & Applied Statistics
University of Virginia

Society for Clinical Trials Webinar
October 5, 2017

University of Virginia Health System | Cancer Center

Phase I Trials

- Initial safety trials
- Goal is to recommend a dose for further testing for efficacy in Phase II
- The highest dose with an “acceptable” rate of **dose-limiting toxicity (DLT; yes/no)**, defined by protocol specific adverse events
  - Known as the **maximum tolerated dose (MTD)**
- Ultimate goal is to locate the MTD, while adhering to certain ethical considerations

Overview of dose-finding

- Fundamental assumption: higher doses result in greater chance of DLT (and efficacious response)

Drug X				
Dose label	d_1	d_2	$\cdots$	d_J

- Probability of DLT, $R(d_1) < R(d_2) < \cdots < R(d_J)$
- Assign patient to next highest dose level, only if lower doses are deemed “safe”

Dose-finding Design

Overview

- Phase I and phase I/II
- Small-group-sequential
  - Adapt after every small cohort
- General design strategy
  - Observe a few
  - Estimate a “good” dose
  - Treat at the good dose, and observe

Ethical Considerations in Phase I

- Minimize the number of patients treated at subtherapeutic dose levels
- Minimize the number of patients treated at overly toxic dose levels
- The trial design should quickly escalate through dose levels in the absence of DLT's and quickly de-escalate in the existence of DLT's
- Historically, sample sizes are small (i.e. 15–30).

Existing Phase I Designs

- "Rule-based" methods
  - Variants of the "3+3" method
  - Rolling six design
- "Model-based" methods
  - **Continual reassessment method** (O'Quigley et al., 1990)
  - Escalation with overdose control (EWOC; Babb et al., 1998)
- "Model-assisted" methods (Yan, Madrekar, Yuan, 2017)
  - Cumulative cohort design (CCD; Ivanova et al., 2007)
  - Modified toxicity probability interval design (MTPI; Ji et al., 2010)
  - Bayesian optimal interval design (BOIN; Liu and Yuan, 2015)

Relative performance of competing designs

- Abundance of articles in statistical literature on the poor operating characteristics of 3+3[1]
  - Despite poor operating characteristics, used in $>$ 90% of published phase I oncology trials[2]
- Model-assisted designs have better performance than 3+3, but inferior performance compared to CRM (Horton et al., 2017)
- Model-assisted designs are a special case of semi-parametric dose finding methods (Clertant and O'Quigley; JRSS-B, 2017)

[1] Iasonos A, et al. (2008). *Clin Trials* **5**: 465–477.

[2] Paoletti et al. (2015). *Ann Oncol* **26**: 1808–12.

Continual Reassessment Method (CRM)

O'Quigley, Pepe, and Fisher (*Biometrics*, 1990)

- "A statistical procedure that updates the information on the probabilities of DLT in light of the results obtained for **all** patients already observed"[1]
- "Allocation rule to sequentially assign **each** incoming patient to one of the possible doses, with the intent of assigning doses ever closer to, and eventually recommending, the MTD"[1]
- MTD is defined as the dose level with DLT rate closest to a predetermined target DLT rate θ; i.e. (20%, 25%, 30%, etc.) so that

$$\text{MTD} = \arg\min_j |R(d_j) - \theta|$$

Attributes of CRM

- Working mathematical dose-toxicity model is assumed.
  - Choice in the functional form of model.
- CRM relies upon simple, under-parameterized models
  - Not crucial to accurately estimate entire dose-toxicity curve
- Common choice is 'empiric' model[1]

$$R(d_j) = \Pr(\text{DLT at dose } d_j) \approx \alpha_j^{\exp(a)},$$

where $0 < \alpha_j < 1$ are pre-specified constants (termed **skeleton**) of the working model

[1] **Note:** Other model choices include one-parameter logistic or hyperbolic tangent model

CRM Specifications

- Model-based estimation of DLT probabilities can be done in a Bayesian or frequentist framework
- Bayesian:
  - Estimation based on posterior mean of a
  - What is the prior distribution on the parameter a?
- Frequentist:
  - Estimation based on maximum likelihood estimation of a
  - We need at least 1 DLT and 1 non-DLT to fit the likelihood.
  - How do we begin the trial to obtain this needed data?
- This talk will focus on the Bayesian version.

Inference

Bayesian CRM

- Data: $\mathcal{D} = \{y_j, n_j\}$, # DLT's and patients at each dose
- Likelihood is given by

$$\mathcal{L}(\mathcal{D} \mid a) \propto \prod_{j=1}^{J} \left(\alpha_j^{\exp(a)}\right)^{y_j} \left(1 - \alpha_j^{\exp(a)}\right)^{n_j - y_j}$$

- Denote prior on a by $f(a)$. Estimate DLT probability at each dose

$$\widetilde{R}(d_j) = \alpha_j^{\exp(\widetilde{a})}; \quad \widetilde{a} = \int a \frac{\mathcal{L}(\mathcal{D} \mid a) f(a)}{\int \mathcal{L}(\mathcal{D} \mid a) f(a) da} da$$

Dose-finding algorithm

CRM

1. After each cohort, update dose-toxicity curve, $\widetilde{R}(d_j)$, based on accumulated data at each dose level.
2. Assign next cohort to dose d_j with DLT rate estimated to be closest to target DLT rate; $|\widetilde{R}(d_j) - \theta|$
   - Usually restrict skipping a dose when escalating
3. Continue this process until a fixed number of n patients have been observed (or a stopping rule is triggered).
4. MTD is the recommended dose level for the next $(n+1)$th patient, had one been included

Skeleton Choice

- Skeleton should be chosen to yield robust operating characteristics
- "Reasonable" skeletons in CRM designs are defined by adequate spacing between adjacent levels[1]
- Unreasonable choices:
  - $\alpha_j = \{0.12, \mathbf{0.20}, \mathbf{0.21}, \mathbf{0.22}, 0.36\}$
  - $\alpha_j = \{\mathbf{0.01}, \mathbf{0.20}, \mathbf{0.85}, 0.90, 0.95\}$
- Algorithms available for choosing reasonable skeleton

[1] O'Quigley, Zohar (2010). *J Biopharm Stat* 20: 1013–25

Algorithm for skeleton choice

Lee and Cheung (*Clinical Trials*, 2009)

- Recommendation is to use **getprior**$(\delta, \theta, \nu_0, J)$ function in **R** package **dfcrm**
- θ is target DLT rate (*predetermined*)
- J is the number of test dose levels (*predetermined*)
- ν_0 is the prior MTD
  - Place at median dose level (*recommended*)
- δ is a "spacing measure" termed the half-width
  - Optimal range [0.04-0.08] for common values of θ (*recommended*)

Choice of prior distribution

Lee and Cheung (*Stat in Med*, 2011)

- For empiric model, mean zero normal prior

$$a \sim \mathcal{N}(0, \sigma_a^2)$$

- Prior variance σ_a^2 can calibrated
  - Algorithm yields least informative normal prior
  - Vague in terms of which dose is the MTD
- For a specified skeleton, **R** code available for calibrating least informative prior
  - `http://faculty.virginia.edu/model-based_dose-finding/`

Illustration

- Trial testing $J = 4$ dose levels

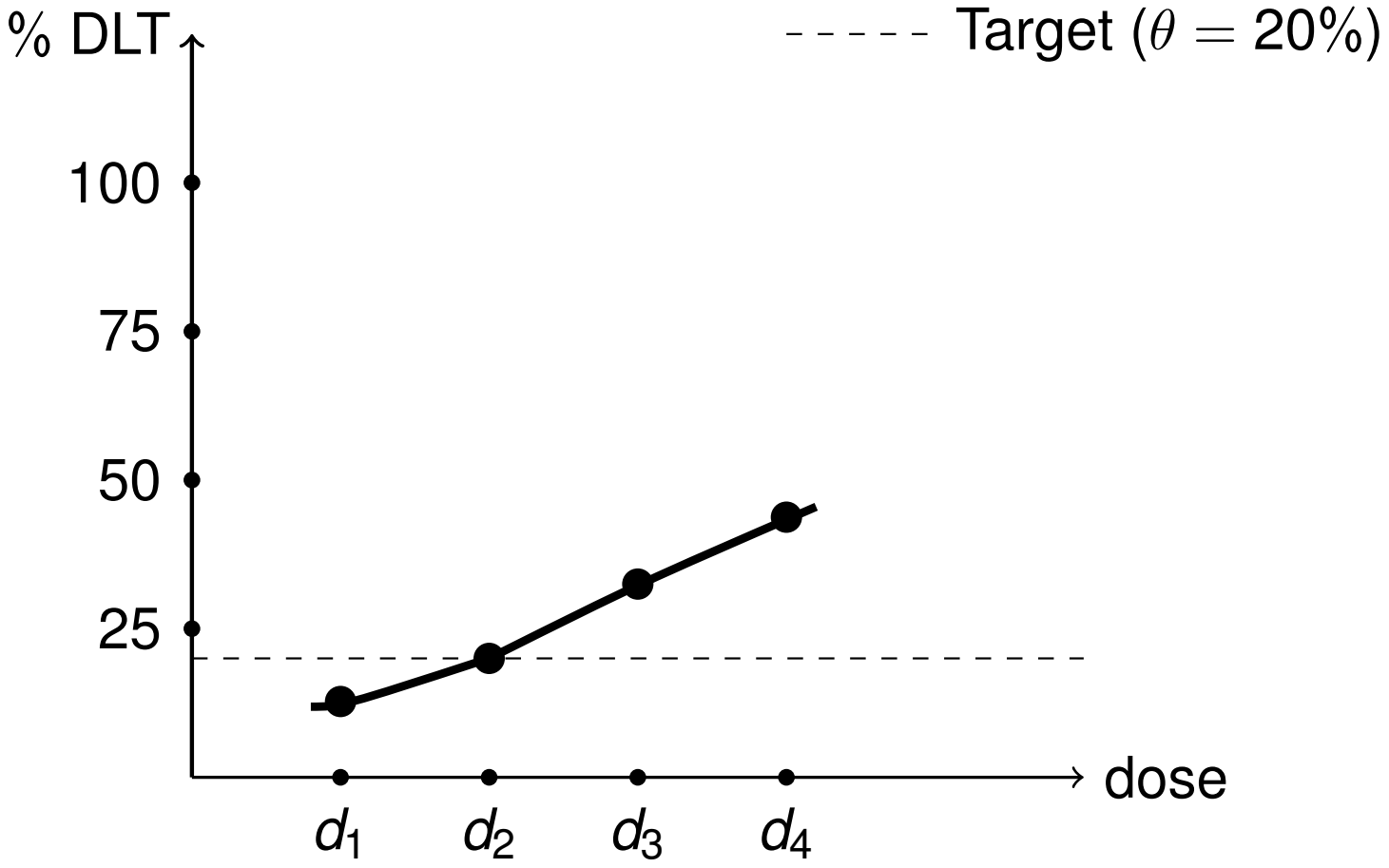
Study dose level	-1	1^*	2	3
Dose label	d_1	d_2	d_3	d_4

***Note:** Starting dose level is always 1, but dose label indexes the possible number of study dose levels explored.

- Predetermined target DLT rate that defines the MTD, $\theta = 0.20$
- Skeleton choice **getprior**$(0.05, 0.20, 2, 4)$
- Prior variance $\sigma_a^2 = 0.27$
- Cohort size 2 patients

Example: skeleton for 4 dose levels

getprior(0.05, 0.20, 2, 4)



Example: working model for 4 dose levels

getprior$(0.05, 0.20, 2, 4)$

	Dose labels			
Model	d_1	d_2	d_3	d_4
$\alpha_j^{\exp(a)}$	$0.11^{\exp(a)}$	$0.20^{\exp(a)}$	$0.31^{\exp(a)}$	$0.42^{\exp(a)}$

Illustration

1st cohort of 2 pts

- Suppose first two patients are put on d_2 and no DLTs are observed
- Accumulated data:

Dose	d_1	d_2	d_3	d_4
# pts	0	2	0	0
# DLT's	0	0	0	0

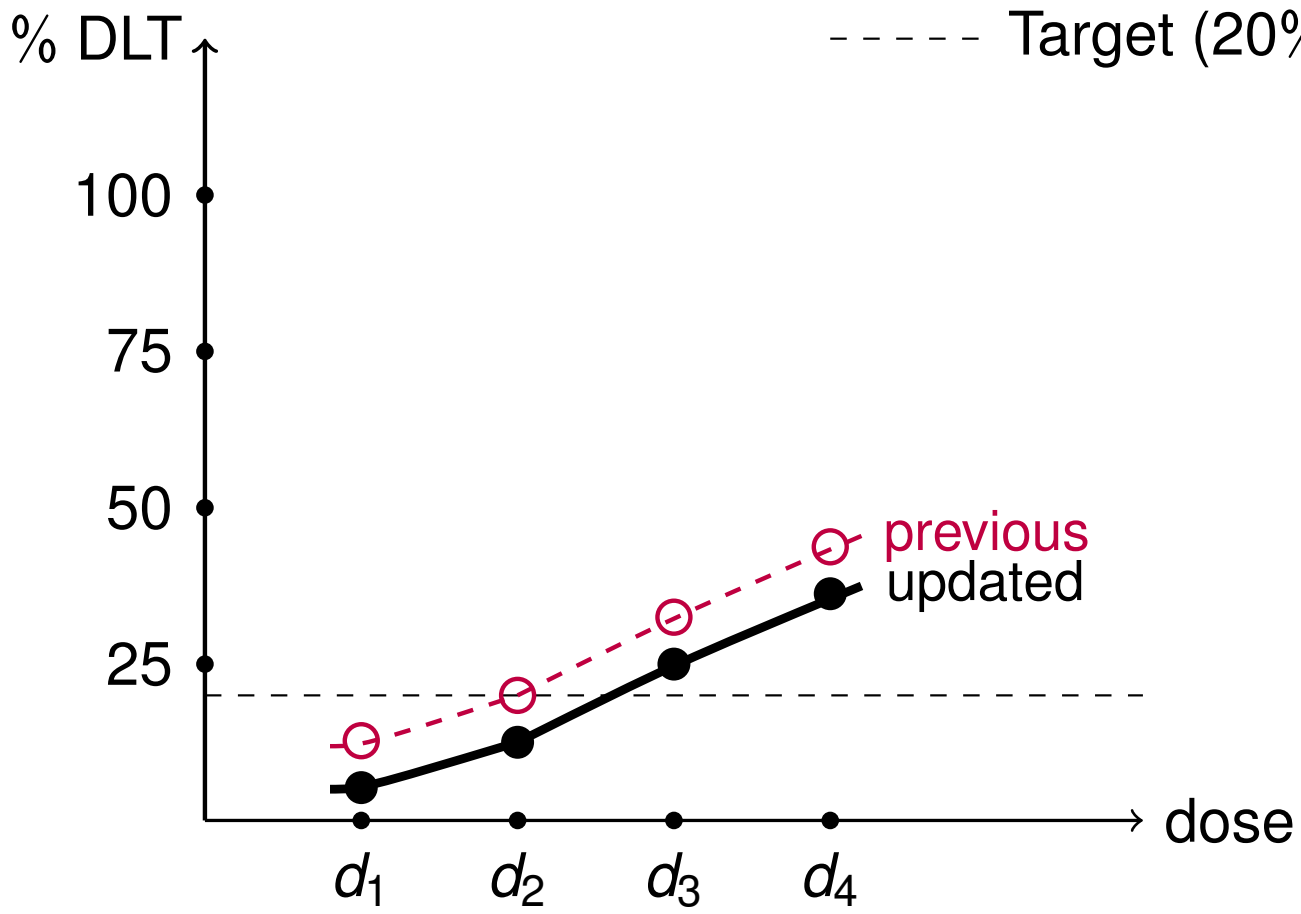
Updated DLT probabilities

After first cohort of 2 pts

- Data: $\mathcal{D} = \{y_2 = 0, n_2 = 2\}$
- Posterior mean of a; $\widetilde{a} = 0.18$

Dose labels			
d_1	d_2	d_3	d_4
$0.11^{\exp(0.18)}$ ≈ 0.07	$0.20^{\exp(0.18)}$ ≈ 0.14	$0.31^{\exp(0.18)}$ ≈ 0.24	$0.42^{\exp(0.18)}$ ≈ 0.36

Updated curve after first cohort of 2 pts



Illustration

2nd cohort of 2 pts

- Suppose second cohort of two patients is put on d_3 and one DLT is observed
- Accumulated data:

Dose	d_1	d_2	d_3	d_4
# pts	0	2	2	0
# DLT's	0	0	1	0

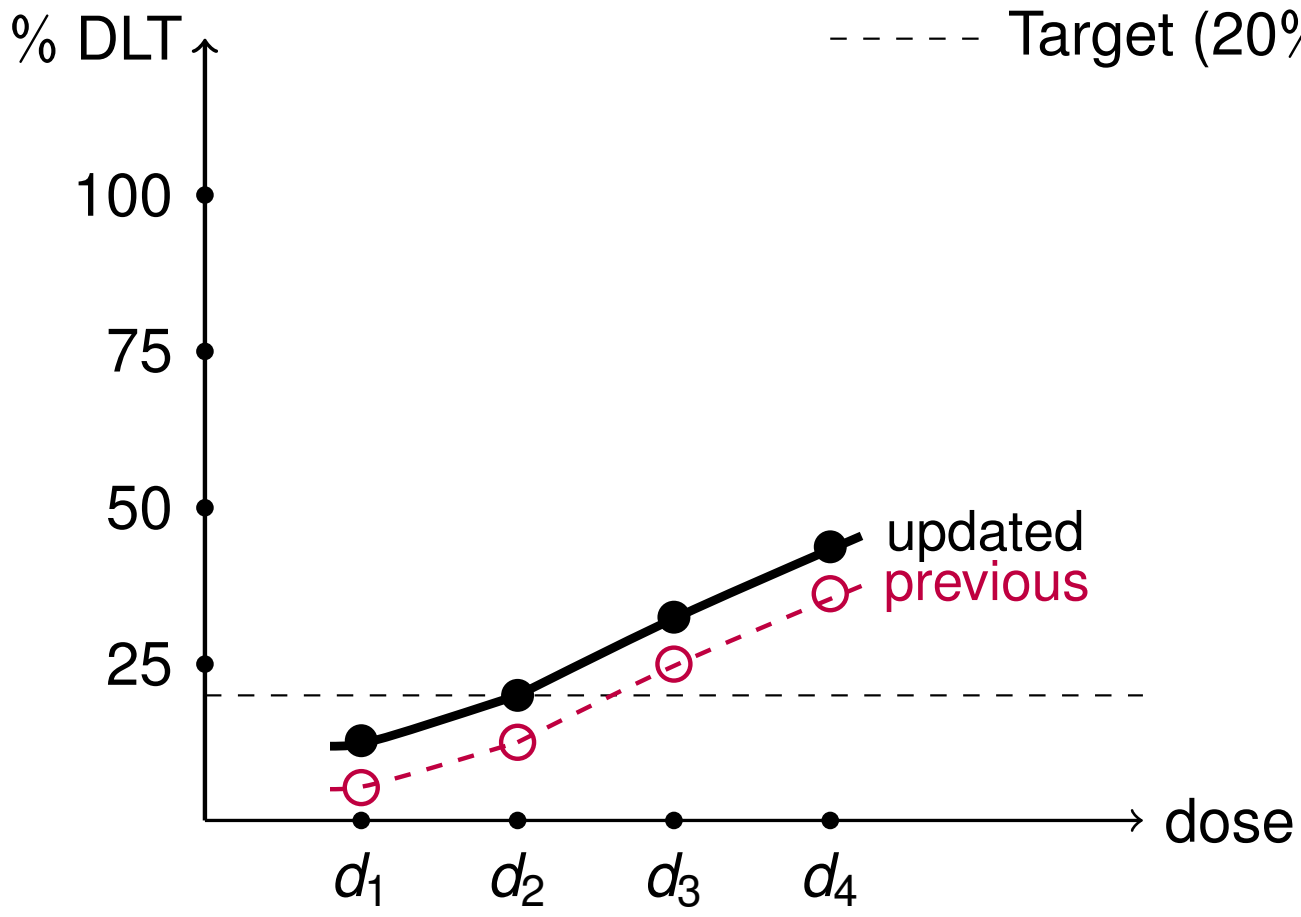
Updated DLT probabilities

After 2nd cohort of 2 pts

- Data: $\mathcal{D} = \{y_2 = 0, y_3 = 1, n_2 = 2, n_3 = 2\}$
- Posterior mean of a; $\widetilde{a} = 0.006$

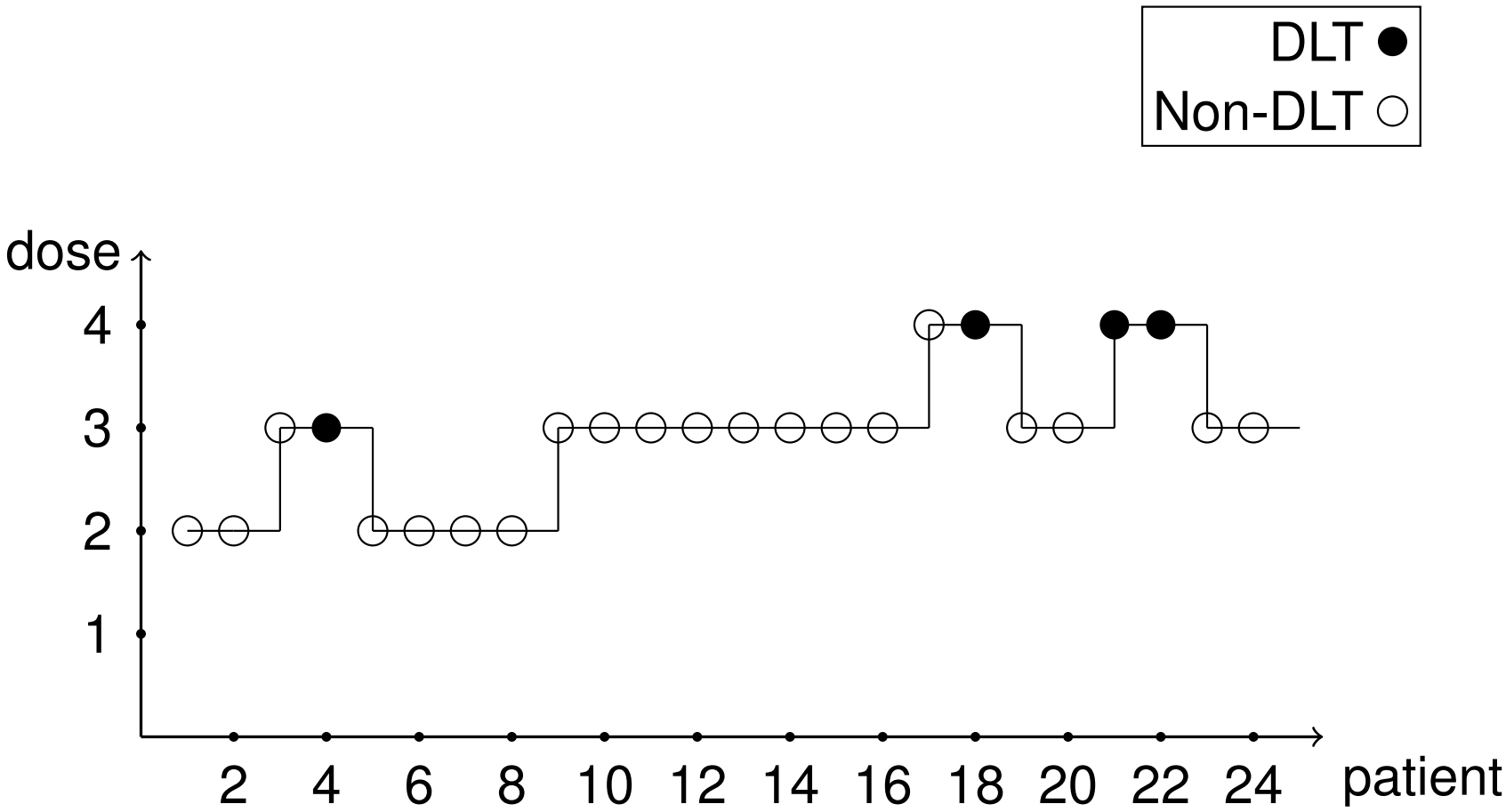
Dose labels			
d_1	d_2	d_3	d_4
$0.11^{\exp(0.006)}$ ≈ 0.11	$0.20^{\exp(0.006)}$ ≈ 0.20	$0.31^{\exp(0.006)}$ ≈ 0.31	$0.42^{\exp(0.006)}$ ≈ 0.42

Updated curve after 2 cohorts



Simulated trial of 24 pts

Assumed true DLT probabilities $\{0.01, 0.09, \mathbf{0.20}, 0.36\}$



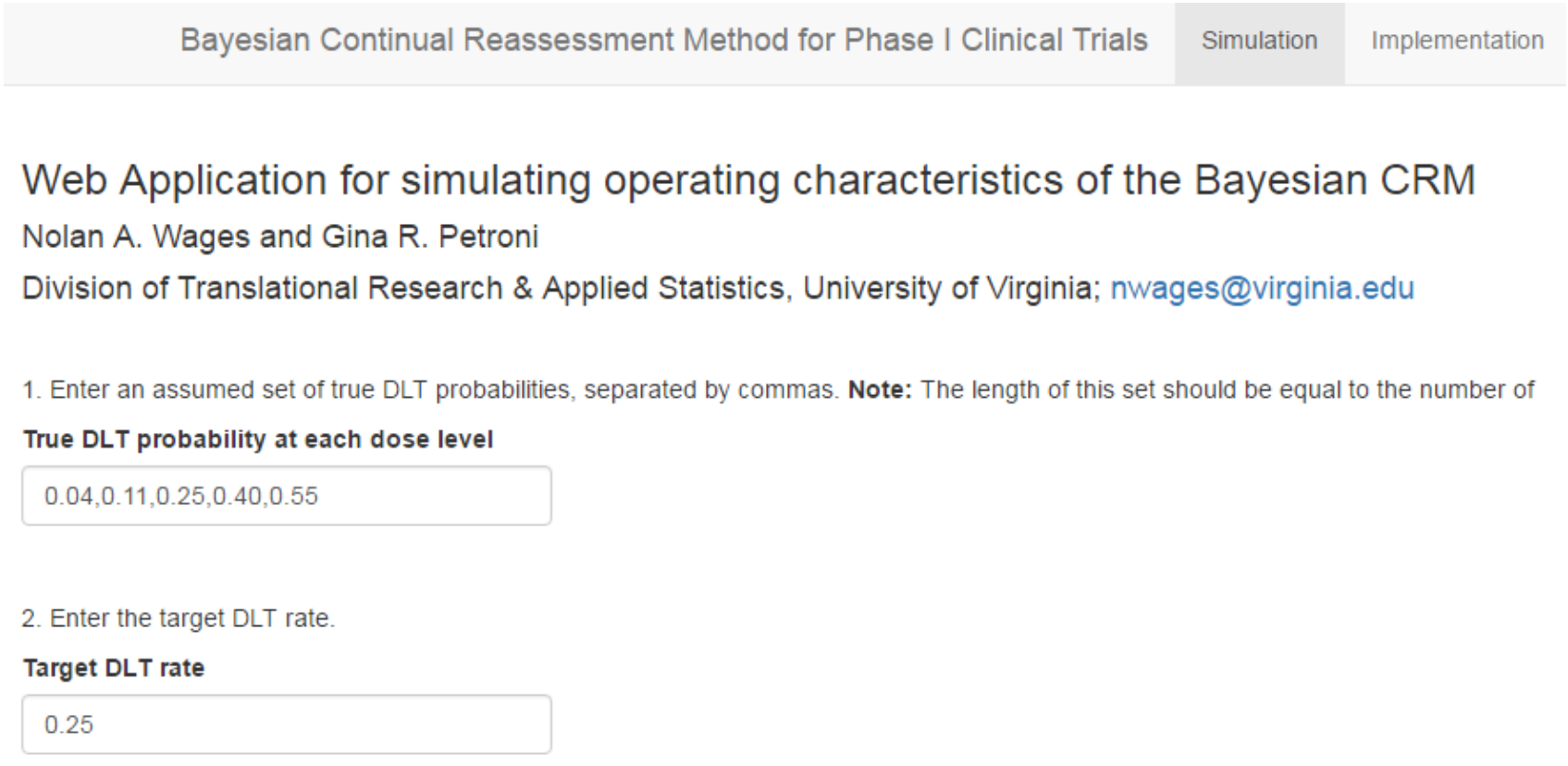
Perceived reasons for infrequent use of CRM[1]

1. "Black-box" mentality, poor understanding of how it works
   - **Truth: Rooted in sound statistical principles**
2. Sensitivity to choice of design specifications
   - Working dose-toxicity model and prior distributions
   - **Truth: Published recommendations with robust choices**
3. Computationally burdensome
   - Requires more time, effort, and personnel; regular interaction between clinical and statistical team
   - **Truth: Bayesian CRM web tool provides computational ease**

[1] Cheung YK. *Dose-finding by the continual reassessment method*; CRC Press: New York, 2011.

Bayesian CRM Web Tool

https://uvatrapps.shinyapps.io/crmb/



Bayesian CRM Web Tool

https://uvatrapps.shinyapps.io/crmb/

- Simulates operating characteristics

Simulation	Implementation

- Computes the recommended dose level for the next patient based on accumulated data

Simulation	Implementation

Bayesian CRM Web Tool

Simulation Input 1–2

1. Enter an assumed set of true DLT probabilities, separated by commas.
   Note: The length of this set should be equal to the number of possible study dose levels

True DLT probability at each dose level

0.01, 0.09, 0.20, 0.36

2. Enter the target DLT rate

Target DLT rate

0.20

Bayesian CRM Web Tool

Simulation Input 3–5

3. Enter the cohort size required before the next model-based update. Cohort size may be 1, 2, or 3 patients.

Cohort size

2

4. Enter the maximum sample size for the study. This number should be a multiple of the cohort size entered above.

Maximum number of patients

24

5. Enter the number of simulations. A minimum of 1000 is recommended.

Number of simulated trials

1000

Bayesian CRM Web Tool

Simulation Input 6–7

6. Enter the index of the starting dose level. **Note:** Index of lowest dose level is always 1. If the design allows for 'minus' dose levels (i.e. -2, -1, etc.), then the index of the starting dose should account for these lower levels (i.e. if -1 dose level allowed, starting dose is 2.)

Index of starting dose level

2

7. Set the seed of the random number generator.

Random seed

580

Run simulation study

Bayesian CRM Web Tool

Simulation Output

Skeleton of working model:	0.11	0.20	0.31	0.42
True DLT probability:	0.01	0.09	**0.20**	0.36
MTD selection percentage:	0.30	26.5	**56.9**	16.3
Average number of DLTs:	0.00	0.70	2.1	0.9
Average number of patients:	2.51	7.81	10.94	2.74
Percent stopped for safety:	0			

Bayesian CRM Web Tool

Implementation Input 2–4

2. Enter number of observed DLTs at each dose level. If none have been observed or a dose level has not yet been tried, enter '0'. **Note:** The length of this set should be equal to the number of possible study dose levels.

Number of observed DLTs at each dose level

0, **0**, 0, 0

3. Enter the number of patients evaluated for DLT at each dose level. If a dose level has not yet been tried, enter '0'. **Note:** The length of this set should be equal to the number of possible study dose levels.

Number of patients evaluated for DLT at each dose level

0, **2**, 0, 0

4. Enter the most recent dose level administered in the study. Get updated recommended dose level.

Current dose level

2

Get updated recommended dose level

Bayesian CRM Web Tool

Implementation Output

Skeleton of working model:	0.11	0.20	0.31	0.42
Number of DLTs:	0	0	0	0
Number of patients:	0	2	0	0
Estimated DLT probabilities:	0.07	0.14	0.24	0.36
Target DLT rate:	0.20			
Recommended dose level:	3			

Bayesian CRM Web Tool

Updated Data

2. Enter number of observed DLTs at each dose level. If none have been observed or a dose level has not yet been tried, enter '0'. **Note:** The length of this set should be equal to the number of possible study dose levels.

Number of observed DLTs at each dose level

0, 0, **1**, 0

3. Enter the number of patients evaluated for DLT at each dose level. If a dose level has not yet been tried, enter '0'. **Note:** The length of this set should be equal to the number of possible study dose levels.

Number of patients evaluated for DLT at each dose level

0, 2, **2**, 0

4. Enter the most recent dose level administered in the study. Get updated recommended dose level.

Current dose level

3

Bayesian CRM Web Tool

Updated Output

Skeleton of working model:	0.11	0.20	0.31	0.42
Number of DLTs:	0	0	1	0
Number of patients:	0	2	2	0
Estimated DLT probabilities:	0.11	0.20	0.31	0.42
Target DLT rate:	0.20			
Recommended dose level:	2			

Notes on web app

https://uvatrapps.shinyapps.io/crmb/

- Utilizes a set of default design specifications based on practical recommendations from literature
- These specifications produce robust operating characteristics.
  - Contains the type of simulation information that aid clinicians and reviewers in understanding operating characteristics for the accuracy and safety of the CRM
- The bottom of the web page contains detailed notes about the design specifications, including the skipping restriction and safety stopping rule.
  - For input in a protocol statistical section.

Conclusions on web app

https://uvatrapps.shinyapps.io/crmb/

- The web tool provides a mechanism for conducting the Bayesian CRM in a timely and reproducible fashion, requiring no programming knowledge.
- Free to access and use on any device with an internet browser, including a smart phone.
- Can easily be used to compare CRM to other methods with available software
- We hope this leads to broader implementation of CRM and will facilitate more efficient collaborations within study teams.

Evaluating performance

- Generate operating characteristics (OC) via simulation studies.
- Accuracy
  - % that correctly identify true MTD
  - Accuracy index (*next slide*)
- Safety
  - expected # of DLTs at each dose level
  - % of patients treated above MTD; i.e., risk of overdosing
- Conducted under a broad range of assumed dose-toxicity curves

Accuracy Index

- For a sample size of n, accuracy index of Cheung (2011)

$$A_n = 1 - J \times \frac{\sum_{j=1}^{J} |R(d_j) - \theta| \Pr(\text{selecting dose } j)}{\sum_{j=1}^{J} |R(d_j) - \theta|},$$

  is a summary of the distribution of the selected dose through its weighted average.
- Weights are the distances from $R(d_j)$ to θ
- Its maximum value is 1 with larger values (close to 1) indicating that the method possesses high accuracy

[1] Cheung YK. *Dose-finding by the continual reassessment method*; CRC Press: New York, 2011.

Comparing methods

Challenges

- A conclusion may be reached under one set of assumed curves that does not old under another set
- Can be the case for Bayesian design where the impact of prior information can be difficult to evaluate
  - How well does the prior align with some chosen truth?
  - Can favor performance in certain situations and hinder in others
  - The choice of curves to show then becomes subjective
- Important to consider how well a design can possibly perform.

Non-parametric optimal benchmark

O'Quigley, Paoletti, and Maccario (*Biostatistics*, 2002)

- Theoretical tool for simulation studies
- Upper bound on the accuracy of MTD selection for a binary toxicity endpoint
  - Gives a sense of the plausibility of a methods operating characteristics
  - Does not account for patient allocation
  - Can only be used as a simulation tool, not in practice
    - Assumes knowledge of true, underlying dose-toxicity curve
- Is it possible to outperform the benchmark?

Super-optimality

- O'Quigley et al. (2002) showed that it is not generally possible to beat the benchmark based on the observations themselves
  - Admissible designs
- Super-optimality requires extraneous knowledge
  - i.e., informative prior that favors the true MTD in a particular scenario
- How would the design "always choose level 3" perform...
  - when the true MTD is level 3?
  - when the true MTD is some other level?

Simulating the benchmark **R** shiny

https://uvatrapps.shinyapps.io/nonparbnch/

1. Enter a set of assumed true DLT probabilities, separated by commas. The length of this set should be equal to the number of dose levels.

True DLT probability at each dose level

0.01, 0.09, 0.20, 0.36

2. Enter the target DLT rate for the study.

Target DLT rate

0.20

3. Enter the number of simulated trials to be generated. A minimum of 1000 is recommended.

Number of simulated trials

2000

4. Set the seed of the random number generator.

Random seed

580

Run simulation study

R shiny app for non-parametric benchmark[1]

Simulation Output

- MTD selection percentage for each dose
  - Percent correct selection (PCS) is 63.4%
- Accuracy Index

True DLT probability:	0.01	0.09	**0.20**	0.36
MTD selection percentage:	0.10	19.7	**63.4**	16.8
Accuracy index:	0.5765			

[1] Wages and Varhegyi (*Clinical Trials*, 2017)

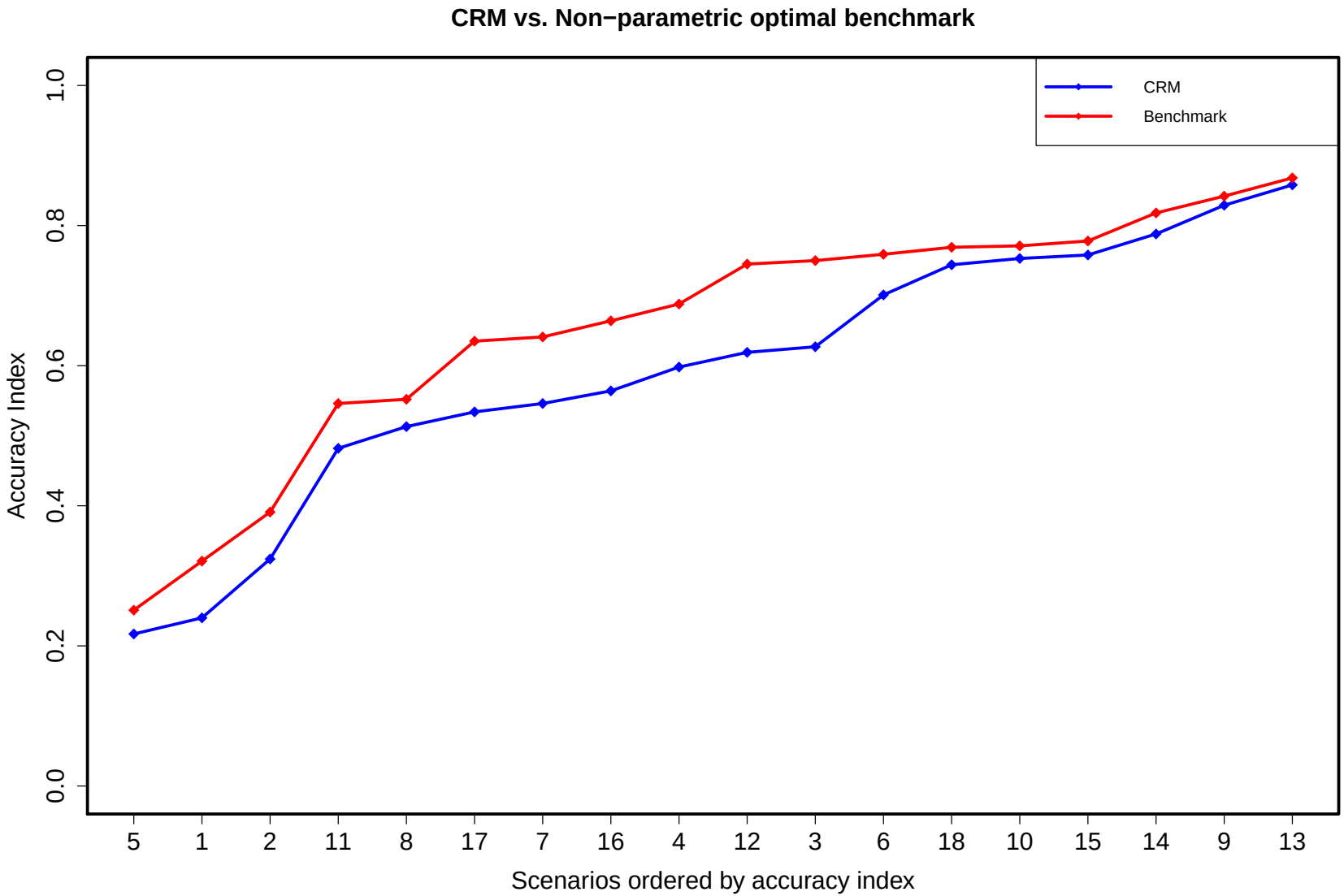
Performance of CRM relative to benchmark

- CRM has excellent statistical properties in terms of correctly identifying, as the MTD, doses at and around the target dose compared to the benchmark[1]
- Simulation study performed over 18 assumed dose-toxicity scenarios
  - Various target DLT rates: $\theta = \{0.20, 0.25, 0.30\}$
  - Various sample sizes: $n = \{20, 25, 30\}$
  - Scenarios reflect a mixture of steep, flat, and intermediate curves
  - Various cohort sizes: 1, 2, or 3 patients.
- The following slide reports the accuracy index over the 18 scenarios.

[1] Wages, Conaway, O'Quigley (*Clinical Trials*, 2013)

Simulation results

Wages, Conaway , O'Quigley (*Clinical Trials*, 2013)



Efficiency

- Over many scenarios, Cheung (2011) reported that CRM is 86% efficient in terms of average PCS relative to the benchmark

$$\frac{\overline{PCS}_{CRM}}{\overline{PCS}_{Optimal}} = 86\%$$

- Wages et al (2013) reported that CRM is 91% efficient in terms of average accuracy index A relative to the benchmark

$$\frac{\overline{A}_{CRM}}{\overline{A}_{Optimal}} = 91\%$$

Other available software

Web apps

- Comparison of simulated operating characteristics for competing methods
  - `https://cqs.mc.vanderbilt.edu/shiny/AdaptiveDesignS/`
  - Cannot be used for implementation of any design
- AplusB - operating characteristics of A+B designs
  - `https://graham-wheeler.shinyapps.io/AplusB/`
- BOIN
  - `http://www.trialdesign.org/`

CRM extensions

- Late-onset DLT
  - Time-to-event (TITE) CRM (Cheung and Chappell, 2000)
- Multidimensional dose-finding problems
  - Drug combinations (Wages et al., 2011)
  - Different treatment schedules (Wages et al., 2014)
  - Patient heterogeneity (Iasonos and O'Quigley, 2014)
  - Toxicity and efficacy (Wages and Tait, 2015)

Thank you!

Questions?