Impact of model complexity on adaptive dose-finding methods

Alexia Iasonos, Ph.D.
Memorial Sloan Kettering Cancer Center
Nolan A. Wages, Ph.D.
University of Virginia

Outline

- Objectives of Phase I or early phase trials
- Model based designs
- Continual Reassessment Method
- Design properties
- Systematic Review; Simulation studies
- Examples
- Evaluating performance: optimal
- More complex settings: combination trials
- Recommendations

Why have Phase I trials become so complicated?

Simplest case	More Complex cases
• Single agent • Single schedule • MTD • 5-6 levels • N=20-25	• Combination agents • 2 schedules • MTD (1 or >1) • OBD • multiple disease groups • DLT definition (onset, attribution AE) • starting dose relative to MTD • N=50 – 60 dose escalation • >120 (25-40 per cohort)

Need to use efficient designs

Is the drug safe and at which dose? Which patient population and which drug/regimen to prioritirize?

- Success with single agent targeted therapies
- Develop resistance because of multiple genetic alterations and advanced metastatic disease
- Regimens with 1 or more targeted agent
- Many single agents/ combination regimens in the pipeline. Competing Resources

- Minimize number of patients and trial duration

What should a design be able to achieve?

- Answer scientific question
  - find the MTD or OBD
  - evaluate safety - DLTs
- Ethical
  - safe
  - optimal, efficient (how to get to the answer)
  - patient allocation: over dose, under dose, min sample size

How to answer the study's objectives

- Scientific valid – answer's primary objective
- How does it get to the right answer?

- Safe
- Efficient / optimal in terms of number of patients and trial duration

Protocol development and approval

- Getting timely IRB approval from multiple sites
- Protocol Scientific Review
  - Iasonos, Gonen, Bosl, JCO 2015
  - Petroni G et al. Stats Med 2016

- Provide Operating Characteristics
  - How accurate ?
  - At which levels will patients be treated?
  - Aggressive vs conservative dose escalation?

Operating characteristics

Scientific valid	Accurate: finds the right dose
Safe	Patient allocation: overdose underdose
Efficient	Sample size/ duration

- Flexible and clinically sensible
- Aggressive/ conservative escalation

Why use a model to guide escalation?

- More accurate (30% vs >60%)
- Meets ethical and scientific criteria :
  - treats fewer patients at suboptimal dose levels by getting to the MTD faster
  - requires the same or fewer total number of patients as other established methods (3+3)
  - safe (1 dose at a time)
  - shorter trial duration

Continual Reassessment Method

$$\psi(d_i, a) = \alpha_i^{\,a}, g(a) = \exp(-a), \textit{prior}$$

$$F_j = \{(d_1, y_1), \ldots (d_{j-1}, y_{j-1})\}, \textit{data}$$

$$f(a, F_{j+1}) = \frac{g(a)\prod_{l=1}^{j} \psi(d_l, a)^{y_l}[1-\psi(d_l, a)]^{1-y_l}}{\int_0^{\infty} g(u)\prod_{l=1}^{j} \psi(d_l, u)^{y_l}[1-\psi(d_l, u)]^{1-y_l}\, du}, \textit{posterior}$$

$$\hat{a}(j) = \int a f(a, F_j) da, \textit{mean}$$

$$\hat{p}_i(j) = \psi(d_i, \hat{a}(j))$$

$$\min \Delta(\hat{p}_i(j) - \theta)$$

Continual Reassessment Method

- Tuning parameters allow for different clinical scenarios
- Simulation studies and systematic review proved to be safe and accurate
- 3/53 problematic trials: cohort size, prior specification, DLT window of observation
- no need for ad-hoc rules to correct the model's recommendation

Design properties

- **Consistency of estimators.** By consistent we mean the usual convergence with probability one (almost sure convergence) to the population value. In particular we would like, under certain conditions, to be able to claim that the recommended dose converges almost surely to the true MTD (O'Quigley, Shen 1996).

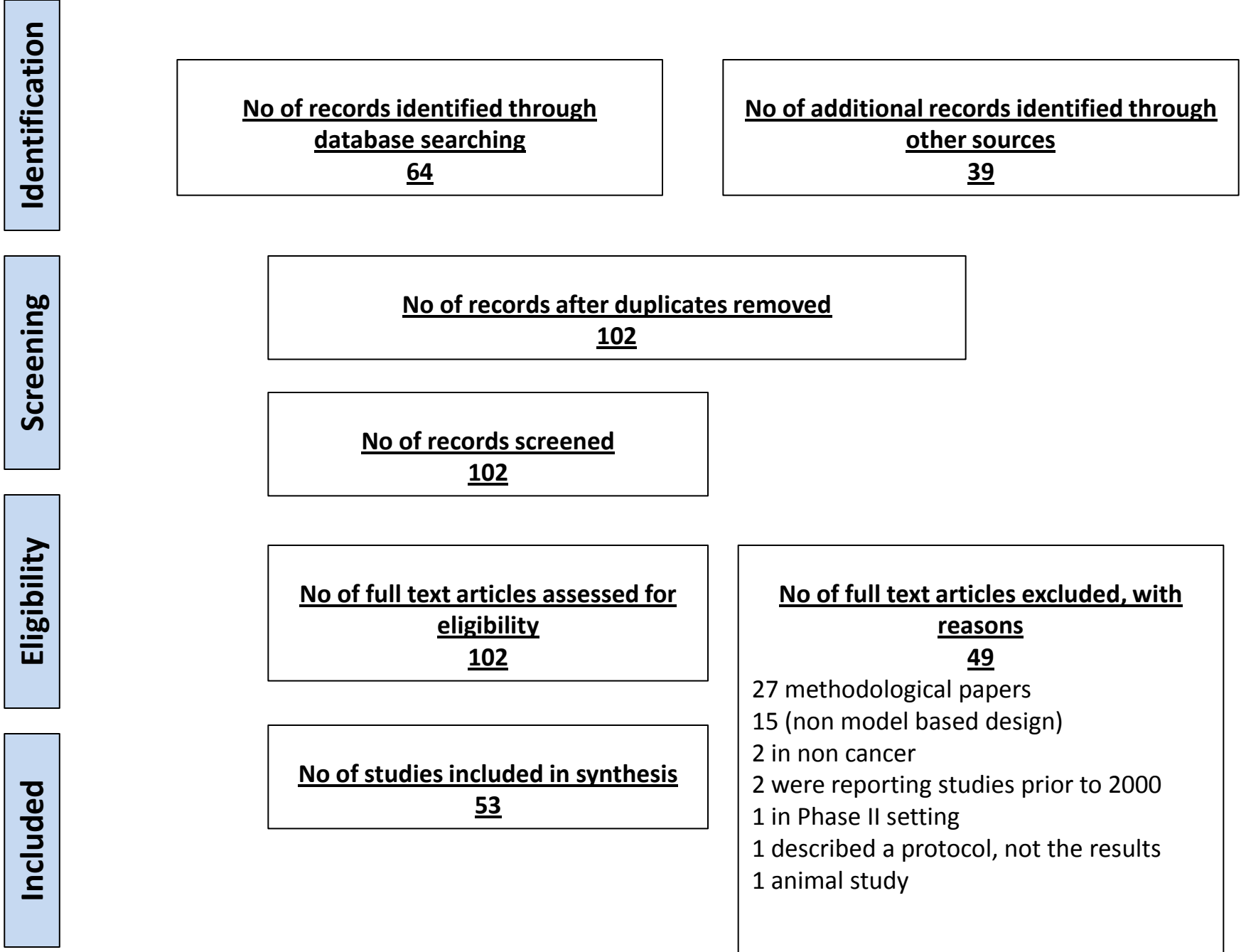
- **Coherence of design**. Given that the current level at which the patient is treated has been recommended by the model, then a toxicity should result in recommending either the same level or a lower one and a non-toxicity should result in a recommendation for the same level or a higher one (Cheung 2005).

- **Rigidity.** Informally, the accumulating information should guide us in the right direction. In particular, we should be able to move away from a level when the data do not support that level being the MTD. (Cheung 2011).

- **Efficiency.** Any design should make efficient use of the data, in particular performance should be satisfactory when contrasted with the non-parametric optimal design (O'Quigley J, Paoletti X, Maccario J. 2002)

Systematic Review
Iasonos and O'Quigley, JCO 2014

- 53 trials (Jan 2003 to Sept 2013)
- Quantitative Review:
  - safety, patients to dose allocation
- Qualitative Review:
  - are they flexible
  - what is the clinical question / objective
  - how do they deal with different schedules,
  - patient populations, drug combinations


Identification
Screening
Eligibility
Included
No of records identified through database searching
64
No of additional records identified through other sources
39
No of records after duplicates removed
102
No of records screened
102
No of full text articles assessed for eligibility
102
No of full text articles excluded, with reasons
49
27 methodological papers
15 (non model based design)
2 in non cancer
2 were reporting studies prior to 2000
1 in Phase II setting
1 described a protocol, not the results
1 animal study
No of studies included in synthesis
53

Review of the literature:
53 trials (Jan 2003 to Sept 2013)

Iasonos and O'Quigley, JCO 2014

- 54% single agent regimen
- 46% combination regimen
- Enrolled 35 patients,
- Evaluable for DLT 25 patients
- 25 months, tested 5 dose levels
- Acceptable toxicity rate of 26% (range 10-33%)
- DLT timeframe 38 days (median= 28 days)

Safety – based on real trials

- 18% DLT rate (target varied from 10-33%)
- 75% of patients treated within MTD +/- level

Treated below	Treated at MTD	Treated above
41%	39%	19% (4.7 pts)

Iasonos and O'Quigley JCO 2014

Qualitative Review

Supplemental material:

Supplemental Table 1: Qualitative review of trials: patient population, DLT definition, trial design and model parameters

Study: Trial Author, Year Aim Number of groups/schedules	Single agent (S) or combination regimen (C)	DLT definition (Endpoint); DLT timeframe N: number of enrolled patients (evaluable) Levels: number of dose levels	Type of Design Model Parameters	Comment
Thornton KA, 2013[28] **Aim:** MTD of Temsirolimus and liposomal doxorubicin for patients with soft tissue sarcoma **No of groups/schedules:** 2 groups - children and adults	C	**DLT:** Fatal toxicity gets score 1; reversible Gr 4 toxicity as 0.5; reversible Gr 3 toxicity as 0.25 **DLT timeframe:** 56 days N=15 **Levels:** 5	CRM with graded toxicities [1, 6] Rate=20% Cohort size: 3 Stopping rule: 6 patients at current level and no change in MTD Starting level 3	
Harvey RD, 2013[29] **Aim:** MTD of Bortezomib (B) and sunitinib (Sun) in patients with sold tumors **No of groups/schedules:** 1	C	**DLT:** Gr 4 neutropenia, anemia or thrombocytopenia, Gr 4 fatigue or 2 point decline in ECOG, Gr ≥3 gastrointestinal, any Gr ≥3 AE **DLT timeframe:** 42 days (6 weeks) N=31 (30) **Levels:** 7	CRM –EWOC (2 stage) [71] Alpha parameter varies Rate=33% Cohort size: 1 Escalation restricted: EWOC	The study escalated each drug. First it found the MTD of Sun with fixed B and then B was increased with fixed Sun. The toxicity ordering is not clear thus patients treated above MTD cannot be reported.
Ben-Josef E, 2012[30] **Aim:** MTD of Radiation + fixed dose gemcitabine in patients with pancreatic cancer **No of groups/schedules:** 1	C	**DLT:** Gastrointestinal toxicity Gr ≥3, neutropenic fever, deterioration in performance status to ≥3 **DLT timeframe:** Day 1-126 (13 weeks) N=51 **Levels:** 6	TITE-CRM (2 p logistic) [72] Rate=25%; Cohort size: 1 Escalation restricted: 1 level	

Iasonos and O'Quigley, JCO 2014

Design Type – modifications of CRM

TYPE OF DESIGN	N=53
CRM (O'Quigley 1990)	23/53
TITE CRM (Cheung 2002)	8/53
CRM with continuous dosing (Piantadosi 1998)	9/53
EWOC (Babb, Rogatko et al, 1999)	12/53
Lower grades (Goodman 1995)	1/53

- Accelerated 1st stage (doubling dose 1 pt per cohort; gr 2 AE)
- TITE CRM: deals with late on set toxicities
  - radiotherapy, targeted agents with late on-set toxicities
- EWOC: escalation with overdose control
  - Chu P, Lin Y, Shih WJ 2009

Escalation with overdose control

- CRM and EWOC are equivalent (25% vs 50%) under certain parameterization (bound); Chu et al. 2009
- EWOC can be too conservative or too aggressive depending on feasibility bound
- Increasing a feasibility bound regardless of DLT responses can lead to incoherent dose-escalations (Wheeler G et al. SIM 2017)
- Controlling the risk of overdosing can be achieved without the addition of extra parameters (Chu et al. 2009)
- Using the point estimate vs the probability that the estimated risk lies within an interval

TITE CRM

- a DLT within the interval counts as a DLT
- a non-DLT if, at interval completion, it is still a non-DLT.

Difference:

- CRM ignores patients without a DLT and for whom follow-up is less than the entire interval, whereas TITE-CRM counts it toward a non-DLT at that level.

TITE

- TITE can be problematic when there is fast accrual and long DLT observation window (Muler, JCO 2004)
- TITE: non DLTs are counted as non DLTs but are down-weighted

Iasonos, Wages, et al 2016 Stats Med

- Prior can be informative
- Over parameterization
- 2 parameter logistic model is not more flexible
  - non identifiable
  - performs worse even when the data are generated by 2 parm model
  - can get stuck
- One source of information

The two parameter logistic model is inconsistent

- Theorem 3 in Iasonos et al SIM 2016 (Dimension of model parameter space)

- Under adaptive sequential sampling, the two parameter logistic model is inconsistent. The determinant of the Fisher Information, instead of increasing without bound as is required for consistency, will converge almost surely to zero.

- By not consistent we mean that the parameter estimates fail to converge almost surely to their population counterparts. This is true whether or not the model is misspecified. It may be true that we have almost sure convergence to the MTD or possibly the weaker result, convergence in probability to the MTD (where we would not settle at the MTD but the probability of sampling there would go to one) but it is not at all clear how either of these could be shown.

Rigidity example 1

- 2 parameter estimation may confine trial to suboptimal doses indefinitely
- NeuSTART (Elkind et al., 2008)
  - 5 dose levels
  - Objective: identify dose with DLT rate closest to 10%
  - 0/3 DLTs on dose level 1
  - 1/3 DLTs on dose level 2
  - Cheung (2011) proves that the 2 parameter model will always recommend dose level 1 form this point on, no matter what outcomes are observed at dose level 1

Rigidity example 2

- Doughtery et al., 2000
  - 4 dose levels
  - Objective: identify dose with DLT rate closest to 20%
  - 0/2 DLTs on dose level 1
  - 1/3 DLTs on dose level 2
  - 1/1 DLTs on dose level 3
  - The 2 parameter model will always recommend dose level 2 form this point on, no matter what outcomes are observed at dose level 2

Evaluating performance

- Generate operating characteristics (OC) via simulation studies
- Accuracy
  - % that correctly identify true MTD
  - Accuracy index (Cheung, 2011)
- 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
- It can sometimes be difficult to compare the relative performance of competing methods

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 method's operating characteristics
  - Does not account for patient allocation
    - Can not be used to evaluate safety
  - 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

Wages and Varhegyi (*Clinical Trials*, 2017)

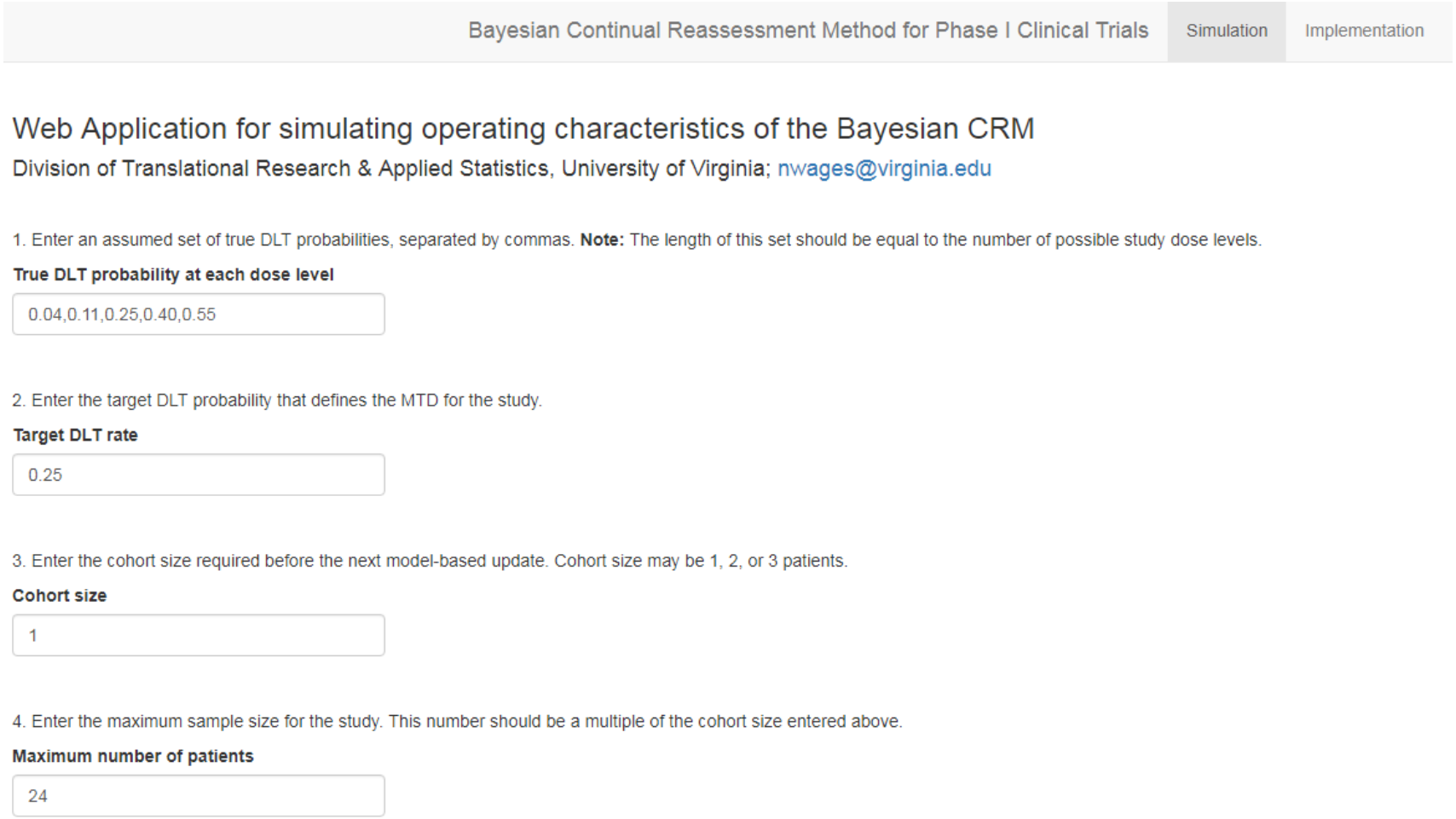
- Available R Shiny web application
  - https://uvatrapps.shinyapps.io/nonparbnch/
- User input
  - Assume DLT probability at each dose
  - Target DLT rate
  - Sample size
  - Number of simulated trials
- Output
  - MTD selection percentage for each dose

CRM Software

- https://cran.r-project.org/web/packages/**CRM**/index.html
- https://cran.r-project.org/web/packages/**dfcrm**/index.html
- R Shiny app
  - https://uvatrapps.shinyapps.io/crmb

Bayesian CRM Web Tool

https://uvatrapps.shinyapps.io/crmb



Simulation Studies

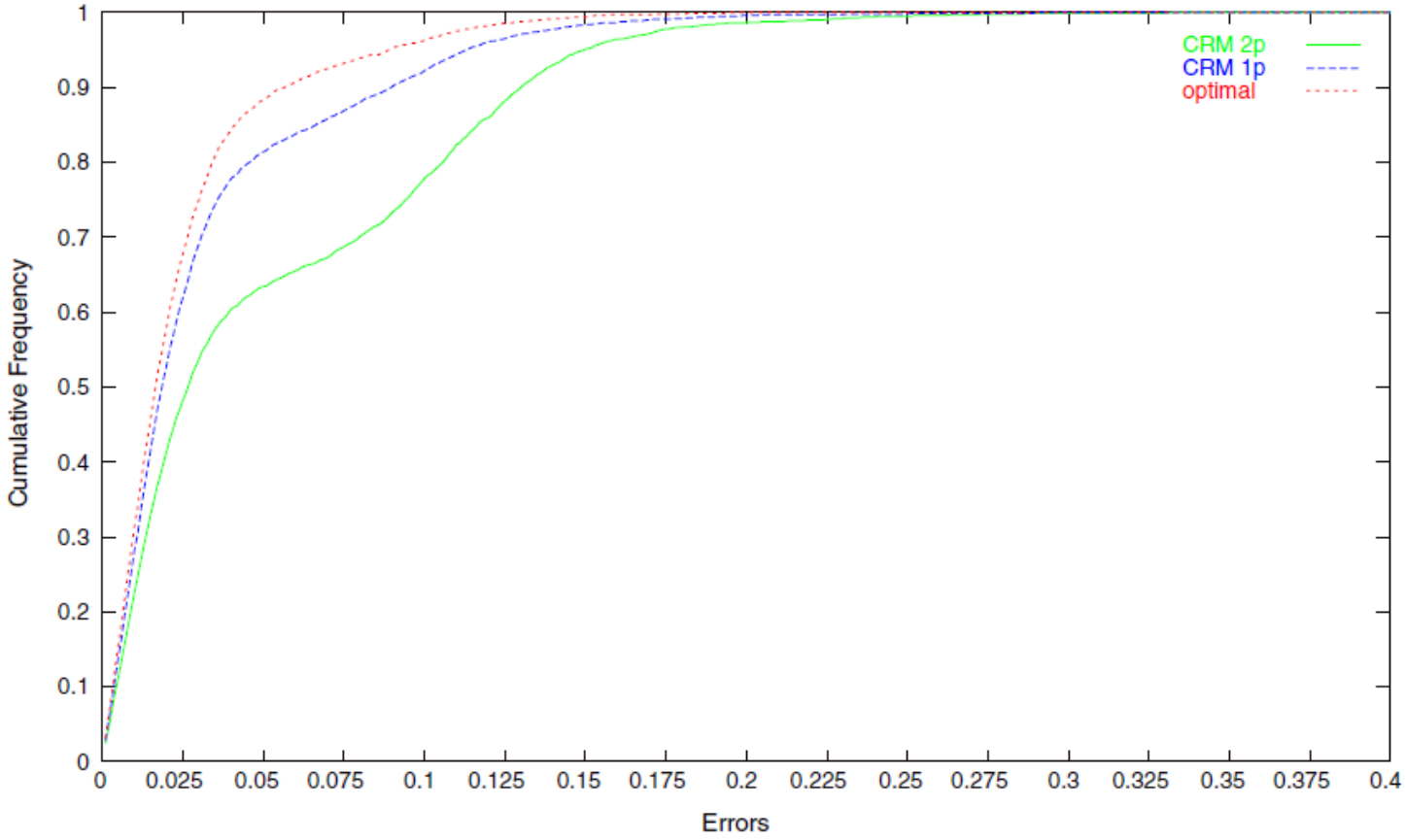


Figure 1. Cumulative distribution of errors. Optimal design, one-parameter continual reassessment method (CRM) and two-parameter CRM as shown in O'Quigley, Paoletti and Maccario [5].

Accuracy Index

Cheung (2011)

- Measures accuracy of entire distribution of doses selected as MTD
- Max value is 1 with larger values indicating better accuracy
- $A_n = 1 - k \times \frac{\sum_{i=1}^{k} |R(d_i) - \theta| \times \text{Probability of selecting dose } i}{\sum_{i=1}^{k} |R(d_i) - \theta|}$

1-Param CRM vs Optimal

Wages, Conaway, O'Quigley (2013)

- 18 total scenarios with various target DLT rates (20%, 25%, 30%)
  - Various sample sizes (n=20, 25, 30)
- 10,000 simulated trials in each scenario
- Average accuracy
  - CRM: 0.595
  - Optimal: 0.655
- A measure of efficiency
  - 0.595/0.655 = 91%

Simulation Studies

Table V. Performance two-parameter logistic versus one-parameter power model when the true underlying model is the two-parameter logistic; entries represent proportion of trials recommending each dose (as shown in O'Quigley, Pepe, and Fisher [1]).

Dose	1	2	3	4	5	6
$R(d_i)$	0.06	0.08	0.12	0.18	0.40	0.71
1-Parm CRM	0.00	0.04	0.23	0.57	0.15	0.00
2-Parm CRM	0.01	0.11	0.16	0.48	0.19	0.05

$R(d_i)$ denotes the true DLT rates at each dose level; the target was 0.2, and true MTD was level 4.

DLT, dose-limiting toxicities; MTD, maximum tolerated dose; Parm, parameter; CRM, continual reassessment method.

Simulation Setting

- k = 4 dose levels
- n = 25 patients
- 200 simulated trials
- Target DLT rate = 20%
- 1p vs 2p CRM
- Skeleton for 1p CRM (0.06, 0.11, 0.20, 0.30)
- 2p CRM used pseudo-data similar to Whitehead et al. (2010)

A Note on Skeleton Choice in 1p CRM

- Algorithms available for choosing the skeleton in order to yield robust operating characteristics for CRM
  - Reasonable skeleton is defined by adequate spacing between adjacent values
- Use **getprior** function in **R** package **dfcrm** with recommended specifications and algorithm of Lee and Cheung (2009)

Simulation Studies

Table III. Percent of trials that selected each dose for five scenarios and accuracy index (γ).

Scenario 1	0.10	0.20	0.30	0.45	γ
1-Parm	0.24	0.51	0.23	0.02	0.53
2-Parm	0.25	0.49	0.21	0.06	0.47
optimal	0.20	0.52	0.26	0.02	0.54
Scenario 2	0.02	0.11	0.20	0.35	
1-Parm	0.01	0.29	0.52	0.17	0.48
2-Parm	0.03	0.35	0.51	0.11	0.49
optimal	0	0.23	0.56	0.2	0.51
Scenario 3	0.01	0.07	0.10	0.20	
1-Parm	0	0.05	0.28	0.67	0.68
2-Parm	0.04	0.09	0.41	0.46	0.42
optimal	0	0.05	0.18	0.76	0.76
Scenario 4	0.20	0.29	0.36	0.44	
1-Parm	0.65	0.27	0.07	0.01	0.70
2-Parm	0.55	0.26	0.08	0.11	0.48
optimal	0.66	0.24	0.09	0.02	0.68
Scenario 5	0.01	0.07	0.20	0.40	
1-Parm	0	0.23	0.64	0.13	0.57
2-Parm	0.01	0.28	0.64	0.08	0.59
optimal	0	0.13	0.75	0.12	0.68

The first row in each scenario shows the true DLT rates. The true MTD is the dose closest to $\theta = 0.20$. DLT, dose-limiting toxicities; MTD, maximum tolerated dose; Parm, parameter.

How can we overcome the clinical challenges?

- Toxicity ordering no longer holds
- Multiple agents /doses / schedules
- Efficacy (+ safety simultaneously)
- Patient populations – heterogeneity
- Disease vs drug related toxicities; attribution
- Late onset AE (fractional DLTs)

A model based design can accommodate these multiple components

Drug combinations

- Partial order CRM (POCRM; Wages et al., 2011) uses a class of one-parameter models to estimate combination probabilities
  - OC for POCRM are generally better than methods that rely on models that attempt to model all marginal and secondary effects
  - Class of 1 parameter model is sufficient for objective of locating and experimenting at and around MTD combination
- Richer models require information at more combinations

Bivariate outcomes

- Binary toxicity and efficacy
- Some methods attempt to model the correlation between the two outcomes
- Simulation results for methods that avoid modeling the association are generally preferred (Cunanan and Koopmeiners, 2014)
  - Results of Eff-Tox (Thall and Cook, 2004) indicate that the percentage of selecting a specific dose decreases from 82% to 14% on the basis of differences in the value of the association parameter

Conclusions

Iasonos and O'Quigley JCO 2014

- Review of case studies confirmed results of simulated trials reported in the statistical literature

- Method is rigid once parameters are selected but it is flexible to deal with clinical problems through the choice of tuning parameters

Recommendations

- Calibrate the prior of the Bayesian CRM to ensure experimentation according to the model
  - Use non-informative prior
  - Lee and Cheung (2011) algorithm
- Historical data can be used as a prior
  - Weight assessed via simulations
  - Design robustness
- Instead of Bayesian, use two-stage likelihood based CRM design
  - Non-model based first stage until at least one DLT and one non-DLT

Questions

iasonosa@mskcc.org
nwages@virginia.edu

Thank you to Dooti Roy

Acknowledgments to Collaborators
John O'Quigley and Mark Conaway
Clinical Investigators at MSKCC and UVA

References

- Cheung YK (2011). *Dose-finding by the continual reassessment method.* Chapman and Hall/CRC press: New York.
- Cunanan K, Koopmeiners J. (2014). Evaluating the performance of copula models in Phase I-II clinical trials under model misspeification, *BMC Research Methodology*, 14: 51.
- **Iasonos A**, O'Quigley J. Integrating the escalation and dose expansion studies into a unified Phase I clinical trial. Contemp Clin Trials. 2016 Sep;50:124-34.
- **Iasonos A**, Wages NA, Conaway MR, Cheung K, Yuan Y, O'Quigley J. Dimension of model parameter space and operating characteristics in adaptive dose-finding studies. Stat Med. 2016 Sep 20;35(21):3760-75.
- **Iasonos A** and O'Quigley J. Sequential Monitoring of Phase I Dose expansion cohorts. Stat Med. 2017 Jan 30;36(2):204-214
- **Iasonos A**, O'Quigley J. Phase I Designs that Allow for Uncertainty in the Attribution of Adverse Events. Journal of the Royal Statistical Society: Series C (Applied Statistics). 2017. doi: 10.1111/rssc.12195 [Epub ahead of print].
- Lee and Cheung (2009). Model calibration in the continual reassessment method, *Clinical Trials;* **6** (3): 227-238.
- Lee and Cheung (2011). Calibration of prior variance in the Bayesian continual reassessment method, *Statistics in Medicine;* 30: 2081-2089.

References

- O'Quigley J, Paoletti X, Maccario J. (2002). Non-parametric optimal design in dose-finding studies. Biostatistics; 3: 51-56.
- O'Quigley J, Paoletti X, Maccario J. (2002). Non-parametric optimal design in dose-finding studies, *Biostatistics;* **3** (1): 51-56.
- O'Quigley J, **Iasonos A**, Bornkamp B. Handbook of methods for designing, monitoring and analyzing dose finding trials. Chapman and Hall/CRC. Taylor and Francis Group. 2017.
- Petroni GR, **Wages NA**, Paux G, Dubois F. (2017). Implementation of adaptive methods in early-phase clinical trials. *Statistics in Medicine* 36: 215-224.
- Thall PF, Cook J. (2004). Dose-finding based on efficacy-toxicity trade-offs. *Biometrics*; 60(3): 684-693.
- **Wages NA**, Conaway MR, O'Quigley J. (2013). Performance of two-stage continual reassessment method relative to an optimal benchmark, *Clin Trials*; 10: 862-75.
- **Wages NA**, Varhegyi N. (2017). A web application for evaluating phase I methods using a non-parametric optimal benchmark. *Clin Trials*; [epub ahead of print].
- Whitehead J, Thygesen H, Whitehead A. (2010). A Bayesian dose-finding procedure for phase I clinical trials based only on the assumption of monotonicity, *Stat Med*, 29: 1808-24.

Extra slides

3 Trials

- imatinib and docetaxel in prostate cancer patients, where 8 out of 10 patients experienced a DLT above the MTD. [Mathew, P., et al., *Platelet-derived growth factor receptor inhibitor imatinib mesylate and docetaxel: a modular phase I trial in androgen-independent prostate cancer.* J Clin Oncol, 2004. **22**(16): p. 3323-9.]
- dose escalation study of cisplatin with gemcitabine in pancreatic cancer, 50% (4/8) of patients treated above the MTD experienced DLTs. [Muler, J.H., 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.* J Clin Oncol, 2004. **22**(2): p. 238-43.]
- A CRM trial where the recommendation was to escalate after observing 2 DLTs out of 2 patients treated at a level [Neuenschwander, B., M. Branson, and T. Gsponer, *Critical aspects of the Bayesian approach to phase I cancer trials.* Stat Med, 2008. **27**(13): p. 2420-39.]

TITE CRM Likelihood

$$\prod_j \phi(x_j, a)^{T_j} (1 - w_j \phi(x_j, a))^{1-T_j}$$

$$w_j = w(C_j; \tau)$$

$$w(0; \tau) = 0, \quad w(C, \tau) = 1, \quad C \geq \tau$$

- The weight is monotone increasing in C_j patient's FU time
- TITE CRM reduces to CRM if all patients are followed up for the entire period τ

TRIAL 3

Table I. Posterior summaries for probabilities of DLT (CRM).

	Doses									
	1	2.5	5	10	15	20	25	30	40	50
No. of patients	3	4	5	4	—	—	2	—	—	—
No. of DLTs	0	0	0	0	—	—	2	—	—	—
(*A*) *Posterior summaries* (*original skeleton*)										
Skeleton (CRM)	0.010	0.015	0.020	0.025	0.030	0.040	0.050	0.100	0.170	0.300
Mean	0.069	0.085	0.099	0.111	0.123	0.144	0.163	0.242	**0.330**	0.465
Std. dev.	0.055	0.062	0.068	0.072	0.076	0.082	0.087	0.101	0.109	0.108
(*B*) *Posterior summaries* (*equidistant skeleton*)										
Skeleton (CRM)	0.063	0.125	0.188	0.250	0.313	0.375	0.438	0.500	0.563	0.625
Mean	0.024	0.054	0.090	0.130	0.176	0.226	**0.281**	0.341	0.405	0.475
Std. dev.	0.030	0.051	0.069	0.084	0.097	0.107	0.115	0.119	0.120	0.117

Columns in boldface highlight the recommended dose for the next cohort.

Supplemental Table 3: Case Study, Trial 3 as described in Supplemental Appendix A.1

Patient	Dose Treated (mg) (Level)	DLT	Actual Dose recommendation (mg) (Level) Initial curve	Updated DLT rate at dose 50mg	Correct Dose recommendation (mg) (Level) Correct curve	DLT
Prior/ Initial			50 (L10)	30	**1 (L1)**	
1	1 (L1)	No	50 (L10)	21	**1 (L1)**	No
2	1 (L1)	No	50 (L10)	17	**5 (L3)**	No
3	1 (L1)	No	50 (L10)	15	**15 (L5)**	No
4	2.5 (L2)	No	50 (L10)	13	**20 (L6)**	No
5	2.5 (L2)	No	50 (L10)	11	**25 (L7)**	Yes
6	2.5 (L2)	No	50 (L10)	10	**15 (L5)**	No
7	2.5 (L2)	No	50 (L10)	9	**20 (L6)**	No
8	5 (L3)	No	50 (L10)	8	**20 (L6)**	No
9	5 (3L)	No	50 (L10)	8	**25 (L7)**	Yes
10	5 (L3)	No	50 (L10)	7	**20 (L6)**	No
11	5 (L3)	No	50 (L10)	7	**20 (L6)**	No
12	5 (L3)	No	50 (L10)	6	**25 (L7)**	Yes
13	10 (L4)	No	50 (L10)	6	**20 (L6)**	No
14	10 (L4)	No	50 (L10)	6	**20 (L6)**	No
15	10 (L4)	No	50 (L10)	6	**20 (L6)**	No
16	10 (L4)	No	50 (L10)	5	**25 (L7)**	Yes
17	25 (L7)	Yes	40 (L9)	38	**20 (L6)**	No
18	25 (L7)	Yes	40 (L9)	47	**20 (L6)**	No
Recommended dose			**40 (L9)**		**25 (L7)**	

Footnote: Initial curve assigns 30% DLT rate at level 10 (50mg) and very low rates at all remaining levels (rates for each respective dose level: 0.01,0.015,0.02,0.025,0.03,0.04,0.05,0.1,0.17,0.3);

Correct curve assigns 30% initial DLT rate (prior to seeing the data) at dose 1 so that experimentation starts at dose 1. Rates under correct curve for each dose level: 0.30, 0.40, 0.48, 0.56, 0.64, 0.72, 0.80, 0.88, 0.92, 0.99,)

Iasonos, O'Quigley, JCO 2014

Intervals based methods

- 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)
- Horton et al. (2017) demonstrated that CRM had superior performance to these interval based methods across a wide range of scenarios.