Cytel
Shaping the Future of Drug Development

Bayesian Model-Based Dose-Finding Designs until they are not Bayesian or even Model-Based

Yuan Ji, PhD, Professor of Biostatistics, University of Chicago

Innovations in Bayesian Trials – Register today!

Symposium

Thursday, November 19, 10AM EDT
Statistical Design and Conduct of Platform Trials
Jason Connor

Thursday, December 3, 10AM EDT
Bayesian Models for Precision Oncology Clinical Trials
Peter Mueller

Thursday, December 10, 11AM EDT
Recent Development on Bayesian Clinical Trial Designs Using Historical Data
Ming-Hui Chen

Interactive Workshop

Thursday, October 29, 9AM EDT
Interactive Workshop 2: Bayesian Meta-analysis and Hierarchical Models

Thursday, November 12, 9AM EDT
Interactive Workshop 3: Prior distributions (Part I)

Thursday, January 21, 9AM EDT
Interactive Workshop 4: Prior distributions (Part II)

Thursday, February 4, 9AM EDT
Interactive Workshop 5: Survival analysis (Part I)

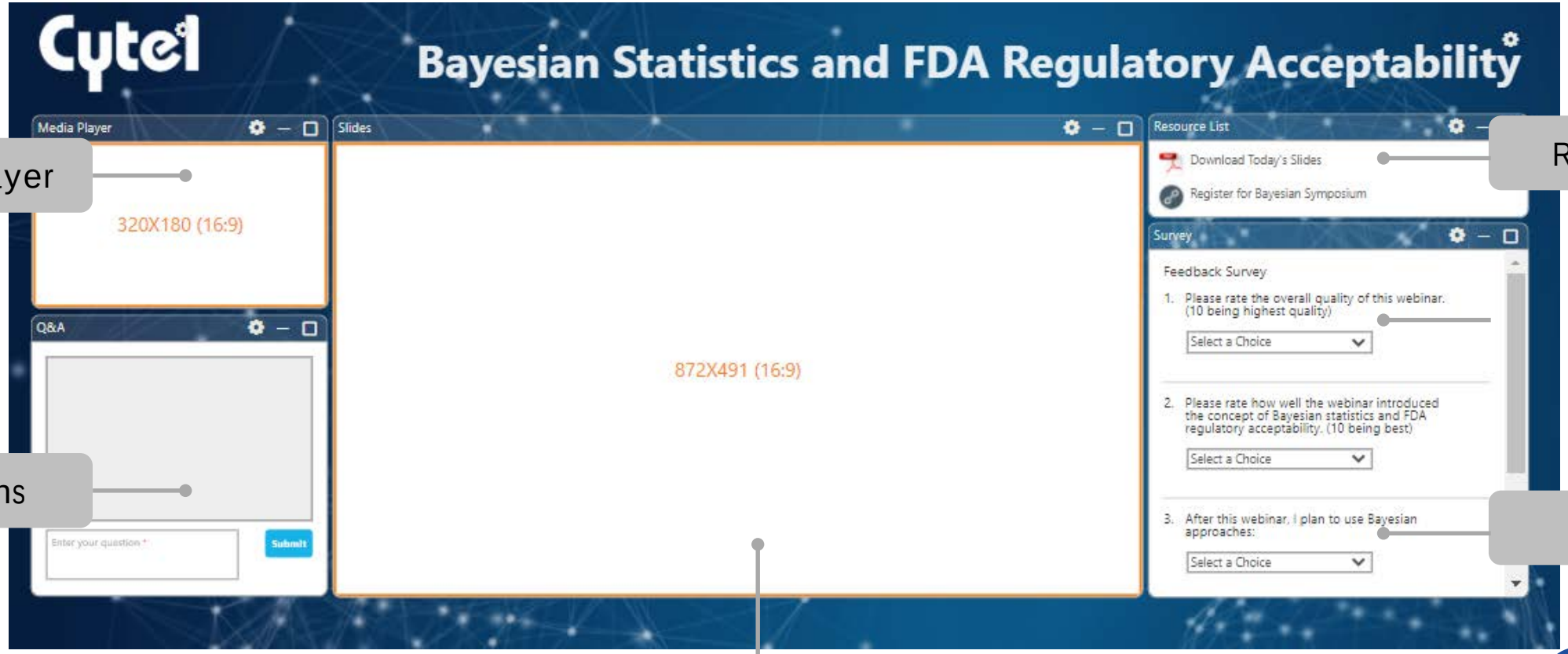
Thursday, February 18, 9AM EDT
Interactive Workshop 6: Survival analysis (Part II)

Thursday, March 4, 9AM EDT
Interactive Workshop 7: Other topics & Wrap-up

Cytel

Console Overview

***Configurable windows – expand or contract as needed**



Media Player

Questions

Slides

Resources

Survey

Cytel

Bayesian model-based dose-finding designs until they are not Bayesian or even model based

Yuan Ji, PhD

Professor of Biostatistics

The University of Chicago

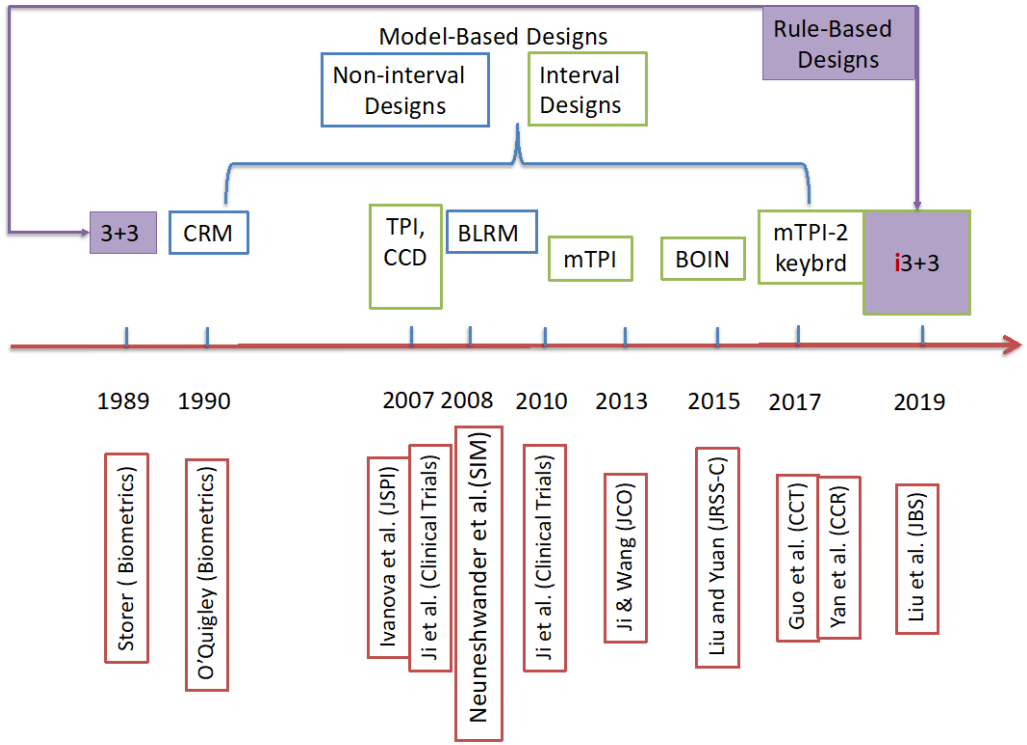
October 22, 2020

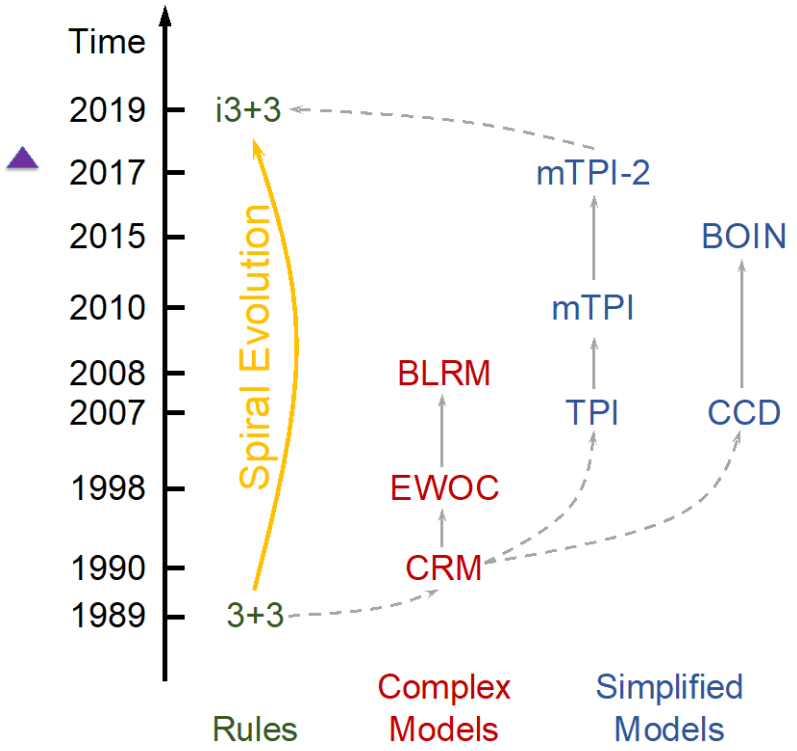
Conflict of Interests

- U-Design is a commercial platform developed by Laiya Consulting, Inc., now part of Cytel Inc.
- Yuan Ji is the co-founder of Bayesoft
  - Also, an IDMC member for an Astellas trial
  - Consultant for Cytel
  - Research contracts with Abbvie

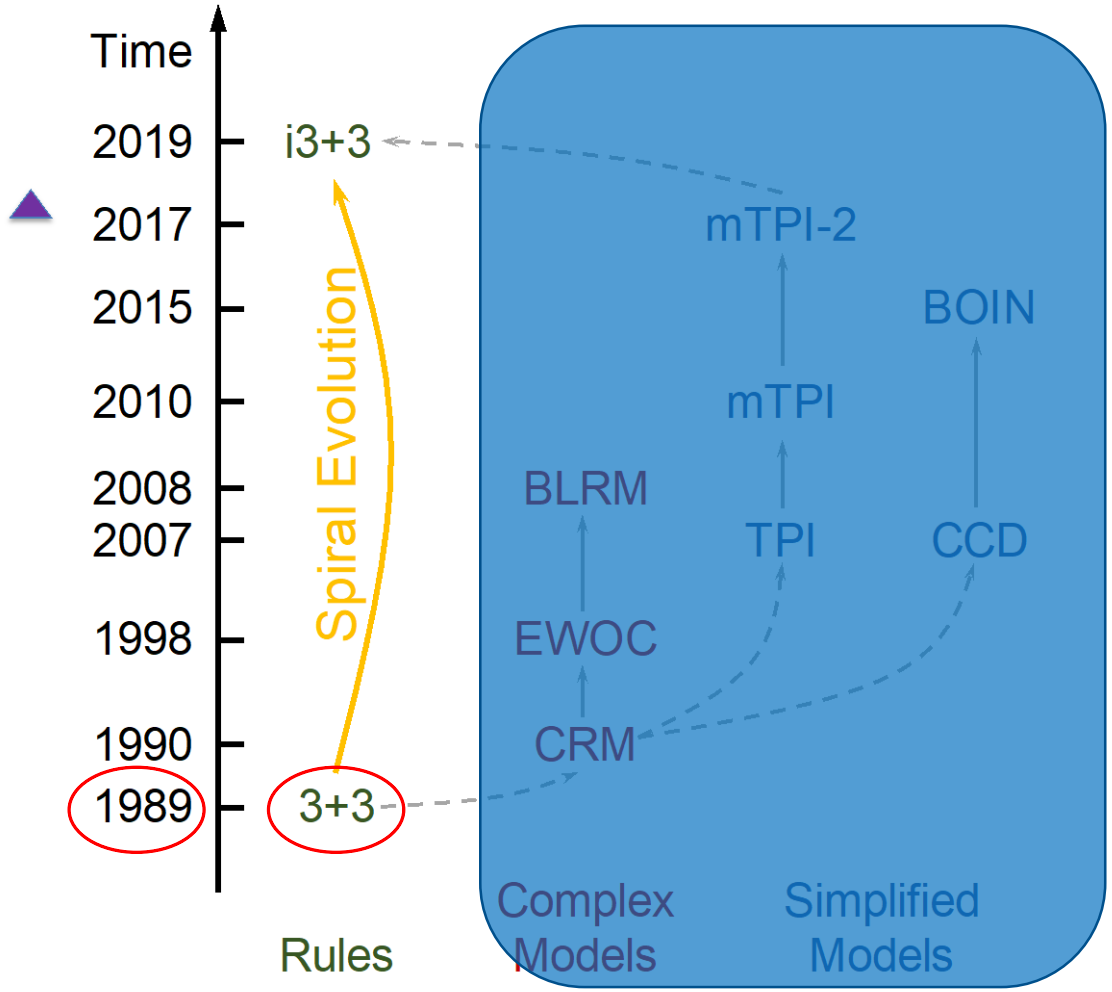
Dose-finding designs over last 30 years

- So many designs are available now. Which one to use?

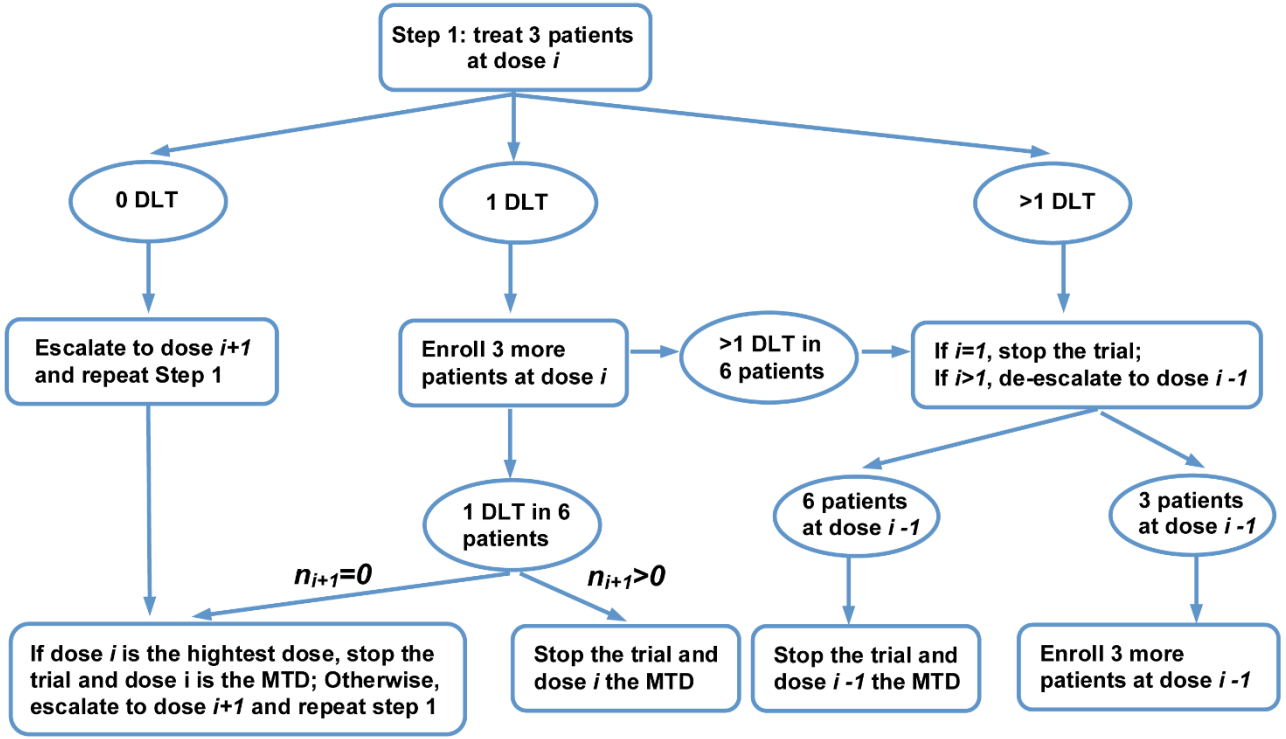




The 3+3 design (1989)



The 3+3 design (1989)



- Rule based
- No statistical models
  - Easy
  - Transparent
  - Societal acceptance
- Naïve/Rigid
  - <= 6 patients per dose
  - MTD wide range (1/6-1/3)
  - Performance depends on the # of doses
  - Large variabilities in MTD identification
  - Often little data supporting RP2D choices

Numerous papers have shown 3+3 is inferior in many ways

J Clin Oncol. 2013 May 10;31(14):1785-91. doi: 10.1200/JCO.2012.45.7903. Epub 2013 Apr 8.

Modified toxicity probability interval design: a safer and more reliable method than the 3 + 3 design for practical phase I trials

Yuan Ji [1], Sue-Jane Wang

Affiliations + expand

PMID: 23569307 PMCID: PMC3641699 DOI: 10.1200/JCO.2012.45.7903

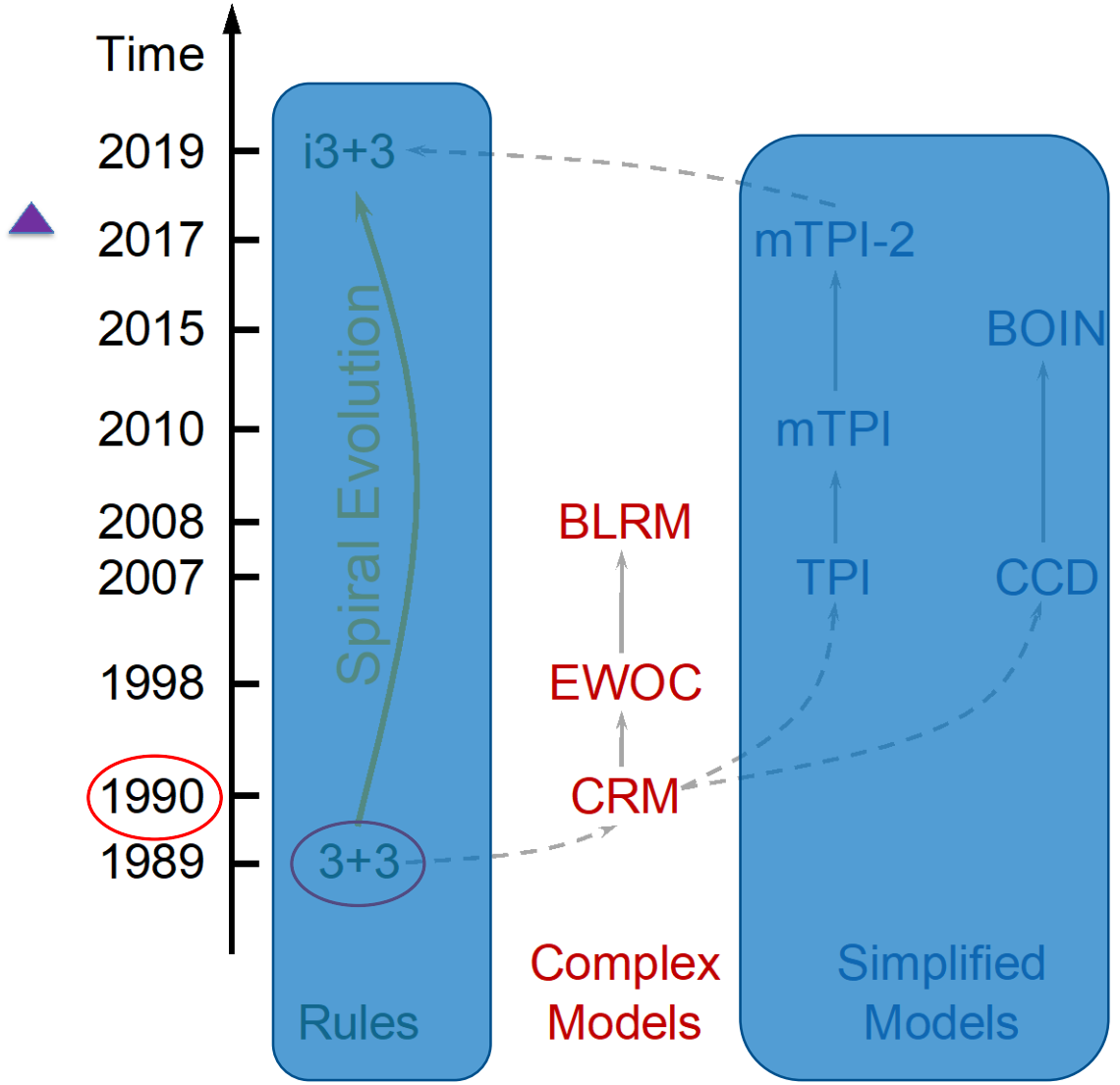
Free PMC article

CCR FOCUS 2016

Rendering the 3 + 3 Design to Rest: More Efficient Approaches to Oncology Dose-Finding Trials in the Era of Targeted Therapy

Lei Nie[1], Eric H. Rubin[2], Nitin Mehrotra[3], José Pinheiro[4], Laura L. Fernandes[1], Amit Roy[5], Stuart Bailey[6], and Dinesh P. de Alwis[7]

The CRM designs (1990-2007)

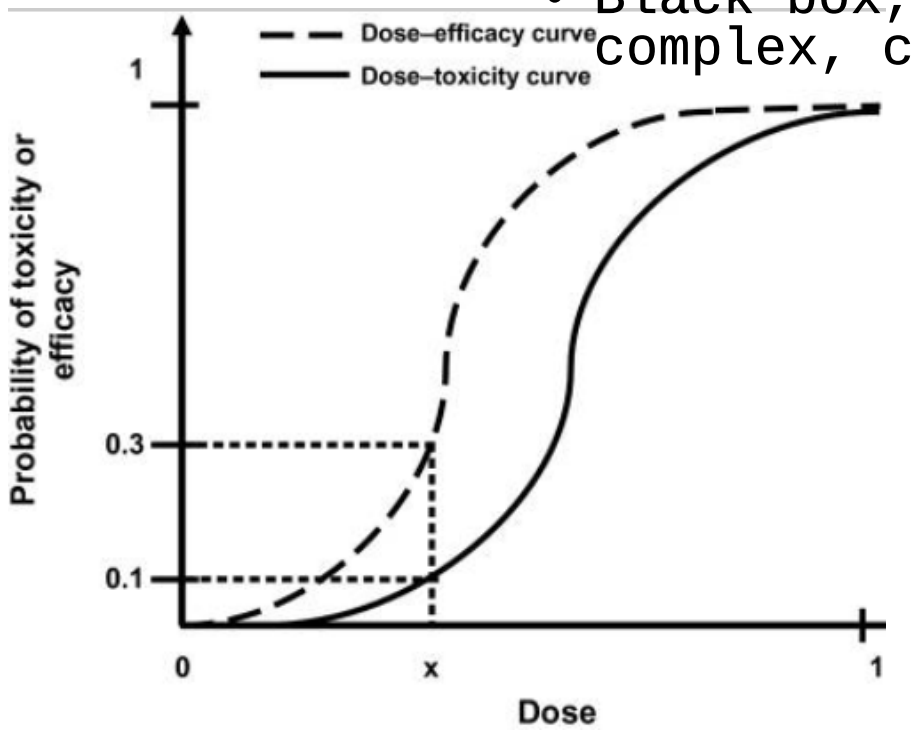


The CRM & BLRM designs

- MTD: a target rate p_T
- BLRM: probability intervals
- Dose-response curve
- $p(x) = p_0(x)^{\exp(\alpha)}$ or $\text{logit}^{(-1)}\ (x\beta)$
  - $\alpha \sim N(0, 1.34)$; or $\beta \sim prior$
  - $p_0(x)$ is the "skeleton"
  - Next dose = $\text{argmin}|\hat{p}(x) - p_T|$ or based on posterior prob. of intervals
- Operation
  - Need a statistical expert for inference and decision making
  - Too complex for the clinical team
  - SMC may override dosing decision
  - Ad-hoc rules for over-dose control

- Model based
  - Account for variability
  - Dose response curves
  - Flexible and powerful

- Lots of modifications
  - Over-dose control
  - Bayesian models
  - # of parameters
  - Black box, complex, costly



Wheeler *et al. BMC Medical Research Methodology* (2019) 19:18
https://doi.org/10.1186/s12874-018-0638-z

BMC Medical Research Methodology

TECHNICAL ADVANCE Open Access

How to design a dose-finding study using the continual reassessment method

Graham M. Wheeler[1*], Adrian P. Mander[2], Alun Bedding[3], Kristian Brock[4], Victoria Cornelius[5], Andrew P. Grieve[6], Thomas Jaki[7], Sharon B. Love[8,9], Lang'o Odondi[8], Christopher J. Weir[10], Christina Yap[4] and Simon J. Bond[2,11]

Hundreds of papers on CRM over the past 3 decades – very popular as a research topic

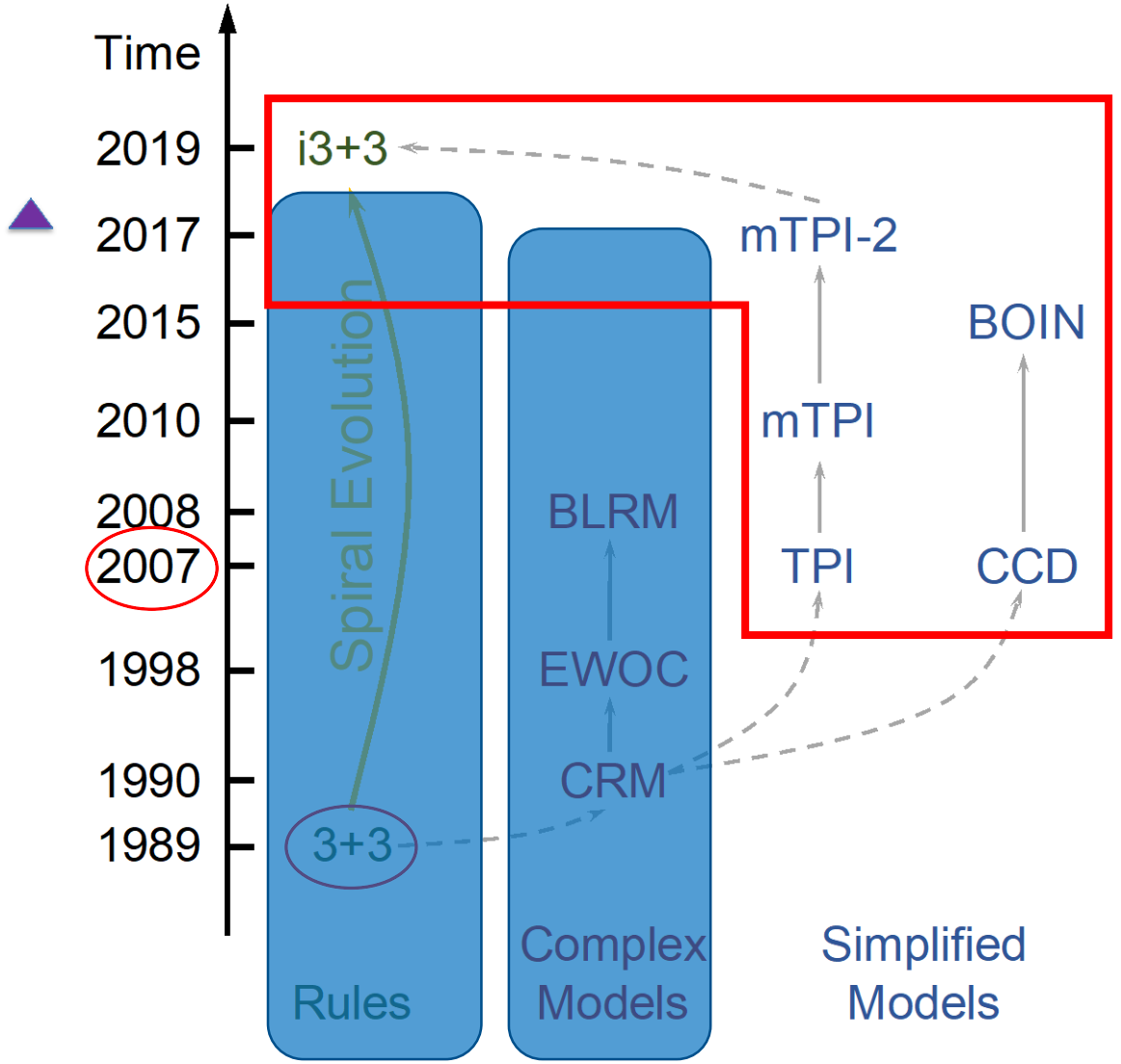
- First paper, O'Quigley, Fisher, Pepe (1990); solid statistical principle: Model-based inference; borrow information across doses
- Wheeler et al. (2019) provide a comprehensive tutorial summarizing the decades of research on CRM.
- CRM is not easy to implement – in 2019 still needing a tutorial

CRM-Software (Wheeler et al., 2019)

Name	Host/Institution	Software/Stand-alone	Free/Commercial	Rule-based/Model-based	Description
bcrm [88]	CRAN	R	Free	Both	Design, run, and simulate trials using the CRM and 3 + 3 design
dfcrm [18]	CRAN	R	Free	Model-based	Design, run, and simulate trials using the CRM and Time-to-event CRM
crmPack [89]	CRAN	R	Free	Both	Design, run, and simulate trials using the CRM (includes other model-based designs, joint toxicity-efficacy modelling)
crm [90]	IDEAS (RePEc)	Stata	Free	Model-based	Run a single trial using the CRM
MoDEsT [91]	Lancaster University	Stand-alone (online)	Free	Model-based	Design, run, and simulate trials using the CRM
Bayesian CRM for phase I trials [92]	University of Virginia	Stand-alone (online)	Free	Model-based	Design, run, and simulate trials using the CRM
AplusB [93]	MRC Biostatistics Unit, University of Cambridge	Stand-alone (online)	Free	Rule-based	Compute exact operating characteristics for 3 + 3 and other rule-based designs
Center for Quantitative Sciences Calculator [94]	Vanderbilt University	Stand-alone (online)	Free	Both	Simulate trials using the CRM (uses bcrm [88] and dfcrm [18]) and other designs (rule-based/model-based)
CRMSimulator [95]	MD Anderson Cancer Center, University of Texas	Stand-alone	Free	Model-based	Simulate trials using the CRM
FACTS [96]	Berry Consultants	Stand-alone	Commercial	Both	Design program for phase I trials using the CRM, plus fixed and adaptive designs for phase II trials
ADDPLAN [97]	ICON PLC	Stand-alone	Commercial	Both	Design, simulate, and analyse trials using the CRM (includes methods for dose-response modelling)
U-Design	Laiya Consulting	Stand-alone	Commercial	Both	Phase I, II, II/III adaptive designs, master protocols sample size cal; etc.
EAST Escalate	Cytel	Stand-alone	Commercial	Both	Phase I adaptive designs,

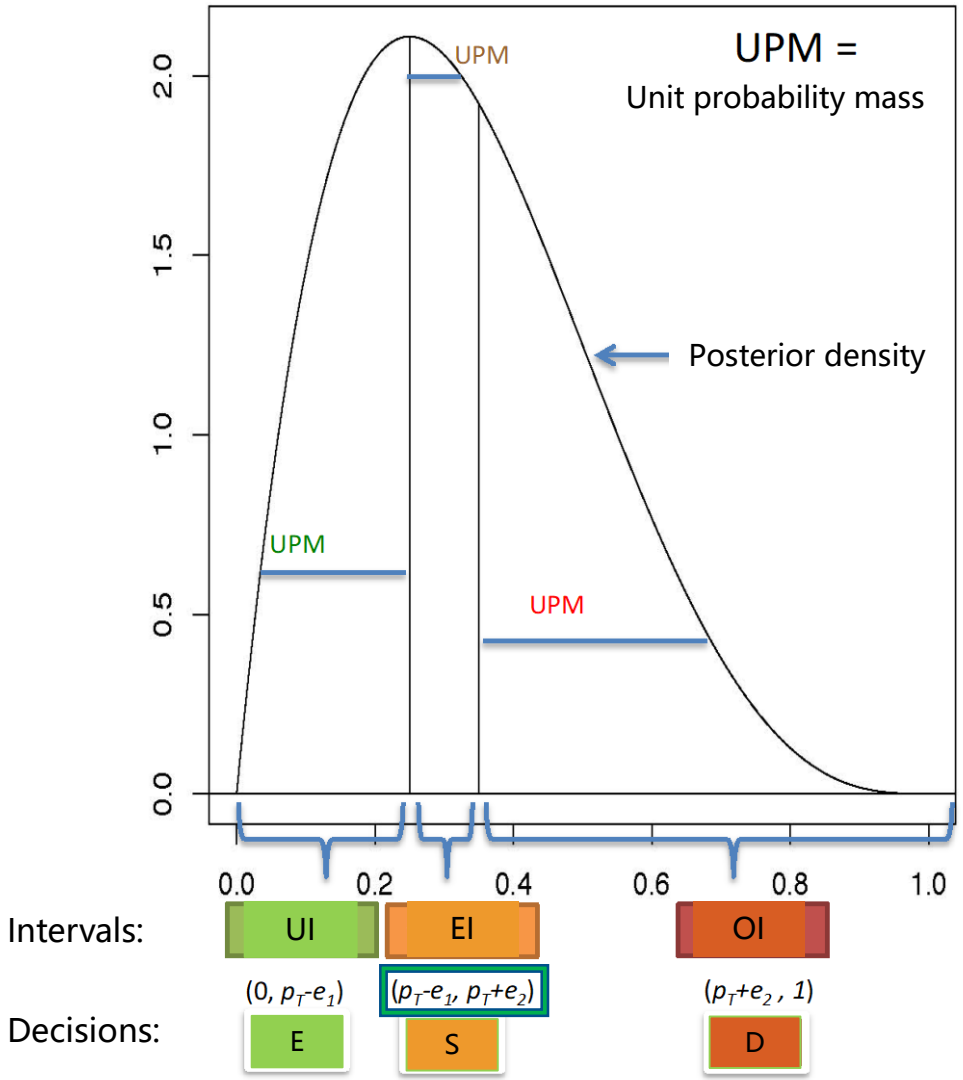
The interval-based designs (2007-now)

- Model-based designs
  - Account for variability
  - ~~Dose response curves~~
  - Flexible and powerful
- Simple & Transparent
  - Over-dose control
  - ~~Simple Bayesian~~ models
  - Decision tables



2) designs: Specify an equivalence interval

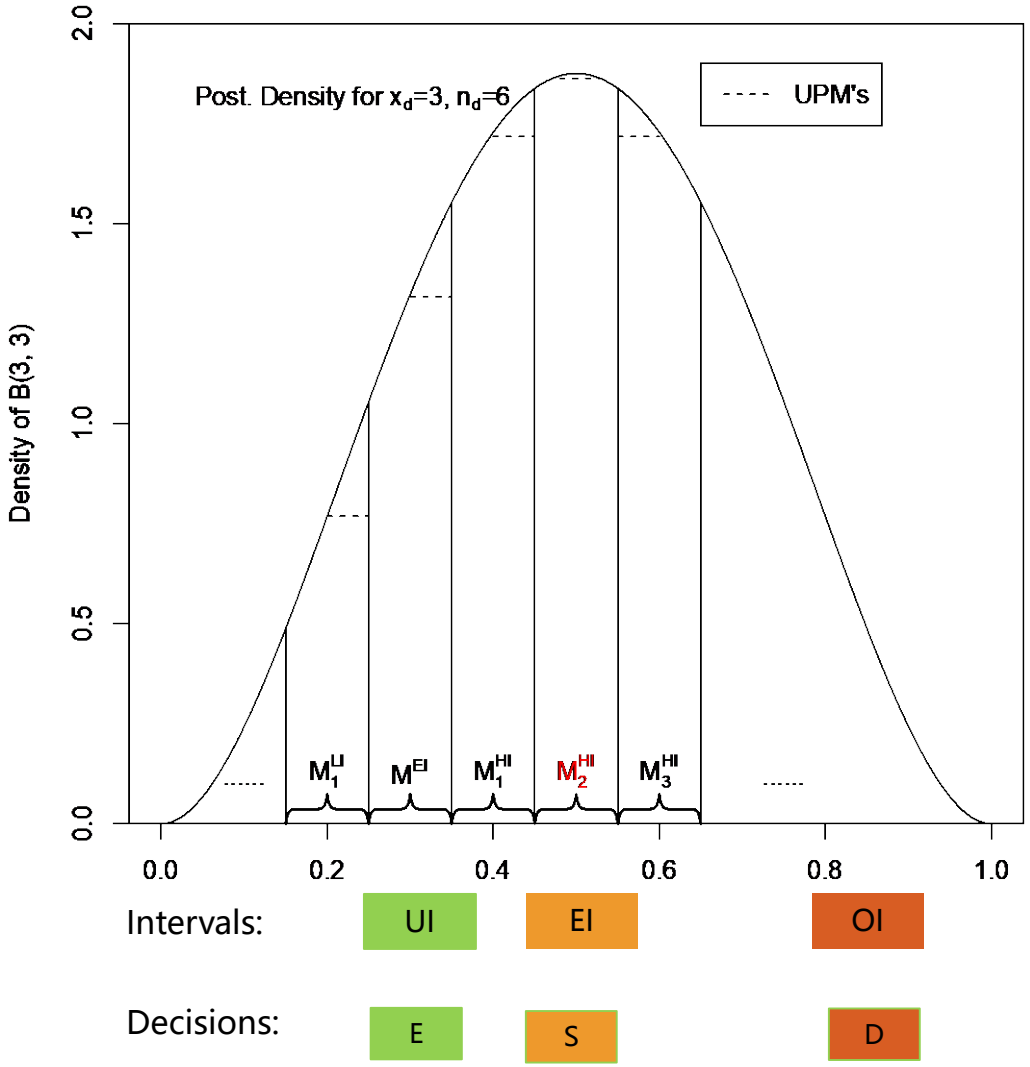
- MTD target is set at p_T, say 0.25.
- An equivalence interval $(p_T - \epsilon_1,\ p_T + \epsilon_2)$, where $(p_T - \epsilon_1)$ and $(p_T + \epsilon_2)$ are the lowest and highest toxicity rates for a dose to be considered as the MTD.
- All the decisions for dose finding can be pre-tabulated (Ji et al., 2007, 2010)
- mTPI-2 (Guo et al, 2017) and keyboard (Yan et al, 2017) are identical.



UPM = Marginal posterior probability of interval (Guo et al., 2017)

The mTPI (mTPI-2) designs: Equal-lengthed subintervals

- MTD target is set at p_T, say 0.25.
- Due to Ockham's razor (Guo et al., 2017), mTPI-2 further divides the three intervals into subintervals with equal length.
- mTPI-2 (Guo et al, 2017) and keyboard (Yan et al, 2017) are identical.



Sample size = 9 ; Target toxicity probability = 30% ; epsilon 1 = 0.05 ; epsilon 2 = 0.05 ;

Number of Patients

Number of DLTs	1	2	3	4	5	6	7	8	9
0	E	E	E	E	E	E	E	E	E
1	D	S	S	S	S	E	E	E	E
2		DU	D	S	S	S	S	S	S
3			DU	DU	D	S	S	S	S
4				DU	DU	DU	D	D	S
5					DU	DU	DU	DU	DU
6						DU	DU	DU	DU
7							DU	DU	DU
8								DU	DU
9									DU

E: Escalate to the next higher dose; **S**: Stay at the same dose; **D**: De-escalate to the previous lower dose; **DU**: De-escalate to the previous lower dose and the current dose will never be used again in the trial;

An mTPI decision table

- Generated based on models
- Presented as rules

Contribution to the society: Interval-based designs (2007-2013) & (2015-2020)

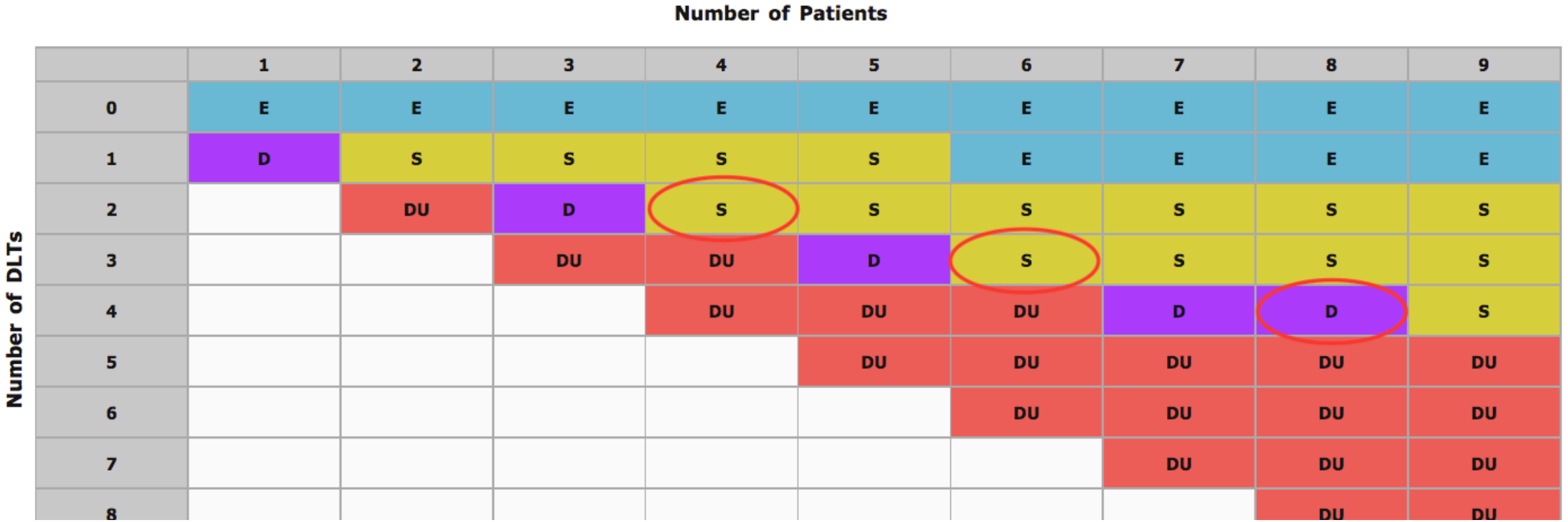
= 9 ; Target toxicity probability = 30% ; epsilon 1 = 0.05 ; epsilon 2 = 0.05 ;

Number of Patients

1	2	3	4	5	6	7	8	
E	E	E	E	E	E	E	E	
D	S	S	S	S	E	E	E	
	DU	D	S	S	S	S	S	
		DU	DU	D	S	S	S	
			DU	DU	DU	D	D	
				DU	DU	DU	DU	
					DU	DU	DU	
						DU	DU	
							DU	

to the next higher dose; **S**: Stay at the same dose; **D**: De-escalate to the previous lower dose; **DU**: De-escalate to the and the current dose will never be used again in the trial;

- For the first time bridged simplicity and model-based inference
- Effectively challenged the 3+3 design as the only clinically popular method
- Widely used in practical trials (publications in Lancet Oncology, JAMA Oncology, etc)
- CCD/BOIN/i3+3 further simplify the approaches



	1	2	3	4	5	6	7	8	9
0	E	E	E	E	E	E	E	E	E
1	D	S	S	S	S	E	E	E	E
2		DU	D	S	S	S	S	S	S
3			DU	DU	D	S	S	S	S
4				DU	DU	DU	D	D	S
5					DU	DU	DU	DU	DU
6						DU	DU	DU	DU
7							DU	DU	DU
8								DU	DU

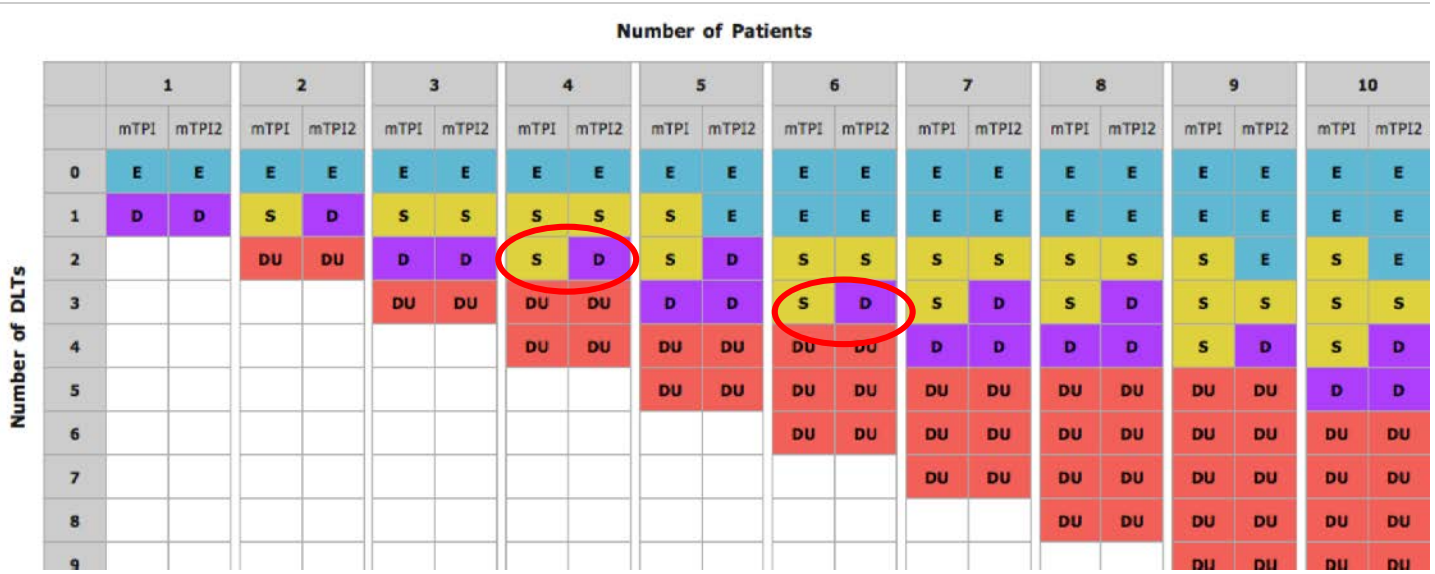
Criticism of the mTPI design table

When 3/6 patients have DLT, how can we “S”, stay at the current dose?

Note: 2/4 – S; but 4/8 – D!

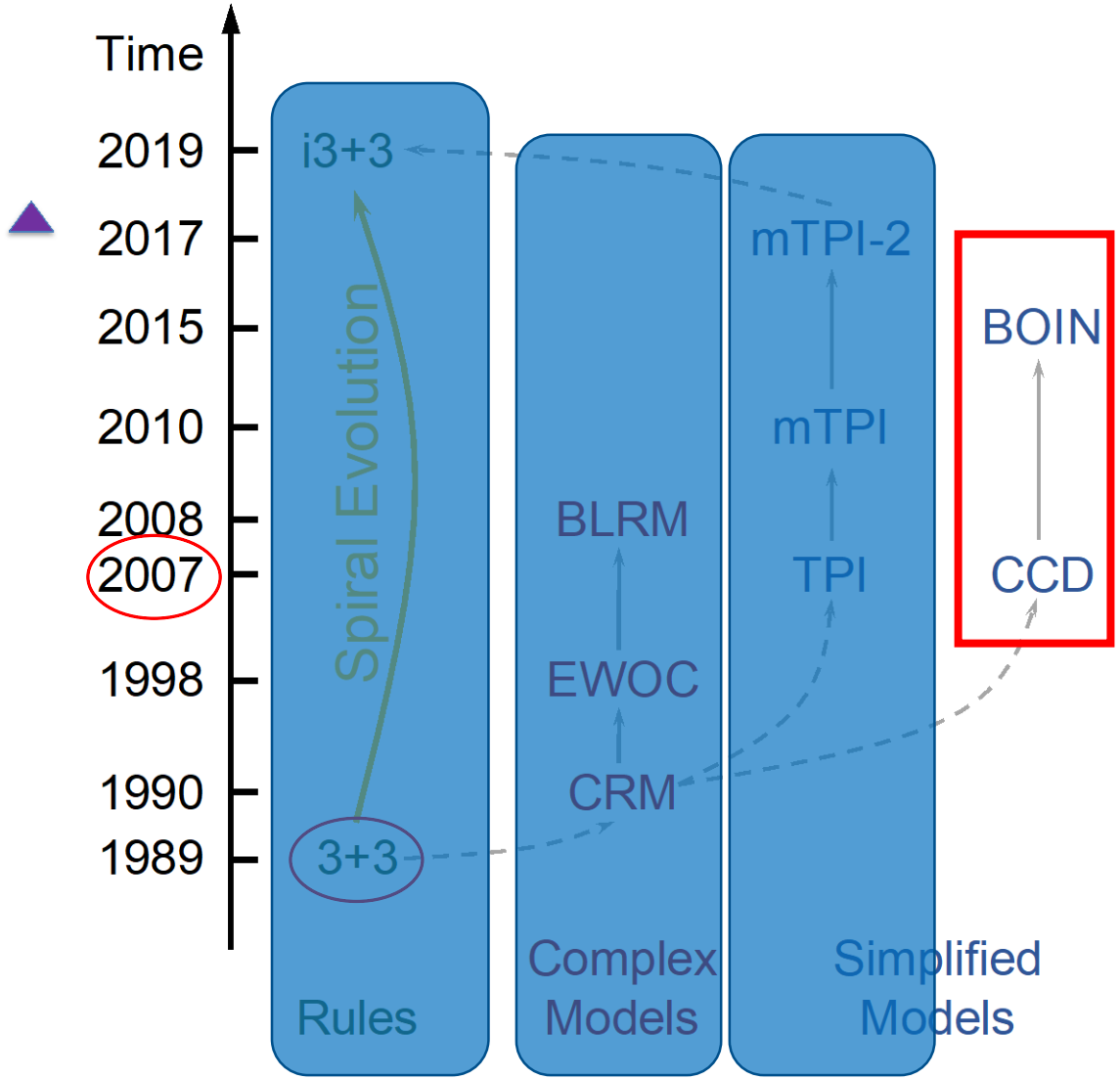
The mTPI decisions are statistically optimal, but

- Guo et al. (2017) show that the decisions in mTPI minimizes the posterior expected 0-1 loss – it is statistically optimal!
- So how can Stay at 3/6 be an optimal decision?
  - 3/6 Stay when $p_T = 0.3$ and EI=(0.25, 0.35). Is it wrong?
  - 4/8 De-escalate based on the same table
  - Statistical variability is the key; 6 patients have larger variability than 8 patients;
  - Alternatives:
    - Change 0-1 loss to a loss based on distance from p_T
    - Ockham's razor: Guo et al. (2017) – the mTPI-2 design
    - mTPI-2 blunts Ockham's razor and makes decisions more "nimble"



The interval designs (2015-now)

- Model-assisted designs
  - Statistical inference using models (simple models)
  - Inference based on point estimate
  - Presentation of the decisions as rules



What is the BOIN design and why is it popular?

- At a dose, **n** (e.g., =3, 6, 9) patients are treated, **y** (=0, 1, 2,…) patients DLT
- Compare $\frac{y}{n}$ with intervals
  - If $\frac{y}{n} \leq p_T - \lambda_1$, Escalate
  - If $p_T - \lambda_1 < \frac{y}{n} < p_T + \lambda_2$, Stay
  - If $\frac{y}{n} \geq p_T + \lambda_2$, De-escalate
- The above rules originally proposed by the CCD design (Ivanova et al., 2007)
- BOIN applies the same safety rules as mTPI/mTPI-2/keyboard

Examples:

0/3, Escalate; 1/3 Stay, 2/3, 3/3 De-escalate

0/6, 1/6, Escalate; 2/6 Stay, 3/6, 4/6, 5/6, 6/6 De-escalate

BOIN is very simple and easy to use. However,
What are the λ_1 and λ_2 and how to decide them?

Quick answer: elicit from physicians.

BOIN: "O" stands for "optimal"

λ_1 and λ_2 are based on an optimization procedure.

- Physicians provide an interval $(p_T - \epsilon_1, p_T + \epsilon_2)$
- BOIN changes it to "optimal" $(p_T - \lambda_1, p_T + \lambda_2)$

How does the optimization work in BOIN? (unpublished)

A decision theoretic framework requires a reasonable model

A classical decision theoretic optimization framework

Data: y; Model: f$(y|\theta)$, Prior: $\pi(\theta)$, Posterior: $p(\theta|y)$

Action: $a \in \{D, S, E\}$; Loss: $l(a,\theta)$, Decision rule: $R(y) \to a$

Optional decision rule: $R^*(y)$ is optimal if it provides the smallest loss (or expected loss); for example, Bayes' rule

$$R^*(y) = \arg\max_a \int l(a,\theta)p(\theta|y)d\theta$$

mTPI/mTPI-2/keyboard are based on Bayes' rules for a model

f$(y|\theta) = Bin(n,\theta)$, $\pi(\theta|interval) = beta(1,1)Ind(interval), \pi(interval) = unf$

and 0-1 loss

BOIN's optimization: prior distribution assumes three point masses

Model: f$(y|\theta) = Bin(n,\theta)$; Prior: $\pi(\theta) = \frac{1}{3}$ if $\theta \in \{\phi_1, p_T, \phi_2\}$

Loss: 0-1 loss $Ind(a = D, \theta \neq \phi_2) + Ind(a = S, \theta \neq p_T) + Ind(a = E, \theta \neq \phi_1)$

Optimal decision: Bayes' rule

if $\frac{y}{n} \leq p_T - \lambda_1$, E(scalate)

if $p_T - \lambda_1 < \frac{y}{n} < p_T + \lambda_2$, S(tay)

if $\frac{y}{n} > p_T + \lambda_2$, D(e-escalate)

Note: The λ_1 and λ_2 can be derived based on optimization using Bayes' rule.

Question: do we believe the probability of toxicity only takes three values?

CCD/BOIN decision rules based on a point estimate $\frac{y}{n}$

if $\frac{y}{n} \leq \phi_1$, E(scalate)
if $\phi_1 < \frac{y}{n} < \phi_2$, S(tay)
if $\frac{y}{n} > \phi_2$, D(e-escalate)

- Turns out the above decision rules in BOIN corresponds to a Bayes' (optimal) rule – optimal for posterior expected 0-1 loss
- However
  - It is only optimal if one assumes that the prior distribution (of toxicity probability of each dose) can only take three values; $\theta =$
    - Target toxicity probability: p_T with prob 1/3
    - Left boundary of the EI: $\phi_1 = (p_T - e_1)$ with prob 1/3
    - Right boundary of the EI: $\phi_2 = (p_T + e_2)$ with prob 1/3
  - The optimization boundaries $p_T - \lambda_1 < \frac{y}{n} < p_T + \lambda_2$ have an unintended consequence (next slide)

An unintended consequence of optimization

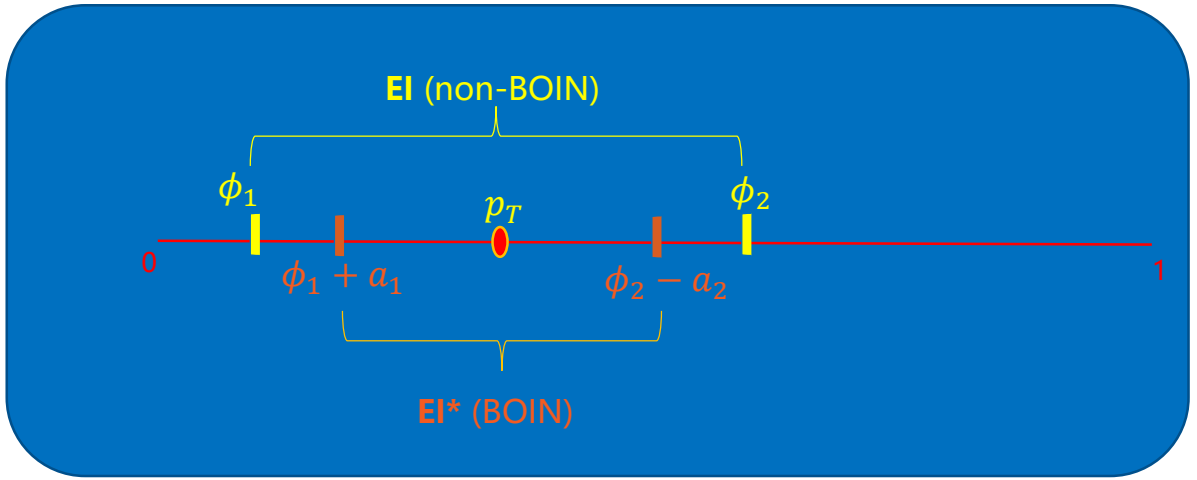
- Except for BOIN, all other iDesigns and ibDesigns prespecify (by users input)

$$\text{EI} = (\phi_1, \phi_2) \equiv (p_T - e_1, p_T + e_2)$$

- BOIN "optimize" EI to an "optimal" EI,

$$\text{EI}^* = (\phi_1 + a_1, \phi_2 - a_2) \equiv (\lambda_1, \lambda_2)$$

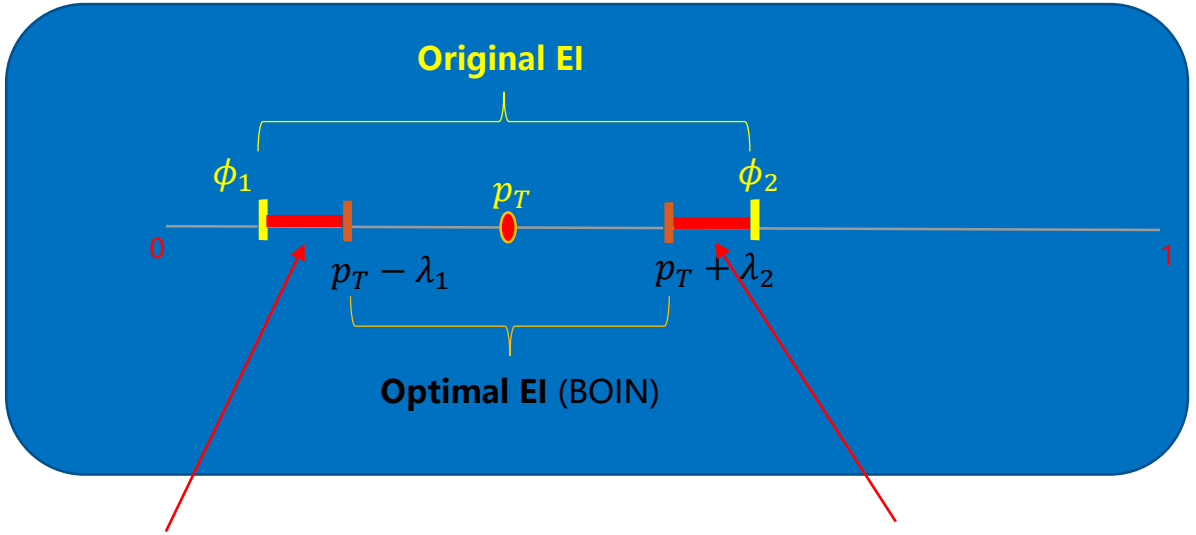
before observing any data



Remarks:

1. All the designs elicit EI from the investigators as the "tolerance" for MTD to be away from p_T
2. EI is a preference, just like p_T
3. So what does it mean to optimize EI before data are observed?

An unintended issue with optimal intervals



- The "optimal" EI and the elicited EI generate gaps due to the optimization framework

1. The gaps are independent of n, the sample size: they will always be there however large the sample size
2. When a dose "falls" into the gaps, the decisions based on the original EI and optimal EI are different! – this directly contradicts the elicitation process with physicians
3. The performance of the designs with original EI and optimal EI is almost identical in small samples (gaps are small)
4. The gaps are due to the over-simplified statistical model: $\theta = (p_T, \phi_1, \phi_2)$ with prob 1/3
5. But θ is the probability of toxicity at each dose – it cannot just take three values

One may not need simulations to evaluate interval-based designs!

Number of Patients

	2		3		4		5		6		7		8		9		10	
	mTPI	mTPI2	mTPI	mTPI2	mTPI	mTPI2	mTPI	mTPI2	mTPI	mTPI2	mTPI	mTPI2	mTPI	mTPI2	mTPI	mTPI2	mTPI	mTPI2
	E	E	E	E	E	E	E	E	E	E	E	E	E	E	E	E	E	E
	S	D	S	S	S	S	S	E	E	E	E	E	E	E	E	E	E	E
	DU	DU	D	D	S	D	S	D	S	S	S	S	S	S	S	E	S	E
			DU	DU	DU	DU	D	D	S	D	S	D	S	D	S	S	S	S
					DU	DU	DU	DU	DU	DU	D	D	D	D	S	D	S	D
							DU	DU	DU	DU	DU	DU	DU	DU	DU	DU	D	D
									DU	DU	DU	DU	DU	DU	DU	DU	DU	DU
											DU	DU	DU	DU	DU	DU	DU	DU
													DU	DU	DU	DU	DU	DU
															DU	DU	DU	DU
																	DU	DU

e next higher dose; **S**: Stay at the same dose; **D**: De-escalate to the previous lower dose; **DU**: De-escalate to the se and the current dose will never be used again in the trial;

- Interval-based designs contain three key components:
  - i. Safety rules that stop trials or exclude doses if $\Pr(p_d \text{ or } p_1 > p_T | data) > 0.95$
  - ii. MTD selection procedure: for example, $\text{argmin}|\hat{p}(x) - p_T|$ where $\hat{p}(x)$ is isotonic transformed posterior mean
  - iii. The pretabulated decision tables

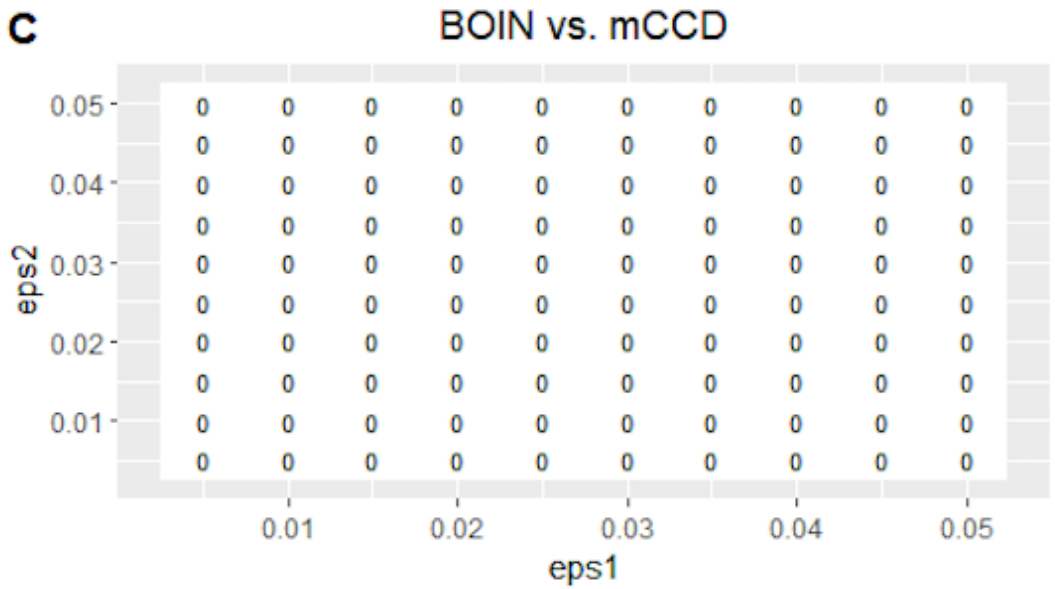
The mTPI, mTPI-2/keyboard, BOIN, i3+3 have identical i and ii. The only differences are in iii.

Without running simulations one can already evaluate interval-based designs – BOIN and CCD

- BOIN and mCCD (mCCD = CCD + safety rules)

  have the same i & ii; and for iii mCCD does not optimize the EI

- The decis e identical (for <=51



Differences = 0 out of a total 1,326 decisions; $p_T = 0.3$

Simulation results can be misleading

- For interval-based designs; just look for the three components:
  - i. Safety rules
  - ii.MTD selection
  - iii.Decision tables

- Simulation results based on repeated computer-generated clinical trials are completed decided by i-iii.

- Read review papers with caution
  - No single design can dominate another design in ALL scenarios

Statistical modeling is about variabilities: 3/6=0.5; 30/60=0.5, 3000/6000=0.5…

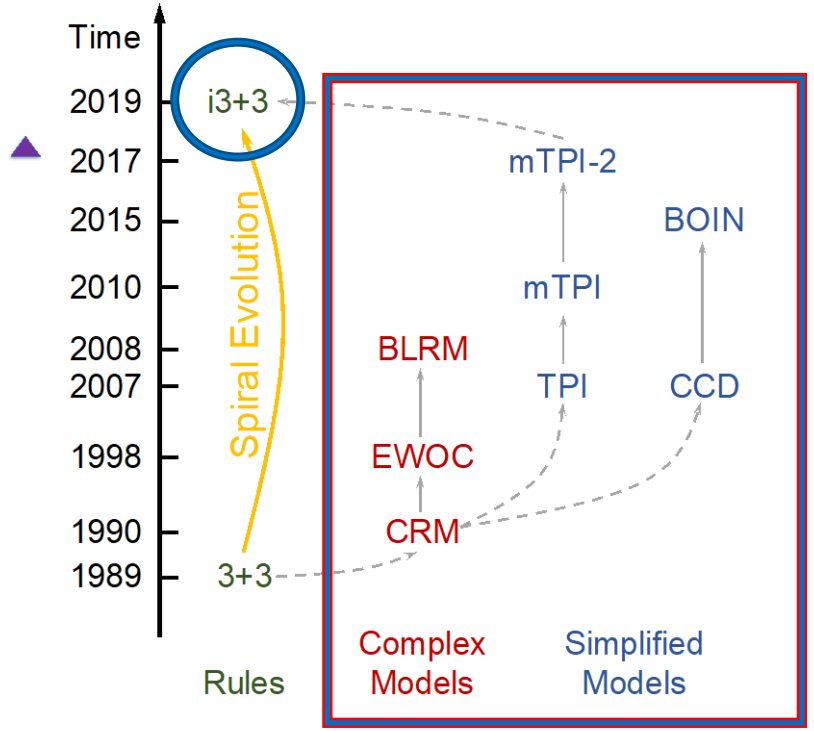
- The hallmark of statistics is variability
  - If no variability, no need for statistics!
  - But the rules below ignore variability!

If $\frac{y}{n} \le p_T - \lambda_1$, Escalate; if $p_T - \lambda_1 < \frac{y}{n} < p_T + \lambda_2$, Stay; if $\frac{y}{n} \ge p_T + \lambda_2$, De-escalate

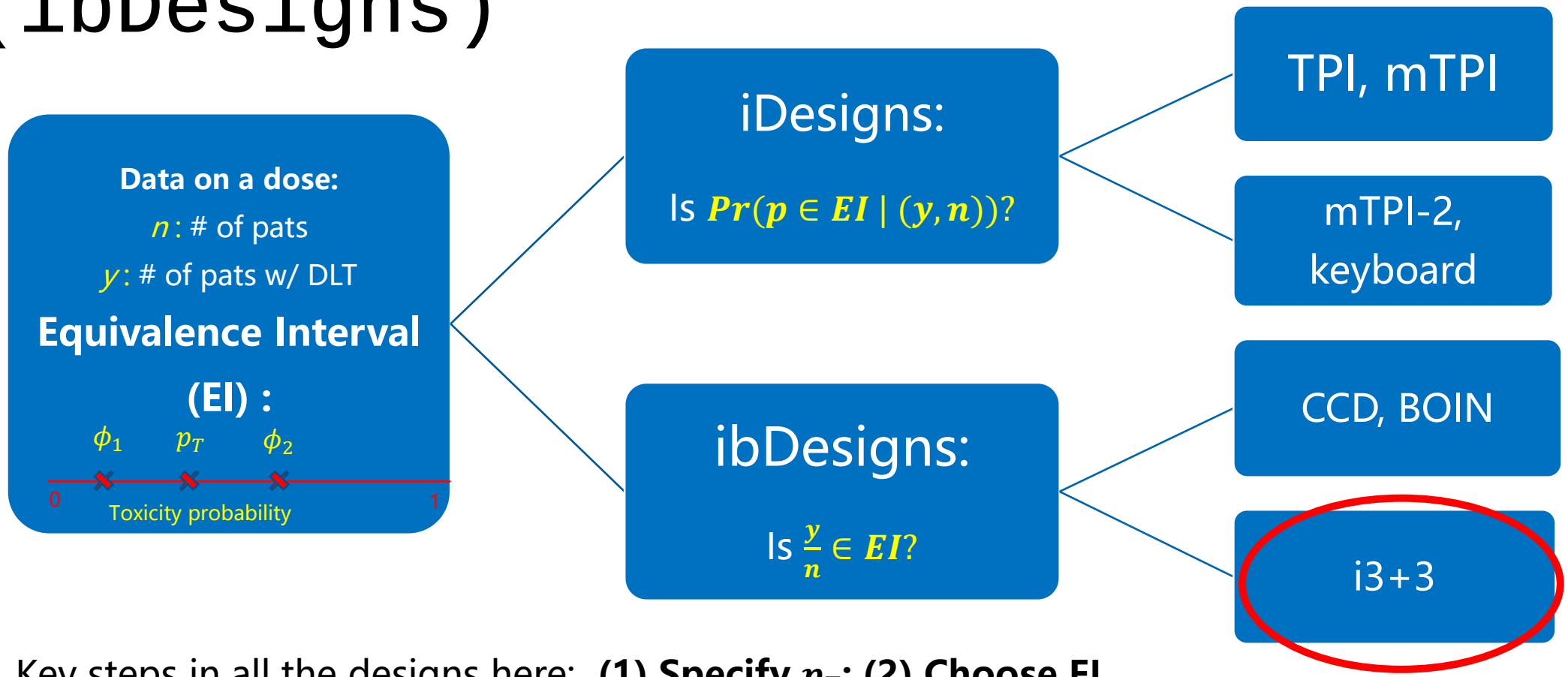
- Remember this picture?
  - We spent 30 years resorting to sta
  - Because models account for variabi
  - But if in the end the decision rul

account for variability; why bother:

The last chapter: back to rule-base



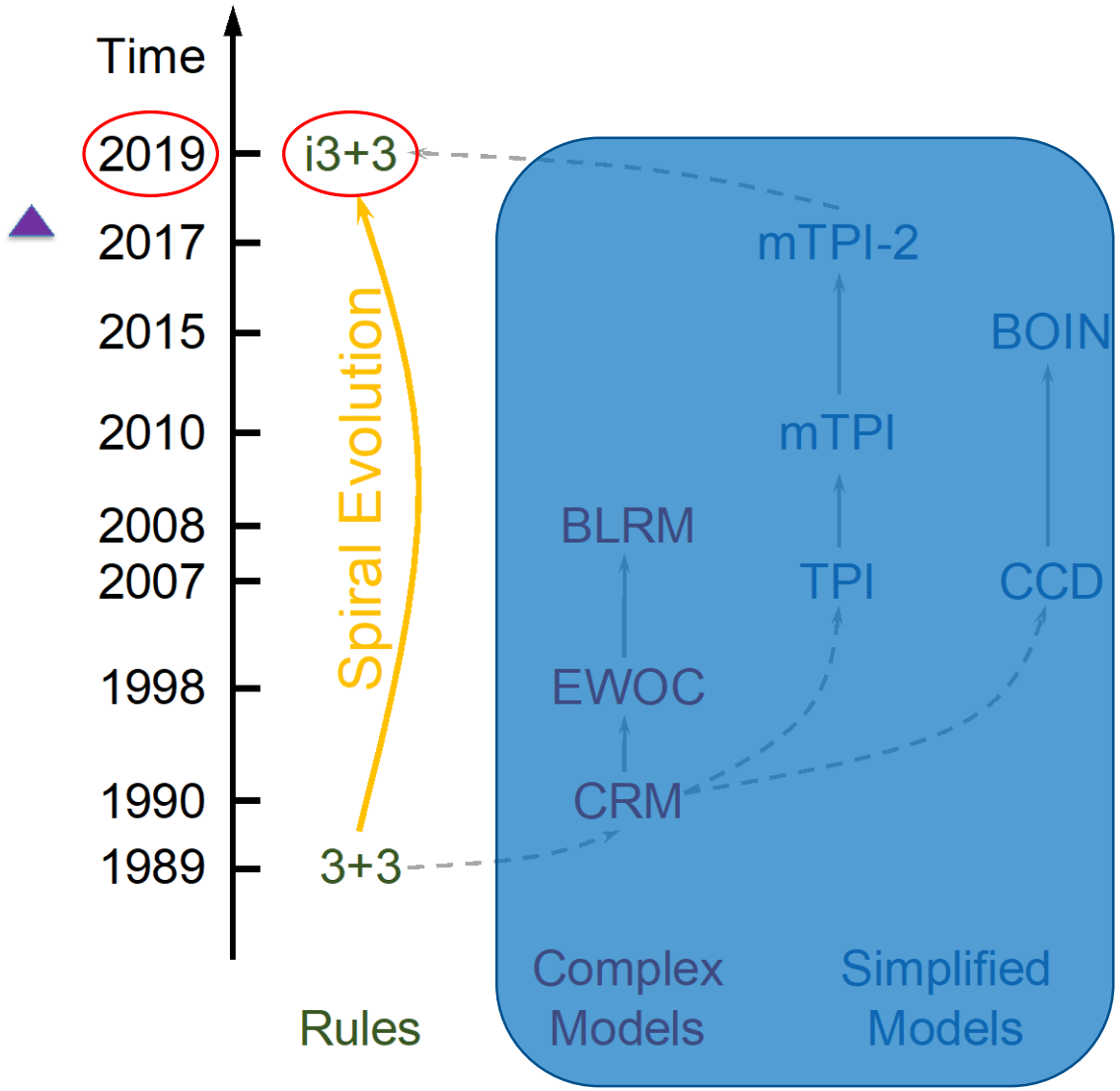
and the interval-boundary designs (ibDesigns)



- Key steps in all the designs here: **(1) Specify p_T; (2) Choose EI**
- All the designs here use **the same isotonic regression** to choose final MTD
- Except CCD, all the designs have **the same safety rules**

The i3+3 design (2019)

- Rule based (smart rules)
- No statistical model
  - Easy
  - Transparent
  - Social acceptance
- Flexible and powerful



The i3+3 design (Liu, Wang, Ji, 2020)

Users provide p_T (e.g., 0.3), and EI = ($p_T - e_1$, $p_T + e_2$). No need to change.

i3+3:Dose-finding algorithm

- If $\frac{y}{n} < p_T - e_1$, Escalate;
- If $p_T - e_1 \le \frac{y}{n} \le p_T + e_2$, Stay;
- If $\frac{y}{n} > p_T + e_2$,
  1. If $\frac{y-1}{n} < p_T - e_1$, Stay;
  2. Else, De-escalate;

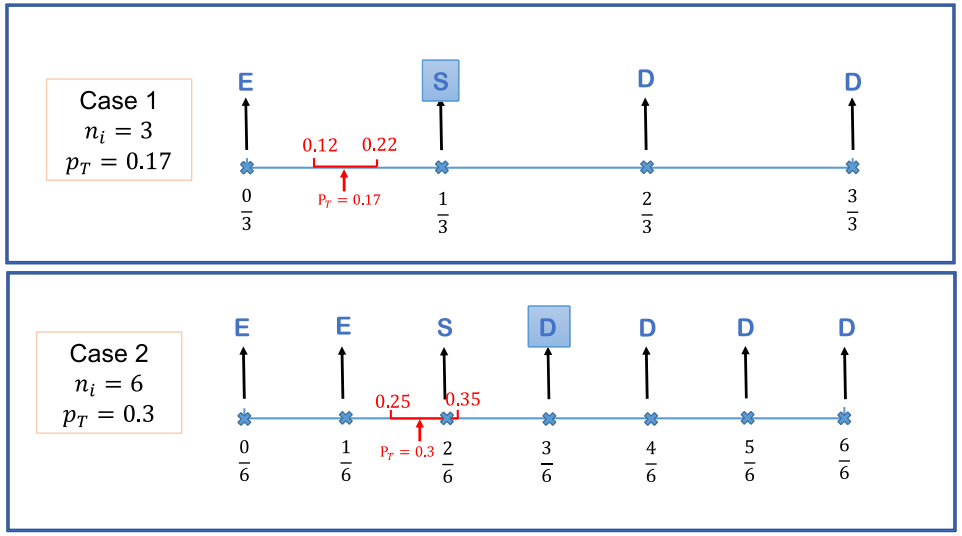
Examples: p_T = 0.25, EI = (0.2, 0.3)

i3+3:	*BOIN:*
0/3 – Escalate;	0/3 – Escalate;
1/3 – Stay;	1/3 – De-escalate;
2,3/3 – De-escalate	2,3/3 – De-escalate

When $\frac{1}{n} > e_1 + e_2$, i3+3, CCD, and BOIN have identical decisions as long as they use the same EI

The i3+3 decision rules (two examples)

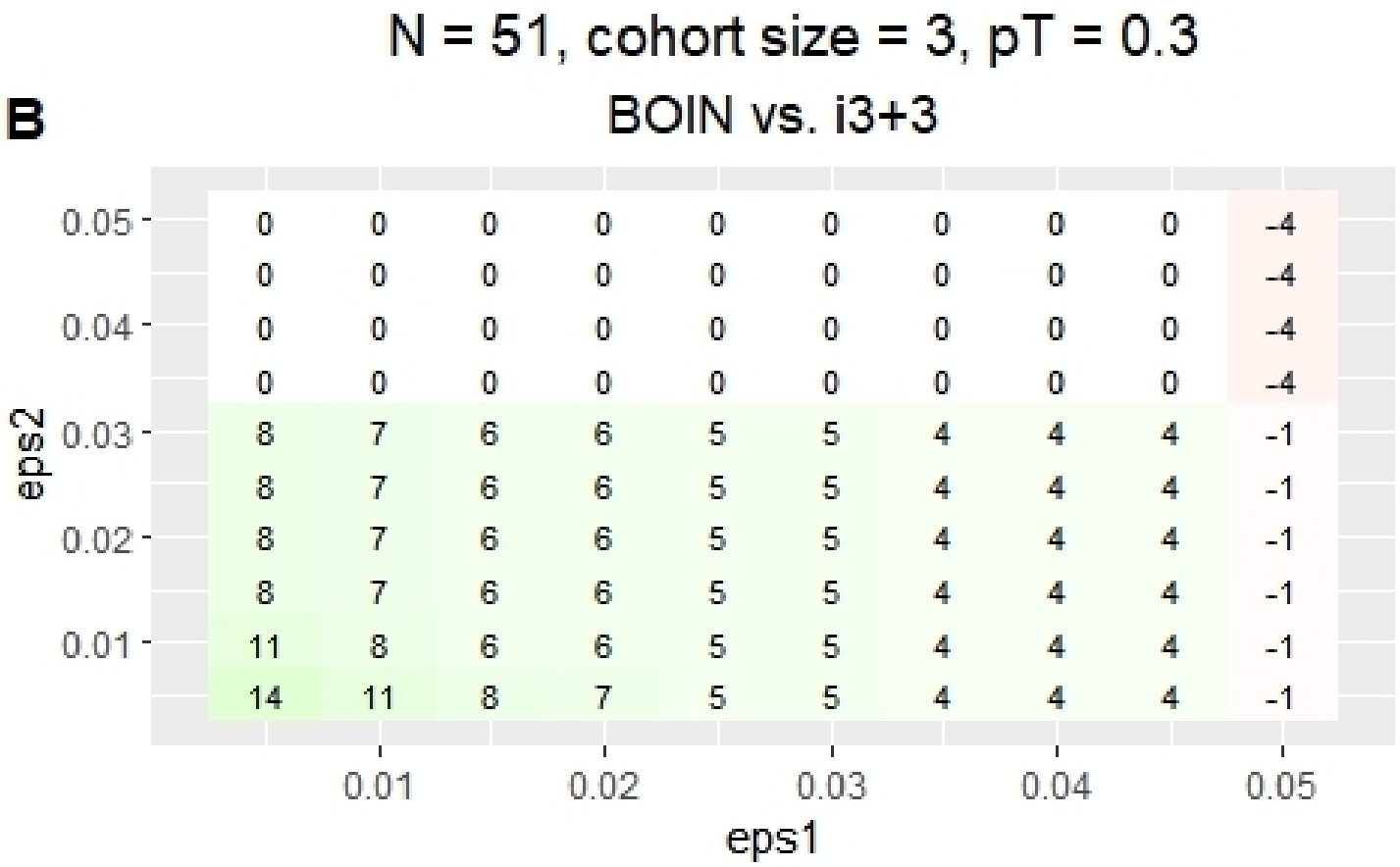
- Comparison to mTPI and BOIN when 1/n is large



Equivalence Interval

Current dose: d **No. enrolled:** $n = 6$	**Target probability:** $p_T = 0.3$ **EI:** $[p_T - \epsilon_1, p_T + \epsilon_2] = [0.25, 0.35]$		
No. DLTs: x	Next dose level		
	i3+3	mTPI	BOIN
0	$d+1$	$d+1$	$d+1$
1	$d+1$	$d+1$	$d+1$
2	d	d	d
3	$d-1$	d	$d-1$
4	$d-1$	$d-1$	$d-1$
5	$d-1$	$d-1$	$d-1$
6	$d-1$	$d-1$	$d-1$

Current dose: d **No. enrolled:** $n = 3$	**Target probability:** $p_T = 0.17$ **EI:** $[p_T - \epsilon_1, p_T + \epsilon_2] = [0.12, 0.22]$		
No. DLTs: x	Next dose level		
	i3+3	mTPI	BOIN
0	$d+1$	$d+1$	$d+1$
1	d	d	$d-1$
2	$d-1$	$d-1$	$d-1$
3	$d-1$	$d-1$	$d-1$



Differences in about 1% out of a total 1,326 decisions

So how does i3+3 perform?

- For interval-based designs; just look for the three components:
  - Safety rules
  - MTD selection
  - Decision tables
- With only tiny differences the two designs perform comparably

How to evaluate the simulation results? Lots of summary statistics; can be difficult to cc

Scenario 1

Simulated Dose Escalation

p_T = 0.25, n_{sim} = 1000		Selection Prob.		Average # of Patients Treated (s.d.)		Average # of Toxicities (s.d.)	
Dose Level	True Tox Prob.	mTPI-2	mCCD	mTPI-2	mCCD	mTPI-2	mCCD
1	0.13	0.375	0.234	12.072 (8.092)	11.781 (7.821)	1.605 (1.961)	1.575 (1.929)
2	0.25	0.491	0.555	11.646 (5.862)	11.634 (5.625)	2.88 (1.912)	2.842 (1.9)
3	0.38	0.112	0.191	5.016 (5.313)	5.28 (5.231)	1.908 (1.987)	2.025 (1.985)
4	0.5	0.008	0.012	1.038 (2.449)	1.077 (2.481)	0.528 (1.228)	0.549 (1.245)
5	0.63	0	0.001	0.102 (0.649)	0.102 (0.649)	0.075 (0.479)	0.075 (0.479)

		mTPI-2	mCCD
MTD Selection	Prob. of Selecting MTD	0.491	0.555
	Prob. of Selecting Dose-over-MTD	0.12	0.204
	Prob. of No Selection	0.014	0.007
Patients Assignment	Prob. of Correct Allocation (s.d.)	0.388 (0.195)	0.388 (0.187)
	Prob. of Overdosing Allocation (s.d.)	0.205 (0.234)	0.215 (0.231)
Trial Toxicity	Prob. of Toxicity	0.234	0.237
Trial Stopping**	Prob. of Early Stopping Trial due to Safety Rule	0.007	0.007
	Prob. of Early Stopping Trial due to Reaching K	0	0
	Prob. of Stopping Trial due to Reaching n	0.993	0.993
Trial Sample Size	Average # of Patients Treated (s.d.)	29.874 (1.622)	29.874 (1.622)

Less

* The row with background color indicates the TRUE MTD

** For further details concerning Trial Stopping Rule, please refer to section 1.2.2 in the User Manual.

The J-Score is a weighted average of MTD selection and patient allocation

$$Utility_{iA} = R_1 \times \%SelAtMTD_i - p_{l_1} \times \%SelBelowMTD_i - p_{h_1} \times \%SelAboveMTD_i$$

and

$$Utility_{iB} = R_2 \times \%PntAtMTD_i - p_{l_2} \times \%PntBelowMTD_i - p_{h_2} \times \%PntAboveMTD_i$$

The total utility for design i is defined as the sum of the two utilities:

$$Utility_i = Utility_{iA} + Utility_{iB}$$

and for $i = 1, ..., I$ designs, $\mathbf{Utility} = \{Utility_i\}$ denotes the vector of the designs that are under comparison.

The J-score, which is the continuous rank index of the total utility, is thus defined as:

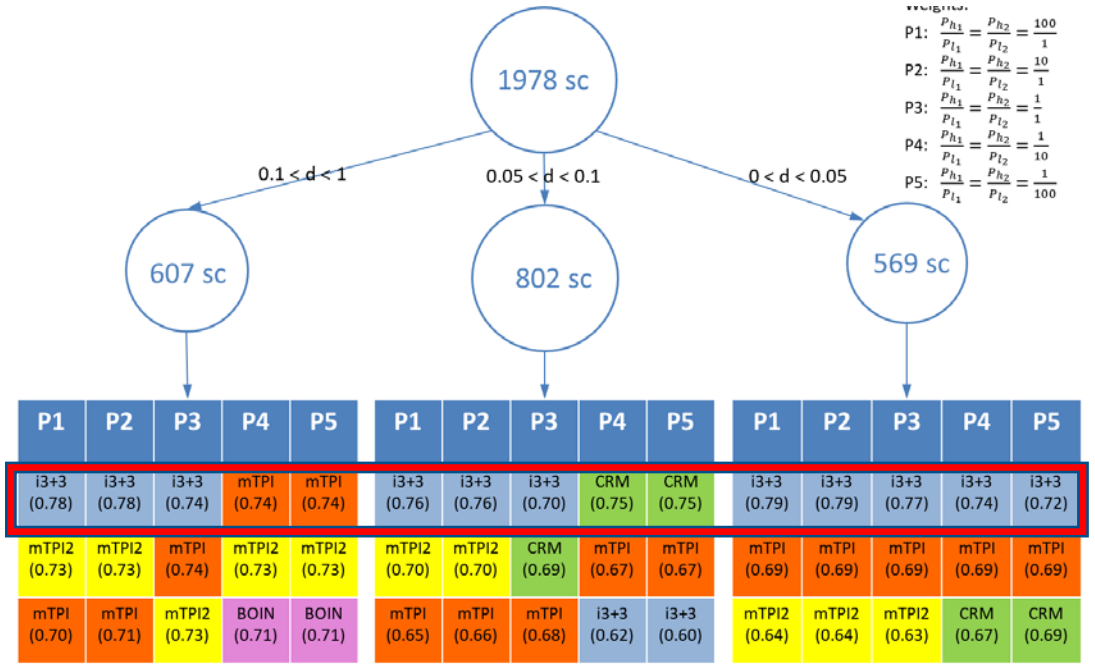
$$Jscore_i = \frac{Utility_i - \min(\mathbf{Utility})}{\max(\mathbf{Utility}) - \min(\mathbf{Utility})}$$

J score is between 0 and 1; the larger value, the better the design

J-Score results: Different designs perform better in different types of scenarios

- Based on 1,978 scenarios from users
- Massive simulations
- A tree summarizes the best designs based on scenarios and drug development preferences
  - Aggressive: P1, P2
  - Neutral: P3
  - Conservative: P4, P5
- A higher score means better
  - High probability of selecting the MTD
  - Safe in allocating patients
  - Safe in selecting the doses

d – distance between MTD and adjacent doses; larger value means easier scenario



Top performers for each category

Conclusions

1. Model-based methods are more powerful
   - Yes, but depends on what you care (simplicity; model-misspecification; 1/3 Stay no matter what)

2. mTPI is not safe
   - Yes, but depends on your loss function (e.g., 3/6 is not too much, but 4/8 is)

3. CCD and BOIN are model-assisted designs
   - Yes, model-assisted is still model-based

4. mTPI, mTPI-2, BOIN, Keyboard
   - They are model-based designs; and mTPI-2 = Keyboard

5. Which designs to use? Depends!
   - For classical single-agent DLT-based cohort—enrollment phase 1 trial, physicians can use i3+3 or mTPI-2, but mostly i3+3, unless your drug is very very safe (future talk)
   - CRM is also really good but requires statistical expertise and support
   - BOIN performs really well and is simple; but it has theoretical flaws
   - It seems that

heavy safety regulation + simple model/inference ≈ Model-free (rule-based) design

Phase I Trials in oncology is becoming more sophisticated and powerful

- Seamless Phase 1a dose finding + Phase 1b cohort expansion
  - Bayesian hierarchical models for borrowing information
- Immune and targeted therapies
  - MTD may not be the RP2D;
  - Multiple candidate doses for expansion
  - Multiple indications (NSCLC, GC, Ovarian, Prostate, RCC, etc)
- Delayed toxicity outcomes
- Combination treatments (novel + novel combo)
- Eff/Tox dose finding (for cell/gene therapies, e.g., CAR-T)
- Rolling enrollment to speed up the trial
- Borrowing historical data for dose finding (Hi3+3; https://arxiv.org/abs/2010.10244)

Cytel
Shaping the Future of Drug Development

Thank you

Questions, Comments?