Novel Trial Designs for Early Phase Clinical Trials: Beyond 3+3

Paul Frankel, PhD, City of Hope

Phase I Traditional 3+3 Design – A Risk-Limiting Design

# on Current Level			Action
DLT	EVAL	EVAL+At Risk	
0	0-2	1-2	**Accrue next patient at this level**
0	0-2	3	**Hold accrual**
0	3	3	**Accrue next patient at the next higher level****
1	1-2	1-2	**Accrue next patient at this level**
1	1-2	3	**Hold accrual**
1	3-5	3-5	**Accrue next patient at this level**
1	3-5	6	**Hold accrual**
1	6	6	**Accrue next patient at the next higher level****
2-3	any	any	**Accrue next patient at the next lower level*****

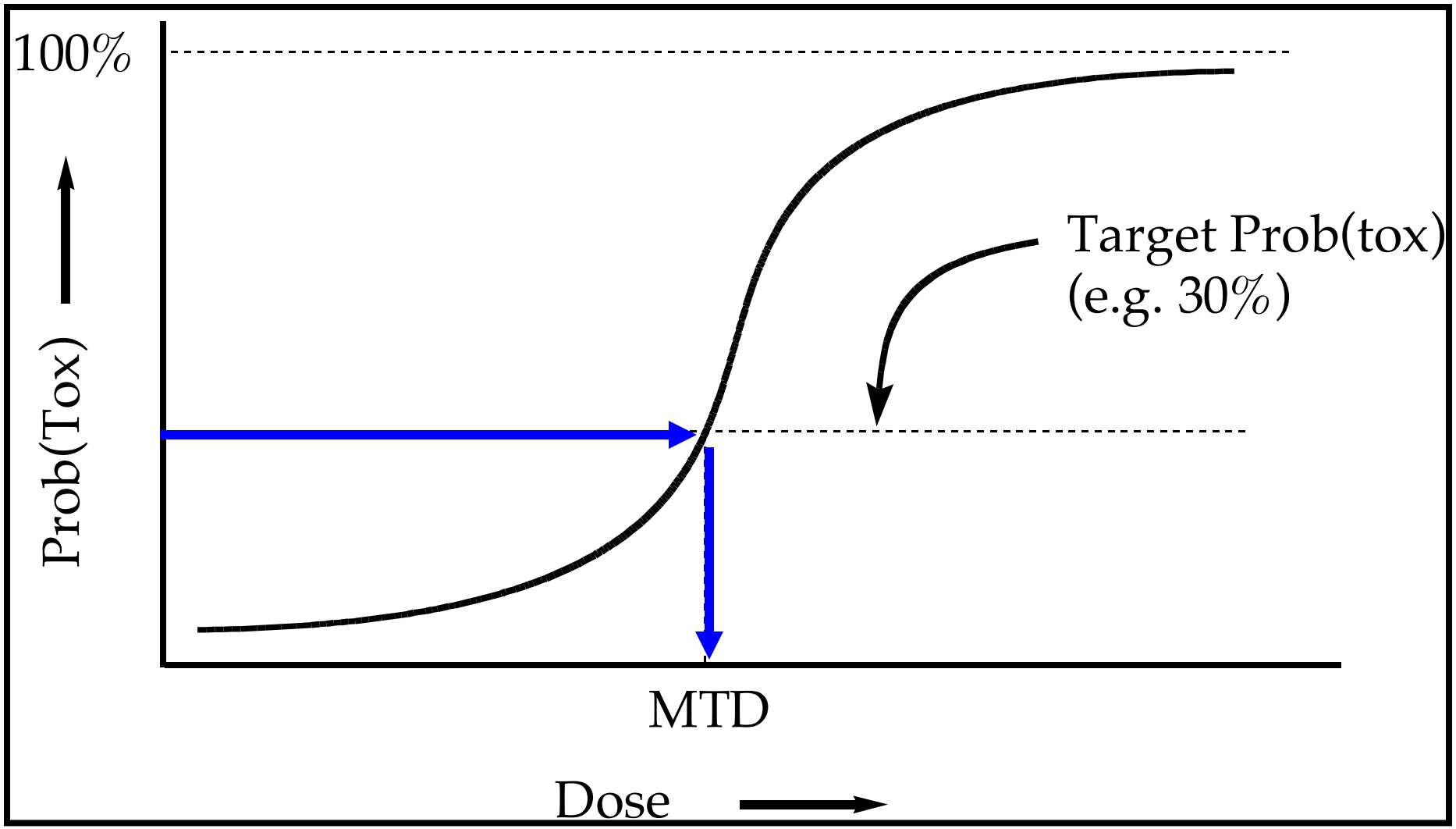
****If the next higher dose level exceeds the MTD, declare the dose the MTD or if there are not sufficient evaluable patients at the current level to declare it the MTD, accrue at the current level.**

*****Current level exceeds the MTD. The MTD is the highest level at which <33% of patients had DLTs, with at least 6 evaluable patients.**

IQ 3+3 (Same Risk-Limits)

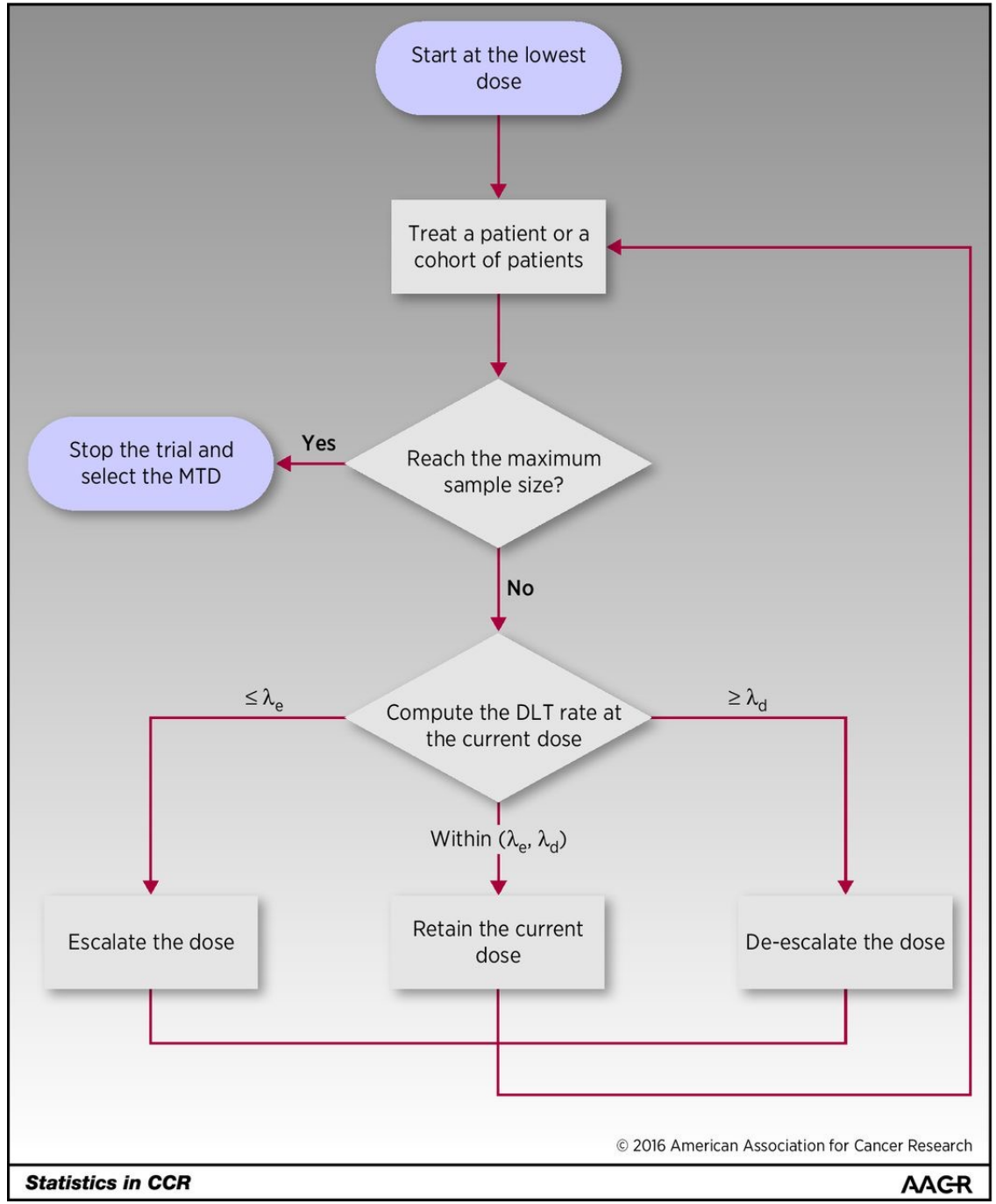
# Patients on Current Level			Action**
DLT	EVAL	EVAL+At Risk	
0	0	1-2	**Accrue next patient at this level**
0	0	3	**Hold accrual**
0	1	1-3	**Accrue next patient at this level**
0	1	4	**Hold accrual**
0	2	2-5	**Accrue next patient at this level**
0	2	6	**Hold accrual**
0	3-6	3-8	**Accrue next patient at the next higher level***
1	1	1-2	**Accrue next patient at this level**
1	1	3	**Hold Accrual**
1	2	2	**Accrue next patient at this level**
1	2	3	**Accrue next patient at this level**
1	2	4	**Hold accrual**
1	3-5	3-5	**Accrue next patient at this level**
1	3	6	**Hold accrual**
1	4	6	**Accrue next patient at this level**
1	5	6	**Accrue next patient at this level**
1	4	7	**Hold accrual**
1	5	7	**Accrue next patient at this level**
1	6-8	6-8	**Accrue next patient at the next higher level***
2	2-6	2-6	**Accrue next patient at the next lower level**
2	6	7	**Accrue next patient at the next lower level**
2	7	7	**Declare MTD**
2	7	8	**Hold accrual**
2	8	8	**Declare MTD**
3	**any**	**any**	**Accrue next patient at the next lower level**

MTD Definition for Risk-Targeting Designs



Y. Yuan, K, Hess, S. Hilsenbeck, M Gilbert, CCR 2016

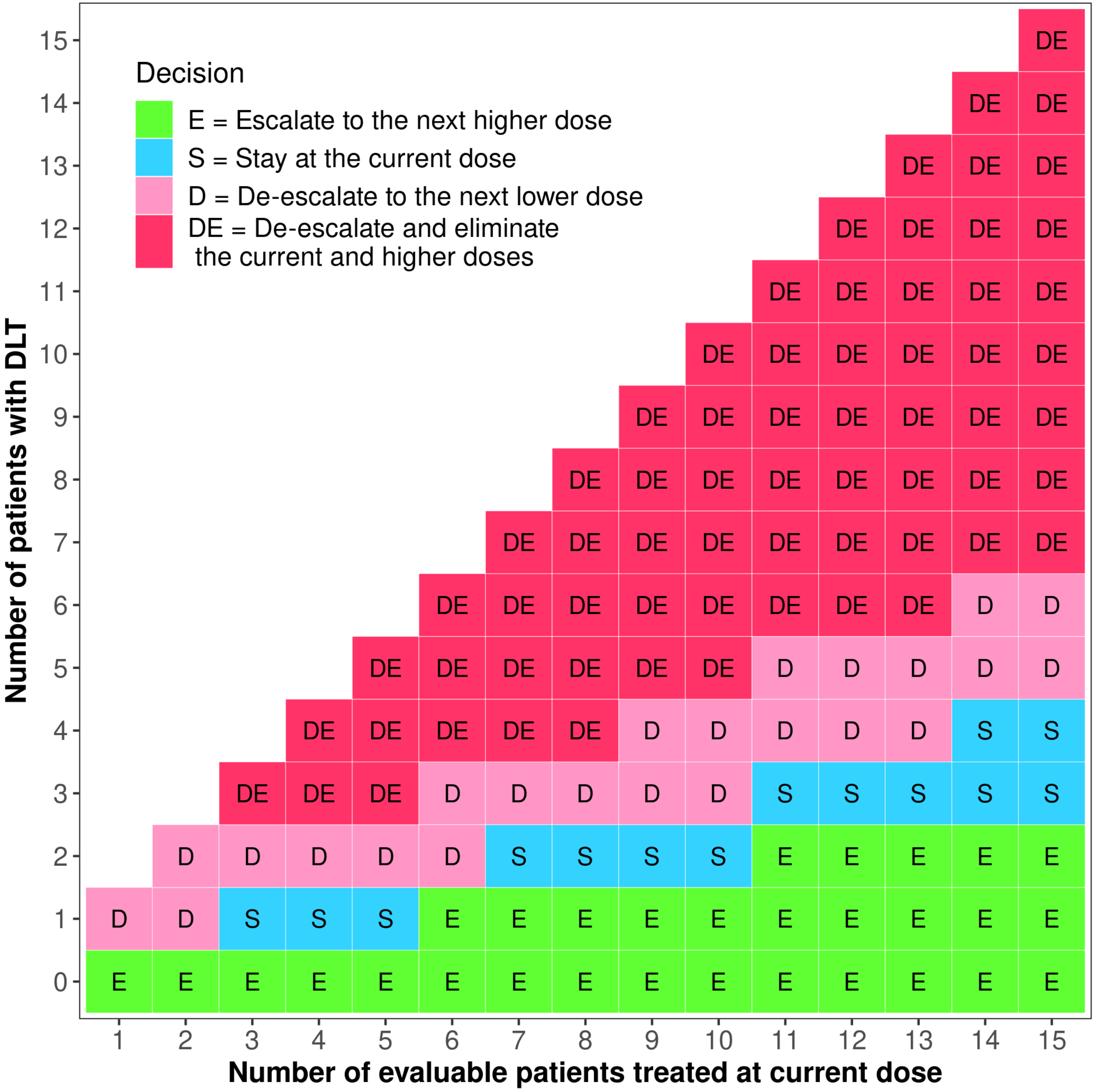
A numerical study shows that the BOIN design ... is more likely than the 3+3 design to correctly select the MTD and allocate more patients to the MTD.


Start at the lowest dose
Treat a patient or a cohort of patients
Reach the maximum sample size?
Yes
Stop the trial and select the MTD
No
Compute the DLT rate at the current dose
≤ λe
≥ λd
Within (λe, λd)
Escalate the dose
Retain the current dose
De-escalate the dose
© 2016 American Association for Cancer Research
Statistics in CCR
AACR

Dose escalation and de-escalation boundaries

	Target toxicity rate for the MTD						
	0.1	0.15	0.2	0.25	0.3	0.35	0.4
λe (escalation)	0.078	0.118	0.157	0.197	0.236	0.276	0.316
λd (de-escalate)	0.119	0.179	0.238	0.298	0.358	0.419	0.479

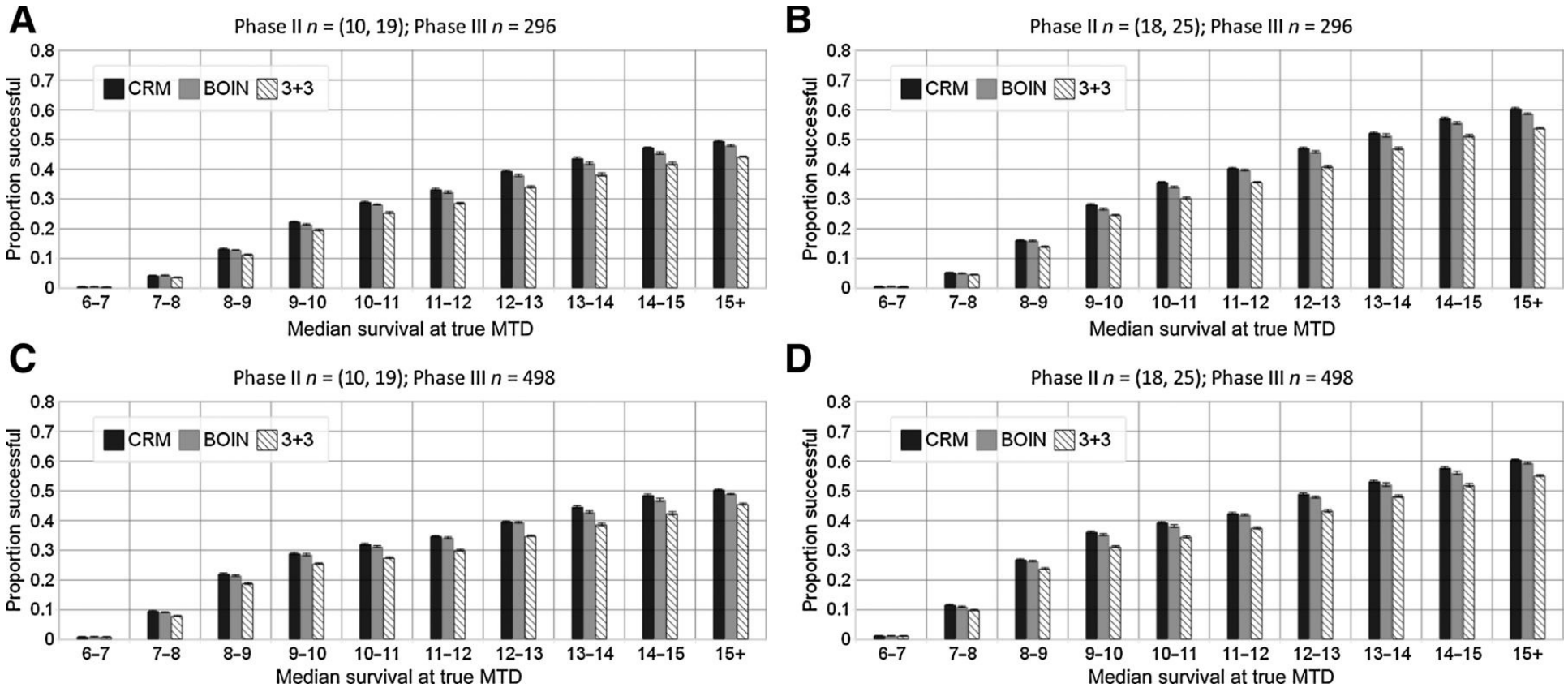
For a 25% DLT target, this translates into the following:


Decision
E = Escalate to the next higher dose
S = Stay at the current dose
D = De-escalate to the next lower dose
DE = De-escalate and eliminate the current and higher doses
Number of patients with DLT
Number of evaluable patients treated at current dose

Conaway and Petroni CCR 2019

"The Impact of Early-Phase Trial Design in the Drug Development Process"

The results underscore the importance of the choice of the early-phase designs. Use of the 3+3 results in fewer agents with successful phase III trials compared with the CRM or BOIN. The difference is more pronounced among highly effective agents. (Conclusion in Abstract)


A
Phase II n = (10, 19); Phase III n = 296
B
Phase II n = (18, 25); Phase III n = 296
C
Phase II n = (10, 19); Phase III n = 498
D
Phase II n = (18, 25); Phase III n = 498
CRM
BOIN
3+3
Proportion successful
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
6–7
7–8
8–9
9–10
10–11
11–12
12–13
13–14
14–15
15+
Median survival at true MTD

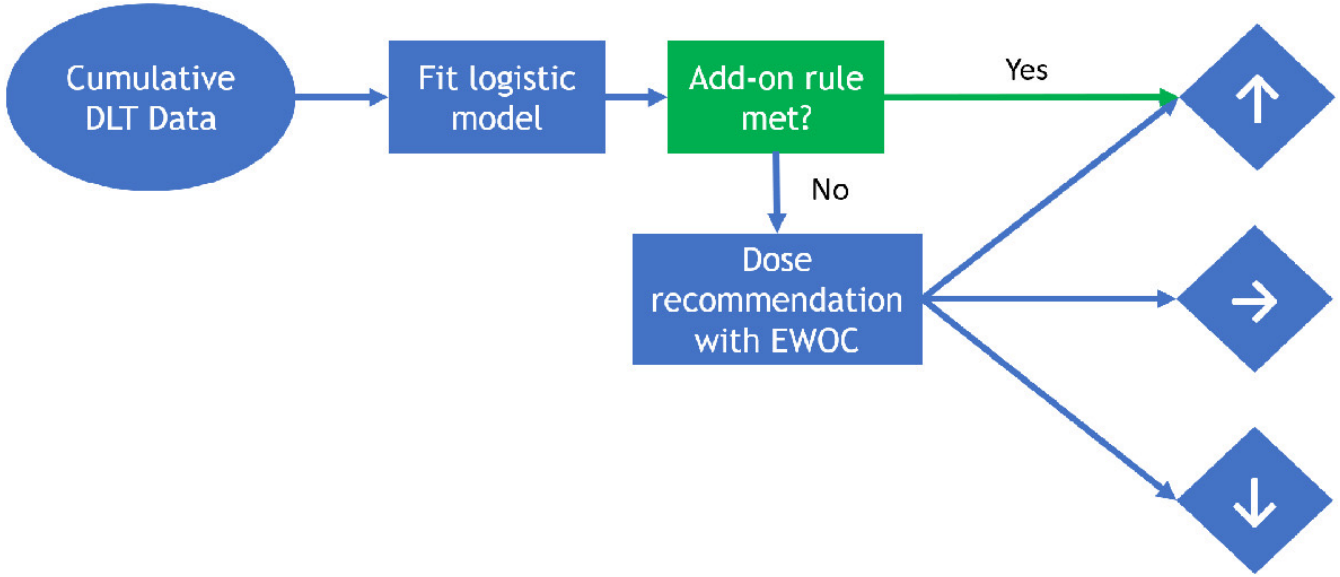
Bayesian logistic regression model with underdose control (and OC): Stat Med 2022

"Simulation results reveal that new designs have better accuracy and treat more patients at the MTD"

BLRM with overdose control is found to be conservative. To overcome this, we ...also account for underdosing.

α(Probability of underdosing) >
(1-α)(Probability of overdosing)

Improving the performance of Bayesian logistic regression model with overdose control in oncology dose-finding studies



Statistics in Medicine, Volume: 41, Issue: 27, Pages: 5463-5483, First published: 15 April 2022, DOI: (10.1002/sim.9402)

Early-Phase Oncology Trials: Why so many designs?

M. Clertant, JCO Oct, 2022

..an overview of the designs' technical differences, **advantages** .. which have a straightforward impact.

*Translational relevance

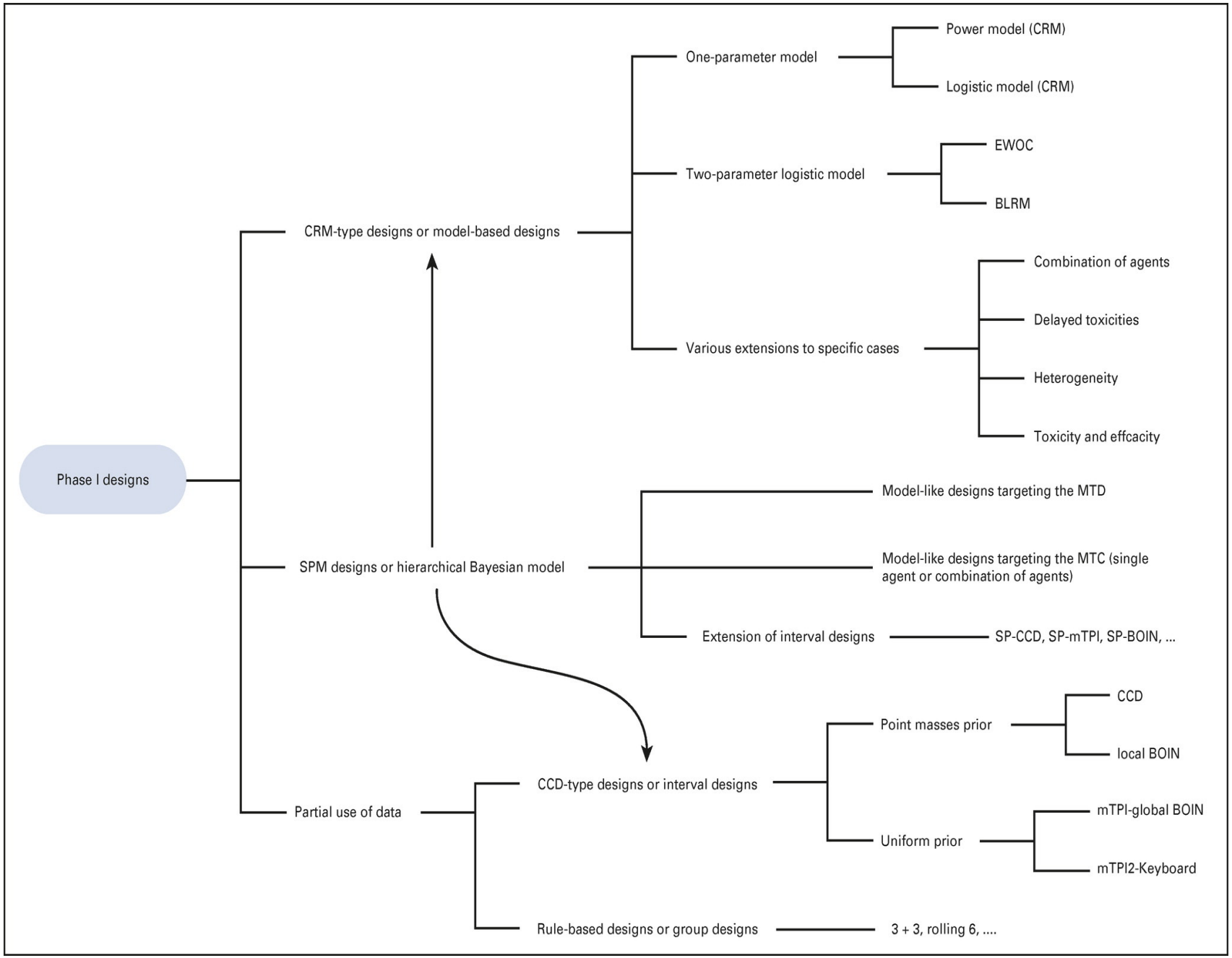

Phase I designs
CRM-type designs or model-based designs
One-parameter model
Power model (CRM)
Logistic model (CRM)
Two-parameter logistic model
EWOC
BLRM
Various extensions to specific cases
Combination of agents
Delayed toxicities
Heterogeneity
Toxicity and effcacity
SPM designs or hierarchical Bayesian model
Model-like designs targeting the MTD
Model-like designs targeting the MTC (single agent or combination of agents)
Extension of interval designs
SP-CCD, SP-mTPI, SP-BOIN, ...
Partial use of data
CCD-type designs or interval designs
Point masses prior
CCD
local BOIN
Uniform prior
mTPI-global BOIN
mTPI2-Keyboard
Rule-based designs or group designs
3 + 3, rolling 6, ...

TABLE 2. Different Designs and Their Properties

Design Class	3 + 3	CRM	BLRM	EWOC	CCD	mTPI	mTPI2	BOIN	Keyboard	SPM
Coherence		✓								✓
Adaptability	✓	✓			✓	✓	✓	✓	✓	✓
No unused observations		✓	✓	✓						✓
Straightforward evaluation	✓	✓			✓	✓	✓	✓	✓	✓
Interval calibration					✓	✓	✓	✓	✓	✓
Interval convergence		✓								✓
MTD estimation										✓
Confidence in MTD estimate		✓	✓	✓						✓

NOTE. Umbrella designs are in bold. By MTD estimation, we mean under general conditions. Only SPM achieves this. CRM will also achieve this but requires some added conditions and hence no tick.

Abbreviations: BLRM, Bayesian logistic reassessment method; BOIN, Bayesian Optimal Interval; CCD, cumulative cohort design; CRM, continual reassessment method; EWOC, escalation with overdose control; MTD, maximum tolerated dose; mTPI2, modified Toxicity Probability Interval; SPM, semiparametric dose finding method.

Summarium Fictus

- 3+3 Design is inferior
- IQ 3+3 inherits the 3+3 limitations
- Rolling 6, IQ Rolling 6 and other rule-based designs are similarly limited
- BOIN design is better per CCR
- BLRM is too conservative but can be improved with underdose control
- SPM – Semiparametric is best - JCO

Summarium Fictus

- **… Let us see if we can reverse this**

M. Clertant, JCO Oct, 2022, “Early-Phase Oncology Trials: Why so many designs?”

Since 1990 “there have been few years in which less than one new alternative approach has been published.”

What are the odds that each publication will claim their method is “best” by some metric?

What are the odds that the JCO review was used by Clertant to self-reference to claim his SPM method was best?

M. Clertant, JCO Oct, 2022, “Early-Phase Oncology Trials: Why so many designs?”

SPM is theoretical and the method is far from transparent.

Clertant had only one paper in PUBMED prior to the JCO review. A theory paper on SPM in Statistica Sinica IF 1.33. None since in PUBMED.

Zero clinical papers.

Superficial (incorrect) Summary

- 3+3 Design is inferior
- IQ 3+3 inherits the 3+3 limitations
- Rolling 6, IQ Rolling 6 and other rule-based designs are similarly limited
- BOIN design is better per CCR
- BLRM is too conservative but can be improved with underdose control
- ~~SPM – Semiparametric is best - JCO~~

BLRM with underdose control added to overdose control:

1) Technical: if 10mg, 25mg, 50mg and 100mg, with no DLTs, the authors complain that 200mg exceeds the overdose control rule so too conservative. Would a 40% increment have been allowed?
2) **“Simulation results reveal that new designs have better accuracy and treat more patients at the MTD”..** Why is it good to treat more patients at the MTD? ..how would the PI have an expectation that the pre-specified DLT target identifies a good RP2D? Does accuracy on an arbitrary target matter? There are no clinical data to support any target DLT-rate”
3) Is underdose control really needed by statistical rules?

(questioned in Stat in Med Nov 2022, P. Frankel, E. Garrett-Mayer, M.Krailo)

Superficial (incorrect) Summary

- 3+3 Design is inferior
- IQ 3+3 inherits the 3+3 limitations
- Rolling 6, IQ Rolling 6 and other rule-based designs are similarly limited
- BOIN design is better per CCR
- ~~BLRM is too conservative but can be improved with underdose control~~
- ~~SPM – Semiparametric is best - JCO~~

Comments on Conaway and Petroni, CCR:

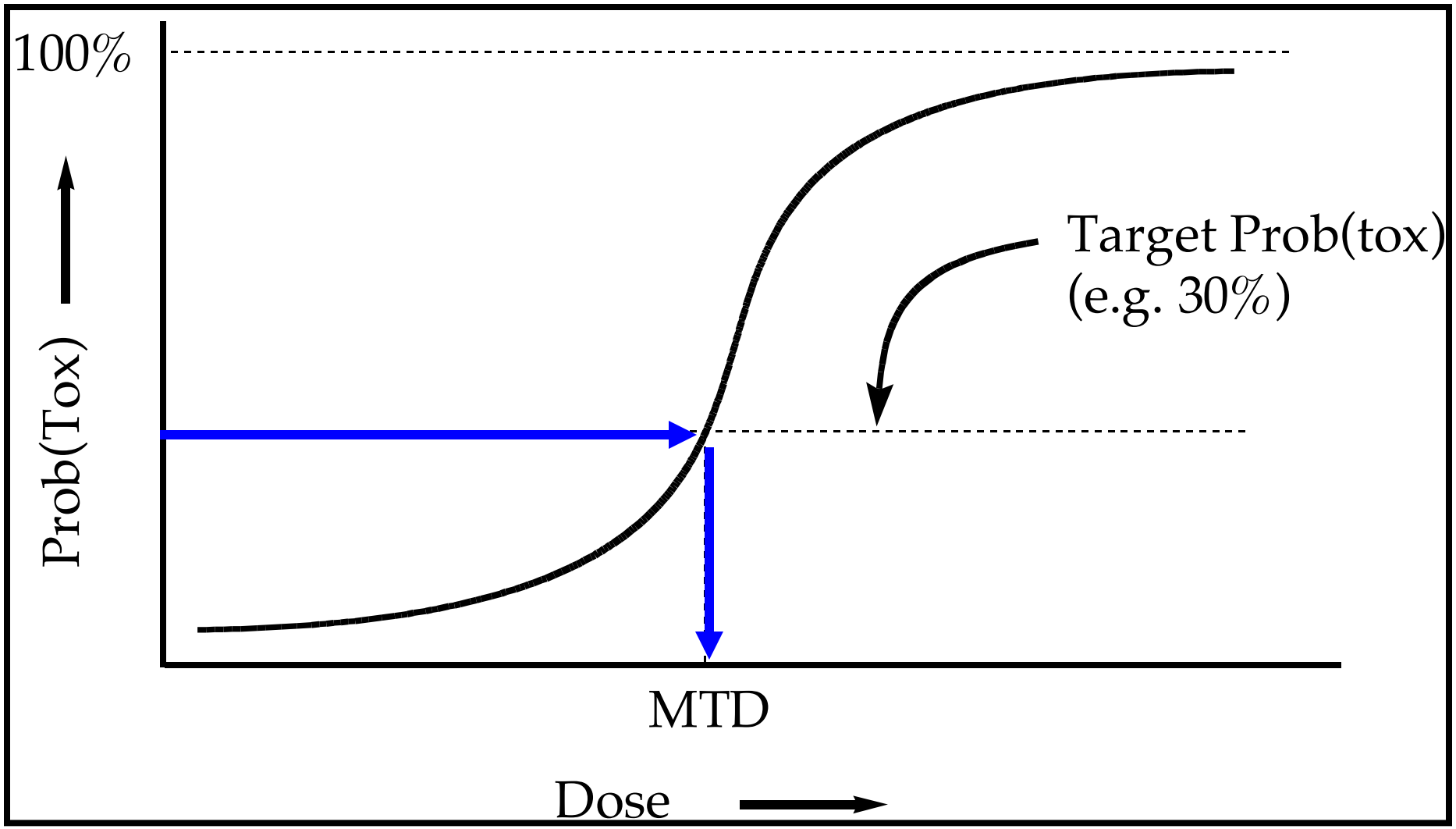
- **Mark Ratain**: “Simulation relies on outdated paradigms… The authors assume that as toxicity increases, efficacy increases”.
- If one **assumes** more toxicity means more activity, then designs that target more toxicity will have more activity. One doesn’t need simulations to know the pre-determined result.
- **Beat Neuenschwander** (author of the original BLRM method) voiced similar concerns.
- **FDA Project Optimus:** “When less is more”.

“ Sponsors should carefully evaluate exposure-response, efficacy, and safety data from early trials to inform dose selection, rather than automatically selecting the maximum tolerate dose”.

Superficial (incorrect) Summary

- 3+3 Design is inferior
- IQ 3+3 inherits the 3+3 limitations
- Rolling 6, IQ Rolling 6 and other rule-based designs are similarly limited
- ~~BOIN design is better per CCR~~
- ~~BLRM is too conservative but can be improved with underdose control~~
- ~~SPM – Semiparametric is best - JCO~~

MTD Definition for Risk-Targeting Designs



Informed Consent?

Ethics and Clinical Research: Improving Transparency and Informed Consent in Phase I Oncology trials (JCO April 2023).

Co-authors: P Frankel, S Groshen, J Beumer, Laura Cleveland, E Kim, J Karp.

Jesse Gelsinger

17yr with an ornithine transcarbamylase (OTC) deficiency causing problems breaking down ammonia. Lived on non-protein diet and controlled his OTC “fairly well”. Enrolled on a gene therapy clinical trial using an adenovirus to deliver the normal OTC gene into his liver. He died 4 days later on 9/17/1999 at University of Pennsylvania.

Jesse Gelsinger

- **Main issue still discussed 20 years later:**
  - **Failure to appropriately disclose risks in the informed-consent documentation.**

- **Effect: The field of gene therapy collapsed for at least a decade. Penn's Institute for Human Gene Therapy was shut down.**

Ethics as a Guide

- Are patients told when the goal is to find the dose (and treat as many as possible) where 1 in 4 will experience an especially concerning severe or life-threatening adverse event in the first 28 days despite dose modification rules?
- Is it documented as part of the informed consent process?
- How does this relate to the consortium?

Actual Summary

- 3+3 Design sets Traditional Risks
- IQ 3+3 inherits the 3+3 Risk-limits saving 20% on study duration
- IQ Rolling 6 similarly accelerates Rolling 6
- Risk-targeting designs are fine if decisions align with physician goals and are acceptable to the subject
- Should be written as a guide, not a rule (other than upper limit on decisions).
- Reminder: Peer-review is no guarantee of quality