

Introduction to Dose Finding Designs

Jay Park

jay@coreclinicalsciences.com

Core Clinical Sciences & McMaster University

February 27, 2025



Learning Objectives

1. To review clinical development process, different clinical trial phases, and terminology for dose-finding designs

2. To gain a conceptual understanding of phase 1 dose-finding designs:
 - 3+3 design
 - Bayesian Optimal INterval (BOIN) design

3. To gain a conceptual understanding of phase 2 dose-finding designs:
 - Multiple Comparisons Procedure - Modelling (MCP-Mod)



Prior presentations for UBC RALS



Prior presentations

1. Presentation from November 7, 2022
2. Presentation from March 5, 2024

These presentations introduce the concept of adaptive trial designs and simulation-guided design, existing standards and reporting guidelines, and practical considerations.

Today I will build on these presentations and discuss phase 1 and phase 2 dose-finding trials



Clinical development

What is clinical development?

- Clinical development, also known as **drug development**, is a blanket term to describe the process of bringing new medical products (e.g., drugs, device, vaccines and etc) to the market
- It includes *drug discovery / product development; preclinical research; and clinical trials*
- The first two topics are beyond the scope of this presentation
- Today we will be talking about clinical trials that are conducted on humans

Overview of clinical trials for clinical development

- Clinical trials are commonly divided into 3 phases (categories)
- The 3rd phase aims to confirm the favorable benefit shown in earlier trials
 - These trials are also known as **confirmatory** or **pivotal** trials that are typically used for regulatory decision making
- The 1st phase aims to estimate the highest dose that can be administered with an acceptable level of toxicity (also referred to as **maximum tolerated dose** or **MTD**)
- The 2nd phase aims to achieve two things:
 1. Determine whether a phase 3 trial is warranted for the product
 2. Identify the most promising dose or doses that ought to be tested in phase 3 trials



Focus for today

- Our focus for this presentation will be on the first 2 phases of clinical development and dose selection
- First 2 phases of trials are commonly referred to as exploratory trials, since regulatory approval decision making *typically* requires phase 3 trial(s) to be conducted.
- Ultimately, the goal of these exploratory trials is to select the right dose(s) to advance into confirmatory (i.e. phase 3) investigation



Challenges with dose selection

- Selection of the right dose is one of the most challenging decisions during clinical development.
- Selecting too high of a dose can result in unfavourable toxicity profile.
 - Manufacturing output capacity is finite. Too high of a dose could mean less number of patients being treated as a result
- Selecting too low of a dose decreases the chance of demonstrating efficacy in phase III trials.
 - We may miss out on a new drug that can offer benefits



Phase 1 trials

Why do we conduct phase I trials

- Typically, phase 1 trials are conducted to estimate the MTD (maximum tolerated dose)
- MTD is defined as the highest dose that can be administered within an acceptable level of toxicity
- Any doses higher than the MTD are “too” toxic to administer



MTD - Maximum tolerated dose

- To determine the MTD, we need to define:
 - Dose limiting toxicity (DLT): Protocol-specific adverse event definitions and their time frame
 - Target toxicity level: Rate of acceptable level of DLT



Dose limiting toxicity (DLT): An illustrative example

- A dose limiting toxicity (DLT) is defined as any of the following Agent X-related adverse event (AE) that occurs during the DLT period, graded according to the NCI Common Terminology Criteria for Adverse Events (CTCAE), Version 4.0:
 - Grade 4 neutropenia lasting ≥ 7 days;
 - Grade 3 or 4 neutropenia complicated by fever $\geq 38.0^{\circ}\text{C}$ or infection;
 - Grade 4 thrombocytopenia;
 - Grade 3 thrombocytopenia complicated by hemorrhage;
 - Grade 3 or 4 anemia; and etc

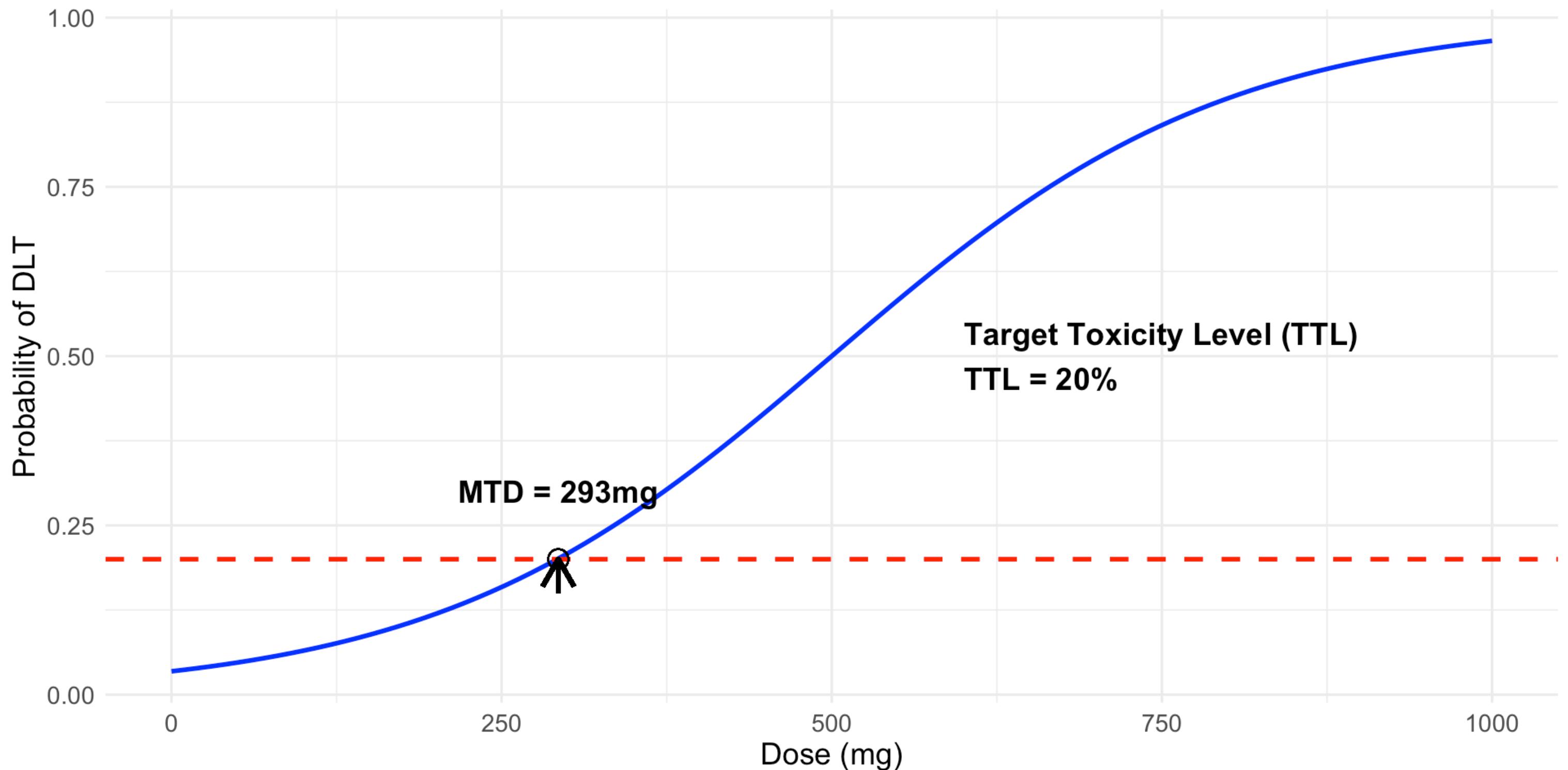


Target toxicity level

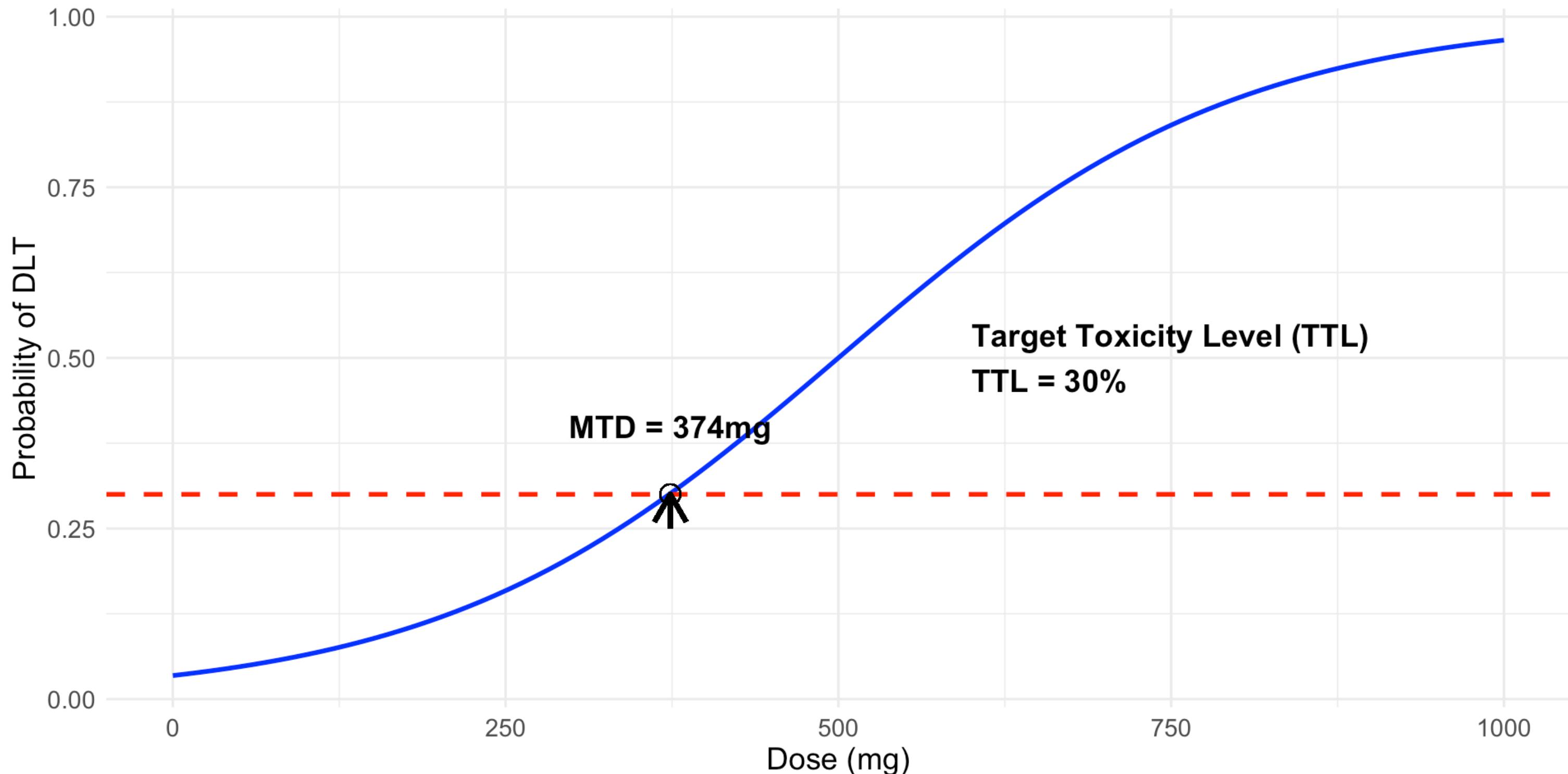
- Target toxicity level depends on the disease, treatment under investigation, target population, and likely associated AEs included in the definition of DLT
- Typically determined by clinical expertise, evidence from existing studies, and guidance from trial statistician
- Often set between 0.20 and 0.33



Graphical illustration of MTD: 0.20 target toxicity level



Graphical illustration of MTD: 0.30 target toxicity level



Starting point of phase 1 trial design

As a starting point, we need to determine the following:

- Dose limiting toxicity (DLT): Protocol-specific adverse event definitions and their time frame
- Target toxicity level (TTL): Rate of acceptable level of toxicity
 - Target DLT rate - usually between 0.20 and 0.33
 - 0.33 has historical precedent but not a strong scientific rationale (at least in my opinion)
- A range of dose to be explored (lowest to highest dose)
 - Determined based on clinical input, preclinical data, and etc.
- Total sample size, number of doses, and number of patients per dose
 - In oncology, usually 20-30 patients in total and 3 patients per cohort



Phase 1 designs

- Typically, we start with the lowest dose, then we observe for the rate of these protocol-specific adverse events
- Escalate the dose if shown to be tolerable
- Continue to escalate until the maximum tolerable dose
- De-escalate (if permitted) or stop the trial if the dose shown to be intolerable
 - Traditional phase 1 trials would allow the doses to be escalated, de-escalation was often not permitted



Traditional 3 + 3 design

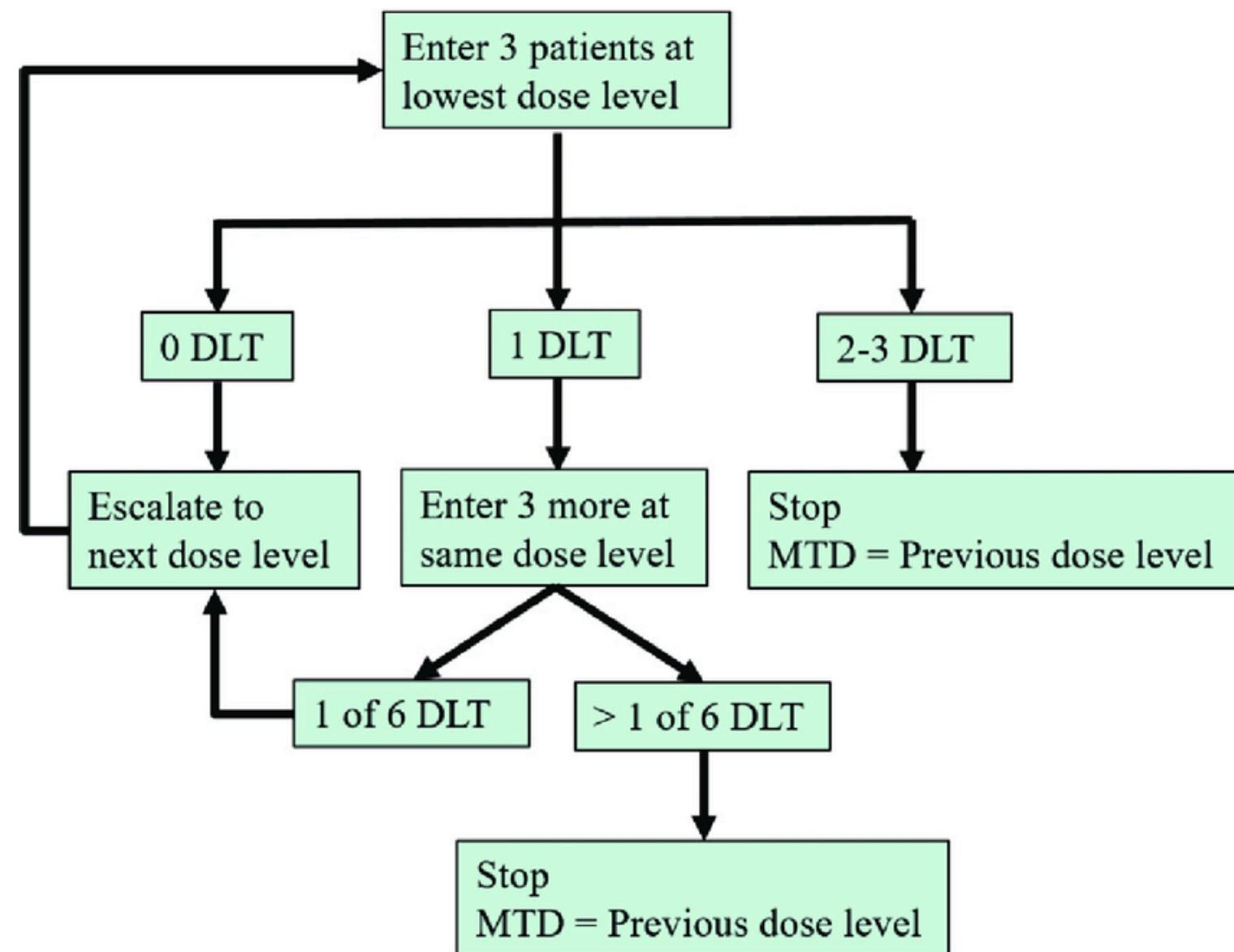


3 + 3 design

- Most common approach for phase 1 design
- It's a simple rule-based design
- Initially enroll 3 patients to the lowest dose and observe for DLT
 - If no DLT, escalate to the next dose with 3 patients
 - If one patient has a DLT, we add 3 more patients to the same dose
 - If no more DLT, we escalate to the next dose
 - If any DLT, we stop dose escalation
 - If more than 2 DLTs observed, we stop dose escalation



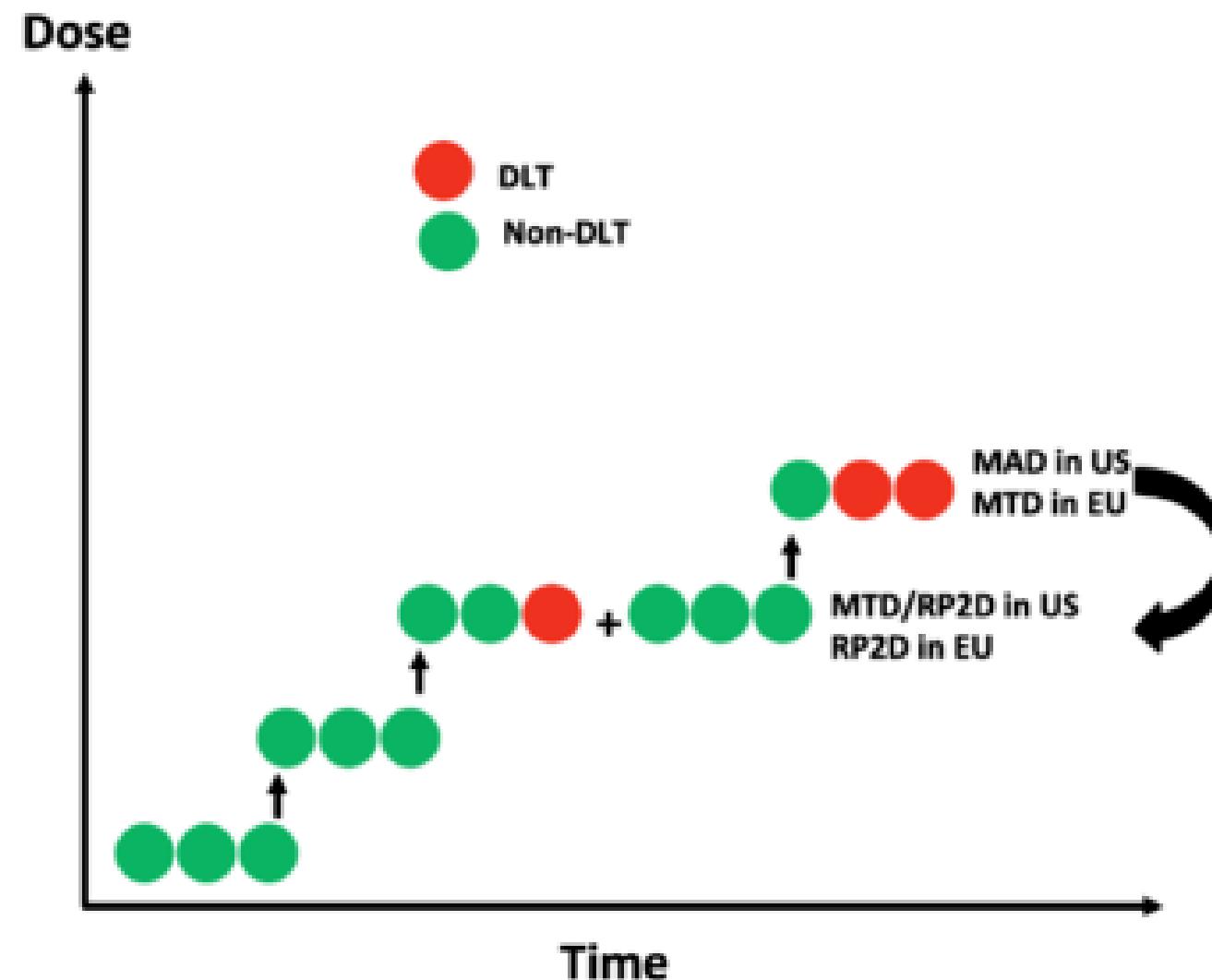
3 + 3 design - continued



- Initially enroll 3 patients to the lowest dose and observe for DLT
 - If no DLT, escalate to the next dose with 3 patients
 - If one patient has a DLT, we add 3 more patients to the same dose
 - If no more DLT, we escalate to the next dose
 - If any DLT, we stop dose escalation
 - If more than 2 DLTs observed, we stop dose escalation

3 + 3 schema

A Standard 3 + 3 Design



- Each circle represents a patient. Green means no DLT and red means DLT
- 1. No DLT in the 1st (lowest) dose
- 2. No DLT in the 2nd dose
- 3. 1 DLT in the 3rd dose triggering another cohort being assigned
- 4. 2 DLTs in the 4th dose
- In this case, 3rd dose would be determined as the MTD

Problems with 3 + 3 design

- 3+3 design is a memory-less design
- When determining the MTD, it doesn't account for what happened in the previous cohorts
- There are many other model-based designs that have shown better statistical properties than the 3+3 design
 - It is difficult to cover every alternate options today. I will be going over the Bayesian Optimal INterval (BOIN) design



Bayesian Optimal INterval (BOIN) design

- A model-assisted design

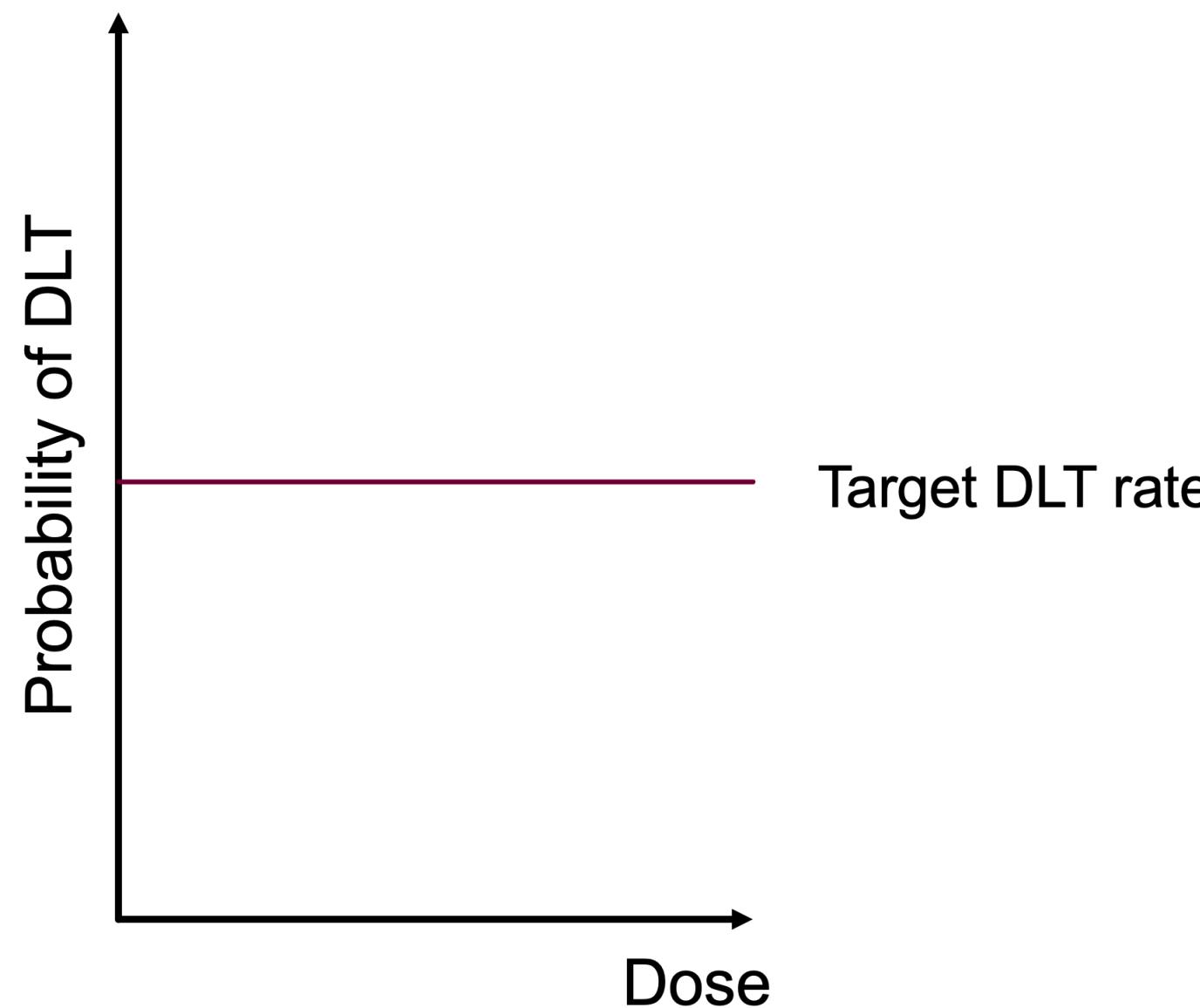


Background on BOIN

- BOIN is a model-assisted design that can be easier to implement than other model-informed designs
- In 2021, the FDA has granted the BOIN design the fit-for-purpose designation (see the letter [here](#)).
- FDA's FFP is program that provides a pathway for regulatory review and acceptance of quantitative tools for drug development (see link for FFP [here](#))



BOIN intro



- With the 3 + 3 design, we are setting the target toxicity (DLT) rate at 0.33 (1/3)
- With BOIN, we can set the target DLT rate to be any rate we want

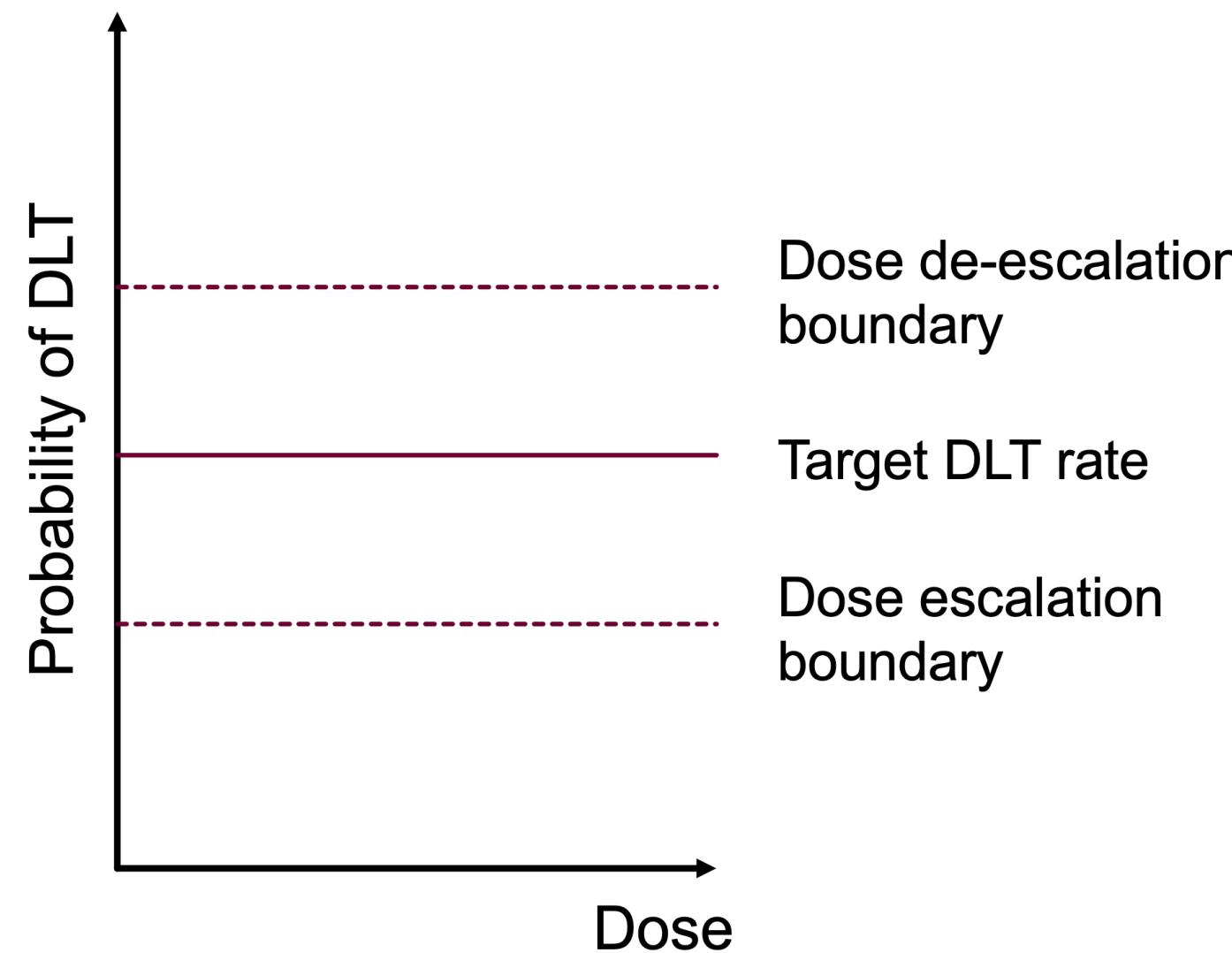
BOIN boundaries

In addition to target DLT rate, we need to specify:

- **Lowest toxicity rate** - doses below this rate would be considered as sub-therapeutic
- **Highest toxicity rate** - doses above this rate are considered to be excessively toxic
- With these rates, we calculate our dose-escalation and dose-de-escalation boundaries.
 - The recommended values for lowest and highest toxicity rates are 0.6 and 1.4 times of the target DLT rate

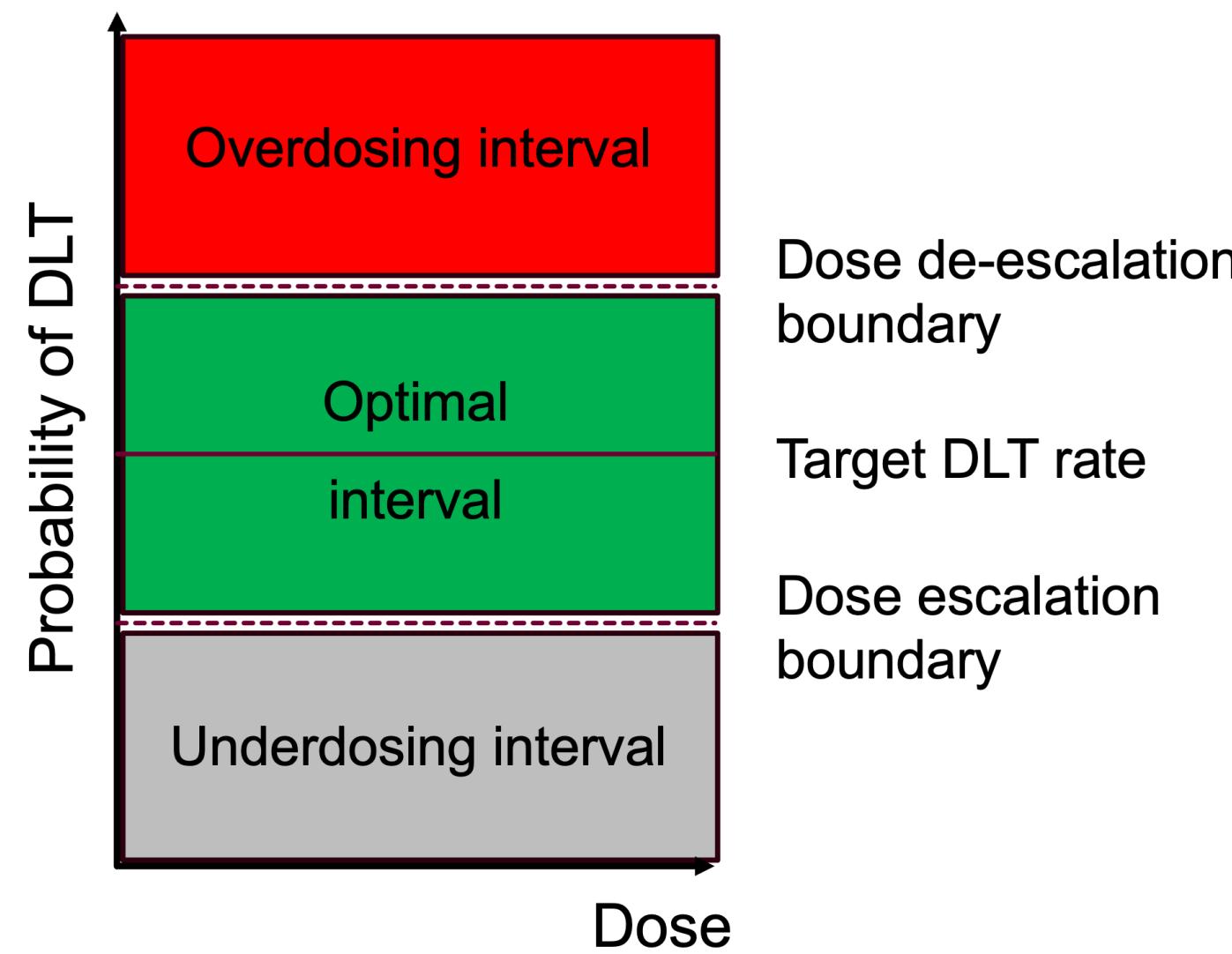


BOIN boundaries



- We want doses **above** the dose-escalation boundary
- We want doses **below** the dose-de-escalation boundary
- Our optimal dose is defined as dose that's between the dose-escalation and dose-de-escalation boundaries

BOIN boundaries - continued



- With these boundaries, we end up creating three intervals:
 - Overdosing interval;
 - Optimal interval; and
 - Underdosing interval
- Escalate if within the underdosing interval & de-escalate within the overdosing interval
- The optimal interval is defined by the dose de-escalation (upper) and dose-escalation (lower) boundary

General implementation of BOIN

1. Treat patients in the first cohort at the pre-specified starting dose level (*doesn't have to be the lowest dose here*)



General implementation of BOIN

1. Treat patients in the first cohort at the pre-specified starting dose level (*doesn't have to be the lowest dose here*)
2. Determine if the stopping rule for safety is met
 - If yes, eliminate the current and higher dose if the dose is shown to be overly toxic
 - If a dose is eliminated, we automatically de-escalate the dose to the next lower level.
 - If the lowest dose is eliminated, the trial is stopped for safety with no dose being selected as the MTD.



General implementation of BOIN

1. Treat patients in the first cohort at the pre-specified starting dose level (*doesn't have to be the lowest dose here*)
2. Determine if the stopping rule for safety is met
 - If yes, eliminate the current and higher dose if the dose is shown to be overly toxic
 - If a dose is eliminated, we automatically de-escalate the dose to the next lower level.
 - If the lowest dose is eliminated, the trial is stopped for safety with no dose being selected as the MTD.
3. (*If stopping rule for safety is not met*) compute the DLT rate at the current dose



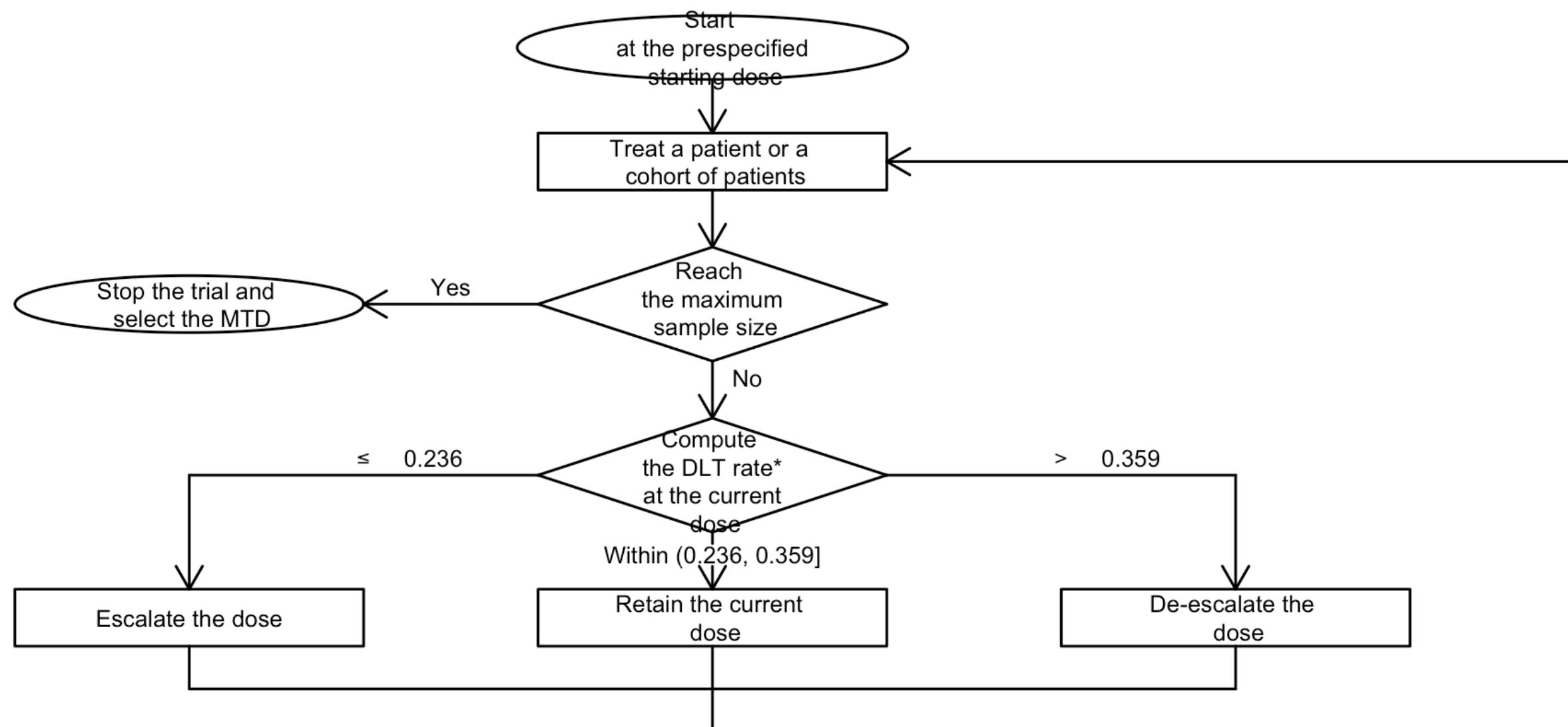
General implementation of BOIN

1. Treat patients in the first cohort at the pre-specified starting dose level (*doesn't have to be the lowest dose here*)
2. Determine if the stopping rule for safety is met
 - If yes, eliminate the current and higher dose if the dose is shown to be overly toxic
 - If a dose is eliminated, we automatically de-escalate the dose to the next lower level.
 - If the lowest dose is eliminated, the trial is stopped for safety with no dose being selected as the MTD.
3. (*If stopping rule for safety is not met*) compute the DLT rate at the current dose
4. Retain, escalate and de-escalate the dose according to the rules established based on overdosing, optimal, and underdosing intervals.
5. Repeat the processes above until the maximum sample size is reached or earlier decision is reached.



BOIN design flowchart

- Let's assume our target DLT rate is 0.30. We will recruit 10 cohorts with cohort size of 3 (30 in total)



$$* \text{ DLT rate} = \frac{\text{Total number of patients who experienced DLT at the current dose}}{\text{Total number of evaluable patients treated at the current dose}}$$

Escalation and de-escalation rules

Escalation and de-escalation rules			
Number of patients treated	Escalate if # of DLT <=	Deescalate if # of DLT >=	Eliminate if # of DLT >=
3	0	2	3
6	1	3	4
9	2	4	5
12	2	5	7
15	3	6	8
18	4	7	9
21	4	8	10
24	5	9	11
27	6	10	12
30	7	11	14

- We start by treating first cohort of 3 patients and escalate or de-escalate accordingly.
- Similar to the 3 + 3 design, easy to implement

BOIN simulation case study

Background information

- Goal: Determine the MTD out of 5 dose tiers
 1. 3.125 mg twice daily
 2. 6.25 mg twice daily
 3. **12.5 mg twice daily: The starting dose of the trial**
 4. 25 mg twice daily
 5. 50 mg twice daily

We expected that Dose Tier 3 would be tolerable based on existing human data



Sample size by cohort size and number of cohorts

Cohort Size	Total # of cohorts	Total N
3	10	30
3	15	45
3	20	60

Trial design assumptions

Assumptions	Values
Target DLT rate	0.150
Escalation (lower) boundary	0.118
De-escalation (upper) boundary	0.179
Dose range	5 dose tiers
Dose elimination cut-off	0.950



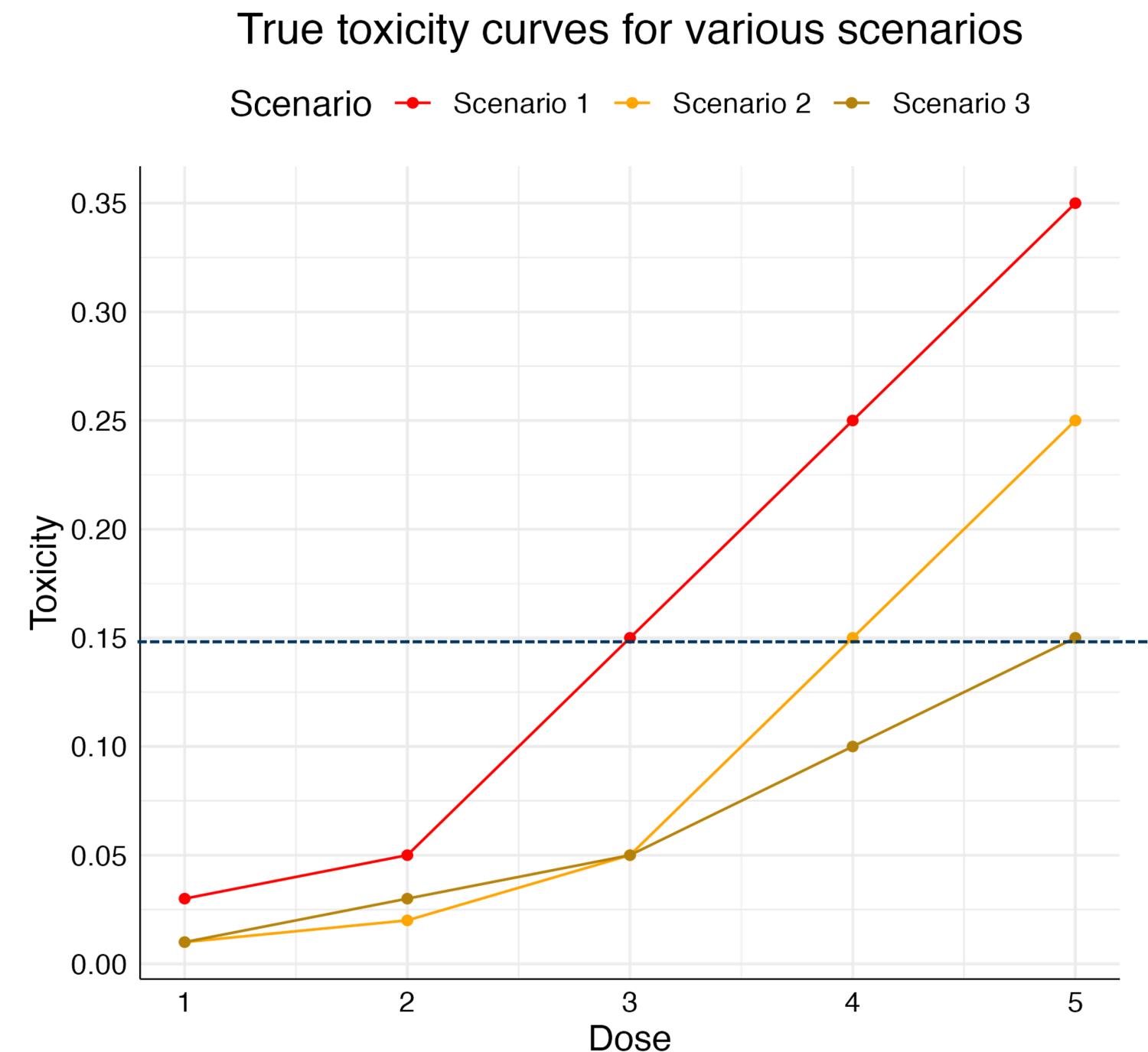
Simulation scenarios

Scenarios	Dose 1	Dose 2	Dose 3	Dose 4	Dose 5
Scenario 1	0.03	0.05	0.15	0.25	0.35
Scenario 2	0.01	0.02	0.05	0.15	0.25
Scenario 3	0.01	0.03	0.05	0.10	0.15



Simulation scenarios - continued

Visually, our scenarios look like this



Simulation results - 30 patients

Simulation results for 30 patients

Operating Characteristics		Dose 1	Dose 2	Dose 3	Dose 4	Dose 5
Scenario 1	True DLT Rates	0.03	0.05	0.15	0.25	0.35
	DLT selection %	2.40	21.40	45.80	25.50	4.50
Scenario 2	True DLT Rates	0.01	0.02	0.05	0.15	0.25
	DLT selection %	0.50	1.50	21.90	48.90	27.20
Scenario 3	True DLT Rates	0.01	0.03	0.05	0.10	0.15
	DLT selection %	0.60	2.10	10.90	32.50	53.90

- Probability of selecting the target DLT ranges from 45.80% to 53.90%

Simulation results - 60 patients

Simulation results for 60 patients

Operating Characteristics		Dose 1	Dose 2	Dose 3	Dose 4	Dose 5
Scenario 1	True DLT Rates	0.03	0.05	0.15	0.25	0.35
	DLT selection %	2.00	25.10	54.60	17.00	0.70
Scenario 2	True DLT Rates	0.01	0.02	0.05	0.15	0.25
	DLT selection %	0.10	2.10	22.20	61.00	14.60
Scenario 3	True DLT Rates	0.01	0.03	0.05	0.10	0.15
	DLT selection %	0.20	1.90	6.50	33.00	58.40

- Probability of selecting the target DLT improved to 54.60% to 58.40%

Comments on the case study

- For the actual project, we explored other scenarios and questions (results not shown)
 - What happens if we targeted a higher DLT rate (e.g., 0.200 instead of 0.150)
 - Different cohort sizes: 3, 6, and 9
 - Different number of cohorts: 10, 15, and 20



Comments on BOIN

- Having the fit-for-purpose designation from the FDA really helped with the uptake of this design
- There are different variations of BOIN, and there are many other dose-finding design methods that have been shown to outperform the 3 + 3 design.



Phase 2 Dose Finding Trials

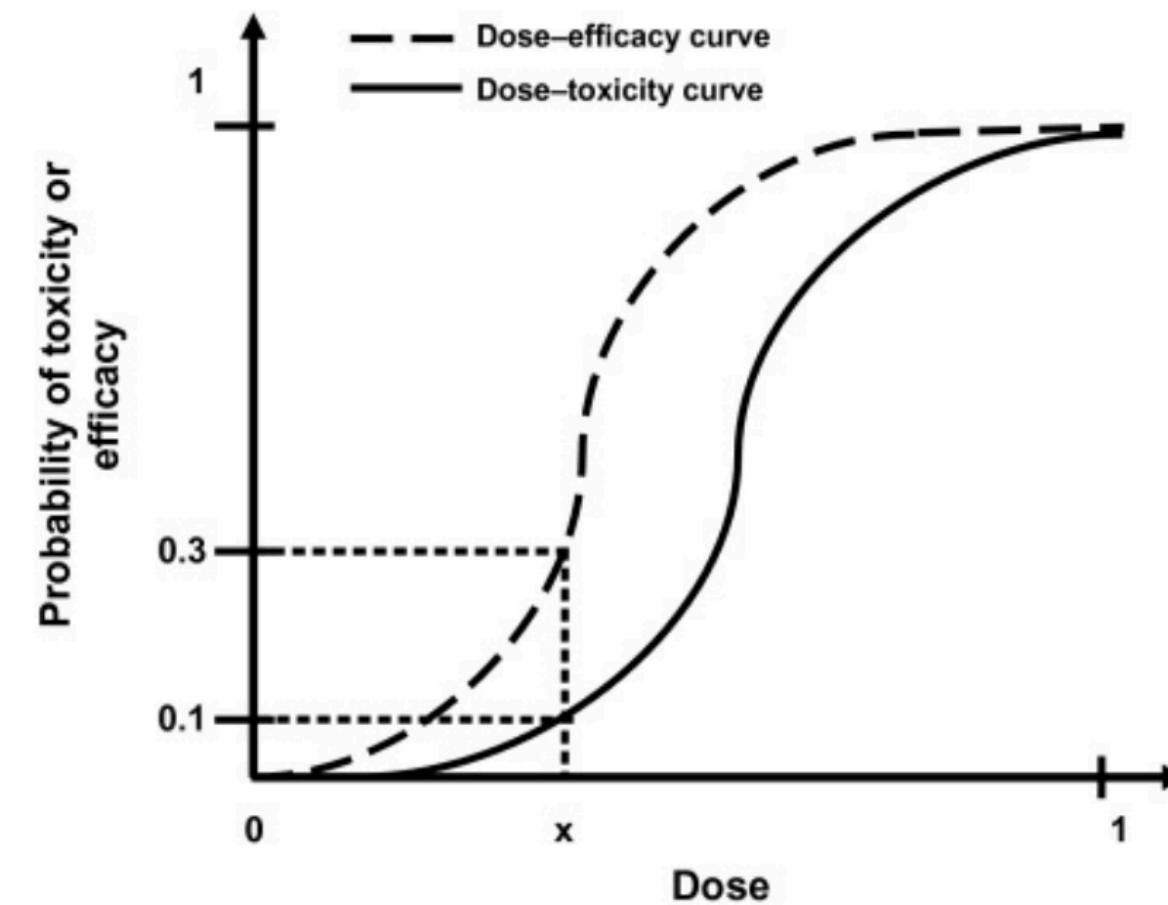


Overview of phase 2 trials

- Phase 2a trials, also known as proof-of-concept studies, aim to determine if one or more doses could demonstrate efficacy over placebo
 - If proof of concept can be demonstrated, then we typically will proceed to the phase 2b trial
- Phase 2b trials, also known as dose-response studies, test several doses with a goal of selecting a dose or doses that should be tested in a phase 3 trial



What do we mean by dose-finding study in a phase 2 setting?



- Think of phase 1 trials as dose-toxicity studies and phase 2 dose-finding studies as dose-response studies
- Dose-response studies are also referred to as dose-ranging studies, where we test a range of different therapeutic doses (that are lower than the MTD)
- Dose-response studies will often have two or more active doses + placebo, with patients being randomly assigned to one

What are we trying to achieve with our phase 2 dose-response studies?

Two broad goals:

1. Determine whether further development, such as a phase 3 trial, is warranted. Here, we want to demonstrate proof of concept
2. Determine what dose or doses should be tested in our phase 3 trial



Analysis of phase 2 dose-response studies

- Statistical analysis of dose-response studies can be divided into two strategies:
 1. Multiple comparison procedures (MCP)
 2. Model-based approach (MoD) to characterize the dose response relationship
- Different designs either do multiple comparison procedures (MCP), model-based approach (MoD), or hybrid approach that combines MCP + Mod together
- Our focus today will be on the hybrid approach, MCP-MoD

Multiple comparison procedure (MCP)

- We wish to identify the **minimum effective dose** that is statistically significant and produces important treatment effect in comparison to the placebo
- We treat different doses as a **qualitative factor**. Each active is compared against the placebo.
- **Minimum of 2 active doses required for the MCP step**



Multiple comparison procedure (MCP) - continued

- There are multiple comparison procedures (e.g., comparisons of contrasts) that can preserve the family wise error rate
 - Family wise error rate is the probability of discovering at least one false positive finding when we make multiple comparisons



Limitation of multiple comparison procedure (MCP)

- Our inference is limited to the dose levels that are tested - recall here we are treating different doses as a qualitative factor
- Modelling the dose-response relationship can mitigate this limitation



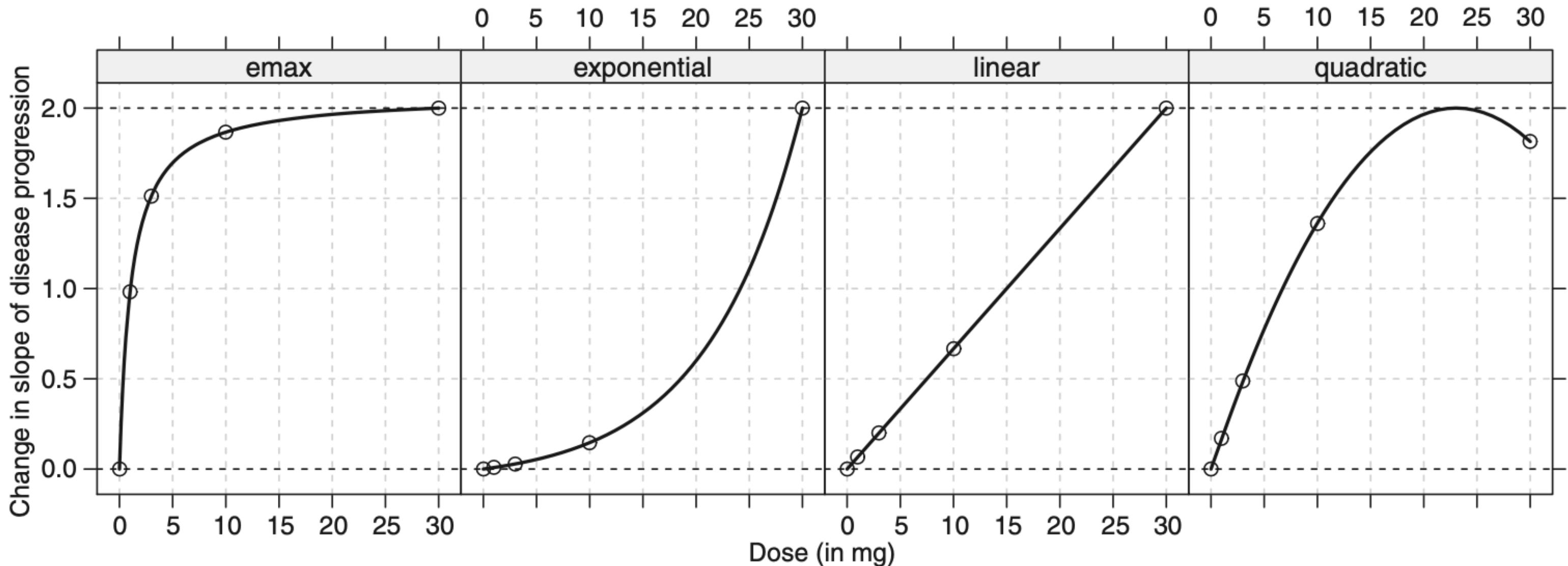
Model-based approach (MoD)

- We treat dose levels as a **quantitative factor**
- We assume there is a functional relationship between the dose levels and response under dose response model(s)



Dose-response models

By dose-response models, we just mean plausible shapes of dose-response curves



- We can work with content experts to come up plausible shapes of dose-response curves (like above).

Dose-response models - continued

- Emax model (shape) assumes that response rate increases quickly at the lowest dose but plateaus
- Exponential shape assumes the response increases exponentially with the dose levels
- Linear shape assumes a linear relationship between dose levels and response (might be unlikely)
- Quadratic shape assumes the response decreases after the optimal dose level



MCP-Mod (Multiple Comparison Procedure - Modelling) methods



MCP-Mod Background

- MCP-Mod is a hybrid approach that combines multiple comparison and modelling into a single trial design
- European Medicines Agency (EMA) has issued a qualification opinion in 2014 that states the MCP-Mod is an efficient statistical methodology for model-based design and analysis of Phase 2 dose-finding studies.
- FDA has granted the MCP-Mod fit-for-purpose (FFP) designation as an adequate and appropriate method for guiding dose selection for Phase 3 testing in 2016 (see the letter [here](#))



Planning for MCP-Mod

Like other dose-finding designs, we need to come up with:

- Endpoint: Can be continuous, binary, count, and time-to-event
- Desired power and type I error rate
- Dose range: minimum and maximum dose amount
- Number of doses
 - Need at least 2 active doses for MCP
 - Need at least 3 active doses for MoD, but generally recommended having 4 to 7 active doses
- Total sample size and how that will be allocated between active doses + placebo
- Dose response models based on plausible dose-response shapes



General flow of MCP-Mod procedure

- Perform multiple comparison first then do the modelling (if required)

General flow of MCP-Mod procedure - continued

- Perform multiple comparison first then do the modelling (if required)
- For our multiple comparisons, we use the dose response curves (models) that we pre-specified at the design in our analyses
 - These dose response models are used to maximize the power of our statistical tests
 - We derive contrast weights (coefficients) that maximize the power for each dose response curve
 - Note each comparison is specific to each dose response model

General flow of MCP-Mod procedure - continued

- Perform multiple comparison first then do the modelling (if required)
- For our multiple comparisons, we use the dose response curves (models) that we pre-specified at the design in our analyses
 - These dose response models are used to maximize the power of our statistical tests
 - We derive contrast weights (coefficients) that maximize the power for each dose response curve
 - Note each comparison is specific to each dose response model
- Then if any of the comparisons is found to be significant, we proceed with the modelling part
 - We estimate the dose response and target dose based on the dose response model(s) that showed significance



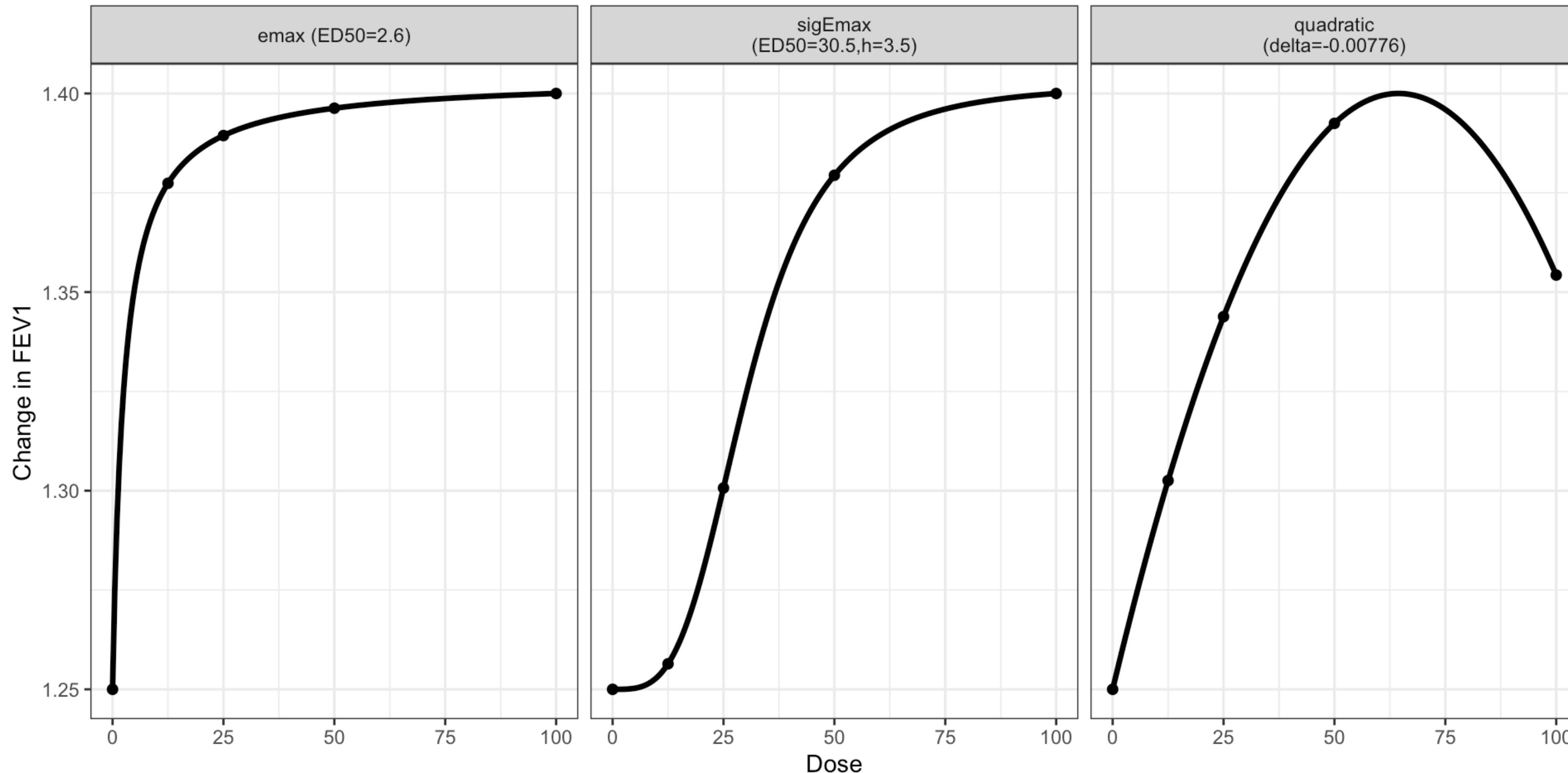
MCP-Mod case study

Case study background

- 5 dosing groups: 4 active doses + placebo
 - 0.0, 12.5, 25.0, 50.0, and 100 mg
- Primary endpoint: Change from baseline in forced expiratory volume in 1 second (FEV1)
- Approximately 260 patients randomly assigned to one of the 5 dosing groups



Plausible dose response models / shapes



MCP-Mod procedure in our case study

MCP stage

- For each of the dose-response models that we have specified, we come up with optimal contrast weights
- We conduct a global hypothesis testing for each of the dose-response model to see if there is a dose-response detected

Mod stage

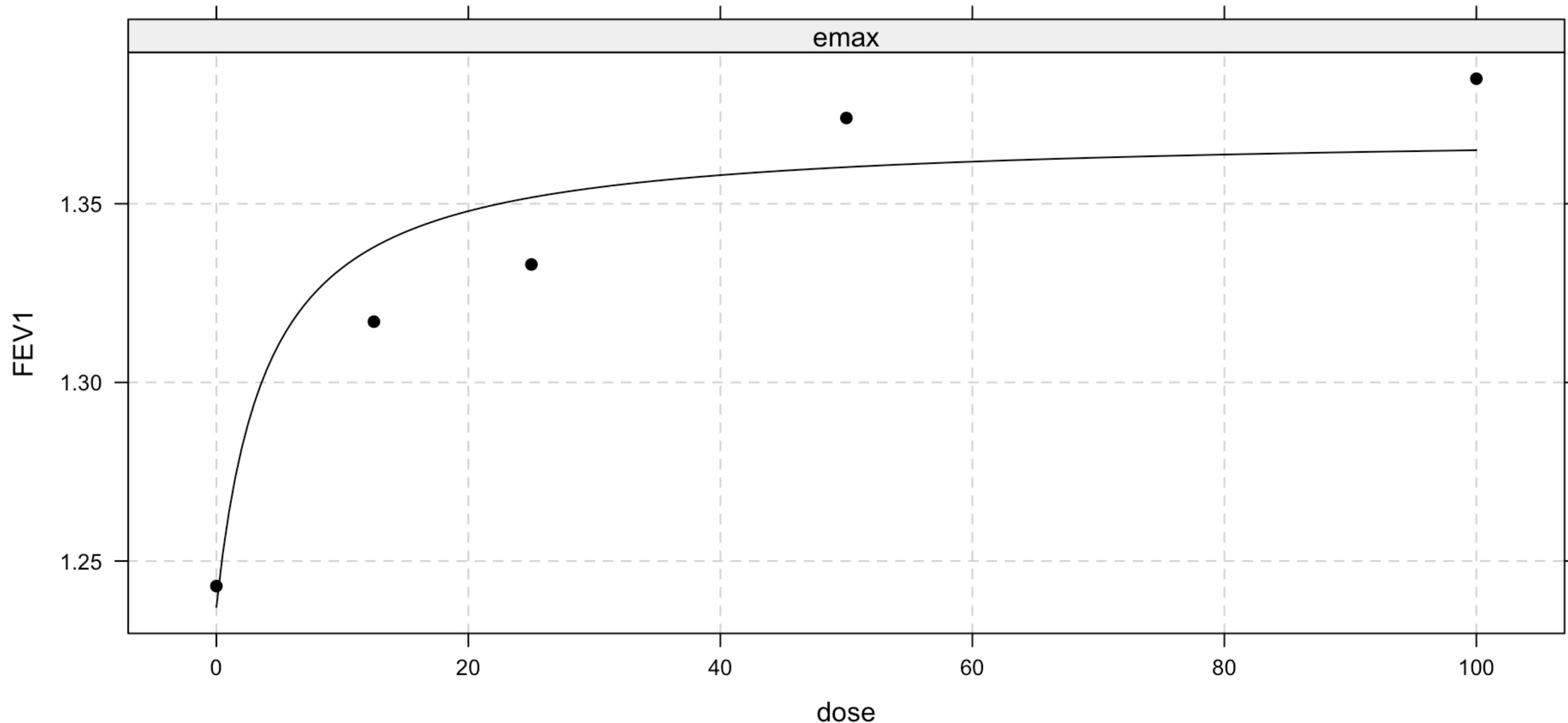
- Then we perform dose-response estimation in the dose-response model(s) that were found to be significant from the MCP-stage

Analysis stage - MCP

FEV1 results

Dose	Average FEV1	SD	Number of patients
0.0	1.243	0.0156	49
12.5	1.317	0.0145	55
25.0	1.333	0.0151	51
50.0	1.374	0.0148	53
100.0	1.385	0.0148	53

Modelling stage - dose-response estimation



Summary

Final comments

It's important to recognize how we characterize an optimal dose is different between different phases

- In phase 1 trials with 3 + 3 or BOIN (the example we went over today) designs, we are characterizing the optimal dose based on highest “serious” adverse events (MTD)
 - There are other examples of phase 1 designs that combine adverse events and efficacy into utility to determine “optimal biological dose”
- In phase 2 trials, such as MCP-Mod designs, we are characterizing the optimal dose based on efficacy and how it stands out against the control arm.



Final comments - continued

(If you have had a chance to look at recordings of my previous slides)

- Planning trials with simulations and collaboration makes the overall trial better, regardless of the final design you end up choosing.
- Utility of simulation-guided design for dose-finding studies is no different.
 - For new novel products, our phase 1 trial might be the first instance where we are exposing un-tested products to humans; and our phase 2 trial might be only the second time for exposure. So we really need to get it right.



References

1. Lee SM, Wages NA, Goodman KA, Lockhart AC. Designing dose-finding phase I clinical trials: top 10 questions that should be discussed with your statistician. *JCO precision oncology*. 2021 Jan;5:317-24.
2. Yuan Y, Hess KR, Hilsenbeck SG, Gilbert MR. Bayesian optimal interval design: a simple and well-performing design for phase I oncology trials. *Clinical Cancer Research*. 2016 Sep 1;22(17):4291-301
3. U.S. Food and Drug Administration. Drug Development Tools: Fit-for-Purpose Initiative
4. Bretz F, Hsu J, Pinheiro J, Liu Y. Dose finding—a challenge in statistics. *Biometrical Journal: Journal of Mathematical Methods in Biosciences*. 2008 Aug;50(4):480-504.
5. Pinheiro J, Bornkamp B, Bretz F. Design and analysis of dose-finding studies combining multiple comparisons and modeling procedures. *Journal of biopharmaceutical statistics*. 2006 Oct 1;16(5):639-56.

