

# *Decision-focused Applications of Modeling and Simulation (M&S) in CNS Drug Development*

William R. Gillespie  
Pharsight Corporation  
Cary, North Carolina, USA

FDA/CDER  
Rockville, MD  
September 10, 2004

# *Decision-focused Applications of Modeling and Simulation (M&S) in CNS Drug Development*

- Objective
  - Discuss the role & value of M&S to support decision-making in clinical development of CNS drugs
- Brief case studies that illustrate a variety of M&S applications in CNS drug development:
  - Integration of pre-clinical & clinical evidence to support a go/no-go decision (GAD)
  - Combined dose-response modeling + decision analysis for dose optimization (MDD)
  - Modeling & clinical trial simulation for decisions regarding PoC trial design & analysis (disease involving gradual neurological deterioration)
  - Modeling, clinical trial simulation & decision analysis in an adaptive clinical trial design (chronic psychiatric disorder)

## *A few qualifications regarding the examples*

- Examples are adapted from works commissioned by Pharsight clients.
- Examples are blinded and partially fictional to prevent disclosure of proprietary information & enhance their illustrative value
- The decision metrics and criteria reflect the values and economic motivations of those clients
- However, the same principles and methods may be adapted to reflect the values & priorities of other constituencies, e.g., patients, prescribers or regulators

# *Why modeling & simulation?*

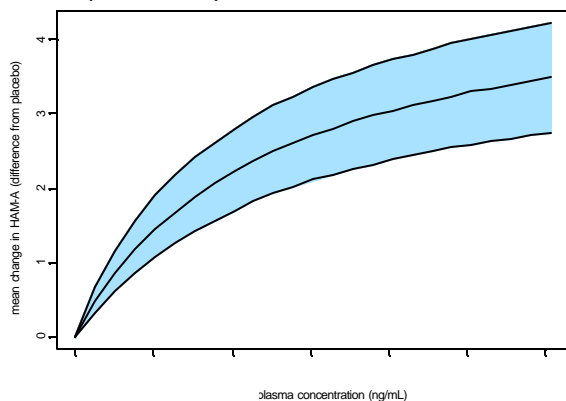
- M&S permits a quantitative synthesis of multiple types of evidence from multiple sources, e.g., information from public sources about other drugs, from in vitro and animal studies, etc.
- M&S provides a low-risk means for quantitative exploration of untested treatment regimens and novel trial designs.
- M&S permits quantitative evaluation of competing strategies from medical, business and ethical perspectives.

## *Example 1: Go/no-go decision for a potential GAD treatment in early Phase II*

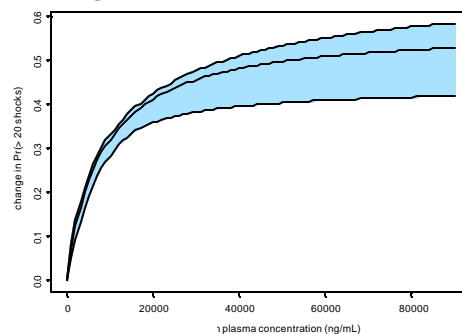
- Illustrative example of the practical application of modeling and simulation for decision-making in early clinical development.
  - Illustrates how models may be used to integrate/leverage prior information from multiple sources, e.g., clinical and nonclinical information from public literature, previous trials, related compounds/ analogues, competitors.
- NoGAD is being developed for treatment of GAD and other anxiety disorders
  - Preclinical studies suggest that NoGAD may be as effective as benzodiazepines without the undesirable side effects of sedation, cognitive and psychomotor impairment, alcohol interaction, and abuse/physical dependence liability.
  - Phase I studies confirm a low incidence of CNS depression-related AE's at multiple doses up to 20X mg BID.
  - Phase II study using 1X mg BID and 3X mg BID doses failed to provide convincing evidence of anxiolytic activity.
- Question posed by the NoGAD team:
  - Is the probability of a therapeutically useful anxiolytic response sufficiently great at higher doses  $\leq 10X$  mg BID to warrant further development?

# Modeling & simulation strategy

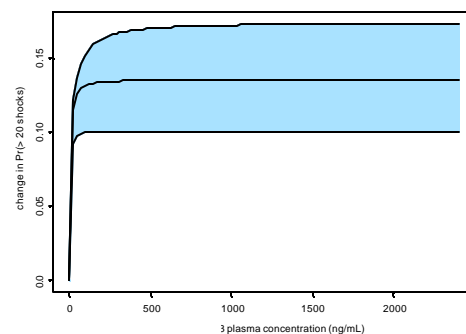
oldazepine clinical PK/PD  
(HAM-A)



oldazepine nonclinical PK/PD  
(Vogel lick suppression)

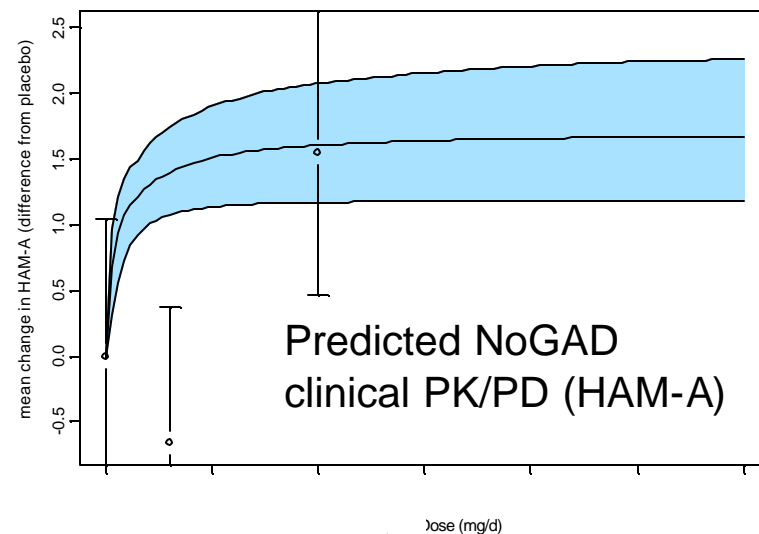


NoGAD nonclinical PK/PD  
(Vogel lick suppression)



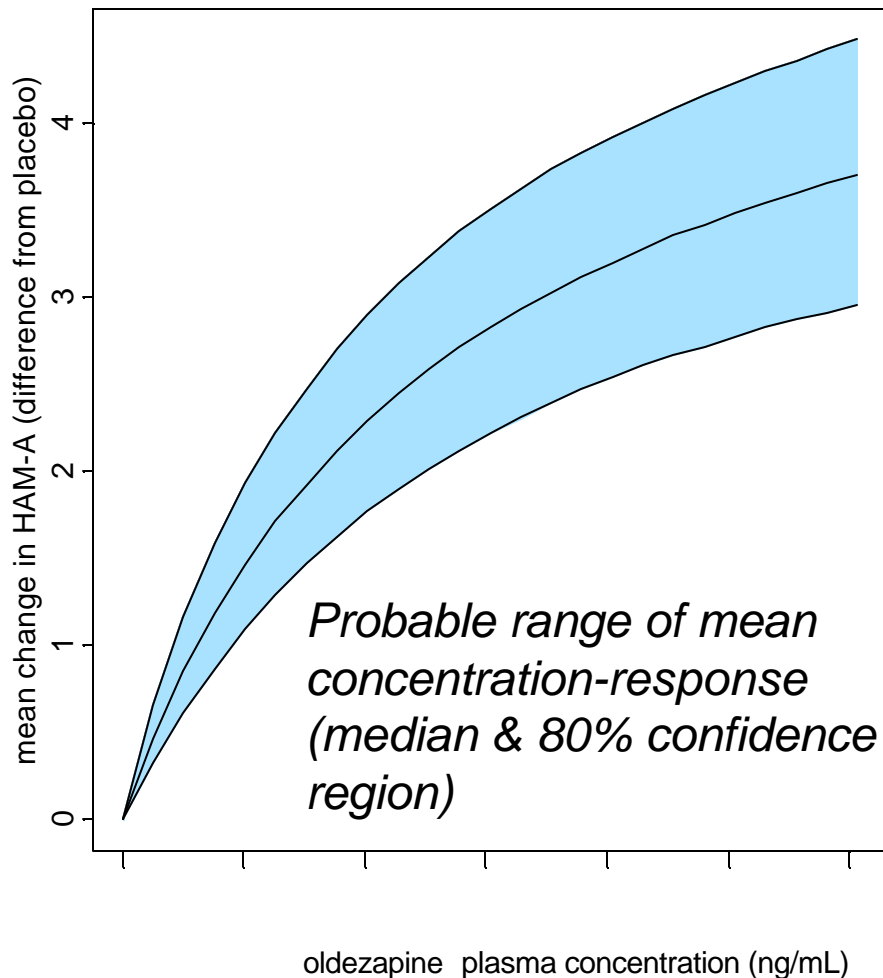
Rescale clinical model based on relative efficacy/potency in nonclinical models

- Construct a clinical dose-response model for oldazepine (another fictional name) in GAD patients
- Construct a non-clinical concentration-response models for oldazepine and NoGAD in the Vogel lick suppression test
- Model the preclinical→clinical relationship
- Predict the probable range of NoGAD dose-response (HAM-A) by combining those models, i.e., rescaling the clinical model based on relative potency and efficacy in the nonclinical models.

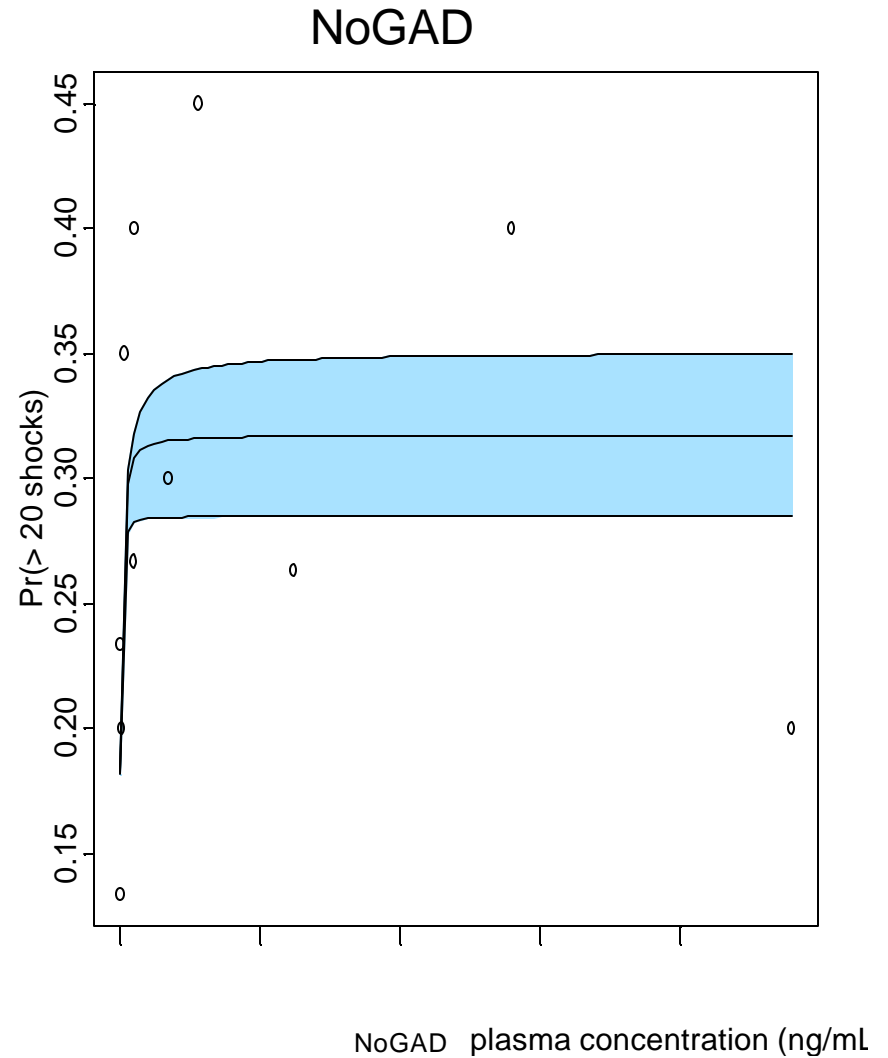
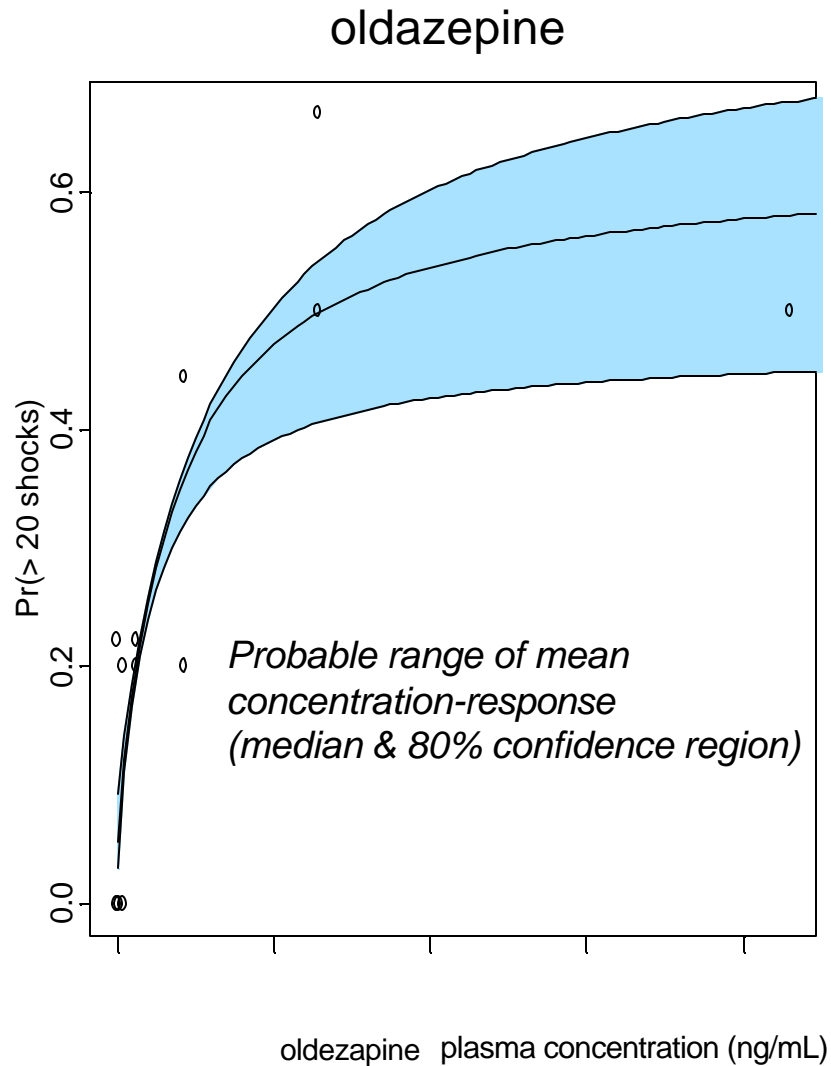


# Oldazepine: Modeling HAM-A in GAD patients

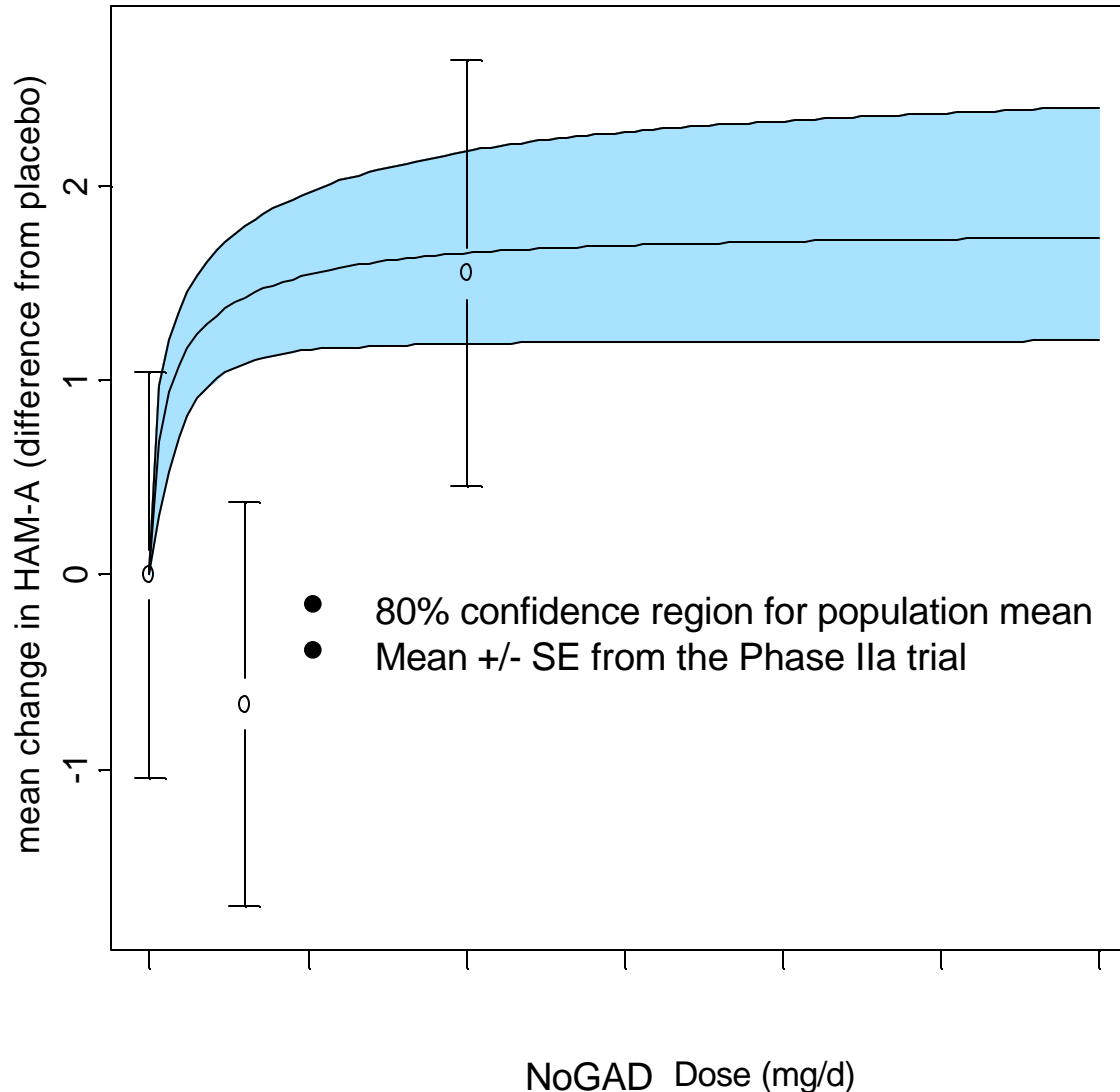
- ~1300 patients
- HAM-A at endpoint (4-6 weeks or last observation)



# Vogel lick suppression in rats in response to oldazepine and NoGAD



# NoGAD: Predicted clinical dose-response in GAD patients



## Conclusions

- The HAM-A scores based on the preclinical-to-clinical model are largely consistent with the observed data from the Phase IIa trial.
- Extrapolation of the model to higher doses such as 20X mg/d (10X mg BID) suggests that:
  - Minimal increases in efficacy are likely to occur.
  - Efficacy comparable to benzodiazepines is unlikely.
- Recommend termination of NoGAD development.

## *Example 2: Dose optimization for a potential antidepressant*

- Nuoxetine is an antidepressant candidate in late Phase II development
- Phase II PoC & dose finding studies have been completed
- The development team wishes to determine what, if any, nuoxetine daily dose:
  - Provides an optimal trade-off between favorable and unfavorable responses to the drug
  - Results in an equivalent or superior benefit/risk profile relative to paroxetine 20-40 mg/d
- The team proposed several treatment attributes that might be considered for dose selection:
  - Efficacy-related: HAMD change from baseline, % responders, % remissions, rate of onset
  - Adverse effects: sexual dysfunction, dropouts due to AE's, anxiety, agitation, weight gain
  - Other: required dosing frequency, drug interactions

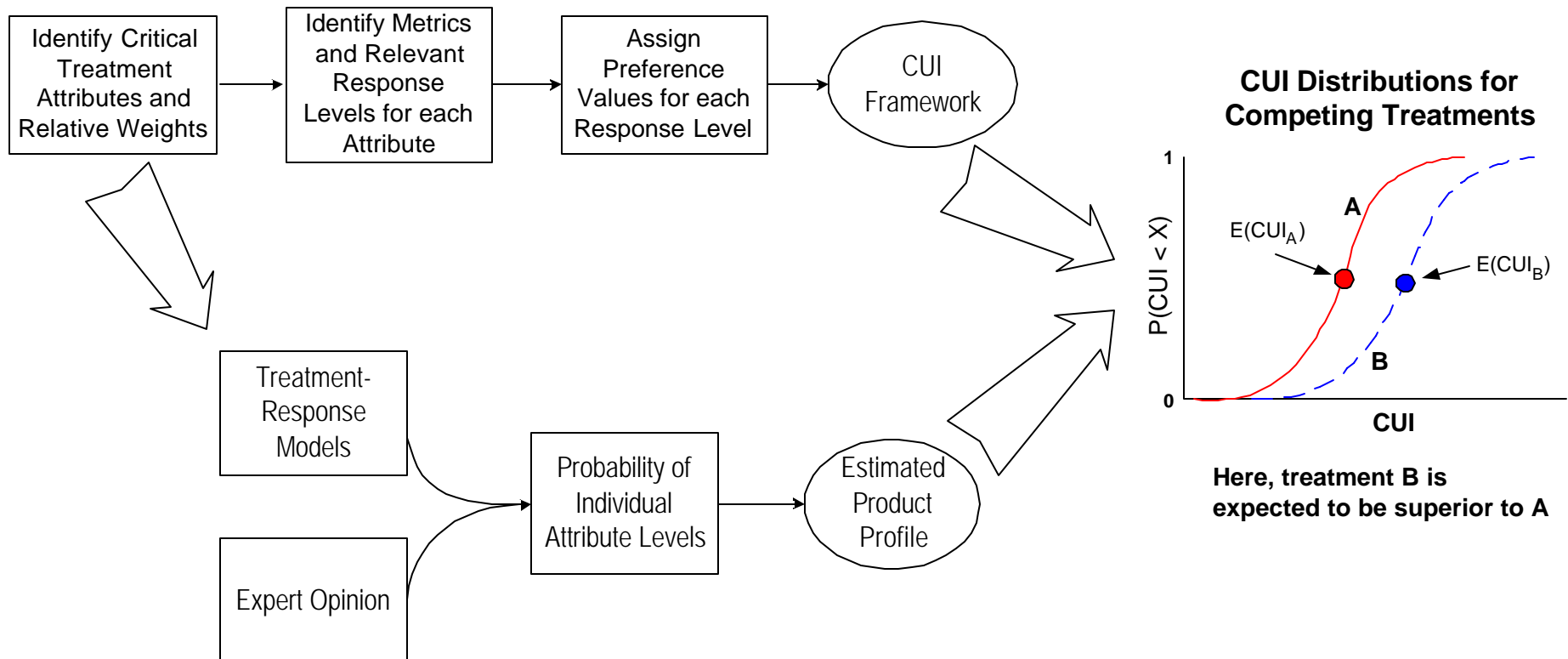
# *A Clinical Utility Index (CUI) is proposed to aid the dose selection decision*

- A CUI provides a means for quantifying the net benefit of a drug treatment.
- It considers the tradeoffs among the various drug effects comprising the product profile, balancing the benefits and risks
- The Clinical Utility Index (CUI) quantifies these tradeoffs by providing a single metric for the multiple dimensions of benefit and risk.
- **It is...** a **systematic** approach to understanding subjective preferences  
a **transparent** way of weighing tradeoffs  
**knowledge-driven**; available data are used; if not rely on expert opinion
- **It is not** an "**objective**" measure in the sense of a physiological measurement

# *The Clinical Utility Index rests on a well-established theoretical foundation*

- The Clinical Utility Index (CUI) combines preferences for outcomes into a single metric
  - It is an example of a “Multi-attribute utility function”
  - Use of these functions has been developed by researchers and practitioners of decision analysis over the past 20 years  
(e.g. Keeney, Ralph L. and Raiffa, Howard. 1976. Decisions with Multiple Objectives: Preferences and Value Tradeoffs. New York: John Wiley & Sons.)
- Construction of a CUI involves identifying and weighing those attributes critical to patient benefit
- Attributes and their levels are elicited and evaluated to ensure:
  - They describe all important and relevant clinical issues
  - The levels effectively capture the range of outcomes possible for the attribute
  - The clinical value of each of the levels is captured in the preference ratios and attribute ratings

*The framework for the CUI is elicited from the project team; when combined with models of response, it provides an relative estimate of the patient benefit*



# *The Clinical Utility Index - Process*

1. Identify product, value proposition, and competition
2. List attributes relevant to prescribing decision
3. Describe and discretize attributes into levels
4. Elicit preferences within attributes
5. Elicit preferences across attributes
6. Assess sensitivity of preference results
7. Use CUI

## *The process led to a CUI involving 11 attributes*

<b>Attribute</b>	<b>Rank</b>	<b>Relative weight</b>
HAMD change from baseline	1	100
% responders	2	95
% remissions	6	70
rate of onset	3 (tie)	85
sexual dysfunction	3 (tie)	85
dropouts due to AE's	5	80
anxiety	7	60
agitation	8	55
weight gain	9	40

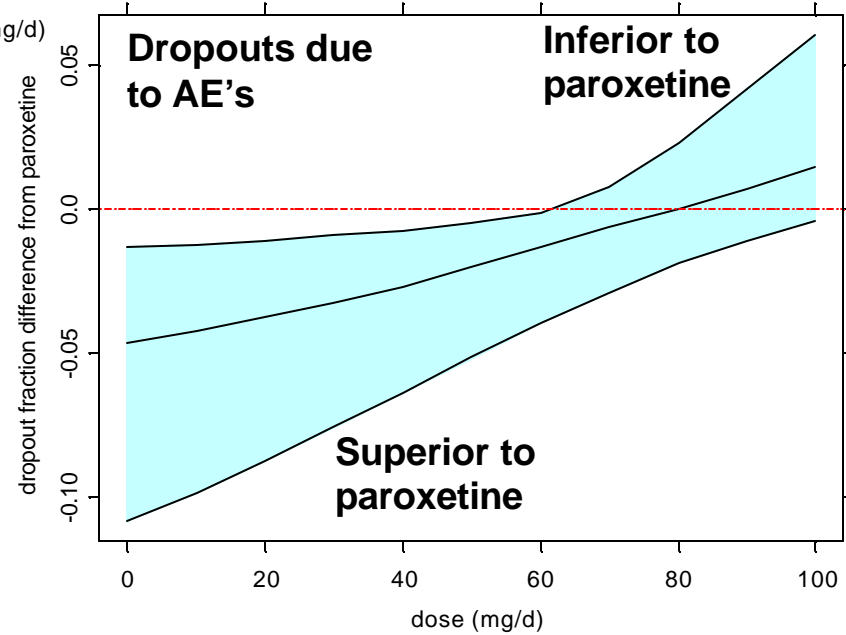
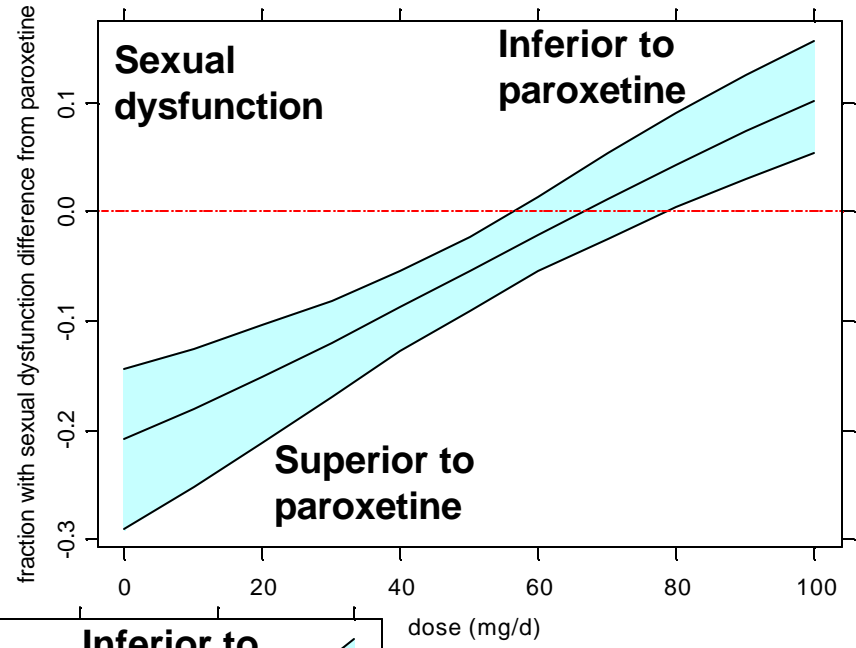
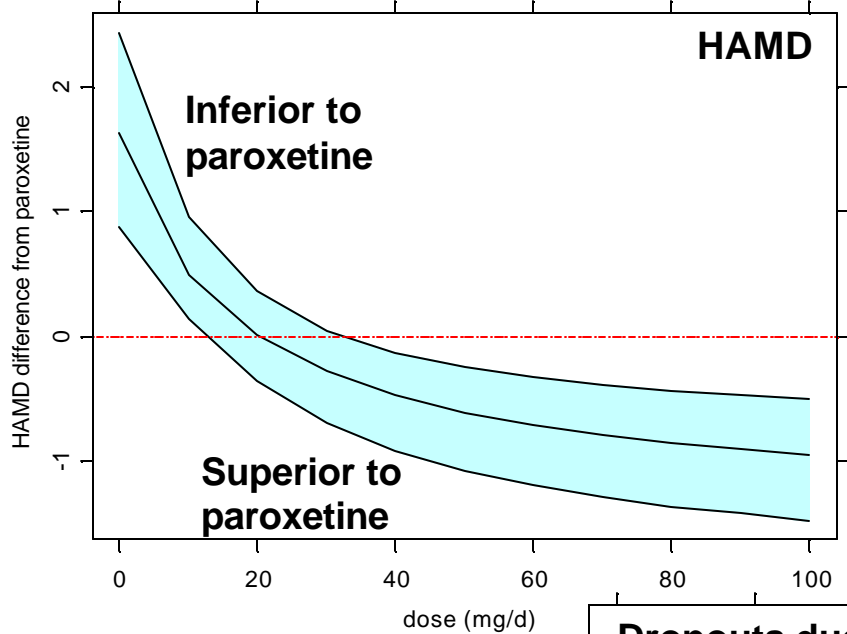
## *A simplified example*

- To simplify the presentation let's consider a CUI involving only 3 attributes:

<b>Attribute</b>	<b>Rank</b>	<b>Relative weight</b>
HAMD change from baseline	1	100
sexual dysfunction	2	85
dropouts due to AE's	3	80

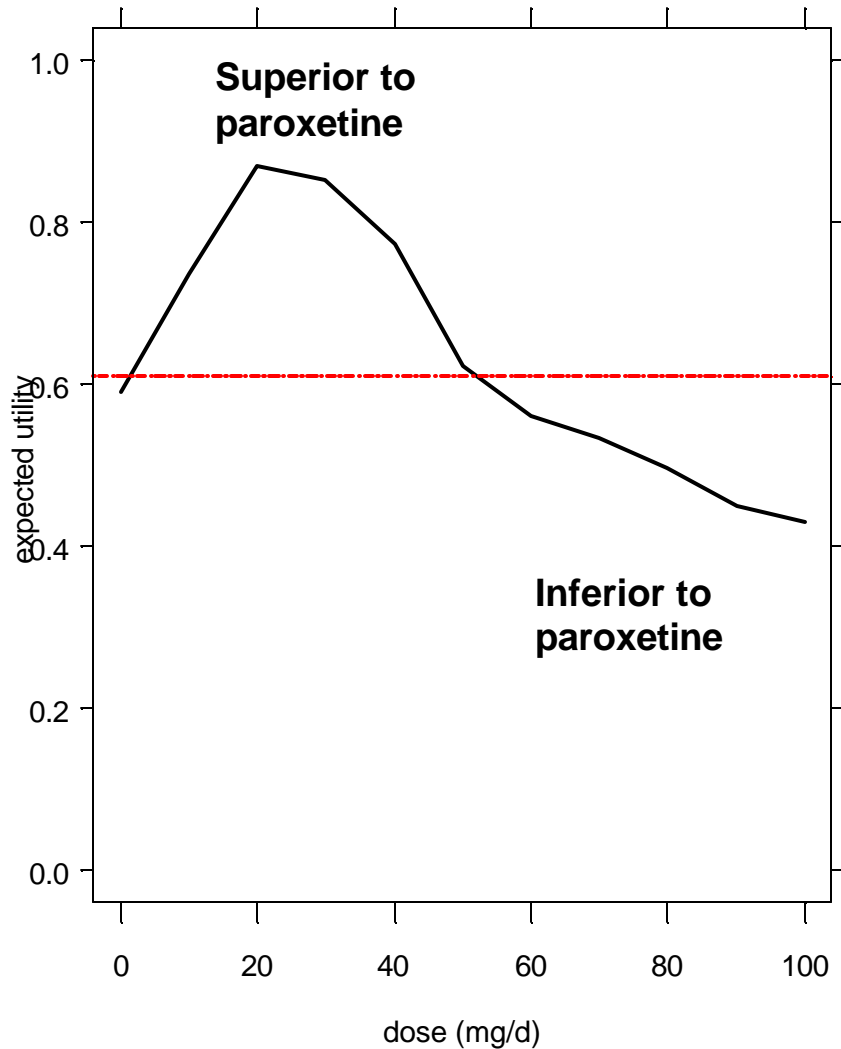
- Now that we have identified the key attributes, we develop dose-response models for each based on available data for nuoxetine and paroxetine.

# Nuoxetine dose-response expressed as differences from paroxetine



**Median & 90% prediction intervals**

*The CUI based on those predicted responses provides a clear measure of the net benefit of various nuoxetine doses*



## Conclusions

- Nuoxetine 10 to 50 mg/d is superior to paroxetine 40 mg/d
  - Supports further development of nuoxetine
- Optimal dose of nuoxetine is ~ 20 mg/d

## *Example 3: Phase II PoC trial design & analysis*

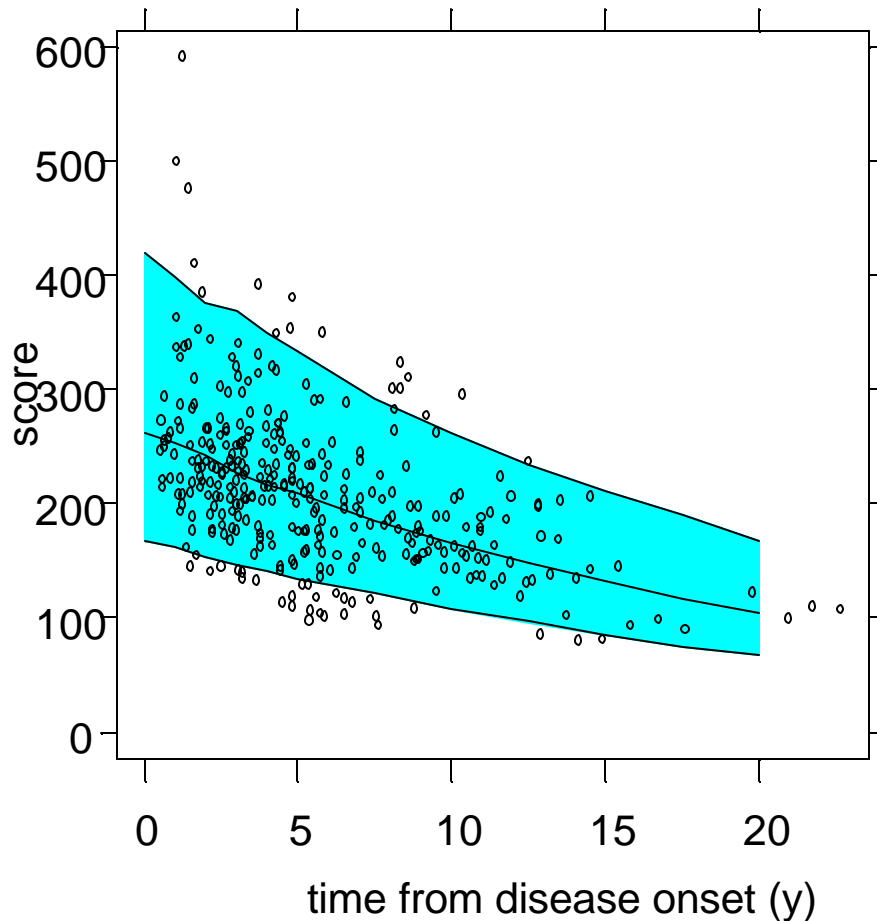
- Simuzine is a NCE for treatment of a slowly progressing neurological disorder.
- Previous exploratory clinical study of simuzine:
  - Primary efficacy score measured at baseline and 6 months.
  - Results encouraging but inconclusive.
  - Longer duration treatment is necessary to reach a decisive outcome.
- Additional longitudinal data available for model development:
  - Efficacy score for patients with observations at various times over durations up to 6 years.
  - Believed to be representative of the placebo group in the new trial.
- The new trial is already underway, so the M&S effort focuses on optimizing analysis of the trial results to support a PoC decision.

## *Example 3: Phase II PoC trial design & analysis*

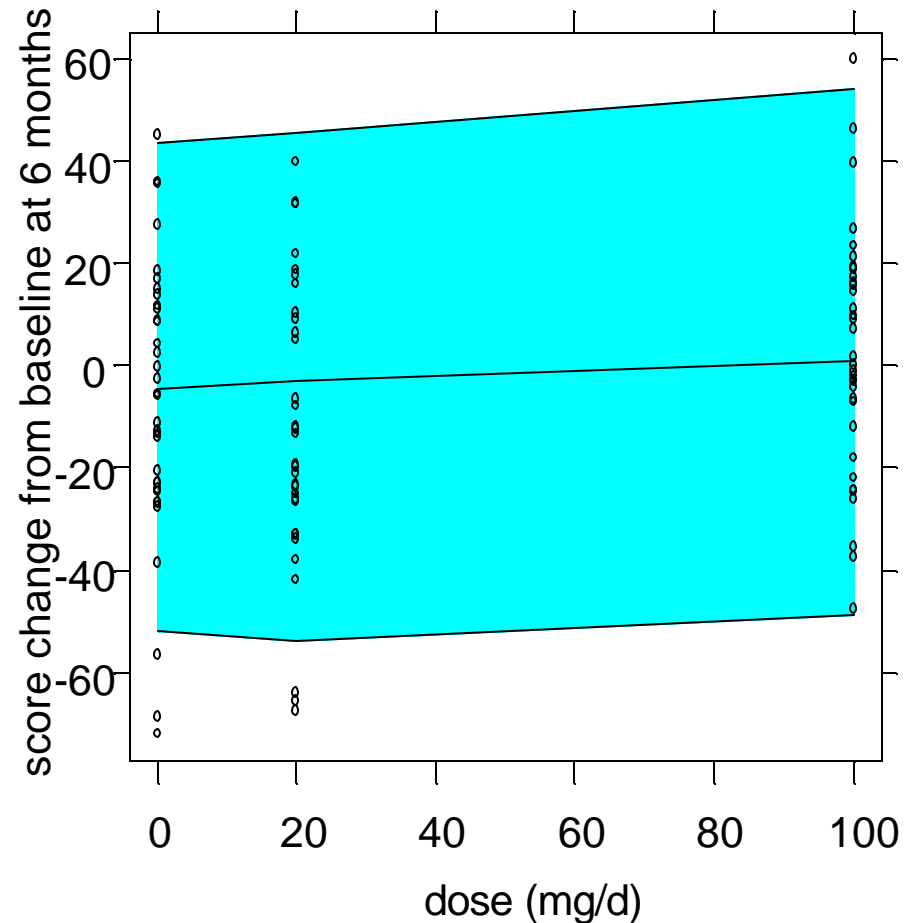
- New trial design:
  - Parallel, 2 treatment arm trial comparing simuzine 100 mg to placebo.
  - Primary endpoint = efficacy score at 2 years (LOCF imputation).
  - 100 patients per treatment arm.
- The example compares the use of 3 different trial analyses:
  - Conventional frequentist analysis of endpoint data (ANCOVA)
  - Bayesian longitudinal analysis with use of prior data.
  - Bayesian longitudinal analysis without prior data (non-informative priors).

# Model fitting results: Posterior predictive intervals

The model captures the decline in score after disease onset: Observed and model predicted (median and 90% prediction intervals) scores (placebo data)



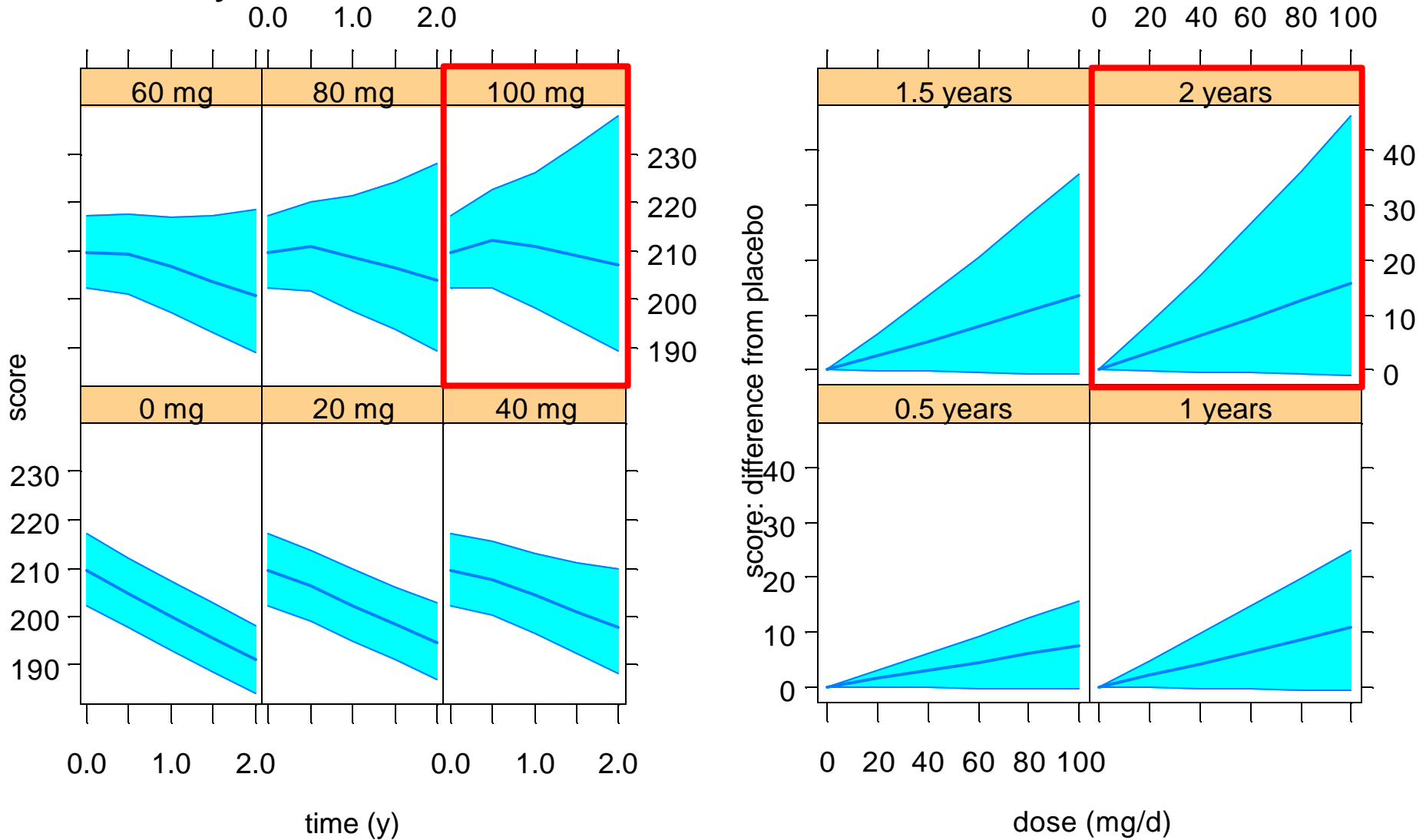
And the effect of simuzine: Observed and model predicted (median and 90% prediction intervals) change from baseline



## *Extending the model for simulating 2 year outcomes*

- The current model is based on data from only 6 months of simazine administration.
  - Model predictions for 1 and 2 years represent major extrapolations from experience.
  - The magnitude of the response at 1 and 2 years is more uncertain than indicated by simple linear extrapolation.
- For example, the available clinical evidence is also consistent with a more pessimistic model in which the drug benefit is not sustained beyond 6 months.
- Extrapolation beyond 6 months is based on expert judgment:
  - Upper bound: Linear extrapolation (constant slope)
  - Lower bound: Slope changes to pretreatment value after 6 months
  - Uncertainty in the post-6 month slope is modeled as a uniform distribution between those extremes.

*Model results indicate a highly uncertain but potentially large drug effect on the efficacy score*

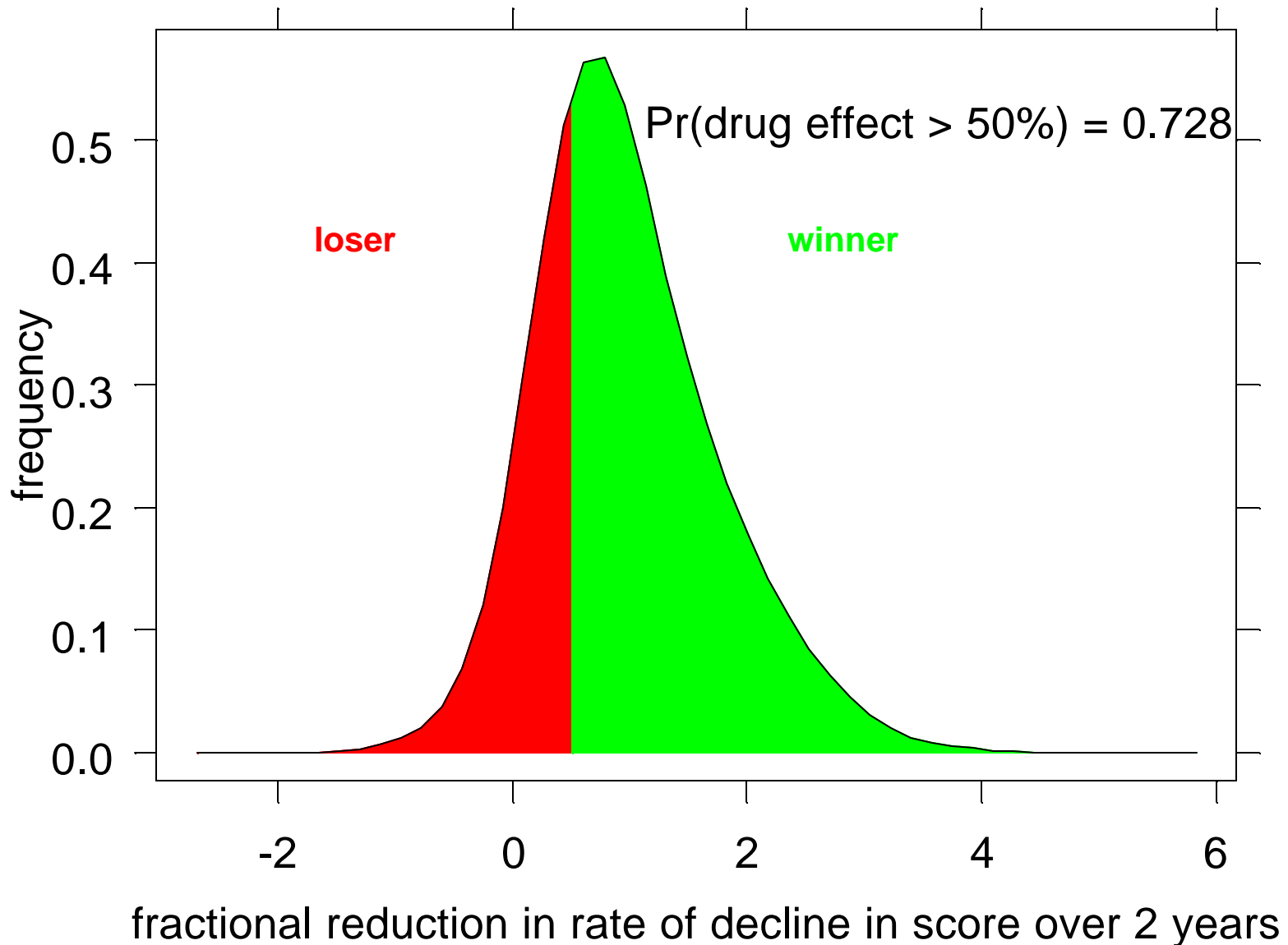


Model-predicted population mean score (median & 90% probability intervals) as a function of dose and time

## *Trial performance is measured by the quality of the proof-of-concept decision*

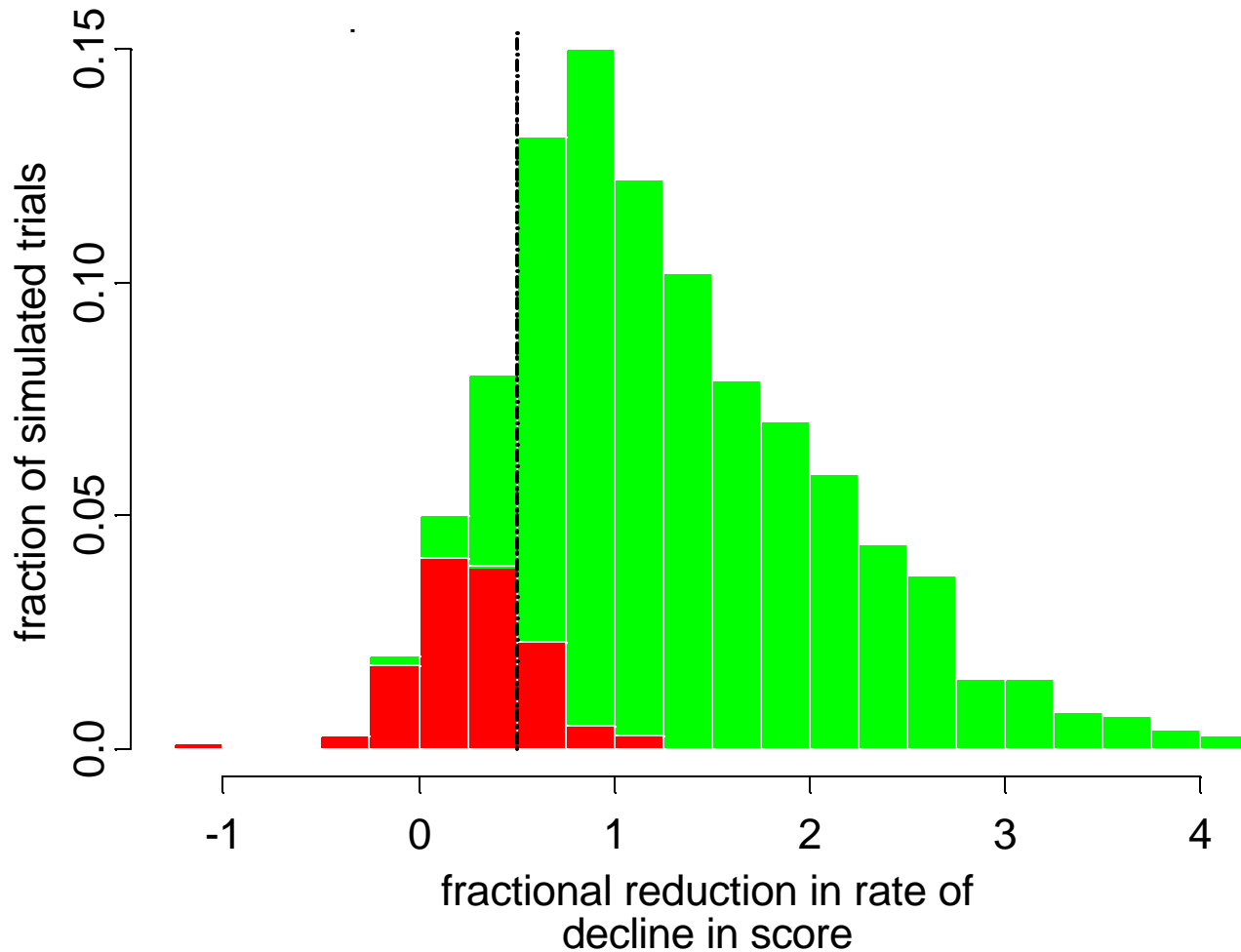
- Probability of reaching the correct (highest value) decision, i.e., go for a “winner” drug and no-go for a “loser” drug.
- You want to choose a trial design and a go/no-go decision method and criteria that minimizes:
  - $\text{Pr}(\text{go}|\text{loser})$ : probability of an incorrect go decision.
  - $\text{Pr}(\text{stop}|\text{winner})$ : probability of a lost opportunity.
- What is a “winner” or “loser” drug treatment?
  - The working definition of a “winner” used for the analyses presented here is a drug treatment that results in at least a 50% reduction in the rate of decline of the efficacy score over 2 years.

*Prior information indicates a 73% probability that simuzine 100 mg is a “winner”*



# Current trial design and per protocol analysis results in too many go decisions for “losers”

- $\Pr(\text{go}|\text{loser}) = 0.34$  high probability of an incorrect go decision.
- $\Pr(\text{stop}|\text{winner}) = 0.037$  low probability of a lost opportunity.



# Results of Bayesian longitudinal analyses

Go criteria:  $\text{Pr}(\geq 50\% \text{ reduction in the rate of decline over 2 years}) \geq p_{\text{crit}}$

Bayesian longitudinal analysis (without prior information) can be calibrated to improve the PoC decision by reducing incorrect go decisions.

Analysis criteria	Simulated trial results		
	$p_{\text{crit}}$	Pr(stop winner)	Pr(go loser)
Bayesian longitudinal analysis without prior information			
0.18	0.037	0.290	
0.5	0.065	0.152	
0.6	0.085	0.122	
0.7	0.097	0.084	
0.8	0.126	0.057	
0.9	0.182	0.041	
0.95	0.216	0.020	
ANCOVA results	0.037	0.338	

## Results of Bayesian longitudinal analyses

Go criteria:  $\Pr(\geq 50\% \text{ reduction in the rate of decline over 2 years}) \geq p_{\text{crit}}$

Incorporation of the prior data offers little or no additional improvement in the quality of the PoC decision.

Analysis criteria	Simulated trial results	
$p_{\text{crit}}$	Pr(stop winner)	Pr(go loser)
Bayesian longitudinal analysis without prior information		
0.18	0.037	0.290
0.5	0.065	0.152
0.6	0.085	0.122
0.7	0.097	0.084
0.8	0.126	0.057
0.9	0.182	0.041
0.95	0.216	0.020
Bayesian longitudinal analysis with prior information		
0.22	0.037	0.281
0.5	0.078	0.149
0.6	0.111	0.115
0.7	0.126	0.088
0.8	0.145	0.057
0.9	0.180	0.027
0.95	0.223	0.017
ANCOVA results	0.037	0.338

## What next?

- The presented approach for assessing trial design performance does not explicitly optimize the tradeoffs between false positives (go|loser) and false negatives (stop|winner).
- That may be addressed by associating values (possibly economic) to the losses due to those competing errors and using Bayesian decision analysis to optimize the choice of analysis criteria ( $p_{\text{crit}}$ ).
- Alternatively, the go/no-go decision method could be based on Bayesian decision analysis rather than the approach shown here.

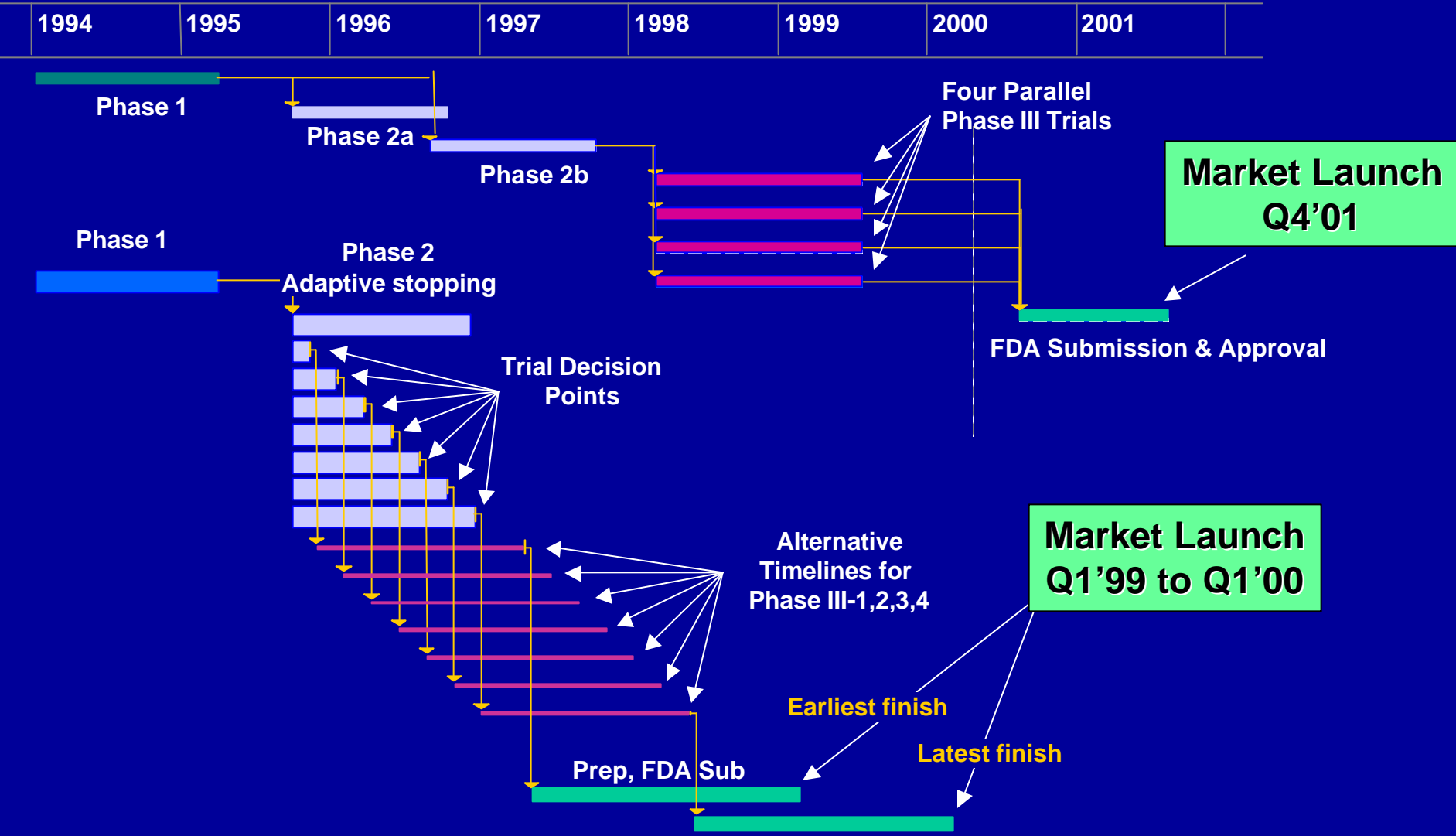
## Example 4: Adaptive trial design for PoC & dose-finding

- Compound:** • **Potential treatment for a chronic psychiatric disorder, novel mechanism of action**
- Status:** • **Phase I completed**
- Competitor:** • **On market for 5 years**
- Target Product Profile:** • **Similar or better efficacy than Competitor with reduced side-effects**
- Issues:** • **What is the best (highest NPV) Program Strategy for determining:**
- **If Drug X is a “dud” and should be killed?**
  - **If “Effective”, how and when to move to Phase III?**

## *Our mission was to find a way to improve decision making and phase II/III trial execution.*

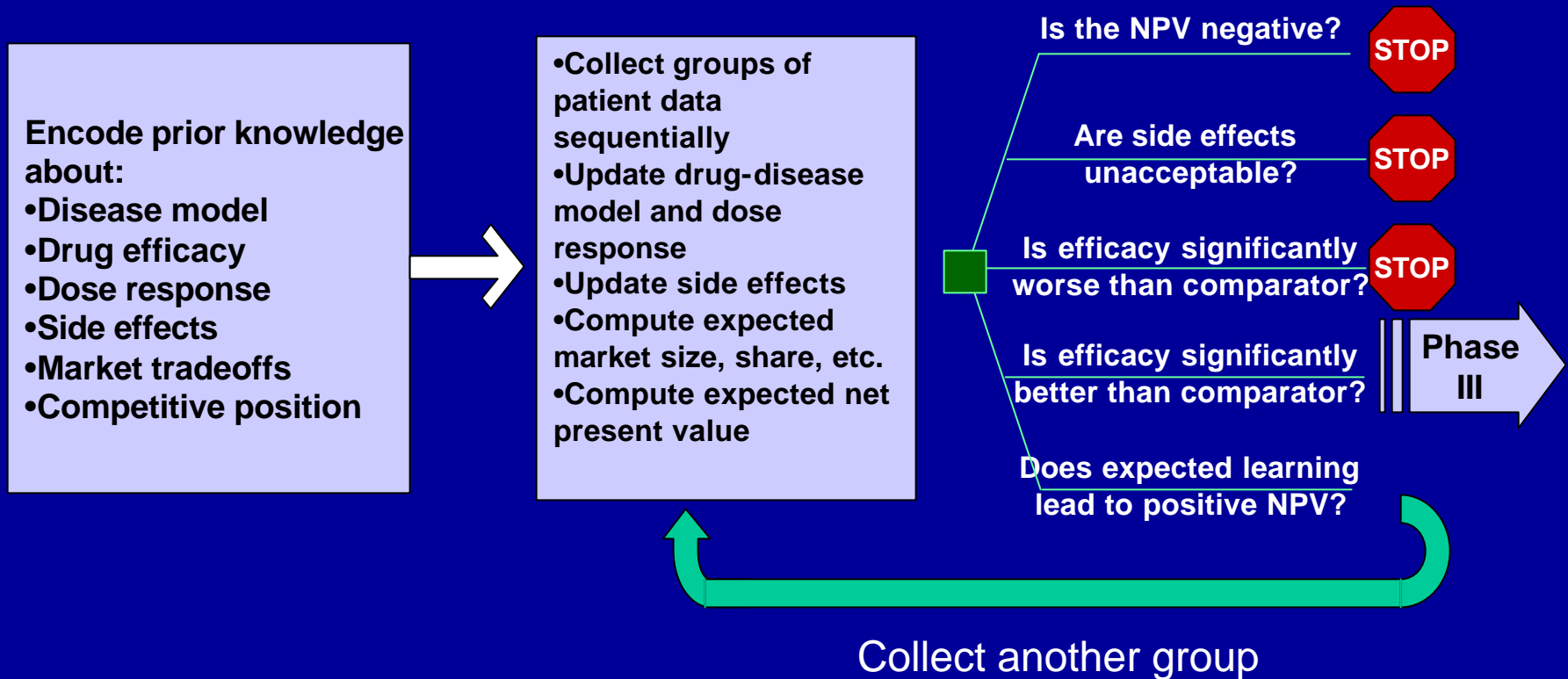
- This was actually a post-mortem of a development program that failed in Phase III.
- We were initially blinded from the results of the Phase II/III trials.
- We proposed a phase II design which we believed would be more effective than the client design.
- The Phase II data was then provided and used for model construction.
- The proposed design was simulated.
- The phase II results, time and cost were compared for competing designs

# Models allowed design of a development program that was up to two years faster

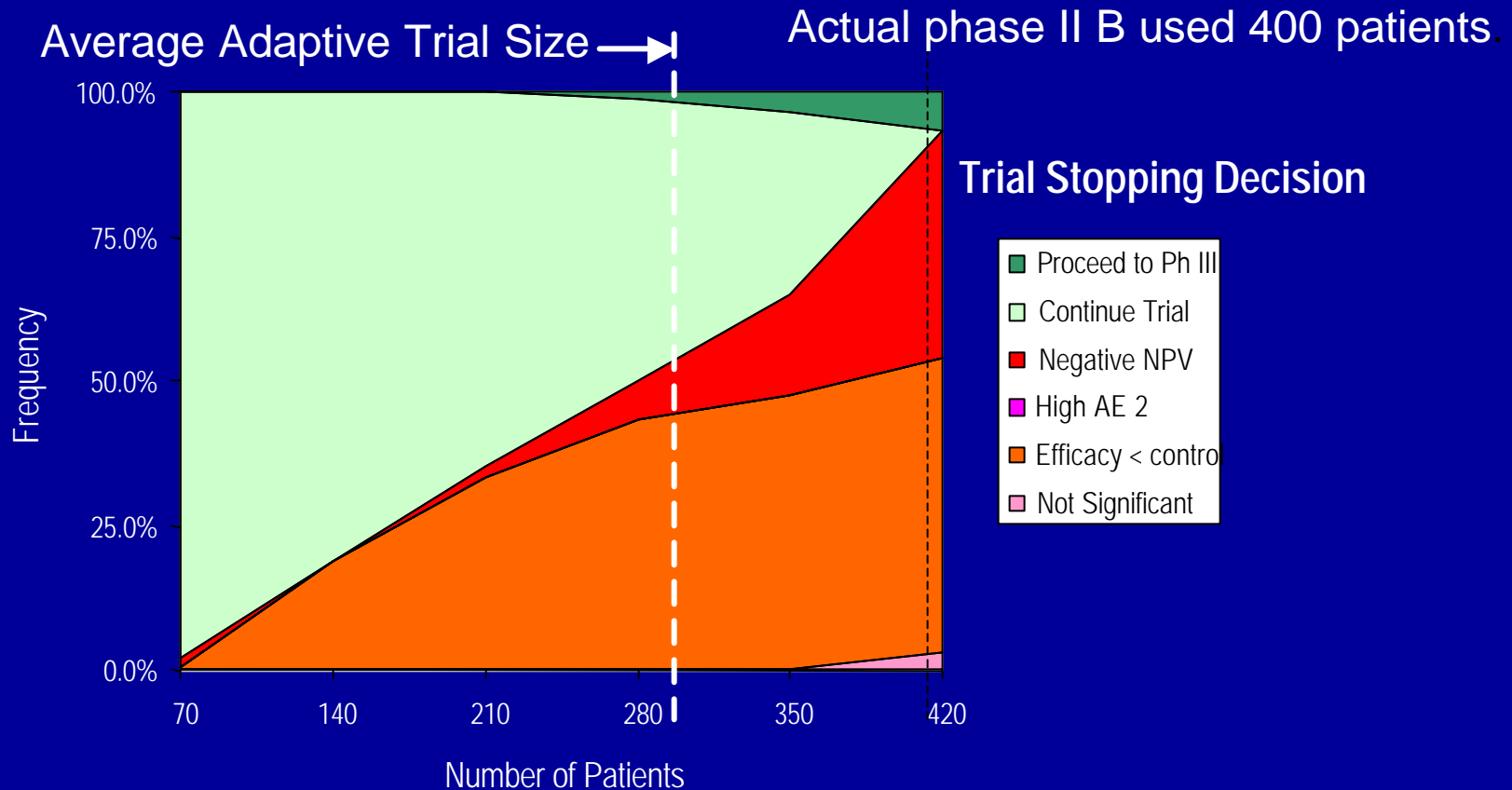


*For phase II, adaptive designs can provide multiple learning and decision opportunities.*

## The Proposed Phase II Design



*Repeated simulations showed that the drug would have failed early in the adaptive trial design.*

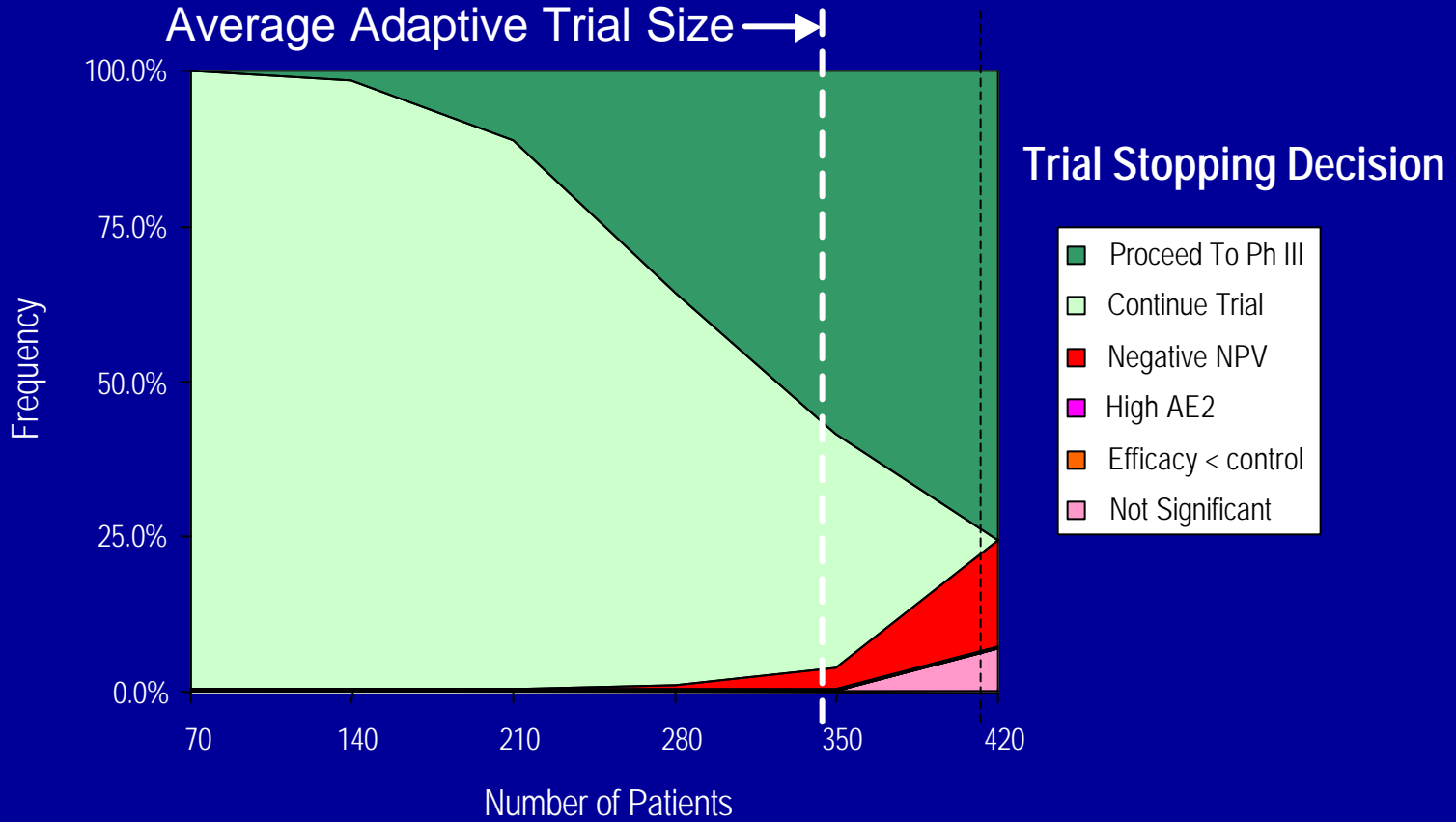


- **Average trial accrued 295 patients in ~10 mo. An erroneous “GO” decision given 6.5% of the time.**
- **The client had equivocal results and proceeded to phase III.**

*What if the client had a more efficacious drug?*

*In 76% of the simulations, the adaptive trial design recommended starting phase III given a drug efficacy profile similar to the positive control.*

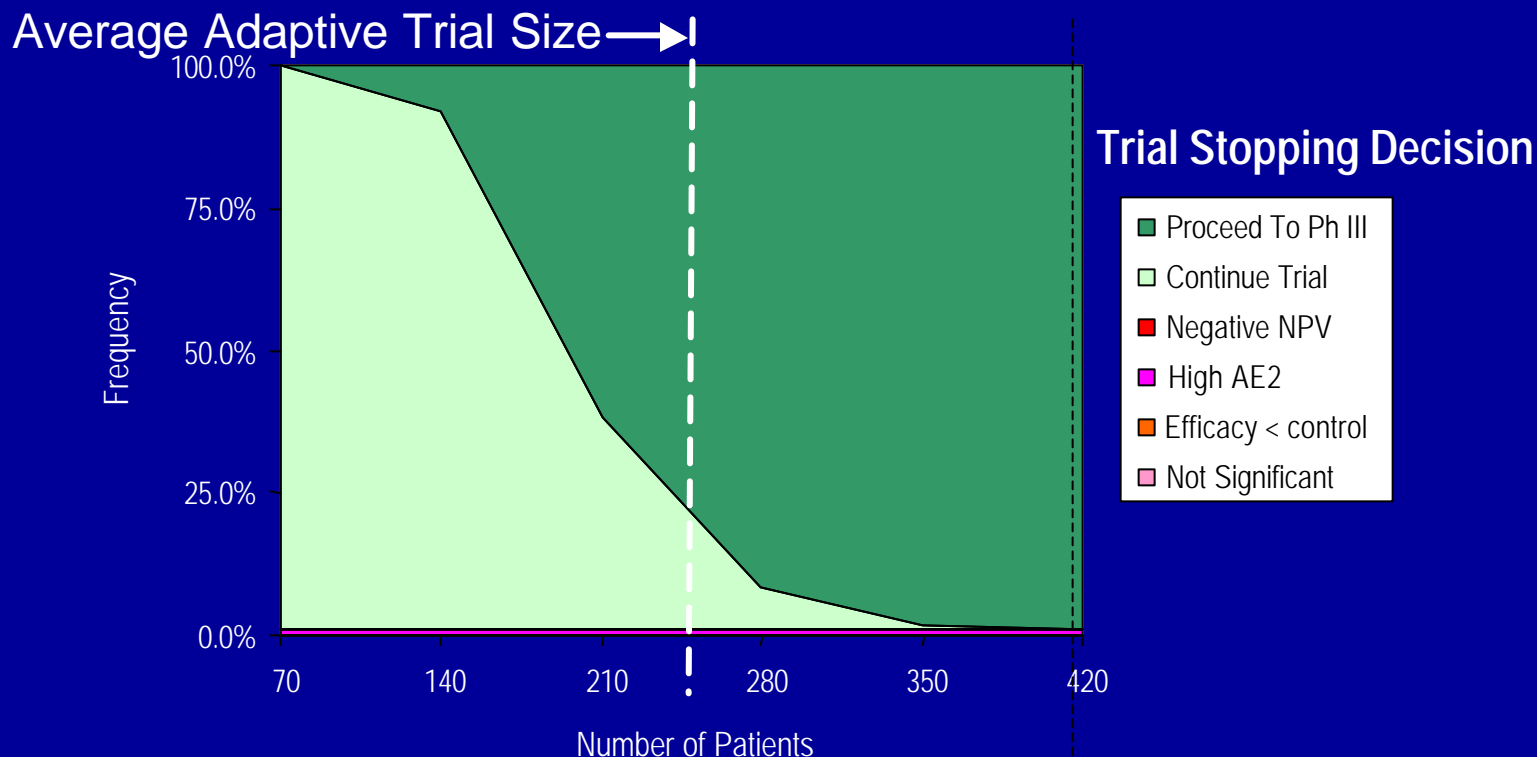
Actual phase II B used 400 patients.



➤ **Average trial accrued 341 patients in ~12 mo.**

*With blockbuster efficacy the proposed trial recommended starting phase III much earlier.*

Actual phase II B used 400 patients.



- **Average trial showed significant efficacy with 235 patients in ~9 mo.**
- **Compared to the original design, the expected value of the compound would increase by 26% due to a quicker time to market.**

# Discussion

- The examples shown here focus on decision-making by pharmaceutical companies.
- Could these or similar methods be adapted to regulatory decisions?
- What is the value of sponsor-performed M&S for supporting regulatory decisions?
  - End of Phase II: dose selection & trial design
  - NDA: labeling, doses for special populations, supportive evidence for approval