Biostatistical Considerations: Biologics and Biomarkers in Oncology

Daniel Normolle, Ph.D.
Graduate School of Public Health Department of Biostatistics
UPCI Biostatistics Facility
University of Pittsburgh

No Relationships to Disclose

September 30, 2010

Outline

Outline

Outline

Traditional Cytotoxic Development Paradigm

- Phase 1: Characterize Safety

Traditional Cytotoxic Development Paradigm

- Phase 1: Characterize Safety
  - $3 + 3$

Traditional Cytotoxic Development Paradigm

- Phase 1: Characterize Safety
  - $3 + 3$
- Phase 2: Characterize Efficacy

Traditional Cytotoxic Development Paradigm

- Phase 1: Characterize Safety
  - $3 + 3$
- Phase 2: Characterize Efficacy
  - Single-arm, historical control

Traditional Cytotoxic Development Paradigm

- Phase 1: Characterize Safety
  - $3 + 3$
- Phase 2: Characterize Efficacy
  - Single-arm, historical control
  - Randomized parallel trials, historical control

Traditional Cytotoxic Development Paradigm

- Phase 1: Characterize Safety
  - $3 + 3$
- Phase 2: Characterize Efficacy
  - Single-arm, historical control
  - Randomized parallel trials, historical control
- Phase 3: Confirm Efficacy & Safety

Traditional Cytotoxic Development Paradigm

- Phase 1: Characterize Safety
  - $3 + 3$
- Phase 2: Characterize Efficacy
  - Single-arm, historical control
  - Randomized parallel trials, historical control
- Phase 3: Confirm Efficacy & Safety
  - Randomized, active concurrent controlled

$3 + 3$ Design for Dose Escalation on Toxicity

- Treat 3 patients at a low (lowest) dose

$3 + 3$ Design for Dose Escalation on Toxicity

- Treat 3 patients at a low (lowest) dose
- Escalate for 0 DLTs/3 patients

$3 + 3$ Design for Dose Escalation on Toxicity

- Treat 3 patients at a low (lowest) dose
- Escalate for 0 DLTs/3 patients
- Treat 3 more patients for 1 DLTs/3 patients

$3 + 3$ Design for Dose Escalation on Toxicity

- Treat 3 patients at a low (lowest) dose
- Escalate for 0 DLTs/3 patients
- Treat 3 more patients for 1 DLTs/3 patients
- Escalate for 1 DLTs/6 patients

$3 + 3$ Design for Dose Escalation on Toxicity

- Treat 3 patients at a low (lowest) dose
- Escalate for 0 DLTs/3 patients
- Treat 3 more patients for 1 DLTs/3 patients
- Escalate for 1 DLTs/6 patients
- Deëscalate for > 1 DLTs/6 patients

$3 + 3$ Design for Dose Escalation on Toxicity

- Treat 3 patients at a low (lowest) dose
- Escalate for 0 DLTs/3 patients
- Treat 3 more patients for 1 DLTs/3 patients
- Escalate for 1 DLTs/6 patients
- Deëscalate for > 1 DLTs/6 patients
- RP2D is highest dose with ≤ 1 DLT/6 patients

Why Trialists Like $3 + 3$

- Requires limited number of participants

Why Trialists Like $3 + 3$

- Requires limited number of participants
- Easy to explain

Why Trialists Like $3 + 3$

- Requires limited number of participants
- Easy to explain
- (Should be) easy to execute

Why Trialists Like $3 + 3$

- Requires limited number of participants
- Easy to explain
- (Should be) easy to execute
- Doesn't require mathematical modeling, or statisticians

Why Trialists Like $3 + 3$

- Requires limited number of participants
- Easy to explain
- (Should be) easy to execute
- Doesn't require mathematical modeling, or statisticians
- Appearance of prudence

Unavoidable Limitations of $3 + 3$

- Intended for treatments where toxicity increases with dose

Unavoidable Limitations of $3 + 3$

- Intended for treatments where toxicity increases with dose
- Treats large proportions of participants at potentially sub-therapeutic doses

Unavoidable Limitations of $3 + 3$

- Intended for treatments where toxicity increases with dose
- Treats large proportions of participants at potentially sub-therapeutic doses
- Can be slow to escalate even if no DLTs are observed

Unavoidable Limitations of $3 + 3$

- Intended for treatments where toxicity increases with dose
- Treats large proportions of participants at potentially sub-therapeutic doses
- Can be slow to escalate even if no DLTs are observed
- No quantitative mechanism for employing prior knowledge about toxicity

Unavoidable Limitations of $3 + 3$

- Intended for treatments where toxicity increases with dose
- Treats large proportions of participants at potentially sub-therapeutic doses
- Can be slow to escalate even if no DLTs are observed
- No quantitative mechanism for employing prior knowledge about toxicity
- Eliminates bad doses from further consideration, but is underpowered for selecting among the remaining doses:

Unavoidable Limitations of $3 + 3$

- Intended for treatments where toxicity increases with dose
- Treats large proportions of participants at potentially sub-therapeutic doses
- Can be slow to escalate even if no DLTs are observed
- No quantitative mechanism for employing prior knowledge about toxicity
- Eliminates bad doses from further consideration, but is underpowered for selecting among the remaining doses:
    - 95% exact 1-sided CI for 0 DLTs in 6 patients: (0,0.39)

Unavoidable Limitations of $3 + 3$

- Intended for treatments where toxicity increases with dose
- Treats large proportions of participants at potentially sub-therapeutic doses
- Can be slow to escalate even if no DLTs are observed
- No quantitative mechanism for employing prior knowledge about toxicity
- Eliminates bad doses from further consideration, but is underpowered for selecting among the remaining doses:
  - 95% exact 1-sided CI for 0 DLTs in 6 patients: (0,0.39)
  - 95% exact 1-sided CI for 1 DLTs in 6 patients: (0,0.58)

Effective Use of $3 + 3$

- Eliminate extremely toxic doses,

Effective Use of $3 + 3$

- Eliminate extremely toxic doses,
- Not choose between doses that are not extremely toxic

Possible Non-Cytotoxic Developmental Contexts

- Biological therapy highly likely to be very low in toxicity

Possible Non-Cytotoxic Developmental Contexts

- Biological therapy highly likely to be very low in toxicity
- Biological therapy unlikely to have increasing toxicity with increasing dose

Possible Non-Cytotoxic Developmental Contexts

- Biological therapy highly likely to be very low in toxicity
- Biological therapy unlikely to have increasing toxicity with increasing dose
- Addition of a component to a known therapy intended to reduce toxicity and, possibly, increase efficacy

Why Not Use $3 + 3$ in the Non-Cytotoxic Context?

- Monitoring toxicity is different from escalating on toxicity

Why Not Use $3 + 3$ in the Non-Cytotoxic Context?

- Monitoring toxicity is different from escalating on toxicity
- If toxicity is low, escalation will just move to highest dose

Why Not Use $3 + 3$ in the Non-Cytotoxic Context?

- Monitoring toxicity is different from escalating on toxicity
- If toxicity is low, escalation will just move to highest dose
- Highest dose may not be best dose

Why Not Use $3 + 3$ in the Non-Cytotoxic Context?

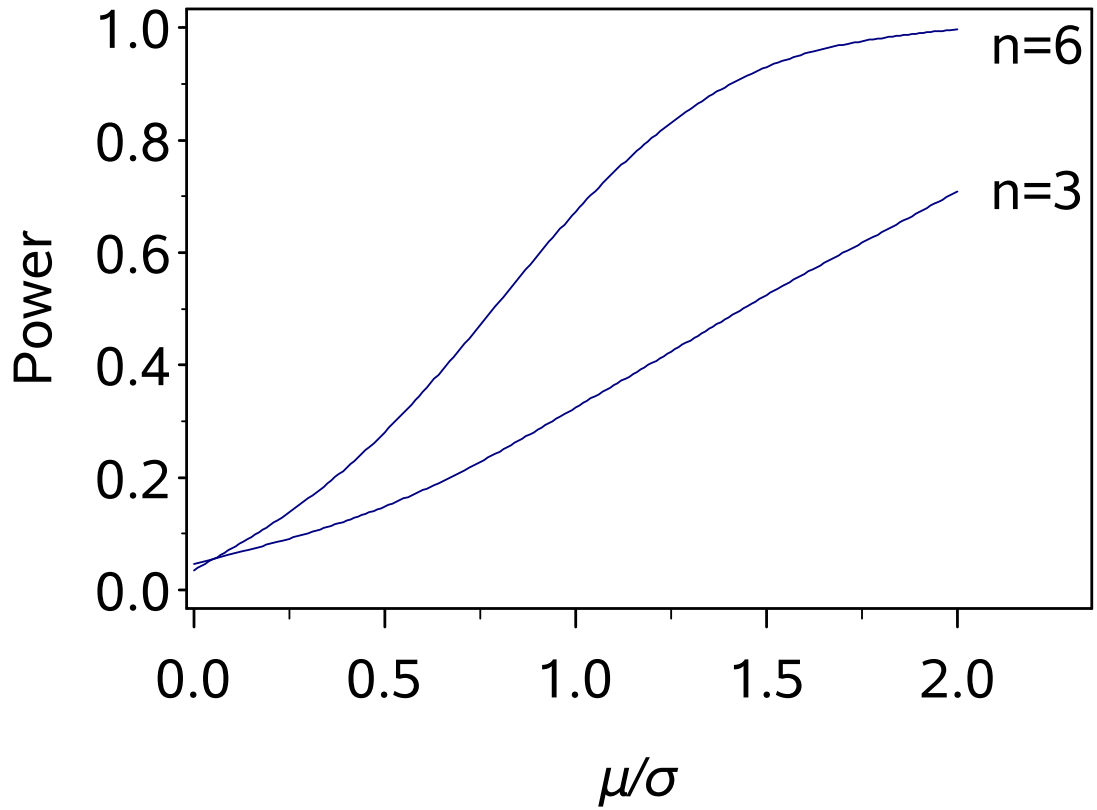
- Monitoring toxicity is different from escalating on toxicity
- If toxicity is low, escalation will just move to highest dose
- Highest dose may not be best dose
- If added component really does reduce toxicity, escalating on toxicity may not choose most useful dose

Why Not Use $3 + 3$ in the Non-Cytotoxic Context?

- Monitoring toxicity is different from escalating on toxicity
- If toxicity is low, escalation will just move to highest dose
- Highest dose may not be best dose
- If added component really does reduce toxicity, escalating on toxicity may not choose most useful dose
- Cohort sizes of 3 and 6 are often too small to be useful for the most relevant objectives

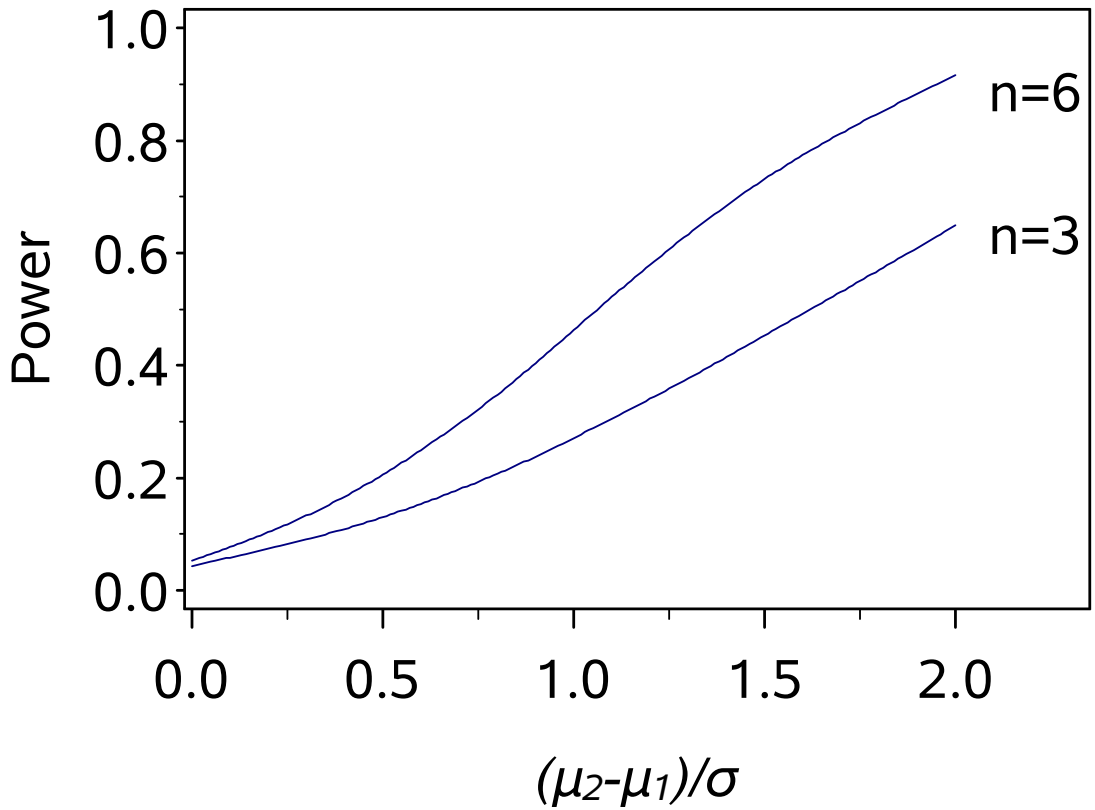
Testing Biomarker Endpoints Between $3+3$ Cohorts

- Test $H_o : \mu = 0$, one-sided, $\alpha = 0.05$



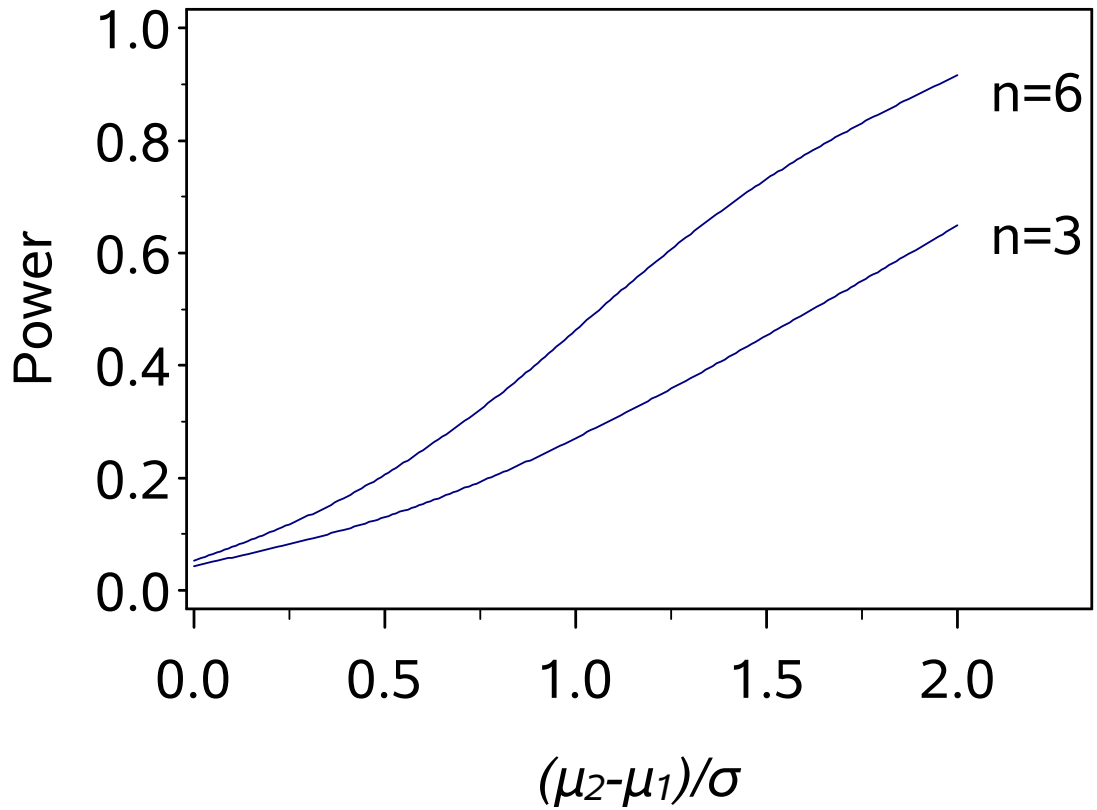
Testing Biomarker Endpoints Between $3+3$ Cohorts

- Test $H_o : \mu_1 = \mu_2$, 1-sided, $\alpha = 0.05$



Testing Biomarker Endpoints Between $3+3$ Cohorts

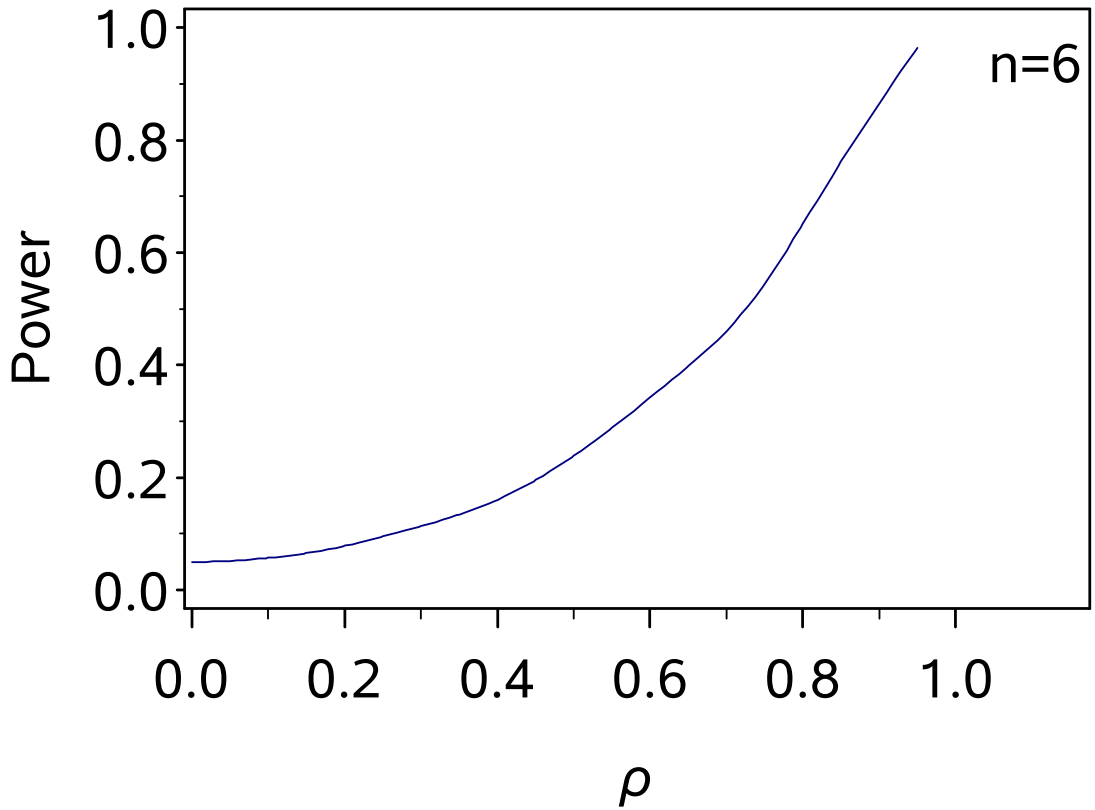
- Test $H_o : \mu_1 = \mu_2$, 1-sided, $\alpha = 0.05$



- Power for $\pi = 0.3$ versus $\pi = 0.05$ (1-sided, $\alpha = 0.05$) is 0.18

Testing Biomarker Endpoints Between $3+3$ Cohorts

- Test $H_o : \rho = 0$, one-sided, $\alpha = 0.05$



Expansion Cohort?

- No provision for incorporation of expansion cohort safety responses into estimate of RP2D

Expansion Cohort?

- No provision for incorporation of expansion cohort safety responses into estimate of RP2D
- High probability of expansion at suboptimal dose

Expansion Cohort?

- No provision for incorporation of expansion cohort safety responses into estimate of RP2D
- High probability of expansion at suboptimal dose
- Maximal (at $p = 0.5$) 95% exact two-sided binomial confidence interval based on 12 patients: (0.21,0.79)

Problem in a Nutshell

- *Monitoring* toxicity is always important, but

Problem in a Nutshell

- *Monitoring* toxicity is always important, but
- Eliminating grossly toxic doses is not always the primary objective

Problem in a Nutshell

- *Monitoring* toxicity is always important, but
- Eliminating grossly toxic doses is not always the primary objective
- Three or six participants per arm is insufficient for decision making other than eliminating grossly toxic doses

Problem in a Nutshell

- *Monitoring* toxicity is always important, but
- Eliminating grossly toxic doses is not always the primary objective
- Three or six participants per arm is insufficient for decision making other than eliminating grossly toxic doses
- *Escalation* on toxicity produces poor operating characteristics when the probability of toxicity is very low or a monotonic dose-toxicity relationship is unlikely

Outline

Criteria for Early Phase Trials of Non-Cytotoxics

- Safe enough

Criteria for Early Phase Trials of Non-Cytotoxics

- Safe enough
- Feasible number of participants

Criteria for Early Phase Trials of Non-Cytotoxics

- Safe enough
- Feasible number of participants
- Informative concerning primary endpoint

Alternative Objectives In Early Phase Combination Therapy Studies

- Proof of principle

Alternative Objectives In Early Phase Combination Therapy Studies

- Proof of principle
- Identify sources of variability in biomarker assessments

Alternative Objectives In Early Phase Combination Therapy Studies

- Proof of principle
- Identify sources of variability in biomarker assessments
- Estimate biologically effective doses

Alternative Objectives In Early Phase Combination Therapy Studies

- Proof of principle
- Identify sources of variability in biomarker assessments
- Estimate biologically effective doses
- Eliminate biologically ineffective doses from further consideration

Alternative Objectives In Early Phase Combination Therapy Studies

- Proof of principle
- Identify sources of variability in biomarker assessments
- Estimate biologically effective doses
- Eliminate biologically ineffective doses from further consideration
- Assess relationships between markers at biologically effective doses

Scenario 1: Autologous Tumor Cell Vaccine

- Non-cytotoxic characteristic

Scenario 1: Autologous Tumor Cell Vaccine

- Non-cytotoxic characteristic
  - Toxicity is very likely to be very low (P(DLT) $< 5\%$) at any testable dose

Scenario 1: Autologous Tumor Cell Vaccine

- Non-cytotoxic characteristic
  - Toxicity is very likely to be very low (P(DLT) $< 5\%$) at any testable dose
- Objectives

Scenario 1: Autologous Tumor Cell Vaccine

- Non-cytotoxic characteristic
  - Toxicity is very likely to be very low (P(DLT) $< 5\%$) at any testable dose
- Objectives
  - Characterize immunological response

Scenario 1: Autologous Tumor Cell Vaccine

- Non-cytotoxic characteristic
  - Toxicity is very likely to be very low (P(DLT) $< 5\%$) at any testable dose
- Objectives
  - Characterize immunological response
  - Establish immunological response at highest dose

Scenario 1: Autologous Tumor Cell Vaccine

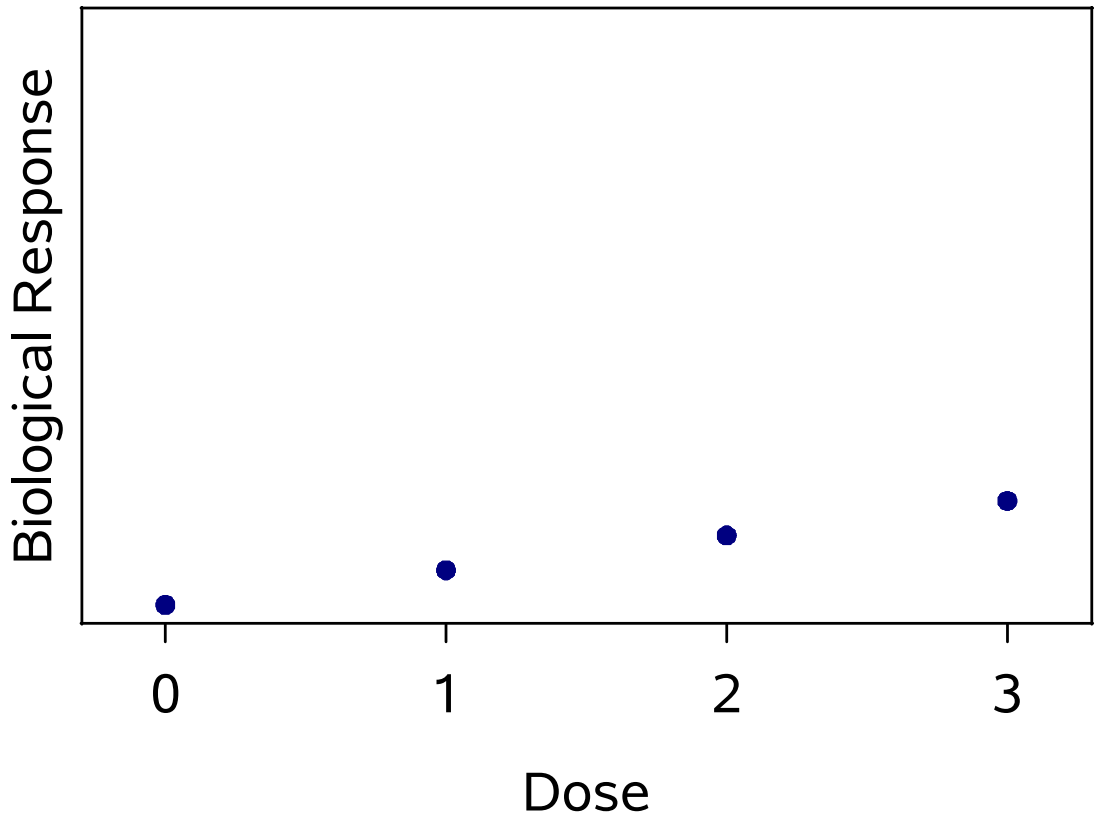
- Non-cytotoxic characteristic
  - Toxicity is very likely to be very low (P(DLT) $< 5\%$) at any testable dose
- Objectives
  - Characterize immunological response
  - Establish immunological response at highest dose
  - Determine if lower doses also induce response

Scenario 1: Autologous Tumor Cell Vaccine

- Non-cytotoxic characteristic
  - Toxicity is very likely to be very low (P(DLT) $< 5\%$) at any testable dose
- Objectives
  - Characterize immunological response
  - Establish immunological response at highest dose
  - Determine if lower doses also induce response
  - Monitor toxicity

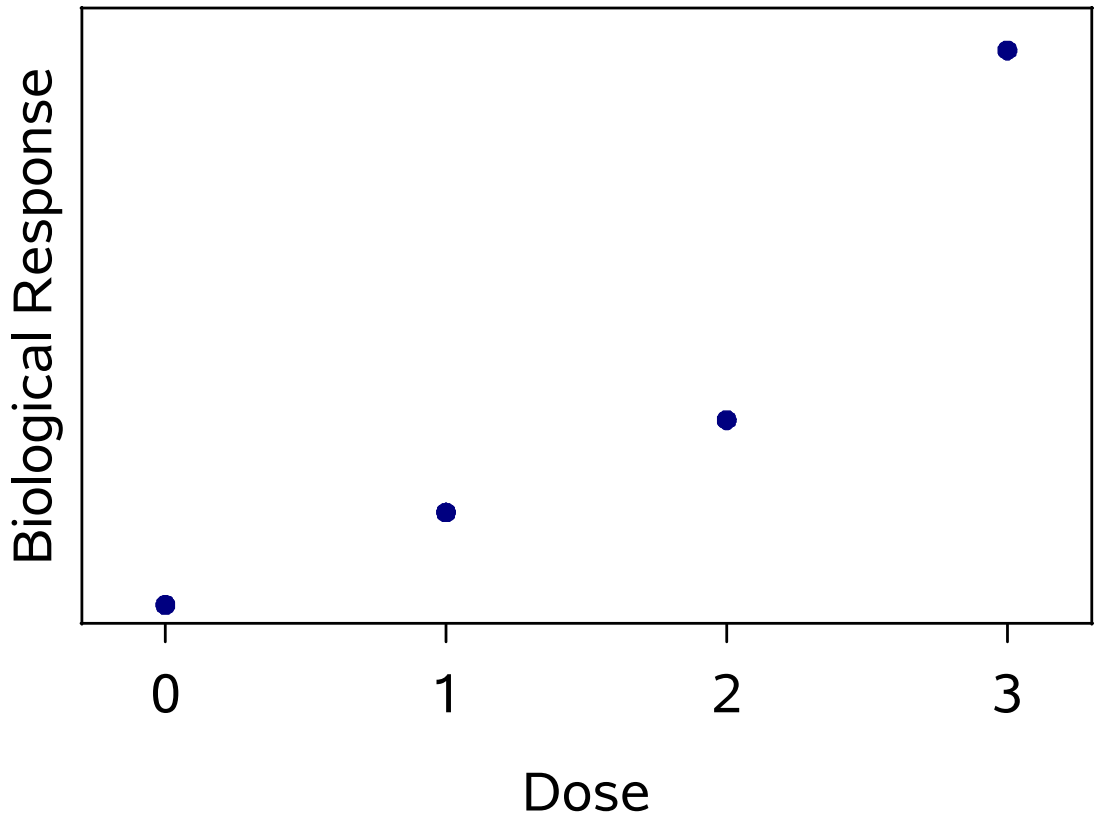
Dose Preference

- No dose is effective



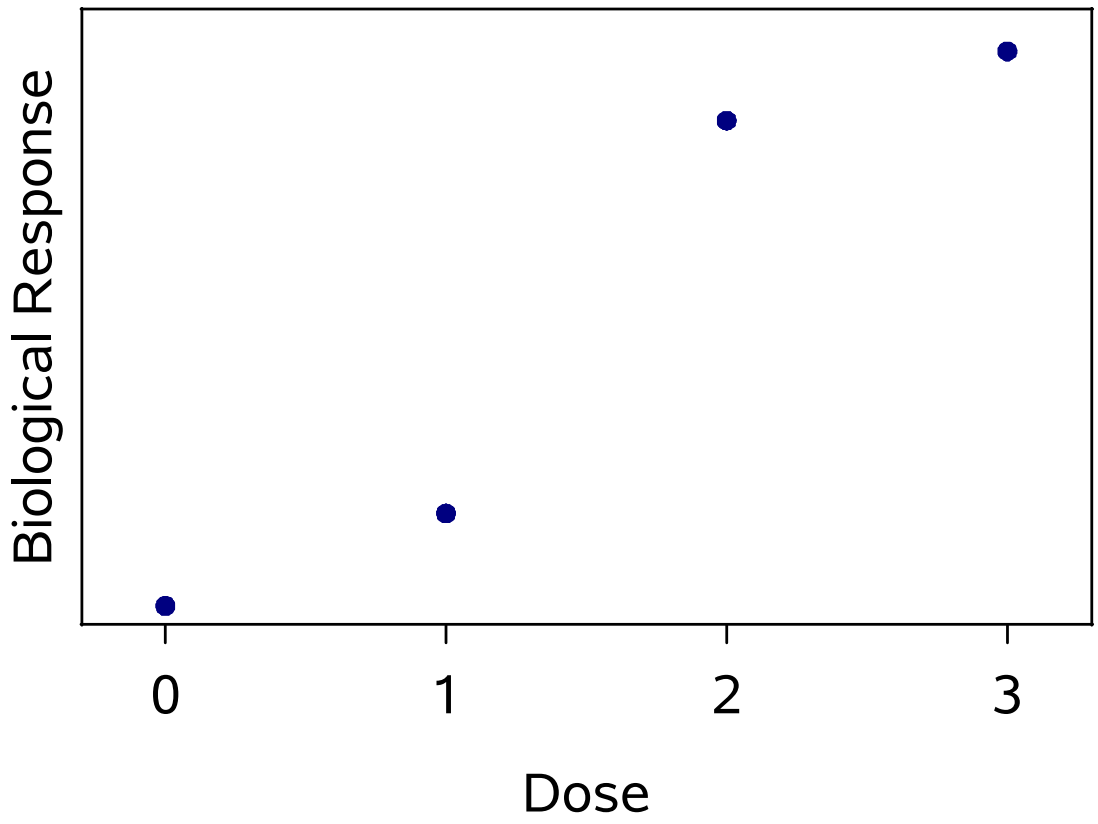
Dose Preference

- Prefer Dose 3



Dose Preference

- Prefer Dose 2



Goals

- Establish immunologic activity at highest dose

Goals

- Establish immunologic activity at highest dose
- Determine if lower doses are as effective as highest dose

Goals

- Establish immunologic activity at highest dose
- Determine if lower doses are as effective as highest dose
- Avoid ineffective doses

Goals

- Establish immunologic activity at highest dose
- Determine if lower doses are as effective as highest dose
- Avoid ineffective doses
- Monitor toxicity

Simple Randomized Trial

- Global stopping rule for toxicity

Simple Randomized Trial

- Global stopping rule for toxicity
- Equal allocation of participants to dosing arms

Simple Randomized Trial

- Global stopping rule for toxicity
- Equal allocation of participants to dosing arms
- No choice of dose until trial ends

Simple Randomized Trial

- Global stopping rule for toxicity
- Equal allocation of participants to dosing arms
- No choice of dose until trial ends
- Inefficient if some doses are similar or highest dose is ineffective

Simple Randomized Trial

- Global stopping rule for toxicity
- Equal allocation of participants to dosing arms
- No choice of dose until trial ends
- Inefficient if some doses are similar or highest dose is ineffective
- Too many participants?

Two- (or Three-) Stage Randomized Trial with Continual Monitoring (Adapted from Su, 2010)

- Global stopping rule for toxicity

Two- (or Three-) Stage Randomized Trial with Continual Monitoring (Adapted from Su, 2010)

- Global stopping rule for toxicity
- Start by establishing response at highest dose

Two- (or Three-) Stage Randomized Trial with Continual Monitoring (Adapted from Su, 2010)

- Global stopping rule for toxicity
- Start by establishing response at highest dose
- Interim analysis (frequentist), or continual assessment (Bayesian)

Two- (or Three-) Stage Randomized Trial with Continual Monitoring (Adapted from Su, 2010)

- Global stopping rule for toxicity
- Start by establishing response at highest dose
- Interim analysis (frequentist), or continual assessment (Bayesian)
- Stop trial if response to highest dose is not significant

Two- (or Three-) Stage Randomized Trial with Continual Monitoring (Adapted from Su, 2010)

- Global stopping rule for toxicity
- Start by establishing response at highest dose
- Interim analysis (frequentist), or continual assessment (Bayesian)
- Stop trial if response to highest dose is not significant
- Once activity is established at highest dose, add randomization to lower dose

Two- (or Three-) Stage Randomized Trial with Continual Monitoring (Adapted from Su, 2010)

- Global stopping rule for toxicity
- Start by establishing response at highest dose
- Interim analysis (frequentist), or continual assessment (Bayesian)
- Stop trial if response to highest dose is not significant
- Once activity is established at highest dose, add randomization to lower dose
- Interim analysis (frequentist), or continual assessment (Bayesian) comparing doses

Two- (or Three-) Stage Randomized Trial with Continual Monitoring (Adapted from Su, 2010)

- Global stopping rule for toxicity
- Start by establishing response at highest dose
- Interim analysis (frequentist), or continual assessment (Bayesian)
- Stop trial if response to highest dose is not significant
- Once activity is established at highest dose, add randomization to lower dose
- Interim analysis (frequentist), or continual assessment (Bayesian) comparing doses
- If lower dose is not as effective, stop, otherwise, choose lower dose as minimal biologically effective or lower the dose further and repeat

Two-Stage Design, Example 1

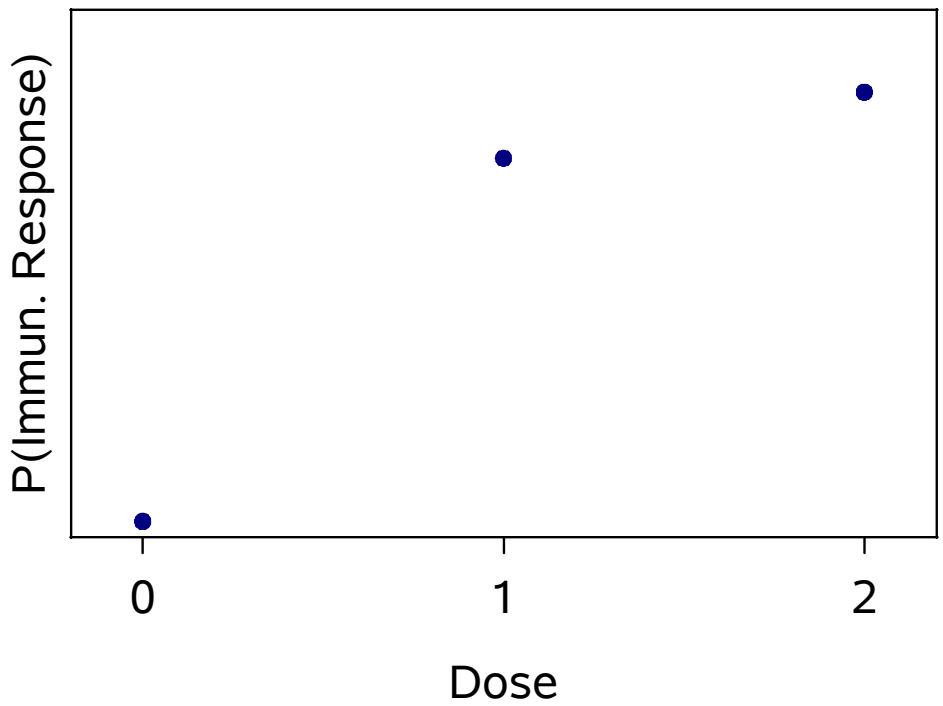
- Two doses

Two-Stage Design, Example 1

- Two doses
- Classify immunologic response ($\pm$)

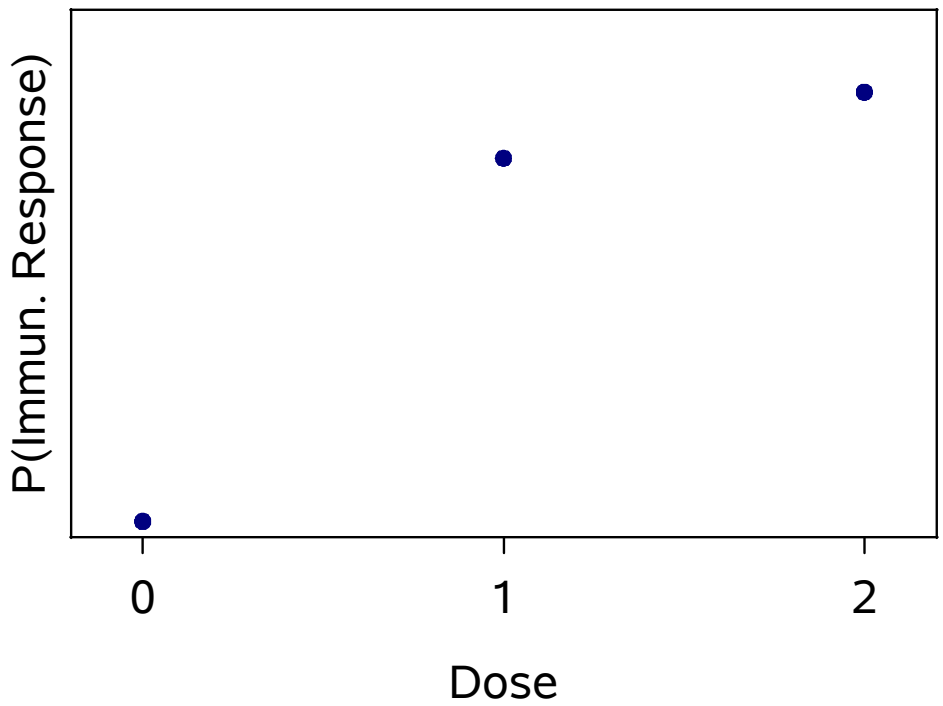
Two-Stage Design, Example 1

- Two doses
- Classify immunologic response ($\pm$)
- True probability of response:



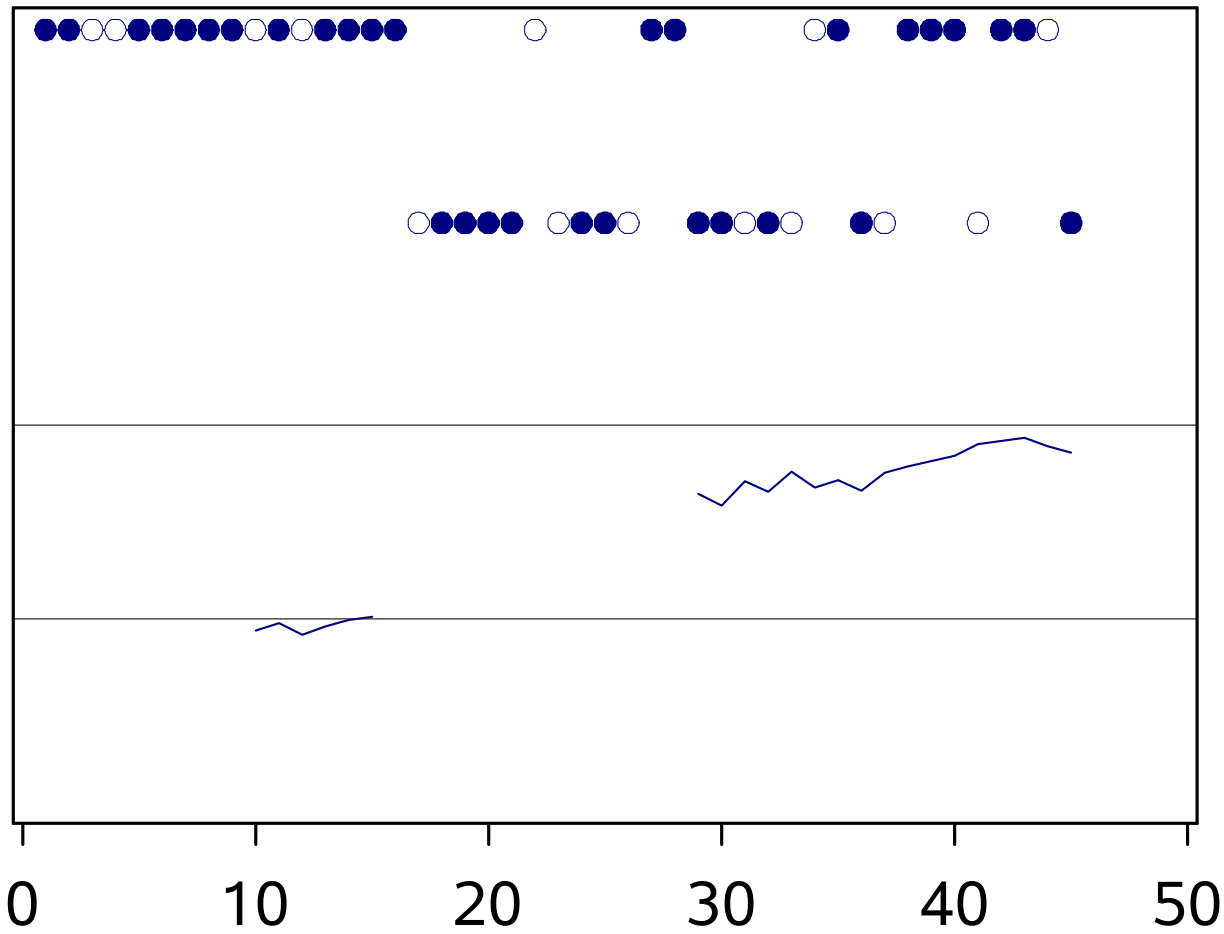
Two-Stage Design, Example 1

- Two doses
- Classify immunologic response ($\pm$)
- True probability of response:



- Dose 1 is preferred

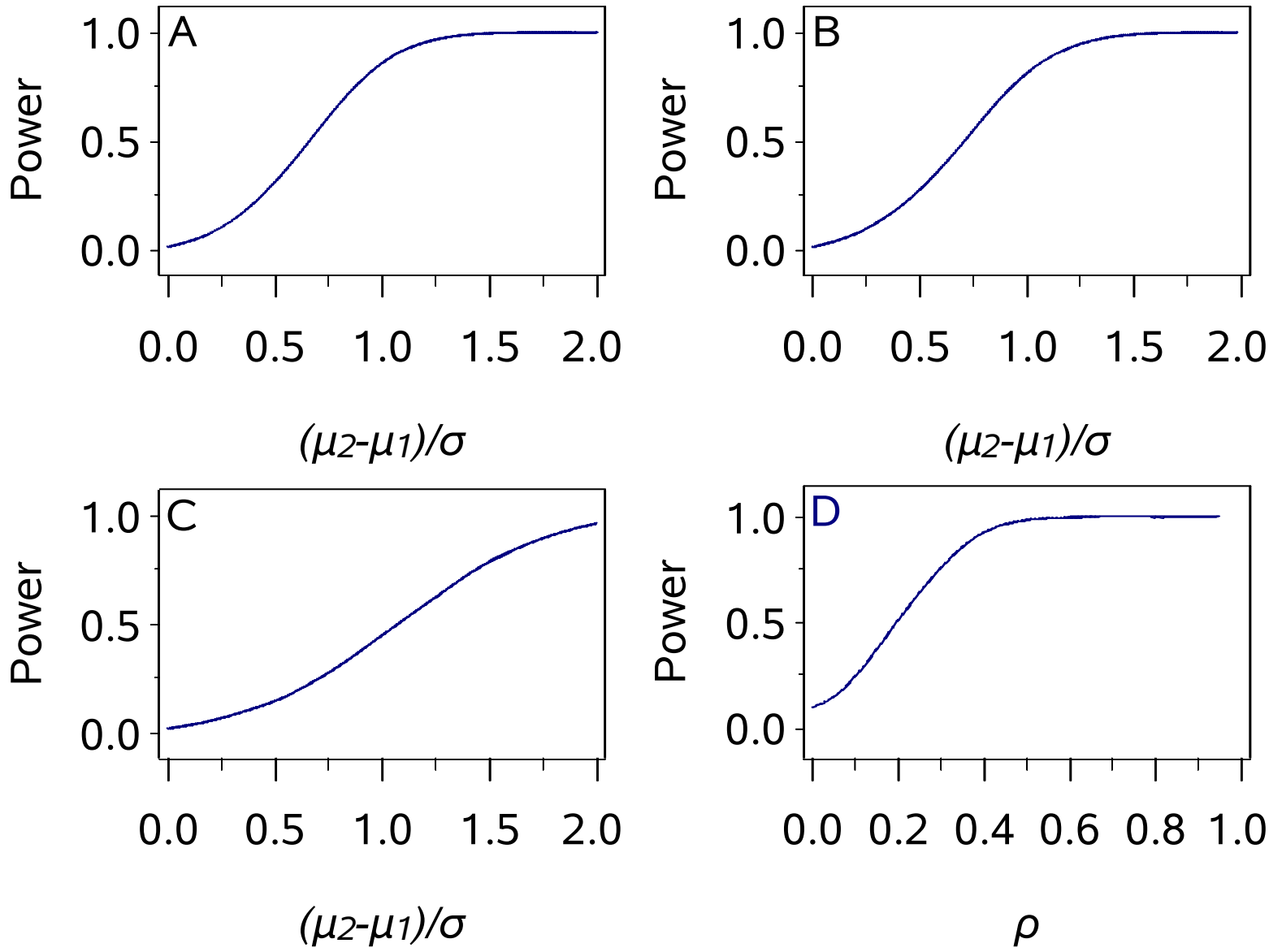
Two-Stage Design, Example 1



Two-Stage Design, Example 1

	# Resp / # Trt	$\hat{p}$	90% CI
High	20/27	0.74	(0.57,0.87)
Low	11/17	0.65	(0.42,0.83)

Two-Stage Design, Example 1



Two-Stage Design, Example 2

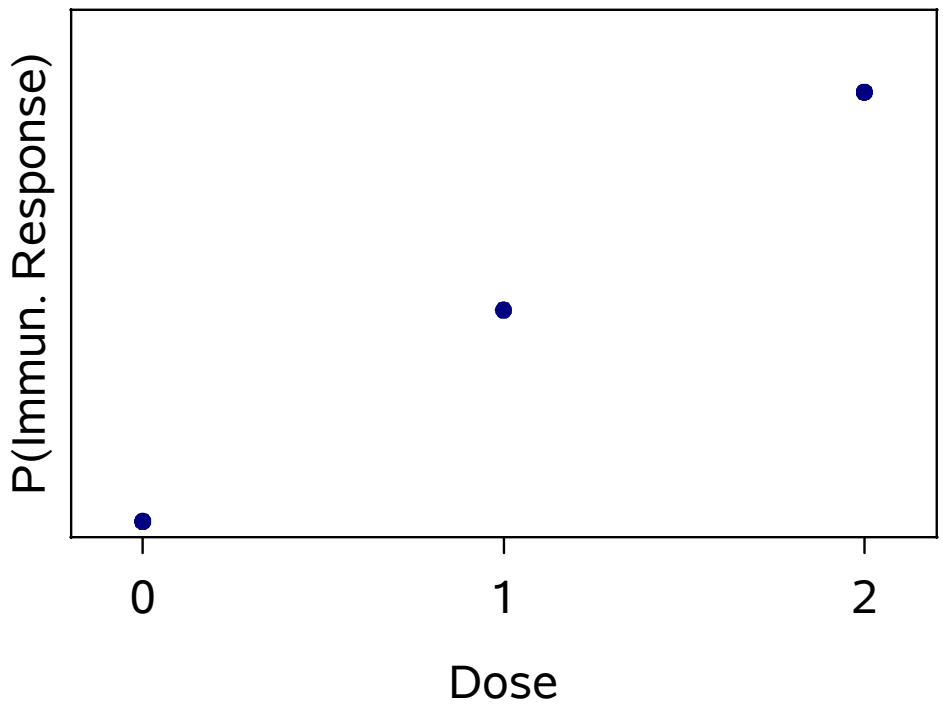
- Two doses

Two-Stage Design, Example 2

- Two doses
- Classify immunologic response ($\pm$)

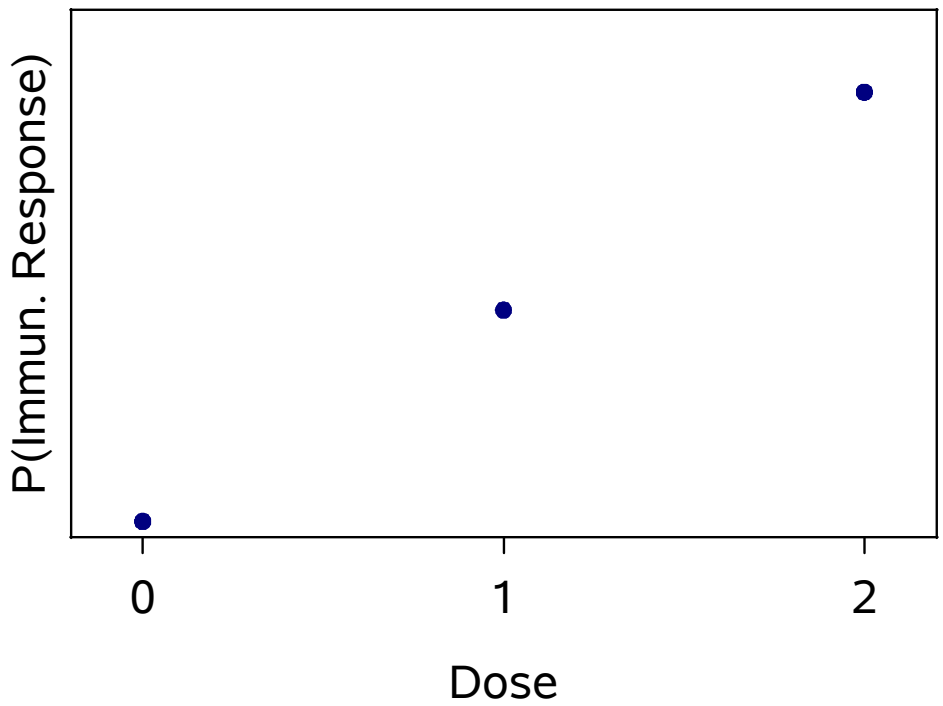
Two-Stage Design, Example 2

- Two doses
- Classify immunologic response ($\pm$)
- True probability of response:



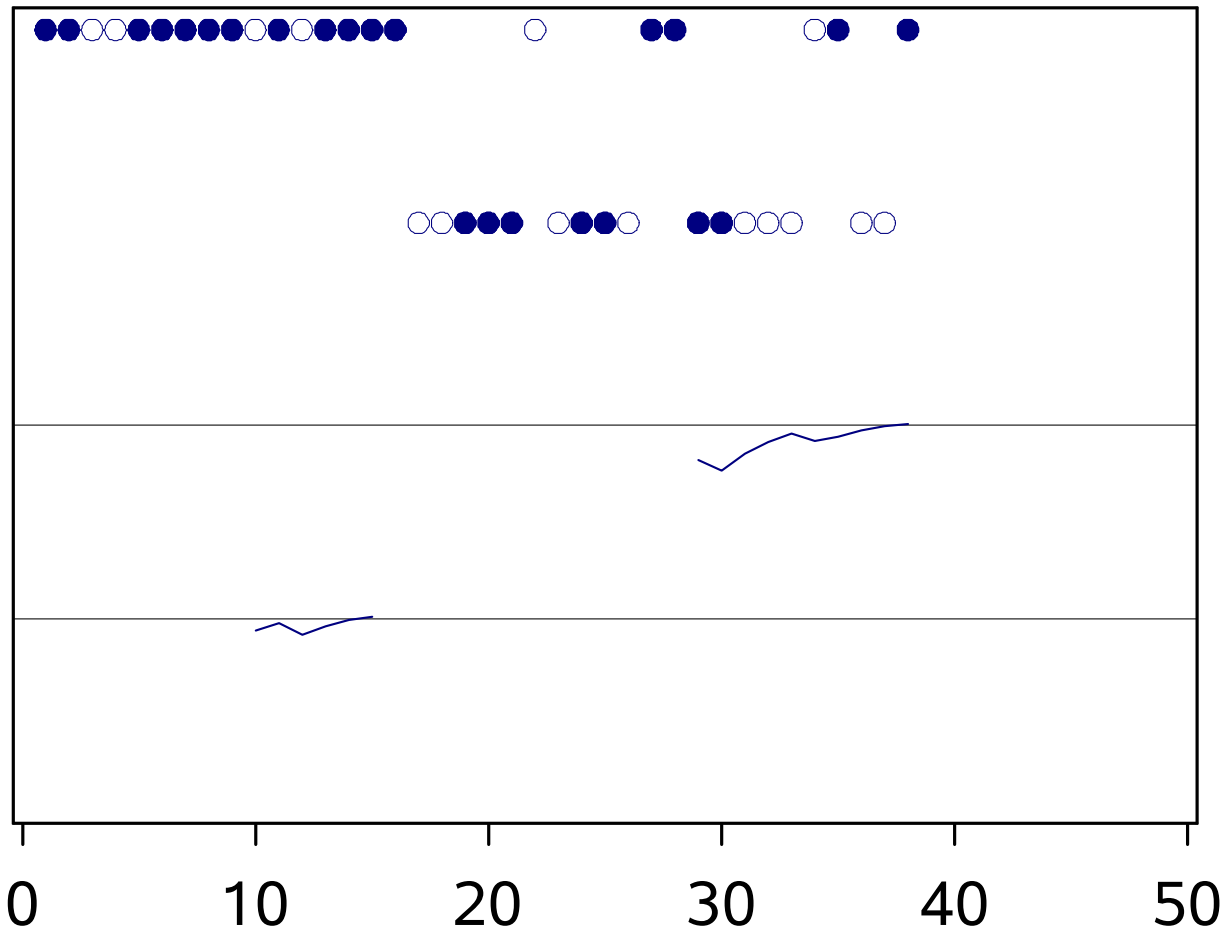
Two-Stage Design, Example 2

- Two doses
- Classify immunologic response ($\pm$)
- True probability of response:



- Dose 2 is preferred

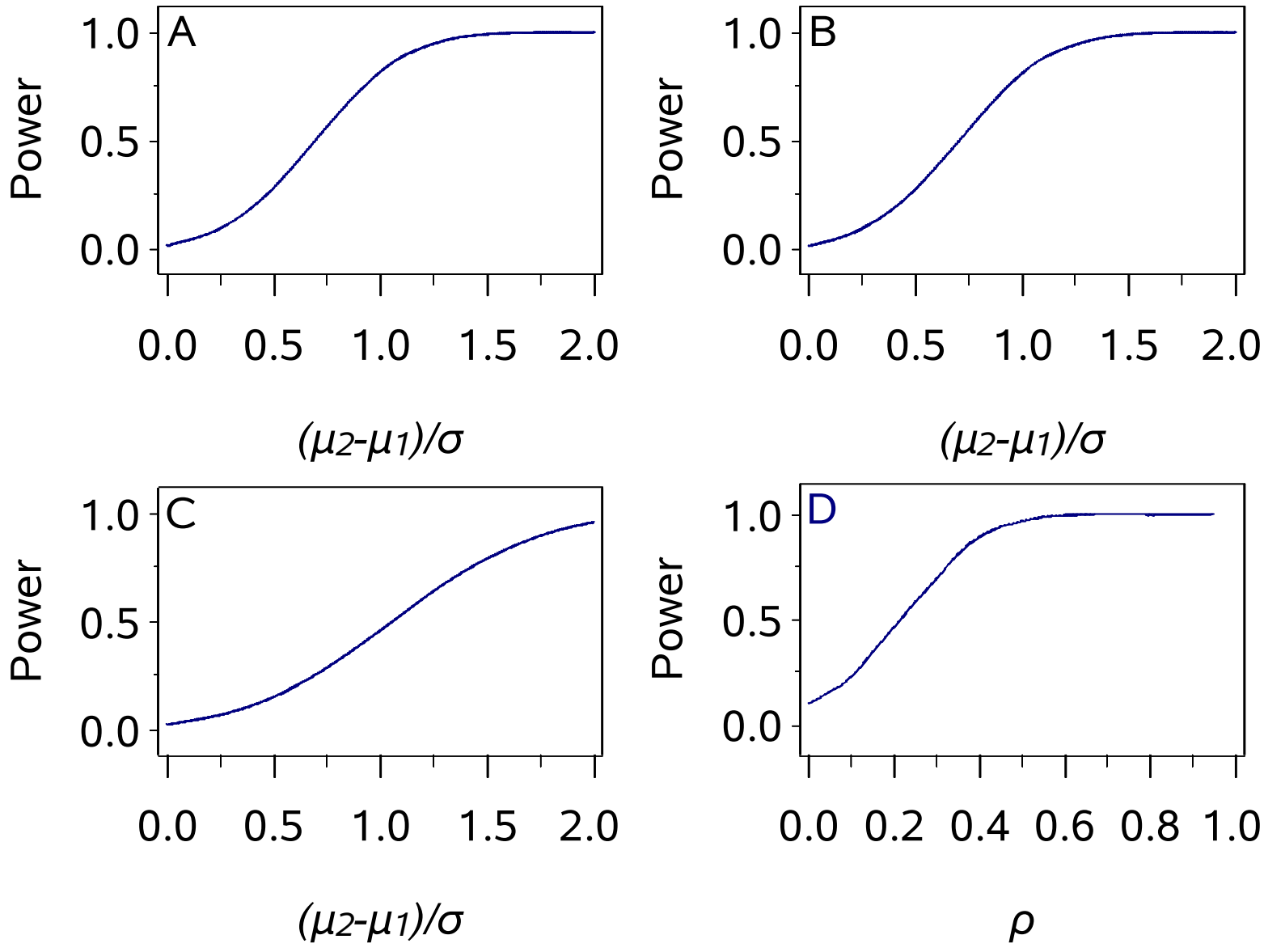
Two-Stage Design, Example 2



Two-Stage Design, Example 2

	# Resp / # Trt	$\hat{p}$	90% CI
High	16/22	0.73	(0.53,0.87)
Low	7/16	0.44	(0.23,0.67)

Two-Stage Design, Example 2



Scenario 2: Combination Therapy to Reduce Toxicity

- Conventional therapy (IL-2) is administered until toxicity occurs

Scenario 2: Combination Therapy to Reduce Toxicity

- Conventional therapy (IL-2) is administered until toxicity occurs
- Add new component to IL-2

Scenario 2: Combination Therapy to Reduce Toxicity

- Conventional therapy (IL-2) is administered until toxicity occurs
- Add new component to IL-2
- Goal: increase number of IL-2 doses administered

Scenario 2: Combination Therapy to Reduce Toxicity

- Conventional therapy (IL-2) is administered until toxicity occurs
- Add new component to IL-2
- Goal: increase number of IL-2 doses administered
- Response-adaptively randomization of dose of new component modulated by number doses of IL-2

Scenario 2: Combination Therapy to Reduce Toxicity

- Conventional therapy (IL-2) is administered until toxicity occurs
- Add new component to IL-2
- Goal: increase number of IL-2 doses administered
- Response-adaptively randomization of dose of new component modulated by number doses of IL-2
- Include control arm (IL-2 only) to verify number of doses can be increased

Scenario 2: Combination Therapy to Reduce Toxicity

- Conventional therapy (IL-2) is administered until toxicity occurs
- Add new component to IL-2
- Goal: increase number of IL-2 doses administered
- Response-adaptively randomization of dose of new component modulated by number doses of IL-2
- Include control arm (IL-2 only) to verify number of doses can be increased
- Accrual goal: 50 participants randomized to four doses (including control)

Outline

Why Do We Persist In Undersizing Studies?

- Expose limited number of patients to potentially ineffective and/or toxic therapies

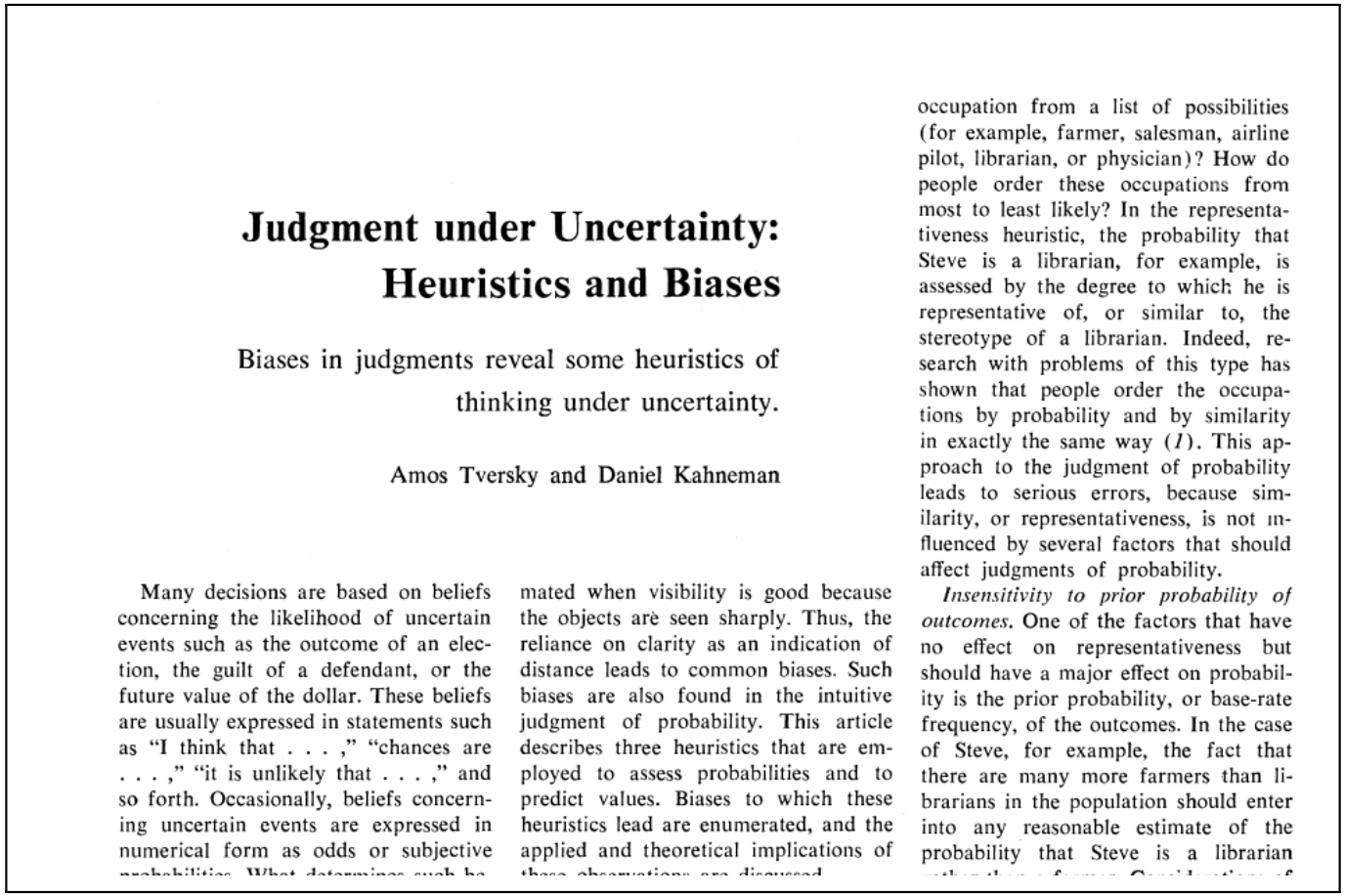
Why Do We Persist In Undersizing Studies?

- Expose limited number of patients to potentially ineffective and/or toxic therapies
- Need to accrue evidence a novel therapy has potential prior to investment

Why Do We Persist In Undersizing Studies?

- Expose limited number of patients to potentially ineffective and/or toxic therapies
- Need to accrue evidence a novel therapy has potential prior to investment
- Institutional pressure to produce trial results quickly

Why Do We Persist In Undersizing Studies?



Judgment under Uncertainty: Heuristics and Biases

Biases in judgments reveal some heuristics of thinking under uncertainty.

Amos Tversky and Daniel Kahneman

Many decisions are based on beliefs concerning the likelihood of uncertain events such as the outcome of an election, the guilt of a defendant, or the future value of the dollar. These beliefs are usually expressed in statements such as "I think that . . . ," "chances are . . . ," "it is unlikely that . . . ," and so forth. Occasionally, beliefs concerning uncertain events are expressed in numerical form as odds or subjective

mated when visibility is good because the objects are seen sharply. Thus, the reliance on clarity as an indication of distance leads to common biases. Such biases are also found in the intuitive judgment of probability. This article describes three heuristics that are employed to assess probabilities and to predict values. Biases to which these heuristics lead are enumerated, and the applied and theoretical implications of

occupation from a list of possibilities (for example, farmer, salesman, airline pilot, librarian, or physician)? How do people order these occupations from most to least likely? In the representativeness heuristic, the probability that Steve is a librarian, for example, is assessed by the degree to which he is representative of, or similar to, the stereotype of a librarian. Indeed, research with problems of this type has shown that people order the occupations by probability and by similarity in exactly the same way (*1*). This approach to the judgment of probability leads to serious errors, because similarity, or representativeness, is not influenced by several factors that should affect judgments of probability.

Insensitivity to prior probability of outcomes. One of the factors that have no effect on representativeness but should have a major effect on probability is the prior probability, or base-rate frequency, of the outcomes. In the case of Steve, for example, the fact that there are many more farmers than librarians in the population should enter into any reasonable estimate of the probability that Steve is a librarian

- Tversky & Kahneman, *Science* 1974
- We use heuristics that tend to overweight the evidence from the first few data points in a series

Novel Designs are Difficult

- Novel designs require extensive preparation

Novel Designs are Difficult

- Novel designs require extensive preparation
- Consensus on primary objectives

Novel Designs are Difficult

- Novel designs require extensive preparation
- Consensus on primary objectives
- Concordance on next step in developmental program

Novel Designs are Difficult

- Novel designs require extensive preparation
- Consensus on primary objectives
- Concordance on next step in developmental program
- Larger trials are difficult for young investigators

Adaptive Trials are Not Magic

- Reductions in sample size may be elusive

Adaptive Trials are Not Magic

- Reductions in sample size may be elusive
- Implementation may be complicated

Adaptive Trials are Not Magic

- Reductions in sample size may be elusive
- Implementation may be complicated
- Data management and analysis must be prompt

Adaptive Trials are Not Magic

- Reductions in sample size may be elusive
- Implementation may be complicated
- Data management and analysis must be prompt
- Comparisons of results from trials with different designs is challenging

Can't We All Just Get Along?

- Use development process to determine the primary objective

Can't We All Just Get Along?

- Use development process to determine the primary objective
- Realistically power the study to achieve primary objective

Can't We All Just Get Along?

- Use development process to determine the primary objective
- Realistically power the study to achieve primary objective
- Assess all the relevant operating characteristics of a given design

Can't We All Just Get Along?

- Use development process to determine the primary objective
- Realistically power the study to achieve primary objective
- Assess all the relevant operating characteristics of a given design
- Employ novel designs if we really think they will help

Can't We All Just Get Along?

- Use development process to determine the primary objective
- Realistically power the study to achieve primary objective
- Assess all the relevant operating characteristics of a given design
- Employ novel designs if we really think they will help
- Mutually engage bench scientists, trialists and statisticians during design, implementation and analysis

Acknowledgements and Support

- Heidi Weiss
- Charity Moore
- Bill Gooding
- Mike Lotze
- NIH/NCI R01CA148713