Bayesian Methods in Regulatory Science

Gary L. Rosner, Sc.D.

Regulatory-Industry Statistics Workshop
Washington, D.C.   13 September 2018

What is Regulatory Science?

- US FDA

  ▸ Regulatory Science is the science of developing new tools, standards, and approaches to assess the safety, efficacy, quality, and performance of all FDA-regulated products.

What Do I Mean by "Bayesian Approaches"?

- Truly Bayes:

  ▸ Choose design points to maximize expected utility

    - E.g., sample size for certain precision at fixed cost

- "Stylized" Bayes (or "calibrated" Bayes, etc.)

  ▸ Bayesian formulation of study data

  ▸ Choose design points to achieve "desirable" frequentist operating characteristics

    - E.g., choose prior parameters to attain monitoring rules good sensitivity, specificity, & expected N

Acceptance When Want to Borrow Information

- Change in device from previous version

  ▸ Historical information reasonable if minor change

- Population pharmacokinetics & pharmacodynamics

  ▸ Population modeling & hierarchical models

- Pediatric drug development

  ▸ Borrow information

- Rare diseases or adverse events

- Early phase studies (e.g., phases 1 & 2)

When Are Bayesian Statistical Methods Not Accepted

- When want to be "objective"

  ‣ Concern about influence of prior distribution

    - *But frequentist methods are subjective, too!*

      ‣ Model,

      ‣ Null & alternative hypotheses,

      ‣ Incorporating current trial in context of others

- Confirmatory drug studies

Sheiner: Two Major Learn-Confirm Cycles in Clinical Drug Development

- 1st cycle: *Clin Pharmacol Ther* 61:275-91, 1997

  ‣ Phase 1: Learn what dose is tolerated

  ‣ Phase 2: Confirm dose has promise of efficacy

  - Make decision based on this learn-confirm cycle

- 2nd cycle:

  ‣ Phase 2B: Learn how to use the drug in patients

  ‣ Phase 3: Confirm in large representative pt pop'n that therapy achieves acceptable benefit:risk ratio

  - If acceptable, approval is granted

Sheiner (cont'd)

- Learning & confirming are distinct

  - Different goals, designs, methods of analysis

    - Analysis choice: Hypothesis testing or estimation?

    - Learning involves estimation

      - "The [B]ayesian view is well suited to this task because it provides a theoretical basis for learning from experience; that is, for updating prior beliefs in the light of new evidence."

        Clin Pharmacol Ther 61:275-91, 1997

    - Confirming involves hypothesis testing

Neyman & Pearson

- Need more than a test based on probability to establish truth of a particular hypothesis

THE MOST EFFICIENT TESTS OF STATISTICAL HYPOTHESES. 291

as far as a particular hypothesis is concerned, no test based upon the theory of probability* can by itself provide any valuable evidence of the truth or falsehood of that hypothesis.

But we may look at the purpose of tests from another view-point. Without hoping to know whether each separate hypothesis is true or false, we may search for rules to govern our behaviour with regard to them, in following which we insure that, in the long run of experience, we shall not be too often wrong. Here, for example, would be such a "rule of behaviour": to decide whether a hypothesis, H, of a given type be

Neyman, J., and Pearson, E. S. (1933). On the Problem of the Most Efficient Tests of Statistical Hypotheses. *Philosophical Transactions of the Royal Society of London. Series A*, 231, 289-337.

R. A. Fisher

- Discusses forming summary statistics & evaluating deviation relative to a dist'n (for "tests of significance").

only once in 370 trials, while Table II. shows that to exceed the standard deviation sixfold would need nearly a thousand million trials. The value for which $P = \cdot 05$, or 1 in 20, is 1·96 or nearly 2 ; it is convenient to take this point as a limit in judging whether a deviation is to be considered significant or not. Deviations exceeding twice the standard deviation are thus formally regarded as significant. Using this criterion,

Fisher, R. A. (1934). *Statistical Methods for Research Workers*, 5th ed. Edinburgh: Oliver and Boyd.

SN Goodman: Toward Evidence-Based Medical Statistics 1 & 2

- A p value seems to provide

  ‣ a measure of evidence against H_0 from this study

  and

  ‣ a means to control Type 1 error regarding rejecting H_0 in the long run.

- But, no single number can do both jobs!

  ‣ Describe long-run behavior & meaning in this study.

  ‣ Treat this study as one of infinitely many & provide a measure of evidence for results in this study.

Goodman, S. N. (1999). Toward Evidence-Based Medical Statistics 1 & 2 *Ann Intern Med* 130, 995–1004 & 1005-1013.

Hypothesis Testing

- Answer a question

  ▸ Do observations agree with predictions based on one hypothesis more than the other? $\Pr(\mathbf{y} \mid H_0) \text{ vs. } \Pr(\mathbf{y} \mid H_A)$

- N-P testing

  ▸ Decision rule to limit long-run risks of errors

- Bayesian testing

  ▸ Compare $\Pr(H_A \mid \mathbf{y})$ to $\Pr(H_0 \mid \mathbf{y})$ Condition on *data*

    - Could be with Bayes factor and Jeffreys's criteria

$$\frac{\Pr(H_A \mid \mathbf{y})}{\Pr(H_0 \mid \mathbf{y})} \bigg/ \frac{\Pr(H_A)}{\Pr(H_0)}$$

Which Do We Want?

- Evidence one trt is superior to the other?

  ‣ Estimate with precision?

- Decision rule regarding hypothesis?

  ‣ "Yes" reject? "No" do not reject?

- Decision rule regarding next step?

  ‣ Continue to next phase of study?

  ‣ Approve the treatment for indication?

Decision Theory & Clinical Trials

- Why decision theory?

  ‣ Clinical trials: Purpose is to lead to decisions

    - What dose(s) to use?

    - How best to apply the therapy?

    - What is the next step for evaluating therapy?

    - Should patients receive this therapy from now on?

    - Which patients receive the most benefit?

  Why not make decisions explicit and coherent?

  ‣ Put results in context via formal decision analysis

13

Decision Theory & Clinical Trials

- Clinical trial design involves decisions, too

  ‣ Sample size,

  ‣ Duration of follow-up,

  ‣ Stopping rules,

  ‣ Whether to run the study in the 1st place

Why not make these decisions explicit and coherent?

How to Make a Big Decision

Have no fear. An emerging science can now help you choose.

By Steven Johnson

Mr. Johnson writes about science and the history of innovation.

- "Value model"

  ▸ Weight each "value" (utility for each outcome)

  ▸ Develop scenarios (simulate the trial)

  ▸ "Multiply each grade by the weight of each value and add up the numbers for each scenario. The scenario with the highest score wins."

Decision Theory

- Consider

  ▸ Set of possible actions: $\mathcal{A}$

    - E.g., stop the study; move to phase 3

  ▸ Set of possible outcomes: $\mathbf{Y}$

  ▸ Parameters characterizing stochastic nature: $\mathbf{\Theta}$

    - E.g., treatment effect, model parameters

  ▸ Utility function: $u(a)$

Bayesian Optimal Design

- "Bayes" action maximizes expected utility

  ▸ Expectation to account for sources of uncertainty

    - Uncertainty in parameters $p(\theta)$

    - Variation in data resulting from action $p_a(y \mid \theta)$

$$\mathcal{U}(a) = \int_{\mathbf{Y}} \int_{\mathbf{\Theta}} u(y) p_a(y \mid \theta) p(\theta) d\theta dy$$

    - Choose: $a^* = \arg\max \mathcal{U}(a)$

Application

- Cancer & Leukemia Group B wanted to study Taxol

  ‣ Large population of women

  ‣ 3-hour infusion

 ‣ Many participating hospitals

  ‣ Outpatient

Problem

- Cannot carry out extensive sampling

    ▸ Large study

    ▸ Many institutions

- Devise limited-sampling scheme

    ▸ Optimal sampling times

Stroud JR, Müller P, Rosner GL. Optimal sampling times in population pharmacokinetic studies. Applied Statistics. 2001;50(3):345-59.

Objective: Maximize Precision

Toxicity risk = $f(AUC)$ or $g(T_{cv})$

AUC

T_{cv}
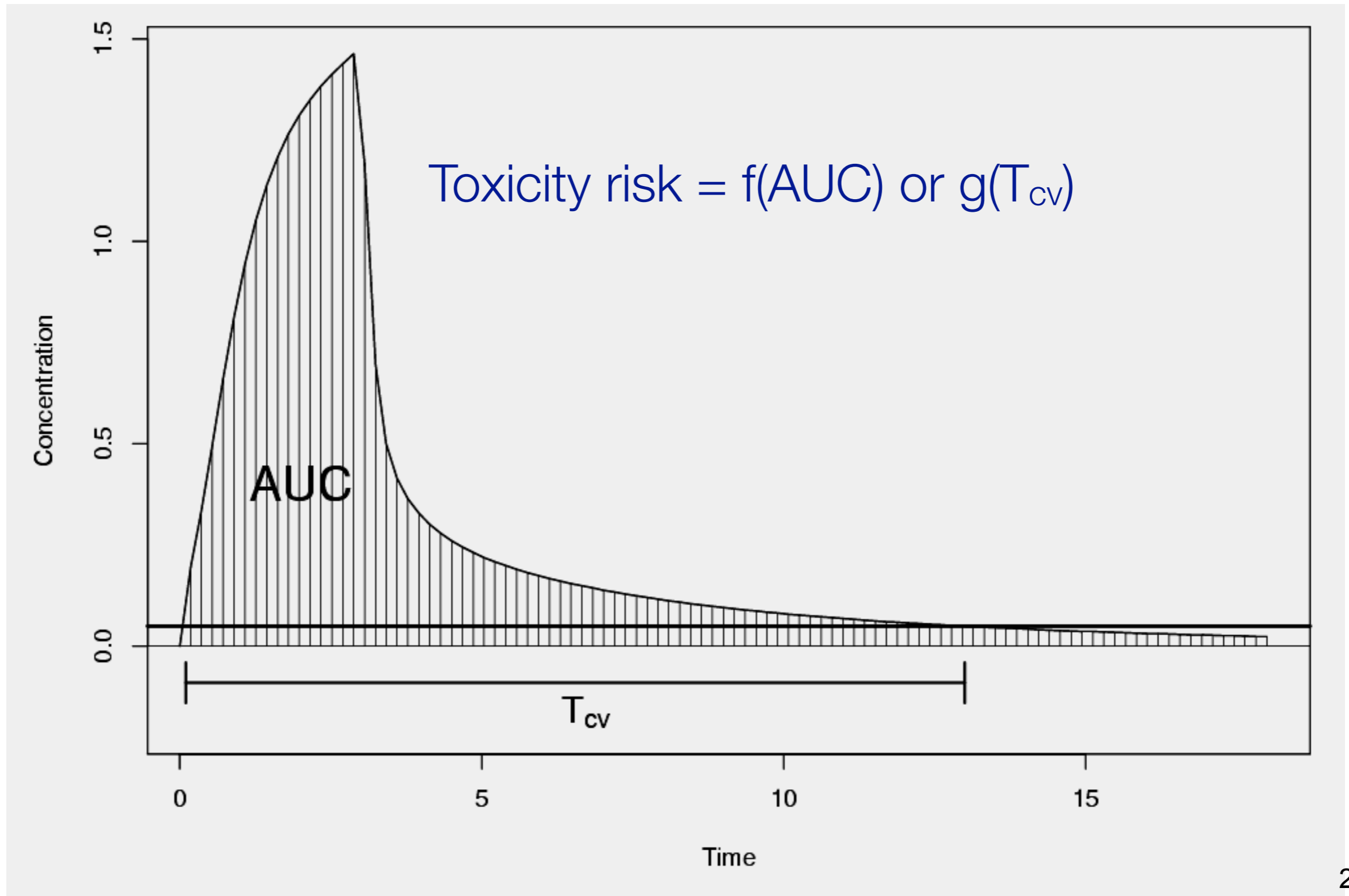
Paclitaxel PK Sampling Times
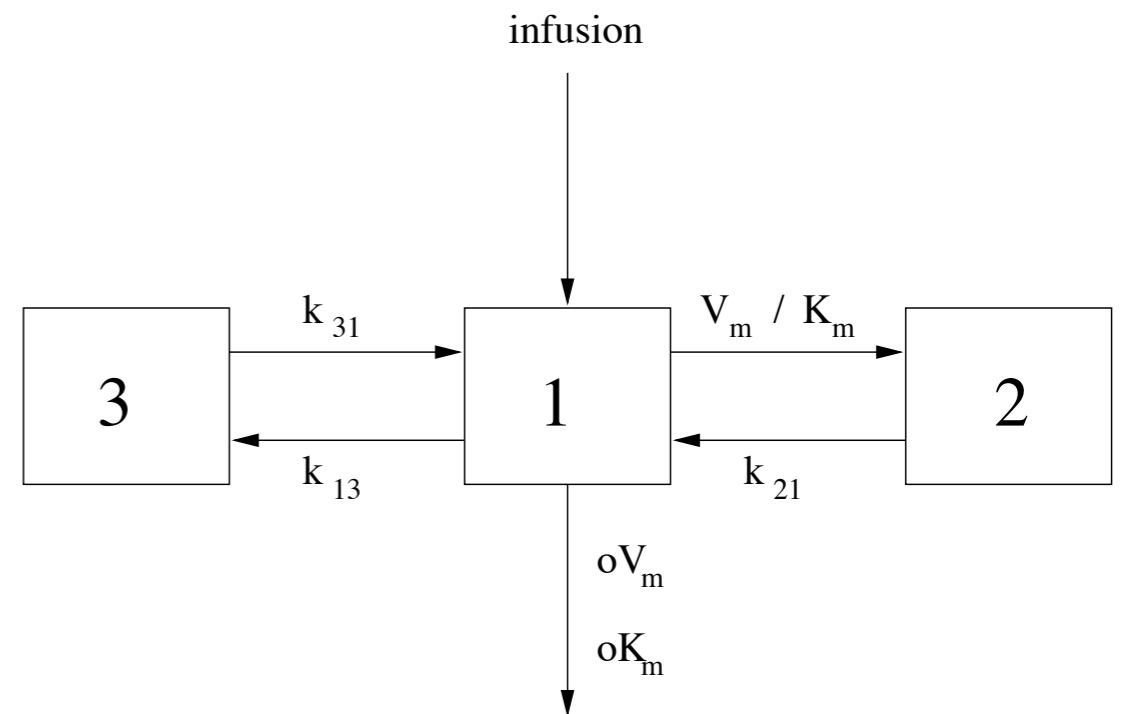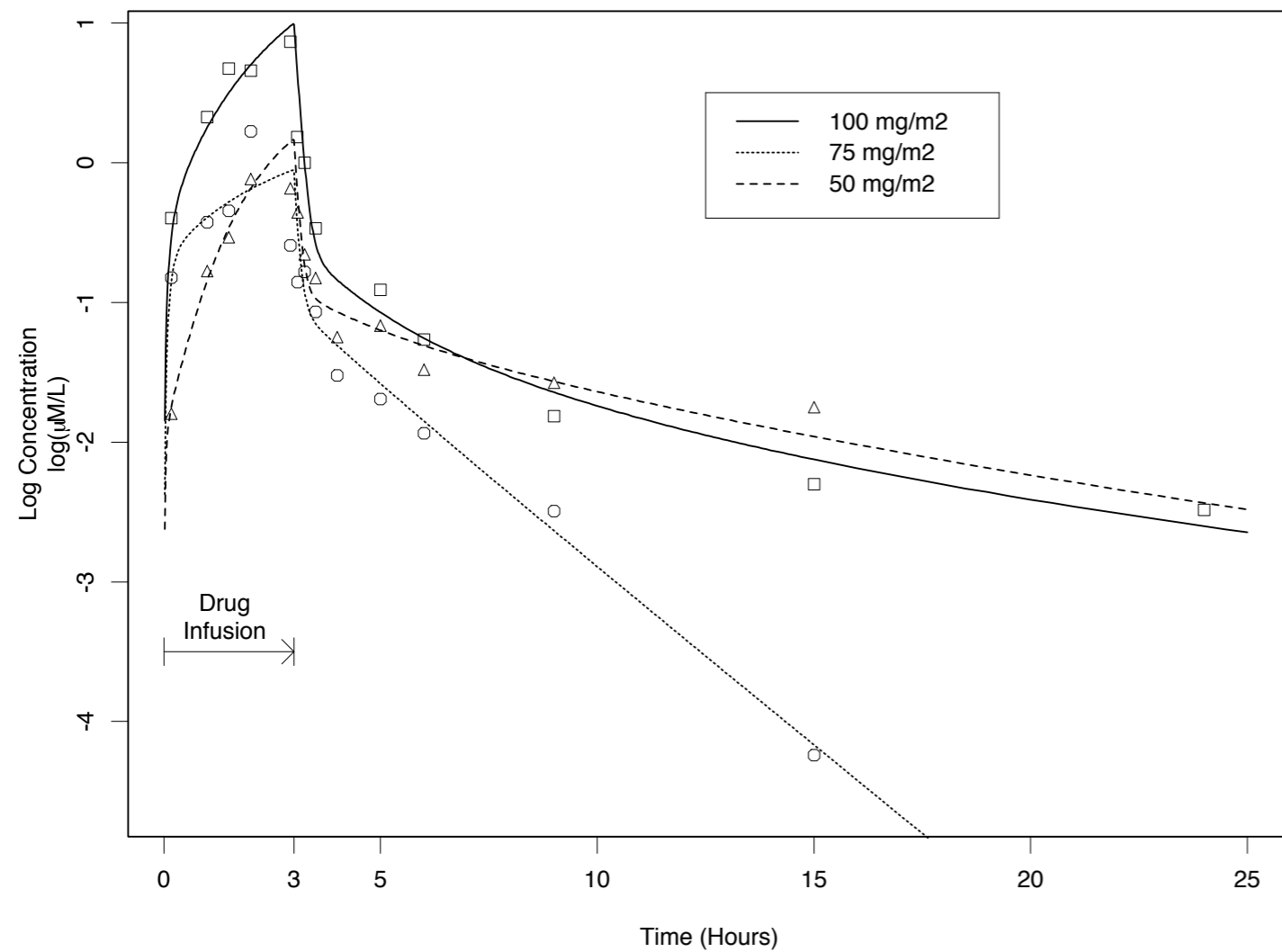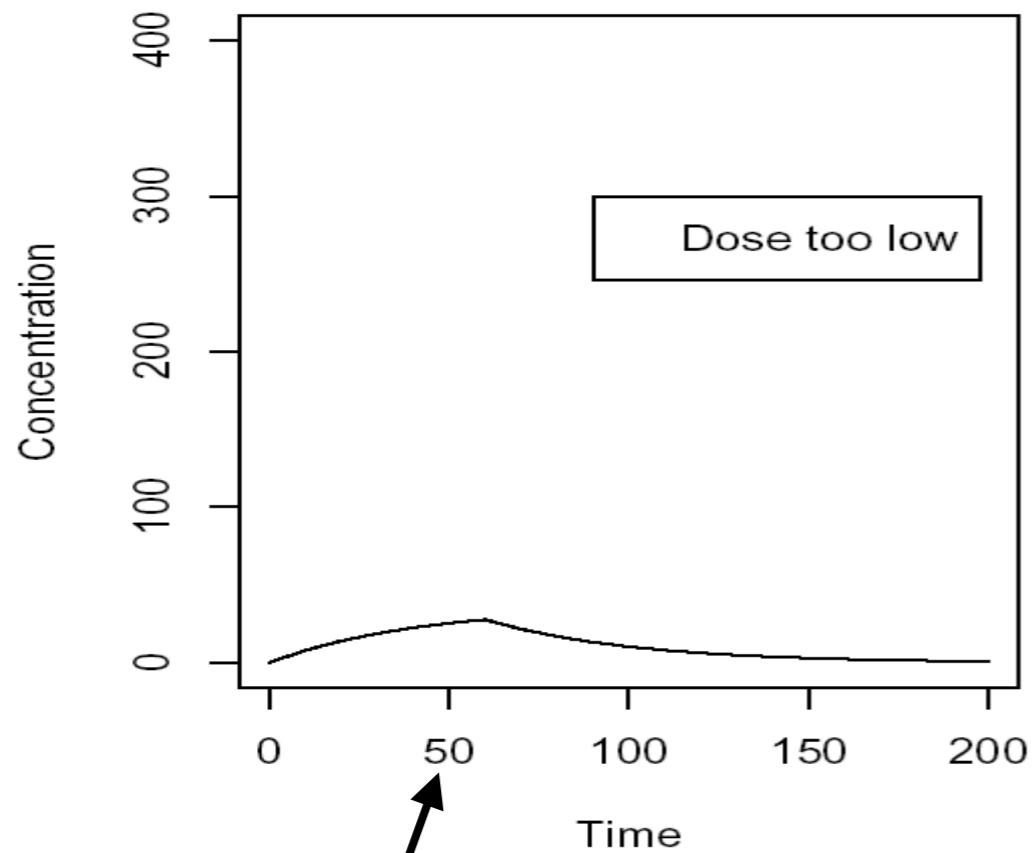
Optimal Sampling Times for AUC

$$u(t, y) =$$

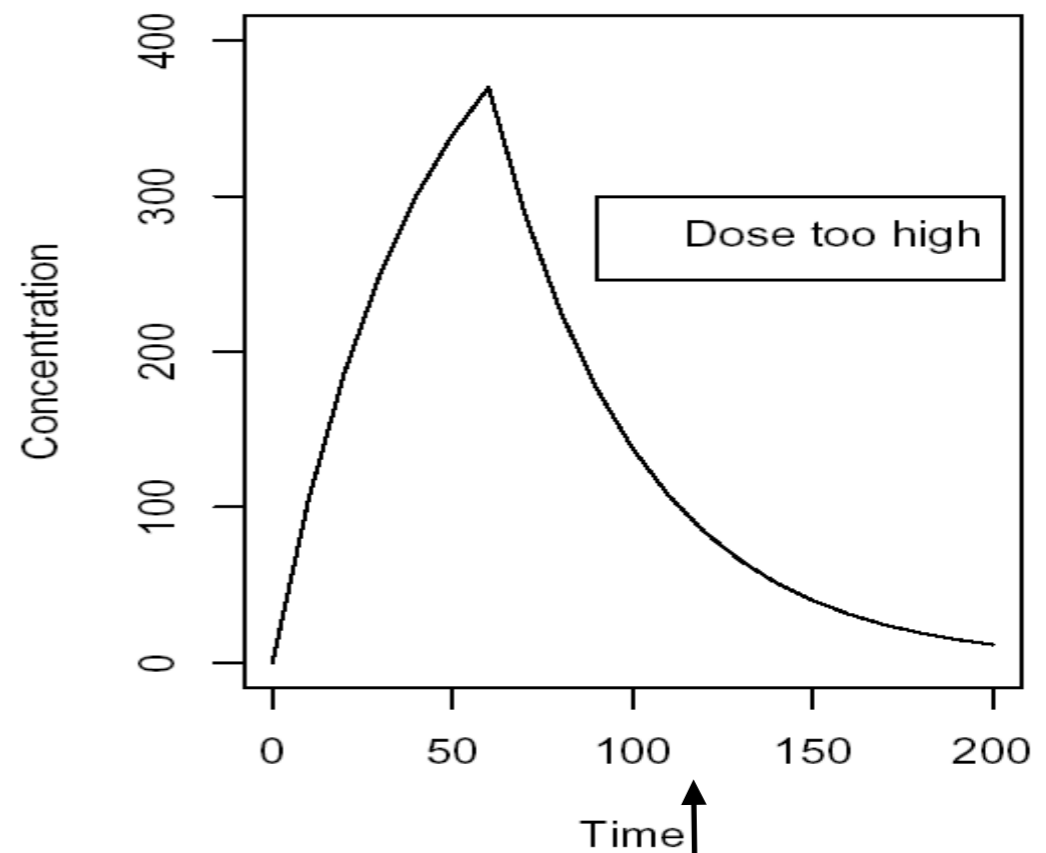$$\left[\int \{\phi(\theta) - E[\phi(\theta) \mid y, Y, t]\}^2 p(\theta \mid y, Y, t) d\theta \right]^{-1}$$

$$- k \sum_{i=1}^{p} t_i^2 I_{\{t_i > 8\}}$$

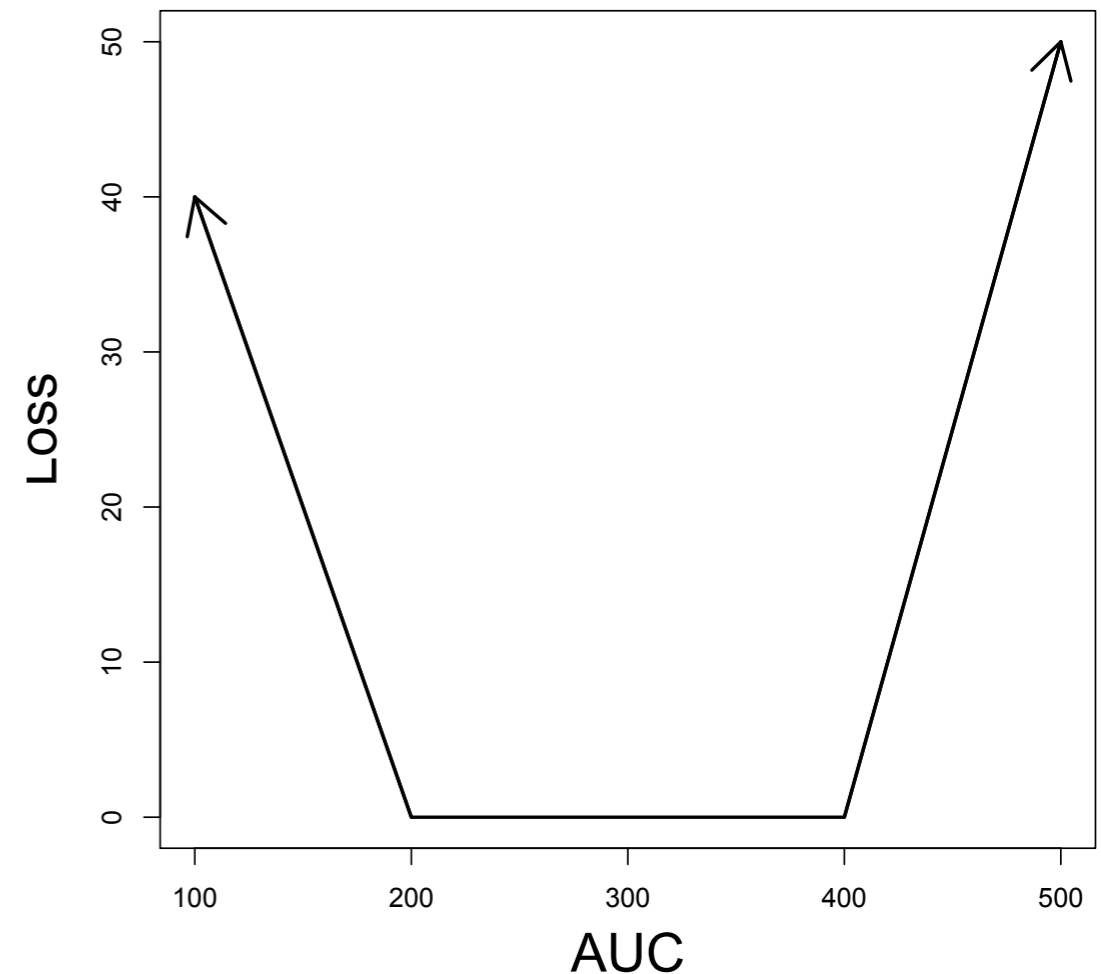| Cost Coeff k	n=1		n=2	
	t^*	U^*	t^*	U^*
0.00000	(3)	0.53	(3,25)	0.90
0.00004	(3)	0.53	(3,10)	0.74
0.00008	(3)	0.53	(3,7)	0.70

Dose Optimization

Asymmetric Loss Function
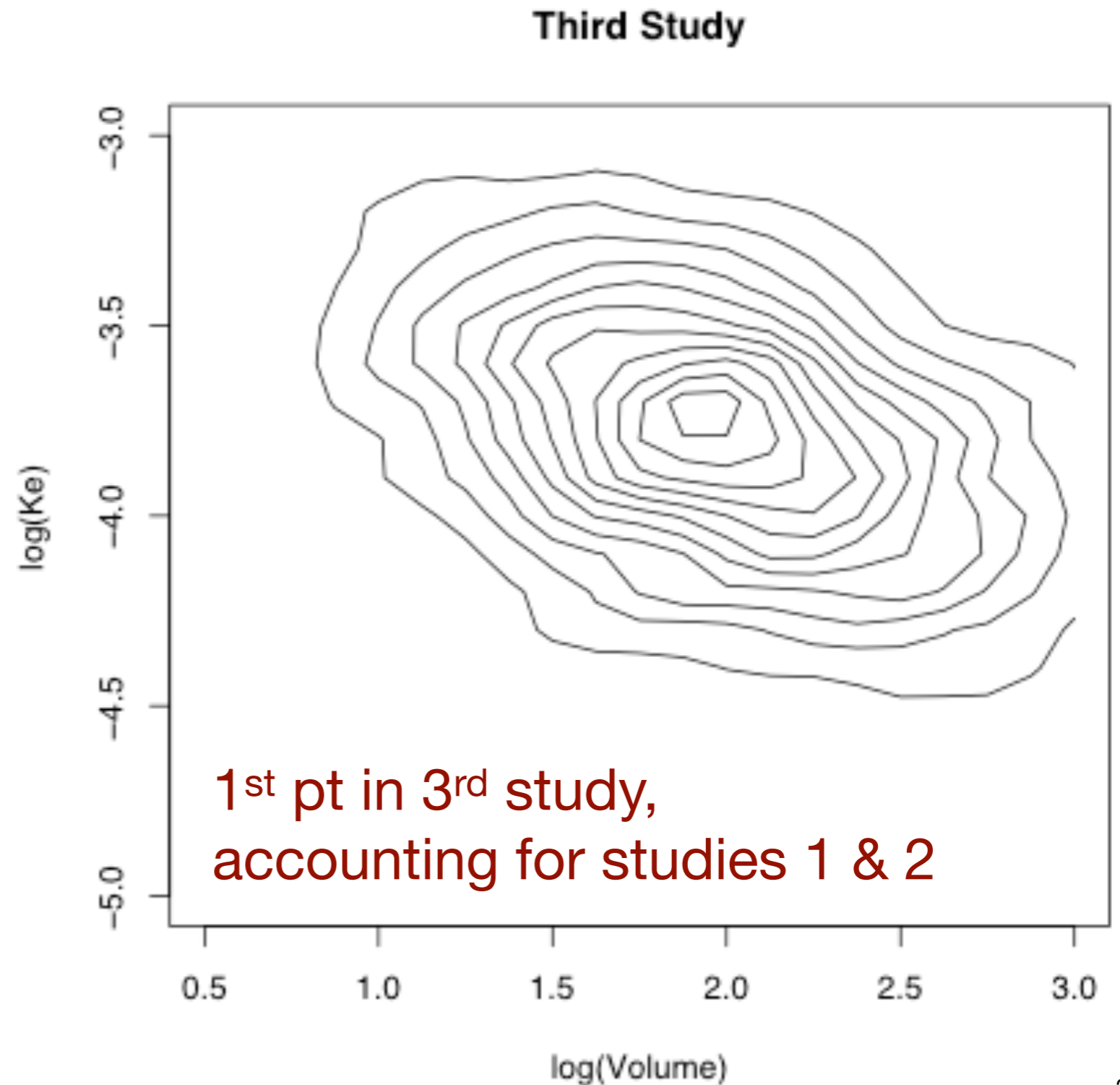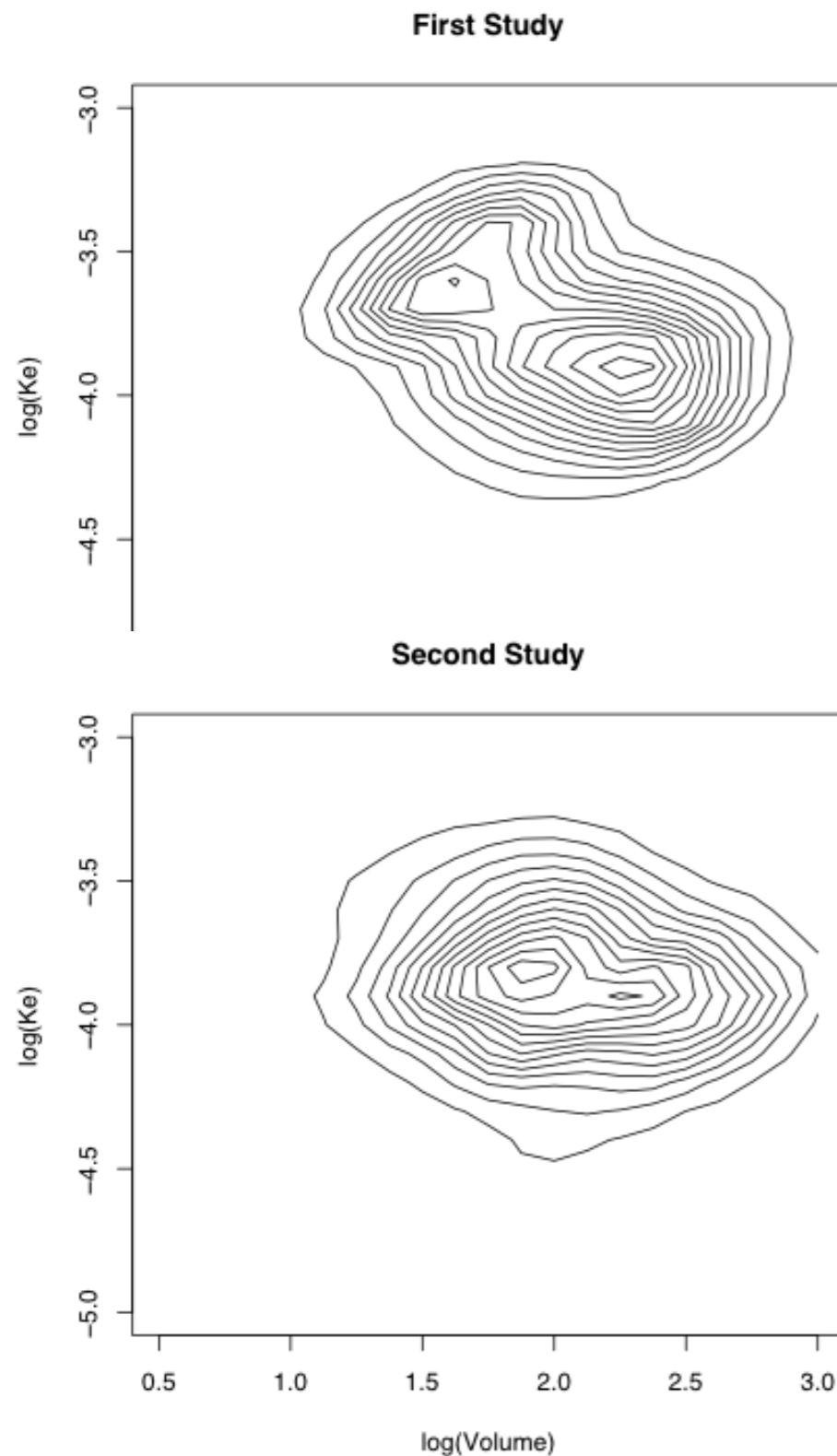
- Want AUC in "optimal" range

$$AUC_{ll} \leq AUC \leq AUC_{ul}$$

- Loss function

$$L(auc) = \begin{cases} L^-(auc, \text{AUC}_{ll}) & \text{if } auc < \text{AUC}_{ll} \\ 0 & \text{if } \text{AUC}_{ll} \leq auc \leq \text{AUC}_{ul} \\ L^+(auc, \text{AUC}_{ul}) & \text{if } auc > \text{AUC}_{ul} \end{cases}$$

Posterior for Pt's PK Parameters

First Study

Second Study

Third Study

1st pt in 3rd study, accounting for studies 1 & 2

Optimal Dose w.r.t. Posterior

$$E[u(y, d, \theta)] = \int \mathrm{L}\left[AUC(y)\right] p(y \mid d, \theta) p(\theta \mid \mathrm{Data}) \, dy \, d\theta$$
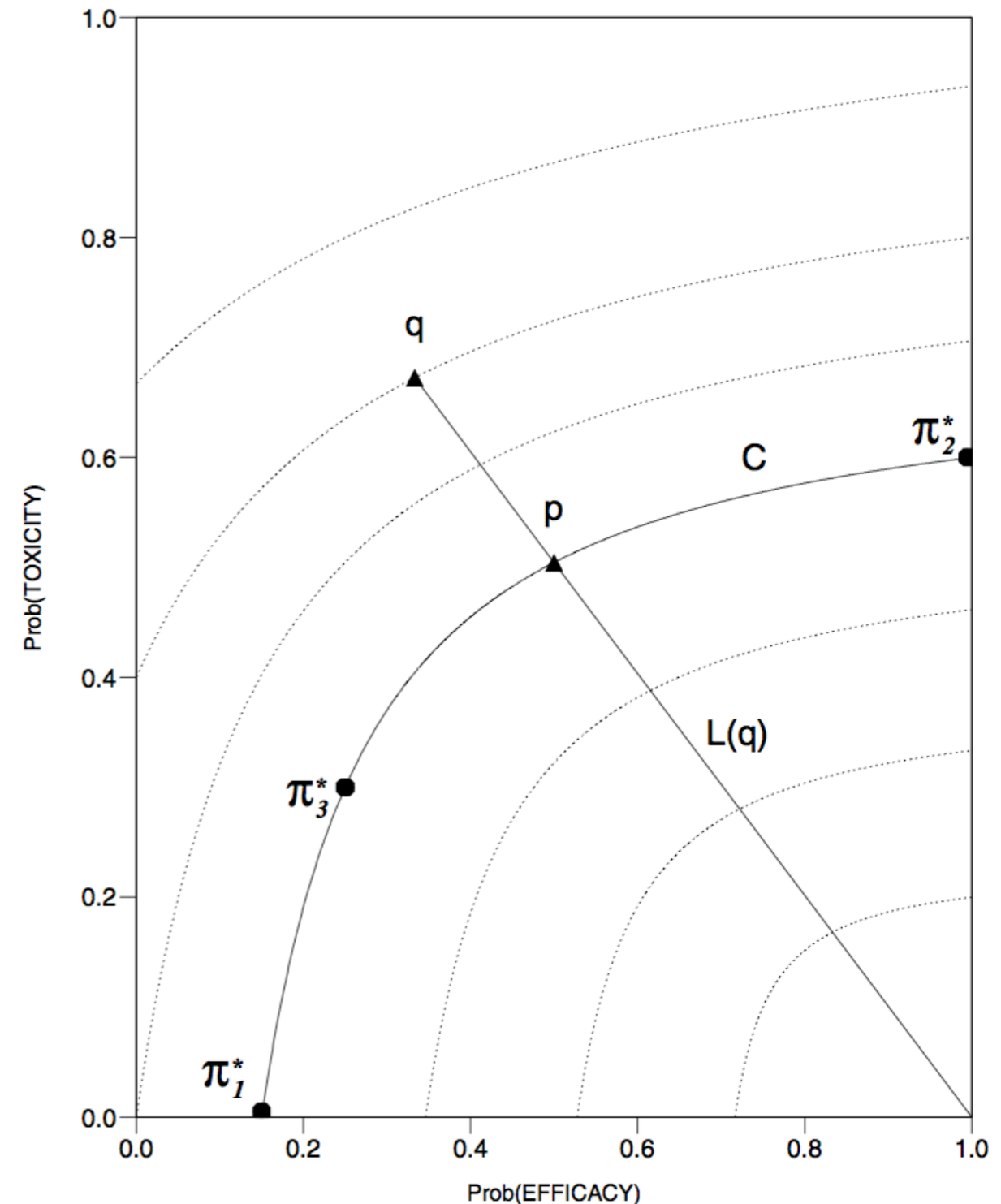
Loss Function

Expected Loss

Optimal Dose = 1.9
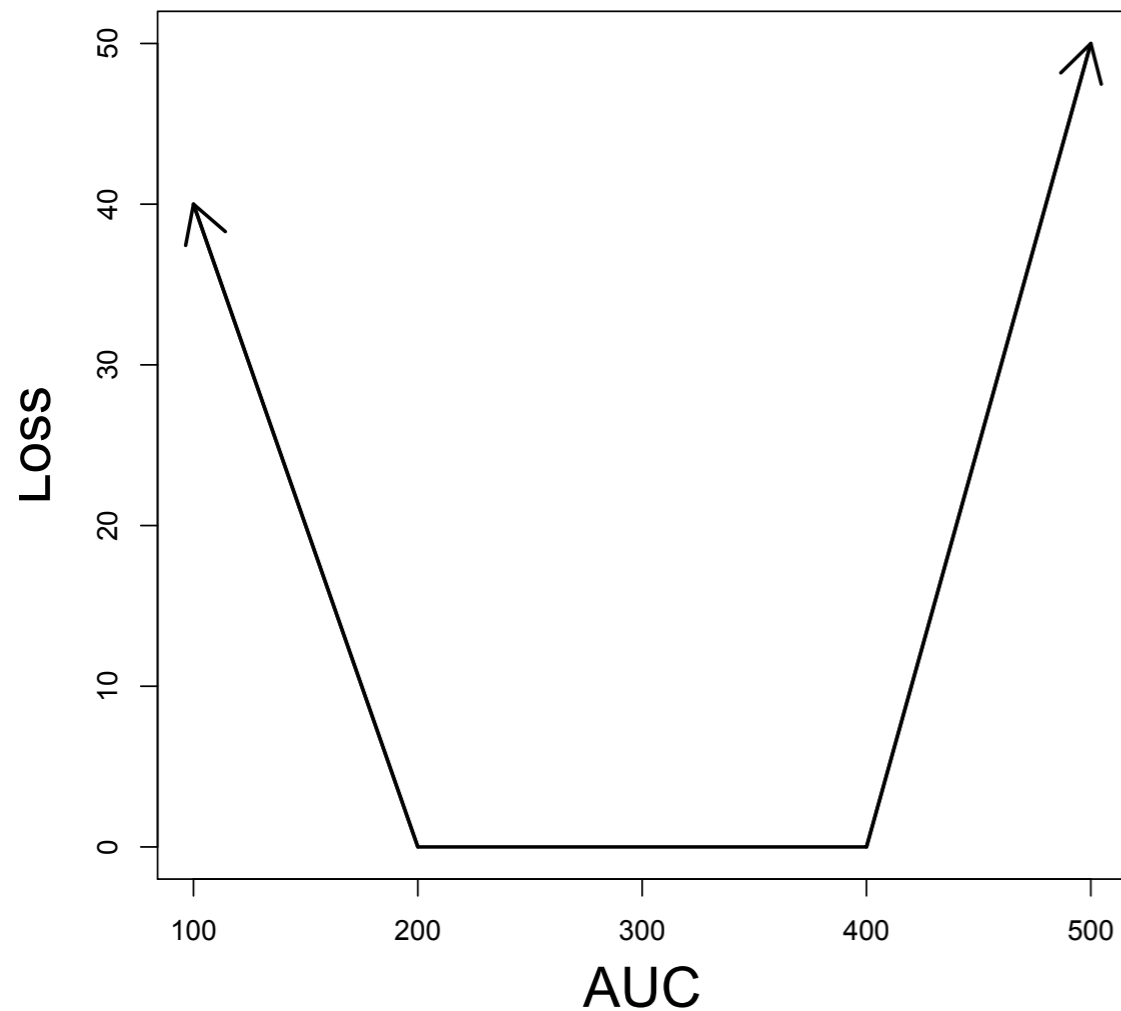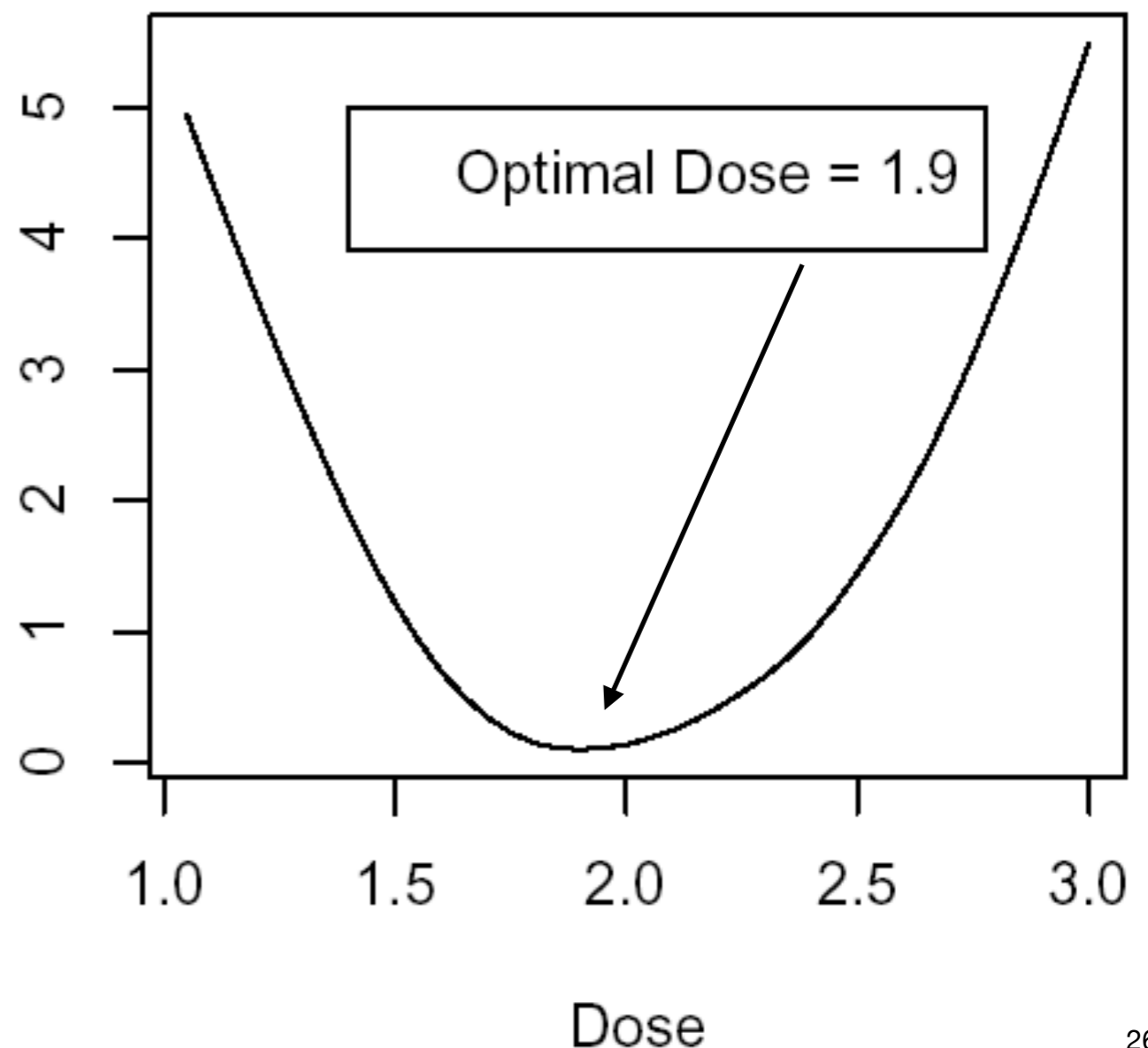
Learn About Benefit:Risk

- Weigh chance of benefit against risk of adverse event

  ▸ Dose finding based on trade-offs between probs of efficacy and toxicity

Thall PF, Cook JD (2004). Dose-Finding Based on Efficacy-Toxicity Trade-Offs. *Biometrics* 60, 684-693.

Trippa: Response-Adaptive Designs for Basket Trials

- Strategies for choosing among baskets

- Objective function is Bayesian decision rule (Exp. utility)

  ‣ Decisions during study work toward maximum information at the end of the study

  ‣ Fully sequential hard!

  ‣ Bayesian Uncertainty Directed (BUD) designs

    - Approximate optimal decision rules (myopic)

    - Goal: Design rule that approximates Bayesian decision and that one can compute

Utility Function: Information Measure

- Information measure consistent with goal of study

  ▸ E.g., entropy of estimate, posterior variance, entropy of posterior distribution of an indicator

    - Entropy of posterior distribution

$$X \sim p(X), \quad H[p(X)] = -E[\log p(X) \mid \text{Data}]$$

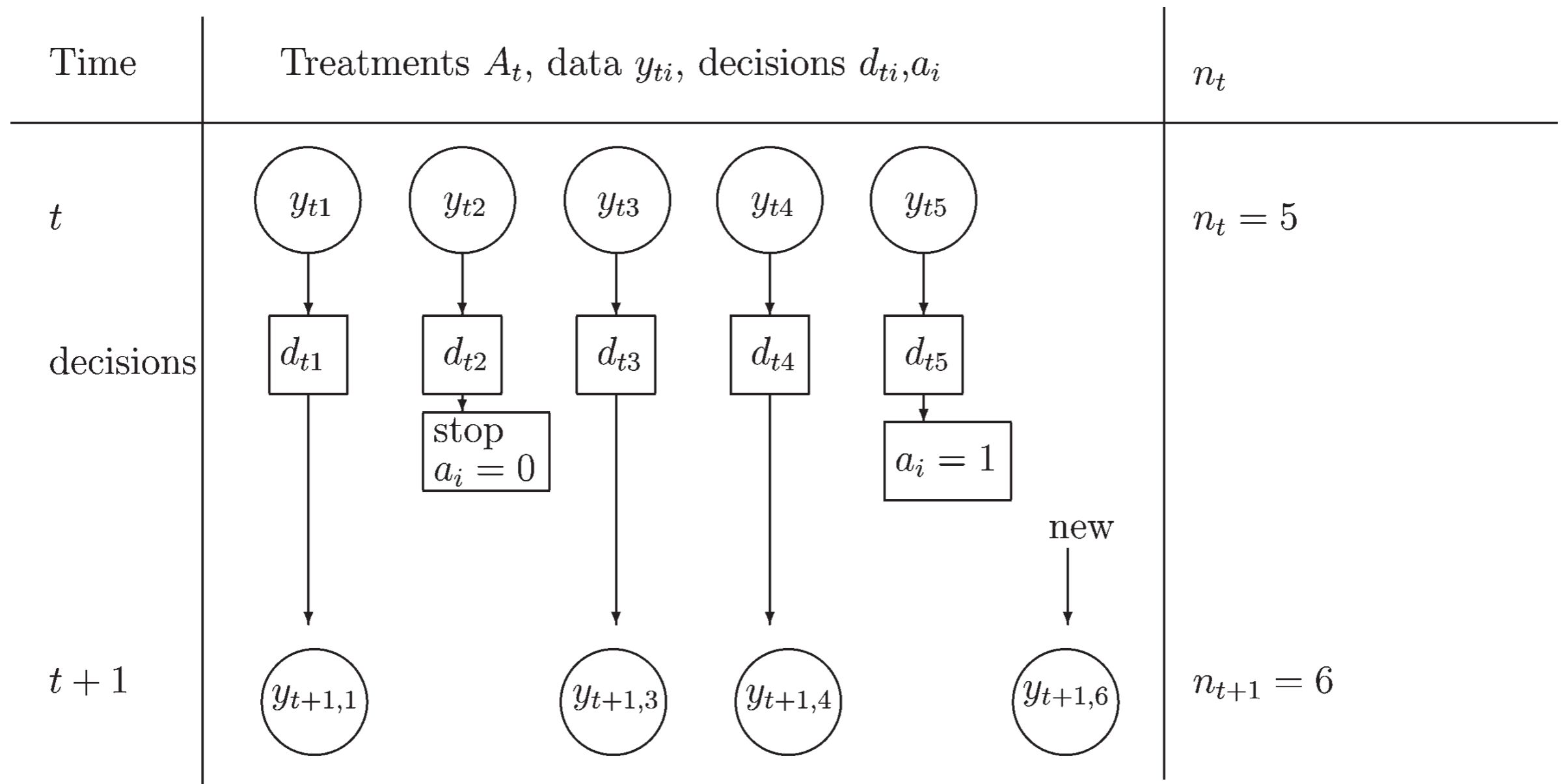see also Piantadosi (2005) *Clinical Trials* **2**:182-192

Can Incorporate Frequentist Criteria in Utility Function

- Bayesian design optimizes expected utility

  ▸ Utility function can include different considerations

    - Sample size or cost

    - Precision

    - Number of patients who benefit

    - Prediction of future study outcomes

  ▸ E.g., Anscombe ('63), Berry & Ho ('88), Lewis & Berry ('94), Carlin, Kadane, & Gelfand ('98), Stallard, Thall, & Whitehead ('99), Lewis, Lipsky, & Berry ('07), Trippa, Rosner, & Müller ('12), Ventz & Trippa ('15)

Platform or Master Protocols

- **At any one time, multiple phase II studies**

Time	Treatments A_t, data y_{ti}, decisions d_{ti}, a_i	n_t
t	y_{t1} y_{t2} y_{t3} y_{t4} y_{t5}	$n_t = 5$
decisions	d_{t1} d_{t2} d_{t3} d_{t4} d_{t5} / stop $a_i = 0$ / $a_i = 1$ / new	
$t+1$	$y_{t+1,1}$ $y_{t+1,3}$ $y_{t+1,4}$ $y_{t+1,6}$	$n_{t+1} = 6$

Decisions

- At each analysis-decision time t

  ‣ Make decision d_{ti} for current study or trt

    - $d_{ti} = 1$: Abandon current study trt

    - $d_{ti} = 2$: Stop study and move to phase 3

    - $d_{ti} = 3$: Continue current study

- Could have multiple trt-studies at t

  ‣ Then have $d_t = (d_{t1}, d_{t2}, \ldots, d_{tk})$

  ‣ Only considering one at a time now $d_t = d_{t1}$

Utility Function

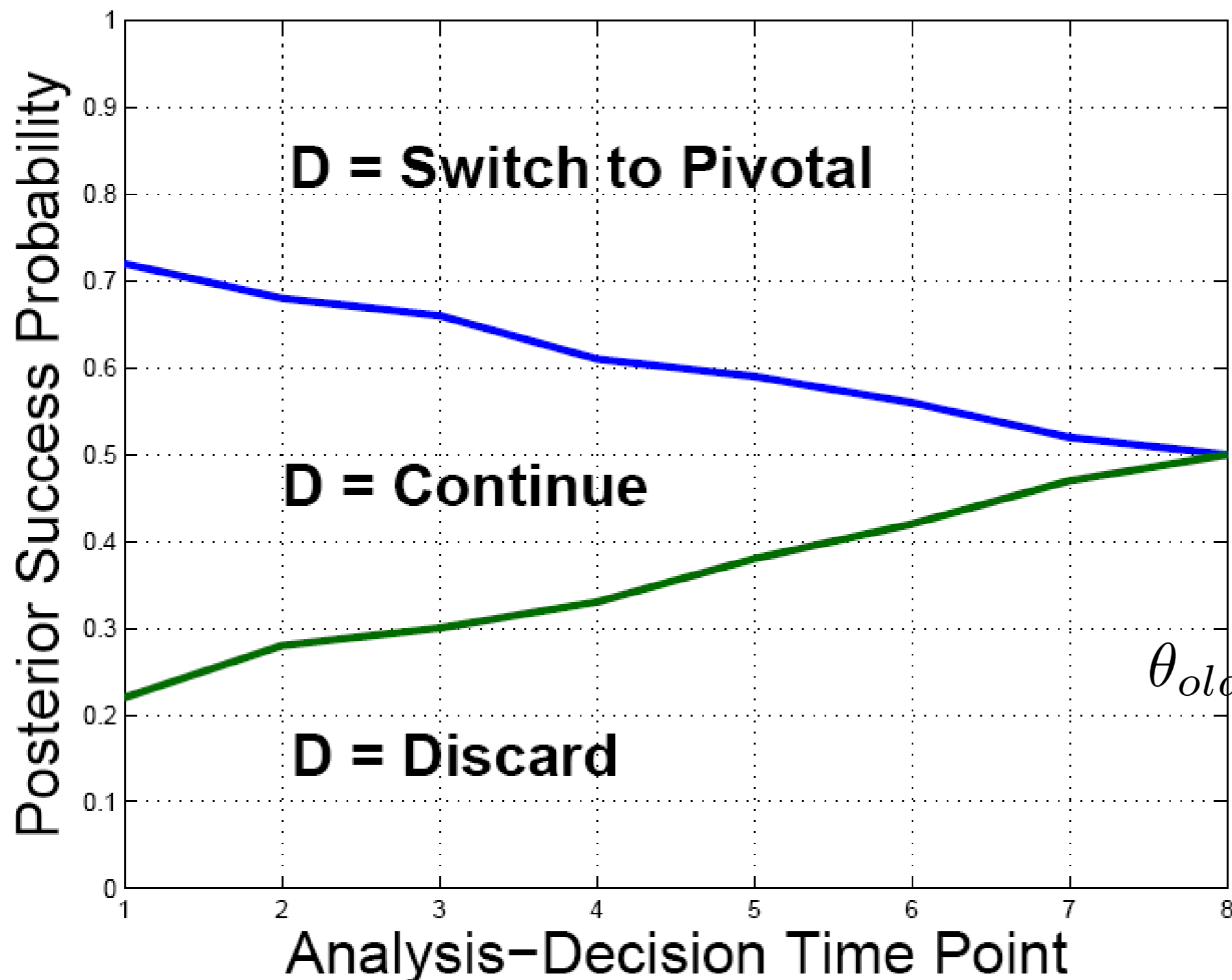- Utility at decision-time **t** (for current trt)

$$u_t(d_t, \theta, Y_t, Y_{III}) = \begin{cases} -c_1 \times n_1 \times t & \textbf{if stop \& discard} \\[2em] -\{c_1 \times n_1 \times t + c_2 \times n_2\} + \\[1em] b \times \{\theta_{new} - \theta_{old}\} I_{[z > z_{1-\alpha}]} \end{cases}$$

Phase 3 sample size & cost

Gain if "significant" phase 3;
gain proportional to effect

Predict phase 3 outcome

Stopping Boundaries

5 per cohort
$c_1 = 0.14$
$c_2 = 0.7$
$b = 290$
$n_2 = 96$
$\theta_0 = 0.2$
$\theta_A = 0.5$
$\theta_{old} \sim \text{Beta}(20, 80)$

2000 sims

Our Colleagues in Other Disciplines Want Our Help

- ## Work with colleagues in other fields

News & Views   the Clinical Chemist

Bayesian Inference Dilemma in Medical Decision-Making:
A Need for User-Friendly Probabilistic Reasoning Tools

Ronald R. Henriquez[†] and Nichole Korpi-Steiner[*†]

Clin Chem 62:1285-6, 2016

- ## They want our help

MEDICINE

Risk literacy in medical decision-making

How can we better represent the statistical structure of risk?

Operskalski JT, Barbey AK (2016). Risk literacy in medical decision-making. *Science* 352:413-414

Opportunities

- Work with colleagues in other fields

Summary: Why Bayes?

- Easier to combine or incorporate information

  ▸ Bayesian paradigm corresponds to learning

    - External information feeds priors

- Frequentist methods seem impractical in some cases

  ▸ Evidence of treatment benefit for rare diseases

- Interest in complex designs & decision making

  ▸ Outcome adaptive randomization

  ▸ Matching treatments to subgroups

Conclusions

- Clinical drug development involves learning & confirming

- Bayesian inference has a place in regulatory science

- Hypothesis testing in and of itself is not evil

  ▸ Need better measures of weight of evidence than p

- Clinical research involves decisions

  ▸ Incorporate statistical decision theory

  ▸ Include predictive dist'n for frequentist tests

Thank You!