Using Diverse Types of Statistical Evidence

Robert Bell
AT&T Labs-Research

Workshop on Statistical Issues
in Analyzing Information from Diverse Sources

Rutgers University
May 6 - 7, 2010

Overview

- Evidence about a “question of interest”
  - Direct evidence
  - Prior information
  - Indirect evidence (B. Efron, “The Future of Indirect Evidence,” *Statistical Science* website, Future Papers)
- Relationships among evidence types
- Bayesian framework
- Subjectivity and objectivity

Notation

- θ = question of interest (scalar)
- γ = other parameter(s)
- $p(\theta, \gamma)$ = prior distribution of parameters

- Y = outcomes

- $\ell(Y \mid \theta, \gamma)$ = likelihood function

Starting Principles

- Use all relevant evidence
  - Data
  - Other information
- Weight evidence properly
- Don’t double count

Direct Evidence

- Efron's definition: "data … [that] directly bear on the question of interest"

- Y such that $\ell(Y \mid \theta, \gamma)$ depends on θ

Direct Evidence about the Quantitative Ability of a College Applicant

- SAT Q
  - Multiple takings, perhaps
- ACT Math
- Other tests
- Math grades
- Reference letters

Combine Direct Evidence Using the Likelihood Function

- Theoretically-correct way to combine evidence
  - Weights pieces correctly
  - Accounts for dependence
  - Frequentist and Bayesian analyses
- May be difficult to determine
- Usually subjective

Prior Information: A Simple Bayesian Example

- $Y_1, \ldots, Y_n \sim N(\theta, \sigma^2)$ for known σ^2
- $\theta \sim N(\mu, \tau^2)$ for known μ and τ^2
- $post(\theta^* \mid \bar{y}) = \dfrac{p(\theta^*)\ell(\bar{y} \mid \theta^*)}{\int p(\theta)\ell(\bar{y} \mid \theta)d\theta}$
- $\theta \mid \bar{Y} \sim N\left(\dfrac{n\bar{y} + (\sigma^2/\tau^2)\mu}{n + (\sigma^2/\tau^2)}, \dfrac{\sigma^2}{n + (\sigma^2/\tau^2)}\right)$

- Prior is equivalent to (σ^2/τ^2) obs centered at μ

Why I Didn't Become a Bayesian

- Stanford, late 1970's
- Pre MCMC
- Subjective priors
  - Limited appeal in my experience
  - Loss of "objectivity"
  - Hamper communication of what's in data
- No killer app

Indirect Evidence

- Efron: “indirect evidence, my catchall term for useful information that isn’t of obvious direct application to a question of interest”

- Illustrated by several examples

Stein Estimation of Baseball Batting Averages (1970 Season)

Name	Hits/AB	Obs.
1. Clemente	18/45	.400
2. F. Robinson	17/45	.378
3. F. Howard	16/45	.356
4. Johnstone	15/45	.333
...		
...		
14. Petrocelli	10/45	.222
15. E. Rodriguez	10/45	.222
16. Campaneris	9/45	.200
17. Munson	8/45	.178
18. Alvis	7/45	.156
Grand Average		.265

Stein Estimation of Baseball Batting Averages (1970 Season)

Name	Hits/AB	Obs.	"Truth"
1. Clemente	18/45	.400	.346
2. F. Robinson	17/45	.378	.298
3. F. Howard	16/45	.356	.276
4. Johnstone	15/45	.333	.222
...			
...			
14. Petrocelli	10/45	.222	.264
15. E. Rodriguez	10/45	.222	.226
16. Campaneris	9/45	.200	.286
17. Munson	8/45	.178	.316
18. Alvis	7/45	.156	.200
Grand Average		.265	.265

Stein Estimation of Baseball Batting Averages (1970 Season)

Name	Hits/AB	Obs.	“Truth”	James-Stein
1. Clemente	18/45	.400	.346	0.294
2. F. Robinson	17/45	.378	.298	0.289
3. F. Howard	16/45	.356	.276	0.285
4. Johnstone	15/45	.333	.222	0.280
...				
...				
14. Petrocelli	10/45	.222	.264	0.256
15. E. Rodriguez	10/45	.222	.226	0.256
16. Campaneris	9/45	.200	.286	0.252
17. Munson	8/45	.178	.316	0.247
18. Alvis	7/45	.156	.200	0.242
Grand Average		.265	.265	0.265

Stein Estimation (Empirical Bayes)

- Approximation of Bayesian estimation
- Estimates μ and τ^2 from combined data
- Very close to Bayesian analysis with μ and τ^2 drawn from hyperprior distributions
- Excellent frequentist properties
  - Always lower risk than observed values

Drug Experiment with Multiple Doses

- Placebo, single dose, double dose
- Efron writes: "Even if the double dose yields strongly significant results in favor of the new drug, a not-quite significant result for the single dose, say p-value .07, will not be enough to earn FDA approval. The single dose by itself must prove its worth."
- Efron continues: "My own feeling at this point would be that the single dose is very likely to be vindicated in any subsequent testing. The strong result for the double dose adds *indirect evidence* to the direct, nearly significant, single dose outcome."

Direct vs. Indirect Evidence

- Direct Evidence

$$\ell(Y^{Dir}|\theta)$$

$$Y^{Dir} \Rightarrow \theta$$

Direct vs. Indirect Evidence

- Direct Evidence

$$\ell(Y^{Dir}|\theta)$$

$$Y^{Dir} \Rightarrow \theta$$

- Indirect Evidence

$$\ell(Y^{Ind}|\gamma)$$

$$Y^{Ind} \Rightarrow \gamma$$

Direct vs. Indirect Evidence

- Direct Evidence

$$\ell(Y^{Dir}|\theta)$$

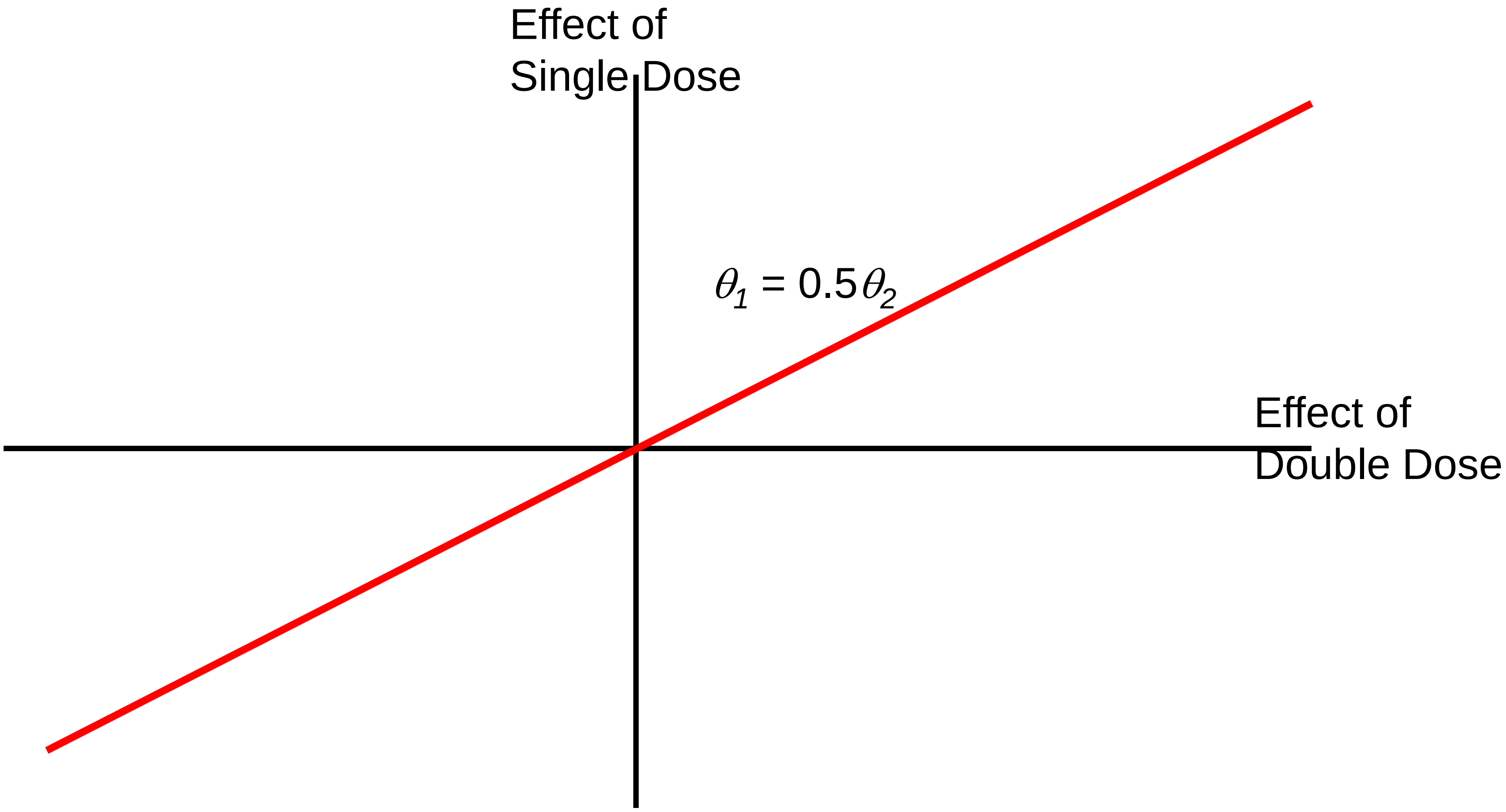
$$Y^{Dir} \Rightarrow \theta$$

- Indirect Evidence

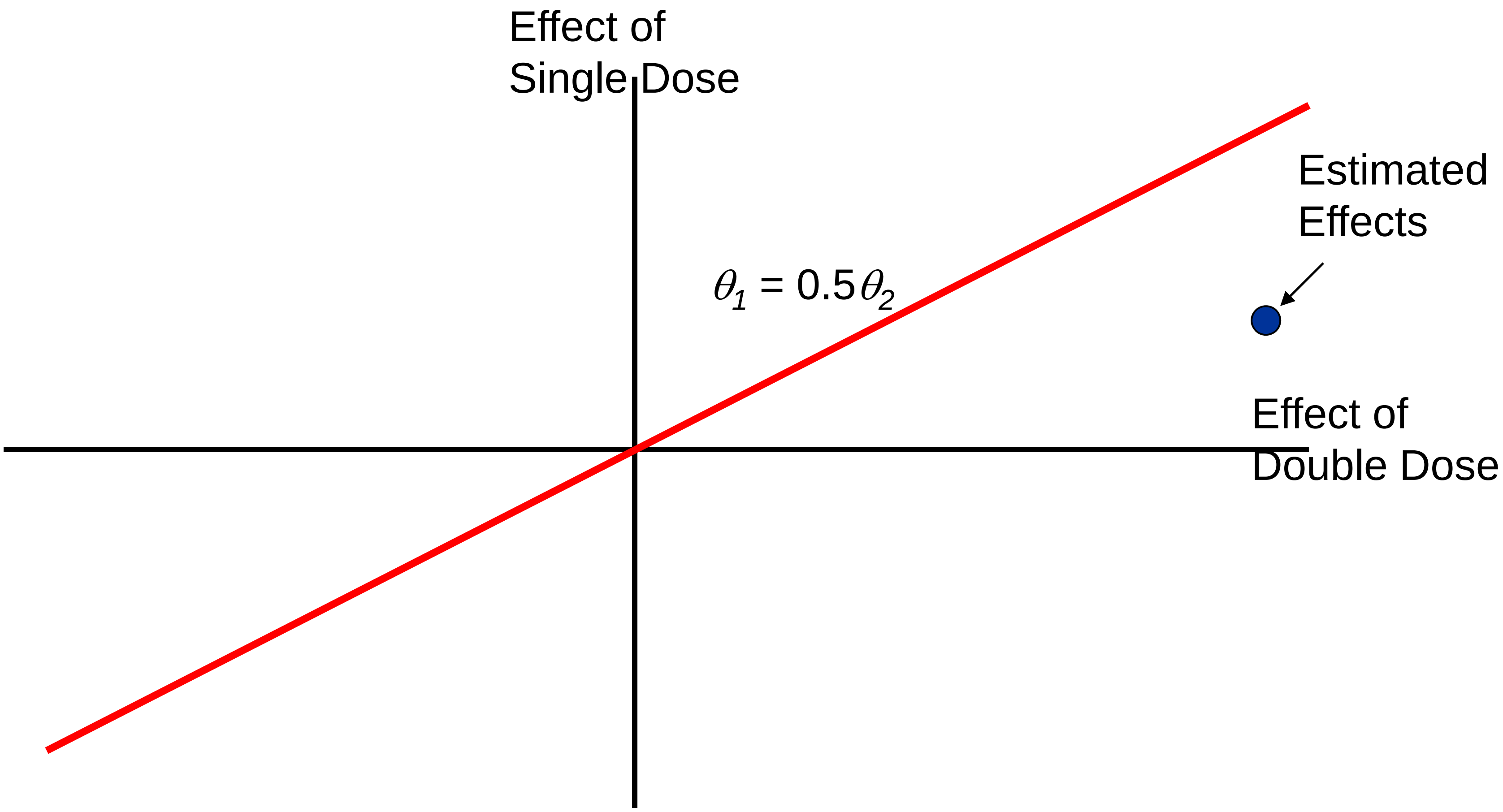
$$\ell(Y^{Ind}|\gamma) \qquad p(\theta,\gamma)$$

$$Y^{Ind} \Rightarrow \gamma \Rightarrow \theta$$

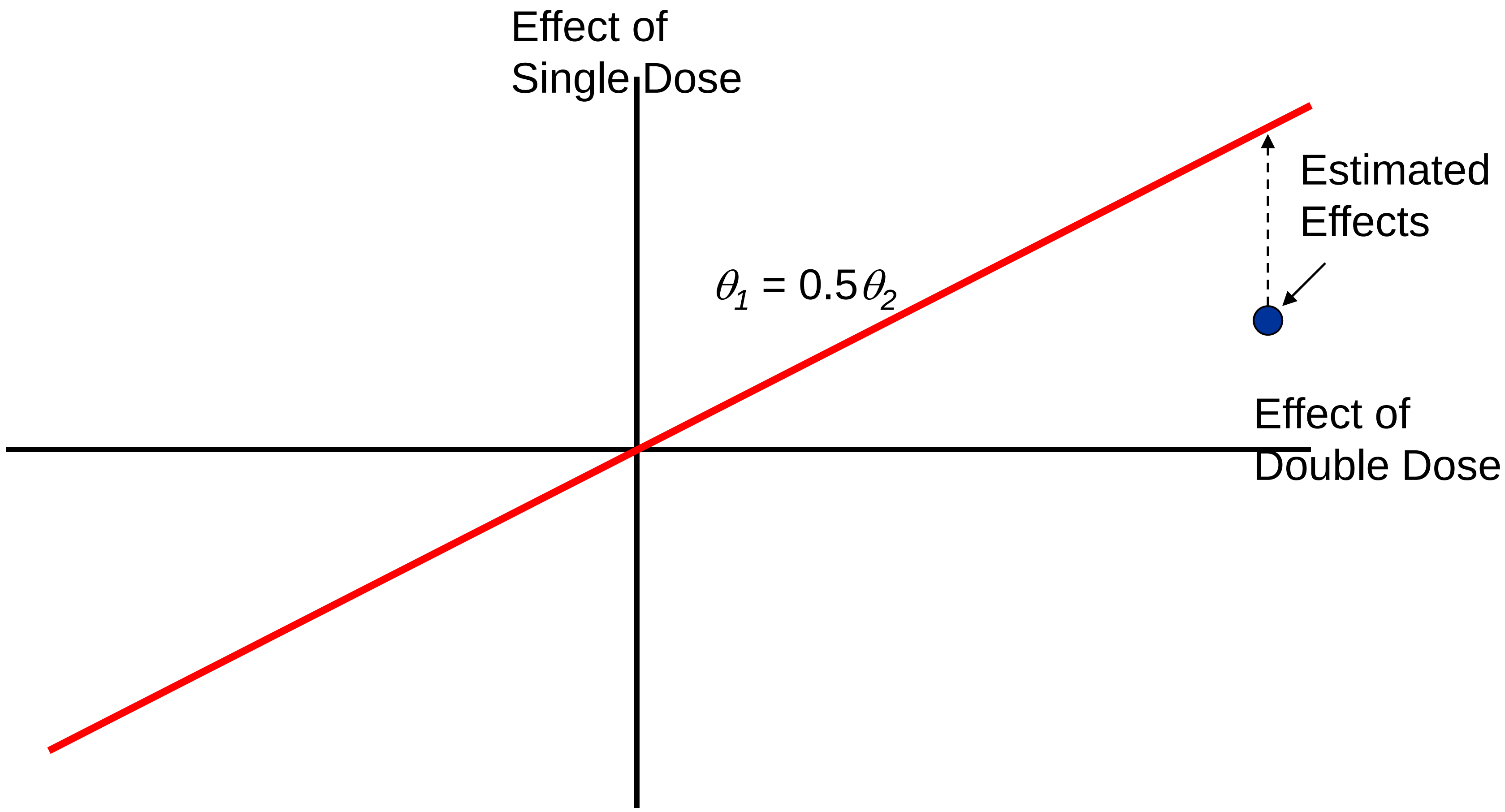
Prior about Effects of Single and Double Doses (Linear Model)



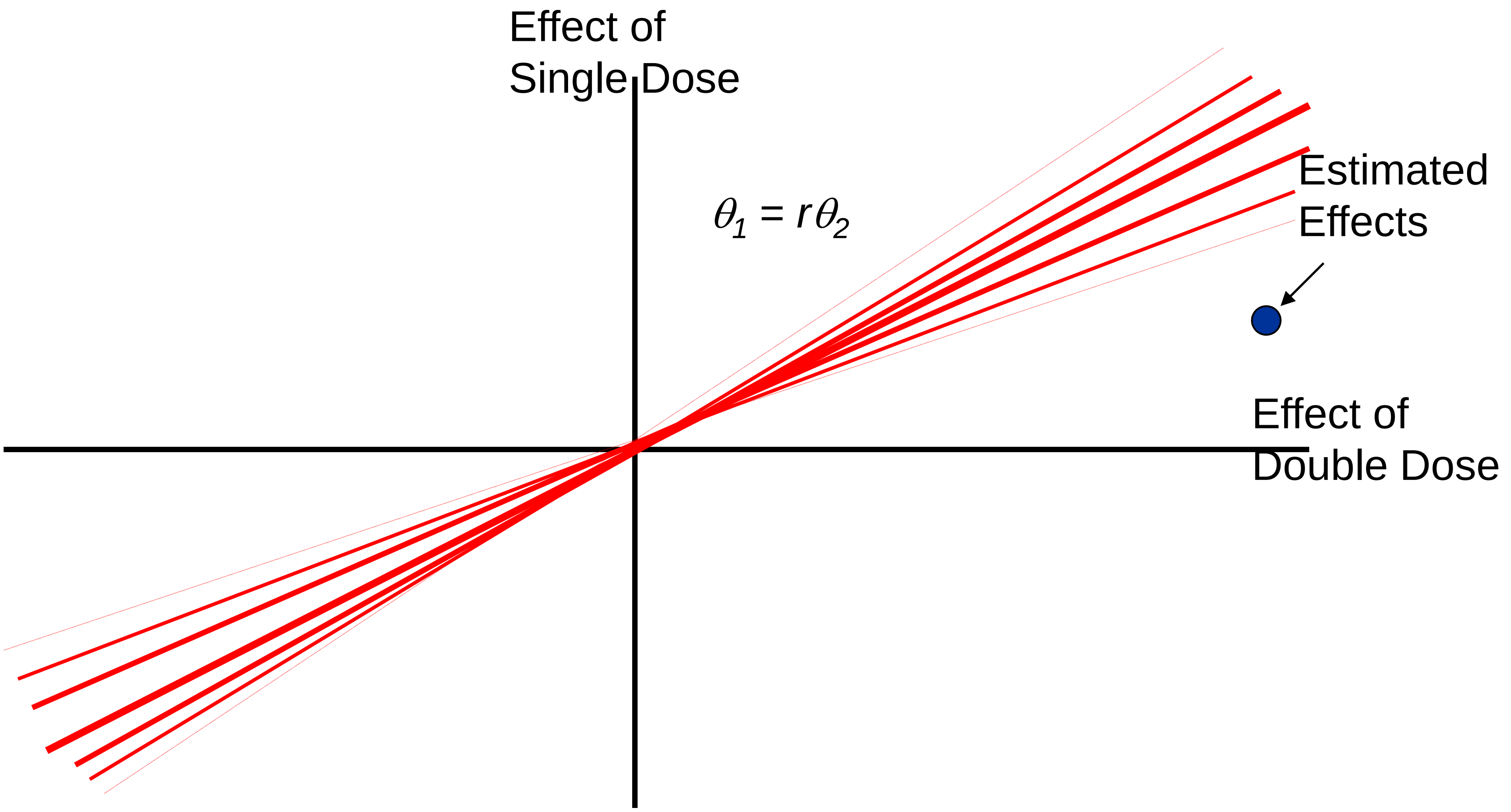
Prior about Effects of Single and Double Doses (Linear Model)



Prior about Effects of Single and Double Doses (Linear Model)



More Flexible Prior about Effects of Single and Double Doses



Towards Objective Bayes

- Substantial efforts to find priors that are
  - Objective
  - Non-informative
- Informative priors
  - Experience based
  - Data driven
  - Still somewhat subjective

Hierarchical Bayesian Models

- The killer app
  - Baseball example
  - Random coefficient models
  - Matrix factorization for recommender systems
- Lots of exchangeable parameters
- Power of subjective priors in an objective package
  - 1000 coins
- Never fully objective

Conclusions

- Use all relevant data
- Indirect evidence is a valuable concept
  - Highlights relationship among parameters
- Fuzzy boundaries between types of evidence
- Objectivity is a worthy goal
- Complete objectivity is a fantasy
  - Subjectivity in any prior, or lack of one
  - Likelihood is also subjective
- Hierarchical models and extensions are the killer app for Bayesian analysis