Predicting Low Dose Effects for Chemicals in High Through-Put Studies

Edward J. Stanek III and
Edward Calabrese
UMass-Amherst, USA

Outline

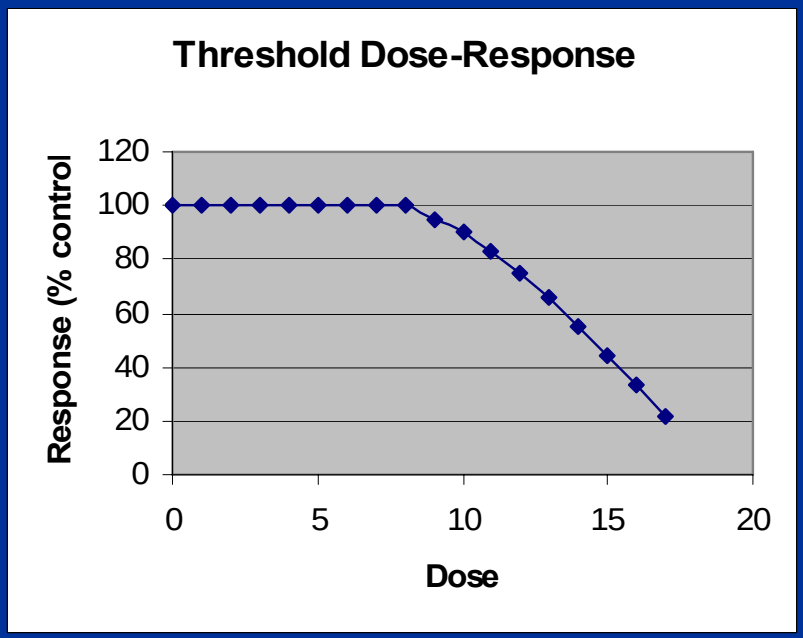
1. Introduction
2. The Example and Study Data
3. Testing Hypotheses about Low Dose Response
4. Estimating Response at Low Doses
5. Better Estimates: Predictors of Realized Random Effects
6. Conclusions

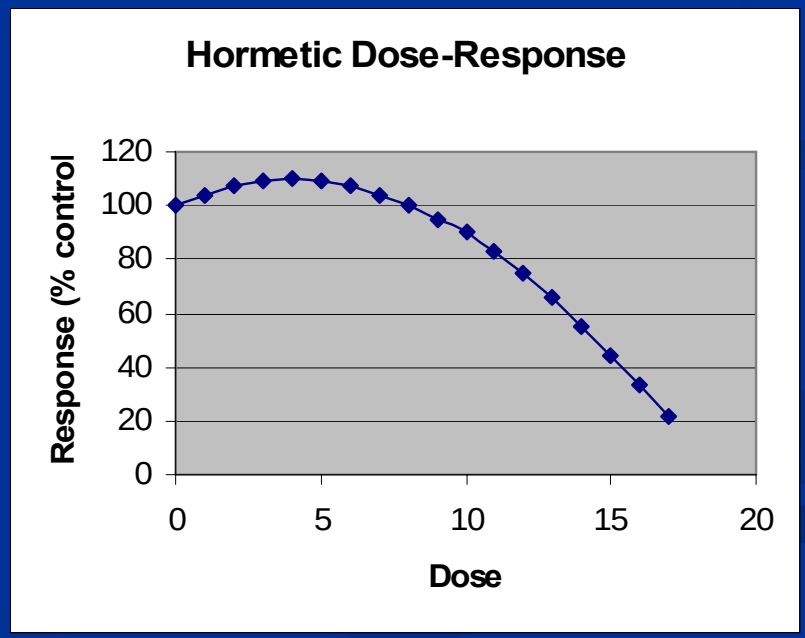
Introduction

- High through-put studies common
- Can data be 'harvested'?
  - Cons: Ad-hoc, Post-hoc, Low Power, Bias, not significant
  - Pros: Low cost, Common in 'observational' studies, Can learn something
- Illustrate use to assess response at low doses-
- Focus: Statistical Models and Interpretation

Example and Study Data

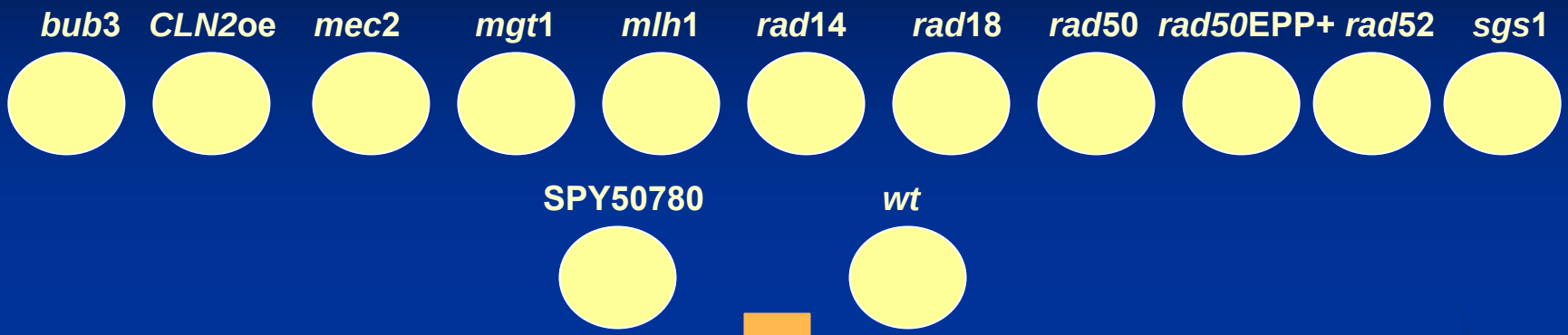
NCI Yeast data- 2189 chemicals, 13 yeast strains
5 doses x 2 replications-
Focus on doses in Low Dose Region





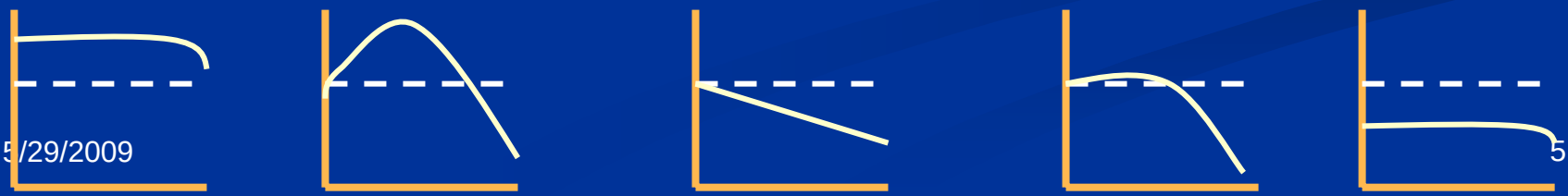
NCI Methods

13 Yeast Strains



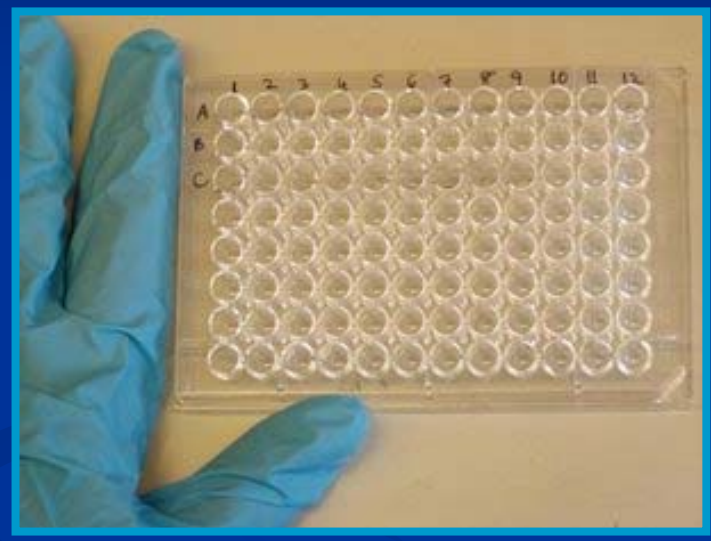
2,189 Anti-tumor Agents
(1.2, 3.7, 11, 33 & 100 µM + Control)

28,457 Replicated Dose-Response Experiments

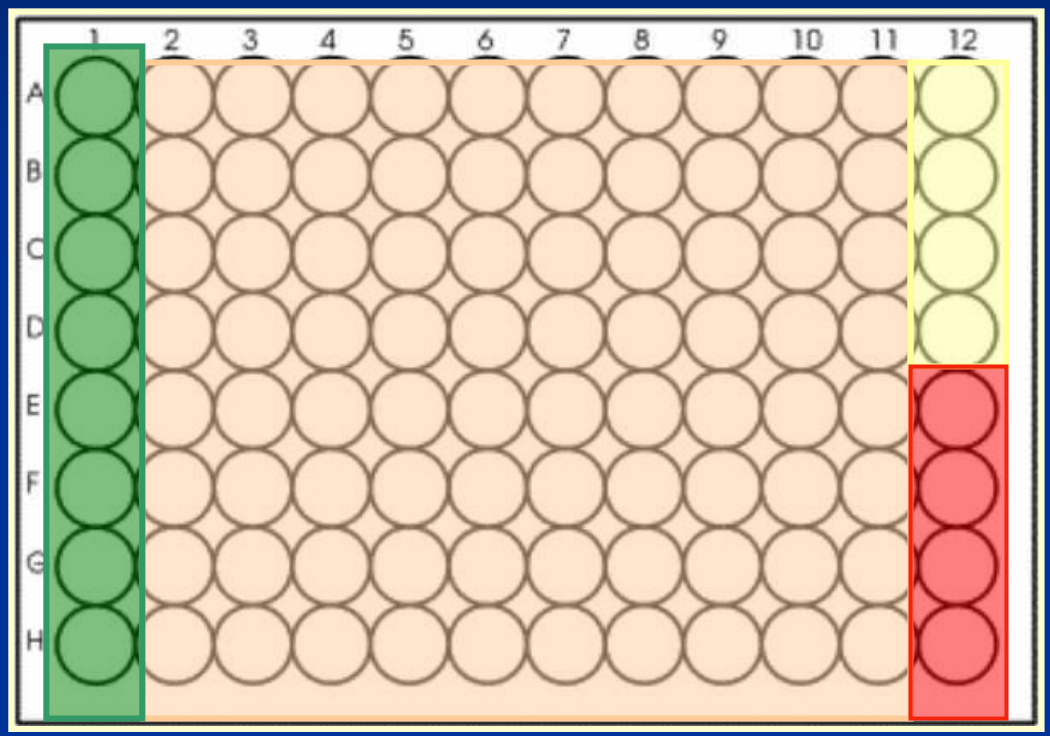


Methods: Cell Line

- Yeast cells in exponential growth inoculated onto microtiter plates
- 10^4 cells/well with 200µl of medium
- 80 different anti-tumor agents added per plate
- Each plate run at one concentration (1.2, 3.7, 10, 33, or 100 µM)
- Each (of 5) conc. incubated over the same 12 hrs.
- Cell number measured with spectrophotometer (OD_{600})



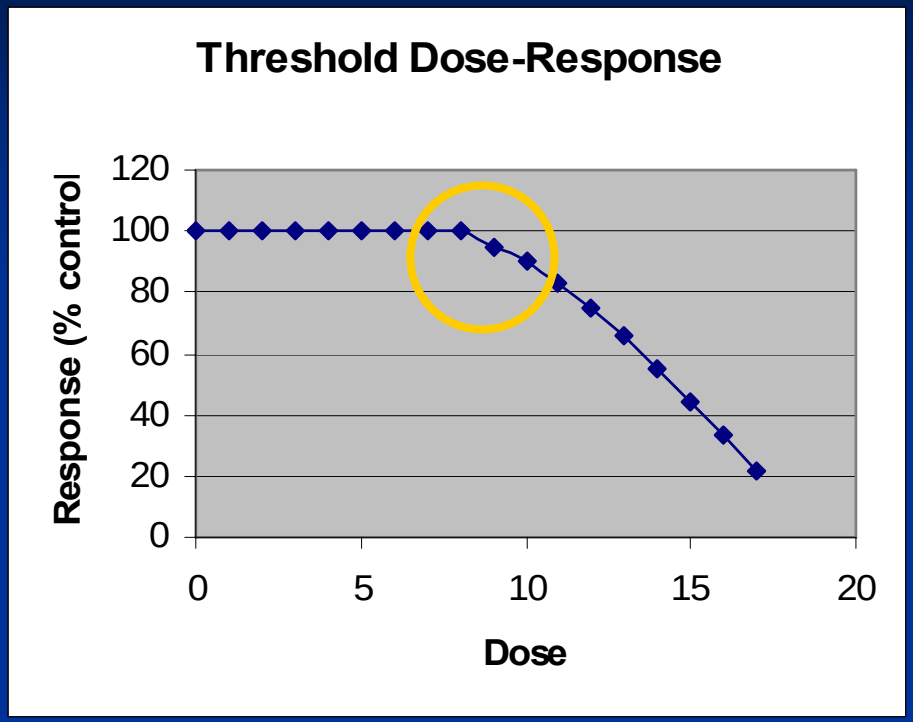
Methods: 96-well Plate



- 80 different chemicals at same concentration
- 8 wells solvent (DMSO) controls
- 4 wells cycloheximide controls
- 4 wells unexposed controls

Reported response is the growth of yeast in respective concentration relative to growth in solvent control

Where is the Low Dose Region?



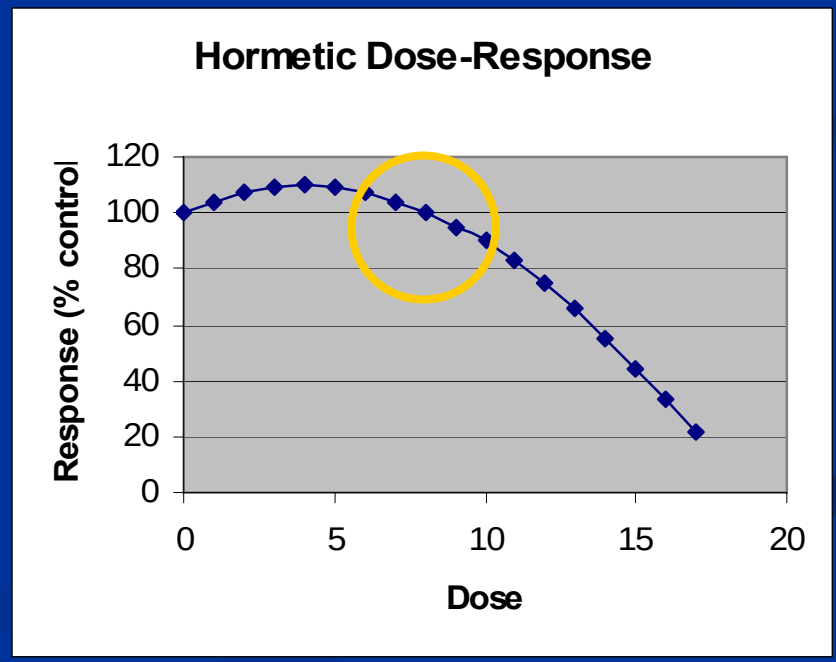
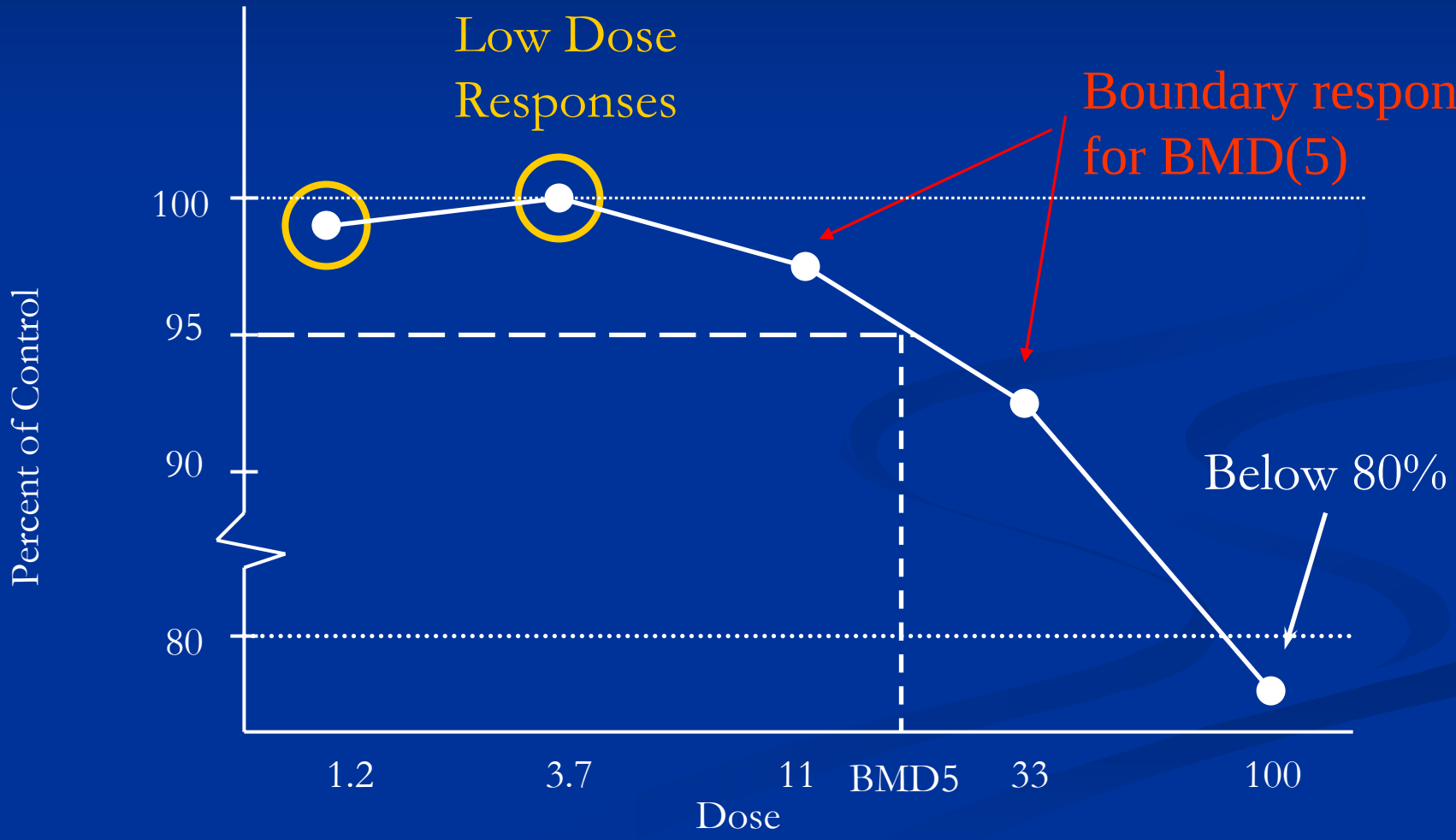


Figure 1. General Scheme Used for the Derivation of the BMD(5)



START

Selection Process

28,457 **dose-responses included in the database:**
2,189 anticancer agents x 13 yeast strains

6,480 removed (23%)

21,977 **(77%) have an inhibitory response ≤ 80 of the control**

9,375 removed (43%)

12,602 **(44%) also have a dose below the BMD(5)**

10,151 removed

2,451 **(9%) also have 3 doses below the BMD(5)**

FINISH

253 Wild Type (wt)

Example Data
Wild Type- NSC#1928

Dose	Ave Response	Difference
1.2	97	0
3.7	98	12
11	91	15

Example Data
Wild Type- NSC#1928

Dose	Rep	Ave Response	Difference	Response (% Diff from 100)
1.2	1	97	0	-3
1.2	2	97	0	-3
3.7	1	98	12	4
3.7	2	98	12	-8
11	1	91	15	-1.5
11	2	91	15	-16.5

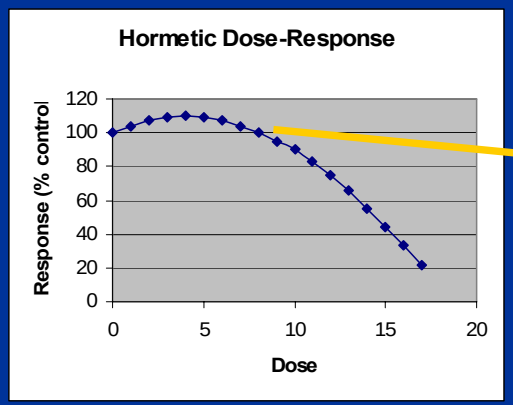
How should we evaluate Response in Low Dose Region?

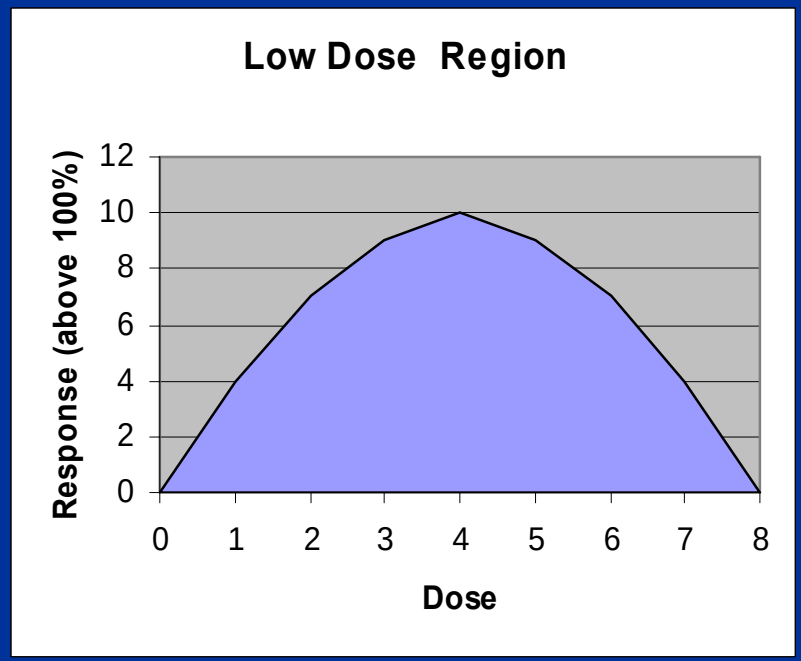
Data:

Selected Set of chemicals (n=253)

b) For each chemical, 3 doses

c) For each dose, replicate responses





Approach 1: Test a Hypothesis
Approach 2: Estimate 'True Response'

Need a statistical Model: Response

Chemical s is "NSC#1928"

Dose (t)	Rep (k)	Y_{stk}
1	1	-3
1	2	-3
2	1	4
2	2	-8
3	1	-1.5
3	2	-16.5

What is 'True Response'?

$$Y_{stk} = \mu_{st} + E_{stk}$$

$$E_R\left(Y_{stk}\right) = \mu_{st}$$

True Response:

$$\mu_s = \frac{1}{m}\sum_{t=1}^{m}\mu_{st}$$

Expected Response

Response

Dose (t)	Rep (k)	Y_{stk}	μ_{st}
1	1	-3	μ_{s1}
1	2	-3	μ_{s1}
2	1	4	μ_{s2}
2	2	-8	μ_{s2}
3	1	-1.5	μ_{s3}
3	2	-16.5	μ_{s3}

Model of Response for Chemical "s"

Chemical $s = 1, \ldots, n$

Dose $t = 1, \ldots, m$

Replication $k = 1, \ldots, r (= 2)$

where $Y_{stk} = \mu_{st} + E_{stk}$

$= \mu_s + \delta_{st} + E_{stk}$

$$\mu_s = \frac{1}{m} \sum_{t=1}^{m} \mu_{st}$$

Hypothesis Testing

Model: $Y_{stk} = \mu_s + \delta_{st} + E_{stk}$

Test the Null Hypothesis: $H_0 : \mu_s = 0$

$H_A : \mu_s \neq 0$

"NSC#1928" $\hat{\mu}_s = -4.67$

$se(\hat{\mu}_s) = 3.20$

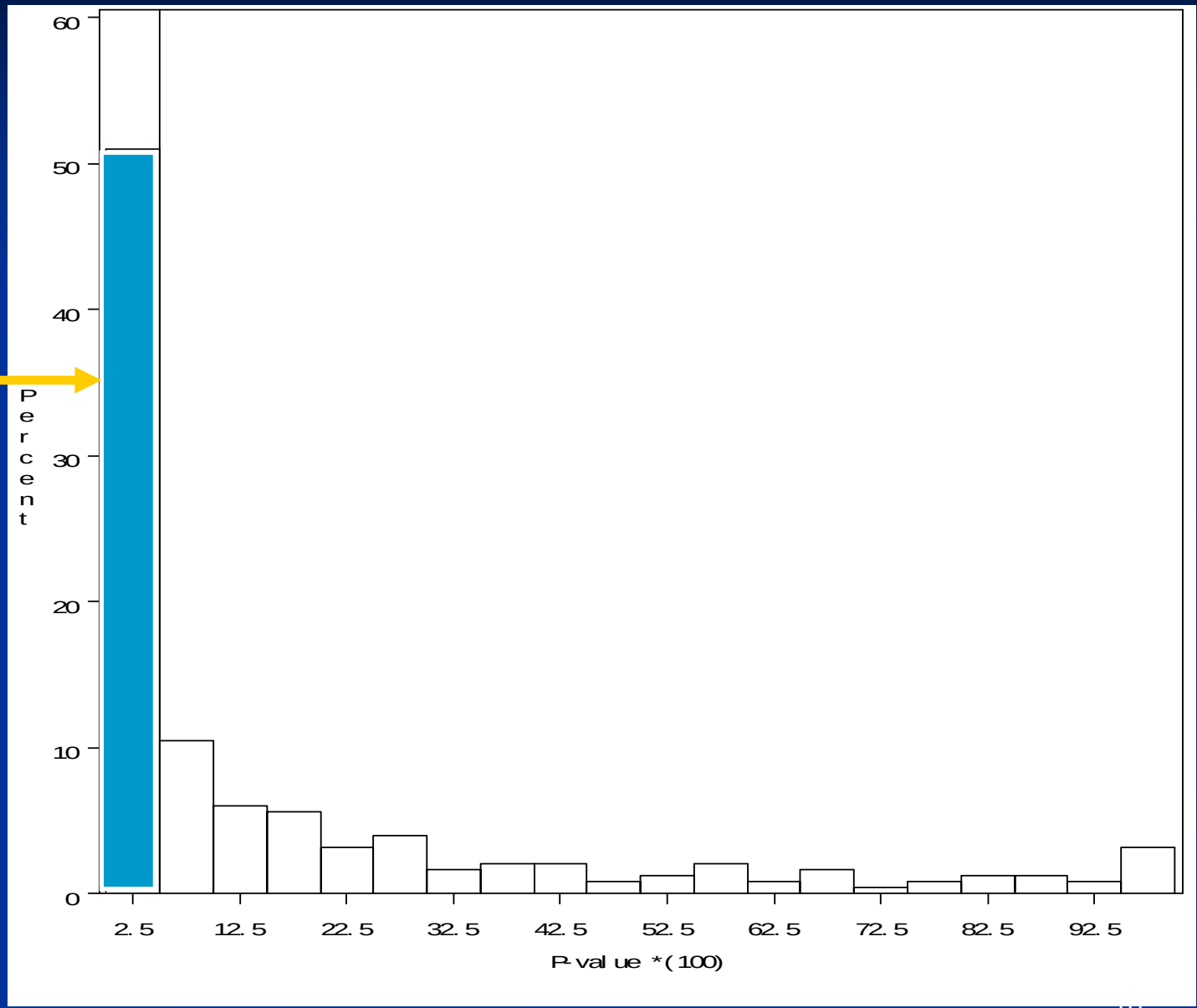
$t = -1.46$

$p - value = 0.241$

Decision: Fail to reject H_0

Histogram of P-values for 249 Chemicals:

P-value < 0.05
127 Chemicals
(50.8%)



Excluding 4
Chemicals with
"0" stderr

Interpretation of Tests
All 249 Chemicals

Expected # Significant by Chance: 12 to 13

Observed 127 Significant: Mean Response

3 Negative

124 Positive

If 1 Sided test: $H_0 : \mu_s \leq 0$

$H_A : \mu_s > 0$

P-value < 0.05
146 Chemicals
(58.6%)

Multiple Comparisons?

Preserve Type I Error:

Prob(False Rejection of H_0)

Change the critical value to $\alpha = \frac{0.05}{249} = 0.0002$

P-value < 0.0002

11 Chemicals (4.4%)

Mean response:

0 Negative

11 Positive

Conclusions:

- Studies are not designed for low-dose range
- Low power to evaluate response
- If controlling for multiple comparisons, few chemicals reject null hypothesis (4.4%)
- Data are not suitable to learn about response at low doses

Approach 2: Estimate Response at Low Doses

Fit the model to chemical "s":

$$Y_{stk} = \mu_s + \delta_{st} + E_{stk}$$

Estimate μ_s Estimator $\bar{Y}_s = \frac{1}{m}\sum_{t=1}^{m}\left(\frac{1}{r}\sum_{k=1}^{r} Y_{stk}\right)$

Estimate a 95% Confidence Interval for μ_s

Example: Wild Type

Response

Dose (t)	Rep (k)	Y_{stk}
1	1	-3
1	2	-3
2	1	4
2	2	-8
3	1	-1.5
3	2	-16.5

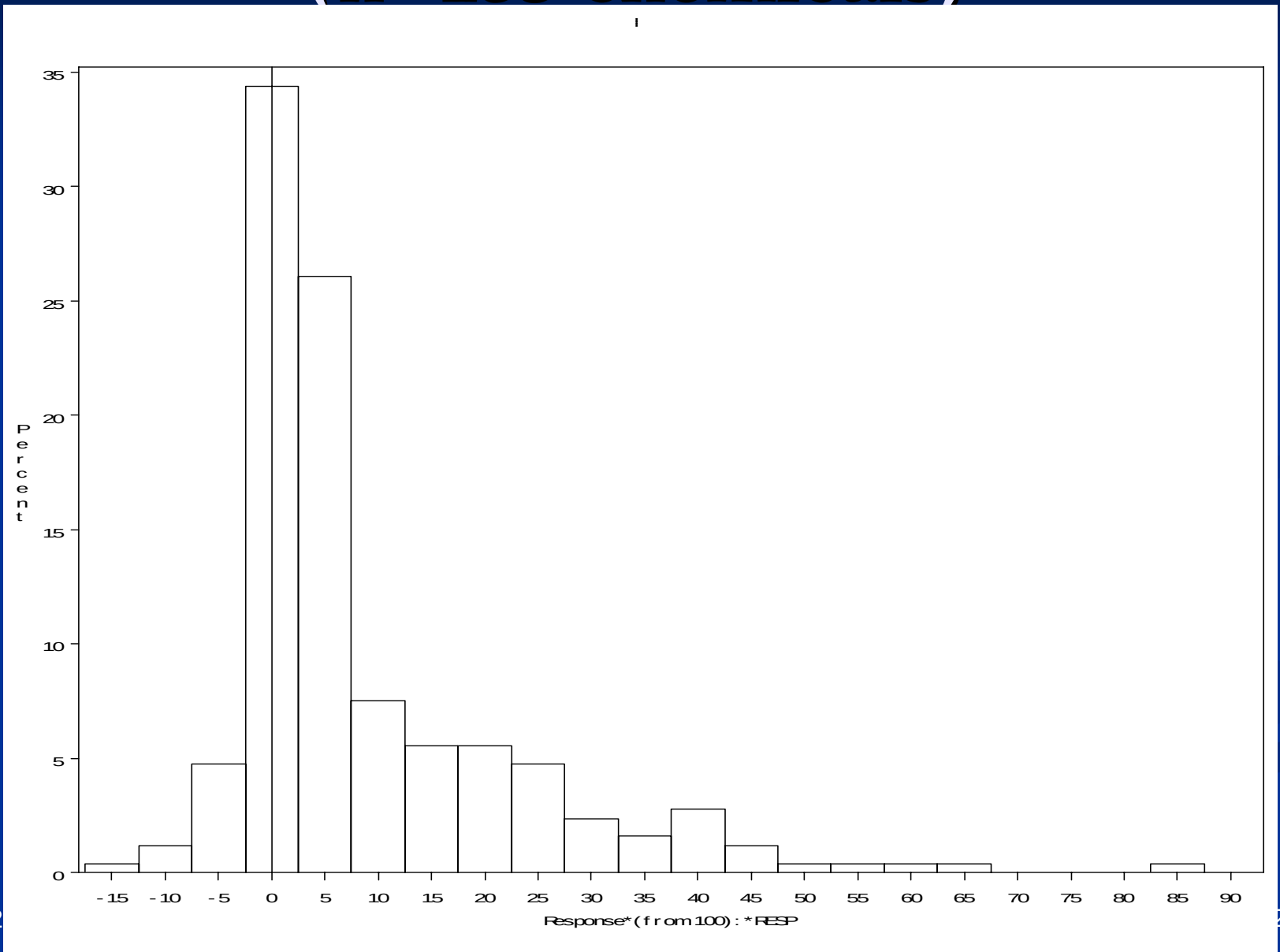
"NSC#1928"

$$\bar{Y}_s = -4.67$$

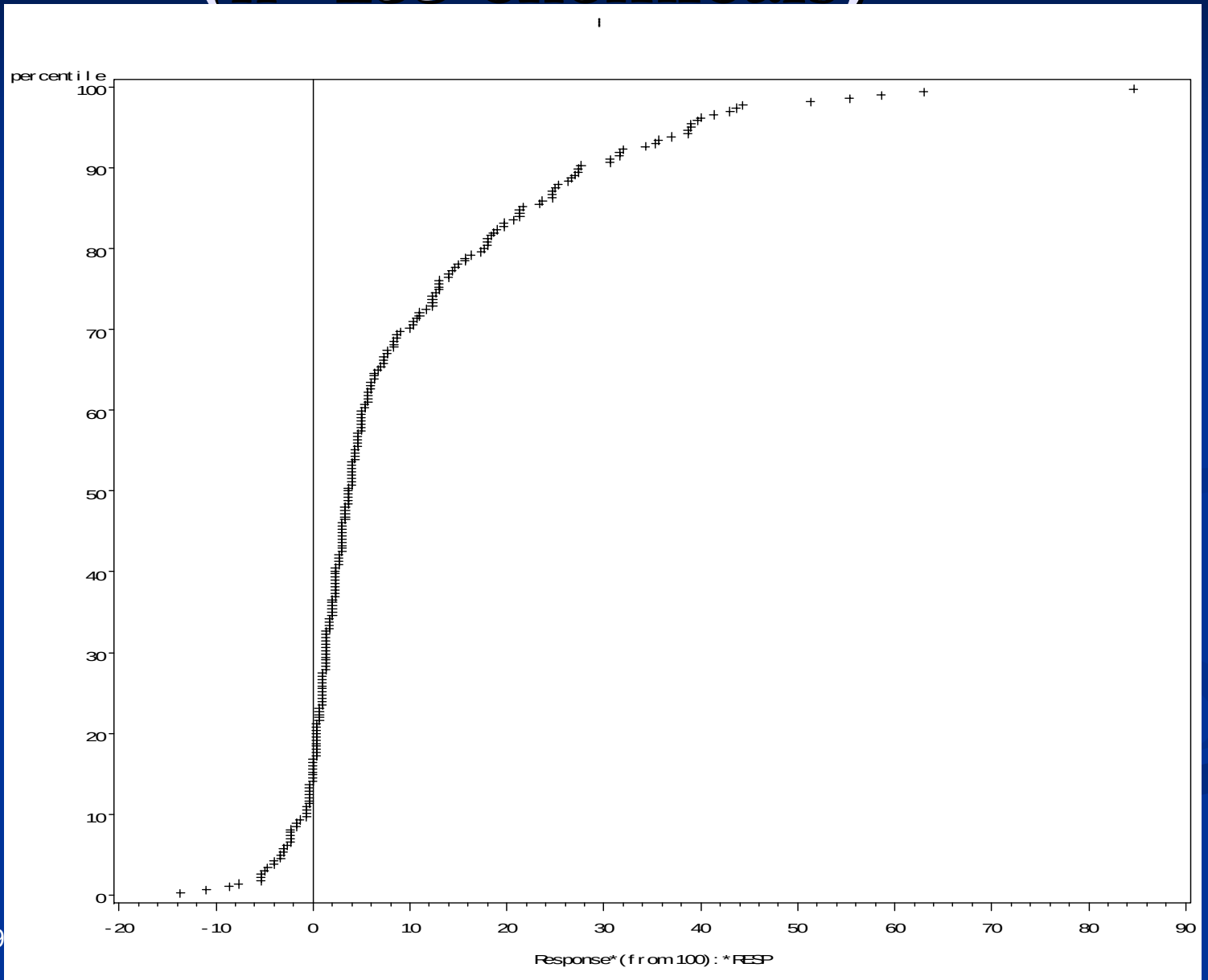
$$se\left(\bar{Y}_s\right) = 3.20$$

$$\left(\bar{Y}_s - t_{df,\left(1-\frac{\alpha}{2}\right)} se\left(\bar{Y}_s\right), \bar{Y}_s + t_{df,\left(1-\frac{\alpha}{2}\right)} se\left(\bar{Y}_s\right)\right) \quad (-14.9, 5.5)$$

Histogram of Estimates (n=253 chemicals)



Cumulative Distribution of Estimates (n=253 chemicals)



What about Control for Multiple Chemicals?

$$\left(\bar{Y}_s - t_{df,\left(1-\frac{\alpha}{2}\right)} se\left(\bar{Y}_s\right), \bar{Y}_s + t_{df,\left(1-\frac{\alpha}{2}\right)} se\left(\bar{Y}_s\right)\right) \qquad (-14.9 \ , \ 5.5)$$

Should we use $\alpha = \frac{0.05}{253} = 0.0002$

Using $t_{3,(0.0001)} = -22.2$ $\qquad (-75.8 \ , \ 66.4)$

Conclusion: CI is so large, the estimate seems useless.

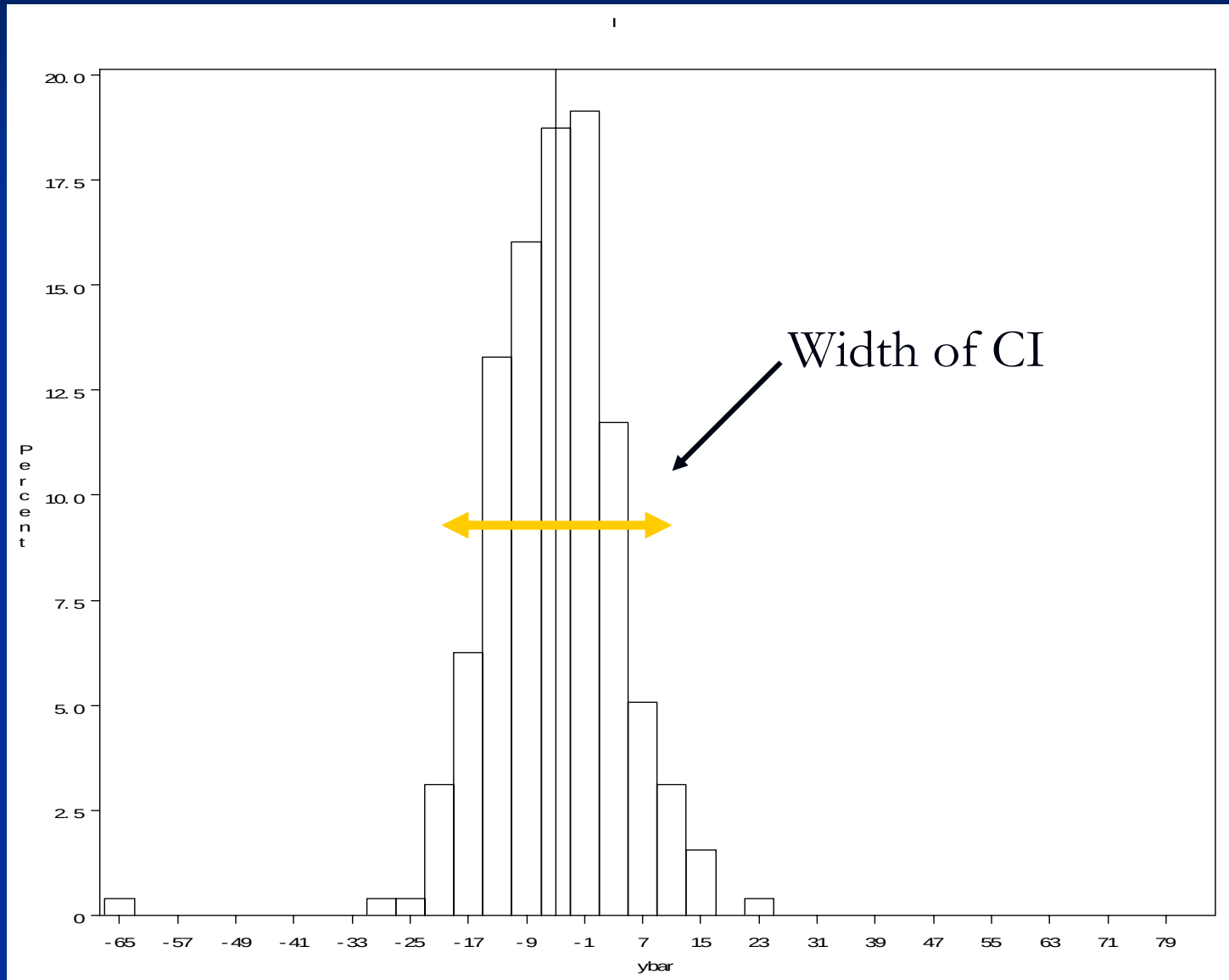
Review: What is a Confidence Interval?

Assume $\mu_s = -5.0$

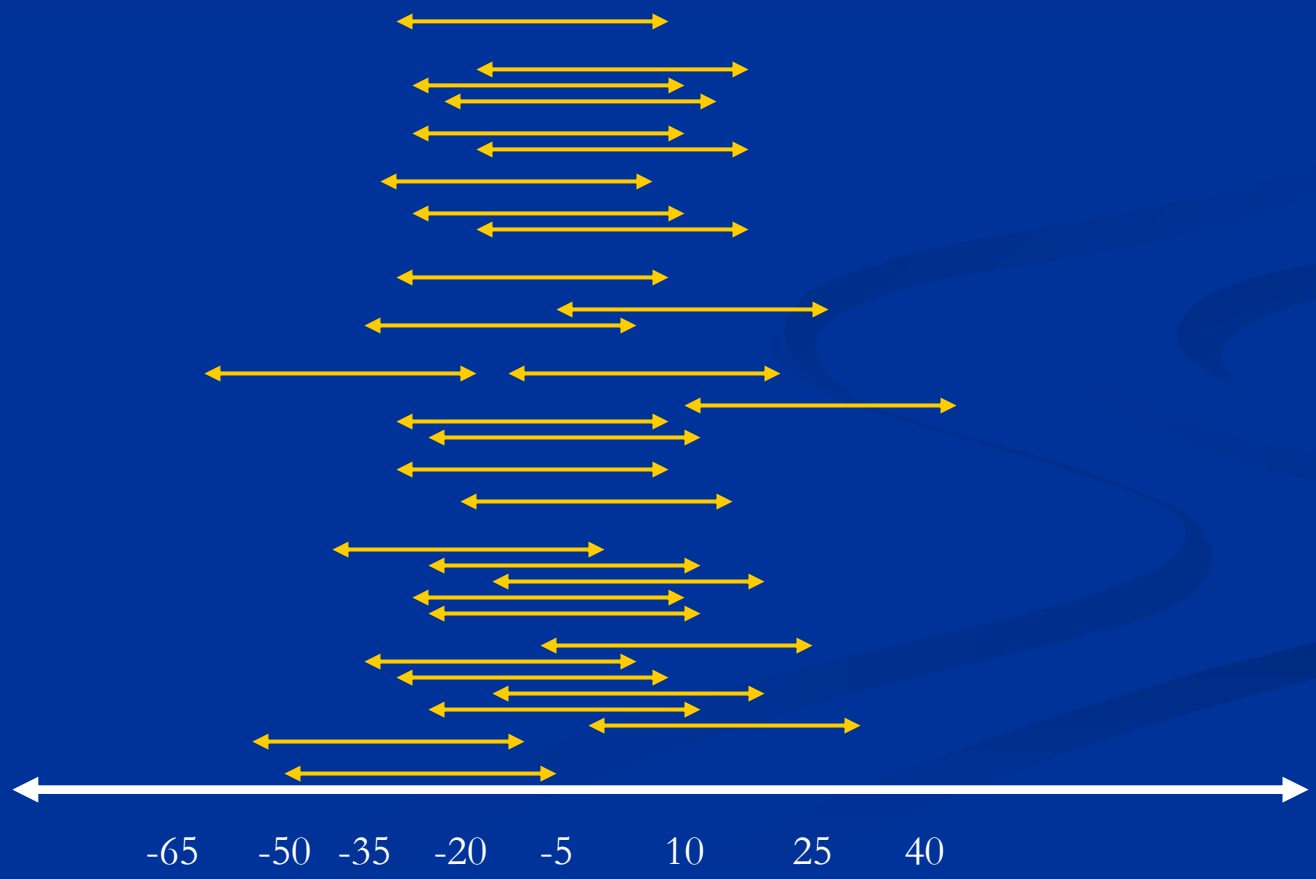
$se\left(\bar{Y}_s\right) = 8.0$

Suppose we repeat the study many times and summarize the Estimator $\bar{Y}_s$

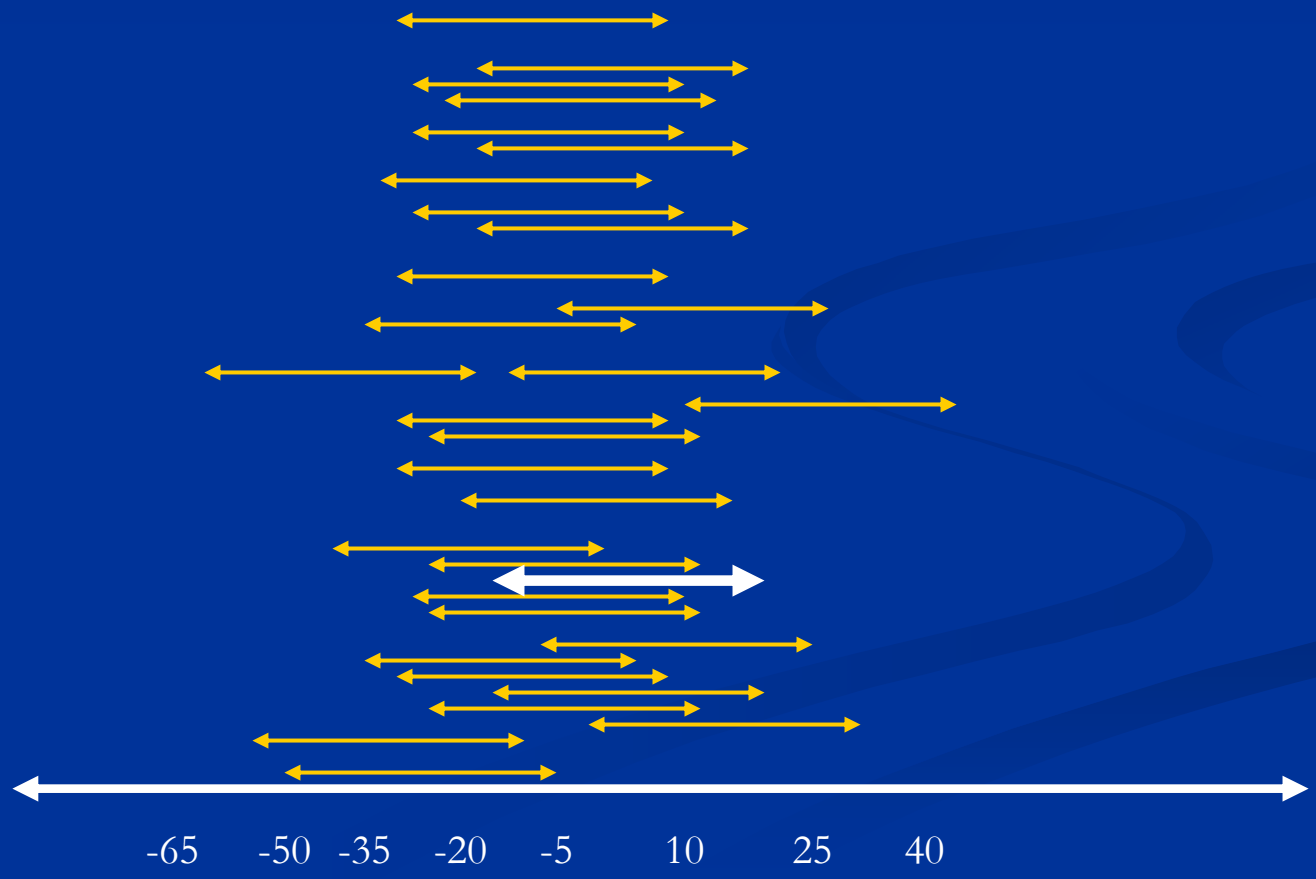
Distribution of Estimator



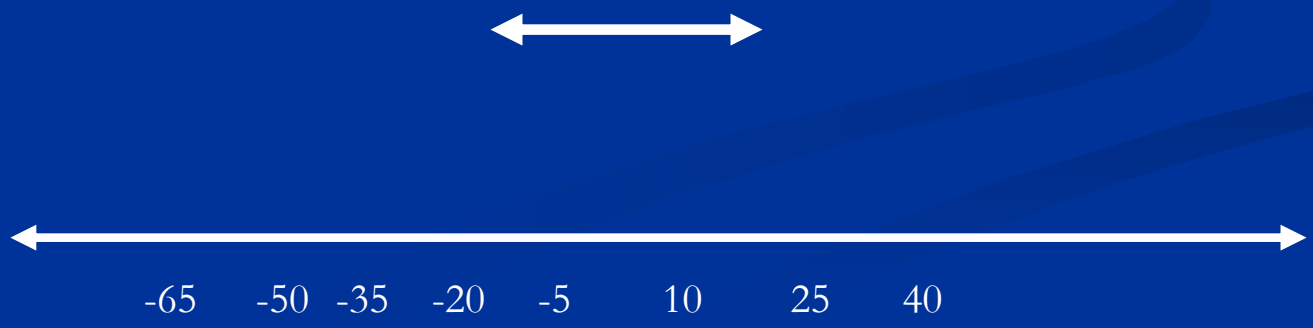
Distribution of CI Estimator



Observed Confidence Interval



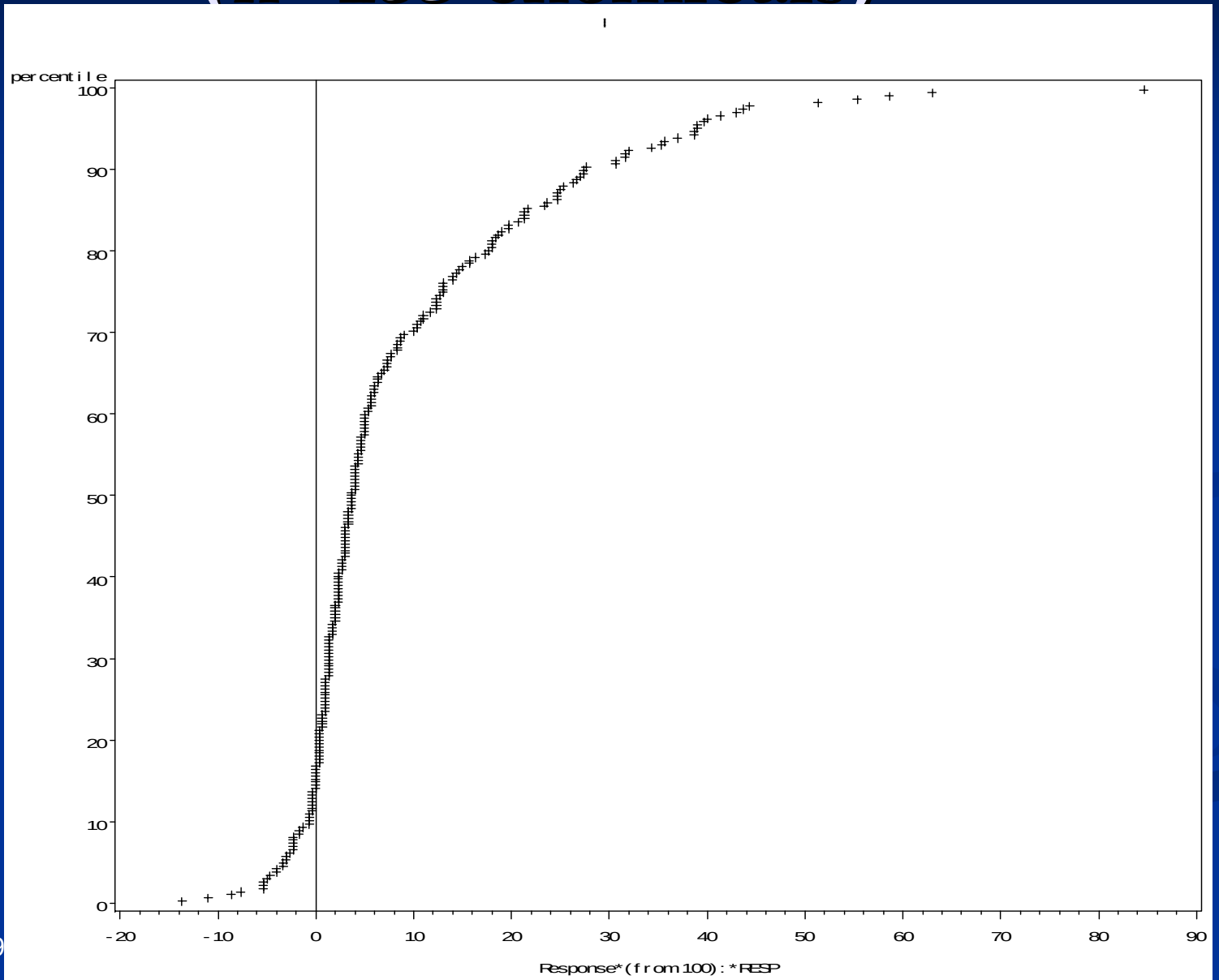
Observed Confidence Interval



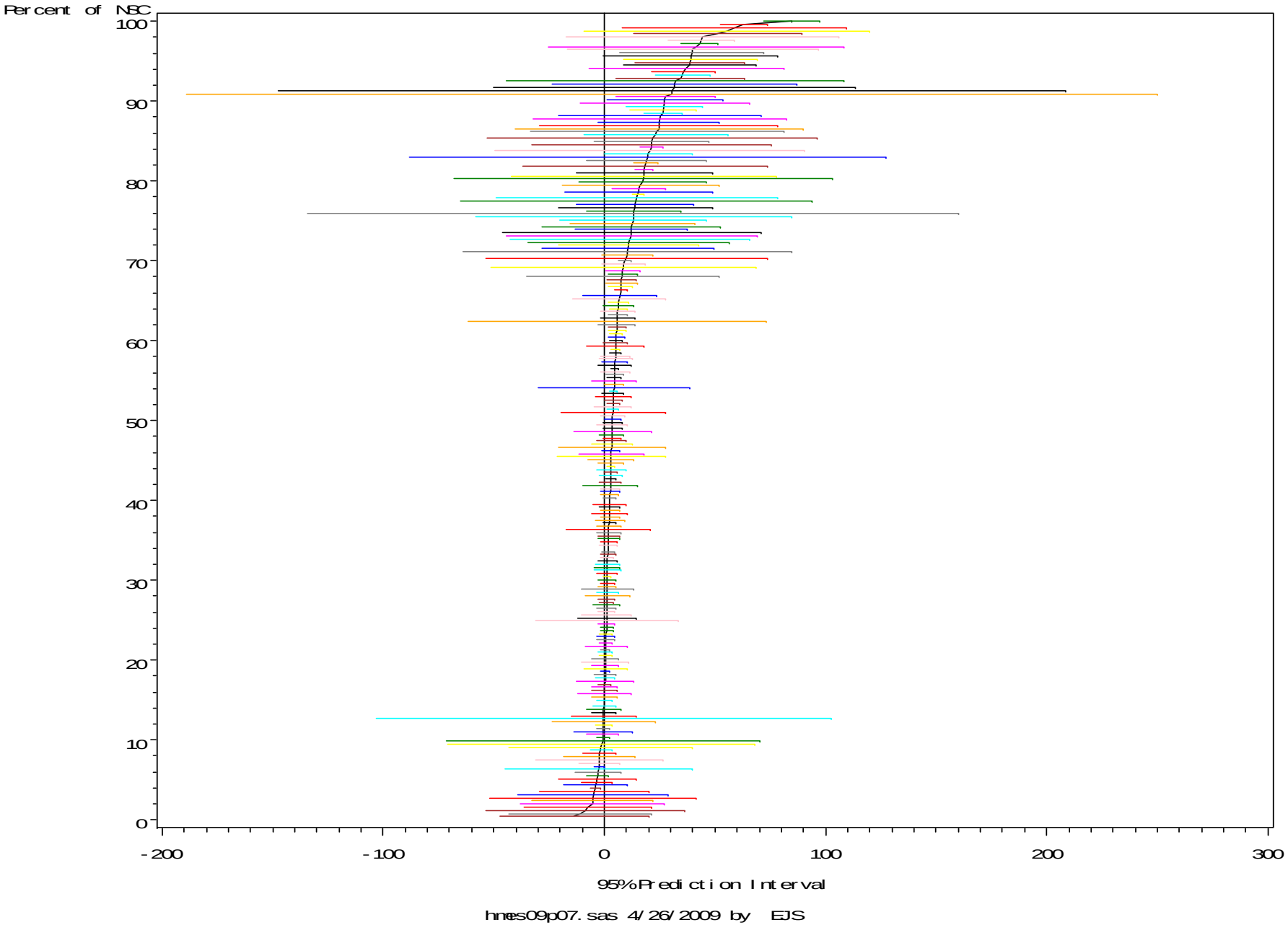
Conclusion:

- Confidence Intervals indicate the 'central' width of the distribution of the Estimator
- The 'width' doesn't change when Estimating different chemicals
- No adjustment for the length of confidence intervals is needed.
- Different from Hypothesis testing.

Cumulative Distribution of Estimates (n=253 chemicals)



Wild Type: Confidence Intervals by Chemical with 3 Doses below BMD(5)(n= 253)



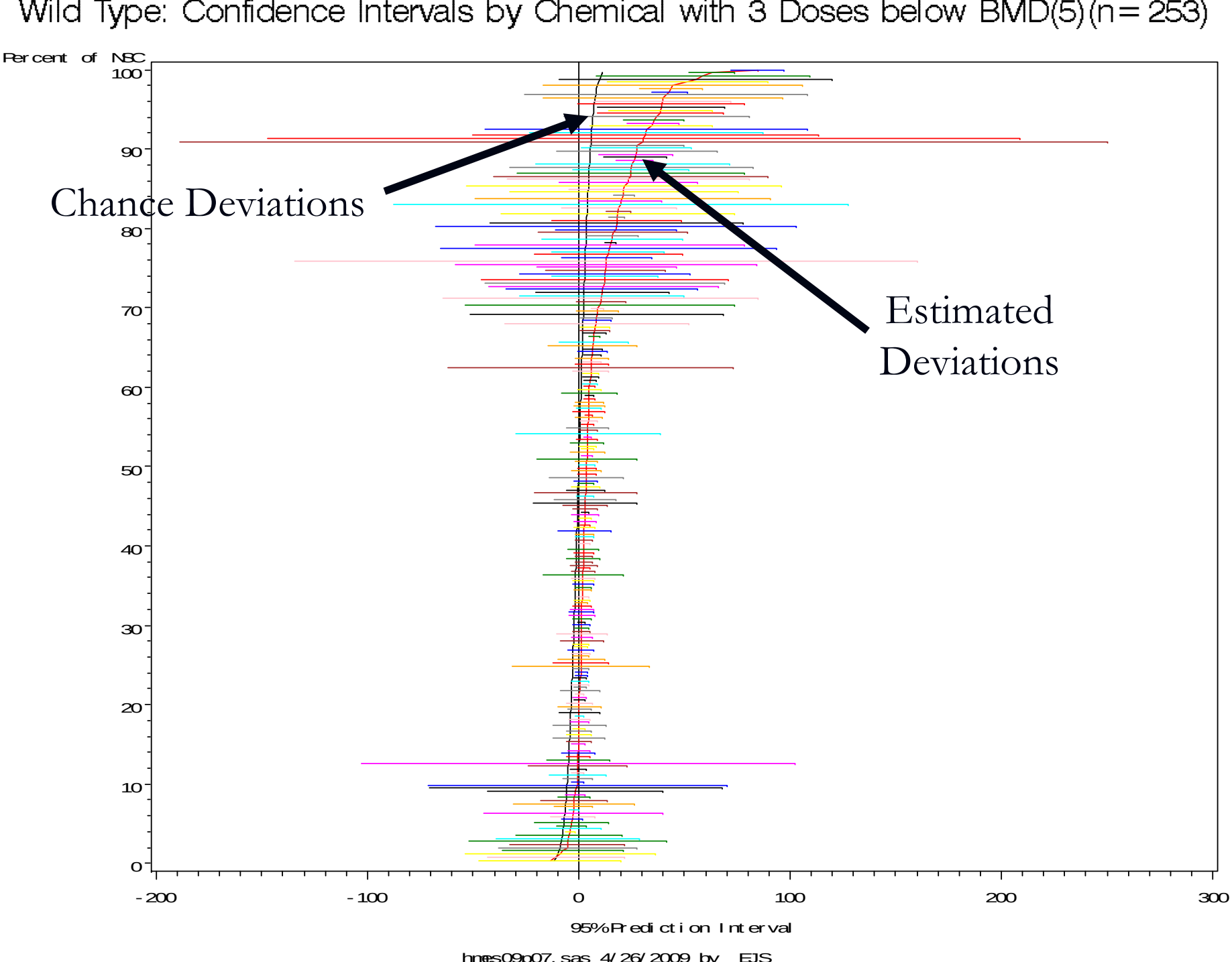
What would we expect if Threshold Model was the True Model?

- Pool Estimates of the Response Error Variance across Chemicals

$$\hat{\sigma}_e = 10.49$$

Use the estimate to estimate the standard error of the estimator

$$se\left(\overline{Y}\right) = \frac{\hat{\sigma}_e}{\sqrt{n}} = 4.28$$


Wild Type: Confidence Intervals by Chemical with 3 Doses below BMD(5)(n=253)
Percent of NSC
100
90
80
70
60
50
40
30
20
10
0
-200
-100
0
100
200
300
95% Prediction Interval
Chance Deviations
Estimated Deviations
hnes09p07.sas 4/26/2009 by EJS

Can we improve the estimates?

- Account for higher Response Error at extremes.
- Improve Accuracy of estimates-

YES:

- Motivate Estimator by Considering Study Chemicals to have arisen from a sample

Mixed model

Model of Response for Chemical "s"

Chemical $s = 1,\ldots,n$

Dose $t = 1,\ldots,m$

Replication $k = 1,\ldots,r(=2)$

$$Y_{stk} = \mu_{st} + E_{stk}$$
$$= \mu_s + \delta_{st} + E_{stk}$$

Define a Population of Chemicals $s = 1,\ldots,N$

Mean for Chemical "s"

Define Mean for Population $\mu = \sum_{s=1}^{N} \mu_s$

$$Y_{stk} = (\mu + \beta_s) + \delta_{st} + E_{stk}$$

Consider Study Chemicals as a sample

Population of Chemicals $s = 1, \ldots, N$

Sample of Chemicals $i = 1, \ldots, n$

For Chemical "s" $Y_{stk} = (\mu + \beta_s) + \delta_{st} + E_{stk}$

For the i^{th} Selected Chemical $Y_{itk} = (\mu + B_i) + D_{it} + E_{itk}$

Fixed Effects

Random Effects

Summary of Mixed Model

$$Y_{ijk} = (\mu + B_i) + E_{ijk}$$

Fixed

Random

i = chemical

J = *dose*

k = *replication*

Best Linear Unbiased Predictor (BLUP) from Mixed Model

$\mu + B_i$ Latent Response of ith Selected Chemical

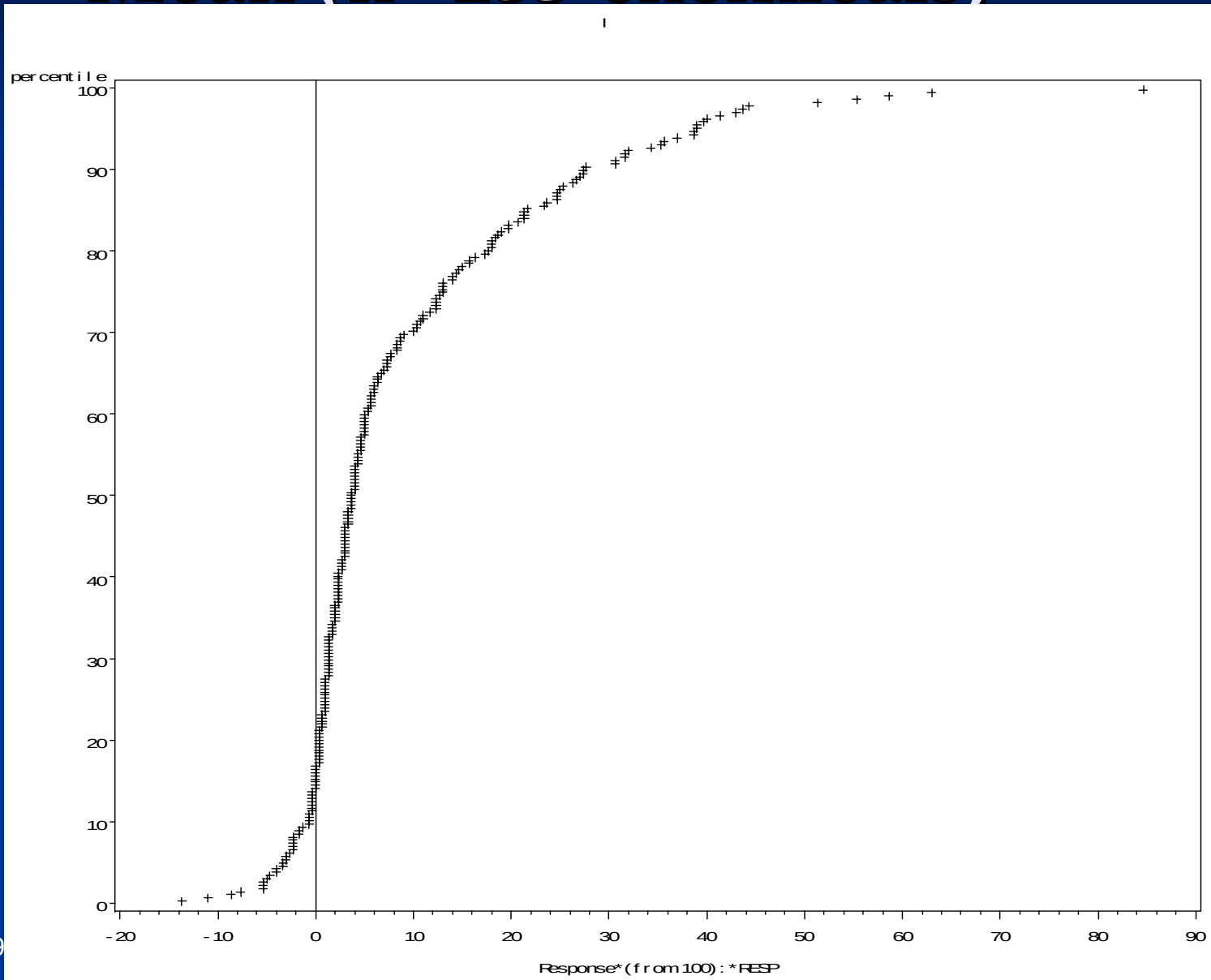
Predictor

$$\hat{P}_i = \hat{\mu} + k\left(\overline{Y}_i - \hat{\mu}\right)$$

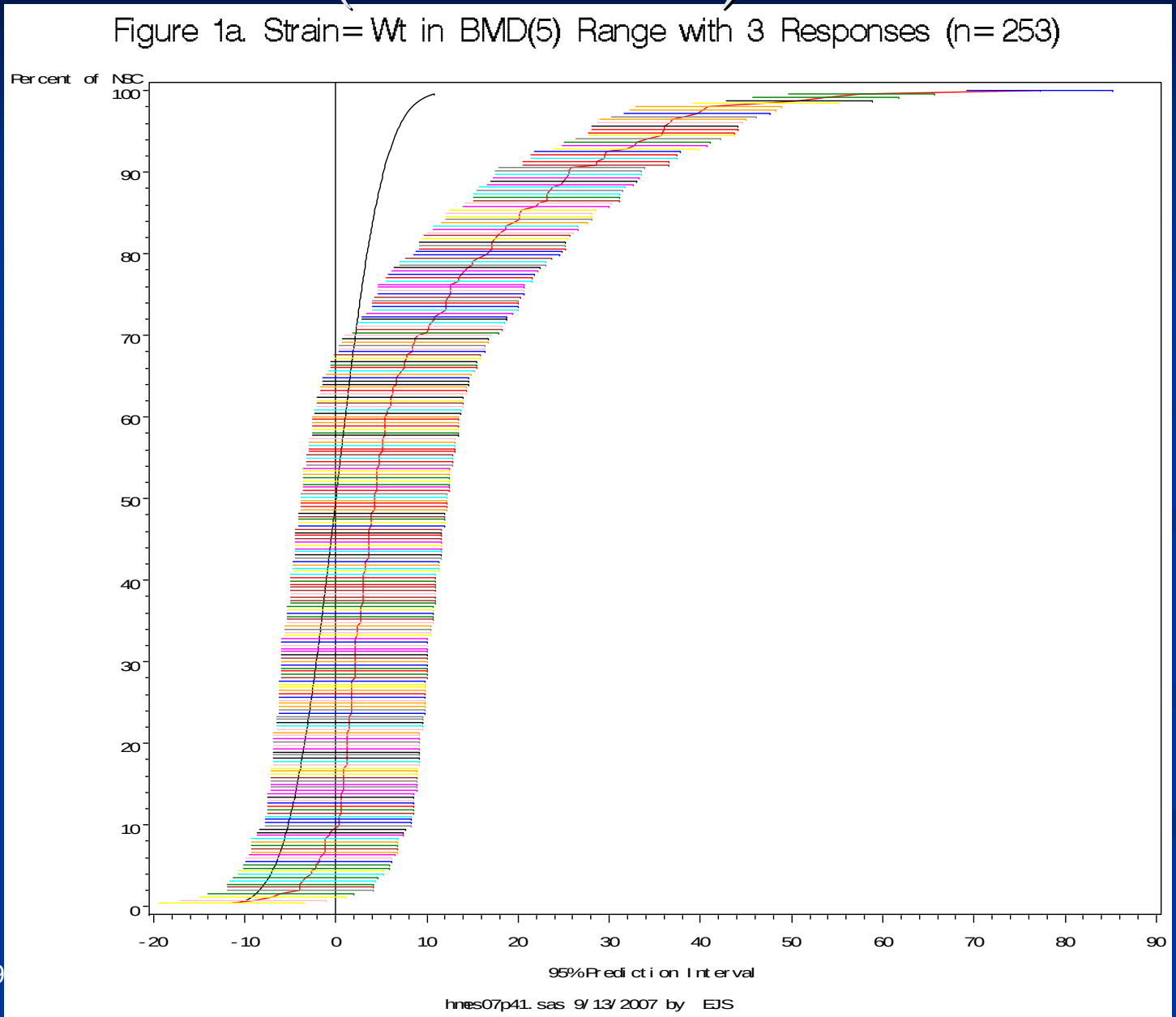
Shrinkage Constant

$$k = \frac{\sigma^2}{\sigma^2 + \sigma_e^2 / m}$$

Cumulative Distribution of Simple Mean (n=253 chemicals)



Cumulative Distribution of BLUP Estimates (n=253 chemicals)



Conclusions

- We can learn from data- even if it isn't an experiment.
- Estimation is the key- Hypothesis testing may lead us astray.
- More accurate estimates are possible with Mixed models
  - Require minimal additional assumptions
  - Dampen some of the response error
- Predictors are related to Bayesian methods

Thanks

- Questions?