Quantification of length-bias in screening trials with covariate-dependent test sensitivity

Dissertation Research

Sonya L. Heltshe, MS, Karen Kafadar, PhD

University of Colorado at Denver and Health Sciences Center

June 6$^{\mathrm{th}}$, 2007

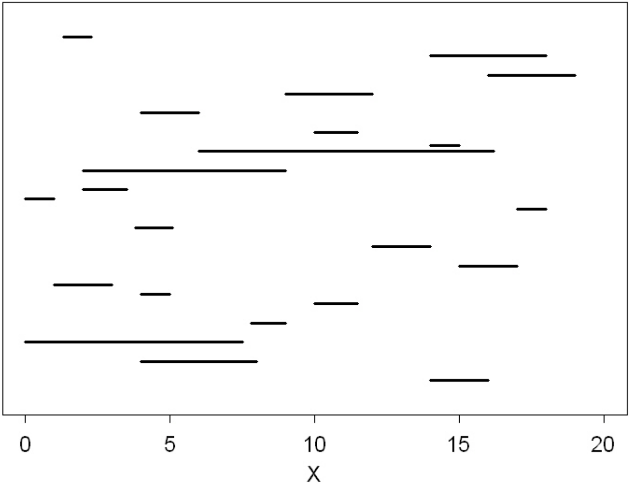
Outline

1. Length Bias/Inspection Paradox
2. Background: randomized screening trials
3. Length bias model in periodic screening
4. Results
5. Conclusion

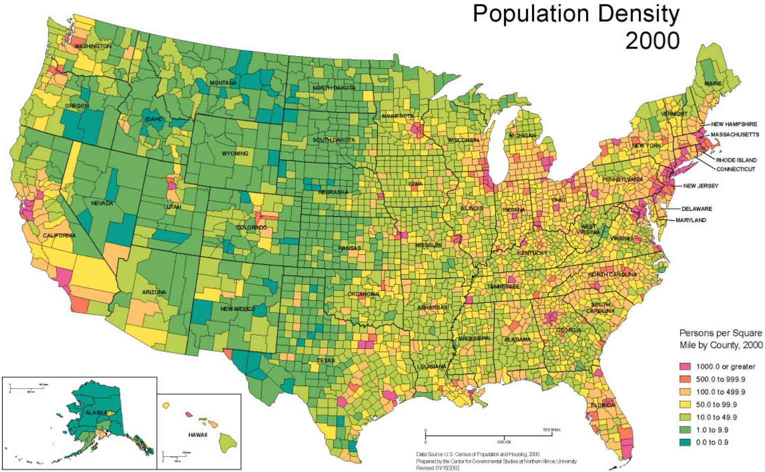
1. Length Bias or Inspection Paradox

- Sampling bias
- Probability of being sampled is proportional to size
  - lengths of yarn
  - traffic
  - gas lines
  - warranty database
  - family size
  - residential sampling
  - disease screening
  - proteomics (mass spectrometry)

Length bias



Cluster size bias

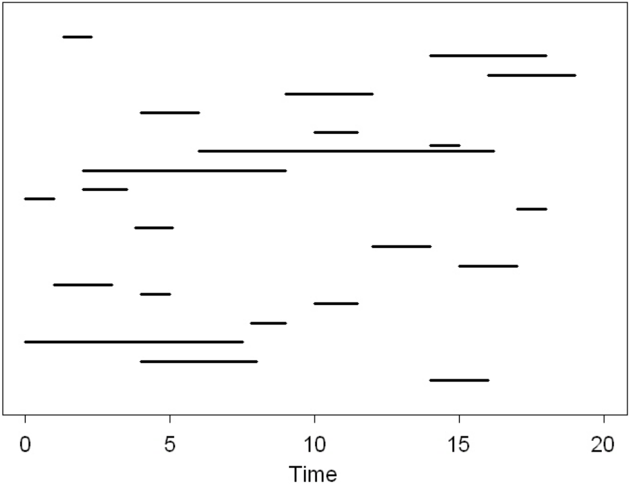


Length Bias or Inspection Paradox

- Industrial data
  - sampled variable is estimated
    - string
    - warranty
    - protein
  - true size/length is observable
- Disease/health data
  - sampled variable is correlated to estimable variable (Y, Z)
    - Hospital data: sample Stay, estimate $
    - Disease data: sample Pre-clinical, estimate Clinical duration

Length Bias: Example

Sample hospital stays to estimate final bill ($)



2. Randomized screening trials

- Screening mechanisms are used to detect disease early
- Participants are randomized to one of two arms
  - Study Arm -offered periodic screening.
  - Control Arm -standard medical care.
- Outcome of interest
  - disease specific mortality rates
  - extension of survival time

Randomized Screening Trials

- Disease progression model for unscreened and screen-detected cases if screening benefit exists

Unscreened Case:

| ⟵ Preclinical Duration ⟶ | ← Clinical Duration → |

Screen-detected case at X:

X

|LeadTime | ← Clinical Duration → | Benefit|

| ⟵ Preclinical Duration ⟶ |

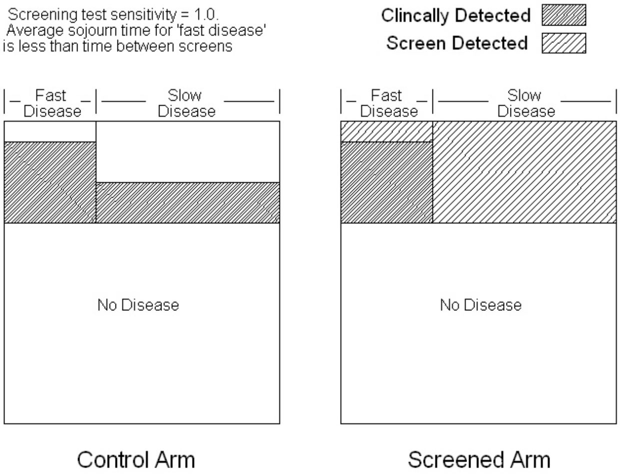
Randomized Screening Trials

- Survival since Dx in *unscreened* arm is an unbiased estimate of clinical duration
- Average lead time can be estimated using time since entrance into study
  - b/w arm difference in average survival time since entrance into study: $\Delta_{t_0} = \overline{benefit}$
  - b/w arm difference in average survival time since Dx: $\Delta_{D_x} = \overline{lead} + \overline{benefit}$
  - $\overline{lead} = \Delta_{D_x} - \Delta_{t_0}$

Randomized Screening Trials

- Survival time since entry into the *screened* arm confounds average benefit time, and average length bias
- Length bias:
  - when the probability of selection is proportional to the size (length) of the variable being sampled
  - Zelen (*1976*) recognized this phenomenon in screening. "Cases found by screening tend to be less advanced… slower-growing disease"
  - of particular concern when pre-clinical duration is positively correlated with clinical duration

Distribution of disease and detection in the two study arms



3. Length bias in periodic screening

A. Single Sample

- (Y, Z) denote preclinical and clinical duration of disease in the diseased general population.
- $E(Y) = \mu_Y$, $E(Z) = \mu_Z$, $SD(Y) = \sigma_Y$, $SD(Z) = \sigma_Z$, k.
- (Y^*, Z^*) denote the same in the length-biased sample. Cox and Lewis (1972) show:

$$f_{Y^*}(y) = \frac{y}{\mu_Y} f_Y(y)$$

- Schotz & Zelen, 1971; Kafadar & Prorok, 2005

$$f_{Y^*,Z^*}(y, z) = \frac{y}{\mu_Y} f_{Y,Z}(y, z)$$

Length bias in periodic screening

B. Successive Sampling

- Goal: Estimate $E(Z^*)/E(Z)$:
  - Find density function of Y_k^*: $f_{Y_k^*}(y) = g_k(y)f_Y(y)$
  - Calculate $E(Y_k^*) = \int_0^\infty y g_k(y) f_Y(y) dy$.
  - Calculate expected sojourn time for screen-detected cases:

$$E(Y^*) = \sum_{k=1}^{K} \beta(1-\beta)^{k-1} E(Y_k^*)$$

  - Calculate $E(Z^*)/E(Z)$:

$$E(Z^*) = \sum_{k=1}^{K} \beta(1-\beta)^{k-1} E(g_k(Y)Z)$$

Length bias in periodic screening

Estimate $E(Z^*)/E(Z)$ under a number of conditions:

- Distribution of pre-clinical duration
  - exponential, gamma, lognormal, weibull
- Correlation between pre-clinical and clinical duration, ρ
- Proportions of fast versus slow disease, ϕ
- moments of fast and slow disease (μ_ϕ, σ^2_ϕ, $\mu_{(1-\phi)}$, $\sigma^2_{(1-\phi)}$)
- Ratio of time between screens, δ, and mean preclinical duration
- Variable test sensitivity: $\beta(y)$

Disease mixture scenarios:

- **A:** Fast ($\mu_Y = 1.5$, $\mu_Z = 2$)
  & Slow ($\mu_Y = 5$, $\mu_Z = 6$)
- **B:** Fast ($\mu_Y = 1$, $\mu_Z = 1.5$)
  & Moderate ($\mu_Y = 3$, $\mu_Z = 2$)
- **C:** Moderate ($\mu_Y = 3$, $\mu_Z = 2$)
  & Slow ($\mu_Y = 7$, $\mu_Z = 9$)
- **E:** Very Fast ($\mu_Y = 0.5$, $\mu_Z = 0.5$)
  & Moderate ($\mu_Y = 3$, $\mu_Z = 2$)

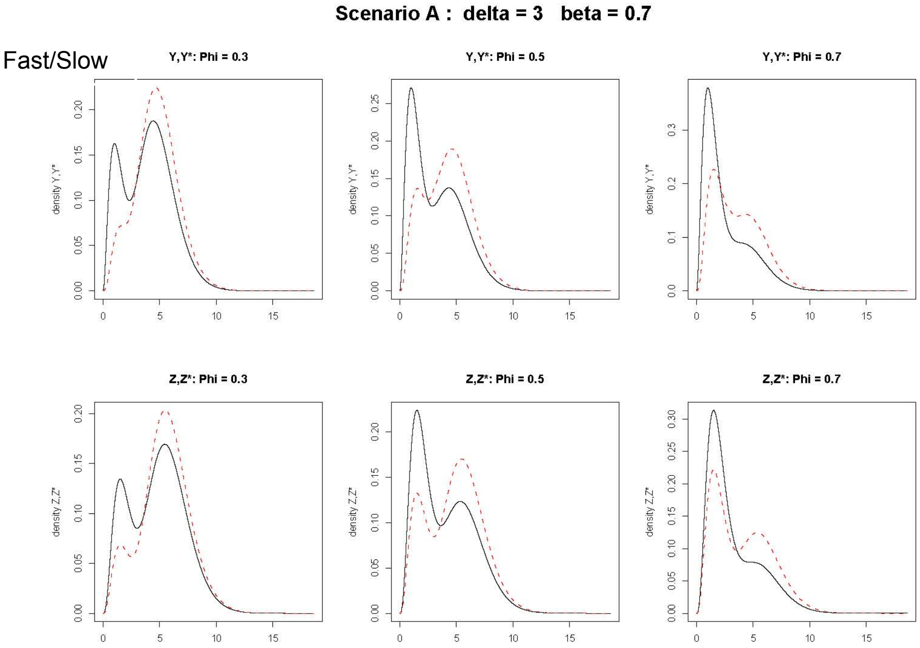
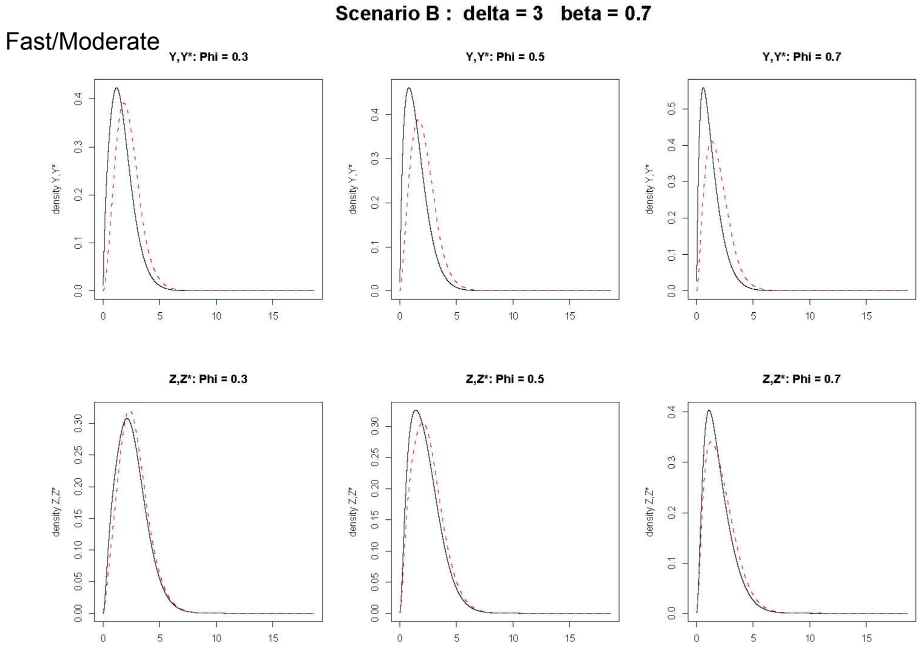
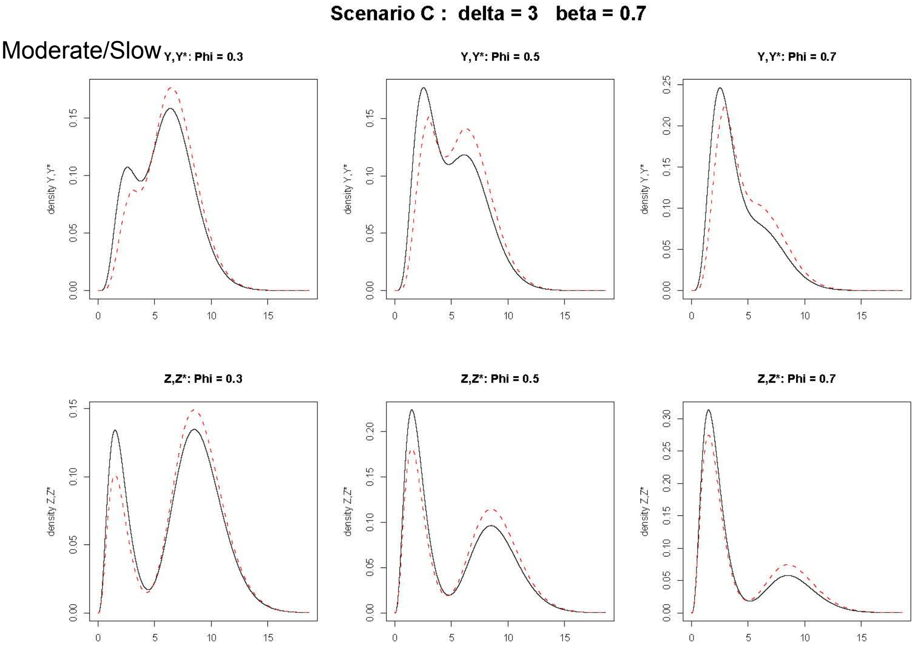


Figure 22: Scenario A: Plots of densities, unsampled and subjected to length-biased sampling. Top



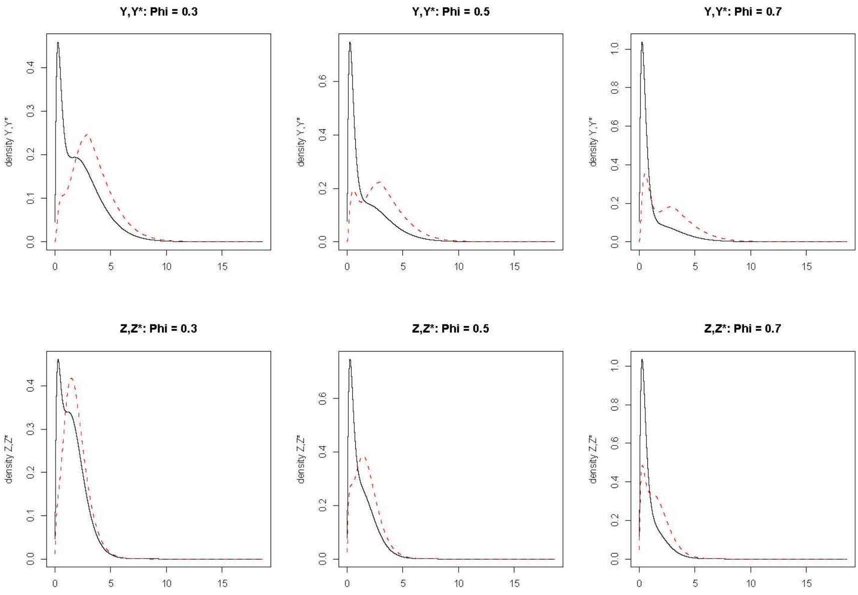
Scenario B : delta = 3 beta = 0.7
Fast/Moderate
Y,Y*: Phi = 0.3
Y,Y*: Phi = 0.5
Y,Y*: Phi = 0.7
density Y,Y*
Z,Z*: Phi = 0.3
Z,Z*: Phi = 0.5
Z,Z*: Phi = 0.7
density Z,Z*



Scenario C : delta = 3 beta = 0.7
Moderate/Slow
Y,Y*: Phi = 0.3
Y,Y*: Phi = 0.5
Y,Y*: Phi = 0.7
Z,Z*: Phi = 0.3
Z,Z*: Phi = 0.5
Z,Z*: Phi = 0.7
density Y,Y*
density Z,Z*

V. Fast/Moderate

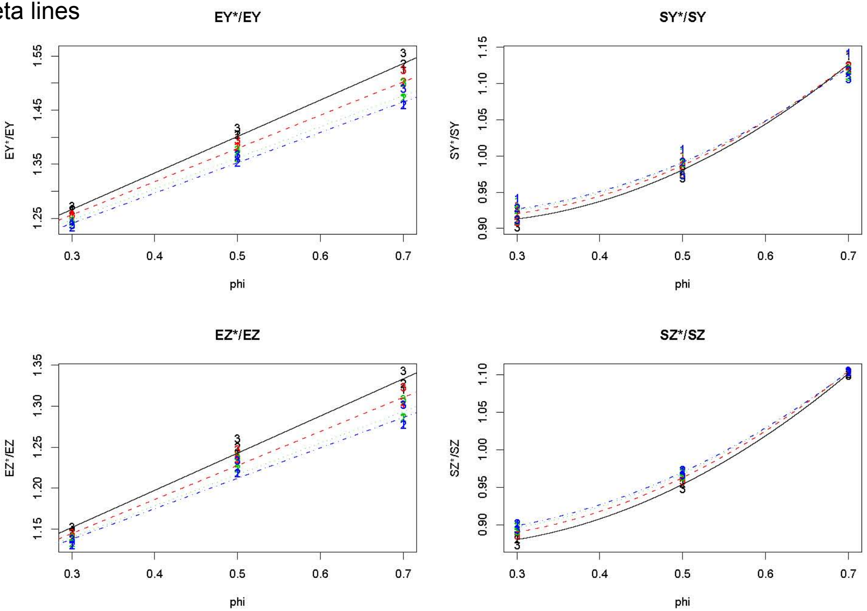
Scenario E : delta = 3 beta = 0.7



Fast/Slow

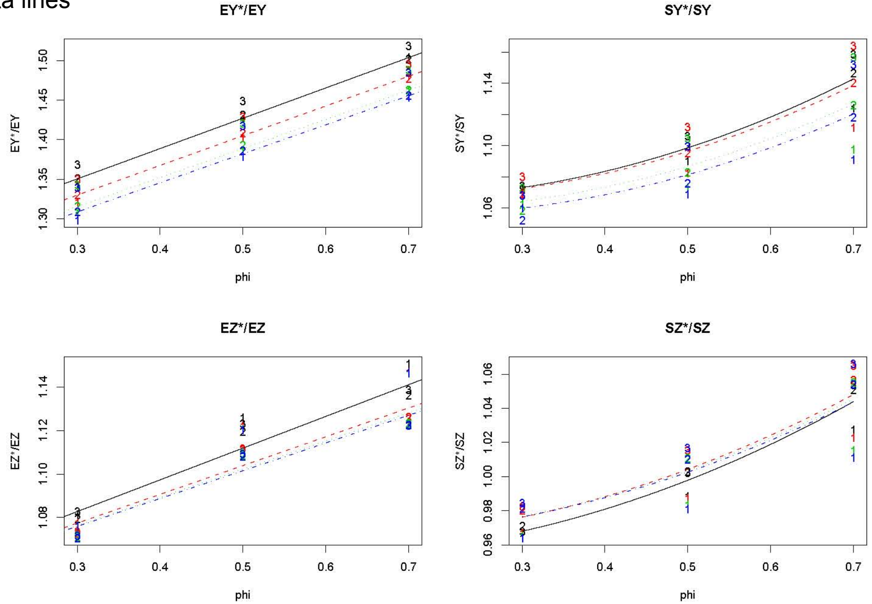
Beta lines

Scenario A



Fast/Moderate
Beta lines

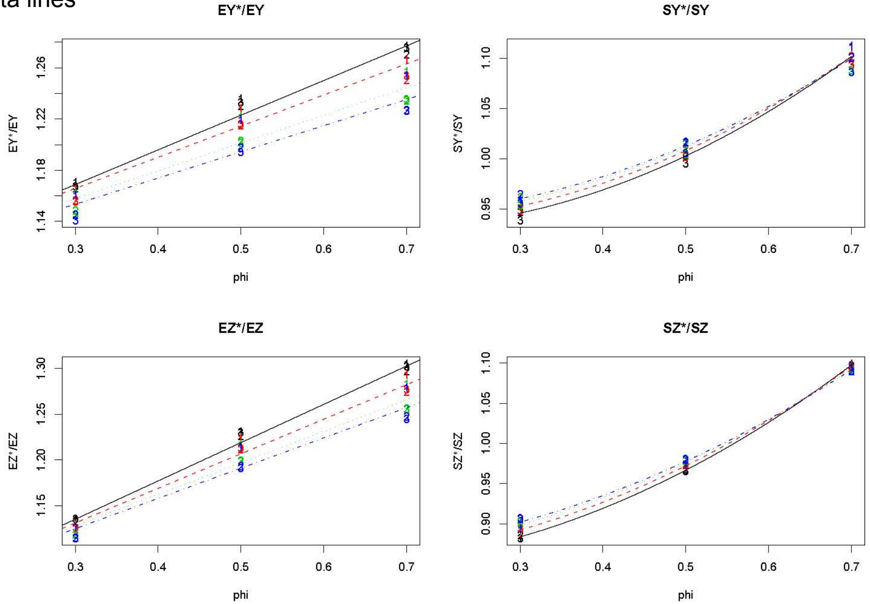
Scenario B

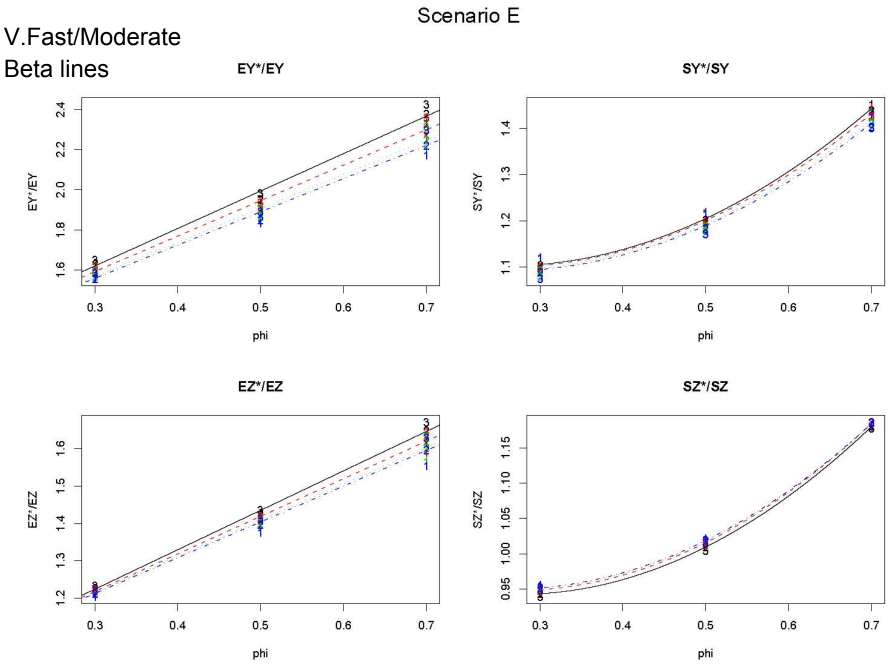


Moderate/Slow

Beta lines

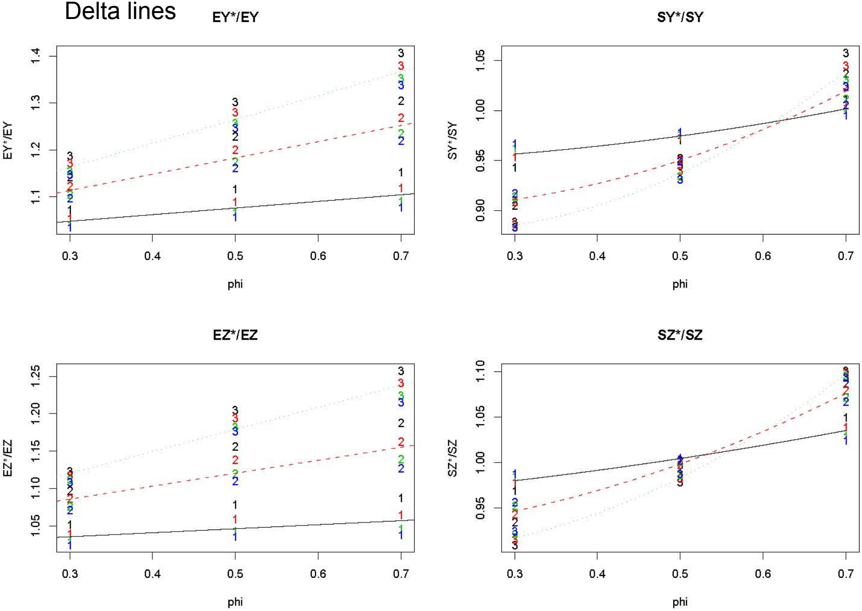
Scenario C




Scenario E
V.Fast/Moderate
Beta lines
EY*/EY
SY*/SY
EZ*/EZ
SZ*/SZ
phi

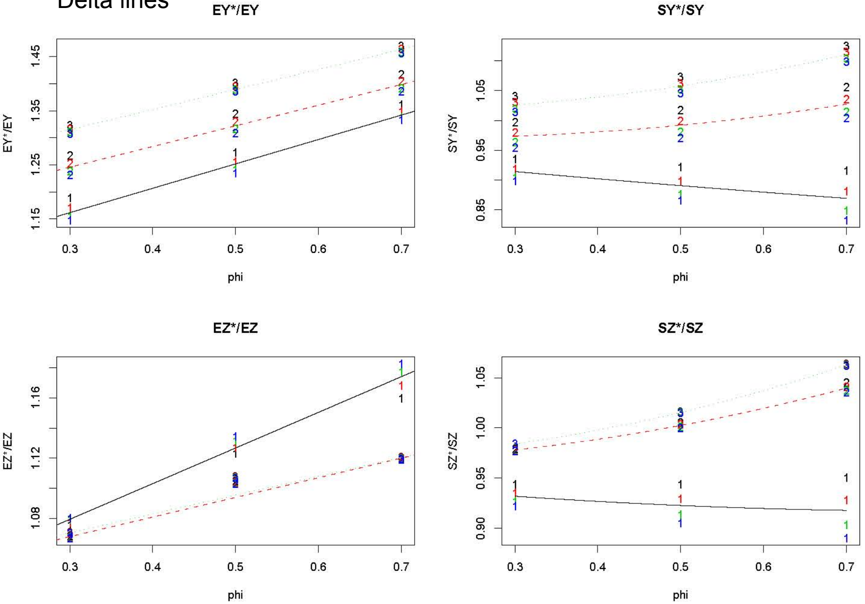
Scenario A (5 periodic screens)

Fast/Slow
Delta lines



Scenario B (5 periodic screens)

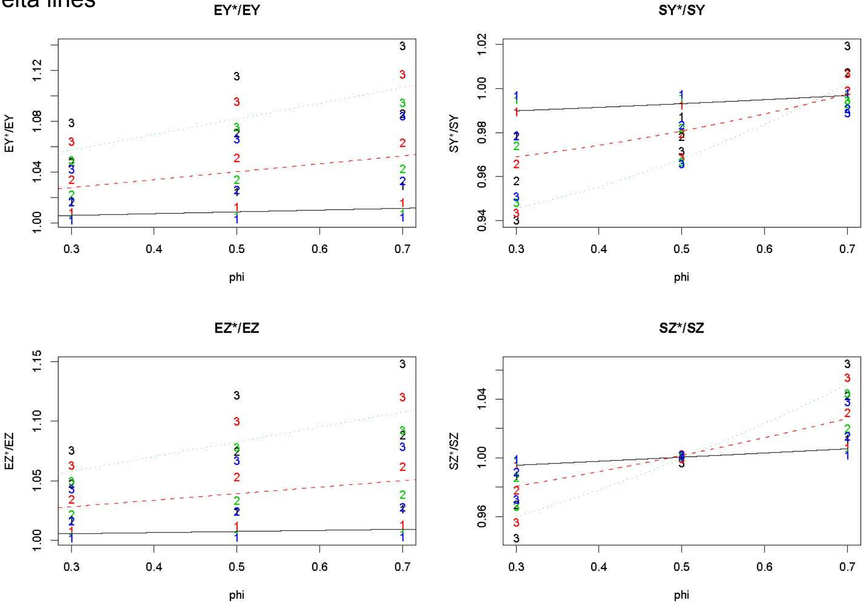
Fast/Moderate
Delta lines



Moderate/Slow

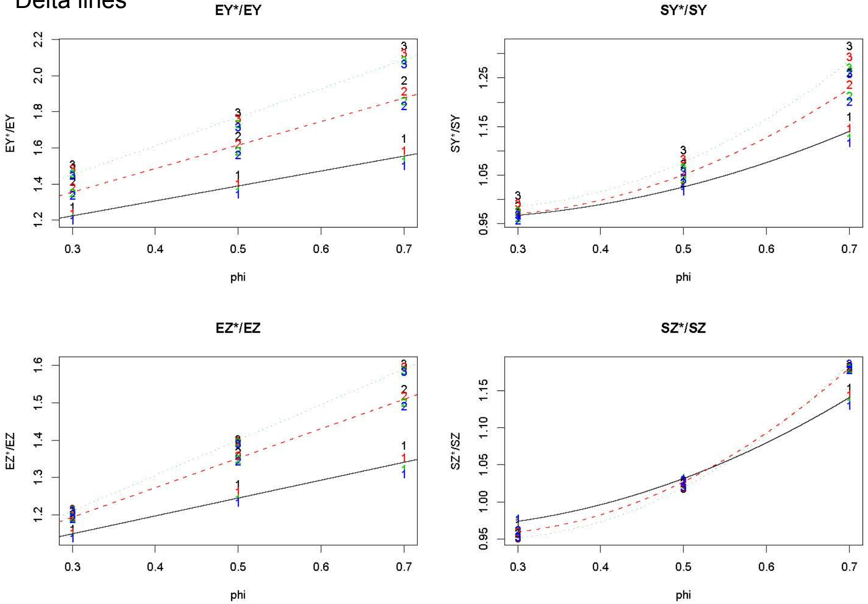
Delta lines

Scenario C (5 periodic screens)



Scenario E (5 periodic screens)

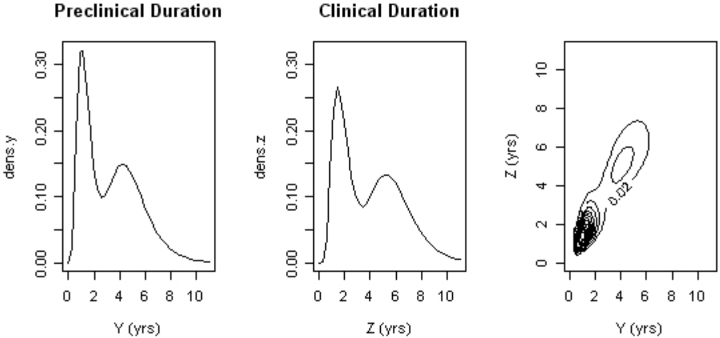
V. Fast/Moderate
Delta lines



Variable Test Sensitivity

- Start with mixture of correlated lognormals:

$$\begin{aligned} f(y,z) &= \phi\mathcal{LN}(y,z;\mu_Y=1.5,\sigma_Y=1,\mu_Z=2,\sigma_Z=1) \\ &+ (1-\phi)\mathcal{LN}(y,z;\mu_Y=5,\sigma_Y=2,\mu_Z=6,\sigma_Z=2) \end{aligned}$$



Example

- $\mathcal{LN}(y, z; \mu_Y = 1.5, \sigma_Y = 1, \mu_Z = 2, \sigma_Z = 1)$
- $\mathcal{LN}(y, z; \mu_Y = 5, \sigma_Y = 2, \mu_Z = 6, \sigma_Z = 2)$
- $\delta = 1.0,\ k = 4$
- As ρ increases, so does $E(Z^*)/E(Z)$

ρ	β_ϕ	$\beta_{(1-\phi)}$	$\frac{E(Z^*)}{E(Z)}$	
0.70	.9	.9	1.10	
0.70	.7	.7	1.11	
0.70	.7	.9	1.12	
0.80	.9	.9	1.16	
0.80	.7	.7	1.19	
0.80	.7	.9	1.21	$\longleftarrow$

- Higher β decreases bias unless it's disproportionably higher for slow disease

5. Conclusion

- Incorporate variable test sensitivity $\beta(y)$
- Quantify size-biased sampling under various conditions and distributions
- Determine most influential factors
  - distribution
  - ratio $\delta/E(Y)$
  - test sensitivity
  - correlation between Y and Z
  - ratio $E(Y)_\phi/E(Y)_{(1-\phi)}$
- Adjust survival time for length-bias
  $\rightarrow$ unbiased estimates of screening benefit