To Screen or Not to Screen?
Evaluating Risks & Biases in Cancer Screening Trials

Karen Kafadar

Department of Statistics

University of Virginia

kkafadar@virginia.edu

http://statistics.as.virginia.edu/faculty-staff/profile/kk3ab

In collaboration with Philip C. Prorok

Division of Cancer Prevention, National Cancer Institute

OUTLINE

1. What conditions indicate screening may be beneficial?
2. Benefits/Risks of large-scale screening
   - Potential biases: lead time, slow vs fast, overdiagnosis
   - Differential treatment of screened/not-screened cases
3. Examples: Breast/Prostate Cancer Screening
4. Summary

1. Considerations for Screening

Conditions favoring screening for disease

1. Serious public health condition

   obesity, *breast cancer*, *prostate cancer*, *colon cancer*

2. Well-defined target population at risk

   teens, *women* > 50, *men* 55–75, *all* > 60

3. Recognizable pre-clinical phase correlated w/disease progression

   high BMI, *lump*, *elevated PSA*, *polyp*

4. Exam is not harmful, non-invasive, convenient, "accurate"

5. Test results clearly indicate next steps

6. When disease is detected in pre-clinical phase, the treatment has clear benefits (*reduced mortality, extended survival)*

7. Acceptable sensitivity (Sn), high **specificity** (Sp), high prevalence $\Rightarrow$ high PPV

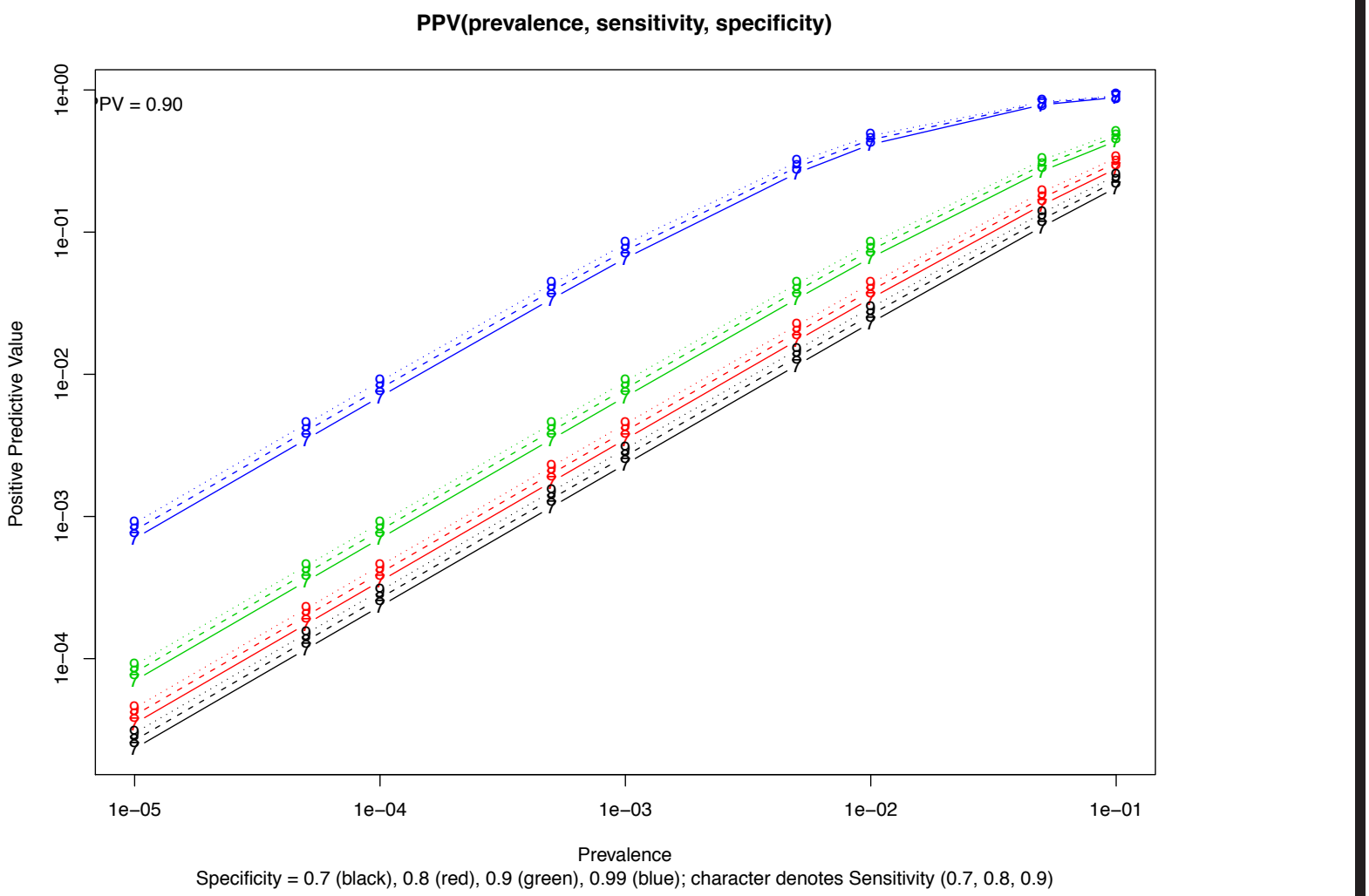
$$PPV = [(Sn)(p)]/[(Sn)(p) + (1 - Sp)(1 - p)]$$
$$= 1/[1 + 1/(OR \cdot LR)]$$

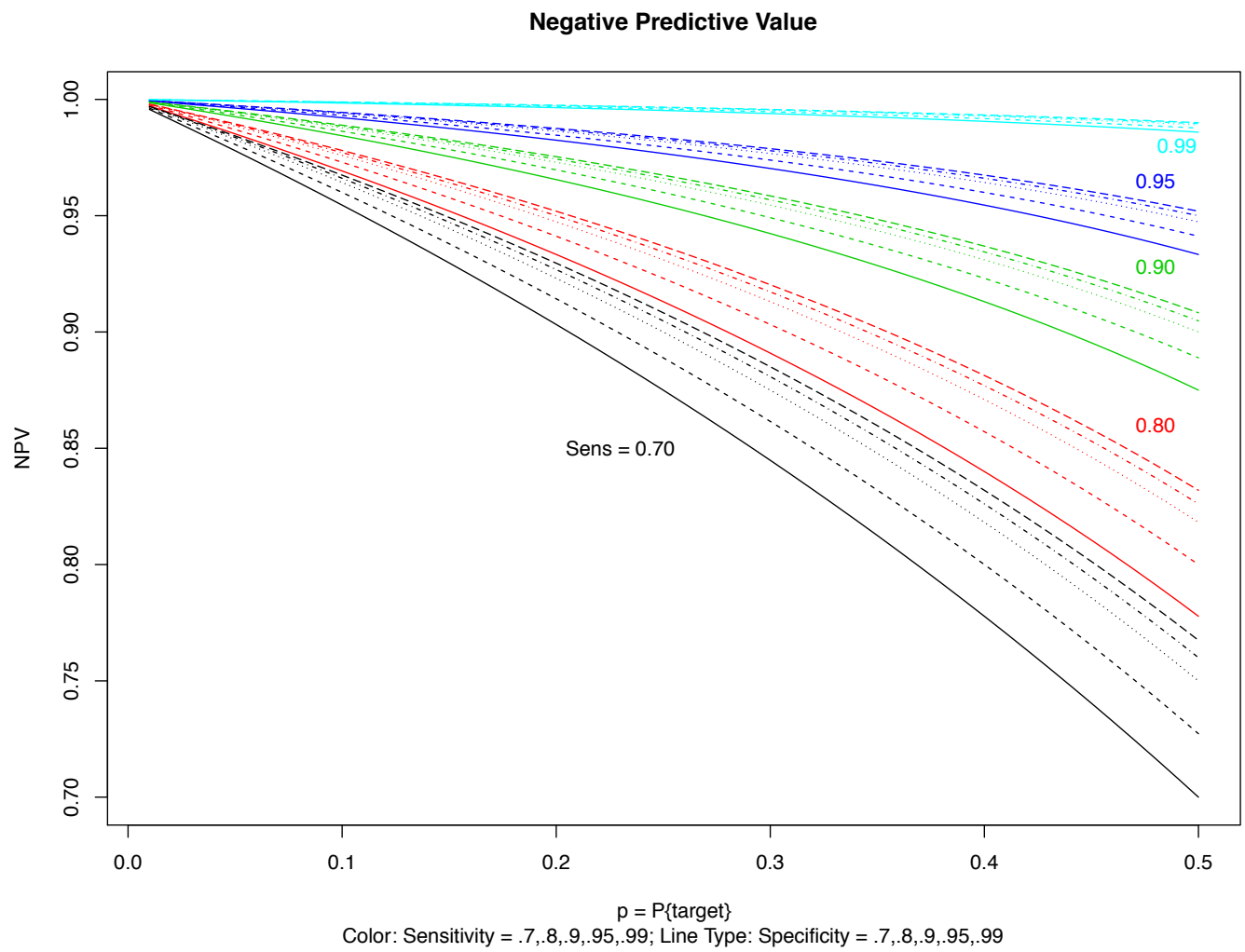
$LR = Sn/(1 - Sp)$, $OR = p/(1 - p)$

Note: compare two methods via ROC curves:

Same p, LR = tangent of ROC curve, so higher $LR \Rightarrow$ higher PPV.

For higher NPV, need higher $1/LR_{-} = Sp/(1 - Sn)$


PPV(prevalence, sensitivity, specificity)
PPV = 0.90
Positive Predictive Value
1e+00
1e−01
1e−02
1e−03
1e−04
1e−05
1e−04
1e−03
1e−02
1e−01
Prevalence
Specificity = 0.7 (black), 0.8 (red), 0.9 (green), 0.99 (blue); character denotes Sensitivity (0.7, 0.8, 0.9)


Negative Predictive Value
NPV
1.00
0.95
0.90
0.85
0.80
0.75
0.70
0.99
0.95
0.90
0.80
Sens = 0.70
0.0
0.1
0.2
0.3
0.4
0.5
p = P{target}
Color: Sensitivity = .7,.8,.9,.95,.99; Line Type: Specificity = .7,.8,.9,.95,.99

Advantages that screening offers

1. Simpler treatment (less invasive/costly/painful)
2. Better prognosis for *true* positives (if screening is beneficial)
3. Higher quality of life (less aggressive disease)
4. Some reassurance for true negatives (no disease)

Potential disadvantages

1. cost of exam (e.g., MRI to screen for breast cancer)
2. longer morbidity / attention to problem
3. increased costs for false positives
   (financial, physiological, psychological)
4. potential for overdiagnosis (strong evidence for
   neuroblastoma, breast/lung/prostate/cervical cancers)

False Positives: NLST *Croswell et al. Ann Int Med 2010*

Feasibility study for National Lung Screening Trial:
`ClinicalTrials.gov` NCT00006382

- Randomized to low-dose CT vs chest radiography (CR)
- 1 baseline, 1 annual screen, 1 year follow-up
- Current/former smokers ($\geq$ 30 pack-years), 55–74
- Stratified by age (55–64/65–74), gender (M/F), site (6)
- $n_{CT} = 1610$, $n_{CR} = 1580$; 66% 55–64; 42% female; 42% former smoker; 42% $\geq$ 60 pack-years

"False Positive": Subsequent testing confirmed negative

Logistic regression for FP after first screen/both screens

- age (55–64 vs 65–74), smoking status (current vs former), amount (30–59 vs 60+ pack-years) have little effect
- lowest: 55–64 former smokers; highest: 65–74 current smokers
- false positive rate varies by site (center):

Center	FP(CT)	FP(CR)
1	30–40%	14–19%
2	8–11%	8–11%
3	9–13%	5– 8%
4	9–13%	10–14%
5	21–30%	2– 3%
6	12–17%	8–12%

2. Evaluating Benefits of Screening: RCTs

Most effective method for evaluating risks and benefits of screening = RCT = **randomized controlled trial**:
Consenting participants are randomized to one of two arms:

- Study arm: **Offered** periodic screening exams
- Control arm: Follow usual medical care

(Note: potential for non-compliance in both arms)

Example: PLCO (Prostate, Lung, Colorectal, Ovarian)

- Recruitment Dec 1993 – Jun 2001 in 10 sites (incl CO, HI)
- Age at initial screen: 55–74
- Baseline screen + five annual screens
- Each arm: 37,000 men + 37,000 women, designed to detect 20%/15%/22%/35% reduction in P/L/C/O mortality (10 yrs)

Does screening reduce mortality, or extend survival?

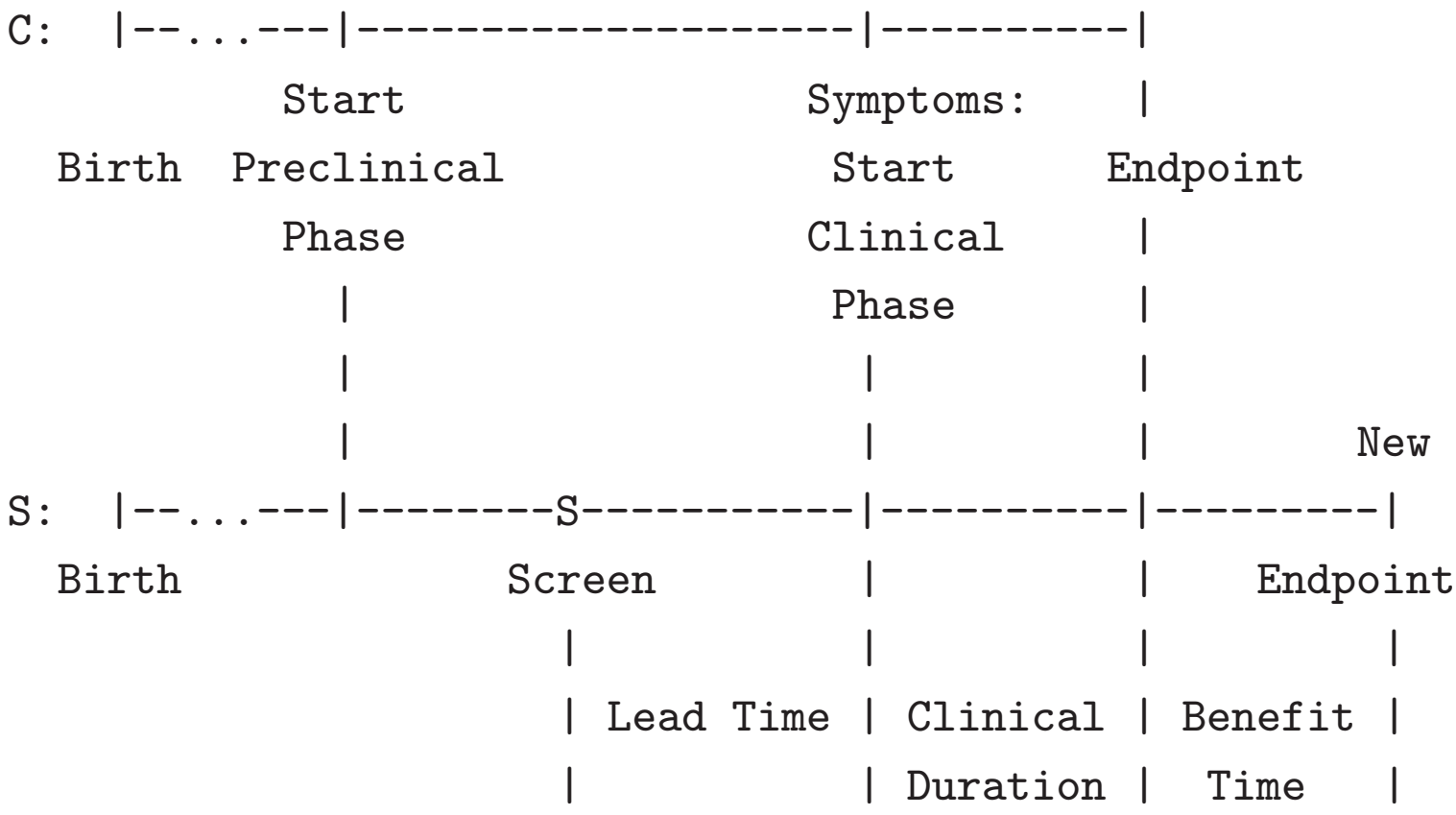
Metrics for evaluating screening:

1. Reduction in mortality: What proportion of deaths (from disease of interest) in control arm would be eliminated if they had been screened?
   Ex: 200 deaths in control arm & 180 deaths in study arm $\rightarrow$ 10% reduction in mortality

2. Extended survival: How much longer does a screen-detected case live over a comparable control case?
   Ex: average survival since diagnosis of screened cases = 2 yrs; average survival since diagnosis of control cases = 1.8 years $\rightarrow$ 15% (or 10 weeks) longer survival

(1) is not affected by *lead time* (see below)
(2) may be more interpretable (individual's longevity)

(2): Components of *survival since diagnosis* in Control/Study Case



True benefit of screening = "Benefit Time" (without lead time)

How to estimate *average benefit time*, accounting *jointly* for:

(a) **Case Group Comparability**: **When** to compare?

(b) **Lead time**: How much sooner is disease detected?

(c) **Overdiagnosis**: Would disease have ever surfaced?

(d) **Slower-growing**: More likely to be screen-detected

Additional issue in screening trials: **Comparable case groups**

(a) *When* to compare mortality rates/survival times? [KKPP03]

- "Limited time offer" (e.g., 5 years for PLCO)
- During screening: # study cases > # control cases
  (screening detects them sooner)
- After screening: # study cases < # control cases
  (screening has detected them already; control cases continue to accrue at steady rate)
- Long after screening ends: incidence in two arms about equal

At what time point are two case groups *comparable* for purposes of assessing screening effectiveness?

HIP Breast Cancer Screening Trial (Shapiro et al. 1988)

- Health Insurance Plan (HIP) of New York, 1963–1969
- 30,565 "usual medical care"
  30,131 Mammography **plus** Clinical Breast Exam
- Initial screen (Dec 1963 – June 1966)
  3 annual screens (to June 1969)
- 10,800 refused screening altogether
  20,200 screened at least once
  12,000 screened all four times
- Follow-up mail surveys at 5, 10, 15 years after entry

Cumulative incidence/mortality; Mortality reduction

HIP results

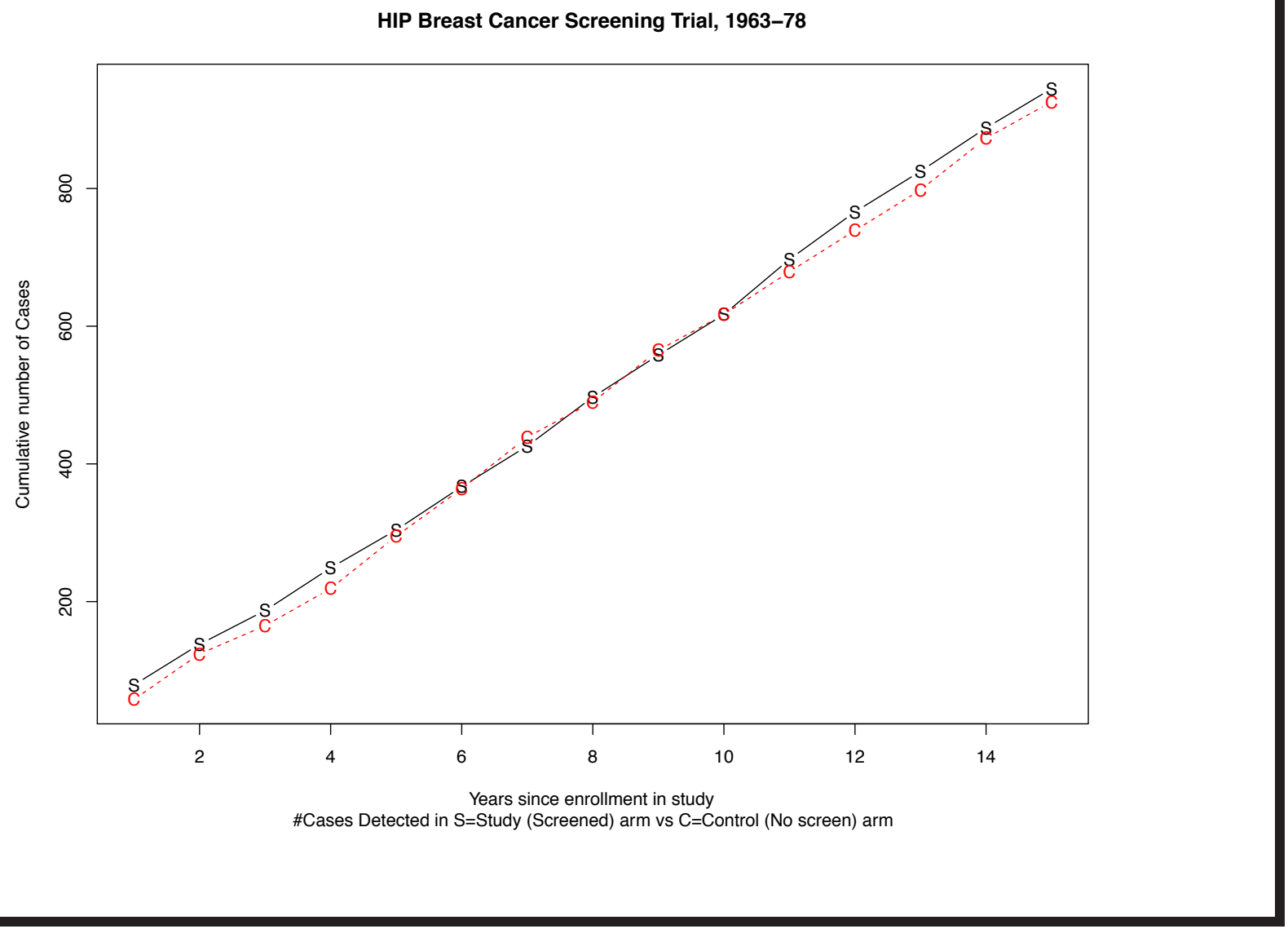
Yr	Incidence		Mortality		Rate/100,000		Ratio
	S	C	S	C	S	C	S/C
1	79	58	6	2	2.00	0.66	3.04
2	138	124	11	8	1.83	1.32	1.39
3	187	165	17	19	1.90	2.09	0.91
4	249	219	24	38	2.02	3.14	0.64
5	304	295	39	63	2.62	4.18	0.63
6	367	364	58	95	3.26	5.28	0.62
7	426	439	81	124	3.92	5.92	0.66

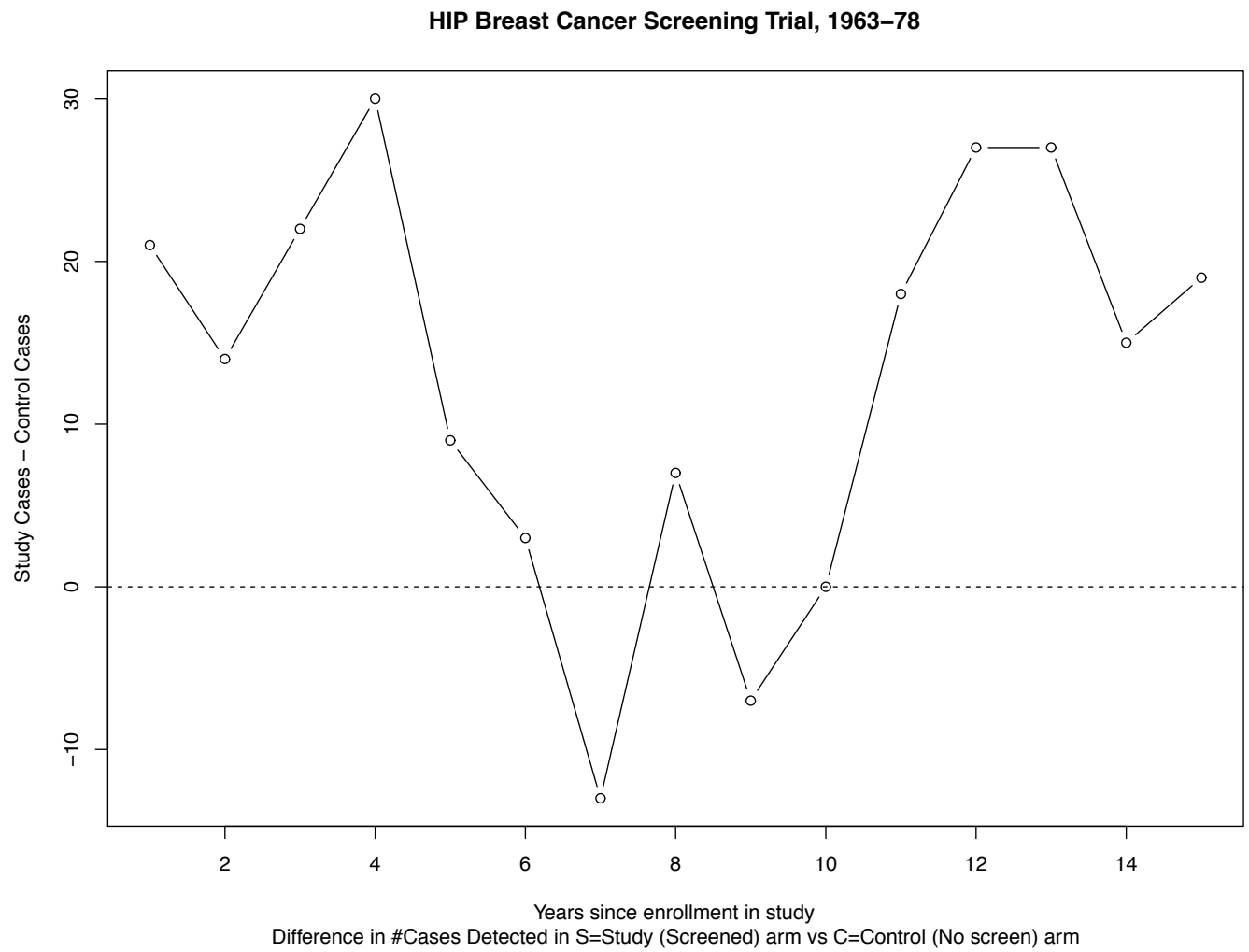
Connor and Prorok (1984 *Controlled Clinical Trials*)

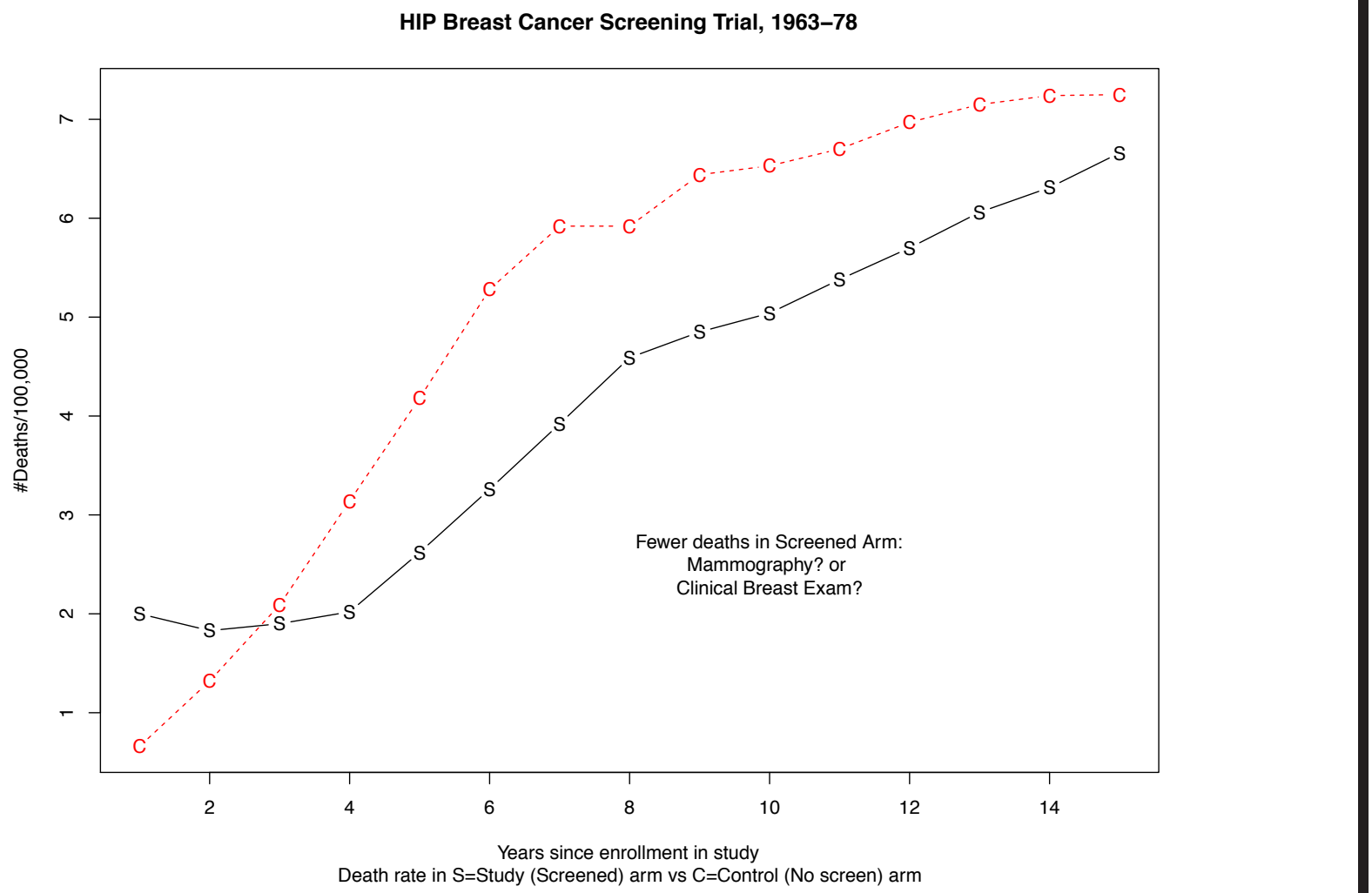
	Incidence		Mortality		Rate/100,000		Ratio
	S	C	S	C	S	C	S/C
7	426	439	81	124	3.92	5.92	0.66
8	497	490	108	141	4.59	5.92	0.78
9	558	565	128	172	4.85	6.44	0.75
10	617	617	147	193	5.04	6.53	0.77
11	697	679	172	217	5.38	6.70	0.80
12	766	739	198	245	5.70	6.97	0.82
13	825	798	227	271	6.06	7.15	0.85
14	888	873	253	294	6.31	7.24	0.87
15	945	926	285	314	6.66	7.25	0.92

Year	Fisher Exact[a]	Poisson Rate Test[b]	Logrank Test[c]	Gehan-Wilcoxon[c]
7	0.004	0.004	0.004	0.004
8	0.049	0.047	0.046	0.044
9	0.017	0.015	0.015	0.014
10	0.019	0.018	0.017	0.016
11	0.033	0.031	0.031	0.029
12	0.040	0.036	0.036	0.032
13	0.072	0.067	0.067	0.060
14	0.112	0.109	0.108	0.095
15	0.324	0.304	0.303	0.252

Answer: "catch-up" (related to mean sojourn time; KK+PP 2003)

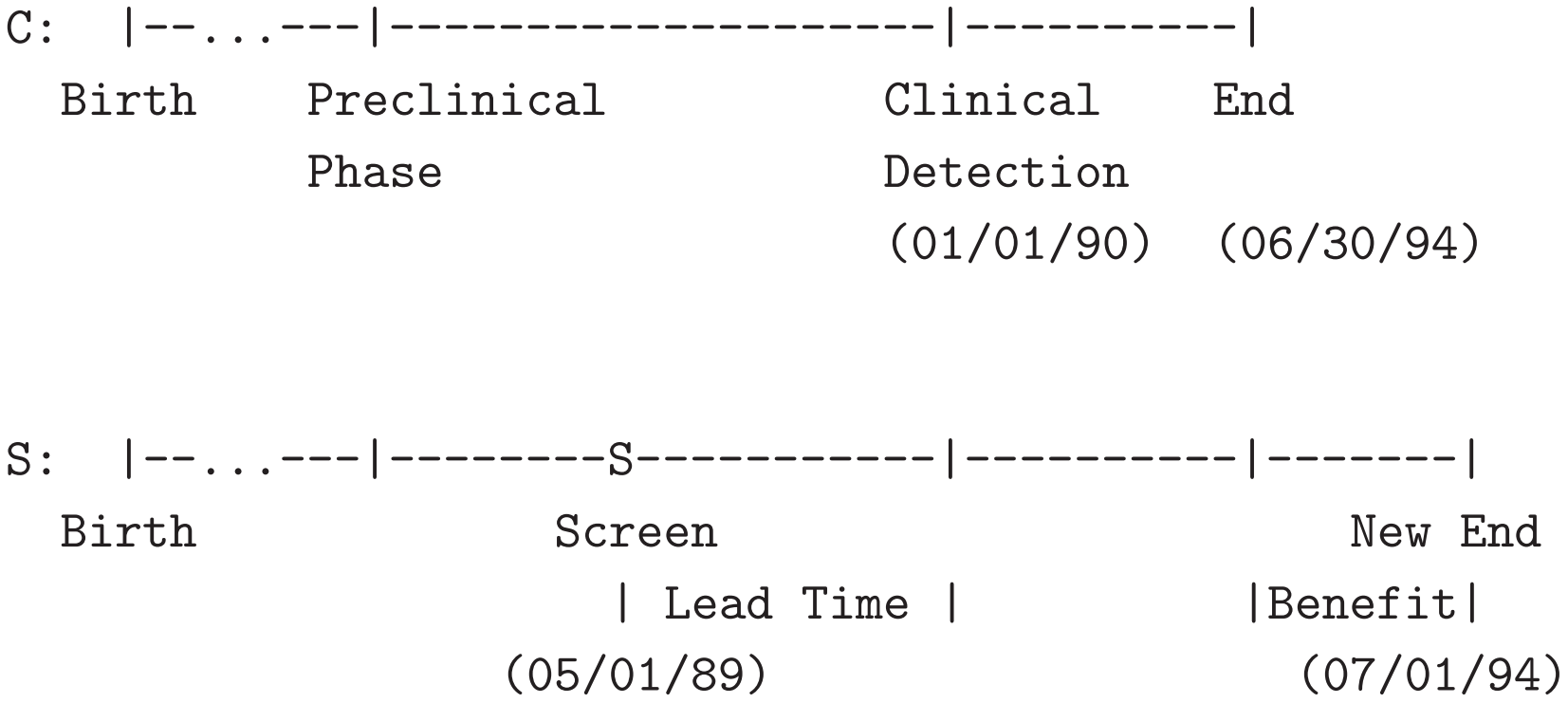

HIP Breast Cancer Screening Trial, 1963–78
Cumulative number of Cases
200
400
600
800
2
4
6
8
10
12
14
Years since enrollment in study
#Cases Detected in S=Study (Screened) arm vs C=Control (No screen) arm


HIP Breast Cancer Screening Trial, 1963–78
Study Cases – Control Cases
30
20
10
0
−10
2
4
6
8
10
12
14
Years since enrollment in study
Difference in #Cases Detected in S=Study (Screened) arm vs C=Control (No screen) arm


HIP Breast Cancer Screening Trial, 1963–78
#Deaths/100,000
1
2
3
4
5
6
7
2
4
6
8
10
12
14
Fewer deaths in Screened Arm:
Mammography? or
Clinical Breast Exam?
Years since enrollment in study
Death rate in S=Study (Screened) arm vs C=Control (No screen) arm

Do screen-detected cases actually live longer?

(b) Lead time bias:



- **S** detected 7 months earlier — *but lived only 1 more day*
- Need randomized trial to estimate lead/benefit times

Randomized Screening trial: survival *since time of entry into trial* not biased by lead time (survival *since diagnosis* confounds lead time and benefit time)

Estimates of average lead time (*avg time to Dx in screened arm – avg time to Dx in control arm*) can have large standard errors

- Health Insurance Plan, NY (1966-70, Mammography + CBE)
  Program lead time: 3.0 months (SE 1.6 months);
  Adjusted for screened-detected cases: 8.5 months (SE 4.5 mo)
- European Randomized Study of Screening for Prostate Cancer (Rotterdam): Lead time $\approx$ 12.3 yrs (age 55), 6 yrs (age 75)
- Prostate part of PLCO: Lead time $\approx$ 2 yrs (SE 1 yr)

(c) Overdiagnosis & Overtreatment

Diagnosis of condition that would not have been diagnosed during person's lifetime had screening not been conducted: condition was not harmful/life-threatening, death arose from another cause

- Mayo Lung Project, 1971–1983:
  Annual X-rays/Sputum cytology (6 yrs), 9211 males
  200 cases in study arm; 160 cases in control arm $\Rightarrow$ 20% excess
  (Marcus et al. JNCI 2006)
- Ovarian in PLCO, 15 years after screening ended:
  236 cases in study arm, 209 cases in control arm $\Rightarrow 27/236 = 11.4\%$ cases may have been overdiagnosed *(Buys et al. 2011)*

Consumer Reports on Health, Feb 2010 (p.7):

- "**False Positives.** When women start mammograms at age 40 instead of 50 there's a jump in false positives, or worrisome findings that prove harmless after additional testing, such as follow-up mammograms and biopsies. That testing can cause anxiety and expose women to unneeded radiation."
- "**Overdiagnosis and unnecessary treatments.** Mammography sometimes detects slow-growing tumors that left alone might never result in death or even bothersome symptoms. But doctors cannot say for certain which cancers are harmless, so all usually get treated. And those treatments, including chemotherapy, lumpectomy, mastectomy, and radiation, pose some risks."

(d) Aggressiviness of disease

- Periodic screening (annual, biannual) will miss aggressive cases (start & end very fast: screening never had a chance)
- Cases that arise in screened arm: more favorable prognoses

Length-biased sampling (KK + PP 2012):
Selection probability $\propto$ length (size) of observation (sojourn time)

Zelen (1976): *"People who are diagnosed by an early detection program do not constitute a random sample of preclinical cases. Cases found by screening tend to be less advanced....Women who are found earlier in a detection program tend to ... have slower-growing disease."*

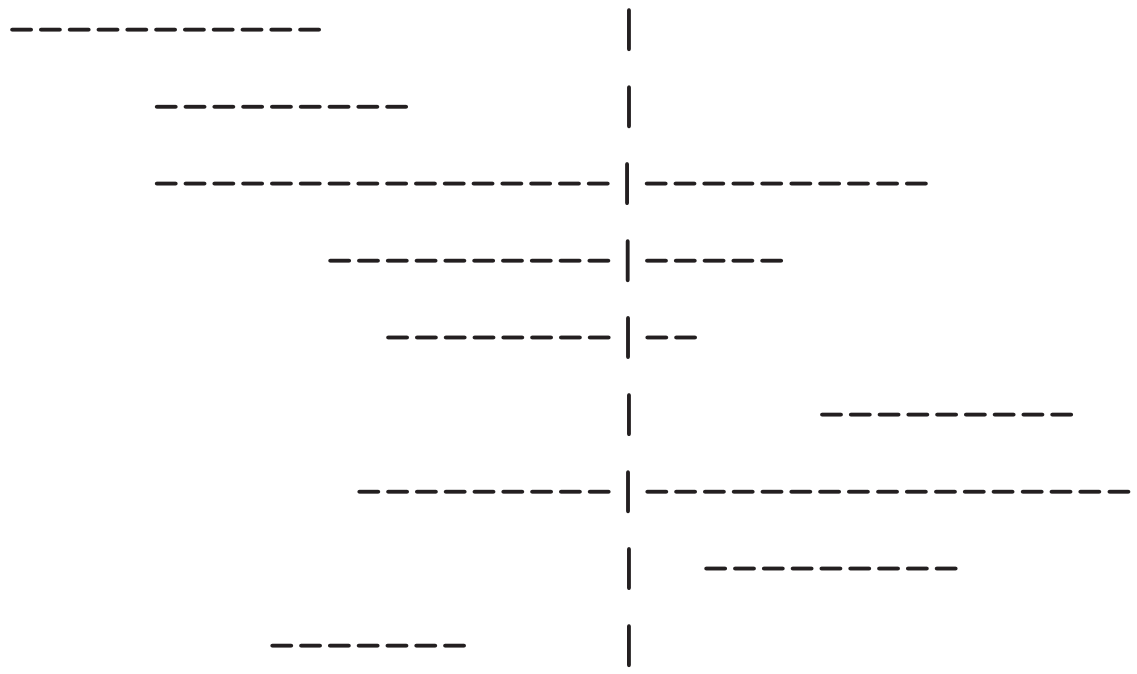
Examples of Length Biased Sampling

Arises when measurement process favors experimental units whose lengths (sizes) are proportional to selection probability $\Rightarrow$
Selection probability $\propto$ length (size) of observation

- Sample particles from mixture (heavier ones more likely)
- Survey hospital patients (longer stays $\Rightarrow\uparrow$ probability)
- Select units in database (longer in system $\Rightarrow\uparrow$ probability)
- **Longer sojourn times more likely to cross screen point**

Bias from length biased sampling is not eliminated by randomized design

Single screen (Cox 1969, Cox and Lewis 1972)



Cases with longer preclincal durations more likely to be "caught"

Preclinical durations (PD) Y_1, ..., Y_n

$f_Y(\cdot)$ = pdf of unsampled PDs

$f_{Y^*}(\cdot)$ = pdf of length-biased sampled PDs

$n_y = \#$ of Y_i's with lengths $y \leq Y_i \leq y + dy$

$f_{Y^*}(y) = \lim_{n\to\infty}(\text{Prop of } \sum Y_i \text{ due to intervals of length } y)$
$= \lim_{n\to\infty}(y \cdot n_y / \sum Y_i)$
$= \lim_{n\to\infty}(y \cdot n_y/n / \sum Y_i/n)$
$= y \cdot f_Y(y)/\mu_y \equiv g(y) \cdot f_Y(y)$

$E(Y^*)/E(Y) = (1 + \sigma_y^2/\mu_y^2) = (1 + CV_y^2)$

Effect on apparent benefit time/extended survival: 5-20%

Longer sojourn times tend to be positively correlated with longer clinical durations and hence less aggressive disease

Compare survival experiences of screen-detected cases with non-screen-detected cases when SD cases may be less aggressive control cases? (SD cases may be less aggressive)

Magnitude of effect of length biased sampling depends on mean sojourn time & screening frequency

Ex: HIP Breast cancer screening trial:

132 **Screen-detected** study cases: **63%** node-negative

91 study **interval** / 73 study **refusers** cases: **47%** node-negative

284 **Control** cases: **46%** node-negative

Ex: Proportion of Prostate PLCO cases by Clinical Stage I/II/III/IV

Study*	1500	97.0	1.5	1.0	0.2
Control	2074	95.5	1.4	2.1	0.4

**Excludes: 1952 cases: 154 never screened; 549 at baseline screen; 374 interval; 875 after screening*

How to adjust for this effect?

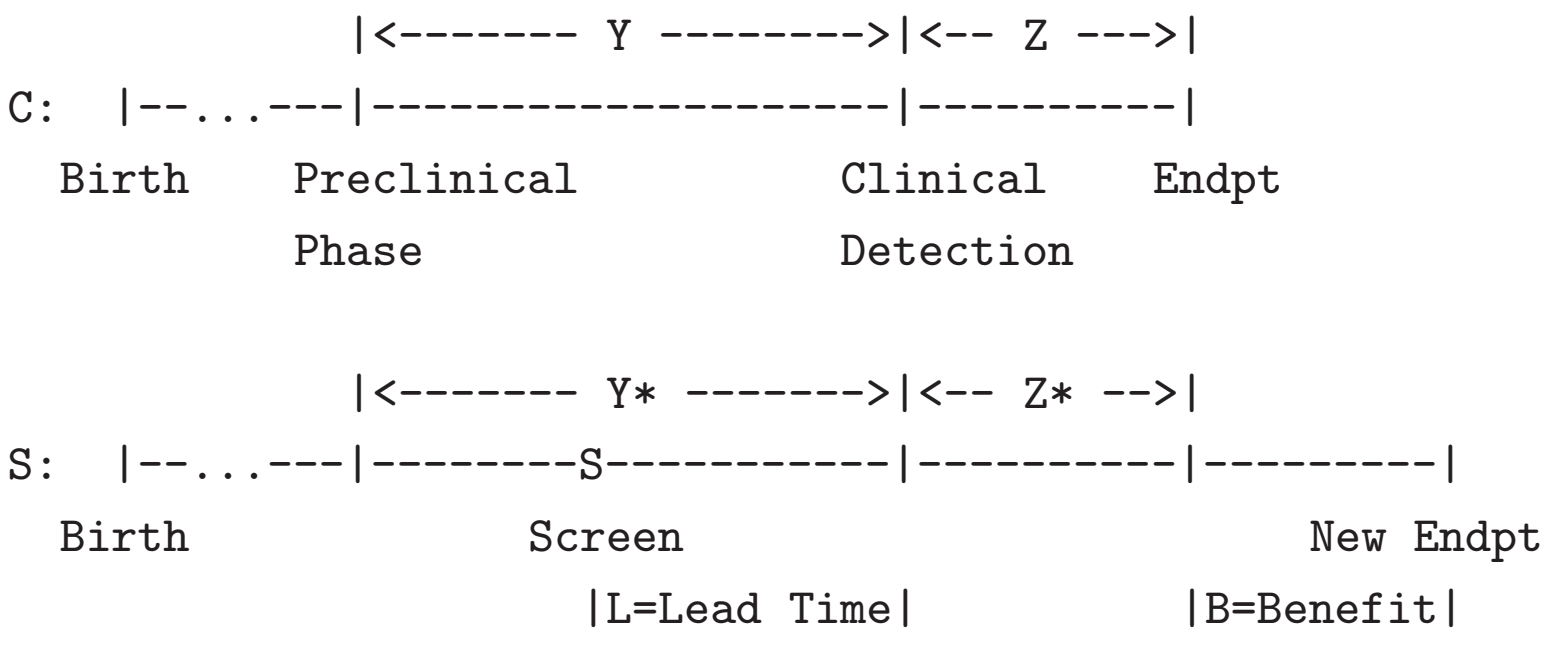
General size-biased sampling (Scheaffer 1972):
geographic areas, volumes, ...

$$f_{Y^*}(y) = y^\alpha \cdot f_Y(y)/\mu_\alpha, \mu_\alpha \equiv \int_0^\infty y^\alpha f_Y(y) dy$$

Typical goal in most problems: Estimate $E(Y^*)$

Our problem:

- (Y_i, Z_i) = (preclinical, clinical) durations, joint pdf $f_{YZ}(\cdot, \cdot)$
- Length biased sampling is on Y, but
  - we cannot observe Y
  - variable of real interest is Z
- How does the tendency for screening to select longer Ys (sojourn times) affect distribution of Zs (clininal durations)?



Jointly estimate average L, B Z^*, *when affected by overdiagnosis*

Estimating Components of Survival: Model

- Y = duration of preclinical phase (unobserved)
- Z = duration of clinical phase
- $f_{Y,Z}(\cdot,\cdot)$ = joint pdf of Y, Z
- μ_y, σ_y^2, CV_y = mean, variance, CV of Y
- μ_z, σ_z^2, CV_z = mean, variance, CV of Z
- Y^* = length-biased sampled Y
- Z^* = Clinical duration corresponding to Y^*

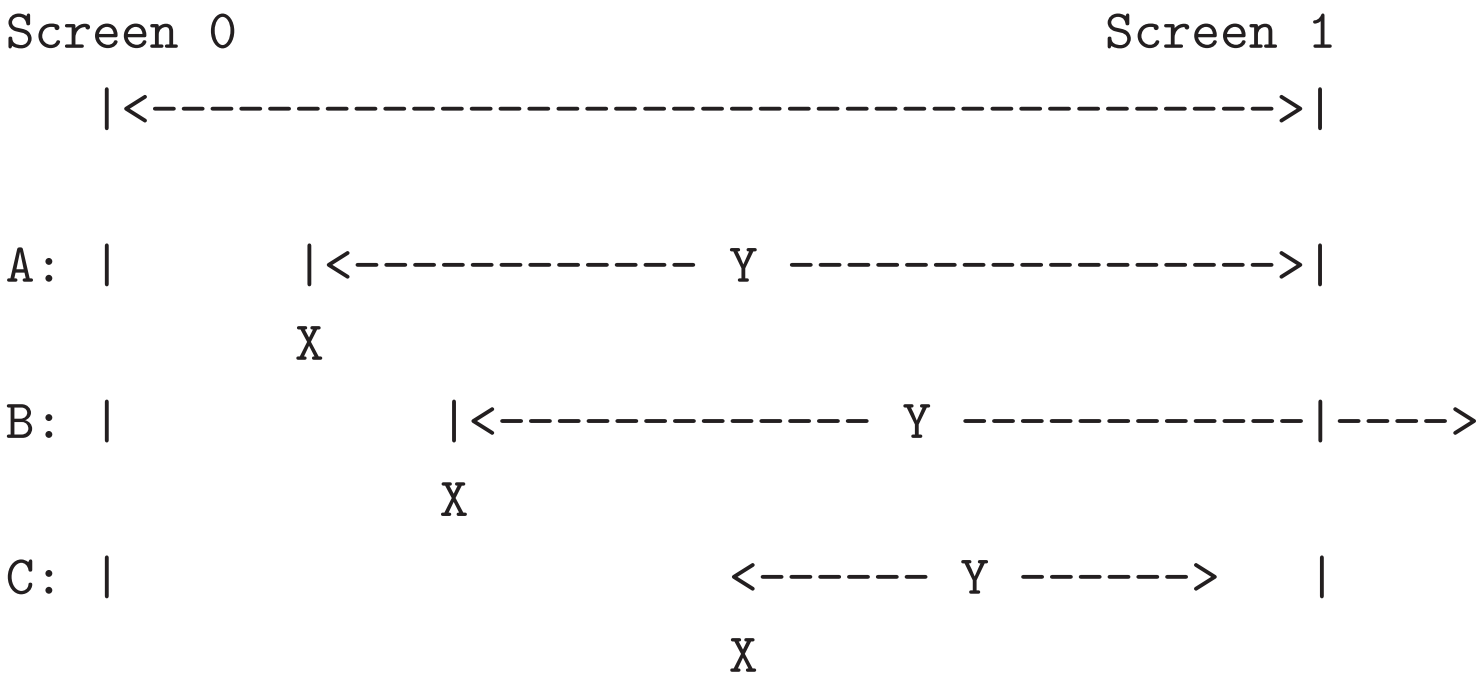
Periodic screening, general pdfs

Target: $E(Z^*)/E(Z) = E(g(Y) \cdot Z)/\mu_z$

Periodic screening case, screening interval δ

X = time at which preclinical duration begins $\sim$ Unif(0,δ)

Y = duration of preclinical disease



Cases A and B are screen-detectable

Case C's preclinical duration is too short to be detected

Equate observed mean *survival since Dx* in study & control arms, accounting for overdiagnosed cases, *after screening ends*:

$$\bar{z}_C n_C = n_{int}\bar{z}_{int} + n_{post}\bar{z}_{post} + n_{ref}\bar{z}_{ref} + n_{SD}\bar{z}^*$$

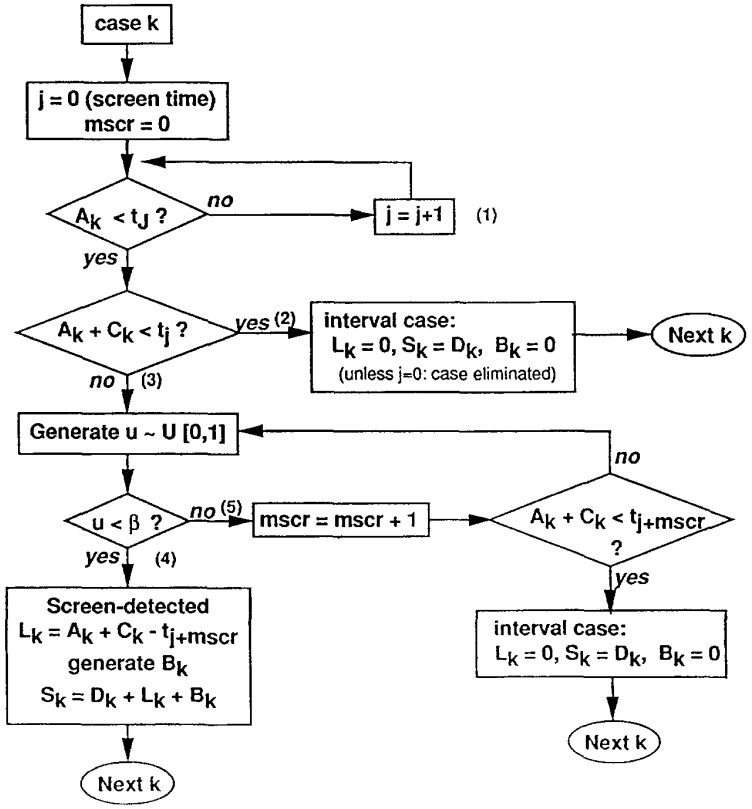
All but $\bar{z}^*$ can be estimated from observed data

$$\bar{z}_S n_S = n_{int}\bar{z}_{int} + n_{post}\bar{z}_{post} + n_{ref}\bar{z}_{ref}$$
$$+ (n_{SD} - n_{ODx}) \cdot [(\bar{z}^* + \mu_B + \mu_L) + n_{ODx}\mu_{ODx}]$$

Estimate μ_L = avg lead time (difference in avg time **to** Dx) and n_{ODx} = difference in # of diagnosed cases after screening ends
$\Rightarrow$ estimate of μ_B = Average Benefit Time

(KK+PP, in preparation)

C_k = duration of preclinical disease (exponential, mean = Z)
D_k = duration of clinical disease (exponential, mean = 5 + B_k)
L_k = lead time
B_k = benefit time
S_k = survival time (time of death - time of diagnosis)
t_j = time of j^{th} screen, j = 0, 1, 2, 3, 4, 5
β = Prob{ Screen detects case | disease }
mscr = number of missed screens



Flowchart for simulating a case history for a screened participant in a randomized trial. Numbers in parentheses refer to steps of the process described in Section 4.3.

To find $E(Y_k^*)/E(Y)$, $E(Z_k^*)/E(Z)$ (and ratio of SDs):

1. Density function of $Y_{(k)}^*$: $f_{Y_{(k)}^*}(y) = g_{(k)}(y) f_Y(y)$,
   $g_{(k)}(y) = [(y-(k-1)\delta) I_{[(k-1)\delta, j\delta]}(y) + \delta I_{[k\delta,\infty)}(y)]/D_{01k\delta}$
   where $D_{01k\delta} = \mathrm{P}\{0 \cdot \delta \le X \le 1 \cdot \delta, X+Y > k\delta\}$
   $= \delta - [J(k\delta) - J((k-1)\delta)]$

2. Sojourns for cases arising before 0, detected at screen k:
   $\mathrm{P}\{Y \le y | -L < X < 0, X+Y > k\delta\}$
   $\Rightarrow f_{(k^+)}(y) = (y-k\delta) f_Y(y) I_{(k\delta,\infty)}(y)/D_{01k^+\delta}$,
   $D_{01k^+\delta} = \mu_Y - k\delta + J(k\delta)$

3. Sojourns detected **at** screen k: wtd avg of cond'l pdfs:
   $f_{Y_k}(y) = [\sum_{j=1}^k w_j f_{(j)}(y) + w_{k^+} f_{(k^+)}(y)]/W$
   $= f_Y(y)[\sum_{j=1}^k w_j g_{(j)}(y) + w_{k^+} g_{(k^+)}(y)]/W$
   $\equiv f_Y(y) \cdot g_k(y)$
   where $W = \sum w_j$, $w_j = \beta(1-\beta)^{j-1} D_{01j\delta}$

4. Expected sojourn time for cases detected at screen k:

$$E(Y_k^*) = \int_0^\infty y f_{Y_k}(y)dy = \int_0^\infty y g_k(y) f_Y(y)$$

5. pdf of Z^*: $\int_0^\infty f_{Y^*,Z^*}(y,z)dy = \int_0^\infty g_k(y) f_{Y,Z}(y,z)dy$

6. $E(Z^*)/E(Z) = E(g_k(Y)Z)/\mu_z$,
   $g_k(y) = [\sum_{j=1}^k w_j g_{(j)}(y) + w_{k^+} g_{(k^+)}(y)]/W$
   $g_{(j)}(y) = [(y-(j-1)\delta)I_{[(j-1)\delta,j\delta]}(y) + \delta I_{[j\delta,\infty)}(y)]/D_{01j\delta}$

Explicit calculations possible when $f_{Y,Z}(\cdot) \sim$ bivariate gamma

Calculate average Z^* when $f_{Y,Z}(\cdot) \sim$ bivariate gamma:

preclinical (Y) & clinical (Z) durations are bimodal

$$f_{Y,Z} = e^{-(\lambda_1 y + \lambda_2 z)}[\phi h_a(y,z) + (1-\phi)h_b(y,z)]$$
$$h_a(y,z) = \lambda_1^{r_1} y^{r_1-1} \lambda_2^{r_2} z^{r_2-1} / [\Gamma(r_1)\Gamma(r_2)]$$
$$h_b(y,z) = \lambda_3^{r_3} y^{r_3-1} \lambda_4^{r_4} z^{r_4-1} / [\Gamma(r_3)\Gamma(r_4)]$$

With probability ϕ: $Y \sim Gamma(r_1, \lambda_1)$, $Z \sim Gamma(r_2, \lambda_2)$

With probability $1-\phi$: $Y \sim Gamma(r_3, \lambda_3)$, $Z \sim Gamma(r_4, \lambda_4)$

Different combinations of parameters allow modeling of different disease scenarios

Different parameter choices ⇒ Six scenarios:

A: Combination of Short & Long durations ($\rho = 0.57, 0.66, 0.67$)
Preclinical: means 1.5, 5; Clinical: means 2, 6; 11-28%

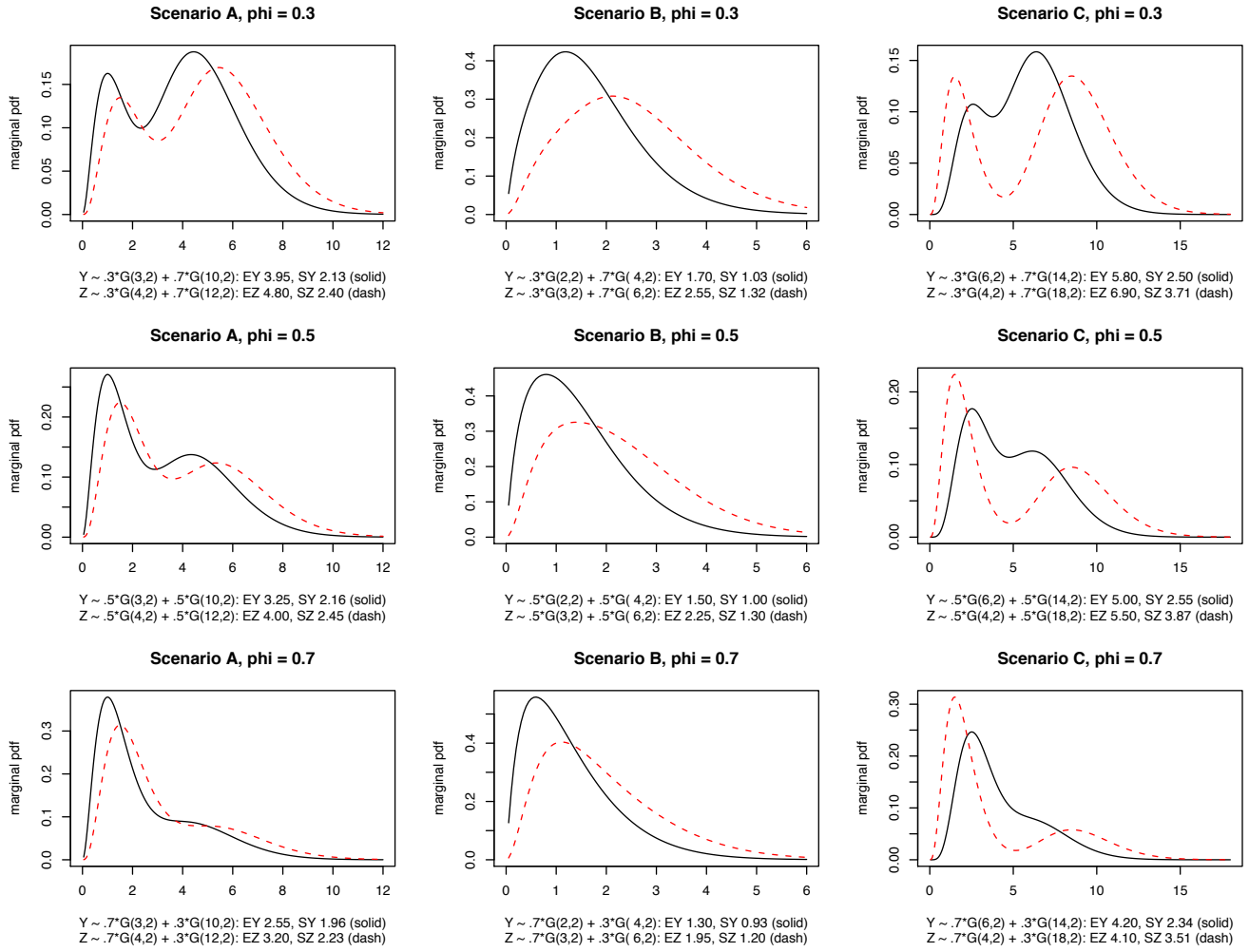
B: Combination of Short & Moderate ($\rho = 0.23, 0.28, 0.29$)
Preclinical: means 1, 2; Clinical: means 1.5, 3; 28-55%

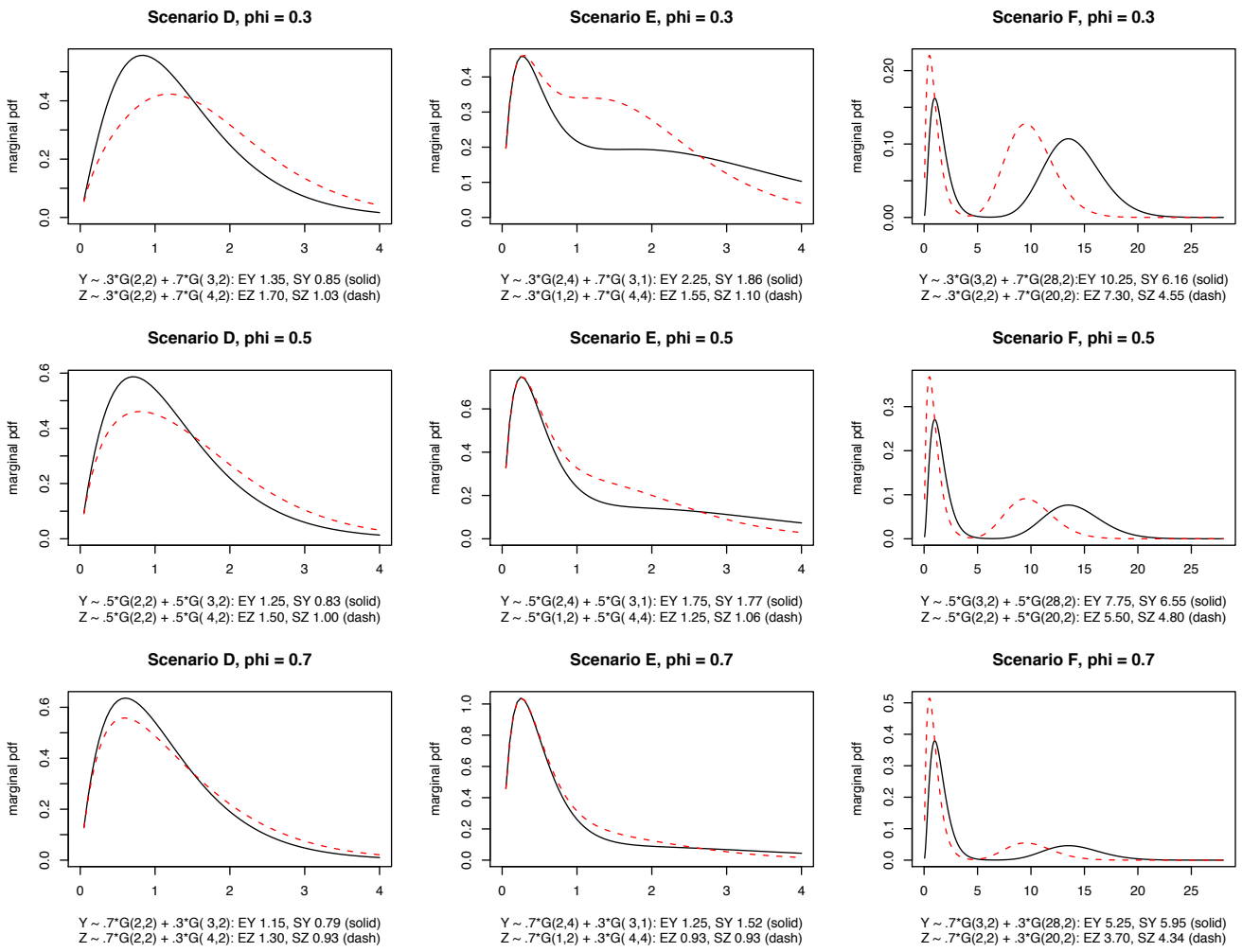
C: Combination of Moderate & Long ($\rho = 0.63, 0.71, 0.72$)
Preclinical: means 3, 7; Clinical: means 2, 9; 6-14%

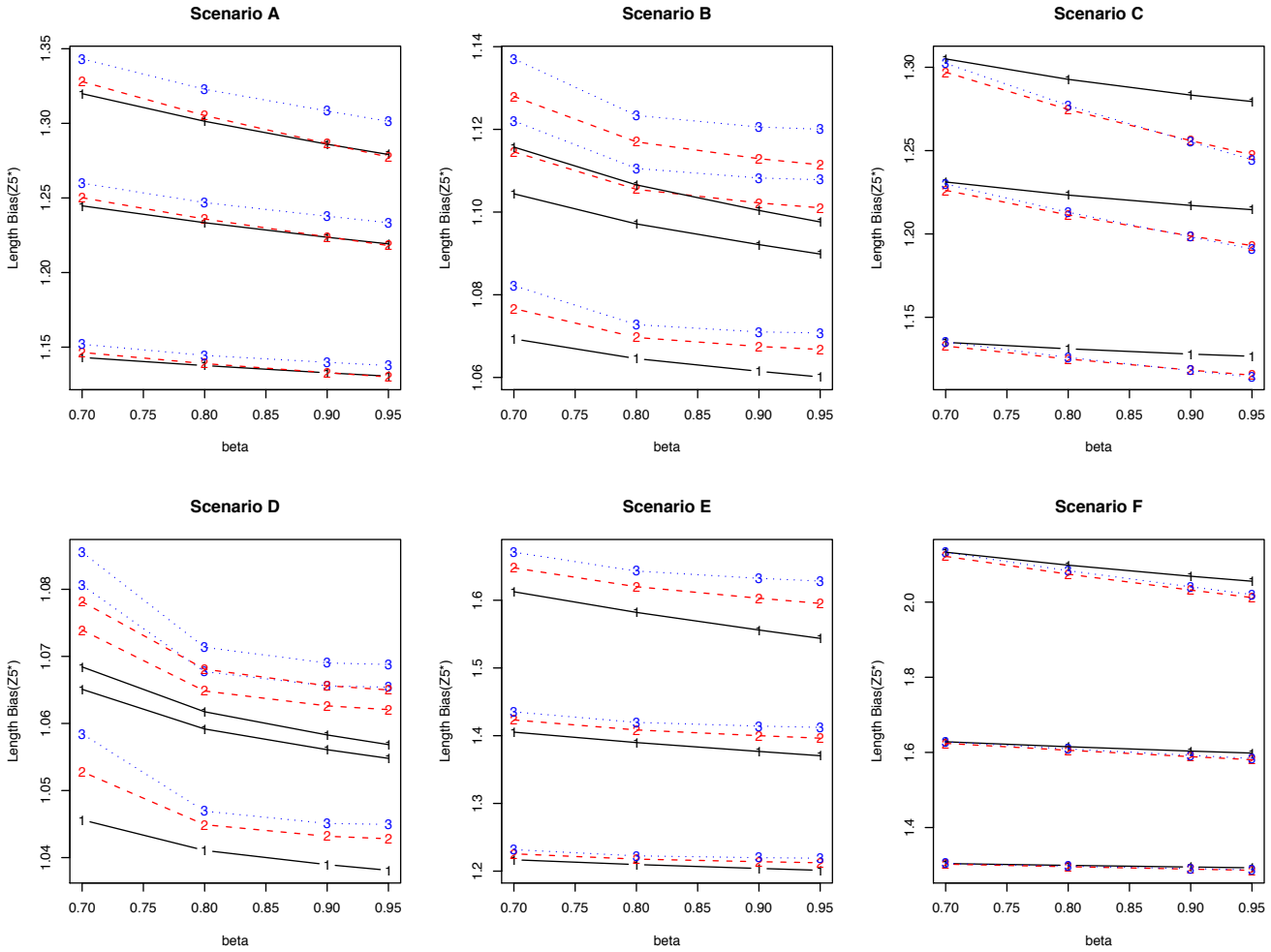
D: Combination of Short & Speedy ($\rho = 0.12, 0.14, 0.15$)
Preclinical: means 1, 1.5; Clinical: means 1, 2; 29-48%

E: Combination of Speedy & Moderate ($\rho = 0.39, 0.50, 0.56$)
Preclinical: means 0.5, 3; Clinical: means 0.5, 2; 14-60%

F: Combination of Speedy & Very long ($\rho = 0.77, 0.84, 0.86$)
Preclinical: means 1.5, 14; Clinical: means 1, 10; 30-120%


Scenario A, phi = 0.3
marginal pdf
Y ~ .3*G(3,2) + .7*G(10,2): EY 3.95, SY 2.13 (solid)
Z ~ .3*G(4,2) + .7*G(12,2): EZ 4.80, SZ 2.40 (dash)
Scenario B, phi = 0.3
marginal pdf
Y ~ .3*G(2,2) + .7*G(4,2): EY 1.70, SY 1.03 (solid)
Z ~ .3*G(3,2) + .7*G(6,2): EZ 2.55, SZ 1.32 (dash)
Scenario C, phi = 0.3
marginal pdf
Y ~ .3*G(6,2) + .7*G(14,2): EY 5.80, SY 2.50 (solid)
Z ~ .3*G(4,2) + .7*G(18,2): EZ 6.90, SZ 3.71 (dash)
Scenario A, phi = 0.5
marginal pdf
Y ~ .5*G(3,2) + .5*G(10,2): EY 3.25, SY 2.16 (solid)
Z ~ .5*G(4,2) + .5*G(12,2): EZ 4.00, SZ 2.45 (dash)
Scenario B, phi = 0.5
marginal pdf
Y ~ .5*G(2,2) + .5*G(4,2): EY 1.50, SY 1.00 (solid)
Z ~ .5*G(3,2) + .5*G(6,2): EZ 2.25, SZ 1.30 (dash)
Scenario C, phi = 0.5
marginal pdf
Y ~ .5*G(6,2) + .5*G(14,2): EY 5.00, SY 2.55 (solid)
Z ~ .5*G(4,2) + .5*G(18,2): EZ 5.50, SZ 3.87 (dash)
Scenario A, phi = 0.7
marginal pdf
Y ~ .7*G(3,2) + .3*G(10,2): EY 2.55, SY 1.96 (solid)
Z ~ .7*G(4,2) + .3*G(12,2): EZ 3.20, SZ 2.23 (dash)
Scenario B, phi = 0.7
marginal pdf
Y ~ .7*G(2,2) + .3*G(4,2): EY 1.30, SY 0.93 (solid)
Z ~ .7*G(3,2) + .3*G(6,2): EZ 1.95, SZ 1.20 (dash)
Scenario C, phi = 0.7
marginal pdf
Y ~ .7*G(6,2) + .3*G(14,2): EY 4.20, SY 2.34 (solid)
Z ~ .7*G(4,2) + .3*G(18,2): EZ 4.10, SZ 3.51 (dash)


Scenario D, phi = 0.3
marginal pdf
Y ~ .3*G(2,2) + .7*G(3,2): EY 1.35, SY 0.85 (solid)
Z ~ .3*G(2,2) + .7*G(4,2): EZ 1.70, SZ 1.03 (dash)
Scenario E, phi = 0.3
marginal pdf
Y ~ .3*G(2,4) + .7*G(3,1): EY 2.25, SY 1.86 (solid)
Z ~ .3*G(1,2) + .7*G(4,4): EZ 1.55, SZ 1.10 (dash)
Scenario F, phi = 0.3
marginal pdf
Y ~ .3*G(3,2) + .7*G(28,2):EY 10.25, SY 6.16 (solid)
Z ~ .3*G(2,2) + .7*G(20,2): EZ 7.30, SZ 4.55 (dash)
Scenario D, phi = 0.5
marginal pdf
Y ~ .5*G(2,2) + .5*G(3,2): EY 1.25, SY 0.83 (solid)
Z ~ .5*G(2,2) + .5*G(4,2): EZ 1.50, SZ 1.00 (dash)
Scenario E, phi = 0.5
marginal pdf
Y ~ .5*G(2,4) + .5*G(3,1): EY 1.75, SY 1.77 (solid)
Z ~ .5*G(1,2) + .5*G(4,4): EZ 1.25, SZ 1.06 (dash)
Scenario F, phi = 0.5
marginal pdf
Y ~ .5*G(3,2) + .5*G(28,2): EY 7.75, SY 6.55 (solid)
Z ~ .5*G(2,2) + .5*G(20,2): EZ 5.50, SZ 4.80 (dash)
Scenario D, phi = 0.7
marginal pdf
Y ~ .7*G(2,2) + .3*G(3,2): EY 1.15, SY 0.79 (solid)
Z ~ .7*G(2,2) + .3*G(4,2): EZ 1.30, SZ 0.93 (dash)
Scenario E, phi = 0.7
marginal pdf
Y ~ .7*G(2,4) + .3*G(3,1): EY 1.25, SY 1.52 (solid)
Z ~ .7*G(1,2) + .3*G(4,4): EZ 0.93, SZ 0.93 (dash)
Scenario F, phi = 0.7
marginal pdf
Y ~ .7*G(3,2) + .3*G(28,2): EY 5.25, SY 5.95 (solid)
Z ~ .7*G(2,2) + .3*G(20,2): EZ 3.70, SZ 4.34 (dash)


Scenario A
Scenario B
Scenario C
Scenario D
Scenario E
Scenario F
Length Bias(Z5*)
beta
0.70
0.75
0.80
0.85
0.90
0.95

Real data?

- LBS effect depends primarily on μ_Y, σ_Y, μ_Z, σ_Z, $\rho_{Y,Z}$
- Estimate μ_Z, σ_Z from control arm cases
- How to estimate μ_Y, σ_Y, $\rho_{Y,Z}$?
- Recall: We can estimate $\bar{L}$ = ave lead time
- *Exponential* sojourn time: $\mu_L = \mu_Y / prevalence$
- Would a function $g(\rho_{\bar{L},\bar{Z}}) \Rightarrow \rho_{Y,Z}$?
- Would a function $h(\bar{L}, \bar{Z}) \Rightarrow \mu_Y$?

3. Example: Prostate Cancer Screening

NCI Prostate, Lung, Colorectal, Ovarian Cancer Screening

(Design: *Prorok et al Control Clin Trials 2000*)

(Results: *Andriole et al NEJM 2009*)

- Subjects: $n_S = 38,343$, $n_C = 38,350$, 10 centers, 1993–2001
- Screening: Annual PSA testing (6 yrs), DRE (4 yrs)
- Age at initial screen: 55–64 (63.6%); 65-74 (36.4%)
- 85% Caucasian, 4.5% African-American, 10.5% other

Non-compliance:

- Study (data): 15% had no PSA, 14% had no DRE
- Control (estimate): 40-52% received some screening in first 5 yrs
  [estimated via weighted average of 1% random surveys
  and 3758 (9.8%) w/repeated screens before trial]
- Potential screening-related risks (usually rare):
  - DRE
  - PSA
  - Diagnostic procedures (infection, bleeding, clots, ...)
  - Treatment consequences (e.g. infection)

PLCO Results per 10,000 person-years (Year 7):

	Screen	Control	Ratio (95%CI)
Person-yrs	254,295	253,317	
# PrCa cases	3297	2790	
Incidence	116	95	1.22 (1.16, 1.29)
# PrCa deaths	50	44	
Mortality Rate	**2.0**	**1.7**	**1.13 (0.75, 1.70)**
Other deaths	3953	4058	0.97 (0.93, 1.01)

No differences in treatment by stage

Prostate Cancer Screening via PSA, DRE shows no reduction in mortality (hence no extended survival)

Colorectal Screening (flex sig) **showed ∼26% reduction in mortality: Extended benefit time TBD**

Canadian National Breast Cancer Screening Study:
13-Yr results of randomized trial, women 50–59

$N = 39{,}405$, randomized Jan 1980 – Mar 1985;
active followup to Jun 30, 1996

19,711 Study: Physical Exam + BSE + Mammography
19,694 Control: Physical Exam + BSE

$T = 4$ or 5 (first 62%) annual screens
Compliance: 100% ($T = 1$), 90% ($T = 2$); 86% ($T = 5$)

Detection: In-situ: 71 (Study), 16 (Control)
Invasive: Screen-detected + Interval + Incident

CNBSS Detection, invasive cancers

(Screen-detected + Interval + Incident)

Year	Study	Control
Year 1	118 + 114 + 0	64 + 16 + 0
Years 2–5	149 + 36 + 32	84 + 72 + 47
Years 6–9	0 + 0 + 175	0 + 0 + 217
Total	276 + 50 + 507	148 + 88 + 264
	= 524	= 500

CNBSS Mortality:

Type	Study	Control
Breast cancer	88	90
Other cancer	376	313
Other non-cancer	270	287

A.B. Miller et al. JNCI 20 Sep 2000: 1490–1499

Reduction in Mortality, cases diagnosed through:

Year	Study	Control	Ratio	95% CI
Year 6	84	76	1.10	(0.81, 1.51)
Year 7	93	83	1.12	(0.83, 1.50)
Year 8	99	89	1.10	(0.84, 1.48)
Year 9	104	97	1.07	(0.81, 1.41)
All Years	107	105	1.02	(0.78, 1.33)

Benefit of mammography over CBE/BSE may be slight for women 50–59

HIP: (Mammography + CBE) vs (Usual medical care)

Screening benefit may be due to breast self exam + physical exam

4. Summary

- Screening must be evaluated using well-designed studies
- Metric can be *reduction in mortality* or *extended lifetime*
- Potential for benefit if treatment is effective and extends life (with high quality)
- Potential for harm if ineffective treatment (shortened lifespan)
- Analogue for environmental monitoring (e.g., leakage in radioactive waste containers or oil drilling equipment): Periodic inspection requires **reliable indicators** with **high specificity** and **long sojourn times**

References

Andriole et al: Mortality results from a randomized prostate cancer screening trial, *NEJM* 26 Mar 2009

Cox DR, Lewis PAW. *The Statistical Analysis of Discrete Time Events*. Oxford Press: London, 1972.

Croswell et al: Cumulative incidence of false positive test results in lung cancer screening, *Ann Int Med* 20 Apr 2010

Kafadar K; Prorok P: Alternative definitions of comparable case groups and estimates of lead time and benefit time in randomized cancer screening trials, *Stat Med* 22:83–111, 2003.

Kafadar K; Prorok P: Effect of length-biased sampled sojourn times on the survival distribution in randomized cancer screening trials,, *Stat Med* 28:2116-2146, 2009.

Marcus et al: Extending cancer incidence follow-up in the Mayo lung project and overdiagnosis, *JNCI* 7 Jun 2006

Miller AB, To T, Baines CJ, Wall C: Canadian National Breast Screening Study-2: 13-year results of a randomized trial in women aged 50–59 years, *JNCI* 2000: 1490–1499.

Prorok et al: Design of the PLCO Cancer Screening Trial, *Control Clin Trials* 21:273S-309S, 2000.

Scheaffer RL: Size-biased sampling, *Technometrics* **14**:635–644, 1972.

Schotz WE, Zelen M: Effect of length sampling bias on labeled mitotic index waves, *J Theoretical Bio* **32**, 383–404, 1971.

Schröder et al: Screening and prostate cancer mortality in a randomized European study, *NEJM* 26 Mar 2009

Wolters et al: The effect of study arm on prostate cancer treatment in the large screening trial ERSPC, *Int J Cancer* 126:3287–2393, 2010.