

# **Computational Methods in Medical Decision Making: To Screen or Not To Screen?**

Karen Kafadar

Department of Mathematics  
University of Colorado-Denver

Philip C. Prorok

Chief, Biometry Research Group  
National Cancer Institute

1. Introduction
  - Context: Screening for disease
  - Examples of screening trials
  - Relation to vaccination programs
2. Evaluation of Screening Programs
  - Factors
  - Potential biases
  - Computational methods
3. Disease Progression Model
  - lead time, benefit time, length bias
4. Discrete event simulation
5. Conclusions and further work

## 1. Introduction: Screening for disease

Widespread testing of apparently healthy individuals  
(no clinically apparent symptoms)

Encounter is initiated by those who administer test,  
not by patient

*Purpose:*

Separate the population into two groups:  
high / low probability of given disorder

*Implicit assumption:*

**Benefit to early diagnosis**

(better prognosis, safer treatment, ...)

*Conditions calling for screening:*

- Serious public health condition
- Well-defined target population
- Recognizable asymptomatic phase
- Reliable screening test
- Treatment in pre-symptomatic stage has benefit  
(e.g. reduction in mortality, longer survival)

*Vaccination programs:*

Similar issues -- except focus is on disease  
*prevention, not detection*

*Potential advantages of screening:*

- Improved prognosis
- Less radical treatment
- Reassurance to those with negative results
- Potential health cost savings

*Potential disadvantages:*

- Cost of screening test
- Longer morbidity
- False negative results
- Increased costs in false positives

	Test result	
	Positive	Negative
Disease Present	True Positive Potential Benefit	False Negative Potential Harm
Disease Absent	False Positive Increased Costs	True Negative Reassurance

*Does potential benefit outweigh potential harm?*

## 2. Evaluation of screening programs

*Depends upon many factors:*

- Sensitivity:  $\text{Prob}\{\text{True} + | \text{Disease} +\}$   
For diagnostic purposes, want *high sensitivity*
- Specificity:  $\text{Prob}\{\text{True} - | \text{Disease} -\}$   
For screening purposes, want *high specificity*  
(low specificity  $\rightarrow$  false positives  $\rightarrow$  high costs)
- Age of subjects being screened (affects sens/spec)
- Prevalence of disease
- Potential for bias

*Potential biases:*

- Self-selection bias (BCDDP)
- Case group bias
- Lead time bias
- Length bias

*How to evaluate benefits of screening in face of*

- varying factors
- changing population demographics  
(more/fewer persons in specific age groups)
- biases
- new screening modalities

*To screen or not to screen?*

*Some specific cancer screening studies:*

*Breast Cancer Detection Demonstration Project*  
(Georgetown University, 1970s)

- Advertise for subjects
- Compare survival rates with general population rates
- Empirical observations:
  - Those at *greater* risk for breast cancer tend to present themselves for screening
  - Those at *lower* risk for ovarian cancer tend to present themselves for screening

*Randomized cancer screening trial:*

- Study arm: offer screening at regular intervals
- Control arm: “usual medical care”
- Contamination in both arms (→ "intention to treat")

*Examples of randomized screening trials:*

- HIP = Health Insurance Plan of New York (1963-66)
- CNBSS = Canadian National Breast Screening Study (1980-85)
- PLCO = Prostate, Lung, Colorectal, Ovarian (now)

*Measures for evaluation:*

- Reduction in mortality
- Extended benefit times
- Lead time relative to benefit time
- Potential for length bias

*How to evaluate benefits of screening in face of varying factors, changing demographics, and biases?*

Require computational methods and simulations to

- develop method for comparable case groups
- assess methods of estimating average lead time and average benefit time (extended survival)
- estimate effect of length biased sampling on benefit

Trials are expensive ---

Design, analysis, evaluation depend upon computational methods and simulations

### **3. Disease Progression Model: Identifying comparable case groups Estimating lead time and benefit time**

#### *Traditional clinical trials:*

Strict protocols ensure comparability of cases  
in both arms of trial (treatment; standard)

#### *Screening trials:*

Cases *evolve* at different times as trial proceeds

At what times do cases appear in control arm,  
cases that are comparable to those in study arm?

During screening: *more* study cases than control cases  
(screening detects them sooner)

After screening ends: *fewer* study than control cases  
(study cases already detected;  
control cases continue to accrue at "steady" rate)

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

*When* to compare mortality rates and survival times?

## HIP Breast Cancer Screening Study:

(Shapiro et al. 1988, p.19)

Health Insurance Plan of NY, 1963-1969

Randomized Clinical Trial

30,565 Control: "usual health care"

30,131 Study: Mammography+Clinical Breast Exam

10,800 refused screening

20,200 screened at least once

12,000 screened all four times

Initial screen (Dec 1963 - June 1966)

plus  $T = 3$  annual screens (to June 1969)

Follow-up mail surveys at 5, 10, 15 years after entry

Mortality reduction; Cumulative incidence/mortality

Year	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
8	497	490	108	141	4.59	5.92	0.78
9	558	565	128	172	4.85	6.44	0.75

Incidence curves cross at  $\approx 6.2$  years (367 vs 364 cases)

For women under 50 years of age at start of study:

Incidence curves cross at  $\approx 11.5$  years (then over 50)

Reduction in mortality: 0.77 (0.50-1.16)

Ideally: Compare survival experiences between:

- (a) those that *could* benefit from screening  
(screen-detected cases, plus refusers)
- (b) those that *could not* benefit from screening  
("counterparts" from control arm)

"Counterparts" from control arm:

When diagnosed? Within  $C = ???$  years?

Which ones would have "refused if offered"?

Some possible solutions:

1. Compare survival experiences of:
  - only cases diagnosed in study arm through  $T$  years
  - all cases diagnosed in control arm through followup

Problem: aging population

HIP:  $T = 3$  years;  $F = 15$  years

PLCO:  $T = 5$  years;  $F = 25$  years

2. Use all cases in both arms up to year  $C$   
(keeps both populations as much alike as possible)

What is  $C$ ? ( $T < C < F$ )

$C$  too small: missing control arm counterparts

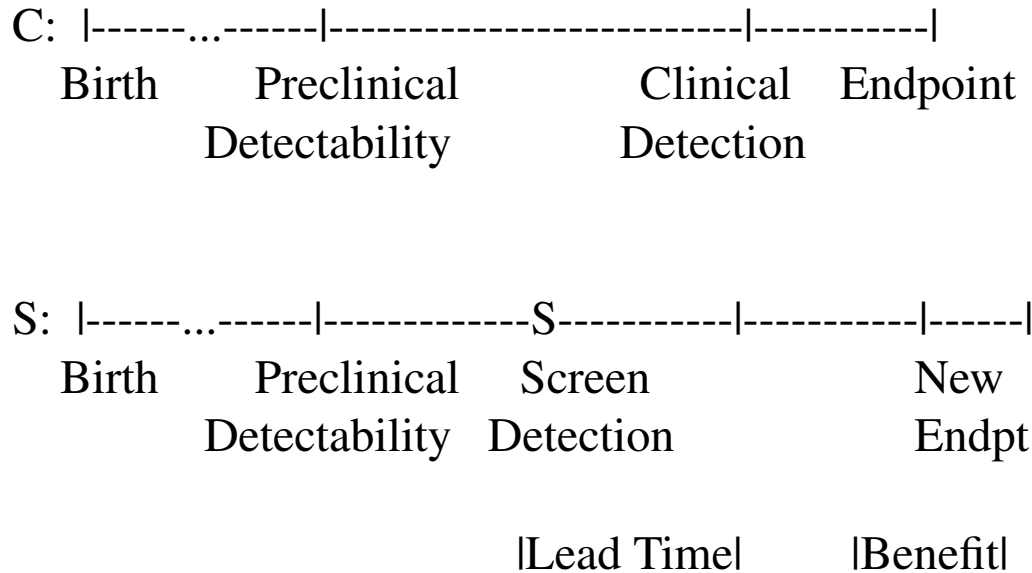
$C$  too large: diluted effect of screening

$C$  about right: same # of cases diagnosed in 2 arms

Good estimates of average lead/benefit time require good method for obtaining comparable case groups

**Lead time and Benefit time:**

Even with a method for obtaining comparable case groups, *lead time* affects survival comparison (control, study):



*Solution to unbiased (by lead time) comparison:*

*Randomized screening trial*

*Compare survival times since start of study*

*(survival since time of diagnosis confounds lead time and benefit time)*

Several proposed estimates of average lead time:

Zelen and Feinleib 1964

Morrison 1972

Kafadar and Prorok 1996 (also benefit time)

Which estimate performs best?

-- Computer simulated randomized screening trial

*Computer simulation parameters:*

$N = 20,000$  participants in each arm

$\lambda = 20$  incident cases/year ( $\approx$  breast cancer),  
Poisson case arrival process

$T = 5$  years of screening (initial + 5 annual)

$M = 20$  years of case accrual (follow-up)

$\beta = 0.20 =$  false negative probability

$(X, Y) =$  (preclinical, clinical) durations

Bivariate gamma, correlation = 0.30

$(\mu_x, \sigma_x^2) = (2, 1), \quad (\mu_y, \sigma_y^2) = (4, 4)$

$b_i =$  individual benefit time for screen-detected case

= virtually none if preclin duration  $X < 1$  or  $X > 5$

= maximum benefit about 6 years ( $X \approx 3$ )

$= 4k(X - a)(b - X) / (b - a)^2$

$k \sim N(6, 0.25)$  (average max benefit  $\approx 6$  years)

$a \sim N(1, 0.25)$  (average min PD about 1 year)

$b \sim N(5, 1.00)$  (average max PD about 5 years)

*Accuracy of estimators of average lead time L and average benefit time B, using different proposals for defining ‘comparable case groups’:*

Aron and Prorok 1974  
Connor and Prorok 1994  
Etzioni and Self 1995  
KK and PP 2003

Evaluation: simulate the screening process

500 simulated randomized screening trials, 4 scenarios:

Scenario	Sojourn time		Clinical Duration		Correlation
	Mean	Var	Mean	Var	
(a)	2	1	2	1	0.9
(b)	2	1	4	4	0.3
(c)	4	4	2	1	0.3
(d)	4	4	4	4	0.9

Kafadar and Prorok (*Stat Med* 2003):

$$C_{\mu} = T + 2\hat{\mu}$$

= final year of screening + 2(mean sojourn time)

Importance of comparable case groups

*Estimating average lead/benefit time for HIP trial:*

$$\hat{\mu}_0 = \text{mean sojourn time} = 1.3 \text{ years}$$

$$C_\mu = 3 + 1.3 + \sqrt{1.3} = 6.0 \text{ years}$$

By year 6 following start of study:

367 Study cases

364 Control cases

132 cases found by screening

	Estimator	Program Time	Screened Time
Benefit Time	$\hat{B}_1$	1.18 yr	3.28 yr
	$\hat{B}_2$	1.14 yr	3.19 yr
	(SE)	0.46 yr	1.29 yr
Lead Time	$\hat{L}_1$	3.04 mo	8.45 mo
	$\hat{L}_2$	3.30 mo	9.18 mo
	(SE)	1.62 mo	4.50 mo

Screened time (screen-detected cases only)

$$= (367/132) \times \text{Program time (all cases)}$$

*Aside: Dilution of benefit effect*

Suppose that, by year  $C$ , both arms have (roughly) same number of cases ( $n$ ).

Study arm survival times:  $X_1, \dots, X_n$

$n_1$  cases could benefit from screening: mean  $\mu_1 + B$

$n_2$  cases could not benefit from screening

(diagnosed too late, interval cases, refusers): mean  $\mu_2$

Control arm survival times:  $Y_1, \dots, Y_n$ , mean  $\mu_0$

*Randomization* ensures, under  $H_0$ ,

$$\begin{aligned}\mu_0 &= (n_1\mu_1 + n_2\mu_2) / (n_1 + n_2) \\ \rightarrow \mu_2 &= \mu_0 + (n_1 / n_2) (\mu_0 - \mu_1)\end{aligned}$$

A crude estimate of average benefit time is  $\hat{B} = \bar{X} - \bar{Y}$ :

$$\begin{aligned}E(\bar{X} - \bar{Y}) &= [n_1(\mu_1 + B) + n_2\mu_2] / (n_1 + n_2) - \mu_0 \\ &= [n_1 / (n_1 + n_2)] \cdot B \equiv [n_1 / (n_1 + n_2)] \cdot B \equiv \alpha B\end{aligned}$$

Intensity of 'signal'  $\downarrow$  as  $n_2 \uparrow$  ('dilution effect')

*Empirical estimate of B for HIP trial:*

Using all cases in both arms diagnosed up to year C:

$$\hat{B}(C) = \bar{X}(C) - \bar{Y}(C)$$

$$= \alpha B \equiv n_1 B / (n_1 + n_2)$$

$$\log \hat{B}(C) + \log(\# \text{ study cases to year } C) = \log n_1 + \log B$$

(should be constant eventually)

$n_1$ : How many people could benefit from screening?

- cases in study arm, *not* refusers
- cases diagnosed before screening ends
- cases detected by screening

So  $n_1 = 132$  cases (detected by screening)

1906 cases diagnosed to year 15 after start of trial

35 cases: time of Dx = exit (4 screened, 12 refusers, 19 control)

2 cases: time of Dx later than time of exit

Result: 1869 cases, of which 132 were detected by screening.

$$\log \hat{B}(C) + \log(\# \text{ study cases to year } C) = 11.95$$

so

$$\log B = 11.95 - \log(132) \rightarrow B = 1173 \text{ days} = 3.2 \text{ years}$$

**Length (size) biased sampling:**

measurement process favors those experimental units whose lengths (sizes) are proportional to the probability of selection

Sample particles from mixture (heavier ones more likely)

Survey hospital patients (longer stays => greater probability)

Select units in database (those in system longer more likely to be sampled)

Longer sojourn times => more likely to cross screening point

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 be asymptomatic longer, i.e. have slower-growing disease."*

HIP study: Proportion of cases having negative nodes:

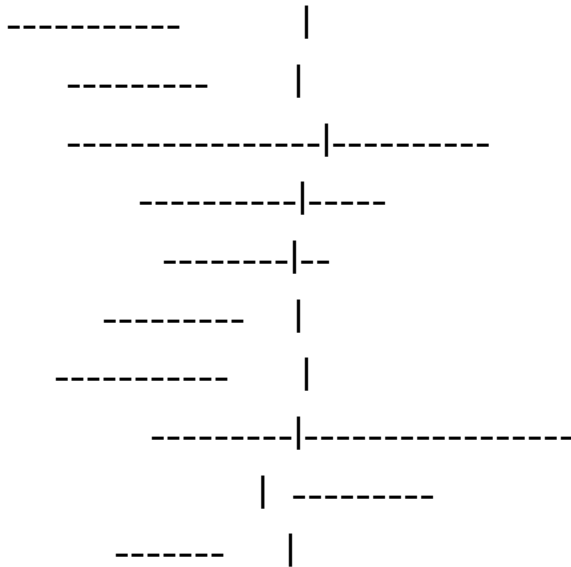
132 Study Cases detected by screening:	63%
91 Study interval + 73 Study refusers cases:	47%
284 Control cases:	46%

Bias that results from length biased sampling  
*is not eliminated by randomized design*

How much of the survival time (since diagnosis) is due to *length biased sampling*, after having accounted for lead time and benefit time?

One-shot screening program: Single screen

Cox (1969), Cox and Lewis (1972):



Preclinical durations (PD)  $Y_1, \dots, Y_n$

$f_Y(\cdot)$  = pdf of PDs

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

$f_{Y^*}(y)$  = pdf of sampled PDs

$$= \lim_{n \rightarrow \infty} (\text{Proportion of } \sum_{i=1}^n Y_i \text{ due to intervals of length } y)$$

$$= \lim_{n \rightarrow \infty} (y \cdot n_y \cdot dy / \sum Y_i)$$

$$= \lim_{n \rightarrow \infty} (n_y/n \cdot y \cdot dy) / (\sum Y_i / n)$$

$$= y \cdot f_Y(y) / \mu_y$$

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

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

$$f_{Y^*}(y) = y^\alpha \cdot f_Y(y) / \mu_\alpha$$

$$\mu_\alpha \equiv \int_0^\infty x^\alpha f_Y(x) dx$$

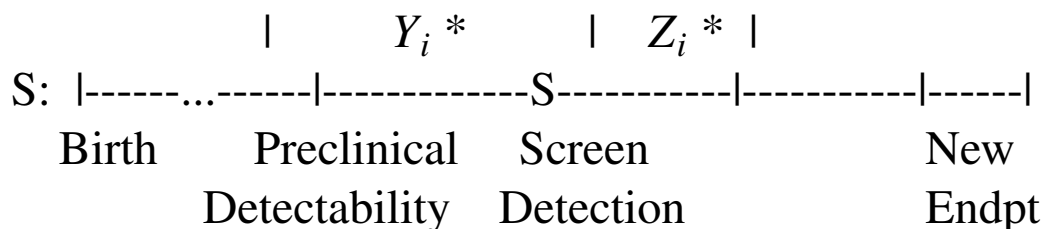
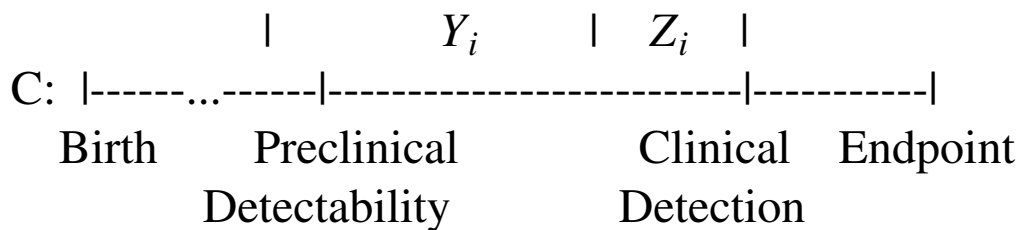
Goal in most of these problems: estimate  $E(Y^*)$ .

*Our real interest: not Y, but Z = clinical durations*

Length biased sampling occurs on a variable that affects the main variable of interest

$Z_1, \dots, Z_n =$  clinical durations correspond to  $Y_1, \dots, Y_n$ , joint pdf  $f_{YZ}(\cdot, \cdot)$ .

We want  $E(Z^*) / E(Z)$ : How much does the length biased sampling of  $Y$  affect  $E(Z)$ ?



Most studies of length (size) - biased sampling assume that the biased sampling variable is observed (e.g., Cnaan 1985; Nair and Wang 1989)

Schotz and Zelen (*J. Theor. Biol.* 1971): *Effect of Length Sampling Bias on labeled Mitotic Index Waves*

Cell proliferation: Four phases

$G_1$ : pre-DNA synthesis (RNA, proteins)

$S$ : DNA synthesis (duplicate DNA)

$G_2$ : post-DNA synthesis

$M$ : mitosis

Only  $M$  phase is observable, and only if cell has been "labeled" ( $H^3$  or  $C^{14}$ ) during  $S$  phase -- i.e., length biased sampling

Derives moments of phase durations of labeled cells

Examine effects of bias in

observed labeling index (proportion of labeled cells)

mitotic index (proportion of mitotic cells)

observed labeled mitotic index

(= proportion of mitotic cells that are labeled)

Conditions:

- Steady-state population
- Joint pdf = mixture of gamma densities
- Single labeling time

**Screening context:**

- Preclinical duration phase *not observed*
- Periodic "labeling" (screening opportunities)
- General density

**Model:**

$X$  = start time of preclinical phase (unobserved)  
Assume uniformly distributed over interval

$Y$  = duration of preclinical phase

$Z$  = duration of clinical phase

$f_{YZ}(\cdot, \cdot)$  = joint pdf of  $Y, Z$

$\mu_y, \sigma_y^2, CV_y$  = mean, variance,  $CV$  of  $Y$

$\mu_z, \sigma_z^2, CV_z$  = mean, variance,  $CV$  of  $Z$

$Y^*$  = length-biased sampled  $Y$

$Z^*$  = Clinical duration corresponding to  $Y^*$

**Target:**  $E(Z^*) / E(Z)$

## Single screen

$$\begin{aligned} f_{Y^*,Z^*}(y, z) &= f_{Z^*|Y^*}(z|y) f_{Y^*}(y) \\ &= f_{Z|Y}(z|y) y f_Y(y) / \mu_y \quad (*) \\ &= y \cdot f_{Y,Z}(y, z) / \mu_y \end{aligned}$$

$$\begin{aligned} E(Y^*) / E(Y) &= \int_0^\infty \int_0^\infty y^2 f_{Y^*,Z^*}(y, z) dy dz / \mu_y \\ &= (1 + CV_y^2) = (1 + \sigma_y^2 / \mu_y^2) \end{aligned}$$

$$\begin{aligned} E(Z^*) / E(Z) &= E(YZ) / (\mu_y \mu_z) = (\rho \sigma_y \sigma_z + \mu_y \mu_z) / \mu_y \mu_z \\ &= (1 + \rho CV_y CV_z) \quad (**) \end{aligned}$$

Example:  $CV_y = CV_z = 1, \rho = 0.5 \Rightarrow 50\%$  increase

Good news:

We can assess the effect of LBS at prevalence screen

Bad news:

We have no data for estimating either  $\rho$  or  $CV_y$

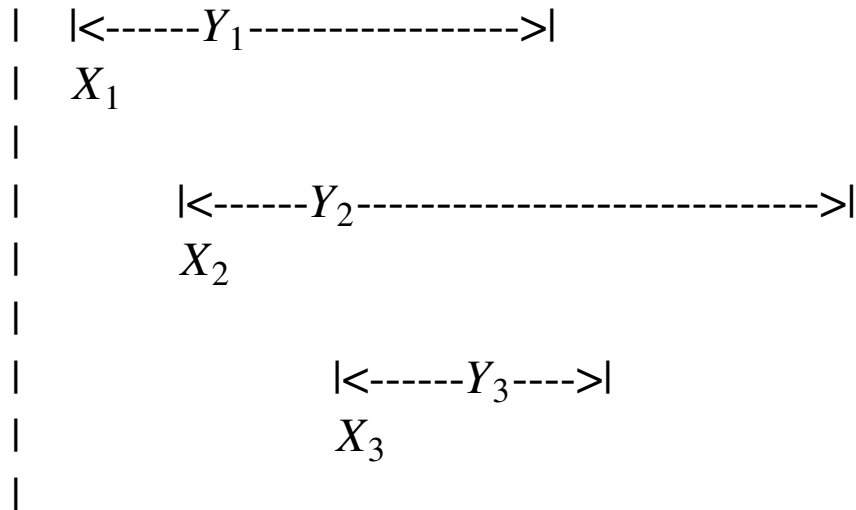
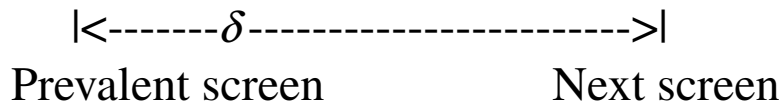
(\*) Note:

$$f_{Y^*}(y) = g(y) \cdot f_Y(y), \text{ where } g(y) = y / \mu_y$$

$$E(Y^*) = \int_0^\infty \int_0^\infty yg(y) f_{Y,Z}(y, z) dy dz = \int_0^\infty yg(y) f_Y(y) dy$$

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

Periodic screening case:  $X \sim \text{Unif}(0, \delta)$



Case that is eligible to be screen-detected at first screen following initial (prevalence) screen satisfies:

$$0 < X \leq \delta \text{ and } X + Y > \delta$$

(starts before  $\delta$ , sojourn time lasts at least until  $\delta$ )

Otherwise, case starts too late, or is detected clinically

$Y_1$  = sojourn time among cases detected at screen 1:

$$F_{Y_1}(y) = P \{ Y \leq y \mid X + Y > 1 \cdot \delta \}$$

$$= [ y \cdot F_Y(y) - I(y) ] / A_\delta \quad y \leq \delta$$

$$= [ \delta \cdot F_Y(y) - I(y) ] / A_\delta \quad y > \delta$$

$$I(y) \equiv \int_0^y F(u) du \quad A_\delta \equiv \delta - I(\delta)$$

$$f_{Y_1}(y) = \min(y, \delta) \cdot f_Y(y) / A_\delta$$

$\mathbf{Y}_k$  = sojourn time among cases that **start** in  $(0, \delta]$  but are detected **at**  $k^{th}$  screen following initial screen

$\beta$  = Test sensitivity = Probability that actual disease is detected at the screen (may depend on  $k$ ; here we treat as a constant, 0.90 or 0.95)

$$E(\mathbf{Y}_1) = A_\delta^{-1} \cdot \left[ \int_0^\delta y^2 f_Y(y) dy + \delta \int_\delta^\infty y f_Y(y) dy \right]$$

To find  $E(Z^*) / E(Z)$ :

1. Find the density function of  $\mathbf{Y}_k$ .

$$f_{\mathbf{Y}_k}(y) = g_k(y) f_Y(y)$$

2. Find  $E(\mathbf{Y}_k)$  for  $k = 1, 2, 3, 4$ .

Ex:  $\mathbf{Y}_2$ =sojourn time of case not detected at screen 1, long enough to be screen-detected at screen 2

3. Expected sojourn time for those cases that are actually screen-detected:

$$E(Y^*) = \sum_{k=1}^4 \beta(1 - \beta)^{k-1} E(\mathbf{Y}_k)$$

4.  $E(Z^*) / E(Z) = \sum_{k=1}^4 \beta (1 - \beta)^{(k-1)} E(g_k(Y) Z) / \mu_z$

Find density of  $\mathbf{Y}_k$ , when sojourn time distribution is

- Exponential = Gamma(1,1): median = 0.69
- Gamma(2,1): median = 1.68
- Gamma(4,1): median = 3.67

Step (2): Distribution of preclinical durations  
for  $k - 1$  false negative screens:

$T$  = start of screening program

$\delta$  = interval between screens

$$\begin{aligned}
 F_{Y_k}(y) &= P\{ Y \leq y \mid X + Y > k\delta, 0 \leq X \leq \delta \} \\
 &= \frac{[y - (k - 1)\delta]F_Y(y) - [J(y) - J((k - 1)\delta)]}{\delta - [J(k\delta) - J((k - 1)\delta)]} \quad y \leq k\delta \\
 &= \frac{\delta F_Y(y) - [J(k\delta) - J((k - 1)\delta)]}{\delta - [J(k\delta) - J((k - 1)\delta)]} \quad y > k\delta
 \end{aligned}$$

$$\begin{aligned}
 F_{Y_k}(y) &= P\{ Y \leq y \mid X + Y > k\delta, i\delta \leq X \leq j\delta \} \\
 &= \frac{[y - (k - j)\delta]F_Y(y) - [J(y) - J((k - j)\delta)]}{(j - i)\delta - [J((k - i)\delta) - J((k - j)\delta)]} \quad (k - j)\delta < y \leq (k - i)\delta \\
 &= \frac{(j - i)\delta F_Y(y) - [J(y) - J((k - j)\delta)]}{(j - i)\delta - [J((k - i)\delta) - J((k - j)\delta)]} \quad y > (k - i)\delta
 \end{aligned}$$

$$f_{Y_k}(y) = (d/dy)F_{Y_k}(y)$$

$$\begin{aligned}
 &= \frac{[y - (k - j)\delta]f_Y(y)}{(j - i)\delta - [J((k - i)\delta) - J((k - j)\delta)]} \quad (k - j)\delta < y \leq (k - i)\delta \\
 &= \frac{(j - i)\delta f_Y(y) - [J(y) - J((k - j)\delta)]}{(j - i)\delta - [J((k - i)\delta) - J((k - j)\delta)]} \quad y > (k - i)\delta
 \end{aligned}$$

**Specific calculations: Exponential( $\mu$ ) sojourn time**

$$E(\mathbf{Y}_1) = \mu^{-1} [2 - (\mu\delta)e^{-\delta\mu} / (1 - e^{-\delta\mu})]$$

$$\rightarrow \mu \text{ as } \delta \rightarrow 0$$

$$Var(\mathbf{Y}_1) = (2 / \mu^2) [1 - \gamma(\mu\delta)^2 / 2(1 - \gamma)^2] \delta \rightarrow 0$$

$$\rightarrow \mu^2 \text{ as } \delta \rightarrow 0$$

Relative error (%) in Mean sojourn time (function of  $\mu / \delta$ )

	$\delta$				
$\mu$	1	2	3	4	5
$\frac{1}{2}$	70	93	98	100	100
1	42	70	84	93	97
2	23	42	57	70	77
5	10	19	27	35	42
10	5	10	14	19	23

Relative error (%) in standard deviation (function of  $\mu / \delta$ )

	$\delta$				
$\mu$	1	2	3	4	5
$\frac{1}{2}$	13	30	38	41	42
1	4	13	23	30	35
2	1	4	8	13	18
5	0	1	1	3	4
10	0	0	0	1	1

Ratio  $E(Y^*) / E(Y)$ ,  $\beta = 0.90$   
 four screens following prevalent screen

$\mu$	$\delta$				
	1	2	3	4	5
$\frac{1}{2}$	1.91	2.37	2.66	2.89	3.12
1	1.53	1.91	2.18	2.37	2.53
2	1.29	1.53	1.74	1.91	2.06
5	1.12	1.23	1.34	1.44	1.53
10	1.06	1.12	1.18	1.23	1.29

Ratio  $E(Y^*) / E(Y)$ ,  $\beta = 0.95$

$\mu$	$\delta$				
	1	2	3	4	5
$\frac{1}{2}$	1.79	2.14	2.30	2.42	2.53
1	1.47	1.79	2.00	2.14	2.23
2	1.26	1.47	1.65	1.79	1.91
5	1.10	1.21	1.30	1.40	1.47
10	1.05	1.11	1.16	1.21	1.26

Ratio  $SD(Y^*) / SD(Y)$ ,  $\beta = 0.95$

$\mu$	$\delta$				
	1	2	3	4	5
$\frac{1}{2}$	1.13	1.30	1.38	1.41	1.41
1	1.04	1.13	1.23	1.30	1.35
2	1.01	1.04	1.08	1.13	1.18
5	1.00	1.01	1.01	1.00	1.04
10	1.00	1.00	1.00	1.01	1.01

Density of  $\mathbf{Y}_k$  with  $\Gamma(r, \lambda)$  sojourn time:

$$f_{\mathbf{Y}_k}(y; r, \lambda) = [y - (k - 1)\delta] \lambda^r y^{r-1} e^{-\lambda y} / (\text{denom}) \quad (k - 1)\delta \leq y \leq k\delta$$

$$= \delta \lambda^r y^{r-1} e^{-\lambda y} / (\text{denom}) \quad y > k\delta$$

$$(\text{denom}) = \Gamma(r)[e^{\delta\lambda}(2 - \delta\lambda + k\delta\lambda) - k\delta\lambda - 2] / \lambda e^{k\delta\lambda}$$

$$E(\mathbf{Y}_k) = [-2 - k\delta\lambda + e^{\delta\lambda}(2 - \delta\lambda + k\delta\lambda)] / [\lambda(e^{\delta\lambda} - 1)]$$

Ratio  $100 \cdot [E(Y^*) / E(Y) - 1]$ ,  $\beta = 0.95$ ,

Gamma(2,1), four screens following prevalent screen

	$\delta$				
$\mu$	1	2	3	4	5
$\frac{1}{2}$	49	68	79	90	100
1	26	49	61	68	74
2	11	26	40	49	56
5	3	8	14	20	26
10	1	3	5	8	11

Ratio  $100 \cdot [E(Y^*) / E(Y) - 1]$ ,  $\beta = 0.95$ ,

Gamma(4,1), four screens following prevalent screen

	$\delta$				
$\mu$	1	2	3	4	5
$\frac{1}{2}$	31	43	53	64	74
1	15	31	37	43	48
2	4	15	24	31	34
5	0	2	6	10	15
10	0	0	1	2	4

Little effect on standard deviation

**Main issue: Ratio of clinical durations E(Z\*)/E(Z)**

Model joint distribution (preclinical, clinical) durations:  
*bivariate gamma distribution*

$$f_{YZ}(y, z) = \frac{e^{-(\lambda_1 y + \lambda_2 z)}}{[\psi L(y, z; a_1, a_2, \lambda_1, \lambda_2) + (1 - \psi)L(y, z; b_1, b_2, \lambda_1, \lambda_2)],}$$

$$L(y, z; a_1, a_2, \lambda_1, \lambda_2) = \lambda_1^{a_1} y^{a_1-1} \lambda_2^{a_2} z^{a_2-1} / [\Gamma(a_1)\Gamma(a_2)]$$

With probability  $\psi$ :  $Y \sim \Gamma(\lambda_1, a_1), Z \sim \Gamma(\lambda_2, a_2)$

With probability  $1 - \psi$ :  $Y \sim \Gamma(\lambda_1, b_1), Z \sim \Gamma(\lambda_2, b_2)$

$$\begin{aligned} \mu_x &= [\psi a_1 + (1 - \psi)b_1] / \lambda_1, & \sigma_x^2 &= (\omega_1^2 + \lambda_1 \mu_x) / \lambda_1^2 \\ \mu_y &= [\psi a_2 + (1 - \psi)b_2] / \lambda_2, & \sigma_y^2 &= (\omega_2^2 + \lambda_2 \mu_y) / \lambda_2^2 \\ \omega_1^2 &= (a_1 - b_1)^2 \psi(1 - \psi), & \omega_2^2 &= (a_2 - b_2)^2 \psi(1 - \psi) \\ \rho_{xy} &= (\omega_1 \omega_2) / (\lambda_1 \lambda_2 \sigma_x \sigma_y) \end{aligned}$$

One screen only:

$$E(YZ) / \mu_z = \frac{\psi a_1 a_2 + (1 - \psi)b_1 b_2}{[\psi a_1 + (1 - \psi)b_1][\psi a_2 + (1 - \psi)b_2]}$$

Ex:  $\mu_x = 4, \sigma_x = 3, \mu_y = 2, \sigma_y = 1, \rho = 0.5$ :

Non-unique solutions for  $a_1, a_2, b_1, b_2, \lambda_1, \lambda_2, \psi$ :

$$a_1 = 8.108, b_1 = 2.775, \lambda_1 = 0.889$$

$$a_2 = 14.828, b_2 = 6.828, \lambda_2 = 4.000, \psi = 0.1464$$

Result:  $E(Z^*)/E(Z) = 1.1874$ , or 18.74% increase

**Canadian national Breast Cancer Screening Study:  
13-Yr Results of Randomized Trial, Women, 50-59**

A.B. Miller et al. (JNCI 20 Sep 2000: 1490-1499)

*N* = 39,405; randomized Jan 1980 - Mar 1985;  
active follow-up to 6/30/96

Study: Ann Mammography + Phys Exam + BSE (19,711)

Control: Phys Exam + BSE (19,694)

*T* = 4 or 5 (first 62%) annual screens

Compliance: 100% (*T*=1), 90% (*T*=2); 86% (*T*=5)

<b>Detection:</b>	<b>Study</b>	<b>Control</b>
In-situ	71	16
Invasive: Screen-detected + Interval + Incident:		
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	267 + 50 + 207 = 524	148 + 88 + 264 = 500

<b>Mortality:</b>	<b>Study</b>	<b>Control</b>
Breast Cancer	88	90
Other Cancer	376	313
Other non-cancer	270	287

**Reduction in Mortality for cases diagnosed through:**

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)

## **Conclusions and further work**

Decision to implement screening programs depends upon methods of evaluation

Evaluation depends upon factors such as test sensitivity/specificity and outcome of *randomized* trials

Even randomized trials can involve biases, such as length biased sampling and overdiagnosis bias, which are hard to measure in terms of extended benefit times

Some 'rule' for defining *comparable case groups* is essential for accurate estimation of average lead time and average benefit time

The best chance for detecting a non-zero average benefit time occurs when the comparison includes few cases that could not have benefited from the screening (maximizing the 'signal to noise ratio')

Length biased sampling of preclinical durations result in extended survival times -- even after accounting for lead time, in the absence of screening benefit

Some of the analysis could be parametric, but more general analyses will require discrete event simulation and computer intensive estimates of uncertainty

HIP trial: Mammography plus clinical breast exam:

Screened benefit time estimate = 3.2 yr ( $\pm$  1.3 yr)

Screened lead time estimate = 9 mo ( $\pm$  4.5 mo)

These methods are expected to apply to future randomized screening trials (e.g., PLCO)

Applications in reliability; e.g. inspection preventive maintenance schedules (tail of Boeing 737: jackscrew that raises and lowers a control surface; periodic inspection of radioactive waste containers for leakage)

**Estimate of Average Lead Time for HIP trial:**

*Exponential model for preclinical durations:*

Median < 1 year

over 70% of lead times < 2 years

Mean lead time 13.3 months

*Average time to diagnosis for 392 cases (6.5 years):*

Program Lead time (all 392 cases):

Study: 36.3 months; Control: 39.8 months

Difference 3.5 months (se = 2.5)

Screened Lead time (only 132 of 392 screen-detected):

$(392/132) \times 3.5 = 10.4$  months (se = 4.5)

Little of extra survival time is due to lead time

**Estimate of Average Benefit Time:**

*Average survival time since start of study for 392 cases:*

3.2 years (+ 1.3 years)