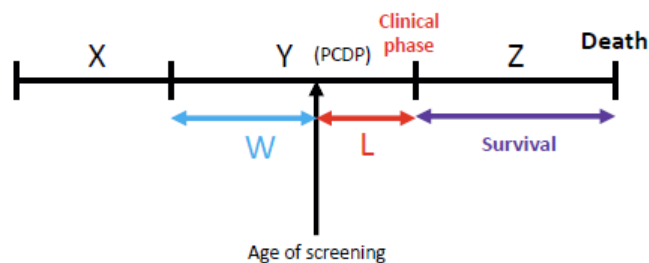


## Module 5 Bias Adjustment in Cancer and Chronic Disease Screening

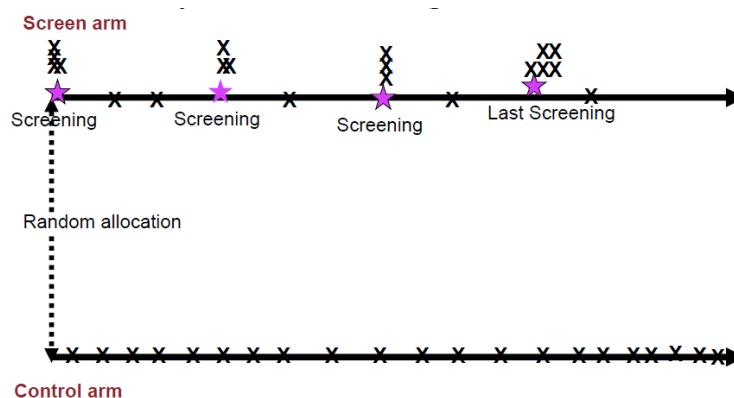
### 5.1 Lead-time bias

#### 5.1.1 The definition of lead time

The lead time is the interval between asymptomatic disease detected by screening and time to clinical phase with symptoms or signs.

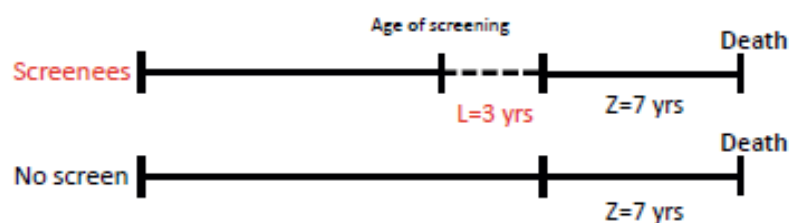


From the temporal natural history of disease, screen may detect the tumor/cancer at earlier time before presentation to clinical signs and symptoms even when screen cannot prolong life (lead-time bias).



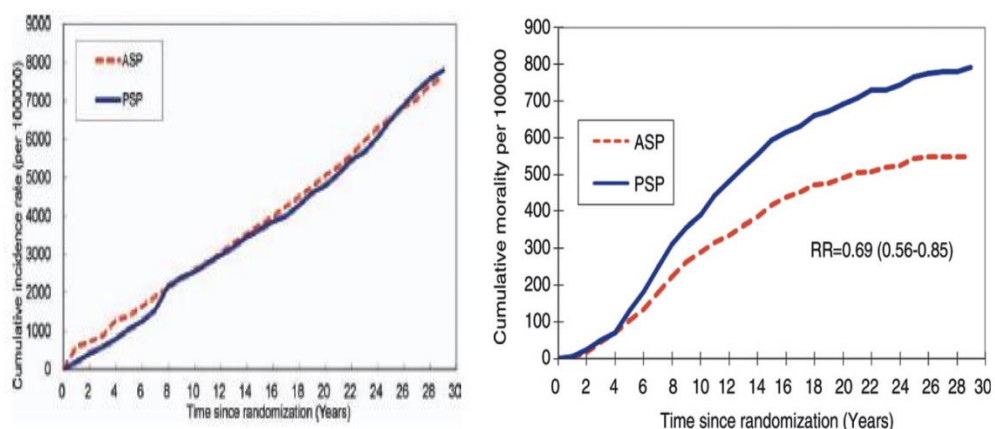
### 5.1.2 Survival analysis

The survival is the interval between date of disease diagnosis and date of death of this disease. Intuitively, about the survival calculation, there is no difference between screened and unscreened subjects. But, compared with no screen, we count more time interval in screenees because the subjects were diagnosed earlier by screening, which means the lead time was included into the survival.



### 5.1.3 Cumulative mortality of breast cancer

The evidence from breast cancer screening empirical data showed the more cancer cases (or more advanced cancer) were detected during the initial period of programme. However, the mortality of screening arm was not different from control arm. It indicated that more time interval we gained and lead to a lacking of difference in mortality during the initial period. It (lead time of breast cancer) seems around 4 years in cumulative mortality rate (Tabár et al. Radiology 2011 & Yen et al. Cancer 2012).



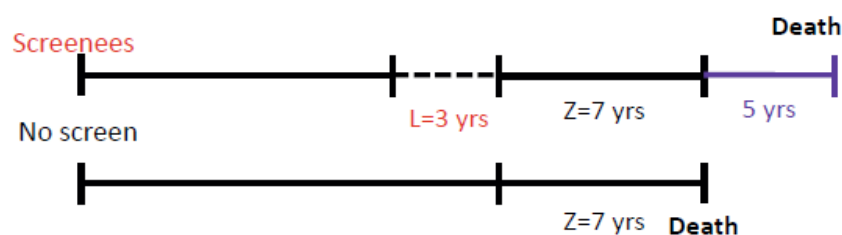
#### 5.1.4 Factors affecting the lead-time

- (A) time-point of screening intake
- (B) sensitivity of screening tool
- (C) Both (A) and (B) are correlated

#### 5.1.5 How to eliminate lead-time bias

- (A) From survival aspect → naïve method

Lead time cannot be directly observed by screening for ethical consideration. But, we could get from mathematical estimation. For example, we estimated survival of 15 years for the screenee's and 7 years for unscreened. We obtained information from modelling that the lead-time of breast cancer is about 3 years. therefore, the calibrated survival for screenee is  $15-3=12$  years. So, actually, the benefit of screening is 5 years.

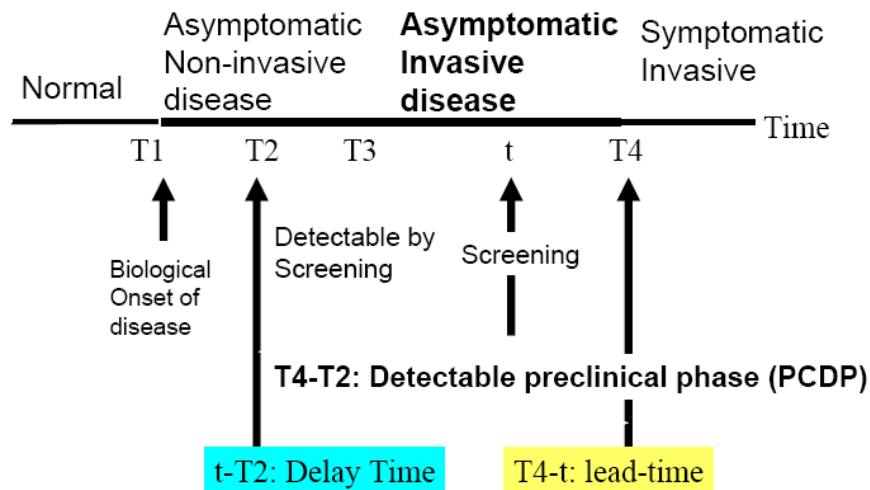


- (B) Mortality as observed endpoint

Mortality is not influenced by the timing of diagnosis. But, survival will vary by date of diagnosis depending on early detection by screening (lead-time gain).

## 5.2 How lead-time and length bias affect the survival by detection mode?

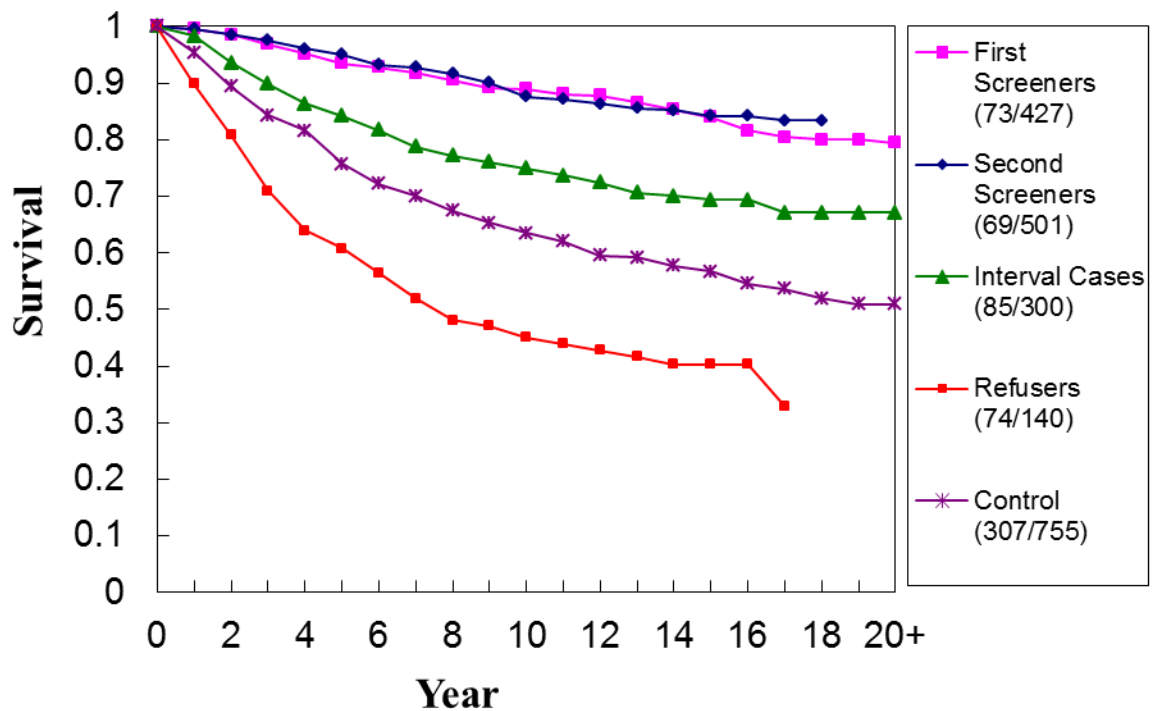
### 5.2.1 Model of disease progression



The interval between  $T_2$  and  $T_4$  is usually called the *pre-clinical detectable phase (PCDP)*. The duration  $T_4 - T_2$  is called the '*sojourn time*'.  $T_4 - t$  is the lead time gained in the screening programme and  $t - T_2$  is the *delay time*. It is noted that the time  $t$  of the screen plays an important role in deciding how much lead time can be gained. The earlier the time at  $t$  the more lead time gained and the shorter the delay time.

### 5.2.2 Survival curves by detection modes

The cumulative survival from the Swedish Two-county Trial is as follows.



(1) The cumulative survival rate by detection mode is as follow.

Screen-detected cancer (SD) > Interval cancer (IC) > Control > Refuser:

(2) Which comparisons would be valid for effectiveness of screening? What

bias may arise in these comparisons?

(a) SD vs. Clinically-detected cases (IC+Refuser)

(b) SD vs. Refuser

(c) SD+IC vs. Refuser

### **5.2.3 The effect of leadtime and length bias on survival time by detection mode (Wu et al, *Biom J* 2012)**

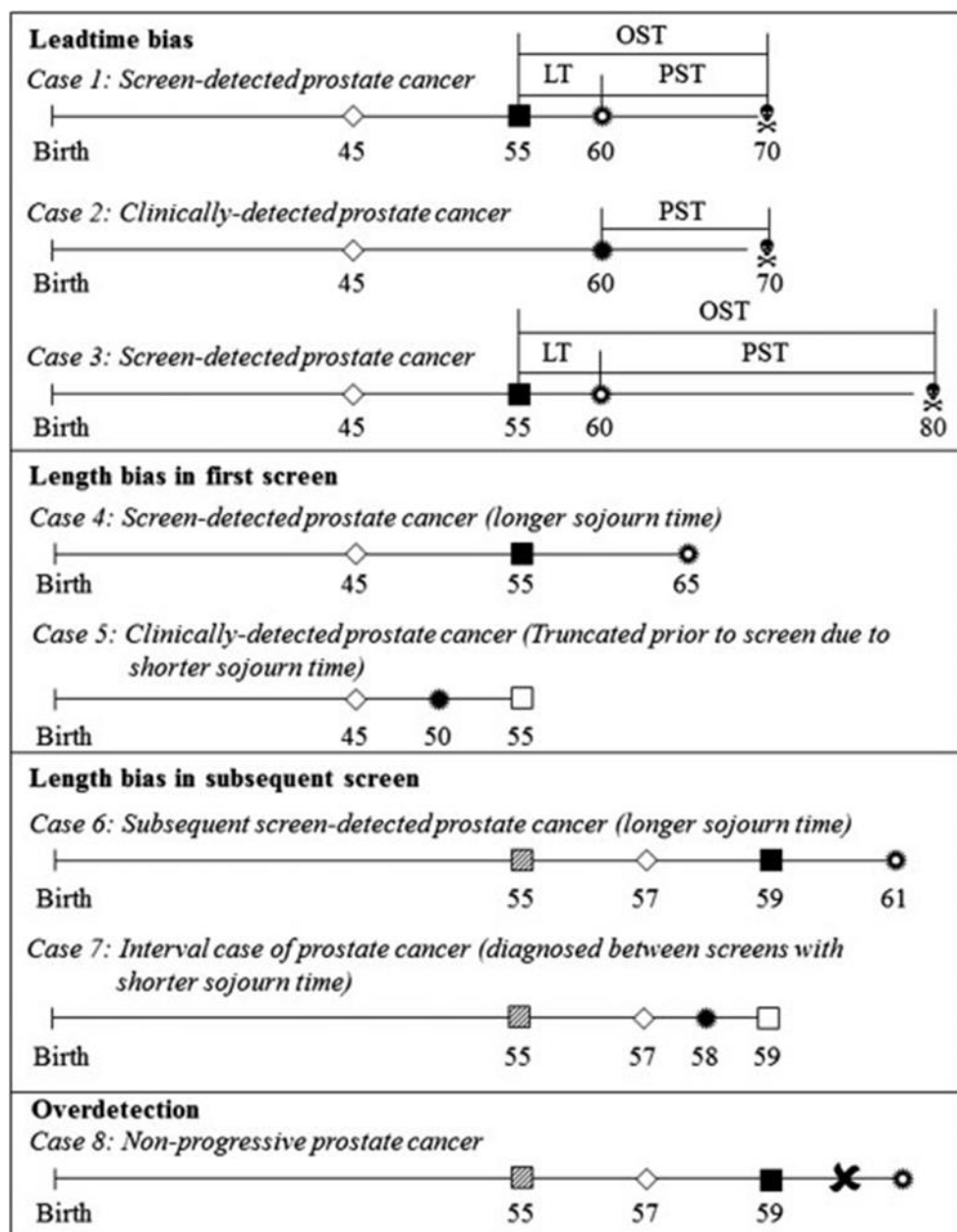
Leadtime bias, length bias, and over-detection of cancers are important issues in the evaluation of prostate cancer screening. They arise from the sojourn time, which is the duration of the pre-clinical detectable phase (PCDP), assuming that the temporal natural history of the disease follows a three-state model in which an individual's disease status is normal prior to the development of the disease, then passes through a PCDP and finally to the CP when the disease becomes symptomatic. Leadtime is the amount of time by which the detection of a cancer is advanced by screening. Length bias is inherent from the fact that tumors have different sojourn times, depending on their aggressiveness, and leads to the phenomenon that screen-detected cancers, particularly those detected at first screen, tend to have longer sojourn times than interval cancers (cancers diagnosed symptomatically between screens). Over-detected cases are defined as cancers with a sojourn time equal to infinity i.e., cases that would not have been diagnosed if there had been no screening.

Prostate cancer is the most common cancer in men in many industrialized countries and has a slow natural course. Analyses comparing the survival of clinically-detected cases with the survival of those screen-detected, therefore, need to be adjusted for the above biases. Firstly, the early detection of cases may simply advance the date of diagnosis without prolonging life (see cases 1 and 2 in Figure 1 where the earlier diagnosis of case 1 leads to 5 years of artificial leadtime), resulting in leadtime bias. The mean leadtime for prostate cancer has been estimated to be between 5 and 12 years. Even if early

detection due to screening with the prostate specific antigen (PSA) test does genuinely prolong life, when no adjustment is made for leadtime the associated survival benefit will be exaggerated (for example, case 3 as opposed to case 2 in Figure 1 has additional 10 years of survival after correction for a 5-year leadtime instead of 15 years without correction). Secondly, empirical data on screen-detected and interval cases of prostate cancer ascertained within a population-based screening program provide valuable information as regards length bias. Screen-detected prostate cancers, particularly those detected at the first screen, tend to have a longer sojourn time (see case 4 in Figure 1) than those arising clinically before the first screen (case 5 in Figure 1). Similarly, prostate cancers arising after the first screen are more likely to be detected at subsequent screens if they have a longer sojourn time (case 6 compared with case 7 in Figure 1). Interval cancers (cases diagnosed clinically in the interval between screens following a negative screen) are not affected by leadtime bias and have shorter sojourn times (case 7 in Figure 1). These scenarios suggest that the distribution of sojourn time for screen-detected cases is different to that for interval cancers. Thirdly, previous statistical models to adjust for these biases when comparing the survival of screen-detected and clinical cases of prostate cancer have not taken over-detection into account. This is a major weakness as some screen-detected prostate cancers progress so slowly that they would never produce symptoms (case 8 in Figure 1), a major issue in prostate cancer screening with the PSA test.

Unfortunately, leadtime, sojourn time and over-detection cannot be directly observed because medical treatment interrupts the natural course of screen detected cases, leaving the key details (the times at which the disease entered

the PCDP and CP) unknown. Sophisticated statistical models are therefore required to estimate the unknown variables.



◇: Date of biological onset (entering pre-clinical phase); \* : date of surfacing to clinical phase due to occurrence of symptoms or signs; ○ : hypothetical date of surfacing to clinical phase if no screen had taken place; ✕: date of death from prostate cancer; ■: date of detection by screen; □: hypothetical date of screen if no clinical symptoms or signs of prostate cancer had taken place; ▨: date of screen as normal; OST: observed survival time; LT: lead time; PST: post-leadtime survival time.



Case 1 vs. case 2: The early detection of cases may simply advance the date of diagnosis without prolonging life where the earlier diagnosis of case 1 leads to 5 years of artificial leadtime, resulting in leadtime bias.

Case 3 vs. case 2: Case 3 has additional 10 years of survival after correction for a 5-year leadtime instead of 15 years without correction).

Case 4 vs. case 5: Case 4 have shorter sojourn time and be truncated prior to screen. Therefore those detected at the first screen (case 5) tend to have a longer sojourn time.

Case 7 vs Case 6: Interval cancers (Case 7) are not affected by leadtime bias and have shorter sojourn time These scenarios suggest that the distribution of sojourn time for screen-detected cases is different to that for interval cancers.

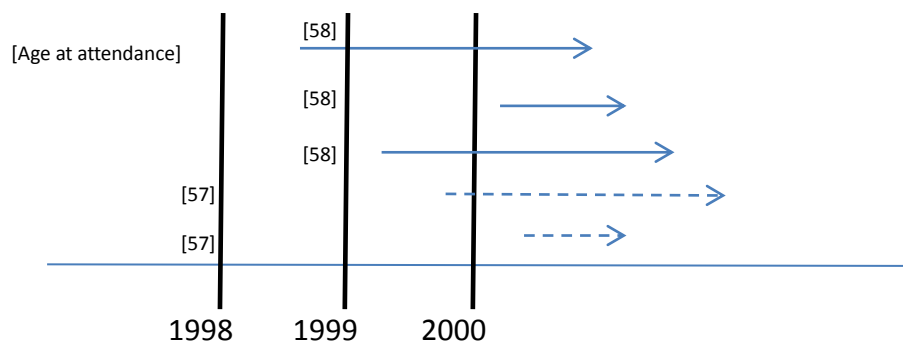
### **5.3 Mean sojourn time estimation from non-randomized breast cancer screening program (Wu et al, *Breast Cancer Research Treatment*, 2010)**

#### **5.3.1 For estimation of mean sojourn time (MST), several issues raised from non-randomized breast cancer screening which should be further considered**

- (1).Selection-bias: The progressions of breast cancer between participants and non-participants.
- (2).Measurement errors: the role of sensitivity and specificity
- (3).Truncated screening period: inherent lead-time bias

### 5.3.2 A demonstration of truncated problem from Finnish breast cancer service screening program (Wu et al., 2010)

- (1). Screening program: Mammography screening was offered to women aged between 50 and 59 years in Finland between 1988 and 2000
- (2). Average numbers of screen in fixed study period.: 2.53 ( 55-59 years) < 3.07 (50-54 years)
- (3). Slowly-growing breast tumor with long sojourn time (i.e., small but still undetectable by mammography when they were invited to screen) for women aged 55–59 years would not be detected given less rounds of screen offered



- (4). Solution: post-screening cancers (PSC), Clinical breast cancers diagnosed after 60 years of age and occurring after last invitation with follow-up time.
- (5). Detection modes for estimation of natural history of breast cancer
  - A. Prevalent screen
    - I. Normal: True negative + false negative cases
    - II. Prevalent Cancer: PCDP breast cancer detected at first screen
  - B. Later screen

- I. Normal: True negative + false negative cases (staying in the PCDP) at subsequent screen
- II. Incident Cancer: PCDP breast cancer detected at subsequent screen
- C. Interval cancer: Clinical breast cancer (newly diagnosed cases and false negative cases surfacing to clinical phase) developed between screens
- D. Refuser cancer: Clinical breast cancers arising from non-participant
- E. Post-screening cancer: Breast cancers occurring after last invitation with follow-up until the end of study

### 5.3.3 Shorter Estimation of MST with consideration of sensitivity

Parameters	Estimates	95% CI
Three-state model <sup>a</sup>		
50–59		
Normal → preclinical cancer ( $\lambda_1$ )	0.0025	(0.0022, 0.0028)
Preclinical cancer → clinical cancer ( $\lambda_2$ )	0.4956	(0.3816, 0.6097)
Mean sojourn time ( $1/\lambda_2$ )	2.02	(1.64, 2.62)
Sensitivity	84.83%	(74.88%, 94.79%)
Specificity	99.97%	(99.89%, 100%)
50–54		
Normal → preclinical cancer ( $\lambda_1$ )	0.0025	(0.0022, 0.0027)
Preclinical cancer → clinical cancer ( $\lambda_2$ )	0.5207	(0.4057, 0.6356)
Mean sojourn time ( $1/\lambda_2$ )	1.92	(1.57, 2.46)
Sensitivity	83.75%	(71.26%, 96.23%)
55–59		
Normal → preclinical cancer ( $\lambda_1$ )	0.0025	(0.0021, 0.0029)
Preclinical cancer → clinical cancer ( $\lambda_2$ )	0.4269	(0.3131, 0.5408)
Mean sojourn time ( $1/\lambda_2$ )	2.34	(1.85–3.19)
Sensitivity	89.48%	(76.56%, 100%)

2.02 years of MST for women aged 50-59 due to (1) the truncation of slow-growing breast tumor (2) sensitivity and MST are negatively correlated the higher the sensitivity the shorter the sojourn time (3) other biological and organized factors

## **5.4 Lead-time adjusted survival with stochastic models (Chen et al, *JASA* 2012)**

Comparison of the survival of clinically detected and screen-detected cancer cases from either population-based service screening programs or opportunistic screening is often distorted by both lead-time and length biases.

**5.4.1** Although lead-time bias and length bias are both related to sojourn time and a consequence of screening the two issues have a fundamental difference.

- (1) Lead-time bias is inherent, a function of the sojourn time
- (2) Length-bias results from the oversampling of cancers with long sojourn times.

### **5.4.2 Lead-time bias vs. Measurement error**

The false-negative rate is positively correlated with the MST in breast cancer. Again, the lower sensitivity the longer mean sojourn time

### **5.4.3 Why should we consider length-bias adjustment after correction of lead-time?**

The individual heterogeneity of sojourn time could be captured by including information about disease aggressiveness (e.g., lymph node involvement).

The more severe the pre-clinical stage of the tumor the closer the disease is to entering the clinical phase. The comparison of survival between screen-detected and clinically-detected cases would be affected by a

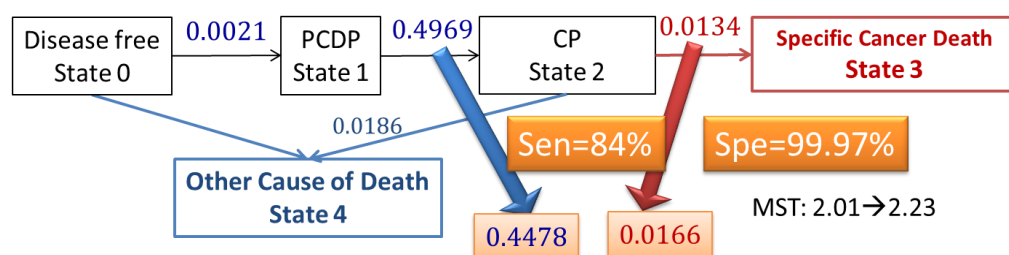
degree of length-bias even after correction for lead-time.

#### 5.4.4 Lead-time bias vs. attribute of the tumor

Because both the sojourn time and false negative rate are related to the attributes of the tumor (such as lymph node involvement) lead-time bias is thought to be smaller for larger, more aggressive tumors (e.g., those with lymph node involvement or poor differentiation) than for smaller, less aggressive tumors.

#### 5.4.5 Estimation results from two-county breast cancer screening program

(1) Based on 25-year follow-up, assuming 100% sensitivity and specificity, an estimated MST of 2.01 years. The estimate of the MST increased to 2.23 years when the sensitivity (83.66%) and specificity (99.95%) were accounted.



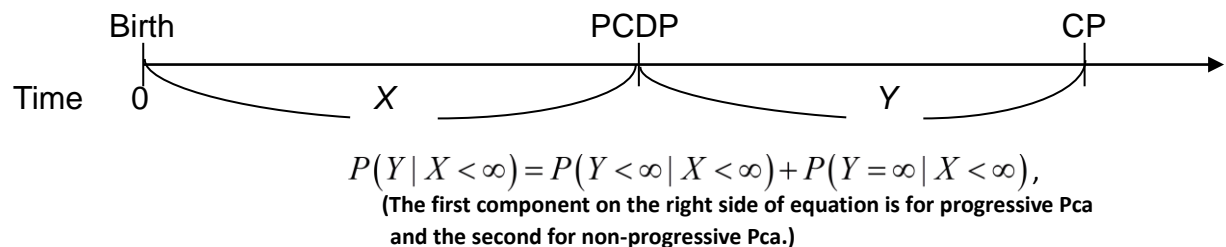
(2) The shorter MST (1.47 years) for women aged 40-49 compared with those aged 50-59 (2.41 years). When measurement errors were taken into account, the estimates of the MST were increased to 1.95 years for the younger women (68.68% sensitivity) and to 2.60 years for the older women (88.79% sensitivity).

- (3) The estimate of 1.82 years for the MST for those without nodal involvement was longer than 0.82 years of MST for those with nodal involvement. Considering measurement error, the estimates of MST were 2.56 years for node-negative and 1.04 years for node-positive.
- (4) The MSTs for women aged 40–49 years, without and with nodal involvement, were estimated to be 1.34 years and 0.47 years. Considering measurement error, The MSTs for women aged 40–49 years, without and with nodal involvement, were estimated to be 2.57 years and 0.63 years (57.7% sensitivity for N(-), 94.19% sensitivity for N(+)).
- (5) The MSTs for women aged 50–59 years, without and with nodal involvement, were estimated to be 2.21 years and 1.14 years. Considering measurement error, The MSTs for women aged 50–59 years, without and with nodal involvement, were estimated to be 2.69 years and 1.41 years (87.13% sensitivity for N(-), 88.47% sensitivity for N(+)).

## 5.5 Lead-time bias and length bias adjustment for prostate cancer screening

### 5.5.1 The characteristics of the natural history for prostate cancer (PCa)

(Wu et al, *IJC* 2012)



- A. slow natural disease course  $\rightarrow Y$  is longer than other diseases
- B. elderly onset age  $\rightarrow X$  is larger than other diseases
- C. existing indolent disease  $\rightarrow P(Y = \infty | X < \infty) > 0$
- D. incidence rate increasing with age  $\rightarrow X$  is not exponential distributed

### 5.5.2 An application to screening policy of prostate-specific antigen (PSA)

screening (Wu et al, *Eu Urology* 2012; 61: 1011-1018)

The conflicting results of the population-based screening for PCa using PSA test have been reported between two main RCT, PLCO in the USA and ERSPC in the Europe. The relative mortality rates varied across study centers in the ERSPC. The efficacy of population-based PSA screening is affected by (1) screening interval, (2) age at start and termination of screening, (3) attendance, and (4) contamination of the controlled group in the RCT.

The computer simulation considering sensitivity of PSA test, the natural course of PCa, and competing cause of death for different screening policies could help us to assess the impact of above-mentioned

factors on screening efficacy. The efficacy was measured in terms of the reduction in advanced PCa (stage III or worse) and Pca mortality.

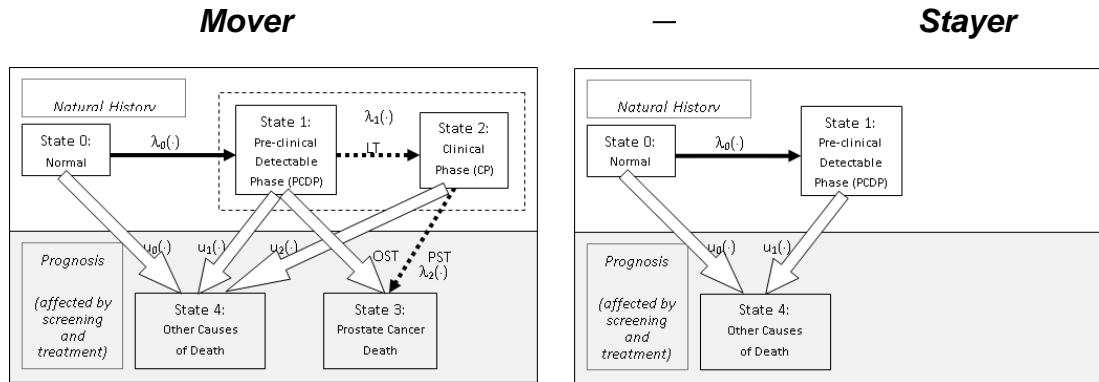
Wu et al's study showed that the screening interval had a greater impact on mortality reduction than did the age to begin screening from 55 yr onwards. The good results of both internal and external validation, except Swenden in ERSPC, indicated the adequacy of their model. The external validation helped us identify factors accounting for the conflicting findings of the ERSPC and PLCO. The implication suggests that the benefit of annual PSA screening may be offset, to a large extent, by low biopsy compliance and high cut-off PSA and, to a lesser degree, by high contamination. The intensive screening protocol used at the Swedish center with a lower PSA cut-off level may increase the sensitivity of the test and lengthen the sojourn time, thus providing additional time for early detection and greater reduction in mortality.

### **5.5.3 Use survival of early prostate cancer with adjustments for leadtime, length bias, and over-detection to demonstrate the screening benefit (Wu et al, *Biometrical J*, 2012; 54: 20-44)**

#### **(1) Strategy for calibrating biases**

- A. Leadtime bias: estimate sojourn time and applied the transition rate from PCDP to the CP to project the adjusted survival curve
- B. Length bias: applying left-truncation for screen-detected case in the prevalent screen, and including interval cancers
- C. Over-diagnosis: using mover-stayer model to treat overdiagnosed cases as stayer for them never moving to the CP





## (2) Empirical example of the Finnish center in the ERSPC

### A. Comparison between screen-detected and clinical detected PCa

Crude hazard ratio (HR): 0.24 (95% CI: 0.16-0.35)

### B. -1 Calibrating for leadtime and length biases with constant hazards

$(\widehat{MST} = 5.24, 95\% \text{ CI: } 4.82 - 5.74)$

aHR: 0.61 (95% CI: 0.45-0.81)

### -2 Calibrating for leadtime and length biases with non-constant hazards

$(\widehat{MST}_{55-62} = 3.54, 95\% \text{ CI: } 3.06 - 4.20,$

$\widehat{MST}_{63+} = 8.21, 95\% \text{ CI: } 7.09 - 9.76)$

aHR: 0.76 (95% CI: 0.58-1.00)

### C. Further adjusting for over-detection (40.45%)

$(\widehat{MST}_{55-62} = 7.27, 95\% \text{ CI: } 5.60 - 10.37$

$\widehat{MST}_{63+} = 7.46, 95\% \text{ CI: } 6.37 - 9.01)$

aHR: 1.03 (95% CI: 0.79-1.33)

Figure Prostate cancer survival curves of screen-detected and clinically detected prostate cancer cases adjusted for leadtime bias, length bias, and over-detection in the Finnish population-based prostate cancer screening randomized controlled trial, 1996–2005

