Scientific
Research

An Exact Formula for Estimation of Age-Specific Sensitivity for Screening Tests

Ning Jia, Sandra J. Lee[*]

Department of Biostatistics and Computational Biology, Dana-Farber Cancer Institute, Boston, USA
Email: ning@jimmy.harvard.edu, [*]sjlee@jimmy.harvard.edu

Abstract

There has been a growing interest in screening programs designed to detect chronic progressive cancers in the asymptomatic stage, with the expectation that early detection will result in a better prognosis. One key element of early detection programs is a screening test. An accurate screening test is more effective in finding cases with early-stage diseases. Sensitivity, the conditional probability of getting a positive test result when one truly has a disease, represents one measure of accuracy for a screening test. Since the true disease status is unknown, it is not straightforward to estimate the sensitivity directly from observed data. Furthermore, the sensitivity is associated with other parameters related to the disease progression. This feature introduces additional numerical complexity and limitations, especially when the sensitivity depends on age. In this paper, we propose a new approach that, through combinatorial manipulation of probability statements, formulates the age-dependent sensitivity. This formulation has an exact and simple expression and can be estimated based on directly observable probabilities. This approach also helps evaluate other parameters associated with the natural history of disease more accurately. The proposed method was applied to estimate the mammography sensitivity for breast cancer using the data from the Health Insurance Plan trial.

Keywords

1. Introduction

Screening of asymptomatic individuals for chronic diseases is a rapidly growing public health initiative. Early detection programs are aimed at detecting the disease in the stage when a disease is present without symptoms.

---

[*]Corresponding author.

For example, in breast cancer, there had been eight randomized screening trials demonstrating that mammography screening is beneficial in finding breast cancer at an earlier stage and consequently leads to a decrease in mortality among women of 50 - 65 years of age ([1]-[7]).

One key element of these programs is a screening test. An important measure for the effectiveness of a screening test is sensitivity (β), the conditional probability of getting a positive screening test result given that one has the disease. Ideally it should be evaluated in the setting of natural history of disease model ([8]-[11]). Often parameters like the sojourn time distribution in early stage of disease (pre-clinical state), transition probability from no disease state to early-stage disease state are needed to estimate the sensitivity.

These parameters are not directly observable and add to the complexity of the problem formulation and numerical computation of the sensitivity. Furthermore, the sensitivity can be age-dependent. For example in breast cancer the mammogram sensitivity in younger women is lower than that in older women [12]. The formulation for estimating age-dependent sensitivity becomes complex and does not always guarantee a numerical solution.

In this paper, we derive a formula that expresses age-dependent sensitivity in terms of probabilities that are directly observable in screening trials/programs and discuss the characteristics and generalization of this formula. We apply the formula to breast cancer screening trial (Health Insurance Plan) data to estimate the mammography sensitivity [1].

2. Method: An Exact Expression for Age-Dependent Sensitivity β_a

Consider the following health states in the natural history of disease progression:
- S_0, the disease-free state, when the disease cannot be detected by any current early detection program;
- S_p, the pre-clinical state, when the disease is detectable by an early detection program, but no symptoms are shown; and
- S_c, the clinical state, when symptoms show.

A progressive disease model assumes that the disease progresses in the direction $S_0 \to S_p \to S_c$ unless it is interrupted by a medical intervention. In the case of chronic progressive diseases, the goal of screening program in detecting the disease in S_p. The age is denoted by a. Let
- β_a be the sensitivity of the screening exam (*i.e.*, the conditional probability that a disease is detected by the screening exam given disease is present) if the exam is taken at age a;
- w_a be the transition rate from S_0 to S_p at age a;
- $q_a(t)$ be the sojourn time distribution in S_p if the transition $S_0 \to S_p$ happens at age a; and
- P_a be the probability for case being in the pre-clinical state S_p at age a, *i.e.* the proportion of people in S_p at age a in absence of screening examinations.

Note that β, w, q and P are *functions* of age a, but to simplify notations, we use a in the subscripts. We also do this in part because our method will give estimates of β for discrete values of age a.

We now introduce two important probabilities that all the subsequent derivations are based upon. Suppose there are no screening examinations before age a, and a series of examinations are scheduled (but not necessarily taken) at times $0 \leq t_1 \leq t_2 \leq \cdots \leq t_n$ at ages $a, a+t_1, a+t_2, \cdots, a+t_n$, with no examinations at other times. Consider a person of age a at time 0, taking m examinations, at times $t_{i_1} \leq t_{i_2} \leq \cdots \leq t_{i_m}$. Let
- $X_a(t)$ be the probability of entering S_p before a, and still be in S_p at age $a+t$ if there has been no screening examinations; and
- $T_{i_1,i_2,\cdots,i_m}$ be the probability for a case undetected by the first $m-1$ exams and being detected in the pre-clinical state at the last examination at time t_{i_m}.

The importance of $T_{i_1,i_2,\cdots,i_m}$ is obvious as it is the probability that can be estimated directly from screening programs by taking proportions. The notion of $X_a(t)$ is more subtle, but it will facilitate derivations which will be shown later, especially when determining w and $q(t)$.

Note that it is implied in the definition of T and X that by the time of the last examination, one would either be in S_0 or S_p, but not S_c. Thus if we were to use such definitions, we need to limit the cohort to only those who will still be in S_0 or S_p by the last examination taken. In fact, the last examination would be at the same time when multiple X's and T's appear in an equation together in all our derivations. Also note that $X_a(0) = P_a$.

We derive a formula for β_a via a recursion based on the number of examinations. Consider a cohort of age a at time 0, with examinations scheduled at times $0 = t_0 \leq t_1 \leq t_2$. We have $T_i = P_{a+t_i}\beta_{a+t_i}$ for $i = 0, 1, 2, 3 \cdot n$

$T_{0,1}$, or the probability of a case undetected at examination at time 0 but at t_1, involves the probability to enter S_p

- before time 0 (or age a), go undetected at time 0, and stay in S_p until getting detected t_1, which is $X_a(t_1)(1-\beta_a)\beta_{a+t_1}$; and
- between time 0 and t_1, and stay in S_p until at least t_1. This can be thought of as the difference between probability to be in S_p at age $a+t_1$ if there are no exams, or P_{a+t_1}, and the probability to enter S_p before age a and stay in it until age $a+t_1$, or $X_a(t_1)$. To be detected at t_1, just multiply by β_{a+t_1}, which gives $\left(P_{a+t_1} - X_a(t_1)\right)\beta_{a+t_1}$.

Therefore,

$$T_{0,1} = X_a(t_1)(1-\beta_a)\beta_{a+t_1} + \left(P_{a+t_1} - X_a(t_1)\right)\beta_{a+t_1}$$
$$= -X_a(t_1)\beta_a\beta_{a+t_1} + P_{a+t_1}\beta_{a+t_1}$$
$$= -X_a(t_1)\beta_a\beta_{a+t_1} + T_1.$$

Or,

$$X_a(t_1)\beta_a\beta_{a+t_1} = T_1 - T_{0,1}. \tag{1}$$

Similarly, for $T_{0,1,2}$, we need three terms, involving the probability to enter S_p during the following times and stay in S_p until being detected at t_2:

- before age a and go undetected at times 0 and t_1, which is $X_a(t_2)(1-\beta_a)(1-\beta_{a+t_1})\beta_{a+t_2}$;

- between ages a and $a+t_1$ and go undetected at age $a+t_1$, which is $\left(X_{a+t_1}(t_2-t_1) - X_a(t_2)\right)(1-\beta_{a+t_1})\beta_{a+t_2}$; and

- between ages $a+t_1$ and $a+t_2$, which is $\left(P_{a+t_2} - X_{a+t_1}(t_2-t_1)\right)\beta_{a+t_2}$.

Therefore,

$$T_{0,1,2} = X_a(t_2)(1-\beta_a)(1-\beta_{a+t_1})\beta_{a+t_2}$$
$$+ \left(X_{a+t_1}(t_2-t_1) - X_a(t_2)\right)(1-\beta_{a+t_1})\beta_{a+t_2}$$
$$+ \left(P_{a+t_2} - X_{a+t_1}(t_2-t_1)\right)\beta_{a+t_2}$$
$$= -X_a(t_2)\beta_a\beta_{a+t_2} + X_a(t_2)\beta_a\beta_{a+t_1}\beta_{a+t_2}$$
$$- X_{a+t_1}(t_2-t_1)\beta_{a+t_1}\beta_{a+t_2} + P_{a+t_2}\beta_{a+t_2}.$$

From (1), we have $X_{a+t_1}(t_2-t_1)\beta_{a+t_1}\beta_{a+t_2} = T_2 - T_{1,2}$ and $T_i = P_{a+t_i}\beta_{a+t_i}$, substituting both to the above formulation, we get

$$T_{0,1,2} = -\left(T_2 - T_{0,2}\right) + X_a(t_2)\beta_a\beta_{a+t_1}\beta_{a+t_2} - \left(T_2 - T_{1,2}\right) + T_2,$$

which gives

$$X_a(t_2)\beta_a\beta_{a+t_1}\beta_{a+t_2} = T_{0,1,2} - T_{0,2} - T_{1,2} + T_2 \tag{2}$$

Observing (1) and (2), we arrive at our main result in this section:

Theorem 2.1.

$$\beta_{a+t_1} = \frac{X_a(t_2)\beta_{a+t_2}\beta_{a+t_1}\beta_a}{X_a(t_2)\beta_a\beta_{a+t_2}} = \frac{T_{0,1,2} - T_{0,2} - T_{1,2} + T_2}{T_2 - T_{0,2}} = 1 - \frac{T_{1,2} - T_{0,1,2}}{T_2 - T_{0,2}} \tag{3}$$

Therefore by writing out a few recursions, we are able to eliminate all $X_a(t)$ terms and express the age- dependent β as a simple expression of the T terms, which are directly observable. A special case for application of Theorem 2.1 is when $t_1 = t_2 = 0$. This implies a situation where repeated tests are performed on the same individual to measuring the sensitivity. For example, assume up to three repeated tests with the same sensitivity are performed at the same time, at three testing centers with same equipment. We label the testing centers as

Center 1, Center 2 and Center 3.

Let T_0 be the probability of being detected at Center 1, regardless of what the results from Centers 2 and 3. Let $T_{0,0}$ be the probability of being detected by Center 2, but not Center 1, regardless of the result from Center 3. This is equivalent to the probability of being detected by Center 3, but not Center 2, regardless of the result from Center 1. Lastly, let $T_{0,0,0}$ be probability of being detected by Center 3, but not Center 1 or 2. The subtlety here is that $T_{0,0,0}$ is not the probability of being detected by one center out the three, but just 1/3 of it, because we have ordered the centers beforehand. Then we have

Corollary 2.2. *When at least three repeated tests with the same sensitivity β can be performed at the same time*, *we have*

$$\beta = 1 - \frac{T_{0,0} - T_{0,0,0}}{T_0 - T_{0,0}}$$

The result is achieved by simply taking limits $t_1 \to 0$ and $t_2 \to 0$ in the derivation for Theorem 2.1.

Theorem 2.1 is far from being intuitive, but corollary is easily verified, when we realize that when $t_1 = 0$ and $t_2 = 0$, there is $T_0 = P_a \beta_a$, $T_{0,1} = P_a (1 - \beta_a) \beta_a$, and $T_{0,1,2} = P_a (1 - \beta_a)^2 \beta_a$.

Advantage of Theorem 2.1. One main advantage of Theorem 2.1 is that it enables us to ignore the consideration for transition rate w and sojourn time distribution $q(t)$ in our set-up, which, even in the simplest cases, creates a complicated relationship. For example, consider the cohort of age 50, all getting their first screening exam. The probability of being detected, if we write it all out as an integral [13], is

$$T_0 = \beta_{50} \int_0^{50} w_x \int_{50-x}^{\infty} q_x(y) \, dy \, dx$$

where w_x is the transition probability from S_0 to S_p at age x and $q_x(t)$ is the sojourn time distribution if the transition $S_0 \to S_p$ happens at age x. The interpretation of this expression is straight-forward: to be detected at age 50, the transition $S_0 \to S_p$ needs to happen at some age x before 50, therefore the first integral; and S_p needs to last from age x to at least age 50, therefore the second integral.

Thus in just this one term, all the w_x's and q_x's for x less than age 50 are involved, when we do not even know what *type* of distribution q_x is, which may vary for different x's. This is just the simplest case—when there are multiple exams involved, the expressions for the probability of detection will involve several double integrals. One of the more rigorous approaches by Shen and Zelen ([10]) uses such expressions to form likelihood function to find estimates for β, w and q. But due to the complexity of such expressions, even when age-dependency is ignored, and q is assumed to be exponential (so there are only three parameters to determine: β, w and the parameter for the exponential q), there is no guarantee that a global maximum exists, or can be determined [11].

However Theorem 2.1 allows us to bypass w and q_x and find an expression for β in as an exact expression of T's. This, of course, does not mean that the effects of β on detection can be isolated from that of w and q_x, but that the information of w and q_x are already incorporated in the T's. We also want to point out that the T's involved in Theorem 2.1 are naturally observable probabilities in the case of breast cancer screening (which is the main area of study for the authors): it is recommended that women to be screened either once every year, which gives terms $T_{1,2}$, $T_{0,1,2}$, or once every other year, which gives term $T_{0,2}$.

Another important advantage of Theorem 2.1 is that once (age-dependent) β is determined, we will be able to determine age-dependent w and q with various approaches in a more exact manner. This we will explore in another paper.

One limitation of the Theorem 2.1 is that it does not provide a way to apply the "interval cases", or the cases that enter clinical stages between two scheduled screening exams. These cases, however, can be used later on to determine w and $q(t)$ once β is determined.

3. Symmetry between *X* and *T*

An interesting mathematical result involves certain symmetry between *X* and *T*. We use the same setup as in the last section: suppose there are no examinations before age a, and a series of exams are scheduled (but not necessarily taken) at times $0 = t_0 \leq t_1 \leq t_2 \leq \cdots \leq t_n$ $(t_1 \geq 0)$ at ages $a, a + t_1, a + t_2, \cdots, a + t_n$, with no examinations at other times. Consider a person of age a at time $t_0 = 0$, taking m examinations at times $t_{i_1} \leq t_{i_2} \leq \cdots \leq t_{i_m}$.

Recall that we let $T_{i_1 i_2 \cdots i_m}$ be the probability for the person to be undetected by all of the first $m-1$ exams and only detected to be in the pre-clinical stage at the last exam at time t_{i_m}.

Under the same set up, we introduce the new notation:

$$Y_{i_1 i_2 \cdots i_m} = X_{a+t_{i_1}}\left(t_{i_m} - t_{i_1}\right) \cdot \beta_{a+t_{i_1}} \beta_{a+t_{i_2}} \cdots \beta_{a+t_{i_m}}, \tag{4}$$

where $X_{a+t_{i_1}}\left(t_{i_m} - t_{i_1}\right)$ is the probability for someone to enter S_p before age $a+t_{i_1}$ and still being in S_p at age $a+t_{i_m}$, if there has not been any medical intervention. Thus $Y_{i_1 i_2 \cdots i_m}$ is the probability for such a patient being detected at examinations taken at times $t_{i_1} \leq t_{i_2} \leq \cdots \leq t_{i_m}$. This is a hypothetical probability that will facilitate our derivation.

We also recall how we have defined T: let $T_{i_1 i_2 \cdots i_m}$ be the probability for the person to be undetected by all of the first $m-1$ exams and only detected to be in the pre-clinical stage at the last examination at time t_{i_m}. So the onset of S_p can happen anytime before t_{i_m}, as compared to in the definition of $X_{a+t_{i_1}}\left(t_{i_m} - t_{i_1}\right)$ (and therefore $Y_{i_1 i_2 \cdots i_m}$), where the onset of S_p occurs before age $a+t_1$.

Recall we derived the following expressions in the last section (notice that $P_{a+t_n} \beta_{a+t_n} = X_{a+t_n}(0) \beta_{a+t_n} = Y_n$):

$$T_{0,1} = -X_a(t_1) \beta_a \beta_{a+t_1} + T_1$$

$$T_{0,1,2} = -X_a(t_2) \beta_a \beta_{a+t_2} + X_a(t_2) \beta_a \beta_{a+t_1} \beta_{a+t_2}$$
$$\qquad - X_{a+t_1}(t_2 - t_1) \beta_{a+t_1} \beta_{a+t_2} + P_{a+t_2} \beta_{a+t_2}$$

$$X_a(t_1) \beta_a \beta_{a+t_1} = T_1 - T_{0,1}$$

$$X_a(t_2) \beta_a \beta_{a+t_1} \beta_{a+t_2} = T_{0,1,2} - T_{0,2} - T_{1,2} + T_2$$

or, after some reorganizing of the terms, and recognizing that $T_0 = X_a(0) \beta_a = Y_0$, we have

$$T_0 = Y_0$$
$$T_{0,1} = -Y_{0,1} + Y_1 \qquad Y_{0,1} = -T_{0,1} + T_1$$
$$T_{0,1,2} = Y_{0,1,2} - Y_{0,2} - Y_{1,2} + Y_2 \qquad Y_{0,1,2} = T_{0,1,2} - T_{0,2} - T_{1,2} + T_2$$

Therefore it is clear that there is a symmetric relationship between the T terms and the corresponding Y terms, in the case of 1, 2 and 3 examinations. The pattern of these expressions becomes more obvious in case of 4 examinations: (which will be proved along with the general case) to make it more obvious:

$$T_{0,1,2,3} = -Y_{0,1,2,3} + Y_{0,1,3} + Y_{0,2,3} + Y_{1,2,3} - Y_{0,3} - Y_{1,3} - Y_{2,3} + Y_3$$
$$Y_{0,1,2,3} = -T_{0,1,2,3} + T_{0,1,3} + T_{0,2,3} + T_{1,2,3} - T_{0,3} - T_{1,3} - T_{2,3} + T_3$$

In the expression of $T_{0,1,2,3}$ in terms of the Y terms (and vice versa), the subscripts of Y terms go though all the eight subsets of $\{0,1,2\}$ union 3, and the sign in front of the Y terms alternate according to the size of the subset. These facts turn out to be universal. We have:

Theorem 3.1. *For any n, there is*

$$Y_{0,1,2,\cdots,n} = \sum_{S \subset [n-1]} (-1)^{|S|} T_{S \cup n}$$

and

$$T_{0,1,2,\cdots,n} = \sum_{S \subset [n-1]} (-1)^{|S|} Y_{S \cup n},$$

where S goes over all subsets of $[n-1] = \{0,1,2,\cdots,n-1\}$, including the empty set, and $|S|$ is the size of S.

Proof. Because we will be applying the inclusion-exclusion principle later, to make the corresponding arguments, we revise the definition of the T's slightly. Assume a series of examinations are scheduled at times $0 \leq t_1 \leq t_2 \leq \cdots \leq t_n$ $(t_1 \geq 0)$ at ages a, $a+t_1$, $a+t_2, \cdots, a+t_n$, with no examinations at other times. However, assume that all the examinations are taken (again it is implied that one is still in either S_0 or S_p by the time t_n, *i.e.*, we limit to the cohort of such), and not all results are taken into consideration. Therefore even if S_p has

been detected at time t_i $(i < n)$, all the subsequent examinations will still be taken. Let $T_{i_1, i_2, \cdots, i_m}$ be the probability for the person be the probability for an individual undetected by the examinations at times $t_{i_1}, t_{i_2}, \cdots, t_{i_{m-1}}$, and detected in S_p at exam t_{i_m}, *regardless of what happens at other examinations*. Although the definition for T's has changed, the numerical expression of T's are the same as before.

Let

- C_i $(i < n)$ be the event of being detected at t_n but undetected at t_i;
- C_n be the event of being detected at t_n;
- A be the event of being detected at all exams at $t_1, t_2, \cdots, t_n$.

Therefore we have

$$C_n = \bigcup_{i=0,1,\cdots,n-1} C_i \cup A$$

and $Pr(C_n) = T_n$ (which is also equal to Y_n), while the size of the right-hand-side would be $Pr\left(\bigcup_{i=0,1,\cdots,n-1} C_i\right) + Pr(A)$, since A and each C_i are mutually exclusive events. We also have $Pr(A) = Y_{0,1,\cdots,n}$.

By inclusion-exclusion principle [14],

$$Pr\left(\bigcup_{i=0,1,\cdots,n-1} C_i\right) = \sum_{i=0,\cdots,n-1} Pr(C_i) - \sum_{0 \le i < j \le n-1} Pr(C_i \cap C_j) + \sum_{0 \le i < j < k \le n-1} Pr(C_i \cap C_j \cap C_k) + \cdots,$$

But $Pr\left(C_{j_1} \cap C_{j_2} \cap \cdots \cap C_{j_m}\right)$ is simply $T_{j_1, j_2, \cdots, j_m}$, therefore we have

$$T_n = Pr\left(\bigcup_{i=0,1,\cdots,n-1} C_i\right) + Y_{0,1,2,\cdots,n}$$

$$= \sum_{i=0,1,\cdots,n-1} T_{i,n} - \sum_{0 \le i < j \le n-1} T_{i,j,n} + \sum_{0 \le i < j < k \le n-1} T_{i,j,k,n} + \cdots + Y_{0,1,2,\cdots,n}$$

$$= \sum_{S \subsetneq [n-1]} (-1)^{|S|-1} T_{S \cup n} + Y_{0,1,2,\cdots,n}$$

Or, $Y_{0,1,2,\cdots,n} = \sum_{S \subset [n-1]} (-1)^{|S|} T_{S \cup n}$.

The second expression can be proved in a similar manner. With the same set up as the above, let

- D_i $(i < n)$ be the event of being detected by two examinations at t_i and at t_n;
- B be the event of being undetected by every examination taken at $t_1, t_2, \cdots, t_{n-1}$, but detected at t_n.

Recall that C_n is the subset of all those detected at t_n, then

$$C_n = \bigcup_{i=0,1,\cdots,n-1} D_i \cup B$$

and $|C_n| = T_n = Y_n$. While this time the size of the right-hand-size is $\left|\bigcup_{i=0,1,\cdots,n-1} D_i\right| + |B|$, since B and each D_i are mutually exclusive events. We also have $|B| = T_{0,1,\cdots,n}$.

Again by inclusion-exclusion principle,

$$Pr\left(\bigcup_{i=0,1,\cdots,n-1} D_i\right) = \sum_{i=0,\cdots,n-1} Pr(D_i) - \sum_{0 \le i < j \le n-1} Pr(D_i \cap D_j) + \sum_{0 \le i < j < k \le n-1} Pr(D_i \cap D_j \cap D_k) + \cdots,$$

However $Pr\left(D_{j_1} \cap D_{j_2} \cap \cdots \cap D_{j_m}\right)$ is simply $Y_{j_1, j_2, \cdots, j_m}$, therefore we have

$$T_n = Y_n = Pr\left(\bigcup_{i=0,1,\cdots,n-1} D_i\right) + T_{0,1,2,\cdots,n}$$

$$= \sum_{i=0,1,\cdots,n-1} Y_{i,n} - \sum_{0 \le i < j \le n-1} Y_{i,j,n} + \sum_{0 \le i < j < k \le n-1} Y_{i,j,k,n} + \cdots + T_{0,1,2,\cdots,n}$$

$$= \sum_{S \subsetneq [n-1]} (-1)^{|S|-1} Y_{S \cup n} + T_{0,1,2,\cdots,n}$$

Or, $T_{0,1,2,\cdots,n} = \sum_{S \subset [n-1]} (-1)^{|S|} Y_{S \cup n}$.

Corollary 3.2. *There are infinitely many expressions of age-specificity in terms of* T.

Proof. This is because each β_{a+t_i} $(i > 0)$ can be expressed as

$$\beta_{a+t_i} = \frac{Y_S}{Y_{S \setminus i}}$$

where S is any subset of $[n] = 0, 1, 2, \cdots, n$ that contains i, and $S \setminus i$ means S without i. For example, we saw in the last section that $\beta_{a+t_1} = \frac{X_a(t_2) \beta_{a+t_2} \beta_{a+t_1} \beta_a}{X_a(t_2) \beta_a \beta_{a+t_2}} = \frac{Y_{0,1,2}}{Y_{0,2}}$, and it is also equal to $\frac{Y_{1,2,3}}{Y_{2,3}} = \frac{Y_{1,4}}{Y_4} = \cdots$, etc.

Each Y can be expressed in terms of T's. □

Given expressions like Theorem 3.1, the proof for Theorem 2.1 is not as straightforward. However the purpose of that proof was to relate some of the traditional considerations and reasoning's to our new notions. This will be useful when evaluating important parameters such as age-dependent disease transition probabilities and sojourn time distribution.

4. An Example

We illustrate the proposed method using the data from the Health Insurance Plan Project, or HIP data [1]. HIP is a randomized screening trial of mammography screening vs. no screening for the women who did not have previous mammography. Even though HIP is a large-scale screening trial, because breast cancer incidence rate is relatively low, we do not have sufficient number of screen-detected and interval cases to readily estimate age-specific β. Thus for an illustration of the Theorem 2.1, we group all the women together as one age group, and assume that all patients had the same age at the initial examination. Because the cohort is required to have annual check-ups, we set $t_1 = 1$ and $t_2 = 2$ in the Theorem 2.1, and get

$$\beta = 1 - \frac{T_{1,2} - T_{0,1,2}}{T_2 - T_{0,2}}.$$

To calculate, saying $T_{0,1}$, we simply use the ratio between the number of cases detected at the first annual examination after the initial examination and all the cases who had taken both examinations. Using the HIP data, we get

$$\beta = 1 - \frac{\dfrac{32}{15936} - \dfrac{20}{13679}}{\dfrac{55}{20166} - \dfrac{1}{1085}} = 0.6977.$$

Note that this estimate is essentially identical to Shen and Zelen's published estimate of 0.7 ([10] [11]) based on the maximum likelihood method, but is computationally trivial.

Acknowledgements

References

[1] Shapiro, S., Venet, W., Strax, P., *et al.* (1988) Periodic Screening for Breast Cancer: The Health Insurance Plan Project and Its Sequelae, 1963-1986. The Johns Hopkins University Press, Baltimore.

[2] Tabar, L., Gunnar, F., Duffy, S.W., *et al.* (1992) Update of the Swedish Two-County Program of Mammographic Screening for Breast Cancer. *Radiologic Clinics of North America*, **30**, 187-210.

[3] Andersson, I., Aspegren, K., Janzon, L., *et al.* (1988) Mammographic Screening and Mortality from Breast Cancer: The Malmo, Mammographic Screening Trial. *British Medical Journal*, **297**, 943-948.
http://dx.doi.org/10.1136/bmj.297.6654.943

[4]   Frisell, J., Glas, U., Hellstrom, L., *et al.* (1986) Randomized Mammographic Screening for Breast Cancer in Stockholm. *Breast Cancer Research and Treatment*, **8**, 45-54. http://dx.doi.org/10.1007/BF01805924

[5]   Bjurstam, N., Bjorneld, L., Duffy, S.W., *et al.* (1997) The Gothenburg Breast Screening Trial: First Results on Mortality, Incidence, and Mode of Detection for Women Ages 39 - 49 Years at Randomization. *Cancer*, **80**, 2091-2099. http://dx.doi.org/10.1002/(SICI)1097-0142(19971201)80:11<2091::AID-CNCR8>3.0.CO;2-#

[6]   Roberts, M.M., Alexander, F.E., Anderson, T.J., *et al.* (1990) Edinburgh Trial of Screening for Breast Cancer: Mortality at Seven Years. *Lancet*, **335**, 241-246. http://dx.doi.org/10.1016/0140-6736(90)90066-E

[7]   Miller, A.B., Baines, C.J., To, T., *et al.* (1992) Canadian National Breast Screening Study 1: Breast Cancer Detection and Death Rates among Women Aged 40 to 49 Years. *Canadian Medical Association Journal*, **147**, 1459-1476.

[8]   Duffy, S.W., Chen, H., Tabar, L. and Day, N. (1995) Estimation of Mean Sojourn Time in Breast Cancer Screening Using a Markov Chain Model of Both Entry to and Exit from the Preclinical Detectable Phase. *Statistics in Medicine*, **14**, 1531-1553. http://dx.doi.org/10.1002/sim.4780141404

[9]   Paci, E. and Duffy, S.W. (1991) Modeling the Analysis of Breast Cancer Screening Programmes: Sensitivity, Lead Time and Predictive Value in the Florence District Programmes (1975-1986). *International Journal of Cancer*, **20**, 852-858.

[10]   Shen, Y. and Zelen, M. (1999) Parametric Estimation Procedures for Screening Programmes: Stable and Non-Stable Disease Models for Multimodality Case Finding. *Biometrika*, **86**, 503-515. http://dx.doi.org/10.1093/biomet/86.3.503

[11]   Shen, Y. and Zelen, M. (2001) Screening Sensitivity and Sojourn Time from Breast Cancer Early Detection Trials: Mammograms and Physical Examinations. *Journal of Clinical Oncology*, **19**, 3496-3499.

[12]   Tabar, L., Fagerberg, G., Chen, H., *et al.* (1995) Efficacy of Breast Cancer Screening by Age. *Cancer*, **75**, 2507-2517. http://dx.doi.org/10.1002/1097-0142(19950515)75:10<2507::AID-CNCR2820751017>3.0.CO;2-H

[13]   Zelen, M. (1993) Optimal Scheduling of Examinations for the Early Detection of Disease. *Biometrika*, **80**, 279-293. http://dx.doi.org/10.1093/biomet/80.2.279

[14]   Stanley, R.P. (1999) Enumerative Combinatorics, Vol. 1. Cambridge University Press, Cambridge. http://dx.doi.org/10.1017/CBO9780511609589

Scientific Research Publishing (SCIRP) is one of the largest Open Access journal publishers. It is currently publishing more than 200 open access, online, peer-reviewed journals covering a wide range of academic disciplines. SCIRP serves the worldwide academic communities and contributes to the progress and application of science with its publication.

Other selected journals from SCIRP are listed as below. Submit your manuscript to us via either submit@scirp.org or Online Submission Portal.