*Biostatistics* (2004), **5**, 4, *pp.* 603–613
doi: 10.1093/biostatistics/kxh012

# Overdiagnosis in early detection programs

ORI DAVIDOV[†]
*Department of Statistics, University of Haifa, Mount Carmel, Haifa 31905, Israel*
davidov@stat.haifa.ac.il

MARVIN ZELEN
*Department of Biostatistics, Harvard University and the Dana Farber Cancer Institute, Boston MA 02115, USA*
zelen@hsph.harvard.edu

SUMMARY

Overdiagnosis refers to the situation where a screening exam detects a disease that would have otherwise been undetected in a person's lifetime. The disease would have not have been diagnosed because the individual would have died of other causes prior to its clinical onset. Although the probability of overdiagnosis is an important quantity for understanding early detection programs it has not been rigorously studied. We analyze an idealized early detection program and derive the mathematical expression for the probability of overdiagnosis. The results are studied numerically for prostate cancer and applied to a variety of screening schedules. Our investigation indicates that the probability of overdiagnosis is remarkably high.

*Keywords*: Early detection programs; Overdiagnosis; Screening.

## 1. INTRODUCTION

Advances in diagnostic testing procedures have made screening and early detection programs for many chronic diseases widely available. In particular, testing for prostatic specific antigen (PSA) to detect prostate cancer is in routine use and an integral part of current medical practice. Overdiagnosis, also known as overdetection, refers to the situation where a screening exam detects disease that would have otherwise been undetected because the individual would have died of other causes prior to the onset of clinical disease. Although common in practice, the role of overdiagnosis and the costs associated with it are not widely appreciated. For example, overdiagnosis often complicates the evaluation of medical investigations of the benefit of early detection programs. This phenomenon is particularly important in retrospective studies on early detection where overdiagnosis results in misclassification of the exposure. The exposure in these settings is defined as an early detection exam (Davidov and Zelen, 2003). However, some of the cases diagnosed by an early detection program would have never developed the disease (i.e. overdiagnosis). Comparing them with cases diagnosed by routine care, which by definition are not overdiagnosed, may be misleading. In addition, overdiagnosis distorts our understanding of the official statistics of cancer because it changes the age-specific incident curves and affects the estimates of sensitivity and specificity (Black, 2000). Most importantly overdiagnosis may be regarded as the most serious side-effect of cancer

[†]To whom correspondence should be addressed.

screening because it means that individuals who would have never developed the disease, are informed of a distressing diagnosis, and receive treatment with possible negative affects on the quality and duration of their residual life. This phenomen is often called over-treatment.

Overdiagnosis has been recognized as a potential problem for prostate cancer (Zappa *et al.*, 1998; Bostwick and Chang 1999; Ciatto *et al.*, 2000). It is of particular interest in diseases with long lead-times. Empirical and simulation-based estimators of the magnitude of overdiagnosis of prostate cancer are given by McGregor *et al.* (1998), Etzioni *et al.* (2001) and Draisma *et al.* (2003). They report an overdiagnosis rate of roughly 30–50%, depending on race and additional inputs and report that the overdiagnosis rate increases with age. The implication for a cohort of individuals can be quantified given the prevalence and intensity of screening in a population (Pinsky, 2001).

In this communication we evaluate the probability of overdiagnosis using a mathematical model for screening. The probability of overdiagnosis is calculated for individuals as well as for a general early detection program. The paper is organized in the following way. In Section 2 we describe our notation and derive the probability of overdiagnosis. The results are investigated numerically for prostate cancer in Section 3. Section 4 contains a summary and discussion.

## 2. THE PROBABILITY OF OVERDIAGNOSIS

### 2.1 *Preliminaries*

Consider an idealized model for the natural history of disease. Suppose that an individual can be in one of three states which we denote by $S_h$, $S_p$ and $S_c$. An individual in the healthy state ($S_h$) is either disease-free, or has disease which cannot be detected. An individual in the pre-clinical state ($S_p$) has asymptomatic disease which may be detected by a special exam. An individual in the clinical state ($S_c$) has been diagnosed with disease through usual medical care, i.e. the disease is symptomatic and the individual seeks medical attention. The natural history of disease is assumed to be progressive and is represented by the sequence $S_h \rightarrow S_p \rightarrow S_c$. Let $q_s(x)$ denote the sojourn time probability density function in the $s$th state where $s = h, p, c$. The corresponding survival functions are $Q_s(x)$. We denote by $\tau$ the screening schedule, which refers to the age(s) at which an individual has had early detection exams. If multiple exams are administered then we label them by $\tau = (\tau_1, \ldots, \tau_m)$. The probability of detecting the disease conditional on being in $S_p$ is denoted by $\beta$. In practice the sensitivity may be a random variable with a distribution centered at $\beta$. Finally, let $T$ denote the time-to-death for all other causes. Its survival function is denoted $R(t) = P[T > t]$.

The fundamental idea of overdiagnosis is that the screening exam diagnoses the disease at an earlier age than it would have been diagnosed under usual care. In addition, the individual would have died of other causes before the clinical onset of the disease. Thus overdiagnosis may be viewed as a competing risk problem. Our development emphasizes the program aspect of early detection programs followed by the implications in some special cases including the single exam. Note that the model for a single exam compares the forward recurrence time (the time from early diagnosis to when the disease would have been clinically diagnosed) with the residual life time, both measured from age $\tau$. When the residual life time is shorter than the forward recurrence time the individual will be overdiagnosed.

We introduce two measures of overdiagnosis which we call individual overdiagnosis and schedule overdiagnosis. Schedule overdiagnosis reflects the overall overdiagnosis rate associated with a particular screening schedule. Individual overdiagnosis reflects the risk of being overdiagnosed in an upcoming exam given an individual's screening history. These quantities are related. In fact they coincide if a single early detection exam is administered.

### 2.2 *Schedule overdiagnosis*

Consider a screening schedule $\tau = (\tau_1, \ldots, \tau_m)$ with $m$ exams. Let $\tau_0 = 0$. An individual belongs to the $i$th generation $i = 1, \ldots, m$, if he/she enters $S_p$ in the interval $(\tau_{i-1}, \tau_i]$. Let $G_i$ be the event of an $i$th generation individual being in the pre-clinical state at age $\tau_i$. Its probability is denoted by $P_i = P(G_i)$. Let $V_i$ denote the forward recurrence time in the pre-clinical state for the $i$th generation. We denote its probability density function by $f_i(v) = f(v|\tau_i)$. Note that the forward recurrence time is the remaining sojourn time in $S_p$ conditional on being in $S_p$ at age $\tau_i$. It is well known (e.g. Zelen, 1993), that

$$P_i = \int_{\tau_{i-1}}^{\tau_i} q_h(u)\, Q_p(\tau_i - u)\, \mathrm{d}u$$

and

$$f_i(v) = \frac{\int_{\tau_{i-1}}^{\tau_i} q_h(u)\, q_p(\tau_i + v - u)\, \mathrm{d}u}{\int_{\tau_{i-1}}^{\tau_i} q_h(u)\, Q_p(\tau_i - u)\, \mathrm{d}u}, \text{ for } v \geqslant 0.$$

For $j \geqslant i$, let the random variable $V_{j|i}$ denote the conditional forward recurrence time of an $i$th generation individual given that he/she is in $S_p$ at age $\tau_j$. More formally,

$$\begin{aligned} f_{j|i}(v)\, \mathrm{d}v &= P\left[V_i \in (\tau_j - \tau_i + v, \tau_j - \tau_i + v + dv) \,|\, V_i \geqslant \tau_j - \tau_i\right] \\ &= \frac{f_i(\tau_j - \tau_i + v)}{\overline{F}_i(\tau_j - \tau_i)}\, \mathrm{d}v, \text{ for } v \geqslant 0, \end{aligned}$$

where $\overline{F}_i(v) = \int_v^\infty f_i(t)\, \mathrm{d}t$ is the tail probability associated with the $i$th forward recurrence time. Define the random variable $T_j$ to be the residual length of life measured from age $\tau_j$. Its tail probability is given by

$$R_j(t) = R(t|\tau_j) = P\left[T > \tau_j + t | T > \tau_j\right] = \frac{R(\tau_j + t)}{R(\tau_j)}.$$

The conditional probability that an $i$th generation individual who is diagnosed at age $\tau_j$ develops disease before the individual dies is given by

$$P\left[V_{j|i} < T_j\right] = \int_0^\infty f_{j|i}(v)\, R_j(v)\, \mathrm{d}v = \frac{A_{ij}}{B_{ij}} \tag{2.1}$$

where

$$\begin{aligned} A_{ij} &= \int_0^\infty R(\tau_j + v) \int_{\tau_{i-1}}^{\tau_i} q_h(u)\, q_p(\tau_j + v - u)\, \mathrm{d}u\, \mathrm{d}v, \\ B_{ij} &= \int_{\tau_j - \tau_i}^\infty R(\tau_j) \int_{\tau_{i-1}}^{\tau_i} q_h(u)\, q_p(\tau_i + v - u)\, \mathrm{d}u\, \mathrm{d}v. \end{aligned}$$

Consequently the conditional probability of overdiagnosis for an $i$th generation individual who is diagnosed at age $\tau_j$, denoted by $\omega_{ij}$, is

$$\omega_{ij} = P\left[V_{j|i} > T_j\right] = \frac{B_{ij} - A_{ij}}{B_{ij}}.$$

Consider an individual who is diagnosed at age $\tau_j$. The probability that he/she belongs to the $i$th generation is given by $P\left[G_i|D_j\right]$ where $D_j$ is the event of being diagnosed at age $\tau_j$. Hence the probability of being overdiagnosed, conditional on being diagnosed at age $\tau_j$, is simply

$$\sum_{i=1}^{j} \omega_{ij} P\left[G_i|D_j\right].$$

The probability of being diagnosed at $\tau_j$ conditional on being diagnosed in any of the early detection exams is $P\left[D_j|D\right]$ where $D = \cup_{j=1}^{m} D_j$ is the event of being diagnosed. Consequently, the probability of overdiagnosis for a screening schedule $\tau$, which we denote $\Omega_S$, is

$$\Omega_S = \sum_{j=1}^{m} \sum_{i=1}^{j} \omega_{ij} \lambda_{ij}, \tag{2.2}$$

where

$$\lambda_{ij} = P\left[G_i|D_j\right] P\left[D_j|D\right] = \frac{P\left[G_i, D_j\right]}{P\left[D\right]}. \tag{2.3}$$

Finally we express the RHS of (2.3) in terms of the natural history of the disease, the screening schedule and the exam sensitivity. Recall that an $i$th generation individual diagnosed at $\tau_j$ entered $S_p$ in the $i$th interval, had a forward recurrence time longer than $\tau_j - \tau_i$, was missed on $j - i$ early detection exams and detected at $\tau_j$. Hence,

$$P\left[G_i, D_j\right] = P_i \overline{F}_i\left(\tau_j - \tau_i\right) \beta \left(1 - \beta\right)^{j-i}.$$

The events $D_j$ are mutually exclusive, therefore

$$P\left[D\right] = \sum_{j=1}^{m} P\left[D_j\right] = \sum_{j=1}^{m} \sum_{i=1}^{j} P\left[G_i, D_j\right] = \sum_{j=1}^{m} \sum_{i=1}^{j} P_i \overline{F}_i\left(\tau_j - \tau_i\right) \beta \left(1 - \beta\right)^{j-i}.$$

Substituting (2.1) and (2.3) in (2.2) we obtain

$$\Omega_S = \frac{\beta}{P\left[D\right]} \sum_{j=1}^{m} \sum_{i=1}^{j} \frac{B_{ij} - A_{ij}}{R\left(\tau_j\right)} \left(1 - \beta\right)^{j-i},$$

which is the formula for the probability of schedule overdiagnosis. The above equation is easily generalized to include the situation where $\beta$ may depend on age.

2.2.1 *Some special cases.* Consider the case of a single screening exam administered at age $\tau$. Using the formulae in the previous section it is easy to see that the probability of overdiagnosis reduces to

$$\frac{B_\tau - A_\tau}{B_\tau}$$

where

$$A_\tau = \int_0^\infty R\left(\tau + v\right) \int_0^\tau q_h\left(u\right) q_p\left(v + \tau - u\right) \, \mathrm{d}u \, \mathrm{d}v, \tag{2.4}$$

$$B_\tau = \int_0^\infty R\left(\tau\right) \int_0^\tau q_h\left(u\right) q_p\left(v + \tau - u\right) \, \mathrm{d}u \, \mathrm{d}v. \tag{2.5}$$

We note that the situation where a single exam is administered at age $\tau$ is fundamental and helps clarify the more general case. We therefore briefly repeat the derivation of the formula for this case to bring out the role of the forward recurrence time. Let $P(\tau)$ denote the probability that the individual is in $S_p$ at age $\tau$. The joint probability of being in $S_p$ at age $\tau$ and remaining in $S_p$ for an additional $v$ units of time is

$$P(\tau) f(v|\tau) = \int_0^\tau q_h(u) q_p(v + \tau - u) \, \mathrm{d}u$$

where $u \leqslant \tau$ is the age of entry into $S_p$ and $f(v|\tau)$ is the probability density function for the forward recurrence time conditional on being diagnosed at age $\tau$. Note that the lead time gained by early diagnosis is the forward recurrence time. We denote this lead time by $V_\tau$. Define $T_\tau$ to be the residual survival measured from age $\tau$; its tail probability function (as a function of $v$) is given by $R(\tau + v)/R(\tau)$. Therefore the probability of being overdiagnosed, conditional on being diagnosed at age $\tau$, is the probability that the forward recurrence time is greater than the residual survival,

$$P(V_\tau > T_\tau) = \int_0^\infty \left(1 - \frac{R(\tau + v)}{R(\tau)}\right) f(v|\tau) \, \mathrm{d}v.$$

A short calculation shows that the expressions above agree. Note that this quantity is independent of $\beta$, the sensitivity of the early detection exam.

More generally, if the early detection exam has unit sensitivity, then (2.2) simplifies considerably and is equal to

$$\sum_{i=1}^m P_i \omega_i / \sum_{i=1}^m P_i,$$

where $\omega_i \equiv \omega_{ii} = \left(B_{\tau_i} - A_{\tau_i}\right)/B_{\tau_i}$ and $A_{\tau_i}$ and $B_{\tau_i}$ are given by substituting $\tau_i$ in (2.4) and (2.5) respectively.

Another interesting case arises when $q_h(u) = h_i$ for all $u \in (\tau_{i-1}, \tau_i]$, i.e. the density of the transition from $S_h$ to $S_p$ is piecewise constant. This assumption is reasonable when the inter-exam intervals are relatively short compared to the overall age. With this assumption (2.1) simplifies to

$$\omega_{ij} = 1 - \frac{\int_0^\infty R(\tau_j + v)\left[Q_p(\tau_j - \tau_i + v) - Q_p(\tau_j - \tau_{i-1} + v)\right] \mathrm{d}v}{\mu R(\tau_j)\left[\overline{F}(\tau_j - \tau_i) - \overline{F}(\tau_j - \tau_{i-1})\right]}$$

where $\overline{F}(v) = \int_v^\infty Q_p(t)/\mu \, \mathrm{d}t$ and $\mu = \int_0^\infty Q_p(t) \, \mathrm{d}t$. Note that $\overline{F}(v)$ is the tail area of the forward recurrence time obtained in the so-called *steady state* (e.g. Zelen, 1993; Davidov, 1999), and $\mu$ is the mean sojourn time in the clinical state. Moreover,

$$P_i = \mu h_i \left[1 - \overline{F}(\tau_i - \tau_{i-1})\right],$$
$$\overline{F}_i(\tau_j - \tau_i) = \frac{\overline{F}(\tau_j - \tau_i) - \overline{F}(\tau_j - \tau_{i-1})}{1 - \overline{F}(\tau_i - \tau_{i-1})},$$

which further simplifies the quantities $\lambda_{ij}$. Further simplifications are obtained under the well known *stable disease model* for which the incidence and the prevalence of the disease are constant over time implying that $h_i = h$ and $P_i = P$ for all $i$. In this situation the forward recurrence time distributions are all the same, i.e. $\overline{F}_i(v) = \overline{F}(v) = \int_v^\infty Q_p(t)/\mu \, \mathrm{d}t$.

### 2.3 *Individual overdiagnosis*

Let $\tau = (\tau_1, \ldots, \tau_m)$ denote the screening history of an individual and let $\overline{\tau}$ denote the age of the next scheduled exam. Clearly $\overline{\tau} > \tau_m$ and all $m$ exams have been negative. Let $\overline{T}$ denote the residual length of life measured from $\overline{\tau}$. Let $\overline{V}_i$ denote the conditional forward recurrence time for an $i$th generation individual that has not entered $S_c$ by age $\overline{\tau}$. Thus $\overline{\omega}_i = P\left[\overline{V}_i > \overline{T}\right]$ is the conditional probability of overdiagnosis for an $i$th generation individual who is diagnosed at age $\overline{\tau}$. Denote by $\overline{\lambda}_i$ the probability that an individual who is diagnosed at $\overline{\tau}$ belongs to the $i$th generation. Thus,

$$\overline{\lambda}_i = P\left[G_i, \overline{D} | \overline{D}\right] = \frac{P\left[G_i, \overline{D}\right]}{P\left[\overline{D}\right]} = \frac{P_i \overline{F}_i\left(\overline{\tau} - \tau_i\right)(1-\beta)^{m+1-i}}{\sum_{i=1}^{m+1} P_i \overline{F}_i\left(\overline{\tau} - \tau_i\right)(1-\beta)^{m+1-i}},$$

where $\overline{D}$ is the event of being diagnosed at $\overline{\tau}$. Note that the definitions of $\overline{\omega}_i$ and $\overline{\lambda}_i$ mimic those of $\omega_{ij}$ and $\lambda_{ij}$. Consequently the probability of individual overdiagnosis, denoted $\Omega_I$, is

$$\Omega_I = \sum_{i=1}^{m+1} \overline{\omega}_{i+1} \overline{\lambda}_{i+1}. \tag{2.6}$$

Clearly, in the absence of a screening history, individual overdiagnosis coincides with schedule overdiagnosis with one exam. Furthermore, if the sensitivity of the screening exam is unity, individual overdiagnosis reflects the overdiagnosis rate for generation $m + 1$.

## 3. APPLICATION TO PROSTATE CANCER

In this section we calculate the probability of overdiagnosis for hypothetical early detection programs for prostate cancer. We assume that the screened population is a random sample from the general population. Let $p_k$ for $k = 1, 2, \ldots$ be the probability that an individual of age $k$ dies in the age interval $(k, k+1]$. We obtain the quantities $p_k$ from a Period Life Table published by the Social Security Administration describing the mortality of US males in 1997. It immediately follows that the survival probabilities $R_k$ may be calculated as

$$R_k = P\left[T > k\right] = \prod_{i \leqslant k} \left(1 - p_i\right).$$

At non-integer values the survival function $R(t)$ is computed by simple linear interpolation,

$$R(t) = R_{k-1} + (t-k)\left(R_k - R_{k-1}\right) I_{(k-1,k]}(t)$$

for $t \leqslant 120$ and $R(t) = 0$ otherwise. We assume that the sojourn time distribution in the pre-clinical state follows an exponential distribution with mean $\mu$. Although there is some biological motivation for choosing an exponential form (i.e. exponential rate of doubling time for tumor cells) the main reason for choosing the exponential distribution is for convenience. Note that the pre-clinical sojourn time cannot be directly observed; moreover, to our knowledge there are no studies indicating its form. The exponential assumption implies that the time gained by early diagnosis by a screening exam is the same as the mean sojourn time in the pre-clinical state. The estimation of the pre-clinical mean sojourn time for prostate cancer and the lead time has been of considerable interest (e.g. Whittemore *et al.*, 1991, 1995; Pearson and Carter, 1994; Stenman *et al.*, 1994; Gann *et al.*, 1995; Etzioni *et al.*, 1998; Hugosson *et al.*, 2000; Draisma *et al.*, 2003). Values ranging from five to 15 years have been reported depending on the method and the data used. In our calculations we consider the values $\mu = 5, 7.5, 10, 12.5$ and 15 years. Note that

the assumption of an exponential sojourn time distribution in the pre-clinical state somewhat simplifies the calculations as $f_i(\cdot)$ and $f_{j|i}(\cdot)$ reduce to the exponential distribution $q_p(\cdot)$.

Calculating the probability of overdiagnosis requires the knowledge of $q_h(v)$, the sojourn time distribution in the healthy state, which cannot be observed directly. However, data on age-specific incidence of invasive prostate cancer are available from the SEER data base collected by the National Cancer Institute. Let $I(t)\,\mathrm{d}t$ be the probability of developing prostate cancer in the age interval $(t, t+\mathrm{d}t)$. Note that $I(t)$ is the point incidence function. Clearly,

$$I(t) = \int_0^t q_h(x)\, q_p(t-x)\,\mathrm{d}x.$$

In the SEER data base the incidence is grouped into 5-year age intervals. Denote by $I_k$ the observed incidence in the age group $[5(k-1), 5k]$. Then we can write

$$I_k = \int_{t \in J_K} I_k(t)\,\mathrm{d}t, \;\; J_k = [5(k-1), 5k] \tag{3.1}$$

where $I_k(t)$ is the value of the incidence function on the $k$th interval. It is impossible to deduce uniquely the function $q_h$ from equations (3.1) without further assumptions. In fact this is an example of an ill-posed inverse problem. Mezzetti and Robertson (1999) suggest a Bayesian approach to this problem. Here we follow Lee and Zelen (1998) and model the sojourn time in $S_h$ as a piecewise constant function on the intervals $J_k$, hence $q_h(x) = \sum_j h_j I_{\{t_{k-1} < x \leqslant t_k\}}$. This is the simplest possible model. Let $I_k(t)$ be the value of $I(t)$ on the $k$th interval. It is seen that

$$I_k(t) = \sum_{k=1}^{k-1} h_k \left[Q_p(t-t_k) - Q_p(t-t_{k-1})\right] + h_k \left[1 - Q_p(t-t_{k-1})\right]. \tag{3.2}$$

Using (3.2), an algebraic expression for (3.1) may be derived. The constants $h_j$ are the solution to a system of linear equations that can be solved in a recursive manner. Note that more complicated models for the sojourn time distribution in $S_h$ are also a possibility, for example a piecewise linear and continuous model for $q_h$. Our calculations used the SEER prostate cancer incidence data for the years 1993–97.

Figure 1 shows the probability of overdiagnosis for a single scheduled exam for various mean lead times. Recall that the mean lead time and the mean sojourn time are the same for an exponential distribution. The probability of overdiagnosis increases with age and with the pre-clinical mean sojourn time. In fact our calculations suggest that the probability of overdiagnosis is high and in the range of 20–40% for most realistic mean values and ages of screening. This result is in general agreement with the clinical reports of a prevalence as high as 30% for indolent prostatic cancer in autopsy data.

Table 1 presents the overdiagnosis rate for three different multiple exam schedules. In the first schedule, exams are administered at the ages $\tau = (50, 55, 60)$. In the subsequent schedules a decade of exams at five year intervals is added. The probability of overdiagnosis is computed for five values of $\mu$ and three sensitivity levels, low, medium and high, i.e. 0.3, 0.7 and 0.9. Recall that the schedule overdiagnosis rate, $\Omega_S$, reflects the probability of overdiagnosis for a given schedule. Our calculations show that overdiagnosis is a major problem in early detection programs for prostate cancer (or other diseases with long pre-clinical duration). Populations in which individuals are repeatedly screened will have a high proportion of individuals likely of being overdiagnosed and consequently overtreated. For example if $\mu = 10$ and individuals are screened for prostate cancer every five years from age 50 through 80, then just over 40% of the individuals diagnosed will be overdiagnosed. This is an exceedingly high figure. Our calculations show that the probability of overdiagnosis is not dependent on the sensitivity of the screening test. Note that the probability of overdiagnosis increases with $\mu$ and the number of exams.

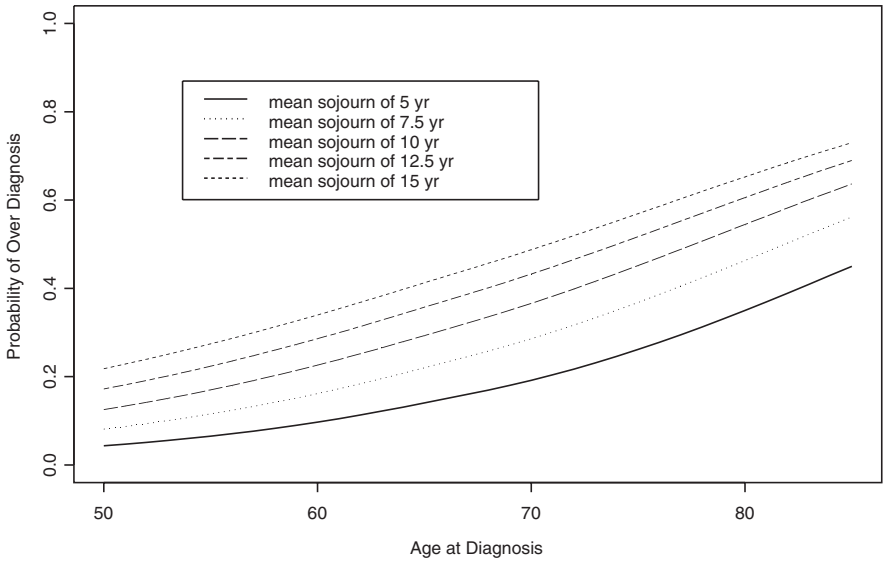


Fig. 1. The probability of overdiagnosis as a function of age and the mean sojourn time in the preclinical state for a single exam.

Table 1. *Overdiagnosis probabilities* ($\times 100\%$) *for prostate cancer as a function of the mean preclinical sojourn time, the sensitivity and the exam schedule*

| Mean Sojourn Time ($m$) | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 5 | | | 7.5 | | | 10 | | | 12.5 | | | 15 | | | |
| Sensitivity ($\beta$) | | | Sensitivity ($\beta$) | | | Sensitivity ($\beta$) | | | Sensitivity ($\beta$) | | | Sensitivity ($\beta$) | | | |
| 0.3 | 0.7 | 0.9 | 0.3 | 0.7 | 0.9 | 0.3 | 0.7 | 0.9 | 0.3 | 0.7 | 0.9 | 0.3 | 0.7 | 0.9 | |
| Exam Schedule $\tau = (50, 55, 60)$ | | | | | | | | | | | | | | | |
| 8.52 | 8.50 | 8.48 | 14.4 | 14.4 | 14.4 | 20.6 | 20.6 | 20.4 | 26.3 | 26.2 | 26.1 | 31.5 | 31.4 | 31.4 | |
| Exam Schedule $\tau = (50, 55, 60, 65, 70)$ | | | | | | | | | | | | | | | |
| 15.2 | 15.1 | 15.1 | 23.5 | 23.4 | 23.3 | 31.0 | 30.7 | 30.6 | 37.4 | 37.1 | 37.0 | 42.8 | 42.6 | 42.5 | |
| Exam Schedule $\tau = (50, 55, 60, 65, 70, 75, 80)$ | | | | | | | | | | | | | | | |
| 23.6 | 23.3 | 23.0 | 33.6 | 33.0 | 32.7 | 41.7 | 41.0 | 40.6 | 48.3 | 47.4 | 47.0 | 53.6 | 52.7 | 52.2 | |

Figure 2 plots the individual overdiagnosis rate. At each age $\tau$ we condition on the past screening history. The curves are computed assuming a sensitivity of $\beta = 0.9$. The results indicate that individual overdiagnosis rates are very high. For example if $\mu = 10$ then conditional on being screened every five years an individual who is diagnosed at age 80 has a 54% chance of being overdiagnosed! In fact the individual overdiagnosis probabilities are very close to the probability of overdiagnosis for a single exam. This observation implies that the probability of overdiagnosis is independent of the previous history of exams.

## 4. SUMMARY AND DISCUSSION

Overdiagnosis of a disease is defined as the diagnosis of an asymptomatic disease having no signs or symptoms, which would have never become symptomatic during an individual's remaining life time.

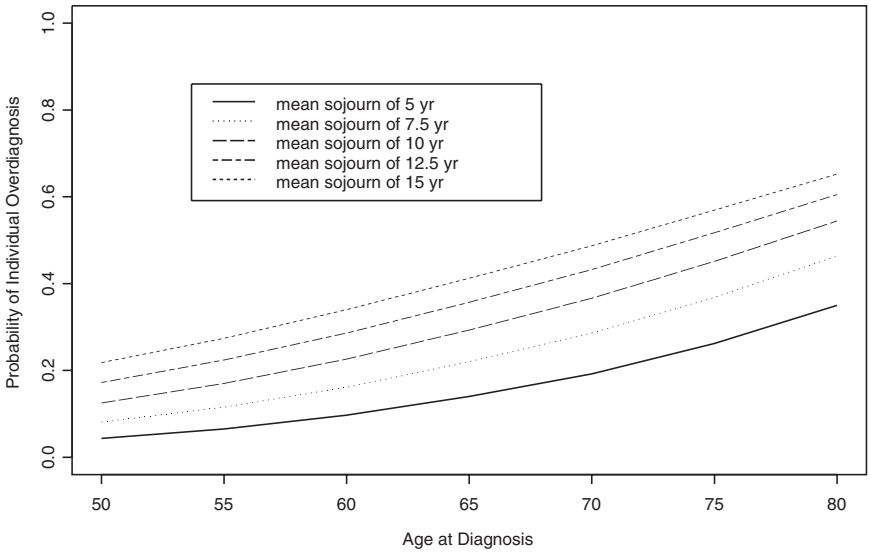


Fig. 2. The probability of individual overdiagnosis as a function of age, screening history and the mean sojourn time in the preclinical state. The screening history is assumed to have started at age 50 and further exams are given every five years. The exam sensitivity is 90%.

Ordinarily the early diagnosis would be made using a special case-finding modality (a screening test) which is capable of leading to the diagnosis of the disease. In this paper we have derived the basic equations for the probability of overdiagnosis. Two cases are envisioned. One is the overdiagnosis of an individual conditional on the age of diagnosis. The other is the overdiagnosis associated with a program of periodic screening exams.

The fundamental idea of overdiagnosis is that the special examination diagnoses the disease at an earlier age than it would have been diagnosed under usual care. Ordinarily, a diagnosis under usual care is made when the disease manifests signs or symptoms leading the individual to seek medical attention. The difference in the two ages is often called the lead time. The lead time is a function of the pre-clinical sojourn time distribution and possibly other covariates—principally age. Overdiagnosis occurs when the lead time is greater than the residual survival conditional on the age of early diagnosis. The probability of this event is independent of the examination sensitivity as the individual has already been diagnosed. However, if we consider a program of periodic examinations, the probability of overdiagnosis associated with the program depends only to a minor extent on the examination sensitivity.

Overdiagnosis of prostate cancer is one of the most important issues in the early diagnosis of any cancer. The mean sojourn time is believed to be long and, since the disease mainly affects older men, the residual survival is relatively short. The American Cancer Society recommends that the prostate specific antigen test (PSA) and digital rectal exam (DRE) should be offered annually beginning at age 50 to men having a life expectancy of at least 10 years. Current life tables indicate that male life expectancy drops below ten years at age 75. Our calculations show that the prostate overdiagnosis rate of 33% is associated with annual screening for men between the ages of 50–75 assuming a mean pre-clinical sojourn time of 10 years and exam sensitivity of 0.9. Note that this is only slightly higher than the overdiagnosis rate associated with an exam taken every five years (see Table 1). Men at higher risk, including those of African decent and with a first degree relative diagnosed at a younger age should begin prostate testing at age 45. High-risk men (those with multiple first-degree relatives with the disease) should begin at age 40; see Smith *et al.* (2003). However, these recommendations are not widely carried out. The National

Health Interview Survey for 2000 estimates that 34% of men between 50–64 and 51% of men over 65 had a PSA test within the past year. The rate of PSA testing varies. There is a distinct trend for increased testing among men with higher family incomes; see Swan *et al.* (2003).

Our numerical results depend on the distribution of the lead time. The paper by Draisma *et al.* (2003) used a simulation model (called Miscan) to estimate the mean lead time based on partial results from the European Randomized Study of Screening for Prostate Cancer (ERSPC). They reported that the mean lead time decreases with age from 12.3 years at age 55 to 6.0 years at age 75. This trend is somewhat surprising as for many cancers the mean lead time is expected to increase with age or stay constant (it is generally thought that cancers develop more slowly as age increases). They report overdetection of 27% for men at age 55 with a mean lead time of 12.3 years. Our calculations show a value of 21.9% for the same mean lead time. Other estimates of the mean lead time have been obtained from stored blood samples of healthy men who eventually develop prostate cancer. These lead times range from 5.5–9.2 years and depend on the age of the cohort group as well as the hypothetical threshold level of the PSA that indicates prostate cancer. However, these estimates may seriously underestimate the lead time; the statistical estimation problem is complicated as it is necessary to account for the possible future incidence of prostate cancer for men who are still free of cancer at the last follow-up as well as those who die of other causes who could have developed prostate cancer. There are other methods for estimating the mean lead time which rely on the increased incidence of prostate cancer due to earlier diagnosis: see Etzioni *et al.* (2002).

Regardless of the value of the lead time, the relatively high probability for the overdiagnosis of prostate cancer raises important issues of whether to treat a disease which has a significant probability of overdiagnosis. This is particularly important for prostate cancer as treatment may reduce the quality of life. The principal side effects of prostatectomy are impotence and incontinence. There is no scientific evidence available which demonstrates whether or not screening for prostate cancer will lower mortality. There are currently two randomized clinical trials addressing this issue, the European Randomized Study of Screening for Prostate Cancer (de Koning *et al.*, 2002) and the National Cancer Institute's trial on prostate screening (Gohagan *et al.*, 2000). However, to date there have been no reported results.

Finally, we note that it is an open issue of how one should design early detection programs which may result in a high probability of overdiagnosis. Depending on the behavior of individuals, two people presented with the same overdiagnosis probability may take different actions.

## REFERENCES

BLACK, W. C. (2000). Overdiagnosis: an under recognized cause of confusion and harm in cancer screening. *Journal of the National Cancer Institute* **92**, 1280–1281.

BOSTWICK, D. G. AND CHANG, L. (1999). Overdiagnosis of prostatic adenocarcinoma. *Semin Urological Oncology* **17**, 199–205.

CIATTO, S., ZAPPA, M., BONARDI, R. AND GERVASI, G. (2000). Prostate cancer screening, the problem of overdiagnosis and lessons to be learned from breast cancer screening. *European Journal of Cancer* **39**, 1347–1350.

DAVIDOV, O. (1999). The steady state probabilities for regenerative semi Markov processes with application to prevention and screening. *Applied Stochastic Models* **15**, 55–63.

DAVIDOV, O. AND ZELEN, M. (2003). The theory of case-control studies for early detection programs. *Biostatistics* **4**, 411–421.

DE KONING, H. J., LIEM, M. K., BAAN, C. A., BOER, R., SCHRODER, F. H. AND ALEXANDER, F. E. (2002). Prostate cancer mortality reduction by screening: power and time frame with complete enrollment in the European Randomized Screening for Prostate Cancer (ERSPC) trial. *International Journal of Cancer* **98**, 268–273.

ETZIONI, R., CHA, R., FEUER, E. J. AND DAVIDOV, O. (1998). Asymptomatic incidence and duration in prostate cancer. *American Journal of Epidemiology* **148**, 775–785.

ETZIONI, R., PENSON, D. F., LEGLER, J. M., DI TOMMASO, D., BOER, R., GANN, P. H. AND FEUER, E. J. (2002). Overdiagnosis due to prostate-specific antigen screening: lessons from US prostate cancer incidence trends. *Journal of the National Cancer Institute* **94**, 981–990.

GANN, P. H., HENNEKENS, C. H. AND STAMPFER, M. J. (1995). A prospective evaluation of plasma prostate specific antigen for detection of prostatic cancer. *Journal of the American Medical Association* **273**, 289–294.

GOHAGAN, J. K., PROROK, P. C., HAYES, R. B., KRAMER, B. S. AND PROSTATE, LUNG, COLORECTAL AND OVARIAN CANCER SCREENING TRIAL PROJECT TEAM (2000). The Prostate, Lung, Colorectal and Ovarian (PLCO) Cancer Screening Trial of the National Cancer Institute: history, organization, and status. *Control Clinical Trials* **21**, 251S–272S.

HUGOSSON, J., AUS, G., CARLSSON, S., ERIKSSON, H., LILJA, H. AND LODDING, P. (2000). Would prostate cancer detected by screening with prostate specific antigen develop into clinical cancer if left undiagnosed? *British Journal of Urology* **85**, 1078–1084.

LEE, S. J. AND ZELEN, M. (1998). Scheduling periodic examinations for the early detection of disease: applications to breast cancer. *Journal of the American Statistical Association* **93**, 1271–1281.

MEZZETTI, M. AND ROBERTSON, C. (1999). A hierarchical Bayesian approach to age-specific back calculation of cancer incidence rates. *Statistics in Medicine* **18**, 919–933.

MCGREGOR, M., HANLEY, J. A., BOIVIN, J. F. AND MCLEAN, R. G. (1998). Screening for prostate cancer estimating the magnitude of overdetection. *Canadian Medical Association Journal* **159**, 1368–1372.

PARMIGIANI, G., SKATES, S. AND ZELEN, M. (2002). Modelling and optimization in early detection programs with a single exam. *Biometrics* **58**, 30–36.

PEARSON, J. D. AND CARTER, H. B. (1994). Natural history of changes in prostate specific antigen in early stage prostate cancer. *Journal of Urology* **154**, 1743–1748.

PINSKY, P. F. (2001). Estimation and prediction for cancer screening models using deconvolutions and smoothing. *Biometrics* **57**, 389–395.

STENMAN, H., HAKAM, M., KNEKT, P., AROMAA, A., TEPPO, L. AND LEINONEN, J. (1994). Serum concentrations of prostate specific antigen and its complex with alpha 1-antichymotrypsin before diagnosis of prostate cancer. *The Lancet* **344**, 1594–1598.

SMITH, R. A., COKKINIDES, V. AND EYRE, H. J. (2003). American Cancer Society guidelines for the early detection of cancer. *CA Cancer Journal Clinical* **53**, 27–43.

SWAN, J., BREEN, N., COATES, R. J., RIEMER, B. K. AND LEE, N. C. (2003). Progress in cancer screening practices in the United States. Results from the 2000 National Health Interview Survey. *Cancer* **97**, 1528–1540.

WHITTEMORE, A. S., KELLER, J. B. AND BETENSKY, R. (1991). Low grade latent prostate cancer volume: predictor of cancer incidence? *Journal of the National Cancer Institute* **83**, 1231–1235.

WHITTEMORE, A. S., LELE, C. AND FRIEDMAN, G. D. (1995). Prostate specific antigen as a predictor of prostate cancer in black men and white men. *Journal of the National Cancer Institute* **87**, 345–360.

ZAPPA, M., CIATTO, S., BONARDI, R. AND MAZZOTTA, A. (1998). Overdiagnosis of prostate carcinoma: an estimate based on the results of the Florence screening pilot study. *Annals of Oncology* **9**, 1297–1300.

ZELEN, M. (1993). Optimal scheduling of examinations for the early detection of disease. *Biometrika* **80**, 279–293.

[*Received September 15, 2003; revised February 22, 2004; accepted for publication March 18, 2004*]