

A Conditional Approach to Measure Mortality Reductions Due to Cancer Screening

Zhihui (Amy) Liu^{1,2}, James A. Hanley³, Olli Saarela² and Nandini Dendukuri³

¹Cancer Care Ontario, Toronto, Canada

E-mail: amy.liu@cancercare.on.ca

²Dalla Lana School of Public Health, University of Toronto, Toronto, Canada

³Department of Epidemiology, Biostatistics and Occupational Health, McGill University, Montreal, Canada

Summary

The prevailing lack of consensus about the comparative harms and benefits of cancer screening stems, in part, from the inappropriate calculations of the expected mortality impact of a sustained screening programme. There is an inherent, and often substantial, time lag from the time of screening until the resulting mortality reductions begin, reach their maximum and ultimately end. However, the cumulative mortality reduction reported in a randomised screening trial is typically calculated over an arbitrarily defined follow-up period, including follow-up time where the mortality impact is yet to realise or where it has already been exhausted. Because of this, the cumulative reduction cannot be used for projecting the mortality impact expected from a sustained screening programme. For this purpose, we propose a new measure, the time-specific probability of being helped by screening, given that the cancer would have proven fatal otherwise. This can be decomposed into round-specific impacts, which in turn can be parametrised and estimated from the trial data. This represents a major shift in quantifying the benefits due to a sustained screening programme, based on statistical evidence extracted from existing trial data. We illustrate our approach using data from screening trials in lung and colorectal cancers.

Key words: Cancer screening; conditional likelihood; early detection; randomised screening trial.

1 Introduction

1.1 Motivation

Developed countries spend considerable resources on cancer screening. Policy makers, funders and the public rely on experts who can readily tally the financial outlays and the harms to individuals. Unfortunately, despite many long, costly randomised screening trials involving very large numbers of people, the evidence produced on mortality impacts of cancer screening is commonly misinterpreted. It has not always been recognised that the mortality impact of cancer screening is not constant over time and thus cannot be fully characterised by a single-number summary statistic, such as the commonly reported cumulative mortality reduction, which averages the impact of screening over the entire follow-up period in the trial.

Single-number summaries would be more suitable for describing interventions that result in virtually immediate and long-lasting reductions in disability, infection, sickness or death rates, such as phenylketonuria screening at birth, adult circumcision, human papillomavirus vaccine or screens for abdominal aneurysms. However, the benefit of breast cancer screening seen in 2014 is not due to mammograms carried out in 2014, but to some specific earlier ones. Similarly, the returns on prostate-specific antigen screens for prostate cancer carried out in 2014 will begin to emerge in the late 2010s and early 2020s. The cumulative mortality reduction ignores this critical timing information and thus does not characterise the impact of a sustained cancer screening programme. Naively pooling results from several trials, as many meta-analysts and task forces do (e.g. Djulbegovic *et al.*, 2010; CTFPHC, 2011), is even less meaningful.

The fact that the mortality reduction is not constant over time after the initiation of the screening has been recognised and discussed for instance by Morrison (1992, p. 36), Miettinen *et al.* (2002), Hanley (2005, 2010, 2011), Miettinen & Karp (2012, p. 81) and most recently by us (Hanley *et al.*, 2013). Despite the obvious non-proportionality of the hazard functions, the cumulative mortality reduction statistic is the one that is routinely used in reports of screening trials. It may be a legitimate basis for a test statistic aimed at rejecting the null hypothesis that m rounds of screening, spaced Δ years apart, will produce zero reduction in mortality over the person years accumulated during an average of τ years of follow-up. However, this statistic cannot then be carried forward into projections of mortality reductions under some other screening regimens.

For this purpose, we propose replacing the cumulative mortality reduction with a time-specific analogue. This measure and the proposed estimation method enable reporting the evidence of screening effectiveness in a form that is directly relevant to decision-making.

1.2 Model-based Approaches for Measuring the Benefits of Cancer Screening

Screening for a cancer in pursuit of early (pre-clinical) diagnosis (Miettinen, 2011, p. 26) enables early treatment (in lieu of later ones) and eventually potential mortality reduction at the population level. Randomised screening trials, in which subjects asymptomatic of the cancer are randomly assigned to receive a number of screening examinations, are carried out to produce evidence of efficacy. Diagnostic data collected from the screening period can be used to derive early but imperfect indicators of the possible benefits. Two such indicators are the sensitivity of the screening examinations and the lead time. The statistical estimation of these parameters from trial data has been addressed by a number of authors, such as Walter & Day (1983), Day & Walter (1984) and Shen & Zelen (1999, 2005), among others.

However, because earlier diagnosis does not necessarily translate into more successful treatments, a reduction in cancer-specific mortality is considered more definitive evidence of the benefit of screening. Building on early works (Zelen & Feinleib, 1969; Zelen, 1993), Hu & Zelen (1997) developed a probability model for planning early detection trials using mortality as the end point, by modelling the full disease history through state transitions. Since then, Zelen and co-authors have refined this model to accommodate, for example, dependence between the incidence or prevalence of the cancer and age (Shen & Zelen, 1999; Lee & Zelen, 2008). However, these models typically require a large number of parameter inputs, which are not always available from published trial reports nor obtainable from a single data source, as well as many assumptions that are generally unverifiable. Moreover, the focus in these modelling efforts has been on planning trials, rather than on estimation of the mortality reduction due to screening. Microsimulations used for the latter purpose (e.g. Berry *et al.*, 2005; Zuber *et al.*, 2008; Mandelblatt *et al.*, 2009) are also based on modelling of the entire disease history (e.g. from disease-free to pre-clinical disease state to clinically diagnosed to dead).

In contrast, the conditional approach we propose eliminates parameters characterising prevalence, incidence, sensitivity, state transitions, stage shifts or sojourn time and produces evidence-based, probabilistic projections. Our proposed approach is the first attempt to statistically decompose the bathtub-shaped mortality reduction function, described by Liu *et al.* (2013), into round-specific components. These components can then be compounded to project the reductions that a screening programme, with a possibly different and longer schedule than in a trial, would produce.

We specify our estimand of interest as the conditional probability of being helped by screening (through earlier treatment) given that the cancer would have proven fatal in the absence of screening. This estimand is equivalent to the ‘factor-conditional etiogenetic proportion’ of cancer deaths due to lack of screening-associated early treatments (Miettinen & Karp, 2012, p. 82). We show that this conditional probability has a direct interpretation as the time-specific reduction in cancer mortality and that it can be decomposed into a function of round-specific reductions. We suggest a parametric form for the round-specific reduction, based on which we then formulate a likelihood function.

The remainder of this paper is organised as follows. The estimand and the assumptions necessary to identify it are specified in Section 2. In Section 3, we formulate a parametric model to characterise the round-specific impact and the resulting likelihood expressions for individual-level and aggregated data. In Section 4, we fit our model to data from screening trials in lung and colorectal cancers and illustrate the resulting projections. The paper concludes with a discussion in Section 5.

2 Specifying the Estimand

2.1 Notation

In a randomised screening trial, subjects asymptomatic of cancer are randomly assigned to either a screening or non-screening arm at time $s_0 = 0$, and all are followed up for death due to the cancer or another cause or until the end of follow-up at time τ , whichever comes first. During the interval $[0, \tau]$, a total of m screening examinations are carried out at the ordered time points $s_1 < s_2 < \dots < s_m$ in the screening arm, with the j -th interval denoted by $[s_{j-1}, s_j)$ and its length by $\Delta_j = s_j - s_{j-1}$ for $j = 1, 2, \dots, m$.

We define a screening assignment indicator Z_i taking the value 1 if individual i is assigned to the screening arm, with $Z_i = 0$ otherwise. Let T_i denote the observed time of the event (i.e. death due to the cancer, death due to another cause or type I censoring due to the end of the follow-up period at τ). We take this to be $T_i = Z_i T_{1i} + (1 - Z_i) T_{0i}$, where T_{1i} and T_{0i} denote the potential/counterfactual event times under screening and in the absence of it, respectively (this corresponds to assuming either a ‘stable unit treatment value’, e.g. Angrist *et al.*, 1996, or ‘consistency’, e.g. Cole & Frangakis, 2009). Similarly, let E_i denote the observed event type, taking the value of 1 for cancer-specific death, 2 for death due to another cause and 0 for censoring. This is given by $E_i = Z_i E_{1i} + (1 - Z_i) E_{0i}$, where E_{1i} and E_{0i} are indicator variables for the potential/counterfactual event types under screening and in the absence of it, respectively. The unobservable gained survival time due to screening for individual i is $G_i \equiv T_{1i} - T_{0i}$.

2.2 Object of Inference

We take the estimand to be the probability that a cancer-specific death in the absence of screening was indeed ‘caused’ by the absence of screening-associated early treatments (cf.

Miettinen & Karp, 2012, p. 48). This probability in turn is equivalent to the probability of being helped by screening, had it been available. This is specified as the conditional probability

$$H(t) \equiv P(T_{1i} > t \mid T_{0i} = t, E_{0i} = 1) \quad (2.1)$$

of surviving beyond time t under screening, given a cancer death at time t in the absence of it. Because an individual's cancer can be detected, and subsequently successfully treated, as a result of only one screening examination, we introduce a random variable $S_i \in \{s_1, s_2, \dots, s_m, \infty\}$ to represent the time of being detected, and subsequently successfully treated, with $S_i = \infty$ taken to mean that the cancer was not detected in any of the scheduled screenings or was not successfully treated. Furthermore, because only the screening examinations before the time of death $T_{0i} = t$ can potentially be helpful, we take $m(t) \equiv \max\{j \in \{1, 2, \dots, m\} : s_j < t\}$ to index the last screening examination before t . Thus, we have that $H(t) = P(S_i \in \{s_1, s_2, \dots, s_{m(t)}\} \mid T_{0i} = t, E_{0i} = 1)$ and can express (2.1) as

$$\begin{aligned} H(t) &= \sum_{j=1}^{m(t)} P(S_i = s_j \mid T_{0i} = t, E_{0i} = 1) \\ &= \sum_{j=1}^{m(t)} P(S_i = s_j \mid T_{0i} = t, E_{0i} = 1, S_i \geq s_j) P(S_i \geq s_j \mid T_{0i} = t, E_{0i} = 1). \end{aligned} \quad (2.2)$$

The first term inside the sum (2.2) is the probability of being helped as a result of the j -th screening, given that the previous screenings at times s_1, s_2, \dots, s_{j-1} failed to detect the cancer. Because only new or previously undetected cancers can be detected in the j -th screening, we take the probability

$$Q_j(t) \equiv P(S_i = s_j \mid T_{0i} = t, E_{0i} = 1, S_i \geq s_j) \quad (2.3)$$

as our measure to quantify the mortality impact of a single round of screening; modelling of the round-specific impact is needed to project the mortality impact of a sustained screening programme. The probability (2.2) is fully specified in terms of (2.3), $j = 1, \dots, m$, as

$$H(t) = \sum_{j=1}^{m(t)} Q_j(t) \prod_{k=1}^{j-1} \{1 - Q_k(t)\} = 1 - \prod_{j=1}^{m(t)} \{1 - Q_j(t)\}, \quad (2.4)$$

which follows from the failure probability function for a discrete failure time random variable (e.g. Kalbfleisch & Prentice, 2002, p. 9). The representation (2.4) in turn enables likelihood construction through parametrisation of the functions $Q_j(t)$ (Section 3).

2.3 Identifying Assumptions

Because (2.1) is defined in terms of unobservable quantities, further assumptions are needed to identify it based on observed data. One possible approach would be to assume a deterministic accelerated failure time model, such as $T_{1i} = T_{0i} e^{g(T_{0i})}$ for the potential outcomes (e.g. Hernán *et al.*, 2005). In this case,

$$\begin{aligned} P(T_{1i} > t \mid T_{0i} = t, E_{0i} = 1) &= P(T_{0i} (e^{g(t)} - 1) > 0 \mid T_{0i} = t, E_{0i} = 1) \\ &= P(g(t) > 0 \mid T_{0i} = t, E_{0i} = 1), \end{aligned}$$

which equals 1 whenever the acceleration/deceleration function g is positive and 0 otherwise. This suggests that direct modelling of the gained survival time G_i is unhelpful in addressing the probability of being helped by screening. Instead, we pursue modelling in terms of cause-specific subdensity functions $f_k(t) \equiv P(T_{ki} \in dt, E_{ki} = 1)/dt$ (cf. Kalbfleisch & Prentice, 2002, p. 252), $k = 0, 1$, for individual i dying of the cancer at time t in the absence and presence of screening, respectively. (We use dt to denote both an infinitesimally small interval around t and the infinitesimal length of this interval.)

In order to estimate (2.1), four identifiability assumptions we make in this paper are (i) monotonicity $T_{0i} \leq T_{1i}$; (ii) strongly ignorable assignment, that is, $\{(T_{1i}, E_{1i}), (T_{0i}, E_{0i})\} \perp\!\!\!\perp Z_i$ and $0 < P(Z_i = 1) < 1$ (cf. Rosenbaum & Ruben, 1983, p. 43); (iii) curative early treatments, in the sense that

$$P(T_{1i} > t \mid T_{0i} = t, E_{0i} = 1) = P(T_{1i} > t, E_{1i} \neq 1 \mid T_{0i} = t, E_{0i} = 1); \quad (2.5)$$

and (iv) screening specificity, that is, $E_{0i} = 2 \Rightarrow T_{1i} = T_{0i}, E_{1i} = 2$.

Assumption (i) (cf. Angrist *et al.*, 1996) states that the potential time of death of any cause for an individual in the screening arm is at least as long as that in the non-screening arm, that is, screening cannot shorten anyone's life. Assumption (ii) states that the allocation in the trial is randomised. Assumption (iii) states that the screening-associated early treatments cure the cancer, in the sense of delaying the cause-specific death beyond a death due to a competing cause (or censoring). Assumption (iv) states that the screening technique is specific in the sense that it does not lead to early detection and treatment of conditions other than the site-specific cancer of interest.

Nine different types of event histories, possible under assumptions (i) and (ii), are illustrated in Figure 1. Subject 1 would die of another cause, which could not have been prevented by screening. The death of subject 2 due to another cause was delayed because of screening. This could occur if the screening can also lead to detection of conditions other than the site-specific cancer of interest (which is unlikely). The same applies to histories for subjects 3 and 4. Subject 5 would be alive at the end of the follow-up time, and the time of death due to the cancer for subject 7 would be the same with and without screening; thus, during the follow-up, neither of them would have benefited from screening. Subjects 6, 8 and 9 would die of the cancer in the absence of screening, but in the presence of screening, they would die of another cause or die later as a result of the cancer or be censored at the end of the follow-up, respectively; thus, they could benefit from early detection of the cancer and consequent therapy. Introducing assumption (iii) rules out the event histories of type 8. As will be shown in Section 2.4, this is required for identification of (2.1), because delayed cancer deaths in the screening arm cannot be distinguished from the non-delayed ones based on the observed data. Similarly, we need to use assumption (iv) to rule out histories of type 3, because these would show as excess cancer mortality in the screening arm.

Further, we note that under continuous time, no two cause-specific counting processes can jump simultaneously (Aalen *et al.*, 2008, p. 55), unless they are in fact the same process (which would occur if there is no screening effect). In the present setting, this means that $T_{0i} = T_{1i} \Rightarrow E_{0i} = E_{1i}$, ruling out event histories of the type $(E_{0i} \neq E_{1i}, G_i = 0)$ not present in Figure 1.

2.4 Equivalence between the Probability of Being Helped and Mortality Reduction

We show that under the assumptions stated in Section 2.3, the probability (2.1) of being helped by screening is equivalent to the time-specific reduction in cancer mortality, a quantity

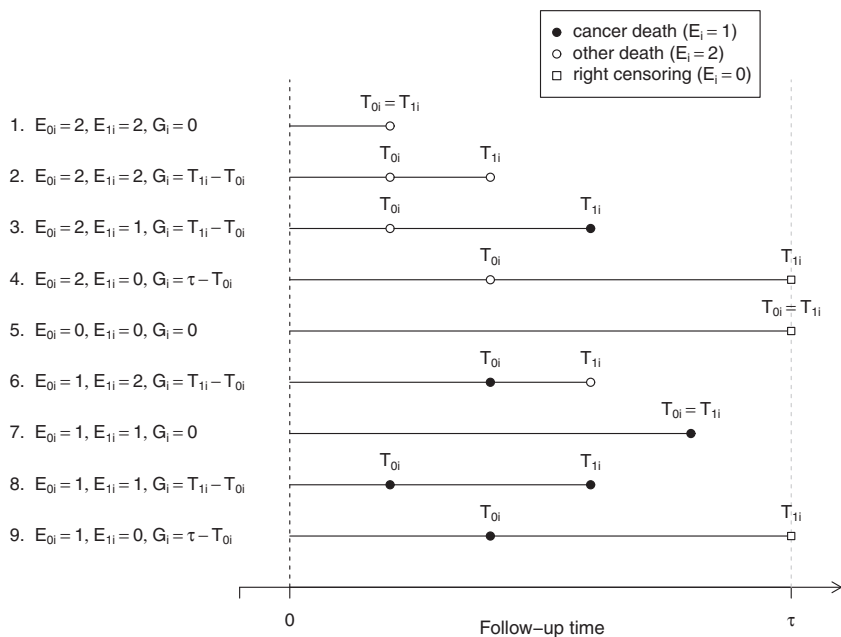


Figure 1. Illustration of nine different possible event histories, one row per individual, possible under the identifying assumptions (i) and (ii) of Section 2.3. For identifying the causal estimand (2.1), event histories of types 3 and 8 need to be further ruled out by the identifying assumptions (iii) and (iv) of Section 2.3. Only event histories of types 6, 7 and 9 eventually contribute to the conditional likelihood expression (3.4).

that can be estimated based on the trial data. We may express (2.1) as

$$\begin{aligned}
 P(T_{1i} > t \mid T_{0i} = t, E_{0i} = 1) &= 1 - P(T_{1i} \leq t \mid T_{0i} = t, E_{0i} = 1) \\
 &= 1 - \sum_{k=0}^2 P(T_{1i} \in dt, E_{1i} = k \mid T_{0i} = t, E_{0i} = 1) \\
 &= 1 - P(T_{1i} \in dt, E_{1i} = 1 \mid T_{0i} = t, E_{0i} = 1) \\
 &= 1 - \frac{P(T_{0i} \in dt, E_{0i} = 1 \mid T_{1i} = t, E_{1i} = 1) P(T_{1i} \in dt, E_{1i} = 1)/dt}{P(T_{0i} \in dt, E_{0i} = 1)/dt} \\
 &= 1 - \frac{f_1(t)}{f_0(t)}.
 \end{aligned}$$

The second equality is due to the monotonicity assumption (i). The third equality follows from the continuous-time model for the counting processes. The fifth equality is due to $P(T_{0i} \in dt, E_{0i} = 1 \mid T_{1i} = t, E_{1i} = 1) = 1$, which follows from assumptions (i), (iii) and (iv). In Appendix A, we point out that the measure obtained previously is a time-specific analogue of the more familiar cumulative mortality reduction.

While the estimand is specified in terms of potential outcome variables, the ignorability assumption (ii) enables its estimation using the observed outcomes in the two trial arms because

$$1 - \frac{f_1(t)}{f_0(t)} = 1 - \frac{P(T_i \in dt, E_i = 1 \mid Z_i = 1)/dt}{P(T_i \in dt, E_i = 1 \mid Z_i = 0)/dt} = 1 - \frac{f(t \mid Z_i = 1)}{f(t \mid Z_i = 0)},$$

where $f(t | Z_i) \equiv P(T_i \in dt, E_i = 1 | Z_i)/dt$. We further note that if we were only interested in the time-specific mortality reduction curve $1 - f_1(t)/f_0(t)$, only the identifying assumption (ii) would be needed. The other identifying assumptions are required to connect this to the probability of being helped (2.1), which can be decomposed into round-specific impacts.

3 Methods

In this section, we present a parametric model characterising the effect of a single round of screening. The mortality reduction at time t can then be obtained as a compound of the impacts of each screen that persons have received up to t , as shown in Section 2.2.

3.1 Model Formulation

As emphasised by several authors referred to in the introduction, a quintessential feature of cancer screening is the non-constancy of its impact over time. According to Miettinen (2013), it is a fundamental truism that the mortality reduction ‘cannot be constant over successive intervals of time after the screening’s initiation; that it is initially nil, then increases and later declines, and ultimately totally vanishes’. This can be understood in terms of a detectability–curability trade-off; tumours in earlier stage are difficult to detect but presumably easier to cure, while late-stage tumours are more detectable, but treatment may come too late. However, both detection and successful treatment are required for cure. A cancer that would prove fatal within months from now is not likely to be cured by screening today, while a cancer that is cured today as a result of early detection would otherwise have proven fatal several years from now.

3.1.1 Stationarity property

To start with, we assume that the probabilities of being helped by each round of screening as functions of time are shifted versions of each other, that is,

$$Q_1(t) = Q_2(t + \Delta_1) = \dots = Q_m \left(t + \sum_{k=1}^{m-1} \Delta_k \right). \tag{3.1}$$

For simplicity, take the screenings to be equally spaced so that $\Delta_1 = \Delta_2 = \dots = \Delta_{m-1} = \Delta$, and take the successive examinations j and $j + 1$ as an example, in which case

$$\begin{aligned} Q_j(t) = Q_{j+1}(t + \Delta) &\Leftrightarrow P(S_i = s_j | T_{0i} = t, E_{0i} = 1, S_i \geq s_j) \\ &= P(S_i = s_j + \Delta | T_{0i} = t + \Delta, E_{0i} = 1, S_i \geq s_j + \Delta). \end{aligned}$$

Thus, modelling assumption (3.1) can be interpreted as a stationarity property for the functions $Q_j(t)$. This is plausible because the length $T_{0i} - S_i$ of the interval from time of screen detection to the potential time of death without screening is kept constant. Furthermore, the preceding probabilities are conditional on not being detected in the previous screening examinations, and if the sensitivity of the screening test and participation rates are high, the cancers to be detected at any s_j are mainly ‘new’ ones, having progressed to the detectable state in the interval $[s_{j-1}, s_j)$. While this applies to the repeat screenings, the first or ‘prevalence’ screening might involve a different stage distribution of cancers; we address this question briefly in Section 5.

3.1.2 Examples of possible parametrisations

The effect of one round of screening could be characterised in terms of maximal reduction (γ), the time lag between the time of screening and the maximal reduction (location parameter μ) and the spread of the reductions over time (scale parameter σ). Using these three parameters, a possible formulation for the time-specific reduction due to one screen is

$$Q_j(t; \gamma, \mu, \sigma) \equiv \gamma \exp \left\{ - \left(\frac{t - (\mu + \sum_{l=1}^j \Delta_l)}{\sigma} \right)^2 \right\}, \quad (3.2)$$

where $0 \leq t < \infty$, $0 \leq \gamma \leq 1$, $\mu > 0$ and $\sigma > 0$. Function (3.2) characterises how deep, how far into the future and how wide the mortality reductions produced by a single screen are.

A possible limitation of formulation (3.2) is that it does not enforce the restriction $\lim_{t \rightarrow s_j^+} Q_j(t) = 0$ (if the death in the absence of screening would have resulted immediately after the detection, any therapy would be unlikely to help the patient). An alternative formulation that satisfies this restriction could be

$$\begin{aligned} Q_j(t; \gamma, \alpha, \beta) &\equiv \gamma \frac{f(t - \sum_{l=1}^j \Delta_l; \alpha, \beta)}{f((\alpha - 1)\beta; \alpha, \beta)} \\ &= \gamma \left\{ \frac{t - \sum_{l=1}^j \Delta_l}{(\alpha - 1)\beta} \right\}^{\alpha-1} \exp \left\{ (\alpha - 1) - \frac{t - \sum_{l=1}^j \Delta_l}{\beta} \right\}, \end{aligned} \quad (3.3)$$

where $0 \leq t < \infty$, $0 \leq \gamma \leq 1$, $\alpha > 1$ and $\beta > 0$. Here, $f(t; \alpha, \beta)$ is the probability density function of a gamma distribution with the mode $t = (\alpha - 1)\beta$. By scaling down the density function by its maximum value, we restrict the time-specific mortality reductions to be between 0 and 1. Possible shapes with various parameter inputs for the two formulations can be found in Figure 2. This also demonstrates that when the delay in the mortality reduction is long, shape (3.3) approximates shape (3.2). Because the former ensures a continuous overall reduction curve, in the examples of Section 4, we use only the gamma kernel parametrisation.

Taking θ to be the collection of the three parameters, appropriately transformed, the compound reduction $H(t; \theta)$ resulting from $m(t)$ screens before time t can now be obtained by substituting $Q_j(t; \theta)$'s into Equation (2.4).

3.2 Likelihood Formulation

3.2.1 Individual-level data

We adopt a conditional approach, where the conditional likelihood contribution of individual i is the probability of the screening assignment Z_i given that there was a cancer death at t , that is, $Z_i \mid (T_i = t, E_i = 1) \sim \text{Bernoulli}\{\pi(t)\}$, resulting in likelihood contributions of the form $\pi(t)^{Z_i} (1 - \pi(t))^{1-Z_i}$ for each cancer death. Here, with equal allocation $P(Z_i = 1) =$

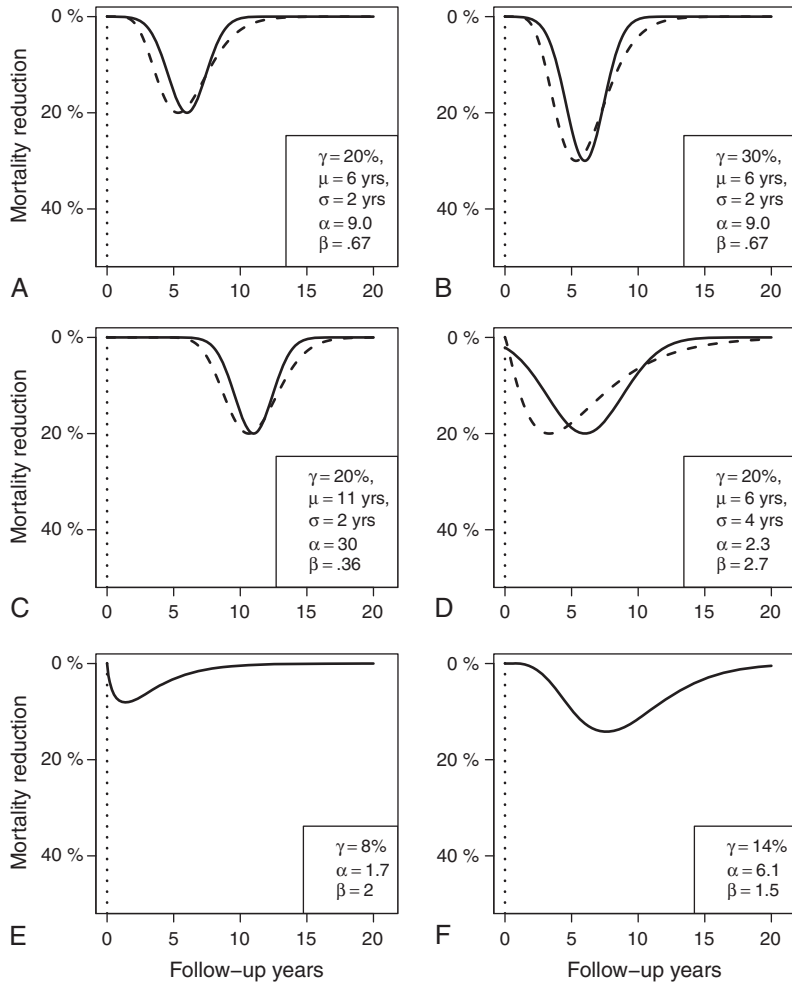


Figure 2. Impact of a single round of screening at time $s_1 = 0$, with different patterns determined by different parameter inputs. Solid and dashed lines correspond to Equations (3.2) and (3.3), respectively. Panels E and F correspond to the fitted reduction patterns in the examples of Sections 4.1 and 4.2, respectively.

$P(Z_i = 0) = 0.5$ between the two arms,

$$\begin{aligned}
 \pi(t) &\equiv P(Z_i = 1 \mid T_i = t, E_i = 1) \\
 &= \frac{P(T_i \in dt, E_i = 1 \mid Z_i = 1)P(Z_i = 1)}{P(T_i \in dt, E_i = 1 \mid Z_i = 0)P(Z_i = 0) + P(T_i \in dt, E_i = 1 \mid Z_i = 1)P(Z_i = 1)} \\
 &= \frac{f(t \mid Z_i = 1)/f(t \mid Z_i = 0)}{1 + f(t \mid Z_i = 1)/f(t \mid Z_i = 0)} \\
 &= \frac{1 - H(t; \theta)}{1 + 1 - H(t; \theta)}.
 \end{aligned} \tag{3.4}$$

3.2.2 Aggregated data

If the individual-level mortality data are not reported or accessible, our model can be fitted to aggregated (e.g. yearly) numbers of deaths in each arm, extractable from the

cumulative mortality curves in the published trial reports (Liu *et al.*, 2014). Let $D_{0j} = \sum_i \mathbf{1}_{\{E_i=1, t_{j-1} < T_i < t_j, Z_i=0\}}$ and $D_{1j} = \sum_i \mathbf{1}_{\{E_i=1, t_{j-1} < T_i < t_j, Z_i=1\}}$ denote the numbers of cancer-specific deaths during the interval $[t_{j-1}, t_j]$, $j = 1, 2, \dots, J$, in the non-screening and screening arms, respectively. Thus, the distribution of D_{1j} conditional on the total deaths during interval j is $D_{1j} \mid (D_{0j} + D_{1j} = d_j) \sim \text{Binomial}(d_j, \pi_j)$, where

$$\begin{aligned} \pi_j &= \frac{N_1[F(t_j \mid Z_i = 1) - F(t_{j-1} \mid Z_i = 1)]}{N_0[F(t_j \mid Z_i = 0) - F(t_{j-1} \mid Z_i = 0)] + N_1[F(t_j \mid Z_i = 1) - F(t_{j-1} \mid Z_i = 1)]} \\ &= \frac{[F(t_j \mid Z_i = 1) - F(t_{j-1} \mid Z_i = 1)]/[F(t_j \mid Z_i = 0) - F(t_{j-1} \mid Z_i = 0)]}{1 + [F(t_j \mid Z_i = 1) - F(t_{j-1} \mid Z_i = 1)]/[F(t_j \mid Z_i = 0) - F(t_{j-1} \mid Z_i = 0)]} \\ &\approx \frac{1 - \int_{t_{j-1}}^{t_j} H(t; \theta) \frac{1}{t_j - t_{j-1}} dt}{1 + 1 - \int_{t_{j-1}}^{t_j} H(t; \theta) \frac{1}{t_j - t_{j-1}} dt}, \end{aligned} \quad (3.5)$$

where $N_1 = N_0$ are the numbers of individuals randomised to screening and control arms, respectively, and $F(t \mid Z_i) \equiv \int_0^t f(v \mid Z_i) dv$. Notably, $\lim_{t_j \rightarrow t_{j-1}} \pi_j = \pi(t_{j-1})$, reducing to the individual-level formulation in (3.4). The resulting log-likelihood function is the sum of contributions from the entire duration of the follow-up time, given by $l(\theta) \equiv \sum_{j=1}^J \{D_{1j} \log(\pi_j) + D_{0j} \log(1 - \pi_j)\}$.

3.3 Estimation

The likelihood functions for individual-level or aggregated data in Sections 3.2.1 and 3.2.2 can be maximised with respect to parameters specifying the mortality reduction function $H(t; \theta)$ using standard numerical optimisation methods, such as those implemented by the `optim` function of the R statistical environment. Standard errors for the parameter estimates may be obtained by inverting the numerically differentiated observed information matrix at the maximum likelihood point. Because all the parameters in (3.2) or (3.3) are positive, reparametrisations should be used when applying a normal approximation to the likelihood in order to obtain standard errors for the parameter estimates. However, rather than the individual parameters, our main interest is in obtaining measures of uncertainty for the mortality projections. Because the projections are based on a probability model fitted using maximum likelihood, time-specific confidence bands may be constructed straightforwardly by randomly drawing parameter estimate values from the approximate large-sample sampling distribution $N(\hat{\theta}, i(\hat{\theta})^{-1})$, where $i(\hat{\theta})$ is the observed information matrix at the maximum likelihood point, and calculating the projection curve at each value. The 2.5% and 97.5% limits at each time point can then be obtained as sample quantiles.

3.4 Generalisations

In this subsection, we extend our model to accommodate an unequal allocation of person-time between the screening and non-screening arms, less than full compliance and multiple screening arms within a trial.

If the randomisation ratio between the screening arm and the non-screening arm is $N_1/N_0 \equiv \phi : 1$ instead of 1:1, such as in the Swedish two-county trial (Tabár *et al.*, 1985), as well as two other mammographic screening trials in Stockholm (Frisell *et al.*, 1997) and Gothenburg (Bjurstam *et al.*, 2003), then Equation (3.5) becomes

$$\pi_j = \frac{\phi \left\{ 1 - \int_{t_{j-1}}^{t_j} H(t; \theta) \frac{1}{t_j - t_{j-1}} dt \right\}}{1 + \phi \left\{ 1 - \int_{t_{j-1}}^{t_j} H(t; \theta) \frac{1}{t_j - t_{j-1}} dt \right\}}.$$

Sometimes, multiple screening arms are employed within the same trial, such as the Minnesota colorectal cancer study (Shaikat *et al.*, 2013) in which participants were randomly assigned to be screened annually, biennially or not at all. To accommodate this, let D_{kj} denote the number of cancer-specific deaths in arm k , where $k = 0, \dots, K$, during the j -th interval. Given the total number of deaths $d_j = \sum_{k=0}^K D_{kj}$, the split into the K study arms is distributed as

$$D_{0j}, \dots, D_{Kj} \mid \left(\sum_{k=0}^K D_{kj} = d_j \right) \sim \text{Multinomial}(d_j, \pi_{0j}, \dots, \pi_{Kj}),$$

resulting in a log-likelihood function $l(\theta) \equiv \sum_{j=1}^J \sum_{k=0}^K D_{kj} \log(\pi_{kj})$, where $\pi_{0j} = 1 - \sum_{k=1}^K \pi_{kj}$.

While our estimand (2.1) should be interpreted as an intention-to-treat type of effect, with the potential outcome (T_{1i}, E_{1i}) corresponding to being assigned to the screening arm of the trial, as opposed to actually undergoing screening as scheduled, in the projection task, it might be appropriate to upscale or downscale the mortality impact of the screening programme by the expected participation rate. In addition, a relevant quantity for decision-making at the individual level would be the mortality impact conditional on participation in the screening. Assuming that the participation in the screening arm in round j , denoted as $C_{ij} = 1$, is completely at random in the sense that $P(C_{ij} = 1 \mid T_{0i} = t, E_{0i} = 1, S_i \geq s_j) = P(C_{ij} = 1) \equiv c_{Tj}$ and that no one is screened in the control arm, the participant probability of being helped by this round is simply

$$P(S_i = s_j \mid T_{0i} = t, E_{0i} = 1, S_i \geq s_j, C_{ij} = 1) \equiv Q_j^*(t) = \frac{1}{c_{Tj}} Q_j(t). \quad (3.6)$$

This follows from $P(S_i = s_j, C_{ij} = 0 \mid T_{0i} = t, E_{0i} = 1, S_i \geq s_j) = 0$. Differential participation between the successive rounds of screening may now be accounted for by using the relationship (3.6) in fitting the likelihood (3.4) or (3.5), by replacing $Q_j(t)$ in Equation (2.4) with $c_{Tj} Q_j^*(t)$, with the parameter estimates then representing the effects of screening under completely random non-participation. Now, if the expected participation in the screening programme round j is c_{Pj} , the mortality impact of this round can be projected simply as $c_{Pj} Q_j^*(t)$, with the compound impact given by formula (2.4). We demonstrate this approach in the example of Section 4.2. A full treatment of possibly non-random non-participation in our modelling framework is a topic for further work.

3.5 Checking the Model Fit

The appropriateness of the modelling assumptions of Section 3.1 can be checked through comparing the predictions from the fitted model with the observed aggregated death counts. Because the aggregated counts D_{1j} in the screening arm are binomially distributed given the total count d_j in the j -th interval, we can construct a goodness-of-fit test statistic of the form

$$\sum_{j=1}^J \frac{(D_{1j} - d_j \hat{\pi}_j)^2}{d_j \hat{\pi}_j (1 - \hat{\pi}_j)} \sim \chi_{J-p}^2,$$

where p is the number of estimated parameters in the model and $\hat{\pi}_j$ is the estimated split derived from substituting the parameter estimates $\hat{\theta}$ into formula (3.5). We carried out a simulation study to verify that the test produces the correct type I error rate; the results are reported in Appendix B.

4 Examples

4.1 The US National Lung Screening Trial

We illustrate our methods using data from the US National Lung Screening Trial (Aberle *et al.*, 2011), which compared lung cancer mortality among 53 454 heavy smokers randomised to either low-dose computed tomography (CT) scans or chest X-rays. The screening regimen in the trial comprised three annual rounds, the first one soon after randomisation. A 20% cumulative mortality reduction was reported in the CT arm after 7 years of follow-up, compared with the X-ray arm. On the other hand, we are interested in the mortality reductions that would be produced by a sustained screening programme targeted to such high-risk individuals.

Because lung cancer is rapidly progressing, the mortality reductions might manifest with a short delay, and consequently, a separate parameter to characterise this delay might not be needed. Initial model fits with the three-parameter formulation (3.3) also suggested this, with the parameter α estimating to the boundary value 1, motivating reduction of the number of parameters to two. A very parsimonious model still producing a reasonable reduction pattern for a single round of screening can be obtained by fixing $\beta = 2$ in (3.3), giving a two-parameter model based on the χ^2 -kernel. The fitted reduction curve due to one round of screening is shown in Figure 2E. Because the individual-level data from the trial were provided to us by the National Cancer Institute, we could fit this model to both the exact times of death (Equation (3.4)) and the yearly and half-yearly aggregated numbers (Equation (3.5)). The yearly numbers of deaths and fitted reduction curves due to three screenings are presented in Figure 3A, which suggests that the aggregated numbers are near-sufficient statistics for the mortality reduction: the curves fitted to aggregated data are almost identical to the individual-level fit. Applying the goodness-of-fit test statistic of Section 3.5 to the yearly observed and expected counts gave a p -value of 0.55, indicating no evidence against the fitted model. The maximum mortality reduction produced by the three rounds of screening is around 20%, which fades after the screening was discontinued. However, the projected reduction pattern in Figure 3B based on 10 rounds of annual screening and 90% compliance demonstrates that the mortality reductions would plateau at a nadir of around 30%, should the screening be continued long enough.

4.2 The Minnesota Colorectal Cancer Screening Study

Shaukat *et al.* (2013) reported that the mortality from colorectal cancer in the screening arm with 11 annual and 6 biennial faecal occult blood (FOB) tests is 32% and 22% lower than that in the non-screening arm, respectively. The study involved 46 551 participants equally allocated to the three arms and followed up for 30 years. These mortality reductions were achieved despite a 4-year funding-related hiatus in screening and averaging over the entire 30-year follow-up. Presumably, the reductions would have been larger without such an interruption.

To study this, we extracted the yearly numbers of deaths from the published figure of cumulative colorectal cancer mortality and present the observed and fitted mortality reductions in Figure 4A. The fitted model was specified using the parametrization (3.3), and the pattern of reduction due to one round of screening is shown in Figure 2F. The goodness-of-fit test of

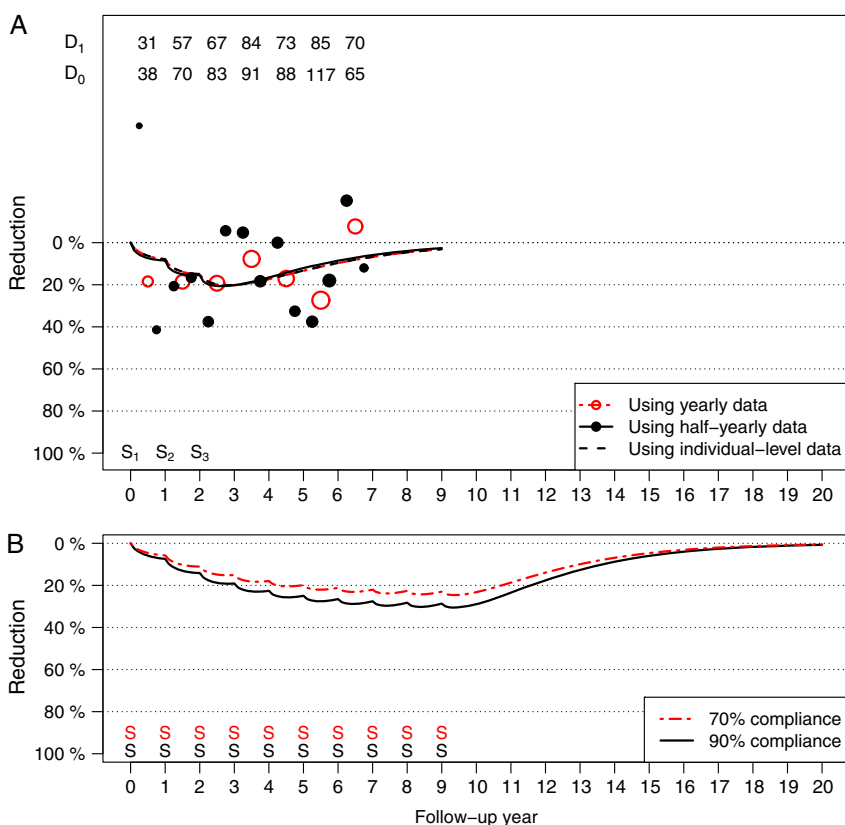


Figure 3. Panel A: Empirical ($100\% \times (1 - D_{1j} / D_{0j})$) and fitted mortality reductions based on individual-level, as well as aggregated yearly and half-yearly, data from the National Lung Screening Trial. The size of each dot is proportional to the information contribution of the empirical year-specific mortality ratio. Panel B: Projection of time-specific lung cancer mortality reductions that would be generated by 10 years of annual computed tomography (versus chest X-ray) screening.

Section 3.5, applied to 2-year aggregated death counts due to the small yearly numbers, gave p -values of 0.87 and 0.57 in the annual and biennial screening arms, respectively, indicating no evidence against the fitted model. While the impact of the hiatus in the screening is not obvious in the cumulative mortality curves (Figure 1 of Shaukat *et al.*, 2013), our time-specific ones indeed exhibit a W shape, showing the lagged responses to the two phases of screening: after a delay of some years, a nadir of around 40% reduction for annual and 30% for biennial schedules were reached before beginning to revert back to zero; this pattern is repeated when screening was resumed.

Figure 4B shows the projected reductions due to 16 years of continuous (annual and biennial) screening. The time patterns generated by these two regimens are similar in that benefits start to emerge some 5 years after the initiation of screening, continue to manifest until reaching the nadir in year 15 and continue onwards. However, the projected sustained reduction is close to 60% for annual screening and 40% for biennial in the time window affected, assuming the same compliance rates, 75% and 78% (annual and biennial, respectively; Mandel *et al.*, 2000), as in the trial. The 95% time-specific confidence bands in Figure 4B are obtained, as outlined in Section 3.3, for the biennial regimen based on 10 000 random draws.

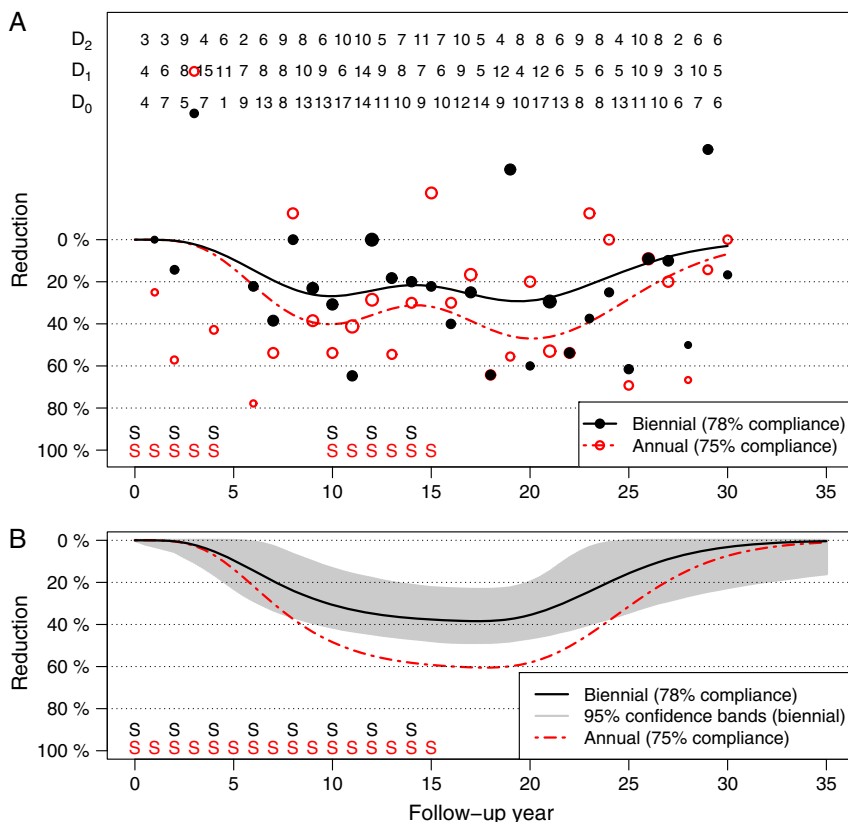


Figure 4. Panel A: Empirical ($100\% \times (1 - D_{1j}/D_{0j})$ and $100\% \times (1 - D_{2j}/D_{0j})$) and fitted mortality reductions based on the yearly numbers of colorectal cancer deaths in the two screening arms of the Minnesota Colorectal Cancer Screening Study, with the 4-year hiatus. The size of each dot is proportional to the information contribution of the empirical year-specific mortality ratio. Because the hiatus was in calendar time rather than follow-up time and entries were staggered, the timing of the screens, each denoted by an S, is only approximate. Panel B: Projection of yearly mortality reductions in colorectal cancer that would be generated by 15 years of uninterrupted annual and biennial faecal occult blood screening. The grey area represents time-specific 95% confidence bands under the biennial screening regimen.

5 Discussion

Although we did not make distinctions between the impact pattern of the first round of screening and that of the subsequent ones, more parameters could easily be added for modelling the effect of the first, or prevalence, screen, provided that there are sufficient data to enable estimation of the added parameters. For instance, one option would be to model the maximal reduction γ as a function of the time of the screening examination. Another option, motivated by the FOB testing for colorectal cancer, would be to employ six parameters to characterise two modes for each round, corresponding to immediate and remote mortality impacts of removing colorectal cancers and polyps, respectively. However, our experience is that the parametric models should be fairly simple to ensure identifiability of the estimation problem, at least if the trial involved only one screening regimen. For identifiability of further parameters, it might be helpful to have data available on different individual-level screening histories along with their exact timings.

Instead of explicitly modelling sensitivity of the screening examinations or the effectiveness of the subsequent treatment, we concentrate on modelling the probability of being helped by

screening; the former two are components of the latter. For instance, a probability of being helped resulting from a high sensitivity of detecting the cancer combined with ineffective treatment is not distinguishable from one resulting from a low sensitivity combined with an effective treatment. In particular, estimating sensitivity of the screening would be problematic; the true disease status is inherently unobservable, as the true positives are those screen-detected cancers that would eventually have proven to be fatal in the absence of screening. Our conditional approach circumvents the overdiagnosis problem by focusing on cancer deaths instead of cancer diagnoses.

To summarise, our conditional approach addresses the mortality impact directly by parametrising the time-specific conditional probability of being helped by screening, given that the cancer would have proven fatal otherwise. This, under the assumptions stated in Section 2.3, is equivalent to the time-specific mortality reduction, a quantity estimable from trial data. By fitting our model to data from lung and colorectal cancer screening trials, we illustrated how the parameter estimates can be used to project and compare reduction curves that could be produced by long-term screening programmes. Our methods can provide policy makers and funders more relevant evidence on how effective cancer screening programmes are and could be.

Acknowledgements

This research was funded by the Canadian Institutes of Health Research (grant number 115204). The authors thank the National Cancer Institute (NCI) for access to NCI's data collected by the National Lung Screening Trial. The statements contained herein are solely those of the authors and do not represent or imply concurrence or endorsement by the NCI.

References

- Aalen, O., Borgan, Ø., Gjessing, H.K. & Gjessing, S. (2008). *Survival and Event History Analysis: A Process Point of View*. New York: Springer.
- Aberle, D.R., Adams, A.M., Berg, C.D., Black, W.C., Clapp, J.D., Fagerstrom, R.M., Gareen, I.F., Gatsonis, C., Marcus, P.M. & Sicks, J.D. (2011). Reduced lung-cancer mortality with low-dose computed tomographic screening. *New Engl. J. Med.*, **365**, 395–409.
- Angrist, J.D., Imbens, G.W. & Rubin, D.B. (1996). Identification of causal effects using instrumental variables. *J. Amer. Statist. Assoc.*, **91**, 444–455.
- Berry, D.A., Cronin, K.A., Plevritis, S.K., Fryback, D.G., Clarke, L., Zelen, M., Mandelblatt, J.S., Yakovlev, A.Y., Habbema, J.D. & Feuer, E.J. (2005). Effect of screening and adjuvant therapy on mortality from breast cancer. *New Engl. J. Med.*, **353**, 1784–1792.
- Bjurstam, N., Bjorneld, L., Warwick, J., Sala, E., Duffy, S.W., Nystrom, L., Walker, N., Cahlin, E., Eriksson, O., Hafstrom, L.O., Lingaas, H., Mattsson, J., Persson, S., Rudenstam, C.M., Salander, H., Save-Soderbergh, J. & Wahlin, T. (2003). The Gothenburg Breast Screening Trial. *Cancer*, **97**, 2387–2396.
- Cole, S.R. & Frangakis, C.E. (2009). The consistency statement in causal inference: a definition or an assumption? *Epidemiology*, **20**, 3–5.
- CTFPHC. (2011). The Canadian Task Force on Preventive Health Care: recommendations on screening for breast cancer in average-risk women aged 40–74 years. *Can. Med. Assoc. J.*, **183**, 1991–2001.
- Day, N.E. & Walter, S.D. (1984). Simplified models of screening for chronic disease: estimation procedures from mass screening programmes. *Biometrics*, **40**, 1–13.
- Djulgovic, M., Beyth, R.J., Neuberger, M.M., Stoffs, T.L., Vieweg, J., Djulgovic, B. & Dahm, P. (2010). Screening for prostate cancer: systematic review and meta-analysis of randomised controlled trials. *BMJ*, **341**, c4543.
- Frisell, J., Lidbrink, E., Hellstrom, L. & Rutqvist, L.E. (1997). Followup after 11 years—update of mortality results in the Stockholm mammographic screening trial. *Breast Cancer Res. Tr.*, **45**, 263–270.

- Hanley, J.A. (2005). Analysis of mortality data from cancer screening studies: looking in the right window. *Epidemiology*, **16**, 786–790.
- Hanley, J.A. (2010). Mortality reductions produced by sustained prostate cancer screening have been underestimated. *J. Med. Screen.*, **17**, 147–151.
- Hanley, J.A. (2011). Measuring mortality reductions in cancer screening trials. *Epidemiologic Reviews*, **33**, 36–45.
- Hanley, J.A., McGregor, M., Liu, Z., Strumpf, E.C. & Dendukuri, N. (2013). Measuring the mortality impact of breast cancer screening. *C. J. Public Health*, **104**, e437–e442.
- Hernán, M.A., Cole, S.R., Margolick, J., Cohen, M. & Robins, J.M. (2005). Structural accelerated failure time models for survival analysis in studies with time-varying treatments. *Pharmacoepidem. Dr. S.*, **14**, 477–491.
- Hu, P & Zelen, M. (1997). Planning clinical trials to evaluate early detection programmes. *Biometrika*, **84**, 817–830.
- Kalbfleisch, J.D. & Prentice, R.L. (2002). *The Statistical Analysis of Failure Time Data, Second Edition*. New Jersey: Wiley.
- Lee, S.J. & Zelen, M. (2008). Mortality modeling of early detection programs. *Biometrics*, **64**, 386–395.
- Liu, Z., Hanley, J.A. & Strumpf, E.C. (2013). Projecting the yearly mortality reductions due to a cancer screening program. *J. Med. Screen.*, **20**, 156–164.
- Liu, Z., Rich, B. & Hanley, J.A. (2014). Recovering the raw data behind a non-parametric survival curve. *Systematic Reviews*, **3**(151), 1–10.
- Mandel, J.S., Church, T.R., Bond, J.H., Ederer, F., Geisser, M.S., Mongin, S.J., Snover, D.C. & Schuman, L.M. (2000). The effect of fecal occult-blood screening on the incidence of colorectal cancer. *New Engl. J. Med.*, **343**, 1603–1607.
- Mandelblatt, J.S., Cronin, K.A., Bailey, S., Berry, D.A., de Koning, H.J., Draisma, G., Huang, H., Lee, S.J., Munsell, M., Plevritis, S.K., Ravdin, P., Schechter, C.B., Sigal, B., Stoto, M.A., Stout, N.K., van Ravesteyn, N.T., Venier, J., Zelen, M. & Feuer, E.J. (2009). Effects of mammography screening under different screening schedules: model estimates of potential benefits and harms. *Ann. Intern. Med.*, **151**, 738–747.
- Miettinen, O.S. (2011). *Epidemiological Research: Terms and Concepts*. Dordrecht: Springer.
- Miettinen, O.S. (2013). Screening for breast cancer: what truly is the benefit. *Can. J. Public Health*, **104**, e435–e436.
- Miettinen, O.S., Henschke, C.I., Pasmantier, M.W., Smith, J.P., Libby, D.M. & Yankelevitz, D.F. (2002). Mammographic screening: no reliable supporting evidence? *Lancet*, **359**, 404–405.
- Miettinen, O.S. & Karp, I. (2012). *Epidemiological Research: An Introduction*. Dordrecht: Springer.
- Morrison, A.S. (1992). *Screening in Chronic Disease*, second edition. New York: Oxford University Press.
- Rosenbaum, P.R. & Rubin, D.B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, **6**, 41–55.
- Shaukat, A., Mongin, S.J., Geisser, M.S., Lederle, F.A., Bond, J.H., Mandel, J.S. & Church, T.R. (2013). Long-term mortality after screening for colorectal cancer. *New Engl. J. Med.*, **369**, 1106–1114.
- Shen, Y. & Zelen, M. (1999). Parametric estimation procedures for screening programmes: stable and nonstable disease models for multimodality case finding. *Biometrika*, **86**, 503–515.
- Shen, Y. & Zelen, M. (2005). Robust modeling in screening studies: estimation of sensitivity and pre-clinical sojourn time distribution. *Biostatistics*, **6**, 604–614.
- Tabár, L., Fagerberg, C.J., Gad, A., Baldetorp, L., Holmberg, L.H., Grontoft, O., Ljungquist, U., Lundstrom, B., Manson, J.C. & Eklund, G. (1985). Reduction in mortality from breast cancer after mass screening with mammography. Randomised trial from the Breast Cancer Screening Working Group of the Swedish National Board of Health and Welfare. *Lancet*, **325**, 829–832.
- Walter, S.D. & Day, N.E. (1983). Estimation of the duration of a pre-clinical disease state using screening data. *Am. J. Epidemiol.*, **118**, 865–886.
- Zauber, A.G., Lansdorp-Vogelaar, I., Knudsen, A.B., Wilschut, J., van Ballegooijen, M. & Kuntz, K.M. (2008). Evaluating test strategies for colorectal cancer screening: a decision analysis for the US preventive services task force. *Ann. Intern. Med.*, **149**, 659–669.
- Zelen, M. (1993). Optimal scheduling of examinations for the early detection of disease. *Biometrika*, **80**, 279–293.
- Zelen, M. & Feinleib, M. (1969). On the theory of screening for chronic diseases. *Biometrika*, **56**, 601–614.

[Received April 2014, accepted December 2014]

Appendix A: Relationship to Cumulative Mortality Reduction

Our estimand, the probability of being helped, which equals the time-specific mortality reduction, has a natural connection to the cumulative mortality reduction, a measure commonly used as a descriptive or test statistic in randomised trials (Section 4). With the same assumptions as stated in Section 2.3, we can express the probability of surviving beyond the potential time of death in the absence of screening, had the cancer proven fatal without screening *before* time t , as

$$\begin{aligned}
 & P(T_{1i} > T_{0i} \mid T_{0i} \leq t, E_{0i} = 1) \\
 &= 1 - P(T_{1i} \leq T_{0i} \mid T_{0i} \leq t, E_{0i} = 1) \\
 &= 1 - \frac{\int_{v \in [0, t]} P(T_{1i} \leq v, T_{0i} \in dv, E_{0i} = 1)}{\int_{v \in [0, t]} P(T_{0i} \in dv, E_{0i} = 1)} \\
 &= 1 - \frac{\int_{v \in [0, t]} P(T_{1i} \leq v \mid T_{0i} = v, E_{0i} = 1) P(T_{0i} \in dv, E_{0i} = 1)}{\int_{v \in [0, t]} P(T_{0i} \in dv, E_{0i} = 1)} \\
 &= 1 - \frac{\int_{v \in [0, t]} P(T_{0i} \in dv, E_{0i} = 1 \mid T_{1i} = v, E_{1i} = 1) P(T_{1i} \in dv, E_{1i} = 1)}{\int_{v \in [0, t]} P(T_{0i} \in dv, E_{0i} = 1)} \\
 &= 1 - \frac{\int_0^t f_1(v) dv}{\int_0^t f_0(v) dv} \equiv 1 - \frac{F_1(t)}{F_0(t)},
 \end{aligned}$$

where $F_k(t)$, $k = 0, 1$, are the cause-specific cumulative incidence functions for cancer mortality under no screening and screening, respectively. In the context of planning a trial, Hu & Zelen (1997, p. 823) use the risk difference $F_0(\tau) - F_1(\tau)$ at the end of the follow-up period as the measure of the impact of the planned screening regimen used in the trial. As demonstrated here, under the assumptions of Section 2.3, the mortality reduction, $1 - F_1(\tau)/F_0(\tau)$, is equivalent to the probability of being helped by screening given a cancer death during the follow-up window $[0, \tau]$ in the absence of screening.

Appendix B: Simulation Study

To validate the under-the-null behaviour of the testing procedure described in Section 3.5, we considered a 15-year follow-up with expected control arm death counts given by $D_{0j} = 100$, $j = 1, \dots, 15$, and expected total counts $d_j = D_{0j} + [1 - H((t_j - t_{j-1})/2; \theta)]D_{1j}$, rounded to the nearest integer, where the function $H(t; \theta)$ was specified through formula (2.4), with three screening rounds in total, located at the start of the first, second and third years, and the per-round mortality impact specified through formula (3.3), choosing the parameter values as $\text{logit}(\gamma) = -1$, $\text{log}(\alpha - 1) = 2$ and $\text{log}(\beta) = 0$. The yearly deaths D_{1j} in the screening arm were simulated from the binomial distribution, where π_j is given by (3.5). The conditional likelihood was maximised numerically with respect to $\theta = (\text{logit}(\gamma), \text{log}(\alpha - 1), \text{log}(\beta))$ using the `optim` function of R to obtain $\hat{\theta}$ and $\hat{\pi}_j = [1 - H((t_j - t_{j-1})/2; \hat{\theta})]/[1 + 1 - H((t_j - t_{j-1})/2; \hat{\theta})]$, and the expected event counts $d_j \hat{\pi}_j$. This procedure was repeated 10 000 times, with the sampling distributions of the maximum likelihood estimators and the goodness-of-fit test p -values displayed in Figure B1. The estimators do behave like maximum likelihood estimators, with approximately normal sampling distribution centred at the true values. Further, the goodness-of-fit test p -values are uniformly distributed under the null, with the Monte Carlo type I error rate at a 5% significance level of 0.0514.

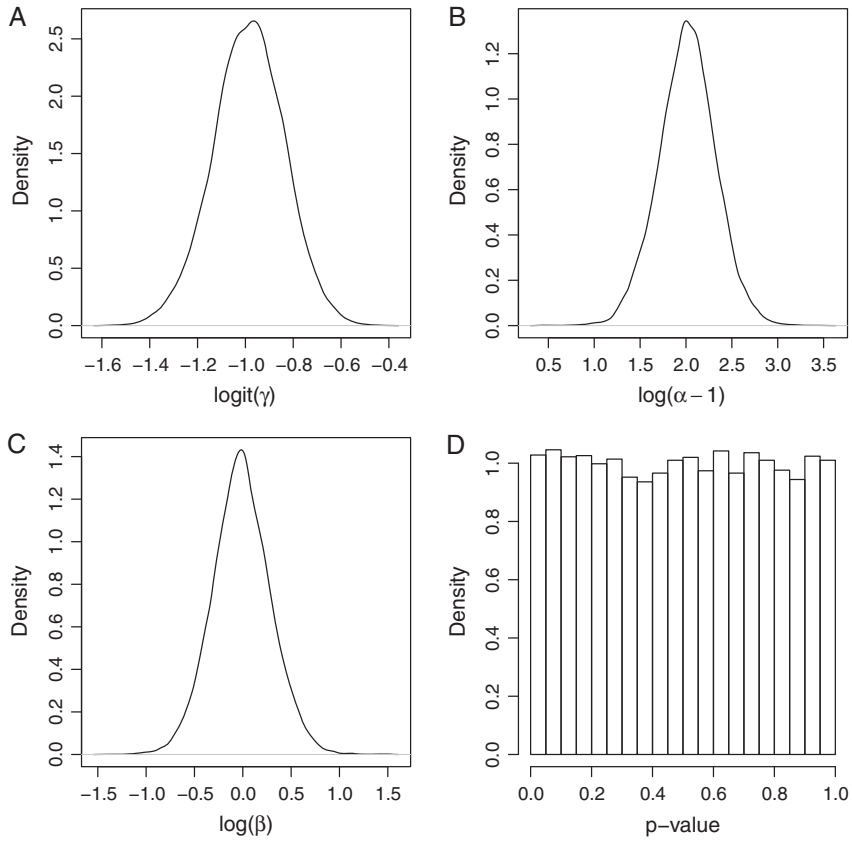


Figure B1. Panel A: Sampling distribution of $\text{logit}(\gamma)$. Panel B: Sampling distribution of $\log(\alpha - 1)$. Panel C: Sampling distribution of $\log(\beta)$. Panel D: Sampling distribution of the goodness-of-fit test p -values over 10 000 replications.