

Research article

Open Access

Comparing breast cancer mortality rates before-and-after a change in availability of screening in different regions: Extension of the paired availability design

Stuart G Baker*¹, Barrett S Kramer² and Philip C Prorok¹

Address: ¹Biometry Research Group, Division of Cancer Prevention, National Cancer Institute, Bethesda, MD 20892-7354, USA and ²Office of Disease Prevention, National Institutes of Health, Bethesda, MD 20892, USA

Email: Stuart G Baker* - sb16i@nih.gov; Barrett S Kramer - KramerB@OD.NIH.GOV; Philip C Prorok - pp2g@nih.gov

* Corresponding author

Published: 18 May 2004

Received: 21 January 2004

BMC Medical Research Methodology 2004, **4**:12

Accepted: 18 May 2004

This article is available from: <http://www.biomedcentral.com/1471-2288/4/12>

© 2004 Baker et al; licensee BioMed Central Ltd. This is an Open Access article: verbatim copying and redistribution of this article are permitted in all media for any purpose, provided this notice is preserved along with the article's original URL.

Abstract

Background: In recent years there has been increased interest in evaluating breast cancer screening using data from before-and-after studies in multiple geographic regions. One approach, not previously mentioned, is the paired availability design. The paired availability design was developed to evaluate the effect of medical interventions by comparing changes in outcomes before and after a change in the availability of an intervention in various locations. A simple potential outcomes model yields estimates of efficacy, the effect of receiving the intervention, as opposed to effectiveness, the effect of changing the availability of the intervention. By combining estimates of efficacy rather than effectiveness, the paired availability design avoids confounding due to different fractions of subjects receiving the interventions at different locations. The original formulation involved short-term outcomes; the challenge here is accommodating long-term outcomes.

Methods: The outcome is incident breast cancer deaths in a time period, which are breast cancer deaths that were diagnosed in the same time period. We considered the plausibility of the basic five assumptions of the paired availability design and propose a novel analysis to accommodate likely violations of the assumption of stable screening effects.

Results: We applied the paired availability design to data on breast cancer screening from six counties in Sweden. The estimated yearly change in incident breast cancer deaths per 100,000 persons ages 40–69 (in most counties) due to receipt of screening (among the relevant type of subject in the potential outcomes model) was -9 with 95% confidence interval (-14, -4) or (-14, -5), depending on the sensitivity analysis.

Conclusion: In a realistic application, the extended paired availability design yielded reasonably precise confidence intervals for the effect of receiving screening on the rate of incident breast cancer death. Although the assumption of stable preferences may be questionable, its impact will be small if there is little screening in the first time period. However, estimates may be substantially confounded by improvements in systemic therapy over time. Therefore the results should be interpreted with care.

Background

The paired availability design is a study design and method of analysis that reduces selection bias when using data from historical controls [1-4]. With standard historical controls, one compares (i) outcomes in subjects in the current time period who received treatment with (ii) outcomes in subjects in an earlier time period who did not receive treatment. Because of self selection (e.g., less healthy subjects might be more likely to receive treatment than more healthy subjects), results from standard historical controls may be substantially biased. In contrast, in the paired availability design there is no self-selection bias because a comparison is made between (i) outcomes in *all* subjects in the current time period when the intervention is more widely available and (ii) outcomes in *all* subjects in the previous time period when intervention was less widely available. To account for the change in availability of intervention between the current and previous time periods, Baker and Lindeman [1] proposed a potential outcomes model based on the intervention subjects would have received had they entered the study in a different time period. The model makes it possible to estimate efficacy, the effect of intervention among the type of subjects who would have only received the intervention during a period of increased availability, as opposed to effectiveness, the effect of a change in availability. The model requires various assumptions, best described in [4], that are plausible in many situations.

Estimating efficacy, as opposed to estimating effectiveness, is important when combining estimates from different locations (hospitals or regions). If the fraction of subjects who receive intervention differs among locations, it is difficult to interpret the overall estimate of effectiveness. In contrast, the overall estimate of efficacy is not confounded by varying the fraction of subjects who receive intervention in different locations.

Heretofore the paired availability design has only been formulated for evaluating the effect of an intervention on a *short-term* endpoint, namely the effect of epidural analgesia on the probability of Caesarian section [1-4]. To extend the paired availability design to breast cancer screening, we need to consider the implications of *long-term* endpoints.

Methods

Basic requirements

The first step in extending the paired availability design to the evaluation of breast cancer screening is to identify various geographic regions with a change in the availability of breast cancer screening from time period 0 to time period 1. To simplify this discussion, we presume that screening is more available in time period 1 than time period 0. The methodology is also applicable in the

unlikely situation in which the reverse were true in some or all regions. The change in availability is a change in the fraction of the eligible population who are invited for screening. Following Duffy et al [5], there are three basic design requirements to which we have added a fourth.

Requirement 1

The time periods should be sufficiently long to give screening sufficient time to maximize (or almost maximize) its impact on breast cancer mortality rates.

Requirement 2

For each geographic region, time periods 0 and 1 should be the same length.

Requirement 3

The outcome in each time period is incident breast cancer deaths, namely deaths from breast cancer in the specified time period arising from diagnosis of breast cancer during the same time period.

Requirement 4

We consider only situations in which most screening occurs at regular intervals of the same length during each time period.

Requirement 1 can be relaxed in the special case when the two time periods are separated by a time interval. In that case one need only require that non-overlapping observation times from the start of each time period past the end of each time period be sufficiently long to maximize the impact on breast cancer screening. However, as with randomized trials, if follow-up after the last screening is too long, there could be considerable dilution from breast cancers that could not have benefited from screening, and that would reduce the efficiency of the estimates [6].

The rationale for *Requirement 2* is that one wants the breast cancer mortality rates to be the same in the two time periods if screening has no effect and if there are no time-varying changes that could confound the results.

The rationale for *Requirement 3* is that by using incident breast cancer deaths (instead of all breast cancer deaths) as an outcome measure, one can avoid dilution from the breast cancer deaths that could not have benefited from screening [5,7-10]. However when using incident breast cancer deaths instead of all breast cancer deaths, there will be a preferential selection for screening evaluation of those breast cancers that cause death soon after diagnosis (as these deaths are more likely to occur in the same time period as diagnosis). With time periods substantially longer than the mean time between breast cancer diagnosis and death, this preferential effect is mitigated, as breast cancers occurring in a larger fraction of the time period (as

compared to the situation with short time periods) have a greater potential to cause death a long time after diagnosis in the absence of screening and still be included in the evaluation. Nevertheless, it is worth bearing this preferential selection in mind.

The rationale for *Requirement 4* is that the screening intervention must be comparable in the two time periods.

Potential outcomes model

For each before-and-after geographic region, our goal is to estimate the efficacy of breast cancer screening, which we define as the change in average yearly probability of incident breast cancer deaths due to the receipt of screening. (We later discuss combining estimates over all regions.) As proposed in the paired availability design [1-4], we use the following thought experiment to set the groundwork for estimating efficacy. Under this thought experiment, there are four types of subjects:

A, always-receivers, who would receive screening in either time period,

C, consistent-receivers, who would not receive screening in the time period with less availability and would receive it in the time period with greater availability,

I, inconsistent-receivers, who would receive screening in the time period with less availability and would not receive it in the time period with greater availability,

N, never-receivers, who would not receive screening in either time period.

For the sake of simplicity, we assume two conditions: (1) all-or-none behavior (i.e. an individual either receives all screens at the recommended interval or none, but does not switch back and forth), and (2) there is a single dominant screening test rather than a choice among screening tests of varying efficacy. In our application, there was only one screening modality.

Let π_{iAz} , π_{iCz} , π_{iIz} and π_{iNz} denote the probabilities of subject types *A*, *C*, *I*, *N*, respectively, in region *i* and time period *z*. Let β_{iAz} , β_{iCz} , β_{iIz} and β_{iNz} denote the probability of incident breast cancer death in time period *z* and region *i*, for subject types *A*, *C*, *I*, and *N*, respectively. The probability of incident breast cancer death in each time period is a mixture of the probabilities over all subject types in each time period,

$$\theta_{i0} = \pi_{iN0} \beta_{iN0} + \pi_{iC0} \beta_{iC0} + \pi_{iI0} \beta_{iI0} + \pi_{iA0} \beta_{iA0}, \text{ for time period 0,}$$

$$\theta_{i1} = \pi_{iN1} \beta_{iN1} + \pi_{iC1} \beta_{iC1} + \pi_{iI1} \beta_{iI1} + \pi_{iA1} \beta_{iA1}, \text{ for time period 1.} \quad (1)$$

As with the standard paired availability design, to ensure identifiably we restrict the estimation of efficacy to type *C* subjects. Let T_i denote the length of follow-up for time periods 0 and 1 for region *i*. We define the efficacy (for type *C* subjects) in region *i* as

$$\Delta_i = \frac{\beta_{iC1} - \beta_{iC0}}{T_i}. \quad (2)$$

The probability in (2) differs from a naive comparison of the effect of screening between subjects who receive screening in time period 1 and subjects who do not receive screening in time period 1. Instead Δ_i is the effect of receiving screening among type *C* subjects. Related potential outcome models were independently formulated for randomized trials with all-or-none compliance [11,12].

Assumptions

In order to estimate (2) we require the following assumptions adapted from the standard paired availability design [4].

Assumption 1. (Stable population)

The characteristics of the population that affect the probability of incident breast cancer death are constant over time.

Assumption 2. (Stable treatment)

The screening modality and therapy following diagnosis do not change over time.

Assumption 3. (Stable evaluation)

The outcome measure, which is incident cancer breast deaths, does not change in definition over time.

Assumption 4. (Stable preferences)

Factors affecting the decision to receive screening do not change over time.

Assumption 5. (Stable screening effects)

The effect of screening on the probability of incident breast cancer death rates does not change over time.

Assumption 1 is plausible if there is little immigration or out-migration related to screening. However, substantial immigration or out-migration of subpopulations with different underlying health or cancer risk can affect results over long time periods.

Assumption 2 is problematic in evaluating some screening modalities because the reduction in incident breast cancer deaths could result from better systemic therapies, such as chemotherapy and hormonal therapy [13]. It is possible that these changes in therapy could explain the decrease in

cancer mortality rates over time, even if screening has no benefit. Therefore it is particularly important to consider the plausibility of *Assumption 2*.

Assumption 3 is plausible, absent major changes in death code systems.

The basic idea of *Assumption 4* is that the screening intervention (including any campaigns to increase public awareness) should be the same in both time periods. If *Assumptions 1–4* hold, the probability of each subject type does not change over time period. In other words $\pi_{iAz} = \pi_{iA}$, $\pi_{iCz} = \pi_{iC}$, $\pi_{iIz} = \pi_{iI}$, and $\pi_{iNz} = \pi_{iN}$. In addition, by virtue of *Assumption 4*, there are basically no inconsistent receivers, i.e. $\pi_{iI} = 0$. If there is no screening in time period 0, so $\pi_{iAz} = \pi_{iCz} = 0$, and *Assumption 4* only requires $\pi_{iCz} = \pi_{iC}$ and $\pi_{iNz} = \pi_{iN}$, which is very plausible especially if one views public awareness as part of the screening intervention.

Assumption 5 likely holds for type *N* subjects because the same prior history of no screening applies to both time periods 0 and 1. Thus we can reasonably assume that the probability of incident breast cancer death among type *N* subjects does not depend on time period, i.e., $\beta_{iN0} = \beta_{iN1} \equiv \beta_N$. However, unless there is no screening in time period 0, *Assumption 5* will not hold for type *A* subjects. The reason is that (i) screening is generally more available *before* time period 1 than *before* time period 0, and (ii) prior screening may affect the probability of incident cancer death if screening confers benefit.

As a consequence of the above assumptions (and not applying *Assumption 5* to type *A* subjects), we can write (1) as

$$\begin{aligned} \theta_{i0} &= \pi_{iN} \beta_{iN} + \pi_{iC} \beta_{iC0} + \pi_{iA} \beta_{iA0} \\ \theta_{i1} &= \pi_{iN} \beta_{iN} + \pi_{iC} \beta_{iC1} + \pi_{iA} \beta_{iA1}. \end{aligned} \quad (3)$$

Because $\theta_{i1} - \theta_{i0} = \pi_{iC} (\beta_{iC1} - \beta_{iC0}) + \pi_{iA} (\beta_{iA1} - \beta_{iA0})$, we obtain from (3)

$$\Delta_i = \frac{\theta_{i1} - \theta_{i0} - \pi_{iA} (\beta_{iA1} - \beta_{iA0})}{\pi_{iC} T_i}. \quad (4)$$

If *Assumption 5* held for type *A* subjects, as in the usual paired availability design, $\beta_{iA1} = \beta_{iA0}$, and we would obtain the standard formula, averaged over the duration of the time period, for efficacy in the paired availability design, $\Delta_i = (\theta_{i1} - \theta_{i0}) / (\pi_{iC} T_i)$. We would also obtain the standard formula if there were no screening in time period 0 (and thus no type *A* subjects).

Estimates

In order to estimate (4) we need to estimate θ_{iz} , π_{iC} , π_{iA} , and β_{iAz} . Following the standard paired availability design we can estimate the first three parameters as follows. Let $s = 1$ if screening was received during the time period and 0 otherwise. Following *Requirement 4*, we assume most screening occurs at regular intervals during the time period. Let $\gamma = 1$ if incident breast cancer death, and 0 otherwise.

In the ideal scenario (Scenario I) the investigators would report data n_{izsy} , which is the number of subjects in region *i* and time period *z* with indicator of receipt of screening *s* and outcome γ . In the typical scenario (Scenario II), the only data in published reports are the numbers who received or did not receive screening n_{izs+} and the numbers with a given outcome γ but unknown screening status $n_{iz+\gamma}$ where "+" denotes summation over the indicated subscript. For both scenarios, we obtain the following estimates,

$$\hat{\theta}_{iz} = n_{iz+1} / n_{iz++} = \text{fraction of subjects in time period } z \text{ with incident cancer death} \quad (6)$$

$$\hat{\pi}_{iA} = n_{i01+} / n_{i0++} = \text{fraction who received screening in time period 0}, \quad (7)$$

$$\begin{aligned} \hat{\pi}_{iC} &= n_{i11+} / n_{i1++} - n_{i01+} / n_{i0++} \\ &= \text{fraction who received screening in time period 1 (a combination of types A and C) minus fraction who received screening in time period 0 (type A)}. \end{aligned} \quad (8)$$

If we had the full data n_{izsy} we could estimate β_{iA0} . However because subjects in time period 1 who receive screening are a combination of types *C* and *A*, we cannot uniquely estimate β_{iA1} . We discuss how to circumvent this difficulty in the two scenarios.

Scenario I: Full reporting of data

When there are full reporting of data, we can estimate θ_{izs} = $pr(Y = 1 | i, z, s)$ by $\hat{\theta}_{izs} = n_{izs1} / n_{izs+}$. Under the potential outcomes model, we write

$$\begin{aligned} \theta_{i00} &= \pi_{iN} \beta_{iN} + \pi_{iC} \beta_{iC0}, \quad \theta_{i01} = \pi_{iA} \beta_{iA0}, \\ \theta_{i10} &= \pi_{iN} \beta_{iN}, \quad \theta_{i11} = \pi_{iC} \beta_{iC1} + \pi_{iA} \beta_{iA1}. \end{aligned} \quad (9)$$

We introduce an exogenous parameter *h*, which is the relative risk for incident breast cancer death among type *A* subjects in time period 1 versus time period 0, namely,

$$\beta_{iA1} = h \beta_{iA0}. \quad (10)$$

We discuss specification of h in the section below on lead-time adjustment. Combining (9), (10) and (2) gives

$$\hat{\Delta}_i(Full) = \frac{\hat{\beta}_{iC1} - \hat{\beta}_{iC0}}{T_i} = \frac{\hat{\theta}_{i1+} - (\hat{\theta}_{i00} + \hat{\theta}_{i01}/h)}{T_i \hat{\pi}_{iC}} \tag{11}$$

The asymptotic variance is approximately

$$var(\hat{\Delta}_i(Full)) \approx \frac{\frac{\hat{\theta}_{i1+}(1-\hat{\theta}_{i1+})}{n_{i1}} + \frac{\hat{\theta}_{i00}(1-\hat{\theta}_{i00})}{n_{i0}} - 2\frac{\hat{\theta}_{i00}\hat{\theta}_{i01}}{n_{i0}h} + \frac{\hat{\theta}_{i01}(1-\hat{\theta}_{i01})}{n_{i0}h^2}}{(T_i)^2 (\hat{\pi}_{iC})^2} \tag{12}$$

Scenario II. Limited reporting of data

With limited reporting of data we introduce a second exogenous parameter k to essentially create the same estimates as with the full reporting of data. In particular we write

$$\theta_{i01} = k \theta_{i0}, \tag{13}$$

where $k = pr(S = 1 | Y = 1, Z = 0)$ = fraction of incident cancer deaths in time period 1 that are attributed to screening. If the full data were available, we would have an estimate of k and the methodology would be equivalent to that for Scenario I. In the absence of reported data, we propose a sensitivity analysis for k . A lower bound on k is 0 and an upper bound, assuming screening does not cause cancer deaths, is an estimate of the fraction screened in time period zero, namely $pr(S = 1 | Z = 0)$. Substituting (13) into (11) gives

$$\hat{\Delta}_i = \frac{\hat{\beta}_{iC1} - \hat{\beta}_{iC0}}{T_i} = \frac{\hat{\theta}_{i1} - \hat{\theta}_{i0}(1-k+k/h)}{T_i \hat{\pi}_{iC}} \tag{14}$$

where h is the same as in (10). We approximate the asymptotic variance by

$$var(\hat{\Delta}_i) \approx \frac{\frac{\hat{\theta}_{i1}(1-\hat{\theta}_{i1})}{n_{i1}} + \frac{\hat{\theta}_{i0}(1-\hat{\theta}_{i0})}{n_{i0}} (1-k+k/h)^2}{(T_i)^2 (\hat{\pi}_{iC})^2} \tag{15}$$

Using actual reported data from the limited data scenario, we checked the approximate variances in (12) and (15) by making reasonable assumptions to impute n_{iZSY} and then also computed the asymptotic variance using the delta method. The agreement was excellent: using the data in the example and assuming relative risk of incident cancer death of .7 for screened versus not screened, the approximate and exact asymptotic variance agreed to three significant digits.

Lead time adjustment related to prior screening

We specify a value for h in (10) by using the following argument based on lead time, which is the time from screen detection to clinical detection in the absence of screening. Duffy et al [5] discussed a related lead time

adjustment. The type A incident cancer deaths in a time period are composed of two subgroups: (i) subjects who would have been screen-diagnosed in the previous time period if there were screening in the previous time period and (ii) subjects who would *not* have been screen diagnosed in the previous time period if there were screening in the previous time period. Generally there is no screening prior to time period 0, so type A incident cancer deaths in time period 0 are composed of subgroups (i) and (ii). Because there is screening in time period 0 (as we are discussing type A subjects), type A incident cancer deaths in time period 1 are composed of only subgroup (ii). The parameter h in (10) is the ratio of type A incident cancer deaths in time period 1 to type A incident cancer deaths in time period 0. We approximate h by the number of subjects in subgroup (ii) divided by the number of subjects in the combination of subgroups (i) and (ii). Let L denote the mean lead time, which is approximately 2 years for breast cancer screening [14]. Subjects in subgroup (i) are, on average, detected on screening in the last L years of the time period. Assuming uniform detection rates, we further approximate h by T_i , the length of time of screen-detection in subgroup (ii), divided by $T_i + L$, the average length of time of screen detection in the combination of subgroups (i) and (ii), giving $h \approx T_i / (T_i + L)$. Importantly, if L is short relative to T_i there will be little bias as h would approximately equal one.

Lead time adjustment related to age-range at diagnosis

Because incident cancer cases are defined based on age at diagnosis, there is also a subtle bias [9] related to lead time. Suppose the age-range at diagnosis of breast cancer for incident breast cancer deaths is 40–69 in both time periods. Consider type C subjects who would be clinically diagnosed if in time period 0 and screen-detected if in time period 1. Let L denote the average lead time. Due to lead time, a type C subject clinically diagnosed with cancer in the age range (40, 40 + L - 1) in time period 0 would (on average) be screen detected in the age range (40 - L , 39) in time period 1 and would contribute to incident cancer deaths in time period 0 but not time period 1. Similarly a type C subject clinically diagnosed with cancer in the age range (70, 70 + L - 1) in time period 0 would (on average) be screen-detected with cancer in the age range (70 - L , 69) in time period 1 and would contribute to incident cancer deaths in time period 1 but not in time period 0. All other type C subjects would contribute to incident cancer deaths in both time periods. To approximately correct for the bias from a mean lead time of L years, suppose for simplicity that we divide age into three age groups 40–49, 50–59, 60–69, and assume the same number of subjects in each age group. A fraction (10 - L)/10 of type C subjects in the 40–49 age group in time period 0 would be counted toward incident cancer deaths in time period 0, but would not be counted toward incident cancer deaths

in time period 1. Also a fraction $(10 + L)/10$ type C subjects in the 60–69 age group in time period 0 would not be counted toward incident cancer deaths in time period 0, but would be counted toward incident cancer deaths in time period 1. Consequently, for this adjustment, we multiply θ_{i1} by $b \pi_{iC} + (1 - \pi_{iC})$, where

$$b = \frac{(10 - L)/10 D_{40} + D_{50} + D_{60} + (10 + L)/10}{D_{40} + D_{50} + D_{60}}, \tag{5}$$

and D_a is five-year cumulative mortality following breast cancer diagnosis at age a . With $L = 2$ [14] and approximating $D_{40} = .167$, $D_{50} = .131$, and $D_{60} = .124$, based on US population data [15], we obtain $b = .98$. In our example, the effect of lead-time bias due to a specified age-range is negligible and is therefore ignored.

Combining estimates over regions

To obtain a combined estimate of efficacy over all regions, we use a simple random effects meta-analysis [4,16,17].

Let $w_i = 1 / \text{var}(\hat{\Delta}_i)$ and let $\sigma^2 =$ the larger of $(Q - (r - 1)) / (\sum_i w_i - \sum_i w_i^2 / \sum_i w_i)$ and 0, where $Q = \sum_i w_i (\hat{\Delta}_i - m)^2$, $m = \sum_i \hat{\Delta}_i w_i / \sum_i w_i$. The random-effects weights are $w_i^* = 1 / (s_i^2 + \sigma^2)$, and the summary statistic is $\hat{\Delta} = \sum_i \hat{\Delta}_i w_i^* / \sum_i w_i^*$, with standard error $se(\hat{\Delta}) = \sqrt{1 / \sum_i w_i^*}$. The 95% confidence interval is $(\hat{\Delta} - t_{r-1} se(\hat{\Delta}), \hat{\Delta} + t_{r-1} se(\hat{\Delta}))$, where t_{r-1} is the value of the 97 / 2 percentile of a t -distribution with $r - 1$ degrees of freedom, where r is the number of regions. For an example of these calculations, see Table 1.

Results

We applied the methodology to before- and after-data on breast cancer screening in various Swedish counties [5]. The original data involved 7 counties. However, because one-third of the population of Dalarna county enrolled in a randomized trial, some of the assumptions might be violated for women screened in Dalarna county. The main difficulty with using data from Dalarna county is that subjects who refused screening in the randomized trial may not be comparable to subjects outside of the trial who did not obtain screening due to lack of availability, and the methodology does not allow for this difference in comparability. Therefore, in our analysis, we dropped data from Dalarna county. The age ranges were 40–74 for three counties, 40–69 for two counties, and 50–59 for one county.

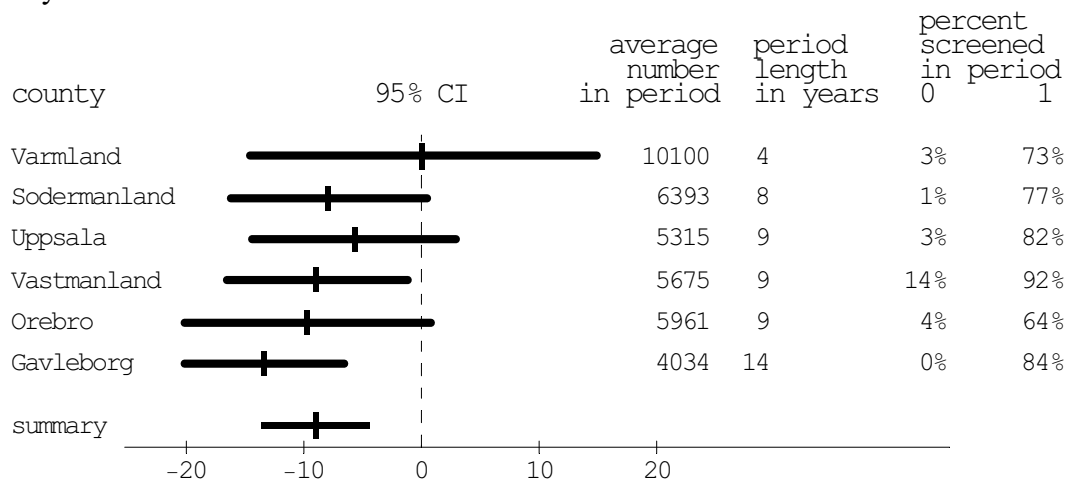
The data in [5] were reported in terms of person years of receiving screening. Dividing person years by the length of the time period we obtained the approximate number of persons eligible for screening in each region and group, n_{iz++} . Using these data, we estimated the change in the average yearly death rate of incident breast cancer among type C subjects ages 40–69 due to receipt of breast cancer screening as -9 per 100,000 with 95% confidence interval of $(-14, -4)$ per 100,000 for $k = 0$ and similarly -9 per 100,000 with 95% confidence interval of $(-14, -5)$ per 100,000 when k equaled the fraction screened. See Table 1 and Figure 1. The estimates were similar for the two values of k because only in Vastmanland County was there substantial screening in time period 0, and that was only 14%. We caution that Assumption 2 may not hold due to improvements in available systemic therapy over the periods of interest [14]. Therefore the results must be interpreted with caution, as they may overestimate the benefit of screening.

Table 1: Data and calculations

county	time period 0			time period I			yearly change in incident breast cancer deaths per 100,000	
	number eligible	incident breast cancer deaths	fraction screened	number eligible	incident breast cancer deaths	fraction screened	estimate	standard error
Varmland	39946	36	.03	40853	37	.73	.16	7.55
Sodermanland	49526	98	.007	53760	79	.77	-7.89	4.28
Uppsala	43340	110	.03	52426	112	.82	-5.73	4.43
Vastamanland	49827	112	.14	52329	85	.92	-8.88	3.93
Orebro	52936	133	.04	54366	108	.64	-9.74	5.36
Gavleborg	56794	311	.00	56145	219	.84	-13.40	3.64

Fraction screened is from Table 1 of [5]. Incident breast cancer deaths are from Table 3 of [5]. Number eligible is person years from Table 3 of [5] divided by years in time period from Table 2 of [5]. Calculations assume $k = 0$ for sensitivity analysis. Above yields $Q = 3.72$, $\sigma^2 = 0$, and overall estimate of -9 with 95% confidence $(-14, -4)$ per 100,000.

Analysis with $k = 0$



Analysis with $k = \text{fraction screened}$

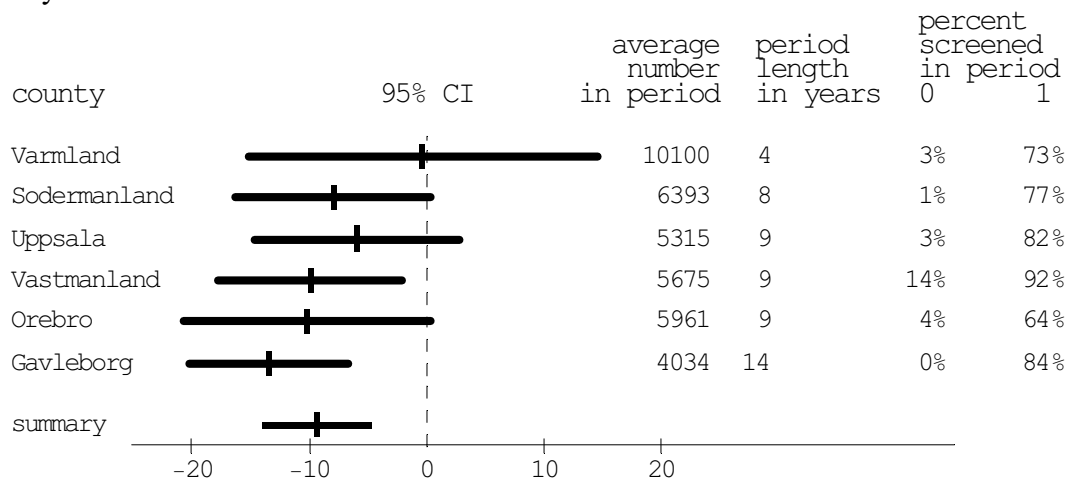


Figure 1

Estimated change (and 95% confidence intervals) in average yearly probability of incident cancer death due to receipt of screening per 100,000 type C subjects ages 40–69

Discussion

Our methodology complements that of Duffy et al. [5], who obtained qualitatively similar results (i.e. a statistically significant reduction in breast cancer mortality rates in time period 1 versus time period 0) based on data from seven Swedish counties. Unlike our paper, Duffy et al. estimated relative risks instead of risk differences. Duffy et al. [5] fit a Poisson regression model to data from all subjects in both time periods as well as data on those screened and not screened in the current time period. To adjust for

self-selection bias, Duffy et al. [5] fit a separate model to data from refusers and participants in a randomized trial in Dalarna County. An implicit assumption is that the self-selection adjustment based on refusers in the randomized trial would apply to women who did not receive screening outside the trial. This assumption is not used in the paired availability design, which does not use data from a randomized trial. However the paired availability design is subject to bias from changes in therapy over

time, as in any analysis comparing outcome in one period versus another.

Another approach to the analysis of before- and after- cancer screening data is to regress the change in cancer mortality rate over time for each region on the change in screening rates over time in each region. Sometimes the change in cancer incidence rates is used as a proxy for the change in screening rates [18]. This approach is an attempt to extract more information from the data, because larger changes in cancer screening rates should ideally correspond to larger changes in cancer mortality rates. However this type of regression based on population-level data can give different results than a regression based on individual-level data, a phenomenon known as the ecological fallacy [19].

With additional data, it may be possible to adjust for the effect of changes in therapy over time, if an additional assumption is reasonable. Suppose we had additional data on incident cancer deaths in time periods 0 and 1 in regions in which there was *no* screening in either time period. If (i) the therapies in these regions are representative of the therapies in the regions with screening and (ii) the population characteristics are similar to the population characteristics in regions with screening, one could reasonably estimate the effect of changes in therapy on the probability of incident cancer death.

Although data were reported on person-years of eligibility for screening, we did not use a survival analysis. A survival analysis can be incorporated into the potential outcomes model for all-or-none compliance [20]. However such an analysis requires more data than were reported. Also, in this framework, even a constant hazard model would involve a complicated likelihood calculation.

Besides this methodology based on the paired availability design, one could also analyze observational screening data using the method of periodic screening evaluation [21]. However this method requires regular screenings, data on the number of cancers detected on screening and in the intervals between screenings, and follow-up of subjects detected with cancer. The definitive method to evaluate cancer screening is a randomized trial [6]. Observational approaches have a role because such trials are very expensive and difficult to implement. Thus this extension of the paired availability design to the evaluation of cancer screening could play an important role in cancer screening evaluation, but only if there were no change in therapy over time or if one could adequately adjust for any effect of a change in therapy over time.

Conclusion

The paired availability design can be extended to the evaluation of breast cancer screening by using incident breast cancer deaths as the outcome and requiring sufficiently long equal-length time periods before and after a change in availability of periodic screening. However the assumptions should be examined carefully. The assumption of stable preferences may be violated by a campaign to encourage screening, although the impact would be greatly mitigated if there were little screening in time period 0. Also the assumptions regarding changes in therapy over time may also be violated.

Authors' contributions

SGB wrote the initial draft. BSK and PCP made substantial improvements to the manuscript.

Competing interests

None declared.

Acknowledgment

We thank Ping Hu for helpful comments.

References

1. Baker SG, Lindeman KS: **The paired availability design: A proposal for evaluating epidural analgesia during labor.** *Statistics in Medicine* 1994, **13**:2269-2278.
2. Baker SG: **The paired availability design: an update.** In: *Nonrandomized Comparative Clinical Studies* Edited by: Abel U, Koch A. Dusseldorf: Medinform-Verlag; 1998:79-84.
3. Baker SG, Lindeman KS: **Rethinking historical controls.** *Biostatistics* 2001, **2**:383-396.
4. Baker SG, Lindeman KL, Kramer BS: **The paired availability design for historical controls.** *BMC Medical Research Methodology* 2001, **1**:9 [<http://www.biomedcentral.com/1471-2288/1/9>].
5. Duffy SW, Tabar L, Chen HH, Holmqvist M, Yen MF, Abdallah S, Epstein B, Frodis E, Ljungberg E, Hedborg-Melander C, Sundbom A, Tholin M, Wiege M, Akerlund A, Wu HM, Tung TS, Chiu YH, Chiu CP, Huang CC, Smith RA, Rosen M, Stenbeck M, Holmberg L: **The impact of organized mammography service screening on breast carcinoma mortality in seven Swedish counties.** *Cancer* 2002, **95**:458-469.
6. Baker SG, Kramer BS, Prorok PC: **Statistical issues in randomized trials of cancer screening.** *BMC Medical Research Methodology* 2002, **2**:11 [<http://www.biomedcentral.com/1471-2288/2/11>].
7. Anntila A, Koskela J, Hakama M: **Programme sensitivity and effectiveness of mammography service screening in Helsinki, Finland.** *Journal of Medical Screening* 2002, **9**:153-8.
8. Hakama M, Pukkala E, Heikkila M, Kallio M: **Effectiveness of the public health policy for breast cancer screening in Finland: population based cohort study.** *British Medical Journal* 1997, **314**:864-7.
9. Jonsson H, Nystrom L, Tomberg S, Lenner P: **Service screening with mammography of women aged 50-69 years in Sweden: effects on mortality from breast cancer.** *Journal of Medical Screening* 2001, **8**:152-60.
10. Paci E, Duffy SW, Giorgi D, Zappa M, Crocetti E, Vezzosi V, Bianchi S, del Turco MR: **Quantification of the effect of mammographic screening on fatal breast cancers: The Florence Programme 1990-96.** *British Journal of Cancer* 2002, **87**:65-9.
11. Angrist JD, Imbens GW, Rubin DB: **Identification of causal effects using instrumental variables.** *Journal of the American Statistical Association* 1996, **92**:444-455.
12. Cuzick J, Edwards R, Segnan N: **Adjusting for non-compliance and contamination in randomized clinical trials.** *Statistics in Medicine* 1987, **16**:1017-1029.

13. Pater JL, Pritchard KI: **Optimal chemotherapy for women with breast cancer: The plot thickens.** *Journal of Clinical Oncology* 2003, **6**:963-964.
14. Walter SD, Day NE: **Estimation of the duration of pre-clinical disease state using screened data.** *American Journal of Epidemiology* 1983, **118**:865-86.
15. **SEER 1973–2001 Public-Use Data** [<http://www.seer.cancer.gov/publicdata/>]
16. DerSimonian R, Laird N: **Meta-analysis in clinical trials.** *Controlled Clinical Trials* 1986, **7**:177-188.
17. Follman DA, Proschan MA: **Inference in random effects meta-analysis.** *Biometrics* 1999, **55**:732-737.
18. Perron L, Moore L, Bairati I, Beranrd P, Meyer F: **PSA screening and prostate cancer mortality.** *Canadian Medical Association Journal* 2002, **166**:586-91.
19. Piantadosi S, Byar DP, Green SB: **The ecological fallacy.** *American Journal of Epidemiology* 1988, **127**:893-904.
20. Baker SG: **Analysis of survival data from a randomized trial with all-or-none compliance: estimating the cost-effectiveness of a cancer screening program.** *Journal of the American Statistical Association* 1998, **93**:929-934.
21. Baker SG, Erwin D, Kramer BS, Prorok PC: **Using observational data to estimate an upper bound on the reduction in cancer mortality due to periodic screening.** *BMC Medical Research Methodology* 2003, **3**:4 [<http://www.biomedcentral.com/1471-2288/3/4>].

Pre-publication history

The pre-publication history for this paper can be accessed here:

<http://www.biomedcentral.com/1471-2288/4/12/prepub>

Publish with **BioMed Central** and every scientist can read your work free of charge

"BioMed Central will be the most significant development for disseminating the results of biomedical research in our lifetime."

Sir Paul Nurse, Cancer Research UK

Your research papers will be:

- available free of charge to the entire biomedical community
- peer reviewed and published immediately upon acceptance
- cited in PubMed and archived on PubMed Central
- yours — you keep the copyright

Submit your manuscript here:

http://www.biomedcentral.com/info/publishing_adv.asp

