International Epidemiological Association

International Journal of Epidemiology, 2022, 1910–1919
https://doi.org/10.1093/ije/dyac096
Advance Access Publication Date: 13 May 2022
Original article

OXFORD

Methods

Evaluating the effectiveness of population-based breast cancer service screening: an analysis of parsimonious patient survival information with the time-varying Cox model

Rene Wei-Jung Chang,[1] Grace Hsiao-Hsuan Jen,[2] Kuan-Chia Lin,[3] Tsung-Chi Cheng,[4] Shao-Yuan Chuang,[5] Shin-Liang Pan ,[6] Tony Hsiu-Hsi Chen[1] and Amy Ming-Fang Yen[2]*

[1]Institute of Epidemiology and Preventive Medicine, College of Public Health, National Taiwan University, Taipei, Taiwan, [2]School of Oral Hygiene, College of Oral Medicine, Taipei Medical University, Taipei, Taiwan, [3]Institute of Hospital and Health Care Administration, Community Medicine Research Center, Preventive Medicine Research Center, National Yang-Ming University, Taipei, Taiwan, [4]Department of Statistics, National Chengchi University, Taipei, Taiwan, [5]Institute of Population Health Science, National Health Research Institutes, Miaoli, Taiwan and [6]Departments of Physical Medicine and Rehabilitation, National Taiwan University Hospital and College of Medicine, National Taiwan University, Taipei City, Taiwan

*Corresponding author. School of Oral Hygiene, College of Oral Medicine, Taipei Medical University, No. 250, Wuxing St., Xinyi Dist., Taipei City 110, Taiwan. E-mail: amyyen@tmu.edu.tw

Received 23 May 2021; Editorial decision 6 April 2022; Accepted 24 April 2022

Abstract

Background: This study is aimed at estimating the unbiased effectiveness of population-based breast cancer service screening based on case survival information alone rather than large-scale individual screening data pursuant to the intention-to-treat principle of a randomized–controlled trial.

Methods: A novel time-dependent switched design with two modalities of cancer detection (screen-detected vs clinically detected) was proposed to evaluate the effectiveness of breast cancer screening. We used data on 767 patients from Kopparberg in the Swedish Two-County trial and on 78 587 patients in the Taiwan population-based service screening. We estimated the relative rate of the screen-detected vs the clinically detected with adjustment for both truncation and lead-time biases. The absolute effectiveness in terms of the number needed to screen (NNS) for averting one death from breast cancer was estimated.

Results: The relative rate of effectiveness was estimated as 33%, which was consistent with the 37% reported from the original Swedish randomized–controlled trial. The corresponding estimate for the Taiwan screening programme was 42%, which was also very close to that estimated using individual screening history data (41%). Both relative estimates were further applied to yield 446 and 806 of NNS for averting one death from breast cancer for the corresponding two data sets.

Conclusion: The proposed time-dependent switched design and analysis with two modalities of case survival information provides a very efficient means for estimating the unbiased estimates of relative and absolute effectiveness of population-based breast cancer service screening dispensing with a large amount of individual screening history data.

Key words: ITT effectiveness, time-dependent, survival analysis, mammography screening

Key Messages

- This is the first study to evaluate the unbiased effectiveness of population-based service screening as observed in a randomized–controlled trial or the equivalent trial with a time-dependent switched design relying on case survival information only.
- The results of the estimated intention-to-treat effectiveness using the intervention arm of the Swedish randomized–controlled trial only and Taiwan service screening programme were comparable to their corresponding findings based on large-scale individual data and the control group required for traditional evaluation.
- Evaluating case survival information alone with the proposed time-dependent switched design answers the true effectiveness of early detection and dispenses with the randomized–controlled design and large-scale individual screening data.

Background

Evaluating the effectiveness of population-based cancer service screening programmes is essential to the principle of evidence-based medicine for cancer prevention.[1–4] Despite a series of randomized–controlled trials that have already demonstrated the benefit of screening at the population level,[1–3] whether the effectiveness of reducing specific-cause mortality in an organized service screening programme can be reproducible from randomized–controlled trials is still uncertain because the factors associated with effectiveness are manifold, including the basic characteristics (such as participation rate) of screening, early detection in relation to the natural disease history and survival related to the prognosis of patients. The failure to consider these characteristics based on individual and time-stamped screening history data often renders the estimated results of effectiveness vulnerable to different kinds of biases, including selection bias, lead-time bias and length bias, when evaluating population-based service screening beyond a randomized–controlled trial. However, collecting a large amount of population-based screening history data is very costly and laborious.

To obtain an unbiased estimate of the effectiveness of population-based service screening with efficiency, one solution is to evaluate the survival of two main modalities, screen-detected cancers and clinically detected cancers, merely relying on patient survival data. Screen-detected cancers include prevalent screen-detected cancers and subsequent screen-detected cancers, and clinically detected cancers include interval cancer and cancers in non-participants. Unfortunately, numerous previous studies have overestimated the survival benefit for patients with screen-detected compared with clinically detected cancers as a result of lead-time bias and length bias. Lead-time bias is defined as an artificial extension of survival time because of only advancing the date of diagnosis after the administration of screening rather than prolonging the life that is actually conferred by screening and early detection. The length bias pertains to oversampling cancers with long sojourn time cancers at the point of screening, particularly the first round of prevalent cases, leading to the overestimation of survival benefit, as the longer the sojourn time, the better the prognosis.[5]

Several previous studies focusing on survival by detection mode have been proposed to address lead-time and length bias when the survival of patients with screen-detected vs clinically detected cancers was compared. Duffy[6] proposed an idea of excluding the prevalent screen-detected cases from the invited group to reduce length bias arising from prevalent cases to render an 'unbiased set' comparable to cases arising from the control group. Paci[7] compared the survival between the invited and the control using not-yet-invited patients to approximate the effectiveness as close as possible to the effectiveness with intention-to-treat (ITT) analysis. However, these studies are still dependent on the use of an external control group. Duffy[8] further compared the survival of patients with screen-detected and clinically detected cancers through sensitivity analysis with different magnitudes of length bias and lead time when the control group was waived. However, the adjustment of lead-time

and length bias in their study was treated as two independent factors and was also entirely subject to the distribution of the sojourn time applied to the parametric model.

To relax the assumption of the parametric model, the semi-parametric time-dependent Cox regression model may be considered.[5,9–11] Warwick[12] has already used a time-dependent model to assess whether the effect of tumour attributes on the hazard of breast cancer death varied with the follow-up time, using the analogue of the landmark analysis[13,14] to elucidate how the hazard ratio of each time-dependent covariate varies with the landmark time. These findings suggest that length bias cannot be overemphasized.

In addition, using only survival data may also hardly be amenable to estimating the ITT efficacy of screening at the population level, as self-selection bias involved with the behaviour of the uptake of screening is not considered in previous survival models. The idea we come up with here is to make better use of truncated cases being diagnosed clinically before attending the first screen. These truncated cases may indirectly convey messages on the self-selection behaviour of attending screening on the premise that non-participants are disproportionally more prone to having advanced disease and poor survival, and because of rapid progression from the pre-clinical detectable phase to the clinical phase.[4] In an organized service screening, those non-participants may be switched to participants in a time-dependent manner if they are not diagnosed with cancer before attending the first screen. It is therefore interesting to develop a switched design and time-dependent Cox regression model for tackling the problem of self-selection and taking lead-time and length bias into account when evaluating the effectiveness of an organized service screening only relying on survival information dispensing with individual screening history and the control group following the ITT principle.

The aim of this study was to estimate the ITT efficacy of population-based screening with the proposed switched design and time-dependent Cox proportional hazards regression model using survival data available from the invited group for the uptake of screening. We then applied this proposed model to the Kopparberg arm of the Swedish Two-County trial and Taiwanese service screening data. The ITT efficacy estimated based on survival data was further validated by the empirical results on a population-based randomized–controlled trial and service screening data published previously on the basis of individual screening history data.

Methods

Time-dependent switched design

Basically, it is customary to estimate the effectiveness of population-based cancer screening (such as breast cancer) with an ITT analysis under a randomized–controlled trial (RCT) design, as shown in the left panel of Figure 1, for which the eligible population for screening was randomized into an active study population and a passive study population, as seen from the Kopparberg arm of the Swedish Two-County trial.[3] The effectiveness of cumulative mortality between the two groups was compared and the effect size regarding the relative rate of mortality from breast cancer between the two groups was estimated with an ITT analysis. However, such a population-based design for the unbiased ITT analysis would require individual screening history data that may involve enormous costs and the difficulty of logistic follow-up. Most importantly, the estimate of effectiveness with an ITT principle based on a randomized–controlled trial also needs the control group, but it is impractical to have the control group in a population-based organized service screening programme, although the use of a pre-screen period or non-participant group can be used as a proxy for the comparator, an analogue to the control group in an RCT. However, the use of such a proxy is subject to different threats, such as varying baseline risks in different periods while using the pre-screening period and self-selection bias while using the non-participants.

Here, we propose a binary time-varying design with two main cancer detection modes—screen detection and clinical detection—using only cancer case survival information to estimate the effectiveness of population-based screening under the ITT principle. As shown in the right panel of Figure 1, we used cancers for the following binary time-dependent Cox model. Instead of using exposed or unexposed to screen often used in population-based screening data, the effectiveness of survival was evaluated by two detection modes: screen-detected cases ($X = 1$) or clinically detected cases ($X = 0$). Among screen-detected cancers, those detected by the first screen are defined as prevalent screen-detected cancers and those detected by the subsequent screen are defined as subsequent screen-detected cancers. Among clinically detected cancers, those diagnosed by the presence of symptoms or signs between two screens and from non-participants are defined as interval cancer cases and refusers, respectively. As our goal is to provide an unbiased estimate of effectiveness for one-arm population-based service screening, we postulate that the comparison of survival between the screen-detected group and the clinically detected group is analogous to the comparison of survival between the invited group and the uninvited group, as seen in the theoretical RCT design (left panel of Figure 1).

The comparison of screen-detected and clinically detected cases for reaching unbiased effectiveness relies on two pairs: (i) subsequent screen-detected cancers vs interval cancers and (ii) prevalent screen-detected cancers vs refusers. However, such a comparison without the control group as in the RCT design is often compromised by

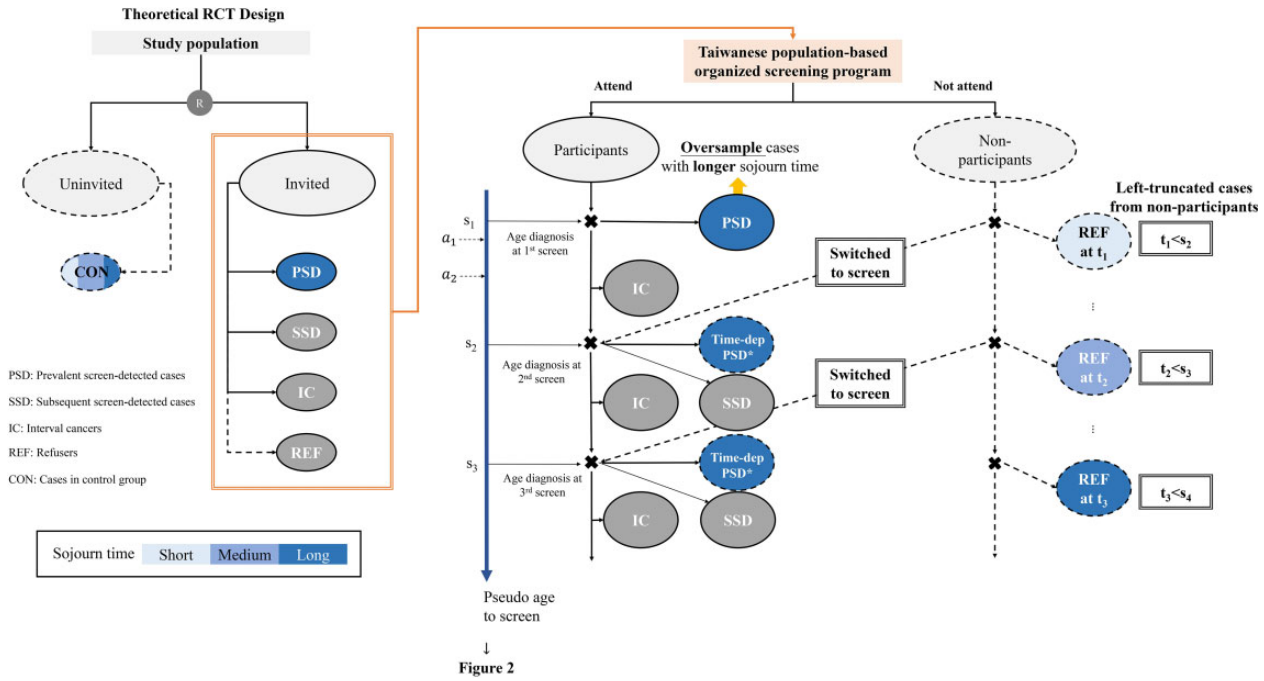

Figure 1 Theoretical randomized–controlled trial design (left side) and time-dependent switched screen with case survival design (right side)

selection bias between participants and non-participants. Compared with the second pair, the first pair may be less likely to be affected by selection bias, as both modes have experienced the same screening trajectory but are merely diagnosed by different detection modes. Unfortunately, it is not such a case for the second pair if a binary switched mode for prevalent screening from non-participants cannot be allowed for (the right panel of Figure 1) as each invited individual may have the staggered entry to attend screening that turns the detection of prevalent screen-detected cancers via first screen into a binary switched mode, depending on when individuals were first screened. Non-participants with cancers detected at first screen were classified as prevalent screen-detected cancers (X = 1) in contrast to those who had been diagnosed with breast cancer (X = 0) before the chance of being switched to first screen (truncated from first screen). The failure to consider the waiting time for the first screen clearly involves serious length bias caused by such a selective process between prevalent screen-detected cases and refusers because prevalent screen-detected cases from participants or those who switched from non-participants had longer sojourn times compared with refusers. The better survival for prevalent screen-detected cases compared with refusers may not be entirely attributed to screening, but to selection bias for cases with longer sojourn time. The earlier the age at diagnosis at first screening, the longer the sojourn time of the cases. To tackle this issue, we flagged pseudo age to screen (a_k) according to the ranking of age of death from breast cancer to allow for when prevalent screen-detected cases switched from clinical detection (X = 0) to screen detection (X = 1).

Figure 2 shows how to apply the time-dependent switched mode from clinical detection (X = 0) to screen detection (X = 1), depending on the pseudo age of entry flagged by each age of death, to obtain an unbiased risk set of partial likelihood for both screen and clinical detection modes. The use of pseudoage to screen given the failure time expressed by age at death has the advantage of removing the influence of age of entering the pre-clinical detectable phase, equivalent to screen-detected status and sojourn time for surfacing to clinical detection and lead time for screen detection, both of which are supposed to be incorporated into baseline hazards as age at death covering these two time horizons and survival time. Such a baseline hazard is well known to be able to be cancelled out between the event of numerator and the risk set of denominator given each of failure time while the conditional partial likelihood is formed. Therefore, such a binary time-dependent switched design can be adopted to estimate the unbiased effectiveness of screening in population-based organized service screening with only information on the survival of two detection modes of breast cancer dispensing with population-based individual screening history.

Data sources

To apply our proposed design and method to estimate the unbiased estimate of effectiveness, particularly in the

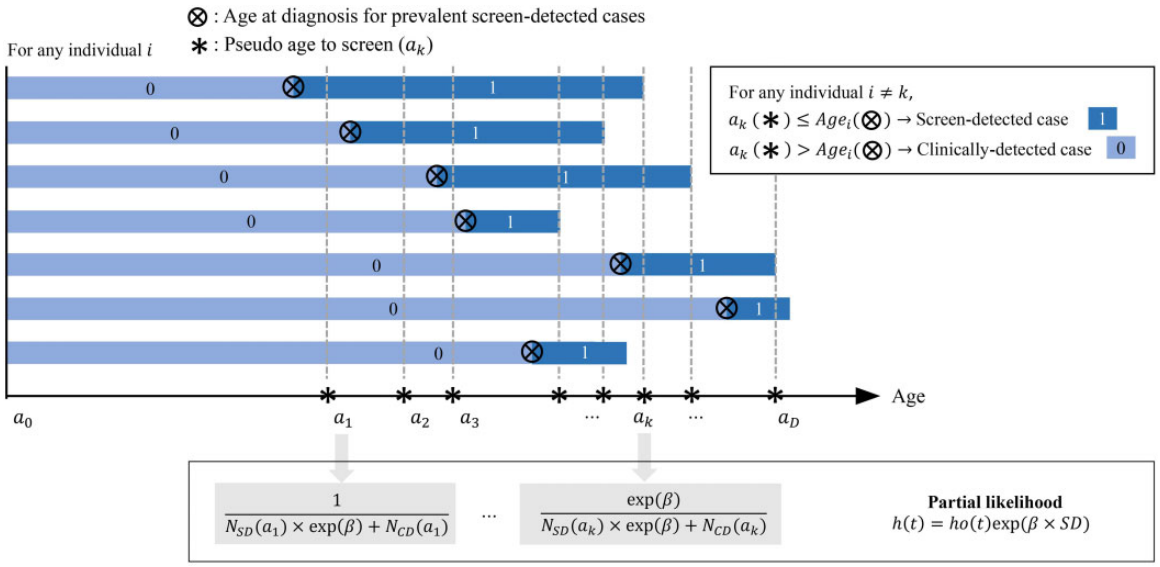

Figure 2 Illustration of time-varying exposure to a prevalent screen by pseudo age to screen

scenario of a population-based organized service screening programme, we first exploited the one-arm data from the previous population-based randomized–controlled trial and dropped the data on the control group in the absence of screening. We then estimated the effectiveness of mortality reduction using our switched design and compared the result with the result obtained from the same follow-up time of the trial. We then applied the same design and methodology to another population-based organized service screening. Two data sources are therefore used and described as follows.

Swedish breast cancer screening

A randomized–controlled trial for breast cancer screening with mammography started between 1977 and 1978 for Kopparberg and Östergötland, respectively, in Sweden. This trial is also called the Swedish Two-County trial. The eligible age for screening was between 40 and 74 years old, and the inter-screening intervals were 24 and 33 months for women younger or older than 50 years old. The original study reported the population-based effectiveness of mortality reduction based on individual screening history data on mammography screening after an 8-year study period.[2] The unbiased effectiveness following the ITT analysis after 30 years of follow-up was also reported based on individual follow-up data.[3]

In this analysis, instead of collecting data from an eligible population of ~50 000 women in the Kopparberg arm of the Swedish Two-County trial, we used survival data only to evaluate the effectiveness of screening on mortality reduction. Because women aged 70–74 years had a lower attendance rate during the trial period, the screen was stopped in the second round. We only evaluated the efficacy for women aged 40–69 years and included all the identified breast cancer patients in the trial period. A total of 1053 women were diagnosed with breast cancer in the original survival data, but we excluded cases in the control group ($n = 255$) and those who were aged 70–75 years ($n = 17$) or had a breast cancer diagnosis between randomization and invitation from our analysis ($n = 14$). The reason for excluding the latter is that such prior cancers detected before randomization would not occur in the scenario of population-based service screening, which is the major goal of the targeted group under evaluation. Nonetheless, the results of ITT effectiveness in reducing breast cancer mortality had only a minor change (33–34%) when these prior cancers and the oldest age group were also included in the following analysis. Data on the date of diagnosis, date of death, cause of death and detection mode were collected at the patient level. Information on tumour attributes, including size and node, was also included.

Taiwan breast cancer screening

To demonstrate whether the proposed methods can be applied to an organized service screening, we used data from the Taiwan breast cancer service screening programme from which a biennial mammography screening programme has been provided for women aged 50–69 years since 2004. The lower bound of screening age was adjusted to 45 years old in 2012. Survival information on breast cancer cases was retrieved from the Taiwanese Cancer Registry to include the date of diagnosis, treatment, death and cancer staging.

Regarding the detection mode, patients attending screens between 2004 and 2015 and identified upon screen

(prevalent and subsequent screen detection), diagnosed due to clinical symptoms after a negative screen (interval cancer) and those never attending the screening programme (refusers) were included to evaluate the effectiveness of the Taiwan mass screening programme in our analysis.

Statistical analysis

As indicated earlier, a binary time-dependent switched design was adopted to estimate the unbiased estimate of effectiveness in population-based breast cancer screening. The corresponding time-dependent Cox proportional hazards regression model was therefore applied here for adjusting self-selection of the truncation issue and the lead-time bias inherent from the population-based screening programme. The truncation issue is derived from oversampling slow-progression breast cancer in prevalent screening among participants because of the left truncation of clinically detected breast cancer before they were switched to first screen among non-participants as mentioned earlier.

The switched detection mode of prevalent screen-detected cases is illustrated in Figure 2. After sorting age of death from breast cancer, there are D distinct ages at death used to flag pseudoage to screen (a_1–a_D). Given any pseudoage to screen (a_k) ($a_1 < \ldots < a_k < \ldots < a_D$), our idea of this methodology is to compare the age of diagnosis for prevalent screen-detected cases at risk with pseudoage to screen. Among prevalent screen-detected cases, if the age of diagnosis for the ith individual was $\geq a_k$, the detection mode of this individual was switched from clinical to screen detection. However, if the age of diagnosis for the ith individual was $< a_k$, the detection mode was treated as clinical detection at a_k.

Therefore, at each pseudoage to screen (a_k), the detection mode of prevalent screen-detected cases was time-dependent. The time-dependent Cox model can be written as follows:

$$h(a_k) = h_o(a_k)\exp\left(\beta \times X(a_k)\right)$$

As shown in Figure 2, the construction of partial likelihood at each pseudoage to screen depends on the number of screen-detected and clinically detected cases at that time point denoted by $N_{SD}(a_k)$ *and* $N_{CD}(a_k)$. The partial likelihood at a_k can be written as:

$$L(\beta) = \begin{cases} \dfrac{1}{N_{SD}(a_k) \times \exp(\beta) + N_{CD}(a_k)} & \text{if } X(a_k) = 0 \\ \dfrac{\exp(\beta)}{N_{SD}(a_k) \times \exp(\beta) + N_{CD}(a_k)} & \text{if } X(a_k) = 1 \end{cases}$$

where $X(a_k)$ is the detection mode for an individual at risk at a_k. More importantly, when age at death is flagged as the pseudoage of entry to screen that covers three time epochs: age for developing pre-clinical detectable breast cancer, the sojourn time for surfacing to clinical detection and lead time for screen detection, and the survival time, the merit of using the comparison between screen and clinical detection would render the baseline risk of being in the pre-clinical detectable phase and sojourn time or lead time adjusted through the conditional detection mode encoded in the partial likelihood as mentioned earlier.

In screening scenarios, overdiagnosed cases may also influence the evaluation of effectiveness.[15] However, as the failure time used here is the age of death, overdiagnosis cases would be treated as right-censoring in the partial likelihood function in our analysis and would not be flagged as one of pseudoage of entry to screen, as shown in the last case of the bottom panel of Figure 2.

Number needed to screen to avert one breast cancer death

Once the unbiased effectiveness can be obtained using a time-dependent switched design, the additional merit of the study design is the further application of the self-selection bias formula to yield an estimate of absolute cumulative mortality of the control group (M_c) in the absence of screening, provided that the attendance rate (C) and person-years of the exposed group and the unexposed group can be approximated given the assumption of rare events of death to calculate the cumulative mortality rate for the exposed group (M_E) and the unexposed group (M_{UE}), respectively:

$$M_c = \frac{M_E \times \mathrm{C} + M_{UE} \times (1 - \mathrm{C})}{\mathrm{ITT} - \text{Relative rate of mortality}} \quad (1)$$

Using the cumulative mortality of the control group as obtained from Equation (1), the number needed to screen (NNS) to avert one breast cancer death can be calculated using Equation (2):

$$\begin{aligned}\mathrm{NNS} = [1/(&\text{Cumulative mortality rate for control group} \\ &- \mathrm{ITT} - \text{Cumulative adjusted mortality rate})]\end{aligned} \quad (2)$$

Results

Table 1 shows the numbers of breast cancer cases and deaths by detection mode in the Kopparberg arm of the Swedish Two-County trial and Taiwan service screening programme.

The Swedish Two-County trial—Kopparberg arm

Recall that to examine whether our proposed methods can be applied to an organized service screening programme when there is a lack of a control group, the fair comparator,

Table 1 Frequencies of breast cancer cases and deaths by detection mode in two studies

Study and detection modes	Number of patients	Breast cancer death	
		Number	%
Swedish RCT data[a]			
Screen-detected cases	586	97	16.55
Clinically detected cases	212	78	36.79
Taiwan service data[b]			
Screen-detected cases	24 739	1574	6.36
Clinically detected cases	53 848	9192	17.07

[a]Patients diagnosed between 1977 and 1987, and the vital status followed up in 2003 (mean follow-up time = 13.69 years).

[b]Patients diagnosed between 2004 and 2015, and the vital status followed up in 2017 (mean follow-up time = 6.76 years).

RCT, randomized–controlled trial.

we discarded data from the control group and estimated the hazard ratios of screen-detected cases compared with clinically detected cases. Table 2 shows that the unadjusted hazard ratio for screen-detected cases was 0.33 (95% CI: 0.24, 0.44), whereas the adjusted relative risk after considering truncation and lead-time bias was inflated to 0.67 (95% CI: 0.48, 0.93) after calibration. This estimate was close to the unbiased effectiveness extracted from the original randomized–controlled trial study [0.63 (95% CI: 0.42, 0.96)].[2] As both estimates were very close, the current finding suggests the proposed time-dependent switched design can be applied to the cases only for approximating the unbiased effectiveness with an ITT analysis dispensing with the collection of a large amount of individual screening history data and without relying on the control group.

With further adjustment for tumour attributes, therapies and surgery, the adjusted hazard ratio of screen-detected cases was further increased to 0.99 (95% CI: 0.70, 1.39), as shown in Table 3, suggesting that >90% of the survival benefit of screen-detected cases can be explained by early detection of breast cancer by screening through the surrogate endpoint of favourable tumour attributes for the primary endpoint of mortality.

Taiwan service screening programme

Although the proposed time-dependent switched method was applied to the Taiwan service screening programme, the unadjusted effectiveness in terms of mortality reduction was estimated as 66% (95% CI: 64%, 68%) and was increased to 42% (95% CI: 39%, 45%) after calibration (Table 2). This estimate was also close to the value estimated from a previous study using population-based individual screening history data that used the physical examination group as the comparator [41% (95% CI: 27%, 52%)].[4]

With further adjustment of tumour attributes by stage, therapies and surgery, the calibrated effectiveness was inflated to 1% (95% CI: –6%, 5%), suggesting that 99% of the survival benefit of screen-detected cases in the Taiwan screening programme can be explained by the surrogate endpoint of tumour attributes, as screen-detected cases had favourable tumour attributes, as shown in Table 4.

Table 5 shows the results of the NNS estimates to avert one breast cancer death based on the adjusted relative risk of reducing breast cancer mortality on the Kopparberg arm of the Swedish Two-County trial and Taiwan service programme, which were 0.67 and 0.58, respectively, as shown in Table 2. By combining information on the compliance rate and the cumulative mortality rate of the exposed group (M_E) and the unexposed group (M_{UE}) using the formula for the ITT – relative rate of mortality based on Equation (1), the cumulative mortality rate of the control group (M_C) was ~679 per 100 000 person-years for the Swedish RCT after 25 follow-up years since randomization. The NNS was 446.

The corresponding figure of the NNS was 806 for the Taiwan service programme given that the estimated cumulative mortality rate of the control group (M_C) was ~295 per 100 000 person-years when the follow-up since diagnosis was comparable to the follow-up since diagnosis of breast cancer patients in the Swedish RCT.

Discussion

We proposed a novel time-dependent switched design and Cox survival model design to estimate the ITT effectiveness

Table 2 Univariate analysis for unadjusted and adjusted hazard ratios by detection mode

	Per-protocol analysis Estimation (95% CI)	Intention-to-treat analysis Estimation (95% CI)	True efficacy Estimation (95% CI)
Swedish RCT data			
Screen-detected cases	0.33 (0.24, 0.44)	0.67 (0.48, 0.93)	0.63 (0.42, 0.96)[a]
Clinically detected cases	1.00	1.00	1.00
Taiwan service data			
Screen-detected cases	0.34 (0.32, 0.36)	0.58 (0.55, 0.61)	0.59 (0.48, 0.73)[b]
Clinically detected cases	1.00	1.00	1.00

[a]Ref. [2]; [b]Ref. [4].

RCT, randomized–controlled trial.

Table 3 Multivariate analysis for unadjusted and adjusted hazard ratios by detection mode for Swedish randomized–controlled trial data

	Per-protocol analysis		Intention-to-treat analysis	
	Estimation	95% CI	Estimation	95% CI
Detection mode				
Screen-detected cases	0.51	(0.37, 0.71)	0.99	(0.70, 1.39)
Clinically detected cases	1.00	–	1.00	–
Tumour information				
Node				
Node-positive	2.49	(1.66, 3.74)	2.18	(1.46, 3.26)
Node-negative	1.00	–	1.00	–
Size				
$10 \leq$ Size <15	1.26	(0.58, 2.77)	1.29	(0.59, 2.82)
$15 \leq$ Size < 20	1.48	(0.68, 3.21)	1.52	(0.70, 3.32)
$20 \leq$ Size < 30	2.01	(0.94, 4.28)	2.33	(1.10, 4.97)
Size ≥ 30	3.72	(1.72, 8.04)	3.78	(1.74, 8.19)
$1 \leq$ Size < 10	1.00	–	1.00	–
In situ	0.31	(0.07, 1.43)	0.37	(0.08, 1,73)
Missing	3.89	(1.72, 8.79)	3.58	(1.59, 9.06)
Surgery				
Mastectomy	1.43	(0.82, 2.50)	1.74	(1.00, 3.03)
Breast conservative surgery	1.00	–	1.00	–
Therapy[a]				
Yes	1.06	(0.74, 1.53)	1.38	(0.97, 1.95)
No	1.00	–	1.00	–

[a]Chemotherapy, radiotherapy and hormone therapy.

Table 4 Multivariate analysis for unadjusted and adjusted hazard ratios by detection mode for Taiwan service data

	Per-protocol analysis		Intention-to-treat analysis	
	Estimation	95% CI	Estimation	95% CI
Detection mode				
Screen-detected cases	0.54	(0.51, 0.57)	0.99	(0.94, 1.05)
Clinically detected cases	1.00	–	1.00	–
Cancer stage				
0	0.53	(0.44, 0.63)	0.53	(0.45, 0.64)
I	1.00	–	1.00	–
II	2.81	(2.56, 3.07)	2.87	(2.62, 3.15)
III	9.00	(8.23, 9.84)	9.23	(8.44, 10.09)
IV	30.52	(27.71, 33.62)	26.02	(23.63, 28.66)
Missing	13.90	(12.71, 15.20)	13.04	(11.93, 14.26)
Surgery				
Mastectomy	1.82	(1.70, 1.94)	1.58	(1.48, 1.69)
Breast conservative surgery	1.00	–	1.00	–
Therapy[a]				
Yes	1.17	(1.10, 1.24)	1.18	(1.11, 1.25)
No	1.00	–	1.00	–

[a]Chemotherapy, radiotherapy and hormone therapy.

of a population-based organized service screening programme (lacking the control group) that would be expected to be achieved if a population-based randomized–controlled trial with the control group is applied.

To demonstrate its application to the empirical data, we first applied the proposed model to the Swedish randomized–controlled trial by only using the survival information of breast cancer from the study arm to compare the ITT

Table 5 Number needed to screen for Swedish randomized–controlled trial data and Taiwanese service data

	Swedish RCT data	Taiwan service data
Cumulative mortality rate for invited group, per 100 000	455	171
Estimated cumulative mortality rate for uninvited group, per 100 000	679	295[a]
Compliance rate	85%	26%
Relative risk for mortality reduction	0.67	0.58
Number needed to screen	446	806

[a]Annual clinical breast examination from 1999 to 2001.

estimate based on our proposed method and analysis with that which we have already obtained from the randomized–controlled trial. We then applied it to a Taiwan organized breast cancer service screening to demonstrate how the proposed design and analysis can feasibly be applied to population-based organized service screening data for estimating the unbiased effectiveness of population-based screening programmes.

The advantages of using the proposed switched design analysed using the time-dependent Cox survival analysis are 2-fold. First, information required for the proposed model can dispense with the collection of a large amount of individual screening history data, which is very time-consuming and there is difficulty in the logistics of follow-up. Making use of such a model and design would be very cost-effective in terms of the manpower and costs involved in a mass screening registry. Second, the proposed design and analysis enable one to evaluate the power of surrogate endpoints for the primary endpoint of mortality, as illustrated in the current study. The results can also shed light on why the screening works and what is the relative contribution between screening, treatment and therapy as we did in our multivariable analysis with adjustment for tumour attributes, and treatment and therapy.

There are two major limitations of this study. The first is that the proposed time-dependent switched design and Cox model are restricted to predicting absolute survival, as the baseline hazard rate cannot be explicitly estimated. Unlike the Nelson-Aalen estimate that is often used for predicting the baseline hazard, our baseline hazard is too complicated to be imputed. The second is that although we have already applied the proposed method to two data sets, representing high and low incidence of breast cancer regions, whether this method can be generalized to different scenarios should be warranted by more external applications to different types of population-based screening programmes.

In conclusion, the proposed time-dependent switched design and analysis is efficient and useful for providing evidence-based effectiveness of population-based screening dispensing with a large amount of individual screening history data.

Ethics approval

This study was approved by the Ethical Committee of Taipei Medical University (TMU-JIRB No.: N201607008) and National Taiwan University Hospital (NTUH-REC No.: 202103098 W).

Data availability

The data underlying this article will be shared on reasonable request to the corresponding author.

Author contributions

A.M.-F.Y. had full access to all the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis. R.W.-J.C. and A.M.-F.Y. conducted and are responsible for the data analysis. Study concept and design: T.H.-H.C., A.M.-F.Y. Acquisition of data: T.H.-H.C., A.M.-F.Y. Drafting of the manuscript: R.W.-J.C., G.H.-H.J. Critical revision of the manuscript for important intellectual content: K.-C.L., T.-C.C., S.-Y.C., S.-L.P., T.H.-H.C., A.M.-F.Y. Statistical analysis: R.W.-J.C., A.M.-F.Y., G.H.-H.J. Obtained funding: T.H.-H.C., A.M.-F.Y. Administrative, technical or material support: T.H.-H.C., A.M.-F.Y. Study supervision: K.-C.L., T.-C.C., S.-Y.C., S.-L.P. Final approval of the version to be published: all authors.

Funding

This study was supported by the Ministry of Science and Technology, Taiwan (MOST 108–2118-M-002 –002 -MY3; MOST 110-2811-M-038-003; MOST 108–2118-M-038 –001 -MY3; 110-2314-B-002-081-MY3; MOST 110–2811-B-002–586). Funders/sponsors had no role in the design and conduct of the study; collection, management, analysis and interpretation of the data; preparation, review or approval of the manuscript; and decision to submit the manuscript for publication.

Acknowledgements

We thank Professor László Tabár for professional comments on the evaluation of the breast cancer screening programme.

Conflict of interest

None declared.

References

1. Chen TH, Yen AM, Fann JC *et al*. Clarifying the debate on population-based screening for breast cancer with mammography:

a systematic review of randomized controlled trials on mammography with Bayesian meta-analysis and causal model. *Medicine (Baltimore)* 2017;**96**:e5684.

2. Tabár L, Fagerberg CJ, Gad A *et al.* 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* 1985;**325**:829–32.
3. Tabár L, Vitak B, Chen TH *et al.* Swedish two-county trial: impact of mammographic screening on breast cancer mortality during 3 decades. *Radiology* 2011;**260**:658–63.
4. Yen AM, Tsau HS, Fann JC *et al.* Population-based breast cancer screening with risk-based and universal mammography screening compared with clinical breast examination: a propensity score analysis of 1429890 Taiwanese women. *JAMA Oncol* 2016;**2**:915–21.
5. Walter SD, Stitt LW. Evaluating the survival of cancer cases detected by screening. *Stat Med* 1987;**6**:885–900.
6. Duffy SW, Tabar L, Fagerberg G *et al.* Breast screening, prognostic factors and survival: results from the Swedish two county study. *Br J Cancer* 1991;**64**:1133–38.
7. Paci E, Ponti A, Zappa M *et al.* Early diagnosis, not differential treatment, explains better survival in service screening. *Eur J Cancer* 2005;**41**:2728–34.
8. Duffy SW, Nagtegaal ID, Wallis M *et al.* Correcting for lead time and length bias in estimating the effect of screen detection on cancer survival. *Am J Epidemiol* 2008;**168**:98–104.
9. Joensuu H, Lehtimäki T, Holli K *et al.* Risk for distant recurrence of breast cancer detected by mammography screening or other methods. *JAMA* 2004;**292**:1064–73.
10. Wishart GC, Greenberg DC, Britton PD *et al.* Screen-detected vs symptomatic breast cancer: is improved survival due to stage migration alone? *Br J Cancer* 2008;**98**:1741–44.
11. Shen Y, Yang Y, Inoue LY, Munsell MF, Miller AB, Berry DA. Role of detection method in predicting breast cancer survival: analysis of randomized screening trials. *J Natl Cancer Inst* 2005; **97**:1195–203.
12. Warwick J, Tabàr L, Vitak B, Duffy SW. Time-dependent effects on survival in breast carcinoma: results of 20 years of follow-up from the Swedish Two-County Study. *Cancer* 2004;**100**: 1331–36.
13. Jansson S, Bendahl PO, Larsson AM, Aaltonen KE, Rydén L. Prognostic impact of circulating tumor cell apoptosis and clusters in serial blood samples from patients with metastatic breast cancer in a prospective observational cohort. *BMC Cancer* 2016; **16**:433.
14. Earl HM, Hiller L, Vallier AL *et al.* 6 versus 12 months of adjuvant trastuzumab for HER2-positive early breast cancer (PERSEPHONE): 4-year disease-free survival results of a randomised phase 3 non-inferiority trial. *Lancet* 2019;**393**:2599–612.
15. Fann JC, Chang KJ, Hsu CY *et al.* Impact of overdiagnosis on long-term breast cancer survival. *Cancers (Basel)* 2019;**11**:325.