The effect on survival of early detection of breast cancer in South Australia

by

G.M.Tallis[1], P.Leppard[1] & T.J. O'Neill[2*]

1. Honorary Visiting Research Fellow,
School of Mathematical Sciences,
The University of Adelaide

2. Head of School & Professor of Applied Statistics
School of Finance & Applied Statistics
Faculty of Economics & Commerce
Australian National University
*** Corresponding author**

Summary

Early detection of breast cancer is an important public health policy. Programs of regular screening examinations have been widely established in an attempt to detect the disease when the primary tumour diameter is small. In South Australia, BreastScreen SA suggests that women between the ages of 50 and 70 years be screened every 24 months. Our aim in this paper is to make assessments of various screening procedures by using statistical models with parameters estimated exclusively from South Australian data. We establish a relationship between primary tumour diameter and ultimate survival time. We estimate an advantage of 2.9 (.7) years in median survival time for those women detected with the disease by BreastScreen SA, compared with an unscreened population. We construct a computer model from which we determine the consequences of using a 12 month screening interval, and also the effect of beginning screening at the age of 40 rather than the current conventional commencement age of 50 years.

Key words: breast cancer; South Australia; early detection; breast cancer screening; median survival time; computer model.

1. Introduction

Breast cancer remains a severe health risk. In South Australia a woman has an estimated lifetime chance of 1 in 13 of developing the disease, and this chance is doubled or tripled if there have been cases of breast cancer in close relatives.

The prognosis at first detection involves many factors, including the size of the primary tumour, the disease stage, the tumour histology, the age and general health of the woman. However, the main methods of detection involve the primary breast tumour, the size of which is now regarded as a prime single predictor of the future course of the disease. There is a large body of literature concerning this issue spread over an extended period of time; for example, Fisher, Slack & Bross (1969); Duncan & Kerr (1976); Zurrida *et al* (1999). Detection programs attempt to find primary tumours early, before they have spread to local lymph glands, and beyond. Such programs carried out on a regular basis are referred to as "screening", with an individual set of examinations being called a "screen".

The principles of early detection have changed very little over the years. The present mechanics of screening women is, apart from technological advances in radiography, essentially the same as that used in the first major randomised clinical trial testing the effectiveness of regularly screening female populations for breast cancer. This trial was established in New York City in 1963, and it was known as the HIP (Hospital Insurance Plan) Study. Conclusions based on this trial, and others, have been summarised in a review paper by the US Department of Health and Human Services entitled "Screening for Breast Cancer" (USDHHS, (1996)). It was in the analyses of the results of such studies that technically challenging problems were discovered.

It is intuitive that screen detected disease is generally at an earlier stage than "naturally reported" disease. The difference between the average age of breast cancer patients whose disease is self detected and the average age at which they would have had the tumour detected under a screening program is called the lead-time. Lead-time for breast

cancer is of the order of six to eighteen months and it must be taken into account in survival analyses following any trial to avoid systematic biases.

With regard to screening recommendations, these too have not altered greatly over recent years. It is strongly advised to screen women between the ages of 50 years and 70 years every 24 months using mammography, clinical breast examination and breast self-examination as a joint detection modality. It is unclear whether women outside this age range benefit significantly from screening every 24 months. Nevertheless, women of any age who are thought to be at a higher risk of getting breast cancer due to its presence in close relatives are advised to enrol in an early detection program.

It is reported in USDHHS *(loc cit)* that screening appears to lead to a reduction in breast cancer mortality of at least 22%. However, this criterion of assessing the impact of a screening program has recently been criticised; Black, Haggstrom & Welch (2002); Tabar *et al* (2002); Juffs & Tannock (2002). We believe that a more useful and direct interpretation of the benefits of breast cancer screening for a woman who is considering entering a program is in terms of the extra lifetime that she may gain. Hence, all our results in this paper are expressed in terms of additional survival time past the age at reporting. We note that survival times have also been used by Jansen & Zoetelief (1995).

In 1989 BreastScreen SA was established to join a National Programme of Early Detection of Breast Cancer. The South Australian affiliate services a total population of about 142,000 women aged between 50 years and 70 years. Annually between 60,000 and 70,000 screens are carried out, providing a screening service with a screening interval of 24 months (Robinson *et al* (1996)).

The main aim of this paper is to demonstrate the possibility of evaluating the impact of a screening service using only data from the service itself, and additional data from the local population-based cancer registry. Our population consists of the women in South Australia and the criterion of evaluation is in terms of additional life-time past the age at which breast cancer was first diagnosed. No data external to South Australia has been

used in our evaluation, thus eliminating the concerns of bias due to population mismatch. We emphasise that the models developed in this paper are specifically designed for the data that has been routinely collected in South Australia. It is not anticipated that these models will necessarily be useful elsewhere. The results reported here are the culmination of efforts spanning some 30 years and we believe this to be an original way to assess the effect on survival of early detection of breast cancer.

In brief, Section 2 is a glossary of acronyms, notation and definitions. In Section 3, we establish the method of estimating median additional survival time following the detection of a primary tumour which is the measure we use to assess the advantages of early detection. Section 4 demonstrates the survival advantages of early detection and quantifies the achievement of BreastScreen SA. Section 5 consists of a technical description of breast cancer screening, develops procedures for some necessary modelling and estimation processes, and extrapolates the results of Section 4. Section 6 summarises the findings of this paper and discusses potential biases in the estimation of survival time.

2. Notation, definitions and data sets

This section is designed to help readers by defining terms and important acronyms here for quick reference, and describing the various data sets used in this paper**.**

1. *a*; the age in years at which a primary breast tumour is first detected.
2. *x*; the diameter in centimetres of the tumour reported at *a.*
3. *m(a,x)*; the median additional survival time in years of a woman at *a* with *x*. A fitted Cox proportional hazards model, $\hat{m}(a,x)$, specified in Section 3 is used to calculate values for *m(a,x)*. The survival time of a woman past *a* is a random quantity with a mean that cannot be accurately estimated from our current data sets.
4. $G = \hat{m}(a_1, x_1) - (a_2 - a_1) - \hat{m}(a_2, x_2)$; the calculated gain in median survival for tumours detected at a_1 with x_1 and at a_2 with x_2, after correcting for lead-time $a_2 - a_1$

5. $x(t;\beta)$; the diameter of a primary tumour of age *t* defined by the standard biological model $x(t;\beta) = x_0 e^{\beta t}$; see Kusama *et al* (1972). Here, x_0 is the diameter of a single cell, approximately .002cm, and β is the tumour growth rate.
6. SACR; South Australia Cancer Registry
7. BSSA ; BreastScreen SA
8. SACR diameter data set; consisting of 1346 cases of breast cancer accumulated by the SACR in the period 1980-86. For each case in this data set, the age of the woman reporting the disease, the diameter of the reported tumour and the survival time from reporting until death or surviving until December 2002, is available. The survival times of these cases have minimum and maximum periods of observation of 16 and 22 years. For this data set, $\bar{a} = 58.8$ and $\bar{x} = 3.2$.
9. SACR conditioned data set; consisting of a sub-set of 627 cases of the SACR diameter data set with $55 < a < 75$ and $x \leq 5.5$. The restrictions on the ranges of *a* and *x* make it possible to accurately determine median survival time for the specific values of *a* and *x* that are appropriate for screening designs presented later in this paper. For this data set, $\bar{a} = 64.5$ and $\bar{x} = 2.60$.
10. BSSA data set; consisting of the 2769 cases of breast cancer detected in women who entered the BreastScreen SA screening program between 50 and 70 years of age during the period 1990-2002. For this data set, $\bar{a} = 60.9$ and $\bar{x} = 1.65$.
11. SEER data set; consisting of cases extracted from the publicly accessible data base of the SEER program (2003) which consists of over 400,000 cases of breast cancer. A sub-group of 3057 SEER cases from 1989-90 was matched as closely as possible to the same *a* and *x* specifications of the SACR conditioned data, giving an approximate 5:1 ratio in the sample sizes of the two data sets. Prior to 1989 reporting diameters were not recorded for breast cancer cases in the SEER data base. The choice of the SEER data from 1989-90 was an attempt to make the survival characteristics of the two matched groups as comparable as possible, although there are possibly unknown effects of case-

mix and temporal shifts between the two sets. As the maximum possible survival time for the SEER data is approximately 12 years, the estimation of median survival time conditioned on *a* and *x* is currently restricted for the SEER data. For this data set, $\overline{a} = 64.6$ and $\overline{x} = 2.55$.

12. SACR incidence data set; consisting of the year of reporting and *a* for 8582 cases of breast cancer registered by the SACR for the period 1977-92. (Epidemiology of Cancer in South Australia (1993)). The time period 1977-92 essentially precedes the introduction of formal breast cancer screening in South Australia after which the concept of random self-reporting no longer can be assumed. The frequency distribution of *a*, accumulated over 1977-92, is shown in Table 3 in Section 5.2. For this data set, $\overline{a} = 60.5$.

3. Estimation of median survival time

In this section we consider the estimation of median survival time in some detail, since this is the measure by which we assess the performance of screening designs of the type currently practiced in South Australia. It is also the measure that we use to quantify the advantages of alternate screening designs.

Breast cancer survival data is available for two distinct but generally comparable populations. The primary data used for survival analysis is the SACR conditioned data (see 2.9). A Cox proportional hazards model was fitted to these survival data, and linear, quadratic and interaction terms for *a* and *x* were investigated. Full diagnostic checks for proportionality led to a model with linear and interaction terms for *a* and *x*. Influence plots confirmed that there were no data points with undue influence. Scaled Schoenfeld residual plots for the interaction model show no departures from proportionality.

The adequacy of calculating median survival time $\hat{m}(a,x)$ from this specific parameterization of the Cox model fitted to the SACR conditioned data is illustrated in Table 1. The investigations detailed in the previous paragraph do not specifically address the estimation of median survival. For various diameter groupings, median survival time is calculated directly by the non-parametric actuarial lifetable method and each value is

closely reproduced by $\hat{m}(\bar{a},\bar{x})$ determined from the Cox model. The technical reasons for the particular choice of diameter groupings in Table 1 have been previously discussed in Tallis *et al* (2003).

Table 1: Comparison of actuarial and Cox model median survival times

x	N	$\bar{a}$	$\bar{x}$	Actuarial median (se)	$\hat{m}(\bar{a},\bar{x})$ (se)	z-score
≤ 1.2	67	64.2	0.94	18.71 (1.08)	18.97 (1.82)	-.12
1.3-1.7	107	64.9	1.50	16.84 (1.45)	16.50 (1.11)	.19
1.8-2.2	118	64.7	2.00	13.60 (1.13)	14.74 (0.96)	-.77
2.3-2.7	102	64.2	2.50	11.67 (0.95)	13.24 (0.81)	-1.26
2.8-5.5	233	64.6	3.93	9.04 (0.57)	9.41 (0.60)	-.45

The agreement between each pair of median survival times, actuarial and $\hat{m}(\bar{a},\bar{x})$, is assessed by a z-score. It can be seen in Table 1 that only one z-score is greater than 1 in absolute value. By squaring and summing these z-scores, we get an approximate and perhaps conservative χ^2_2=2.42, with P = .30. Thus we believe that $\hat{m}(a,x)$ calculated from the fitted Cox model adequately represents median survival time in the SACR conditioned data over these ranges of *a* and *x*.

The survival distributions of the SACR conditioned data and the SEER data *(see* 2.11) were examined by using actuarial survival curves, and the results of these comparisons are shown in Table 2, conditioned on the diameter groupings used in Table 1. In order to

make the survival times from the two data sets comparable, the survival times in the SACR conditioned data were re-coded so that deaths recorded after 12 years became censorings at 12 years. The result of accumulating the probabilities from the five diameter groups by using the Fisher method is $\chi^2_{10} = 8.07$ with P=.62; Cox & Hinkley, (1974). Two summary statistics, the 75th percentile and the survival after 11 years S(11), are also shown in Table 2 to illustrate the magnitude of the differences in survival between the two groups. The conclusion is that overall survival, and survival conditioned on diameter, is not materially nor demonstrably different in the two data sets matched on *a* and *x*, but not on other factors which may influence survival. It was reassuring that the Australian and American breast cancer survival experiences showed a close agreement, which was anticipated a priori, since a substantial divergence would have lead to an uncomfortable inferential impasse.

We have compared the survival distributions determined from the SACR conditioned data against the SEER data because of the critical role that estimated median survival time plays in our assessments of screening designs. Nevertheless, it must be noted that, since the survival distributions can only be compared over 12 years, the power of discrimination may be weak and for reasons previously stated, the two groups cannot be regarded as absolutely exchangeable. (see 2.11)

Table 2: Comparison of South Australian & SEER Actuarial Survival Curves

x	Mantel-Cox P-value	75th Percentile			S(11)		
		SA	SEER	z-score	SA	SEER	z-score
≤5.5	.267	5.58 (0.38)	5.39 (0.19)	.45	.533 (.055)	.569 (.071)	-.40
≤1.2	.904	10.55 (0.75)	10.82 (0.60)	-.28	.716 (.023)	.743 (.077)	-.34
1.3-1.7	.139	7.55 (0.91)	9.25 (0.61)	-1.55	.626 (.046)	.696 (.031)	-1.26
1.8-2.2	.248	5.07 (0.77)	6.35 (0.57)	-1.34	.551 (.031)	.612 (.051)	-1.02
2.3-2.7	.730	5.38 (1.03)	5.20 (0.41)	.16	.539 (.115)	.551 (.076)	-.09
2.8-5.5	.779	4.02 (0.49)	3.47 (0.20)	1.04	.425 (.073)	.439 (.109)	-.11

4. Survival advantage of early detection of breast cancer

4.1 An illustration

In order to illustrate the effect of early detection on survival, consider the example of a woman with a tumour which remains undetected until it reaches x=3 by a=65, with $\hat{m}(65,3)=10.9$. This combination of a, x and $\hat{m}(a,x)$ is used as a benchmark in this section, with a and x being a reasonable approximation to the historical average reporting age and diameter of the SACR diameter data (*see* 2.8).

Suppose that the tumour was discovered at a less than 65 when x=1. For the present purposes we take β=.07, determined later in Section 5. Applying the model for tumour growth $x(t;\beta)$ (see 2.5) it is easily found that x=1 when a = 63.7, and consequently

$\hat{m}(63.7,1) = 19.2$. With $a_1 = 63.7$, $x_1 = 1$ and $a_2 = 65$, $x_2 = 3$, then $G = 7.1$ (see 2. 4). Similarly, if the tumour was detected at $x = 2$, or $x = 4$ or $x = 5$, the same argument gives values of G as 3.5 – 2.0 and -3.8, all with respect to the benchmark $\hat{m}(65,3)$. These figures clearly suggest the possible advantages of early detection of the disease.

4.2 Results using Breastscreen SA data

Now we apply the concepts introduced in Section 4.1 to assess the impact of the breast screening program of BreastScreen SA during 1990-2002. The first task is to establish the appropriate benchmark for this period.

From the SACR diameter data, naturally reported cases of breast cancer in South Australia have the historical average tumour diameter of 3.20cm. However, work reported in Tallis, Leppard & O'Neill (2003) suggests that this naturally reported diameter may have been reduced over recent years to 2.44cm. The SACR has independently published in an annual report results in a tabular form for 1997-99, from which we estimate $\overline{a} = 60.5$ and $\overline{x} = 2.41$ for an unscreened group of women. These last two estimates are in substantial agreement, and we use our estimate of x=2.44 to define the benchmark in subsequent analyses. The precise reasons for the reduction over time in average reporting diameter in an unscreened group are not known. This same phenomenon has been observed in the United States where $\overline{x}$ varied from 3 to 2.1 over the period 1970 to 2000; Fisher *et al* (*loc cit*), Cady (1997).

A woman having the average detection age and tumour diameter of the BSSA data (*see* 2.10) has $\hat{m}$ (60.9,1.65)=19.79 (1.11), where the estimated standard error in parentheses has been calculated by using bootstrap techniques with 1000 repeated samplings; Efron & Tibshirani (1993). In the absence of screening, a tumour with x=1.65 would grow to x=2.44 in .5 years, using the arguments of the previous section, and hence defines a benchmark of x=2.44 at a=61.4 with $\hat{m}$ (61.4,2.44)=16.36 (0.83). Thus for this situation G=2.93 (0.68).

Since, as discussed above, it appears that the average tumour diameter of an unscreened populations is decreasing, we have also used for comparison the benchmark tumour diameter of 2.1 cm reported by Cady *(loc cit)*, giving a survival gain of G=1.70 (0.47). These assessments of the advantages of screening in terms of increased gain in survival time emphasise the key role played in such assessments by the choice of a, x and $\hat{m}$ that define the benchmark.

5. Statistical modelling of breast cancer screening for South Australia

In Section 4.2 we have used G as an assessment of screening as performed by BreastScreen SA. We conclude that their screening program with a 24 month period between examinations leads to a gain in survival of approximately 3 years for the average woman detected under the screening program. This is an encouraging and important result. In the following sections we develop a structural framework and statistical estimates required to investigate the effect on G of reducing the screening interval from 24 to 12 months, and of beginning screening at 40 rather than 50 years of age.

5.1 Definitions of components for the screening process

It is straightforward to heuristically review the broadest aspects of screening. For example, suppose a woman enters a screening program at the age of 50 years. She has been continuously subjected to the risk of initiating a breast tumour during the previous 30 years or so. There are three possible outcomes for a woman entering the screening program with an undetected tumour. She could have the tumour detected by screening, she could self-detect the tumour between screens, or she could die before the tumour is detected in either way. These events are assumed to be dictated largely by the rate of initiating tumours, tumour size and rate of growth, the sensitivity of the screening procedure for detecting tumours, and the likelihood of the woman finding and reporting the disease herself in the presence of the competing risk of natural mortality. If a tumour of a particular size is found, then the patient's survival prospects will depend on, at the very least, the age of the woman and the size of the tumour.

In this way we identify six basic components which must be known in order to develop a basic understanding and quantification of screening. These components are

(1) The probability of surviving to *a* for South Australian women, *S(a);*
(2) The probability of self-detecting a primary tumour less than *x, P(x);*
(3) The probability of detecting a primary tumour of diameter *x* when subjected to screening, *r(x);*
(4) The distribution of growth rates, β , of primary tumours, $b(\beta)$;
(5) The probability of a woman ever contracting breast cancer, γ ; in the absence of natural mortality, and given that the woman contracts the disease, the probability of initiating the disease by age *y, H(y).*
(6) The median survival time of a women at *a* with *x, m(a,x).*

The function *S(a)* is obtained from population life tables published by the Australian Bureau of Statistics. The function *m(a,x)* has been estimated in Section 3. The functions *P(x)* and *r(x)* were estimated in Tallis *et al (loc cit),* and here we use the notation $\hat{P}$ and $\hat{r}$ respectively for these functions with parameters set at the maximum likelihood values given in that paper. The functions $b(\beta)$ and *H(y)*, and γ , are estimated in the following section.

5.2 A model for breast cancer incidence

The SACR incidence data (*see* 2.12) is shown in Table 3 with the frequency distribution denoted as I_a . To derive a tailored and credible model to simulate the distribution of I_a is a detailed and subtle process. However, here we restrict the derivation to a discussion of the mean of I_a and a brief outline of distributional properties.

Given a woman does contract the disease, let A be the age at which the woman first detects a primary breast tumour, the age at reporting for short. Clearly, A is a random variable, and we let Pr (A=*a) = f(a*), $\sum_{20}^{100} f(a) = 1$. Under the same conditioning, the probability that a woman reports at A=*a,* in the presence of natural mortality, is

f(a) S(a), from 5.1(1). Thus, the unconditional probability that a woman reports at A=*a* is γ *f(a)S(a),* from 5.1(5).

Now define N_a as the total number of women born *a* years before either 1977, or 1978, …, or 1992. Then $\lambda_a = E\, I_a = N_a \gamma\, f(a)S(a)$ is the expected total number of women who report at A=*a* during the 16 years 1977-1992. We now briefly show that for all *a*, the I_a are Poisson distributed with parameter λ_a, and that the I_a are all mutually independent.

Referring to Table 3, consider I_{40}. The 16 years of data collection is 1977-92, and I_{40} is built up from reported cases from each of these years, since results have been pooled. Now, only women born in 1937 can contribute to the 40 year old incidence figure for 1977. Women born in 1938 contribute to the 40 year incidence for 1978, and so on, to women born in 1952 who contribute to the incidence figure for 1992. The 16 contributions are obviously independent, and since γ *f(a)S(a)* is small, the random parts adding to I_{40} are Poisson distributed, and consequently so is I_{40}.

Table 3: Age-specific breast cancer incidence 1977 - 1992

Age(a)	I_a	λ_a	$d_a = \frac{I_a - \lambda_a}{\sqrt{\lambda_a}}$	Age(a)	I_a	λ_a	$d_a = \frac{I_a - \lambda_a}{\sqrt{\lambda_a}}$
≤ 30	92	91.1	0.091	61	215	199.6	1.093
31	26	36.0	-1.672	62	187	200.1	-0.925
32	45	43.2	0.275	63	230	203.7	1.845
33	49	50.5	-0.209	64	206	204.7	0.093
34	64	60.8	0.416	65	194	211.5	-1.202
35	74	67.2	0.836	66	221	204.9	1.125
36	85	76.2	1.004	67	213	197.0	1.141
37	71	85.5	-1.567	68	214	199.6	1.021
38	97	95.9	0.111	69	201	195.4	0.398
39	106	110.1	-0.386	70	191	198.1	-0.504
40	120	115.5	0.420	71	181	188.4	-0.542
41	133	120.8	1.114	72	185	182.0	0.223
42	130	135.2	-0.446	73	164	175.5	-0.869
43	140	139.0	0.084	74	159	167.0	-0.622
44	131	149.2	-1.487	75	139	163.9	-1.946
45	161	152.4	0.697	76	158	156.8	0.100
46	154	154.9	-0.072	77	139	143.3	-0.357
47	161	160.3	0.058	78	132	134.5	-0.218
48	178	159.8	1.442	79	109	126.4	-1.551
49	162	167.0	-0.386	80	114	119.6	-0.512
50	194	177.6	1.234	81	94	107.2	-1.274
51	156	167.4	-0.880	82	109	96.1	1.312
52	173	179.6	-0.492	83	106	86.2	2.130
53	187	182.5	0.336	84	79	78.9	0.010
54	161	184.1	-1.703	85	75	70.2	0.569
55	188	193.6	-0.402	86	66	63.0	0.383
56	170	191.3	-1.543	87	58	50.2	1.099
57	196	191.5	0.328	88	52	42.1	1.522
58	206	198.6	0.524	89	27	37.0	-1.650
59	203	197.8	0.371	≥ 90	128	121.1	0.624
60	211	211.7	-0.045				
							$D = \sum_{a=30}^{90} d_a^2 = 56.6$
							$\Pr(D = \chi^2_{55} \geq 56.6) = .42$

Of course, the women born in 1937 contribute to I_{41} in 1978, I_{42} in 1979, and so on. These contributions are generated from a multinomial with 61 probabilities, $\gamma f(a)S(a)$, and, since the number of women in the 1937 cohort is large, the random variables generated by the multinomial, in the limit, are distributed as independent Poisson variables. The same is true for all cohorts who contribute to the 1977-92 incidence years. Finally, $I_a \sim Poisson(\lambda_a)$ and $I_a, I_b, a \neq b$, are independent for all *a* and *b*. A detailed mathematical demonstration of these results can be achieved using probability generating function techniques outlined in Chapter 12, Volume 1 of Feller (1960).

We have chosen the standard biological model for tumour diameters, $x(t;\beta)$ (see 2.5). Now we choose a density function $b(\beta)$ for the growth rate β, of the form $b(\beta_1) = \theta$ and $b(\beta_2) = 1-\theta$. This form allows just two values of β mixed by θ and $1-\theta$, reflecting the current medical thinking of "slow" and "fast" growing tumours. The most recent discussion of tumour growth rates in a human population are reported in Kusama *et al (loc cit)* and Vorherr (1981).

Given a woman does contract the disease at some time during her life, there are two additional variables to model: Y, the age at initiation of the disease; T, the time delay to reporting the breast tumour. The discrete probability density function (pdf) of T is $p(t;\beta) = P(x(t;\beta)) - P(x(t-1;\beta))$, for integer values of *t*, since $x(t;\beta)$ is a strictly monotone increasing function of the delay to reporting. The function *P* is the cumulative distribution function (cdf) of self-reported diameters, *x*; see Tallis *et al loc cit*. Since the model $b(\beta)$ has two values of β, β_1 and β_2, this gives $p(t) = \theta\, p(t;\beta_1) + (1-\theta)\, p(t;\beta_2)$. Similarly, we let the discrete pdf of Y be $h(y) = H(y) - H(y-1)$, where the function *H* is the cdf of the age at initiation of a breast cancer tumour. Since A is the age at reporting, A = Y + T, where Y and T are assumed independent. Since Pr (A = a) = *f(a),* then $f = h * p$, *i.e.* $f(a) = \sum_{y=20}^{a} h(y)\, p(a-y)$, with *h(y)*=0 for $y < 20$ by assumption.

To complete the definition of λ_a, we need to specify $f(a)$ in greater detail. We use

$$P(t) = 1 - \exp(-\alpha_1 x(t;\beta)^{\alpha_2}) \text{ and } H(y) = \frac{H^*(y) - H^*(20)}{H^*(100) - H^*(20)}$$

where $H^*(y) = \log(1 + \exp(\delta_1 + \delta_2 y))$. The choice of the form of *P* was suggested by years of past experience, and an empirical analysis of the SACR diameter data. The form of $H^*(y)$ was suggested by a graphical display of the distribution of I_a in Table 3, taking into account the form of *P(t)*. This analysis suggested that the pdf of *H(y)* had a sigmoid shape and this particular property of the data is necessarily reflected in our parameterised version of *H (y).* Truncation at ages 20 and 100 keeps estimates within realistic bounds.

The set of parameters $\pi = (\gamma, \delta_1, \delta_2, \alpha_1, \alpha_2, \theta, \beta_1, \beta_2)$ are estimated by standard maximum likelihood techniques. The log-likelihood function, $\log L$, is expressed as two components

$$\log L = \log L_1 + \log L_2$$
$$= \sum_{a=30}^{90} (I_a \log \lambda_a(\pi) - \lambda_a(\pi)) + \sum_{j=1}^{1345} \left(\log \alpha_1 + \log \alpha_2 + (\alpha_2 - 1) \log x_j - \alpha_1 x_j^{a_2} \right)$$

In the expression for $\log L_1$, $\lambda_a(\pi)$ requires values for $N_a S(a)$, the size of the population at age *a* accumulated over the period 1977 to 1992. This data is obtained from population figures published by the Australian Bureau of Statistics. The expression $\log L_2$ is obtained by differentiating *P* with respect to *x*, and evaluated using the SACR diameter data. The likelihood parameter estimates and standard errors are given in Table 4. The estimates $\hat{\beta}_1$ and $\hat{\beta}_2$ conform to the results reported by Kusama *et al* (*loc cit*).

It should be noted that the breast cancer cases defining $\log L_2$ are a subset of the cases defining $\log L_1$. Since *a* and *x* are essentially independently distributed (see Tallis *et al* (*loc cit*)) the likelihood procedure specified above is appropriate. The issue of independence is further demonstrated by a correlation of .031 (.027) between *a* and *x* in the SACR diameter data, where there is also no evidence of higher order regression of *x* on *a*.

Let $\hat{\pi}$ be the maximum likelihood estimator for π. The theory of section 4(b) of Tallis & Chesson (1976) was modified to deal with parameter estimates which are generated externally to the main data set, whose agreement with the model is to be tested. From this work, it can be shown that conservatively,

$$D = \sum_{a=30}^{90} d_a^2 = \sum_{a=30}^{90} (I_a - \lambda_a(\hat{\pi}))^2 / \lambda_a(\hat{\pi}) \sim \chi^2_{55} \text{ and } \Pr(D = \chi^2_{55} \geq 56.6) = .42.$$

Analysis of the standardised residuals d_a produced a mean and standard deviation of .01 and .97. There is no regression of d_a on a; all terms of a polynomial regression of order 3 are non-significant with p-values greater than .3 and $R^2 = 2.2\%$. There is no suggestion of non-normality in the distribution of d_a as assessed by both the Anderson-Darling and Shapiro-Wilk tests of normality. Hence, overall, there is no evidence that our model does not fit the SACR incidence data.

Table 4: Parameter estimates (se) of cancer incidence model

Parameter	Estimate	se
γ	0.187	0.008
δ_1	-6.890	0.555
δ_2	0.202	0.017
α_1	0.121	0.007
α_2	1.636	0.032
θ	0.354	0.023
β_1	0.018	0.002
β_2	0.101	0.014

5.3 A computer model for screening in South Australia

The six breast cancer screening components detailed in Section 5.1, which have been estimated using South Australian data, have been incorporated into a computer model of the screening process specifically for South Australia. We examine the consequences of both reducing the screening interval from 24 months to 12 months, and beginning screening at 40 rather than 50. There is an extensive literature of mathematical modelling and computer simulation of breast cancer screening processes, some of the earliest being

Knox (1973), Schwartz (1978) and the evolving general cancer screening simulation project MISCAN (1984). More recently Jansen & Zoetelief *(loc cit)* have produced a simulation program which calculates the benefit of breast cancer screening based on the reduction in tumour size at detection and survival as a function of tumour diameter. The technical features of breast cancer screening have been extensively reviewed in O'Neill, Tallis & Leppard (1995) where further references are given.

The logical structure of the model is briefly outlined, using the notation of the six components discussed in Section 5.1. A cohort of (computer generated) women is progressively aged from birth with the cohort continuously subjected to natural mortality specified by *S(a).* Tumours are initiated at each age *y* according to $\hat{\gamma}$ and $\hat{H}(y)$, with a growth rate assigned to these tumours according to $\hat{b}(\beta)$. Thereafter these tumours increase in diameter over time following initiation *t,* according to $x_0 e^{\beta t}$, for β either $\hat{\beta}_1$ or $\hat{\beta}_2$. Women who initiate a tumour may self-report at $a=y+t$ as specified by $\hat{P}(x(t))$, (s*ee* Section 5.1) subject to surviving to *a* specified by *S(a).* A screening procedure defined by $\hat{r}(x)$ (*see* Section 5.1) is introduced for those women who are alive at a designated age a_S and who have not previously self-reported a tumour before this age. An initial screening examination is made at a_S, with further screening examinations made at intervals of *l* months until the women reach age a_F. The general screening design is designated as D(a_S , l, a_F). Within the period a_S to a_F each woman will either

a) self-report a tumour at a known a and with a known x
b) have a tumour detected at a known a and with a known x by a screening examination
c) die through natural mortality before either (a) or (b) occurs
d) survive to a_F without either (a) or (b) occurring.

The group of women who have a tumour detected at any periodic examination or who self-report a tumour at any time between two successive examinations are considered detected under the screening design. The average detection age, $\overline{a}_1$, and the average reporting diameter, $\overline{x}_1$, are determined from this group to represent a screened population,

with median survival time $m(\bar{a}_1, \bar{x}_1)$. The appropriate comparison is made using the group of women with tumours detected using screening design D(a_S, *, a_F), representing only self-reported tumours over the period a_S to a_F and without the imposition of any screening examinations. The average reporting age, $\bar{a}_0$, and the average reporting diameter, $\bar{x}_0$, are determined from this group to represent an unscreened population, with median survival time $m(\bar{a}_0, \bar{x}_0)$. The gain of the screening design is assessed by $G = m(\bar{a}_1, \bar{x}_1) - (\bar{a}_0 - \bar{a}_1) - m(\bar{a}_0, \bar{x}_0)$.

An initial basic assessment of the performance of the model was made by specifying a design that accumulates self-reported cases over all ages without involving screening. The results for this design, labelled Model-unscreened, are shown in Table 5. Equivalent observed values from a SACR 1997-99 report are also shown in Table 5. We tested the model further. For the BSSA data, $\bar{a}_S = 57.7$ and $\bar{a}_F = 65.3$. These values arise because the data from BreastScreenSA are produced by a complex mixture of screening designs. Each woman examined has an age between 50 and 70 years at which she began her screening program, and a time under observation between one and twelve years, depending on the calendar year between 1990 and 2002 at which she began her particular screening program. These features of the data specify a value of a_S and a_F for each woman, from which $\bar{a}_S$ and $\bar{a}_F$ were calculated. The results for the model with these settings are shown in Table 5 labelled Model-24. Because of the close agreement between the results from the observed data and the model, we confidently use the model in Section 5.4 to investigate various screening designs.

Table 5: Verification of computer screening model

	Design	$\bar{a}$	$\bar{x}$	$\hat{m}(\bar{a},\bar{x})$	Difference in median survival (lead-time corrected)
SACR 1997-99		60.5	2.41	17.26	
Model-Unscreened	D(0,*,100)	60.8	2.49	16.87	.09 (.13)
BSSA data		60.9	1.65	19.79	
Model-24	D(57.7,24,65.3)	61.1	1.53	20.24	.35 (.24)

5.4 Applications of the computer screening model

We firstly consider what might have been achieved had BreastScreen SA used a 12 month screening interval rather than a 24 month interval. This is calculated using the design labelled Model-12. The results are compared to those for Model-24 with the values of *G* shown in Table 6. We estimate that there may have been a marginal advantage in using a 12 month screening interval.

Table 6: Estimated G for BreastScreen SA using 12 month screening interval

	Design	$\bar{a}$	$\bar{x}$	$\hat{m}(\bar{a},\bar{x})$	*G*
Model-24	D(57.7,24,65.3)	61.1	1.53	20.24	
Model-12	D(57.7,12,65.3)	61.0	1.19	21.51	1.25 (.58)

The results presented in Table 6 are based on the current average screening exposure time of approximately 7 years. We now use the model to estimate the advantages of long term screening. In Table 7, results are shown using the model with the self-reporting function $\hat{P}$. The model is specified for an unscreened population over the age of 50, and for screening designs beginning at age 50, with 24 and 12 month screening intervals. The model with a 24 month screening interval is compared to the model for an unscreened population, and the model with a 12 month screening interval is compared to the model

with a 24 month screening interval. For illustration, the parameters defining $\hat{P}$ have been modified to produce a variation in self-reporting, P^*, which has an average value of 2. This change approximates the self-reported diameter of Cady (*loc cit*), substantially lowers the benchmark and thereby reduces the possible effect of screening, as previously discussed in Section 4.2. Corresponding results are produced for P^* as for $\hat{P}$.

Table 7: Comparison of screening designs beginning at age 50

	Screening interval	Design	$\overline{a}$	$\overline{x}$	$\hat{m}(\overline{a},\overline{x})$	G
	Unscreened	D(50,*,100)	67.30	2.49	10.37 (0.81)	
$\hat{P}$	24 month	D(50,24,100)	66.79	1.52	14.49 (1.16)	3.6 (0.82)
	12 month	D(50,12,100)	66.52	1.09	16.30 (1.44)	1.5 (0.64)
	Unscreened	D(50,*,100)	67.26	2.00	12.12 (0.99)	
P^*	24 month	D(50,24,100)	66.82	1.36	14.90 (1.03)	2.4 (0.59)
	12 month	D(50,12,100)	66.58	1.04	16.29 (1.24)	1.1 (0.49)

There are a number of issues that arise from Table 7.

- The value of G=3.6 achieved for $\hat{P}$ and a 24 month screening interval agrees reasonably well with G=2.9 (.68) calculated for BreastScreen SA in Section 4.2. The value of $\overline{a}$ from BSSA data is about 6 years less than $\overline{a}$ of the model. This occurs because BreastScreen SA currently has a period of observation of twelve years whereas the design setting summarises a cohort screened every 24 months after fifty.
- The results using $\hat{P}$ suggest that there is an additional survival advantage of 1.5 years for a screening program with a 12 month screening interval when compared to a program with a 24 month interval. This agrees with the value of 1.25 shown for Model-12 in Table 6.

- The results using P^* parallel those found for $\hat{P}$, but to a lesser amount, again emphasising the role that the benchmark plays in these assessments.

We use our model to generate results equivalent to those shown in Table 7, but using 40 rather than 50 as the age of initial screening. The corresponding values of *G* are 3.41 (0.78) and 1.69 (0.68), which are essentially the same as those shown in Table 7.

A prominent question in the breast cancer screening literature concerns the efficacy of beginning screening at the age of 40 rather than 50, the currently accepted age at which screening programs start ; USDHHS (*loc cit*). In order to examine this question, it is necessary to compare the results from a group of women who begin screening regularly from the age of 40, with the results from a group reaching the same age of 40 but who defer screening until age 50. This comparison produces a value G=.62 (.23) in favour of the group who begin screening at 40, for a screening interval of 24 months. A value of *G*=1.05(.49) is found for the group starting at 40, for a 12 month screening interval.

6. Discussion and Summary

We believe that we have achieved the main aims of this project as outlined in the Introduction. A random sample of women who reported with breast cancer during the period 1980-86 defined the SACR diameter data. This set specified the reporting age *a*, the diameter of the primary tumour *x*, and the survival time and status of the patient at the end of 2002. With this information we were able to restrict *a* and *x* to form the SACR conditioned data which was used to develop a Cox proportional hazard model with covariates *a* and *x* from which we estimated the median additional survival time after reporting, $m(a,x)$. Using this quantity, we provided in Section 4.1 some simple illustrations of the effect on survival of early detection of the disease, and introduced the concept of benchmark from which early and late detection can be measured and assessed in terms of additional survival time. Screening data from BreastScreen SA was evaluated in Section 4.2. It was found that the reduction in primary tumour diameter at detection due to the screening program produces an estimated gain in median survival time of

about three years. This estimate is encouraging, but perhaps a more conservative way of interpreting it is to use the standard error to conclude that the gain should be at least 1.5 years. Further, to put these results into perspective, the benchmark median survival represents a loss of 6 years when compared to normal median survival. It appears that under screening, about 3 of these years are recovered.

The interval between screens in the BreastScreen SA program is 24 months and the target group of women are between 50 and 70 years of age. What may have been the outcome had BSSA used different specifications? In order to address this question, we developed a computer program in Section 5 which closely approximates the BSSA screening experience. The modelling to develop this computer program involved six basic components and each component required estimation. We restricted ourselves entirely to data held by the SACR and BSSA for this task. Part of the estimation problem is dealt with in this paper, and the rest is published in Tallis *et al (loc cit)*. The computer model, using well defined estimates of its parameters, was run first with the current BSSA screening interval of 24 months for women in the age range of 50 to 70. The computer results, in terms of the average age and primary tumour diameter, were very close to the observed BSSA figures. Since the computer program appeared to faithfully reproduce reality, we examined the outcomes for other settings.

Using the computer program, for screening over 50, we estimate a survival advantage of 1.5 years in favour of a screening interval of 12 months compared to an interval of 24 months. For a screening interval of 24 months, a small advantage of approximately half a year was found for screening starting at 40 compared to the standard age of 50. For a screening interval of 12 months, an advantage of one year was found for screening starting at 40 rather than at 50.

We have identified two potential sources of bias which could influence estimates of median survival time. The first is a registry problem. Inevitably some cases are lost to follow-up, thereby leading to over-estimates of median survival time, Tallis *et al* (1988). Fortunately, this is not a major issue in a professionally run registry such as the SACR

but, in any case, since we use the difference between two median estimates, any small biases would tend to negate each other.

The second source of bias is less obvious and more technical. An unscreened population such as that leading to the SACR conditioned data is in a steady state. In terms of *x*, the population is a mixture of two different growth rates, β_1 and β_2, but, conditionally on β, the cdf of *x* is invariant, see Tallis *et al (loc cit)*. Screening changes this steady state by detecting the disease at considerably smaller diameters. It can be shown that when a screened population is referred to a steady state, unscreened population for an estimate of survival time based on *a* and *x,* an over-estimate of perhaps half a year can be anticipated.

We have examined both types of bias in detail and conclude that together, they should amount to less than .5 years. In view of the size of the standard errors, and keeping in mind that the historical data of 1980-86 may underestimate the survival experiences of 1990-2002, no attempt at any correction has been made.

Acknowledgements

We wish to thank BreastScreen SA for providing data from breast cancer cases detected under their screening program of South Australian women over the period 1990-2002. We also wish to thank the staff of the South Australian Cancer Registry for their assistance over an extended period of time in providing incidence and survival data. A joint collaboration of the authors and Dr. Martin Cook, Institute of Medical and Veterinary Science, over the period 1980-86 produced the breast tumour diameter data which have since been incorporated into the South Australian Cancer Registry, and which we labelled as the SACR diameter data set.

References

Black, W.C., Haggstrom, D.A. & Welch, H.G. (2002) All-cause mortality in randomised trials of cancer screening. Journal of the National Cancer Institute, 94,167-173.

Cady, B. (1997) New era in breast cancer. Impact of screening on disease presentation. Surgical Oncology Clinics of North America, 6 (2), 195-202.

Cox, D.R. & Hinkley, D.V. (1974) Theoretical Statistics. Chapman & Hall, London.

Duncan, W. & Kerr, G.R. (1976) The curability of breast cancer. British Medical Journal, 2(6039), 781-783.

Efron, B. & Tibshirani, R.J. An introduction to the bootstrap. Chapman & Hall, New York.

Feller, W. (1960) An introduction to probability theory and its applications. V1, 2nd edition, John Wiley & Sons, New York.

Fisher, B., Slack, W.D. & Bross, I.D.J. (1969) Cancer of the breast: size of neoplasm and prognosis. Cancer, 24, 1071-1080.

Jansen, J.T. & Zoetelief, J. (1995) MBS: a model for risk benefit analysis of breast cancer screening. The British Journal of Radiology, 68, 141-149.

Juffs, H.G. & Tannock, I.F. (2002) Screening trials are even more difficult than we thought they were. Journal of the National Cancer Institute, 94, 156-157.

Knox, E.G. (1973) A simulation system for screening procedures. In : McLachlan G., Future and present indicatives, problems and progress in medical care, 9th series. Oxford, London. Nuffield Provincial Hospitals Trust.

Kusama, S., Spratt, J.S. Jr, Donegan, W.L., Watson, F.R. & Cunningham, C. (1972) The gross rates of growth of human mammary carcinoma. Cancer, 30, 594-599.

MISCAN (1984) Habbema, J.D.F., van Oortmarssen, G.J., Lubbe, J.T.N. & van der Maas, P.J. The MISCAN simulation program for the evaluation of screening for disease. Computer Methods and Programs in Biomedicine, 20, 79-83.

O'Neill, T.J., Tallis, G.M. & Leppard, P. (1995) A review of the technical features of breast cancer screening illustrated by a specific model using South Australian cancer registry data. Statistical Methods in Medical Research, 4, 55-72.

Robinson, J.I., Crane, C.E.B., King, J.M., Scarce, D.I. & Hoffmann, C.E.J. (1996) The South Australian Breast X-Ray Service: Results from a state-wide mammographic screening program. British Journal of Cancer, 73, 837-842.

SEER (2003) Surveillance, Epidemiology and End Results program, US National Cancer Institute. http//www.seer.cancer.gov

Schwartz, M. (1978) An analysis of the benefits of serial screening for breast cancer based upon a mathematical model of the disease. Cancer, 41, 1550-1564.

South Australian Cancer Registry (1993) Epidemiology of Cancer in South Australia.

Tabar, L., Duffy, S.W., Yen, M.F., Warwick, J., Vitak, B., Chen, H.H. & Smith, R.A. (2002) All-cause mortality among breast cancer patients in a screening trial: support for breast cancer mortality as an end point. Journal of Medical Screening, 9(4), 159-167.

Tallis, G.M. & Chesson, P.L. (1976) The relationship between a class of asymptotically normal estimators and goodness of fit tests. Australian Journal of Statistics, 18, 53-61.

Tallis, G.M., Leppard, P. & O'Neill, T.J. (1988) The analysis of survival data from a central cancer registry with passive follow-up. Statistics in Medicine, 7, 483-490.

Tallis, G.M., Leppard, P. & O'Neill, T.J. (2003) A note on the sensitivities of self-reporting and screen detection of primary breast tumours. Australian & New Zealand Journal of Statistics, 45(1), 7-18.

Vorherr, H. (1981) Pathobiology of breast cancer; hypothesis of biological predetermination and long-term survival. Klin Wochenschr, 59 (15), 819-829.

USDHHS (1996) DHHS Publication: Guide to Clinical Preventative Services, 2nd edition, Screening for breast cancer. (US Department of Health and Human Services)

Zurrida, S., Morabito, A., Galimberti, V., Luini, A, *et al* (1999) Importance of the level of auxiliary involvement in relation to traditional variables in the prognosis of breast cancer. International Journal of Oncology, 15(3), 475-480.