

Research article

Open Access

## Using observational data to estimate an upper bound on the reduction in cancer mortality due to periodic screening

Stuart G Baker\*<sup>1</sup>, Diane Erwin<sup>2</sup>, Barnett S Kramer<sup>3</sup> and Philip C Prorok<sup>1</sup>

Address: <sup>1</sup>Biometry Research Group, Division of Cancer Prevention, National Cancer Institute, USA, <sup>2</sup>Information Management Services, USA and <sup>3</sup>Office of Disease Prevention, National Institutes of Health, USA

Email: Stuart G Baker\* - sb16i@nih.gov; Diane Erwin - derwin@btinternet.com; Barnett S Kramer - KramerB@OD.NIH.GOV; Philip C Prorok - pp2g@nih.gov

\* Corresponding author

Published: 6 March 2003

Received: 15 November 2002

*BMC Medical Research Methodology* 2003, **3**:4

Accepted: 6 March 2003

This article is available from: <http://www.biomedcentral.com/1471-2288/3/4>

© 2003 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:** Because randomized cancer screening trials are very expensive, observational cancer screening studies can play an important role in the early phases of screening evaluation. Periodic screening evaluation (PSE) is a methodology for estimating the reduction in population cancer mortality from data on subjects who receive regularly scheduled screens. Although PSE does not require assumptions about natural history of cancer it requires other assumptions, particularly progressive detection – the assumption that once a cancer is detected by a screening test, it will always be detected by the screening test.

**Methods:** We formulate a simple version of PSE and show that it leads to an upper bound on screening efficacy if the progressive detection assumption does not hold (and any effect of birth cohort is minimal) To determine if the upper bound is reasonable, for three randomized screening trials, we compared PSE estimates based only on screened subjects with PSE estimates based on all subjects.

**Results:** In the three randomized screening trials, PSE estimates based on screened subjects gave fairly close results to PSE estimates based on all subjects.

**Conclusion:** PSE has promise for obtaining an upper bound on the reduction in population cancer mortality rates based on observational screening data. If the upper bound estimate is found to be small and any birth cohort effects are likely minimal, then a definitive randomized trial would not be warranted.

### Background

Because randomized cancer screening trials are very expensive and sometimes difficult to implement, observational cancer screening studies can play an important role in estimating the efficacy of cancer screening during early phases of evaluation of the screening test. However the standard methodology for observational cancer screening studies has various limitations. Case-control studies require adequate case identification, eligibility criteria for

equal access of cases and controls to screening, distinguishing symptomatic and diagnostic tests, and adjustments for self-selection bias [1]. Cohort studies often involve natural history models which rest upon assumptions about the duration of preclinical cancer or the growth rate of the tumor, the sensitivity of the screening test, and how screening affects cancer mortality. Some examples can be found in [2–6]. Importantly natural history

models based only on observational data must implicitly assume no selection bias, a very tenuous assumption.

In contrast, periodic screening evaluation (PSE), which combines estimates from screened subjects to estimate the reduction in population cancer mortality associated with periodic cancer screening [7–9], does not involve natural history models and the associated assumptions. However a different set of assumptions is required. In certain situations these assumptions may be more plausible than the natural history assumptions, so in some circumstances, the method may be complementary and possibly superior, to the natural history modeling approach.

PSE starts with the following estimates based directly on observed data from a few screenings at regular intervals over various ages (1) age-specific rates of cancer detection on first screening, interval cancers, and cancer detection on subsequent screenings, and (2) cancer fatality rates following cancer detection at first screening, interval cases, and cancer detection on subsequent screening. For evaluation it is also necessary to estimate (1) the age-specific rate of cancer detection in the *absence* of screening and (2) the cancer mortality rate following detection in the *absence* of screening. Because there are no randomized controls the challenge is to estimate rates in the absence of screening in a manner that mitigates selection bias.

In estimating the cancer detection rate in the absence of screening PSE mitigates selection bias in a unique manner. As will be discussed, PSE estimates the age-specific detection rate in the absence of screening as the sum of the age-specific rates of detection for cancers on the first screening, interval cancers, and cancers on subsequent screenings, minus the age-specific rate of detection for cancer on the first screening in subjects one year older. This estimation assumes progressive detection, namely that once a cancer is detected on screening it will always be detected on screening. Previous versions of PSE made this assumption. However progressive detection is not likely to hold for many types of screening modalities. Fortunately, as we discuss, if progressive detection is violated, the estimated detection rate in the absence of screening is an upper bound and this can lead to useful estimates.

In estimating the cancer fatality rate following detection in the absence of screening, earlier versions of PSE used data from refusers and simply assumed no selection bias. To avoid this assumption (and the need to collect data from refusers), we estimated the cancer mortality rate following detection in the absence of screening by the cancer mortality rate in interval cancers. As we discuss, this also leads to an upper bound (i.e. optimistic) estimate of screening efficacy.

Thus, this version of PSE circumvents the problem of selection bias by estimating an upper bound. The specific estimates of cancer detection rates and cancer fatality rates after cancer detection are not meaningful as separate quantities. Fortunately, one can longitudinally combine the estimates to estimate an upper bound on the reduction in population cancer mortality associated with periodic screening. (Given these data, it is not possible to estimate reduction in population cancer mortality for other intervals between screenings or after periodic screening has stopped). The longitudinal combination of cross-sectional estimates, which also appears in earlier versions of PSE [7–9], is similar to G-computation [10] and the method of Flanders and Longini [11].

If an upper bound estimate of screening efficacy is small, a definitive randomized trial to evaluate the effect of screening on cancer mortality would not be warranted. Thus the upper bound estimate is helpful only if is not unreasonably large. To determine if the upper bound is reasonable, we estimated its value using data from screened subjects in randomized trials of colorectal cancer screening [12,13], breast cancer screening [14], and lung cancer screening [15,16]. We then compared this estimate to a modified PSE estimate using data from all subjects, so that estimates of age-specific cancer detection in the absence of screening and cancer mortality in the absence of screening are based on data from randomized controls and refusers.

We also compared the PSE estimates with estimates based on a comparison of outcomes in the two randomized groups, adjusting for refusers and mitigating the effect of dilution after stopping screening [17]. It is important to bear in mind that the two estimates are answering different questions. For PSE, the question is "What is the effect of periodic screening starting at a given age and ending at a later age?" For comparing randomized groups, the question is "What is the effect of the particular screening program in the intervention group?"

## Methods

### Simple formulation of PSE

We derive a simple formulation of PSE and show that it gives an upper bound on the estimated reduction in population cancer mortality.

PSE requires two types of data from subjects who receive two or more screenings at regular intervals. The first type of data are the numbers of subjects who receive each screen and who are detected with cancer as a result of screening or in the interval between screens. (See Tables 1,2,3,4) The second type of data are numbers of subjects with cancer who die from cancer and are in the risk set each year after diagnosis. (See Table 5)

**Table 1: Age-specific breast cancer detection data from HIP study.**

age	number detected					number at risk				
	F	I	S	R*	C*	F	I	S	R*	C*
40	0	0	0	0	1	302	302	204	196	550
41	1	0	0	0	2	838	1032	619	552	1839
42	4	1	1	0	3	1044	1789	1170	970	3219
43	1	0	1	3	5	1084	2563	1578	1396	4642
44	0	2	1	2	8	1062	3134	1816	1619	5523
45	3	2	2	1	10	1040	3412	1831	1687	5696
46	1	3	2	5	8	985	3407	1748	1650	5618
47	3	3	3	3	8	1014	3336	1765	1653	5478
48	2	5	6	3	9	933	3207	1700	1583	5434
49	1	1	0	1	13	978	3211	1689	1602	5349
50	1	3	0	3	14	915	3082	1607	1635	5306
51	2	3	3	2	10	958	3079	1615	1618	5319
52	3	4	4	3	12	907	3021	1632	1665	5255
53	3	1	3	3	8	885	3032	1525	1629	5189
54	4	3	3	2	8	926	2954	1547	1593	5151
55	3	1	3	6	8	872	2891	1511	1559	5085
56	3	2	3	2	5	872	2841	1512	1559	5018
57	4	2	4	1	12	765	2746	1427	1548	4873
58	3	2	5	6	12	719	2540	1293	1495	4610
59	3	4	1	4	5	635	2378	1177	1400	4279
60	3	2	1	2	12	548	2125	1047	1294	3873
61	4	1	4	3	8	502	1890	911	1199	3487
62	0	0	1	0	10	534	1745	845	1055	3190
63	2	0	3	2	12	434	1565	787	949	2868
64	0	1	3	2	7	376	1422	704	869	2615

F = first screen, I = interval between screens, S = subsequent screen, R = refuser, C = control \*data excluded for PSE with only screened subjects

PSE involves three steps: (1) estimate the age-specific incidence of cancer associated with different types of detection: first screen, interval between screens, subsequent screen, refusers, and controls, if available, (2) estimate cancer fatality rates after cancer detection, and (3) combine these estimates to estimate the reduction in population cancer mortality associated with periodic cancer screening.

Because PSE requires regular intervals, the analysis is restricted to screenings that occur "on-time", namely, within a window of time close to the length of interval. The length of the window for on-time screenings is somewhat arbitrary. A very wide window discards relatively few subjects; however it may introduce bias into calculations that are based on assuming the screening interval equals the midpoint of the window. Alternatively, a narrow window might discard too much data, increasing the chance for bias from nonrandom exclusion of subjects. In designing a study for PSE analysis, the screening intervals should be as regular as possible. Without loss of generality in the following discussion, we assume a regular interval between screens of 1 year.

**Step 1: Age-specific cancer incidence**

PSE requires estimates of age-specific cancer incidence for the following types of cancer detection: type F, detection on the first screening; type I, detection in the interval between screenings; type S, detection on screening subsequent to the first; and type A, detection in the absence of screening. The incidence of type F detection at age a is  $q_F(a) = x_F(a) / n_F(a)$ , where  $x_F(a)$  is the number of subjects detected with cancer as a result of a first screening at age a and  $n_F(a)$  is the number who received the first screening at age a. The incidence of type I detection at age a is  $q_I(a) = x_I(a) / n_I(a)$  where  $x_I(a)$  is the number of cases in the interval after screening at age a and  $n_I(a)$  is the number at risk at the start of the interval. The incidence of type S detection at age a is  $q_S(a) = x_S(a) / n_S(a)$  where  $x_S(a)$  is the number of subjects detected with cancer as a result of an "on-time" screening after a previous screening at age a and  $n_S(a)$  is the number of subjects who received an "on-time" screening after the previous screening at age a. Although a type S detection occurs on screening at age a + 1, for mathematical convenience, it is associated with screening at age a. We cannot observe type A detection from data on sub-

**Table 2: Age-specific colorectal cancer detection data from Minnesota Colon Cancer Control Study with annual screening**

age	number detected					number at risk				
	F	I	S	R*	C*	F	I	S	R*	C*
50	0	0	0	0	0	397	527	418	71	818
51	0	0	0	0	0	492	919	821	122	1377
52	0	0	0	0	0	518	1311	1186	165	1973
53	0	0	2	0	0	502	1675	1542	211	2509
54	0	2	2	0	1	536	2058	1923	248	3093
55	1	0	0	0	2	535	2251	2092	307	3711
56	0	0	2	1	3	551	2318	2157	359	4355
57	1	0	0	0	2	602	2442	2270	395	4965
58	0	1	1	0	6	603	2550	2405	446	5567
59	1	0	0	0	6	573	2820	2652	482	6189
60	0	2	2	0	5	651	3350	3125	527	6840
61	2	0	4	1	6	637	3823	3633	584	7480
62	1	0	5	1	11	655	4273	4077	633	8151
63	2	0	2	0	8	651	4747	4563	684	8770
64	2	1	7	3	7	635	5138	4910	719	9314
65	1	2	4	1	19	613	5231	5032	760	9854
66	2	1	4	0	26	596	5313	5124	779	10196
67	2	0	8	1	17	574	5318	5136	765	10185
68	2	2	3	1	19	501	5170	4998	767	10081
69	2	2	4	0	11	429	4944	4761	754	9911
70	1	1	2	3	18	411	4831	4673	753	9706
71	3	2	2	5	18	380	4658	4500	743	9376
72	0	1	5	5	15	313	4376	4243	730	8979
73	1	1	7	2	24	271	4140	4029	713	8474
74	0	0	7	1	22	252	3865	3746	673	8013
75	2	1	4	2	13	206	3574	3492	664	7512

F = first screen, I = interval between screens, S = subsequent screen, R = refuser, C = control \* data excluded for PSE with only screened subjects

jects screened. However, as derived in [8], one can estimate the probability of type A detection by

$$q_A(a) = q_F(a) + q_I(a) + q_S(a) - q_F(a + 1), \quad (1)$$

if the following key assumption holds.

*Assumption 1. Progressive Detection*

Once a subject is detectable on screening the subject will always be detectable on screening.

(The quantity  $q_F(a + 1)$  in (1) is the cancer detection rate on the first screening among subjects age  $a + 1$  at first screening.) The graphical proof of (1) in Figure 1 generalizes the graphical proof in [8] to allow some subjects with preclinical cancer to be missed on screening. In Figure 1, *Assumption 1* corresponds to  $\delta = 0$ , where  $\delta$  is the probability that some individuals with preclinical cancer would be detected if screened at age  $a$  but missed if screened at age  $a + 1$ . As an example of *Assumption 7*, consider a woman who would have been detected with breast cancer at age

50, but is not screened at age 50 for reasons unrelated to the screening or any possibility of cancer. Under *Assumption 1*, if the woman were screened at age 51, she would be detected with cancer. There is sometimes confusion about how this relates to sensitivity of the screening test. As shown in Figure 1, *Assumption 1* implies that the sensitivity of the screening test equals 1 if a previous screening test would have detected cancer.

Ideally  $q_F(a + 1)$  estimates the probability of detecting cancer on a first screen at age  $a + 1$  one year after the start of the study. Because there are no data on subjects first screened after a one year delay, we compute  $q_F(a + 1)$  from subjects age  $a + 1$  at the start of the study. This procedure requires the following additional assumption.

*Assumption 2. No Birth Cohort Effect*

Given age, year of birth provides no additional information for predicting cancer incidence on the first screen.

**Table 3: Age-specific colorectal cancer detection data from Minnesota Colon Cancer Control Study with biennial screening**

age	number detected					number at risk				
	F	I	S	R*	C*	F	I	S	R*	C*
50	1	0	0	0	0	891	1015	658	102	1377
52	0	3	1	0	0	997	1789	1410	196	2510
54	2	1	1	0	3	1047	2594	2139	291	3723
56	1	1	2	1	5	1141	2986	2521	405	4979
58	1	2	3	1	12	1145	3279	2754	483	6212
60	2	1	3	1	11	1230	4227	3717	590	7508
62	3	6	3	3	19	1295	5123	4543	688	8774
64	6	2	3	2	26	1267	5914	5328	783	9734
66	5	5	7	3	43	1160	5924	5387	813	9821
68	3	2	7	4	30	961	5740	5240	822	9548
70	3	8	5	2	36	761	5253	4881	821	8958
72	4	10	4	4	39	621	4640	4307	787	8097
74	3	10	2	0	35	460	4171	3871	727	7133

F = first screen, I = interval between screens, S = subsequent screen, R = refuser, C = control \* data excluded for PSE with only screened subjects

**Step 2: Cancer-fatality rates among cases**

PSE also requires estimates of cancer fatality-rates among cases. The estimated probability of death from cancer within 5 years of type *d* cancer detection at age *a* is

*pr* (cancer death | type *d* detection at age *a*)

$$= Surv(a, a + 1)h_{d1} + \sum_{i=2}^5 Surv(a, a + i) \prod_{t=1}^{i-1} (1 - h_{dt}) h_{di}, \quad (2)$$

where  $h_{di}$  is the estimated case-fatality rate from cancer in year *i* after type *d* detection, and  $Surv(a, a + i)$  is the probability of surviving competing risks from age *a* to age *a + i*. See also Gooley et al [18]. For year *j* after type *d* detection (*d* = F, I, S), the estimated case-fatality rate from the cancer under study is  $h_{dj} = x_{dj} / r_{dj}$ , where  $x_{dj}$  is the number of cancer deaths among cases at year *j* after type *d* detection, and  $r_{dj}$  the number of cases with type *d* detection who are at risk at year *j* since detection.

We approximate  $Surv(a, a + i)$  by the probability of surviving from age *a* to age *a + i* obtained from demographic data stratified by sex [19]. We approximate the probability of surviving competing risks each year over five years as the probability of surviving competing risks at the midpoint of two years, i.e.,  $Surv(a, a + i) = Surv(a, a + 3)$  for *i* = 1,2,3,4,5. This lets us approximate (2) by

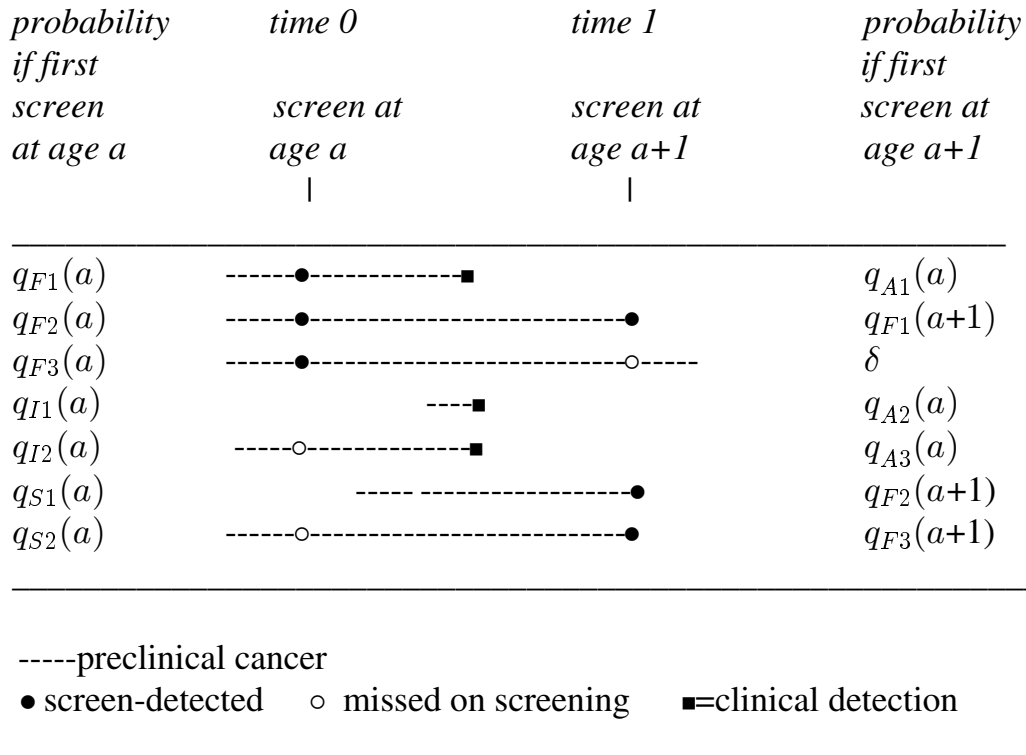
$$pr \text{ (cancer death in cases | type } d \text{ detection at age } i) = Surv(a, a + 3) m_d, \quad (3)$$

where  $m_d = 1 - \prod_{t=1}^5 (1 - h_{dt})$  is the estimated probability of cancer death within five years of type *d* detection

conditional on no death from competing risk and  $Surv(a, a + 3)$  is the approximate probability of surviving competing risks within five years of type *d* detection.

A major challenge is how to estimate  $m_A$ , the probability of cancer fatality within five years of type A detection (i.e. in the absence of screening) conditional on no death from competing risk. Previous approaches [1-3] used data from refusers, substituting  $m_R$  for  $m_A$ . However this requires a strong unreasonable assumption as well as data from refusers, which is often not available.

As an alternative, we estimate an upper bound on the reduction in the population cancer mortality rate from screening by estimating the cancer fatality rate in the absence of screening using data from interval cancers, namely, substituting  $m_I = m_A$ . The reason this is an upper bound is that cancers arising in the absence of screening are composed of cancers that would have arisen in the interval after a negative screening (had there been screening) and cancers that would have been detected on a previous screening (had there been screening). The latter cases are presumably slower growing (a type of length-biased sampling) with better survival, so using only the interval cancers artificially increases the estimated cancer-fatality rate in the absence of screening. (One caveat is that the survival of interval cancers may be improved due to increased awareness of treatment options that would occur as part of a screening program. If the effect of length bias is relatively small, substituting  $m_I$  for  $m_A$  might not be an upper bound, although we believe it would be a reasonable approximation.)



**Figure 1**

Graphical proof that  $q_A(a) + \delta = q_F(a) + q_I(a) + q_S(a) - q_F(a+1)$  The seven lines represent all possible situations in which cancer is diagnosed over the interval  $(a, a + 1)$ . Therefore, the sum of all their probabilities represents the incidence of cancer over that interval. Each row corresponds to a unique potential history of cancer detection, so the probabilities on the left and right side of each row are identical. The labels on each side are different because the realized events differ depending on whether screening started at time 0 (left side) or time 1 (right side). Defining  $q_d(a) = \sum_j q_{dj}(a)$  and setting the left side sum of probabilities equal to the right side sum gives the formula on top. Under Assumption 1, that once a subject is detectable on screening, the subject will always be detectable on screening,  $\delta = 0$ .

**Step 3. Reduction in population cancer mortality rates due to periodic screening**

PSE estimates the reduction in population cancer mortality rates due to starting periodic cancer screening at age  $a$  instead of age  $b$ , where ages  $a$  and  $b$  lie in the range of ages at initial screening. (This estimate accounts for competing risks through the use of  $Surv(a, a + 3)$ ). To avoid different

rates of overdiagnosis between comparison groups, PSE compares population cancer mortality rates in two hypothetical scenarios involving full compliance: Scenario  $\mathcal{P}$ , periodic screening from age  $a$  until age  $b$  and Scenario  $\mathcal{N}$ , no periodic screening from age  $a$  to age  $b-1$  followed by

**Table 4: Age-specific lung cancer detection data from Mayo Lung Project**

age	number detected					number at risk				
	F	I	S	R*	C*	F	I	S	R*	C*
51	2	1	1	0	0	375	3086	2363	54	706
52	5	1	2	0	0	392	3147	2444	88	1264
53	4	0	3	0	1	365	3113	2414	117	1732
54	4	1	1	1	1	374	3125	2417	114	1684
55	5	2	4	0	4	354	3106	2419	139	1736
56	4	1	2	0	2	364	3126	2414	322	1722
57	3	0	6	0	1	366	3100	2417	136	1663
58	5	0	4	0	3	317	2985	2353	110	1664
59	4	1	3	1	1	330	3005	2385	101	1641
60	4	2	5	0	10	315	2933	2330	74	1601
61	5	0	4	0	2	301	2836	2275	61	1526
62	5	1	2	0	3	302	2742	2162	57	1485
63	5	0	3	0	5	248	2563	2072	60	1504
64	6	2	9	1	3	242	2410	1903	59	1389
65	3	4	2	1	5	232	2224	1768	65	1355
66	5	1	2	0	0	170	2004	1612	65	1272
67	3	1	6	0	2	176	1834	1446	64	1176
68	5	3	3	0	2	149	1550	1213	54	1099
69	3	0	5	0	3	124	1357	1068	29	963

F = first screen, I = interval between screens, S = subsequent screen, R = refuser, C = control (after 6 year washout period) \* data excluded for PSE with only screened subjects

screening at age  $b$ . Scenario  $\mathcal{P}$  involves either detection on the first screen at age  $a$  or detection in the interval or on subsequent screens to age  $b$ . Scenario  $\mathcal{N}$  involves either detection in the absence of screening from ages  $a$  to  $b - 1$  or detection on a first screen at age  $b$ . Screening at age  $b$  in both Scenarios  $\mathcal{N}$  and  $\mathcal{P}$  avoids differential overdiagnosis rates because, if Assumption 1 holds, both scenarios  $\mathcal{N}$  and  $\mathcal{P}$  specify equal probabilities of detecting cancer by age  $b$ .

More formally we can write the reduction in population cancer mortality rates associated with starting periodic cancer screening at age  $a$  instead of age  $b$  as

$$g = pr(\text{cancer mortality under Scenario } \mathcal{N}) - pr(\text{cancer mortality under Scenario } \mathcal{P}), \quad (4)$$

where

$$pr(\text{cancer mortality under Scenario } \mathcal{N}) = \sum_{i=a}^b q_A(i) pr(\text{cancer death in cases | type A detection at age } i)$$

$$+ q_F(b) pr(\text{cancer death in cases | type } F \text{ detection at age } b) pr(\text{cancer mortality under Scenario } \mathcal{P})$$

$$= q_F(a) pr(\text{cancer death in cases | type } F \text{ detection at age } a) + \sum_{i=a}^b q_I(i) pr(\text{cancer death in cases | type } I \text{ detection at age } i) + \sum_{i=a}^b q_S(i) pr(\text{cancer death in cases | type } S \text{ detection at age } i)$$

and cancer death in cases refers to death from cancer in cases within five years of cancer detection. Accounting for deaths from competing risks, the estimated probability of type  $d$  detection at age  $i$ , conditional on being alive at age  $a$ , is

$$p_r^{\wedge}(\text{type } d \text{ detection at age } i) = Surv(a, i) q_d(i) \quad (5)$$

Substituting (5) and (3) into (4) gives

**Table 5: Data on numbers of cancer deaths following detection**

year since detection	number of cancer deaths					number at risk at start of year				
	F	I	S	R*	C*	F	I	S	R*	C*
Minnesota Colon Cancer Control Study (annual screening)										
1	2	1	1	3	36	26	19	77	27	259
2	3	2	2	1	10	24	18	74	24	217
3	0	2	2	2	17	21	16	72	23	202
4	0	1	3	0	7	20	13	69	20	177
5	0	1	0	2	6	20	12	64	19	165
Minnesota Colon Cancer Control Study (biennial screening)										
1	2	9	0	7	36	34	51	41	21	259
2	0	4	2	0	10	31	42	40	14	217
3	4	3	3	0	17	30	37	38	13	202
4	2	0	1	1	7	26	33	33	12	177
5	0	0	3	1	6	23	31	29	11	165
HIP study of breast cancer screening										
1	2	2	1	9	19	54	47	57	59	210
2	2	6	2	5	27	52	44	56	49	189
3	0	5	0	2	12	50	38	54	44	161
4	1	1	1	2	12	50	33	54	41	148
5	3	2	3	1	9	48	31	53	38	133
Mayo Lung Project										
1	32	11	11	2	21	80	21	67	4	48
2	7	5	13	1	4	46	10	56	2	20
3	6	0	2	0	4	36	5	41	0	13
4	3	0	1	0	2	29	4	37	0	4
5	1	0	1	0	0	25	3	29	0	1

F = first screen, I = interval between screens, S = subsequent screen, R = refuser, C = control \*data excluded for PSE with screened subjects

$$\hat{g}_{basic} = \left( \sum_{i=a}^b \text{Surv}(a,i)q_A(i)\text{Surv}(i,i+3)m_A + \text{Surv}(a,b)q_F(b)\text{Surv}(b,b+3)m_F \right) - \left( \text{Surv}(a,a)q_F(a)\text{Surv}(a,a+3)m_F + \sum_{i=a}^b \text{Surv}(a,i)q_I(i)\text{Surv}(i,i+3)m_I + \sum_{i=a}^b \text{Surv}(a,i)q_S(i)\text{Surv}(i,i+3)m_S \right). \tag{6}$$

To simplify (6), we define

$$Q_d = \sum_{i=a}^b \text{Surv}(a,i+3)q_d(i), \text{ for } d = S, I, A.$$

$$q_{F0}(basic) = \text{Surv}(a,a+3)q_F(a)$$

$$q_{F0}(basic) = \text{Surv}(a,b+3)q_F(b) \tag{7}$$

Substituting (7) into (4) gives the following simple estimate,

$$\hat{g}_{basic} = (Q_A m_A + q_{F1}(basic) m_F) - (q_{F0}(basic) m_F + Q_I m_I + Q_S m_S). \tag{8}$$

To increase the stability of  $\hat{g}_{basic}$ , we used averages over  $k = 3$  intervals, for the probabilities of detection on the first screenings at age  $a$  and at age  $b$ ,

$$q_{F0} = \sum_{i=a}^{a+k-1} \text{Surv}(a,i+3)q_F(i)/k,$$

$$q_{F1} = \sum_{i=b}^{b-k+1} \text{Surv}(a,i+3)q_F(i)/k. \tag{9}$$

Substituting (9) into (8) gives the modified estimate

$$\hat{g}_{modified} = (Q_A m_A + q_{F1} m_F) - (q_{F0} m_F + Q_I m_I + Q_S m_S). \tag{10}$$

Invoking (1), we substitute  $q_{F0} + Q_I + Q_S - q_{F1}$  for  $Q_A$  (10). As discussed previously to obtain an upper bound, we also set  $m_A = m_I$ . This gives the following estimated upper bound in the reduction in population cancer mortality from periodic screening and its asymptotic variance

$$\hat{g}_{upp} = (q_{F0} - q_{F1}) (m_I - m_F) + Q_S (m_I - m_S), \tag{11}$$



$$\text{var}(\hat{g}_{\text{upp}}) = (q_{F1} - q_{F0})^2 (v_I + v_F) + Q_S^2 (v_I + v_S) + (V_{F0} + V_{F1})(m_I - m_F)^2 + V_S (m_I - m_S)^2. \quad (12)$$

where  $V_d = \text{var}(Q_d) = \sum_{j=a}^b \text{Surv}(a, j + 3)^2 q_d(j) / n_d(j)$ ,  
 $V_{F0} = \text{var}(q_{F0}) = \sum_{j=a}^{a+k} \text{Surv}(a, j + 3)^2 q_F(j) / k^2$ ,  
 $V_{F1} = \text{var}(q_{F1}) = \sum_{j=b-k}^b \text{Surv}(a, j + 3)^2 q_F(j) / k^2$ ,  
 $v_d = \text{var}(m_d) = m_d^2 \sum_{j=a}^b x_{dj} / (r_{dj} - x_{dj})$  from Greenwood's formula [20].

**Upper bound if Assumption 1 is violated**

There are two basic scenarios in which Assumption 1 could be violated. First the cancer, or at least the detectable part of cancer, could regress over time. This would most likely occur if the cancer were at a very early stage. Of course, early lesions are the principal targets of screening tests. Second, chance fluctuations in the results of the screening test might mask cancer detection, particularly if the interval between the screening tests were small. For example if the screening test were based on a sampling of cells, the screening test may, by chance, not include any of the tumor cells. For many screening modalities Assumption 1 may not hold.

As shown in Figure 1, if Assumption 1 did not hold, PSE would estimate the cancer incidence rate in the absence of screening as  $q_A(a) + \delta$ , for  $\delta > 0$ , instead of  $q_A(a)$ . Thus if Assumption 1 were not satisfied, PSE would overestimate the cancer incidence rate in the absence of screening which overestimates the reduction in the population cancer mortality rate. On the other hand, a violation of Assumption 1 would also imply that some cancers detected on screening in  $\mathcal{P}$  would not have been detected on the last screening in  $\mathcal{N}$  which would lower the reduction in the population cancer mortality rate in (4). However we think this latter situation would have a small impact relative to the former which involves the entire span of ages at screening and not just the last age at screening.

Thus (11) is an upper bound for two reasons. First it uses interval cancers to estimate fatality rates following cancer diagnosis in the absence of screening. Second it is an upper bound if Assumption 1 is violated.

**Validation methodology**

If an upper bound is too large it will not be useful. To investigate the upper bound, we used data from three randomized screening trials to compare PSE estimates based on screened subjects with PSE estimates based on all subjects.

In computing PSE estimates for all subjects, the progressive detection assumption is not necessary. To estimate  $q_A(j)$  we use a simple noncompliance adjustment for randomized trials (see [17] and references therein),  $q_A(j) = (q_C(j) - q_R(j) \pi) / (1 - \pi)$  where  $q_C(j)$  is the age-specific cancer incidence rate in controls,  $q_R(j)$  is the age-specific cancer incidence rate in refusers, and  $\pi$  is the fraction of subjects who refused screening. Also with data from a randomized trial it is not necessary to use interval cancer cases to estimate the case fatality rate in the absence of screening. Instead we estimate  $m_A = (m_C - m_R \pi) / (1 - \pi)$ , where subscript C refers to randomized controls and R refers to refusers. Substituting  $(Q_C - Q_R \pi) / (1 - \pi)$  for  $Q_A$  in (10) and using  $m_A$  gives the following estimated reduction in population mortality from cancer screening and its variance,

$$\hat{g}_{\text{all}} = \left( \frac{Q_C - Q_R \pi}{1 - \pi} m_A + q_{F1} m_F \right) - (q_{F0} m_F + Q_I m_I + Q_S m_S). \quad (14)$$

$$\text{var}(\hat{g}_{\text{all}}) \approx (q_{F1} - q_{F0})^2 v_F + Q_I^2 v_I + Q_S^2 v_S + Q_A^2 \frac{v_C + v_R \pi^2}{(1 - \pi)^2} + (V_{F0} + V_{F1}) m_F^2 + V_I m_I^2 + V_S m_S^2 + \frac{V_C + V_R \pi^2}{(1 - \pi)^2} m_A^2. \quad (15)$$

where  $\pi$  is treated as known, which is reasonable due to the large sample size.

We computed (12)-(15) using data from the following three randomized screening trials.

**Minnesota Colon Cancer Control Study (MCCCS)**

Between 1975 and 1978 investigators randomized approximately 45,000 subjects to either 5 annual fecal occult blood screenings, 3 biennial screenings, or no screening [6,7]. Due to a lower than expected death rate among controls, the investigators resumed screening between 1982 and 1986. After a hiatus of screening of between 3 and 5 years, the annual screened group received 5 additional annual screenings and the biennial screened group received 3 additional biennial screenings. Approximately 14 percent of subjects randomized to screening did not receive screening. Each screening cycle consisted of six Hemocult slides with planned definitive work-up if any slide showed evidence of occult blood. Screenings for the annual groups were labeled as on time if they were done in the 9 to 15 month time window since the previous screening. Screenings for the biennial group were labeled as on-time if they were done 20 to 28 months since the previous screening. (The longer time window was used to keep the loss of data arbitrarily to no more than 15%) Excluding the first screening after the resumption of screening, approximately 93 percent of the annual subsequent screen-

ings and 85 percent of the biennial subsequent screenings were on-time. The age range for the analysis was 50 to 75. For estimating the age-specific incidence of cancer among controls, we used data collected up to the time of the last screen, which was 16 years after the start of the study. We increased the precision of the estimated age-specific cancer incidence on the first screen by pooling data on the first screening in the annual and biennial arms. For annual cancer screening, age was divided into intervals of 1 year. For biennial screening, age was divided into intervals of 2 years.

#### **Health Insurance Plan of Greater New York (HIP) Study**

Starting in 1963, approximately 60,000 women were randomly assigned to either a study group invited for four annual mammograms and physical examinations or to a control group that received no screening within the study [8]. Approximately 1/3 of the subjects in the study group refused the first screening and received no screenings. Screenings were labeled on-time if they were done 9 to 15 months after a previous screening. Approximately 79 percent of second screenings, 76 percent of third screenings, and 73 percent of fourth screenings were on time. The age range for the analysis was 40 to 64. For estimating the age-specific incidence of cancer among controls, we used data collected up to the time of the last screen, which was 4 years after the start of the study.

#### **Mayo Lung Project (MLP)**

Between 1971 and 1976 approximately 9,200 male heavy smokers who tested negative on a prevalence (initial) screening were randomized to either a study group urged to undergo radiologic and cytological screening examinations every 4 months for 6 years or a control group that at study entry received a recommendation for annual chest X-rays with no further reminders [9,10]. Approximately 7 percent of the study group subjects did not receive any screenings.

Because PSE requires a single screening time at each round of screening, we restricted PSE to the screenings in which the time between cytology and x-ray was less than 3 weeks. Screenings were labeled as on-time if they were done within 3.5 to 5.5 months of the previous screening and the time between cytology and x-ray was less than 3 weeks. Approximately 85 percent of the subsequent screenings were on-time.

Only yearly age data were available. Because the screenings in the Mayo Lung Project were scheduled at 4 month intervals, yearly cancer incidence data for types *I* and *S* detection are approximated using the sums of counts for three successive screens. Because all subjects in the control group had an initial screening, we pooled data for detec-

tion rates on initial screenings in the study and control groups.

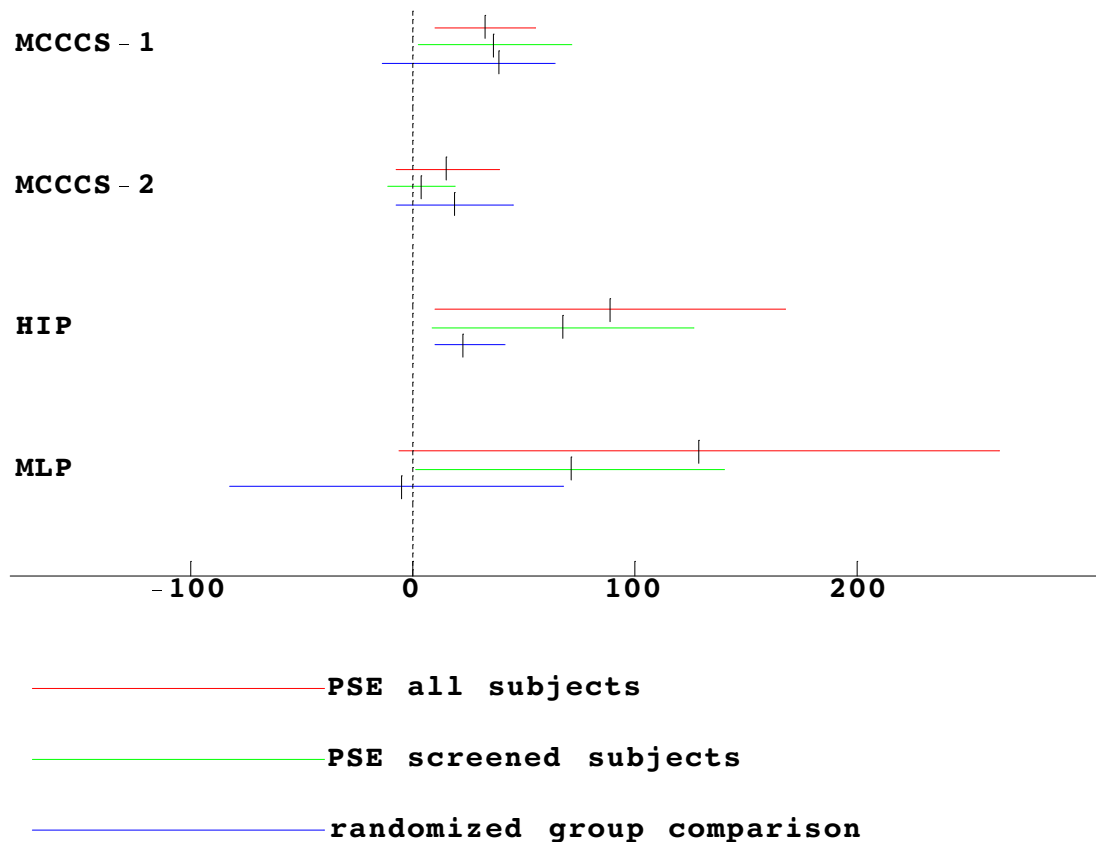
Unfortunately the initial screening in the control group greatly complicated the validation, which requires that no screening be performed in the control group. In order to better approximate a control group that received no screening, we only used data in controls starting 6 years after randomization. The underlying assumption is that by 6 years, most cancers detected on the prevalence screening would have progressed to clinical cancer in the absence of intervention. (This may not be true because of the likely possibility of overdiagnosis [10] and lead times that may exceed 6 years, but it may serve as a useful approximation if the amount of overdiagnosis is small). Due to the 6-year wash-out period, we start the age range for PSE at 51 instead of 45. We chose 6 years for the washout period as a compromise. We thought a longer washout period would have greatly restricted the age range under study and a shorter washout period would have had a much more limited effect.

We illustrate the calculations for the analysis of the HIP data on breast cancer screening. The probability of a women surviving competing risk from age 40 to each successive age up to age 64 is 1., 1., 0.99, 0.99, 0.99, 0.99, 0.99, 0.98, 0.98, 0.98, 0.97, 0.97, 0.96, 0.96, 0.95, 0.95, 0.94, 0.93, 0.92, 0.92, 0.91, 0.9, 0.89, 0.87. Using these probabilities and data from Tables 1 and 5, we computed  $q_F = .00166$ ,  $q_{FS} = .00373$ ,  $Q_I = .0151$ ,  $Q_S = .0338$ ,  $m_I = .334$ , and  $m_S = .123$ , Substituting into (11) gave  $\hat{g}_{upp} = .00676$ . To estimate the variance we computed  $v_I = .00497$ ,  $v_S = .00189$ ,  $V_{F0} = .00000056$ ,  $V_{F1} = .0000023$ . Substituting into (12) gave  $var(\hat{g}_{upp}) = .0000090$ .

#### **Results**

To determine if the PSE estimated upper bound from screened subjects is reasonable, we computed its value (and 95% confidence interval) along with the PSE estimate (and 95% confidence interval) based on all subjects (Figure 2). To account for the correlation we also computed the estimated difference (and 95% confidence interval) between the two types of PSE estimates (Figure 3).

In addition, we computed the estimated cancer mortality reduction between the two randomized groups (and 95% confidence intervals) (Figure 2). To compute this estimate we first computed the estimated efficacy of receiving screening among subjects who would receive screening if offered. This equals the intent-to-treat estimate divided by the fraction in the screening group who did not refuse any screening [17]. For the HIP study and the Mayo Lung Project, in which screening stopped well before the end of follow-up, we also computed an adaptive estimate to mit-



**Figure 2**

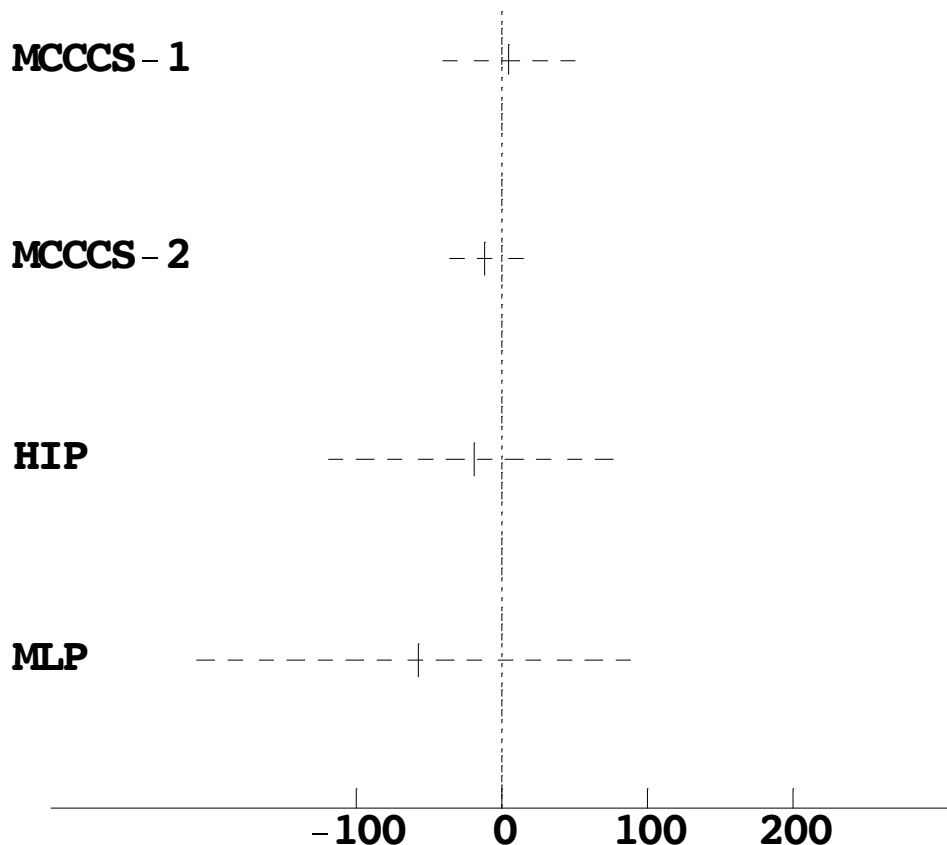
Estimated reduction in cancer deaths per 10,000 due to periodic cancer (with 95% confidence interval). MCCC-1 is Minnesota Colorectal Cancer Control Study, ages 50 to 71, annual screens, MCCC-2 is Minnesota Colorectal Cancer Control Study ages 50 to 71, biennial screens, HIP is Health Insurance Program study of annual breast cancer screening, ages 40 to 64, MLP is Mayo Lung Project, ages 51 to 69 with screens every 4 months. Positive numbers on the horizontal axis represent cancer deaths averted; negative numbers represent excess cancer deaths from screening.

igate the effect of dilution [17]. The adaptive estimate is the estimate at the time after screening stops when the estimate divided by its standard error is largest. Confidence intervals for the adaptive estimate are based on bootstrapping.

**Minnesota Colon Cancer Control Study (MCCC)**

There is a large overlap in the confidence intervals for the PSE estimated upper bound from screened subjects and PSE estimates from all subjects (Figure 2). Also, the 95%

confidence interval for the estimated difference includes zero (Figure 3), indicating that the upper bound estimate is reasonable. The estimated mortality reduction in the screening program is similar to the PSE estimates (Figure 2) because the long-duration of screening (5 annual or 3 biennial screenings followed by 3–5 years hiatus followed by 5 annual or 3 biennial screenings) in the trial approximated the 21 years of periodic screening.



**Figure 3**

*Difference per 10,000 deaths between PSE randomized estimate and PSE nonrandomized upper bound estimate (with 95% confidence interval). MCCC-1 is Minnesota Colorectal Cancer Control Study, ages 50 to 71, annual screens, MCCC-2 is Minnesota Colorectal Cancer Control Study, ages 50 to 71, biennial screens, HIP is Health Insurance Program study of annual breast cancer screening, ages 40 to 64, MLP is Mayo Lung Project, ages 51 to 69 with screens every 4 months.*

**Health Insurance Plan of Greater New York (HIP) Study**

As in the previous example, there is a large overlap in the confidence intervals for the PSE estimated upper bound from screened subjects and PSE estimates from all subjects (Figure 2). Also, similar to the previous example, the 95% confidence interval for the difference included zero (Figure 3) indicating that the upper bound estimate is reasonable. The PSE estimates were higher than the estimated

effect of the screening program in the trial because the former was based on 24 annual screenings and the latter was based on only 4 annual screenings.

**Mayo Lung Project (MLP)**

Unlike the other examples the confidence intervals for PSE estimates from screened subjects and PSE estimates from all subjects differed considerably (Figure 2). We

think *Assumption 1* may not have held due to the short interval between screens and to the fact that the performance of sputum cytology screening depends on sampling of the tumor cells. Although the 95% confidence interval for the difference included zero (Figure 3), its large width means that a substantial bias cannot be ruled out. The PSE estimates were higher than the estimated effect of the screening program in the trial because (i) the former is based on 24 years of screening while the latter is based on only 6 years and (ii) the effect in the latter was reduced by a prevalence screening in the controls. PSE for screened subjects does not use any data from the control group. With PSE for all subjects, we assumed a wash-out period to try to remove the effect of the prevalence screen.

The results indicate that a PSE estimated upper bound based on subjects screened is not unreasonable when compared to the PSE estimate based on all subjects in the randomized trial. Because of sampling variability it is not surprising that the point estimate of the upper bound can be smaller than the point estimate based on all subjects.

We caution that violations of *Assumption 2* could have a substantial impact. *Assumption 2* depends on the cumulative effect of birth cohort from ages  $a$  to  $b$ . According to Moran [21] the relative bias due to violation of *Assumption 2* is particularly large if the age-specific incidence on the first screen changes little with age and interval and subsequent cancers are relatively rare. In that case Moran advised that other methods be applied. One way to reduce bias from *Assumption 2* is to only estimate the effect of screening for at most 5 years. That way the cumulative birth cohort effect would be limited to only 5 years.

## Conclusion

We think the major role of PSE is to rule out screening modalities that have little benefit. This information is useful when making policy decisions about screening, or when considering a large randomized trial to definitively compare benefits and harms of screening strategies. Because PSE is estimating an upper bound when *Assumption 1* is violated, if PSE estimates little reduction in population cancer mortality, the true reduction in population cancer mortality due to periodic screening is likely small. If any effects of birth cohort are minimal, further evaluation with a randomized trial would not be warranted.

## Competing interests

None declared.

## Authors' contributions

SGB wrote the initial draft. DE created Tables 12345 from the data files and provided many helpful suggestions. BSK and PCP substantially improved the manuscript.

## Acknowledgments

The authors thank Pamela Marcus and Ping Hu for their assistance and comments. The manuscript reflects the opinions of the authors and does not necessarily reflect the official opinion of the U.S. Department of Health and Human Services.

## References

1. Cronin KA, Weed DL, Connor RJ and Prorok PC **Case-control studies of cancer screening: Theory and Practice.** *J Natl Cancer Inst* 1998, **90**:498-504
2. Duffy SW, Chen H, Prevost TC and Tabar L **Markov chain models of breast tumour progression and its arrest by screening.** In: *Quantitative Methods for the Evaluation of Cancer Screening* (Edited by: Duffy SW, Hill C, Esteve J) London: Edward Arnold Limited 2001, 42-60
3. van Oortmarssen GJ, Boer R and Habbema JDF **Modelling issues in cancer screening.** *Stat Methods Med Res* 1995, **4**:33-54
4. Stevenson CE **Statistical models for cancer screening.** *Stat Methods Med Res* 1995, **4**:18-32
5. Paci E, Boer R, Zappa M, de Koning HJ, van Oortmarssen GJ, Crocetti E, Giorig D, Rosselli Del Turco M and Habbema JDF **A model-based prediction of the impact on reduction in mortality by a breast cancer screening programme in the city of Florence, Italy.** *Eur J Cancer* 1995, **31A**:348-353
6. Parmigiani G **Decision models in screening for breast cancer.** *Bayesian Statistics* 1999, **6**:525-546
7. Baker SG and Chu KC **Evaluating screening for the early detection and treatment of cancer without using a randomized control group.** *Journal of the American Statistical Association* 1990, **85**:321-327
8. Baker SG **Evaluating the age to begin periodic breast cancer screening using data from a few regularly scheduled screens.** *Biometrics* 1998, **54**:1569-1578
9. Baker SG **Evaluating periodic cancer screening without a randomized control group: a simplified design and analysis.** In: *Quantitative Methods for the Evaluation of Cancer Screening* (Edited by: Duffy SW, Hill C, Esteve J) London: Edward Arnold Limited 2001, 34-41
10. Robins JM **A new approach to causal inference in mortality studies with sustained exposure periods – application to control of the health worker survivor effect.** *Mathematical Modelling* 1986, **7**:1393-1512
11. Flanders WD and Longini IM **Estimating benefits of screening from observational cohort studies.** *Stat Med* 1990, **9**:969-980
12. Mandel JS, Church TR, Ederer F and Bond JH **Colorectal cancer mortality: effectiveness of biennial screening for fecal occult blood.** *J Natl Cancer Inst* 1999, **91**:434-7
13. Mandel JS, Bond JH, Church TR, Snover DC, Bradely M, Schuman LM and Ederer F **Reducing mortality from colorectal cancer by screening for fecal occult blood.** *N Engl J Med* 1993, **328**:1365-1371
14. Shapiro S, Venet W, Strax P and Venet L **Periodic Screening for Breast Cancer, The Health Insurance Plan Project and Its Sequelae, 1963–1986.** Baltimore, Johns Hopkins University Press 1988.
15. Fontana RS, Sanderson DR, Woolner LB, Taylor WF, Miller WE and Muhm JR **Screening for lung cancer: a critique of the Mayo Lung Project.** *Cancer* 1991, **67(4 Suppl)**:1155-64
16. Marcus PM, Bergstralh EJ, Fagerstrom RM, Williams DE, Fontana R, Talyor WF and Prorok PC **Lung Cancer Mortality in the Mayo Lung Project: Impact of Extended Follow-up.** *J Natl Cancer Inst* 2000, **92**:1308-1316
17. Baker SG, Kramer BS and 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>]
18. Gooley TA, Liesenring W, Crowley J and Storer BE **Estimation of failure probabilities in the presence of competing risks: new representations of old estimators.** *Stat in Med* 1999, **18**:695-706
19. Anderson RN **United States Life Tables, National Vital Statistics Reports, from the Centers for Disease Control and Prevention, 47, Number 2.** 1999,
20. Cox DR and Oakes D **Analysis of Survival Data.** London: Chapman and Hall 1984,
21. Moran M **Periodic Screening Evaluation Under Cohort Effect.** Master's Thesis, School of Public Health, University of South Carolina 1998,

### Pre-publication history

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

<http://www.biomedcentral.com/1471-2288/3/4/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](http://www.biomedcentral.com/info/publishing_adv.asp)

