THE BREAST CANCER SCREENING CONTROVERSY: WHEN AND HOW OFTEN

by

Donald Berry
M.D. Anderson Cancer Center
University of Texas

Anirban DasGupta*
Purdue University

and

Giovanni Parmigiani
Duke University

Technical Report #99-05

Department of Statistics
Purdue University
West Lafayette, IN USA

May 1999
Revised August 1999

*

THE BREAST CANCER SCREENING CONTROVERSY: WHEN AND HOW OFTEN

by

Donald Berry
M.D. Anderson Cancer Center
University of Texas

Anirban DasGupta*
Purdue University

and

Giovanni Parmigiani
Duke University

Abstract

This article looks at the current controversy on breast cancer screening by regular mammograms. Using available data on age specific incidence, for carriers as well as non-carriers of a BRCA1 mutation, we do a number of statistical and probabilistic investigations of various screening schedules. A screening schedule would give a starting age and an interval between screenings for mammograms. The analysis includes examination of screening schedules as regards their proneness to producing incorrect results, and as regards their efficiency in early detection. We try to balance the two factors. In particular, we do a markov chain analysis of the sequence of mammograms and work out certain relevant quantities like the mean transition time to the first false positive result and the equilibrium distribution. The article lists a number of screening schedules (pp 17) one of which may be used for a specific individual after using additional information such as family history available for that individual. Some of the analysis is accompanied by a sensitivity study.

* Research supported by a grant from the National Security Agency.

1. INTRODUCTION

Approximately 181,000 new cases of female breast cancer are diagnosed per year (TIME 1999 Almanac, pp. 814). Among white females, the 5 year survival rate is about 88%; it is about 72% among African American women. The debate about advisability, benefits, and risks of regular mammograms and the debate about the proper starting age and screening interval have been continuing; see, in particular, the overviews in Berry (1998), Fletcher (1997), Kopans (1997), and Taubes (1997). The National Cancer Institute (NCI) as well as the American Cancer Society (ACS) now recommend regular mammograms starting at the age of 40, at intervals of 1 to 2 years. This recommendation does not take into account recent discoveries on connections of mutations at the BRCA1 and BRCA2 genes to development of breast cancer. At this time, roughly a third of US women over 50 choose to undergo regular mammograms.

It is explained in the above articles that the entire issue of regular mammogram screenings is quite complicated. There are numerous factors affecting the clinical benefits of regular screening, and in quantifying such benefits to a population as a whole, and to individual women. Some major clinical trials spanning roughly the period 1960 - 1990 have provided differing but still very useful information. Eight such trials, their structures, and summary findings can be seen in Table 1 in Berry (1998). A meta-analysis of the trials gives an approximate 18% reduction in mortality rates for the screening group against the control group. Berry discusses, in elaborate detail, the extreme care needed in interpreting the 18% figure, and in particular the statistically cumbersome issue of if the 18% reduction can be extrapolated to the years after the follow up period(s) of the trials. Another important issue is how to apply the 18% figure to individual women, and even what is the correct population to which an overall figure such as this may be applied. It now seems that information, if available, on mutations of the BRCA1 and BRCA2 genes would be crucial information for individual cases; Berry et al (1997) describe a model and calculations for obtaining such information from family history, without actually performing genetic testing. Also see Couch et al (1997).

From these and other scientific writings, it seems that quantifying clinical benefits of regular mammograms to individuals or a population as a whole remains a difficult problem. Typically, subjects would want to know if regular screenings can prevent death among significant proportions of women, or about the magnitude of increase in life expectancy due to regular screenings. Some figures, for instance, are given in Berry (1998). Depending on how one perceives the continuity of benefits of regular screenings at the end of a follow up period, the average increase in life expectancy due to screening for 15 years seems to be between 1.5 and 5 days per woman. It is not clear if individual cases will be close to such an average figure. Berry compares this to a lottery. See also Kerlikowske (1997).

A central reason that advisability of regular mammograms continues to remain a controversial issue is that there are some well known negative aspects of regular screening. Many of these negative aspects may be grouped together as psychological factors. Primary are the realities of false positives, the stress and anxiety they cause, and the discomfort and pain of a biopsy that is likely to follow a positive mammogram report. There are medical

factors as well. Berry (1998) explains how mammograms may detect harmless lesions or other non-invasive tumors, and the difficulty of knowing which are harmless. There is also the rather widespread public concern that the radiation itself might cause development of tumors, particularly among certain young women. To summarize, the psychological as well as the medical effects of incorrect mammography reports are important reasons that the entire issue remains an intellectually, medically, and ethically challenging one.

Specifically, in spite of the existing recommendations from the NCI and the ACS, the questions of the appropriate starting age of possible regular mammograms and the appropriate screening interval remain interesting questions. This article makes a set of statistical calculations to understand these two questions. The content of the following pages may be summarized as follows :

a. We use existing data, including data separately available for carriers of mutations at the BRCA1 site and noncarriers;

b. Using such data, we fit certain parametric models to the probability of developing breast cancer as a function of age;

c. We also use some available data on the specificity and sensitivity of mammograms;

d. Combining these and the previous data, we provide some calculations on frequency of incorrect and frequency of positive mammogram reports;

e. We also provide some calculations on the probabilities of having a tumor remain undetected for a period of one year or more as a function of the starting age of regular screenings and the screening interval;

f. We attempt to deal with the issues of an appropriate starting age and interval by balancing frequencies of incorrect reports with likelihoods of having undetected tumors;

g. We write a markov chain model to describe the sequence of mammogram reports for an individual, and as part of the analysis of this model, we derive the mean transition time to the first false positive result. We see that the mean transition time is about 20 mammograms. We also derive the equilibrium distribution of this markov chain and we see that the chain is in the false positive state 5% of the times. A sensitivity analysis shows these numbers to remain quite unaffected by subjective inputs.

h. Based on these analyses, we make some recommendations about screening schedules, one of which may be considered for a specific individual by using additional information such as family history available for that individual.

Many of the findings are reported in the form of tables and plots.

The tools required for these calculations we have provided are certain results from theory of runs and patterns, standard combinatorics, and some markov chain theory. Almost all of these are available in Feller (1973), Blom et al (1994), and Norris (1997).

2.1. NOTATION

In the sequel, 't' will denote age of a generic subject, and let A, B, and D stand for:

$$\begin{cases} A = & \text{A tumor exists (at a given time } t) \\ B = & \text{Test sample does not include any abnormal mass} \\ D = & \text{Test result reports presence of abnormal mass;} \end{cases}$$

the dependence of A, B, D on the specific time t will be suppressed in the notation.

We will also use the additional notation:

$$\left.\begin{array}{ll} p = & p(t) = P(A) \\ \mu = & P(B/A) \\ \pi = & P(D/B) \\ \nu = & P(D^C/B^C); \end{array}\right\} \tag{1}$$

Of course, D^C means the complement of D. We do in fact assume that μ, π, and ν are constants independent of t. This amounts to assuming that the efficiency of mammograms as a screening device is stable over time. The assumption seems reasonable for the kind of approximate calculations we are doing. It also seems reasonable to make the following assumptions:

$$\left.\begin{array}{l} P(D/B) = P(D/B, A) = P(D/B, A^C) \\ P(D/B^C) = P(D/B^C, A) = P(D/B^C, A^C) \end{array}\right\} \tag{2}$$

This amounts to saying that given $B(B^C), D$ and A are conditionally independent. This also seems reasonable. We shall make assumption (2).

2.2. MODELING

Easton et al (1995) and Berry et al (1997) give estimates for the cumulative probability $F(t)$ of developing breast cancer by age t, separately for carriers of a BRCA1 mutation and noncarriers. As in Berry et al (1997), only the BRCA1 information is used in this article.

From Figure 1 in Berry et al (1997), we note that about 82% of the carriers develop breast cancer by the age of (approximately) 80, and the modal value for age specific incidence is (approximately) 49. For noncarriers, these two figures are approximately 10% and 71. They also note that the allelic proportion of mutations is between .0004 and .002 and use the value .0006. Therefore, the probability of an individual carrying the mutation is between $2 \times .0004 - (.0004)^2$ and $2 \times .002 - (.002)^2$. Berry et al primarily used the value .0012 for this probability; we use the slightly more conservative value .0015.

For carriers as well as the noncarriers, we use a Type II Beta to model the function $p(t)$. Using $i = 1$ for the carriers and $i = 2$ for the noncarriers, we use

$$\begin{aligned} p_1(t) =& \lambda_1 (t - a_1)^{\alpha_1} \quad (b_1 - t)^{\beta_1}, \\ \text{and } p_2(t) =& \lambda_2 (t - a_2)^{\alpha_2} \quad (b_2 - t)^{\beta_2}. \end{aligned} \tag{3}$$

We will set $a_1 = a_2 = 22$ and $b_1 = b_2 = 80$. This leaves us with 3 parameters for each group. Separately for each group, these 3 parameters were fitted by equating the theoretical mode, the theoretical penultimate cumulative incidence, and the theoretical incidence at the mode to their estimated values. The theoretical values come from (3); the estimated values were taken from Berry et al (1997). Here is the sketch for the carrier group:

$$\left.\begin{array}{ll} \text{Theoretical mode } = \frac{80\alpha_1+22\beta_1}{\alpha_1+\beta_1} & = 49 \\ F_1(80) = \lambda_1(58)^{\alpha_1+\beta_1+1}\ B(1+\alpha_1, 1+\beta_1) & = .82, \\ \text{and } \lambda_1(27)^{\alpha_1}(31)^{\beta_1} & = .024, \end{array}\right\} \tag{4}$$

where $B(\cdot,\cdot)$ denotes the Beta function.

The solutions to (4) are

$$\left.\begin{array}{ll} \lambda_1 & = 1.03944 \times 10^{-6} \\ \alpha_1 & = 1.388 \\ \beta_1 & = 1.594 \end{array}\right\} \text{Carriers} \tag{5}$$

and similarly,

$$\left.\begin{array}{ll} \lambda_2 & = 6.39404 \times 10^{-6} \\ \alpha_2 & = 1.432 \\ \beta_2 & = .263 \end{array}\right\} \text{Noncarriers} \tag{6}$$

For $p = p(t)$, we use the combined (mixture) Type II Beta

$$p(t) = .0015p_1(t) + .9985p_2(t), \tag{7}$$

with the respective parameters as in (5), (6). Figures 1 and 2 describe the fitted Type II Betas for the two groups; they are very close to the figures given in Berry et al (1997).

Fig. 1

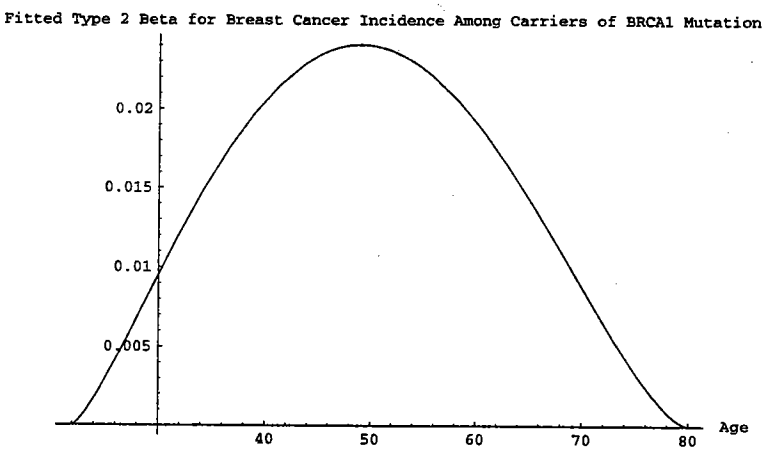
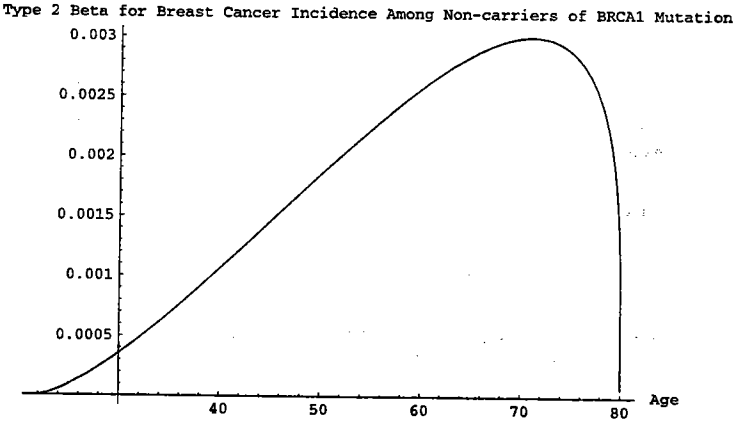


Fig. 2



2.3. SOME BASIC FORMULAS

2.3.1. PROBABILITY OF A POSITIVE REPORT

It will be instructive to compare the actual value of $p(t)$ with the probability of receiving a positive test report at age t. If the test was perfect, they would be equal.

Now, P(A positive test report)

$$
\begin{aligned}
&= P(D) \\
&= P(DBA) + P(DBA^C) + P(DB^CA) + P(DB^CA^C) \\
&= P(D/B,A) \cdot P(B/A) \cdot P(A) + P(D/B,A^C) \cdot P(B/A^C) \cdot P(A^C) \\
&+ P(D/B^C,A) \cdot P(B^C/A) \cdot P(A) + 0 \\
&= \pi\mu p + \pi \cdot 1 \cdot (1-p) + (1-\nu) \cdot (1-\mu) \cdot p;
\end{aligned}
$$

that is, P(Positive test report at age t)

$$= \pi\mu p + \pi(1-p) + (1-\nu)(1-\mu)p. \tag{8}$$

A similar calculation shows that

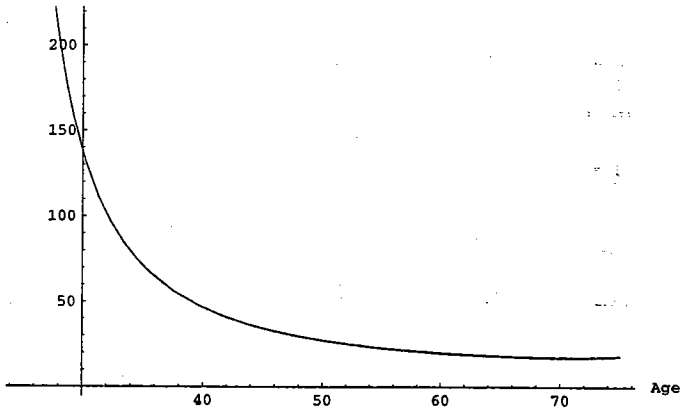
$$
\begin{aligned}
P(D/A^C) &= \pi, \\
\text{and } P(D^C/A) &= \mu(1-\pi) + \nu(1-\mu).
\end{aligned} \tag{9}
$$

An approximate value for π is .05 and an approximate value for $P(D^C/A)$ is .3 (Don Berry, personal communication). Using these, on some algebraic reformatting of (8), one gets

$$
\begin{aligned}
&P(\text{Positive test report at age } t) \\
&= .05 + .65p(t)
\end{aligned} \tag{10}
$$

(10) shows that, generally, positive test reports will be much more frequent than actual incidence of disease in the overall population as $p(t)$ is generally small. Figure 3 gives the relevant likelihood ratio $\frac{P(D)}{p(t)}$; they are in the range of 20 to several hundred. This is relevant in counseling patients after a positive test report.

Fig. 3
Likelihood Ratio of Positive Mammography and Presence of Tumor As Function of Age



2.3.2. PROBABILITY OF AN INCORRECT REPORT

Of more direct concern to an individual and her physician is the correctness of a mammogram report. Incorrectness in either form, i.e., a tumor present not detected or abnormality reported when there is no tumor, is not desirable. We then see

$$
\begin{aligned}
&P(\text{An incorrect test result }) \\
&= P(DA^C) + P(D^C A) \\
&= P(DA^C B) + P(DA^C B^C) + P(D^C AB) + P(D^C AB^C) \\
&= \pi \cdot 1 \cdot (1-p) + 0 + (1-\pi) \cdot \mu \cdot p + \nu \cdot (1-\mu) \cdot p \\
&= \pi + p\{\mu(1-\pi) + \nu(1-\mu) - \pi\}; \qquad (11)
\end{aligned}
$$

and so by using the previous values for $\mu(1-\pi)+\nu(1-\mu) = .3$ (equation (9)) and $\pi = .05$, we see that

$$
\begin{aligned}
&P(\text{An incorrect test result at age } t) \\
&= .05 + .25p(t). \qquad (12)
\end{aligned}
$$

DISCUSSION

Two observations are in order. First, since in the overall population $p(t)$ is numerically quite small, the likelihood of an incorrect result (or a positive result as well; see equation (10)) is pretty much constant over t; second, although it is not apriori obvious, inspection of equations (10) and (12) shows that an incorrect result is always (slightly) less likely than a positive result due to the smaller coefficient .25 in equation (12). This will be the case whenever $P(D^C/A) = \mu(1-\pi) + \nu(1-\mu) \leq \cdot 5$.

In Table 1 we give the values of $p(t)$, the probability of a positive test result, and the

probability of an incorrect test result at some selected ages.

Table 1

Age	P(Presence of tumor)	P(Positive Report)	P(Incorrect Report)
40	.0011	.0507	.0503
45	.0015	.0510	.0504
50	.0019	.0512	.0505
55	.0023	.0515	.0506
60	.0026	.0517	.0506
65	.0029	.0519	.0507
70	.0030	.0520	.0508
75	.0029	.0519	.0507

2.4. HOW MANY INCORRECTS IN A LIFETIME

It is well known that many women receive one or more false positive reports from their mammograms during their lifetime. It might be useful to study more generally the number of incorrect reports in a lifetime. Evidently this depends on the duration of the screening period and the interval between screenings. Below and in the rest of this article, we will denote

$$\left.\begin{array}{ll} \theta & = \text{starting age of screening} \\ b & = \text{age at end of screening} \\ r & = \text{(constant) screening interval ;} \end{array}\right\} \tag{13}$$

in practice, r is not a constant for a given individual but taking it to be a constant will let us avoid more arbitrary assumptions and seems reasonable.

For a specific individual, there would be some amount of dependence between the test results. The dependence seems likely to be not too much. So the total number of incorrect test results seems to be the sum of some weakly dependent Bernoulli variables and we may reasonably conclude therefore (see Barbour et al (1992)) that it will be approximately Poisson distributed. If the mammograms are conducted at ages $\theta, \theta + r, \theta + 2r, \ldots$, up to the age b, then the mean of this Poisson distribution is

$$\lambda(\theta, b, r) = \sum_{i=0}^{[\frac{b-\theta}{r}]} P(\text{An incorrect result at age } \theta + ir); \tag{14}$$

in the above $[x]$ denotes the integer part of a real number x. By using equation (12), one can compute $\lambda(\theta, b, r)$ for any given values of θ, b, r. In Table 2, 3, 4 below, we summarize the Poisson approximations to probabilities of k or more incorrect reports in a lifetime if the starting age of screening is 40, 45, or 50, and if the screening interval is 1, 1.5, or 2 years. In these tables, we have taken b to be 75. In addition, in Figures 4, 5, and 6, we have plotted the expected number of incorrect results separately for the overall population and carriers of a BRCA1 mutation for three starting ages $\theta = 40, 45$, and 50. Interestingly, for each value of θ, the plots for the overall population and carriers are roughly parallel; that is, independent of the screening interval, the carriers of the BRCA1 mutation are

expected to have a constant larger amount of incorrect test results than women in the overall population.

Table 2: Starting age = 40

$P(k$ or more incorrect results)

k	$r = 1$year	$r = 1.5$years	$r = 2$years
1	.838	.703	.598
2	.543	.342	.231
3	.275	.123	.065

Table 3: Starting age = 45
$P(k$ or more incorrect results)

k	$r = 1$year	$r = 1.5$years	$r = 2$years
1	.792	.655	.555
2	.465	.287	.195
3	.209	.092	.049

Table 4: Starting age = 50
$P(k$ or more incorrect results)

k	$r = 1$year	$r = 1.5$years	$r = 2$years
1	.732	.577	.482
2	.379	.213	.142
3	.147	.057	.029

We note that if mammograms are performed every 18 months starting at 40 until the age of 75, a third of women will receive at least two incorrect reports. If the starting age is 50, still one fifth of the women receive at least two incorrect reports. At first glance, this seems to suggest against an early starting age or frequent screenings. But we have to balance low number of incorrect reports with the medical concern of having tumors remain undetected due to a late starting age or infrequent screenings. We study this in a subsequent section. But we continue with the topic of incorrect results in the next section.

Fig. 4

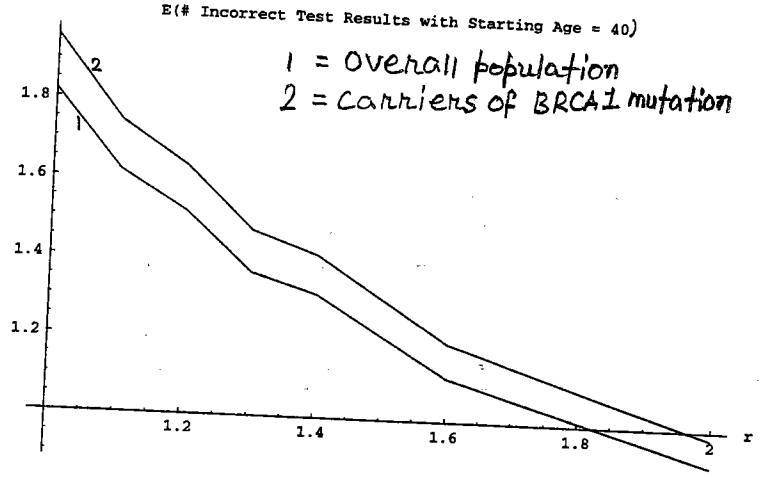


Fig. 5

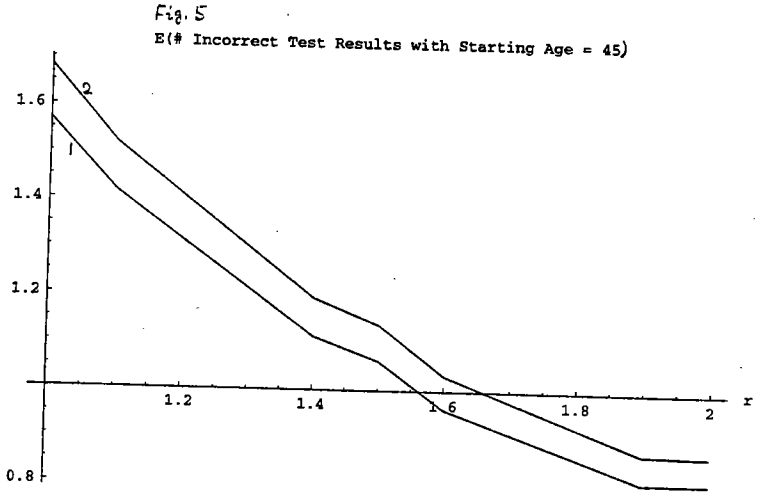
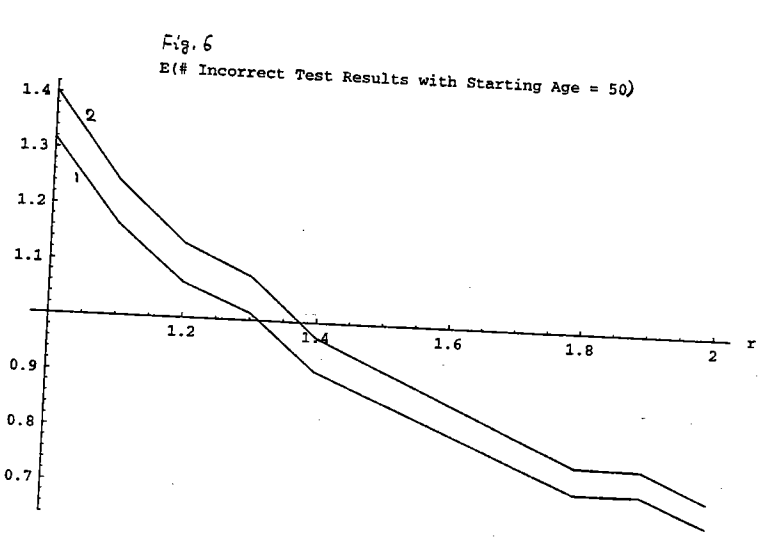


Fig. 6



2.5. COMMUNICATION AND COUNSELING WITH WOMEN

For purposes of advising women, it seems to us that it would be helpful to have some combinations of θ and r for which the expected number of incorrect results in a lifetime will be 1. The value 1 is a nice round number that most women will probably consider acceptable and rational. In a specific case, one combination may be picked at the beginning depending on the woman's family history and other background. Table 5 below summarizes the value of the screening interval for which the expected number of incorrect results in a lifetime is 1. For convenience, we report the values in number of months. Note that we take b to be 75.

Table 5
Screening intervals at which
E(# incorrect results per lifetime)= 1

Starting age	Screening interval
40	22 months
42	21 months
43	20 months
45	19 months
47	18 months
48	17 months
50	16 months

3. PROBABILITY OF HAVING AN UNDETECTED TUMOR

3.1 HARMLESS VS. HARMFUL TUMORS

Obviously, giving long gaps between successive mammograms or starting the screening at a late age will increase the likelihood of having a tumor remain undetected for a longer period of time, even if self examination were to occur. The medical issues are complex. The undiagnosed tumor may be noninvasive and harmless; it could also be an aggressive tumor. Berry et al (1997) give a clear discussion of this issue, and the discovery of DCIS by mammograms.

But, still, it seems reasonable from ethical as well as medical considerations, that a recommended starting age θ and a recommended screening interval r ought to be such that the likelihood of ever having a tumor remain undetected for a substantial period of time is low. At this time, we do not have enough credible information to treat the likelihood of having harmful tumors remain undetected. It makes sense to try to keep small the likelihood of undetected tumors of any kind.

3.2. THE CALCULATION

We now present a brief derivation of the formula for having a tumor remain undetected for one year or longer if screening starts at an age θ and if the screening interval is r. The goal is to find recommended values for θ and r. Why one year? First, a specific choice is needed to agree on which probability we would like to keep small; one year may be a generally acceptable choice when a specific choice is needed. Second, the derivation presented below may be easily extended to cover the case of a general time period, say δ years or longer, instead of one year or longer.

Below, τ will denote the age at onset and $F(t)$ will denote the cumulative probability (in the overall population) of developing breast cancer by age t.

Now, a tumor can remain undetected for one year or longer if

a. it develops a year or earlier before screening starts;

b. it develops in between two screenings such that the later screening is at least a year away from the time of onset;

c. it develops within one year of a screening but that future screening produces a clean (i.e., a negative) report.

Thus for $r \geq 1$ (we will not consider $r < 1$ as neither the NCI nor the ACS are recommending $r < 1$),

P(A tumor remains undetected for one year or longer)

$$
\begin{aligned}
&= P(\tau \leq \theta - 1) + \sum_{i=0}^{[\frac{b-\theta}{r}]} P(\theta + ir < \tau \leq \theta + (i+1)r - 1) \\
&\quad + \sum_{i=0}^{[\frac{b-\theta}{r}]} P(\theta + ir - 1 \leq \tau \leq \theta + ir, \\
&\quad \text{test at time } \theta + ir \text{ gives a negative report}) \\
&= F(\theta - 1) + \sum_{i=0}^{[\frac{b-\theta}{r}]} \{F(\theta + (i+1)r - 1) - F(\theta + ir)\} \\
&\quad + \sum_{i=0}^{[\frac{b-\theta}{r}]} P \text{ (Test at time } \theta + ir \text{ gives a negative report} \\
&\quad \mid \text{tumor is present at time } \theta + ir) \\
&\quad \cdot \{F(\theta + ir) - F(\theta + ir - 1)\} \\
&= F(\theta - 1) + \sum_{i=0}^{[\frac{b-\theta}{r}]} \{F(\theta + (i+1)r - 1) - F(\theta + ir)\}
\end{aligned}
$$

$$+ \cdot 3 \sum_{i=0}^{[\frac{b-\theta}{r}]} \{F(\theta + ir) - F(\theta + ir - 1)\} \tag{15}$$

(see the paragraph following equation (9) for the value .3).

It turns out that the cumulative probability $F(t)$ from the fitted model (7) is actually a closed form expression and so expression (15) is easily programmed and computable. The expression for $F(t)$ itself is not relevant for our purpose.

We now report the probability of having a tumor undetected for a year or longer for selected cases.

3.3. NUMERICAL VALUES

The notation $\xi(\theta, b, r)$ is used for the probability of having a tumor undetected for one year or longer if screening occurs between the ages θ and b at an interval of r years. As before, we will take b to be 75. Table 6 gives the value of this probability for selected values of θ and r.

Table 6
P(Having a tumor undetected for $\geq$ 1 year)

Starting age	$r = 1$ year	$r = 1.25$ years	$r = 1.5$ years	$r = 2$ years
40	.032	.044	.051	.061
42	.033	.044	.053	.061
43	.034	.045	.053	.064
45	.036	.047	.055	.065
47	.038	.048	.055	.065
50	.042	.052	.057	.066

DISCUSSION

Clearly, everyone would want to have a small probability for having an undetected tumor for a substantial time period. We previously chose the time period to be 1 year or longer. But what is a small probability in this context? No universal agreement would be possible. We suggest that the probability should be 5% or smaller. Certainly this is a subjective choice, and many others may like a more conservative value, such as 1%.

The following is a list of pairs (θ, r) for which the probability of having a tumor undetected for one year or longer is equal to 5%. The pairs are found by solving for r

when the expression in (15) is equal to .05, with a specific value for θ.

Table 7
Starting age and screening intervals at which P(Tumor remaining undetected for $\geq$ 1 year) = .05

Starting age	40	42	43	45	47	48	50
Screening interval (in months)	17	17	16.5	16.5	16	15	14.5

By inspection of Table 7 and Table 5, it seems it is not possible to simultaneously have an expected number of incorrect results equal to 1 and a probability of an undetected tumor equal to .05. The actual expected number of incorrect results at the (θ, r) combinations of Table 7 are as follows:

$(\theta, r)(r$ in months)	$(40, 17)$	$(42, 17)$	$(43, 16.5)$	$(45, 16.5)$	$(47, 16)$	$(48, 15)$	$(50, 14.5)$
$E(\#$ incorrect results)	1.26	1.21	1.21	1.11	1.11	1.11	1.06

RECOMMENDATION: From Table 7 and the above list, one may recommend that mammograms be conducted

Every $16 - 17$ months starting at 45
or Every 16 months starting at 47
or Every $14 - 15$ months starting at 50,

with a specific choice for each woman made depending on her history, background, psychological status, and other information known to her counselor and physician. For carriers of a BRCA1 mutation, this recommendation does not apply.

4. CLOSER EXAMINATION OF CURRENT RECOMMENDATIONS

The NCI and the ACS now both recommend that women start regular mammograms at the age of 40 with a screening interval of 1 to 2 years. We previously saw (see Table 2) that if annual mammogram starts at the age of 40, about 55% of women will receive 2 or more incorrect reports in their lifetime. This seems a bit too much. We now investigate the existing recommendation and some other plausible choices of θ and r in a bit more

detail. The ensuing description is more technical than the previous material, but has an element of mathematical interest. It is also medically relevant.

4.1. UNDESIRABLE RUNS

If two successive mammograms result in incorrect reports, it will be called an incorrect run of length 2. Similarly, one can talk of positive runs, etc. Incorrect runs can be medically and/or emotionally dangerous. It is quite instructive, as we describe below, that incorrect runs will not be infrequent under certain common screening plans. One of our findings is that 8 in 100 women will experience at least one incorrect run in their lifetime if they do annual mammograms starting at 40. This does seem quite surprising and a matter of some concern.

The probability of an incorrect result in a single mammogram was previously derived in equation (12). The dependence of this probability on the exact age t is very very small. Practicaly it would be sensible to work with the simplifying assumption that the probability of an incorrect result in a single mammogram equals the constant value

$$p = \frac{1}{80-22}\int_{22}^{80}(.05 + .25p(t))dt = .0504 \tag{16}$$

As we remarked before, successive mammograms for a specific woman are likely to have some dependence; for now , this dependence will be ignored.

Therefore we are able to use the elegant classical theory of runs in independent Bernoulli trials. If screening starts at age θ, and continues till an age b, with a screening interval equal to r, then a woman will undergo $n = 1 + [\frac{b-\theta}{r}]$ mammograms in total (recall that [ ] denotes the integer part of a number).

If θ_n denotes the probability that a woman will receive at least one run of consecutive incorrect reports in her lifetime, then from the classical theory of runs,

$$\theta_n = u_2 + u_3 + \ldots + u_n, \tag{17}$$

where the numbers u_i can be found from the recurrence relation

$$u_2 = p^2, u_3 = qp^2, u_i = qu_{i-1} + pqu_{i-2}, \tag{18}$$

where, of course, $q = 1 - p = .9496$. See section 14.14 in Blom et al (1994).

Algorithm (17) is not computationally efficient for large n (e.g., if $\theta = 40, b = 75$, and $r = 1$, then $n = 36$, and (17) becomes an inefficient algorithm). Feller (1973, chapter 13) shows an accurate and rapid approximation to θ_n:

$$\theta_n \approx 1 - \frac{1-px}{qx^{n+1}(3-2x)}, \tag{19}$$

where x is the unique positive root of the cubic $qp^2s^3 - s + 1 = 0$ (in smaller print, Feller shows that $1 < x < \frac{1}{p}$).

With $p = .0504$, this unique root $x = 1.00243$. By (19), we are immediately able to find the probability of at least one run of consecutive incorrect reports for any given screening plan (θ, b, r). We summarize these probabilities in Table 8; again we take b to be 75.

Table 8
P(Receiving consecutive incorrect reports in a lifetime)

Starting age	$r = 1$ year	$r = 1.25$ years	$r = 1.5$ years	$r = 2$ years
40	.08	.06	.05	.04
45	.07	.05	.045	.03
50	.06	.045	.04	.03

DISCUSSION

It is quite serious and surprising that 8 in 100 women can be expected to receive consecutive incorrect results in their lifetime with annual mammograms starting at 40. If annual mammogram starts at 45, the percentage drops only marginally to 7%. A strict schedule of annual mammograms starting at 40 seems to cause substantial negative effects, perhaps mostly psychological. It is likely to be mostly psychological because an incorrect report is likely to be a false positive report. Assuming that a positive report causes a biopsy, there may be other consequences as well, for example, rise of insurance costs.

To make the description complete, below we list the probability of consecutive incorrect reports in a lifetime for the recommended combinations at the end of section 3.4.

Table 9
P(Consecutive incorrect reports in a lifetime)

(Starting age, Screening interval)	
(45, 16.5 months)	.047
(47, 16 months)	.047
(50, 14.5 months)	.045

REINFORCEMENT: So the three combinations (45, 16-17 months), (47, 16 months), and (50, 14.5 months) have the following characteristics: for each of these schedules,

$$\begin{cases} P(\text{Having a tumor remain undetected for} \geq 1 \text{ year}) = .05 \\ E(\#\text{ incorrect reports in a lifetime}) \approx 1 \\ P(\text{Ever receiving consecutive incorrect reports}) \approx .045 < .05 \end{cases}$$

These characteristics are easy to communicate and look reasonable. For a woman expected to be a noncarrier of a BRCA1 mutation, a schedule close to one of the above three may be recommended.

5. FALSE POSITIVES AND MARKOV CHAINS

5.1. PROS, CONS AND GOAL

As is well known, a large proportion of women having mammograms receive at least one false positive report in their lifetime. We will now write and analyze a Markov chain model for the sequence of test reports for a given woman.

If we let X_n denote the states of the individual at time n, then X_n is in one of the following four states:

$$\left.\begin{array}{l} \text{State 1} = \text{True positive } (DA) \\ \text{State 2} = \text{False negative } (D^cA) \\ \text{State 3} = \text{False positive } (DA^c) \\ \text{State 4} = \text{True negative } (D^CA^C) \end{array}\right\} \tag{20}$$

We take X_n to be a homogeneous Markov chain. Although this model, clearly, is subject to valid criticisms, it has at least two positive features: it is a reasonable way to break loose from an iid model, and it allows us to use the established mathematical machinery of homogeneous Markov chains to answer a number of important questions.

The questions investigated are the following:

Question 1. How long does it take for the first false positive report to occur depending on what the very first report was for the given individual?

Question 2. In the long run, about what percentage of the mammogram reports for a given individual are false positives?

5.2. WRITING THE TRANSITION MATRIX

5.2.1. NOTATION

Evidently, one has to apply care and use specific information about the disease in writing the transition matrix. Some sensitivity analysis on the values of the entries of the transition matrix is also necessary and is in fact part of our analysis. This care is necessary because the transition matrix will drive the analysis in a Markov chain model.

First some notation needs to be fixed. In discussing the transitions of our homogeneous Markov chain, the subscript 1 will be used for the earlier time and 2 for the latter time. Thus, for example, $P(D_2A_2|D_1A_1)$ will mean the (1,1) diagonal element of the transition matrix.

We will denote the transition matrix by P, and p_{ij} will mean the probability of transition to state j from state i in one step $(i, j = 1, 2, 3, 4)$.

5.2.2. KEY QUANTITIES GOVERNING THE TRANSITION MATRIX

Some particular quantities arise in expressions for the entries p_{ij} of the transition matrix. Let us briefly discuss the clinical background of how these quantities will arise in the Markov chain analysis.

The key quantities and some clinical background are now presented. Some of these are already familiar from the previous analysis in sections 1 through 4. The quantities are:

$$\left.\begin{aligned} \pi &= P(D_1|A_1^c) = P(D_2|A_2^c) \\ \eta &= P(D_1|A_1) = P(D_2|A_2) \\ \varepsilon &= P(A_2^c|A_1D_1) \\ \delta &= P(D_2^c|A_2) - P(D_2^c|A_2D_1^cA_1) \end{aligned}\right\} \tag{21}$$

Of these, we encountered π and η before and used $\pi = .05$ and $\eta = .7$ (see the paragraph after equation (9)). ε and δ are fresh quantities and use of clinical information would be necessary.

Regarding ε, if at the previous mammogram, the subject had breast cancer and the mammogram showed it, an array of consequences is possible. The physician may choose to simply watch the status if the tumor seems harmless; if treatment starts, there would be a variety of possible treatments depending on the type of the cancer. If, say, we are thinking of the next mammogram in a year or two, and if the treatment is with tamoxifen, then for such patients ε may be taken as 0. On the other hand, if surgery or chemotherapy for a number of months is the treatment, then the probability of a local recurrence in a year or two may be 2 to 5%, i.e., for such patients ε may be close to 1. We want in our Markov chain model an overall single ε. It seems the best option is to keep ε as it is for the calculations, leading to a sensitivity analysis in the numerical part with a number of values of ε. The same logic would apply to the other quantity δ as well. Note that $P(D_2^c|A_2D_1^cA_1)$ is expected to be smaller than $P(D_2^c|A_2)$ $(= 1 - \eta)$ because if the previous mammogram failed to detect an existing tumor, then it should be less likely that in the

next mammogram it would be undetected still. For instance, a 50% reduction in this likelihood would imply $\delta = \frac{3}{2} = .15$.

5.3. SKETCH OF DERIVATION OF THE TRANSITION MATRIX

In equation (7), a mixture type II Beta formula was given for $P_t(A)$, the probability of a tumor being present at a specific age t. The magnitude of $P_t(A)$ for varying t is small. For the homogeneous Markov chain model, we will use the average value

$$\tilde{p} = \frac{1}{80-22}\int_{22}^{80} p(t)dt = .0017 \tag{22}$$

for $P_t(A)$ at any t.

One expression that would be used in deriving the transition matrix is the formula

$$P(A_2|A_1^c) = \frac{\tilde{p}\varepsilon\eta}{1-\tilde{p}}. \tag{23}$$

To see (23), write

$$\begin{aligned} & P(A_2^c A_1^c) \\ =& 1 - P(A_1 \cup A_2) \\ =& 1 - 2\tilde{p} + \tilde{p}(1-\eta) + (1-\varepsilon)\tilde{p}\eta \\ =& 1 - \tilde{p}(1+\varepsilon\eta), \end{aligned} \tag{24}$$

and thus $P(A_2 A_1^c) = \tilde{p}\varepsilon\eta$, giving (23).

So, for instance, the element p_{31} in the transition matrix P is

$$\begin{aligned} & p_{31} \\ =& P(D_2 A_2 | D_1 A_1^c) \\ =& P(D_2|A_2 D_1 A_1^c) \cdot P(A_2|D_1 A_1^c) \\ =& \eta \cdot \frac{\tilde{p}\varepsilon\eta}{1-\tilde{p}} \\ =& \frac{\tilde{p}\varepsilon\eta^2}{1-\tilde{p}} \end{aligned} \tag{25}$$

if we assume

$$\begin{aligned} P(D_2|A_2 D_1 A_1^c) &= P(D_2|A_2), \text{ and} \\ P(A_2|D_1 A_1^c) &= P(A_2|A_1^c). \end{aligned} \tag{26}$$

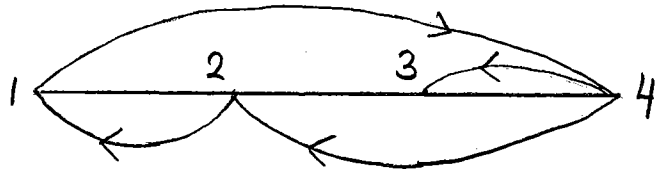
For example, the first assumption says that a false positive in the previous mammogram was probably followed up and discarded and so should not affect the result of the next mammogram. This seems reasonable. To maintain the continuity of the discussion, we

present the transition matrix now. A full set of assumptions like (26) above that are used in writing the transition matrix are listed in an appendix.

The transition matrix is

$$P = \begin{bmatrix} 1-\varepsilon & 0 & 0 & \varepsilon \\ \eta+\delta & 1-\eta-\delta & 0 & 0 \\ \frac{\tilde{p}\varepsilon\eta^2}{1-\tilde{p}} & \frac{\tilde{p}\varepsilon\eta(1-\eta)}{1-\tilde{p}} & \pi(1-\frac{\tilde{p}\varepsilon\eta}{1-\tilde{p}}) & (1-\pi)(1-\frac{\tilde{p}\varepsilon\eta}{1-\tilde{p}}) \\ \frac{\tilde{p}\varepsilon\eta^2}{1-\tilde{p}} & \frac{\tilde{p}\varepsilon\eta(1-\eta)}{1-\tilde{p}} & \pi(1-\frac{\tilde{p}\varepsilon\eta}{1-\tilde{p}}) & (1-\pi)(1-\frac{\tilde{p}\varepsilon\eta}{1-\tilde{p}}) \end{bmatrix} \tag{27}$$

Note that according to this P, it is not possible, for example to have a false positive followed by a true positive in the preceding mammogram. The following diagram shows that in spite of some zero entries in the transition matrix, the Markov chain is irreducible:



The chain is clearly also aperiodic. *Hence, our Markov chain is ergodic.* The consequence of this is described in Section 5.5.

5.4. TRANSITION TIME TO THE FIRST FALSE POSITIVE

It would be useful for health practitioners and women to know something about the number of mammograms after which the first false positive occurs. In Markov chain terminology, this is the transition time to $G_1 = \{DA^c\}$ starting from one of the states in $G_2 = \{DA, D^cA, D^cA^c\}$. We present below the three mean transition times:

$$\left.\begin{aligned} E(N_1) &= E(\text{Transition time to } DA^c \text{ starting at } DA), \\ E(N_2) &= E(\text{Transition time to } DA^c \text{ starting at } D^cA), \\ E(N_4) &= E(\text{Transition time to } DA^c \text{ starting at } D^cA^c). \end{aligned}\right\} \tag{28}$$

$E(N_4)$ is probably the most useful of the three. To evaluate these mean transition times, we partition P as

$$P = \begin{bmatrix} H & S \\ R & Q \end{bmatrix},$$

where H is the scalar p_{33}, $Q_{3\times3}$ is the matrix of transition probabilities between the states DA, D^cA, and D^cA^c of G_2, $S_{1\times3}$ is the vector of transition probabilities from DA^c to the states in G_2, and $R_{3\times1}$ is the vector of transition probabilities from the states in G_2 to DA^c. H, S, R, and Q are directly available from the full transition matrix P (equation

(27)). In particular

$$Q = \begin{bmatrix} 1-\varepsilon & 0 & \varepsilon \\ \eta+\delta & 1-\eta-\delta & 0 \\ \frac{\tilde{p}\varepsilon\eta^2}{1-\tilde{p}} & \frac{\tilde{p}\varepsilon\eta(1-\eta)}{1-\tilde{p}} & (1-\pi)(1-\frac{\tilde{p}\varepsilon\eta}{1-\tilde{p}}) \end{bmatrix} \tag{29}$$

The mean transition times will then be

$$\left.\begin{aligned} E(N_1) &= \text{First row sum in } (I-Q)^{-1} \\ E(N_2) &= \text{Second row sum in } (I-Q)^{-1} \\ E(N_4) &= \text{Third row sum in } (I-Q)^{-1} \end{aligned}\right\} \tag{30}$$

where I is the identity matrix of order 3 (see Blom et al (1994)). The inverse matrix $(I-Q)^{-1}$ is computable in closed form; this symbolic computation was performed on *Mathematica* and one has the following expressions for the mean transition times of equation (30):

Mean transition time to a false positive starting at a true negative

$$= E(N_4)$$
$$= \frac{.70083 + .00036\varepsilon + 1.00119\delta}{.035 - .00004\varepsilon + .05\delta - .00006\varepsilon\delta}; \tag{31}$$

Mean transition time to a false positive starting at a true positive

$$= E(N_1)$$
$$= \frac{.035 + .00036\varepsilon^2 + .05\delta + \varepsilon(.70079 + 1.00113\delta)}{\varepsilon(.035 - .00004\varepsilon + .05\delta - .00006\varepsilon\delta)}; \tag{32}$$

Mean transition time to a false positive starting at a false negative

$$= E(N_2)$$
$$= \frac{.035 + .00030\varepsilon^2 + .05\delta + \varepsilon(.75079 + 1.00113\delta)}{\varepsilon(.035 - .00004\varepsilon + .05\delta - .00006\varepsilon\delta)}. \tag{33}$$

As we remarked before, $E(N_4)$ might be the most useful to health practitioners and the public at large. Examination of formula (31) immediately reveals the following:

$$E(N_4) \approx \frac{.70 + \delta}{.035 + .05\delta} = 20,$$

essentially for any choices of ε and δ. The conclusion is that

> On an average, starting at a true negative, the first false positive will occur at the 21st mammogram. This is for the overall population without using any information about family history or genetic testing.

Table 10 gives values of $E(N_i)$ for selected values of ε and δ.

Table 10

	$\varepsilon = .1$			$\varepsilon = .2$			$\varepsilon = .5$		
δ	$E(N_1)$	$E(N_2)$	$E(N_4)$	$E(N_1)$	$E(N_2)$	$E(N_4)$	$E(N_1)$	$E(N_2)$	$E(N_4)$
0	30.03	31.46	20.03	25.03	26.46	20.03	22.04	23.47	20.04
.05	30.03	31.36	20.03	25.03	26.36	20.03	22.04	23.37	20.04
.15	30.03	31.20	20.03	25.03	26.21	20.03	22.04	23.22	20.04

So, for example, if annual mammograms to age 75 are done starting at age 40, then each of $E(N_1)$, $E(N_2)$, and $E(N_4)$ will be quite a bit less than 36, the total number of mammograms performed. A false positive would be likely. It would be helpful to formally see this by working out the equilibrium distribution of our markov chain. This is done next.

5.5. THE EQUILIBRIUM DISTRIBUTION

Due to the ergodic nature of our Markov chain, the equilibrium distribution will exist and be a left eigenvector of the transition matrix P (see Norris (1997)). The exact numerical values in the equilibrium distribution will depend on the exact choice of ε and δ. We choose $\varepsilon = .1, .2$, and $.5$ and $\delta = 0, .05$, and $.15$ in order to conduct a sensitivity analysis. Remarkably, when rounded to 3 places after the decimal, the equilibrium distribution is the same for each of these values of ε and δ and it is:

$$\text{Equilibrium distribution} = (.001, .000, .050, .949)$$

In other words, under our Markov chain model, 94.9% in a sufficiently long sequence of mammograms are true negative (D^cA^c) and 5% are false positive (DA^c). These figures, interestingly, are not much affected by the choice of the quantities ε and δ of equation (21). The 5% false positive rate in the equilibrium distribution is consistent with the conditional probability $\pi = P(D|A^c)$ which we took to be .05. But the false negative rate in the equilibrium distribution (zero to three decimals) is very different from the conditional probability $1 - \eta = P(D^c|A)$ which we took to be .3.

6. SUMMARY

The controversy about breast cancer screening has been a persistent one in the medical community. The clinical benefits as measured by reduction in mortality rate and/or increase in life expectancy due to regular screenings have been the subject of

some international clinical trials. The National Cancer Institute and the American Cancer Society both recommend, at the present time, regular mammograms starting at 40 with an interval of one to two years. Still, the controversy and even more, the debate on what to recommend to a specific individual, continue.

We used data available for carriers of a BRCA1 mutation and the noncarriers to fit a model for age specific incidence for the overall population. The model was a mixture of Type II Betas. We also used some available information on the error rates of mammograms. Using all of these information, we examined some common screening schedules, for example, a schedule of annual mammograms starting at 40 or 50, with respect to some relevant factors: how susceptible they are to giving incorrect reports, runs of incorrect reports, or leaving tumors undetected for a period of one year or longer.

We also wrote and analyzed a markov chain model for the sequence of mammogram reports vis-a-vis the cancer status of an individual. As part of this analysis, we derived the mean transition time to a false positive result, and derived the equilibrium distribution. A sensitivity analysis indicated that these are not much affected by inputs required for the transition matrix of the markov chain.

Based on these calculations, we recommended a number of specific screening schedules one of which may be followed approximately depending on additional information available for a particular individual. We also saw some evidence that a schedule of annual mammograms starting at 40 may have some undesirable properties.

7. APPENDIX

We list the assumptions that were made for writing the transition matrix P in equation (27).

ASSUMPTIONS

a $P(A_2^c|D_1^c A_1) = 0$

b $P(A_2|D_1 A_1^c) = P(A_2|D_1^c A_1^c) = P(A_2|A_1^c)$

c $P(D_2|D_1 A_1 A_2) = 1$

d $P(D_2|D_1 A_1 A_2^c) = 0$

e $P(D_2|D_1 A_1^c A_2) = P(D_2|D_1^c A_1^c A_2) = P(D_2|A_2)$

f $P(D_2|D_1 A_1^c A_2^c) = P(D_2|D_1^c A_1^c A_2^c) = P(D_2|A_2^c)$

REFERENCES

[1] Time Almanac(1999), Information Please LLC, Boston.

[2] Barbour A. et al (1992), Poisson Approximation, Clarendon Press, Oxford.

[3] Berry, Donald A. et al (1997), Probability of carrying a mutation of breast–ovarian cancer gene BRCA1 based on family history, Jour. Nat. Cancer Inst., **89**, 227–238.

[4] Berry, Donald A.(1998), Benefits and risks of screening mammography for women in their forties: A statistical appraisal, Preprint.

[5] Blom, G et al (1994), Problems and snapshots from the World of Probability, Springer–Verlag, New York.

[6] Couch, F.J. et al (1997), BRCA–1 mutations in women attending clinics that evaluate the risk of breast cancer. New England Journal of Medicine, **336**, 1409–1415.

[7] Easton D.F. et al (1995), Breast and ovarian cancer incidence in BRCA1 mutation carriers, Am. Jour. Human Genet., **56**, 265–271.

[8] Feller, William (1973), An Introduction to Probability Theory and Its Applications, Vol. 1, Wiley, New York.

[9] Fletcher, S.(1997), Breast cancer screening among women in their forties: an overview of the issues, Monogr. Nat. Cancer Inst., **22**, 5–9.

[10] Kerlikowske, K.(1997), Efficacy of screening mammography among women aged 40-49 years and 50-69 years : Comparison of relative and absolute benefit, Monogr. Nat. Cancer Inst., **22**, 79–86.

[11] Kopans, D.(1997), An overview of the breast cancer screening controversy, Monogr. Nat. Cancer Inst., **22**, 1–3.

[12] Norris, J.R. (1997), Markov Chains, Cambridge University Press, UK.

[13] Taubes, G.(1997), The breast–screening brawl, Science, **275**, 1056–1059.

Fig. 1

Fitted Type 2 Beta for Breast Cancer Incidence Among Carriers of BRCA1 Mutation

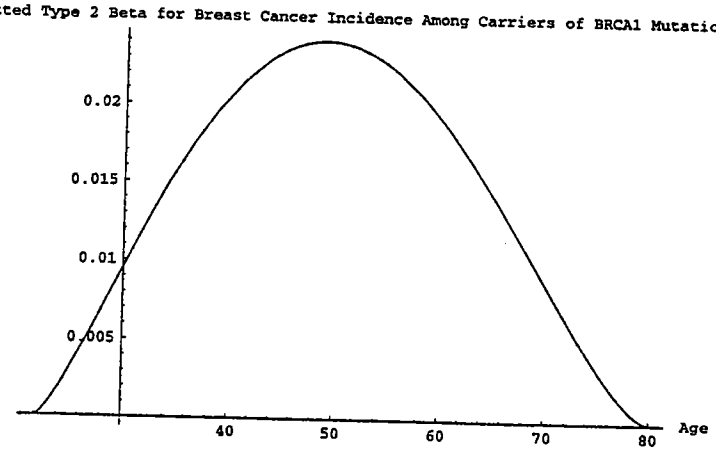


Fig. 2

Fitted Type 2 Beta for Breast Cancer Incidence Among Non-carriers of BRCA1 Mutation

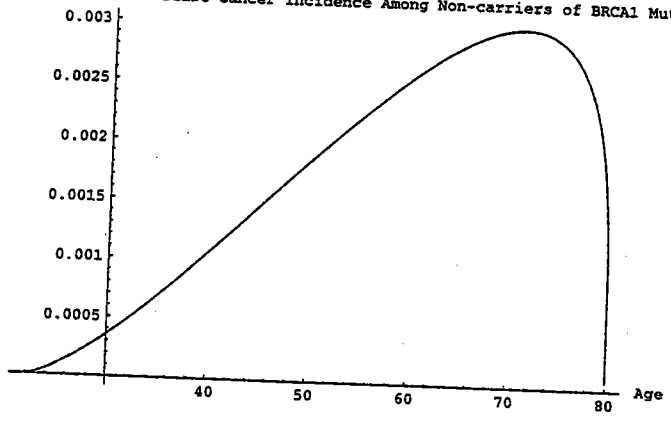


Fig. 3

Likelihood Ratio of Positive Mammography and Presence of Tumor As Function of Age

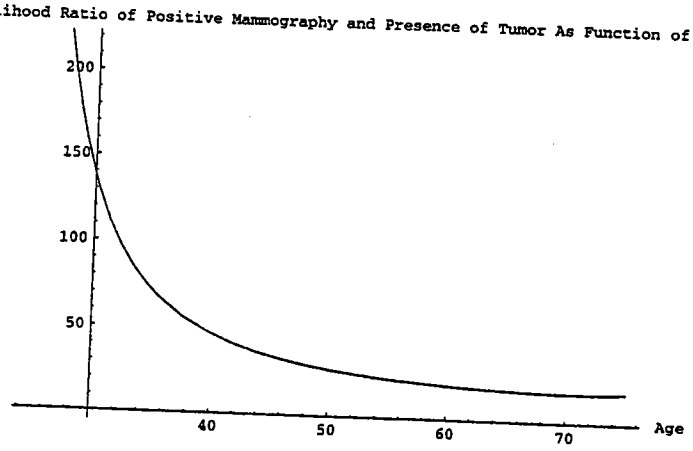


Fig. 4

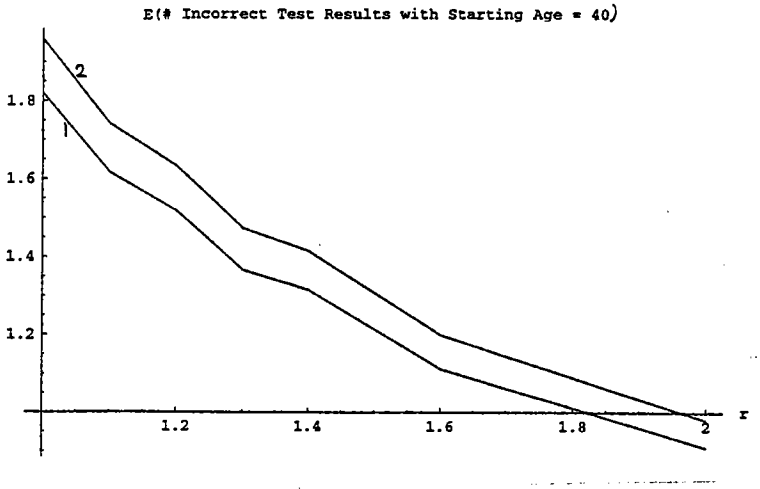


Fig. 5

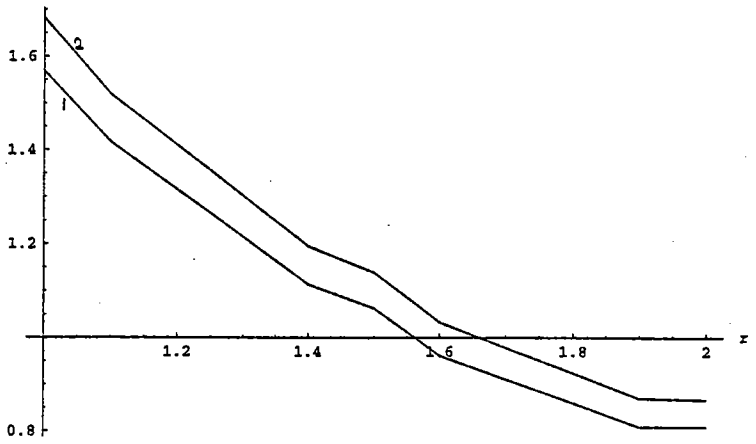


Fig. 6