Markov Models of Breast Tumor Progression: Some Age-Specific Results

Stephen W. Duffy, Nicholas E. Day, László Tabár, Hsiu-Hsi Chen, Teresa C. Smith*

Researchers have noted that mammographic screening has a reduced effect on breast cancer mortality in women in their forties compared to older women. Explanations for this include poorer sensitivity in younger women due to denser breast tissue, as well as more rapid tumor progression, giving a shorter mean sojourn time (the average duration of the preclinical screen-detectable period). To test these hypotheses, we developed a series of Markov-chain models to estimate tumor progression rates and sensitivity. Parameters were estimated using tumor data from the Swedish two-county trial of mammographic screening for breast cancer. The mean sojourn time was shorter in women aged 40–49 compared to women aged 50–59 and 60–69 (2.44, 3.70, and 4.17 years, respectively). Sensitivity was lower in the 40–49 age group compared to the two older groups (83%, 100%, and 100%, respectively). Thus, both rapid progression and poorer sensitivity are associated with the 40–49 age group. We also modeled tumor size, node status, and malignancy grade together with subsequent breast cancer mortality and found that, to achieve a reduction in mortality commensurate with that in women over 50, the interscreening interval for women in their forties should be less than two years. We conclude that Markov models and the use of tumor size, node status, and malignancy grade as surrogates for mortality can be useful in design and analysis of future studies of breast cancer screening. [Monogr Natl Cancer Inst 1997;22:93–97]

In assessing the early detection of a disease through screening, a first model is often the following:

1. Every subject begins with no detectable disease at all. Some subjects will develop the disease of interest, some will remain free of the disease all their lives.
2. For a subject who develops disease, at a certain time \( t_1 \), the person will pass to a state in which the disease is asymptomatic but can be detected by a screening test. This phase is often called the preclinical detectable period (PCDP).
3. For this subject, at a certain time \( t_2 (t_2 > t_1) \), the disease will become clinically symptomatic. In the absence of screening, this is defined as the time of diagnosis (although in practice there may be a delay from symptoms to diagnosis). The period \( t_2 - t_1 \) is known as the sojourn time.

Screening might take place as part of an immunization program, to prevent the spread of a communicable disease or to identify cases in time to effectively treat them. Here we will concentrate on the last purpose, and the specific area we will focus on is breast cancer screening with mammography. For screening to be effective in this context, disease needs to be diagnosed some time before \( t_2 \), while it is still treatable with less aggressive methods and while it is curable in the long term. This means a substantial lead time and good sensitivity are required. Lead time = \( t_2 - t_3 \), where \( t_2 \) is time of clinical diagnosis as above and \( t_3 \) is actual time of detection by screening. Sensitivity is the probability that a case of preclinical, detectable disease is actually diagnosed by the screening test. The sensitivity and the average length of the preclinical detectable period (= mean sojourn time [MST]) are therefore crucial parameters in assessing the ability of screening to affect subsequent mortality.

Note that the sojourn time is an upper limit on the lead time achievable, but if sojourn time is assumed to be exponentially distributed, the expected lead time of a screen-diagnosed cancer is equal to the mean sojourn time. The seminal papers on this subject are by Zelen and Feinlieb (1), Prorok (2), and Day and Walter (3). Usually, in the modeling of tumor progression and its arrest by early detection, estimation has to be heavily supported by assumptions, constraints, and analytic strategies one would prefer to avoid. These include estimation in several stages—for example, the underlying preclinical incidence may be estimated as the clinical incidence in an unscreened population and the progression rate to clinical disease estimated thereafter (3), or the rate of progression to clinical disease may be estimated first assuming a 100% sensitivity of the early-detection tool and the sensitivity thereafter estimated with the progression rate assumed constant at the estimated value (4). It has also often been necessary to make sweeping assumptions about sensitivity (5). Loss of information due to blocking of time into discrete years or screening rounds is also common (3,6). One well-designed method employed in the past is that of Day and Walter (3), which estimates the sensitivity and progression rate simultaneously but which requires a prior estimate of the underlying disease incidence to be specified as fixed beforehand.

It seems intuitively desirable to develop a comprehensive model that would simultaneously estimate all parameters, if a data set of adequate design could be found. In particular, it would be desirable to estimate the preclinical incidence from the

*Affiliations of authors: S. W. Duffy, N. E. Day, T. C. Smith, MRC Biostatistics Unit, Institute of Public Health, Cambridge, United Kingdom; L. Tabár, Director, Mammography Department, Central Hospital, Falun, Sweden; H. H. Chen, Taiwan National University, Taipei, Taiwan.

Correspondence to: Stephen W. Duffy, M.D., MRC Biostatistics Unit, Institute of Public Health, Robinson Way, Cambridge CB2 2SR, UK.

© Oxford University Press
same data set used to estimate the progression from the preclinical state. Here we demonstrate the use of Markov-chain models to estimate the progression rates from empirical screening data and point up some applications in assessing the likely age-specific effect of screening on future mortality from breast cancer.

Data and Methods

We used the data from the Swedish two-county trial of mammographic screening for breast cancer (7); 77,080 women aged 40–74 were randomized to invitation to screening (Active Study Population [ASP]), and 55,985 to no invitation (Passive Study Population [PSP]), for seven to eight years. The PSP was given a single screen at the last screen of the ASP. We shall concentrate on women aged 40–69 at randomization, as screening was abandoned after the second screen in women aged 70–74 due to poor attendance rates. The cancers diagnosed in the trial are shown by detection mode in Table 1.

Progression of the disease was modeled as a Markov chain (8). In this model, individuals occupy states for random, exponentially distributed periods of time and move from state to state independently of each other. The major assumption of this model is that if we know the state at time t for a given individual, knowledge of that individual’s states at times prior to t is of no additional benefit in assessing the individual’s likely future progression.

A simple example is a three-state model where states 0, 1, and 2 represent no detectable disease, preclinical screen-detectable disease, and clinical symptomatic disease, respectively. Associated with such a Markov model is a transition matrix of instantaneous probabilities of moving from state to state. For the above three-state model we posit the following transition matrix:

$$
\begin{pmatrix}
-\lambda_1 & \lambda_1 & 0 \\
0 & -\lambda_2 & \lambda_2 \\
0 & 0 & 0
\end{pmatrix}
$$

Here $\lambda_1$ denotes the birth rate into the PCDP and $\lambda_2$ the transition rate from preclinical to clinical disease. We assume spontaneous regression to be impossible. We also assume that to reach the clinical phase, a tumor must pass through the preclinical phase. A property of this model is that 1/$\lambda_2$ is the MST (9). The instantaneous transition rates need to be converted into probabilities of transition in noninstantaneous periods of time, by solving a potentially complex set of algebraic equations known as Kolmogorov equations (8). In this simple model, the solution can be derived by hand, giving the formula for probabilities of transition in a non-negligible time $t$ as:

$$
\begin{bmatrix}
e^{-\lambda_1 t} & \frac{\lambda_1 (e^{-\lambda_2 t} - e^{-\lambda_1 t})}{(\lambda_1 - \lambda_2)} & 1 - e^{-\lambda_1 t} - \frac{\lambda_1 (e^{-\lambda_2 t} - e^{-\lambda_1 t})}{(\lambda_1 - \lambda_2)} \\
0 & e^{-\lambda_2 t} & 1 - e^{-\lambda_2 t} \\
0 & 0 & 1
\end{bmatrix}
$$

The formulas will become further complicated by the fact that those with a previous history of clinical breast cancer were excluded, so we have to condition on this at the first screen of the ASP and at entry to the trial of the PSP. Also, inclusion of false-positive and false-negative screening error probabilities render the probabilities very complex indeed.

Also, introducing more states—for example, “node negative” and “node positive” within each of the preclinical and clinical phases—brings about considerable increases in algebraic complexity. In the latter case, hand solution of the Kolmogorov equations is not feasible, so the following strategy was implemented:

1. We used the computer program Mathematica to solve the Kolmogorov equations to produce transition probabilities from the transition rates (10).
2. For each type of transition observed, we used (1) and the error probabilities to calculate the expected number of transitions.
3. We then solved, as a nonlinear regression, the equation: observed transitions = expected transitions + error.

This means we did not actually maximize the likelihood but instead estimated the transition rates and error probabilities as a solution of a complex set of generalized estimating equations. For further details of algebra and statistical methods, see Duffy et al. (9) and Chen et al. (11,12).

Results

Three-State Model

Table 2 shows the results for a model of progression among three states, as described above: no detectable disease, preclinical...

Table 1. Cancers in the two-county study by detection mode (%) and age

<table>
<thead>
<tr>
<th>Detection mode</th>
<th>Age</th>
<th>ASP prior*</th>
<th>ASP screen 1</th>
<th>ASP screen 2+</th>
<th>ASP interval</th>
<th>ASP refuser</th>
<th>Total ASP</th>
<th>PSP pre-screen</th>
<th>PSP screen</th>
<th>Total PSP</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>40–49</td>
<td>50–59</td>
<td>60–69</td>
<td>70–74</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ASP prior*</td>
<td>6 (2)</td>
<td>5 (1)</td>
<td>13 (2)</td>
<td>4 (1)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ASP screen 1</td>
<td>39 (15)</td>
<td>103 (27)</td>
<td>184 (35)</td>
<td>101 (39)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ASP screen 2+</td>
<td>110 (43)</td>
<td>156 (41)</td>
<td>183 (35)</td>
<td>52 (20)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ASP interval</td>
<td>91 (36)</td>
<td>90 (24)</td>
<td>96 (18)</td>
<td>52 (20)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ASP refuser</td>
<td>10 (4)</td>
<td>28 (7)</td>
<td>53 (10)</td>
<td>50 (20)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total ASP</td>
<td>256</td>
<td>382</td>
<td>529</td>
<td>259</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PSP pre-screen</td>
<td>115 (71)</td>
<td>221 (71)</td>
<td>277 (66)</td>
<td>142 (96)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PSP screen</td>
<td>47 (29)</td>
<td>94 (29)</td>
<td>140 (34)</td>
<td>6 (4)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total PSP</td>
<td>162</td>
<td>315</td>
<td>417</td>
<td>148</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Prior = cancers diagnosed clinically between randomization and first screen.
†Includes 30 cancers diagnosed after screening was abandoned in this age group.

Table 2. Three state model results—instantaneous transition rates, MST, sensitivity, and PPV

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Age</th>
<th>Preclinical incidence rate per 100,000 person-years (95% CI)</th>
<th>MST in years (95% CI)</th>
<th>Sensitivity (95% CI)</th>
<th>PPV</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>40–49</td>
<td>50–59</td>
<td>60–69</td>
<td></td>
<td>(95% CI)</td>
</tr>
<tr>
<td>Preclinical incidence rate per</td>
<td>89</td>
<td>155</td>
<td>240</td>
<td>(84–95)</td>
<td>150–160 (230–251)</td>
</tr>
<tr>
<td>100,000 person-years (95% CI)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MST in years (95% CI)</td>
<td>2.44</td>
<td>3.70</td>
<td>4.17</td>
<td>(2.12–2.86)</td>
<td>(3.44–4.17)</td>
</tr>
<tr>
<td>Sensitivity (95% CI)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PPV</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>85%</td>
<td>100%</td>
<td>100%</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
cal screen-detectable disease, and symptomatic clinical disease. The rate of progression from preclinical to clinical disease (the reciprocal of the mean sojourn time) is much faster in the 40–49 age group than in women aged 50 or more. Note that in Table 2, we present the false-positive probability in terms of positive predictive value (PPV)—that proportion of screen-detected tumors that would have arisen clinically in the future had screening not taken place. PPV is commonly used to evaluate diagnostic tests, and should not be confused with biopsy predictive value. In women aged 50 or more, both sensitivity and PPV were around 100%, whereas in the 40–49 group, sensitivity was 83% and PPV 85%.

**Five-State Model**

Consider a model including axillary lymph node status. There are five states:

(1) No detectable disease (0);
(2) Preclinical node negative (pre −);
(3) Preclinical node positive (pre +);
(4) Clinical node negative (clin −);
(5) Clinical node positive (clin +);

The transition matrix is:

\[
\begin{pmatrix}
-\lambda_1 & \lambda_1 & 0 & 0 & 0 \\
0 & -\lambda_2 & -\lambda_3 & \lambda_2 & \lambda_3 \\
0 & 0 & -\lambda_4 & \lambda_4 \\
0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0
\end{pmatrix}
\]

We assume no regression, as before, that all tumors are born node negative and in the preclinical phase and that transition in two dimensions at exactly the same instant is not possible. Note that we cannot estimate transitions within the clinical phase, as once a tumor is diagnosed, it is excised and further assessment of natural history thereafter is impossible.

Table 3 shows the estimated instantaneous transition rates. Note the more rapid progression from node-negative to node-positive tumors in the preclinical phase in younger women and the more rapid progression from preclinical to clinical. In all age groups, the inclusion of more states to pass through once a cancer is in the preclinical state leads to a faster rate of progression into the preclinical state.

**Table 3. Results for a five-state model for progression with respect to node status model**

<table>
<thead>
<tr>
<th>Transition</th>
<th>Rate (95% CI) for age 40–49</th>
<th>Rate (95% CI) for age 50–59</th>
<th>Rate (95% CI) for age 60–69</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 → preclinical N−</td>
<td>0.00122 (0.00120–0.00125)</td>
<td>0.00176 (0.00175–0.00177)</td>
<td>0.00263 (0.00260–0.00267)</td>
</tr>
<tr>
<td>preclinical N− → preclinical N+</td>
<td>0.35 (0.22–0.55)</td>
<td>0.23 (0.15–0.35)</td>
<td>0.15 (0.10–0.23)</td>
</tr>
<tr>
<td>preclinical N+ → clinical N−</td>
<td>0.26 (0.16–0.42)</td>
<td>0.18 (0.11–0.29)</td>
<td>0.20 (0.13–0.30)</td>
</tr>
<tr>
<td>clinical N− → preclinical N+</td>
<td>2.11</td>
<td>0.85</td>
<td>0.61</td>
</tr>
<tr>
<td>preclinical N+ → clinical N+</td>
<td>(1.09–4.08)</td>
<td>(0.54–1.33)</td>
<td>(0.41–0.91)</td>
</tr>
</tbody>
</table>

Implications of the three transition probability matrices include:

(1) In the age group 40–49, a tumor which is node negative and preclinical now has a 46% chance of progression to node positive or clinical phase or both within the next year. The corresponding figures for the 50–59 and 60–69 age groups are 34% and 29%, respectively.

(2) For preclinical node-positive tumors in the 40–49 age group, 88% progress to the clinical phase within a year—that is, there is very little opportunity for detection by screening thereafter. For the 50–59 and 60–69 age groups, the figures are 57% and 54%, respectively.

(3) A preclinical node-negative tumor in the 40–49 age group is about three times as likely to progress to clinical node positive than a corresponding preclinical node-negative tumor in the 50+ age groups.

Similar patterns are observed for tumor size and for a model that includes progression with respect to both variables.

**Malignancy Grade**

This is a histological measure of aggressive potential of the tumor comprising differentiation, nuclear size, pleomorphism, and mitotic rate. Tumors are graded as 1 (good prognosis), 2 (intermediate prognosis), or 3 (poor prognosis). The malignancy grade used to be thought of as an innate unchanging quantity, but this may be an oversimplification because of a phenomenon known as “dedifferentiation,” or “phenotypic drift” (13). It is
well documented that some tumors are internally heterogeneous with respect to grade (and indeed phenotypic character). In this case, the pathologist has to score the grade based on what is the dominant component of the tumor examined. One might suspect that in such a case, the more aggressive component would grow faster than the less aggressive if the tumor were left untreated—that is, if the malignancy grade changes (the tumor differentiates) as the cancer ages. This would be manifested by more grade 3 tumors in control series than a screened series, after elimination of length-bias cases. Length bias is removed by excluding the first screen from both the ASP and the PSP. This is because: first, theory tells us that the length bias cases remain in the preclinical screen-detectable phase for a long time and therefore need only one screen to detect them; second, in the two-county study, the excess incidence in the ASP vanished after a single screen of the PSP; and third, the experience of clinicians working with screening is that the first screen contains a disproportionate number of dubious malignancies.

Table 4 shows the percentage of grade 3 tumors by age and study group, after removal of the length-bias cases. There does indeed seem to be a tendency for tumors to dedifferentiate.

There may be a further complication in that some tumors have this heterogeneity and therefore the potential to ‘‘dedifferentiate’’ and others do not. We therefore propose a mover-stayer mixture of models (12). Suppose we have a five-state model:

(1) No detectable disease.
(2) Preclinical grade 1–2.
(3) Preclinical grade 3.
(4) Clinical grade 1–2.
(5) Clinical grade 3.

For an unknown proportion p of tumors, the transition rates from state (2) to (3) and from state (4) to (5) are zero (i.e., changing grade with time is impossible), and for the remaining 1−p of all tumors, nonzero transition rates apply (i.e., progression with respect to grade is possible). Fitting this model to the two-county data, we estimated the proportion of tumors with the propensity to dedifferentiate to be 81%, 48%, and 51% for the age groups 40–49, 50–59, and 60–69, respectively. Thus, there is a larger proportion of tumors whose malignancy grade may deteriorate in women aged under 50.

### Discussion

The overwhelming implication of the above results is that progression to the clinical phase, and with respect to node status and tumor size (data available from the authors), is faster in the age group 40–49 than in older age groups. In addition, the potential for dedifferentiation or phenotypic drift is stronger in the 40–49 age group than in women aged 50 or more. This is consistent with previous suggestions that a shorter interscreening interval is required in this age group. It is of some value to quantify this further, in terms of the mortality expected from different screening frequencies. We used the survival data from the 2,468 tumors in the two-county study to predict mortality as follows:

1. Using the Markov models, we predicted the numbers of tumors by node status, size, and grade in an unscreened population and in a population screened every one, two, or three years.
2. Using survival data to estimate the Cox regression parameters for the various categories, we estimated the 10-year survival probability in each category.
3. We then multiplied the expected numbers of tumors in each category by the proportion expected to die of breast cancer in that category, thus giving the expected 10-year mortality.
4. Finally, we obtained the predicted relative mortality by dividing the predicted mortality for the screened population by that for the unscreened.

Table 5 shows the predicted relative mortality using a model incorporating both size and node status. The effects of annual two-year and three-year screening are given. In our calculations, we assumed that 90% of those invited actually attended for screening, and we used the sensitivities estimated in Table 2. Major points to note are that the predicted effects are close to those observed in the two-county study, that a shorter interscreening interval is required in the age group 40–49, and that the interval is less crucial for older women.

We can validate the use of the two-county study survival data by applying them to other trials to predict the relative mortality. Figure 1 shows the relative mortality observed in the Malmö, Gothenburg, Edinburgh, two-county, Stockholm, and Canadian trials, compared with that predicted using the node status, tumor size, and (where available) malignancy grade of tumors diagnosed within each trial, coupled with the survival rates pertaining to node status, size, and grade from the two-county study (14). The line of perfect agreement is also shown. Clearly the agreement between predicted and observed relative mortality is good, and in five out of the six trials, the predicted mortality gives a slightly conservative result.

The above has implications for study design. First, the Markov models and predicted mortality methods may be used for power and sample size calculations. Second, because of the greater information, predicted mortality from tumors diagnosed has a lower variance than observed mortality. We might therefore consider using the predicted mortality from the tumors diagnosed as a surrogate in studies to evaluate breast cancer screening strategies. Predicted mortality is to be used in the UK

### Table 4. Percent of grade 3 tumors by age and study group, two-county trial

<table>
<thead>
<tr>
<th>Group</th>
<th>40–49</th>
<th>50–59</th>
<th>60–69</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bias-free ASP</td>
<td>45%</td>
<td>38%</td>
<td>39%</td>
</tr>
<tr>
<td>Bias-free PSP</td>
<td>51%</td>
<td>49%</td>
<td>46%</td>
</tr>
</tbody>
</table>

### Table 5. Expected relative ten-year mortalities from 3-yearly, 2-yearly, and annual screening by age group*

<table>
<thead>
<tr>
<th>Interval between screens</th>
<th>40–49 (83% sensitivity)</th>
<th>50–59 (100% sensitivity)</th>
<th>60–69 (100% sensitivity)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 year</td>
<td>0.64</td>
<td>0.54</td>
<td>0.56</td>
</tr>
<tr>
<td>2 years</td>
<td>0.82 (0.87)</td>
<td>0.61</td>
<td>0.61</td>
</tr>
<tr>
<td>3 years</td>
<td>0.96</td>
<td>0.66 (0.66)</td>
<td>0.66 (0.60)</td>
</tr>
</tbody>
</table>

*Relative mortality calculated as deaths/person-years for the invited group divided by the same figure for the control group, assuming sensitivities as in Table 2 and 90% attendance rates. Figures in parentheses represent the observed relative mortalities in the Swedish Two-County Study.
Breast Screening Frequency Trial rather than actual mortality, and preliminary analysis indicates that this will double its power (15). Predicted mortality also provides results some 10 years earlier than observed mortality. This is particularly relevant to the case of breast cancer screening in the age group 40–49, for whom the actual mortality effect is often far off in the future, but the need for an answer is relatively urgent.

Conclusions

We draw the following conclusions from our analysis:

(1) Progression to a more advanced state is considerably more rapid in the 40–49 age group.

(2) This progression also occurs with respect to the malignancy grade of the tumor. The proportion of tumors capable of dedifferentiation appears to be greater in women aged 40–49.

(3) In this age group, the best indicator of future benefit is the relative rate of advanced tumors, or the predicted deaths from these. These can reasonably be used in trials.

(4) There is a potential for the effect on advanced tumors to be used to assess the likely future effect on mortality, but only if there are good data available on the stage, size, or node status of tumors before screening.

(5) The above results do not tell us whether or not to screen in this age group. They do, however, tell us something of the biological background that screening in this age group is up against and indicate that a shorter interscreening interval is more likely to be effective.

References


