Online Supplement to “Combined DES/SD Model of Breast Cancer Screening for Older Women, I: Natural-History Simulation”

In Section A1 of this Online Supplement to the main article titled “Combined DES/SD Model of Breast Cancer Screening for Older Women, I: Natural-History Simulation,” we provide a complete explanation of the cancer incidence submodel, including a detailed discussion of the BCSC data set and the Barlow logistic regression equation for predicting a woman’s risk of receiving a confirmed diagnosis of breast cancer within one year of receiving a screening mammogram. In Section A2 we discuss the disease progression submodel, including the tumor-growth equation and the equation for predicting the cancer stage at diagnosis as a function of tumor size. In Section A3 we provide a comprehensive set of results generated by the natural history simulation. Throughout this Online Supplement we survey the relevant literature in detail, including additional information regarding alternative submodels for representing key simulation input processes; and we discuss areas of potential future work.

A1. Cancer Incidence Submodel

The cancer incidence submodel consists of the Barlow risk equation (Barlow et al., 2006), whose parameters were estimated from a data set provided by the Breast Cancer Surveillance Consortium (BCSC, 2006). It is important to understand the sources of the inputs used in this simulation, so that the limitations and appropriate applications of the simulation are clear.

A1.1. BCSC Data

The following is a description of the BCSC that has been taken from the organization’s Web site http://breastscreening.cancer.gov/:

The Breast Cancer Surveillance Consortium (BCSC) is a research resource for studies designed to assess the delivery and quality of breast cancer screening and related patient outcomes in the United States. The BCSC is a collaborative network of seven mammography registries with linkages to tumor and/or pathology registries. The network is supported by a central Statistical Coordinating Center. Currently, the Consortium's database contains information on 7,521,000 mammographic examinations, 2,017,869 women, and 86,700 cancer cases (72,800 invasive cancers and 13,800 In Situ). A full description of the history and early work of the BCSC can be found at: http://breastscreening.cancer.gov/espp_report.html.

The BCSC provides data for women of ages 35+, and thus it is possible to incorporate younger women into the natural-history simulation by altering the logic in many of the simulation’s submodels and using the entire data set to sample attributes for individual women. For premenopausal and postmenopausal
women, Barlow et al. (2006) formulated separate logistic regression equations to predict the risk of receiving a confirmed diagnosis of breast cancer within one year of a screening mammogram. In the rest of this section, the discussion is focused primarily on the Barlow risk equation for postmenopausal women.

### A1.2. Barlow Risk Equation for Postmenopausal Women

A logistic regression equation (Hosmer, Lemeshow, and Sturdivant, 2013) is typically used to predict the probability \( p \) of occurrence of a specific event using a logistic function of an appropriate linear combination of relevant predictor variables (covariates) \( X_1, X_2, \ldots, X_n \). In general the logistic function of the real variable \( u \) is defined as

\[
f(u) \equiv \frac{e^u}{1+e^u} = \frac{1}{1+e^{-u}} \quad \text{for} \quad -\infty < u < \infty,
\]

so that \( f(u) \) is S-shaped and strictly increasing with \( \lim_{u \to -\infty} f(u) = 0 \) and \( \lim_{u \to \infty} f(u) = 1 \). These properties of the logistic function suggest trying to use the logistic regression equation

\[
p = f(L) = \frac{e^L}{1+e^L}, \quad \text{where} \quad L = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_n X_n
\]

(2)

to represent the dependence of \( p \) on the predictor variables \( X_1, X_2, \ldots, X_n \). To estimate the regression coefficients \( \beta_0, \beta_1, \ldots, \beta_n \) in Equation (2), we are naturally led to consider the logit (or log-odds) function

\[
\text{logit}(p) = \ln \left( \frac{p}{1-p} \right) = L = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_n X_n
\]

(3)

so that the regression coefficients can be estimated from sample data by the method of maximum likelihood (Hosmer, Lemeshow, and Sturdivant, 2013, pp. 8–10).

In their logistic regression equation for predicting a postmenopausal woman’s risk of receiving a confirmed diagnosis of breast cancer within one year of a screening mammogram, Barlow et al. (2006) tested all eleven of the risk factors in Table 1 for statistical significance. Because of the large size of the BCSC data set, effects judged to be practically insignificant could appear to be statistically significant at the 0.05 level of significance; and for this reason the authors used the significance level 0.0001 as the criterion for declaring that a risk factor is statistically significant and should therefore be included in the logistic regression equation. No interaction terms were included so as to keep the logistic regression equation simple enough for use in practice. Although there were statistically significant interactions between breast density and age and between body mass index (BMI) and current use of hormone therapy, the authors chose not to include these interactions in their logistic regression equation because including those interactions would only increase the \( c \)-statistic from 0.624 to 0.626 while requiring 28 additional covariates. To verify their risk equation, the authors used 75% of the original BCSC data as the training data set from
which they estimated the equation’s parameters; and they tested the equation’s prediction accuracy in the remaining 25% of the original BCSC data, which therefore served as the validation data set.

Table 1. Risk factor attributes of individuals in the BCSC Risk Model Data Set (BCSC, 2006).

| Factor Symbol | Factor Definition and Coding |
|---------------|-----------------------------|
| $X_1$         | Menopausal status: 0 = premenopausal; 1 = postmenopausal or age>=55; 9 = unknown |
| $X_2$         | Age group: 1 = 35–39; 2 = 40–44; 3 = 45–49; 4 = 50–54; 5 = 55–59; 6 = 60–64; 7 = 65–69; 8 = 70–74; 9 = 75–79; 10 = 80–84 |
| $X_3$         | BI-RADS breast density: 1 = Almost entirely fat; 2 = Scattered fibroglandular densities; 3 = Heterogeneously dense; 4 = Extremely dense; 9 = Unknown or different measurement system |
| $X_4$         | Race: 1 = white; 2 = Asian/Pacific Islander; 3 = black; 4 = Native American; 5 = other/mixed; 9 = unknown |
| $X_5$         | Hispanic: 0 = no; 1 = yes; 9 = unknown |
| $X_6$         | Body mass index: 1 = 10–24.99; 2 = 25–29.99; 3 = 30–34.99; 4 = 35 or more; 9 = unknown |
| $X_7$         | Age at first birth: 0 = Age < 30; 1 = Age 30 or greater; 2 = Nulliparous; 9 = unknown |
| $X_8$         | Number of first-degree relatives with breast cancer: 0 = zero; 1= one; 2 = 2 or more; 9 = unknown |
| $X_9$         | Previous breast procedure: 0 = no; 1 = yes; 9 = unknown |
| $X_{10}$      | Result of last mammogram before the current mammogram: 0 = negative; 1 = false positive; 9 = unknown |
| $X_{11}$      | Surgical menopause: 0 = natural; 1 = surgical; 9 = unknown or not menopausal |

For postmenopausal women there were eleven statistically significant risk factors from Table 1: age $X_1$; breast density $X_2$; race $X_3$; Hispanic ethnicity $X_4$; BMI $X_5$; age at first birth $X_6$; family history $X_7$; history of breast procedures $X_8$; result of last mammogram $X_9$; surgical menopause $X_{10}$; and current hormone replacement therapy $X_{11}$. Given specific values of these risk factors, a postmenopausal woman has conditional probability $p_{\text{POST}}$ of receiving a confirmed diagnosis of breast cancer within one year of a screening mammogram; and the associated logit function $L_{\text{POST}} = \logit(p_{\text{POST}})$ having the general form (3) is specifically given by

\[
L_{\text{POST}}(x) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_{11} x_{11}
\]
where in general a logical expression of the form \((X_i = j)\) takes the value one if the predictor variable \(X_i\) has the value \(j\); and otherwise this expression takes the value 0. Finally the woman’s probability of receiving a confirmed diagnosis of breast cancer within one year of a screening mammogram is estimated by the logistic regression equation

\[
L_{\text{POST}} = -7.0713 + (X_1 = 3) \times (0) + (X_1 = 4) \times (0.283) + (X_1 = 5) \times (0.6743) + (X_1 = 6) \times (0.8196) + (X_1 = 7) \times (0.9034) + (X_1 = 8) \times (1.026) + (X_1 = 9) \times (1.1086) + (X_1 = 10) \times (1.2033) + (X_2 = 1) \times (0) + (X_2 = 2) \times (0.7377) + (X_2 = 3) \times (1.0799) + (X_2 = 4) \times (1.1482) + (X_2 = 9) \times (0.2251) + (X_2 = 3) \times (0.7377) + (X_3 = 1) \times (0) + (X_3 = 2) \times (-0.2251) + (X_3 = 3) \times (0.0935) + (X_3 = 4) \times (-0.6108) + (X_3 = 5) \times (-0.0269) + (X_3 = 9) \times (-0.00037) + (X_4 = 0) \times (0) + (X_4 = 1) \times (-0.3072) + (X_4 = 9) \times (-0.0641) + (X_5 = 1) \times (0) + (X_5 = 2) \times (0.1337) + (X_5 = 3) \times (0.2463) + (X_5 = 4) \times (0.3827) + (X_5 = 9) \times (0.0295) + (X_6 = 0) \times (0) + (X_6 = 1) \times (0.1871) + (X_6 = 2) \times (0.1637) + (X_6 = 9) \times (0.0203) + (X_7 = 0) \times (0) + (X_7 = 1) \times (0.2722) + (X_7 = 2) \times (0.5080) + (X_7 = 9) \times (-0.0488) + (X_8 = 0) \times (0) + (X_8 = 1) \times (0.262) + (X_8 = 9) \times (0.0585) + (X_9 = 0) \times (0) + (X_9 = 1) \times (0.5241) + (X_9 = 9) \times (0.1840) + (X_{10} = 0) \times (0) + (X_{10} = 1) \times (-0.1701) + (X_{10} = 9) \times (-0.0603) + (X_{11} = 0) \times (0) + (X_{11} = 1) \times (0.171) + (X_{11} = 9) \times (0.124), \tag{4}
\]

\[
\begin{align*}
L_{\text{POST}} &= -7.0713 + (X_1 = 3) \times (0) + (X_1 = 4) \times (0.283) + (X_1 = 5) \times (0.6743) + (X_1 = 6) \times (0.8196) + (X_1 = 7) \times (0.9034) + (X_1 = 8) \times (1.026) + (X_1 = 9) \times (1.1086) + (X_1 = 10) \times (1.2033) + (X_2 = 1) \times (0) + (X_2 = 2) \times (0.7377) + (X_2 = 3) \times (1.0799) + (X_2 = 4) \times (1.1482) + (X_2 = 9) \times (0.2251) + (X_2 = 3) \times (0.7377) + (X_3 = 1) \times (0) + (X_3 = 2) \times (-0.2251) + (X_3 = 3) \times (0.0935) + (X_3 = 4) \times (-0.6108) + (X_3 = 5) \times (-0.0269) + (X_3 = 9) \times (-0.00037) + (X_4 = 0) \times (0) + (X_4 = 1) \times (-0.3072) + (X_4 = 9) \times (-0.0641) + (X_5 = 1) \times (0) + (X_5 = 2) \times (0.1337) + (X_5 = 3) \times (0.2463) + (X_5 = 4) \times (0.3827) + (X_5 = 9) \times (0.0295) + (X_6 = 0) \times (0) + (X_6 = 1) \times (0.1871) + (X_6 = 2) \times (0.1637) + (X_6 = 9) \times (0.0203) + (X_7 = 0) \times (0) + (X_7 = 1) \times (0.2722) + (X_7 = 2) \times (0.5080) + (X_7 = 9) \times (-0.0488) + (X_8 = 0) \times (0) + (X_8 = 1) \times (0.262) + (X_8 = 9) \times (0.0585) + (X_9 = 0) \times (0) + (X_9 = 1) \times (0.5241) + (X_9 = 9) \times (0.1840) + (X_{10} = 0) \times (0) + (X_{10} = 1) \times (-0.1701) + (X_{10} = 9) \times (-0.0603) + (X_{11} = 0) \times (0) + (X_{11} = 1) \times (0.171) + (X_{11} = 9) \times (0.124),
\end{align*}
\]

\[
\text{A1.3 Calibration and Discrimination of the Barlow Risk Equation}
\]

To formulate standard measures of calibration and discrimination for a risk prediction model, we adapt the notation of Gail and Pfeiffer (2005). For a randomly selected individual from a designated population of patients who are at risk of developing a certain disease in a specified time interval, let the random variable \(Y\) denote the indicator for the patient’s developing the disease \((Y = 1)\) or not developing the disease \((Y = 0)\) in the specified time interval; let \(\pi = \Pr\{Y = 1\}\) denote the individual’s true probability of developing the disease in the specified time interval; and let \(X = (X_1, \ldots, X_n)\) denote the individual’s associated random vector of predictor variables. For a given realization \(X = x\) of the predictor variables, a risk prediction model \(r(x)\) is said to be perfectly calibrated if

\[
r(x) = \mathbb{E}[\pi | X = x]\quad \text{for each } x, \tag{6}
\]
where the conditional expectation in Equation (6) is the true probability of developing the disease in the specified time interval computed over all individuals in the designated population for which the condition \( X = x \) holds.

One commonly used measure of discrimination for a risk model is the concordance statistic, \( c \), which is the probability that a randomly selected patient who develops the disease in the specified time interval (i.e., a case) will have a higher predicted risk than a randomly selected patient who does not develop the disease in the specified time interval (i.e., a noncase or control). For a test based on the cutoff value \( r^* \) for the predicted risk \( r(X) \) of a randomly selected patient, that individual receives a positive diagnosis if \( r(X) \geq r^* \); otherwise, the individual receives a negative diagnosis. The associated receiver operating characteristic (ROC) curve is a plot of \( \Pr \{ r(X) \geq r^* \mid Y = 1 \} \) (the test’s sensitivity) on the vertical axis against \( \Pr \{ r(X) \geq r^* \mid Y = 0 \} \) (the complement of the test’s specificity) on the horizontal axis for all \( r^* \in [0, 1] \); and the area under the ROC curve is equal to the \( c \)-statistic. The area under the ROC curve ranges from 0.5 (no discrimination) to a theoretical maximum of 1.0 (perfect discrimination).

The \( c \)-statistic is not the probability that a randomly sampled patient will be classified correctly as a case or a control (Cook 2007). Moreover, the \( c \)-statistic is not the probability that a randomly sampled patient with a high predicted risk \( r(X) \) will eventually become a case. Cook (2007) argues that the latter probability is a better measure of the predictive value of the risk model, because in diagnostic testing patients (and their examining physicians) are interested in whether they have the disease given their test result; by contrast, sensitivity and specificity are, respectively, the probabilities of having a positive or negative test result given the presence or absence of the disease. There is a trade-off between calibration and discrimination; and Diamond (1992) argues that a perfectly calibrated risk model cannot achieve a \( c \)-statistic equal to 1.0 under normal circumstances. Gail and Pfeiffer (2005) demonstrate that the maximum value of the \( c \)-statistic is dependent upon the distribution of risk in the population.

In Section A1.3.1, we show that although the Barlow risk equation for postmenopausal women is well calibrated, the actual upper limit for the area under its ROC curve (and hence the largest possible value of its \( c \)-statistic) is much less than 1.0. In fact we demonstrate that a perfectly calibrated risk model for predicting a postmenopausal woman’s risk of receiving a confirmed diagnosis of breast cancer within one year of a screening mammogram has an upper limit of approximately 0.61 for its associated \( c \)-statistic.

A1.3.1 Calibration of the Barlow risk equation

For the logistic regression equation of Barlow et al. (2006), the degree to which that risk equation is well calibrated can be determined by computing the ratio \( O / E \), where: (i) \( O \) denotes the number of confirmed diagnoses (i.e., cases) of breast cancer within one year of a screening mammogram in the validation data set; and (ii) \( E \) denotes the expected number of cases predicted by the logistic regression equation for the validation data set. Table 6 in Barlow et al. (2006) shows that the observed incidence rate of breast cancer in the validation data set is 5.612 cases per 1000 screening mammograms, and the expected
incidence rate based on Equation (4) is 5.668. Therefore $O/E = 0.9901$ in the validation data set; and under the plausible assumption that $O$ has a Poisson distribution, an approximate $100(1-\alpha)\%$ confidence interval (CI) for the expected value of $O/E$ based on the variance-stabilizing transformation $W = \sqrt{O/E}$ (Box, Hunter, and Hunter, 2005) is

$$\left(\frac{O}{E} \pm \left(\frac{z_{1-\alpha/2}}{4E}\right)\left(\frac{\sqrt{O}}{E}\right)\right)$$

so that an approximate 95% CI for the expected value of $O/E$ is $[0.9501, 1.031]$. Moreover, the Hosmer-Lemeshow goodness-of-fit test, which compares the observed and expected numbers of cases for selected subsets of the validation data set, has a $p$-value of 0.23; see Section 5.2.2 of Hosmer, Lemeshow, and Sturdivant (2013). On the basis of these results, we concluded that the Barlow risk equation is well calibrated. Because the main performance measures generated by the natural-history and screening-and-treatment simulations are averages taken over all simulated patients, in our study good calibration is the most critical property of the Barlow risk equation.

A1.3.2 Discrimination of the Barlow risk equation

On the other hand, the logistic regression equations of Barlow et al. (2006) have been criticized on the grounds that these risk equations cannot adequately discriminate women who will receive a confirmed diagnosis of breast cancer within one year of a screening mammogram from those who will not, because the $c$-statistics for premenopausal and postmenopausal women are 0.631 and 0.624, respectively. If $r(x)$ is a perfectly calibrated risk model so that Equation (6) holds, then for a randomly sampled patient who will develop the disease in the specified time period the cumulative distribution function (c.d.f.) of the random variable $r = r(X)$ is

$$F_{\text{case}}(v) = \Pr\{r \leq v \mid Y = 1\} = \mu^{-1} \int_0^v r \, dF(r) \quad \text{for } 0 \leq v \leq 1,$$

where: $F(r) = \Pr\{r \leq v\}$ is the c.d.f. of $r$ for $0 \leq r \leq 1$ taken over the entire target population (including both cases and noncases); and

$$\mu = \int_0^1 r \, dF(r)$$

is the mean risk computed over the entire target population (see Equation (2.5) of Gail and Pfeiffer (2005)). Similarly we see from Equation (2.6) of Gail and Pfeiffer (2005) that for a randomly sampled patient who will not develop the disease in the specified time interval, the c.d.f. of $r$ is

$$F_{\text{control}}(v) = \Pr\{r \leq v \mid Y = 0\} = (1 - \mu)^{-1} \int_0^v (1 - r) \, dF(r) \quad \text{for } 0 \leq v \leq 1.$$

If for a perfectly calibrated risk model the random variable $r$ is continuous with a probability density function (p.d.f.) $f(r)$ defined on the interval $[a, b]$, then we must have $0 \leq a < b \leq 1$; and we seek to com-
pute the area under the associated ROC curve in order to evaluate the associated c-statistic. Given a cutoff value $r^*$ (where $0 \leq r^* \leq 1$) so that a randomly sampled patient receives a positive diagnosis if $r \geq r^*$ and a negative diagnosis if $r < r^*$, the sensitivity of the risk model is

$$\text{sens}(r^*) = \Pr \{ r \geq r^* \mid Y = 1 \} = 1 - F_{\text{case}}(r^*) = 1 - \mu^{-1} \int_0^{r^*} rf(r) \, dr ;$$

and the specificity of the risk model is

$$\text{spec}(r^*) = \Pr \{ r < r^* \mid Y = 0 \} = F_{\text{control}}(r^*) = (1 - \mu)^{-1} \int_0^{r^*} (1-r) f(r) \, dr .$$

Using the natural-history simulation, we generated a random sample of size 100,000 from the c.d.f. $F(r)$ associated with the Barlow risk model for postmenopausal women. Then using the distribution-fitting software package ExpertFit (Law 2007), we found that the best-fitting bounded distribution for this data set was a generalized beta distribution with p.d.f. having the form

$$f(r)=\begin{cases} 
0, & \text{if } r \leq a, \\
\frac{\Gamma(\alpha_1 + \alpha_2)}{\Gamma(\alpha_1)\Gamma(\alpha_2)(b-a)^{\alpha_1-1}}(r-a)^{\alpha_2-1}(b-r)^{\alpha_1-1}, & \text{if } a \leq r \leq b, \\
0, & \text{if } b \leq r,
\end{cases}$$

and the estimated parameters

$$a = 0.0, \quad (14)$$
$$b = 0.03516, \quad (15)$$
$$\alpha_1 = 5.15884, \quad (16)$$
$$\alpha_2 = 20.29302. \quad (17)$$

Figure 1 displays a plot of the fitted p.d.f. superimposed on a histogram constructed using Scott’s (1979) rule.
Starting from Equations (11), (12), and (13), we take a direct Riemann-sum approach to computing the \( c \)-statistic for a perfectly calibrated risk model for predicting the probability that a postmenopausal woman will receive a confirmed diagnosis of breast cancer within one year of a screening mammogram. Let \( K \) denote the number of subintervals in the Riemann sum we have to compute as an approximation to the integral representing the area under the associated ROC curve. Let

\[
\frac{r^*_k}{K} = \frac{k}{K} \quad \text{for } k = 0,1,\ldots,K
\]

(18)

so that we have the following horizontal coordinates delimiting subintervals on the ROC curve,

\[
t_k = 1 - \text{spec} \left( r^*_k \right) = 1 - \left( 1 - \mu \right)^{-1} \int_0^{r^*_k} \left( 1 - r \right) f \left( r \right) dr \quad \text{for } k = 0,1,\ldots,K;
\]

(19)

and in terms of the auxiliary quantities

\[
\Delta_k = t_{k-1} - t_k \quad \text{for } k = 1,\ldots,K
\]

(20)

and

\[
\tilde{r}_k = \frac{1}{2} \left( r^*_k + r^*_{k-1} \right) \quad \text{for } k = 1,\ldots,K,
\]

(21)
the vertical coordinate on the approximation to the ROC curve over the subinterval \([t_{k-1}, t_k]\) is
\[
u_k = \text{sens}(\bar{r}_k^*) = 1 - \mu^{-1} \int_0^{\bar{r}_k^*} r f(r) \, dr \quad \text{for } k = 1, \ldots, K.
\] (22)

Thus for a perfectly calibrated model of the risk that a postmenopausal woman will receive a confirmed diagnosis of breast cancer within one year of a screening mammogram, our Riemann-sum approximation to the \(c\)-statistic is approximately given by
\[
c = \sum_{k=1}^{K} \Delta_k u_k.
\] (23)

Figure 2 is a graphical depiction of the ROC curve and our method for computing \(c\)-statistic, which is simply the area under the curve.

![Figure 2. Calculation of Riemann Sum for ROC Curve](image)

We implemented the procedure outlined above for alternative values of \(K\), and Figure 3 provides a plot of the computed maximum value of the \(c\)-statistic as a function of \(K\).
Figure 3. Convergence of the Maximum Value of the \(c\)-statistic as the Accuracy of the Riemann Sum Increases

We can see from examination of Figure 3 that the maximum value of the \(c\)-statistic converges to a value of 0.61 as the Riemann-sum approximation increases in accuracy and converges on the true maximum value of the \(c\)-statistic. One should keep in mind that this is the maximum value for the \(c\)-statistic under the following assumptions:

A1: The Barlow risk model is perfectly calibrated; and

A2: The exact distribution of risk in the designated population is known.

**Remark 1.** It can be shown that an alternative expression for the area under the ROC curve is

\[
1 - \frac{1}{\mu(1-\mu)} \int_0^1 \int_0^r wf(w) dw \cdot (1-r)f(r) dr ;
\]

and Equation (24) may provide the basis for a more computationally efficient method to evaluate the maximum value of the \(c\)-statistic for a perfectly calibrated risk. Using Equation (24) and a similar (but computationally more efficient) Riemann sum approximation procedure, the maximum value of the \(c\)-statistic again converges to a value of 0.61.

As previously stated, there is a trade-off between calibration and discrimination; and while we concluded in Section A1.3.1 that the Barlow risk equation is well calibrated, it is not perfectly calibrated according to the criterion of Equation (6). Thus, we would expect the maximum value of the \(c\)-statistic for the Barlow risk equation to be slightly greater than the computed value of 0.61 based on the assumption A1 of perfect calibration. Moreover, upon examining the plot of the fitted beta distribution superimposed
on a histogram of the risk data for the population, we see that the mode of the fitted beta distribution does not fall in the histogram bin with the highest observed relative frequency; and from Table 2, we see that the standard deviation of the fitted beta distribution is slightly smaller than the sample standard deviation of the data set. As noted in Gail and Pfeiffer (2005) and Cook (2007), the maximum value of the $c$-statistic increases as the standard deviation of risk in the designated population increases; and thus we would expect the maximum value of the $c$-statistic to be slightly greater than the computed value of 0.61 under the assumption $A_2$ of a perfectly fitted distribution of risk. Even after accounting for these deviations from assumptions $A_1$ and $A_2$, it is reasonable to conclude the maximum possible value of the $c$-statistic for the Barlow risk equation is much less than 1.0.

Because of the deviations from assumptions $A_1$ and $A_2$, we performed a sensitivity analysis by reperforming the analysis given above with an alternative beta distribution having the same endpoints and mean, but the standard deviation was rescaled by factors of 0.50, 1.50, and 2.00 to yield alternative scenarios in which the standard deviation of risk had the values 0.001375, 0.004125, and 0.00550, respectively. Note that altering the standard deviation by a factor of less than 0.5 or more than 2.0 results in beta distributions with excessively large or small shape parameters when all other parameters are held constant.

Table 3 presents the numerical results of this sensitivity study, and Figure 4 is a one-way sensitivity plot of the maximum value of the $c$-statistic as a function of the standard deviation of the distribution of risk in the target population. We can see that when the standard deviation is rescaled by the factor 1.5 so that the standard deviation has the value 0.004125, the maximum value of the $c$-statistic is 0.662; and this can be considered a conservative upper bound for the Barlow risk equation.

**Table 2. Statistical Properties of Breast Cancer Risk in the Target Population and for the Fitted Beta Distribution**

| Measure | Data       | Fitted Beta Distribution |
|---------|------------|---------------------------|
| $\mu$   | 0.007121   | 0.007126                  |
| $\sigma^2$ | 7.561E-06 | 7.552E-06                |
| $\sigma$ | 0.002749   | 0.002748                  |
| Min     | 0.00105    | 0                        |
| Max     | 0.03317    | 0.03516                   |

**Table 3. Numerical Results for Sensitivity Analysis on the Maximum Value of the $c$-statistic as a Function of the Standard Deviation of the Beta Distribution with All Other Parameters Identical**

| $K$  | $\mu$   | $\sigma^2$ | $\sigma$ | $\sigma$ Ratio | Min  | Max  | $\alpha_1$ | $\alpha_2$ | $c$  |
|------|---------|------------|----------|----------------|------|------|-------------|------------|------|
| 1000 | 0.007126| 1.89E-06   | 0.001375 | 0.50           | 0.0000 | 0.03516 | 21.200      | 83.400     | 0.554159 |
| 1000 | 0.007127| 7.55E-06   | 0.002748 | 1.00           | 0.0000 | 0.03516 | 5.159       | 20.293     | 0.607668 |
| 1000 | 0.007111| 1.70E-05   | 0.004124 | 1.50           | 0.0000 | 0.03516 | 2.170       | 8.560      | 0.662006 |
| 1000 | 0.007095| 3.02E-05   | 0.005493 | 2.00           | 0.0000 | 0.03516 | 1.130       | 4.470      | 0.713044 |
The value of the $c$-statistic for the Barlow risk equation is 0.624. If we accept the conservative upper bound of 0.662, then the Barlow Risk Model achieves 94% of the maximum discriminatory power it can achieve when it is perfectly calibrated. In order for the model to have a higher value for the $c$-statistic, one of two things would need to occur: (i) the distribution of risk in the population would need to be changed; or (ii) the model would need to be calibrated less accurately. The distribution of risk in the population is a function of the data and cannot be controlled, so using this mechanism to increase the value of the $c$-statistic is not practical. Similarly, in our study it is not practical to sacrifice calibration (correct cancer rates) for discriminatory power (who gets cancer) for the reasons detailed at the end of Section A1.3.1. The risk equation should be as well calibrated as possible; and given that, it should achieve a reasonably high $c$-statistic when compared with the maximum value of the $c$-statistic for a given distribution of risk in the population. The Barlow risk equation does exactly this as demonstrated by the discussion in this section, and thus we have demonstrated its appropriateness for use in our simulation modelling context.

**A1.3.3 The Gail Risk Model**

The Gail model (Gail et al., 1989) is another popular breast cancer risk model that is prevalent in breast cancer literature and is used by the National Cancer Institute to educate the public. More information can be found at [http://www.cancer.gov/bcrisktool/about-tool.aspx#gail](http://www.cancer.gov/bcrisktool/about-tool.aspx#gail). The Gail Model predicts the probability that a woman will be diagnosed with cancer within the next five years. This is different from the Barlow model, which has a one-year time span. The risk factors used in the Gail model are also different than those used in the Barlow model. The Gail model uses the following risk factors: age, personal history, age at the start of menstruation, age at the first live birth, and the family history of breast cancer. Unfortunate-
ly, we were unable to find an up-to-date, readily available data set that contains these attributes for a
group of older US women. The Gail also model uses risk factors that were known in 1989. In contrast, the
Barlow risk model was published in 2006; and its associated study was started in 1996, when recent in-
formation about new relevant risk factors had been discovered. For the reasons discussed, the Barlow
model was chosen for use in the natural history simulation, but it is important to explore and acknowledge
other valid models for calculating risk also exist in the literature (Meads et al., 2012)

A2. Disease Progression Submodel

The disease progression submodel consists of a Gompertzian tumor growth model (Norton, 1988) that
tracks the size of the primary tumor from onset to death, and the Plevritis stage progression model
(Plevritis, et al., 2007) that is used to determine the stage of invasive breast cancer at diagnosis as a func-
tion of primary tumor size. The Gompertzian tumor growth model was found in the literature and was
validated against three data sets for untreated tumor growth; however, we extended the basic model in
several ways that we felt made the model more realistic and better suited to our purposes. The Plevritis
model allows us to determine whether the invasive cancer is in the local, regional, or distant stage at the
time of diagnosis. The stage of invasive breast cancer at diagnosis has a dramatic effect on the type of
treatment and on the patient's quality of life for the rest of her life. Different screening policies will lead
to different stage distributions at diagnosis, and this is an important measure of screening effectiveness
that we are able to capture. We are also able to validate the natural history model by comparing the stage
distribution resulting from perfect screening to SEER's reported 65-and-older stage distribution, which
reflects the actual level of screening in the population.

A2.1. Gompertz Tumor Growth Model and Extensions

One of the most important pieces of the natural history simulation is a model of breast cancer tumor
growth, which describes the growth of cancer over time within the body. A Gompertz model proposed by
Norton of human breast cancer growth is used to describe the size of a single breast cancer tumor at time
t by \(N(t)\), the number of malignant cells in the tumor at time \(t\). Other tumor growth models (Shumate &
El-Shenawee, 2007, Speer, et al., 1984) were considered for use in the natural history model, but Norton’s
approach was chosen because there are abundant data that suggest breast cancer growth in an individual
woman can be accurately represented by a general Gompertz function (McManus & Waelsch, 1980,
Norton, 1988, Pearlman, 1976, Rae-Venter & Reid, 1980, Surborne & Norton, 1993). Equation (25) was
derived from Norton and describes tumor growth from some initial tumor size \(N(t_1)\) at time \(t_1\) to the size
\(N(t_2)\) at some future time \(t_2\),

\[
N(t_2) = N(t_1) \exp \left( k \left( 1 - e^{-b(t_2-t_1)} \right) \right), \quad \text{for } (t_2 \geq t_1),
\]

(25)
where:

- \( N(t_1) \) = Individual’s tumor size (cells) at the time \( t_1 \)
- \( N(t_2) \) = Individual’s tumor size (cells) at the time \( t_2 \)
- \( N(\infty) \) = Maximum size of tumor (cells),
- \( N_L \) = Tumor size at which breast cancer is considered lethal,
- \( b \) = Tumor growth constant,
- \( t = t_2 - t_1 \) = Time delay between tumor sizes \( N(t_1) \) and \( N(t_2) \) (months),
- \( k = \ln\left[\frac{N(\infty)}{N(t_1)}\right] \).

Norton assumes \( N(t_1) = N(t_{cd}) \), the size at clinical (or symptomatic) detection, and \( N(t_2) = N_L \). He was interested in determining the best values of \( N(t_{cd}) \), \( N_L \), and \( N(\infty) \); and he was also interested in determining the distribution of \( b \), the tumor growth rate. Norton uses a data set provided by Bloom, et al. (1962) that consists of data on 250 women who had untreated breast cancers and were followed in Middlesex Hospital of London between 1805 and 1933. Norton uses an iterative scheme to solve for the following parameter values and distributions when fitting the Gompertzian model to the Bloom data:

- \( N(t_{cd}) = 4.8 \times 10^9 \) cells,
- \( N_L = 10^{12} \) cells,
- \( N(\infty) = 3.1 \times 10^{12} \) cells,
- \( b \sim \text{Lognormal}(E[b] = 0.0709, V[b] = 0.00328) \).

Norton acknowledges that in reality, the values of \( N(t_{cd}) \), \( N_L \), and \( N(\infty) \) are random variables and not constants. This analysis did not provide information about the size of breast cancer at mammographic detection, which is a parameter that is necessary for our simulation. Thus, we extended some of the assumptions of this rather simplistic model in a number of ways in order to account for newly acquired information about breast cancer, such that the growth rate is dependent upon age. In the following section, we describe how we extended and updated this tumor growth model for use in the natural history simulation. It should also be noted that in reality, tumors reach a terminal size; and breast cancer death is caused by metastasis to other organs, which prevents those organs from functioning properly (Michaelson, et al., 1999). In this research, we refer to \( N_L \) as the size at which the tumor becomes lethal because in reality, the cancer will never reach that size, but it represents the size the cancer would be when it becomes lethal if it kept growing and the body could support an ambiguously large tumor.
A2.1.1. Extensions to the Tumor Growth Model

Breast cancer is believed to behave less aggressively in older women (Crivellari et al. 2007, Diab et al. 2000, Downey et al. 2007, Mandelblatt et al. 1992, Peer et al. 1993). Thus, to accurately capture the behavior of tumor growth in older women, the tumor growth rate needs to be age-dependent. We match the 25th percentile of the original distribution of $b$ to the expected value of the lognormal distribution of $b$ for a 25-year-old woman and the 75th percentile of the original distribution of $b$ as the expected value of the lognormal distribution of $b$ for a 75-year-old woman; and we let the mean of $b$ vary linearly for all ages between 25 and 75. Figure 5 is a graph of Norton's original lognormal distribution for $b$ (x 1,000 for clarity) and Table 4 shows the percentiles of this distribution. Figure 6 and Figure 7 are graphs of the probability distributions of tumor growth constant $b$ for a 65-and a 75-year-old woman after using the method described above to determine the age dependency. Other approaches for modeling tumor growth as a function of age may be considered in the future.

![Figure 5. Original Distribution of Tumor Growth Constant $b \times 1,000$](image.png)
Table 4. Percentiles for Original Distribution of Tumor Growth Constant $b$

| Percentile | Value  |
|------------|--------|
| 5          | 0.01711|
| 10         | 0.02215|
| 15         | 0.02336|
| 20         | 0.03027|
| 25         | 0.03408|
| 30         | 0.03791|
| 35         | 0.04185|
| 40         | 0.04596|
| 45         | 0.05032|
| 50         | 0.05502|
| 55         | 0.06015|
| 60         | 0.06586|
| 65         | 0.07233|
| 70         | 0.07984|
| 75         | 0.08882|
| 80         | 0.10001|
| 85         | 0.11484|
| 90         | 0.13668|
| 95         | 0.17690|

Figure 6. Distribution of Tumor Growth Constant $b$ for a 65-Year-Old Woman x 1,000
There are two manners by which tumors are deemed "detected" in the natural history simulation. The first is when the Barlow risk model determines the one-year risk of being diagnosed with breast cancer, and the simulation determines she will get breast cancer. The second is by clinical detection from a self-exam, symptomatic cancer, and physician’s clinical exam. As stated in Tejada, et al. (2013a), a generalized beta distribution was fit to parameters given by Norton having a minimum $10^9$, a mode of $4.8 \times 10^9$, and maximum of $5 \times 10^9$, and a standard deviation equal to one-sixth of the range, which leads to the first shape parameter having the value of 4.149 and the second shape parameter having the value of 1.166. Figure 8 is a graphical depiction of this distribution.

Figure 7. Distribution of Tumor Growth Constant $b$ for a 75-Year-Old Woman x 1,000

Figure 8. Distribution of Tumor Size at Clinical Detection, $N(t_{cd})$
Onset ages are of interest, and our method for calculating them is as follows. In back-calculating a woman’s age at the onset of the tumor, we define the following quantities:

\[ N(t_1) = N(t_{on}) \] is the tumor size (cells) at the time of cancer onset \( t_1 = t_{on} \).

For a woman whose cancer is detected by mammography \( t_{md} < t_{cd} \):

\[ N(t_2) = N(t_{md}) \] is the tumor size (cells) at the time \( t_2 = t_{md} \) of mammographic detection.

For a woman whose cancer is detected clinically \( t_{md} < t_{cd} \):

\[ N(t_2) = N(t_{cd}) \] is the tumor size (cells) at the time \( t_2 = t_{cd} \) of clinical detection.

Given values of \( N(t_1), N(t_2), N(\infty), \) and \( b \), we can calculate the time delay \( t_0 = t_2 - t_1 \) (in months) required for the cancer to grow from its initial size to the size at detection:

\[
t_0 = t_2 - t_1 = (-b)^{-1} \ln \left\{ 1 - k^{-1} \ln \left[ \frac{N(t_2)}{N(t_1)} \right] \right\} \text{ months.}
\]  

We can then subtract this time delay from the woman’s current age, and thus find \( AGE_{on} \), the theoretical onset age of the cancer:

\[
AGE_{on} = \text{CurrentAge} - \left( \frac{t_0}{12} \right) \text{ years.}
\]  

Similarly, by letting \( N(t_1) = N(t_{on}) \) but now letting \( N(t_2) = N_L \), we calculate the time delay for the cancer to grow from its initial size to lethal size, and the age at which death from breast cancer would occur if the cancer remains untreated. In the natural history model, all cancers are assumed to be left untreated, and all cancers which reach lethal size are assumed to cause breast cancer death. However, after computation of the breast cancer death age, the model determines whether or not a death from other causes would have occurred prior to breast cancer death; and if so, then the death is recorded as a death not caused by breast cancer.

Other tumor growth models were considered including the Speer model (Speer, et al., 1984) and a model proposed by Shumate and El-Shenawee (Shumate & El-Shenawee, 2007), and this model was chosen because it offered advantages over the others. The Speer model is a slightly more complicated version of the Gompertz model. Essentially, it assumes the same underlying Gompertzian kinetics (small tumors grow quickly while large tumors grow slowly) but allows for the tumor to grow in "spurts" instead of steadily throughout the course of a woman's life. These spurts occur stochastically throughout the life of the tumor. The equation which governs the Speer model is more complicated than the equation that governs the Gompertz model, and it provides the same goodness-of-fit. Lastly, a model proposed by Shumate and El-Shenawee was evaluated. This is the most complicated model reviewed, and the model went into a level of detail beyond what was required for the simulation model. The model has three dimensions, and tracks the volume and shape of the tumor as well as the number of cells in the tumor at any given time. This specific model was not tailored specifically to breast cancer tumors, but is a representation of general tumor growth that was applied to breast cancer.
A2.2. Stochastic Model of Cancer Stage at Diagnosis

Breast cancer is typically defined in terms of three stages: local, regional, and distant. Refer to Table 5 and Table 6 for a definition of these stages. Plevritis, et al. (2007) uses SEER data to construct a stochastic model of the natural history describing the progression in the stages of breast cancer, and Figure 9 is a graphical depiction of the model.

Table 5. Local, Regional, and Distant Stages of Breast Cancer (Weiss, 2008)

| Stage     | Definition                                      |
|-----------|------------------------------------------------|
| Local     | The cancer is confined within the breast.       |
| Regional  | The lymph nodes, primarily those in the armpit, and possibly those near the collarbone, are involved. |
| Distant   | The cancer is found in other parts of the body and other organs are involved. |

Table 6. Detailed Stages of Breast Cancer (Weiss, 2008)

| Stage  | Definition                                                                 |
|--------|---------------------------------------------------------------------------|
| Stage 0| Cancer cells remain inside the breast duct, without invasion into normal adjacent breast tissue. |
| Stage I| Cancer is 2 centimeters or less and is confined to the breast (lymph nodes are clear). |
| Stage IIA| No tumor can be found in the breast, but cancer cells are found in the auxiliary lymph nodes (the lymph nodes under the arm) OR the tumor measures 2 cm or smaller and has spread to the auxiliary lymph nodes OR the tumor is larger than 2 cm but no larger than 5 cm and has not spread to the auxiliary lymph nodes. |
| Stage IIB| The tumor is larger than 2 cm but no larger than 5 cm and has spread to the auxiliary lymph nodes OR the tumor is larger than 5 cm but has not spread to the auxiliary lymph nodes. |
| Stage IIIA| No tumor is found in the breast. Cancer is found in auxiliary lymph nodes that are sticking together or to other structures, or cancer may be found in lymph nodes near the breastbone OR the tumor is any size. Cancer has spread to the auxiliary lymph nodes, which are sticking together or to other structures, or cancer may be found in lymph nodes near the breastbone. |
| Stage IIIB| The tumor may be any size and has spread to the chest wall and/or skin of the breast AND may have spread to auxiliary lymph nodes that are clumped together or sticking to other structures or cancer may have spread to lymph nodes near the breastbone. Inflammatory breast cancer is considered at least stage IIIB |
| Stage IIIC| There may either be no sign of cancer in the breast or a tumor may be any size and may have spread to the chest wall and/or the skin of the breast AND the cancer has spread to lymph nodes either above or below the collarbone AND the cancer may have spread to auxiliary lymph nodes or to lymph nodes near the breastbone. |
| Stage IV| The cancer has spread — or metastasized — to other parts of the body. |
Table 7 shows the maximum likelihood estimates for the parameters $\gamma$, $\eta$, $\omega$, $\beta$, and $\alpha$ with asymptotic confidence intervals conditioned on $\alpha = \beta$. Using the parameter values given in Table 7 with the exception of setting $\gamma = 0$ produces a set of equations that estimate breast cancer stage distribution only conditioned on the tumor size.

Table 7. Maximum Likelihood Estimates of Parameters, with Asymptotic Confidence Intervals Conditioned on $\alpha = \beta$ (Plevritis, et al., 2007)

| Parameter | Estimates of the Natural Log of Parameter Values | 95% CI Endpoints of Natural Log of Parameter Values | Estimates of Parameter Values |
|-----------|-----------------------------------------------|-----------------------------------------------|-------------------------------|
| $\hat{\gamma}$ | $-9.602$ | $[-9.624, -9.580]$ | $6.759 \times 10^{-5}$ |
| $\hat{\eta}$ | $-9.636$ | $[-9.661, -9.610]$ | $6.533 \times 10^{-5}$ |
| $\hat{\omega}$ | $-11.765$ | $[-11.816, -11.713]$ | $7.771 \times 10^{-6}$ |
| $\hat{\beta}$ | $-0.165$ | $[-0.187, -0.143]$ | $0.8478$ |
| $\hat{\alpha}$ | $\hat{\beta}$ | $-$ | $-$ |

In another model of breast cancer metastasis, Michaelson, et al. (1999) estimate the distribution of the probability $1/P$ that a single cell will leave the primary tumor and form a distant metastasis. On each day $t$ we know the number of cells in the tumor, $N(t)$, and the probability $1/P$ that each of those cells will independently form a distant metastasis at that time. Michaelson estimates that 25% of women have a $1/P$ value of $10^{-11}$, 50% of women have a $1/P$ value of $10^{-12}$, and 25% of women have a $1/P$ value of
On each day given the current values of $N(t)$ and $1/P$, the number of cells that metastasize, $X(t)$, is a random variable whose conditional distribution is binomial with the number of trials $N(t)$ and the success probability $1/P$ so that $X(t)$ has the following conditional probability mass function (PMF) given $N(t)$ and $1/P$,

$$\Pr\{X(t) = x \mid N(t), 1/P\} = \frac{[N(t)]!}{x![(N(t) - x)]!} \left(\frac{1}{P}\right)^x \left[1 - \left(\frac{1}{P}\right)\right]^{N(t) - x}$$

for $X(t) = 0, 1, \ldots, N(t)$. (28)

To sample the number of cells that metastasize according to the conditional PMF (6), we would need to iteratively compute successive points on that conditional PMF, accumulate the corresponding points on the cumulative distribution function, and then exploit the inverse transform method (Kelton et al., 2010). We considered using this approach to compute the time at which the first distant metastasis occurred and the size of the tumor at that time. However, this method requires computation of the probability of metastasis on each day from initial size to lethal size, which would drastically increase the run time of the natural history simulation model. In addition, the model requires our simulation software to realize extremely small probabilities (on the order of $10^{-13}$) an extremely large number of times, and Arena does not handle these situations well. Note that in a simulation, we could exploit the Poisson approximation to the binomial PMF,

$$\Pr\{X(t) = x \mid N(t), 1/P\} \approx \frac{[N(t)/P]^x \exp[-N(t)/P]}{x!} \text{ for } x = 0, 1, \ldots, N(t);$$

see Section 2.5.1 of Hoel (1984). It follows immediately from Equation (29) that the random variable $X(t)$ could be sampled from a Poisson distribution with mean $N(t)/P$, but the execution time for this approximation for Michelson’s approach is still prohibitively large. Lastly, this model considers regional metastasis only and provides no way to measure the point at which regional metastasis occurs. For these reasons, we use Plevritis’ model of cancer staging.

Carter et al. (1989) describe the relationship of tumor size to lymph node status and overall survival by analyzing SEER data. This allows us to estimate the stage of cancer on a more detailed level, particularly when it comes to cancers in the regional stage. The authors divide the number of positive lymph nodes into three categories: 0 nodes positive, 1–3 nodes positive, and 4+ nodes positive. Table 8 provides a summary of this data, where we have factored out unknown cases because they do not provide useful information. Table 9 shows the probabilities of lymph node involvement as a function of tumor size that are used in the natural history simulation.
Table 8. Distribution of Breast Cancer Cases by Size and Lymph Node Status (Carter, et al., 1989)

| Tumor Size (mm) | Cases | 0 Nodes | 1–3 Nodes | 4+ Nodes | Unknown | Total | Total–Unknown |
|-----------------|-------|---------|-----------|----------|---------|-------|--------------|
| 0–5             | 21,530| 9,721   | 563       | 197      | 11,049  | 21,530| 10,481       |
| 5–10            | 37,075| 23,816  | 2,261     | 541      | 10,457  | 37,075| 26,618       |
| 10–20           | 93,875| 58,654  | 14,035    | 4,475    | 16,711  | 93,875| 77,164       |
| 20–30           | 54,610| 27,139  | 12,194    | 6,440    | 8,837   | 54,610| 45,773       |
| 30–40           | 23,880| 9,497   | 5,704     | 4,112    | 4,267   | 23,880| 19,618       |
| 40–50           | 11,786| 3,866   | 2,692    | 5,735    | 4,275   | 11,786| 9,311        |
| 50–100          | 17,015| 4,120   | 3,135    | 4,706    | 17,015  | 12,309|              |
| 100+            | 2,580 | 382     | 290       | 1,238    | 2,580   | 1,342 |              |

Table 9. Probabilities of Lymph Node Involvement Used in the Natural History Model

| Tumor Size (mm) | 0 Nodes | 1–3 Nodes | 4+ Nodes | Total 1+ Nodes | Number of Nodes Given Node–Positive 1–3 Nodes | 4+ Nodes |
|-----------------|---------|-----------|----------|----------------|----------------------------------------------|----------|
| 0–5             | 0.9275  | 0.0537    | 0.0188   | 0.0725         | 0.74079                                      | 0.25921  |
| 5–10            | 0.8947  | 0.0849    | 0.0203   | 0.1053         | 0.80692                                      | 0.19308  |
| 10–20           | 0.7601  | 0.1819    | 0.0580   | 0.2399         | 0.75824                                      | 0.24176  |
| 20–30           | 0.5929  | 0.2664    | 0.1407   | 0.4071         | 0.65440                                      | 0.34560  |
| 30–40           | 0.4842  | 0.2908    | 0.2250   | 0.5158         | 0.56386                                      | 0.43614  |
| 40–50           | 0.4152  | 0.2891    | 0.2957   | 0.5848         | 0.56386                                      | 0.43614  |
| 50–100          | 0.3347  | 0.2547    | 0.4106   | 0.6653         | 0.38283                                      | 0.61717  |
| 100+            | 0.2846  | 0.2161    | 0.4993   | 0.7154         | 0.30208                                      | 0.69792  |

In Table 9, the two columns on the far right represent the conditional probability of having 1–3, or 4+ positive nodes as a function of tumor size, given that there are positive lymph nodes. If it is determined there are 1–3 positive nodes, then we assume that the number of positive nodes is uniformly distributed on the integers 1–3. If more than three nodes are positive, then we simply record that fact that there are four or more nodes positive. We use this information and the size of the tumor at diagnosis to estimate the stage at diagnosis at the more detailed level (refer to Table 6). Tejada (2012) shows our method for determining the detailed stage at diagnosis as a function of the tumor size (i.e., diameter) and lymph node involvement.

One of the concerns about using the Michelson model for metastasis was the fact that it did not provide a method for determining whether the cancer was in the local or regional stage at diagnosis. If the cancer is not in the distant stage, then the “Total 1+ Nodes” column in Table 9 is an estimate of the probability that lymph nodes are involved as a function of tumor size at diagnosis. If lymph nodes are in-
volved, then the cancer is in the regional stage; otherwise the cancer is in the local stage. If it is determined that the cancer is in the regional stage, the number of nodes that are positive can be assigned as previously described, and the detailed stage of cancer can again be determined. However, the computational inefficiency of the Michaelson model still makes its use infeasible.

A3. Population Growth Submodel

As stated in Tejada et al. (2013a), the target population of US women ages 65 and older will be rapidly increasing over the next decade. In order to account for this population growth in our simulation model, we used US Census data to determine the percentage growth in the target population from 2001–2009, and then fit several mathematical models to this data in order to predict the trend from 2010–2020 when data are unavailable. After fitting several mathematical models, including linear models and high order polynomial models, the two best fits were the linear model and the simple quadratic model. Figure 10 and Figure 11 are graphical depictions of these fits for the time periods 2000–2009 and 2000–2020 respectively. Visual inspection of Figure 10 shows little difference between the two models in the area of the fit (2000–2009). However, upon examining Figure 11, we can clearly see that there is a significant difference in how these two models project population growth from 2010–2020. The linear model projects continuing growth while the quadratic model projects a leveling off and then less growth beginning in 2018. Given that the target population of US females 65 and older is projected to continue growing rapidly through 2020, the linear model is more appropriate. Additionally, the linear model has a higher adjusted \( R^2 \) value (0.7806 compared with 0.7193). Thus, we choose to use a linear model of population growth in order to drive the underlying position process used to govern the growing population size in the natural history model.
Figure 10. Graphs of Fits of both Linear and Quadratic Models to Census Data for the Percentage Population Growth for US Females 65 and Older (2000 – 2010)

Figure 11. Graphs of Fits of both Linear and Quadratic Models to Census Data for the Percentage Population Growth for US Females 65 and Older (2000 – 2020)
Table 10. Data from Fits of both Linear and Quadratic Models to Census Data for the Percentage Population Growth for US Females 65 and Older

| Year | Census Data (C) | Linear Fit ($F_1$) | Quadratic Fit ($F_2$) |
|------|----------------|---------------------|-----------------------|
| 2000 | 0.453          | 0.3724              | 0.2965                |
| 2001 | 0.468          | 0.5379              | 0.5132                |
| 2002 | 0.733          | 0.7033              | 0.7171                |
| 2003 | 0.687          | 0.8688              | 0.9082                |
| 2004 | 0.947          | 1.0342              | 1.0865                |
| 2005 | 1.122          | 1.1996              | 1.2519                |
| 2006 | 1.489          | 1.3651              | 1.4045                |
| 2007 | 2.108          | 1.5305              | 1.5443                |
| 2008 | 1.685          | 1.6960              | 1.6713                |
| 2009 | 1.488          | 1.8614              | 1.7855                |
| 2010 |                | 2.0269              | 1.8869                |
| 2011 |                | 2.1923              | 1.9754                |
| 2012 |                | 2.3578              | 2.0511                |
| 2013 |                | 2.5232              | 2.1140                |
| 2014 |                | 2.6886              | 2.1641                |
| 2015 |                | 2.8541              | 2.2014                |
| 2016 |                | 3.0195              | 2.2259                |
| 2017 |                | 3.1850              | 2.2375                |
| 2018 |                | 3.3504              | 2.2363                |
| 2019 |                | 3.5159              | 2.2223                |
| 2020 |                | 3.6813              | 2.1955                |

| $R^2$ | Adjusted $R^2$ | $\Sigma (C - F_i)$ | $\Sigma |C - F_i|$ | $\Sigma (C - F_i)^2$ | $\Sigma (C - F_i)^2/N$ |
|-------|----------------|---------------------|---------------|---------------------|------------------------|
|       |                | 0.8050              | 6.27E-09      | 1.612403            | 0.546906               |
|       |                | 0.7806              |               | 1.667506            | 0.525210               |
|       |                | 0.7593              |               | 0.054691            | 0.052521               |

A4. Analysis of Simulation Results, Discussion, and Validation Considerations

There are several important results from the natural history model presented in this section. There are a few different classes of results, including annual 95% CIs on population cancer statistics, 95% CIs on cancer statistics for the population as a whole over the entire model time horizon, 95% CIs on stage distribution, 95% CIs on invasive cancer outcome percentages, and distributions that were inferred from individual observations of certain individual cancer attributes. Many of the results are presented graphically in this section, but the raw data for the confidence intervals and SEER data are available in Appendix C of Tejada (2012).

A4.1. Annual 95% CI's on Population Cancer Statistics

The following is a list of important results for which 95% confidence intervals were computed for each of the years 2001–2020, and their definitions.
1. Cancer Incidence - the expected value of the number of women that are diagnosed with cancer in a population over a certain finite period of time, typically chose to be one year. Usually presented in the form of incidence rates per 100,000 women.

2. Cancer Prevalence - the expected value of the total number of women in a population that have breast cancer regardless of the time of diagnosis. Usually presented in the form of incidence rates per 100,000 women or percentages of the designated population.

3. Deaths - the expected value of the number of cancer deaths, deaths from causes other than breast cancer, and the total number of deaths that occur each year.

4. Population Size - the expected values of the size of the population each year

**Incidence**

We begin with cancer incidence rates per 100,000 women, which we divide into invasive cancer incidence rates, DCIS incidence rates, and total incidence rates. Each graph contains the mean, upper and lower limits of a 95% CI about the mean, and SEER data (age adjusted) for comparison and validation purposes. Figure 12 is a graph of annual invasive cancer incidence rates, Figure 13 is a graph of annual DCIS incidence rates, and Figure 14 is a graph of annual total breast cancer incidence rates.

![Invasive Cancer Incidence Rate Per 100,000 Women](image)

*Figure 12. Annual Invasive Cancer Incidence Rates per 100,000 Women*
Figure 13. Annual DICS Incidence Rates per 100,000 Women

Figure 14. Annual Total Cancer Incidence Rates per 100,000 Women
Prevalence

We move now to cancer prevalence rates per 100,000 women, which we again divide into invasive cancer prevalence rates, DCIS prevalence rates, and total prevalence rates. Each graph again contains the sample mean as well as upper and lower limits of a 95% CI for the true prevalence rates; however, SEER data for prevalence is not presented on an annual basis. SEER prevalence data from a few different studies simply present the prevalence percentages for certain age groups, and we use a recent US study for comparison with our results. Figure 15 and Figure 16 are graphs of annual invasive cancer prevalence rates and percentages respectively. Figure 17 and Figure 18 are graphs of DCIS prevalence rates and percentages, respectively. Finally, Figure 19 and Figure 20 are graphs of total breast cancer prevalence rates and percentages, respectively.

![Annual Invasive Cancer Prevalence Rate Per 100,000 Women](image)

**Figure 15.** Phase I Annual Invasive Cancer Prevalence Rates per 100,000 Women
Figure 16. Phase I Annual Invasive Cancer Prevalence Percentages

Figure 17. Phase I Annual DCIS Prevalence Rates per 100,000 Women
Figure 18. Phase I Annual DCIS Prevalence Percentages

Figure 19. Phase I Annual Total Prevalence Rates per 100,000 Women
Deaths

Death rates per 100,000 women are divided into breast cancer death rates, death rates from other causes, and total death rates. Each graph contains the sample mean, upper and lower limits of a 95% CI for the theoretical death rate about the mean, and SEER data for comparison and validation purposes. Figure 21 is a graph of annual breast cancer death rates, Figure 22 is a graph of annual non breast cancer death rates, and Figure 23 is a graph of annual total breast cancer death rates.
Figure 21. Phase I Annual Cancer Death Rates per 100,000 Women with SEER Data for Comparison

Figure 22. Phase I Annual Non Breast Cancer Death Rates per 100,000 Women with SEER Data for Comparison
This section provides 95% confidence intervals on the expected values of the main cancer performance measures and other population characteristics. The goal of this research is to make recommendations for future breast cancer screening policies, so Table 12 contains performance measures for the full twenty years from 2001–2020, for the 11 year "warm-up" period 2001–2011, and for the future period 2012–2020. There are some interesting results in this table that merit a brief discussion.

We reported the number of cancers that were detected by mammography as opposed to clinical detection; and as expected with perfect annual screening, nearly all cancers are detected by mammography before symptoms become present. However, there were still some fast-growing cancers that caused symptoms to become present before mammographic detection. Perfect annual screening over twenty years leads to 470,000 mammograms being administered over this time period; and since the population size is 0.1% of the actual US population, this number can be multiplied by 1,000 to estimate the actual number of mammograms that would be given in the United States if there were perfect annual screening. Equivalent information to that given in Table 12 is presented for scenarios in the screening-and-treatment simulation (Tejada et al., 2013b).
### Table 12. 95% CIs on Population Characteristics Based on Natural History Model

| Performance Measure                        | Years     | Mean      | HW       | LL       | UL       |
|-------------------------------------------|-----------|-----------|----------|----------|----------|
| Number of Invasive Cancers Detected       | 2001–2020 | 2,633.90  | 43.41    | 2,590.49 | 2,677.31 |
|                                           | 2001–2011 | 1,344.60  | 25.99    | 1,318.61 | 1,370.59 |
|                                           | 2012–2020 | 1,289.30  | 23.09    | 1,266.21 | 1,312.39 |
| Number of DCIS Cancers Detected           | 2001–2020 | 641.70    | 18.78    | 622.92   | 660.48   |
|                                           | 2001–2011 | 321.30    | 14.36    | 306.94   | 335.66   |
|                                           | 2012–2020 | 320.40    | 14.84    | 305.56   | 335.24   |
| Method of Detection - Mammography         | 2001–2020 | 2,590.20  | 42.85    | 2,547.35 | 2,633.05 |
|                                           | 2001–2011 | 1,324.50  | 25.63    | 1,298.87 | 1,350.13 |
|                                           | 2012–2020 | 1,265.70  | 23.28    | 1,242.42 | 1,288.98 |
| Method of Detection - Clinical Presentation| 2001–2020 | 43.70     | 3.44     | 40.26    | 47.14    |
|                                           | 2001–2011 | 20.10     | 2.69     | 17.41    | 22.79    |
|                                           | 2012–2020 | 23.60     | 3.11     | 20.49    | 26.71    |
| Number of Cancer Deaths                   | 2001–2065 | 1,348.00  | 27.00    | 1,310.00 | 1,407.00 |
|                                           | 2001–2011 | 264.00    | 11.62    | 244.00   | 294.00   |
|                                           | 2012–2020 | 513.70    | 17.76    | 482.00   | 551.00   |
| Number of Non-BC Deaths with Invasive Cancer | 2001–2020 | 1,285.90  | 33.40    | 1,193.00 | 1,353.00 |
|                                           | 2001–2011 | 327.90    | 15.29    | 289.00   | 352.00   |
|                                           | 2012–2020 | 519.70    | 15.11    | 490.00   | 543.00   |
| Number of Mammograms Given                | 2001–2020 | 468,833.20| 1,439.76 | 467,393.44| 470,272.96|
|                                           | 2001–2011 | 235,216.60| 606.86   | 234,609.74| 235,823.46|
|                                           | 2012–2020 | 233,616.60| 922.14   | 232,694.46| 234,538.74|
| Number of Mammograms per Invasive Detection | 2001–2020 | 178.09    | 3.09     | 175.00   | 181.18   |
|                                           | 2001–2011 | 731.71    | 21.38    | 710.33   | 753.09   |
|                                           | 2012–2020 | 143.21    | 2.61     | 140.60   | 145.82   |
| Number of Mammograms per DCIS Detection   | 2001–2020 | 178.09    | 3.09     | 175.00   | 181.18   |
|                                           | 2001–2011 | 731.71    | 21.38    | 710.33   | 753.09   |
|                                           | 2012–2020 | 143.21    | 2.61     | 140.60   | 145.82   |
| Number of Mammograms per Cancer Detection | 2001–2020 | 178.09    | 3.09     | 175.00   | 181.18   |
|                                           | 2001–2011 | 731.71    | 21.38    | 710.33   | 753.09   |
|                                           | 2012–2020 | 143.21    | 2.61     | 140.60   | 145.82   |
| Number of Non-BC Deaths                   | 2001–2020 | 21,137.00 | 78.04    | 21,058.96| 21,215.04|
|                                           | 2001–2011 | 11,022.70 | 86.32    | 10,936.38| 11,109.02|
|                                           | 2012–2020 | 10,114.30 | 73.18    | 10,041.12| 10,187.48|
| Number of Survivors                       | 2001–2020 | 28,827.50 | 147.23   | 28,680.27| 28,974.73|
|                                           | 2001–2011 | 20,582.00 | -        | 20,582.00| 20,582.00|
|                                           | 2012–2020 | 22,529.99 | 72.07    | 22,457.92| 22,602.06|
| Initial Population Size                   | 2001–2020 | 28,827.50 | 147.23   | 28,680.27| 28,974.73|
| Average Population Size                   | 2001–2020 | 20,582.00 | -        | 20,582.00| 20,582.00|
| Final Population Size                     | 2001–2020 | 22,529.99 | 72.07    | 22,457.92| 22,602.06|
| Total Number of Women Simulated           | 2001–2020 | 50,799.30 | 156.58   | 50,642.72| 50,955.88|

### A4.3. 95% CIs on Stage Distribution at Diagnosis

The stage distribution at diagnosis reveals the percentage of cancers that were detected in the local, regional, and distant stages, respectively. This is a very important statistic, because breast cancer mortality is largely dependent on the stage at diagnosis. According to SEER data, if invasive breast cancer is diagnosed in the local stage, then there is a 20% chance of dying from breast cancer. This increases to 45% if invasive cancer is diagnosed in the regional stage, and to 80% if invasive cancer is diagnosed in the distant stage. Table 13 shows 95% CIs on the number of invasive breast cancers diagnosed in each stage and the percentages of invasive breast cancers diagnosed in each stage from the natural history simulation.
Table 13 also includes SEER data on the percentages of invasive breast cancer diagnosed in each stage for comparison purposes. Figure 24 and Figure 25 are pie charts of the stage distribution at diagnosis from the natural history simulation and from SEER data, respectively. The percentage of cancers diagnosed in the local stage is 20% higher in the natural history model than in SEER data. The comparable percentage of cancers diagnosed in the regional stage was less by 13%, and the comparable percentage diagnosed in the distant stage was less by 4.5%. Clearly cancers are being diagnosed much earlier in the natural history simulation than they are in the actual population, which is expected with perfect annual screening. As would be expected, most cancers were diagnosed in the local stage, and very few were diagnosed in the regional and distant stages. However, cancers remain untreated in the natural history model; and thus the benefits of these earlier stages at detection cannot be determined until the execution of the screening-and-treatment simulation, when treatment is considered. We assert that this estimate of the stage distribution at diagnosis from the natural history simulation is a reasonable estimate of the stage distribution at diagnosis with perfect annual screening. It is impossible to actually know the stage distribution at diagnosis with perfect annual screening, but cancers should be detected much earlier than in the actual population, and the natural history model captures this expected behavior.

Table 13. Stage Counts and Distribution at Diagnosis: Results of Natural History Model vs. SEER Data for Comparison

| COUNTS   | Mean | HW  | LL  | UL  |
|----------|------|-----|-----|-----|
| Local    | 2191.2 | 29.1 | 2162.1 | 2220.3 |
| Regional | 410.5  | 6.3  | 404.2 | 416.8 |
| Distant  | 45.3   | 6.5  | 38.8 | 51.8 |

| DIST     | Mean | HW  | LL  | UL  |
|----------|------|-----|-----|-----|
| Local    | 82.78 | 0.38 | 82.40 | 83.16 |
| Regional | 15.51 | 0.32 | 15.19 | 15.83 |
| Distant  | 1.71  | 0.25 | 1.46 | 1.96 |

| SEER DIST | Mean | Sim Mean | Diff |
|-----------|------|----------|------|
| Local     | 29.5 | 85.027   | 55.5 |
| Regional  | 44.0 | 13.457   | -30.5|
| Distant   | 16.3 | 1.516    | -14.8|
| Unstaged  | 10.2 | -        | -    |
Another result of interest is the distribution of invasive cancer outcomes. There are three potential outcomes for a woman who is diagnosed with breast cancer in the natural history simulation: breast cancer death, death from a cause other than breast cancer, or survival through the year 2020. Of particular interest is the percentage of women with invasive cancers who die from a cause other than breast cancer. Since there is no treatment in the natural history simulation, women who are diagnosed with breast cancer and die from a cause other than breast cancer would not benefit from treatment. Ideally, we would like to maximize the percentage of women with invasive cancers who survive through 2020, and minimize the percentage of breast cancers resulting in death. These outcomes from the natural history simulation are later compared to outcomes from the screening-and-treatment simulation. Figure 26 is a pie chart of the distribution of invasive cancer outcomes, and Table 14 shows the 95% CI's on those distribution percentages.
**Figure 26.** Pie Chart of Distribution of Invasive Cancer Outcomes from Natural History Model

**Table 14.** 95% CIs on Distribution of Invasive Cancer Outcomes from Natural History Model

| Performance Measure                                      | Mean | HW  | LL  | UL  |
|-----------------------------------------------------------|------|-----|-----|-----|
| Percentage of Invasive Cancers Leading to BC Death        | 31.7 | 0.69| 31.0| 32.4|
| Percentage of Women with Invasive Cancers Still Surviving (2020) | 36.2 | 0.66| 35.5| 36.9|
| Percentage of Invasive Cancers Leading to Non BC Death    | 32.1 | 0.57| 31.5| 32.7|

**A4.5. 95% CIs on Tumor Growth-Rate Classifications**

As previously stated, there is evidence that suggests breast cancers grow more slowly in older women than they do in younger women. To account for this, the tumor growth-rate distribution has a mean that decreases with age, leading to slower tumor growth rates than were implied by the original distribution specified by Norton (1988). Tumor growth rates were divided into six classifications: very slow, slow, slightly slow, slightly fast, fast, very fast. Table 15 contains 95% CIs on the expected number of tumors in each class and the percentage of tumors in each class. Figure 27 is a pie chart for the percentage of tumors in each class. These results show that the majority of tumors are slow growing, and this is expected with an older population. This also implies that the majority of tumors are detected at small sizes with perfect annual screening, and again this is something that is expected.
Table 15. Phase I Tumor Growth Rate Counts and Distribution CI Data After Age Adjustment

|                  | COUNTS   |   |   |   |
|------------------|----------|---|---|---|
|                  | Mean     | HW| LL| UL|
| Very Slow        | 1376.5   | 34.19| 1342.31 | 1410.69 |
| Slow             | 557.2    | 12.58| 544.62   | 569.78  |
| Slightly Slow    | 419.3    | 16.26| 403.04   | 435.56  |
| Slightly Fast    | 198.2    | 5.48 | 192.72   | 203.68  |
| Fast             | 62.7     | 7.50 | 55.20    | 70.20   |
| Very Fast        | 20.0     | 3.92 | 16.08    | 23.92   |

|                  | DISTRIBUTION | Mean | HW | LL | UL |
|------------------|--------------|------|----|----|----|
| Very Slow        | 52.25        | 0.75 | 51.5038 | 53.0038 |
| Slow             | 21.16        | 0.52 | 20.6418 | 21.6818 |
| Slightly Slow    | 15.92        | 0.54 | 15.3781 | 16.4581 |
| Slightly Fast    | 7.53         | 0.23 | 7.2978   | 7.7578  |
| Fast             | 2.38         | 0.27 | 2.1076   | 2.6476  |
| Very Fast        | 0.76         | 0.15 | 0.6108   | 0.9108  |

Figure 27. Phase I Categorical Distribution of Tumor Growth Rate for Women over 65

A4.6. Inferred Distributions of Interest

The Stat::Fit 2.0 statistical distribution fitting software (Geer Mountain Software Corporation, 2012) was used to fit probability distributions to cancer attributes that may be of interest to clinicians and other researchers. In all ten simulated populations, each woman who was diagnosed with cancer produced a cancer history; and in addition, certain cancer attributes of interest were written to a separate text file that could easily be loaded into Stat::Fit for analysis. The following criteria were used to evaluate the fit of alternative distributions to data: visual inspection of the fitted c.d.f.; visual inspection of the fitted proba-
bility density function (p.d.f.); a P-P plot of the fit; \( p \)-values for the chi-squared and Kolmogrov-Smirinov goodness-of-fit tests; and the ability of the fitted distribution to capture the first four sample moments. In this section we simply present Table 5, which contains the list of cancer attributes to which we fitted distributions, the best-fitting distribution in each individual case, and the parameters of the best-fitting distribution. For more detailed information, see the Online Supplement and Appendix D of Tejada (2012).

### Table 5. Inferred distributions based on data generated by natural-history simulation

| Performance Measure                                      | Distribution | Parameters                               |
|-----------------------------------------------------------|--------------|------------------------------------------|
| Time for Tumors to Grow to Minimum Size Detectable by Mammography (months) | Lognormal    | \( \mu = 38.6, \sigma = 29.4, \text{Min (Offset)} = 4.0 \) |
| Time for Tumors to Become Clinically Detectable or Symptomatic (months) | Pearson 6    | \( \beta = 106.4, p = 2.8, q = 4.7, \text{Min (Offset)} = 3.0 \) |
| Time Until Tumors to Become Lethal (months)               | Pearson 6    | \( \beta = 235.4, p = 2.82, q = 4.81, \text{Min (Offset)} = 7.0 \) |
| Size When Tumor Becomes Detectable by Mammography (mm)    | Lognormal    | \( \mu = 2.27, \sigma = 0.639, \text{Min (Offset)} = 1.0 \) |
| Size When Tumor Becomes Clinically Detectable (mm)        | Beta         | \( \alpha_1 = 2.27, \alpha_2 = 0.639, \text{Min (Offset)} = 1.0 \) |
| Cancer Onset Age for Women 65 and Older                   | Pearson 6    | \( \beta = 383.7, p = 23.8, q = 248.6, \text{Min (Offset)} = 36.0 \) |
| Tumor Growth Rate for Women 65 and Older                  | Lognormal    | \( \mu = -3.84, \sigma = 0.75, \text{Min (Offset)} = 0.0 \) |

### References

Barlow, W.E., White, E., Ballard-Barbash, R., Vacek, P.M., Titus-Ernstoff, L., Carney, P.A., Tice, J.A., Buist, D.S.M., Geller, B.M., Rosenberg, R., Yankaskas, B.C. and Kerlikowske, K.  (2006) Prospective breast cancer risk prediction model for women undergoing screening mammography. *Journal of the National Cancer Institute, 98*(17), 1204–1214.

Bloom, H., Richardson, M. and Harries, B. (1962). Natural history of untreated breast cancer (1804–1933): Comparison of treated and untreated cases according to histological grade of malignancy. *British Medical Journal, 2*, 213-221.

Box, G.E.P., Hunter, J.S. and Hunter, W.G. (2005) Statistics for Experimenters: Design, Innovation, and Discovery, second edition, John Wiley & Sons, New York.

Brannick, M. T. (2007). *Logistic Regression*. Available at: [http://luna.cas.usf.edu/~mbrannic/files/logistic/Logistic.html](http://luna.cas.usf.edu/~mbrannic/files/logistic/Logistic.html) (accessed October 12, 2011).

Breast Cancer Surveillance Consortium (BCSC) (2006) Risk Model Dataset Documentation. National Cancer Institute, Bethesda, MD. Available at: [http://breastscreening.cancer.gov/rfdataset/rmdataset_docs.pdf](http://breastscreening.cancer.gov/rfdataset/rmdataset_docs.pdf) (accessed October 21, 2012).
Breast Cancer Surveillance Consortium (BCSC) (2009) Performance measures for 1,960,150 screening mammography examinations from 2002 to 2006 by age—based on BCSC data as of 2009. National Cancer Institute, Bethesda, MD. Available at: http://breastscreening.cancer.gov/data/performance/screening/2009/perf_age.html (accessed October 21, 2012).

Carter, C. L., Allen, C., & Henson, D. E. (1989). Relation of Tumor Size, Lymph Node Status, and Survival in 24,740 Breast Cancer Cases. Cancer, 63, 181-187.

Cook, N. R. (2007). Use and Misuse of the Receiver Operating Characteristic Curve in Risk Prediction. Circulation, 115, 928-935.

Crivellari, D., Aapro, M., Leonard, R., von Minckwitz, G., Brain, E., Goldhirsch, A., Veronesi, A. and Muss, H. (2007) Breast cancer in the elderly. Journal of Clinical Oncology, 25(14), 1882–1890.

Diab, S.G., Elledge, R.M. and Clark, G.M. (2000) Tumor characteristics and clinical outcome of elderly women with breast cancer. Journal of the National Cancer Institute, 92(7), 550–556.

Diamond, G. A. (1992). What price perfection? Calibration and discrimination of clinical prediction models. Journal of Clinical Epidemiology, 45(1), 85–89.

Downey, L., Livingston, R. and Stopeck, A. (2007) Diagnosing and treating breast cancer in elderly women: A call for improved understanding. Journal of the American Geriatrics Society, 55, 1636–1644.

Gail, M.H. and Pfeiffer, R.M. (2005) On criteria for evaluating models of absolute risk. Biostatistics, 6(2), 227–239.

Gail, M. H., Brinton, L. A., Byar, D. P., Corle, D. K., Green, S. B., Schairer, C., & Mulvihill, J. J. (1989). Projecting Individualized Probabilities of Developing Breast Cancer for White Females Who Are Being Examined Annually. Journal of the National Cancer Institute, 81(24), 1879-1886.

Hoel, P. G. (1984). Introduction to Mathematical Statistics (5th ed.). New York: Wiley.

Hosmer, D.W., Lemeshow, S., and Sturdivant, R. X. (2013) Applied Logistic Regression, third edition, John Wiley & Sons, Inc., New York.

Kelton, W. D., Sadowski, R. P., & Sturrock, D. T. (2010). Simulation with Arena, fifth edition. McGraw-Hill, New York, NY.

Mandelblatt, J.S., Wheat, M.E., Monane, M., Moshief, R.D., Hollenberg, J.P. and Tang, J. (1992) Breast cancer screening for elderly women with and without comorbid conditions. Annals of Internal Medicine, 116, 722–730.

McManus, M. J., & Waelsch, C. (1980). DNA Synthesis of Benign Human Breast Tumors in Untreated "Nude" Mouse: An In Vivo Model to Study Hormonal Influences on Growth of Human Breast Tissues. Cancer, 25, 2160-2165.

Meads, C., Ahmed, I. and Riley, R. (2012). A systematic review of breast cancer incidence risk prediction models with meta-analysis of their performance. Breast Cancer Research and Treatment, 132, 365 - 377.
Michaelson, J. S., Halpern, E., & Kopans, D. B. (1999). Breast Cancer: Computer Simulation Method for Estimating Optimal Intervals for Screening. *Radiology*, 551-560.

National Cancer Institute. (1970-2007). *Cancer Query Systems*. Available at: [www.seer.cancer.gov](http://www.seer.cancer.gov) (accessed March 21, 2011).

Norton, L. (1988). A Gomertzian Model of Human Breast Cancer Growth. *Cancer research*, **48**, 7067-7071.

Pearlman, A. W. (1976). Breast Cancer: Influence of Growth Rate on Prognosis and Treatment Evaluation. *Cancer*, **38**, 1826-1833.

Peer, P.G.M., van Dijck, J.A.A.M., Hendriks, J.H.C.L., Holland, R., Verbeek, A.L.M. (1993) Age-dependent growth rate of primary breast cancer. *Cancer*, **71**(11), 3547–3551.

Plevritis, S. K., Salzman, P., Sigal, B. M., & Glynn, P. W. (2007). A Natural History Model of Stage Progression Applied to Breast Cancer. *Statistics in Medicine*, **26**, 581-595.

Rae-Venter, B., & Reid, L. M. (1980). Growth of Breast Cancer Carcinomas in Nude Mice and Subsequent Establishment in Tissue Culture. *Cancer Research*, **40**, 95-100.

Rockhill, B., Spiegelman, D., Byrne, C., Hunter, D. J., & Colditz, G. A. (2001). Validation of the Gail et al. Model of Breast Cancer Risk Prediction and Implications for Chemoprevention. *Journal of the National Cancer Institute*, **93**(5), 358-366. Available at: [http://jnci.oxfordjournals.org/content/93/5/358.full.pdf](http://jnci.oxfordjournals.org/content/93/5/358.full.pdf) (accessed November 21, 2013).

Scott, D. W. (1979) On optimal and data-based histograms. *Biometrika*. **66**, 605–610.

Shumate, S. D., & El-Shenawee, M. (2007). Computational Model of Breast Cancer Tumor Growth. *IEEE Region 5 Technical Conference*, 13-15.

Speer, J. F., Petrosky, V. E., Retsky, M. W., & Wardwell, R. H. (1984). A Stochastic Numerical Model of Breast Cancer Growth that Simulates Clinical Data. *Cancer Research*, **44**, 4124-4130.

Surborne, A., & Norton, L. (1993). Kinetic Concepts in the Treatment of Breast Cancer. *Annals of NY Academic Science*, **698**, 48-62.

Tejada, J.J. (2012) An integrated discrete-event/system dynamics simulation model of breast cancer screening for older US women. Ph.D. dissertation, Edward P. Fitts Department of Industrial and Systems Engineering, North Carolina State University, Raleigh, NC. Available at [http://repository.lib.ncsu.edu/ir/bitstream/1840.16/7935/1/etd.pdf](http://repository.lib.ncsu.edu/ir/bitstream/1840.16/7935/1/etd.pdf) (accessed October 21, 2012).

Tejada, J.J., Ivy, J.S., King, R.E., Wilson, J.R., Ballan, M.J., Kay, M.G., Diehl, K. and Yankaskas, B.C. (2014). Combined DES/SD model of breast cancer screening for older women, II: screening-and-treatment model. *IIE Transactions*, **46**(7), 707–727.

Tejada, J.J., Ivy, J.S., Wilson, J.R., Ballan, M.J., Kay, M.G., Diehl, K. and Yankaskas, B.C. (2013). Combined DES/SD model of breast cancer screening for older women, I: natural-history simulation. *IIE Transactions*, forthcoming.

Weiss, M. (2008). *Understanding Breast Cancer*. Available at: [http://www.breastcancer.org/symptoms/understand_be/](http://www.breastcancer.org/symptoms/understand_be/) (accessed March 3, 2009).
Data collection for this work was supported by the National Cancer Institute-funded Breast Cancer Surveillance Consortium co-operative agreement (U01CA63740, U01CA86076, U01CA86082, U01CA63736, U01CA70013, U01CA69976, U01CA63731, U01CA70040). The collection of cancer data used in this study was supported in part by several state public health departments and cancer registries throughout the U.S. For a full description of these sources, please see: http://www.breastscreening.cancer.gov/work/acknowledgement.html.