**Important note:** This document will remain archived as a technical appendix for publications. New versions will be added periodically as model refinements and updates are completed. The most current version is available at [http://cisnet.cancer.gov/profiles](http://cisnet.cancer.gov/profiles). Note that unlike most PDF documents, the CISNET model profiles are not suitable for printing as they are not typically written or read in sequential fashion.

We recommend you let your interests guide you through this document, using the navigation tree as a general guide to the content available.

The intent of this document is to provide the interested reader with insight into ongoing research. Model parameters, structure, and results contained herein should be considered representative but preliminary in nature.

We encourage interested readers to contact the contributors for further information.

Go directly to the: **Reader's Guide**.
READERS GUIDE

Core Profile Documentation
These topics will provide an overview of the model without the burden of detail. Each can be read in about 5–10 minutes. Each contains links to more detailed information if required.

Model Purpose
This document describes the primary purpose of the model.

Model Overview
This document describes the primary aims and general purposes of this modeling effort.

Assumption Overview
An overview of the basic assumptions inherent in this model.

Parameter Overview
Describes the basic parameter set used to inform the model, more detailed information is available for each specific parameter.

Component Overview
A description of the basic computational building blocks (components) of the model.

Output Overview
Definitions and methodologies for the basic model outputs.

Results Overview
A guide to the results obtained from the model.

Key References
A list of references used in the development of the model.

Further Reading
These topics will provide an intermediate level view of the model. Consider these documents if you are interested gaining in a working knowledge of the model, its inputs and outputs.

JNCIMonograph Outline
This topic provides links to profile content organized according to the JNCI Monograph Outline for Model Description Chapters. Use this outline for comparisons focused on the CISNET Base Case simulations.

Advanced Reading
These topics denote more detailed documentation about specific and important aspects of the model structure.
MODEL PURPOSE

SUMMARY
Here we describe the historical and current purposes of our model of breast cancer.

PURPOSE
Historically, this model was developed to carry out cost–effectiveness analysis of various approaches to breast cancer screening. Among the questions originally addressed were: what are the costs and benefits associated with various population–based interventions aimed at increasing the utilization of mammography in minority populations; what are the costs and benefits of continuing to screen elderly and very elderly women?

The original development of the model focused on simulating the experience of a hypothetical cohort of women of a certain age. For the CISNET project, the inner core of the simulation logic is "re–packaged" to simulate the entire female population of the United States between 1975 and 2000.

The logic of the model is based on the distributions of times to various events in the life of a simulated subject. There are no particular assumptions made about the mechanism by which breast cancer progresses or kills people. As such this model cannot be effectively used to test hypotheses about breast cancer biology, nor to calibrate the parameters of mechanistic models of breast cancer.
MODEL OVERVIEW

SUMMARY
This document describes, at the broadest level, the method by which we simulate the US population incidence and mortality from breast cancer between 1975 and 2000.

PURPOSE
This model, originally developed to perform cost–effectiveness analysis of breast cancer screening programs, has been adapted to simulate the incidence and mortality of breast cancer in the US population between 1975 and 2000. In particular, it is designed to estimate the effects of screening and treatment improvements during that era.

BACKGROUND
Interesting changes in the morbidity and mortality associated with breast cancer have occurred over the past 25 years, but little is known about the causes of this. While at one time mammography was almost universally agreed to be effective in reducing breast cancer mortality, more recent reviews of the original clinical trials have raised serious questions about this. As further large scale trials of mammography are unlikely to be conducted in the near future, it would be helpful if careful analytic approaches could disentangle the effects of increased mammography utilization, improvements in the efficacy of treatment, and other changes in the population.

This model builds on the basic breast–cancer simulation developed and used by this group for cost–effectiveness analysis of screening programs.

MODEL DESCRIPTION
We use an event–driven continuous–time state transition model. Women from different birth cohorts are simulated one at a time, and the times at which relevant events occur are determined by sampling from pre–specified time–interval distributions. We simulate 55 million women to obtain reasonably smooth estimates of the mortality curves. Using US Census data, we begin with women born in or after 1890 to simulate the population distribution of adult women alive in 1975. Women who are destined to develop breast cancer may either be screen detected, present with clinical symptoms, or die of other causes before breast cancer is diagnosed. At presentation, the cancer has a stage assigned, based on whether the tumor is screen or clinically detected. The stage for screen–detected cancers is calculated from what the stage would have been had the tumor presented with symptoms and the lead time gained from screening using a formula derived from Bayes’ theorem. Cancers are designated as being estrogen–receptor (ER) positive or negative. Survival is conditional on age and stage at diagnosis, ER status, and treatment.

Model inputs include:

- age distribution of US women in 1975, age and year–specific projections of breast cancer incidence in the absence of screening (from an age–period cohort model generated by NCI),
• birth–year specific annual US female mortality from all causes other than breast cancer (from the Berkeley mortality database, as modified by M. Rosenberg),
• age–specific distributions of stages of cancers diagnosed clinically (taken from SEER data in 1975),
• age–specific distributions of stages of cancers diagnosed through screening (taken from SEER data in the 1990's),
• sensitivity of mammography screening by age,
• mean tumor sojourn time by age,
• mean tumor dwell time in each clinical stage (DCIS, local, regional, distant),
• age and calendar–year estimates of the pattern of mammography utilization (provided by NCI),
• age–stage–ER specific distributions of treatment choices in different calendar years (provided by NCI),
• age–stage–ER specific breast cancer survival curves
• estimates of the odds–ratios of survival associated with use of adjuvant tamoxifen and adjuvant chemotherapy.

For each woman, the model produces a life history that identifies whether or not a diagnosis of breast cancer is made, and if so, in what stage it presents, what treatment was chosen, as well as a date of death and an indication of whether death is from breast cancer or other causes. The total number of mammography screenings is provided, as is a count of how many such screenings were positive. These life histories are then summarized to produce annual estimates of breast cancer incidence and mortality grouped by decade of age.

Key limitations of the model are that it does not allow for any effect of early detection unless a stage shift results, and that it assumes that all breast cancers (including ductal carcinoma in situ) are progressive. The latter limitation is of particular importance.

CONTRIBUTORS
Jeanne Mandelblatt, MD, MPH
Aimee M. Near, MPH
Clyde B. Schechter, MA, MD
Michael A. Stoto, PhD

All from Georgetown University, except Dr. Schechter who is at Albert Einstein College of Medicine
ASSUMPTION OVERVIEW

SUMMARY
This document gives a broad overview of the assumptions inherent in the model.

BACKGROUND
This model makes no assumptions about the biological mechanisms of breast cancer progression and mortality. Indeed, the model could be used without further modification to simulate any multi–stage failure process for which the appropriate inputs (in particular process step time distributions) were available.

A number of assumptions about the mechanism by which mammography affects the natural history of breast cancer are detailed below.

ASSUMPTION LISTING
This model relies on the following assumptions:

1. The benefit of mammography screening is exactly represented by the effects of shifting diagnosis to an earlier stage of disease. Early detection that is not early enough to result in detection at an earlier stage does not, on average, alter survival. Furthermore, earliness of detection which does result in an earlier stage at diagnosis is "rewarded" with the full difference in stage–specific survivals.

2. In the absence of screening, the distribution of stages of clinically detected tumors would resemble the distribution of stages of clinically detected tumors from the early part of the 1975–2000 era.

3. Breast cancer progresses from a pre–clinical stage to a clinical presence, and then through stages of local, regional, and distant spread. The dwell times in each stage are assumed to have an exponential distribution. All tumors, including all ductal carcinomas in situ, have the potential to progress to metastatic disease and cause death.

4. Dwell times in each successive stage are independent of each other and of the sojourn time.
PARAMETER OVERVIEW

SUMMARY
Focusing on those input parameters which are not common to all of the CISNET models, we describe the sources of our parameters concerning the stage–progression and sojourn time of breast cancer, as well as mammography operating characteristics.

BACKGROUND
Publications which report estimates for the dwell time in any stage of breast cancer, the sojourn time, or the sensitivity and specificity of mammography were reviewed. Studies not carried out in the industrialized world were excluded, as were studies carried out in highly restrictive populations. We did not distinguish studies which fit statistical models to screening data from studies which, in one way or another, directly observed the particular parameter. The median value of reported estimates was initially used as our base case parameter value. Because simulation results with these estimates showed a shortfall in predicted incidence which increased as screening disseminated, we experimented with other values to try to match the observed US population incidence curves. This resulted in selecting values of sensitivity and sojourn time which are about 2/3 of the way between the lowest and highest values in our literature reviews.

PARAMETER LISTING OVERVIEW
The parameters discussed here pertain only to the natural history of breast cancer component of the model. All of these parameters are used to calculate lead time, and to simulate stage at diagnosis with screening.

The parameters are:

---

Sojourn time (mean of exponential distribution)—this is taken to be age dependent.

<table>
<thead>
<tr>
<th>Age</th>
<th>Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>50–54</td>
<td>2.1y</td>
</tr>
<tr>
<td>55–60</td>
<td>3.3y</td>
</tr>
<tr>
<td>60–69</td>
<td>3.9y</td>
</tr>
<tr>
<td>70+</td>
<td>5.2y</td>
</tr>
</tbody>
</table>

Dwell time as DCIS or in local and regional stages (mean of exponential distributions).

<table>
<thead>
<tr>
<th>Stage</th>
<th>Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>DCIS</td>
<td>2.97y</td>
</tr>
<tr>
<td>LOCAL</td>
<td>5.30y</td>
</tr>
<tr>
<td>REGIONAL</td>
<td>11.40y</td>
</tr>
</tbody>
</table>
Sensitivity of initial mammography:

<table>
<thead>
<tr>
<th>Age</th>
<th>Sens</th>
</tr>
</thead>
<tbody>
<tr>
<td>0-49</td>
<td>0.77</td>
</tr>
<tr>
<td>50-59</td>
<td>0.87</td>
</tr>
<tr>
<td>60-69</td>
<td>0.94</td>
</tr>
<tr>
<td>70</td>
<td>0.91</td>
</tr>
</tbody>
</table>

Sensitivity of subsequent mammography:

0.85 at all ages.

Note that in the studies used to derive our base case values, different definitions of stages, or sensitivity have been used. We have ignored these differences and pretended that all studies were estimating the same parameter.
COMPONENT OVERVIEW

SUMMARY
This document gives a broad view of the key components in the model.

OVERVIEW
This model overall includes 4 processes:

1. Subject generation. This part of the model generates simulated women to simulate the age distribution of US women in 1975, their age–specific breast-cancer incidence, and their overall mortality experience.
2. Screening simulation. This part of the model generates a screening schedule for each woman and determines whether it results in the early detection of a breast cancer, and if so, in what stage.
3. Course of Disease. This part of the model identifies what treatment approach the simulated woman will undergo and projects subsequent survival.
4. Bookkeeping. This part of the model doesn't simulate anything—it tallies the results as successive women are simulated.

All of these components operate under the "orchestration" of a general simulation engine. The simulation engine maintains a chronologically ordered event queue. This queue is re–initialized at the start of each simulated subject's processing. The queue contains events such as "gets clinically evident breast cancer," "has next mammogram," "dies of cause other than breast cancer," etc.

COMPONENT LISTING
Population Component
The demographic component generates a population of simulated women having the age distribution of the female population of the United States in 1975. Using SEER
data, the breast cancer incidence component randomly selects which simulated women will develop breast cancer, with what estrogen receptor status, and at what time and in what stage, if the cancer presents clinically.

**Natural History Component**

The screening impact component governs the performance characteristics of screening, including screening test sensitivity and specificity. This portion of the model also calculates a stage shift for the tumor conditional on the lead-time realized by the screening test that detects it.

**Screening Component**

The screening utilization component determines when simulated women undergo breast cancer screening based on a model of the observed diffusion through the population between 1975 and 1999.

**Treatment Component**

The treatment component is activated whenever a tumor is diagnosed (clinically or by screening) and selects a treatment and a corresponding breast-cancer survival time based on SEER data for age, stage, estrogen receptor status, and treatment-specific survival. Competing mortality is estimated using actuarial methods.
OUTPUT OVERVIEW

SUMMARY
This document describes the general types and forms of output from the simulator.

OVERVIEW
There are two major components to the simulation output.

The first component is tallies of incident breast cancers, breast cancer deaths, and total simulated female midyear population by age (single years) and calendar year between 1975 and 2000. (The simulation generates breast cancers and deaths outside the 1975–2000 window of interest, but these are suppressed and excluded from this component of the output.) The incident breast cancer figures are further disaggregated by stage and ER status. These figures are used to calculate agegroup specific incidence and mortality rates.

The second component is a simulated "cancer registry." In this segment of the output, a record is created for each simulated woman with breast cancer which shows are date of birth, date of diagnosis, stage at presentation, ER status, treatment, date of death, and indication of whether death is from breast cancer or not. A unique feature of this "registry" not available in real life is an entry for the date at which the tumor would have presented clinically (= actual date of diagnosis if tumor was not screen detected.) The registry also summarizes the woman's screening history by including the total number of true positive, true negative, positive, and negative mammograms she underwent during her lifetime. These data are used to calculate survival curves for the simulated cancers, and also to estimate mammogram program sensitivity and lead–time distribution. Note that a simulated woman who dies prior to 1975 is not recorded in this "cancer registry," but events occurring outside the 1975–2000 window which take place during the lifetime of a woman who remains alive at any time in this window are recorded in the "cancer registry."
DETAIL

Demographic Component of the Model
Each birth year is selected in our model with a frequency proportional to its prevalence among the US female population in 1975 and inversely proportional to the probability of survival to 1975. This ensures that the 1975 age–distribution we simulate matches that of the given 1975 US female population. Each simulated person’s life is modeled from birth, including the application of cancer incidence functions. Thus, a woman may develop breast cancer before 1975, and if she does not die before 1975, she will be a prevalent case at the start of the model. Women born between 1890 and 1975 but who die before 1975 are also simulated, but we do not include their data in the output.

REFERENCES:

1 Woods & Poole Economic Inc “2001 Regional database Estimated July 1 population by race, sex and single year and 5-year age groups based on 1990 census and post censal census bureau estimates” 2001;
The model makes no explicit assumptions about the biological nature of breast cancer. Rather, all aspects of breast cancer are modeled in terms of stage (SEER historical stages of in–situ, local, regional, and distant), estrogen–receptor (ER) status (positive or negative), and the age of the woman at diagnosis, and treatment selected. This implies, in particular, that any effect of screening on survival is the result of stage–shift (and, to a lesser extent, age–shift in presentation). Screen–detected lesions are assigned the same ER status they would have had if they had presented clinically.¹

Projected age–specific incidence rates in the absence of screening for each birth cohort from 1890 through 1970 were provided by the National Cancer Institute, and include secular trends in incidence for each birth–cohort.² These incidence rates were estimated using an age–period–cohort model which is described elsewhere.³ These incidence rates are used as the hazard in a survival process, where failure consists of incident breast cancer. The corresponding survival function is sampled for each woman, given her year of birth, to determine when she will develop clinical breast cancer. Because the survival function does not go to zero, or even near zero, the majority of women will never develop breast cancer.

For those women for whom a date of incident clinical breast cancer is ascertained, a preclinical sojourn time is also simulated. Sojourn times are assumed to be exponentially distributed, with an age–dependent mean, based on published data.⁷ The appropriate distribution is sampled to determine the preclinical sojourn time. The screening module then determines the actual date of preclinical incidence (if it occurs).

We assume that the dwell time in each stage is exponentially distributed, with mean stage dwell times as input to the program. Dwell times (e.g., from DCIS to invasive cancer, from local to regional disease, or from regional to distant) were estimated based upon data from randomized clinical trials of breast cancer screening,¹² and simulating stage distributions in screened and unscreened settings (personal communication, William Lawrence, 2002).

When a tumor is diagnosed by screening, the lead–time is calculated. The stage at which the tumor would have presented clinically is “known” within the simulation. The conditional probability that a tumor in any given stage would have progressed to that known stage in the obtained lead–time is therefore calculated by convoluting the exponential stage dwell time distributions. A "prior" distribution for stage at screening is taken from the observed distribution of stages among tumors diagnosed recently (personal communication, Breast Cancer Surveillance Consortium, Diane Miglioretti and William Barlow, 2002).¹³ Bayes' theorem is then applied to calculate a "posterior" distribution of stage at screening conditional on the stage at clinical presentation and lead–time. This posterior distribution is then sampled to identify the simulated stage at screen–detection.

One implication of the lead–time is that a woman who is screen–detected several years
earlier than she would have presented clinically may end up getting less intensive treatment because the more intensive treatment, such as multi–drug chemotherapy regimens, were not yet in sufficiently widespread use. Such a woman could actually end up with a worse prognosis as a result of screening, because she got her diagnosis in an earlier era when chemotherapy was not being widely used. While such events do occur in our model, they occur with a low frequency and probably do not have a substantial impact on the results.

We only model the incidence of first breast cancers. Accordingly, the correct denominator for an incidence rate should be the number of women alive who have never had breast cancer. However, in compliance with the procedures adopted by the CISNET collaboration for calculating incidence rates, we actually use the count of all women alive for this denominator. This approach results in a slight underestimate of the incidence rates. The extent of this underestimate increases with a woman’s age, reflecting both the rising rates of breast cancer and the falling surviving population denominator, and tends to increase over time among those over age 50. The underestimate of incidence never exceeds 1% for women under age 50, and only reaches 5.4% for women ages 75 to 79 after 1994.

The preclinical sojourn time is one of the “tunable” parameters of our model. That is, unlike, for example, incidence rates which are directly observable and for which excellent data exist, the sojourn time is a latent variable, and can only be estimated by fitting models to population screening programs. Thus, we varied our estimate within the range of published estimates so as to generate simulated incidence and stage–distribution in screened women that best corresponded to observed incidence.

This calibration was performed using a small number of simulations (5 million per woman). Sojourn time was calibrated together with test sensitivity (see below). Results were inspected for face validity and the final combination selected based on the most reasonable combination of values for each parameter that estimated the observed incidence and stage–distribution as closely as possible.

REFERENCES:

2. CISNET “Female breast cancer incidence rates SEER 9 registries” 2004;
3. CISNET “Breast base case age-specific secular trend parameter” 2004;


10 Chamberlain, J, Coleman, D, Moss, et al “Sensitivity and specificity of screening in the UK trial of early detection of breast cancer’’ 1991; : 3-17


SCREENING COMPONENT

OVERVIEW

Screening Utilization Component of the Model
We use existing data from a model of screening use over time to reflect the dissemination of mammography in the US population. These data were based on fitting parametric frailty models to national screening data. However, these data only covered patterns occurring for women born between 1891 and 1970. Women born in 1890 were assumed to obtain screening at the same ages as women born in 1891. Women born after 1970 were assumed to obtain screen at the same ages as women born in 1970.

Screening Impact Component of the Model
Each screening event (i.e., obtaining a mammography in a given year or not) is simulated by drawing a random number from a uniform distribution between zero and one. If screening occurs the test sensitivity and specificity and the presence or absence of a tumor during its preclinical sojourn time are used to generate a test result (true positive, false positive, true negative, or false negative). True positive test results trigger a diagnosis of breast cancer and calculation of the stage at presentation (and assignment of an ER status). We assume that screen detected and clinically detected interval cancers (false negatives and clinically detected in the absence of screening) have similar tumor characteristics (i.e., distribution of ER) and that conditional on age and stage at diagnosis, ER status, and treatment, they have the same survival functions. To the extent that screen detected tumors are less virulent than interval cases, then mortality reductions associated with screening may be slightly over-estimated.

DETAIL

Mammogram Sensitivity
The sensitivity of mammography is the other "tunable" parameter in our model. In our program, sensitivity is a ratio, with the numerator consisting of positive test results among those with a tumor, and the denominator consisting of those with a tumor. In our model, "with a tumor" is implemented as "occurring during the preclinical sojourn time of a lesion." "With a tumor" therefore is an abstract, unobservable construct whose value cannot be directly measured but can be estimated by fitting statistical models to the data from large screening programs. As a starting point, we relied on published age-specific estimates of sensitivity from different points in time. We assume that sensitivity is greatest for the first screen, and then decreases over time with repeated screenings, but we do not vary the sensitivity according to tumor size or tumor growth over the preclinical sojourn time. Sensitivity is assumed to be age-dependent for the first screening, but age-independent thereafter. We made this choice because there were good data on test performance as a function of age for first screening examinations, but less data on the results for subsequent screens over time by age. We also model test sensitivity as a constant over the period of simulation. While test performance is likely to have improved over time between 1975 and 2000, there was sufficient variability in published estimates from large screening trials by time period that no reasonable time-period-dependent curve could be fit to the observed data.
REFERENCES:

1 CISNET “Breast base case mammography dissemination parameter” 2004;
TREATMENT COMPONENT

Treatment and Survival Component

Adjuvant treatments gradually disseminated into practice after 1975. We used data from Mariotto and colleagues to estimate the dissemination of non-hormonal chemotherapy and tamoxifen. Since the two main surgical options – mastectomy and breast conservation – have equivalent survival we do not include any changes in local treatment approaches over time. For cancers diagnosed before 1975, the 1975 treatment distributions were used.

Data from 1975 were used to estimate survival in the absence of adjuvant treatment with multi-agent chemotherapy or tamoxifen. Women receiving tamoxifen or chemotherapy were assigned a survival time based on a modification of the 1975 survival curve using data from large meta-analyses. For each therapy, the survival function for the base 1975 data is adjusted using the annual reduction in the odds of death associated with each modality. We then sample from the modified survival function to project survival given each therapy. Only women with ER positive tumors are assumed to have survival benefits associated with tamoxifen. For ER positive women receiving both tamoxifen and adjuvant chemotherapy, the two odds ratios are multiplied. This, in effect, assumes that the two treatments are neither synergistic nor interfering.

Because survival is calculated conditional on age at diagnosis, stage at diagnosis, ER status, and treatment, stage shifts can result in improved prognosis. We calculate survival from the date of clinical presentation, even if the lesion was screen detected. As a consequence, death from breast cancer cannot occur during the lead-time. Death from other causes, however, can occur in the lead time. We do not present quality-adjusted survival, as the base model was designed to estimate the potential impact of screening and treatment on observed incidence and mortality in the time period of interest.

Competing Mortality Component

Death from causes other than breast cancer was estimated using birth cohort-specific annual mortality data.

REFERENCES:

1 CISNET “Breast base case treatment dissemination parameter” 2004;
5 CISNET “Breast base case 1975 cause-specific survival parameter”
8 CISNET “Breast base case treatment effect parameter”
9 CISNET “Competing risks 2004” 2004;
KEY REFERENCES

(2004) Breast cancer facts and figure in *American Cancer Society*,
Brekelmans, CTM, Westers, P, Faber, JAJ, Peeters, PHM, Collette, HJA (1996) Age specific sensitivity and sojourn time in a breast cancer screening programme (DOM) in the netherlands: a comparison of different methods in *J Epidem Commun Hlth* 50, p 68-71
CISNET Breast base case 1975 cause-specific survival parameter,
CISNET (2004) Breast base case age specific incidence in the absence of screening,
CISNET (2004) Breast base case age-specific secular trend parameter,
CISNET (2004) Breast base case treatment dissemination parameter,
CISNET (2004) Breast base case mammography dissemination parameter,
CISNET Breast base case treatment effect parameter,
CISNET (2004) Female breast cancer incidence rates SEER 9 registries,
CISNET (2004) Proportion of women receiving breast cancer adjuvant therapy by age, SEER stage and year,
CISNET (2004) Stage Distribution for Female Breast Cancer Cases Diagnosed in SEER 1975-1979, by Age,


Chu, KC, Tarone, RE, Brawley, OW (1999) Breast cancer trends of black women compared with white women in Arch Fam Med 8:6, p 521-528


National Center for Health Statistics, National Cancer Institute (2004) Groups at high risk for cancer deaths,


Soldo, BJ, Agree, EM (1988) America’s elderly,


US Census Bureau, Projections of the resident population by race (2003)
Hispanic origin and nativity: Middle series 2025 to 2045,

Woods & Poole Economic Inc (2001) 2001 Regional database Estimated July 1 population by race, sex and single year and 5-year age groups based on 1990 census and post censal census bureau estimates,