



FLEXKB DOCUMENT
Version: HI.001.07232013.58567
Document generated: 07/23/2013

UNIVERSITY OF WISCONSIN

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>. 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.



[Readers Guide](#)
[Model Overview](#)
[Assumption Overview](#)
[Parameter Overview](#)
[Component Overview](#)
[Output Overview](#)
[Results Overview](#)
[Key References](#)

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.



Readers Guide
Model Overview
Assumption Overview
Parameter Overview
Component Overview
Output Overview
Results Overview
Key References

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 a intermediate level view of the model. Consider these documents if you are interested gaining 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

The Wisconsin model simulates breast cancer in a population over time generating cancer registry-like data sets. By manipulating parametric input assumptions about natural history, screening, and treatment the model can be used to address a number of important policy questions.



[Readers Guide](#)
[Model Overview](#)
[Assumption Overview](#)
[Parameter Overview](#)
[Component Overview](#)
[Output Overview](#)
[Results Overview](#)
[Key References](#)

PURPOSE

The purpose of the NCI Cancer Intervention and Surveillance Modeling Network (CISNET) is to promote simulation modeling as a tool that, in conjunction with the nation's cancer surveillance systems, can help to explain observed changes in cancer incidence and mortality. Wisconsin offers a unique population laboratory to develop and test breast cancer simulation models to meet this goal. We propose a collaboration among simulation and statistics experts, and surveillance and epidemiology experts at the University of Wisconsin, and the state of Wisconsin's Cancer Reporting System to study the use of simulation modeling to better understand trends in breast cancer epidemiology and to enhance the use of simulation modeling for this purpose.

The Wisconsin model evolved from a simulation model constructed by Polun Chang a decade ago for his Ph.D. dissertation¹. Chang asked whether the observed breast cancer incidence and mortality in the state of Wisconsin over the years 1982-92 could be represented by a mechanistic simulation model comprised of reasonable sub-models of population demography, biologic onset and progression of breast cancer, screening for and detection of breast cancer, and breast cancer treatment effectiveness. He programmed a deterministic model which attempted to replicate the Wisconsin Cancer Reporting System (WCRS) data on the annual age- and stage-specific incidence of breast cancer from 1980 to 1992². Chang concluded that a substantial fraction (9%-23%) of all breast cancers are pre-destined from their occult biologic onset to grow only to a limited size (~1 cm diameter), would not present a lethal threat to the woman, and would be indistinguishable from potentially lethal tumors of similar size. He termed these indolent tumors "limited malignant potential (LMP)" tumors.

We had two objectives in redesigning the Chang model as part of the Cancer Intervention and Surveillance Modeling Network (CISNET). The first objective is to answer a question similar to Chang's:

Is it possible to generate a realistic virtual Wisconsin cancer registry of incident breast cancers for women residing in Wisconsin from 1975 to 2000, and to simultaneously replicate age-specific breast cancer mortality in this population during the same time period, with a micro-simulation model comprised of realistically modeled processes representing breast cancer



biologic onset and progression, detection by mammogram screening and case finding outside of screening, and evolving treatment effectiveness over the same time period?

Thus we ask whether observed cancer registry data are compatible at a relatively fine scale over time with the joint product of dynamic processes which most epidemiologists and physicians would agree underlie observed breast cancer data, when those processes are constrained to behave in manner and scale as we think they should. Chang found that he had to add a class of tumors, LMPs, which are indistinguishable from small invasive breast cancers but which in fact do not represent a threat to the host. We began the modeling process prepared to add such assumptions reluctantly, instead exploring many plausible combinations of parameters to improve fit of the virtual cancer registry before resorting to unobservable assumptions about the underlying systems being modeled.

The second objective is to produce a model which can be used to explore ramifications of alternative programs of screening and treatment for breast cancer. Once a simulation model is constructed, it is quite flexible. The model allows output of both simulated cancer registry data and also similar data about breast cancer latent in the population at any given time. This provides the means to answer "What if?" questions about changes in tumor detection and improvements in therapy.

REFERENCES:

-
- ¹ Chang, P., "A Simulation Study of Breast Cancer Epidemiology and Detection since 1982: The Case for Limited Malignant Potential Lesions [Ph.D.]" 1993;
 - ² Bureau of Health Information, Division of Health Care Financing, Wisconsin Department of Health and Family Services "Wisconsin Cancer Incidence and Mortality, 1999" 2002;
-



MODEL OVERVIEW

SUMMARY

This document overviews the modeling effort, the problems it addresses and previous work relevant to this model.



PURPOSE

Model Purpose

BACKGROUND

The National Cancer Institute's Cancer Intervention and Surveillance Modeling Network (CISNET) was formed to use simulation modeling of surveillance data to better understand cancer incidence and mortality. The state of Wisconsin offers a unique population laboratory to develop and test breast cancer simulation models to meet national goals of improving cancer surveillance methods. As part of this consortium we developed and calibrated the Wisconsin Breast Cancer Epidemiology Simulation Model, a discrete–event, stochastic simulation model designed to replicate breast cancer incidence and mortality rates in the Wisconsin female population and applicable to the US population from 1975–2000. The simulation was developed using a systems–science, process modeling approach.

MODEL DESCRIPTION

We have taken a systems engineering approach to construction of our simulation model for breast cancer incidence and mortality in a population. The complex, dynamic biologic and sociodemographic system which results in observed breast cancer statistics is comprised of models of subsystems and specifying the interactions among them in a process analogous to what has been described as "reverse engineering" of complex biologic systems¹. Our model is a discrete–event simulation with a fixed cycle time of 6 months beginning in calendar year 1950. The model is populated by 2.95 million women, divided into birth cohorts, and making up the female population aged 20–100 years of age living in Wisconsin between 1950 and 2000. Women in each birth cohort are individually simulated from calendar year 1950 (or the year in which they were age 20) until they die a simulated death, achieve age 100, or the simulated year 2000 is reached. The processes simulated are:

- A. the natural history of breast cancer from inception to breast cancer death;
- B. detection of breast cancer by screening mammography or clinical surfacing;
- C. improvements in treatment of breast cancer and diffusion of treatments over time; and
- D. death from non–breast cancer causes.



Each of these four major processes is stochastic, unfolding over time in the population, and they jointly result in the observed cancer registry data. These processes form a delicately balanced, interacting system within the population over time. They result in observable consequences unfolding over time as embodied in the statistics collected by a comprehensive cancer surveillance system. When referred to a specific population and time period, these processes result in observed counts of incident breast cancers in each of four distinct stages of disease, in women with known ages, year by year in the reporting system. The model processes also result in counts of deaths in women with known ages across the same years.

CONTRIBUTORS

Dennis G. Fryback, Ph.D.

Natasha K. Stout, Ph.D.

Marjorie A. Rosenberg, Ph.D., F.S.A.

Amy Trentham–Dietz, Ph.D.

Patrick L. Remington, M.D., M.P.H.

Vipat Kuruchittham, M.S.

Acknowledgments

We wish to thank Polun Chang, Ph.D., Associate Professor, National Yang Min University, Taipei, Taiwan, for his assistance in initiating reprogramming of his original model, and Prof. Michael Shwartz has assisted with the tumor progression model. Prof. Miron Livny and his graduate students have provided valuable assistance implementing our model on CONDOR. Discussions with Prof. Polly Newcomb, Dr. Richard Love, Dr. Elizabeth Burnside, Dr. Tara Breslin, and with CISNET collaborators at other institutions have been invaluable in many aspects of the process. Laura Stephenson of the Wisconsin Cancer Reporting System was very helpful in providing WCRS data and responding to questions. We have had programming assistance from Thotsaporn Thanatipanonda, Lorne Tappa, Sriram Ganesan, and Vivek Puttabuddhi.

REFERENCES:

¹ Csete, ME, Doyle, JC “Reverse Engineering of Biological Complexity” in Science 2002; 295: 5560: 1664-1669



ASSUMPTION OVERVIEW

SUMMARY

Our modeling approach assumes that observed incidence and mortality from breast cancer over time can be replicated mechanistically by assembling simulation modules representing basic processes whose general nature and operation is known, but which may be governed by poorly known parameters. This document lists these processes and assumed structure.

BACKGROUND

We assume four major processes, each ultimately stochastic, and each unfolding over time in a population, result in the statistics observed over time in breast cancer surveillance systems:

A. Natural history of breast cancer —

1. Onset: A primary breast cancer may be initiated at some point in a woman's life.
2. Progression/regression: The cancer will grow over time generally progressing in size and spreading to other tissues; depending on their character, tumors may or may not be a lethal threat to the host, and some tumors may regress.
3. Death: Breast cancer deaths occur as an endpoint of a process of uncontrolled growth and spread of the tumor.

D. Detection of breast cancer —

1. Mammogram screening: Breast cancer may be detected by a screening mammogram. A woman's participation in screening may be stochastic or systematic over time. Mammography screening has diffused, and possibly improved over time.
2. Clinical surfacing: Breast cancer may be detected other than by screening mammogram. We do not differentiate these pathways which include diagnosis after symptomatic presentation, self-breast examination, clinical breast examination, or incidental pre-symptomatic detection, terming them collectively as "clinical surfacing."

C. Treatment of breast cancer —

- Potentially lethal breast cancer may in some cases be arrested or retarded by medical intervention. Interventions include surgery of varying extent with or without subsequent radiation, which we take as baseline treatment. Adjuvant therapy with tamoxifen and/or polychemotherapy has been introduced over the time period covered by the model. In the simulation at the individual woman level we use an all or nothing "cure" model for treatment; this sub-model approximates population-level treatment effectiveness statistics.
- Non-breast cancer mortality —
- Women die of non-breast cancer causes, with or without breast cancer present.

Each of these processes is modeled with mathematical functions and stochastic processes. The joint model is referred to as the simulation model or "the model." The parameters governing these functions and processes are constrained by general knowledge, by published data, and by systematic experiments in which the simulation computes estimates of observed surveillance data over time and deviations between the simulation results and observed data are reduced by changing the parameters (we term this latter process calibration, or fitting of the model). Because these processes form a complex and interacting system, some individual parameters may not be identifiable in which case the goal of simulation modeling is to identify feasible sets of values in the parameter space.

ASSUMPTION LISTING

Breast Cancer Natural History Assumptions

1. The probability of breast cancer onset in any given time interval is a function of the woman's individual risk factors and a residual secular trend. ([Cancer Incidence Component](#))
2. Tumor growth is a function of a random initial growth parameter. ([Natural History Component](#))
3. Some tumors have limited malignant potential and will never be a lethal threat to the woman host. ([Natural History Component](#))
4. Breast cancer death can occur only after a tumor has reached the distant stage. ([Survival And Mortality Component](#))

Breast Cancer Detection Assumptions

5. Breast cancer is detected by either screening (characterized in the model as screening mammography) or by other means (characterized in the model as "surfacing"). ([Screening Component](#))
6. The probability an undetected tumor surfaces in a given cycle of the model is an increasing function of tumor diameter. ([Screening Component](#))
7. The sensitivity of a screening mammogram is an increasing function of the woman's age (≥ 50), an increasing function of the diameter of the tumor, and an increasing function of the calendar year of the screening mammogram. ([Screening Component](#))
8. The probability a woman receives a screening mammogram is characterized as increasing over calendar time and her age. ([Screening Component](#))

Treatment Effectiveness Assumptions

9. Treatment outcome is modeled as "cured" or "not cured".
 - a. A cured cancer's growth is arrested instantaneously at the time of treatment.
 - b. An uncured cancer continues to progress at the rate governed by its initially assigned growth parameter to the distant stage and eventual breast cancer death if death from other causes does not intervene. Spherical diameter increase in the model due to "progression" in this case is obviously metaphorical reference to tumor load and not physical diameter. ([Natural History Component](#))



3. The probability of cure is a function of the treatment given and stage of the tumor at detection and possibly the woman's age. ([Treatment Component](#))
4. Treatment modality is a probabilistic function of the calendar year of detection, as treatments have changed and diffused over time. ([Treatment Component](#))

Non Breast Cancer Mortality

12. Non breast cancer mortality occurs independently of whether or not the woman has breast cancer and is a function of the woman's current age and the calendar year of her birth. ([Survival And Mortality Component](#))



PARAMETER OVERVIEW

SUMMARY

The Wisconsin model has many input parameters. This document lists these inputs and their general nature—constants, tables, functions, etc.



[Readers Guide](#)
[Model Overview](#)
[Assumption Overview](#)
[Parameter Overview](#)
[Component Overview](#)
[Output Overview](#)
[Results Overview](#)
[Key References](#)

BACKGROUND

The Wisconsin model is not estimated from any one data source. We have identified processes that are inputs to the epidemiology of breast cancer or which affect statistics that are collected by cancer surveillance in Wisconsin (and nationally) and have attempted to characterize each of these processes mathematically, either with parameters from published or available data sources or parameters estimated by expert judgment.

For demographic data we generally used census data for the US and Wisconsin. Mortality is derived from the Berkeley tables. Cancer surveillance data from the Wisconsin Cancer Reporting System and from SEER were used. Mammography dissemination was from data published in Wisconsin and from the NCI CISNET basecase analysis. Treatment dissemination was supplied by NCI and treatment effectiveness was from the EBCTCG meta-analyses. Breast cancer natural history was patterned after the Shwartz model¹. Mammography characteristics were based on published literature supplemented by expert judgment.

SEER data for Iowa were excluded to allow cross-validation of the Wisconsin model using Iowa data.

Also see [Model Calibration Procedures](#)

PARAMETER LISTING OVERVIEW

The following lists the logical clusters of inputs to the Wisconsin breast cancer simulation model.

Simulation Control

- Starting Simulated Year
- Ending Simulated Year
- Number of Burn-in years
- Termination age
- Cycle Length
- Number of replications

Population Component

- Birth Cohorts
- Number of women in Birth Cohort
- Non-breast cancer mortality tables
- Breast cancer mortality tables



- Population for age-adjusted output

Screening Component

- Mammography screening dissemination model
- Clinical surfacing probabilities
- Mammography operating characteristics

Cancer Incidence Component

- Tumor onset function

Natural History Component

- Tumor type
- Tumor growth rate (exponential growth)
- Positive node probability parameters
- Time to BC death distribution given stage 4
- Tumor size-to-historical-stage translation table

Treatment Component

- Treatment effectiveness by stage
- Adjuvant therapy dissemination model

REFERENCES:

-
- ¹ Shwartz, M. "A Mathematical Model Used to Analyze Breast Cancer Screening Strategies." in *Operations Research* 1978; 6: 26: 937-955
-



COMPONENT OVERVIEW

SUMMARY

The Wisconsin model simulates life histories of individual women from age 20 to age 100, death, or the year 2000. The model utilizes components comprising mortality from non-breast cancer causes, breast cancer onset and progression and mortality, screening and detection, and treatment. These are represented in the flowchart below.



[Readers Guide](#)
[Model Overview](#)
[Assumption Overview](#)
[Parameter Overview](#)
[Component Overview](#)
[Output Overview](#)
[Results Overview](#)
[Key References](#)

OVERVIEW

COMPONENT LISTING

Population Component – Starting the model in 1950, we simulate a number of women equal to each single-year age cohort of women in the Wisconsin population aged 20–99. In each year 1951–1999 we add to the simulation model the number of women aged 20y in that year. The total number of women simulated is approximately 2.95 million.

Cancer Incidence Component – We use "incidence" to mean detection of a tumor; we use "onset" to refer to biological initiation of the tumor. Since tumors grow over time in our model we assumed that tumors which are incident (detected) at one time in fact were biologically onset earlier, and that some women will die of non-breast cancer causes with occult, undiagnosed breast cancer.

Natural History Component – We model breast cancer as a progressive disease, starting with biologic onset at a small focus within the breast and growing spherically in size over time, with probabilistic spread to lymph nodes. The growth model is a simplification of a complicated biologic process, assuming that most breast cancers start as small non-invasive entities and if detected at this time would likely be classified as having historical stage "in situ."

Screening Component – We specify tumor detection probabilities, whether by



mammography or clinical surfacing, as a function of the diameter of the tumor in centimeters, the age of the woman, and the calendar year being simulated. These probabilities were originally specified by a priori expert judgment, then refined by calibrating model outputs to observed surveillance incidence data by tumor stage, not size. The dependence within the model on size is due to an interaction in the model between the growth sub-model and the detection sub-model and the output of these as counts of detected simulated tumors graded into the four stages.

Treatment Component – For simplicity we model treatment as a cure/no-cure process. When a breast cancer is detected, regardless of mode of detection, we assume it is treated. The result of simulated treatment is either "cure" with total arrest of progression at that time and hence no possibility of progressing to a breast cancer death, or the result is "no cure" in which case the tumor continues to progress as if it were undetected and the woman may die of breast cancer, competing causes, or achieve age 100 depending on her individual circumstances.

Survival And Mortality Component – Breast cancer death occurs if the time to a woman's death from non-breast cancer causes is longer than her time to death from breast cancer. The time until death from non-breast cancer causes is chosen at the start of the simulation.



OUTPUT OVERVIEW

SUMMARY

Because the Wisconsin model simulates life histories for individual women, in principle these entire histories are available at the end of the run. In practice the main outputs are age- and historical stage-specific incidence rates of breast cancer and age-specific breast cancer mortality for each year from simulated 1975–2000.



OVERVIEW

The usual outputs saved from a run of the model are as follows:

Means, and standard deviations across N simulations (of the population of interest) for

—

1. 5-yr age group age-, and historical-stage-specific incidence rates of breast cancer in each of the simulated years 1975–2000.
2. 5-yr age-specific breast cancer mortality rates in each of the years simulated.
3. Prevalence of breast cancer in simulated year 1975 (an output constructed for the breast base case)

Alternatively, the same output information can be displayed as age-adjusted rates over calendar years.

With modest reprogramming, the simulation also is capable of outputting the following sorts of quantities —

4. Rates or counts of true positive, false positive, true negative, and false negative mammograms conditioned on screening history and age.
5. Age-specific and stage-specific prevalence of undetected breast cancer in any given calendar year from 1975–2000.

OUTPUT LISTING

See [Output Description](#) for more detail.

RESULTS OVERVIEW

SUMMARY

This document describes results generated by the model.

OVERVIEW

THE FIT OF THE FINAL MODEL AGAINST SEER AND WCRS.

- 
- Readers Guide
 - Model Overview
 - Assumption Overview
 - Parameter Overview
 - Component Overview
 - Output Overview
 - Results Overview
 - Key References

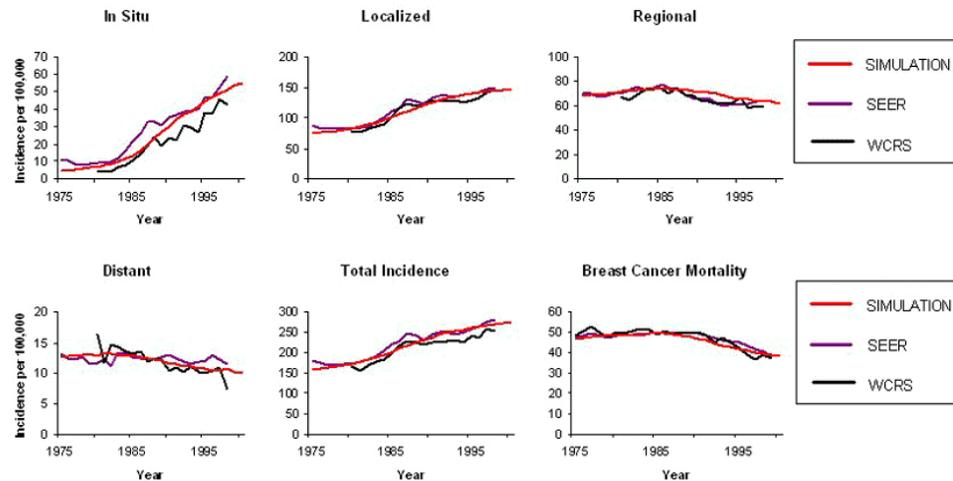


FIGURE 1. The fit of the final model against SEER and WCRS. Incidence rates are age-adjusted to the U.S. population aged 30–79 in year 2000.

RESULTS LIST

To fit observed data, the model required the following assumptions about breast cancer.

1. Breast cancer is a heterogeneous disease, varying in growth rates and in aggressiveness. It is not necessary to assume the heterogeneity is related to age to reproduce many aspects of observed epidemiology of breast cancer with respect to age.
2. There is a class of breast cancer with limited malignant potential, constituting a relatively prevalent latent pool of cancer in the population. If modeled as solid spherical tumors LMP tumors will grow to approximately 1 cm diameter, persist at that size for up to 2 years, and then recede. LMP tumors constituted the followings in the year 2000 (Figure 2).
 - 42% of all cancers at biological onset ($= 126 \div (126+174)$)
 - 28% of all incident tumors ($= (24+45) \div (24+45+30+101+62+10)$)
 - 44% of all incidence in situ tumors ($= 24 \div (24+30)$)
 - 31% of all incident localized tumors ($= 45 \div (45+101)$)
 - There is a small population of breast cancers which are metastatic almost from their beginning. These constitute approximately 4% of non-LMP tumors (or 2% of all tumors).

- The growth rates of tumors which are neither LMP nor the early metastatic type are described by a Gompertz distribution implicit in the growth curves in Figure 3, which shows size of tumors as function of time and percentile of the distribution of means for the gamma distribution of Gompertz growth rates.

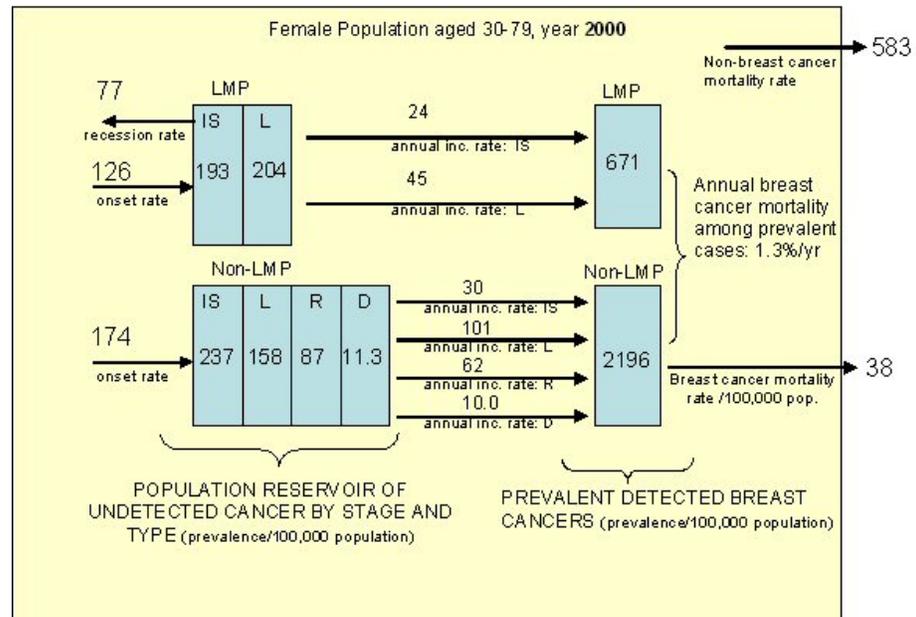


FIGURE 2. A snapshot of occult, incident, and prevalent breast cancer in the U.S. female population aged 30–79 in 2000 as predicted by the simulation model is depicted. Incidence and prevalence are shown as rates per 100,000 in the female population aged 30–79. The annual prevalent case mortality rate is computed as percent per year of all prevalent breast cancers (including LMPs) who die in that year. Of the 671+2196=2867 prevalent cases of breast cancer, 23% are LMP; these women were treated for breast cancer that was not a threat to them. **ABBREVIATIONS:** IS = in situ, L=local, R=regional, D=distant, LMP=Limited Malignant Potential.

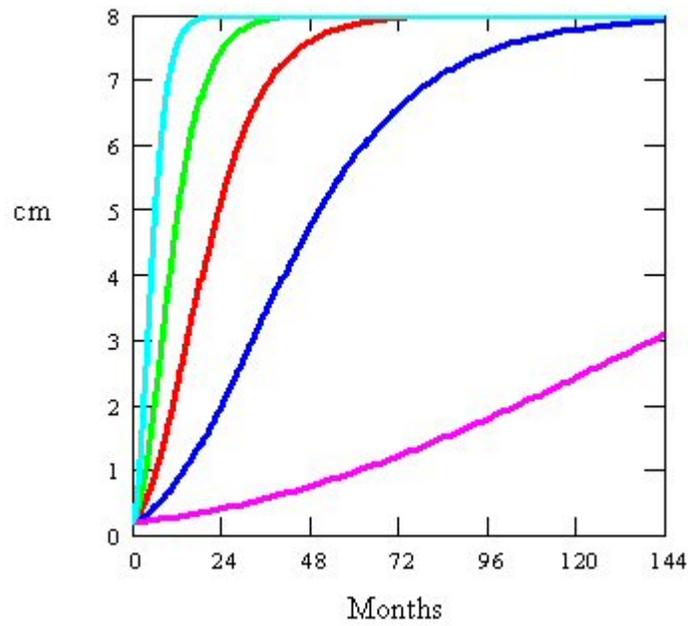


FIGURE 3. Sample growth curves (diameter of ideal spherical tumor as function of time since onset at 0.2 cm diameter) for tumors are shown. The five curves, left to right, represent values of the Gompertz growth rate at the 95th, 75th, 50th, 25th, and 5th percentiles of the gamma distribution for growth rate in the best fitting parameter vector.

Also see: [Model Calibration Procedures](#)



CANCER INCIDENCE COMPONENT

SUMMARY

This document describes the method by which tumor is initiated in the model.

OVERVIEW

We use "incidence" to mean detection of a tumor; we use "onset" to refer to biological initiation of the tumor. Since tumors grow over time in our model we assumed that tumors which are incident (detected) at one time in fact were biologically onset earlier, and that some women will die of non-breast cancer causes with occult, undiagnosed breast cancer. With this consideration in mind, and allowing assumptions about the LMP and hyper-aggressive classes of tumors discussed above, the onset rate for breast cancer is computed in our model as follows.

DETAIL

To derive an onset rate for a woman aged N years in calendar year Y we suppose the observed incidence rate in the absence of screening for a woman in the same birth cohort, but age $N + l$ is I_{N+l} where l is an average lag in years between onset and incidence of tumors in the population. Let p be the proportion of incident (detected) tumors which are not LMP tumors, and let f be the fraction of all onset tumors which are LMP. Then the onset rate for the woman aged N years is given by $O_N = \frac{pI_{N+l}}{(1-f)}$. The numerator of the fraction is the rate of non-LMP tumors as a fraction of all incident tumors l years later, and these are $1 - f$ of the total number of tumors which are onset. The model uses age-period-cohort breast cancer incident rates inferred in the absence of screening, which incorporate an increasing secular trend, provided to the CISNET collaboration by the NCI (see Figure 4)¹. We fit the lag parameter, l , empirically during calibration. The fitted values derived for the three parameters during model calibration are $l=3$, $f=0.42$, and $p=0.95$.



[Readers Guide](#)
[Model Overview](#)
[Assumption Overview](#)
[Parameter Overview](#)
[Component Overview](#)
[Output Overview](#)
[Results Overview](#)
[Key References](#)

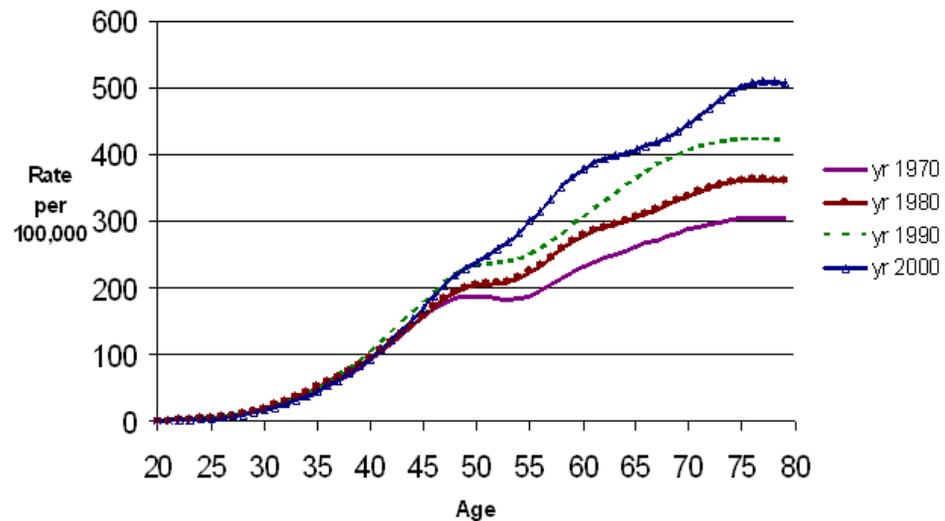


FIGURE 4. An age–period–cohort model for breast cancer incidence in absence of screening. Derived for CISNET consortium.¹ The four lines depict incidence in the indicated years.

Actual incidence observed as output from the simulation model is a function of the underlying pool of tumors that are biologically onset, the nature of the tumors (in particular, tumor diameters), the probability of clinical surfacing as a function of size of tumor, and the operating characteristics for screening mammography as a function of size of the tumors (i.e., screening sensitivity).

With one exception, we do not model recurrence or second primary breast cancers. The exception is that a woman with an undetected LMP tumor which then disappears is again at risk for onset of any type of breast cancer. A woman may not have both an LMP and a "real" tumor simultaneously. Once a woman is diagnosed with breast cancer—LMP or otherwise—in the current simulation she cannot develop a second primary breast cancer. We do not model recurrence per se, as discussed later under "treatment."

REFERENCES:

¹ Holford, TR., Cronin, K., Feuer, EJ., Mariotto, A. "Changing patterns in breast cancer incidence trends, manuscript" 2003;



NATURAL HISTORY COMPONENT

SUMMARY

This document describes the model of tumor progression.

OVERVIEW

We model breast cancer as a progressive disease, starting with biologic onset at a small focus within the breast and growing spherically in size over time, with probabilistic spread to lymph nodes. The growth model is a simplification of a complicated biologic process, assuming that most breast cancers start as small non-invasive entities and if detected at this time would likely be classified as having historical stage "in situ."

DETAIL

Natural history of the disease.

We discuss tumor progression before biological tumor onset here since we found parameters controlling onset depend on the nature of the progression model.

The tumor progression model.

We initially began with a model proposed by Shwartz⁴ which modeled tumor growth as an exponential doubling process. In Shwartz's model, every breast cancer is assumed to have a fixed growth rate once drawn at the tumor's biologic inception from a lognormal distribution. Shwartz modeled the number of involved lymph nodes at a given time as a cumulative Poisson process with rate parameter determined by current tumor diameter and the rate at which the diameter is changing so that the larger the tumor, the more likely it is to have involved nodes and the faster growing the tumor is the more likely it is to have metastatic spread. In Shwartz's model all tumors started at a diameter of 0.5cm.

We altered two aspects of Shwartz's model to better calibrate to surveillance data. Because modern screening is potentially capable of detecting breast cancers less than 0.5cm in size, tumors now enter the simulation with a size of 0.2cm. Second, we implemented a decelerating, Gompertz growth function⁵ to replace Shwartz's exponential growth function. Shwartz had suggested a Gompertz function may fit equally well as the exponential model and did some exploration of this alternative.⁶; (M. Shwartz, personal communication, 2002) The exponential model, characterized by constant tumor volume doubling times, is plausible for early tumor growth, but implausible as tumors become larger in size. The Gompertz growth model is exponential growth with decelerating doubling time shown in equation (1).

$$V(t) = V_0 e^{\frac{\beta}{\alpha}(1-e^{-\alpha t})} \tag{1}$$



- Readers Guide
- Model Overview
- Assumption Overview
- Parameter Overview
- Component Overview
- Output Overview
- Results Overview
- Key References

Here $V(t)$ is the tumor volume at time t , V_0 is the initial tumor volume, the parameter β is the initial tumor growth rate, and α governs deceleration in growth rate over time. We fix one of these parameters by fixing the maximum asymptotic tumor volume as $t \rightarrow \infty$, $V_{max} = V_0 e^{\beta/\alpha}$. Solving for β and substituting into equation (1) we get the volume at time t expressed in terms of one free parameter, α .

$$V(t) = V_0 e^{\ln\left(\frac{V_{max}}{V_0}\right)(1-e^{-\alpha t})} \quad (2)$$

Finally, assuming spherical tumors, tumor volume and diameter are related by $V = \frac{\pi}{6} d^3$. Substituting this expression where appropriate in (2), taking the cube root of both sides, and solving for $d(t)$ as a function of corresponding constants d_{max} , d_0 , and α , we get an equation for tumor diameter at time t .

$$d(t) = d_0 e^{\ln\left(\frac{d_{max}}{d_0}\right)(1-e^{-\alpha t})} \quad (3)$$

The minimum and maximum diameters were specified arbitrarily at 0.2 and 8.0 cm, to be reasonably within bounds observed clinically. The growth parameter α is modeled as a gamma-distributed random variable with a mean of 0.12 and a variance of 0.012, values derived in the model calibration process. The mean and variance of the gamma distribution were determined by model calibration. Figure 1 shows tumor diameter as a function of time (in months) for tumors with growth parameters from various percentiles of the growth rate distribution.

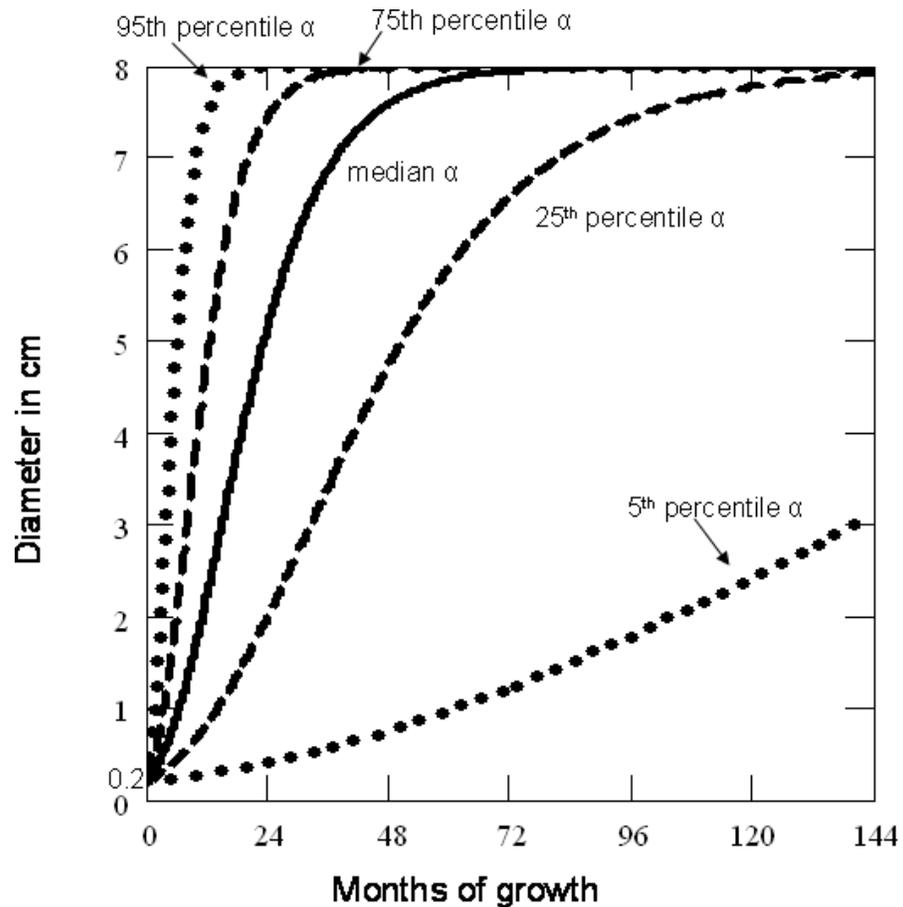


FIGURE 1. Sample growth curves for tumors. The ordinate is diameter in cm, and the abscissa is the number of months since tumor inception. The five curves, left to right, represent values of α at the 95th, 75th, 50th, 25th, and 5th percentiles of the gamma distribution for growth rate (smaller values of represent slower growth)

The rate of additional involved nodes was modeled based on Shwartz's model⁶. In Shwartz's model the instantaneous rate at time t was given by

$n(t) = b_1 + b_2V(t) + b_3V'(t)$ where the b_i are constants with values 0.0058, 0.0053, and 0.0002 respectively, $i=1,2,3$. Integrating $n(t)$ from time $t - 1$ to time t , we derive the additional nodes in that interval,

$$N(t) - N(t-1) = \int_{t-1}^t (b_1 + b_2 \times V(u) + b_3 \times V'(u)) du$$

In calibrating the model to observed incidence surveillance data we found that this formula gave slightly too fast a rate of tumor spread. For a tumor of a given diameter, $d(t)$, and corresponding volume $V(t)$, we used Shwartz's equation for a tumor with a volume corresponding to a diameter 25% smaller, $0.75d(t)$, based on fitting this parameter during calibration. Figure 2 shows the resulting empirical rate of additionally involved lymph nodes in a 6 month period as a function of simulated tumor diameter and growth rate.

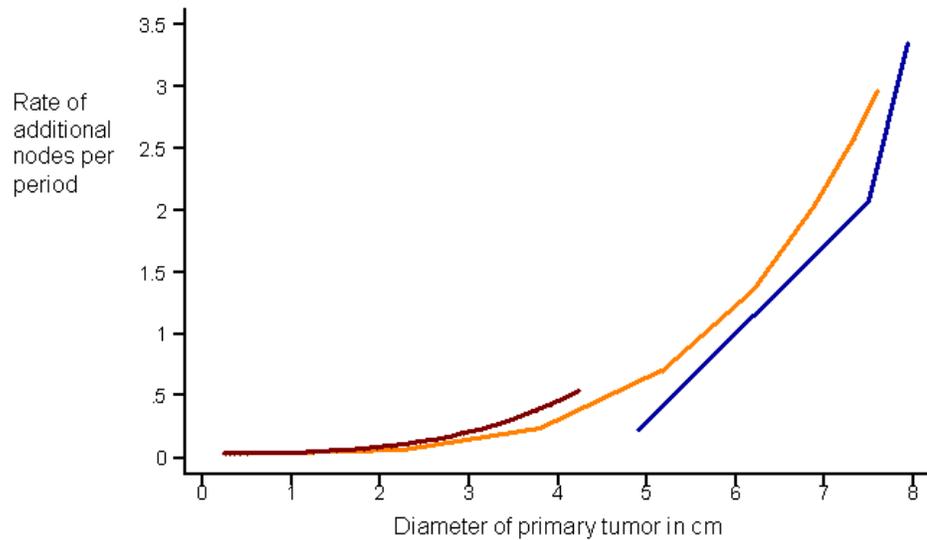


FIGURE 2. Poisson rate of additional involved nodes in the next 6 month period as a function of current tumor size. The three curve segments represent (from left to right) tumors in the 5th, 50th, and 95th percentiles of the growth rate distribution.

Tumors with zero nodes that are less than 0.95 cm in diameter are defined as having in situ stage in our model. Tumors with zero nodes and greater than 0.95 cm are considered to be localized. Tumors with 1 to 4 nodes are considered to be regional stage tumors. Any tumor with 5 or more involved nodes is considered to be in the distant stage and proxy for more widespread involvement.

Additional assumptions about tumor natural history.

Preliminary model results revealed that far too many occult tumors were required to achieve the rise in incidence (detection) that occurred after the widespread use of screening mammography during the late 1980s. In order for breast cancer incidence to be sufficiently high during this time period the model required unrealistically elevated breast cancer mortality rates prior to the dissemination of mammography. For this reason, we incorporated the same assumption that Chang included in his original model: A substantial fraction of all incident breast cancers prior to screening must have been of limited malignant potential (LMP), i.e., of no lethal threat to the host woman. Furthermore, we inferred there must be a reservoir of these occult LMP tumors that would be discovered with the advent of screening programs. The following characterization of LMP tumors evolved during model fitting and calibration:

LMP tumors have the same growth rate distribution as other tumors (the gamma distribution described above), and grow according to the same Gompertz growth function as other tumors. However they cease growth at 1 cm in size, and they disappear after 2 years dwell time at this size. LMP tumors never exhibit metastatic spread, and thus do not lead to breast cancer death.

Each of these characteristics was needed to obtain satisfactory calibration of the model dynamically across time. Since LMP tumors were needed as biologically prevalent but



occult tumors to feed the dramatic rise in incidence of localized breast cancer with the advent of screening, they could not progress beyond the localized stage and do not cause breast cancer death. To remain largely occult in the absence of screening they could not grow too large, hence the 1 cm upper limit. Too large a pool of undiagnosed, prevalent tumors developed in the simulation if the LMP tumors remained indefinitely after reaching the 1 cm limit, so we imposed the 2-year dwell time (this parameter being fit during calibration); future versions of the model will include regression over time, but currently disappearance is modeled as instantaneous at 2 years. The fraction of all tumors that were defined as LMP tumors was found through calibration to incidence surveillance data and the corresponding parameters are discussed below under "onset." The existence of LMP tumors is at present hypothetical. We must assume they are histologically indistinguishable from "real" breast cancer or their existence would be already known. It is our claim, resulting from our modeling efforts, that assuming their existence is necessary to explain the dynamics of the observed facts about breast cancer in the population over the past 25 years.

It is reasonable to ask whether the role that LMP tumors play in the model might be simulated by skewing the growth distribution toward slow-growing tumors. In fact the upper growth limit of about 1 cm is also needed in the characterization of LMP tumors, as is the regression of these tumors. While they may represent an "indolent" end of the growth distribution of breast cancer they appear to be a distinct subpopulation of cancers statistically, with the entire population being a mixture of LMP type and lethal type tumors. An extensive discussion of the need for the LMP tumor type is presented elsewhere⁷.

In the same way that LMP tumors represent an indolent end of the tumor growth spectrum, we found we needed to assume there were hyper aggressive tumors at the other end of the spectrum as well. Since the mid-1990s, the incidence rates for breast cancer diagnosed at the regional or distant stage have leveled off⁸. Our simulation model assumes that a small fraction of tumors have rapid metastatic spread. At initiation in the growth model, when the primary tumor diameter is assumed to be 0.2 cm, 1% of non-LMP tumors are assumed to have 4 positive nodes, and 2% has 5 or more nodes, i.e., these tumors are either regional or distant stage tumors right at initiation and cannot be detected in an earlier stage in the model.

REFERENCES:

- 1 Shwartz, M. "A Mathematical Model Used to Analyze Breast Cancer Screening Strategies." in *Operations Research* 1978; 6: 26: 937-955
- 2 Shwartz, M. "An Analysis of the Benefits of Serial Screening for Breast Cancer Based Upon a Mathematical Model of the Disease." in *Cancer* 1978; 4: 41: 1550-1564
- 3 Shwartz, M. "Estimates of lead time and length bias in a breast cancer screening program" in *Cancer* 1980; 46: 4: 844-51
- 4 Shwartz, M. "Validation and Use of a Mathematical Model to Estimate the Benefits of Screening Younger Women for Breast Cancer." in *Cancer Detection and Prevention* 1981; 4: 595-601
- 5 Spratt, JS., Spratt, JA. "Chapter 21 Growth Rates. In: Donegan WL, editor" in *Cancer of the Breast* 2002; : 443-476
- 6 Shwartz, M. "Analysis of Preventive Screening to Detect Early Breast Cancer [Ph.D.]" 1975;



University of Wisconsin
Natural History Component
References:

-
- ⁷ Fryback, DG, Stout, NK, Rosenberg, MA, Trentham-Dietz, A, Remington, PL, Kuruchittham, V "The case for limited malignant potential breast tumors: Implications of the Wisconsin breast cancer epidemiology simulation model" 2004;
- ⁸ Ries, LAG., Eisner, MP., Kosary, CL., Hankey, BF., Miller, BA., Clegg, L "SEER Cancer Statistics Review, 1975-2000" in National Cancer Institute 2003;
-



SURVIVAL AND MORTALITY COMPONENT

SUMMARY

This document describes the methods by which survival and mortality are determined in the model.



- [Readers Guide](#)
- [Model Overview](#)
- [Assumption Overview](#)
- [Parameter Overview](#)
- [Component Overview](#)
- [Output Overview](#)
- [Results Overview](#)
- [Key References](#)

OVERVIEW

Breast cancer death occurs if the time to a woman's death from non-breast cancer causes is longer than her time to death from breast cancer. The time until death from non-breast cancer causes is chosen at the start of the simulation using mortality probabilities derived from birth cohort-specific, U.S. mortality tables published at UC Berkeley. From these probabilities we have removed breast cancer as a cause of death using breast cancer mortality rates in SEER data, and standard actuarial procedures.

DETAIL

Breast cancer death.

We assume that breast cancer death results only from disease which has progressed to the distant stage. (In our model, tumors continue to grow larger and involve an increasing number of lymph nodes even after clinical detection if they are not "cured." This is obviously a simplistic shadow model for more complicated biologic processes of recurrence and spread of treated breast cancers.) At the time a woman's cancer reaches the distant stage in the simulation, whether detected or not, she is assigned a time of death from breast cancer. Time until death is drawn from an empirical distribution based on survival times for women in the SEER registry diagnosed with distant stage breast cancer between 1975 and 1982 and who died of breast cancer¹. We selected this time period to be prior to the advent of widespread mammography screening. Presumably tumors found in the distant stage after the advent of screening are from the increasingly aggressive end of the growth spectrum and we wished to use survival times to breast cancer death from a more representative sample of the spectrum. The 1975–82 time period also pre-dates modern treatment protocols for distant stage breast cancer. The empirical distribution (figure 3) was estimated by smoothing a life table of years of life from age at diagnosis created for these women from SEER data¹. The median time to breast cancer death after arriving at the distant stage in our model is 1.95 years, and mean time 5.22 years.

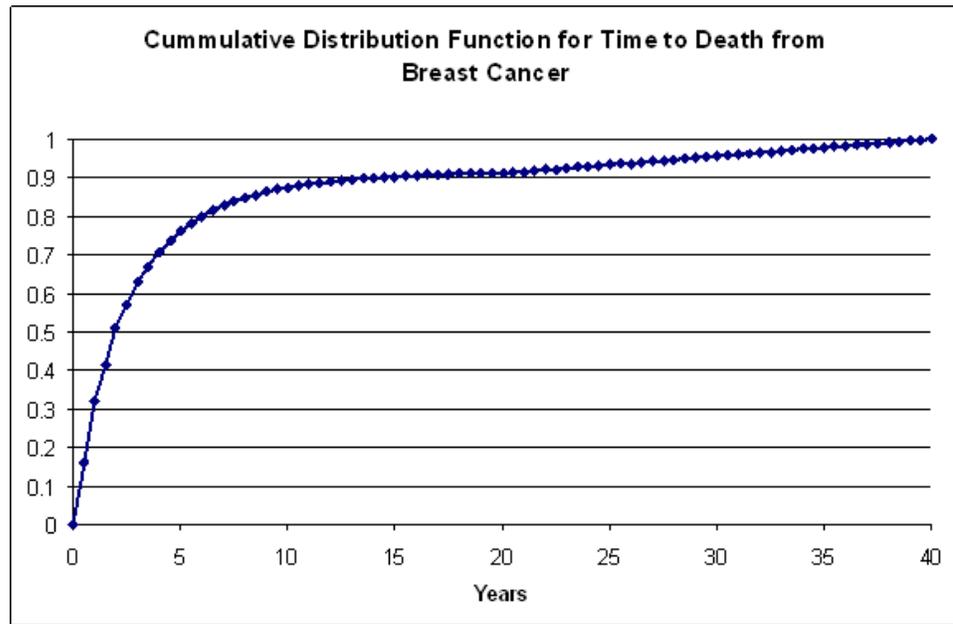


FIGURE 3. Empirical distribution of time from diagnosis to time of breast cancer death for women diagnosed with distant stage breast cancer in SEER between 1975 and 1982.

Mortality from non-breast cancer causes

We actuarially adjusted age-specific all-cause life tables of female mortality by birth cohort to develop mortality rates from non-breast cancer causes. The all-cause mortality by birth cohort from 1891 to 2000 is published by the Berkeley Human Mortality Data Base², while the age-specific breast cancer mortality is from National Center for Health Statistics, Centers for Disease Control and Prevention, vital statistics(see³ for details). The resulting non-breast cancer mortality rates are common input to all the CISNET breast cancer collaboration models.

REFERENCES:

- ¹ Ries, LAG., Eisner, MP., Kosary, CL., Hankey, BF., Miller, BA., Clegg, L "SEER Cancer Statistics Review, 1975-2000" in National Cancer Institute 2003;
- ² Wilmoth, JR., Shkolnikov, V., [Directors] "Human Mortality Database. US Female Death Rates by Year of Birth" 2003;
- ³ Rosenberg, MA "Competing risks" in JNCI Monographs 2004;



SCREENING COMPONENT

SUMMARY

This document describes the methods by which screening is simulated in the model.

OVERVIEW

We specify tumor detection probabilities, whether by mammography or clinical surfacing, as a function of the diameter of the tumor in centimeters, the age of the woman, and the calendar year being simulated. These probabilities were originally specified by a priori expert judgment, then refined by calibrating model outputs to observed surveillance incidence data by tumor stage, not size. The dependence within the model on size is due to an interaction in the model between the growth sub-model and the detection sub-model and the output of these as counts of detected simulated tumors graded into the four stages.

DETAIL

Clinical Surfacing

Clinical surfacing is expressed as an annual probability, and converted to a 6-month interval probability in each cycle of the simulation. We presume the probability of detecting an existing breast cancer in the absence of screening is low for small tumors and the probability should increase with diameter of the tumor. We also believe that women's self-detection of breast lumps has improved, particularly over the past decade, as a result of increasing public awareness of breast cancer. Under these assumptions, the annual probabilities for clinical surfacing of a tumor are shown in Table 1. These are the result of fitting and smoothing during calibration of the model subject to constraints that they are increasing in tumor diameter and over time, and that the probability at ≤ 0.3 cm is zero and at 8 cm is 1.0. Notice that the improvement in detection in the decade of the 1990s appears to be mostly for small to mid-size tumors.

TABLE 1. Annual Probabilities of Clinical Surfacing

Year*	Diameter of the tumor (cm)*							
	≤ 0.3	1.0	1.5	2.0	3.0	4.0	5.0	8.0
≤ 1990	0.0	0.06	0.07	0.15	0.40	0.70	0.80	1.0
2000	0.0	0.06	0.07	0.30	0.55	0.75	0.80	1.0

*Probabilities are linearly interpolated for years between 1990 and 2000, and between tumor diameters.



- Readers Guide
- Model Overview
- Assumption Overview
- Parameter Overview
- Component Overview
- Output Overview
- Results Overview
- Key References

The one exception to this table is that if at the beginning of a simulated 6 month interval a woman with an as-yet undetected tumor is determined to die of breast cancer, we force detection at that time regardless of size of the primary tumor on the presumption that the cause of death would almost surely be diagnosed ante mortem from other signs and symptoms of metastatic disease even though the primary may be small. This proviso is necessitated because of the discrete time steps of the simulation and the fact that we do not model symptoms.

Sensitivity of Mammography Screening

Table 2 shows detection probabilities of a mammogram for an existing breast cancer. These are probabilities of detecting a tumor of a given diameter, in a woman of a given age, in a given calendar year on a given screening mammogram. (These are not rates or annual probabilities.) The values in Table 2 are the result of calibration to incidence data given constraints of monotonicity in tumor diameter, woman's age, and calendar year, and that the sensitivity to a tumor 5 cm is .99, and 8 cm in diameter is 1.0. Note the probability of detecting a tumor between 0.2 and 0.5 cm in diameter is constant in tumor diameter as detection of small tumors is mostly dependent on factors other than size such as calcification.

TABLE 2. Sensitivity of Mammography

Year*	Diameter of the tumor (cm)**						
	≤0.20	0.2-0.5	0.75	1.5	2.0	5.0	8.0
Women aged							
≤1984	0.00	0.06	0.35	0.65	0.85	0.99	1.0
≥2000	0.00	0.20	0.60	0.65	0.85	0.99	1.0
Women aged ≥50y**							
≤1984	0.00	0.1	0.45	0.65	0.85	0.99	1.0
≥2000	0.00	0.30	0.65	0.80	0.90	0.99	1.0

*Probabilities are linearly interpolated for years between 1984 and 2000, and between tumor diameters.

**Sensitivity of mammography is presumed to increase after menopause when breast tissues become more radiolucent with fatty replacement. Our current model uses an abrupt transition at age 50 to represent what is likely a more gradual process.



Diffusion of mammography utilization

Screening mammography before 1982 was sporadic and rare. Originally, we used mammography dissemination data from Wisconsin to specify the probability that a woman of a given age would receive a screening mammogram in a given calendar year from 1982–1991². More recently we have incorporated an age–period model of annual screening probability³. The output of this model is random ages representing the age at first screen for a woman and the ages of subsequent screens up to age 100. The marginal frequencies match national data from the US for screening rates from 1975 to 2000. If a woman with an occult tumor is screened in the simulation, the probabilities in Table 2 are used to determine whether or not the tumor is detected at that screen. When a tumor is detected, the underlying state of the tumor according to the natural history model determines the stage of detection. Note that detection probabilities are not a function of stage, but of simulated tumor size and mammography sensitivity given the woman’s age and the calendar year. Mammography detection rates by tumor stage are emergent properties of the simulation.

REFERENCES:

-
- ¹ Lantz, PM., Weier, A., Remington, PL. “Mammography use in Wisconsin: 1987-1992” 1993; 92: 10: 580-2
 - ² Lantz, P., Bunge, M., Remington, PL. “Trends in mammography in Wisconsin” in Wis Med J 1990; 89: 6: 281-2
 - ³ Cronin, K “The dissemination of mammography in the US” in JNCI in press
-



TREATMENT COMPONENT

SUMMARY

This document describes the treatment of cancer and how its effectiveness is modeled.

OVERVIEW

Treatment and treatment effectiveness

For simplicity we model treatment as a cure/no-cure process. When a breast cancer is detected, regardless of mode of detection, we assume it is treated. The result of simulated treatment is either "cure" with total arrest of progression at that time and hence no possibility of progressing to a breast cancer death, or the result is "no cure" in which case the tumor continues to progress as if it were undetected and the woman may die of breast cancer, competing causes, or achieve age 100 depending on her individual circumstances. Continued simulated growth in this case is used to mark a time-line for progression and size per se is not meant to be biologically representative.

This method for modeling treatment outcomes models survival given treatment as a mixture of two survival curves conditioned on the woman's age and the tumor characteristics at the time of detection. One curve is the survival curve of women that age without breast cancer; the other curve is women that age and with tumors of the same stage but untreated, which is a function of the underlying natural history model. The mixture probability to be fit is termed the "cure" fraction. Although this method gives up fitting the shape of the survival curve given stage, and treatment, it avoids having to make direct assumptions about survival given mode of detection, treatment, etc., which we wish to be emergent properties of the model rather than inputs.

The treatment submodel has three logical parts. First, we specified treatment effectiveness—cure fractions—in the pre-tamoxifen pre-adjuvant polychemotherapy era for tumors treated at different stages with a standard, baseline therapy. These "baseline" cure probabilities metaphorically represent overall mastectomy with or without radiation as was common in the pre-1975 era. Second, we specified the relative improvement in survival with the various combinations of adjuvant therapies added to the baseline therapy. Third, we specified the diffusion of these adjuvant treatments over time as a function of characteristics of the woman and the stage of tumor at diagnosis.

DETAIL

The model assumes that all women receive baseline treatment consisting of standard therapy such as surgery and/or radiation and that the effectiveness of this baseline treatment has not changed over time. In addition women may receive one of five modes of adjuvant therapy. The different modes of adjuvant therapy, which are determinants of the effectiveness of treatment, are chemotherapy only, tamoxifen only for two years, tamoxifen only for five years, chemotherapy in combination with Tamoxifen for two years or chemotherapy in combination with tamoxifen for five years.

Women with breast cancer detected in the localized or regional stages are assigned a





mode of adjuvant treatment based on the calendar year, her current age, tumor size/ stage and revealed estrogen receptor (ER) status. Tumors diagnosed in the in situ or distant stages are not assigned adjuvant therapy. The likelihood of each mode of treatment is based on observed use of the treatment. Data describing the likelihood of treatment were provided by NCI based on analysis of data from the Patterns of Care study as well as combined data from numerous cancer registries¹. These data show increasing use of chemotherapy for large localized tumors and regional tumors over time and increasing use of a 5-year course of tamoxifen for ER positive tumors. All CISNET collaboration models use these data as input.

Revealed ER status is modeled as a function of the true ER status of the tumor as well as the calendar year. True ER status is based on the age of the woman at the time of tumor onset (Table 3)². In the simulation, the treatment probabilities are determined in part by whether the ER status is known. We used SEER data from 1990 forward (the first year this was recorded in the SEER data) to estimate the proportion of tumors with ER status determined; probabilities before this time were based on assessment of a local expert oncologist–breast cancer researcher (Table 4). The treatment administered is in part determined by whether the ER status is known and if so whether it is positive or negative. The treatment effectiveness is determined as a function of the ‘true’ underlying ER status of the tumor and the treatment given.

TABLE 3. Probability that a tumor is ER Positive by Age

Age	Pr(ER+)
	0.6
45–54	0.65
55–64	0.74
65–74	0.77
75+	0.83

TABLE 4. Likelihood that the True ER status of a tumor will be known

Year	Pr(ER Status Known)
	0.1
1975–1979	0.2
1980–1984	0.5
1985–1989	0.63
1990	0.68
> 1991	0.69

No data directly describe the effectiveness of treatment as implemented in the simulation model. Data from randomized clinical trials are usually reported as relative survival gains or decreases in annual mortality odds. Thus the derivation of the likelihood that a particular mode of treatment would be effective was a several step process.

The estimation of treatment effectiveness started with baseline effectiveness for standard treatment by stage at detection in the absence of adjuvant therapy. These likelihoods were estimated by expert opinion and were refined during the calibration process. Observed survival was approximated by a mixture of the survival given treatment is completely effective and the survival given treatment is completely ineffective where the mixture proportion is the probability that treatment is effective.



Using this relationship, 10-year survival probabilities (considering all causes of death) were estimated from the simulation model under conditions that all treatment is completely effective and again assuming all treatment is completely ineffective, and modified 10-year survival probabilities were computed by age group and stage. Annualized mortality rates were calculated assuming that the annual mortality rate was constant over the 10-year interval.

The annual odds of mortality for adjuvant treatment were calculated by applying trial results about the performance of treatment reported as the annual reduction in the odds of mortality to the computed annual odds of mortality for baseline treatment. These annual odds were further adjusted based on the length of the course of treatment for tamoxifen. These adjustments to the baseline odds of mortality assume that the effects of chemotherapy and tamoxifen are independent and that tamoxifen is effective for women with tumors that are ER positive³(pp1-15,71-85).

These adjusted annual odds of mortality by the current age of the woman, tumor stage (localized and regional only), true ER status, and mode of adjuvant treatment were converted back to the annualized mortality rates. Again assuming constant annual all-cause mortality, implied 10-year survival probabilities were computed. Adjusted mixture proportions that treatment is curative were recomputed based on the adjusted 10-year survival probabilities and the two simulated survival probabilities under conditions that all treatment is either completely effective or completely ineffective. These adjusted cure fractions are reported in table 5.



TABLE 5. Cure probabilities used in the model.

Stage	Age	ER status	No Adj Tx	Tam 2yr	Tam 5yr	Chemo only	Tam 2yr +chemo	Tam 5yr +chemo
In Situ		-	0.990	0.990	0.990	0.990	0.990	0.990
In Situ	50-59-		0.990	0.990	0.990	0.990	0.990	0.990
In Situ	60-69-		0.990	0.990	0.990	0.990	0.990	0.990
In Situ	70+	-	0.990	0.990	0.990	0.990	0.990	0.990
In Situ		+	0.990	0.990	0.990	0.990	0.990	0.990
In Situ	50-59+		0.990	0.990	0.990	0.990	0.990	0.990
In Situ	60-69+		0.990	0.990	0.990	0.990	0.990	0.990
In Situ	70+	+	0.990	0.990	0.990	0.990	0.990	0.990
Localized		-	0.820	0.820	0.820	0.882	0.882	0.882
Localized	50-59-		0.820	0.820	0.820	0.867	0.867	0.867
Localized	60-69-		0.820	0.820	0.820	0.864	0.864	0.864
Localized	70+	-	0.820	0.820	0.820	0.957	0.957	0.957
Localized		+	0.820	0.861	0.884	0.882	0.913	0.931
Localized	50-59+		0.820	0.881	0.916	0.867	0.921	0.952
Localized	60-69+		0.820	0.921	0.980	0.864	0.959	1.000
Localized	70+	+	0.820	1.000	1.000	0.957	1.000	1.000
Regional		-	0.400	0.400	0.400	0.527	0.527	0.527
Regional	50-59-		0.400	0.400	0.400	0.470	0.470	0.470
Regional	60-69-		0.400	0.400	0.400	0.446	0.446	0.446
Regional	70+	-	0.200	0.200	0.200	0.268	0.268	0.268
Regional		+	0.400	0.482	0.532	0.527	0.599	0.642
Regional	50-59+		0.400	0.492	0.549	0.470	0.558	0.611
Regional	60-69+		0.400	0.509	0.578	0.446	0.554	0.621
Regional	70+	+	0.200	0.370	0.493	0.268	0.448	0.575
Distant		-	0.050	0.050	0.050	0.050	0.050	0.050
Distant	50-59-		0.050	0.050	0.050	0.050	0.050	0.050
Distant	60-69-		0.050	0.050	0.050	0.050	0.050	0.050
Distant	70+	-	0.025	0.025	0.025	0.025	0.025	0.025
Distant		+	0.050	0.050	0.050	0.050	0.050	0.050
Distant	50-59+		0.050	0.050	0.050	0.050	0.050	0.050
Distant	60-69+		0.050	0.050	0.050	0.050	0.050	0.050
Distant	70+	+	0.025	0.025	0.025	0.025	0.025	0.025

The baseline cure fractions are shown in the "No Adj Tx" column. Initially we held these constant within stage regardless of the age of the woman. However breast cancer mortality among older women was consistently too low. Accordingly, we reduced the cure fractions of the two more advanced stages for women aged 70 years and older. This is consistent with observations that older women appear to be less aggressively treated than younger women ⁶.

Treatment of LMP tumors is assumed to be 100% curative since these tumors are, by our definition, not lethal.



REFERENCES:

- ¹ Mariotto, A., Feuer, E.J., Harlan, L.C., Wun, L-M., Johnson, K., Abrams, J. "The dissemination of adjuvant therapy for breast cancer in the US: 1975-1997" in The National Cancer Institute 2002;
- ² Ashba, J., Traish, AM. "Estrogen and progesterone receptor concentrations and prevalence of tumor hormonal phenotypes in older breast cancer patients" in Cancer Detect Prev 1999; 23: 3: 238-44
- ³ EBCTCG (Early Breast Cancer Trialists' Collaborative Group) "Systemic treatment of early breast cancer by hormonal, cytotoxic, or immune therapy. 133 randomised trials involving 31,000 recurrences and 24,000 deaths among 75,000 women" in Lancet 1992; 339: 8785: 71-85
- ⁴ Newcomb, PA., Carbone, PP. "Cancer treatment and age: patient perspectives" in J Natl Cancer Inst 1993; 85: 19: 1580-4
- ⁵ Kemeny, MM., Peterson, BL., Kornblith, AB., Muss, HB., Wheeler, J., Levine, E. "Barriers to clinical trial participation by older women with breast cancer" in J Clin Oncol 2003; 21: 12: 2268-75
- ⁶ Wyld, L., Reed, MW. "The need for targeted research into breast cancer in the elderly" in Br J Surg 2003; 90: 4: 388-99



MODEL CALIBRATION PROCEDURES

The inputs to the simulation model are numerous. We took the following inputs as given:

1. Estimated breast cancer incidence in the absence of screening;
2. The patterns of mammography dissemination over time;
3. The patterns treatment dissemination over time;
4. Dissemination of ER status determination over time, and the relationship between the woman's age and ER status of her tumor;
5. The relative effectiveness of various modes of adjuvant therapy.

Other inputs to the model were calibrated to make the model outputs conform to surveillance data from 1975–2000. These calibrated inputs include the tumor detection probabilities in Tables 1 and 2 in [Screening Component](#), the baseline treatment effectiveness probabilities in the no adjuvant therapy ("No Adj Tx") column of Table 5 in [Treatment Component](#), and the variables listed in Table 6. Table 6 is entitled "core parameters" because these 10 turned out to be the parameters about which the least is known and which apparently control model output with the most sensitivity.



[Readers Guide](#)
[Model Overview](#)
[Assumption Overview](#)
[Parameter Overview](#)
[Component Overview](#)
[Output Overview](#)
[Results Overview](#)
[Key References](#)



TABLE 6. Calibrated core input parameters.

Parameter	Use in the model	[Sampled Range] (increment size for discrete sampling)		Final value
		Wide ranges*	Focused ranges*	
1. LMP Fraction	Proportion of all biologically incident tumors assumed to be Limited Malignant Potential (LMP)	[0% – 55%] (1%)	[30% – 50%] (1%)	42%
2. Max LMP Size	LMP tumors assumed to grow no larger than this diameter (cm) (this variable was fixed as it is entangled with In Situ Boundary, the Gompertz growth parameters, and LMP Dwell Time)	1 cm	1 cm	1 cm
3. LMP Dwell Time	Maximum sojourn time (years) for LMP tumor after reaching Max LMP Size; after this time without discovery, the LMP tumor disappears next simulation cycle.	[1–3] (0.5)	[1.5–2.5] (0.5)	2 y
4. In Situ Boundary	The diameter (cm) below which the tumor is classified as in situ stage in the simulation if there are no associated positive lymph nodes	[0.75 – 1.0] (0.01)	[0.85 – 0.99] (0.01)	0.95 cm
5. Onset Proportion	Ratio of assumed age-specific biologic onset rate divided by age-specific incidence rate (the latter specified by age-period-cohort model estimated in absence of screening – see text).	0.85 – 1.2 (0.01)	0.8 – 1.0 (0.01)	0.9
6. Onset Lag	Time interval (years) between year of index onset rate and incidence rate used in Onset Proportion. This is to “fill the pipeline” with biologically onset tumors which will be discovered at a given incidence rate some years later. Because the cycle time of the model in 0.5 years, this was taken to be step size.	[1–8] (0.5)	[1.5–4] (0.5)	3 y
7. Percent 4 nodes	Percent of biologically onset, non-LMP tumors which are assigned 4 positive lymph nodes at onset. (This places these tumors at the upper limits of simulated regional tumors, which are presumed to have 1–4 positive nodes.)	[0 – 5%] (1%)	[0 – 1%] (1%)	1%
8. Percent 5 nodes	Percent of biologically onset, non-LMP tumors which are assigned 5 positive lymph nodes at onset. (This simulates these tumors in the distant stage from their initiation in the model.)	[0 – 5%] (1%)	[2 – 4%] (1%)	2%
9. Mean Gamma	The Gompertz growth rate is assumed to have a gamma distribution across all onset tumors. This parameter is the mean of this gamma distribution (see text).	[0.01 – 0.2] (0.01)	[0.08 – 0.18] (0.01)	0.12
10. Var Gamma	The variance of the gamma distribution of Gompertz growth rates	[0.006 – 0.1] (0.001)	[0.01 – 0.05] (0.001)	0.012

*The wide range was used for initial sampling of parameter space; the focused range was used to focus sampling in a sub-region of parameter space closer to where the best solution was thought to be. See text for details.

Our objective in constructing this simulation model of breast cancer epidemiology was to have a computer model whose output mimics the contents of a cancer surveillance registry. Thus, we calibrated to breast cancer incidence data for 10 5-year age groups from 30–34 years to 75–79 years across the 26 calendar years 1975–2000, and 4 historical stages at detection (in situ, local, regional, distant) appearing in the Wisconsin Cancer



Reporting System and in the national SEER data base, exclusive of Iowa¹. Jointly, these comprise a surface of $(10 \times 26 \times 4) = 1040$ points at which model output should approximate data from the two (non-equal) surveillance data sets. Manipulating the inputs in order to maximize fit to this surface is a complex optimization problem with no unique objective function and no closed form solution. Accordingly, we approached calibration heuristically in several stages.

First, to reduce the dimensionality of the problem of comparing model output to known registry data we adjusted incidence rates to the year 2000 standard age population (for ages 30–79 years) and subjectively judged the fit of the model by comparing the 4 stage-specific incidence rate curves and the age-adjusted breast cancer mortality curve over the single years from 1975 to 2000. We concentrated not only on minimizing error in predicting rates (demanding approximately the same relative fit for each curve despite the large variation in absolute rates by stage), but also on the qualitative shapes of these curves compared to WCRS and SEER data.

We began calibration with a set of initial values for input parameters in Tables 1, 2 in [Screening Component](#), and 6. These tables jointly summarize all calibrated inputs to the simulation model and are the parameters controlling incidence rates. Initial values were based on literature searches and expert judgments. Factorial experiments were devised to examine the effects of varying combinations of parameters in reasonable ranges. The output of each combination of parameters was represented as a "calibration plot" such as shown in the first five panels of Figure 5. We assessed these plots "by eye", judging how closely model output fit WCRS and SEER data. A typical factorial experiment varying 3 parameters in a $2 \times 2 \times 4$ design took on the order of 3 days to run and analyze. This process led to modifications of the model structure and inputs. Once model output appeared to fit incidence curves well, the 4 baseline cure fractions (one for each stage) in the "No Adj Tx" column of Table 5 in [Treatment Component](#) were adjusted to bring age-adjusted breast cancer mortality rates predicted by the model to the levels of the observed data.

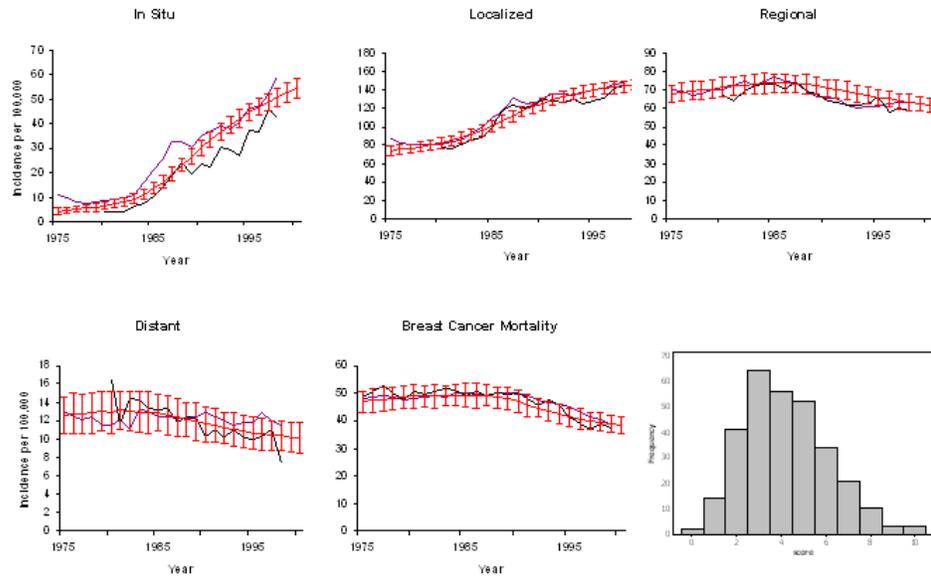


FIGURE 5. This is a calibration plot for the final model inputs. Four panels of the figure show incidence rates per 100,000 women (adjusted to population aged 30–79, US standard population for year 2000) for each stage, in situ, local, regional, or distant, and one panel shows breast cancer mortality rates. The scales have been adjusted to provide comparable visual assessment of fit across the panels; the absolute rates vary considerably across the panels. The two dark lines are data from SEER (excluding data from Iowa) and from the Wisconsin Cancer Registry. Iowa has been held out for later cross-validation as a state similar geographically and demographically to Wisconsin. Vertical bars indicate plus or minus 2 standard deviations for each annual point value across 300 replications using the fixed input parameters. The final panel is a histogram of the evaluation scores (as described in the text) for the 300 replications with the model inputs fixed.

For the next phase of calibration, we performed parameter sampling experiments. The purposes of parameter sampling were to explore whether a better fitting parameter combination(s) may be found than the one at which we arrived heuristically, to assess the likelihood of parameter combinations yielding good fitting solutions, and to be able to make uncertainty statements about our fitted parameter values. In this phase we specified wide ranges around the parameter values resulting from the first phase of calibration and sampled trial parameter values from uniform distributions in these ranges.

It was necessary to develop auxiliary computational tools to carry out the sampling experiments. The major tool was a function to automatically screen the model's "fit" resulting from a particular set of input parameters. Each unique combination of input parameters is a vector in the model parameter space (a 10-dimensional space if we consider only parameters in Table 6, or a higher dimension space if parameter values from Tables 1 and 2 in [Screening Component](#) are also considered). Let \vec{v} be a vector of input parameter values. We wished to evaluate a set of parameter vectors, $\vec{V} = \{\vec{v}_1, \vec{v}_2, \dots, \vec{v}_N\}$, where N is a large number (on the order of tens or hundreds of thousands), sampled uniformly in a pre-specified hypercube of parameter space (ranges for sampling are shown in Table 6). To do this we evaluated the fit of the



simulation output to observed surveillance data using each \vec{v}_i as input. Because there were many parameter vectors to evaluate, a method was required to automatically evaluate "fit", the evaluation we had been doing by "eye" in phase 1. Denote the fit of the simulation at \vec{v}_i by the function, $f(\vec{v}_i)$. The function f must evaluate the closeness of the simulation output to approximately 1040 non-independent points defined by 26 years of surveillance incidence data. Further, given \vec{v}_i , the output of the simulation is complex stochastically. Each execution of the simulation, a "replication", mimics detailed surveillance data from 1975–2000. We did not wish to fit observed data exactly because observed data are only one "sample" from a stochastic real world, but rather to find underlying parameters leading to behavior stochastically similar to observed surveillance data.

Accordingly, we were reluctant to define f solely in terms of an error function such as a sum of squared deviations across the 1040 points. Such an error function de-emphasizes incidence rates with small numerical values thus obscuring the importance of the trends from low to high rates in early stage cancers. Using a squared error penalty function resulted in representing age-adjusted incidence curves which do not "flatten" enough in later years compared to observed surveillance data. We instead elected to define f using acceptance envelopes around the four age-adjusted stage-specific incidence curves shown in Figure 6 and to count number of points at which the simulation output fell outside the envelopes. This equalizes importance of matching general shapes of the observed data curves regardless of the absolute numerical size of the rates in a given year. Envelope widths were defined using replication-to-replication variation in the simulation output (the variation is largely independent of the specific parameter values) such as shown in Figure 5. We set envelope widths to screen for potentially "good" input parameter combinations. That is, given an input parameter vector \vec{v}_i whose *mean* output across many replications would fall near the centers of the envelopes, we expected the 4 stage-specific output incidence curves from a `_single_replication` run would be unlikely to exceed the envelope often and a vector whose mean output was near the edges of the envelope was likely to generate data exceeding the envelope in many places on any given replication. We defined a scoring function, $f(\vec{v}_i)$, as the sum of the number of points at which the 4 curves from one replication executed using the input parameters \vec{v}_i fell outside the envelopes. With 26 years from 1975–2000 and 4 curves, there is a total of $26 \times 4 = 104$ points at which the output from one replication could exceed the envelopes, hence the range of the evaluation function was $0 \leq \vec{v}_i \leq 104$.

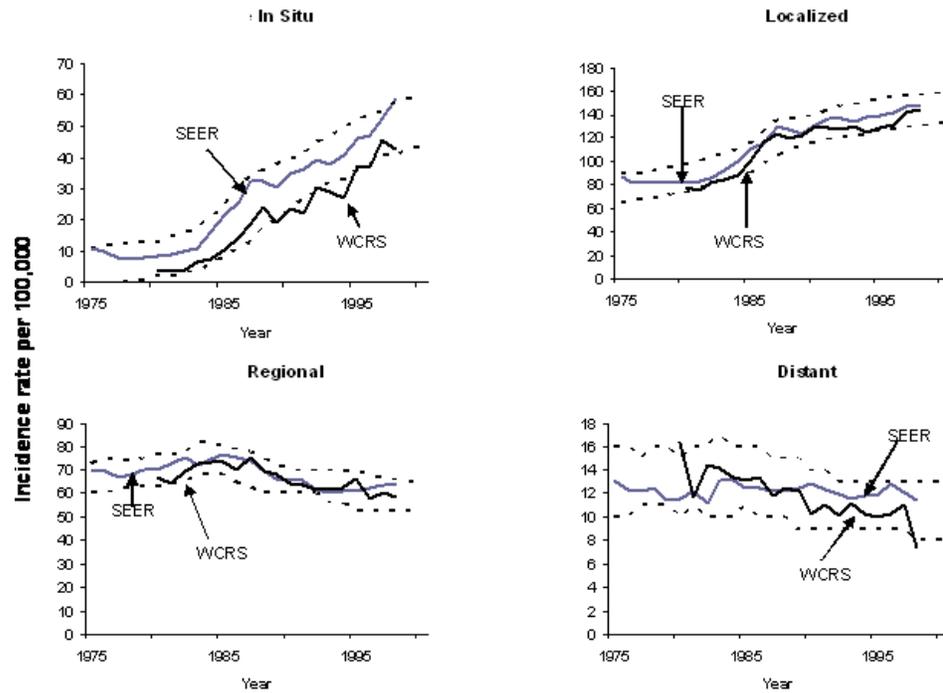


FIGURE 6. The dotted boundaries show the acceptance envelopes used for parameter sampling experiments. Incidence rates are age-adjusted to the U.S. population aged 30–79 in year 2000. The fit of a proposed input parameter vector was scored by counting excursions of one replication’s output of the simulation beyond the envelope bounds at any of the $26 \times 4 = 104$ possible annual points from 1975 to 2000. The envelope for in situ cancer is biased toward the SEER data since those represent a much larger number of cases. Generally envelope widths vary with wider envelopes being associated with lower, and hence more variable, rates.

We used the CONDOR environment for simultaneous execution of the simulation by a large pool of networked computers. With approximately 120 computers in the CONDOR pool to which we had access, we could evaluate approximately 1000–1500 sampled \vec{v}_i s per day.

Three "experiments" were conducted to calibrate and evaluate the model. Experiment 1 was designed to sample the parameter space broadly across all variables from Tables 1, 2 in [Screening Component](#), and 6. The curve cut-points in Tables 1 and 2 in [Screening Component](#) were constrained so that probabilities of detection increased with size of tumor. Table 6 shows ranges sampled. All initial sampling was from uniform distributions (or equal probability distributions for variables sampled discretely). After approximately 57,000 \vec{v}_i s were evaluated with the result being no score under 10 (i.e., $f(\vec{v}_i) \geq 0$ for all \vec{v}_i s evaluated), we constrained the LMP fraction to be under 10% for the next 15,000 samples to ensure dense sampling near zero for this potentially contentious variable. A total of 72,335 \vec{v}_i s were evaluated in this fashion with none scoring less than 10 and only 10 scoring under 16. For the next 30,000 samples we constrained LMP fraction to range from 30–55%, closer to our initial solution of 42%. Of the 30,188 uniformly sampled \vec{v}_i s a total of 363 yielded scores of 10 or less, 91 with a score of 5 or less, 15 with a score of 3, 4 scoring a 2, 1 scoring 1, and none scoring 0.

From Experiment 1, it appeared \vec{v}_i s leading to acceptably calibrated simulated incidence curves are rare in the parameter space defined by plausible marginal ranges of inputs. By sampling, we improved only slightly on our initial "by eye" solution, making small changes in parameters to match a \vec{v}_i with score of 2. We did not select the vector with a score of 1 because of the 4 vectors with score of 2, one of them had slightly better age-specific results than the one scoring 1 point better on age-adjusted evaluation. We also conclude that models with LMP fraction less than 30%, and especially those with LMP fraction near 0%, can be excluded for lack of fit to observed incidence, and this was in spite of testing a very wide range of other inputs to compensate.

Experiment 2 asked whether \vec{v}_i which were identified with low (good) scores in Experiment 1 generally had good solutions also existing in neighborhoods near them, and if those identified with higher (poorer) scores generally had poor solutions in their neighborhoods – i.e., does our scoring function f appear to separate neighborhoods well. We picked four of the \vec{v}_i s with scores in the 0–3 range ("good"), and four \vec{v}_i s with scores in the 11–15 range ("poor"), and sampled and scored an additional 500 \vec{v}_i in the neighborhood of each of these eight vectors. A "neighborhood" was defined by freezing the detection probabilities for each vector (parameters from Tables 1 and 2 in [Screening Component](#)), and sampling remaining parameters (from Table 6) within a range of $\pm 5\%$ of the original vector's values. Examining the distributions of scores from samples around these vectors we generally found about 30% of samples in the neighborhood of a "good" vector to have scores equal to or less than 10, and far fewer than 1% of samples around "poor" vectors to score at 10 or less. We concluded that using our scoring function, f , was a reasonable method to automate exploring the parameter space.

Although broadly speaking our acceptable solutions—i.e., the set of \vec{v}_i with scores less than 10—appear somewhat concentrated in the parameter space, they do not form a connected set. We believe there may be clusters of mutually non-identifiable parameters, hence we see a solution set with some marginal distributions being bi-modal or multimodal. Figure 7 presents marginal plots of selected parameters in the set of \vec{v}_i with scores less than 10. These may be interpreted roughly as posterior marginal distributions for the input parameter values.

Experiment 3 ran 300 replications using our "best fit" solution, denoted \vec{v}_i^* . These replications, when scored using f yielded a score range of 0–10; the distribution of scores is shown in the lower right hand panel of Figure 5. The calibration plot in Figure 5 was drawn using these replications of \vec{v}_i^* . Tables 1, 2 in [Screening Component](#), and 6 present the parameter values for \vec{v}_i^* . Note that \vec{v}_i^* is not simply a joint mean of the marginal distributions in Figure 7 because we have selected this vector based on having a score at the lower end of the range deemed "acceptable" for purposes of producing figure 7.

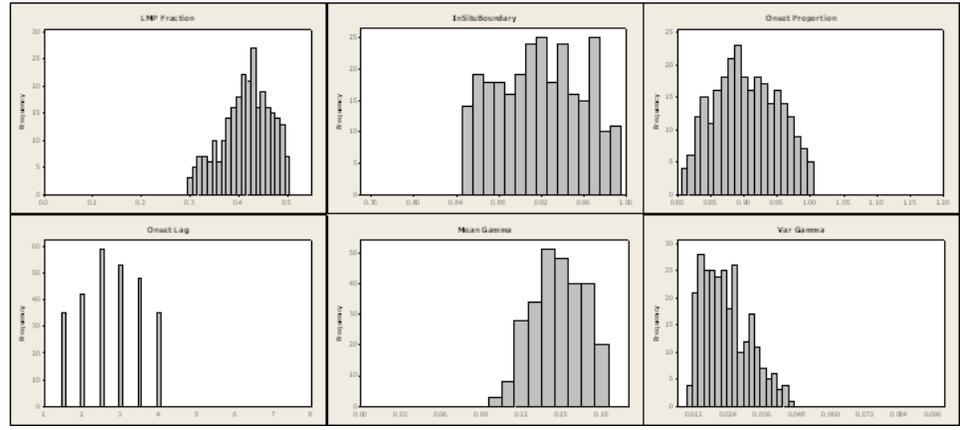


FIGURE 7. Marginal posterior distributions for six of the input parameters in Table 6. These histograms show values of the named parameters for input parameter vectors with scores of 10 or less in the sampling experiment. The ranges shown correspond to the ranges sampled for these parameters. Note that the parameters are not independent.

Additional outputs of the model have been compared to data where they exist (results not shown). For example the survival curves implied by the baseline cure probabilities are a good match to long-term survival published by Fisher, et al, in a follow-up study of modified mastectomy and lumpectomy treatments³. The stage distribution at detection as a function of tumor size is similar to SEER data¹. The age-specific prevalence of breast cancer in the year 1975 appears to match prevalence rates derived by NCI statisticians from SEER data (data not shown).

REFERENCES:

- ¹ Ries, LAG., Eisner, MP., Kosary, CL., Hankey, BF., Miller, BA., Clegg, L. "SEER Cancer Statistics Review, 1975-2000" in National Cancer Institute 2003;
- ² Fisher, B., Jeong, JH., Anderson, S., Bryant, J., Fisher, ER., Wolmark, N. "Twenty-five-year follow-up of a randomized trial comparing radical mastectomy, total mastectomy, and total mastectomy followed by irradiation" in *N Engl J Med* 2002; 347: 8: 567-75
- ³ Fisher, B., Anderson, S., Bryant, J., Margolese, RG., Deutsch, M., Fisher, ER. "Twenty-year follow-up of a randomized trial comparing total mastectomy, lumpectomy, and lumpectomy plus irradiation for the treatment of invasive breast cancer" in *N Engl J Med* 2002; 347: 16: 1233-41



POPULATION COMPONENT

SUMMARY

This document describes the method by which population is simulated in the model.

OVERVIEW

The model was designed to produce counts of incident tumors and breast cancer deaths over time to reproduce corresponding counts in the WCRS from 1980–1998, this being the interval for which WCRS is available and stable. Hence we scaled the simulation to replicate the life histories of a population of women from whom these counts would arise. Starting the model in 1950, we simulate a number of women equal to each single-year age cohort of women in the Wisconsin population aged 20–99. In each year 1951–1999 we add to the simulation model the number of women aged 20y in that year. (We chose age 20 under the assumption that breast cancer is so rare in younger women that we can ignore its occurrence before that age.)

DETAIL

The total number of women simulated is approximately 2.95 million. Each complete simulation of these 2.95 million women is one replication of the simulation and results in data equivalent to the breast cancer cases in the WCRS from 1978–1998 plus the years 1950–1977, and 1999–2000. One replication on a desktop computer with a Pentium 4 processor and 384 Mb of RAM requires approximately 10 minutes. We submit runs consisting of multiple replications, typically 10–50, using a large set of networked computers utilizing the CONDOR sharing software³.

We started the model in 1950 assuming no breast cancer is present in any woman living at that time. The breast cancer onset and progression submodels described above are invoked in 6 month cycles from simulated year 1950 to 1975 as "burn-in" for our model under the assumption that the prevalence has stabilized after the 25 year run-up.

For comparison of output to other CISNET collaborators, when we computed age-adjusted rates, we adjusted results to the US standard population (male and female) aged 30–79 in the year 2000.

REFERENCES:

-
- ¹ Livny, M, Basney, J, Raman, R, Tannenbaum, T "Mechanisms for high throughput computing" in SPEEDUP Journal 1997; 11:
 - ² Basney, J., Livny, M. "Chapter 5: Deploying a high throughput computing cluster" in In: Buyya R, editor
 - ³ Thain, D, Tannenbaum, T., Livny, M. "Condor and the grid. In: Berman F, Hey AJG, Fox G, editors" in Grid Computing: Making the Global Infrastructure a Reality
-



- Readers Guide
- Model Overview
- Assumption Overview
- Parameter Overview
- Component Overview
- Output Overview
- Results Overview
- Key References



OUTPUT DESCRIPTION

When a woman meets the exit criteria for the simulation (i.e., she dies, she reaches age 100, or the ending calendar year of the simulation is reached) her data are tallied in a series of event counters.

The primary counters accumulate counts relevant to numerators and denominators of output quantities of interest. Example counters are:

1. indexed by age and calendar year, counts instances of being alive and a given age in a given calendar year. This count is used as a denominator.
2. similarly indexed, counts instances of being alive and breast-cancer free at a given age in a given calendar year. This too is a potential denominator.
3. indexed by historical stage, age, and calendar year, and mode of detection (mammogram or clinical surfacing), counts instances of incidence breast cancer.
4. indexed by historical stage, age, and calendar year, and nature of the tumor (agressive breast cancer or breast cancer of limited malignant potential) tallies undetected breast cancers

Similar ad hoc counters can be implemented for special runs of the simulation.

The counts in these arrays are converted to rates per 100,000 at the end of 1 simulation replication of the entire population. These rates are then stored for N complete replications and the median, mean, and standard deviations are computed across the N replications.

These single-year outputs can be combined to show age-adjusted outputs across time (see [Results Overview](#))





KEY REFERENCES

- Abrams, K., Ashby, D., et al.** (1996) A Bayesian Approach to Weibull Survival Models – Application to a Cancer Clinical Trial. in *Lifetime Data Analysis 2*, p 159–174
- Ackerman, E.** (1994) Simulation of Micropopulations in Epidemiology: Tutorial 3. Simulation Model Evaluation Methods. in *International Journal of Bio-Medical Computing*:37, p 195–204
- Ackerman, E.** (1994) Simulation of Micropopulations in Epidemiology: Tutorial 1. Simulation: An Introduction. in *International Journal of Bio-Medical Computing*:36, p 229–238
- Ackerman, E.** (1994) Simulation of Micropopulations in Epidemiology: Tutorial 2. Analytic Forms of Event Probabilities. in *International Journal of Bio-Medical Computing 2*:37, p 139–149
- Ackerman, E.** (1994) Simulation of Micropopulations in Epidemiology: Tutorial 4. Evaluations of Simulation Models. in *International Journal of Bio-Medical Computing 2*:39, p 219–229
- Albertsen, PC., Hanley, JA., Gleason, DF., Barry, MJ.** (1998) Competing risk analysis of men aged 55 to 74 years at diagnosis managed conservatively for clinically localized prostate cancer in *Jama* 280:11, p 975–80
- Armstrong, K., Eisen, A., et al.** (2000) Assessing the Risk of Breast Cancer. in *New England Journal of Medicine* 8:342, p 564–571
- Ashba, J., Traish, AM.** (1999) Estrogen and progesterone receptor concentrations and prevalence of tumor hormonal phenotypes in older breast cancer patients in *Cancer Detect Prev* 23:3, p 238–44
- Basney, J., Livny, M.** Chapter 5: Deploying a high throughput computing cluster in *In: Buyya R, editor,*
- Bernstein, L., Ross, R. K.** (1993) Endogenous Hormones and Breast Cancer Risk. in *Epidemiologic Reviews* 1:15, p 48–65
- Black, W. C., Nease, R. F. J., et al.** (1996) Determining Transition Probabilities from Mortality Rates and Autopsy Findings. in *Medical Decision Making* 1:17, p 87–93
- Black, W. C., Welch, H. G.** (1997) Screening for Disease. in *American Journal of Radiology*:168, p 3–11
- Blumenson, L. E. Bross, I. D. J.** (1969) Mathematical Analysis of the Growth and Spread of Breast Cancer. in *Biometrics*, p 95–109
- Brooks, J. M., Chrischilles, E., et al.** (2000) Information Gained From Linking SEER Cancer Registry Data to State-Level Hospital Discharge Abstracts. in *Medical Care* 11:38, p 1131–1140
- Brown, M. L., Fintor, L.** (1993) Cost-Effectiveness of Breast Cancer Screening: Preliminary Results of a Systematic Review of the Literature. in *Breast Cancer Research and Treatment*:25, p 113–118
- Buchanan, J. B., Spratt, J. S., et al.** (1983) Tumor Growth, Doubling Times, and the Inability of the Radiologist to Diagnose Certain Cancers. in *Radiologic Clinics of North America* 1:21, p 115–126
- Bureau of Health Information, D. O. H. C. F.** (2000) Wisconsin Cancer Incidence and Mortality in *Wisconsin Department of Health and Family Services*, p 1–79
- Bureau of Health Information, Division of Health Care Financing, Wisconsin Department of Health and Family Services** (2002) Wisconsin Cancer Incidence and Mortality, 1999,

- Burke, W., Olsen, A. H. , et al** (2001) Misleading Presentation of Breast Cancer in Popular Magazines. in *Effective Clinical Practice* 2:4, p 58–64
- Cameron, DA., Ritchie, AA., Miller, WR.** (2001) The relative importance of proliferation and cell death in breast cancer growth and response to tamoxifen in *Eur J Cancer* 37:12, p 1545–53
- Chang, P.** (1993) A Simulation Study of Breast Cancer Epidemiology and Detection since 1982: The Case for Limited Malignant Potential Lesions [Ph.D.],
- Clemons, M. , Goss, P.** (2001) Estrogen and the Risk of Breast Cancer in *New England Journal of Medicine* 4:344, p 279–285
- Colditz, G. A., Rosner, B.** (2000) Cumulative Risk of Breast Cancer to Age 70 Years According to Risk Factor Status: Data From the Nurses' Health Study. in *American Journal of Epidemiology* 10:152, p 950–964
- Cronin, K** The dissemination of mammography in the US in *JNCI in press*,
- Cronin, K. A., Legler, J. M. , et al.** (1998) Assessing Uncertainty in Microsimulation Modelling With Applications to Cancer Screening Interventions. in *Statistics in Medicine*:17, p 2509–2523
- Crowley, TJ** (2000) Causes of Climate Change Over the Past 1000 Years in *Science* 289, p 270–7
- Csete, ME, Doyle, JC** (2002) Reverse Engineering of Biological Complexity in *Science* 295:5560, p 1664–1669
- Daley, J., Delbanco, T. L. , et al.** (1996) A 47–Year–Old Woman With Ductal Carcinoma In Situ, 1 Year Later [comment]. in *JAMA* 6:276, p 491
- DeVita** (1997) Staging of Breast Cancer. in *Cancer Principles & Practice of Oncology*, p 1569
- Draper, D.** (1995) Assessment and Propagation of Model Uncertainty. in *Journal of the Royal Statistical Society Series B* 1:57, p 45–97
- Drinkwater, M.** (2000) Superclusters—the Largest Structures in the Universe in *Science* 287:5456, p 1271–1218
- EBCTCG (Early Breast Cancer Trialists' Collaborative Group)** (1992) Systemic treatment of early breast cancer by hormonal, cytotoxic, or immune therapy. 133 randomised trials involving 31,000 recurrences and 24,000 deaths among 75,000 women in *Lancet* 339:8785, p 71–85
- Falkowski, P., Scholes, RJ., Boyle, E., Canadell, J., Canfield, D., Elser, J.** (2000) The Global Carbon Cycle: A Test of Our Knowledge of Earth as a System in *Science* 290:5490, p 291–296
- Feuer, E. J. , Wun, L.–M.** (1992) How Much of the Recent Risk in Breast Cancer Incidence Can Be Explained by Increases in Mammography Utilization? in *American Journal of Epidemiology* 12:136, p 1423–1436
- Fisher, B., Anderson, S., Bryant, J., Margolese, RG., Deutsch, M., Fisher, ER.** (2002) Twenty–year follow–up of a randomized trial comparing total mastectomy, lumpectomy, and lumpectomy plus irradiation for the treatment of invasive breast cancer in *N Engl J Med* 347:16, p 1233–41
- Fisher, B., Jeong, JH., Anderson, S., Bryant, J., Fisher, ER., Wolmark, N.** (2002) Twenty–five–year follow–up of a randomized trial comparing radical mastectomy, total mastectomy, and total mastectomy followed by irradiation in *N Engl J Med* 347:8, p 567–75
- Fisher, B., Slack, N. H. , et al.** (1969) Cancer of the Breast: Size of Neoplasm and Prognosis. in *Cancer* 5:24, p 1071–1080
- Fournier, D. V., Weber, E. , et al.** (1980) Growth Rate of 147 Mammary Carcinomas. in *Cancer*:45, p 2198–2207

- Fryback, DG, Stout, NK, Rosenberg, MA, Trentham–Dietz, A, Remington, PL, Kuruchittham, V** (2004) The case for limited malignant potential breast tumors: Implications of the Wisconsin breast cancer epidemiology simulation model,
- Gail, M. H., Brinton, L. A. , et al.** (1989) Projecting Individualized Probabilities of Developing Breast Cancer for White Females Who Are Being Examined Annually. in *Journal of the National Cancer Institute* 24:81, p 1879–1886
- Garber, J.** (1999) A 40–Year–Old Woman with a Strong Family History of Breast Cancer. in *JAMA* 20:282, p 1953–1960
- Gustafson, P.** (1998) Flexible Bayesian Modeling for Survival Data. in *Lifetime Data Analysis*:4, p 281–299
- Habbema, J. D. F., van Oortmarssen, G. J. , et al.** (1984) The MISCAN Simulation Program for the Evaluation of Screening for Disease. in *Computer Methods and Programs in Biomedicine*:20, p 79–93
- Haber, D.** (2000) Roads Leading to Breast Cancer. in *New England Journal of Medicine* 21:343, p 1566–1568
- Holford, TR., Cronin, K., Feuer, EJ., Mariotto, A.** (2003) Changing patterns in breast cancer incidence trends, manuscript,
- Kemeny, MM., Peterson, BL., Kornblith, AB., Muss, HB., Wheeler, J., Levine, E.** (2003) Barriers to clinical trial participation by older women with breast cancer in *J Clin Oncol* 21:12, p 2268–75
- Kerlikowske, K., Carney, P. A. , et al.** (2000) Performance of Screening Mammography Among Women With and Without a First–Degree Relative With Breast Cancer. in *Annals of Internal Medicine* 11:133, p 855–863
- Klawansky, S., Fox, M. C.** (1984) A Growth Rate Distribution Model for the Age Dependence of Human Cancer Incidence: A Proposed Role for Promotion in Cancer of the Lung and Breast. in *Journal of Theoretical Biology* 3:111, p 531–587
- Kopans, D. B.** (1996) Detection and Treatment of Ductal Carcinoma In Situ of the Breast [Letter]. in *JAMA* 11:276, p 896
- Koscielny, S., Tubiana, M. , et al.** (1985) A Simulation Model of the Natural History of Breast Cancer. in *British Journal of Cancer*:52, p 515–524
- Kusama, S., Spratt, J. S. , et al.** (1972) The Gross Rates of Growth of Human Mammary Carcinoma. in *Cancer* 2:30, p 594–599
- Lantz, P., Bunge, M., Remington, PL.** (1990) Trends in mammography in Wisconsin in *Wis Med J* 89:6, p 281–2
- Lantz, P. M., Remington, P. L. , et al.** (1991) Mammography Screening and Increased Incidence of Breast Cancer in Wisconsin. in *Journal of the National Cancer Institute* 21:83, p 1540–1546
- Lantz, PM., Weier, A., Remington, PL.** (1993) Mammography use in Wisconsin: 1987–1992 92:10, p 580–2
- Levin, D. L., Gail, M. H. , et al.** (1986) A Model for Projecting Cancer Incidence and Mortality in the Presence of Prevention, Screening, and Treatment Programs. in *NCI Monographs 2(Cancer Control Objectives for the Nation: 1985–2000)*, p 83–93
- Li, C. I., Anderson, B. O. , et al.** (2000) Changing Incidence Rate of Invasive Lobular Breast Carcinoma among Older Women. in *Cancer* 11:88, p 2561–2569
- Livny, M, Basney, J, Raman, R, Tannenbaum, T** (1997) Mechanisms for high throughput computing in *SPEEDUP Journal* 11,
- Loeve, F., Boer, R. , et al.** (1999) The MISCAN–COLON Simulation Model for the Evaluation of Colorectal Cancer Screening. in *Computers and Biomedical Research*:32, p 13–33



- Mariotto, A., Feuer, E.J., Harlan, L.C., Wun, L-M., Johnson, K., Abrams, J.** (2002) The dissemination of adjuvant therapy for breast cancer in the US: 1975–1997 in *The National Cancer Institute*,
- Markowska, J, Markowska, A.** (1998) Spontaneous tumor regression in *Ginekol Pol* 69:1, p 39–44
- McKay, M. D.** (1995) Evaluating Prediction Uncertainty in *Nuclear Regulatory Commission*,
- McPherson, K., Steel, C. M. , et al.** (2000) Breast Cancer — Epidemiology, Risk Factors, and Genetics. in *British Journal of Medicine* 7261:321, p 624–628
- Mezzetti, M., Robertson, C.** (1999) A Hierarchical Bayesian Approach to Age-Specific Back-Calculation of Cancer Incidence Rates. in *Statistics in Medicine*:18, p 919–933
- Michaelson, J. S., Halpern, E., et al.** (1999) Breast Cancer: Computer Simulation Method for Estimating Optimal Intervals for Screening. in *Radiology*:212, p 551–560
- Newcomb, P. A., Lantz, P. M.** (1993) Recent Trends in Breast Cancer Incidence, Mortality and Mammography. in *Breast Cancer Research and Treatment* 2:28, p 97–106
- Newcomb, PA., Carbone, PP.** (1993) Cancer treatment and age: patient perspectives in *J Natl Cancer Inst* 85:19, p 1580–4
- Osnes, K., Aalen, O. O.** (1999) Spatial Smoothing of Cancer Survival: A Bayesian Approach. in *Statistics in Medicine*:18, p 2077–2086
- Patz, E. F., Goodman, P. C. , et al.** (2000) Screening for Lung Cancer. in *New England Journal of Medicine* 22:343, p 1627–1633
- Piccart, MJ.** (1998) The changing landscape of breast cancer clinical research. ESMO–Award Lecture, ECCO–9 Hamburg, 18 September 1997 in *Ann Oncol* 9:2, p 133–8
- Pike, M. C., Krailo, M. D. , et al.** (1983) 'Hormonal' Risk Factors, 'Breast Tissue Age' and the Age–Incidence of Breast Cancer. in *Nature*:303, p 767–770
- Ries, LAG., Eisner, MP., Kosary, CL., Hankey, BF., Miller, BA., Clegg, L** (2003) SEER Cancer Statistics Review, 1975–2000 in *National Cancer Institute*,
- Rosenberg, MA** (2004) Competing risks in *JNCI Monographs*,
- Ross, CM.** (1991) DYNASIM2 and PRISM: Examples of dynamic modeling in *In: Citro, CF. Hanushek EA, editors. Improving Information for Social Policy Decisions: The Uses of Microsimulation Modeling II*,
- Rutter, C. M., Mandelson, M. T. , et al.** (2001) Changes in Breast Density Associated with Initiation, Discontinuation, and Continuing Use of Hormone Replacement Therapy. in *JAMA* 2:295, p 171–176
- Shwartz, M.** (1978) A Mathematical Model Used to Analyze Breast Cancer Screening Strategies. in *Operations Research* 6:26, p 937–955
- Shwartz, M.** (1978) An Analysis of the Benefits of Serial Screening for Breast Cancer Based Upon a Mathematical Model of the Disease. in *Cancer* 4:41, p 1550–1564
- Shwartz, M.** (1981) Validation and Use of a Mathematical Model to Estimate the Benefits of Screening Younger Women for Breast Cancer. in *Cancer Detection and Prevention*:4, p 595–601
- Shwartz, M.** (1980) Estimates of lead time and length bias in a breast cancer screening program in *Cancer* 46:4, p 844–51
- Shwartz, M.** (1975) Analysis of Preventive Screening to Detect Early Breast Cancer [Ph.D.],



- Smith, B. L.** (2000) Approaches to Breast Cancer Staging. in *New England Journal of Medicine* 8:342, p 580–581
- Soden ,BJ., Wetherald, RT., Stenchikov, GL., Robock, A.** (2002) Global Cooling After the Eruption of Mount Pinatubo: A Test of Climate Feedback by Water Vapor in *Science* 296:5568, p 727–730
- Speer, J. F., Petrosky, V. E. , et al.** (1984) A Stochastic Numerical Model of Breast Cancer Growth That Simulates Clinical Data. in *Cancer Research*:44, p 4124–4130
- Spiegelman, D., Colditz, G. A., et al.** (1994) Validation of the Gail et al. Model for Predicting Individual Breast Cancer Risk. in *Journal of the National Cancer Institute* 8:86, p 600–607
- Spratt, JS., Spratt, JA.** (2002) Chapter 21 Growth Rates. In: Donegan WL, editor in *Cancer of the Breast*, p 443–476
- Stout, NK** (2004) Quantifying the benefits and risks of mammography for women, researchers, and policy makers: An analysis using a simulation model.,
- Tabar, L., Duffy, S. W. , et al.** (1999) The Natural History of Breast Carcinoma: What Have We Learned for Screening? in *Cancer* 3:86, p 449–462
- Thain, D, Tannenbaum, T., Livny, M.** Condor and the grid. In: Berman F, Hey AJG, Fox G, editors in *Grid Computing: Making the Global Infrastructure a Reality*,
- Tsodikov, A. D., Asselain, B. , et al.** (1997) A Distribution of Tumor Size at Detection: An Application to Breast Cancer Data. in *Biometrics*:53, p 1495–1502
- Urban, N., Drescher, C. , et al.** (1997) Use of a Stochastic Simulation Model to Identify an Efficient Protocol for Ovarian Cancer Screening. in *Controlled Clinical Trials*:18, p 251–270
- Walter, L. C., Covinsky, K. E.** (2001) Cancer Screening in Elderly Patients: A Framework for Individualized Decision Making. in *JAMA* 21:285, p 2750–2756
- Welch, H. G., Black, W. C.** (1997) Using Autopsy Series to Estimate the Disease "Reservoir" for Ductal Carcinoma in Situ of the Breast: How Much More Breast Cancer Can We Find? in *Annals of Internal Medicine* 11:127, p 1023–1028
- Welch, H. G., Black, W. C.** (1997) Evaluating Randomized Trials of Screening. in *Journal of General Internal Medicine*:12, p 118–124
- Welch, H. G., Shwartz, L. M. , et al.** (2000) Are Increasing 5–Year Survival Rates Evidence of Success Against Cancer? in *JAMA* 22:283, p 2975–2978
- Welch, H. G.** (2001) Informed Choice in Cancer Screening. in *JAMA* 21:285, p 2776–2778
- Welch, HG., Black, WC.** (1997) Using autopsy series to estimate the disease "reservoir" for ductal carcinoma in situ of the breast: how much more breast cancer can we find? in *Ann Intern Med* 127:11, p 1023–8
- White, E.** (1987) Projected Changes in Breast Cancer Incidence Due to the Trend Toward Delayed Childbearing. in *American Journal of Public Health* 4:77, p 495–497
- Wilmoth, JR., Shkolnikov, V., [Directors]** (2003) Human Mortality Database. US Female Death Rates by Year of Birth,
- Woloshin, S., Shwartz, L. M. , et al.** (1999) Women's Perceptions of Breast Cancer Risk: How You Ask Matters. in *Medical Decision Making* 3:19, p 221–229
- Wyld, L., Reed, MW.** (2003) The need for targeted research into breast cancer in the elderly in *Br J Surg* 90:4, p 388–99



University of Wisconsin
Key References

van den Akker-van Marle, E., de Koning, H. , et al. (1999) Reduction in Breast Cancer Mortality Due to the Introduction of Mass Screening in the Netherlands: Comparison with the United Kingdom. in *Journal of Medical Screening*:6, p 30-34
