PERGAMON

Computers in Biology and Medicine 33 (2003) 345–360

Computers in Biology and Medicine

www.elsevier.com/locate/compbiomed

A computer model for the study of breast cancer

Kimbroe J. Carter*, Frank Castro, Edward Kessler, Barbara Erickson

St. Elizabeth Health Center, The Northeastern Ohio Universities College of Medicine, Youngstown, Rootstown, OH, USA

Received 23 August 2002; received in revised form 25 November 2002

Abstract

A computer model was designed as a relational database to assess breast cancer screening in a cohort of women where the growth and development of breast cancer originates with the first malignant cell. The concepts of thresholds for growth, axillary spread, and distant sites are integrated. With tumor diagnosis, staging was performed that includes clinical and sub-clinical states. The model was parameterized to have staging characteristics similar to data published by the Surveillance, Epidemiology, and End-Results (SEER) Program. Validation was accomplished by comparing simulated staging results with non-SEER sources, and simulated survival with independent clinical survival data.

Keywords: Breast cancer; Breast cancer disease; Breast cancer screening; Breast cancer staging; Computer simulation

1. Introduction

Screening for the presence of breast cancer has become a standard in the medical care of women. The recommendations by different organizations vary from one to another, but each recommends screening in a clearly defined manner as listed in Table 1. However, such rigid recommendations may not be applicable to a diverse group of women. Previous studies have used simulation of risk to examine the characteristics of breast cancer. For example, Eddy has used computer applications that solve the integral equations in his model as well as a time-varying Markov chain [1]. Shwartz [2] analyzed breast cancer screening with a more disease specific approach using exclusively probabilities of events. Habbema et al. [3] developed a simulation method based largely on Monte Carlo and

* Corresponding author. Kimbroe Carter, 539 Hickory Hollow Canfield, OH 44406, USA. Tel.: +1-330-533-3093.
E-mail address: gcrv44a@zoominternet.net (K.J. Carter).

0010-4825/03/$ - see front matter
doi:10.1016/S0010-4825(03)00003-9

Table 1
Guidelines for screening strategies for breast cancer

	Self-exam	Professional-exam	Mammogram
ACP	Not recommended	>Age 40, q12 months	>Age 50, q12 months
ACS	>Age 20, monthly	Ages 20–40, q 36 months; >Age 40, q12 months	Ages 40–49, q12 or 24 months; >Age 50, q12 months
USPSTF	Not recommended	>Age 40, q12 months	Ages 50–75, q12 or 24 months

ACP = American College of Physicians; ACS = American Cancer Society; USPSTF = United States Preventative Services Task Force.

transition probabilities. But none of the above investigators consider the onset of cancer at the first malignant cell. While the present model contains probabilistic aspects, these represent only a portion of the model. The present model was designed to study the onset, growth, and progression of breast cancer over the lifetime of individuals in a cohort employing a relational database design. The current presentation embodies the concept that breast cancers have a finite beginning, namely the first cancer cell and a well-defined end. Within the framework of this concept, breast cancer has Gompertzian growth characteristics. This growth pattern was different from that reported by others as being exponential [2], which indeed may be approximated in a region of a Gompertz growth curve. The model required the incorporation of age dependent incidence and mortality. Moreover, the model considered that once a tumor began to grow, it could spread to the axillary nodes and metastasize to more distant organs. Thresholds based on the volume of the primary tumor defined the appearance of these events. A screening protocol was superimposed on the simulated progression of the disease. Factors considered in the screening protocol included the frequency of screening and the sensitivities and specificities of the screening tests employed. The program used published data when possible although some needed data could only be estimated. From these considerations, a model was developed that demonstrated the screening process and underlying disease as dynamic events.

2. Flow diagram of model

The idea as relating tumor development and growth to screening are illustrated in Fig. 1. The onset of cancer was taken as the appearance of the first malignant cell. The assignment of this cell required inclusion of the incidence at the individual's age. Once the first malignant cell was initiated tumor growth began. As growth continued the disease could progress from in-situ to invasive and spread to regional and distant sites. Each specific event's threshold controlled the progression from one site to another. These progression characteristics were assigned at the time the first malignant cell was determined to be present. Following diagnosis, staging of the tumor was performed.

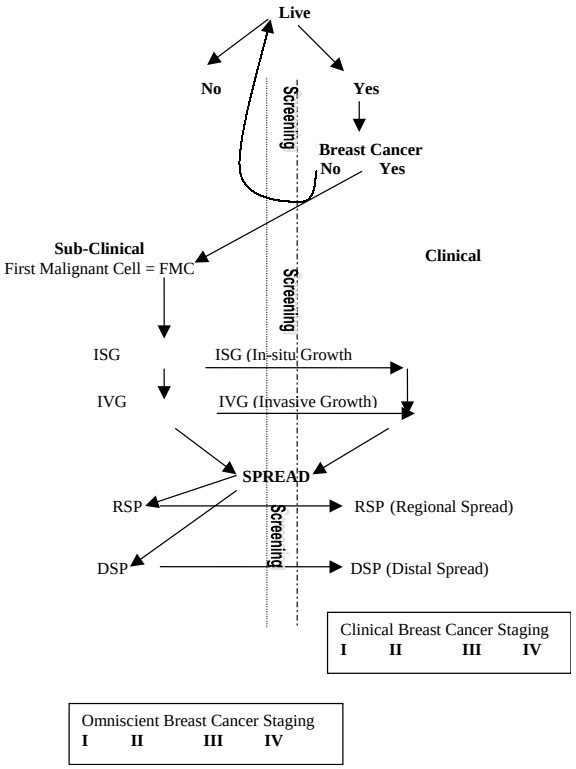

Fig. 1. Flow diagram of breast cancer screening model. Observed was the monthly reiteration of events until the first malignant cell appears. Subsequent to this was the growth and possible spread of the breast cancer until clinical staging is performed. The model includes omniscient staging referring to sub-clinical disease not yet diagnosable.

3. Software considerations

The model was structured in Microsoft ACCESS 2000, a relational database system. Within this application several key program components, including input and result data, exist as illustrated in Fig. 2. The client cohort contains the class of code controlling the cohort server. Additionally input values as defined by the user characterize a specific cohort. Given the specified input, a cohort server component is instantiated. This active server processes each individual through a simulated lifetime. The accumulated information for each simulated individual is managed as a transaction and stored in data tables in a hard drive. Standard results are obtained by queries of the stored data, and the results are used by the client cohort to generate reports.

Table schemas were based on relational database concepts to minimize redundancy. Selected data fields formed keys for tables to facilitate data searches. Forms were designed to collect data used to define the characteristics of cohorts and control data processing. To circumvent database size limitations, conserve data storage space, avoid database corruption, and create a centralized repository for simulation results, a client-server configuration was developed within the local programming environment. Upon execution of the program, the client database creates a supporting cohort server

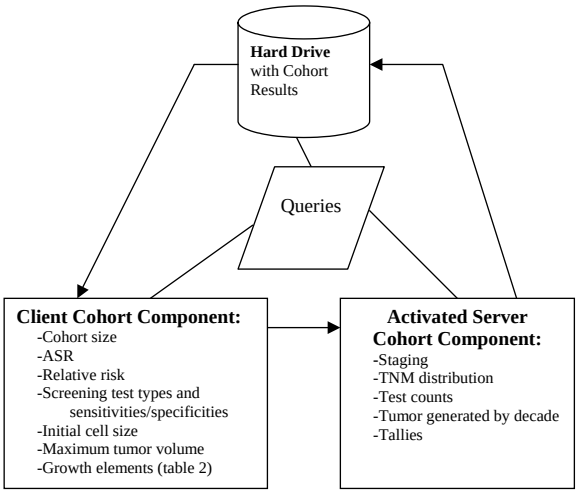

Fig. 2. The configuration of the client–server relationships and the data used in simulating a cohort.

database, and then the client gathers the data from the forms, and transfers this information along with all other necessary database objects to the cohort server. The server then processes the defined cohort, performs all necessary computations, and transfers the results back to the client database where the resultant data was stored. Having obtained the cohort results from the server, the client may then delete the cohort database. Upon completion of a simulation, the client automatically generates a standard report. Reports for specific cohort data also can be obtained on demand. Using automation techniques the program can be set to run a cohort repetitively. The number of computations performed for a cohort of 10,000 individuals requires up to 10 h on a Pentium IV personal computer.

4. Incidence and general mortality

Incidence and mortality, which were followed on a monthly basis, changed with advancing age. To determine if the individual was living and whether cancer would appear a Monte Carlo process was employed. The general mortality, adjusted monthly from actuarial tables based on age, sex, and, race, was taken as the survival threshold for that month [4]. A value was selected from a standard uniform distribution. When that value was greater than the threshold, the individual remained living. Similarly, considering the incidence of disease, the presence or absence of the first cancer cell was determined. When a number from the standard uniform distribution fell between zero and the specific incidence, cancer was initiated.

5. First malignant cell

To establish a standard starting point for tumor development, published incidence data was adjusted to determine first malignant cell (FMC) incidence [5]. Using the median Gompertzian r-value of the breast cancer population and a 1 cm^3 tumor volume, a standard lead-time of approximately 93

Table 2
Tumor progression thresholds for model simulation

Model components	Data source	Value or median (cm^3)	95% Range
Cell size	Constant	1.00E-9	None
Maximum tumor volume	Constant	1.00E + 3	None
Primary tumor growth rate	Lognormal distribution	1.52E-2	(4.23E-3 to 5.49E-2)
Primary tumor symptomatic threshold	Lognormal distribution	5.46E + 1	(7.39 + 0 to 4.03E + 2)
Invasive threshold	Lognormal distribution	1.11E-2	(2.25E-5 to 47E + 0)
Axillary node threshold	Lognormal distribution	1.16E-1	(5.57E-4 to 2.39E + 1)
Axillary node symptomatic threshold	Lognormal distribution	1.00E + 0	(2.02E-2 to 4.94E + 1)
Axillary node diagnostic threshold	Constant	1.00E-6	None
Metastatic threshold	Lognormal distribution	2.36E + 1	(1.20E-1 to 4.45E + 3)
Metastatic growth rate multiplier	Normal distribution	2.50E + 0	(1.50E + 0 to 3.50E + 0)
Metastatic symptomatic threshold	Lognormal distribution	1.00E + 0	(2.02E-2 to 4.94E + 1)
Metastatic diagnostic threshold	Constant	6.54E-2	None

months for the growth from the first malignant cell was calculated, corresponding with the 8-year average growth time as reported by Norton [6]. The incidence data were then adjusted for a given onset 93 months earlier. The shifted incidence was used to determine thresholds for the appearance of the first malignant cell.

6. Mathematics of tumor growth

Tumor growth is one of the most important aspects of the natural progression of breast cancer. The Gompertzian function was used to model tumor growth [6–8]. The Gompertz model for tumor growth can be written as follows:

$$V(t) = V_0 e^{a(1-e^{-rt})},$$

where, $V(t)$ represents the tumor volume at time t; V_0 represents the tumor volume at time zero and was the assumed initial cell diameter of 12.4 μm, and $a = \ln(V_{max}/V_0)$, where V_{max} represents the maximum tumor volume, of 1000 cm^3. The value of r for each individual in the model was a constant determined under Gompertzian growth assumptions. This r-value distribution comes from published data derived from changes in tumor volumes obtained radiographically over a defined time interval. In a cohort of women with breast cancer studied by Norton the r-values were lognormally distributed [6].

Thresholds for the events related to tumor progression in the model are listed in Table 2. These events for growth and spread have either lognormal or normal distributions. The individual thresholds chosen from these distributions were determined using the Box and Muller equation, $y = \mu + \sigma\sqrt{-2\ln(R_1)}\cos(2\pi R_2)$ where y was normally distributed with a mean μ and the standard deviation σ, and R_1 and R_2 are random numbers obtained from a uniform distribution [9,10]. When

necessary, lognormality was obtained by the natural logarithmic transformation of the derived normal value y.

7. In-situ growth

During the in-situ growth phase the only permitted tumor progression event was the development of local invasion. In the absence of data to indicate when this progression might occur, an empirical estimation was made. It was assumed the invasive threshold had a lognormal distribution. Means and variances were systematically changed until a distribution was found consistent with breast cancer staging data reported by the Surveillance, Epidemiology, and End Results Program (SEER) data [11].

8. Regional spread of tumor

The threshold distribution for axillary lymph node spread was derived from published data [12].

The published distributions related primary tumor volume to axillary node tumor positivity as found at surgical excision. However, no data directly related primary tumor volume to the volume of tumor found in the axilla or to the initiation of spread to the axilla. To evaluate when the first malignant cell appeared in the axilla two assumptions were made: (1) 1.0E-6 cm^3 was the minimum detectable tumor volume in the axilla; and (2) axillary tumor growth had Gompertzian characteristics similar to the primary tumor.

Under these assumptions, back extrapolations were performed, first to determine the length of time tumor was present in the axilla. This time was used to calculate the volume of the primary tumor when the first malignant cell appeared in the axilla. This computed primary tumor value represented the threshold for spread to the axilla. This threshold was used monthly to determine initiation of the first malignant cell in the axilla.

9. Distal spread

Over a period of several decades, prior to the use of chemotherapy, Koscielny published data that related primary tumor size variability to the frequency of distant metastasis [13]. From these observations, a threshold distribution for the appearance of distant metastasis was obtained. Using Koscielny's distant metastasis threshold distribution and the Box and Muller selection technique, modeled individuals with breast cancer were assigned a metastatic threshold value. This threshold value for the appearance of the distant metastasis was then compared monthly to the primary tumor volume. As the primary tumor volume reached the predetermined metastatic threshold value, a distant first malignant cell was considered to be present. This distant first malignant cell was assigned Gompertzian growth characteristics. Tumors in distant sites generally grow faster than that of the primary tumor, ranging between 1.5 and 3.5 times that of the primary tumor [14]. This variability was represented as a normal distribution. Sampling from this distribution provided a multiplier that was applied to the r-value of the primary tumor.

10. Compliance

As used by the model, compliance was the average lifetime adherence of a cohort to the recommended interval for test performance. The user of the simulation decides how many groups, up to four, to simulate; the number of individuals in the cohort to assign to each group; and the compliance within each group. To determine into which group an individual belonged, a Monte Carlo based process was utilized.

The pre-assigned compliance threshold controlled, for life, the screening compliance of all members of a group. Whether an individual met or missed a scheduled test was also determined using a Monte Carlo process. A random value was obtained from a standard uniform distribution. A value falling between zero and the compliance threshold represented the completion of a scheduled test; while a value exceeding the threshold represented a missed scheduled test.

11. Screening and diagnostic testing

A screening protocol consisted of a set of tests, recommended intervals between tests and the ages at which testing would begin and end. To simulate this variability, the model was programmed to examine the various screening tests, their sequences and frequencies of performance. Further the model examined test outcomes based on specific levels of compliance to a screening protocol. If a test was negative, the testing sequence was continued per protocol. If the test was positive, confirmatory testing was performed. For example, in the ACS recommendations for breast cancer screening, a positive breast self-examination prompted a professional examination; a positive professional examination prompted mammography, which when positive led to a fine needle biopsy which if positive, was confirmed by an open biopsy. However, if the primary tumor exceeded 1 cm^3 on a professional examination, biopsy was performed, bypassing mammography. Following a positive open biopsy the cancer was staged.

The five tests in the model were assigned values for sensitivity and specificity as obtained from published data [15–19]. The test's sensitivities were dependent upon the primary tumor volume while specificities were independent of tumor volume. Where the tumor was too small to be detected, the sensitivity was zero. Sensitivity increased as tumor size increased to a maximum as listed in Table 3. The sensitivity, based on the individual's underlying primary tumor volume at the time of testing, was used as the threshold to determine test positivity. When the random number, obtained from a standard uniform distribution, was greater than the positivity threshold the test was considered to be negative. The model compared a test's result to the presence or absence of tumor to ascertain test result validity.

12. Staging

Staging criteria as defined by the American Joint Committee on Cancer (AJCC) were employed in the model [11]. For the model, staging was an informational system that included primary tumor, regional node, and metastatic characteristics of the disease regardless of tumor volume. Since simulated disease begins with the first malignant cell, both sub-clinical and clinical states are known, and

Table 3
Optimal screening tests characteristics

Test	Sensitivity (percent)	Specificity (percent)
Self breast exam[a]	80	75
Professional breast exam[b]	90	80
Mammography[c]	71/89[d]	93
Fine needle biopsy	95	95
Open biopsy	100	100

[a]Tumor volume dependent, maximum at 33.5 cm^3.
[b]Tumor volume dependent, maximum at 33.5 cm^3.
[c]Tumor volume dependent, maximum a 1 cm^3.
[d]Premenopause/menopause.

the model extends AJCC clinical staging criteria to include sub-clinical disease. The model's awareness of sub-clinical disease was defined as omniscient. At the conclusion of a simulation, primary tumor, regional and distal spread data were recorded in tables and this data was used to categorize individuals into one of five stages based on sub-clinical and clinical disease. In the report phase of the program, these tables are queried to produce cohort distributions of stage components.

13. Life expectancy

The method of determining mortality in the group that never develops breast cancer was described earlier. For the group that develops breast cancer, the determination of mortality was more complicated. In those individuals in whom Stage IV disease was apparent, a life expectancy of 2 years [20] was assigned provided that the normal life expectancy would be greater than 2 years. The model also considers that the distant metastasis may be sub-clinical at the time of primary tumor diagnosis. Life expectancy in this case was assigned as the time until the distant tumor becomes clinically diagnosable, defined in the model as 1 cm^3, plus 2 years, provided that this exceeded the normal life expectancy, Table 7. The model was constructed such that omniscient Stages 0–III do not progress to Stage IV based on the assumption that treatment resulted in a cure. The model permitted development of a second primary tumor that was followed in the same manner as the original primary tumor.

14. Validation

Validation was based on the concept that model results should correspond to known staging data under similar assumptions. TNM staging was important in model validation since it accounts for the presence of primary, axillary and distant tumor. SEER data from 1983 to 1987 were used for parameterization of the model [21]. After parameterization, the simulated stage data compared favorably with NCDB 12, and Fields et al. [22], Table 8. Furthermore, survival data taken independently from a breast cancer tumor registry of 1800 cases between 1981 and 1999 [23] was compared to

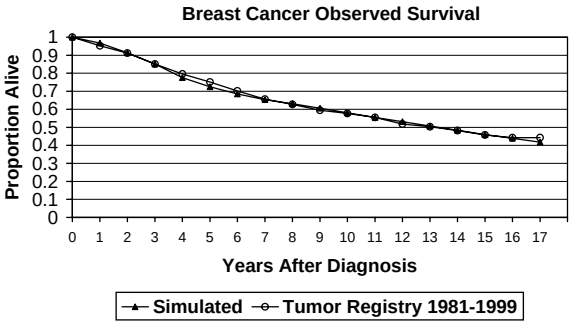

Fig. 3. Survival curves from the breast cancer tumor registry of 1800 patients from 1981 to 1999 compared to simulated survival. The compliance with the ACS recommendations for the tumor registry data was calculated and found to approximate 17%. Accordingly the simulated survival was also performed at the level of 17% compliance. Gehan–Wilcoxon analysis showed $X^2 = 1.27$, $p > 0.26$.

simulated survival, Fig. 3. To obtain equivalent comparison it was first necessary to estimate the level of compliance in the tumor registry data. Using this compliance estimate the simulation for the survival analysis was then run at that level. The close relationship between the tumor registry and the simulated survival curves provides further validation for the model. Using NCSS 97 software, Gehan–Wilcoxon analysis of the two curves was not significantly different ($X^2 = 1.27$, $p > 0.26$).

15. Demonstration of a cancer case in the model

To illustrate the volume of data processed and the interactions of tumor growth and screening, a case is presented that traces the course of the individual from birth through staging of a breast cancer. The ACS breast cancer screening protocol was employed. This analysis was divided into 3 phases where phase 1 represents the absence of cancer; phase 2 where tumor was present but below the limits of mammographic detection; and phase 3 the period of tumor growth from a volume capable of mammographic detection to the point of staging. Tables 4 and 5

detail these phases and illustrate the computations needed each month to define health states.

During phase 1 the model processed monthly the status of screening as follows: (1) whether the screening test was to be scheduled according to protocol, (2) whether there was compliance with the performance of the scheduled tests and (3) the results of the tests performed. Table 4 summarizes the extent of these computations. Phase 2 begins with the first malignant cell on month 754 and continues through month 832. As noted in Table 5, in the last month of phase 2, the tumor volume was still below the mammographic threshold of detectability. The important events that transpired in phase 2 and the numbers of computations are given in Tables 4 and 5. Phase 3 began on month 833 and continued through month 875 at which point the primary tumor was discovered through screening. The primary tumor throughout phase 3 exceeded the mammographic detection threshold of 6.54E-2 cm^3.

The final sequence of screening tests that led to the discovery of the primary tumor was as follows. During phase 3 only 6 of the 43 monthly scheduled self-breast examinations were performed. Three

Table 4
History of testing in a selected case

Phase one[a]		Monthly mortality determinations	Monthly incidence determinations for first cancer	Monthly self-breast exam assessments starting at month 240	Professional breast exam assessments starting at month 240	Mammography assessments starting at month 480
Months 0–753	Threshold range	5.19E-6 to 7.45E-4	0.0 to 4.83E-3	NA	NA	NA
	Scheduled assessments	754	754	514	43	23
	Scheduled exams done+	NA	NA	129	11	6
	Follow up exams done	NA	NA	0	32	8
	Specificity	NA	NA	0.75	0.80	0.93
	Sensitivity	NA	NA	NA	NA	NA
	False positives	NA	NA	32	8	0
Phase two[b]			Assessments for second cancer			
Months 754–832	Threshold range	7.52E-4 to 1.45E-3	2.66E-2 to 2.74E-2	NA	NA	NA
	Scheduled assessments	79	79	79	6	6
	Scheduled exams done+	NA	NA	24	2	1
	Follow up exams done	NA	NA	0	0	0
	Specificity	NA	NA	NA	NA	NA
	Sensitivity	NA	NA	0.0	0.0	0.0
	False negatives	NA	NA	24	2	1

Phase three[c]	Threshold range	1.46E-3 to 2.07E-3	2.74E-2 to 2.70E-2	NA	NA	NA
Months 833–875	Scheduled assessments	43	43	43	4	3
	Scheduled exams done	NA	NA	6	1	0
	Follow up exams	NA	NA	0	3	0
	Specificity	NA	NA	NA	NA	NA
	Sensitivity	NA	NA	0.0–0.50	0.0–0.65	0.68–0.89
	False negatives	NA	NA	3	2	0

[a]Phase one represents months when no breast cancer tumor was present.
[b]Phase two represents time when tumor volume was below the limits capable of detection based against mammography 6.54E-2 cm^3.
[c]Phase three represents time that tumor volume was within limits capable of detection to staging.

of these exams were positive prompting professional breast exams. Four scheduled professional exams were processed, one of which was performed, plus three follow up exams. Two of these professional exams were positive. At the time of the first positive professional exam the tumor volume was 1.88 cm^3 and on the second exam was 4.11 cm^3. During this growth interval, 3 mammographies were scheduled but none were performed. Since on both positive professional exams, the primary tumor volume exceeded the 1 cm^3 threshold, mammograms were bypassed and testing proceeded directly to fine needle biopsy. The initial fine needle assessment at 865 months resulted in a false negative result. The second fine needle biopsy at 875 months was positive and resulted in confirmation by an open biopsy.

Using the data in Table 5, this case was Stage IV. However because the metastatic volume 1.51E-3 cm^3 was below the diagnostic threshold of 6.54E-2 cm^3 and therefore was not detectable, the case was classified clinically as Stage II. The dichotomous view of this case's staging results was due to the application of data filters. With data filters applied only clinical staging was obtained while in the absence of data filters omniscient breast cancer staging was achieved.

The possibility of self-referral was another event analyzed monthly starting with the first malignant cell. The specific tumor volume thresholds for self-referral are in Table 6. However, the self-referral threshold values were not exceeded in the case illustrated. Had any of the thresholds for self-referral been exceeded, staging would have been performed.

An estimate was made of years lost to breast cancer based on the individual's age and the distal tumor volume as noted in Table 7. Given a distal tumor volume of 1.51E-3 cm^3, 20 months were estimated to reach a distal tumor detectable volume of 1 cm^3. Since the median survival with metastatic cancer was 24 months, the case presented has only a total of 44 months of additional life at the time of staging [20]. Had this breast cancer been curable, indicating no distant metastatic disease, life expectancy would have been 168 months. Thus, 124 months were lost due to the presence of distant metastatic disease. Table 8

Table 5
History of tumor development, progression and staging in a selected case

		Invasive event	Axillary event	Metastatic event	Potential detectability	Gompertzian *r*-factor
Phase two	Threshold	3.25E-4 cm^3	1.45E-3 cm^3	4.21E-1 cm^3	6.54E-2 cm^3	NA
Months 754–832	Primary tumor volume at 754 months (1.0E-9 cm^3)	NA	NA	NA	Not detectable	4.39E-4 (initial DT 57.9 days)
	Primary tumor volume at 801 months (3.91E-4 cm^3)	Exceeded threshold	NA	NA	Not detectable	NA
	Primary tumor volume at 808 months (1.46E-3 cm^3)	Exceeded threshold	Exceeded threshold	NA	Not detectable	3.90E-4 (initial DT 65.1 days)
	Primary tumor volume at 832 months (5.81E-2 cm^3)	Exceeded threshold	1.17E-6 cm^3	NA	Not detectable	NA
	Count	48	55	79	NA	NA
Phase three	Primary tumor volume at 833 months (6.61E-2 cm^3)	Exceeded threshold	1.50E-6 cm^3	NA	Detectable	NA
Months 833–875	Primary tumor volume at 849 months (4.21E-1 cm^3)	Exceeded threshold	5.67E-5 cm^3	Exceeded threshold	Detectable	9.12E-4 (initial DT 27.9 days)
	Primary tumor volume at 875 months (4.11 cm^3)	Exceeded threshold	5.47E-3 cm^3 (staging threshold 1.00E-6)	1.51E-3 cm^3 (staging threshold 6.54E-2)	Clinical stage II omniscient stage IV	NA
	Count	NA	NA	17	NA	NA

Table 6
Assessments of self-referral in a selected case

	Primary tumor	Axillary tumor	Distal tumor
Threshold for self-referral (cm^3)	48.9	2.78	1.46
Count	122	122	122

Table 7
Assessments of life expectancy in a selected case

	Assessments	Months
Life expectancy with cure	Determined from actuarial table	168
Life expectancy with progression of disease	Growth of distal tumor from time of clinical staging (1.51E-3 cm^3 at 875 months) to 1 cm^3[a]	20
	Subsequent death from Stage IV disease	24
Life lost to breast cancer	NA	124

[a] 1 cm^3 of distal tumor represents the threshold between sub-clinical and clinical disease.

Table 8
Percentage distribution of stages at the time of clinical diagnosis

Stage	Simulated	SEER	NCDB	Fields et al.
0	7.8	9.1	11.3	12.4
I	34.6	32.5	42.9	41.8
II	44.3	45.2	31.9	33.1
III	4.1	6.0	7.7	8.0
IV	9.2	7.3	6.2	4.7

Simulated data calibrated to approximate SEER data.

16. Discussion

Designing the model presented a challenge due to the numerous events occurring over a lifetime of a woman's participation in breast cancer screening. Meeting this challenge required an understanding of the tenets of screening, biology of breast cancer and concepts of computer programming. Ultimately this knowledge was abstracted into sets of instructions coding the orderly progression of a woman's life related to breast cancer. Thousands of logical decisions and computations for every case in a cohort are processed. Computational emphasis was placed on the defining of events related to screening protocol, compliance, testing, as well as the growth and development of breast cancer. A small change in even a single variable may result in significant changes in results, due to the interaction of computations performed. The complexities of the relationships require extensive data processing, data storage and retrieval capabilities. Analysis of these data sets permits insight

into interactions of variables defined in the model. That the attributes employed in the structure of the model have validity was emphasized by the close relationship of the survival curves of the simulations and that obtained from the data of an independent source.

By examining breast cancer from its onset at the first malignant cell, it was possible to view also the appearance of single cells as harbingers of new events such as axillary and distant spread. Importantly it was noted that the spread of tumors was evident before they could be detected by current technologies. The reason lies in the fact that each of the events in the model has a threshold below which it was present but not evident. Thus, the model considers that sub-clinical spread may occur even before a primary tumor was detectable. The results of simulation help to explain a dichotomy between clinical staging of breast cancer and the true health state.

A standard report of a simulated cohort summaries many relationships. For example it contains interval cancers statistics, distributions of stages according to stage category, the distribution of the components of TNM classifications, and cohort survival. Further the model can forecast the economic feasibility of a screening program when applied to a large population. In addition, with appropriate programming, the model can evaluate the effect of a pharmacologic drug on tumor growth. As new technology develops the model can assess its impact on screening protocols and the effects on the distribution of staging and life-years lost to cancer.

Clinical breast cancer screening studies require extensive amounts of time, are expensive, and are difficult to control. By contrast, the model can control characteristics that clinical studies inherently have difficulty controlling. Moreover information can be obtained more rapidly. As such this model can be used as a tool to assist in the efficient design of clinical studies.

Within the current public recommendations for breast cancer screening, the use of a rigid approach does not account for population diversity. This model is flexible enough to address this diversity. Interested investigators can download a web version of the model at http://www.jeghers.com/mdmweb.

17. Summary

The model's structure was illustrated showing the relationships of tumor growth and development and the interaction with a screening protocol. The concepts are implemented in a relational database. The mathematics of growing tumors starting with the first malignant cell was presented. Further growth and development of tumor are based on concepts of thresholds that indicate the appearance of spread. From the screening of a cohort, according to a defined protocol, staging of tumor was possible. Two cases are described showing the dynamic interaction of staging and tumor growth. Staging, as an information system, identified both clinical and sub-clinical components. Validation of the model was established by comparison of simulated stage data with independent staging data and also by comparing simulated survival with the survival obtained from the cancer registry of a teaching hospital.

Acknowledgements

This research was supported in part through grants from the Saint Elizabeth Development Foundation, Saint Elizabeth Health Center Medical Research Committee and the Northeastern Ohio

Universities College of Medicine, without limitation in designing the study, interpreting data, writing and publishing this paper. The authors wish to thank Leonard P. Caccamo, MD (deceased), Lori M. Gawdyda, MLS, Terry Lisko, Nathan Ritchey, Ph.D., Kenneth Vinciquerra, MA, and Karen Zapko, MSN, RN for their assistance in this project

References

[1] D.M. Eddy, Screening For Cancer. Theory, Analysis, and Design, Prentice-Hall, Inc., Englewood Cliffs, NJ, 1980, p. 161.
[2] M. Shwartz, An analysis of the benefits of serial screening for breast cancer based upon a mathematical model of the disease, Cancer 41 (1978) 1550–1564.
[3] J.D.F. Habbema, G.J. van Oortmarssen, J.Th.N. Lubbe, P.J van der Maas, The MISCAN simulation program for the evaluation of screening for disease, Comput. Methods Programs Biomed. 20 (1984) 79–93.
[4] Vital Statistics of the United States, Rockville (MD), U.S. Department of Health, Education, and Welfare, Public Health Service, Health Resources Administration, National Center for Health Statistics Division of Vital Statistics, SEER Cancer Statistics Review, Tables 6-2, 1973–1995.
[5] National Cancer Institute, SEER Cancer Statistics Review, Table 4-2, 1973–1995.
[6] L. Norton, A Gompertzian model of human breast cancer growth, Cancer Res. 48 (1988) 7067–7071.
[7] C.P. Winsor, The Gompertz curve as a growth curve, Proc. Natl. Acad. Sci. 18 (1932) 1–8.
[8] V.G. Vaidya, F.J. Alexandro, Evaluation of some mathematical models for tumor growth, Int. J. Biomed. Comput. 13 (1982) 19–35.
[9] S. Ross, A First Course in Probability, 3rd Edition, Macmillan, New York, NY, 1988, pp. 386–394.
[10] H.A. Taha, Operations Research: Introduction, 6th Edition, Prentice-Hall, Upper Saddle River, NJ, 1997, p. 688.
[11] American Joint Committee on Cancer, AJCC Cancer Staging Manual, 5th Edition, Lippincott Williams & Wilkins, Philadelphia, PA, 1997, pp. 172–175.
[12] S. Koscielny, M.G. Le, M. Tubiana, The natural history of human breast cancer. The relationship between involvement of axillary lymph nodes and the initiation of distant metastases, Br. J. Cancer 59 (1989) 775–782.
[13] S. Koscielny, M. Tubiana, M.G. Le, A.J. Valleron, H. Mouriesse, G. Contesso, D. Sarrazin, Breast cancer: relationship between the size of the primary tumour and the probability of metastatic dissemination, Br. J. Cancer 49 (1984) 709–715.
[14] J.S. Myers, R.W. McDivitt, Reliability and stability of the thymidine labeling index of breast carcinoma, Lab Invest 54 (2) (1986) 160–164.
[15] S.W. Fletcher, M.S. O'Malley, L.A. Bunce, Physicians' abilities to detect lumps in silicone breast models, J. Am. Med. Assoc. 253 (15) (1985) 2224–2228.
[16] S.W. Fletcher, M.S. O'Malley, C.A. Pilgrim, J.J. Gonzalez, How do women compare with internal medicine residents in breast lump detection? A study with silicone models, J. Gen. Internal Med. 4 (4) (1989) 277–283.
[17] G. Hermann, C. Janus, I.S. Schwartz, B. Krivisky, S. Bier, J.G. Rabinowitz, Nonpalpable breast lesions: accuracy of prebiopsy mammographic diagnosis, Radiology 165 (1987) 323–326.
[18] D.N. Rutledge, Sensitivity and specificity of lump detection in breast models, Am. J. Prev. Med. 8 (5) (1992) 314–318.
[19] M.R. Sarfati, F.A. Fox, J.A. Warneke, L.L. Fajardo, G.C. Hunter, W.D. Rappaport, Stereotiactic fine-needle aspiration cytology of nonpalpable breast lesions: an analysis of 258 consecutive aspirates, Am. J. Surg. 168 (1994) 529–532.
[20] A. Coates, V. Gebski, J.F. Bishop, et al., Improving the quality of life during chemotherapy for advanced breast cancer: a comparison of intermittent and continuous treatment strategies, N. Engl. J. Med. 317 (1987) 1490–1495.
[21] American Joint Committee on Cancer, Manual for Staging of Cancer, 4th Edition J.B. Lippincott, Philadelphia, 1992, p. 151.
[22] K.K. Fields, S.C. Goldstein, R.A. Clark, D.M. Sullivan, Breast Cancer, Decision Making In Oncology Evidence-Based Management, Churchill Livingstone, Philadelphia, 1997, p. 253.
[23] Data on Breast Cancer, Cancer Registry St. Elizabeth Health Center, Youngstown, OH, 1981–1999.

Kimbroe Carter is Assistant Professor of Pathology at Northeastern Ohio Universities College of Medicine and a member of the staff of Saint Elizabeth Health Center, Youngstown, OH. He received his medical degree at Tulane University Medical School. He is currently a graduate student in School of Technology at Kent State University with emphasis on computer technology. His research interests are currently in breast cancer computer simulation.

Frank Castro is a member of the local chapter of the Society for Medical Decision Making. He holds a master's degree in mathematics from Youngstown State University. He currently is a director of logistics for a local firm. His research activities are in the fields of renal disease, medical decision-making and breast cancer.

Edward Kessler is Professor of Internal Medicine at Northeastern Ohio Universities College of Medicine and a member of the staff of Saint Elizabeth Health Center, Youngstown, OH. His medical degree is from Saint Louis University College of Medicine. He was previously Clinical Associate of Medicine at the University of Pittsburgh, Pittsburgh, PA. He is a Fellow of the American College of Physicians. His research interests are in renal disease, medical decision-making and breast cancer.

Barbara Erickson holds a Ph.D. in nursing from the Union Institute and University, Cincinnati, OH and is a member of the local chapter of the Society for Medical Decision Making. Her research interests are in cardiovascular disease, medical decision-making and in breast cancer.