Modeling Tumor Growth and Quantifying the Duration of Time between Metastasis, Detection, and Mortality in Breast Cancer Patients

Daniel Pollack

Roslyn High School

Under the direction of James Michaelson, Ph.D.

Research Science Institute 2011

Personal Section

Since I have always loved both math and science, I was eager to integrate the two fields by conducting a medical research project in applied mathematics. Within the medical field, what inspired me to study cancer were the deaths of my great-grandmother and great-aunt, as well as the death of my former teacher, math team coach, and research mentor, Iftimie Simion, who helped me come to view math not merely as a subject in school, but rather as a beautiful puzzle. His passion for mathematics, coupled with his tragic death, inspired me to research cancer with the hope of improving both tumor detection and treatment. As his former research student, I felt I owed it to his memory to use the mathematics he taught me in an effort to improve our understanding of the disease that took his life far too soon.

My research was conducted under the supervision of Dr. James Michaelson, the Scientific Director at the Laboratory for Quantitative Medicine, in affiliation with Massachusetts General Hospital and Harvard Medical School. A majority of my research was conducted during a summer internship with the Research Science Institute (RSI) held at the Massachusetts Institute of Technology (MIT) in June and July of 2011. I continued to work on my paper and project through December of 2011.

My advice for other high school research students would be to enjoy every part of the research process. Research something you love, and rather than feeling discouraged when you face a challenge or unexpected result, look at it as an opportunity to learn, grow, and work with even greater excitement and determination in search of an answer. Rather than only seeing value in the final result you hope to discover, try to enjoy everything about the process along the way. There is so much to learn through research besides that final result; the skills and knowledge you gain will remain with you long afterwards - and may perhaps even come to use when you least expect it.

One of the most valuable aspects of my experiences with high school research, I believe, was the opportunity to develop a greater understanding and appreciation for the way scientific research is conducted. Though the process can often be long and challenging, the opportunity to problem-solve and think critically to answer unsolved real-world questions is simultaneously humbling, exciting, and incredibly rewarding. Unlike knowledge gained from books and courses, the unique characteristic of research is an eternal quest for the unknown. And the extraordinary characteristic of the unknown is the potential to make the world a better place.

Research Section

Abstract

A greater quantitative understanding of tumor growth is essential to improve cancer screening and treatment. Using a sample of 345 breast cancer patients from California, a novel algorithm was developed to propose a new mathematical model of tumor growth, utilizing essential information from the pre-detection period. Concurrently, this study quantified the median time intervals between metastasis, detection, and mortality by comparing probable time distributions of each event. To develop an accurate theoretical model of pre-detection growth, an iterative approach of parameter optimization was employed. Variations among anatomical locations of metastasis were also examined. When tumors of all localities were considered together, this study found the median time between metastasis and detection was 53.1 months and the median time between metastasis

and mortality was 73.6 months. Perhaps more significantly, the mathematical model developed by this study offers insight into the patterns of tumor growth prior to the attainment of detectable size. These findings are critical for the development of more quantitatively precise cancer treatment regimens and the improvement of various mathematical models of cancer during the pre-detection period. Such information also underscores the need to motivate more frequent cancer screening and develop more sensitive technologies to detect tumors of smaller diameters.

1 Introduction

To improve both cancer screening and treatment, a quantitative understanding of tumor growth is essential [8, 23, 24, 27]. By establishing a novel theoretical model of breast cancer tumor growth, this study dramatically improves previously established probability models of cancer lethality by incorporating essential information from the pre-detection period of tumor growth. With this information, physicians may be able to administer more effective and quantitatively precise treatment regimens by basing the intensiveness of therapy directly on the likelihood that patients experience a lethal metastasis. By quantifying the amount of time prior to mortality among patients with breast cancer metastases in various locations, the results of this study may also enable physicians to predict the likelihood of patient survival more precisely over time among victims of breast cancer metastasis [9]. Furthermore, by quantifying the duration of time between metastasis onset and detection, the results of this study suggest a great need to motivate more frequent cancer screening and to develop more sensitive technologies to detect tumors of smaller diameters.

More generally, tumor growth models might provide an understanding of the factors that naturally retard cancer proliferation, which could be utilized in treatment through molecular growth control [10, 33]. Additionally, since Lee and Spratt [14] showed that tumor growth rates are approximately equal to regression and regrowth rates after radiation, hormones, and chemotherapy, accurate models of tumor growth may allow physicians to predict post-treatment rates of tumor regression and relapse more reliably.

1.1 Modeling Cancer and the Probability of Metastasis

With the understanding that individual patterns may vary, previous studies have aimed to model tumor growth at the population level [38]. As Michaelson indicates, the discrete nature of microscopic biological events gives rise to quantitatively predictable cancer growth and metastasis patterns at the macroscopic level of tumors [2, 16, 18, 19]. Similarly, these quantitative patterns at the individual level give rise to quantitatively predictable patterns at the population level.

One such pattern is the probability of metastasis with respect to tumor size, which has been described by Michaelson's Size-Only Lethality Equation (Equation 1) [2, 18, 19]:

$$L(D) = 1 - e^{-QD^Z}. \tag{1}$$

The model predicts the probabilistic risk L of an eventually lethal metastasis as a function of primary tumor diameter D [2, 18, 19]. Through linear regression of tumor size and lethality data with respect to a transformation of Equation 1, constants Q and Z were empirically determined to be 0.0062 and 1.33, respectively [19]. By replacing the variable D with a function $D(t)$ to represent the size of a primary tumor as a function of time, the probability of metastasis can therefore also be described by a function of time.

1.2 Modeling Tumor Growth

Since cancer cells reproduce by mitosis, the unbounded growth rate is theoretically exponential, exhibiting a constant doubling time [32]. The lognormality of tumor growth rates under certain circumstances reinforces this supposition [5, 30]; however, tumor growth is often poorly fit by logarithmic axes [10]. In practice, the exponential model is only accurate for tumors of small sizes because growth is density-dependent, decelerating as cancer progresses. Early theories about the cause of this growth deceleration have included apoptosis, cell cycle lengthening, and the body's immunological response to cancer [11, 12, 15], whereas more modern theories include the inter-

related factors of growth inhibitory molecules, quiescence, and vascular and metabolic constraints on tumor growth [7, 16].

To address rate deceleration in the exponential model (Equation 2), alternatives such as the logistic model (Equation 3) or the Gompertz model (Equation 4) are often employed, where y_0 is an initial value, k is a rate constant, M is an upper asymptote in the latter two equations, and N is a constant in Equation 2 [23, 27, 39, 40]. In the case of tumor growth, y_0 and M both refer to tumor size expressed as diameter, volume, or number of cancer cells.

$$\frac{dy}{dt} = yk \qquad \text{or} \qquad y(t) = y_0 e^{kt} \tag{2}$$

$$\frac{dy}{dt} = yk\left(1 - \left(\frac{y}{M}\right)^N\right) \qquad \text{or} \qquad y(t) = \frac{M}{\sqrt[N]{1 + Ae^{-Nkt}}},\ \ A = \left(\frac{M}{y_0}\right)^N - 1 \tag{3}$$

$$\frac{dy}{dt} = yk\left(Ae^{-kt}\right) \qquad \text{or} \qquad y(t) = Me^{-Ae^{-kt}},\ \ A = \ln\left(\frac{M}{y_0}\right) \tag{4}$$

Applying these equations to mammographic data of the progression of breast cancer tumor volume with respect to time, Spratt, von Fournier, Spratt, and Weber [28] showed that the logistic function with $N = \frac{1}{4}$ modeled the data most accurately, followed closely by other variations of the logistic and Gompertz functions. It has also been independently demonstrated that the Gompertz function is a useful predictor of tumor growth [1, 23, 36, 37].

1.3 Purpose and Rationale

Although such models provide a general understanding of tumor growth during the post-detection period [3, 6, 36, 37], they may be poor predictors of tumor growth during the pre-detection period. Furthermore, whereas models from previous studies have quantified tumor growth using data from mammograms [22, 26, 29, 31], radiograms [25, 35], and physical examinations [13, 34], measurements obtained directly upon tumor excision can more accurately describe tumor

growth and improve patient treatment [4]. Using data obtained directly upon tumor excision, this study proposed a novel mathematical and compuational methodology to model pre-detection tumor growth and to determine the standard time intervals between metastasis, detection, and mortality at various cancer sites in the body.

2 Materials and Methods

Data were obtained from a sample of 345 breast cancer patients treated in California from 1956 to 2007, and all analyses were completed by an original algorithm developed in MATLAB. Only patients who exhibited metastasis and did not experience neoadjuvant treatment were considered.

Since tumors do not become operationally detectable until they reach a diameter of approximately seven millimeters [17], knowledge of the exact time that a tumor first appears is unattainable. It is possible, however, to approximate the time that a metastasis first appears indirectly based on the size of the primary tumor from which it is derived [2, 18, 19]. Using this information, it is possible to create a probability distribution to quantify the average likelihood of metastasis for a sample of n patients over time before tumor excision, which was defined for all patients as time t_0. Comparing this distribution to other probability distributions obtained from the same dataset, including the range of times of metastasis detection and the range of times of patient mortality, it is possible to quantify the amount of time required for cancer to grow from a single cell to a tumor of detectable size and the amount of time metastases may be present in the body before mortality.

Using the algorithm described on the following pages, the durations of time between metastasis, detection, and mortality were quantified both for the general sample and for specific subsets of patients based on the anatomical locations to which the primary breast tumors metastasized. Simultaneously, the algorithm solved for the parameters of a novel mathematical model of breast cancer tumor growth that incorporates essential information concerning tumor size during the pre-detection period.

2.1 Probability of Metastasis as a Function of Time

To create the distribution of probable times of metastasis, probability functions $L_i(t)$ were calculated from Equation 1 for each patient i. In each equation, the variable D was substituted by a function $D(t)$ in order to incorporate time rather than size as the independent variable:

$$L_i(t) = 1 - e^{-QD(t)^Z}. \tag{5}$$

To calculate the probability with respect to time that a primary tumor metastasized before it was detected and removed, the function $D(t)$ was required to model pre-detection tumor growth. Since previous models of tumor growth, however, have only incorporated the available data of post-detection size, a new model was necessary for the purposes of this study. Although the earlier stages of tumor growth appear more exponential than the latter stages due to the density-dependence factor, exponential growth would be an imprecise model because it assumes a constant doubling time (Equation 2). While the initial volume doubling time should be comparable to the cell cycle time lasting approximately 24 hours, the volume doubling time in a tumor at the time of detection and excision would be approximately 130 days [17, 20]. The generalized logistic model (Equation 3) would therefore offer a more reasonable approximation of tumor growth during the pre-detectable period. Through an iterative process described in Subsection 2.6, the parameters were optimized to develop a novel model of tumor growth in conjunction with these calculations.

2.2 Probability Distribution of Metastasis Time

For each patient i, the logistic model $D(t)$ was horizontally translated such that time t_0 corresponded to the tumor diameter in millimeters upon excision. Thus, each function $L_i(t)$ traced the probability of metastasis from the moment that the primary tumor was a single cell to the moment that the primary tumor was removed at t_0. Unlike in previous studies, pathological tumor size, rather than radiological data, was employed because the measurements obtained from mammograms and x-rays are approximated from tumor shadows, whereas the sizes recorded upon excision

are directly measured. The individual functions for each patient were then averaged according to Equation 6 to obtain a mean function $L(t)$, representing the average probability of metastasis with respect to time before tumor excision:

$$L(t) = \frac{1}{n} \sum_{i=1}^{n} L_i(t). \tag{6}$$

Because all patients in the database exhibited a distant recurrence after the primary breast tumor was removed, the probability that a metastasis occurred before tumor excision was necessarily 100%. Thus, to normalize the function as an average per-patient cumulative distribution of probable times of tumor metastasis $P_{Metastasis}(t)$, the function was multiplied by a constant k, where $k = \frac{1}{L(0)}$, according to Equation 7:

$$P_{Metastasis}(t) = kL(t). \tag{7}$$

All functions were stored as vectors of probability values with indices corresponding to those of their corresponding time values in separate vectors. Subtracting the average cumulative probability of metastasis $P_{Metastasis}$ at each time value t from the probability at the time value immediately following, it was possible to obtain a probability function $p_{Metastasis}(t)$ representing the likelihood of metastasis with respect to time in the interval between the two consecutive time values. This vector included a probability value for each month in the domain before t_0.

2.3 Probability Distributions of Detection and Mortality

Probability distributions of metastasis detection times $p_{Detection}$ and, if applicable, patient mortality times $p_{Mortality}$ were then calculated with respect to time after t_0, which was once again defined as the date of primary tumor excision. Only the 281 patients who had not survived until the conclusion of data collection were included in calculations involving the duration of time between metastasis and mortality. Although the database actually reported the duration of time between primary tumor biopsy, metastasis detection, and patient mortality, the date of primary tumor biopsy

was taken to be the time of excision t_0 because no data were available for the time of excision. For the purposes of this study, such an assumption was justifiable because the time between biopsy and excision, which has a mean of 3.4 months, is negligible when compared to the time scale between metastasis, detection, and mortality [21].

2.4 Deconvolution of the Probability Distributions

The distributions of the duration of time between tumor appearance and detection and between tumor appearance and patient mortality were calculated according to Equations 8 and 9:

$$p_{DetectionInterval}(t) : \ p_{DetectionInterval}(t_j - t_i) = [p_{Metastasis}(t_i)][p_{Detection}(t_j)]; \tag{8}$$

$$p_{MortalityInterval}(t) : \ p_{MortalityInterval}(t_j - t_i) = [p_{Metastasis}(t_i)][p_{Mortality}(t_j)]. \tag{9}$$

Each element t_j in the time vector for detection or mortality was subtracted by each element t_i in the time vector for metastasis. Correspondingly, the probabilities of these time intervals $(t_j - t_i)$ were calculated by multiplying the individual probabilities of t_i and t_j to define the new probability functions, $p_{DetectionIinterval}(t)$ (Equation 8) and $p_{MortalityInterval}(t)$ (Equation 9).

2.5 Statistical Calculations

Based on these probability distributions, the program was then designed to plot the cumulative distributions of the duration of time between metastasis and detection $P_{DetectionInterval}(t)$ or between metastasis and mortality $P_{MortalityInterval}(t)$. The mean time interval was calculated by adding the products of each time value and its corresponding probability (Equation 10), and the median was identified by locating the 50^{th} percentile of the cumulative distribution. The standard deviation was obtained by Equation 11.

$$\mu = \sum tp(t) \tag{10}$$

$$\text{SD} = \sqrt{\sum t^2 p(t) - \mu^2} \tag{11}$$

2.6 Iterative Optimization of Parameters in Logistic Model

When calculating $L_i(t)$ (Equation 5) to obtain the probability distribution of metastasis times, $D(t)$ must reflect the size of the primary tumor with respect to time between origin and excision. Although previous studies have established tumor growth models based on empirical data of tumor size in the post-detection period [3, 6, 28, 36, 37], no empirical knowledge was available concerning tumor size in this pre-detection period. Thus, to identify the most accurate model for the purposes of this study, an iterative appraoch was developed to optimize parameters N and k in the logistic model (Equation 3). Since volume is directly proportional to the number of cells in a tumor, the logistic model was applied to the volume (Equation 12) and converted to diameter (Equation 13).

The maximum value M was defined as 1.1×10^6 mm^3, corresponding to 2^{40} cells, since this value has been established as the theoretical maximum of tumor growth [28]. The initial condition V_0 was set at 10^{-6} mm^3, the volume of a single cell [28].

$$V(t) = \frac{M}{\sqrt[N]{1 + Ae^{-Nkt}}}, \quad A = \left(\frac{M}{V_0}\right)^N - 1 \tag{12}$$

$$D(t) = 2\sqrt[3]{\frac{3V(t)}{4\pi}} \tag{13}$$

To optimize N and k, the algorithm was designed to search for values that, when substituted into the model and iterated through the process previously described, would result in an overall median duration of time between origin and detection of a metastasis that was most nearly identical to the median duration of time that the model would predict between the origin and detection of a primary tumor. By iterating through a reasonable interval of values for both parameters and min-

imizing the difference between the two medians that would result from a given pair of parameters, the most ideal values for N and k in the pre-detection logistic model were determined.

3 Results

3.1 Optimization of Parameters

After iterating through the full domain of N and k, the absolute values of the differences between the median time interval between origin and detection in the primary tumor and the overall median time interval between origin and detection in the metastasis were stored in a matrix Z according to Equation 14:

$$Z(i,j) = \mid \text{Median}_{primary}(k_i, N_j) - \text{Median}_{metastasis}(k_i, N_j) \mid . \tag{14}$$

Z was then plotted with respect to N and k as a surface plot (Figure 1) to identify the general effect of varying the values of N and k on the difference between medians.

The global minimum of Z with respect to N and k was then located to determine the optimal pair of parameters for all calculations. As a frame of reference, the four pairs of parameters resulting in the next smallest Z values were also identified. The mean and median time intervals between metastasis and detection were then calculated for each pair of parameters to determine the sensitivity of the results to the selection of any of these parameter pairs (Table 1).

As the mean overall intervals between metastasis and detection for all five pairs of parameters were all within 0.118 SD of the mean when the optimal pair of parameters, defined by the global minimum of Z, were used (SD $= 47.5$ months), it was determined that the parameters $k = 0.17$ and $N = 0.007$ were acceptable to be used for all results. Substituting these parameters into Equation 12, the following model of tumor growth with respect to time t in months was obtained:

$$V(t) = \frac{1.1 \times 10^6 \text{ mm}^3}{\sqrt[0.17]{1 + 110.4e^{-0.03621t}}}. \tag{15}$$

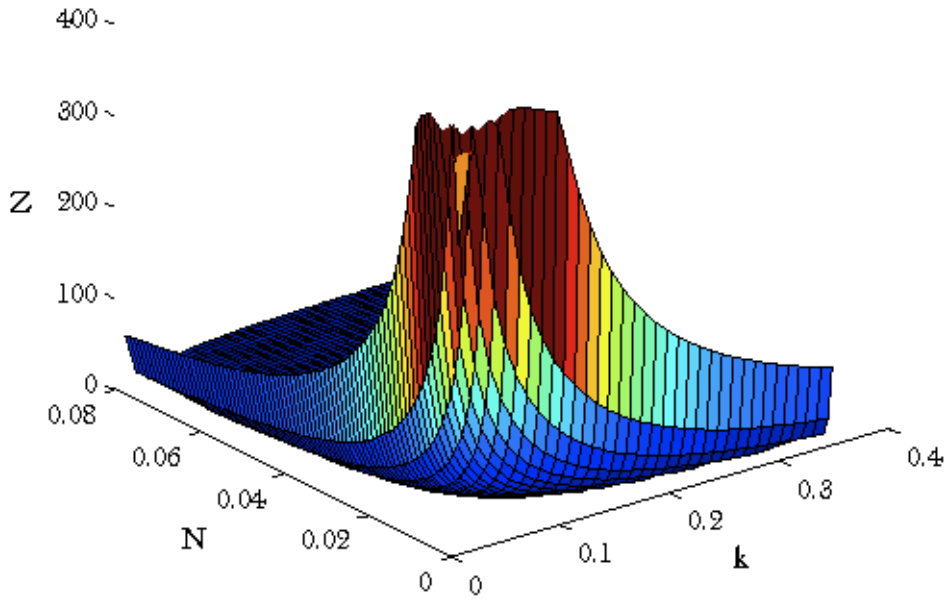

Figure 1: Surface plot of Z with respect to $k \in [0, 0.35]$ and $N \in [0, 0.080]$.

Table 1: Mean and median calculated time intervals between metastasis and detection (with percent deviations from the results based on the optimal pair of parameters listed in the first row) for the five pairs of parameters that resulted in the smallest values of Z.

Parameters	Results (Months)	Deviation
k = 0.17 N = 0.007 Z = 0.0986	Median = 53.1 Mean = 64.7	— —
k = 0.05 N = 0.029 Z = 0.1753	Median = 50.2 Mean = 62.1	5.46% 4.02%
k = 0.20 N = 0.006 Z = 0.1863	Median = 52.8 Mean = 64.5	0.565% 0.309%
k = 0.03 N = 0.063 Z = 0.1863	Median = 46.8 Mean = 59.1	11.9% 8.66%
k = 0.15 N = 0.008 Z = 0.2055	Median = 52.8 Mean = 64.4	0.565% 0.464%

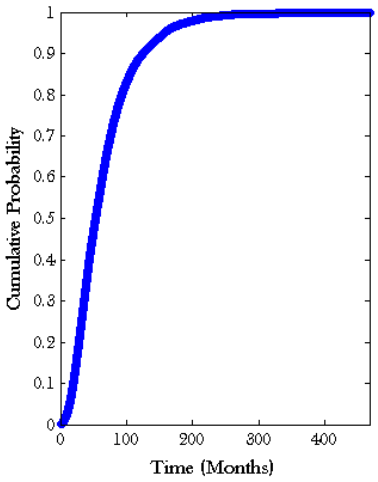

Figure 2: $P_{DetectionInterval}(t)$: Cumulative distribution of the overall time interval between metastasis and detection when all tumor localities were considered together. Mean: 64.7. Median: 53.1. SD: 47.5.

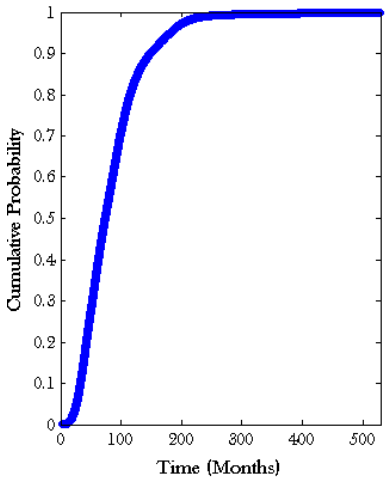

Figure 3: $P_{MortalityInterval}(t)$: Cumulative distribution of the overall time interval between metastasis and mortality when all tumor localities were considered together. Mean: 83.7. Median: 73.6. SD: 50.8.

3.2 Time Intervals Between Metastasis, Detection, and Mortality

After executing the algorithm using $D(t)$ defined with $k = 0.17$ and $N = 0.007$ (Equations 5 and 15), the cumulative distributions of the duration of time between metastasis and detection $P_{DetectionInterval}(t)$ and between metastasis and mortality $P_{MortalityInterval}(t)$ were plotted. Figures 2 and 3 illustrate these distributions when all tumor localities were considered together.

The mean and median durations of time between tumor appearance, detection, and mortality were also calculated based on the anatomical locations to which cancer metastasized. Tables 2 and 3 illustrate these mean and median values, where N represents the sample size for each specific subpopulation of patients whose primary tumor metastasized to the designated location.

3.3 Evaluation of Results

In order to compare the model proposed in this study to existing models of tumor growth, the model proposed here (1) was plotted on the same set of axes in Figure 4 as the logistic model (2) proposed by Spratt *et al.* for post-detection tumor growth [28] and an exponential model (3)

Table 2: Mean and median time interval between metastasis and detection for specified tumor localities. Subpopulations marked with a plus sign ($^+$) indicate groups of patients who exhibited a metastasis in the specified location but who may have also exhibited metastases elsewhere.

Subpopulation	Time (Months)
Total Sample ($N = 345$)	Mean: 64.7 Median: 53.1 SD: 47.5
Bone ($N = 45$)	Mean: 66.7 Median: 52.3 SD: 51.4
Bone$^+$ ($N = 114$)	Mean: 68.8 Median: 55.2 SD: 53.2
Brain ($N = 8$)	Mean: 52.0 Median: 48.0 SD: 25.0
Brain and CNS$^+$ ($N = 46$)	Mean: 61.3 Median: 55.3 SD: 36.5
Liver ($N = 13$)	Mean: 67.4 Median: 55.0 SD: 44.9
Liver$^+$ ($N = 51$)	Mean: 67.8 Median: 54.1 SD: 46.5
Lung ($N = 42$)	Mean: 66.9 Median: 59.0 SD: 40.8
Lung$^+$ ($N = 112$)	Mean: 67.1 Median: 57.2 SD: 49.5
Lymph Nodes$^+$ ($N = 20$)	Mean: 70.0 Median: 56.1 SD: 53.8
Marrow$^+$ ($N = 6$)	Mean: 49.7 Median: 35.6 SD: 33.0
Pleura$^+$ ($N = 7$)	Mean: 91.0 Median: 66.2 SD: 75.4
Skin$^+$ ($N = 4$)	Mean: 55.1 Median: 48.0 SD: 40.9

Table 3: Mean and median time interval between metastasis and mortality for specified tumor localities. Subpopulations marked with a plus sign $(^+)$ indicate groups of patients who exhibited a metastasis in the specified location but who may have also exhibited metastases elsewhere.

Subpopulation	Time (Months)
Total Sample $(N = 281)$	Mean: 83.7 Median: 73.6 SD: 50.8
Bone $(N = 35)$	Mean: 94.1 Median: 84 SD: 55.7
Bone$^+$ $(N = 88)$	Mean: 94.0 Median: 81.9 SD: 60.0
Brain $(N = 6)$	Mean: 58.5 Median: 54.9 SD: 27.0
Brain and CNS$^+$ $(N = 38)$	Mean: 87.7 Median: 78.0 SD: 46.7
Liver $(N = 13)$	Mean: 81.0 Median: 67.8 SD: 48.0
Liver$^+$ $(N = 45)$	Mean: 85.4 Median: 77.6 SD: 45.0
Lung $(N = 35)$	Mean: 95.0 Median: 85.0 SD: 52.8
Lung$^+$ $(N = 91)$	Mean: 90.9 Median: 80.8 SD: 57.6
Lymph Nodes$^+$ $(N = 9)$	Mean: 93.4 Median: 79.3 SD: 51.7
Marrow$^+$ $(N = 5)$	Mean: 67.6 Median: 63.3 SD: 20.8
Pleura$^+$ $(N = 2)$	Mean: 77.9 Median: 76.2 SD: 13.3
Skin$^+$ $(N = 3)$	Mean: 68.2 Median: 42.4 SD: 44.5

assuming the doubling time at detectable size. To maintain consistency, all three models were expressed as tumor diameter in millimeters with respect to time after the origin of the tumor in months. The logistic model proposed here and Spratt's earlier logistic model were given by Equations 12 and 13 with $M = 1.1 \times 10^6$ mm^3 and $V_0 = 10^{-6}$ mm^3. For the model proposed here, $N = 0.17$ and $k = 0.213$ months^{-1} (0.007 days^{-1}), whereas for Spratt's model, $N = 0.25$ and $k = 0.098$ months^{-1} (3.22×10^{-3} days^{-1}). The exponential model was given by Equation 2 with y_0 defined as the diameter of a single cell ($2\sqrt[3]{\frac{3\times10^{-6}}{4\pi}}$ mm) and k based on the doubling time of a breast cancer tumor at the time of tumor excision. According to Michaelson *et al.* [17, 20], the doubling time of breast carcinomas with respect to tumor volume at the appoximate time of tumor excision is 130 days. Thus, since tumor volume is proportional to the cube of the diameter, the doubling time with respect to tumor diameter at the time of tumor excision is approximately 12.8 months (390 days). The rate constant k was therefore chosen to be $\frac{\ln 2}{12.8}$ months^{-1}.

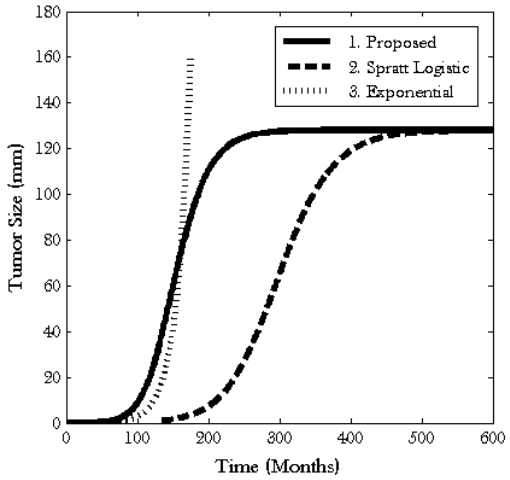

Figure 4: Novel logistic model of tumor growth (1) compared to Spratt's logistic model (2) and exponential model (3).

When all three models were compared graphically (Figure 4), the model proposed in this study (1) adhered relatively closely to the exponential model (3) during the pre-detection period but diverged considerably during the post-detection period, as it began to approach the same theoretical limit of tumor size ($2\sqrt[3]{\frac{3(1.1\times10^6)}{4\pi}}$ mm) as was approached by Spratt's logistic model (2) [28]. In-

terestingly, when the results presented in this study concerning time between metastasis, detection, and mortality were calculated assuming the exponential model illustrated in Figure 4, the results were surprisingly similar, presumably because growth during the pre-detection period precedes significant density-dependent growth deceleration. The logistic model proposed here, however, is still a considerable improvement over the exponential model because it recognizes that the doubling time is not constant and that tumor growth decelerates with respect to time as it approaches a theoretical maximum size.

4 Discussion

4.1 Implications

Although previous studies have proposed models of tumor growth based on data obtained during the post-detection period, this study establishes a new mathematical and computational methodology to model tumor growth more accurately during the pre-detection period. By presenting an improved relation between tumor size and time, the model proposed here may be applied to other mathematical models of cancer to transform such models from functions of tumor size into functions of time. When $D(t)$ is given by the model proposed in this study, other cancer models, such as Michaelson's Size-Only Lethality Equation (Equation 1) [2, 18, 19], which expresses the probability of a lethal metastasis as a function of tumor size, may be expressed as a function of time rather than size (Equation 5). Previously, mathematical models of cancer could be transformed into functions of time either by assuming exponential growth with some constant doubling time or by applying some other model of tumor growth based on data obtained during the post-detection period. The results of this study, however, may improve calculations considerably because the equation proposed here uses mathematical and computational methods to incorporate information about tumor size during the pre-detection period and model tumor growth more accurately.

In addition to proposing a tumor growth model that may be applied to improve other mathematical models of cancer, this study quantified the time interval between metastasis and mortality for

patients with breast cancer metastases in specific locations (Tables 2 and 3). Physicians may utilize this information, along with the improved probability models of lethal metastasis occurrence, to predict cancer patients' survival time more accurately based on primary tumor size and/or the locations to which primary breast tumors metastasized. This information may improve the mathematical precision of current treatment regimens by allowing physicians to base the intensiveness of therapy directly on the probability and timeline of patient survival.

Furthermore, the results concerning time between metastasis and detection indicate that the time between the origin of a breast cancer metastasis and its eventual detection is, on average, well over four years. Because this information suggests that cancer is often detected long after it first appears, the results of this study demonstrate that there is a great necessity to motivate more frequent cancer screening and to develop more effective screening technologies to detect tumors earlier during their progression. Because Michaelson's Size-Only Equation (Equation 1) expresses the probability of metastasis as a monatonically increasing function of tumor size [2, 18, 19], if cancer were detected earlier in its progression, either because of more effective technologies or greater encouragement of individual screening, breast cancer patients would be considerably less prone to experiencing an eventually lethal metastasis before their primary tumors are removed.

4.2 Limitations and Further Study

One limitation of this study was that the deconvolution method described in Subsection 2.4 necessarily overestimated the ranges and standard deviations of the distributions of time intervals between metastasis and detection and between metastasis and mortality. Due to the considerable individual variation that exists in tumor growth rates, the standard deviations of time between metastasis, detection, and mortality would in fact be relatively large. However, the reported standard deviation values were amplified in magnitude by the method of deconvolution presented here because all possible combinations of time values in the distributions of metastasis and detection or mortality were compared. This method overestimates the ranges and standard deviations of the distributions because, in reality, a secondary tumor detected relatively late in the detection distri-

bution would most likely not be derived from a primary tumor that metastasized relatively early in the metastasis distribution. Similarly, a secondary tumor detected relatively early would most likely not be derived from a primary tumor that metastasized relatively late. However, although the ranges and standard deviations were overestimated, the measures of central tendency, especially medians, were likely quite reasonable due to the balance of outliers on both ends of the distributions.

Another limitation was that the sample sizes were quite small for several of the subpopulations of patients with metastases in the specified anatomical locations. In these instances, the findings presented here are not necessarily representative of the general population of breast cancer patients exhibiting metastases in the designated locations, and additional data are necessary to obtain more reliable results. One final limitation of this study was that, in the absence of additional data, the time of tumor biopsy was also taken as the time of tumor excision. In reality, cancer patients do not always undergo surgery immediately after their tumors are detected; however, since the mean time between detection and excision is 3.4 months [21], this assumption likely had a negligible impact on the results, given the large time scale intrinsic to the calculations of this study.

A possible expansion of this study would be to fit continuous differentiable functions to the empirically determined cumulative distributions of detection and mortality times so that the probability distributions would be continuous rather than discretely defined. It may also be desirable to determine the parameters N and k in the logistic model $D(t)$ by minimizing the average difference between all percentiles in the distributions of times between origin and detection for the primary tumor and metastasis, rather than by minimizing only the difference between the medians of the two distributions. Finally, it may be useful to reconcile the post-detection models of tumor growth proposed by previous literature with the pre-detection logistic model proposed in this work, through either linear combination of the models or further addition of parameters. Such a composite model of tumor growth would be crucial for physicians to predict patients' probability of lethality more accurately, enabling them to adjust treatment regimens accordingly.

5 Conclusions

By establishing a mathematical and compuational method to incorporate information about tumor size during the pre-detection period, this study proposes a novel tumor growth model (Equation 15) and provides a more accurate quantitative understanding of breast cancer and metastastic disease. By applying the proposed model to existing mathematical models of cancer, more accurate calculations may be made. In particular, the model proposed here may improve predictions of the probability with respect to time that breast cancer patients will experience one or more lethal metastases. These predictions may dramatically improve cancer treatment by allowing physicians to base intensiveness of therapy directly on patients' likelihood of experiencing a lethal metastasis.

The results of this study also indicate that the median time between metastasis and detection is 53.1 months and that the median time between metastasis and mortality is 73.6 months. The considerable time interval between tumor onset and detection suggests a tremendous need to motivate more frequent cancer screening and to develop more sensitive technologies to detect tumors of smaller diameters.

With additional data, the methodology proposed by this study may be used to compare the time intervals between metastasis, detection, and mortality more reliably based on the anatomical locations of breast cancer metastasis. This information may allow physicians to improve treatment considerably by more precisely predicting the amount of time that breast cancer patients will survive based on the locations to which their primary tumors metastasized. This development would enable physicians to create more mathematically precise treatment regimens by establishing a potential hierarchy of treatment priority among different sites of breast cancer metastasis.

Acknowledgements

I would like to thank Dr. James Michaelson for welcoming me to work in his laboratory and for offering his thoughtful input throughout the course of this study. I thank Ann Ouyang for advising and reviewing the writing of my paper and Dr. Andrew Charman for providing statistical guidance.

I would also like to thank the staff of the Laboratory for Quantitative Medicine and the Research Science Institute (RSI) for their support and feedback regarding my paper and project. Finally, I thank the Center for Excellence in Education (CEE) and the Massachusetts Institute of Technology (MIT) for hosting my research over the summer and making this study possible.

References

[1] I. D. Bassukas. Comparative Gomparative analysis of alterations of tumor growth patterns. Cancer Research 54 (1994), 4385–4392.

[2] L. L. Chen, M. E. Nolan, M. J. Silverstein, M. C. Mihm, A. J. Sober, K. K. Tanabe, B. L. Smith, J. Younger, J. S. Michaelson, and G. Weber. The impact of primary tumor size, lymph node status, and other prognostic factors on the risk of cancer death. Cancer 115 (2009), 5071–5083.

[3] P. W. Durbin, N. Jeung, M. H. Williams, and J. S. Arnold. Construction of a growth curve for mammary tumors of the rat. Cancer Research 27 (1967), no. 1, 1341–1347.

[4] D. v. Fournier, E. Weber, W. Hoeffken, M. Bauer, F. Kubli, and V. Barth. Growth rate of 147 mammary carcinomas. Cancer 45 (1980), 2198–2207.

[5] J. H. Gaddum. Lognormal distributions. Nature 156 (1945), no. 3964, 463–466.

[6] E. Galante, G. Gallus, A. Guzzon, A. Bono, G. Bandieramonte, and S. Di Pietro. Growth rate of primary breast cancer and prognosis: Observations on a 3- to 7- year follow-up in 180 breast cancers. British Journal of Cancer 54 (1986), 833–836.

[7] M. Gyllenberg and G. F. Webb. Quiescence as an explanation of Gompertzian tumor growth. Growth, Development, and Aging 53 (1989), 25–33.

[8] L. Heuser, J. S. Spratt, and H. C. Polk. Growth rates of primary breast cancers. Cancer 43 (1979), 1888–1894.

[9] T. Kuroishi, S. Tominaga, T. Morimoto, H. Tashiro, S. Itoh, H. Watanable, M. Fukuda, J. Ota, T. Horino, T. Ishida, T. Yokoe, K. Enomoto, Y. Kashiki, and M. Ogita. Tumor growth rate and prognosis of breast cancer mainly detected by mass screening. Japanese Journal of Cancer Research 81 (1990), 454–462.

[10] A. K. Laird. Dynamics of tumor growth. British Journal of Cancer 18 (1964), no. 3, 490–502.

[11] A. K. Laird. Dynamics of tumour growth: Comparison of growth rates and extrapolation of growth curve to one cell. British Journal of Cancer 19 (1965), no. 2, 278–291.

[12] A. K. Laird. Dynamics of growth in tumors and in normal organisms. National Cancer Institute Monograph 30 (1969), 15–28.

[13] Y. N. Lee. The lognormal distribution of growth rates of soft tissue metastases of breast cancer. Journal of Surgical Oncology 4 (1972), no. 2, 81–88.

[14] Y. N. Lee and J. S. Spratt, Jr. Rate of growth of soft tissue metastases of breast cancer. Cancer 29 (1972), no. 2, 344–348.

[15] J. A. McCredie, W. R. Inch, J. Kruuv, and T. A. Watson. The rate of tumor growth in animals. Growth 29 (1965), 331–347.

[16] J. Michaelson. The role of molecular discreteness in normal and cancerous growth. Anticancer Research 19 (1999), 4853–4868.

[17] J. Michaelson, S. Satija, R. Moore, G. Weber, E. Halpern, A. Garland, D. B. Kopans, and K. Hughes. Estimates of the sizes at which breast cancers become detectable on mammographic and clinical grounds. Journal of Women's Imaging 5 (2003), no. 1, 3–10.

[18] J. S. Michaelson, L. L. Chen, M. J. Silverstein, J. A. Cheongsiatmoy, M. C. Mihm, Jr., A. J. Sober, K. K. Tanabe, B. L. Smith, J. Younger, G. Weber, and D. Livesto. Why cancer at the primary site and in the lymph nodes contributes to the risk of cancer death. Cancer 115 (2009), 5084–5094.

[19] J. S. Michaelson, L. L. Chen, M. J. Silverstein, M. C. Mihm, A. J. Sober, K. K. Tanabe, B. L. Smith, and J. Younger. How cancer at the primary site and in the lymph nodes contributes to the risk of cancer death. Cancer 115 (2009), 5095–5107.

[20] J. S. Michaelson, M. Silverstein, J. Wyatt, G. Weber, R. Moore, E. Halpern, D. B. Kopans, K. Hughes. Predicting the survival of patients with breast carcinoma using tumor size. Cancer 95 (2002), no. 4, 713–723.

[21] J. Michaelson, G. Weber, J. Wang, J. Younger, and A. A. Gawande. The impact of time on breast carcinoma survival. In preparation.

[22] L. Norton. A Gompertzian model of human breast cancer growth. Cancer Research 48 (1988), 7067–7071.

[23] L. Norton and R. Simon. Tumor size, sensitivity to therapy, and design of treatment schedules. Cancer Treatment Reports 61 (1977), no. 7, 13077–1317.

[24] L. Norton, R. Simon, H. D. Brereton, and A. E. Bogden. Predicting the course of Gompertzian growth. Nature 264 (1976), 542–545.

[25] D. W. Ollila, S. L. Stern, and D. L. Morton. Tumor doubling time: A selection factor for pulmonary resection of metastatic melanoma. Journal of Surgical Oncology 69 (1998), 206–211.

[26] P. G. M. Peer, J. A. A. M. van Dijck, J. H. C. L Hendriks, R. Holland, and A. L. M. Verbeek. Age-dependent growth rate of primary breast cancer. Cancer 71 (1993), no. 11, 3547–3551.

[27] J. F. Speer, V. E. Petrosky, M. W. Retsky, and R. H. Wardwell. A stochastic numerical model of breast cancer growth that simulates clinical data. Cancer Research 44 (1984), 4124–4130.

[28] J. A. Spratt, D. von Fournier, J. S. Spratt, and E. E. Weber. Decelerating growth and human breast cancer. Cancer 71 (1993), no. 6, 2013–2019.

[29] J. A. Spratt, D. von Fournier, J. S. Spratt, and E. E. Weber. Mammographic assessment of human breast cancer growth and duration. Cancer 71 (1993), no. 6, 2020–2026.

[30] J. S. Spratt. The lognormal frequency distribution and human cancer. Journal of Surgical Research 9 (1969), no. 3, 151–157.

[31] J. S. Spratt, R. A. Greenberg, and L. S. Heuser. Geometry, growth rates, and duration of cancer and carcinoma in situ of the breast before detection by screening. Cancer Research 46 (1986), 970–974.

[32] J. S. Spratt, J. S. Meyer, and J. A. Spratt. Rates of growth of human neoplasms: Part I. Journal of Surgical Oncology 60 (1995), 137–146.

[33] J. S. Spratt, J. S. Meyer, and J. A. Spratt. Rates of growth of human neoplasms: Part II. Journal of Surgical Oncology 61 (1996), 68–83.

[34] J. S. Spratt, Jr., M. L. Kaltenbach, and J. A. Spratt. Cytokinetic definition of acute and chronic breast cancer. Cancer Research 37 (1977), 226–230.

[35] J. S. Spratt, Jr., and T. L. Spratt. Rates of growth of pulmonary metastases and host survival. Annals of Surgery 159 (1964), no. 2, 161–171.

[36] P. W. Sullivan and S. E. Salmon. Kinetics of tumor growth and regression in IgG multiple myeloma. The Journal of Clinical Investigation 51 (1972), 1697–1708.

[37] G. W. Swan and T. L. Vincent. Optimal control analysis in the chemotherapy of IgG multiple myeloma. Bulletin of Mathematical Biology 39 (1977), 317–337.

[38] H. Weedon-Fekjr, B. H. Lindqvist, L. J. Vatten, O. O. Aalen, and S. Tretli. Breast cancer tumor growth estimated through mammography screening data. Breast Cancer Research 10 (2008), no. 3, R41.

[39] C. P. Winsor. The Gompertz curve as a growth curve. Proceedings of the National Academy of Sciences 18 (1932), no. 1, 1–8.

[40] S. L. Zeger and S. D. Harlow. Mathematical models from laws of growth to tools for biologic analysis: Fifty years of *Growth*. Growth 51 (1987), 1–21.