The Laboratory of Quantitative Medicine
Technical Report #8
March 9, 2009

HOW AND WHY PRIMARY TUMOR SIZE, NODAL STATUS, AND OTHER PROGNOSTIC FACTORS CONTRIBUTE TO THE RISK OF CANCER DEATH

James S. Michaelson PhD[1,2,5], L. Leon Chen BS[2],

Note: this technical reports accompanies three research paper, but with the full complement of tables and other information that would not be accommodated within the space limitations of the journal

Departments of Pathology[1], Surgery[2], Massachusetts General Hospital, Boston, Massachusetts, USA
Departments of Pathology[5], Surgery [6],, Harvard Medical School, Boston, Massachusetts, USA

Correspondence to James S. Michaelson Ph.D., Division of Surgical Oncology, Cox Building Room 626, Massachusetts General Hospital, 100 Blossom Street, Boston, Massachusetts, 02114
TEL 617 501 0590 FAX 617 724 3895
Email: michaelj@helix.mgh.harvard.edu

ABSTRACT

BACKGROUND: It has long been appreciated that tumor size, nodal status, prognostic factors, and patient survival are related qualities, although it has not been obvious how to isolate the interactions between each of these components of the cancer phenotype. Neither has it been obvious how to integrate information on tumor size, nodal status, and other prognostic factors into predictions of the risk of death for individual patients. The underlying mechanism for the contribution of tumor size, nodal status, and prognostic factor phenotype to cancer lethality has also been obscure. **METHODS:** We describe a new mathematical method, the *binary-biological model of cancer metastasis*, based upon the spread of cancer cells, whose equations can capture the relationships between tumor size, nodal status, and cancer lethality, as well as calculate the probability of the spread of cancer cells from data on the macroscopic manifestations of such spread. **RESULTS:** For melanoma, renal cell carcinoma, and breast carcinoma, the relationship between tumor size and the risk of cancer death is well captured by the *SizeOnly* Equation. For melanoma and breast carcinoma, the relationship between tumor size and the presence of cancer in the nodes is well captured by the *NodalSizeOnly* Equation. For node negative melanoma and breast carcinoma, the relationship between tumor size and risk of death is well captured by the *PrimarySizeOnly* Equation. For breast carcinoma, each positive node has been found to contribute ~6% extra risk of death while each millimeter of primary tumor diameter contributes ~1% risk of death. For melanoma, each positive node has been found to contribute ~23% risk of death while each millimeter of primary melanoma thickness contributes ~8% risk of death. This information is captured by a pair of linked equations, the *Size+Nodes* method. A simple expression, the *ProbabilityEstimation* Equation, could calculate the probabilities of spread of cancer cells from data on tumor size, nodal status, and death rate, revealing that the lethal contributions of cancer at the primary site and nodes can be explained by a simple mechanical process of the spread of cancer cells, occurring with definable probabilities per cell. The presence of cancer in the nodes does not indicate an intrinsic change in a malignancy, but an increased mass of cancer from which spread can emerge. The greater the number of cells at the primary site (tumor size) or the greater the number of cells in the nodes (number of positive nodes) the greater is the aggregate chance that one or more cells has undergone a lethal event of spread, a process captured by the *Size+Nodes* equations. When similar masses of cancer are compared, the chance of lethal spread of a cancer cell to the periphery is about the same whether emerging from a node or from the primary site. The *SizeOnly* equation provided a way to indentify which of a large number of prognostic factors truly make an independent contribution to cancer lethally (the *SizeAssessment* method), while the magnitude of each factor's lethal contribution was quantified by a parameter, *g*, inserted into the *SizeOnly* equation (*PrognosticMeasurement* method). From such values, and with a series of linked equations (the *Size+Nodes+PrognosticFactors* [*SNAP*] method) information on tumor size, nodal status, and other prognostic factors could be combined for each patient to make an accurate estimate of the risk of cancer death. The *SNAP* method was found to accurately estimate the risk of death, and finely stratify patients by risk, for patients with both melanoma and breast carcinoma. **CONCLUSIONS:** The *binary-biological model of cancer metastasis* provides an integrated tool set of equations for answering a range of practical questions about cancer lethality and other manifestations of the spread of cancer cells, including the accurate prediction of the risk of cancer death for each patient.

INTRODUCTION

It has long been appreciated that primary tumor size, nodal status, and survival are related qualities for many cancers, although it has not been obvious as to how to isolate the interactions between each component, nor how to integrate information on tumor size and nodal status into predictions of the risk of death[1,2]. We have found for breast carcinoma that a simple expression, the *SizeOnly* Equation, accurately captures the relationship between primary tumor diameter and lethality[3,4,5,6]. By double-sorting breast carcinoma patients by tumor size and nodal status, we have also found that each positive node is associated with ~6% risk of death, while the lethal contribution from cancer at the primary site can be estimated from tumor diameter with a variant of the *SizeOnly* Equation such that each millimeter of primary tumor diameter is associated with ~1% risk of death. It follows that the overall risk of death is the sum of the risks of death from the nodes and the primary site, as captured by a pair of linked equations, the *Size+Nodes* method[6]. In the absence of information on nodal status, the *Size+Nodes* method reduces to the *SizeOnly* Equation. While the *SizeOnly* and *Size+Nodes* methods have been found to accurately capture the relationship between tumor size, nodal status, and the risk of death for breast carcinoma, their applicability to other cancers has not been examined. Here we outline a more general and efficient mathematical technique (the *Nodal Lethality* and *PrimarySizeOnly* Equations) for isolating the impact of tumor size and nodal status on the risk of cancer death. We examine the applicability of the *SizeOnly* method for relating tumor size to the risk of death for breast carcinoma, melanoma, and renal cell carcinoma, and the applicability of the *Size+Nodes* method for relating tumor size and nodal status to the risk of death for breast carcinoma and melanoma. We also examine the relationship between primary tumor size and the risk of cancer in the nodes for breast carcinoma and melanoma.

Many prognostic factors have been found to be associated with differences in cancer lethality, including tumor size[3,7] , nodal status[6,7], grade[8], histology[9,10], sex, patient age[11], gene expression array pattern[12], immunological marker phenotype, biochemical marker phenotype, and specific mutations.[13] However, it may not be obvious whether the presence of a factor truly makes an independent impact on cancer lethality, or whether the presence of the factor is simply a quality that is correlated with tumor size. Here we describe a new technique, the *SizeAssessment* method, which uses the the *SizeOnly* equation to determine whether a prognostic factor makes an independent contribution to the risk of cancer death, or is merely correlated with tumor size. We then describe a technique, the *PrognosticMeasurement* method, which provides a quantitative measure of each prognostic factor's contribution to cancer lethality, through the introduction of a parameter, *g*, inserted into the *SizeOnly* equation. Finally, we outline a technique, the *Size+Nodes+PrognosticFactors* (*SNAP*) method, which uses the information derived by the *SizeAssessment* and *PrognosticMeasurement* methods to integrate information on tumor size, nodal status, and other prognostic factors into a prediction of the risk of death for each patient. We apply these methods to the analysis of breast cancer and melanoma survival.

The *SizeOnly*, *Size+Nodes*, and *SNAP* methods not only provide highly practical and accurate tools for estimating breast carcinoma outcome, they are also biologically plausible, as their equations have been derived by a consideration of the most generally accepted mechanism of cancer death, which is by the spread of cancer cells occurring with definable probabilities per cell[5,14]. We call this approach the *binary-biological model of cancer metastasis*.

Underlying the macroscopic features of cancer growth, spread, and lethality (the size of a cancer at the primary site and in the nodes, the number of positive nodes, and the risk of death), lie microscopic events affecting the fate of cells, such as the spread of cancer cells from one location to another. Such events of cellular spread are intrinsically discrete, either/or events, because cells are intrinsically discrete entities.[14] Either a cancer cell in a primary tumor will travel to the periphery and give rise to metastatic disease and death, or it won't. Either a cancer cell in a primary tumor will travel to a local lymph node and give rise to a cancer mass seen by the pathologist, or it won't. Either a cancer cell in a lymph node will travel to the periphery and lead to death, or it won't. This either/or quality makes it possible to characterize spread of cancer cells in terms of probabilities. By taking such an approach, which we call the *binary-biological model of cancer metastasis*, it is possible to infer the values of the probabilities of these microscopic events of the spread from data on the macroscopic features of cancer. For example, we shall be able to ask how the probability of the lethal spread of a cancer cell from a primary site compares with the probability of the lethal spread of a cancer cell from a lymph node. As we shall see, applying this

framework to actual data will allow us to determine how cancer at the primary site and in the nodes contributes to lethality, as well as how the presence of various prognostic factors affects the risk of death by altering the underlying chances that a cancer cell will spread.

METHODS

Mathematical Methods

The general theory behind the mathematical methods used here, *the binary-biological model of cancer metastasis*, is described in Technical Report #1, which can be found at: http://www.lifemath.net/cancer/about/techreports/index.php . Only those aspects relevant to data outlined in this Technical Report are repeated in the text below.

TABLE O

Methods of the *Binary-Biological Model of Cancer Metastasis*

Method	Purpose
ProbabilityEstimation Equation (#7)	Estimates the probability of the spread of cancer cells, per cell, from data on the size of the mass from which the spread occurs and the lethal or non-lethal consequences of the spread.
SizeOnly Equation (#1)	Relates tumor size to the chance of cancer death*
TimeOnly Equation (#14)	Relates time to the chance of cancer death
PrimarySizeOnly Equation (#1c)	Relates tumor size to the chance of cancer death for node negative patients
NodalSizeOnly Equation (#1n)	Relates tumor size to the chance of cancer in the lymph nodes
Size+Nodes method	Integrates information on tumor and number of positive nodes into an estimate of the chance of cancer death*
Nodal Lethality Equation (#17)	Calculates the lethal contribution, per positive node, of cancer in the lymph nodes
SizeAssessment method	Determines whether a prognostic factor makes an independent contribution to the risk of cancer death, or is merely correlated with tumor size
PrognosticMeasurement method	Provides a quantitative measure of each prognostic factor's contribution to cancer lethality, through the introduction of a parameter, g, inserted into the *SizeOnly* equation
SNAP (*Size+Nodes+Prognostic Fatcor*) method	Integrates information on primary tumor size, nodal status, and other prognostic factors into an estimate of the chance of cancer death*

* When only tumor size and nodal status are known, the *SNAP* method reduces to the *Size+Nodes* method for estimating the risk of cancer death from information on tumor size and number of positive nodes, while when only size is known, the *Size+Nodes* and *SNAP* methods reduce to the *SizeOnly* method for estimating the risk of cancer death from information on tumor size.

TABLE 00

The *SNAP* (*Size+Nodes+PrognosticMarkers*) Method for Estimating the Risk of Cancer Death from Information on Tumor Size, Nodal Status, and Other Prognostic Factors

$$L = L_{primary} + L_{nodes} - (L_{primary} * L_{nodes})$$

(eq. (4))

Source of Lethality	***Method of Estimation***	***Independent Variable***	***Parameters***	***Interpretation***
The lethal contribution from cancer at the primary site	$L_{primary} = 1 - e^{-(Q^* j_{primary})^*(g_1 * g_2 * g_3 * g_4 * ..) D^Z}$ (1c)	D = Tumor Size: *For Breast Carcinoma:* Diameter (mm) *For Melanoma:* Thickness (mm)	*For Breast Carcinoma:* Q = 0.0118395 Z = 1 $j_{primary}$=0.661 if nodal status is known $j_{primary}$ =1 if nodal status is unknown See Table C2 for g parameter values *For Melanoma:* Q= 0.1428 Z= 0.89 $j_{primary}$ =0.801 if nodal status is known $j_{primary}$ =1 if nodal status is unknown See Table C5 for g parameter values	*The lethal contribution of the primary mass increases gradually with tumor size, and the amount of that lethal contribution is influenced by prognostic factors, as captured by the g parameters in Equation 1d*
The lethal contribution from cancer in the lymph nodes	$L_{nodes} = 1 - e^{-(M * L_{per-node})}$ eq. (2)	M = The Number of Positive Nodes	*For Breast carcinoma:* $L_{per\text{-}node}$ = 0.0608 *For Melanoma:* $L_{per\text{-}node}$ = 0.2253	*The presence of each positive lymph node contributes approximately "$L_{per\text{-}node}$" extra chance of death*

The *SNAP* (*Size+Nodes+PrognosticMarkers*) method reduces to:

- **the *Size+Nodes* method, when only size and nodal status are known.**
- **the *SizeOnly* method, when only size is known.**

Patients

Information on tumor thickness and survival was obtained for 2770 melanoma patients seen at MGH from 1970/01/01 to 2002/05/01; nodal status information was known for 664.

Information on tumor size, nodal status, and breast carcinoma survival was derived from 1352 patients with invasive breast carcinoma diagnosed between 1966 and 1990 at the USC/Van Nuys Breast Center, with survival estimated as of Dec 31 2000[3].

Information on renal cell carcinoma survival was taken from Delahunt et al.[15] and Hafez et al.[16]

Karrison and colleagues[17] had found that little lethality occurs 15 years after diagnosis, and we have found a similar hazard function for melanoma. Thus, we have relied upon the 15-year cancer-specific Kaplan-Meier death rate as our measurements of the cancer death rate (L).

For the analysis of the lethal impact of prognosis factors other than tumor size an nodal status, data were also available on much larger populations of patients seen at the Partners Hospitals (11,271 patients seen at the Massachusetts General Hospital and Brigham and Women's Hospital between 1960 and 2003) and from the SEER national dataset (362,491 patients between 1973 and 2004 with a first malignant tumor of 1-50 mm in diameter and 0-7 positive lymph nodes).[18] Information of the use of these larger datasets to refine the parameters of the equations of *the binary-biological model of cancer metastasis* can be found at http://cancer.lifemath.net/about/techreports/index.php.

Measurement of nodal deposits

The sizes of cancer deposits in the lymph nodes were measured with a Zeiss microscope and Insight digital camera with SPOT-cam software for measuring cross-sectional areas and the widest diameter (Diagnostic Instruments, Inc, Sterling Heights, Michigan). The relationship between melanoma thickness and cross-sectional area was estimated with data by Temple et al[19] ($thickness$=0.6073*$area^{0.5086}$ (R^2=0.65)).

RESULTS

The relationship between tumor size and cancer death is well captured by the *SizeOnly* Equation.

The fit of tumor size/lethality data to the *SizeOnly* Equation (#1):

$$L = 1 - e^{-QD^Z} \quad (1)$$

can be tested by regression by transforming it into Equation #13. As seen in Figure A1, such a regression reveals that for melanoma, Z=0.89 and Q=0.134 (R^2=0.86), while for renal cell carcinoma, Z=1.14 and Q=0.0033 (R^2=0.90), and for breast carcinoma, Z=1.33 and Q=0.0061 (R^2=0.97). Refining the value of Q by a *pseudo-Monte Carlo* method (Equation #1b) yields the slightly more accurate value of Q=0.1428 for melanoma and Q=0.0062 for breast carcinoma (Table A1). The close fit of these data to the *SizeOnly* Equation (#1) can be seen in Figure A2.

The fit of these data to the *SizeOnly* Equation (#1) is also close when Z=1 (Figure A1, right side). This simplifies the *SizeOnly* Equation (#1), and makes it possible to carry out a comparison of the lethalities of different cancers and different subtypes of the same cancer by a comparison of the values of their Q parameters, as we shall see below.

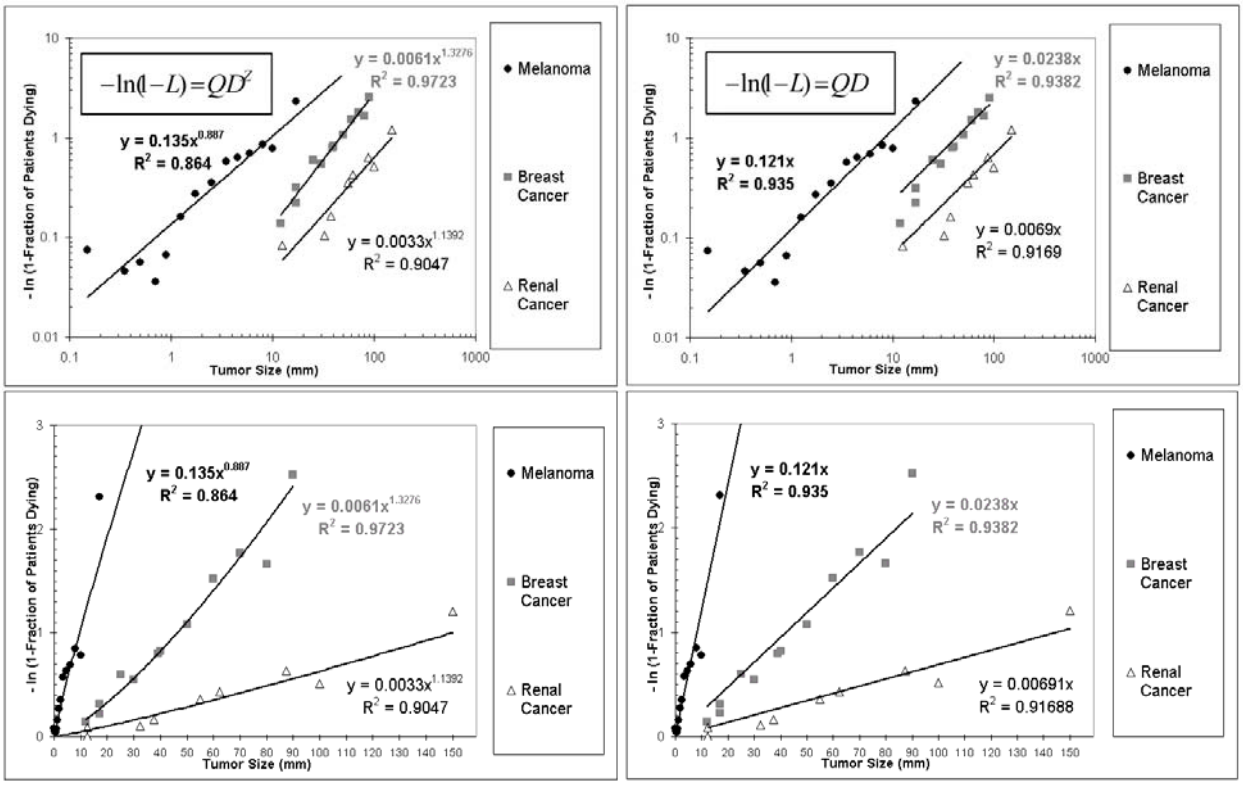

FIGURE A1

FIGURE A1. **Fit of melanoma, breast, and renal cell carcinoma size/lethality data to the *SizeOnly* Equation (#1), as assessed by Equation #13.** On the right, the fit of these data to the *SizeOnly* Equation (#1) is examined with Z=1. Conversion of tumor sizes into values of N (cell number) carried out by Equation #8.

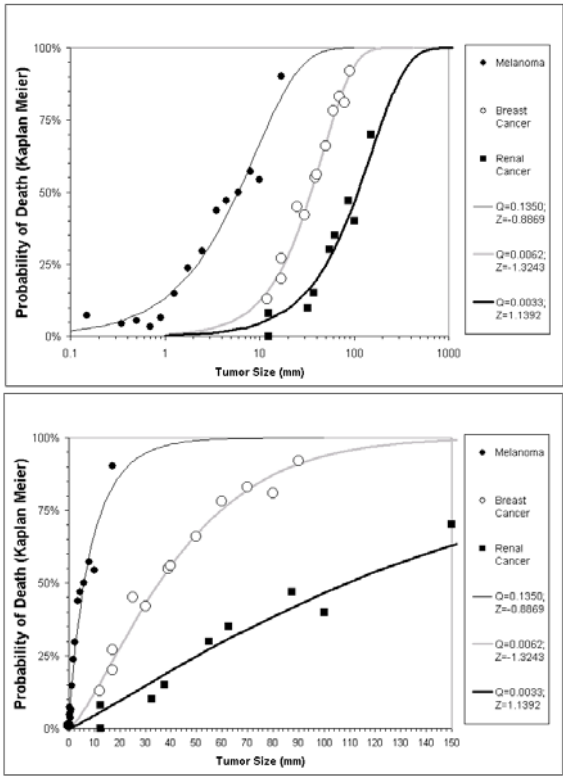

FIGURE * **Fit of Size/Lethality data for melanoma, breast, and renal cell carcinoma to the *SizeOnly* Equation (#1).** TOP: Size on log scale. BOTTOM: Size on conventional scale.
C:\X_MelanomaAnalysisByExcell 1 1 04___Survival Analysis of Basic Dataset by Size\groupsOf300\Renal\[FIGS.xls]Sheet1!$C$1

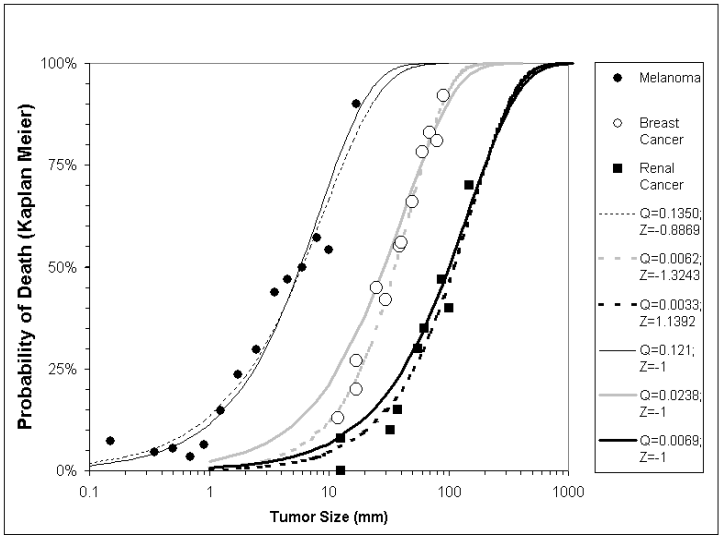

FIGURE *. Fit of Size/Lethality data for melanoma, breast, and renal cell carcinoma to the *SizeOnly* Equation (#1) with and without *Z*=1.

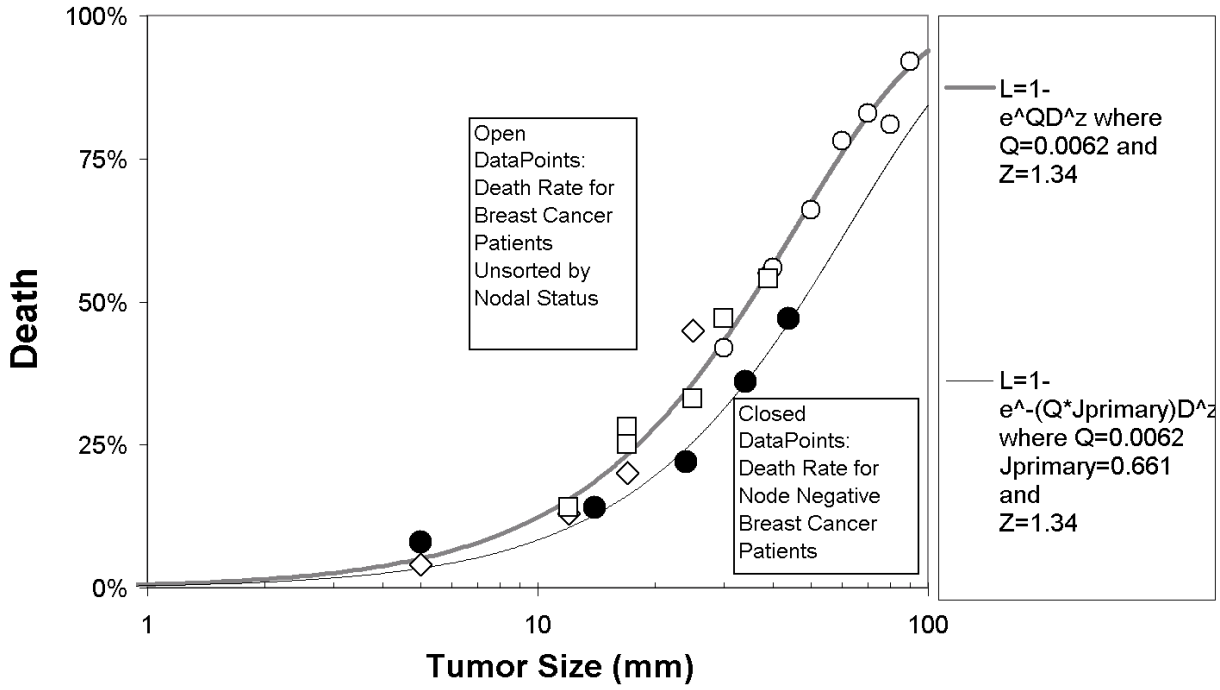

FIGURE A2(a)

FIGURE A2. (a) Relationship between tumor size and cancer death for breast carcinoma patients as a whole (the *NodalSizeOnly* Equation#1) and for node-negative breast carcinoma patients (the *PrimarySizeOnly* Equation #1c). (b) Relationship between tumor size and cancer death for melanoma patients as a whole (the *NodalSizeOnly* Equation #1) and for node-negative melanoma patients (the *PrimarySizeOnly* Equation #1c).

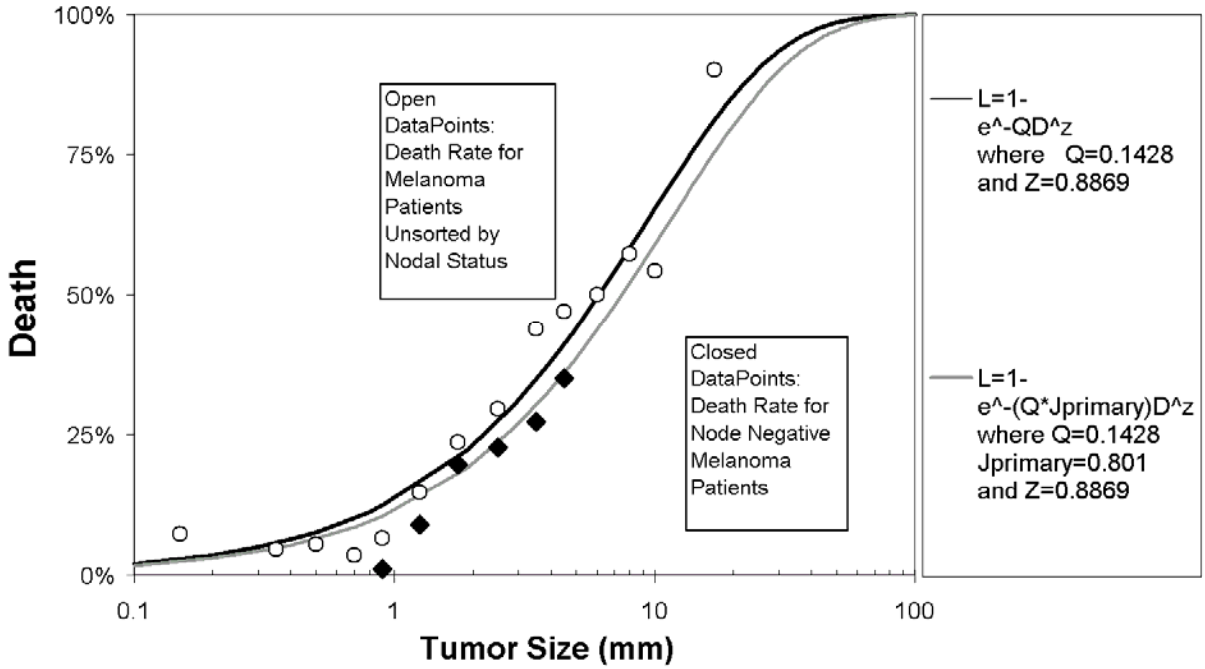

FIGURE A2(b)

The relationship between tumor size and the fraction of patients that are node positive is well captured by the *NodalSizeOnly* Equation.

The fit of tumor size/node positivity data to the *NodalSizeOnly* Equation (#1n):

$$L_{To-Nodes} = 1 - e^{-Q_n D^Z} \quad (1n)$$

can be tested by regression by transforming it into Equation #13. Such a regression reveals that for melanoma, Q_n=0.1018 and Z=1.207 (R^2=0.885), while for breast carcinoma, Q_n=0.019 and Z=1.0041 (R^2=0.90). The capacity of the *NodalSizeOnly* Equation (#1n) using these parameters to capture the relationship between tumor size and the fraction of patients that are node positive can be seen in Figure A3.

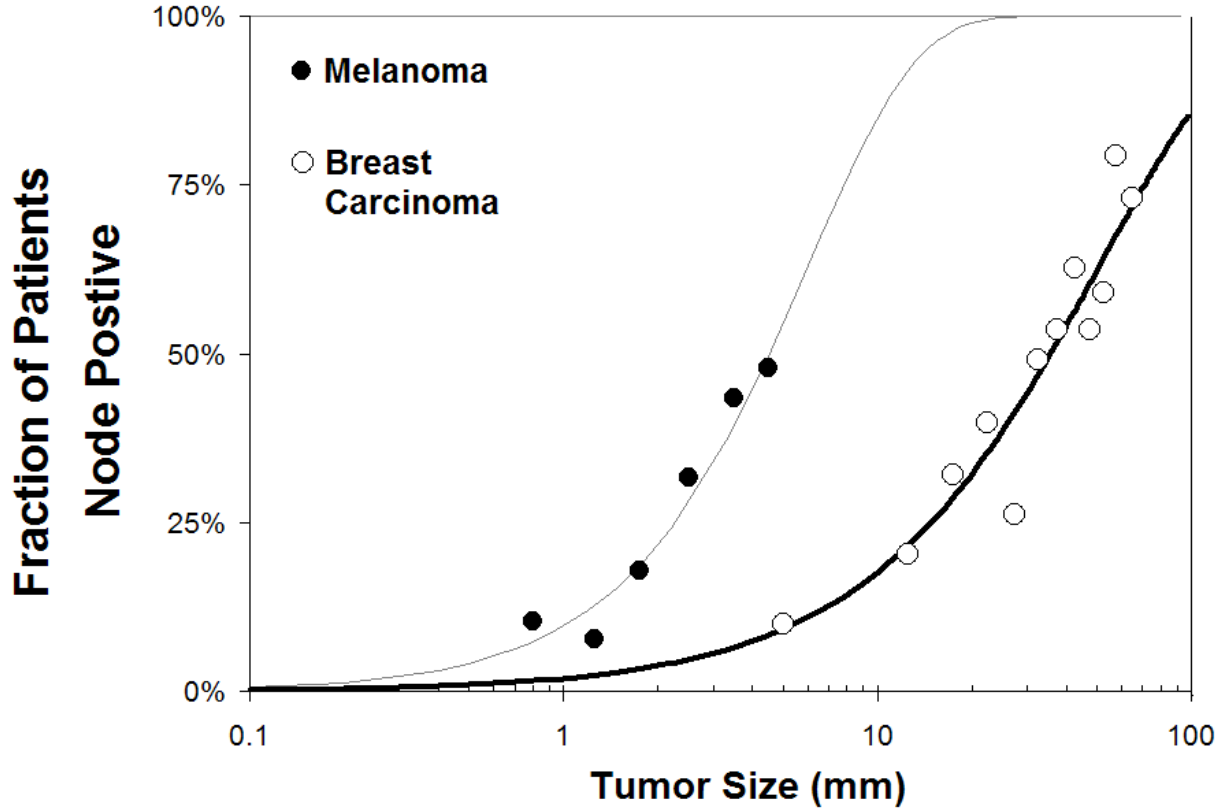

FIGURE A3

FIGURE A3. Relationship between primary tumor size and risk of node positivity, as captured by the *NodalSizeOnly* Equation (#1n). Values of Q_n and Z were determined by regression to Equation #13 (for breast carcinoma, Q_n=0.019, Z= 1.0041, R^2= 0.902; for melanoma, Q_n=0.1018, Z=1.207, R^2= 0.8851)

The relationship between tumor size and cancer death for node negative patients is well captured by the *PrimarySizeOnly* Equation.

We have previously found for breast carcinoma that the relationship between tumor diameter (D) and the risk of cancer death for node-negative patients ($L_{primary}$) is well fit to the *PrimarySizeOnly* Equation (#1c):

$$L_{primary} = 1 - e^{-(Q^* j_{primary}) D^Z} \quad (1c)$$

for which $j_{primary}$ = 0.661, as determined by a *pseudo-Monte Carlo* method (Equation #1c, Figure A2a). As can be seen in Figure A2b, this is also the case for melanoma, such that $j_{primary}$= 0.801.

Each positive node is associated with an extra risk of lethality: ~23% per positive node for melanoma, ~6% per positive node for breast carcinoma

The impact of nodal status on breast carcinoma and melanoma lethality cannot be seen directly from the survival of patients with various numbers of positive nodes because tumor size and nodal status are conflated: as the tumor size increases, both the fraction of patients with positive nodes and the average number of positive nodes increase, while as the number of nodes increases so does the tumor size (Table A3). Equations # 1c and #17 allow us to disentangle the impact of nodal status and tumor size, and to calculate the value of the extra lethality, $L_{per\text{-}node}$, associated with each positive lymph node for patients with various numbers of positive nodes, M.

We can see the general approach by considering the 92 melanoma patients with only 1 positive node (Table A3). This group of patients had a 15-year Kaplan-Meier melanoma death rate of 40%. Equation #1c (see above) tells us that another 92 patients with tumors of the same thicknesses, but who had been node negative, would have had a 25% death rate (Table A3). Thus for patients with one positive node, the lethal contribution ascribable to cancer in that node, $L_{per\text{-}node}$, is approximately 20% ([40%-25%]/[100%-25%]).

Estimating the value of $L_{per\text{-}node}$ for patients with 2 or more positive nodes requires a more complicated calculation, which can be accomplished with the *Nodal Lethality* Equation (#17):

$$L_{per\text{-}node} = (ln\ (1 - ((\ L_W - (1 - e^{-(Q^* jprimary) D\hat{}z})) / ((-e^{-(Q^* jprimary) D\hat{}z}))))) / M \qquad (17)$$

For example, for the 36 melanoma patients with 2 positive nodes, the lethal contribution per node calculated with Equation #17 is 24% ($L_{per\text{-}node} \approx 0.24$), while for the patients with 3 positive nodes, $L_{per\text{-}node} \approx 24\%$, for patients with 4 or 5 positive nodes, $L_{per\text{-}node} \approx 23\%$, and for patients with 6 or 7 positive nodes $L_{per\text{-}node} \approx 24\%$ (Table A3). Thus, we can conclude that no matter how many positive nodes are found in a melanoma patient, each positive node is associated with about a 20-25% extra chance of death.

Similarly, when we examine groups of breast carcinoma patients with various numbers of positive nodes, the *Nodal Lethality* Equation (#17) tells us that each positive node is associated with about an extra 6% chance of death ($L_{per\text{-}node} \approx 0.06$, Table A3). This finding agrees with our previously reported observations[6], made by sorting patients by both tumor size and number of positive nodes, that each positive node is associated with about a 6% extra chance of death.

The values for $L_{per\text{-}node}$ shown in Table A3 reflect subgroups of patients with various numbers of positive nodes. However, we are able to make yet more accurate estimates of the value of $L_{per\text{-}node}$ using all node positive patients by applying a *pseudo-Monte Carlo* method to the *Nodal Lethality* Equation (#17), revealing that $L_{per\text{-}node} = 0.0608$ for breast carcinoma and $L_{per\text{-}node} = 0.22527$ for melanoma.

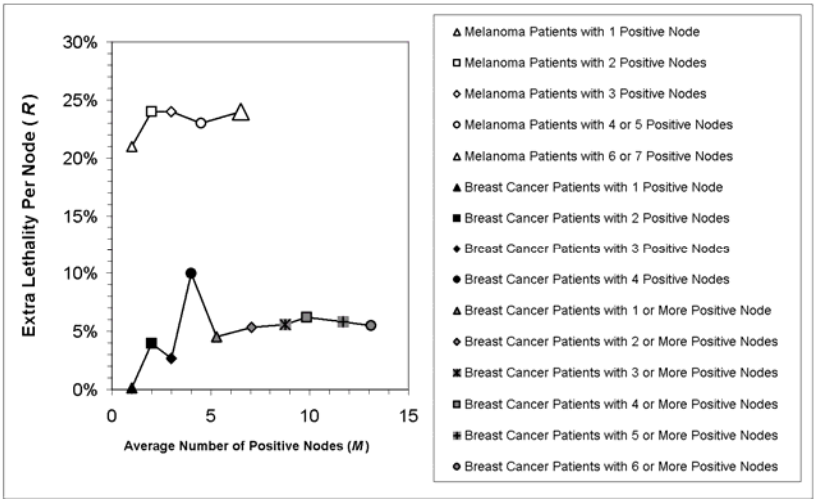

FIGURE Extra lethality ascribable to positive nodes, per node, *R*, estimated for melanoma and breast carcinoma with Equation #10. See TABLE A3 for values.

C:\X_MelanomaAnalysisByExcell 1 1 04_____Fig Value of R 4 25 04\[Fig For R May 25 04.xls]Sheet1 (2)'!$C$28

TABLE A3
Cancer death rates for groups of melanoma breast carcinoma patients sorted by nodal status

Patient Group	Average Number of Positive Nodes (M)	Number of Patients	Mean Tumor Thickness or Diameter (mm) (D)	Actual 15 year Disease Specific Death $L_{overall}$	The Level Of Death Expected for Node-negative Patients with Tumors of these Sizes $L_{primary}= 1-e^{-(Q^* j_{primary})D^Z}$ (Eq. 1c)	Extra Death Ascribable To Positive Nodes, per node ($L_{per\text{-}node}$) (Eq. #17)
Melanoma						
Node Negative	0	487	2.35	21%	21%	-
1 positive node	1	92	3.24	40%	25%	21%
2 positive nodes	2	36	3.43	55%	27%	24%
3 positive nodes	3	21	4.14	67%	32%	24%
4 & 5 positive nodes	4.5	12	5.11	77%	34%	23%
6 & 7 positive nodes	6.5	10	3.64	86%	30%	24%
Breast Carcinoma						
Node Negative	0	790	16.0	20%	20%	-
1 positive node	1	130	25.7	26%	26%	0.16%
2 positive nodes	2	71	29.0	34%	29%	3.96%
3 positive nodes	3	46	30.7	37%	32%	2.66%
4 positive nodes	4	47	35.4	57%	36%	9.97%
1 or more positive node	5.28	443	35.3	49%	35%	4.52%
2 or more positive nodes	7.06	313	39.3	58%	38%	5.35%
3 or more positive nodes	8.76	233	42.9	64%	42%	5.58%
4 or more positive nodes	9.85	196	45.0	69%	42%	6.21%
5 or more positive nodes	11.70	149	48.1	73%	46%	5.84%
6 or more positive nodes	13.11	123	52.3	75%	49%	5.48%

The *Size+Nodes* method accurately predicts the risk of cancer death

Incorporating the information outlined above for estimating the independent contribution of primary tumor size (Equation #1c) and numbers of positive nodes ($L_{per\text{-}node}$ ≈6% per positive node for breast carcinoma, $L_{per\text{-}node}$ ≈23% for melanoma; Equation #2), provides a technique, the *Size+Nodes* method (Tables 1 and 2), for integrating tumor size and nodal status into an estimate of the risk of death (L) for each patient:

$$L = L_{primary} + L_{nodes} - (L_{primary} * L_{nodes}) \quad (4)$$

where:

$$L_{primary} = 1 - e^{-(Q^* j_{primary}) D^Z} \quad (1c)$$

and:

$$L_{nodes} = 1 - e^{-(M^* L_{per-node})} \quad (2)$$

As we have reported previously, the *Size+Nodes* method can accurately estimate the risk of death for breast carcinoma patients, as well as accurately stratify these patients into groups of incrementally increasing lethality.[6] In contrast, classifying breast carcinoma patients by node positivity, or by T group, or by stage, creates groups of women with wide and overlapping levels of lethality. Here we show that this is also true for melanoma (Figures 4 and 5, Tables 4 and 5). For example, when we subdivided the melanoma patients into six separate groups based upon the risk of death predicted by the *Size+Nodes* method (0%-10% estimated risk of death, 10%-19%, 20%-29%, 30%-39%, 40%-59%, 60%-79%, and 80%-99%, Table A4), the actual Kaplan-Meier cancer death rate for each group agrees remarkably closely with the risk of death estimated by the *Size+Nodes* method (R^2=0.97, Figure A4). Furthermore, when tested on groups of melanoma patients sorted into 28 categories by a variety of patient characteristics, including Clark's level, sex, body location, histological sub-type, and ulceration, the estimate of the risk of death made by the *Size+Nodes* method agreed within the 95% confidence interval with the actual death rate measured by the Kaplan-Meier method in every case (Table A5). There was a remarkably linear correlation between the estimations of the risk of death made by the *Size+Nodes* method and actual melanoma death rates for these 28 groups (R^2=0.95, Figure A5).

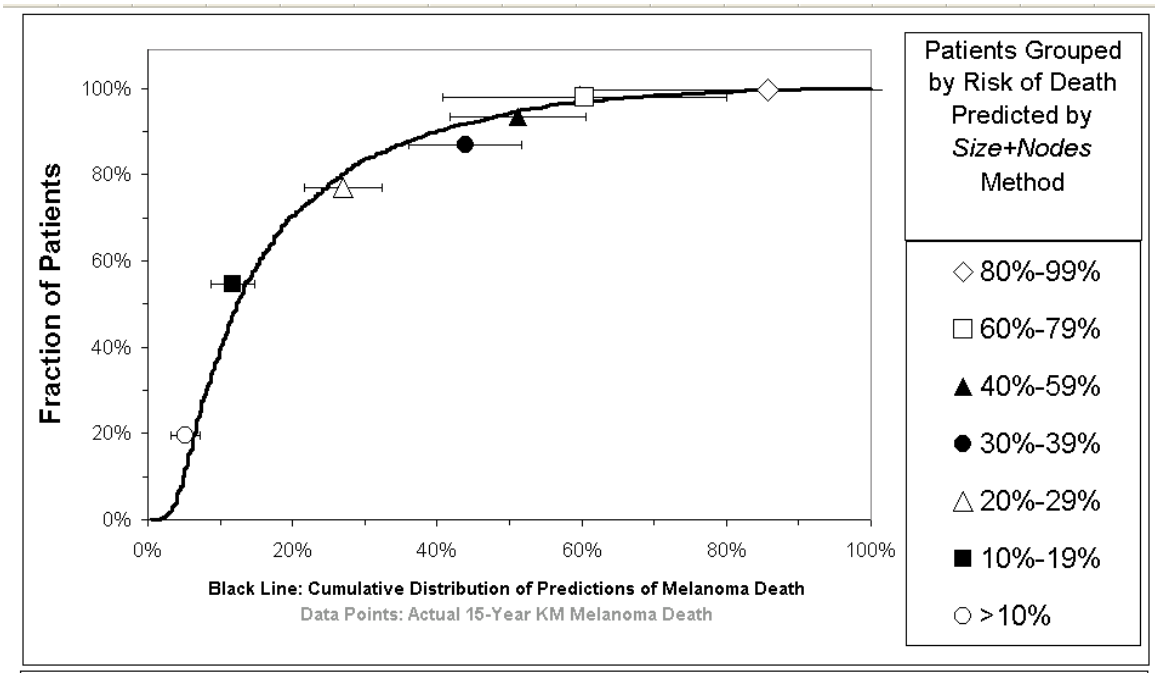

FIGURE A4(a)

FIGURE A4. Capacity of the *Size+Nodes* method (eq. (4)), to stratify melanoma patients according to the risk of death. (a) Cumulative distributions of the estimates values of risk of death estimated from tumor size and nodal status information by the *Size+Nodes* method (Table A1).

TABLE A4
Stratification of Melanoma Patients by Risk of Death Estimated by the *SizeOnly* and *Size+Nodes* Methods

	***SizeOnly* Method**			***Size+Nodes* Method**		
Risk Group	**Risk of Melanoma Death**	**Actual 15 Year Melanoma Death**	**Number of Patients**	**Risk of Melanoma Death**	**Actual 15 Year Melanoma Death**	**Number of Patients**
0%-10%	6.42%	5.17%	1061	6.45%	4.65%	1084
10%-19%	14.31%	11.71%	814	14.19%	12.62%	859
20%-29%	24.56%	27.02%	441	24.77%	28.32%	366
30%-39%	34.63%	43.88%	238	34.89%	38.41%	179
40%-59%	48.60%	51.20%	168	48.61%	50.36%	191
60%-79%	68.12%	60.42%	35	68.16%	62.79%	59
80%-99%	85.26%	85.71%	7	86.72%	91.38%	26

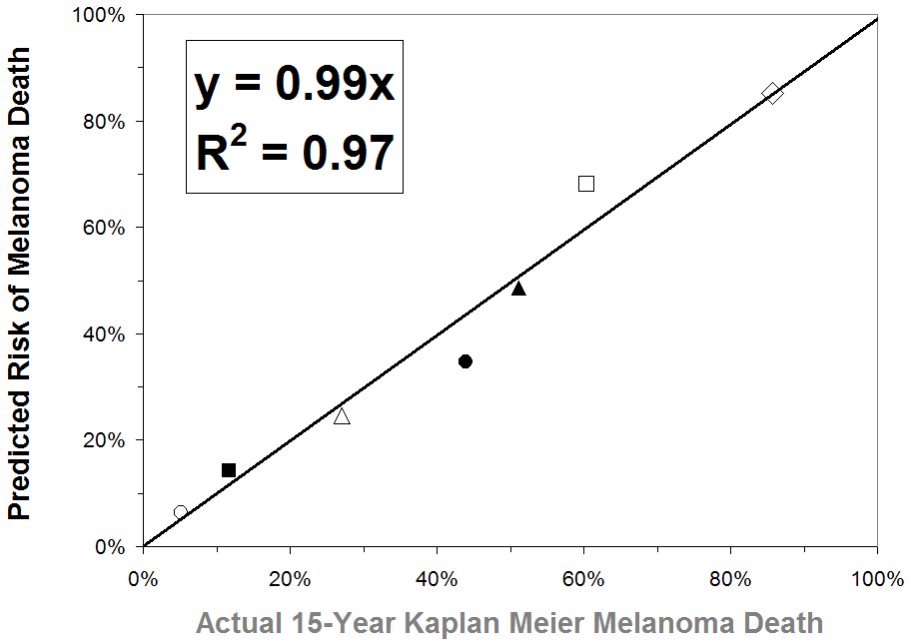

FIGURE A4(b)

(b) Scatter plot comparing the risk of death estimated from tumor size and nodal status information by the *Size+Nodes* method, in comparison to the actual 15-year Kaplan Meier death rates for nine groups stated by virtue of the value the risk of death estimated from tumor size and number of positive nodes by the *Size+Nodes* method. See Table A5 for values. For comparable data for breast carcinoma, see reference 4.

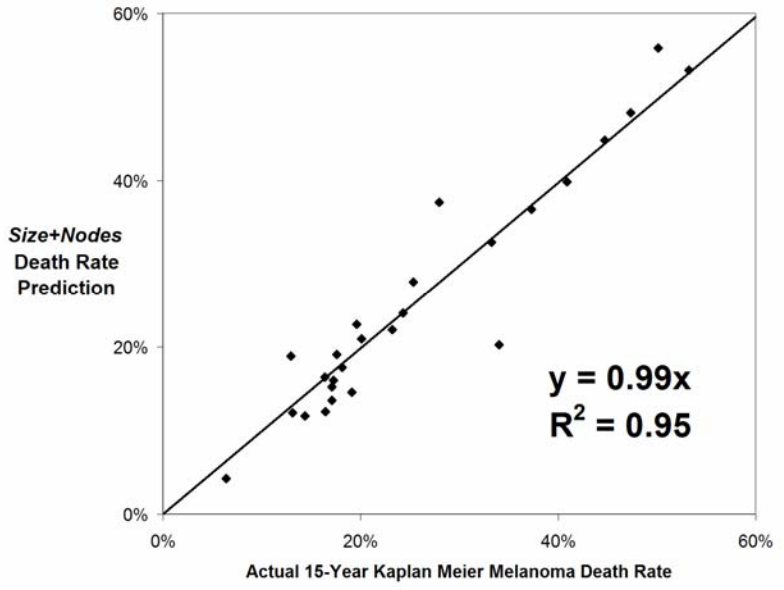

FIGURE A5

FIGURE A5. Comparison of risk of death estimated by the *Size+Nodes* method and actual death rates among patients sorted in various ways (see Tables 4 and 5)

The comparative outcomes of the *Size+Nodes* and *SizeOnly* methods

When used to stratify patients according to the risk of death, the *Size+Nodes* method appears to be slightly better than the *SizeOnly* method. For example, for melanoma, the *Size+Nodes* method placed 26 and 1084 patients in the highest and lowest risk groups (the 80%-99% estimated risk of death group vs. the 0%-10% estimated risk of death group) while the *SizeOnly* method assigned 7 and 1061 patients to each of these groups (Table A4). However, the most striking advantage of the *Size+Nodes* can be seen in its superior ability to estimate the risk of death for groups of patients sorted by nodal status (Table A5). Indeed, in comparison to the *Size+Nodes* method, the *SizeOnly* method over-estimates the chance of death of node-negative patients and under-estimates the chance of death of node-positive patients (Table A5).

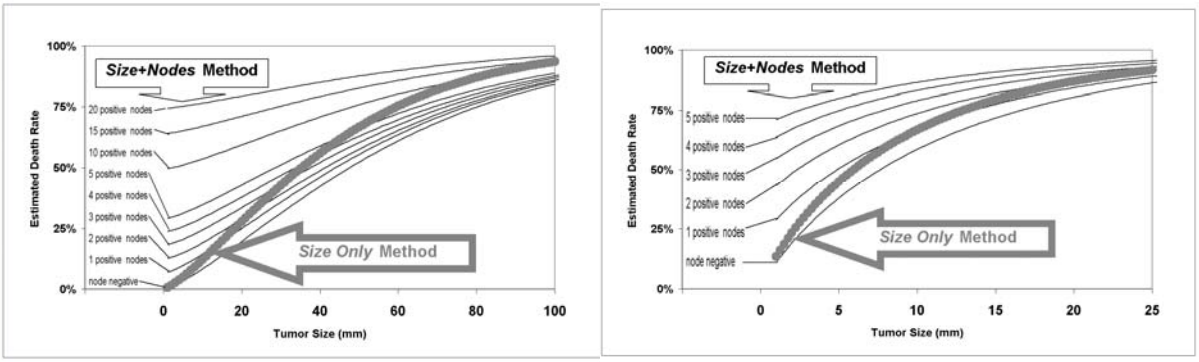

FIGURE 1A The relationship between tumor size and risk of death for the *SizeOnly* Method (Equation #1; thick grey line) and the relationship between tumor size, nodal status, and risk of death for the *Size+Nodes* Method (Equation #4, thin black lines) for breast carcinoma

FIGURE1B

The relationship between tumor size and risk of death for the *SizeOnly* Method (Equation #1; thick grey line) and the relationship between tumor size, nodal status, and risk of death for the *Size+Nodes* Method (Equation #4, thin black lines) for melanoma.

TABLE A5 Comparison of risk estimates made by the *SizeOnly* and *Size+Nodes* methods to empirical death rates among various groups of melanoma patients

Factor (*N*)		$L_{empirical}$ **(95% CI)**		$L_{predicted}$ **SizeOnly (95% CI)**		$L_{predicted}$ **Size+Nodes (95% CI)**	
Nodal Status (664)	Negative (487)	21.02%	(5.42%)	24.77%	(1.06%)	20.12%	(0.91%)
	Positive (177)	53.21%	(9.64%)	33.29%	(2.10%)	53.21%	(2.67%)
	1 Positive (92)	39.87%	(14.71%)	30.27%	(2.81%)	40.91%	(2.07%)
	1,2 Positive (127)	44.85%	(12.08%)	31.04%	(2.35%)	44.71%	(1.99%)
	1,2,3 Positive (148)	48.14%	(10.82%)	31.96%	(2.15%)	47.36%	(2.08%)
	2+ Positive (85)	66.24%	(11.92%)	36.56%	(3.00%)	66.81%	(3.12%)
	3+ Positive (50)	74.46%	(13.69%)	39.03%	(4.05%)	75.67%	(3.23%)
Clark Level (2492)	2 (773)	4.25%	(2.20%)	6.42%	(0.22%)	6.41%	(0.22%)
	3 (655)	11.78%	(3.30%)	14.16%	(0.67%)	14.34%	(0.81%)
	4 (964)	27.83%	(3.94%)	24.53%	(0.78%)	25.32%	(0.98%)
	5 (100)	55.87%	(13.74%)	46.73%	(3.13%)	50.14%	(3.65%)
Site (2747)	Trunk (1017)	19.12%	(3.21%)	16.36%	(0.81%)	17.57%	(1.01%)
	Face (238)	16.39%	(7.23%)	16.72%	(1.71%)	16.40%	(1.90%)
	External Ear (71)	14.61%	(10.52%)	17.54%	(2.75%)	19.09%	(3.15%)
	Upper Limb and Shoulder (594)	15.26%	(3.87%)	17.59%	(1.15%)	17.08%	(1.28%)
	Lower Limb and Hip (647)	16.03%	(3.38%)	18.36%	(1.04%)	17.24%	(1.27%)
	Scalp and Neck (180)	24.12%	(7.36%)	22.31%	(2.45%)	24.30%	(3.05%)
Histology (2742)	Superficial Spreading (1610)	12.17%	(2.16%)	12.84%	(0.47%)	13.12%	(0.56%)
	Lentigo Malignant (221)	18.92%	(8.99%)	13.08%	(1.56%)	12.93%	(1.63%)
	Malignant (453)	22.08%	(4.36%)	22.46%	(1.37%)	23.23%	(1.57%)
	Acral Lentiginous (68)	37.42%	(14.30%)	25.92%	(3.85%)	27.96%	(4.89%)
	Nodular (351)	32.63%	(6.28%)	31.79%	(1.60%)	33.26%	(1.91%)
	Desmoplastic (39)	20.27%	(14.79%)	35.78%	(5.31%)	34.04%	(5.32%)
Ulceration (1040)	Absent (856)	13.63%	(4.93%)	17.03%	(0.85%)	17.09%	(0.83%)
	Present (184)	36.53%	(8.41%)	34.28%	(1.17%)	37.27%	(2.97%)
Sex (2762)	Female (1299)	12.27%	(2.25%)	16.33%	(0.70%)	16.42%	(0.77%)
	Male (1463)	22.71%	(2.89%)	18.76%	(0.75%)	19.61%	(0.88%)
All Patients (2770)		17.59%	(1.84%)	17.59%	(0.51%)	18.11%	(0.59%)

A simple mnemonic for the *Size+Nodes* method

Fortuitously, for melanoma each millimeter in tumor thickness is associated with about an 8% increase in lethality and each positive lymph node is associated with about a 23% extra chance of death, while for breast carcinoma each millimeter of tumor diameter is associated with about a 1% increase in lethality and each positive lymph node is associated with about a 6% extra chance of death (Figure A6). This provides a convenient mnemonic for the *Size+Nodes* calculation (Table A2).

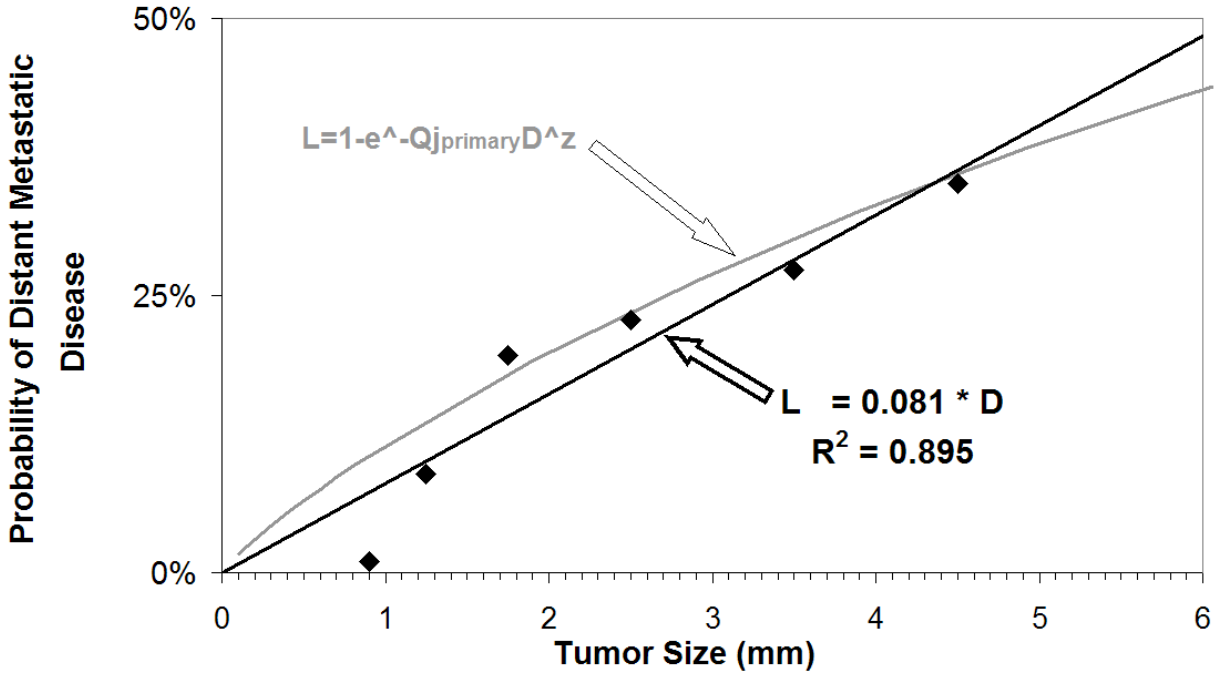

FIGURE A6

FIGURE A6. Approximate linearity of the relationship between the size of the primary mass and the melanoma death rate in node negative patients. Closed diamonds are 15-Year Kaplan-Meier melanoma death rates for groups of node negative patients sorted by primary tumor size (for data, see previous paper). Shown is a linear regression to data on cancer death node-negative melanoma patients sorted by primary tumor size (black line) as well as the fit of Equation #1c to these data (gray line). For comparable data for breast carcinoma, see reference 4.

TABLE III

A simple mnemonic for the *Size+Nodes* method			
	Lethal contribution for each mm of primary tumor size		Lethal contribution for each positive lymph node
Risk of Breast Carcinoma Death=	~1% per mm	+	~6% per node
Risk of Melanoma Death=	~8% per mm	+	~23% per node

Pathways of the spread of cancer cells and their probabilities

The spread of cancer cells can occur along a variety of pathways, each with its characteristic probability: from the primary site to the nodes ($p_{to\text{-}nodes}$), from node to node ($p_{node\text{-}to\text{-}node}$), from a node to the periphery leading to death ($p_{from\text{-}nodes}$), from the primary site to the periphery leading to death in node negative patients ($p_{primary}$), or from the primary site to the periphery leading to death in patients as a whole (p, which is the aggregate of all of these probabilities) (Table B1, Figure B1). Here we shall use clinical data to calculate the probabilities of these events of spread.

To see the general idea for such calculations, consider a group of patients with cancers of ~100 million cells ($N \approx 10^8$). If the cancer death rate in these patients was found to be ~10% ($L \approx 0.1$), then it would follow that the probability of the lethal spread of cancer cells from the primary site to the periphery, p, will be approximately 1 event of spread for every billion cells in the primary mass ($p \approx (L/N) \approx (0.1/10^8) \approx 10^{-9}$). More precisely, the probability of the spread of a cancer cell, p_x, can be calculated with the expression we call the *ProbabilityEstimation* Equation:

$$p_x = -\ln(1 - L_x)/N \tag{7}$$

Let us recall that by definition, p_x does not consider events of spread that do not lead to a macroscopic manifestation, nor does it assume p_x to be constant, nor does it require that every cell in the mass of cancer have the potential for spread, although this may be the case (see the mathematical derivation of the *ProbabilityEstimation* Equation #7). Thus, in the example given above ($L \approx 0.1$, $N \approx 10^8$), whether every cell in the cancer mass has the potential for spread, or whether only 1-in-a-million cells in the cancer mass has the potential for spread, the probability, p, of an event of spread *per cell* in the cancer mass, will still be 1-in-a-billion ($p \approx 10^{-9}$).

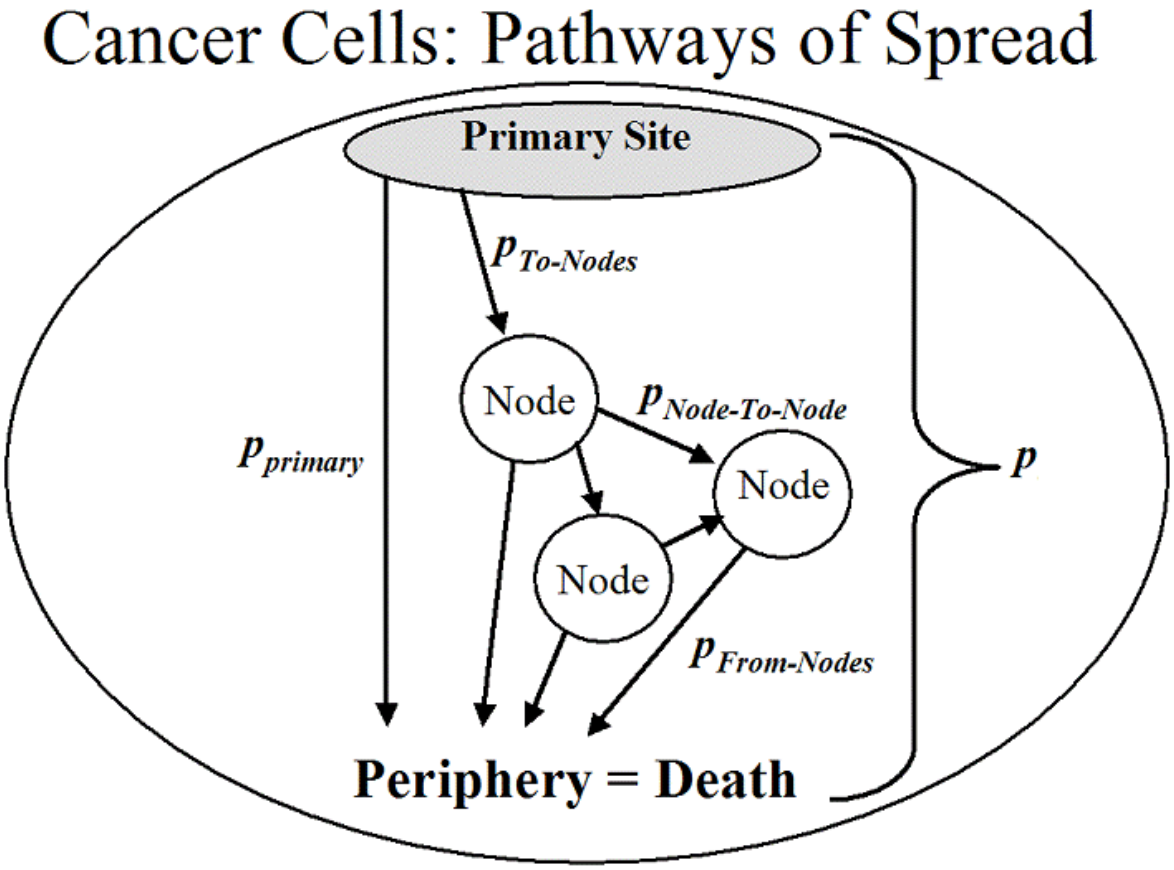

FIGURE B1. Cancer cells: pathways of spread.

TABLE B1
Pathways of The Spread of Cancer Cells and their Probabilities

Pathway of Spread	Seen In	Fraction of patients with a manifestation of spread	Probability of spread, per cell
Lethal spread from the primary site to the periphery (directly only)	Fraction of patients dying (=15-year Kaplan-Meier cancer death rate) among node-negative patients	$L_{primary}$	$p_{primary}$
Non-lethal spread from the primary site to the local lymph nodes	The fraction of patients with positive nodes	$L_{to\text{-}nodes}$	$p_{to\text{-}nodes}$
Lethal spread from the lymph nodes to the periphery	Lethal contribution per positive lymph node	-	$p_{from\text{-}nodes}$
Non-lethal spread from lymph node to lymph node	The numbers of positive nodes (M)	-	$p_{node\text{-}to\text{-}node}$
Lethal spread from the primary site to the periphery (The aggregate consequence of the pathways of spread characterized by $p_{primary}$, $p_{to\text{-}nodes}$, $p_{from\text{-}nodes}$ and $p_{node\text{-}to\text{-}node}$)	Fraction of patients dying (=15-year Kaplan-Meier cancer death rate) among all patients	L	p

The per-cell probability of the spread of cancer declines as tumors get larger

Cancer cells may conceivably spread from a primary mass to the local lymph nodes (resulting in cancer seen in the node upon pathological analysis) or from a primary mass to the periphery (resulting in death). Lethal spread may also conceivably occur directly from the primary site to the periphery, or indirectly from the primary site to a node and then to the periphery (Table B1, Figure B1). The *ProbabilityEstimation* Equation (#7) allowed us to calculate the probability of such events of the spread of cancer cells. These calculations (Figures 2 and 3) reveal the remarkable finding that as tumors increase in size, the per-cell probability of the spread of cancer leaving a mass does not remain constant, but declines in value.[3,5] Furthermore, this decline occurs in a very characteristic fashion, such that the relationship between the probability of the spread of cancer cells and the size of the mass from which the cells emerge, N, is well fit by a power function:

$$p_x = aN^b \tag{9}$$

where $b \approx -2/3$ and a is characteristic of each malignancy. This holds true in each of five different contexts where we have had the data to carry out these calculations: the spread of cancer cells from the primary site to the nodes ($p_{to\text{-}nodes}$, seen in the fraction of patients found to have cancer in the nodes, $L_{to\text{-}nodes}$), as seen in data on breast cancer and melanoma patients; the spread of cancer cells from the primary site to the periphery leading to death in node negative patients ($p_{primary}$, seen in the fraction of node negative patients dying of cancer, $L_{primary}$) , as seen in data on breast cancer patients; and the spread of cancer cells from the primary site to the periphery leading to death in all patients (p, seen in the fraction all patients who die of cancer, L), as seen in data on breast cancer and melanoma patients (Figures B2 and B3).

Why might this decline in the value of p_x occur as tumors get bigger? One possible explanation is that a is the intrinsic probability that a cancer cell will spread, but that only a fraction of cells in the mass are capable of spreading, with the size of that fraction being N^b. Another possible explanation is simple geometry; every cell in a primary mass might be capable of spreading, but as tumors get larger, there are simply more and more cells that must be "pushed aside" before any individual cell can escape from that mass. Indeed, we have shown mathematically[5] that just a process should be expected to result in the reduction in the per-cell probability of spread, such that the relationship between p and N is well fit to Equation #9, $p=aN^b$, with $b \approx -2/3$.

Note that the value of a in Equation #9 is at least five-fold greater for melanoma than breast carcinoma and at least fifteen-fold greater for renal cell carcinoma. Indeed, a describes the probability of spread of cancer cells at the extrapolated point in time when the tumor was but 1 cell in size, and thus when there were no other cells to be "pushed aside" ($p=a$ in Equation #9 when $N=1$). Thus, the component a in Equation #9 appears to be a measure of tumor biology, before the impact of the size of the primary mass can be felt.

TABLE B2
Sizes of Melanoma Nodal Deposits

Node Number	Measured area of the cancerous deposits in this node	Contains approximately the same number of cells as a primary mass of thickness (in mm)	Lethal contribution expected from a primary mass of this size (*SizeOnly* Equation)
1	0.034	0.13	2%
2	0.116	0.25	3%
3	0.133	0.27	3%
4	0.157	0.29	4%
5	0.187	0.32	4%
6	0.201	0.33	4%
7	0.271	0.38	5%
8	0.348	0.44	5%
9	0.424	0.48	6%
10	0.673	0.61	7%
11	0.775	0.66	8%
12	0.853	0.69	8%
13	0.907	0.71	8%
14	1.047	0.76	9%
15	1.285	0.85	9%
16	1.496	0.92	10%
17	1.73	0.99	11%
18	1.768	1.00	11%
19	1.936	1.04	11%
20	2.793	1.26	13%
21	2.798	1.26	13%
22	3.236	1.36	14%
23	3.409	1.39	14%
24	4.095	1.53	15%
25	4.6	1.62	16%
26	5.567	1.79	17%
27	7.868	2.13	20%
28	10.261	2.44	22%
29	12.322	2.68	24%
30	13.33	2.79	25%
31	14.706	2.93	26%
32	16.575	3.11	27%
33	21.894	3.59	30%
34	29.045	4.14	33%
35	29.457	4.17	33%
36	30.372	4.24	34%
37	31.314	4.30	34%
38	33.183	4.43	35%
39	46.069	5.24	39%
40	46.621	5.27	39%
41	57.695	5.87	42%
42	59.6	5.97	43%
43	63.617	6.17	44%
44	86.59	7.22	48%
45	86.59	7.22	48%
46	103.869	7.92	51%
47	117.859	8.44	53%
48	132.732	8.97	55%
49	153.93	9.67	58%
50	153.938	9.67	58%
average	28.005	3.00	23%

TABLE B3
Sizes of Breast Carcinoma Nodal Deposits

	Diameter* (mm)	Estimated area of the cancerous deposits in this node	Fraction of the metastatic area containing cancer**	Contains approximately the same number of cells as a primary mass of diameter (in mm)	Lethal contribution expected from a primary mass of this size (*SizeOnly* Equation)
1	0.3	0.09	1	0.3	0.12%
2	1	1.05	1	1	0.62%
3	1	1.05	1	1	0.62%
4	1.1	1.27	1	1.1	0.70%
5	1.4	2.05	1	1.4	0.97%
6	1.9	3.78	0.65	1.5	1.06%
7	2	4.19	1	2	1.56%
8	2	4.19	0.95	2	1.56%
9	2.1	4.62	1	2.1	1.66%
10	2.5	6.54	0.93	2.4	1.98%
11	2.7	7.63	0.95	2.6	2.21%
12	3	9.42	0.95	2.9	2.55%
13	3	9.42	0.1	3	2.67%
14	3	9.42	0.95	3	2.67%
15	3.2	10.72	0.1	3.2	2.90%
16	3.4	12.11	0.95	3.3	3.02%
17	3.4	12.11	0.93	3.3	3.02%
18	3.9	15.93	0.86	3.6	3.39%
19	3.7	14.34	0.95	3.7	3.52%
20	3.9	15.93	0.1	3.9	3.77%
21	4.3	19.36	0.95	4.2	4.15%
22	4.5	21.21	0.95	4.4	4.41%
23	5.5	31.68	0.76	4.8	4.95%
24	5.5	31.68	0.78	4.9	5.08%
25	5.4	30.54	0.86	5	5.22%
26	5.4	30.54	0.95	5.2	5.49%
27	8	67.02	0.43	5.2	5.49%
28	5.4	30.54	0.1	5.4	5.77%
29	6	37.70	0.08	5.4	5.77%
30	7.1	52.79	0.06	5.5	5.91%
31	6	37.70	0.09	5.7	6.19%
32	6.4	42.89	0.1	6.4	7.19%
33	11	126.71	0.35	6.5	7.33%
34	10	104.72	0.49	7	8.07%
35	7.1	52.79	0.1	7.1	8.21%
36	7.1	52.79	0.1	7.1	8.21%
37	7.7	62.09	0.95	7.5	8.81%
38	9	84.82	0.72	7.6	8.96%
39	10.7	119.89	0.55	7.9	9.42%
40	9	84.82	0.08	8	9.57%
41	10	104.72	0.65	8.1	9.72%
42	12.5	163.62	0.45	8.4	10.18%
43	9	84.82	0.09	8.5	10.34%
44	9.8	100.57	0.84	9	11.11%
45	10	104.72	0.86	9.3	11.58%
46	10	104.72	0.95	9.7	12.21%
47	11	126.71	0.95	10.7	13.80%
48	14.5	220.17	0.85	13.4	18.19%
49	15	235.62	0.1	15	20.83%
average	5.95	51.30	0.644	5.31	5.97%

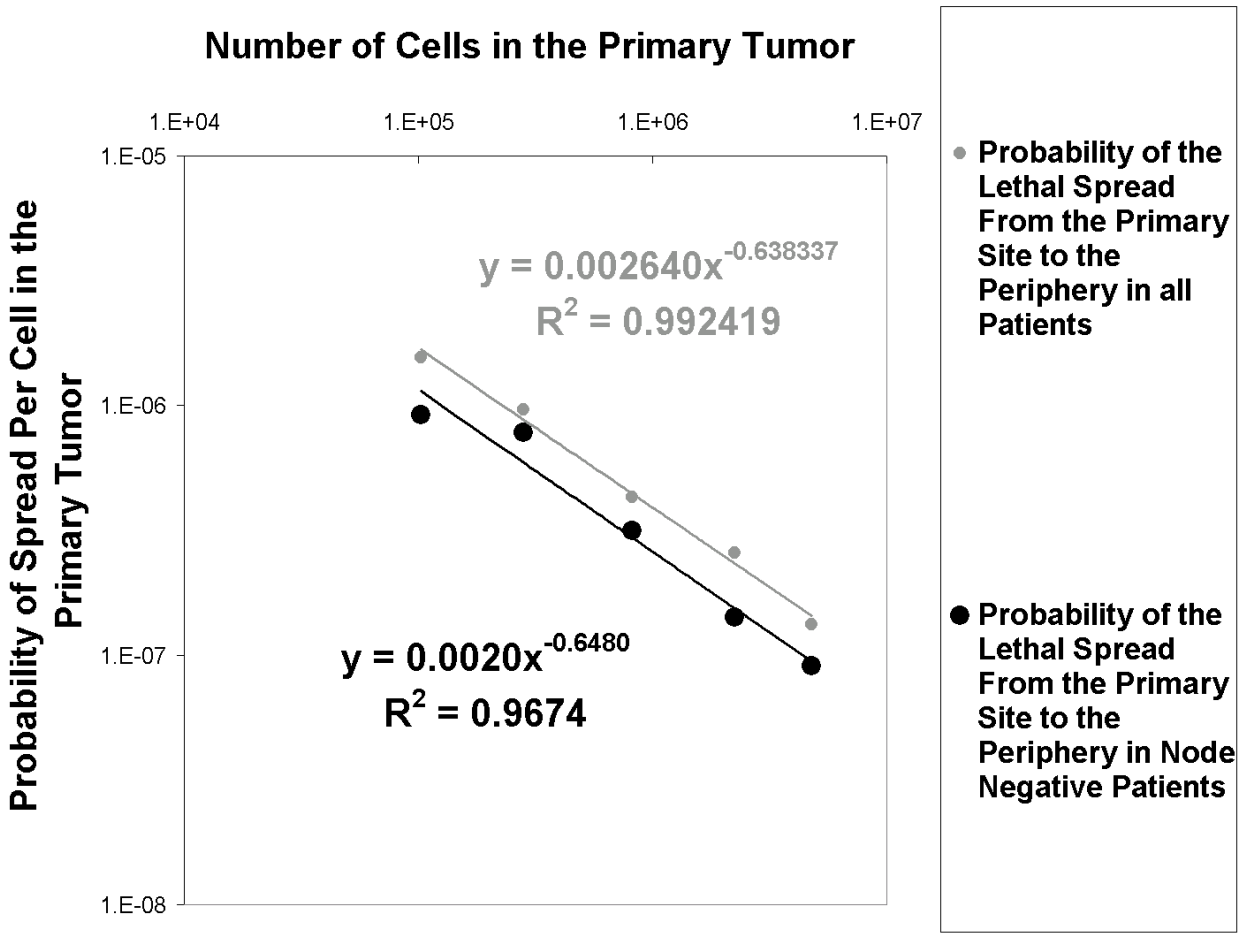

FIGURE B2

FIGURE B2. Calculations with the ProbabilityEstimation Equation (#7) of the probabilities of lethal spread of breast carcinoma from the primary site to the periphery, using tumor size/survival data for all patients, and of the lethal spread of breast carcinoma from the primary site to the periphery with tumor size/survival data for node-negative patients by Equation #13. Note the in both cases the relationship between the probability of spread and tumor size is well fit by a power function. Tumor sizes groups examined: 1 to 10 mm, 11 to 20 mm, 21 to 25 mm, 26 to 30 mm, 31 to 35 mm. Cell numbers estimated by Equation #8 in accompanying paper.

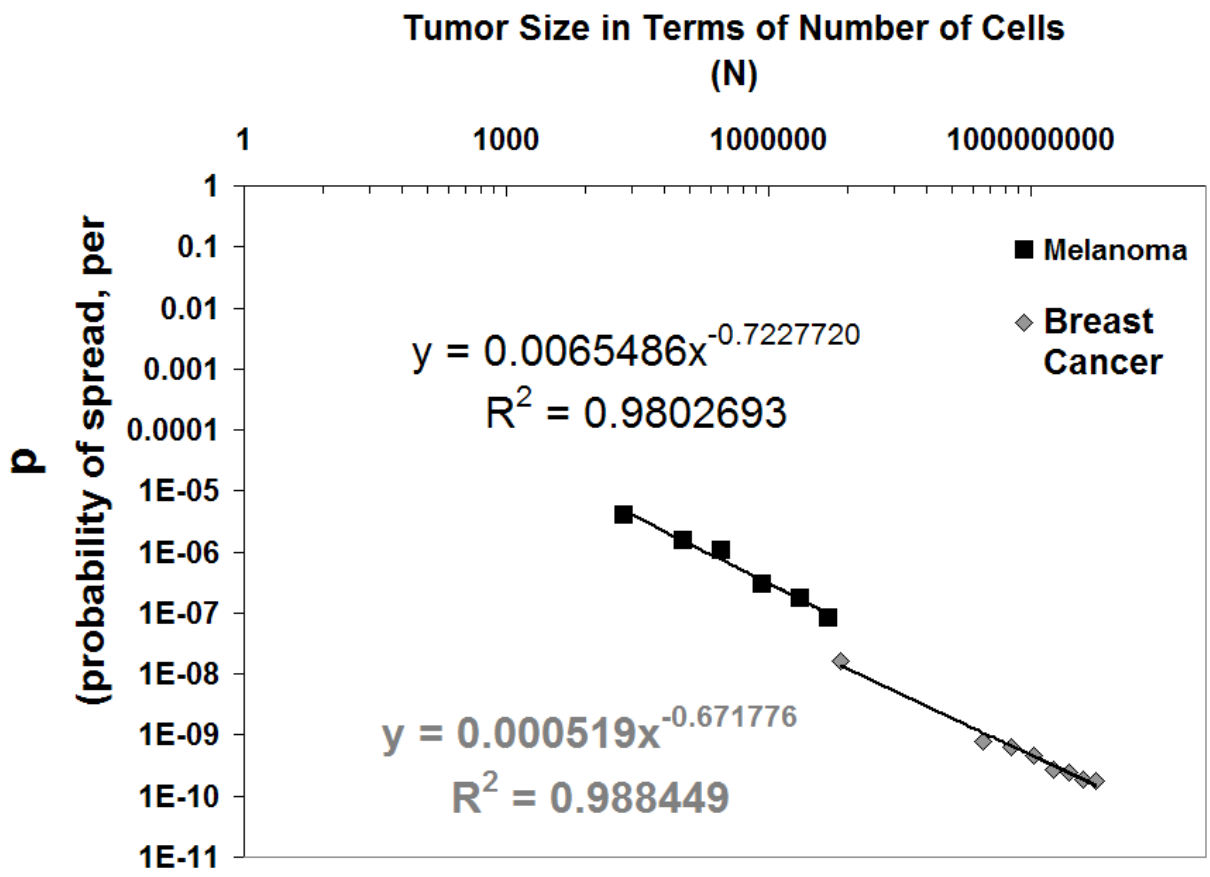

FIGURE B3

FIGURE B3. Calculations with the ProbabilityEstimation Equation (#7) of the probabilities of non-lethal spread of breast carcinoma from the primary site to the nodes, using data on the fraction of patients who are node positive, and the probabilities of non-lethal spread of melanoma from the primary site to the nodes, using data on the fraction of patients who are node positive. Note the in both cases the relationship between the probability of spread and tumor size is well fit by a power function. Tumor sizes groups examined: melanoma 0 to 1 mm, 1 to 1.5 mm, 1.5 to 2 mm, 2 to 3 mm, 4 to 5 mm; Breast Carcinoma: 1 to 10 mm, 11 to 20 mm, 21 to 25 mm, 26 to 30 mm, 31 to 35 mm, 36 to 40 mm, 41 to 45 mm, 45 to 50 mm. Cell numbers estimated by Equation #8 in the accompanying paper.

Spread of cancer cells from node to node.

Cancer in a lymph node may well be the result of the spread of a cell from a primary mass, but is the primary mass the only source for such cells? In fact, there are powerful indications that once cancer gets to a node, it may then form a mass from which progeny may spread to other nodes. Note that we saw from Equation #9 that the probability of non-lethal spread of cancer cells to the local nodes, $p_{\text{-to-nodes}}$, for breast carcinomas of 20-29 mm, is ~$1/3x10^9$. Since tumors of this size contain ~$7x10^8$ cells, it follows that if all of the deposits in the nodes came from cells at the primary site, then we should expect that about $(1/3x10^9 \text{ x } 7x10^8)^M$ patients would have M positive nodes. That is, ~25% of patients should have 1 positive node, ~6% of patients should have 2 positive nodes, ~1% of patients should have 3 positive nodes, ~0.4% of patients should have 4 positive nodes, and ~0.1% should have 5 positive nodes. However, the number of patients with 3, 4, 5 and more positive nodes is much greater than expected for spread directly from the primary site (Figure B4). The most direct inference is that once cancer has spread to a node, there is opportunity for further spread to another node (Table B1, Figure B1).

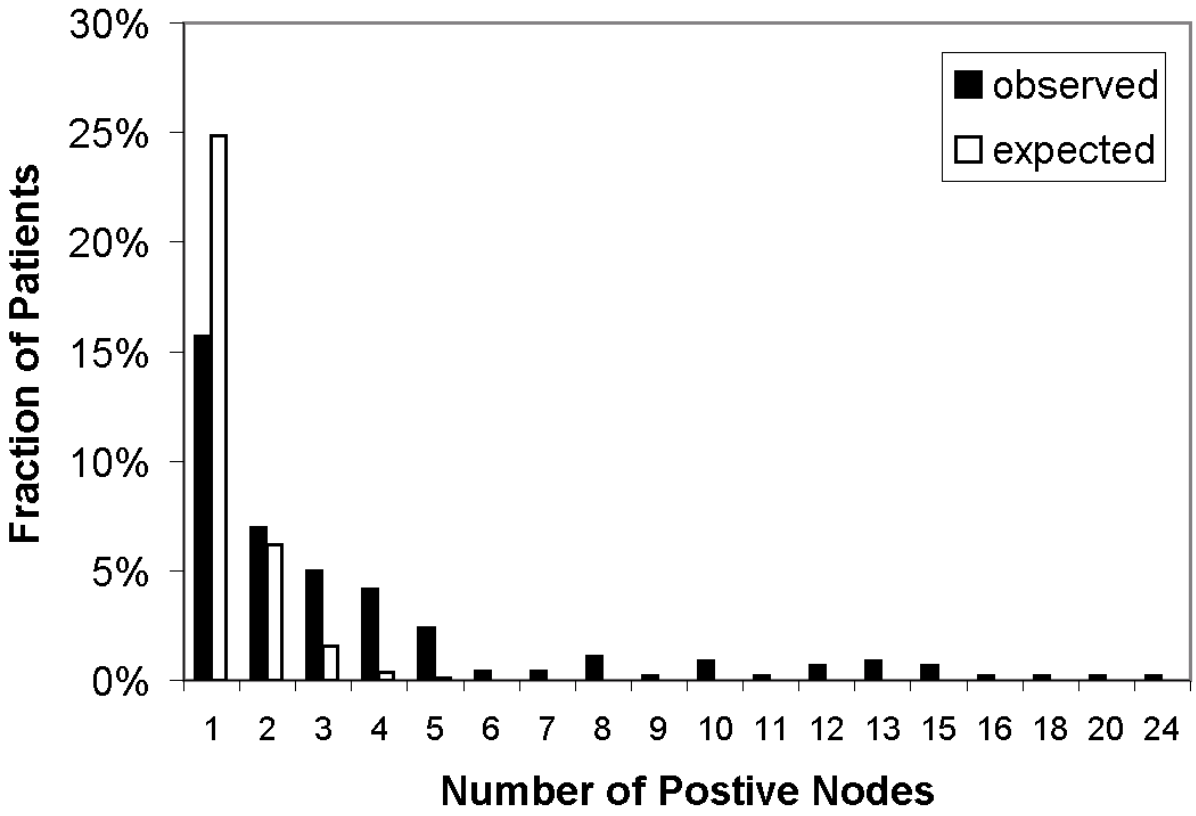

FIGURE B4

FIGURE B4. Histogram of the number of node-positive breast carcinoma patients with tumors of 20 to 29 mm, sorted by the number positive node found, together with the number of nodes expected if cancer in the nodes only originated from the primary site, occurring with per-cell probability of spread of $p=1/3x10^9$. Solid Bars: Observed fraction of patients with various numbers of positive nodes. Open Bars: Fraction of patients with various numbers of positive nodes expected if cancer in the nodes originated only from the primary site.

Spread of cancer cells from a node to the periphery, causing death.

For melanoma, the presence of cancer in a lymph node is associated with ~23% higher chance of death than in node negative patients with primary masses of the same size, while for breast carcinoma, each positive node is associated with ~6% extra chance of death.[6] These values may be used in the *ProbabilityEstimation* Equation (#7) to estimate the value of the probability of the lethal spread of a cancer cell from a lymph node to the periphery, $p_{from\text{-}nodes}$, if we also have information on the size of the cancer in the nodes, N. To provide such information we measured of the sizes of the cancer deposits in 50 positive nodes from melanoma patients and 49 positive nodes from breast carcinoma patients (Tables 2 and 3). These measurements revealed an average size of the nodal deposits of 28 mm^2 for melanoma and 51 mm^2 for breast carcinoma. Translating these values into estimates of the number of cells, N, (Equation #8) allows us to use the *ProbabilityEstimation* Equation (#7) to determine the value of $p_{from\text{-}nodes}$ (Table B4). These calculations revealed that for breast carcinoma, $p_{from\text{-}nodes}=8.0 \times 10^{-9}$, while for melanoma, $p_{from\text{-}nodes}=1.8 \times 10^{-7}$.

TABLE B4

The per cell probability of the lethal spread of cancer cells from the primary site and the nodes

Cancer Type	Average Nodal Size (D_{nodes})	Estimated Number of Cells* (N_{nodes}) (at 10^8 cells/cc)	Lethal contribution per node ($L_{per\text{-}node}$)	Probability of lethal spread per cell from the lymph nodes to the periphery: $p_{From-Nodes} = -\ln(1-L_{per-node})/N_{nodes}$ Equation #7n	Probability of lethal spread per cell from the primary site to the periphery (for the case where the cancer at the primary site is the same size as the average cancer deposits found in a positive lymph node): $p = aN_{nodes}^{\ b}$ Equation #9n
Breast carcinoma	5.31 mm	7.8×10^6	0.0608	8.0×10^{-09}	7.4×10^{-09}
Melanoma	3.00 mm	1.4×10^6	0.22527	1.8×10^{-07}	2.5×10^{-07}

* Estimates of the average number of cells per nodal deposit (N_{nodes}) were made by assuming spherical geometry and 10^8 cells per cc (see Equation #8 in the previous paper and discussion therein).

The chance of lethal spread is about the same whether from a node or from the primary site

The values for the probability of the lethal spread of cancer cells from a lymph node to the periphery described in the previous section ($p_{from\text{-}nodes}=8.0 \times 10^{-9}$ for breast carcinoma; $p_{from\text{-}nodes}=1.8 \times 10^{-7}$ for melanoma) are remarkably similar to the values of the probability of the spread of cancer cells from the primary site to the periphery, p, when the primary masses are the same size as the masses that are seen in the nodes ($p=7.4 \times 10^{-9}$ for breast carcinoma, $p=2.5 \times 10^{-7}$ for melanoma) (Table B4). These calculations reveal that the probability of the lethal spread of cancer cells to the periphery is remarkably similar whether the cells originate from a mass of cancer that is present in a lymph node or from an equally sized mass of cancer at the primary site.

The risk of death associated with cancer at the primary site and nodes reflects the amount of cancer present

Another way to look at the question of the lethal contributions from cancer at the primary site and in the nodes is to ask how the lethal contribution from each site is related to the amount of cancer present there. The *SizeOnly* Equation (#1) allows us to estimate what the risk of cancer death would be if one could magically move the amount of cancer present in a positive node to the cancer's primary site. Such calculations reveal that if the amounts of breast carcinoma found in the 49 breast carcinoma containing nodes had been present instead at the primary sites in the breasts of 49 patients, then we would have expected a breast carcinoma death rate of 5.97% (Table B2). This value is almost the same as the ~6% extra chance of death found to be correlated with the presence of each positive lymph node in breast carcinoma patients, as seen from actual survival data.[6] Similarly, if the amounts of cancer present in the 50 melanoma-containing nodes had been present instead at the primary sites in the skin of 50 patients,

then we would have expected a melanoma death rate of 23% (Table B2). Again, this value is almost the same as the ~23% extra chance of death found to be correlated with the presence of each positive lymph node in melanoma patients, as seen from actual survival data.

The *SizeAssessment* method determines whether a prognostic factor makes an independent contribution to lethality.

The *SizeOnly* equation (#1) makes it possible to determine whether a prognostic factor independently contributes to lethality, or is simply correlated with tumor size. In this test, which we call the *SizeAssessment* method, the actual 15-year cancer specific Kaplan-Meier death rate for a group of patients with a prognostic factor, $L_{empirical}$, is compared the predicted death rate, $L_{predicted}$, that would be expected by the *SizeOnly* equation for patients with tumors of these sizes. The statistical significance of the independent lethal contribution of this prognostic factor is assessed by comparing the difference of $L_{predicted}$ minus $L_{empirical}$ by an independent, two-sample Student's *t*-test with a threshold of 0.05.

The *SizeAssessment* method can be used to not only assess a prognostic factor's impact on the spread of cancer to the periphery, leading to death, but also to assess a prognostic factor's impact on the spread of cancer to the nodes, since both manifestations of the spread of cancer cells are well captured by equations of the form of the *SizeOnly* equation. The expression that relates tumor size to the chance of cancer in the nodes is called the *NodalSizeOnly* equation(#1n).

The *PrognosticMeasurement* method measures the magnitude of a prognostic factor's independent contribution to lethality:

The magnitude of a prognostic factor's impact on lethality can be incorporated into the *SizeOnly* and *NodalSizeOnly* equations by adding multipliers for each prognostic factor, which we call *g* parameters:

$$L_X = 1 - e^{-Q(g_1 * g_2 * g_3 * g_4 * ..) D^Z} \quad (1)$$

The value of the *g* parameter for the lethal contribution of each prognostic factor in the *SizeOnly* equation can be determined by a *pseudo-Monte Carlo* method. We call this technique for quantifying a prognostic factors impact on lethality the *PrognosticMeasurement* method (FIGURE C2).

It follows that the independent contribution of a prognostic factor to the chance of spread of cancer to the nodes can also be considered in terms of g_n parameters inserted into the *NodalSizeOnly* equation.

While the value of a *g* parameter is an abstraction, there are two practical ways to comprehend the nature of its magnitude. *First,* since the *g* parameter sits next to the tumor size, *D*, in the *SizeOnly* equation, patients with tumors with a prognostic factor for which g=2 can be expected to have the same death rate as patients with tumors without the factor but with tumors of twice the size. Likewise, patients with tumors with a prognostic factor for which g=0.5 can be expected to have the same death rate as patients with tumors without the factor but with tumors of half the size. *Second,* the *SizeOnly* equation is roughly linear over most of its range. Thus, patients with tumors with a prognostic factor for which g=2 can be expected to have the roughly twice the death rate as patients with tumors without the factor, while patients with tumors with a prognostic factor for which g=0.5 can be expected to have the roughly half the death rate as patients with tumors without the factor.

It may well be expected that some prognostic factors will be found by the *SizeAssessment* method to make independent contributions to lethality, but whose contributions to lethality, as measured by the *PrognosticMeasurement* method, may be found to be trivial in magnitude. For example, while the lobular and ductal carcinoma phenotypes make statistically significant contributions to lethality, as determined by by the *SizeAssessment* method, as measured by the *PrognosticMeasurement* method, the value of the *g* parameter for lobular was 0.9032 and the value of its *g* parameter for ductal was 1.057. This means that patients with lobular carcinomas have roughly 90% of the death rate of patients with ductal carcinoma (Table C2). For this reason, we have chosen the somewhat arbitrary term "*marked*" to identify those prognostic factors found to have independent contribution to lethality with a *p* value greater than 0.05 by the *SizeAssessment* method, and whose *g* parameter is either <0.74 or >1.33.

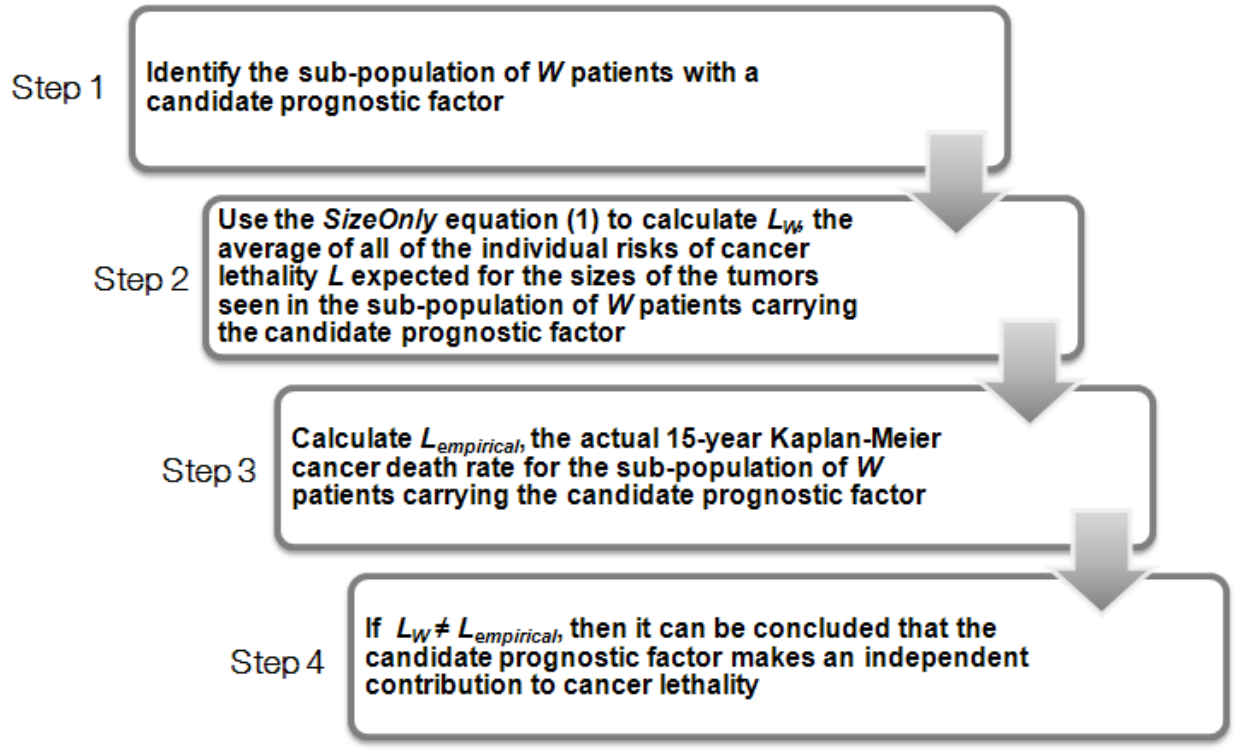

FIGURE C1

FIGURE C1. The *SizeAssessment* method for determining the qualitative impact of prognostic factors suspected of contributing to cancer lethality.

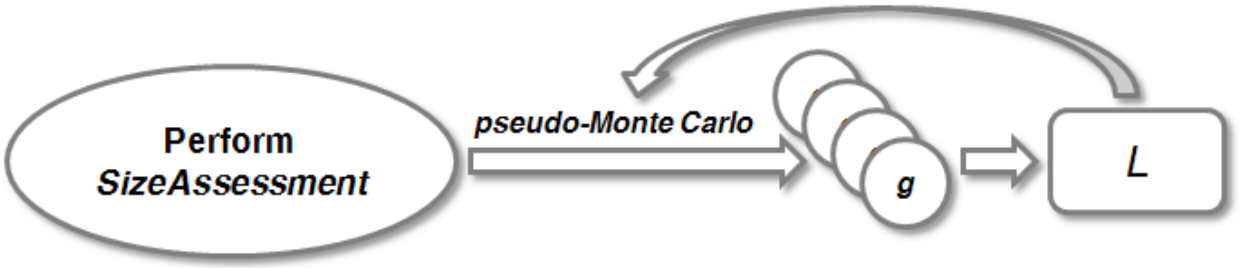

FIGURE C2

FIGURE C2. The *PrognosticMeasurement* method for quantifying the impact of prognostic factors which contribute to cancer lethality. For prognostic factors found to increase or decrease the propensity for lethal spread by the *SizeAssessment* method, we use a pseudo-Monte Carlo method to capture the magnitude of the impact of the prognostic factor with the expanded form of the *SizeOnly* equation. Each *g* parameter in that equation provides a measure of the impact of the corresponding prognostic factor.

The impact of prognostic factors on breast carcinoma lethality

By the *SizeAssessment* method, eight of thirty-two factors investigated (plus nodal status) were found to make *marked* independent contributions to breast cancer lethality (defined above as those prognostic factors found by the *SizeAssessment* method to have independent contribution to lethality with a *p* value greater than 0.05, and whose *g* parameter, measured by the *SizeAssessment* method, is either <0.74 or >1.33, Table C1). Factors examined were: grade, age, histology, ER/PR status, laterality, race, and sex. Factors found to make *marked* positive independent contributions to lethality included Paget's disease, scirrhous adenocarcinoma, and inflammatory disease. Factors found to make *marked* negative independent contributions to lethality included having a medullary or papillary carcinoma, mucinous or tubular adenocarcinoma, or possessing a cancer of grade 1. Similar results were found for the patients in the Partners, Van Nuys, and SEER datasets (Table C2), and the correlation between *g* parameters derived from the Partners and the SEER datasets can be seen in FIGURE C3(a).

Seven of the eight factors found to make a *marked* independent contribution to lethality by the *SizeAssessment* method were also found by the *SizeAssessment* method to make a *marked* independent contribution to the propensity of spread to the nodes (Table C3). Only scirrhous adenocarcinoma was found to make a *marked* independent contribution to lethality but not make a *marked* independent contribution to the propensity of spread to the nodes. The correlation between *g* and g_n correlated with the prognostic factors in the SEER dataset can be seen in FIGURE C3(b).

The impact of prognostic factors on melanoma lethality

Of the five melanoma prognostic factors shown in Table C6 which we investigated by the *SizeAssessment* method - Clark level, the initial site of occurrence, histological sub-type, ulceration, and sex - only sex was found to make statistically significance independent contribution to breast cancer lethality. In addition, nodal status, was also found to make statistically significance independent contribution to lethality. The *g* parameter for female sex was 0.7711, and 1.2062 for male sex, indicating a relatively small contribution to lethality, which did not rise to the criterion of a *marked* independent contribution to breast cancer lethality, as defined above. It is conceivable that some of the other prognostic factors do in reality make a *marked* independent contribution to lethality, but their significance could not be captured by a dataset of this sample size (2,770 patients). For example, the apparently large, negative contribution to cancer death of desmoplastic histology (g=0.4902) but its modest statistical significance (p=0.0568) makes one wonder whether its independent contribution to lethality might achieve significance given a larger sample size.

The *Size+Nodes+PrognosticFactors* (*SNAP*) method combines tumor size, nodal status, and other prognostic factors into estimates of the risk of death.

Once the value of each prognostic factor's *g* parameter is known, we are able to combine information on tumor size, nodal status, and other prognostic factors with three linked equations to estimate of the risk of death, *L* for each patient:

$$L = L_{primary} + L_{nodes} - \left(L_{primary} * L_{nodes}\right) \tag{2}$$

where

$$L_{primary} = 1 - e^{-(Q * j_{primary})(g_1 * g_2 * g_3 * g_4 * ..) D^Z} \tag{3}$$

and

$$L_{nodes} = 1 - e^{-(M * L_{per-node})} \tag{4}$$

where *M* is the number of local lymph nodes found to be positive for cancer, and $L_{per\text{-}node}$ is the lethal contribution for each positive node. We call this technique the *Size+Nodes+PrognosticFactors* (*SNAP*) method (Table 00, Methods section, above).

The validity of the *SNAP* method's predictions of breast carcinoma lethality:

To test the accuracy of the *SNAP* predictions of breast carcinoma lethality, individuals in the SEER and Partners datasets were sorted into groups of various types and the predicted survival value calculated by the *SNAP* method were compared with the actual 15-year cancer specific Kaplan-Meier death rates for each group. For example, when we used the *SNAP* method to stratify the 362,491 patients from the SEER dataset into groups differing by a 2% risk of death (i.e. those patients expected to have 0-2% risk of death, 2%-4% risk of death, 4%-6%, etc), the expected and observed survival values for each group for the 97% of breast cancer patient with predicted risks of death up to a 48% risk of death agreed within 1% (Figure C4(a)). Even for the 3% of patients with the chance of death predicted by the *SNAP* method greater than 48%, the expected and observed survival values for each group agreed within 7%. In other words, the *SNAP* method proved to be remarkably powerful tool for stratifying patients by the risk of death.

When patients are sorted in other ways, the agreement between the expected and observed survival values also proved to be excellent. This could be seen for patients sorted by nodal status, grade, age, histology, ER/PR status, race, and sex (Figure C4(b)). Indeed, we have sorted patients in a great number of ways, with superb agreement between the expected and observed survival values, and interested readers may find these in the technical report available at http://cancer.lifemath.net/about/techreports/index.php (Technical Report #5).

The values of the parameters used here for *SNAP* calculations are derived from the SEER population. Thus, it was satisfying to find that the *SNAP* calculations made with the parameters from the SEER populations also proved very capable of accurately stratifying patients for the Partners population (Figure C5). Indeed, there was excellent agreement between the expected and observed survival values when the Partners patients were sorted in a great variety of ways, and interested readers are again referred to the technical report available at http://cancer.lifemath.net/about/techreports/index.php (Technical Report #5).

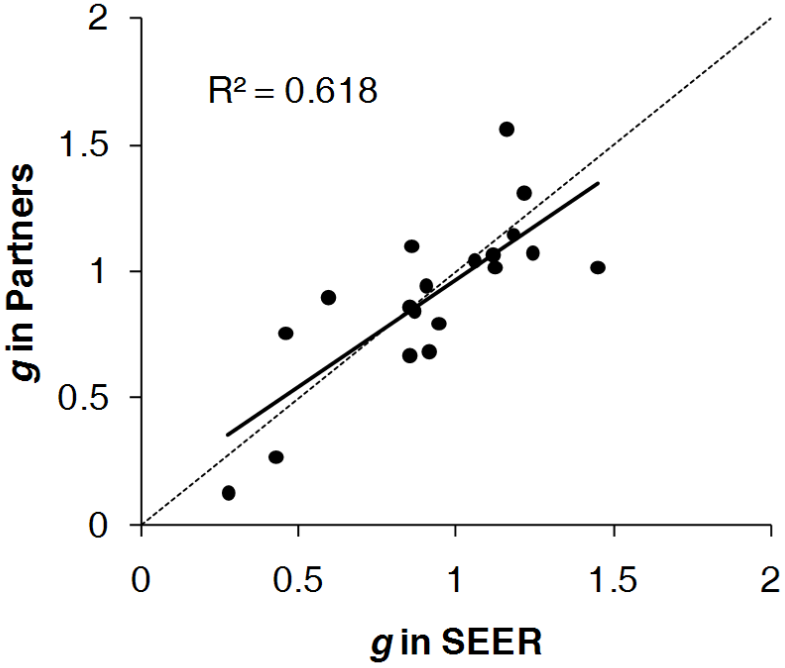

FIGURE C3(a)

FIGURE C3. (a) Scatterplot of *g* parameters derived using the SEER dataset versus *g* parameters derived using the Partners dataset. The inflammatory breast carcinoma group was removed due to being an outlier. Groups that do not achieve significance in SEER were also removed.

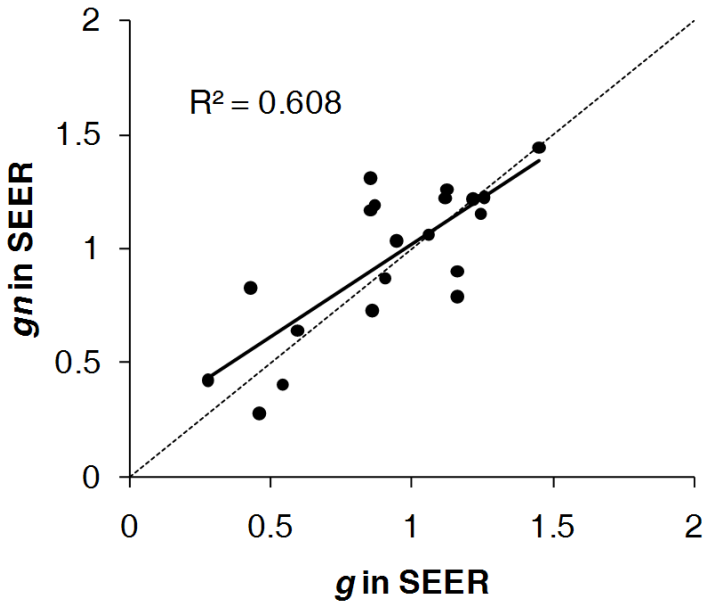

FIGURE C3. (b) Scatterplot of SEER-derived *g* parameters versus g_n parameters. Only groups with both statistically significant *g* and g_n parameters are plotted. In addition, the inflammatory breast carcinoma group was removed due to being an outlier.

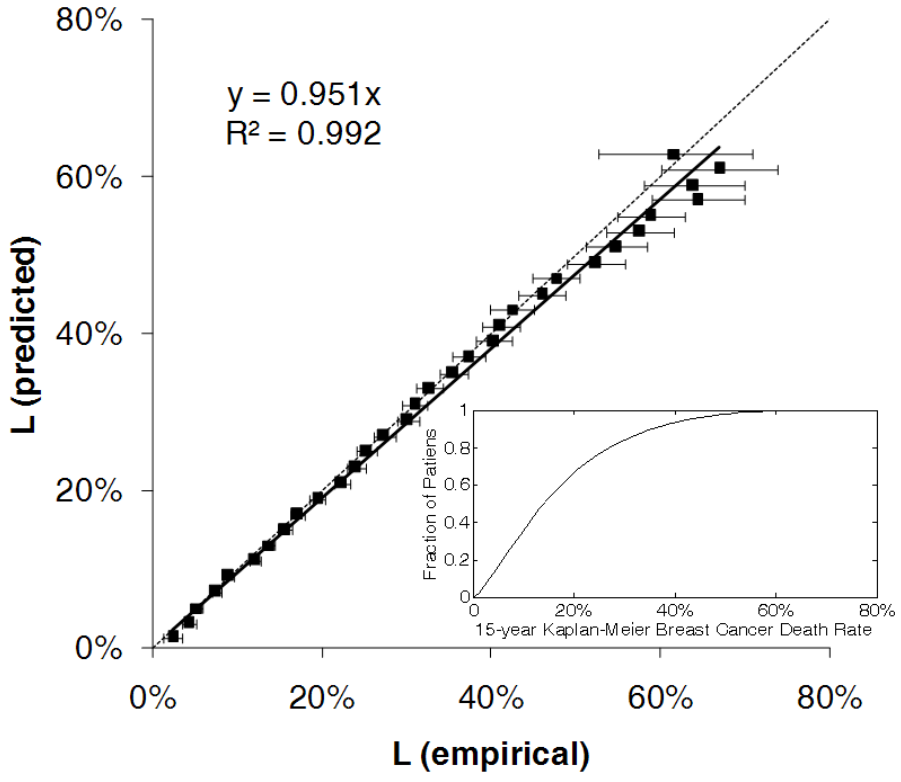

FIGURE C4(a)

FIGURE C4. (a) Stratification of breast carcinoma patients grouped by risk of death, as estimated by the *SNAP* method.

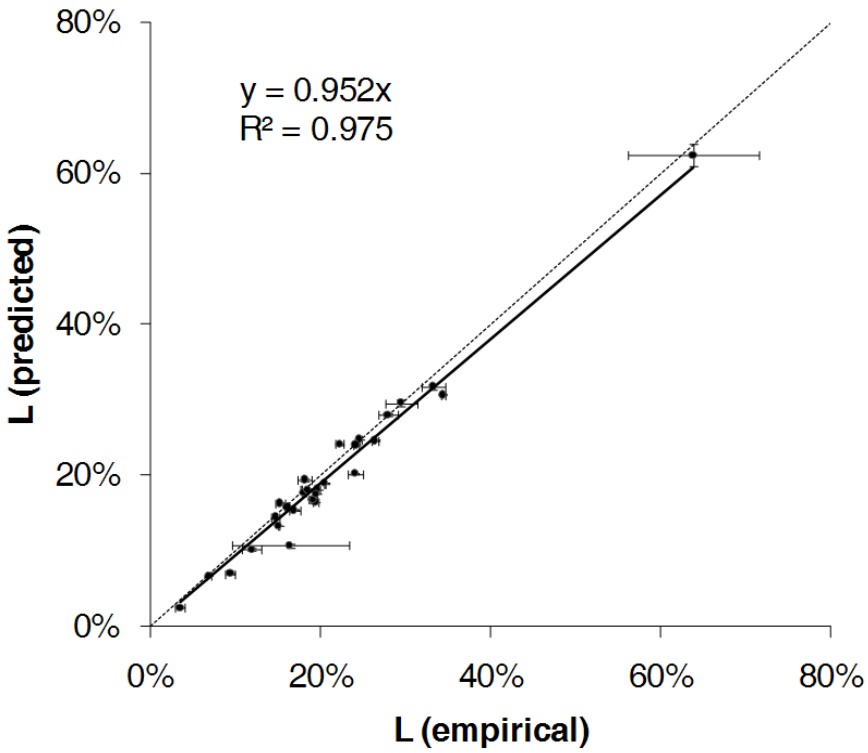

FIGURE C4. (b) $L_{empirical}$ vs. $L_{predicted}$ (*SNAP*) for all the subgroups listed in Table C1 (below) Error bars represent 95% confidence intervals. The y-intercepts of the best-fit lines are set to 0.

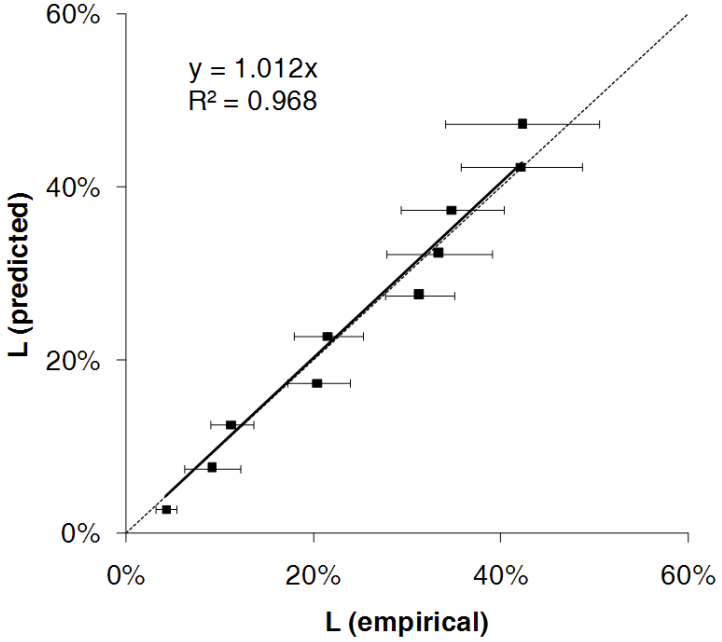

FIGURE C5

FIGURE C5. Patients in the Partners dataset stratified by the *SNAP* method in 5% groups with SEER-derived parameters. Error bars represent 95% confidence intervals. The y-intercepts of the best-fit lines are set to 0. The slope and R^2 values are close to 1, demonstrating the accuracy of the methodology.

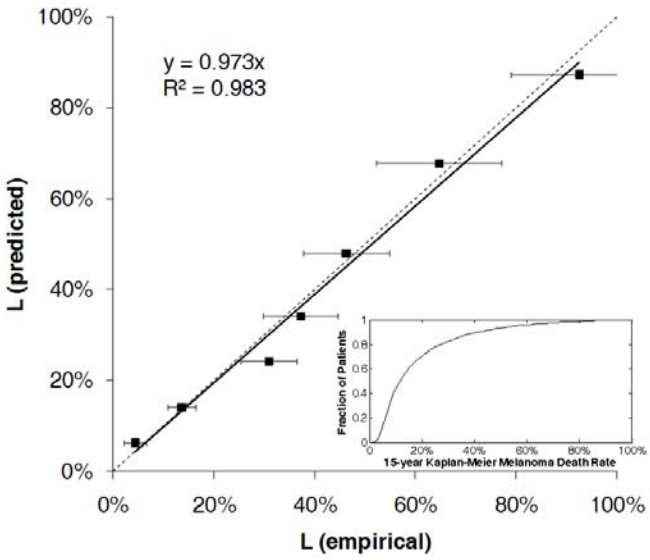

FIGURE C6(a)

FIGURE C6. (a) Stratification of melanoma patients grouped by risk of death, as estimated by the *SNAP* method. Inset depicts the cumulative distribution curve of patients with a certain predicted melanoma death rate.

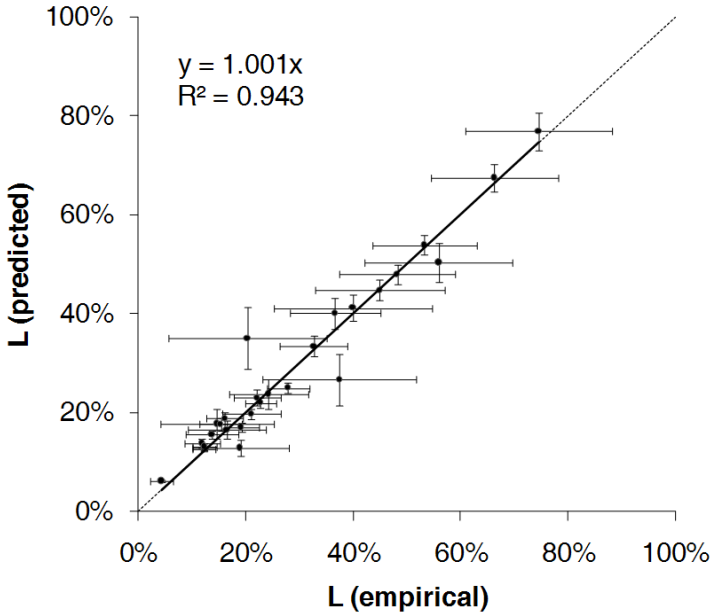

FIGURE C7. (b) $L_{empirical}$ vs. $L_{predicted}$ (*SNAP*) for all the subgroups listed in Table C5. Error bars represent 95% confidence intervals. The y-intercepts of the best-fit lines are set to 0.

Table C1. The lethal impact of various prognostic factors associated with breast carcinoma, using the *SizeAssessment* and *PrognosticMeasurement* methods. Lethality corresponds to the 15-year Kaplan-Meier survival rate. For hormone receptor status, analysis is carried out using a value of Q corresponding to the population for which receptor status is known, which is only after 1990.

Factor (*N*)		Mean Tumor Diameter in mm	$L_{predicted}$ (95% CI)		$L_{empirical}$ (95% CI)		Difference† (pred. – emp.)	*p-value*	*g*††
Nodal Status (362491)	Negative (263544)	18.00	18.28%	(0.02%)	15.09%	(0.12%)	**3.19%**	**<0.0001**	-
	Positive (98947)	23.00	23.49%	(0.03%)	34.42%	(0.29%)	**-10.93%**	**<0.0001**	-
Grade (260666)	1 (51159)	14.01	14.90%	(0.06%)	6.84%	(0.67%)	**8.06%**	**<0.0001**	**0.4324**
	2 (114415)	17.74	18.46%	(0.06%)	16.10%	(0.53%)	**2.35%**	**<0.0001**	**0.8570**
	3 (95092)	21.96	22.32%	(0.06%)	24.61%	(0.53%)	**-2.29%**	**<0.0001**	**1.1224**
Age (361080)	21 – 30 (2840)	22.80	23.10%	(0.33%)	27.91%	(2.18%)	**-4.81%**	**<0.0001**	**1.2545**
	31 – 40 (25272)	21.43	21.84%	(0.12%)	24.17%	(0.76%)	**-2.33%**	**<0.0001**	**1.1267**
	41 – 50 (72296)	19.86	20.39%	(0.06%)	17.99%	(0.45%)	**2.40%**	**<0.0001**	**0.8661**
	51 – 60 (87391)	18.74	19.35%	(0.06%)	19.67%	(0.43%)	-0.32%	0.1486	1.0190
	61 – 70 (85134)	18.11	18.77%	(0.06%)	19.05%	(0.43%)	-0.28%	0.2062	1.0172
	71 – 80 (66843)	18.45	19.09%	(0.06%)	19.42%	(0.61%)	-0.33%	0.2913	1.0201
	81 – 90 (21304)	20.69	21.14%	(0.12%)	24.06%	(1.72%)	**-2.92%**	**0.0009**	**1.1646**
Histology (343186)	Ductal (264692)	18.94	19.55%	(0.04%)	20.51%	(0.25%)	**-0.97%**	**<0.0001**	**1.0573**
	Lobular (25117)	20.46	20.89%	(0.12%)	19.13%	(0.96%)	**1.76%**	**0.0004**	**0.9032**
	Intraductal and LCIS (23449)	18.73	19.35%	(0.12%)	16.90%	(1.25%)	**2.45%**	**0.0001**	**0.8573**
	Mucinous (9374)	18.18	18.84%	(0.18%)	9.39%	(1.10%)	**9.46%**	**<0.0001**	**0.4646**
	Medullary (5675)	23.57	23.87%	(0.22%)	15.21%	(1.12%)	**8.66%**	**<0.0001**	**0.5995**
	Tubular (4992)	10.86	11.84%	(0.18%)	3.46%	(1.08%)	**8.38%**	**<0.0001**	**0.2752**
	Comedo (4184)	20.60	20.96%	(0.31%)	18.48%	(1.41%)	**2.48%**	**0.0008**	**0.8645**
	Scirrhous (1577)	21.80	22.26%	(0.43%)	33.26%	(2.78%)	**-11.01%**	**<0.0001**	**1.6314**
	Inflammatory (147)	28.76	28.12%	(1.67%)	63.85%	(15.17%)	**-35.73%**	**<0.0001**	**3.3130**
	Paget's disease (1266)	21.36	21.67%	(0.55%)	29.51%	(3.59%)	**-7.83%**	**<0.0001**	**1.4535**
	Papillary (1991)	20.18	20.62%	(0.43%)	11.92%	(2.27%)	**8.70%**	**<0.0001**	**0.5414**
	Cribriform (722)	16.16	16.96%	(0.61%)	16.41%	(13.48%)	0.55%	0.9363	0.9636
ER / PR status[a] (230813)	ER+ / PR+ (151742)	17.49	15.89%	(0.04%)	14.68%	(0.84%)	**1.21%**	**0.0048**	**0.9155**
	ER+ / PR- (28880)	18.29	16.51%	(0.10%)	18.51%	(1.71%)	**-2.00%**	**0.0221**	**1.1389**
	ER- / PR+ (5519)	19.42	17.45%	(0.22%)	18.16%	(1.59%)	-0.71%	0.3860	1.0462
	ER- / PR- (44672)	21.52	19.15%	(0.08%)	22.25%	(0.94%)	**-3.10%**	**<0.0001**	**1.1902**
Laterality (362316)	Left (184607)	19.15	19.73%	(0.04%)	19.73%	(0.31%)	0.00%	1	1.0001
	Right (177709)	19.09	19.67%	(0.04%)	19.64%	(0.31%)	0.03%	0.8508	0.9984
Race (337207)	White (310793)	18.89	19.49%	(0.04%)	19.51%	(0.24%)	-0.02%	0.8720	1.0012
	Black (26414)	21.68	22.01%	(0.12%)	26.40%	(0.90%)	**-4.39%**	**<0.0001**	**1.2427**
Sex (362491)	Female (360183)	19.11	19.69%	(0.04%)	19.67%	(0.22%)	0.02%	0.8608	-
	Male (2308)	21.19	21.71%	(0.33%)	25.74%	(3.43%)	**-4.03%**	**0.0219**	**1.2222**

† numbers in bold correspond to prognostic factors that have a significant independent lethal contribution at $p<0.05$, assessed by comparing the difference of the predicted minus the empirical lethalities by an independent, two-sample Student's *t*-test.

†† g parameters that are underlined make a large independent lethal contribution, defined by us to be <0.75 or >1.33.

[a] Q = 0.0101485.

Table C2. Comparison of *g* parameters for breast carcinoma across three datasets (SEER, Partners, and Van Nuys). The p-value indicates the statistical significance level of the difference between the empirical and *SizeOnly*-predicted lethality for a group within each dataset.

Factor		SEER[a]			Partners[b]			Van Nuys[c]		
		N	*g*	p-value	N	*g*	p-value	N	*g*	p-value
Grade	1	51159	**0.4324**	**<0.0001**	1420	**0.2735**	**<0.0001**	636	**0.3205**	**<0.0001**
	2	114415	**0.8570**	**<0.0001**	3425	0.8563	0.0537	524	1.064	0.8259
	3	95092	**1.1224**	**<0.0001**	2946	1.0625	0.3038	350	1.142	0.5507
Age	21 – 30	2840	**1.2545**	**<0.0001**	-	-	-	-	-	-
	31 – 40	25272	**1.1267**	**<0.0001**	1114	1.0196	0.8379	362	**1.3620**	**0.0452**
	41 – 50	72296	**0.8661**	**<0.0001**	2602	**0.8379**	**0.0043**	783	0.8640	0.2225
	51 – 60	87391	1.0190	0.1486	2748	0.9710	0.6345	674	0.9490	0.7416
	61 – 70	85134	1.0172	0.2062	2477	1.0555	0.3835	530	0.7860	0.1591
	71 – 80	66843	1.0201	0.2913	2330	1.1518	0.0564	351	1.5465	0.1856
	81 – 90	21304	**1.1646**	**0.0009**	-	-	-	-	-	-
Histology	Ductal	264692	**1.0573**	**<0.0001**	8864	1.0405	0.2399	2396	1.0455	0.5544
	Lobular	25117	**0.9032**	**0.0004**	795	0.9378	0.6782	317	0.7352	0.1455
	Intraductal and LCIS	23449	**0.8573**	**0.0001**	927	**0.6651**	**0.0400**	-	-	-
	Mucinous	9374	**0.4646**	**<0.0001**	164	0.7615	0.3361	-	-	-
	Medullary	5675	**0.5995**	**<0.0001**	58	0.8930	0.7022	-	-	-
	Tubular	4992	**0.2752**	**<0.0001**	122	**0.1240**	**<0.0001**	-	-	-
	Comedo	4184	**0.8645**	**0.0008**	36	1.1020	0.8159	-	-	-
	Scirrhous	1577	**1.6314**	**<0.0001**	-	-	-	-	-	-
	Inflammatory	147	**3.3130**	**<0.0001**	44	1.3560	0.2804	-	-	-
	Paget's disease	1266	**1.4535**	**<0.0001**	32	1.0205	0.9708	-	-	-
	Papillary	1991	**0.5414**	**<0.0001**	-	-	-	-	-	-
	Cribriform	722	0.9636	0.9363	-	-	-	-	-	-
ER status	ER+	187439	**0.9482**	**<0.0001**	5015	0.7980	0.1923	1449	0.9845	0.8906
	ER-	51221	**1.1808**	**0.0001**	1346	1.1490	0.5197	520	1.3145	0.0503
PR status	PR+	157931	**0.9181**	**<0.0001**	4154	**0.6795**	**<0.0001**	1233	0.9302	0.5034
	PR-	74068	**1.1654**	**<0.0001**	1458	1.5570	0.1653	712	1.2775	0.0610
Laterality	Left	184607	1.0001	1	5825	0.9955	1	-	-	-
	Right	177709	0.9984	0.8508	5523	1.0055	0.9106	-	-	-
Race	White	310793	1.0012	0.8720	9960	**0.9210**	**0.0071**	-	-	-
	Black	26414	**1.2427**	**<0.0001**	553	1.0710	0.5803	-	-	-
Sex	Female	360183	-	0.8608	-	-	-	-	-	-
	Male	2308	**1.2222**	**0.0219**	89	1.3040	0.3684	-	-	-

[a] Q = 0.0118395
[b] Q = 0.015779
[c] Q = 0.01423

Table C3. Nodal g_n parameters for various breast carcinoma subgroups. $Q_{sn} = 0.017446$.

Factor (*N*)		Percent Node Positive (95% CI)		g_n	p-value
Grade (260666)	1 (51159)	17.7%	(0.39%)	**0.8244**	**<0.0001**
	2 (114415)	29.2%	(0.20%)	**1.1723**	**<0.0001**
	3 (95092)	35.9%	(0.39%)	**1.22467**	**<0.0001**
Age (361080)	21 – 30 (2840)	37.0%	(1.76%)	**1.2197**	**<0.0001**
	31 – 40 (25272)	36.1%	(0.59%)	**1.2632**	**<0.0001**
	41 – 50 (72296)	32.5%	(0.39%)	**1.1953**	**<0.0001**
	51 – 60 (87391)	28.6%	(0.39%)	**1.0832**	**<0.0001**
	61 – 70 (85134)	23.6%	(0.20%)	**0.8880**	**<0.0001**
	71 – 80 (66843)	21.9%	(0.39%)	**0.7962**	**<0.0001**
	81 – 90 (21304)	24.0%	(0.59%)	**0.7890**	**<0.0001**
Histology (343186)	Ductal (264692)	28.5%	(0.20%)	**1.0639**	**<0.0001**
	Lobular (25117)	25.9%	(0.59%)	**0.8753**	**<0.0001**
	Intraductal and LCIS (23449)	33.2%	(0.59%)	**1.3076**	**<0.0001**
	Mucinous (9374)	8.2%	(0.59%)	**0.2740**	**<0.0001**
	Medullary (5675)	22.9%	(1.18%)	**0.6468**	**<0.0001**
	Tubular (4992)	7.6%	(0.78%)	**0.4197**	**<0.0001**
	Comedo (4184)	22.2%	(1.18%)	**0.7242**	**<0.0001**
	Scirrhous (1577)	28.4%	(2.16%)	0.9084	0.0512
	Inflammatory (147)	68.7%	(7.45%)	**2.6103**	**<0.0001**
	Paget's disease (1266)	39.4%	(2.74%)	**1.4461**	**<0.0001**
	Papillary (1991)	13.2%	(1.57%)	**0.4108**	**<0.0001**
	Cribriform (722)	17.2%	(2.74%)	**0.6882**	**<0.0001**
ER / PR status[a] (230813)	ER+ / PR+ (151742)	29.6%	(0.20%)	**1.0303**	**<0.0001**
	ER+ / PR- (28880)	30.0%	(0.59%)	1.0050	0.67
	ER- / PR+ (5519)	32.0%	(1.18%)	1.0233	0.3843
	ER- / PR- (44672)	31.6%	(0.39%)	**0.9062**	**<0.0001**
Laterality (362316)	Left (184607)	27.5%	(0.20%)	1.0080	0.1035
	Right (177709)	27.1%	(0.20%)	0.9917	0.1144
Race (337207)	White (310793)	26.6%	(0.20%)	**0.9826**	**<0.0001**
	Black (26414)	33.9%	(0.59%)	**1.1565**	**<0.0001**
Sex (362491)	Female (360183)	-	-	-	-
	Male (2308)	34.9%	(1.96%)	**1.2113**	**<0.0001**

Table C4. Stratification of breast carcinoma patients by risk of death with the *SNP* method.

Predicted Risk Group[†]	N	Cum Percentage	$L_{predicted}$ (95% CI)		$L_{empirical}$ (95% CI)		Difference[†] (pred. – emp.)
0-2%	8435	2.33%	1.28%	(0.02%)	2.41%	(1.14%)	-1.12%
3-4%	22785	8.62%	3.04%	(0.00%)	4.28%	(0.82%)	-1.24%
5-6%	27397	16.19%	4.97%	(0.00%)	5.27%	(0.61%)	-0.29%
7-8%	29112	24.23%	7.02%	(0.00%)	7.56%	(0.61%)	-0.54%
9-10%	27781	31.90%	9.08%	(0.00%)	9.00%	(0.63%)	0.09%
11-12%	25274	38.88%	11.03%	(0.00%)	12.13%	(0.71%)	-1.10%
13-14%	28441	46.74%	13.01%	(0.00%)	13.78%	(0.65%)	-0.77%
15-16%	21946	52.80%	14.96%	(0.00%)	15.66%	(0.80%)	-0.70%
17-18%	19884	58.29%	16.89%	(0.00%)	17.18%	(0.86%)	-0.29%
19-20%	19492	63.68%	18.92%	(0.00%)	19.45%	(0.84%)	-0.53%
21-22%	18656	68.83%	20.93%	(0.00%)	22.37%	(0.88%)	-1.43%
23-24%	14830	72.92%	22.95%	(0.00%)	24.08%	(1.10%)	-1.14%
25-26%	13544	76.66%	24.98%	(0.02%)	25.28%	(1.16%)	-0.30%
27-28%	11202	79.76%	26.95%	(0.02%)	27.42%	(1.29%)	-0.47%
29-30%	11486	82.93%	28.91%	(0.02%)	30.18%	(1.23%)	-1.27%
31-32%	8839	85.37%	30.91%	(0.02%)	30.98%	(1.51%)	-0.07%
33-34%	8486	87.72%	32.99%	(0.02%)	32.76%	(1.55%)	0.24%
35-36%	7048	89.66%	34.93%	(0.02%)	35.55%	(1.72%)	-0.62%
37-38%	6103	91.35%	36.98%	(0.02%)	37.31%	(2.00%)	-0.33%
39-40%	5441	92.85%	38.99%	(0.02%)	40.37%	(2.10%)	-1.38%
41-42%	4688	94.15%	40.93%	(0.02%)	41.12%	(2.27%)	-0.19%
43-44%	3935	95.23%	43.03%	(0.02%)	42.45%	(2.53%)	0.58%
45-46%	3726	96.26%	44.95%	(0.02%)	46.00%	(2.74%)	-1.05%
47-48%	3151	97.13%	47.05%	(0.02%)	47.66%	(2.82%)	-0.62%
49-50%	2625	97.86%	48.95%	(0.02%)	52.38%	(3.35%)	-3.42%
51-52%	2191	98.46%	51.02%	(0.02%)	54.80%	(3.59%)	-3.78%
53-54%	1661	98.92%	52.97%	(0.02%)	57.65%	(3.98%)	-4.68%
55-56%	1456	99.32%	55.00%	(0.02%)	58.83%	(4.04%)	-3.83%
57-58%	919	99.58%	57.02%	(0.04%)	64.39%	(5.43%)	-7.37%
59-60%	781	99.79%	58.84%	(0.04%)	63.89%	(5.94%)	-5.05%
61-62%	492	99.93%	61.02%	(0.06%)	66.97%	(6.90%)	-5.95%
63-64%	261	100.00%	62.90%	(0.08%)	61.73%	(9.02%)	1.17%

[†] 65-66%, 67-68%, and up not included in the calculation of the mean or displayed on graph; 95% CI is greater than 20% or there is insufficient follow-up data for the group

Table C5. Risk estimates made by the *SNP* method to empirical death rates among various groups of breast carcinoma patients.

Factor (*M*)		$L_{predicted}$ (95% CI)		$L_{empirical}$ (95% CI)		Difference[†] (pred. – emp.)
Nodal Status (362491)	Negative (263544)	13.34%	(0.02%)	15.09%	(0.12%)	-1.75%
	Positive (98947)	30.70%	(0.04%)	34.42%	(0.29%)	-3.72%
Grade (260666)	1 (51159)	6.74%	(0.03%)	6.84%	(0.27%)	-0.11%
	2 (114415)	15.77%	(0.03%)	16.10%	(0.27%)	-0.33%
	3 (95092)	24.79%	(0.04%)	24.61%	(0.27%)	0.19%
Age (361080)	21 – 30 (2840)	28.06%	(0.26%)	27.91%	(1.11%)	0.14%
	31 – 40 (25272)	24.07%	(0.09%)	24.17%	(0.39%)	-0.10%
	41 – 50 (72296)	17.76%	(0.05%)	17.99%	(0.23%)	-0.22%
	51 – 60 (87391)	18.18%	(0.04%)	19.67%	(0.22%)	-1.49%
	61 – 70 (85134)	16.77%	(0.04%)	19.05%	(0.22%)	-2.28%
	71 – 80 (66843)	16.50%	(0.05%)	19.42%	(0.31%)	-2.92%
	81 – 90 (21304)	20.23%	(0.09%)	24.06%	(0.88%)	-3.83%
Histology (343186)	Ductal (264692)	18.96%	(0.02%)	20.51%	(0.13%)	-1.56%
	Lobular (25117)	16.84%	(0.07%)	19.13%	(0.49%)	-2.29%
	Intraductal and LCIS (23449)	15.44%	(0.07%)	16.90%	(0.64%)	-1.45%
	Mucinous (9374)	7.00%	(0.06%)	9.39%	(0.56%)	-2.39%
	Medullary (5675)	16.33%	(0.12%)	15.21%	(0.57%)	1.12%
	Tubular (4992)	2.50%	(0.06%)	3.46%	(0.55%)	-0.96%
	Comedo (4184)	18.07%	(0.18%)	18.48%	(0.72%)	-0.41%
	Scirrhous (1577)	31.79%	(0.35%)	33.26%	(1.42%)	-1.48%
	Inflammatory (147)	62.45%	(1.41%)	63.85%	(7.74%)	-1.40%
	Paget's disease (1266)	29.61%	(0.44%)	29.51%	(1.83%)	0.10%
	Papillary (1991)	10.18%	(0.18%)	11.92%	(1.16%)	-1.75%
	Cribriform (722)	10.74%	(0.32%)	16.41%	(6.88%)	-5.67%
ER / PR status[a] (230813)	ER+ / PR+ (151742)	14.49%	(0.03%)	14.68%	(0.43%)	-0.19%
	ER+ / PR- (28880)	18.10%	(0.08%)	18.51%	(0.87%)	-0.41%
	ER- / PR+ (5519)	19.44%	(0.17%)	18.16%	(0.81%)	1.27%
	ER- / PR- (44672)	24.17%	(0.06%)	22.25%	(0.48%)	1.92%
Race (337207)	White (310793)	24.59%	(0.09%)	26.40%	(0.46%)	-1.81%
	Black (26414)	17.55%	(0.02%)	19.51%	(0.12%)	-1.96%

Table C6. The lethal impact of various prognostic factors associated with melanoma, using the *SizeAssessment* and *PrognosticMeasurement* methods. Lethality corresponds to the 15-year Kaplan-Meier survival rate.

Factor (*M*)		Mean Tumor Thickness in mm	$L_{predicted}$ (95% CI)		$L_{empirical}$ (95% CI)		Difference† (pred. – emp.)	*p-value*	*g*
Nodal Status (664)	Negative (487)	2.35	24.77%	(1.06%)	21.02%	(5.42%)	-3.74%	0.1835	-
	Positive (177)	3.69	33.29%	(2.10%)	53.21%	(9.64%)	**19.91%**	**<0.0001**	-
Clark Level (2492)	2 (773)	0.43	6.42%	(0.22%)	4.25%	(2.20%)	-2.16%	0.0546	0.6534
	3 (655)	1.16	14.16%	(0.67%)	11.78%	(3.30%)	-2.38%	0.1662	0.8151
	4 (964)	2.36	24.53%	(0.78%)	27.83%	(3.94%)	3.30%	0.1075	1.1702
	5 (100)	6.00	46.73%	(3.13%)	55.87%	(13.74%)	9.15%	0.2051	1.3316
Site (2747)	Trunk (1017)	1.51	16.36%	(0.81%)	19.12%	(3.21%)	2.76%	0.1024	1.2058
	Face (238)	1.53	16.72%	(1.71%)	16.39%	(7.23%)	-0.33%	0.9748	0.9769
	External Ear (71)	1.56	17.54%	(2.75%)	14.61%	(10.52%)	-2.93%	0.5982	0.8102
	Upper Limb and Shoulder (594)	1.68	17.59%	(1.15%)	15.26%	(3.87%)	-2.33%	0.2582	0.8438
	Lower Limb and Hip (647)	1.72	18.36%	(1.04%)	16.03%	(3.38%)	-2.34%	0.1968	0.8508
	Scalp and Neck (180)	2.28	22.31%	(2.45%)	24.12%	(7.36%)	1.81%	0.6477	1.1056
Histology (2742)	Superficial Spreading (1610)	1.06	12.84%	(0.47%)	12.17%	(2.16%)	-0.67%	0.5525	1.0028
	Lentigo Malignant (221)	1.14	13.08%	(1.56%)	18.92%	(8.99%)	5.85%	0.2103	1.5560
	Malignant (453)	2.22	22.46%	(1.37%)	22.08%	(4.36%)	-0.38%	0.8706	0.9790
	Acral Lentiginous (68)	2.67	25.92%	(3.85%)	37.42%	(14.30%)	11.50%	0.1304	1.6471
	Nodular (351)	3.47	31.79%	(1.60%)	32.63%	(6.28%)	0.84%	0.7995	1.0350
	Desmoplastic (39)	4.05	35.78%	(5.31%)	20.27%	(14.79%)	-15.52%	0.0568	0.4902
Ulceration (1040)	Absent (856)	1.50	17.03%	(0.72%)	13.63%	(4.93%)	-3.40%	0.1812	0.8869
	Present (184)	3.81	34.28%	(2.19%)	36.53%	(8.41%)	2.26%	0.6121	1.2229
Sex (2762)	Female (1299)	1.49	16.33%	(0.70%)	12.27%	(2.25%)	**-4.06%**	**0.0007**	**0.7711**
	Male (1463)	1.81	18.76%	(0.75%)	22.71%	(2.89%)	**3.95%**	**0.0096**	**1.2062**
All Patients (2770)		1.66	17.59%	(0.51%)	17.59%	(1.84%)	-	-	-

† numbers in bold correspond to prognostic factors that have a significant independent lethal contribution at $p<0.05$, assessed by comparing the difference of the predicted minus the empirical lethalities by an independent, two-sample Student's *t*-test.

The validity of the *SNAP* method's predictions of melanoma lethality:

The *SNAP* method also proved capable of combining information on melanoma thickness, nodal status, and other prognostic factors, into estimates of the risk of melanoma death. For example, the *SNAP* method was able to stratify the melanoma patients into groups differing by a 10% risk of death (i.e. those patients expected to have 0-10%, 10-20%, etc. risk of death) (Figure C6(a)). Additionally, when patients are sorted by nodal status, Clark level, the initial site of occurrence, histological sub-type, ulceration, and sex, the agreement between the expected and observed survival values was excellent (Figure C6(b)).

Table C7.
Stratification of melanoma patients by risk of death by the *SNP* method.
Shown lethalities are median values.

Predicted Risk Group	N	*SNP* method $L_{predicted}$	*SNP* method $L_{empirical}$
0%-10%	1215	6.25%	4.30%
10%-19%	734	14.18%	13.62%
20%-29%	332	24.32%	30.75%
30%-39%	194	34.28%	37.14%
40%-59%	183	48.15%	46.26%
60%-79%	77	67.96%	64.58%
80%-99%	29	87.21%	92.39%

Table C8. Comparison of risk estimates made by the *SNP* method to empirical death rates among various groups of melanoma patients.

Factor (*N*)		$L_{empirical}$ (95% CI)		$L_{predicted}$ SNP (95% CI)	
Nodal Status (664)	Negative (487)	21.02%	(5.42%)	19.72%	(0.95%)
	Positive (177)	53.21%	(9.64%)	53.86%	(1.95%)
	1 Positive (92)	39.87%	(14.71%)	41.23%	(2.61%)
	1,2 Positive (127)	44.85%	(12.08%)	44.80%	(2.17%)
	1,2,3 Positive (148)	48.14%	(10.82%)	48.01%	(1.99%)
	2+ Positive (85)	66.24%	(11.92%)	67.54%	(2.81%)
	3+ Positive (50)	74.46%	(13.69%)	76.87%	(3.85%)
Clark Level (2492)	2 (773)	4.25%	(2.20%)	6.24%	(0.25%)
	3 (655)	11.78%	(3.30%)	13.85%	(0.86%)
	4 (964)	27.83%	(3.94%)	25.02%	(1.06%)
	5 (100)	55.87%	(13.74%)	50.40%	(3.96%)
Site (2747)	Trunk (1017)	19.12%	(3.21%)	17.10%	(0.95%)
	Face (238)	16.39%	(7.23%)	16.58%	(1.77%)
	External Ear (71)	14.61%	(10.52%)	17.75%	(3.09%)
	Upper Limb and Shoulder (594)	15.26%	(3.87%)	17.65%	(1.25%)
	Lower Limb and Hip (647)	16.03%	(3.38%)	18.89%	(1.23%)
	Scalp and Neck (180)	24.12%	(7.36%)	23.78%	(2.95%)
Histology (2742)	Superficial Spreading (1610)	12.17%	(2.16%)	12.69%	(0.58%)
	Lentigo Malignant (221)	18.92%	(8.99%)	12.90%	(1.68%)
	Malignant (453)	22.08%	(4.36%)	23.04%	(1.66%)
	Acral Lentiginous (68)	37.42%	(14.30%)	26.71%	(5.17%)
	Nodular (351)	32.63%	(6.28%)	33.45%	(2.06%)
	Desmoplastic (39)	20.27%	(14.79%)	35.05%	(6.21%)
Ulceration (1040)	Absent (856)	13.63%	(4.93%)	15.65%	(0.82%)
	Present (184)	36.53%	(8.41%)	40.15%	(3.11%)
Sex (2762)	Female (1299)	12.27%	(2.25%)	13.15%	(0.69%)
	Male (1463)	22.71%	(2.89%)	22.01%	(0.95%)
	All Patients (2770)	17.59%	(1.84%)	17.84%	(0.62%)

DISCUSSION

The data presented here provide an integrated method for calculating the impact of the size of the primary tumor and the number of positive nodes on the risk of cancer death. For melanoma, renal cell carcinoma, and breast carcinoma, the relationship between tumor size and the risk of cancer death has been found to be well captured by a simple expression, the *SizeOnly* Equation. For melanoma and breast carcinoma, the relationship between tumor size and the presence of cancer in the nodes has also been found to be captured by a variant of the *SizeOnly* Equation, the *NodalSizeOnly* Equation. For node negative melanoma and breast carcinoma, the relationship between tumor size and risk of death is captured by the *PrimarySizeOnly* Equation. For breast carcinoma, each positive node has been found to contribute ~6% risk of death while each millimeter of primary tumor diameter contributes ~1% risk of death. For melanoma, each positive node has been found to contribute ~23% risk of death while each millimeter of primary melanoma thickness contributes ~8% risk of death. This information is captured by a series of linked equations, the *Size+Nodes* method.

The data also provide an integrated explanation for why cancer at the primary site and in the nodes contributes to lethality; either site can provide a source of cancer cells, one or more of which can spread to the periphery, giving rise to lethal distant metastatic disease.

The data also show that there is a characteristic probability, per cell, that an event of spread will occur, and the greater the number of cancer cells at the primary site (as seen in the size of the primary tumor), or the greater the number of cancer cells in the nodes (as seen in the number of positive nodes), the greater is the overall chance that one or more of these cells will undergo such a lethal event of spread. On the other hand, similar masses of cancer appear to make similar lethal contributions, whether present at the primary site or in the nodes. For example, for breast carcinoma, the average size of the cancer in a positive node was found to be ~6 millimeters and the lethal contribution associated with each positive node was ~6%, while the cancer death rate for breast carcinoma patients with primary masses of ~6 millimeter was ~6%. A similar parallel was found for melanoma. Apparently, the probability of an event of lethal spread, per cell, is the same regardless of where the cancer cells are leaving from.

The data also show of a large number of prognostic factors (32), examined in very large datasets of patients with breast carcinoma, that nine were found to make marked independent contributions to breast carcinoma lethality (grade, mucinous, medullary, tubular and scirrhous adenocarcinoma, male sex, inflammatory disease, Paget's disease, nodal status). A similar analysis of smaller populations of patients with melanoma revealed that of a of eight prognostic factors examined, only nodal status was found to an independent contribution to melanoma lethality (although an analysis of a much larger dataset of such patients is now underway, and should reveal with greater precision which prognostic factors truly make an independent contribution to melanoma lethality). The identification of the prognostic factors found to make a contribution to breast carcinoma lethality was made possible by the *SizeOnly* equation, which was used to determine whether a factor truly makes an independent contribution to cancer lethally or is merely associated with tumor size (*SizeAssessment* method), while the magnitude of each factor's lethal contribution was quantified by a parameter, *g*, inserted into the *SizeOnly* equation (*PrognosticMeasurement* method). From such values, and with a series of linked equations (the *Size+Nodes+PrognosticFactors* [*SNAP*] method) combines information on tumor size, nodal status, and other prognostic factors into a single estimate of the risk of death for each patient. The *SNAP* method was found to accurately estimate the risk of death, and finely stratify patients by risk.

The mathematical framework which made it possible to draw these conclusion from the data, the *binary-biological model of cancer metastasis*, whose equations can capture the relationships between primary tumor size, nodal status, prognostic factors, and cancer lethality. The framework is biologically plausible, as it is based upon a mathematical consideration of the most generally accepted mechanism of cancer death, which is by the spread of cancer cells, occurring with definable probabilities of spread peer cell. The framework has also made possible the development of mathematical techniques for quantifying the probability of the spread of cancer cells from clinical data (the *ProbabilityEstimation* Equation), for capturing the relationship between tumor size to the chance of cancer in the nodes (the *NodalSizeOnly* equation), for capturing the relationship between tumor size to the risk of death for node negative melanoma and breast carcinoma patients (the *PrimarySizeOnly* Equation), for calculating the lethal contribution, per positive node, of cancer in the lymph nodes (the *Nodal Lethality Equation*), for teasing out the independent impact of tumor size, nodal status, and prognostic factors on cancer lethality (the

SizeAssessment and *PrognosticMeasurement* methods), and for using these parameters to estimate a patient's risk of death (the *SizeOnly*, *Size+Nodes*, and *Size+Nodes+PrognosticFactors* (*SNAP*) methods). The accuracy of these methods has been confirmed with actual outcome data from large populations of breast carcinoma and melanoma patients.

One of the benefits of building the equations of the *SizeOnly*, *Size+Nodes*, and *Size+Nodes+PrognosticFactors* (*SNAP*) methods from a consideration of the underlying spread of cancer cells is that the parameters in these equations have biological meanings that can be seen by considering the derivation of these expressions. For example, the value of the parameter Q in the *SizeOnly* equation is a measure of the intrinsic propensity of cancer cells for lethal spread. Thus, the fact that the value of Q for melanoma (~0.14) is at least twelve-fold greater than the value of Q for breast carcinoma (~0.012), and at least twenty-fold greater than the value of Q for renal cell carcinoma (~0.007), means that melanoma cells have at least a twelve-fold greater propensity for lethal spread than breast carcinoma cells, and at least twenty-fold greater propensity than renal cell carcinoma cells. Similarly, the fact that value of the equivalent parameter, *Qn*, in the *NodalSizeOnly* equation is also at least five-fold greater for melanoma than for breast carcinoma, indicates that melanoma cells have at least a five-fold greater propensity for non-lethal spread to the nodes than breast carcinoma cells. We see this once again in the values of the parameter $L_{per\text{-}node}$, which captures the lethal contribution for each positive node in the *Size+Nodes* and *Size+Nodes+PrognosticFactors* (*SNAP*) methods, as well as being a measure of the propensity of cancer cells to spread to the lymph nodes; a comparison of the value of $L_{per\text{-}node}$ for melanoma and breast carcinoma reveals about a three-fold greater propensity of melanoma cells for lethal spread from the nodes than for the lethal spread of breast carcinoma cells from the nodes. More subtly, the value of a prognostic factor's g parameter, which sits next to the Q parameter in the *SizeOnly* equation, not only allows us to quantify the independent contribution of the factor to lethality, but also reflects the underlying propensity of the spread of cancer cells in these patients. Thus, as patients with mucinous breast carcinomas have a g parameter of ~0.5, while patients with ductal carcinomas have a g parameter of ~1; it follows that the cancer cells in patients with mucinous carcinoma have about half the chance of spreading to the periphery and causing death as the cancer cells in patients with ductal carcinoma. We see the same phenomenon in g_n parameters in the *NodalSizeOnly* equation, in that patients with mucinous histology have a g_n parameter of ~0.5, while patients with ductal carcinomas have a g_n parameter of ~1. This implies that cancer cells in patients with mucinous histology have about half the chance of spreading to the nodes as cancer cells in patients with ductal histology.

It has often been wondered whether mutation at the time of spread is a requirement for metastasis[20-29], but, following the reasoning outlined previously[5], the values of the probabilities of metastatic spread of breast carcinoma, renal cell carcinoma, and melanoma cells presented here are difficult to reconcile with such genetic change: *First*, the value of the probability of lethal spread for the smallest melanomas (0.1 mm), at ~1 event of spread for every 500 cells, is many orders of magnitude greater than that expected for a genetic change. *Second,* the probability of metastatic spread per cell from the primary site declines as the tumor increases in size. While this decline is consistent with a number of explanations that are mechanical—using this term in the sense in which it is used in physics: "pertaining to the relations of force and matter"—such as the effect of tumor geometry on the escape of cells from the primary mass, it is not what would be expected for genetic events. Indeed, the probability of genetic events over time should be expected either to remain constant (if only a single genetic event is required), or to increase with time (if the accumulation of multiple genetic events is required). *Third,* the occurrence of one event of spread—the spread of a breast cancer cell from the breast to the local lymph nodes—does not appear to increase the probability of a second event of spread—the spread of a breast cancer cell from the local lymph nodes to the periphery. In other words, the occurrence of the initial event of spread does not lead to a cell-heritable change in the tendency of the progeny of that cell to spread. This finding indicates that the presence of cancer in the nodes is not a marker of a genetic change in the tumor, but rather simply a sign that there is more cancer from which cancer cells can emerge.

These data suggest a mechanical model of the spread of cancer cells that is the simplest mechanism one can envisage. The image that comes to mind is of the spread of bricks from the back of a brick truck. Just as each cancer cell in a mass of cancer has certain probability of leaving the primary mass and spreading to the periphery, leading to death, each brick on our truck has certain probability of flying off and killing a pedestrian. The bigger the pile of bricks in the truck, the greater will be the

overall chance of one or more of these bricks flying off the truck, causing death. Curiously, the bigger the pile of bricks, the lower will be the chance that any individual brick will fly; a brick on the bottom of the pile has to "push aside" a lot of other bricks before it could escape. Indeed, one might expect the same decline in the per-brick probability of lethal spread that is correlated with the increase in the size of the brick pile that we saw for the decline in the per-cell probability of lethal spread that is correlated with the increase in the size of the primary cancer mass. Indeed, as we have shown elsewhere, the form of Equation #9, with b=-2/3, which describes the decline in the per-cell probability of lethal spread is correlated with the size of the primary tumor mass is precisely what one would expect for such geometrical considerations.[5] Of course, the analogy would be closer if the bricks increased in number by mitosis, and if the bricks had a chance for non-lethal spread to a local site (a fender?), from which location a second event of lethal spread could occur.

A number of biological models of the development of breast cancer lethality have been proposed that have had widespread impact of thinking about cancer lethality generally, including Halsted's model of contiguous extension[30], Fisher's systemic model[31] and Hellman's spectrum model [32]. The *binary-biological model of cancer metastasis* is based upon the idea that the macroscopic manifestations of cancer can be understood by modeling the underlying microscopic events of spread of cancer cells. This model has allowed us to calculate the values for the probabilities of the spread of cancer cells with which we can assess Halsted's, Fisher's, and Hellman's theories. We have borrowed the inspiration for the *binary-biological model* from statistical mechanics in physics, where the large-scale physical properties of matter are understood as the macroscopic consequences of the underlying microscopic events of molecules. The findings made with this approach - that the relationship between tumor size and the risk of death can be captured with the *SizeOnly* Equation, that the relationship between tumor size and the chance of cancer in the nodes can be captured with the *NodalSizeOnly* Equation, that each positive lymph node contributes a relatively constant amount of extra lethality, as captured by the *Size+Nodes* method, that the events of spread of cancer cells probably do not require mutation at the time of spread - are not assumptions of the *binary-biological model*, but are conclusions that could be extracted from the data with the model. Nonetheless, these findings fit the biological intuitions of Hellman's spectrum model[32] more closely than Halsted's idea that breast cancer is a disease that progresses by contiguous extension[30], or Fisher's suggestion in his systemic model that cancer in the nodes is a sign that the disease has progressed "rather than an instigator of distant disease"[31].

Perhaps the most promising candidate prognostic factors are those detected by gene expression arrays[33,34], and several features of the methods outlined should helpful in analyzing such information. *First*, the analysis of such data has relied upon comparing groups of patients with tumors of the same size. In contrast, the *PrognosticMeasurement* and *SizeAssessment* methods require no such patient matching. In fact, these methods require no control group at all to determine whether patients with a specific phenotype have a higher or lower level of lethality than would be expected for patients in the population as a whole. *Second*, given that the *PrognosticMeasurement* and *SizeAssesment* methods can accurately capture the impact of prognostic factors such as mucinous histology, inflammatory breast disease, and grade 1 breast carcinomas not only in the fraction of patients dying, but also in the fraction of patients with cancer in the nodes (TABLE C3, FIGURE C3(b)), it follows that analysis of gene expression array patterns to detect their impact on the tendency of cancer to spread may be performed using information on nodal status alone. Such nodal status information require essentially no follow-up, is unaffected by censoring, and is available in the great majority of patients. *Third*, the *Size+Nodes+PrognosticFactors (SNAP)* method offers the only technique available that can combine such prognostic factor information together with information on tumor size and nodal status into a single estimate of the risk of death for each patient.

The approach we have taken here, the *binary-biological model of cancer metastasis*, has been to build and test mathematical expressions for capturing the interactions of the macroscopic features of cancer - primary tumor size, nodal status, prognostic factors, and cancer lethality - from a consideration of the underlying microscopic discrete quality of cancer cells. Such an approach is but a specific example of a general project that we have undertaken for understanding the macroscopic features of multi-cellular systems as the aggregate consequences of the many *either/or* events that go on among the discrete components of which we are comprised (http://www.lifemath.net/binbio.html)[14]. Cells are irreducibly discrete, integer, entities. There can be 1, 3, or a million and three cancer cells at a primary site, or in a lymph node, or in the body as a whole, but never 1.3 cells. Thus, when cells move from one location to

another giving rise to new cancer phenotypes, such events of spread must inevitably be discrete, either/or, events. Either a cancer cell has spread from the primary site to the periphery, causing death, or it hasn't. Either a cancer cell has spread from the primary site to a local node, causing cancer in that node, or it hasn't. Either a cancer cell has spread from a node to the periphery, causing death, or it hasn't. This either/or quality of spread of cancer cells from one location to another allowed us to assign probability values for such events of spread,[3,5] and from such a basis, we have been able to derive the equations. Of course, it is not only cells, but all of the microscopic things of which we are made - molecules, atoms, electrons, photons, genes – that have this discrete quality. As we have reported elsewhere, an equivalent binary-biological modeling of the discrete events that underlie all biological processes has shown how multicellular organisms can use this discrete quality to make the normal cellular populations of the body can grow to predictable sizes, at predictable times, and to predictable shapes[14]. Such binary-biological modeling has also provided an explanation for how normal cellular populations become cancerous cellular populations.[14] Thus the empirical verification of the equations of the binary-biological model of cancer metastasis provides an example of the general utility of the binary-biological approach as a way of understanding multicellular systems.

REFERENCES

1. Balch CM, Soong SJ, Gershenwald JE, et al. Prognostic factors analysis of 17,600 melanoma patients: validation of the American Joint Committee on Cancer melanoma staging system. *J Clin Oncol.* 2001;19:3622-3634.

2. White RR, Stanley WE, Johnson JL, Tyler DS, Seigler HF. Long-term survival in 2,505 patients with melanoma with regional lymph node metastasis. *Ann Surg*. 2002;235:879-887.

3. Michaelson JS, Silverstein M, Wyatt J, et al. Predicting the survival of patients with breast carcinoma using tumor size. *Cancer*. 2002;95:713-723.

4. Michaelson JS, Satija S, Kopans D, et al. Gauging the impact of breast carcinoma screening in terms of tumor size and death rate. *Cancer*. 2003;98:2114-2124.

5. Michaelson JS, Cheongsiatmoy JA, Dewey F, et al. Spread of human cancer cells occurs with probabilities indicative of a nongenetic mechanism. *Br J Cancer*. 2005;93:1244-1249.

6. Michaelson JS, Silverstein M, Sgroi D, et al. The effect of tumor size and lymph node status on breast carcinoma lethality. *Cancer*. 2003;98:2133–2143.

7. Carter CL, Allen C, Henson DE. Relation of tumor size, lymph node status, and survival in 24,740 breast cancer cases. *Cancer*. 1989;63:181-187.

8. Elston CW, Ellis IO. Pathological prognostic factors in breast cancer. I. The value of histological grade in breast cancer: experience from a large study with long-term follow-up. *Histopathology*. 1991;19:403-410.

9. Diab SG, Clark GM, Osborne CK, Libby A, Allred DC, Elledge RM. Tumor characteristics and clinical outcome of tubular and mucinous breast carcinomas. *J Clin Oncol*. 1999;17:1442-1448.

10. Wong SL, Chao C, Edwards MJ, et al. Frequency of sentinel lymph node metastases in patients with favorable breast cancer histologic subtypes. *Am J Surg*. 2002;184:492-498.

11. Adami HO, Malker B, Holmberg L, Persson I, Stone B. The relation between survival and age at diagnosis in breast cancer. *N Engl J Med*. 1986;315:559-563.

12. Subramaniam DS, Isaacs C. Utilizing prognostic and predictive factors in breast cancer. *Curr Treat Options Oncol*. 2005;6:147-159.

13. Hermanek P. Prognostic factor research in oncology. *J Clin Epidemiol*. 1999;52:371-374.

14. Michaelson J. The role of molecular discreteness in normal and cancerous growth. *Anticancer Res.* 1999;19:4853-4867.

15. Delahunt B, Kittelson JM, McCredie MR, Reeve AE, Stewart JH, Bilous AM. Prognostic importance of tumor size for localized conventional (clear cell) renal cell carcinoma: assessment of TNM T1 and T2 tumor categories and comparison with other prognostic parameters. *Cancer.* 2002;94:658-664.

16. Hafez KS, Fergany AF, Novick AC. Nephron sparing surgery for localized renal cell carcinoma: impact of tumor size on patient survival, tumor recurrence and TNM staging. *J Urol.* 1999;162:1930-1933.

17. Karrison TG, Ferguson DJ, Meier P. Dormancy of mammary carcinoma after mastectomy. *J Natl Cancer Inst.* 1999;91:80-85.

18. Ries DAG, Melbert D, Krapcho M, et al. SEER Cancer Statistics Review, 1975-2005 , National Cancer Institute. Bethesda, MD, 2008.

19. Temple CF, Huchcroft SA, Hurlbut DJ, Davidson JS. Histologic staging in malignant melanoma: cross-sectional area revisited. *J Surg Oncol*. 1998;69:83-87.

20. Bernards R, Weinberg RA. A progression puzzle. *Nature.* 2002;418:823.

21. Cifone MA, Fidler, IJ. Increasing metastatic potential is associated with increasing genetic instability of clones isolated from murine neoplasms. *Proc Natl Acad Sci USA.* 1981;78:6949-6952.

22. Fidler IJ. The Ernst W. Bertner Memorial Award lecture: the evolution of biological heterogeneity in metastatic neoplasms. *Symp Fundam Cancer Res*. 1983;36:5-26.

23. Couzin J. A clash over genes that foretell metastasis. *Science.* 2003;299:1005.

24. Yokota J. Tumor progression and metastasis. *Carcinogenesis.* 2000;21:497-503.

25. Sobel ME. Metastasis suppressor genes. *J Natl Cancer Inst.* 1990;82:267-276.

26. Welch DR, Steeg PS, Rinker-Schaeffer CW. Molecular biology of breast cancer metastasis. Genetic regulation of human breast carcinoma metastasis. *Breast Cancer Res*. 2000;2:408-416.

27. Bernards R. Cancer: cues for migration. *Nature*. 2003;425:247-248.

28. Van't Veer LJ, Weigelt B. Road map to metastasis. *Nat Med*. 2003;9:999-1000.

29. Yang J, Mani SA, Donaher JL, et al. Twist, a master regulator of morphogenesis, plays an essential role in tumor metastasis. *Cell*. 2004;117:927-939.

30. Halsted WS. The results of radical operations for the cure of carcinoma of the breast. *Ann Surg*. 1907;46:1-19.

31. Fisher B. Laboratory and clinical research in breast cancer--a personal adventure: the David A. Karnofsky memorial lecture. *Cancer Res.* 1980;40:3863-3874.

32. Hellman S. Natural history of small breast cancers. *J Clin Oncol.* 1994;12:2229-2234.

33. West M, Blanchette C, Dressman H, et al. Predicting the clinical status of human breast cancer by using gene expression profiles. *Proc Natl Acad Sci USA*. 2001;98:11462-11467.

34. Van't Veer LJ, Dai H, van de Vijver MJ, et al. Gene expression profiling predicts clinical outcome of breast cancer. *Nature*. 2002;415:530-536.