Original Article

The Impact of Primary Tumor Size, Lymph Node Status, and Other Prognostic Factors on the Risk of Cancer Death

L. Leon Chen, BS[1]; Matthew E. Nolan, BA[1]; Melvin J. Silverstein, MD[2]; Martin C. Mihm, Jr., MD[3,4]; Arthur J. Sober, MD[5,6]; Kenneth K. Tanabe, MD[1,7]; Barbara L. Smith, MD, PhD[1,7]; Jerry Younger, MD[8,9]; James S. Michaelson, PhD[1,3,4]; and Griffin Weber, PhD[1]

BACKGROUND: Although many prognostic factors are associated with differences in cancer lethality, it may not be obvious whether a factor truly makes an independent contribution to lethality or simply is correlated with tumor size. There is currently no method for integrating tumor size, lymph node status, and other prognostic information from a patient into a single risk of death estimate. **METHODS:** The *SizeOnly* equation, which captures the relation between tumor size and risk of death, makes it possible to determine whether a prognostic factor truly makes an independent contribution to cancer lethally or merely is associated with tumor size (*SizeAssessment* method). The magnitude of each factor's lethal contribution can be quantified by a parameter, *g*, inserted into the *SizeOnly* equation (*PrognosticMeasurement* method). A series of linked equations (the *Size+Nodes+PrognosticFactors* [*SNAP*] method) combines information on tumor size, lymph node status, and other prognostic factors from a patient into a single estimate of the risk of death. **RESULTS:** Nine prognostic factors were identified that made marked, independent contributions to breast carcinoma lethality: grade; mucinous, medullary, tubular, and scirrhous adenocarcinoma; male sex; inflammatory disease; Paget disease; and lymph node status. In addition, it was determined that lymph node status made an independent contribution to melanoma lethality. The *SNAP* method was able to accurately estimate the risk of death and to finely stratify patients by risk. **CONCLUSIONS:** The methods described provide a new framework for identifying and quantifying those factors that contribute to cancer lethality and provide a basis for web-based calculators (available at: http://www.CancerMath.net access date) that accurately estimate the risk of death for each patient. **Cancer 2009;000:000–000. © 2009 American Cancer Society.**

KEY WORDS: cancer death, risk, prediction, tumor size, lymph node status, prognostic factors, breast cancer, melanoma.

Corresponding author: James S. Michaelson, PhD, Division of Surgical Oncology, Yawkey Building 7, Massachusetts General Hospital, 32 Fruit Street, Boston, Massachusetts, 02114; Fax: (617) 724 3895; michaelj@helix.mgh.harvard.edu

[1]Department of Surgery, Massachusetts General Hospital, Boston, Massachusetts; [2]Breast Service, Hoag Memorial Hospital Presbyterian, Newport Beach, California; [3]Department of Pathology, Massachusetts General Hospital, Boston, Massachusetts; [4]Department of Pathology, Harvard Medical School, Boston, Massachusetts; [5]Department of Dermatology, Massachusetts General Hospital, Boston, Massachusetts; [6]Department of Dermatology, Harvard Medical School, Boston, Massachusetts; [7]Department of Surgery, Harvard Medical School, Boston, Massachusetts; [8]Department of Medicine, Massachusetts General Hospital, Boston, Massachusetts; [9]Department of Medicine, Harvard Medical School, Boston, Massachusetts

Received: August 12, 2008; **Revised:** February 25, 2009; **Accepted:** March 31, 2009

Published online: Month 00, 2009 © 2009 American Cancer Society

DOI: 10.1002/cncr.24565, www.interscience.wiley.com

Many prognostic factors have been associated with differences in cancer lethality, including tumor size,[1,2] lymph node status[1,3] grade,[4] histology,[5,6] sex, patient age,[7] gene expression array pattern,[8] immunologic marker phenotype, biochemical marker phenotype, and specific mutations.[9] Tumor size long has been directly related to cancer lethality, a relation that we have observed is captured well by a simple expression, the *SizeOnly* equation.[2,3,10,12] For prognostic factors other than size, 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. In this report, we describe a new technique, the *SizeAssessment* method, which uses the *SizeOnly* equation to determine whether a prognostic factor makes an independent contribution to the risk of cancer death or merely is correlated with tumor size. We also 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, lymph node 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.

MATERIALS AND METHODS

Mathematical Methods

The general theory behind the mathematical methods used here, the binary-biologic model of cancer metastasis, is described in the first article in this series.[10]

Data

In addition to the data on tumor size, lymph node status, and survival of the patients with breast carcinoma who were seen at the University of Southern California/Van Nuys Breast Center (Van Nuys) (1352 patients between 1966 and 2006)[2] that also were used in the 2 accompanying articles in this series,[10,12] data were also were available on much larger populations of patients who were seen at the Partners Hospitals (Partners) (11,271 patients who were seen at Massachusetts General Hospital [MGH] and at Brigham and Women's Hospital between 1960 and 2003) and from the Surveillance, Epidemiology, and End Results (SEER) national dataset (362,491 patients between 1973 and 2004 with a first malignant tumor that measured 1-50 mm in greatest dimension and with 0-7 positive lymph nodes).[11]

Data on patients with melanoma for whom complete information was available on tumor thickness and survival were available on 2770 melanoma patients who were seen at MGH from January 1, 1970 to May 1, 2002. Complete information on lymph node status was known for 664 of these patients. Further details are available at www.CancerMath.net (access date). AQ5

RESULTS

The SizeAssessment Method Determines Whether a Prognostic Factor Makes an Independent Contribution to Lethality

We have demonstrated that, for breast cancer and melanoma, the relation between the risk of cancer death, L (defined operationally as the 15-year, Kaplan-Meier, cancer-specific death rate[10]), and tumor size, D, is captured well by a simple expression, the *SizeOnly* equation[2,3,10]:

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

in which Q and Z are constants that capture the metastatic potential and geometry of the tumor.[2,10]

The *SizeOnly* equation makes it possible to determine whether a prognostic factor contributes independently to lethality or simply is correlated with tumor size. In this test, which we call the *SizeAssessment* method (Fig. 1), the actual 15-year cancer-specific Kaplan-Meier death rate for a group of patients with a prognostic factor, $L_{empirical}$, is compared with 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 $L_{predicted}$ and $L_{empirical}$ by an independent, 2-sample Student t test with a threshold of 0.05. F1

J_ID: Z7B Customer A_ID: C2699-08 Cadmus Art: CNCR24565 Date: 13-July-09 Stage: Page: 3

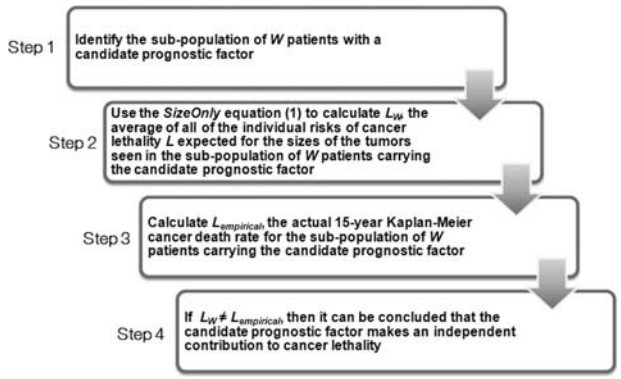

FIGURE 1. The *SizeAssessment* method is used to determine the qualitative impact of prognostic factors suspected of contributing to cancer lethality.

FIGURE 2. The *PrognosticMeasurement* method is used to quantify the impact of prognostic factors that contribute to cancer lethality. For the prognostic factors that increase or decrease the propensity for lethal spread by using the *SizeAssessment* method, a pseudo-Monte Carlo method is used 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 *SizeAssessment* method can be used not only to assess a prognostic factor's impact on the lethal spread of cancer to the periphery but also to assess the impact on the spread of cancer to the lymph nodes, because both manifestations of the spread of cancer cells are captured well by equations in the form of the *SizeOnly* equation. The expression that relates tumor size to the chance of cancer in the lymph nodes is called the *NodalSizeOnly* equation:

$$L_{to\text{-}nodes} = 1 - e^{-Q_{nodes}D^Z} \quad (2)$$

in which $L_{to\text{-}nodes}$ is the fraction of patients identified with cancer in their lymph nodes. Note that this equation captures the spread from the primary site to the lymph nodes, a process that we have considered nonlethal but that, nonetheless, we are able to measure in the same manner that we measure the lethal spread to the periphery (distant metastasis).[12]

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:

$$L_X = 1 - e^{-Q(g_1 * g_2 * g_3 * g_4 * ..)D^Z} \quad (3)$$

The value of the *g* parameter for the lethal contribution of each prognostic factor in the *SizeOnly* equation can be determined by a technique we have called a *pseudo-Monte Carlo* method, as outlined in the accompanying article.[10] We call this technique for quantifying a prognostic factor's impact on lethality the *PrognosticMeasurement* method (Fig. 2). F2

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

Although the value of a *g* parameter is an abstraction, there are 2 practical ways to comprehend the nature of its magnitude. First, because the *g* parameter sits next to the tumor size, *D*, in the *SizeOnly* equation, patients who have tumors with a prognostic factor for which $g = 2$ can be expected to have the same death rate as patients who have tumors without the factor but with tumors of twice the size. Likewise, patients who have tumors with a prognostic factor for which $g = 0.5$ can be expected to have the same death rate as patients who have tumors without the factor but with tumors of half the size. This holds exactly for the case in which $Z = 1$ and is a reasonable approximation for other values of *Z* used by our model (Table 1). T1 Second, because the *SizeOnly* equation is roughly linear over most of its range,[10] patients with a prognostic factor for which $g = 2$ can be expected to have the roughly twice the death rate as patients without the factor, whereas patients with a prognostic factor for which $g = 0.5$ can be expected to have roughly half the death rate as patients without the factor.

Some prognostic factors will be identified that make independent contributions to lethality, but the magnitude may be very small. For example, whereas the lobular and ductal histologies make statistically significant contributions to lethality, as determined by the *SizeAssessment*

ID: jaganm | Black Lining: [ON] | Time: 20:38 | Path: N:/3b2/CNCR/Vol00000/090461/APPFile/C2CNCR090461

Table 1. The *Size+Nodes+PrognosticMarkers* (SNAP) Method for Estimating the Risk of Cancer Death from Information on Tumor Size, Lymph Node Status, and Other Prognostic Factors: $L = L_{primary} + L_{nodes} - (L_{primary}{}^{*}L_{nodes})$†

Source of Lethality	Method of Estimation	Independent Variable	Parameters	Interpretation
The lethal contribution from cancer at the primary site	$L_{primary} = 1 - e^{-(Q^{*}j_{primary})^{*}(g1^{*}g2^{*}g3^{*}g4^{*}\ldots)D^{Z}}$	D = tumor size: For breast carcinoma, greatest dimension (mm); for melanoma, thickness (mm)	For breast carcinoma, $Q = 0.0118395$, $Z = 1$, $j_{primary} = 0.661$ if lymph node status is known and $j_{primary} = 1$ if lymph node status is unknown (see Table 2 for g parameter values); for melanoma, $Q = 0.1428$, $Z = 0.89$, $j_{primary} = 0.801$ if lymph node status is known, and $j_{primary} = 1$ if lymph node status is unknown (see Table 5 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	$L_{nodes} = 1 - e^{-(M^{*}L_{per\text{-}node})}$	M = the no. of positive lymph nodes	For breast carcinoma, $L_{per\text{-}node} = 0.0608$; for melanoma, $L_{per\text{-}node} = 0.2252$	The presence of each positive lymph node contributes approximately "$L_{per\text{-}node}$" extra chance of death

†The *Size+Nodes+PrognosticMarkers* (SNAP) method reduces to: the *Size + Nodes* method, when only size and lymph node status are known and the *SizeOnly* method, when only size is known.

method, the value of the g parameter that was identified by using the *PrognosticMeasurement* method was 0.9032 for lobular histology and 1.057 for ductal histology (Table 2). This means that patients with lobular carcinoma have roughly 90% of the death rate of patients with ductal carcinoma. For this reason, we have chosen the somewhat arbitrary term "marked" to identify those prognostic factors that make an independent contribution to lethality with a P value $>.05$ by using the *SizeAssessment* method and with a g parameter that is either <0.75 or >1.33. This represents an equal negative (factor of 3/4) and positive (factor of 4/3) deviation band around $g = 1$. Here, negative and positive are defined in terms of the direction of deviation from $g = 1$ (toward 0 and infinity, respectively) and should not be confused with qualitative survival outcome. We used lethality rather than survival to be consistent in this regard (ie, a positive impact on lethality represents an increase in lethality).

The Impact of Prognostic Factors on Breast Carcinoma Lethality

By using the *SizeAssessment* method, we identified 8 of 32 investigated factors (plus lymph node status) that made marked, independent contributions to breast cancer lethality (Table 2). The factors examined were grade, age, histology, estrogen receptor/progesterone receptor (ER/PR) status, laterality, race, and sex. The factors that made marked, positive, independent contributions to lethality included Paget disease, scirrhous adenocarcinoma, and inflammatory disease. Factors that made marked, negative, independent contributions to lethality included having a medullary or papillary carcinoma, mucinous or tubular adenocarcinoma, or grade 1 cancer. Similar results were obtained for the patients in the Partners, Van Nuys, and SEER datasets (Table 3), and the correlation between g parameters derived from the Partners and SEER datasets are provided in Figure 3a.

Seven of the 8 identified factors that made a marked, independent contribution to lethality were also made a marked, independent contribution to the propensity of spread to the lymph nodes (Table 4). Only scirrhous adenocarcinoma made a marked, independent contribution to lethality, but not to the propensity, of spread to the lymph nodes. The correlation between g and g_n for various prognostic factors in the SEER dataset is illustrated in Figure 3b.

The Impact of Prognostic Factors on Melanoma Lethality

Of the 5 melanoma prognostic factors listed in Table 5 that we investigated with the *SizeAssessment* method—

Table 2. The Lethal Impact of Various Prognostic Factors Associated With Breast Carcinoma, Using the *SizeAssessment* and *PrognosticMeasurement* Methods*

Factor (No.)	$L_{predicted}$, %	$L_{empirical}$, %	Difference (Pred−Emp), %	*P*	*g*
Lymph node status (362,491)					
Negative (263544)	18.28	15.09	3.19†	<.0001†	—
Positive (98947)	23.49	34.42	−10.93†	<.0001†	—
Grade (260,666)					
1 (51,159)	14.90	6.84	8.06†	<.0001†	0.4324†,‡
2 (114,415)	18.46	16.10	2.35†	<.0001†	0.8570†
3 (95,092)	22.32	24.61	−2.29†	<.0001†	1.1224†
Age (361,080), y					
21-30 (2840)	23.10	27.91	−4.81†	<.0001†	1.2545†
31-40 (25,272)	21.84	24.17	−2.33†	<.0001†	1.1267†
41-50 (72,296)	20.39	17.99	2.40†	<.0001†	0.8661†
51-60 (87,391)	19.35	19.67	−0.32	.1486	1.0190
61-70 (85,134)	18.77	19.05	−0.28	.2062	1.0172
71-80 (66,843)	19.09	19.42	−0.33	.2913	1.0201
81-90 (21,304)	21.14	24.06	−2.92†	.0009†	1.1646†
Histology (343,186)					
Ductal (264,692)	19.55	20.51	−0.97†	<.0001†	1.0573†
Lobular (25,117)	20.89	19.13	1.76†	.0004†	0.9032†
Intraductal and LCIS (23,449)	19.35	16.90	2.45†	.0001†	0.8573†
Mucinous (9374)	18.84	9.39	9.46†	<.0001†	0.4646†,‡
Medullary (5675)	23.87	15.21	8.66†	<.0001†	0.5995†,‡
Tubular (4992)	11.84	3.46	8.38†	<.0001†	0.2752†,‡
Comedo (4184)	20.96	18.48	2.48†	.0008†	0.8645†
Scirrhous (1577)	22.26	33.26	−11.01†	<.0001†	1.6314†,‡
Inflammatory (147)	28.12	63.85	−35.73†	<.0001†	3.3130†,‡
Paget disease (1266)	21.67	29.51	−7.83†	<.0001†	1.4535†,‡
Papillary (1991)	20.62	11.92	8.70†	<.0001†	0.5414†,‡
Cribriform (722)	16.96	16.41	0.55	.9363	0.9636
ER/PR status (230,813)§					
ER+/PR+ (15,1742)	15.89	14.68	1.21†	.0048†	0.9155†
ER+/PR− (28,880)	16.51	18.51	−2.00†	.0221†	1.1389†
ER−/PR+ (5519)	17.45	18.16	−0.71	.3860	1.0462
ER−/PR− (44,672)	19.15	22.25	−3.10†	<.0001†	1.1902†
Laterality (362,316)					
Left (184,607)	19.73	19.73	0.00	1	1.0001
Right (177,709)	19.67	19.64	0.03	.8508	0.9984
Race (337,207)					
White (310,793)	19.49	19.51	−0.02	.8720	1.0012
Black (26,414)	22.01	26.40	−4.39†	<.0001†	1.2427†
Sex (362,491)					
Women (360,183)	19.69	19.67	0.02	.8608	—
Men (2308)	21.71	25.74	−4.03†	.0219†	1.2222†

$L_{empirical}$ indicates the actual 15-year, cancer-specific Kaplan-Meier death rate for a group of patients with a prognostic factor; $L_{predicted}$, the predicted death rate that would be expected; Pred, predicted; Emp, empirical; LCIS, lobular carcinoma in situ; ER, estrogen receptor; PR, progesterone receptor; +, positive; −, negative.

*Lethality corresponds to the 15-year Kaplan-Meier survival rate. For hormone receptor status, the analysis was performed using a value of *Q* corresponding to the population for which receptor status was known, which was available only after 1990.

†These prognostic factors made a significant, independent, lethal contribution at $P < .05$ as assessed by comparing the difference of the predicted lethalities minus the empirical lethalities using an independent, 2-sample Student *t* test.

‡These *g* parameters made a marked, independent, lethal contribution, which was defined as <0.75 or >1.33.

§$Q = 0.0101485$.

Table 3. Comparison of *g* parameters for breast carcinoma across 3 datasets (Surveillance, Epidemiology, and End Results; Partners Hospitals; and Van Nuys Breast Center)

	SEER*			Parnters Hospitals†			Van Nuys Breast Center‡		
Factor	No.	*g*	*P*§	No.	*g*	*P*§	No.	*g*	*P*§
Grade									
1	51,159	0.4324‖	<.0001‖	1420	0.2735‖	<.0001‖	636	0.3205‖	<.0001‖
2	114,415	0.8570‖	<.0001‖	3425	0.8563	.0537	524	1.064	.8259
3	95,092	1.1224‖	<.0001‖	2946	1.0625	.3038	350	1.142	.5507
Age, y									
21-30	2840	1.2545‖	<.0001‖	—	—	—	—	—	—
31-40	25,272	1.1267‖	<.0001‖	1114	1.0196	.8379	362	1.3620‖	.0452‖
41-50	72,296	0.8661‖	<.0001‖	2602	0.8379‖	.0043‖	783	0.8640	.2225
51-60	87,391	1.0190	.1486	2748	0.9710	.6345	674	0.9490	.7416
61-70	85,134	1.0172	.2062	2477	1.0555	.3835	530	0.7860	.1591
71-80	66,843	1.0201	.2913	2330	1.1518	.0564	351	1.5465	.1856
81-90	21,304	1.1646‖	.0009‖	—	—	—	—	—	—
Histology									
Ductal	264,692	1.0573‖	<.0001‖	8864	1.0405	.2399	2396	1.0455	.5544
Lobular	25,117	0.9032‖	.0004‖	795	0.9378	.6782	317	0.7352	.1455
Intraductal and LCIS	23,449	0.8573‖	.0001‖	927	0.6651‖	.0400‖	—	—	—
Mucinous	9374	0.4646‖	<.0001‖	164	0.7615	.3361	—	—	—
Medullary	5675	0.5995‖	<.0001‖	58	0.8930	.7022	—	—	—
Tubular	4992	0.2752‖	<.0001‖	122	0.1240‖	<.0001‖	—	—	—
Comedo	4184	0.8645‖	.0008‖	36	1.1020	.8159	—	—	—
Scirrhous	1577	1.6314‖	<.0001‖	—	—	—	—	—	—
Inflammatory	147	3.3130‖	<.0001‖	44	1.3560	.2804	—	—	—
Paget disease	1266	1.4535‖	<.0001‖	32	1.0205	.9708	—	—	—
Papillary	1991	0.5414‖	<.0001‖	—	—	—	—	—	—
Cribriform	722	0.9636	.9363	—	—	—	—	—	—
ER status									
ER+	187,439	0.9482‖	<.0001‖	5015	0.7980	.1923	1449	0.9845	.8906
ER−	51,221	1.1808‖	.0001‖	1346	1.1490	.5197	520	1.3145	.0503
PR status									
PR+	157,931	0.9181‖	<.0001‖	4154	0.6795‖	<.0001‖	1233	0.9302	.5034
PR−	74,068	1.1654‖	<.0001‖	1458	1.5570	.1653	712	1.2775	.0610
Laterality									
Left	184,607	1.0001	1	5825	0.9955	1	—	—	—
Right	177,709	0.9984	.8508	5523	1.0055	.9106	—	—	—
Race									
White	310,793	1.0012	.8720	9960	0.9210‖	.0071‖	—	—	—
Black	26414	1.2427‖	<.0001‖	553	1.0710	.5803	—	—	—
Sex									
Women	360,183	—	.8608	—	—	—	—	—	—
Men	2308	1.2222‖	.0219‖	89	1.3040	.3684	—	—	—

SEER indicates Surveillance, Epidemiology, and End Results; LCIS, lobular carcinoma in situ; ER, estrogen receptor; +, positive; −, negative; PR, progesterone receptor.

*$Q = 0.0118395$.

†$Q = 0.015779$.

‡$Q = 0.01423$.

§*P* values indicate the statistical significance level of the difference between the empirical and *SizeOnly*-predicted lethality for a group within each dataset.

‖These prognostic factors made a significant, independent, lethal contribution at $P < .05$ as assessed by comparing the difference of the predicted lethalities minus the empirical lethalities using an independent, 2-sample Student *t* test.

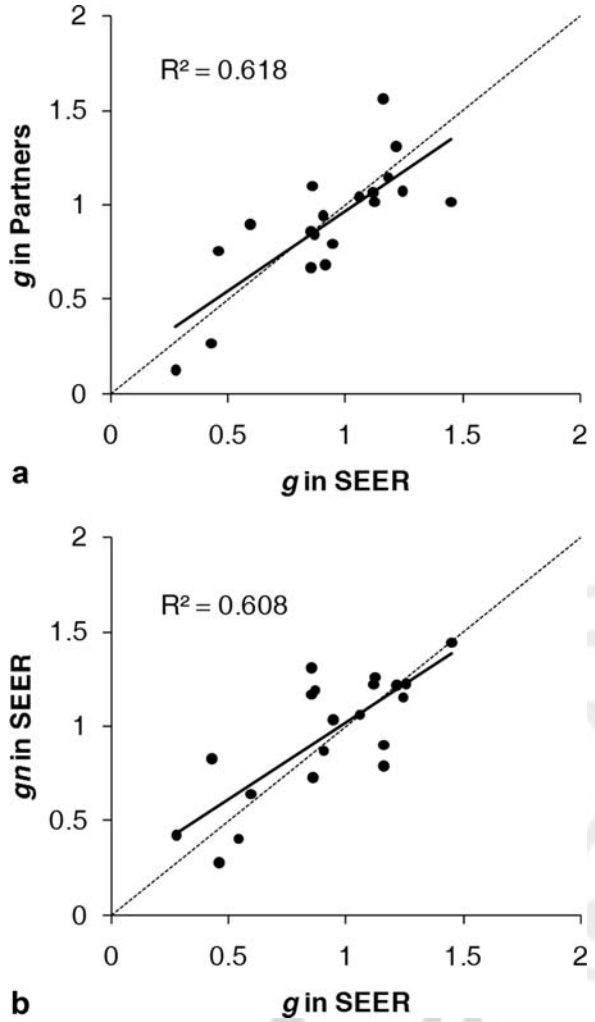

FIGURE 3. (a) This scatterplot of *g* parameters was derived by using the Surveillance, Epidemiology, and End Results (SEER) dataset versus *g* parameters that were derived by using the Partners Hospitals dataset. The inflammatory breast carcinoma group was removed because it was an outlier. Groups that do not achieve significance in SEER also were removed. (b) This is a scatterplot of SEER-derived *g* parameters versus g_n parameters. Only groups with both statistically significant *g* and g_n parameters were plotted. In addition, the inflammatory breast carcinoma group was removed because it was an outlier.

Table 4. Lymph Node g_n Parameters for Various Breast Carcinoma Subgroups*

Factor (No.)	Lymph Node Positive, %	g_n	*P*
Grade (260,666)			
1 (51,159)	17.7	0.8244†	<.0001†
2 (114,415)	29.2	1.1723†	<.0001†
3 (95,092)	35.9	1.22467†	<.0001†
Age (361080), y			
21-30 (2840)	37	1.2197†	<.0001†
31-40 (25,272)	36.1	1.2632†	<.0001†
41-50 (72,296)	32.5	1.1953†	<0001†
51-60 (87,391)	28.6	1.0832†	<.0001†
61-70 (85,134)	23.6	0.8880†	<.0001†
71-80 (66,843)	21.9	0.7962†	<.0001†
81-90 (21,304)	24	0.7890†	<.0001†
Histology (343,186)			
Ductal (264,692)	28.5	1.0639†	<.0001†
Lobular (25,117)	25.9	0.8753†	<.0001†
Intraductal and LCIS (23,449)	33.2	1.3076†	<.0001†
Mucinous (9374)	8.2	0.2740†	<.0001†
Medullary (5675)	22.9	0.6468v	<.0001†
Tubular (4992)	7.6	0.4197†	<.0001†
Comedo (4184)	22.2	0.7242†	<.0001†
Scirrhous (1577)	28.4	0.9084	.0512
Inflammatory (147)	68.7	2.6103†	<.0001†
Paget disease (1266)	39.4	1.4461†	<.0001†
Papillary (1991)	13.2	0.4108†	<.0001†
Cribriform (722)	17.2	0.6882†	<.0001†
ER/PR status (230,813)‡			
ER+/PR+ (15,1742)	29.6	1.0303†	<.0001†
ER+/PR− (28,880)	30	1.0050	.67
ER−/PR+ (5519)	32	1.0233	.3843
ER−/PR− (44,672)	31.6	0.9062†	<.0001†
Laterality (362,316)			
Left (184,607)	27.5	1.0080	.1035
Right (177,709)	27.1	0.9917	.1144
Race (337,207)			
White (310,793)	26.6	0.9826†	<.0001†
Black (26,414)	33.9	1.1565†	<.0001†
Sex (362,491)			
Women (360,183)	—	—	—
Men (2308)	34.9	1.2113†	<.0001†

LCIS, lobular carcinoma in situ; ER, estrogen receptor; PR, progesterone receptor; +, positive; −, negative.

* $Q_n = 0.017446$.

† These prognostic factors had a significant, independent, lethal contribution at $P < .05$ as assessed by comparing the difference of the predicted lethalities minus the empirical lethalities using an independent, 2-sample Student *t* test.

‡ $Q = 0.0101485$. AQ10

Clark level, the initial site of occurrence, histologic subtype, ulceration, and sex—only sex made statistically significant, independent contribution to breast cancer lethality. The *g* parameter was 0.7711 for women and 1.2062 for men, indicating a relatively small contribution to lethality. It is conceivable that some of the other prognostic factors in reality do make a marked, independent contribution to lethality (eg, desmoplastic histology), but

Table 5. The Lethal Impact of Various Prognostic Factors Associated With Melanoma Using the *SizeAssessment* and *PrognosticMeasurement* Methods*

Factor (No.)	$L_{predicted}$, %	$L_{empirical}$, %	Difference (Pred−Emp), %	*P*	*g*
Lymph node status (664)					
Negative (487)	24.77	21.02	−3.74	.1835	—
Positive (177)	33.29	53.21	19.91†	<.0001†	—
Clark level (2492)					
2 (773)	6.42	4.25	−2.16	.0546	0.6534
3 (655)	14.16	11.78	−2.38	.1662	0.8151
4 (964)	24.53	27.83	3.30	.1075	1.1702
5 (100)	46.73	55.87	9.15	.2051	1.3316
Site (2747)					
Trunk (1017)	16.36	19.12	2.76	1024	1.2058
Face (238)	16.72	16.39	−0.33	.9748	0.9769
External ear (71)	17.54	14.61	−2.93	.5982	0.8102
Upper limb and shoulder (594)	17.59	15.26	−2.33	.2582	0.8438
Lower limb and hip (647)	18.36	16.03	−2.34	.1968	0.8508
Scalp and neck (180)	22.31	24.12	1.81	.6477	1.1056
Histology (2742)					
Superficial spreading (1610)	12.84	12.17	−0.67	.5525	1.0028
Lentigo malignant (221)	13.08	18.92	5.85	.2103	1.5560
Malignant (453)	22.46	22.08	−0.38	.8706	0.9790
Acral lentiginous (68)	25.92	37.42	11.50	.1304	1.6471
Nodular (351)	31.79	32.63	0.84	.7995	1.0350
Desmoplastic (39)	35.78	20.27	−15.52	.0568	0.4902
Ulceration (1040)					
Absent (856)	17.03	13.63	−3.40	.1812	0.8869
Present (184)	34.28	36.53	2.26	.6121	1.2229
Sex (2762)					
Women (1299)	16.33	12.27	−4.06†	.0007†	0.7711†
Men (1463)	18.76	22.71	3.95†	.0096†	1.2062†
All patients (2770)	17.59	17.59	—	—	—

$L_{empirical}$ indicates the actual 15-year, cancer-specific Kaplan-Meier death rate for a group of patients with a prognostic factor; $L_{predicted}$, the predicted death rate that would be expected; Pred, predicted; Emp, empirical.

* Lethality corresponds to the 15-year Kaplan-Meier survival rate.

† These prognostic factors made a significant, independent, lethal contribution at $P < .05$ as assessed by comparing the difference of the predicted lethalities minus the empirical lethalities using an independent, 2-sample Student *t* test.

their significance could not be captured by a dataset of this sample size.

The Size+Nodes+PrognosticFactors *Method Combines Tumor Size, Lymph Node 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, lymph node status, and other prognostic factors with 3 linked equations to estimate of the risk of death, *L*, for each patient:

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

in which

$$L_{primary} = 1 - e^{-(Q * j_{primary})(g_1 * g_2 * g_3 * g_4 * ..)D^Z} \quad (5)$$

and

$$L_{nodes} = 1 - e^{-(M * L_{per-node})} \quad (6)$$

in which M is the number of local lymph nodes identified as positive for cancer, and $L_{per\text{-}node}$ is the lethal contribution for each positive lymph node.[3] We call this technique the *Size+Nodes+PrognosticFactors* (*SNAP*)[10] method (Table 1).

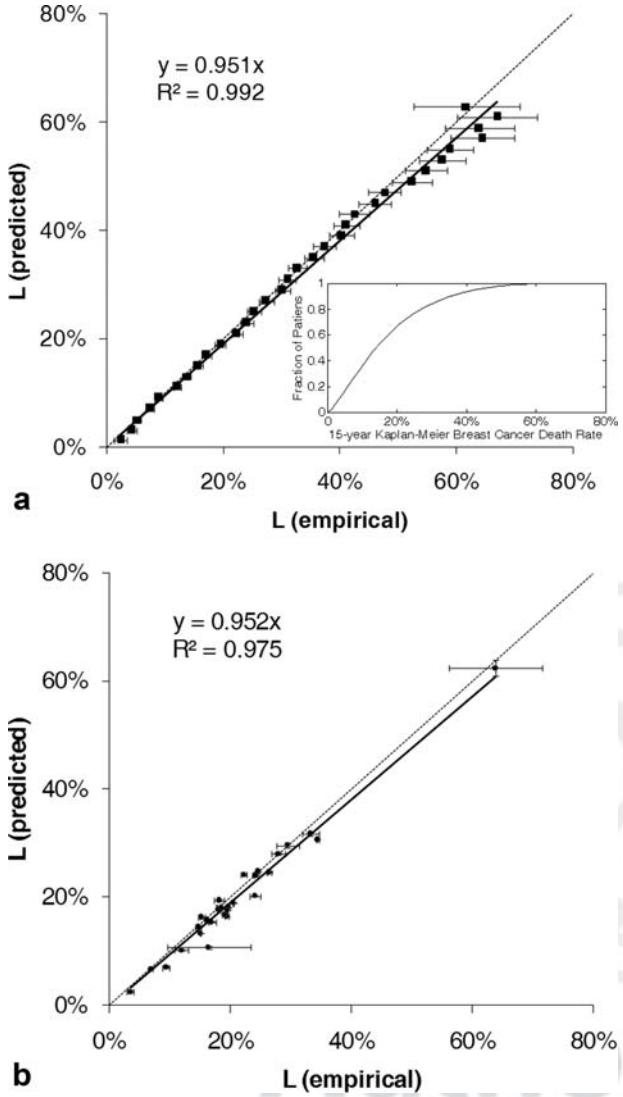

FIGURE 4. (a) Patients with breast carcinoma were stratified according to their risk of death, as estimated by using the *Size+Nodes+PrognosticMarkers* (*SNAP*) method. (b) This chart illustrates the actual 15-year, cancer-specific Kaplan-Meier death rates for groups patients that had a prognostic factor ($L_{empirical}$) versus the predicted death rate ($L_{predicted}$) (*SNAP*) for all subgroups listed in Table 2. Error bars represent 95% confidence intervals. The y-intercepts of the best-fit lines are set to 0.

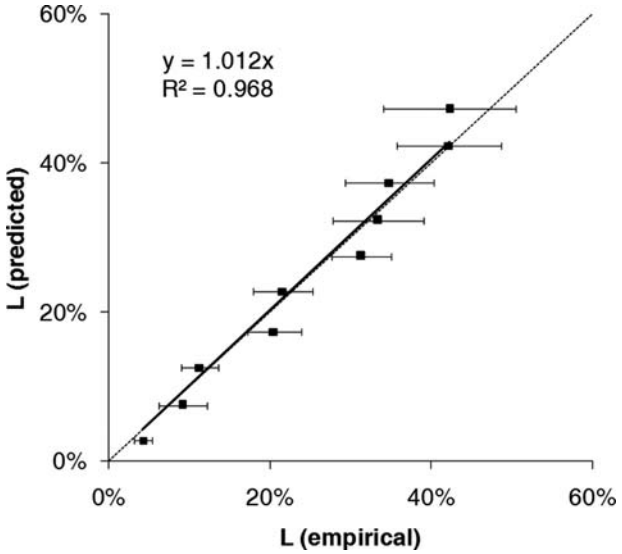

FIGURE 5. Patients in the Partners Hospitals dataset were stratified by using the *Size+Nodes+PrognosticMarkers* (*SNAP*) method in 5% of groups with Surveillance, Epidemiology, and End Results (SEER)-derived parameters. Error bars represent 95% confidence intervals. The y-intercepts of the best-fit lines are set to 0.

The Validity of the SNAP *Method 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 values that we calculated with 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 that differed by a 2% risk of death (ie, patients expected to have a risk of death between 1% and 2%, between 3% and 4%, between 5% and 6%, etc), the expected and observed survival values for each group up to 48% (which comprises 97% of patients) agreed within 1% (Fig. 4a). Even for the 3% of patients who had a chance of death >48% predicted by the *SNAP* method, the expected and observed survival values for each group agreed within 7%. In other words, the *SNAP* method proved to be a powerful tool for stratifying patients according to their risk of death. F4

Patients also can be sorted by lymph node status, grade, age, histology, ER/PR status, race, and sex (Fig. 4b). Interested readers may find the tables and graphs for 38 different sorting schemes, including various permutations of the above groups, in Technical Report 5 (available at: http://cancer.lifemath.net/about/techreports/index.php; access date). AQ6

The values of the parameters used here for *SNAP* calculations were derived from the SEER population but also proved capable of accurately stratifying patients for the Partners population (Fig. 5). There was excellent agreement between the expected and observed survival F5

J_ID: Z7B Customer A_ID: C2699-08 Cadmus Art: CNCR24565 Date: 13-July-09 Stage: Page: 10

values when the Partners patients were sorted in a variety of ways; and, again, interested readers again referred to the technical report (Technical Report 5).

For Figures 4 and 5, the y-intercept of the best-fit line was set at 0 to quantify the extent of over-prediction or under-prediction. A slope of >1 represents over-prediction, and a slope of <1 represents under-prediction. Here, the slope is the more relevant metric for accuracy assessment rather than the R^2 value, which only captures the extent of intergroup variability.

The Validity of the SNAP *Method Predictions of Melanoma Lethality*

The *SNAP* method also proved capable of combining information on melanoma thickness, lymph node status, and other prognostic factors into estimates of the risk of melanoma death. For example, the *SNAP* method was able to stratify patients with melanoma into groups that differed by a 10% risk of death (ie, those patients expected to have a risk of death between 0% and 10%, between F6 10% and 20%, etc) (Fig. 6a). In addition, when patients were sorted by lymph node status, Clark level, the initial site of occurrence, histologic subtype, ulceration, and sex, the agreement between the expected and observed survival values was excellent (Fig. 6b).

Web-Based Calculators Based on the SNAP *Method*

By using the *SNAP* method, we developed a set of web-based calculators for breast cancer and melanoma (avail-AQ7 able at: http://www.CancerMath.net; access date) that can be used by medical professionals to estimate, for each patient, the mortality risk from cancer and the impact that adjuvant treatment will have on that risk of death. The full details of the calculators will not be presented here. In brief, they accept patient data as input (age, tumor size, number of positive lymph nodes, ER, PR, human epidermal human epidermal growth factor 2, grade, histology), and they output several useful measures of patient outcome (15-year, Kaplan-Meier, cancer-specific death rate, cancer/noncancer/overall risk of death for each of the next 15-years, life expectancy with and without adjuvant therapy, years of life saved by adjuvant therapy). In addition, a conditional outcome calculator calculates mortality risk conditional upon the number of years the patient has been alive after diagnosis.

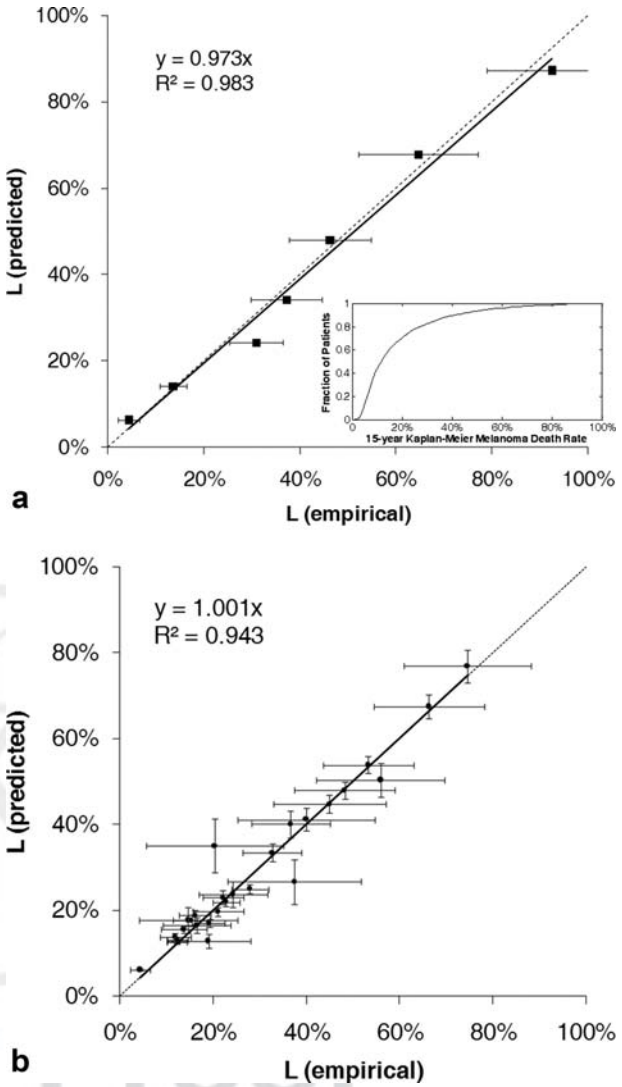

FIGURE 6. (a) Patients with melanoma were stratified according to their risk of death, as estimated by using the *Size+Nodes+PrognosticMarkers* (*SNAP*) method, is illustrated. The *inset* depicts the cumulative distribution curve of patients with a certain predicted melanoma death rate. (b) This chart illustrates the actual 15-year, cancer-specific Kaplan-Meier death rates for groups of patients that had a prognostic factor ($L_{empirical}$) versus the predicted death rate ($L_{predicted}$) (*SNAP*) for all subgroups listed in Table 5. Error bars represent 95% confidence intervals. The y-intercepts of the best-fit lines are set to 0.

DISCUSSION

In this and 2 accompanying articles in this series[10,12] we have outlined a mathematical framework, the binary-

ID: jaganm | Black Lining: [ON] | Time: 20:38 | Path: N:/3b2/CNCR/Vol00000/090461/APPFile/C2CNCR090461

J_ID: Z7B Customer A_ID: C2699-08 Cadmus Art: CNCR24565 Date: 13-July-09 Stage: Page: 11

biologic model of cancer metastasis, with equations that can capture the relations between primary tumor size, lymph node status, prognostic factors, and cancer lethality. The framework is biologically plausible, because it is based on 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 per cell. The framework also has 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 relation between tumor size and the chance of cancer in the lymph nodes (the *NodalSizeOnly* equation); for capturing the relation between tumor size and the risk of death for lymph node-negative patients (the *PrimarySizeOnly* equation); for calculating the lethal contribution, per positive lymph node, of cancer in the lymph nodes (the *NodalLethality* equation); for teasing out the independent impact of 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 *SNAP* methods).

One of the benefits of building the equations of the *SizeOnly*, *Size+Nodes*, and *SNAP* methods from a consideration of the underlying spread of cancer cells is that the parameters in these equations have biologic meaning that can be observed through 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.[10] Thus, the finding that the value of Q for melanoma ($\approx$0.14) is at least 12-fold greater than the value of Q for breast carcinoma ($\approx$0.012) means that melanoma cells have at least a 12-fold greater propensity for lethal spread than breast carcinoma cells. Similarly, the finding that the value of the equivalent parameter, *Qn*, in the *NodalSizeOnly* equation (Equation 2) also is at least 5-fold greater for melanoma than for breast carcinoma indicates that melanoma cells have at least a 5-fold greater propensity for nonlethal spread to the lymph nodes than breast carcinoma cells. We observe this once again in the values of the parameter $L_{per\text{-}node}$, which captures the lethal contribution for each positive lymph node in the *Size+Nodes*[3] and *SNAP*[10] methods and also is a measure of the propensity of cancer cells to spread to the lymph nodes[10]; a comparison of the value of $L_{per\text{-}node}$ for melanoma and breast carcinoma reveals an approximately 3-fold greater propensity of melanoma cells for lethal spread from the lymph nodes than for the lethal spread of breast carcinoma cells from the lymph 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, patients with mucinous breast carcinomas have a g parameter of $\approx$0.5, whereas patients with ductal carcinomas have a g parameter of $\approx$1; it follows that the cancer cells in patients with mucinous carcinoma have approximately half the chance of spreading to the periphery and causing death as the cancer cells in patients with ductal carcinoma. We observe the same phenomenon in g_n parameters in the *NodalSizeOnly* equation, in that patients with mucinous histology have a g_n parameter of $\approx$0.5, whereas patients with ductal carcinomas have a g_n parameter of $\approx$1. This implies that cancer cells in patients with mucinous histology have approximately half the chance of spreading to the lymph nodes as cancer cells in patients with ductal histology.

Perhaps the most promising candidate prognostic factors are those detected by gene expression arrays,[13,14] and several features of the methods outlined should be helpful in analyzing such information. First, the analysis of such data has relied on 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, because these methods can accurately capture the impact of prognostic factors not only on death but also on lymph node positivity (Table 4) (Fig. 3b), it follows that an analysis of gene expression arrays to detect their impact on the spread of cancer may be performed using information on lymph node status alone. Lymph node status information require essentially no follow-up (because it generally is ascertained shortly after diagnosis), is unaffected by censoring (in contrast to death, in which patients who die of other causes or remain alive must be censored), and is available in the majority of patients.

ID: jaganm | Black Lining: [ON] | Time: 20:39 | Path: N:/3b2/CNCR/Vol00000/090461/APPFile/C2CNCR090461

Third, to our knowledge, the *SNAP* method offers the only technique available that can combine such prognostic factor information together with information on tumor size and lymph node status into a single estimate of the risk of death for each patient.

Our modeling approach holds certain advantages over more traditional statistical methods. One shortcoming of empirical models is that they are only as precise as the data from which they are extracted—larger numbers of variables lead to smaller numbers of patients for each subgroup permutation. An undesired consequence of this approach is the considerable uncertainty that inevitably results. These confidence intervals become rather large as more than 1 or 2 factors are specified, especially if 15-year survival outcomes are the focus. By separating out 2 continuous variables (tumor size and the number of positive lymph nodes) through the application of a mechanistic model, we are able to achieve better stratification ability with greater accuracy. Figure 4a illustrates this ability of our method. Indirectly or not, more accurate predictions lead to better and more informed clinical decisions. Clinical tools such as Adjuvant!Online[15] exist and we have built similar web-based calculators based on our mathematical methods (available at: www.CancerMath.net). Because such clinical decision-support tools are used regularly, we believe that improved accuracy and stratification ability will be greatly beneficial.

The approach we have taken here has been to build and test mathematical expressions for capturing the interactions of the macroscopic features of cancer—primary tumor size, lymph node status, prognostic factors, and cancer lethality—from a consideration of the underlying, microscopic, discrete quality of cancer cells. We have called this approach the binary-biologic model of cancer metastasis. Such an approach is only a specific example of a general project that we have undertaken for understanding the macroscopic features of multicellular systems as the aggregate consequences of the many *either/or* events that go on among the discrete components of which we are comprised (available at: http://www.lifemath.net/binbio.html).[16] Cells are irreducibly discrete, integer entities. There can be 1, 3, or 1,000,003 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 1 location to another, giving rise to new cancer phenotypes, such events of spread inevitably must be discrete, either/or events. Either a cancer cell has spread from the primary site to the periphery causing death, or it has not. Either a cancer cell has spread from the primary site to a local lymph node, causing cancer in that lymph node, or it has not. Either a cancer cell has spread from a lymph node to the periphery, causing death, or it has not. This either/or quality of the spread of cancer cells from 1 location to another allowed us to assign probability values for such events of spread[2,17]; and from such a basis, we have been able to derive the equations used herein.[10,12] 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. An equivalent binary-biologic modeling of the discrete events that underlie all biologic processes has illustrated how multicellular organisms can use this discrete quality to make the normal cellular populations of the body grow to predictable sizes, at predictable times, and to predictable shapes, as we have reported elsewhere.[15] Such binary-biologic modeling also has provided an explanation for how normal cellular populations become cancerous cellular populations.[15] Thus the empirical verification of the equations of the binary-biologic model of cancer metastasis provides an example of the general utility of the binary-biologic approach as a way of understanding multicellular systems.

Conflict of Interest Disclosures

Supported by institutional funding to the Massachusetts General Hospital.

References

1. 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.
2. Michaelson JS, Silverstein M, Wyatt J, et al. Predicting the survival of patients with breast carcinoma using tumor size. *Cancer*. 2002;95:713-723.
3. 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.
4. 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.
5. 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.

6. 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.
7. 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.
8. Subramaniam DS, Isaacs C. Utilizing prognostic and predictive factors in breast cancer. *Curr Treat Options Oncol*. 2005;6:147-159.
9. Hermanek P. Prognostic factor research in oncology. *J Clin Epidemiol*. 1999;52:371-374.
10. Michaelson JS, Chen LL, Silverstein MJ, et al. How cancer at the primary site and in the lymph nodes contributes to the risk of cancer death. *Cancer*. 2009;115:000-000.
11. Ries DAG, Melbert D, Krapcho M, et al. SEER Cancer Statistics Review, 1975-2005. Bethesda, Md: National Cancer Institute; 2008.
12. Michaelson JS, Chen LL, Silverstein MJ, et al. Why cancer at the primary site and in the lymph nodes contributes to the risk of cancer death. *Cancer*. 2009;115:000-000. ED1
13. 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 U S A*. 2001;98:11462-11467.
14. 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.
15. Ravdin PM, Siminoff LA, Davis GJ, et al. Computer program to assist in making decisions about adjuvant therapy for women with early breast cancer. *J Clin Oncol*. 2001;19:980-991.
16. Michaelson J. The role of molecular discreteness in normal and cancerous growth. *Anticancer Res*. 1999;19:4853-4867.
17. 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.

0000

The Impact of Primary Tumor Size, Lymph Node Status, and Other Prognostic Factors on the Risk of Cancer Death

L. Leon Chen, Matthew E. Nolan, Melvin J. Silverstein, Martin C. Mihm Jr., Arthur J. Sober, Kenneth K. Tanabe, Barbara L. Smith, Jerry Younger, James S. Michaelson, and Griffin Weber

A new framework is presented that identifies and quantifies the factors that contribute to cancer lethality and combines information on tumor size, lymph node status, and other prognostic factors into estimates of the risk of death. This mathematics drives web-based calculators that accurately estimate the risk of death for each patient.

AQ2

AQ1: Note that the running title was shortened to fit the journal maximum of 40 characters (including spaces): Okay as shortened?

AQ2: Okay to delete the website here: it also appears in the abstract and text.

AQ3: Please check the affiliations carefully to verify that the correct department/division is listed at for each author.

AQ4: Please indicate the most recent date you viewed the CancerMath website.

AQ5: Please indicate the most recent date you viewed the CancerMath website.

AQ6: Please indicate the most recent date you viewed the Technical Report 5 website.

AQ7: Please indicate the most recent date you viewed the CancerMath website.

AQ8: Please indicate the most recent date you viewed the lifemath website.

AQ9: Please check all tables carefully, including titles, headings, data, data alignment, abbreviations, and footnotes, and mark your corrections where needed.

AQ10: Is this Q value footnote correct for ER/PR status (the same as in Table 2)?

ED1: Please update page no.