

Novel Prognostic Staging System for Patients With De Novo Metastatic Breast Cancer

Jennifer K. Plichta, MD, MS^{1,2,3}; Samantha M. Thomas, MS^{4,5}; Daniel F. Hayes, MD^{6,7}; Mariana Chavez-MacGregor, MD, MSc^{8,9}; Kimberly Allison, MD¹⁰; Jennifer de los Santos, MD¹¹; Amy M. Fowler, MD, PhD^{12,13,14}; Armando E. Giuliano, MD¹⁵; Priyanka Sharma, MD¹⁶; Benjamin D. Smith, MD¹⁷; Elizabeth van Eycken, MD¹⁸; Stephen B. Edge, MD^{19,20}; and Gabriel N. Hortobagyi, MD²¹

PURPOSE Given the heterogeneity and improvement in outcomes for metastatic breast cancer (MBC), we developed a staging system that refines prognostic estimates for patients with metastatic cancer at the time of initial diagnosis, de novo MBC (dnMBC), on the basis of survival outcomes and disease-related variables.

METHODS Patients with dnMBC (2010-2016) were selected from the National Cancer Database (NCDB). Recursive partitioning analysis (RPA) was used to group patients with similar overall survival (OS) on the basis of clinical T category, grade, estrogen receptor (ER), progesterone receptor, human epidermal growth factor receptor 2, histology, organ system site of metastases (bone-only, brain-only, visceral), and number of organ systems involved. Three-year OS rates were used to assign a final stage: IVA: >70%, IVB: 50%-70%, IVC: 25 to <50%, and IVD: <25%. Bootstrapping was applied with 1,000 iterations, and final stage assignments were made based on the most commonly occurring assignment. Unadjusted OS was estimated. Validation analyses were conducted using SEER and NCDB.

RESULTS At a median follow-up of 52.9 months, the median OS of the original cohort (N = 42,467) was 35.4 months (95% CI, 34.8 to 35.9). RPA stratified patients into 53 groups with 3-year OS rates ranging from 73.5% to 5.7%; these groups were amalgamated into four stage groups: 3-year OS, A = 73.2%, B = 61.9%, C = 40.1%, and D = 17% (log-rank $P < .001$). After bootstrapping, the survival outcomes for the four stages remained significantly different (log-rank $P < .001$). This staging system was then validated using SEER data (N = 20,469) and a separate cohort from the NCDB (N = 7,645) (both log-rank $P < .001$).

CONCLUSION Our findings regarding the heterogeneity in outcomes for patients with dnMBC could guide future revisions of the current American Joint Committee on Cancer staging guidelines for patients with newly diagnosed stage IV disease. Our findings should be independently confirmed.

J Clin Oncol 41:2546-2560. © 2023 by American Society of Clinical Oncology

INTRODUCTION

Of the estimated 287,850 women with newly diagnosed breast cancer in 2022 in the United States, more than 17,000 (approximately 5.9%) will have metastatic disease at the time of initial presentation (de novo metastatic breast cancer [dnMBC]).¹ While most of these women ultimately succumb to breast cancer, survival varies widely with notable improvements for select subgroups in recent years.² Prior publications have proposed prognostic indices for patients who develop subsequent metastases,^{3,4} but few have sought to develop such tools for patients with dnMBC. A recently developed model for estimating prognosis incorporating clinical and tumor factors performed reasonably well (C-index, 0.731; 95% CI, 0.724 to 0.739)⁵; the model, however, included patients with both dnMBC and recurrent metastatic breast cancer (rMBC). Other risk scoring systems for stratifying patients with dnMBC have been developed, though some included treatment-

related variables,⁶ and others were built on broad populations without information on treatment.⁷ All of these studies confirm heterogeneity in outcomes for patients with dnMBC, which are currently all assigned the same stage at initial diagnosis: stage IV.

For patients with non-MBC, American Joint Committee on Cancer (AJCC) breast cancer staging system was revised in 2016 to include the nonanatomic factors, estrogen receptor (ER), progesterone receptor (PR), human epidermal growth factor receptor 2 (HER2), tumor grade, and genomic assays, in addition to tumor size and nodal status to define the prognostic stage group.^{8,9} Although prior versions of breast cancer staging were based solely on anatomic information, many studies have demonstrated the added value of these other factors in defining prognosis.^{10,11} These same variables, as well as other disease-related factors, such as the sites and number of sites of metastases, have been shown to be associated with duration of survival for patients with

ASSOCIATED CONTENT

Data Supplement

Author affiliations and support information (if applicable) appear at the end of this article.

Accepted on February 13, 2023 and published at ascopubs.org/journal/jco on March 21, 2023; DOI <https://doi.org/10.1200/JCO.22.02222>

CONTEXT

Key Objective

To develop a staging system on the basis of survival outcomes and disease-related variables for patients with de novo metastatic breast cancer (dnMBC) that refines prognostic estimates.

Knowledge Generated

Patients with dnMBC have highly variable survival outcomes, which are associated with select disease-related variables.

These associations can be used to stratify patients into four distinct subgroups, and our findings could be used to revise the current staging guidelines for patients with dnMBC.

Relevance (K.D. Miller)

Patients with de novo metastatic disease are often considered together with those whose disease has recurred after initial therapy. This analysis defines the unique and varied prognosis of patients with de novo metastatic disease, facilitating shared decision making.*

*Relevance section written by JCO Senior Deputy Editor Kathy D. Miller, MD.

MBC.^{2,5,12,13} On the basis of the changes in staging and on improving survival, we sought to develop a staging model for patients with dnMBC that would refine prognostic estimates, and we propose a novel staging system for stratifying patients with dnMBC on the basis of survival outcomes and disease-related variables.

METHODS

The 2004-2017 National Cancer Database (NCDB) Breast Cancer Participant User File (PUF) (Fall 2020 release) was used for the primary analysis. The NCDB is a hospital-based tumor registry in the United States, estimated to capture >80% of all patients diagnosed with breast cancer each year.¹⁴ Patients diagnosed before 2010 were excluded because HER2 data were not routinely collected in the NCDB until 2010 and to account for treatment improvements that have occurred over the past decade, particularly for those with HER2-positive disease. At the time of the primary analysis, the NCDB required exclusion of patients diagnosed in 2017, due to administrative censoring of survival data; thus, the final diagnosis years included in the primary analysis were 2010-2016. This study was restricted to patients diagnosed with clinical or pathologic M1 disease at diagnosis (dnMBC), which excludes patients with breast cancer cells in blood or bone marrow. Patients with unknown/missing survival time or vital status were excluded, as were those with missing or unknown clinical characteristics.

Staging Model Development

Overall survival (OS) was defined as the time from diagnosis to death due to any cause. Patients who did not die were censored at the last follow-up. Recursive partitioning analysis (RPA)¹⁵ was used to group patients with similar OS on the basis of clinical T category (T4 or not T4), grade (1-2 or 3), ER status (positive or negative), PR status (positive or negative),

HER2 status (positive or negative), histology (ductal, lobular, or other), bone-only metastases (yes or no), brain-only metastases (yes or no), visceral metastases (yes or no), and number of organ systems with metastatic disease (assigned S1-4, defined as the number of organ systems involved such that a patient with bone-only metastases would be assigned S1 regardless of the number of bone metastases and a patient with bone and liver metastases would be assigned S2 regardless of the number of bone and liver metastases). Given that staging is performed at the time of diagnosis, and as with other breast cancer staging criteria, treatment was not included as a variable. Unfortunately, the NCDB does not capture data on the amount/volume of disease at the distant sites (only location). The splitting criteria were set to a *P* value of .10 with minimum node size of $N = 100$. In general, recursive partitioning is a classification method that relies on covariates in a regression model to divide data. It is recursive, meaning that it first divides the data on the basis of the most significant covariate split and then continues to partition the newly created subgroups on the basis of the influence of the remaining covariates. This algorithm stops when a pre-specified criterion is met—in this case, the log-rank test *P* value for an attempted partition is >.10 for all possible covariates for all subgroups, while maintaining a minimum subgroup sample size of 100. After recursive partitioning, 3-year OS rates were estimated for each terminal node. Patients were grouped based on the association of disease characteristics with OS. The final staging of patients with homogeneous survival was created based on 3-year survival rates: group A as >70%, group B as 50% to 70%, group C as 25 to <50%, and group D as <25%.

Bootstrapping was then applied, and this algorithm was repeated 1,000 times, as a form of internal validation.¹⁶ For each iteration, a sample was drawn with replacement from the original data set equal to the sample size of the original data set, the RPA was repeated, and patient groups were

created based on that sample. The results were summarized, and final stage assignments were determined based on the group each disease characteristic combination fell into most often within the bootstrapped analysis. There were 768 possible combinations of the final variables selected for the proposed model, and 639 characteristic profiles were found in the NCDB cohort. The remaining 129 characteristic profiles were not included in the final bootstrapped stage assignment (Data Supplement [Table S1], online only). Notably, the original RPA created stage assignments for all 768 combinations, and thus, the 129 profiles not found within the NCDB cohort were still assigned a stage on the basis of the RPA with the original data (before bootstrapping).

Patient demographic, disease, and available treatment variables were summarized with No. (%) for categorical variables and median (IQR [defined as quartile 1 to quartile 3]) for continuous variables. Differences between groups were tested using the chi-square test for categorical variables and analysis of variance for continuous variables. Follow-up time was estimated using the reverse Kaplan-Meier method, and unadjusted OS was estimated using the Kaplan-Meier method. Median OS and survival rates are presented with 95% CIs. Since the NCDB does not have data on recurrence or cause of death, OS was selected as the primary end point, although most patients diagnosed with dnMBC ultimately die from their disease.² Given the potential impact of age on OS, an additional stratified analysis was performed for patients on the basis of age group <50 years versus ≥50 years. The effect of patient groupings determined by recursive partitioning with bootstrapping on OS was then estimated using the Cox proportional hazards model, after adjustment for known covariates (age, race/ethnicity, sex, surgery receipt, chemotherapy receipt, radiation therapy receipt, endocrine therapy receipt, and immunotherapy receipt). To account for the correlation of patients treated at the same hospital, a robust sandwich covariance estimator was included in the adjusted survival model.

Validation 1 (SEER data set)

Patients with breast cancer diagnosed with dnMBC were selected from the SEER-18 database (November 2020 release). Patients diagnosed before 2010 were excluded for the same reasons as those excluded from the NCDB analyses. Patients with missing or unknown survival status or time were excluded, as were those with missing data for T category, grade, ER status, PR status, HER2 status, histology, or metastatic site. Similar to the NCDB, SEER does not capture data on the amount/volume of disease at the distant sites (only location). Follow-up time was estimated using the reverse Kaplan-Meier method, and unadjusted OS was estimated using the Kaplan-Meier method. Median OS and survival rates are presented with 95% CIs. SEER is a population-based registry including virtually 100% of cases in select regions of the United States that collectively include approximately 48% of the US population.¹ The demographics

of the SEER patients are more likely to be representative of the US population. However, there is significant overlap with those patients captured in the NCDB.

Validation 2 (Updated NCDB data set)

Patients with breast cancer diagnosed in 2017-2018 with dnMBC were selected from the 2004-2019 NCDB Breast Cancer PUF (Spring 2022 release). Patients with missing or unknown survival time or vital status were excluded, as were those with missing data for T category, grade, ER status, PR status, HER2 status, histology, or metastatic site. Follow-up time was estimated using the reverse Kaplan-Meier method, and unadjusted OS was estimated using the Kaplan-Meier method. Median OS and survival rates are presented with 95% CIs.

Effective sample sizes are reported for each table/figure. No adjustments were made for multiple comparisons. All statistical analyses were conducted using SAS, version 9.4 (SAS Institute, Cary, NC), or R, version 4.0.4 (R Foundation for Statistical Computing, Vienna, Austria). Given the data in NCDB and SEER are deidentified, this study was deemed exempt by the corresponding author's institutional review board. Of note, most of the coauthors are members of the AJCC breast committee, but this work does not represent an official position by the AJCC.

RESULTS

In the 2004-2017 NCDB PUF, there were 2,981,828 patients with breast cancer. After applying the specified inclusion and exclusion criteria (Data Supplement [Figs S1A and S1B], online only), the final cohort sample size of those with dnMBC was N = 42,467 with a median follow-up of 52.9 months (95% CI, 52.4 to 53.5) and a median OS of 35.4 months (95% CI, 34.8 to 35.9). The median patient age was 61 years (IQR, 51-70 years). Most patients (70.2%) were non-Hispanic White, compared with 17.3% non-Hispanic Black and 5.4% Hispanic.

Staging Model Development (NCDB, diagnosis years 2010-2016)

The RPA stratified patients into 53 groups (Data Supplement [Fig S2], online only) with a range of 3-year OS rates from 73.5% to 5.7% (Table 1). In general, patients with better survival rates had a lower disease burden (S1) and were more likely to have ER-positive disease, while patients with the worst outcomes had more organ systems involved (S3-4) and were more likely to have ER-negative and HER2-negative disease. On the basis of the prespecified 3-year OS rates (A: > 70%, B: 50%-70%, C: 25-< 50%, and D: < 25%), patients in the 53 original groups from the RPA were amalgamated into four stages (IV A/B/C/D) with distinct survival outcomes (Fig 1A, log-rank $P < .001$). With increasing stage (A>>D), survival rates steadily declined: median 3-year OS rates, A: 73.2%, B: 61.9%, C: 40.1%, and D: 17%.

TABLE 1. Characteristics and 3-Year Overall Survival Rates for Terminal Nodes From the Recursive Partitioning Analysis (Data Supplement [Fig S2]), Based on Patients With de Novo Metastatic Breast Cancer, Diagnosed 2010-2016, in the National Cancer Database

Rank	3-Year Survival Rate, %	Node ^a	Sample Size	Characteristics	Proposed Stage Group
1	0.735	34	746	S1/ER+/T0-3/PR+/HER2-/grade 1-2/ductal histology/bone-only mets: no/visceral mets: no	A
2	0.731	39	2,216	S1/ER+/T0-3/PR+/HER2+	A
3	0.687	36	3,728	S1/ER+/T0-3/PR+/HER2-/grade 1-2/ductal histology/bone-only mets: yes	B
4	0.681	47	311	S1/ER+/T4/HER2-/grade 1-2/ductal or other histology/PR+/bone-only mets: no/visceral mets: no	B
5	0.663	35	780	S1/ER+/T0-3/PR+/HER2-/grade 1-2/ductal histology/bone-only mets: no/visceral mets: yes	B
6	0.657	58	948	S1/ER+/T4/HER2+/PR+	B
7	0.657	37	2,858	S1/ER+/T0-3/PR+/HER2-/grade 1-2/lobular or other histology	B
8	0.651	28	891	S1/ER+/T0-3/PR-/HER2+	B
9	0.619	18	1,601	S1/ER-/HER2+/T0-3	B
10	0.605	48	378	S1/ER+/T4/HER2-/grade 1-2/ductal or other histology/PR+/bone-only mets: no/visceral mets: yes	B
11	0.596	95	337	S2/ER+/PR+/visceral mets: yes/grade 3/HER2+/T0-3	B
12	0.586	25	384	S1/ER+/T0-3/PR-/HER2-/grade 1-2/bone-only mets: no	B
13	0.586	38	3,067	S1/ER+/T0-3/PR+/HER2-/grade 3	B
14	0.580	57	446	S1/ER+/T4/HER2+/PR-	B
15	0.574	49	1,166	S1/ER+/T4/HER2-/grade 1-2/ductal or other histology/PR+/bone-only mets: yes	B
16	0.557	89	461	S2/ER+/PR+/visceral mets: yes/grade 1-2/HER2+	B
17	0.553	19	1,095	S1/ER-/HER2+/T4	B
18	0.526	84	1,207	S2/ER+/PR+/visceral mets: yes/grade 1-2/HER2-/ductal or other histology/T0-3	B
19	0.523	26	814	S1/ER+/T0-3/PR-/HER2-/grade 1-2/bone-only mets: yes	B
20	0.497	50	663	S1/ER+/T4/HER2-/grade 1-2/lobular histology	C
21	0.484	44	280	S1/ER+/T4/HER2-/grade 1-2/ductal or other histology/PR-	C
22	0.479	85	798	S2/ER+/PR+/visceral mets: yes/grade 1-2/HER2-/ductal or other histology/T4	C
23	0.473	77	550	S2/ER+/PR-/HER2+	C
24	0.468	55	1,638	S1/ER+/T4/HER2-/grade 3/PR+	C
25	0.455	96	258	S2/ER+/PR+/visceral mets: yes/grade 3/HER2+/T4	C
26	0.455	87	360	S2/ER+/PR+/visceral mets: yes/grade 1-2/HER2-/lobular histology/T0-3	C
27	0.446	65	964	S2/ER-/HER2+	C
28	0.434	9	651	S1/ER-/HER2-/T0-3/visceral mets: no/bone-only mets: no/brain-only mets: no/ductal histology	C
29	0.399	53	317	S1/ER+/T4/HER2-/grade 3/PR-/visceral mets: no	C
30	0.398	105	460	S3-4/ER+/PR+/HER2+	C
31	0.3960	27	857	S1/ER+/T0-3/PR-/HER2-/grade 3	C
32	0.395	92	730	S2/ER+/PR+/visceral mets: yes/grade 3/HER2-/T0-3	C
33	0.381	88	135	S2/ER+/PR+/visceral mets: yes/grade 1-2/HER2-/lobular histology/T4	C
34	0.356	93	643	S2/ER+/PR+/visceral mets: yes/grade 3/HER2-/T4	C
35	0.348	79	412	S2/ER+/PR+/visceral mets: no	C
36	0.333	72	332	S2/ER+/PR-/HER2-/grade 1-2/ductal histology	C
37	0.295	102	230	S3-4/ER+/PR-/HER2+	C
38	0.286	104	1,281	S3-4/ER+/PR+/HER2-	C
39	0.267	12	859	S1/ER-/HER2-/T0-3/visceral mets: no/bone-only mets: yes	C

(continued on following page)

TABLE 1. Characteristics and 3-Year Overall Survival Rates for Terminal Nodes From the Recursive Partitioning Analysis (Data Supplement [Fig S2]), Based on Patients With de Novo Metastatic Breast Cancer, Diagnosed 2010-2016, in the National Cancer Database (continued)

Rank	3-Year Survival Rate, %	Node ^a	Sample Size	Characteristics	Proposed Stage Group
40	0.250	10	129	S1/ER-/HER2-/T0-3/visceral mets: no/bone-only mets: no/brain-only mets: no/lobular or other histology	C
41	0.249	13	1,154	S1/ER-/HER2-/T0-3/visceral mets: yes	D
42	0.248	15	293	S1/ER-/HER2-/T4/grade 1-2	D
43	0.237	66	479	S3-4/ER-/HER2+	D
44	0.234	75	207	S2/ER+/PR-/HER2-/grade 3/T0-3	D
45	0.226	54	130	S1/ER+/T4/HER2-/grade 3/PR-/visceral mets: yes	D
46	0.219	73	128	S2/ER+/PR-/HER2-/grade 1-2/lobular or other histology	D
47	0.194	76	180	S2/ER+/PR-/HER2-/grade 3/T4	D
48	0.184	100	114	S3-4/ER+/PR-/HER2-/grade 1-2	D
49	0.182	16	1,472	S1/ER-/HER2-/T4/grade 3	D
50	0.175	11	122	S1/ER-/HER2-/T0-3/visceral mets: no/bone-only mets: no/brain-only mets: yes	D
51	0.112	101	172	S3-4/ER+/PR-/HER2-/grade 3	D
52	0.085	62	1,430	S2/ER-/HER2-	D
53	0.057	63	605	S3-4/ER-/HER2-	D

Abbreviations: ER, estrogen receptor; HER2, human epidermal growth factor receptor 2; mets, ; PR, progesterone receptor; S, system; T, tumor.
^aNode number correlates with the node numbering in the tree from the recursive partitioning analysis (Data Supplement [Fig S2]).

Bootstrapping was then applied (Data Supplement [Table S2], online only), and final stages were assigned. For those combinations not represented in the primary data set, stage assignments were based on the RPA of the original data (N = 129 profiles; Data Supplement [Table S1]). In total, 25 profiles were assigned to stage A, 200 to stage B, 225 to stage C, and 189 to stage D. Of note, 25 characteristic profiles were assigned to a different stage after bootstrapping compared with the original RPA-based assignment. Once completed, final stage assignments were made for all 768 possible combination profiles (Data Supplement [Table S2]), and survival outcomes remained significantly different across the four final stages (Fig 1B; log-rank $P < .001$). After adjustment for known covariates, including demographic and treatment variables, the association between OS and the newly defined stage group remained statistically significant ($P < .001$; Data Supplement [Table S3], online only). Notably, age was also significantly associated with survival in our adjusted model, and as such, a stratified analysis was performed based on age <50 years versus ≥ 50 years, which demonstrated similar significant differences between the stage groups (Data Supplement [Fig S3], online only; for both age groups, log-rank $P < .001$).

When comparing patients in the newly assigned stage groups (A, B, C, and D; Table 2), patients in the lowest stage tended to be younger (age 18-45 years, A: 20.3%, B: 13.9%, C: 13%, and D: 13.3%, $P < .001$) and more likely to be non-Hispanic White (A: 72.9%, B: 73.3%, C:

68%, and D: 62.6%, $P < .001$). Given that tumor and disease characteristics were used for grouping patients, there were many expected differences between patients in the four stage groups (Table 3). On the basis of their poor survival, the RPA categorized patients with triple negative tumors (ER-negative/PR-negative/HER2-negative) as stages C or D, and 73.2% of patients in the stage D group had triple negative disease. In stage A, all patients had only one site of disease, while all patients with three or four sites of disease were in stages C or D. Not surprisingly, patients in the lower stages were more likely to receive any surgery (A: 48.4%, B: 36.8%, C: 28.8%, and D: 31.5%, $P < .001$), while radiation receipt was not meaningfully different across groups (A: 35.2%, B: 32.9%, C: 34.8%, and D: 33%; Data Supplement [Table S4], online only). Chemotherapy was used more frequently in stage D as part of the first-line treatment (A: 66%, B: 53.2%, C: 61.2%, and D: 73.7%, $P < .001$), which may be related to the tumor phenotype that is more common in this stage (ie, triple negative).

Validation 1 (SEER, diagnosis years 2010-2018)

The final validation cohort sample size was N = 20,469 (Data Supplement [Table S5], online only), and the median follow-up cohort was 47 months (95% CI, 46 to 48). After applying our staging model, patients were successfully stratified into four distinct stages with statistically different survival outcomes (median 3-year OS rates, A: 72.5%, B: 58.4%, C: 35.7%, and D: 14.6%; log-rank $P < .001$; Fig 2A).

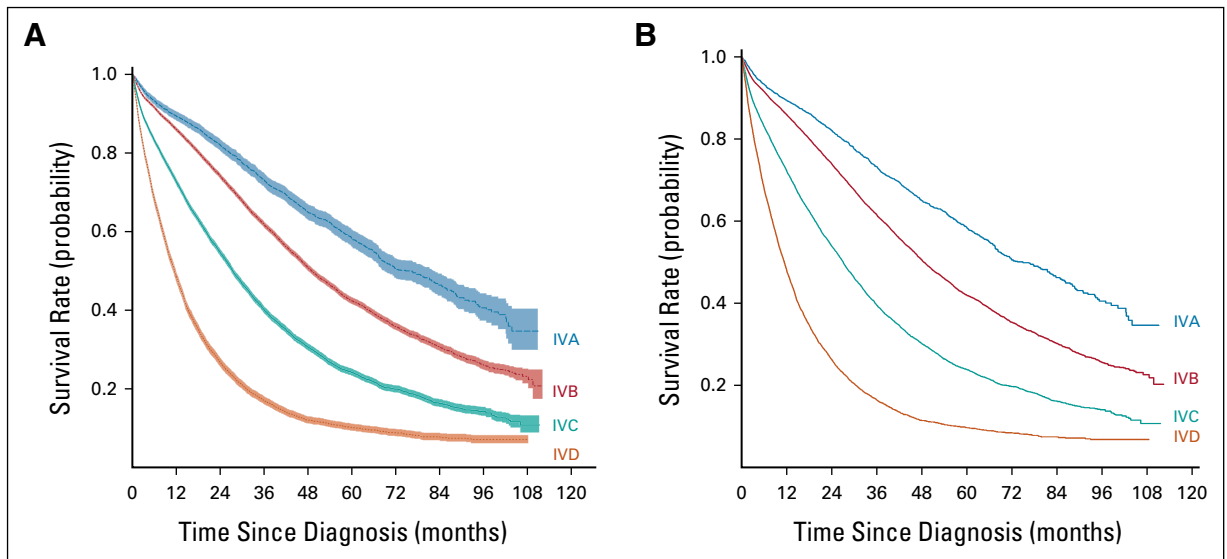


FIG 1. Unadjusted overall survival stratified by assigned stage IVA-D (3-year survival: IVA: >70%, IVB: 50%-70%, IVC: 25% to <50%, and IVD: <25%) shown for (A) the original recursive partitioning analysis and (B) the bootstrapped analysis. Analyses on the basis of patients with de novo metastatic breast cancer, diagnosed 2010-2016, in the National Cancer Database.

Validation 2 (NCDB, diagnosis years 2017-2018)

The final validation cohort sample size was $N = 14,958$ (Data Supplement [Table 6], online only), and the median follow-up was 32.7 months (95% CI, 32.4 to 33.0). Application of our proposed staging model again successfully stratified patients into four stages with distinct survival outcomes (median 3-year OS rates, A: 76.6%, B: 64.2%, C: 43.5%, and D: 21.1%; log-rank $P < .001$; Fig 2B). Of note, changes in systemic therapy may have been considerable during this time. Although this may have impacted the absolute survival estimates, it did not appear to alter the relative stratification of the cohort into four distinct stage groups, as the differences between the groups remained statistically significant.

DISCUSSION

Cancer staging systems were originally developed to concisely and accurately summarize a patient's disease and prognosis.¹⁷ In 2016, the AJCC revised staging guidelines for patients with breast cancer.⁸ However, this did not include any refinement for patients with dnMBC in that all patients, regardless of disease features, remained grouped into one stage IV category. Our data confirm that the outcomes in this patient population are heterogeneous.^{2,5,6} Using large population data sets and including ER, PR, HER2, and grade, we identified four unique subgroups of patients with dnMBC with distinctly different survival. We propose designating these groups as prognostic stages IVA, IVB, IVC, and IVD. Although the creation of this staging system relied largely on a statistical approach (RPA, as opposed to a clinically driven a priori approach), the results were consistent with what is generally observed in practice: patients with less extensive burden of

metastatic disease and/or hormone receptor positive cancer tend to live longer, while those with a higher disease burden and/or triple negative cancer tend to have a shorter survival. Moreover, our findings were reproducible using multiple data sets, thus validating our proposed staging system. If implemented, our novel staging proposal will fulfill a currently unmet need to refine prognostic estimates for this unique population. As with other staging guidelines, it will concisely summarize a patient's disease status and prognosis and, thus, facilitate communication between patients, providers, and researchers.

The AJCC already subdivides stage IV in some other disease sites. For example, non-small-cell lung cancer is categorized as stage IVA or IVB, and colorectal cancer is categorized as stage IVA, IVB, or IVC on the basis of the location and extent of metastatic disease.⁸ On the basis of our findings, the AJCC should consider updating the breast cancer prognostic stage groups to reflect this heterogeneity in outcomes for patients with dnMBC. Furthermore, it will allow oncology providers to relay important prognostic information with relative ease. As our understanding of breast cancer has evolved, we now know that the term "breast cancer" represents a diverse disease entity, and contemporary staging guidelines must accurately reflect this heterogeneity.

However, stratifying patients with dnMBC should only be undertaken if significant differences in outcomes can be consistently demonstrated based on disease factors that are reliably tested. To this end, we selected clinically relevant variables that have been repeatedly shown to be associated with survival outcomes^{12,13,18,19} and were readily available in large national tumor registries. While other factors not included in our analyses, such as genomic assays and

TABLE 2. Patient Demographics by Bootstrap Stage IVA-D

Demographic Variable	All Patients (N = 42,467)	Stage IVA (n = 2962)	Stage IVB (n = 21,118)	Stage IVC (n = 12,117)	Stage IVD (n = 6,270)	P
Age, years						< .001
18-45	5,933 (14)	601 (20.3)	2,926 (13.9)	1,575 (13)	831 (13.3)	
46-74	29,084 (68.5)	1,907 (64.4)	14,248 (67.5)	8,547 (70.5)	4,382 (69.9)	
≥75	7,450 (17.5)	454 (15.3)	3,944 (18.7)	1,995 (16.5)	1,057 (16.9)	
Median (IQR)	61 (51-70)	59 (48-69)	61 (51-71)	61 (52-70)	60 (51-69)	< .001
Sex						< .001
Female	41,857 (98.6)	2,915 (98.4)	20,784 (98.4)	11,930 (98.5)	6,228 (99.3)	
Male	610 (1.4)	47 (1.6)	334 (1.6)	187 (1.5)	42 (0.7)	
Race/ethnicity						< .001
Non-Hispanic White	29,807 (70.2)	2,159 (72.9)	15,480 (73.3)	8,243 (68)	3,925 (62.6)	
Non-Hispanic Black	7,335 (17.3)	386 (13)	3,016 (14.3)	2,332 (19.2)	1,601 (25.5)	
Hispanic	2,296 (5.4)	183 (6.2)	1,091 (5.2)	671 (5.5)	351 (5.6)	
Other	1,665 (3.9)	123 (4.2)	847 (4)	504 (4.2)	191 (3)	
Missing	1,364 (3.2)	111 (3.7)	684 (3.2)	367 (3)	202 (3.2)	
Income level, USD						< .001
<\$48,000	16,449 (38.7)	1,053 (35.6)	7,728 (36.6)	4,872 (40.2)	2,796 (44.6)	
≥\$48,000	22,819 (53.7)	1,627 (54.9)	11,580 (54.8)	6,434 (53.1)	3,178 (50.7)	
Missing	3,199 (7.5)	282 (9.5)	1,810 (8.6)	811 (6.7)	296 (4.7)	
Education level						< .001
≤87% HS graduation rate	17,643 (41.5)	1,183 (39.9)	8,216 (38.9)	5,278 (43.6)	2,966 (47.3)	
>87% HS graduation rate	21,640 (51)	1,497 (50.5)	11,101 (52.6)	6,033 (49.8)	3,009 (48)	
Missing	553 (1.3)	282 (9.5)	1,801 (8.5)	806 (6.7)	295 (4.7)	
Insurance type						< .001
Private	17,902 (42.2)	1,460 (49.3)	9,177 (43.5)	4,832 (39.9)	2,433 (38.8)	
Government	21,849 (51.4)	1,359 (45.9)	10,777 (51)	6,349 (52.4)	3,364 (53.7)	
Not insured	2,163 (5.1)	111 (3.7)	898 (4.3)	773 (6.4)	381 (6.1)	
Missing	553 (1.3)	553 (1.3)	32 (1.1)	266 (1.3)	163 (1.3)	
CD comorbidity score						< .001
0	34,494 (81.2)	2,396 (80.9)	17,244 (81.7)	9,918 (81.9)	4,936 (78.7)	
1	5,806 (13.7)	426 (14.4)	2,834 (13.4)	1,611 (13.3)	935 (14.9)	
≥2	2,167 (5.1)	140 (4.7)	1,040 (4.9)	588 (4.9)	399 (6.4)	
Facility type						.007
Academic	14,493 (34.1)	976 (33)	7,112 (33.7)	4,141 (34.2)	2,264 (36.1)	
Integrated network	5,842 (13.8)	393 (13.3)	2,905 (13.8)	1,672 (13.8)	872 (13.9)	
Comprehensive	17,742 (41.8)	1,305 (44.1)	8,911 (42.2)	5,027 (41.5)	2,499 (39.9)	
Community	4,390 (10.3)	288 (9.7)	2,190 (10.4)	1,277 (10.5)	635 (10.1)	

(continued on following page)

TABLE 2. Patient Demographics by Bootstrap Stage IVA-D (continued)

Demographic Variable	All Patients (N = 42,467)	Stage IVA (n = 2962)	Stage IVB (n = 21,118)	Stage IVC (n = 12,117)	Stage IVD (n = 6,270)	P
Facility location						< .001
Midwest	10,887 (25.6)	720 (24.3)	5,612 (26.6)	3,072 (25.4)	1,483 (23.7)	
Northeast	8,925 (21)	611 (20.6)	4,380 (20.7)	2,615 (21.6)	1,319 (21)	
South	16,110 (37.9)	1,120 (37.8)	7,735 (36.6)	4,624 (38.2)	2,631 (42)	
West	6,545 (15.4)	511 (17.3)	3,391 (16.1)	1,806 (14.9)	837 (13.3)	
Year of diagnosis						< .001
2010	5,136 (12.1)	350 (11.8)	2,440 (11.6)	1,501 (12.4)	845 (13.5)	
2011	5,609 (13.2)	364 (12.3)	2,748 (13)	1,637 (13.5)	860 (13.7)	
2012	5,496 (12.9)	378 (12.8)	2,705 (12.8)	1,613 (13.3)	800 (12.8)	
2013	5,902 (13.9)	402 (13.6)	2,914 (13.8)	1,704 (14.1)	882 (14.1)	
2014	6,530 (15.4)	452 (15.3)	3,232 (15.3)	1,888 (15.6)	958 (15.3)	
2015	6,759 (15.9)	527 (17.8)	3,435 (16.3)	1,825 (15.1)	972 (15.5)	
2016	7035 (16.6)	489 (16.5)	3,644 (17.3)	1,949 (16.1)	953 (15.2)	

NOTE. Percentages may not add up to 100 due to rounding. Data are presented as No. (%) unless otherwise specified. Analyses on the basis of patients with de novo metastatic breast cancer, diagnosed 2010-2016, in the National Cancer Database.

Abbreviations: CD, Charlson/Deyo; HS, high school, USD, US dollars.

TABLE 3. Disease Characteristics by Bootstrap Stage IVA-D

Disease Characteristic	All Patients (N = 42,467)	Stage IVA (n = 2,962)	Stage IVB (n = 21,118)	Stage IVC (n = 12,117)	Stage IVD (n = 6,270)	P
Histology						
Ductal	32,705 (77)	2,657 (89.7)	15,544 (73.6)	9,583 (79.1)	4,921 (78.5)	a
Lobular	6,905 (16.3)	258 (8.7)	4,554 (21.6)	1,690 (13.9)	403 (6.4)	
Other	2,857 (6.7)	47 (1.6)	1,020 (4.8)	844 (7)	946 (15.1)	
Clinical T category						
T0/IS	208 (0.5)	18 (0.6)	100 (0.5)	55 (0.5)	35 (0.6)	a
T1	6,827 (16.1)	795 (26.8)	4,056 (19.2)	1,350 (11.1)	626 (10)	
T2	14,199 (33.4)	1,536 (51.9)	8,166 (38.7)	3,014 (24.9)	1,483 (23.7)	
T3	6,848 (16.1)	613 (20.7)	3,660 (17.3)	1,647 (13.6)	928 (14.8)	
T4	14,385 (33.9)	0 (0)	5,136 (24.3)	6,051 (49.9)	3,198 (51)	
Clinical N category						
N0	11,074 (26.1)	1,083 (36.6)	6,287 (29.8)	2,459 (20.3)	1,245 (19.9)	< .001
N1	18,243 (43)	1,230 (41.5)	9,226 (43.7)	5,190 (42.8)	2,597 (41.4)	
N2	5,045 (11.9)	236 (8)	2,188 (10.4)	1,739 (14.4)	882 (14.1)	
N3	5,744 (13.5)	282 (9.5)	2,251 (10.7)	2,031 (16.8)	1,180 (18.8)	
NX	2,249 (5.3)	125 (4.2)	1,108 (5.2)	665 (5.5)	351 (5.6)	
Missing	112 (0.3)	6 (0.2)	58 (0.3)	33 (0.3)	15 (0.2)	
Clinical M category						
M0	2,048 (4.8)	291 (9.8)	1,196 (5.7)	373 (3.1)	188 (3)	< .001
M1	40,153 (94.6)	2,648 (89.4)	19,776 (93.6)	11,681 (96.4)	6,048 (96.5)	
Missing	266 (0.6)	23 (0.8)	146 (0.7)	63 (0.5)	34 (0.5)	
Inflammatory breast cancer (cT4D)						
No	39,077 (92)	2,962 (100)	19,884 (94.2)	10,890 (89.9)	5,341 (85.2)	< .001
Yes	3,390 (8)	0 (0)	1,234 (5.8)	1,227 (10.1)	929 (14.8)	
Pathologic T category						
T0/IS	954 (2.2)	92 (3.1)	481 (2.3)	233 (1.9)	148 (2.4)	< .001
T1	3,244 (7.6)	517 (17.5)	1,838 (8.7)	619 (5.1)	270 (4.3)	
T2	4,892 (11.5)	563 (19)	2,803 (13.3)	1,016 (8.4)	510 (8.1)	
T3	2,242 (5.3)	159 (5.4)	1,285 (6.1)	499 (4.1)	299 (4.8)	
T4	2,808 (6.6)	51 (1.7)	1,052 (5)	1,027 (8.5)	678 (10.8)	
TX	12,375 (29.1)	681 (23)	5,982 (28.3)	3,821 (31.5)	1,891 (30.2)	
Missing	15,952 (37.6)	899 (30.4)	7,677 (36.4)	4,902 (40.5)	2,474 (39.5)	
Pathologic N category						
N0	3,439 (8.1)	480 (16.2)	1,707 (8.1)	751 (6.2)	501 (8)	< .001
N1	4,563 (10.7)	419 (14.1)	2,457 (11.6)	1,066 (8.8)	621 (9.9)	

(continued on following page)

TABLE 3. Disease Characteristics by Bootstrap Stage IVA-D (continued)

Disease Characteristic	All Patients (N = 42,467)	Stage IVA (n = 2,962)	Stage IVB (n = 21,118)	Stage IVC (n = 12,117)	Stage IVD (n = 6,270)	P
N2	2,921 (6.9)	247 (8.3)	1,626 (7.7)	725 (6)	323 (5.2)	
N3	2,455 (5.8)	199 (6.7)	1,384 (6.6)	578 (4.8)	294 (4.7)	
NX	12,978 (30.6)	711 (24)	6,209 (29.4)	4,032 (33.3)	2,026 (32.3)	
Missing	16,111 (37.9)	906 (30.6)	7,735 (36.6)	4,965 (41)	2,505 (40)	
Pathologic M category						.04
M0	256 (0.6)	35 (1.2)	129 (0.6)	63 (0.5)	29 (0.5)	
M1	13,690 (32.2)	1,182 (39.9)	7,293 (34.5)	3,431 (28.3)	1,784 (28.5)	
Missing	28,521 (67.2)	1,745 (58.9)	13,696 (64.9)	8,623 (71.2)	4,457 (71.1)	
ER status						a
ER+	31,613 (74.4)	2,962 (100)	18,422 (87.2)	9,298 (76.7)	931 (14.8)	
ER-	10,854 (25.6)	0 (0)	2,696 (12.8)	2,819 (23.3)	5,339 (85.2)	
PR status						a
PR+	26,152 (61.6)	2,962 (100)	15,970 (75.6)	6,912 (57)	308 (4.9)	
PR-	16,315 (38.4)	0 (0)	5,148 (24.4)	5,205 (43)	5,962 (95.1)	
HER2 status						a
HER2+	11,007 (25.9)	2,216 (74.8)	5,784 (27.4)	2,528 (20.9)	479 (7.6)	
HER2-	31,460 (74.1)	746 (25.2)	15,334 (72.6)	9,589 (79.1)	5,791 (92.4)	
Tumor subtype						< .001
HER2+	11,007 (25.9)	2,216 (74.8)	5,784 (27.4)	2,528 (20.9)	479 (7.6)	
HR+/HER2-	25,162 (59.3)	746 (25.2)	15,334 (72.6)	7,878 (65)	1,204 (19.2)	
TNBC	6,298 (14.8)	0 (0)	0 (0)	1,711 (14.1)	4,587 (73.2)	
Grade						a
1	3,326 (7.8)	241 (8.1)	2,488 (11.8)	512 (4.2)	85 (1.4)	
2	18,213 (42.9)	1,583 (53.4)	11,813 (55.9)	3,899 (32.2)	918 (14.6)	
3	20,928 (49.3)	1,138 (38.4)	6,817 (32.3)	7,706 (63.6)	5,267 (84)	
Metastatic site						
Lung						< .001
No	29,896 (70.4)	2,673 (90.2)	17,533 (83)	7,011 (57.9)	2,679 (42.7)	
Yes	12,571 (29.6)	289 (9.8)	3,585 (17)	5,106 (42.1)	3,591 (57.3)	
Liver						< .001
No	32,326 (76.1)	2,623 (88.6)	18,287 (86.6)	7,840 (64.7)	3,576 (57)	
Yes	10,141 (23.9)	339 (11.4)	2,831 (13.4)	4,277 (35.3)	2,694 (43)	
Brain						< .001
No	39,537 (93.1)	2,880 (97.2)	20,852 (98.7)	10,645 (87.9)	5,160 (82.3)	
Yes	2,930 (6.9)	82 (2.8)	266 (1.3)	1,472 (12.1)	1,110 (17.7)	

(continued on following page)

TABLE 3. Disease Characteristics by Bootstrap Stage IVA-D (continued)

Disease Characteristic	All Patients (N = 42,467)	Stage IVA (n = 2,962)	Stage IVB (n = 21,118)	Stage IVC (n = 12,117)	Stage IVD (n = 6,270)	P
Bone						< .001
No	15,008 (35.3)	1,735 (58.6)	7,226 (34.2)	2,973 (24.5)	3,074 (49)	
Yes	27,459 (64.7)	1,227 (41.4)	13,892 (65.8)	9,144 (75.5)	3,196 (51)	
Other						< .001
No	36,855 (86.8)	1,937 (65.4)	18,564 (87.9)	10,640 (87.8)	5,714 (91.1)	
Yes	5,612 (13.2)	1,025 (34.6)	2,554 (12.1)	1,477 (12.2)	556 (8.9)	
Metastatic sites						^a
1	29,994 (70.6)	2,962 (100)	19,108 (90.5)	4,969 (41)	2,955 (47.1)	
2	9,132 (21.5)	0 (0)	2,010 (9.5)	5,177 (42.7)	1,945 (31)	
3	2,909 (6.9)	0 (0)	0 (0)	1,731 (14.3)	1,178 (18.8)	
4	432 (1)	0 (0)	0 (0)	240 (2)	192 (3.1)	
Lung-only metastasis						< .001
No	38,055 (89.6)	2,673 (90.2)	18,740 (88.7)	11,593 (95.7)	5,049 (80.5)	
Yes	4,412 (10.4)	289 (9.8)	2,378 (11.3)	524 (4.3)	1,221 (19.5)	
Liver-only metastasis						< .001
No	39,391 (92.8)	2,623 (88.6)	19,185 (90.8)	11,835 (97.7)	5,748 (91.7)	
Yes	3,076 (7.2)	339 (11.4)	1,933 (9.2)	282 (2.3)	522 (8.3)	
Brain-only metastasis						^a
No	41,957 (98.8)	2,880 (97.2)	20,918 (99.1)	12,061 (99.5)	6,098 (97.3)	
Yes	510 (1.2)	82 (2.8)	200 (0.9)	56 (0.5)	172 (2.7)	
Bone-only metastasis						^a
No	26,083 (61.4)	1,735 (58.6)	9,075 (43)	9,487 (78.3)	5,786 (92.3)	
Yes	16,384 (38.6)	1,227 (41.4)	12,043 (57)	2,630 (21.7)	484 (7.7)	
Visceral metastasis ^b						^a
No	23,217 (54.7)	2,334 (78.8)	14,829 (70.2)	4,682 (38.6)	1,372 (21.9)	
Yes	19,250 (45.3)	628 (21.2)	6,289 (29.8)	7,435 (61.4)	4,898 (78.1)	

NOTE. Percentages may not add up to 100 due to rounding. Data are presented as No. (%) unless otherwise specified. Analyses on the basis of patients with de novo metastatic breast cancer, diagnosed 2010-2016, in the National Cancer Database.

Abbreviations: ER, estrogen receptor; HER2, human epidermal growth factor receptor 2; HR, hormone receptor; IBC, inflammatory breast cancer; IS, in situ; N, node; M, metastasis; PR, progesterone receptor; T, tumor; TNBC, triple negative breast cancer.

^aVariable included in staging algorithm.

^bVisceral metastasis includes liver and lung metastatic sites.

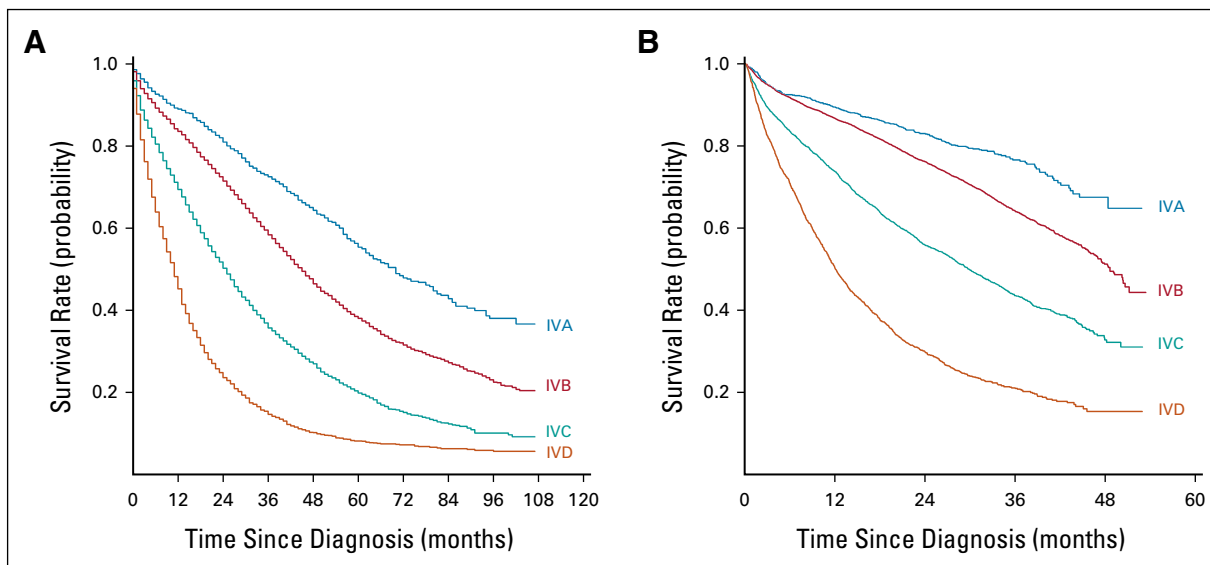


FIG 2. Unadjusted overall survival stratified by assigned stage IVA-D (assignments on the basis of recursive partitioning and bootstrapped analyses) for separate validation cohorts from (A) SEER (patients diagnosed with de novo metastatic breast cancer 2010-2018) and (B) the National Cancer Database (patients diagnosed with metastatic breast cancer 2017-2018).

circulating tumor cells, may also be associated with outcomes,^{20,21} sufficient data sets are not available to confirm or refute their reliability for use in our staging guidelines. Demographic factors could also be considered as prognostic variables, such as age with differentiated thyroid cancer.⁸ However, staging guidelines for patients with non-MBC are limited to only tumor/disease-related factors, and thus, we opted to focus on similar variables for our modeling. Although the treatments received may also impact outcomes, the staging guidelines for breast cancer focus on baseline disease and patient characteristics.^{22,23} It is also important to note that systemic therapy recommendations changed during our study period, and this may have resulted in improved outcomes for select subgroups, which will require future follow-up. Regardless, our proposed system is meant to serve as a starting point upon which future versions with additional variables will continue to refine the prognostic estimates associated with each category.

Others have also proposed nomograms, online tools, and staging models for patients with dnMBC, which could also be used for providing prognostic estimates for patients if more broadly validated.^{5,24,25} On comparison, staging models, such as the one proposed here, categorize patients with similar outcomes, which imply that any patient in any given group will likely have a survival expectation similar to others in that group, thus facilitating communication between providers and patients. One recent study by two of our coauthors used a similar methodology and categorized patients into three distinct subgroups.²⁵ In contrast, our team used a more contemporary data set and elected to divide patients into four groups, thus providing more granular prognostic information. Our model also included more clinically relevant variables (such as histology, brain-only

metastases, and visceral metastases). Further refinements will likely be necessary as our understanding of MBC evolves and treatments continue to improve.

Importantly, staging for dnMBC relays information different from that provided for those with nonmetastatic cancer. Staging in those with nonmetastatic cancer estimates the chance of long-term survival or cure from cancer. Unfortunately, because the large majority of those with MBC will ultimately succumb to their cancer,² the value of stratifying patients with dnMBC is to provide information on the potential duration of survival. This does not diminish the value of developing these stage groupings, as this information may be of great value to patients, providers, and caregivers. While we recognize that our staging system is complex, we believe that it can be relatively easily implemented with the use of electronic clinical support tools, such as those currently used for staging patients with non-MBC.

There were several limitations to our study. While the NCCDB is the largest national tumor registry in the country, it is not population-based and has a slightly smaller proportion of patients with dnMBC compared with SEER. This may be attributed to the fact that hospitals participating in the NCCDB are Commission on Cancer–accredited facilities, and patients with advanced disease may not seek out these types of hospitals. Data from the NCCDB are also observational and prone to errors in data entry. Furthermore, when using these types of retrospective data for unintended purposes, any analyses performed are subject to indeterminate and uncertain selection bias. Notably, >26,000 patients with dnMBC were excluded for missing staging variables (76% due to missing clinical T category and/or tumor grade), which may have influenced our findings.²⁶

We were able to validate these findings using SEER data, a population-based registry of a representative population in the United States, although there is significant overlap with those patients captured in the NCDB. While our contemporary NCDB cohort was a completely separate data set from the original cohort used to build the model, it had a much shorter median follow-up, which may reduce its ability to independently validate our findings. All cancer registries have a limited set of variables available, thus restricting which variables could be included in our modeling. Furthermore, there is known heterogeneity in the analyses performed to assess certain disease characteristics, such as HER2 status.²⁷ In addition, treatment variables may not be reliably recorded, and systemic therapy in particular may have changed considerably over the study periods, which may have impacted our findings. This may contribute to observed discrepancies between outcomes from tumor registry and clinical trial data.^{22,23} It is also important to recognize that our staging guidelines assume that most patients with dnMBC will receive appropriate treatment, similar to the staging guidelines for non-MBC.⁸

Regardless, our findings are consistent with those observed in clinical practice, and we were able to validate our proposed staging guidelines using multiple methods.

In conclusion, we recommend that the AJCC expands the prognostic stage groups for patients with dnMBC to include four subgroups termed stages IVA, IVB, IVC, and IVD on the basis of 3-year OS rates (>70%, 50%-70%, 25% to <50%, and <25%, respectively). This proposed system acknowledges and more accurately communicates the heterogeneous outcomes observed for those with dnMBC and will provide patients, caregivers, and providers useful information in treatment planning, patient understanding of their options, and focused discussions of patient values and end-of-life goals. We subsequently plan to further validate our staging system in collaboration with international tumor registries and organizations. Although further evaluation and refinement will undoubtedly be necessary, this proposal could serve as an excellent foundation for future work and potentially serve as an example for other disease sites.

AFFILIATIONS

¹Department of Surgery, Duke University Medical Center, Durham, NC

²Department of Population Health Sciences, Duke University Medical Center, Durham, NC

³Duke Cancer Institute, Durham, NC

⁴Biostatistics Shared Resource, Duke Cancer Institute, Durham, NC

⁵Duke University, Department of Biostatistics & Bioinformatics, Durham, NC

⁶University of Michigan Rogel Cancer Center, Ann Arbor, MI

⁷Department of Internal Medicine, University of Michigan, Ann Arbor, MI

⁸Department of Health Services Research, The University of Texas MD Anderson Cancer Center, Houston, TX

⁹Department of Breast Medical Oncology, The University of Texas MD Anderson Cancer Center, Houston, TX

¹⁰Department of Pathology, Stanford University School of Medicine, Stanford, CA

¹¹Grandview Cancer Center, Department of Radiation Oncology, Birmingham, AL

¹²Department of Radiology, University of Wisconsin School of Medicine and Public Health, Madison, WI

¹³Department of Medical Physics, University of Wisconsin School of Medicine and Public Health, Madison, WI

¹⁴University of Wisconsin Carbone Cancer Center, Madison, WI

¹⁵Cedars-Sinai Medical Center, University of California—Los Angeles, Los Angeles, CA

¹⁶Department of Internal Medicine, Division of Medical Oncology, University of Kansas Medical Center, Westwood, KS

¹⁷Division of Radiation Oncology, The University of Texas MD Anderson Cancer Center, Houston, TX

¹⁸Belgian Cancer Registry, Brussels, Belgium

¹⁹Department of Surgical Oncology and Cancer Prevention and Control, University at Buffalo, Buffalo, NY

²⁰Roswell Park Comprehensive Cancer Center, Buffalo, NY

²¹Department of Breast Medical Oncology, Division of Cancer Medicine, The University of Texas MD Anderson Cancer Center, Houston, TX

CORRESPONDING AUTHOR

Jennifer K. Plichta, MD, MS, DUMC 3513, Durham, NC 27710; Twitter: @JenniferPlichta; e-mail: jennifer.plichta@duke.edu.

DISCLAIMER

The content of this manuscript is solely the responsibility of the authors and does not necessarily represent the official views of the NIH.

SUPPORT

Supported by the National Institutes of Health Office of Women's Research Building Interdisciplinary Research Careers in Women's Health K12HD043446 (J.K.P.) (PI: Amundsen). Supported in part by Duke Cancer Institute through NIH grant P30CA014236 (PI: Kastan) for the Biostatistics Core.

AUTHORS' DISCLOSURES OF POTENTIAL CONFLICTS OF INTEREST

Disclosures provided by the authors are available with this article at DOI <https://doi.org/10.1200/JCO.22.02222>.

DATA SHARING STATEMENT

The data sets generated during and/or analyzed during the current study are available from the corresponding author on reasonable request and approval by the American College of Surgeons Commission on Cancer. The NCDB and SEER databases can also be obtained directly from the associated managing institutions.

AUTHOR CONTRIBUTIONS

Conception and design: Jennifer K. Plichta, Samantha M. Thomas, Daniel F. Hayes, Mariana Chavez-MacGregor, Amy M. Fowler, Priyanka Sharma, Benjamin D. Smith, Stephen B. Edge, Gabriel N. Hortobagyi

Financial support: Jennifer K. Plichta

Administrative support: Jennifer K. Plichta, Gabriel N. Hortobagyi

Provision of study material or patients: Priyanka Sharma

Collection and assembly of data: Jennifer K. Plichta, Samantha M. Thomas, Gabriel N. Hortobagyi

Data analysis and interpretation: Jennifer K. Plichta, Samantha M. Thomas, Kimberly Allison, Jennifer de los Santos, Armando E. Giuliano, Benjamin D. Smith, Elizabeth van Eycken, Stephen B. Edge, Gabriel N. Hortobagyi

Manuscript writing: All authors

Final approval of manuscript: All authors

Accountable for all aspects of the work: All authors

ACKNOWLEDGMENT

We thank Elaine Alexander, PhD, for all of her work in keeping our group organized and on task. Without her tireless efforts and organization, this publication would not have been possible.

The National Cancer Data Base (NCDB) is a joint project of the Commission on Cancer (CoC) of the American College of Surgeons and the American Cancer Society. The CoC's NCDB and the hospitals participating in the CoC NCDB are the source of the deidentified data used herein; they have not verified and are not responsible for the statistical validity of the data analysis or the conclusions derived by the authors.

This study used the SEER cancer incidence database. SEER is supported by the Surveillance Research Program (SRP) in NCI's Division of Cancer Control and Population Sciences (DCCPS). The interpretation and reporting of these data are the sole responsibility of the authors. The authors acknowledge the efforts of the National Cancer Institute and the SEER Program tumor registries in the creation of the SEER **database.

REFERENCES

1. Surveillance, Epidemiology, and End Results (SEER): Cancer Stat Facts: Female Breast Cancer. Bethesda, MD, National Cancer Institute, 2022
2. Taskindoust M, Thomas SM, Sammons SL, et al: Survival outcomes among patients with metastatic breast cancer: Review of 47,000 patients. *Ann Surg Oncol* 28:7441-7449, 2021
3. Clark GM, Sledge GW Jr., Osborne CK, et al: Survival from first recurrence: Relative importance of prognostic factors in 1,015 breast cancer patients. *J Clin Oncol* 5:55-61, 1987
4. Yamamoto N, Watanabe T, Katsumata N, et al: Construction and validation of a practical prognostic index for patients with metastatic breast cancer. *J Clin Oncol* 16:2401-2408, 1998
5. Barcenas CH, Song J, Murthy RK, et al: Prognostic model for de novo and recurrent metastatic breast cancer. *JCO Clin Cancer Inform* 5:789-804, 2021
6. He ZY, Lian CL, Wang J, et al: Incorporation of biologic factors for the staging of de novo stage IV breast cancer. *NPJ Breast Cancer* 6:43, 2020
7. Chavez-MacGregor M, Mittendorf EA, Clarke CA, et al: Incorporating tumor characteristics to the American Joint Committee on Cancer breast cancer staging system. *Oncologist* 22:1292-1300, 2017
8. AJCC Cancer Staging Manual (ed 8th). New York, NY, Springer International Publishing, 2016
9. Giuliano AE, Connolly JL, Edge SB, et al: Breast Cancer—major changes in the American Joint Committee on Cancer eighth edition cancer staging manual. *CA Cancer J Clin* 67:290-303, 2017
10. Weiss A, Chavez-MacGregor M, Lichtensztajn DY, et al: Validation study of the American Joint Committee on cancer eighth edition prognostic stage compared with the anatomic stage in breast cancer. *JAMA Oncol* 4:203-209, 2018
11. Plichta JK, Ren Y, Thomas SM, et al: Implications for breast cancer restaging based on the 8th edition AJCC staging manual. *Ann Surg* 271:169-176, 2020
12. Leone BA, Vallejo CT, Romero AO, et al: Prognostic impact of metastatic pattern in stage IV breast cancer at initial diagnosis. *Breast Cancer Res Treat* 161:537-548, 2017
13. Press DJ, Miller ME, Liederbach E, et al: De novo metastasis in breast cancer: Occurrence and overall survival stratified by molecular subtype. *Clin Exp Metastasis* 34:457-465, 2017
14. Mallin K, Browner A, Palis B, et al: Incident cases captured in the National Cancer Database compared with those in U.S. population based central cancer registries in 2012-2014. *Ann Surg Oncol* 26:1604-1612, 2019
15. Strobl C, Malley J, Tutz G: An introduction to recursive partitioning: Rationale, application, and characteristics of classification and regression trees, bagging, and random forests. *Psychol Methods* 14:323-348, 2009
16. Lankham I, Slaughter M: Simple and Efficient Bootstrap Validation of Predictive Models Using SAS/STAT Software, SAS Global Forum 2020. <https://www.sas.com/content/dam/SAS/support/en/sas-global-forum-proceedings/2020/4647-2020.pdf>
17. Manual for Staging of Cancer (ed 1). Philadelphia, PA, Lippincott-Raven, 1977, pp 101-108
18. Lobbezoo DJA, van Kampen RJW, Voogd AC, et al: Prognosis of metastatic breast cancer subtypes: The hormone receptor/HER2-positive subtype is associated with the most favorable outcome. *Breast Cancer Res Treat* 141:507-514, 2013
19. Tao L, Chu L, Wang LI, et al: Occurrence and outcome of de novo metastatic breast cancer by subtype in a large, diverse population. *Cancer Causes Control* 27:1127-1138, 2016
20. Nguyen B, Fong C, Luthra A, et al: Genomic characterization of metastatic patterns from prospective clinical sequencing of 25,000 patients. *Cell* 185:563-575.e11, 2022
21. Cristofanilli M, Pierga JY, Reuben J, et al: The clinical use of circulating tumor cells (CTCs) enumeration for staging of metastatic breast cancer (MBC): International expert consensus paper. *Crit Rev Oncol Hematol* 134:39-45, 2019
22. Malin JL, Kahn KL, Adams J, et al: Validity of cancer registry data for measuring the quality of breast cancer care. *J Natl Cancer Inst* 94:835-844, 2002
23. Kumar A, Guss ZD, Courtney PT, et al: Evaluation of the use of cancer registry data for comparative effectiveness research. *JAMA Netw Open* 3:e2011985, 2020
24. Zhao W, Wu L, Zhao A, et al: A nomogram for predicting survival in patients with de novo metastatic breast cancer: A population-based study. *BMC Cancer* 20:982, 2020

25. Plichta JK, Thomas SM, Sergesketter AR, et al: A novel staging system for de novo metastatic breast cancer refines prognostic estimates. *Ann Surg* 275:784-792, 2022
 26. Plichta JK, Rushing CN, Lewis HC, et al: Implications of missing data on reported breast cancer mortality. *Breast Cancer Res Treat* 197:177-187, 2023
 27. Wolff AC, Hammond MEH, Allison KH, et al: Human epidermal growth factor receptor 2 testing in breast cancer: American Society of Clinical Oncology/College of American Pathologists clinical practice guideline focused update. *Arch Pathol Lab Med* 142:1364-1382, 2018
-

ASCO® Journals

Save Time with EZSubmit

All ASCO journals have adopted a format-free submission policy (EZSubmit). New submissions are not scrutinized for compliance with our formatting guidelines (reference style, word limits, order of components, etc.).

Visit ascopubs.org



AUTHORS' DISCLOSURES OF POTENTIAL CONFLICTS OF INTEREST**Novel Prognostic Staging System for Patients With De Novo Metastatic Breast Cancer**

The following represents disclosure information provided by authors of this manuscript. All relationships are considered compensated unless otherwise noted. Relationships are self-held unless noted. I = Immediate Family Member, Inst = My Institution. Relationships may not relate to the subject matter of this manuscript. For more information about ASCO's conflict of interest policy, please refer to www.asco.org/rwc or ascopubs.org/jco/authors/author-center.

Open Payments is a public database containing information reported by companies about payments made to US-licensed physicians ([Open Payments](#)).

Daniel F. Hayes

Stock and Other Ownership Interests: InBiomotion

Honoraria: Tempus

Consulting or Advisory Role: Cepheid, Freenome, Epic Sciences, Cellworks, BioVica, Oncocyte, Turnstone Bio, Predictus Biosciences, Guardant Health, L-Nutra, Macrogenics, Tempus, xilis, Exact Sciences

Research Funding: AstraZeneca (Inst), Pfizer (Inst), Menarini Silicon Biosystems (Inst), Cepheid/Danaher (Inst)

Patents, Royalties, Other Intellectual Property: Royalties from licensed technology, Diagnosis and Treatment of Breast Cancer. Patent No. US 8,790,878 B2. Date of Patent: July 29, 2014. Applicant Proprietor: University of Michigan. Dr D.F. Hayes is designated as inventor/co-inventor, Circulating Tumor Cell Capturing Techniques and Devices. Patent No. US 8,951,484 B2. Date of Patent: February 10, 2015. Applicant Proprietor: University of Michigan. Dr D.F. Hayes is designated as inventor/co-inventor, Title: A Method for Predicting Progression Free and Overall Survival at Each Follow-Up Timepoint During Therapy of Metastatic Breast Cancer Patients Using Circulating Tumor Cells. Patent No. 05725638.0-1223-US2005008602

Other Relationship: Menarini, UpToDate

Uncompensated Relationships: UpToDate

Mariana Chavez-MacGregor

Employment: MD Anderson Physicians Network

Consulting or Advisory Role: Exact Sciences, AstraZeneca/Daiichi Sankyo, Pfizer, Abbott Laboratories, Exact Sciences, Pfizer, Lilly, AstraZeneca

Research Funding: Novartis (Inst), Genentech/Roche (Inst), Pfizer (Inst), Lilly

Expert Testimony: Pfizer, AstraZeneca, Genentech, Lilly

Travel, Accommodations, Expenses: AstraZeneca

Uncompensated Relationships: Legacy Healthcare Services, The Hope Foundation

Kimberly Allison

Consulting or Advisory Role: Mammotome

Expert Testimony: Kaiser Permanente

Amy M. Fowler

Consulting or Advisory Role: GE Healthcare

Patents, Royalties, Other Intellectual Property: Book chapter royalty from Elsevier, Inc

Armando E. Giuliano

Stock and Other Ownership Interests: Black Light Surgical, Inc

Consulting or Advisory Role: Merck, Sharp & Dohme Corp

Priyanka Sharma

Stock and Other Ownership Interests: Amgen Astellas BioPharma

Consulting or Advisory Role: Novartis, Merck, AstraZeneca, Pfizer, Gilead Sciences, GlaxoSmithKline, Sanofi

Research Funding: Novartis (Inst), Bristol Myers Squibb (Inst), Merck (Inst), Novartis, Gilead Sciences

Patents, Royalties, Other Intellectual Property: UpToDate

Benjamin D. Smith

Patents, Royalties, Other Intellectual Property: Through my employer, I have an equity interest in Oncora Medical as part of a partnership agreement (Inst)

Other Relationship: Varian Medical Systems

Uncompensated Relationships: American Society for Radiation Oncology

Open Payments Link: <https://openpaymentsdata.cms.gov/physician/730600>

Elizabeth van Eycken

Employment: MSD Belgium

Stock and Other Ownership Interests: MSD Belgium

Stephen B. Edge

Honoraria: North American Center for Continuing Medical Education, North American Center for Continuing Medical Education

Gabriel N. Hortobagyi

Consulting or Advisory Role: Novartis, Seagan, Blueprint Medicines, AstraZeneca

Research Funding: Novartis (Inst)

No other potential conflicts of interest were reported.