Modeling effective tumor burden of primary lesion and metastatic lymph node in breast cancer patients from the SEER database

Taobo Hu[1#^], Jinbo Wu[1#], Mengping Long[2#^], Xiaojie Zhou[3], Shu Wang[1^]

[1]Department of Breast Surgery, Peking University People's Hospital, Beijing, China; [2]Department of Pathology, Peking University Cancer Hospital, Beijing, China; [3]Hunan Provincial Key Laboratory of Intelligent Manufacturing Technology for High-Performance Mechanical Equipment, Changsha University of Science and Technology, Changsha, China

Contributions: (I) Conception and design: S Wang; (II) Administrative support: S Wang, T Hu; (III) Provision of study materials or patients: T Hu, J Wu, M Long; (IV) Collection and assembly of data: T Hu, J Wu, M Long, X Zhou; (V) Data analysis and interpretation: All authors; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

[#]These authors contributed equally to this work.

Correspondence to: Shu Wang. Department of Breast Surgery, Peking University People's Hospital, 11 Xizhimen South Street, Beijing 100044, China. Email: shuwang@pkuph.edu.cn.

Background: Characterizing the role of the primary tumor and axillary lymph nodes (ALNs) in metastasis progression can provide a rational basis for effective local control and systemic treatments.

Methods: We employed data from the Surveillance, Epidemiology, and End Results (SEER) database to perform survival analysis and modeling to estimate breast cancer (BC) progression in both untreated and treated patients. Effective tumor burden was defined as the number of cancer cells that can lead to distant metastasis. Quantified analysis was conducted for primary lesion and metastatic nodes, respectively. A tumor growth model was built based on mathematical modeling: $N(t) = mN_T(t) + nN_N(t)$. In this model, the parameters "m" and "n" represent the contributions of the primary lesion and metastatic lymph nodes (LNs) to distant metastasis, respectively.

Results: We found that both the primary lesion and the metastatic LNs contribute to the effective tumor burden. Through fitting the built tumor growth model, "m" in treated groups was determined to be 1, and "n" was determined to be 1.5; while in the untreated group, "m" was determined to be 1, and "n" was 0.15, which was 10 times smaller than in treated groups.

Conclusions: Our study revealed that the prognostic value of the anatomic stage in BC progression varied in different historical periods, due to the development of oncological treatments. While tumor size was a significant prognostic factor in both untreated and treated patients, the prognostic value of nodal status was only significant in patients who received locoregional treatment and systemic treatment.

Keywords: Breast cancer (BC); natural history; anatomic stages; mathematical modeling; metastasis

Introduction

Breast cancer (BC) is the most common malignancy in women, and distant metastasis is the eventual cause of breast cancer-specific death (BCSD) in most BC patients (1).

Prediction and prevention of distant metastasis rely on our biological perception of BC, including knowledge of its molecular and clinical characteristics and progression.

In the past two decades, our understanding of BC biology has undergone tremendous changes (2), both physicians and

---

researchers are paying more attention to molecular markers for the diagnosis and treatment of BC, while the importance of traditional anatomic staging is less frequently discussed. However, it remains unclear whether the anatomic stage at BC progression might have shifted over time due to continuous improvements in oncological treatments. The role and relationship of primary lesion and node metastasis in cancer progression have also been the subject of decades-long debate.

To address these issues, in this study, we used the Surveillance, Epidemiology, and End Results (SEER) database to analyze the clinicopathological characteristics and survival data in three groups of patients, i.e., untreated BC patients, BC patients who had received locoregional treatment without systemic treatment, and BC patients who had received locoregional treatment and systemic treatment. Mathematical models were built to explain the differences between the three groups. The prognostic value of tumor size and lymph node (LN) status among these three groups were compared, and mathematical models quantifying effective tumor burden were established. We present the following article in accordance with the STROBE reporting checklist (available at https://gs.amegroups.com/article/view/10.21037/gs-21-919/rc).

Methods

Data source

The SEER 18 registries research database (November 2018 submission) was used for the analysis, including patients diagnosed from 1975 to 2016 and covering approximately 27.8% of the U.S. population. For estrogen receptor (ER) status, ER$^+$ status was defined as having either positive or borderline ER. The human epidermal growth factor receptor 2 (HER2) receptor status of the breast tumors was not included in our analyses, as this information was only included in cases after 2010. The study was conducted in accordance with the Declaration of Helsinki (as revised in 2013).

Survival analysis

Overall survival (OS) was defined as the interval from diagnosis to death due to any cause (including BC) or the date of the last follow-up. The Kaplan-Meier method was used to estimate the survival outcomes of all patients according to different categories, and groups were compared using the log-rank statistic. The OS was calculated from the date of the latest diagnosis of BC to the date of death. Breast cancer-specific survival (BCSS) was defined and calculated as the survival time from the date of BC diagnosis to the date of death due to BC.

Statistical analysis

Student's *t*-test, chi-square test, and log-rank test were used for statistical analysis. All P values were two-sided, with statistical significance indicated by P values less than 0.05. All statistical analyses and case selection were performed in R (version 3.6.0; R Foundation for Statistical Computing, Vienna, Austria). Survival analyses were performed using the "survival" package (version 2.41) (3).

Mathematical modeling

MATLAB (MathWorks, Natick, MA, USA) and R software were used to build models of effective tumor burden and time. The estimations of the number of tumor cells and effective tumor burden are listed in Table S1. The estimation of "time" was calculated by subtracting the specific survival time of each patient from the longest survival time in the same group, which could also be taken as the "life loss" of the patient. For the step-by-step explanation of the mathematical modeling process, please refer to the "Mathematical modeling of the effective tumor burden of BC" section in the "Results".

Results

Patient cohort

A total of 763,873 individual BC patients diagnosed between 1975 and 2016 were identified from 840,660 BC entries in the SEER database. Of these, 508,058 female patients had complete information on tumor staging, surgery, radiotherapy, chemotherapy, and prognosis, and had only 1 primary tumor occurring in their lifetime. Among these patients, 12,687 did not receive surgical treatment, radiotherapy, or chemotherapy and were included in our study as "Group I" or the untreated group; 258,268 patients had undergone either mastectomy or lumpectomy surgery, with or without radiotherapy, but did not receive chemotherapy and were included as "Group II" or the locoregional treatment group; 120,441 of the patients who had received both surgical treatment and chemotherapy were included as "Group III" or the systemic treatment group (*Figure 1*). Information on hormonal therapy and targeted therapy was not provided.
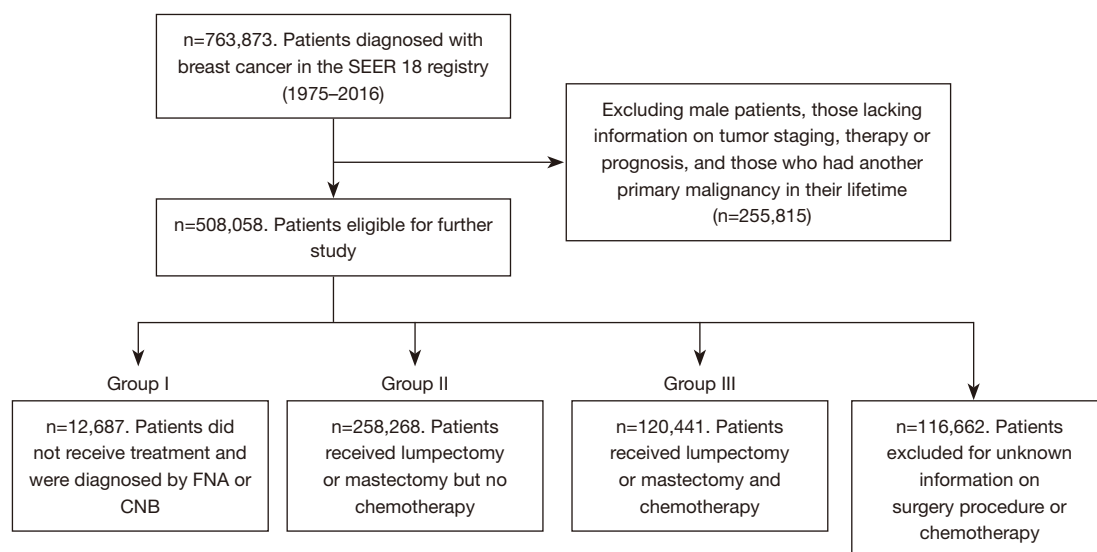
Figure 1 Flowchart of the patient selection process. SEER, Surveillance, Epidemiology, and End Results; FNA, fine-needle aspiration; CNB, core needle biopsy.

Characteristics of the patients enrolled in the study

The T and N stages of each patient were based on the best available pathological information. When pathological information was not available, clinical information was used. The clinicopathological characteristics of Group I patients are summarized in *Table 1*. Among them, 1,390 (11.0%) patients had stage I BC, 1,702 (13.4%) had stage II BC, 1,096 (8.6%) had stage III BC, and 6,265 (49.4%) had stage IV BC. Among the 7,871 (62%) patients with known ER status, 84.6% were ER-positive. The clinicopathological characteristics of Groups II and III patients are also listed in *Table 1*.

Mathematical modeling of the effective tumor burden of BC

To have an objective view of the natural history of BC, a mathematical model was built to evaluate the natural progression. Functions were built with two variables: one was the effective tumor burden "N(t)", which indicates the number of cancer cells that can lead to distant metastasis; the other was time "t", which was derived from the estimated life loss of patients. The effective tumor burden of the primary lesion and metastatic LNs was estimated from the total number of tumor cells and is listed in Table S1. Since the primary lesion and the metastatic LNs contribute to the effective tumor burden,

the functions were built based on: $N(t) = mN_T(t) + nN_N(t)$. The parameters "m" and "n" in the equation represent the contributions of the primary lesion and metastatic LNs to distant metastasis, respectively, which can also be called "relative metastagenicity". The word "metastagenicity" was first introduced by Dr. Heimann to describe the metastatic proclivity of a tumor (4,5). Here, we used it to represent the intrinsic metastatic proclivity of a specific population of cancer cells. For Groups II and III patients, the following function could be built.

$$N(t) = mN_T(t) + nN_N(t) = a \cdot e^{k \cdot (1+b \cdot t)} \quad [1]$$

The fitting curves are shown in *Figure 2*. Through fitting, "m" in Groups II and III was determined to be 1, and "n" in the two groups was determined to be 1.5.

For Group I patients, since the primary lesion and the affected LN are not removed and are thus constantly contributing to the metastasis, it is logical that the total effective tumor number N(t) should be the integral of "t". Thus, we integrated the previous function and found that the data in Group I could be fitted using the integral function.

$$N'(t) = mN'_T(t) + nN'_N(t) = \int_0^t a \cdot e^{k \cdot (1+b \cdot t)} = a \cdot e^k \cdot \frac{e^{b \cdot k \cdot t} - 1}{b \cdot k} = a' \cdot \left(k' \cdot e^{b' \cdot t} - 1 \right) \quad [2]$$

Moreover, through fitting, "m" in this group was determined to be 1, and "n" was 0.15, which was 10 times smaller than the value in Groups II and III.

Table 1 Clinicopathological characteristics of the patient cohort including Groups I, II, and III patients

Characteristics	Group I, n (%)	Group II, n (%)	Group III, n (%)	Total, n (%)
Patients, n	12,687	258,268	120,441	391,396
Age, y				
<40	329 (2.6)	7,771 (3.0)	14,851 (12.3)	22,951 (5.9)
40–54	2,322 (18.3)	70,236 (27.2)	55,641 (46.2)	128,199 (32.8)
55–69	3,665 (28.9)	92,940 (36.0)	40,009 (33.2)	136,614 (34.9)
≥70	6,367 (50.2)	87,319 (33.8)	9,938 (8.3)	103,624 (26.5)
Ethnicity				
Caucasian	9,648 (76.0)	213,084 (82.5)	93,588 (77.7)	316,320 (80.8)
Asian	744 (5.9)	21,187 (8.2)	10,957 (9.1)	32,888 (8.4)
Black	1,970 (15.5)	21,070 (8.2)	14,175 (11.8)	37,215 (9.5)
Other	325 (2.6)	2,927 (1.1)	1,721 (1.4)	4,973 (1.3)
ER status				
Positive	6,661 (52.5)	172,625 (66.8)	75,551 (62.7)	254,837 (65.1)
Negative	1,210 (9.5)	24,946 (9.7)	36,263 (30.1)	62,419 (15.9)
Unknown	4,816 (38.0)	60,697 (23.5)	8,627 (7.2)	74,140 (18.9)
PR status				
Positive	5,366 (42.3)	148,693 (57.6)	64,441 (53.5)	218,500 (55.8)
Negative	2,320 (18.3)	44,351 (17.2)	46,629 (38.7)	93,300 (23.8)
Unknown	5,001 (39.4)	65,224 (25.3)	9,371 (7.8)	79,596 (20.3)
Tumor grade*				
I	1,204 (9.5)	54,821 (21.2)	9,540 (7.9)	65,565 (16.8)
II	3,208 (25.3)	93,472 (36.2)	41,002 (34.0)	137,682 (35.2)
III	2,584 (20.4)	56,105 (21.7)	59,107 (49.1)	117,796 (30.1)
IV	173 (1.4)	6,586 (2.6)	2,230 (1.9)	8,989 (2.3)
Unknown	5,518 (43.5)	47,284 (18.3)	8,562 (7.1)	61,364 (15.7)
Staging				
Stage 0	2,234 (17.6)	71,169 (27.6)	407 (0.3)	73,810 (18.9)
Stage I	1,390 (11.0)	122,848 (47.6)	28,548 (23.7)	152,786 (39.0)
Stage II	1,702 (13.4)	50,339 (19.5)	58,253 (48.4)	110,294 (28.2)
Stage III	1,096 (8.6)	10,999 (4.3)	29,364 (24.4)	41,459 (10.6)
Stage IV	6,265 (49.4)	2,913 (1.1)	3,869 (3.2)	13,047 (3.3)

*, based on the *WHO International Classification of Diseases for Oncology, Second Edition*. Grade I represents well-differentiated tumors; grade II represents moderately differentiated tumors; grade III represents poorly differentiated tumors, and grade IV represents undifferentiated tumors. ER, estrogen receptor; PR, progesterone receptor.
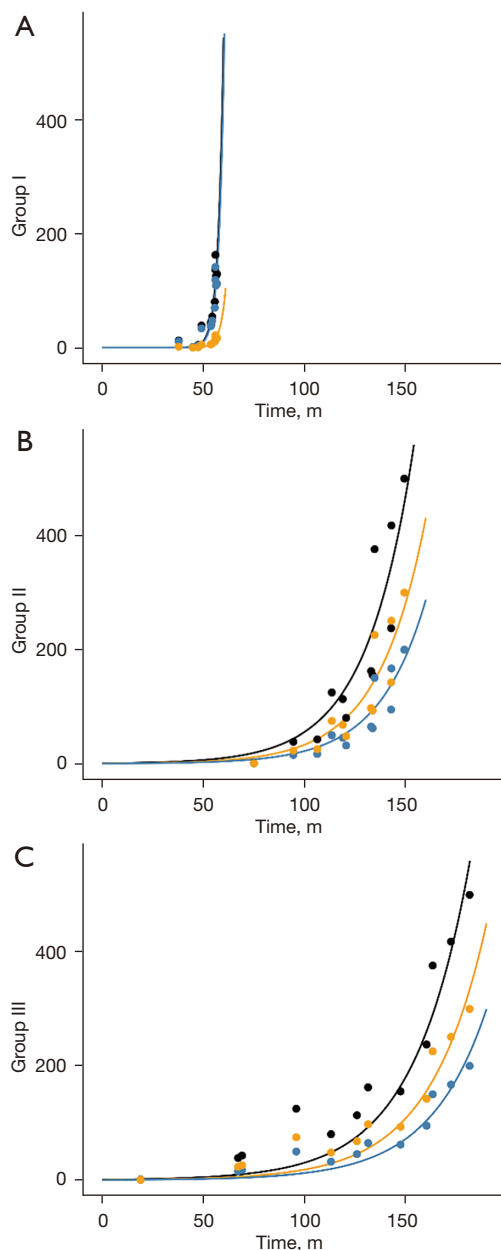
Figure 2 Mathematical modeling of cancer progression in three groups. Curve fitting and resulting models between "effective tumor burden (Y axis)" and "time (X axis)" in the three groups. (A-C) Models were built for the three groups of patients with R^2 values all above 0.8. Black lines represent the total effective tumor burden while the burden of primary tumor and LN was represented by blue and orange lines respectively. In Group I, the fitted model was the integral of the functions of Groups II and III. LN, lymph node.

Comparisons of survival by T and N status in the untreated, locoregionally treated, and systemically treated groups

To observe the direct effect of tumor size and LN status on prognosis, survival analysis was conducted for untreated patients by tumor size, clinical LN status, and clinical stage. The survival rate decreased gradually as the clinical stage increased (Figure S1). The median survival time of untreated patients decreased from 8.0 years for stage I patients to 0.5 years for stage IV patients. Certain patients achieved a long survival time even while untreated, which has been reported previously (6). When the untreated patients were stratified by tumor size (T), the survival rate markedly decreased as the T stage increased from T1 to T3 ($P<0.0001$) (Figure S2). However, when survival was compared among the three nodal statuses (N1, N2, and N3) with a fixed T stage, there were no significant differences among the N stages within each T stage, although minor differences were noted (Figure S3A-S3C). When survival was compared between negative LN status (N0) and positive LN status (N1–N3, denoted N+), no significant differences were seen among any of the three T stages (Figure S4A-S4C). Overall, we found that in untreated patients, the prognostic value of tumor size was significant, while the prognostic value of nodal status was not.

We then explored the prognostic value of tumor size and LN status for survival in patients who received treatment. Unsurprisingly, in Groups II and III, tumor size and nodal status both had significant prognostic value for survival (Figure S5A-S5F). Moreover, the differences in the median survival time among N stages were more significant in Group III than in Group II, as indicated by the wider interval between the dotted lines for Group III in Figure S5A-S5F.

Comparisons of the estimated survival time (EST) by T and N stages within and among the three groups of patients

The above comparisons were carried out between pairs of survival curves. To obtain a more direct and quantified result, the EST for each specific T and N stages was calculated (*Figure 3*). An accelerated failure time parametric model was used for the calculation. Consistent with previous findings, in Group I patients, the EST showed no significant differences as the N stage increased from N0 to
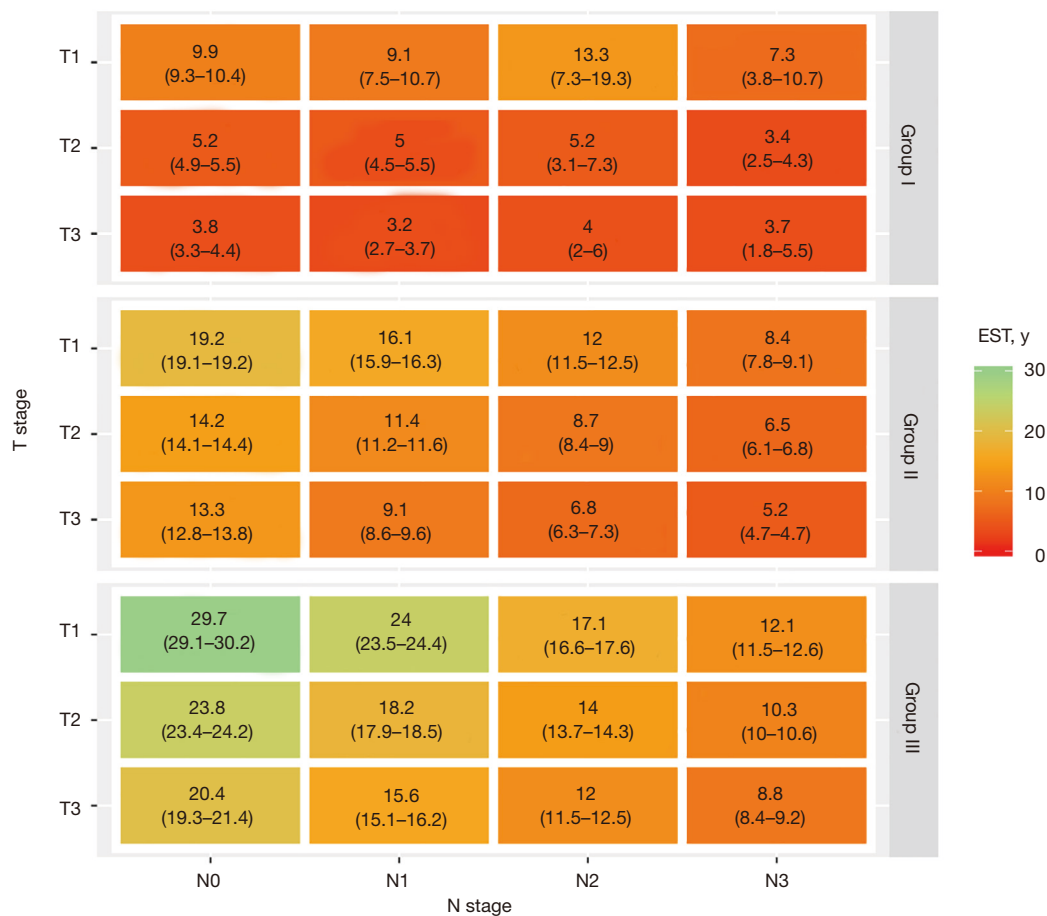
Figure 3 Heatmaps of the EST in each T and N stages of three groups. The estimated survival duration is represented by a color scale with green indicating the longest duration and red indicating the shortest duration. The numbers in each cell are the specific estimated survival duration and its 95% CI. EST, estimated survival time; CI, confidence interval.

N3 when the T stage was fixed while in Groups II and III, the difference was obvious. The differences in the effect of T stage on EST were obvious in all columns and among all groups.

Discussion

In this study, we analyzed the natural progression of BC in untreated and treated patients to explain the role of traditional anatomic staging, especially node staging, in modern oncology. Our data suggest that nodal metastasis has a complex effect on prognosis that can be altered by treatment, while tumor size was a significant prognostic factor in both untreated and treated patients, the prognostic value of nodal status was much higher in patients who received locoregional treatment and systemic treatment.

Whether nodal metastasis is simply a natural course of BC (7) or whether it represents a more aggressive phenotype (8) has long been a topic of discussion, and both supportive and contradictory evidence exists. Our study did not yield direct evidence to show whether the effective tumor burden associated with LN metastasis planted the distant metastasis directly. However, our mathematical modeling showed the different effective tumor burdens in treated and untreated patients and thus can also guide the staging, treatment, and prognostic prediction of BC. The staging system of BC should change with the improvement of treatment methods. For example, in the process of curve fitting, we found that in Group III patients, the American Joint Committee on Cancer (AJCC) 8th anatomic stages could be adjusted by downgrading T1N2M0 from stage IIIA to IIB and upgrading T3N2M0 from stage IIIA to IIIB
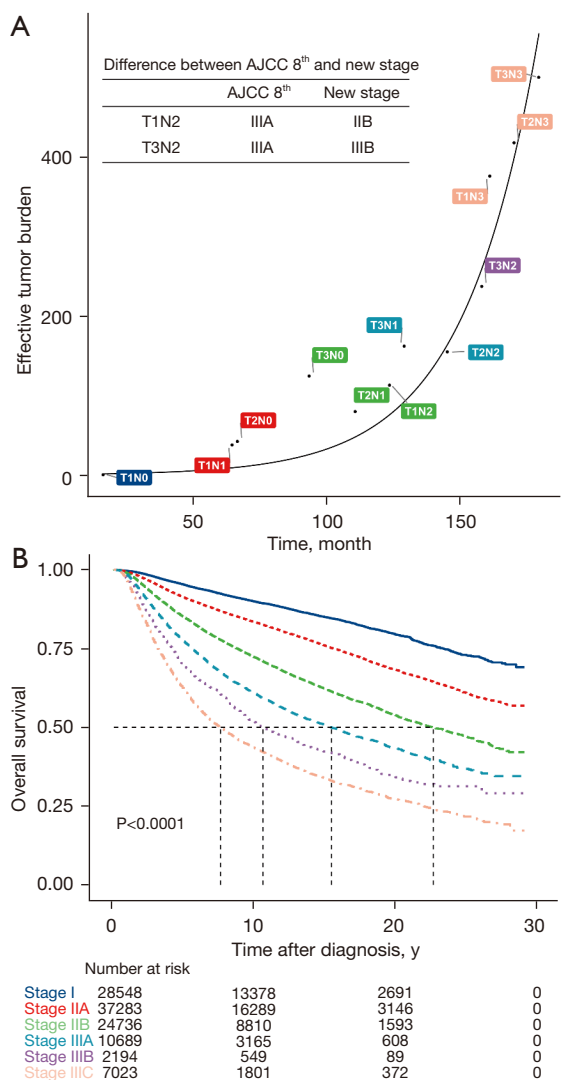
Figure 4 Curve fitting and new staging system of each TN status in Group III patients. (A) Curve fitting for "effective tumor burden" and "time" in Group III patients with each data point indicated. (B) The OS curve plot based on the new staging system with T1N2 downgraded to IIB and T3N2 upgraded to IIIB. OS, overall survival; AJCC, American Joint Committee on Cancer.

to achieve more evenly distributed clustering (*Figure 4A*). An additional adjustment was made by upgrading stage IIIB in the AJCC 8th anatomic stages to stage IIIC. These adjustments made the survival curve of the systemically treated patients more evenly distributed (*Figure 4B*).

Although LN metastasis is a robust prognostic factor in BC (8), there has been decades-long debate regarding its role in cancer progression (9). One hypothesis is that LN

metastasis is a natural step of disease progression, cancer cells can escape the LN and colonize distant organs, and LNs are way stations for metastatic seeding (9). Numerous studies have shown that removal of metastatic LNs or radiotherapy can improve the prognosis; on the contrary, genomic analyses have revealed that synchronous axillary lymph node (ALN) metastasis was not involved in seeding the distant metastasis (10). The NSABP B04 study (11) showed that late resection of LN metastasis did not affect the prognosis, thus de-escalation of the axilla is feasible which is also one of the foundations of the ACOSOG Z0011 study (12). Our research showed that the effective tumor burden associated with LN metastasis was very important, even outweighing the size of the primary tumor in patients who received locoregional treatment and systemic treatment. It was suggested that in modern BC management, the de-escalation of local axillary treatment should be cautious. Although de-escalation is popular now, axilla with low-load metastasis can also consider axillary preservation. We must note that the International Breast Cancer Study Group Trial 23-01 (IBCSG 23-01) clinical trial (13), which is consistent with the ACOSOG Z0011 trial (12) and AMAROS trial (14), involving patients with limited LN metastasis involvement showed no disease-free survival benefit in patients that underwent axillary lymph node dissection (ALND) compared with those that did not. However, all of these patients received traditional systemic adjuvant therapy and radiation that potentially eliminated any residual disease in the LNs. Thus, radiation and systemic therapies may be sufficient to control nodal disease, making ALND unnecessary for BC patients with limited LN involvement, which is different from the de-escalation in the SLN-NSABP B32 study (15). We analyzed N0 patients who did not have ALND from the SEER database and found that the patients who underwent ALND had better survival prognosis. The possible reason is that in clinically LN-negative patients, 30% of them have positive LNs, and these patients will benefit from ALND.

It is widely accepted that the clearance of involved LNs improves local disease control, but whether it influences systemic dissemination is less clear. Studies have shown that LN resection and radiotherapy are valuable for prognosis, but the NSABP B04 trial showed that late resection did not affect prognosis (11,16). Although de-escalation of axillary surgery is fairly common, for patients with negative LN status (N0), many studies have stated that ALND is not beneficial for sentinel LN-negative patients (17-19). Our mathematical modeling results suggest that nodal metastasis

is not merely a straightforward process of BC, but has a complex effect on prognosis that can also be altered by treatment. It could be that there is positive feedback between T and N since a larger number of tumor cells in T can promote the metastasis to N, and the metastasized tumor cells in the LN may induce immunoevasion in the tumor. This positive feedback could be the reason for the rapid increase in tumor burden in Group I patients, which can be eradicated by locoregional and/or systematic treatment, as shown in Groups II and III patients. This could be based on the hypothesis that the presence of a primary tumor in untreated patients could constantly initiate immune attacks against metastatic LNs, thereby suppressing the metastagenicity of LNs (20). Or based on the alteration in the tumor microenvironment since the microenvironment has been known to contribute greatly to metastasis (20-25).

Although we followed the checklist of items for observational research described in the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement (26), there were several potential limitations to our study. First, the use of the SEER database instead of cohort study data and the lack of exact treatment regimens may have limited the strength of evidence. Second, since the data source came from a single country, analysis of data from another country or database may be necessary to further confirm our findings. Last, regarding the analysis of molecular traits, our findings were limited by the incompleteness of the molecular subtyping, as the information of HER2 status and *BRCA1/2* gene mutations were unknown.

Conclusions

Our study found that while tumor size was a significant prognostic factor in both untreated and treated patients, the prognostic value of nodal status was only significant in treated patients which strengthened the prognostic role of LN status in modern medicine. Our results indicated that in the modern era, the predictive value of nodal status should be further emphasized and de-escalating the local treatment of nodes needs cautious evaluation.

Acknowledgments

Footnote

Reporting Checklist: The authors have completed the STROBE reporting checklist. Available at https://gs.amegroups.com/article/view/10.21037/gs-21-919/rc

Conflicts of Interest: All authors have completed the ICMJE uniform disclosure form (available at https://gs.amegroups.com/article/view/10.21037/gs-21-919/coif). The authors have no conflicts of interest to declare.

Ethical Statement: The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. The study was conducted in accordance with the Declaration of Helsinki (as revised in 2013).

References

1. Weigelt B, Glas AM, Wessels LF, et al. Gene expression profiles of primary breast tumors maintained in distant metastases. Proc Natl Acad Sci U S A 2003;100:15901-5.
2. Harbeck N, Penault-Llorca F, Cortes J, et al. Breast cancer. Nat Rev Dis Primers 2019;5:66.
3. Therneau TM, Grambsch PM. Modeling Survival Data: Extending the Cox Model. New York: Springer, 2000.
4. Heimann R, Hellman S. Clinical progression of breast cancer malignant behavior: what to expect and when to expect it. J Clin Oncol 2000;18:591-9.

5. Heimann R, Ferguson D, Recant WM, et al. Breast cancer metastatic phenotype as predicted by histologic tumor markers. Cancer J Sci Am 1997;3:224-9.

6. Steckler RM, Martin RG. Prolonged survival in untreated breast cancer. Am J Surg 1973;126:111-3.

7. Mittra I, MacRae KD. A meta-analysis of reported correlations between prognostic factors in breast cancer: does axillary lymph node metastasis represent biology or chronology? Eur J Cancer 1991;27:1574-83.

8. Jatoi I, Hilsenbeck SG, Clark GM, et al. Significance of axillary lymph node metastasis in primary breast cancer. J Clin Oncol 1999;17:2334-40.

9. Zhou H, Lei PJ, Padera TP. Progression of Metastasis through Lymphatic System. Cells 2021;10:627.

10. Ullah I, Karthik GM, Alkodsi A, et al. Evolutionary history of metastatic breast cancer reveals minimal seeding from axillary lymph nodes. J Clin Invest 2018;128:1355-70.

11. Fisher B, Redmond C, Fisher ER, et al. Ten-year results of a randomized clinical trial comparing radical mastectomy and total mastectomy with or without radiation. N Engl J Med 1985;312:674-81.

12. Guidi AJ, Berry DA, Broadwater G, et al. Association of angiogenesis in lymph node metastases with outcome of breast cancer. J Natl Cancer Inst 2000;92:486-92.

13. Galimberti V, Cole BF, Zurrida S, et al. Axillary dissection versus no axillary dissection in patients with sentinel-node micrometastases (IBCSG 23-01): a phase 3 randomised controlled trial. Lancet Oncol 2013;14:297-305.

14. Donker M, van Tienhoven G, Straver ME, et al. Radiotherapy or surgery of the axilla after a positive sentinel node in breast cancer (EORTC 10981-22023 AMAROS): a olonizati, multicentre, open-label, phase 3 non-inferiority trial. Lancet Oncol 2014;15:1303-10.

15. Krag DN, Anderson SJ, Julian TB, et al. Sentinel-lymph-node resection compared with conventional axillary-lymph-node dissection in clinically node-negative patients with breast cancer: overall survival findings from the NSABP B-32 randomised phase 3 trial. Lancet Oncol 2010;11:927-33.

16. Wolmark N, Fisher B. Surgery in the primary treatment of breast cancer. Breast Cancer Res Treat 1981;1:339-48.

17. Piltin MA, Hoskin TL, Day CN, et al. Oncologic Outcomes of Sentinel Lymph Node Surgery After Neoadjuvant Chemotherapy for Node-Positive Breast Cancer. Ann Surg Oncol 2020;27:4795-801.

18. Gong J, Yu Y, Wu G, et al. Should internal mammary lymph node sentinel biopsy be performed in breast cancer: a systematic review and meta-analysis. World J Surg Oncol 2019;17:135.

19. Lyman GH, Somerfield MR, Bosserman LD, et al. Sentinel Lymph Node Biopsy for Patients With Early-Stage Breast Cancer: American Society of Clinical Oncology Clinical Practice Guideline Update. J Clin Oncol 2017;35:561-4.

20. Garner H, de Visser KE. Immune crosstalk in cancer progression and metastatic spread: a complex conversation. Nat Rev Immunol 2020;20:483-97.

21. Costa A, Kieffer Y, Scholer-Dahirel A, et al. Fibroblast Heterogeneity and Immunosuppressive Environment in Human Breast Cancer. Cancer Cell 2018;33:463-479.e10.

22. Pelon F, Bourachot B, Kieffer Y, et al. Cancer-associated fibroblast heterogeneity in axillary lymph nodes drives metastases in breast cancer through complementary mechanisms. Nat Commun 2020;11:404.

23. Chatterjee G, Pai T, Hardiman T, et al. Molecular patterns of cancer colonisation in lymph nodes of breast cancer patients. Breast Cancer Res 2018;20:143.

24. Jones D, Pereira ER, Padera TP. Growth and Immune Evasion of Lymph Node Metastasis. Front Oncol 2018;8:36.

25. Ogiya R, Niikura N, Kumaki N, et al. Comparison of tumor-infiltrating lymphocytes between primary and metastatic tumors in breast cancer patients. Cancer Sci 2016;107:1730-5.

26. Vandenbroucke JP, von Elm E, Altman DG, et al. Strengthening the Reporting of Observational Studies in Epidemiology (STROBE): explanation and elaboration. PLoS Med 2007;4:e297.

(English Language Editor: J. Jones)

Table S1 Estimation of the number of tumor cells and effective tumor burden for each T and N stages

T/N stages	Number of tumor cells*	Effective tumor burden[#]
T1	1×10^9	1
T2	42.875×10^9	42.875
T3	125×10^9	125
N0	0	0
N1	0.25×10^9	25
N2	0.75×10^9	75
N3	2.5×10^9	250

*, to estimate tumor burden, the number of tumor cells in T1 was estimated as 1×10^9 by assuming a cubic tumor with a size of 1 cm^3. The number of tumor cells in T2 and T3 was estimated to be 42.875×10^9 and 125×10^9, respectively, by assuming tumor sizes of 3.5 and 5 cm. The number of tumor cells in N1 was estimated to be 0.25×10^9 and that in N2 and N3 were estimated as 0.75×10^9 and 2.5×10^9, respectively; #, in the final calculation, effective tumor burden was used instead of the estimated number of tumor cells. Since several studies showed that cancer cells in metastatic LNs harbor higher metastatic capacity than those in primary lesions, the effective tumor burden for LNs was estimated via model fitting to be 100 times the original cell number. LNs, lymph nodes.
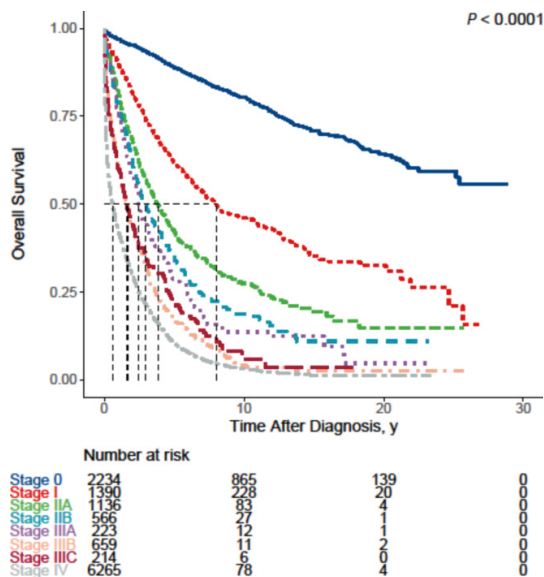
Figure S1 Survival analysis and comparison by clinical stage in untreated patients. The median survival time is marked as a dotted black line.
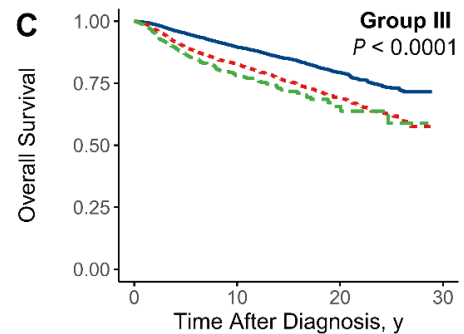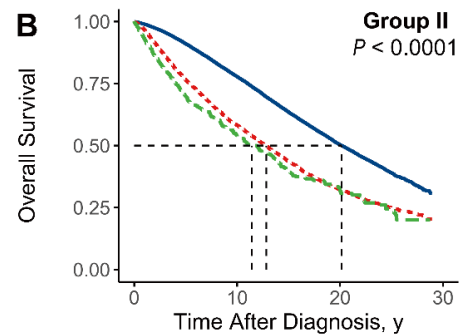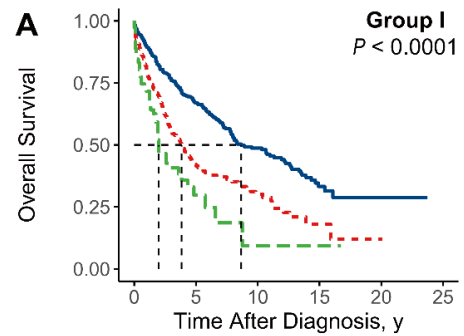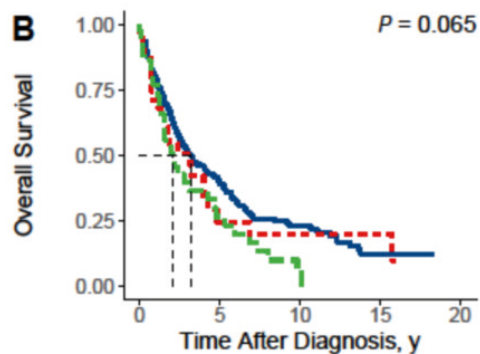

Figure S2 Survival analysis and comparison by T status in three groups of patients. (A) Kaplan-Meier survival curves of untreated patients (Group I) stratified by tumor size. (B) Kaplan-Meier survival curves of patients who received only locoregional treatment (Group II) stratified by tumor size. (C) Kaplan-Meier survival curves of patients who received systemic treatment (Group III) stratified by tumor size. The median survival time is marked as a dotted black line.
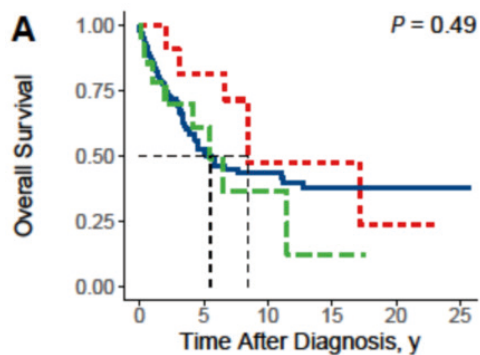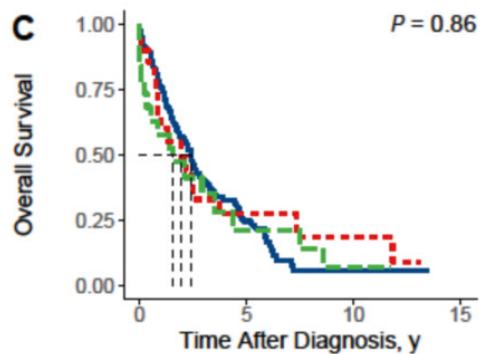
Figure S3 Survival analysis and comparison by nodal status in untreated patients. (A-C) Kaplan-Meier survival curves of untreated patients stratified by nodal status from the T1 to T3 groups. The median survival time is marked as a dotted black line.
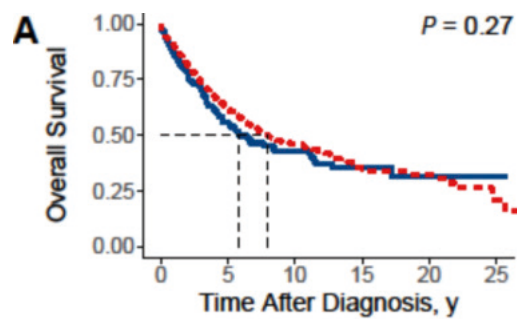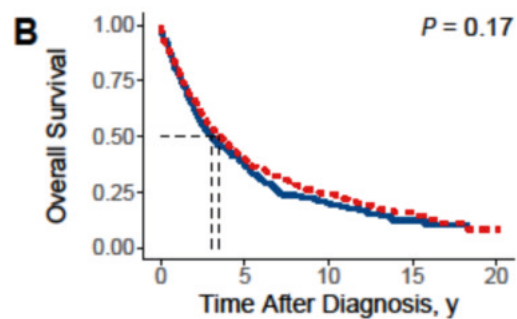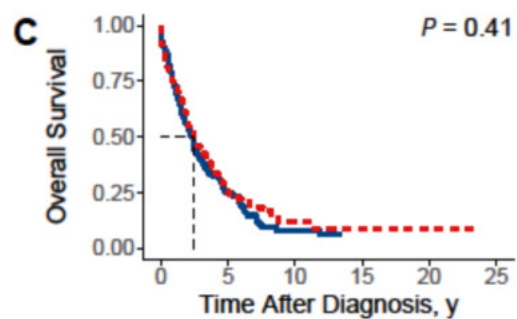
Figure S4 Survival analysis and comparison by N status in untreated patients. (A) Kaplan-Meier survival curves of T1N+(N1–N3)M0 and T1N0M0 patients. (B) Kaplan-Meier survival curves of T2N+M0 and T2N0M0 patients. (C) Kaplan-Meier survival curves of T3N+M0 and T3N0M0 patients. The median survival time is marked as a dotted black line.
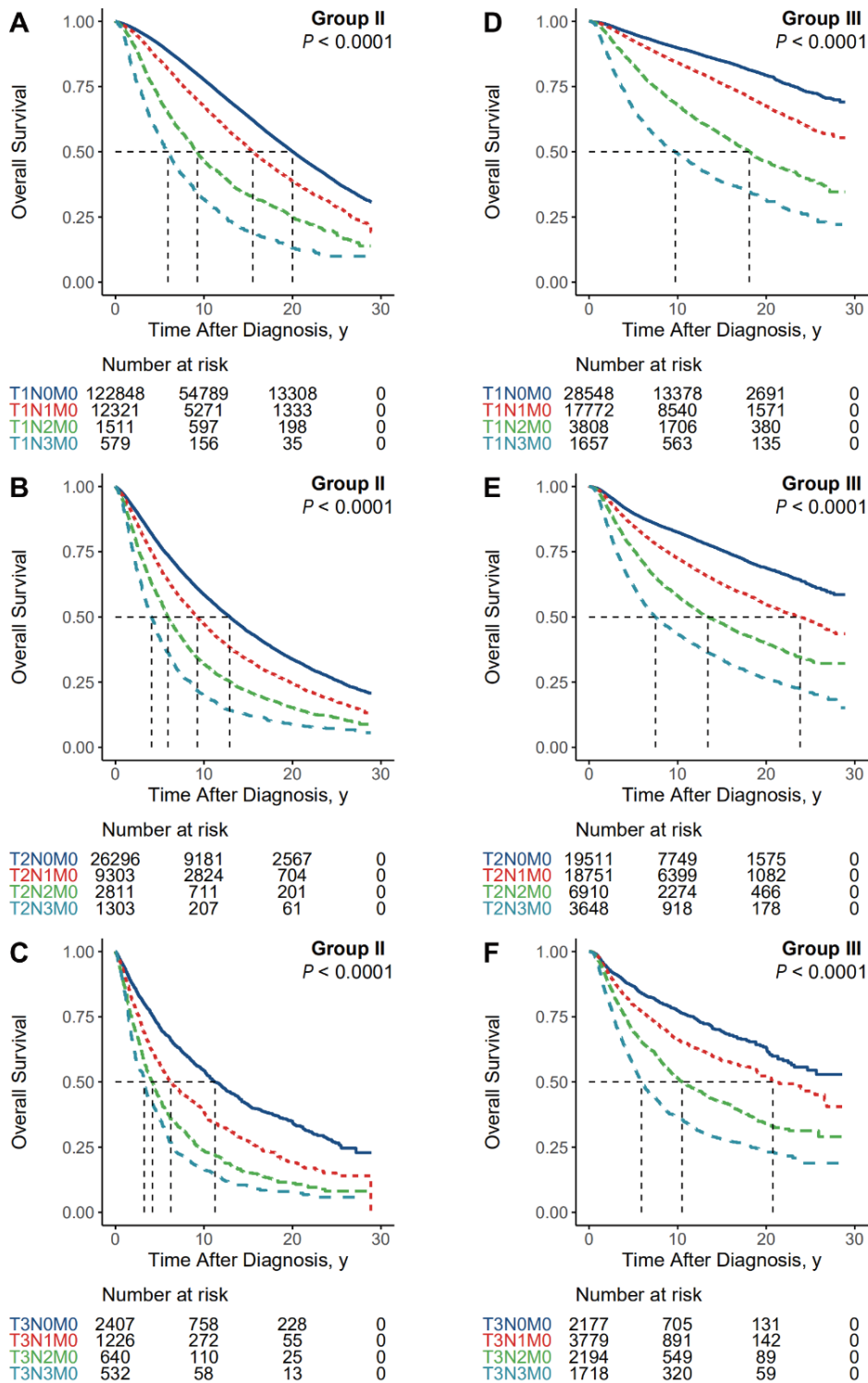
Figure S5 Survival analysis and comparison by N status in patients who received treatment. (A-C) Kaplan-Meier survival curves of patients who received only locoregional treatment (Group II) were stratified by nodal status within each T stage. (D-F) Kaplan-Meier survival curves of patients who received systemic treatment (Group III) stratified by nodal status within each T stage. The median survival time is marked as a dotted black line.