The natural history of breast cancer: a chronological analysis of breast cancer progression using data from the SEER database

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

[1]Department of Breast Disease, 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, M Long, J Wu; (IV) Collection and assembly of data: T Hu, M Long, J Wu, 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 the work.

Correspondence to: Shu Wang. Department of Breast Disease, Peking University People's Hospital, Beijing, China. Email: shuwang@pkuph.edu.cn.

Background: Numerous mathematical models and retrospective analyses have been conducted to explore the natural progression of breast cancer, but there is no precise timeline as to how long it takes breast cancer to progress from stages 0 to IV. Studying the natural history of breast cancer requires the follow-up of a large cohort of patients who received no treatment following diagnosis for a long period, but this has become unpractical in the modern era. In this study, we quantified the natural progression of breast cancer using the Surveillance, Epidemiology, and End Results (SEER) database which was collected the clinical-pathological characteristics and survival data of untreated breast cancer patients.

Methods: We included 12,687 female patients who did not receive surgical treatment, radiotherapy, or chemotherapy, had complete information on tumor staging, and had only 1 primary tumor in their lifetime in our study as the untreated group. Overall survival (OS) was calculated from the date of the latest diagnosis of breast cancer to the date of death. Medium survival time (MST) was defined as the time at which 50% of the patients reached the endpoint. The medium survival time (MST) was calculated for each stage. The progression time was calculated by subtracting the MST of a higher stage from that of a lower stage, which also indicated "life loss". Estrogen receptor (ER) negative and positive patients were analyzed as two separate groups.

Results: For the ER positive patients, the median progression times from stages 0 to I, I to II, II to III, and III to IV were 5.2±1.2, 4.2±0.5, 2±0.3, and 0.9±0.2 years, respectively, which were similar to those of the ER negative patients of 5.4±2.9, 4.7±1.3, 2.5±0.7, and 1.1±0.4 years (P=0.49). The survival of stage IV patients was longer in ER positive patients than ER negative patients (2.7±0.1 *vs.* 0.9±0.2 year, P<0.001).

Conclusions: ER negative patients appear to have the same progression time as that of ER positive patients in the primary site, but had a much worse process in the metastasis site.

Keywords: Breast cancer; natural history; cancer progression; estrogen receptor (ER)

Introduction

The medical phrase "natural history" refers to the behavior of a disease in the absence of treatment. Numerous mathematical models and retrospective analyses have been conducted to explore the natural progression of breast cancer, but there is no precise timeline as to how long it takes breast cancer to progress from stages 0 to IV. Additionally, under the current system of treatment for

breast cancer, early detection means early treatment, but treatment means that we do not know what occurs naturally. Similarly, careful and systematic observation in the days before active treatment is also exceptional.

In the past 2 decades, our understanding of breast cancer biology has undergone tremendous changes (1), but the essential questions involving the factors or mechanisms determining the distant metastasis of breast cancer remain unanswered (2). Thus, reassessing the natural history of breast cancer progression using a real-world data set might provide novel insights into better screening strategies, treatment strategies, and health care policies for breast cancer.

One way to determine the rate of growth of breast cancer is by examining the growth rate or volume doubling time, which is the period that it takes for a tumor to double in size. Previous studies have examined the growth rate via serial mammograms or ultrasounds with widely varying results—reports of the average doubling time range from 44 to 1,800 days (3,4). In a 2016 study that similarly examined growth based on ultrasound between diagnosis and surgery over a 31-day period, tumors increased from 1.47 to 1.56 cm in diameter (5). One older study (6) found that the doubling time of breast cancers was more rapid and fell into the following three categories: (I) ≤25 days; (II) 25–76 days; (III) ≥76 days. However, the above-mentioned studies lacked a comprehensive analysis of the progression of breast cancer over different stages.

Studying the natural progression of breast cancer requires a long-term follow-up period with a large cohort of diagnosed but untreated patients, which is impractical in clinical trials. This study aimed to assess the anatomic stage progression of untreated patients to address these issues. The Surveillance, Epidemiology, and End Results (SEER) database which was collected the breast cancer patients without surgical treatment, radiotherapy, or chemotherapy was used to analyze the clinical-pathological characteristics and survival data of untreated breast cancer patients. We present the following article in accordance with the STROBE reporting checklist (available at https://atm. amegroups.com/article/view/10.21037/atm-22-918/rc) (7).

Methods

Data source

Data analyzed in the manuscript is retrieved as described in our previous paper (8). The SEER 18 registries research database (Nov 2018 submission) was used for the analysis and includes patients diagnosed from 1975 to 2016, covering approximately 27.8% of the United States population. In relation to estrogen receptor (ER) status, ER positive status was defined as having either positive or borderline ER results. The human epidermal growth factor receptor 2 (HER2) status of breast tumors was not included in our analyses, as this information was only included in cases after 2010. The tumor (T) and node (N) stages of each patient were based on the best available pathological information. When pathological information was not available, clinical information was used. The study was conducted in accordance with the Declaration of Helsinki (as revised in 2013).

Survival analysis

The Kaplan-Meier method was used to estimate the survival outcomes. Overall survival (OS) was calculated from the date of the latest diagnosis of breast cancer to the date of death. Medium survival time (MST) was defined as the time at which 50% of the patients reached the endpoint. In conditions where <50% of patients reached the endpoint, MST was calculated using the "survfit" function in the "survival" package. Progression time was calculated by subtracting the MST of a higher stage from that of a lower stage, which also indicated "life loss". Both ER negative and positive patients were analyzed.

Statistical analysis

Kaplan-Meier survival analyses were performed. The Student's t-test, chi-square test, and log-rank test were used for the statistical analysis. All the P values were 2-sided, and a P value <0.05 indicated statistical significance. All the statistical analyses and case selections were performed in R (version 3.6.0, R Foundation for Statistical Computing, Vienna, Austria). The survival analyses were performed using the "survival" package (version 2.41) (9).

Results

Patient characteristics

A total of 763,873 unique breast cancer patients diagnosed between 1975 and 2016 were identified from the 840,660 breast cancer entries in the SEER database. Of these, 508,058 female patients had complete information on
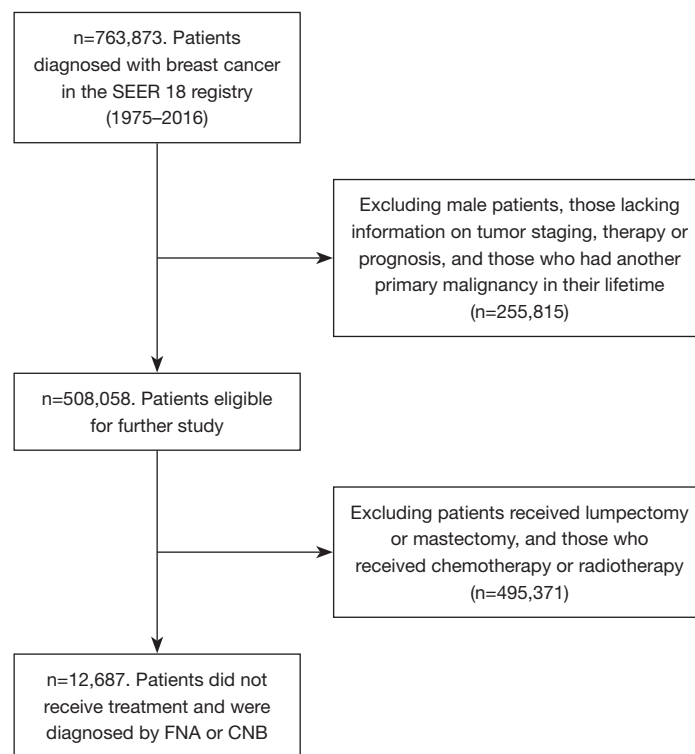
Figure 1 Flowchart of the patient selection process. SEER, Surveillance, Epidemiology, and End Results; FNA, fine-needle aspiration; CNB, core needle biopsy.

tumor staging, surgery, radiotherapy, chemotherapy, and prognosis, and had only 1 primary tumor in their lifetime. Among these patients, 12,687 did not receive surgical treatment, radiotherapy, or chemotherapy, and were included in our study as the untreated group (see *Figure 1*).

The clinicopathological characteristics of Group I patients are summarized in *Table 1*. Among them, 1,390 (11.0%) patients had stage I breast cancer, 1,702 (13.4%) had stage II breast cancer, 1,096 (8.6%) had stage III breast cancer, and 6,265 (49.4%) had stage IV breast cancer. Among the 7,871 (62%) patients with known ER status, 84.6% were ER positive.

Survival analysis

Kaplan-Meier survival analyses were performed in patients with stages I–III disease (see *Figure 2*). The gap of estimated survival time (EST) from breast cancer-specific survival (BCSS) between stage 0, I, IIA, IIB, IIIA, IIIB, IIIC, and IV were 5, 4.2, 1.2, 0.3, 1.3, 0.3, and 1.0 years, respectively (see *Table 2*). For ER positive patients, the median progression times from stage 0 to I, I to II, II to III, and III to IV were

5.2±1.2, 4.2±0.5, 2±0.3, and 0.9±0.2 years, respectively, and those for ER negative patients were 5.4±2.9, 4.7±1.3, 2.5±0.7, and 1.1±0.4 years, respectively, which were similar (P=0.49; see *Figure 3*). The survival difference of the ER positive and negative patients mainly depended on the survival difference of stage IV patients, which was longer in ER positive patients than ER negative patients (ER positive *vs.* ER negative, 2.7 *vs.* 0.9 year, P<0.001). Ninety-seven patients in the SEER database had been enrolled multiple times. The mean interval between stage 0 to stage I was 4.8±4.2 years. A scatter plot is shown in *Figure 4*.

Discussion

Characterizing the evolutionary history of breast cancer could provide a rational basis for effective treatments. In this study, we precisely defined the natural course of breast cancer anatomical staging and progression for the first time. We quantified the natural progression of breast cancer using the SEER database. In untreated patients, the progression times from stage 0 to I, I to II, II to III, and III to IV were 5.3, 4.6, 2.3, and 1.0, respectively. ER positive patients

Page 4 of 7

Wu et al. How long does it take for BC to go from stage 0 to IV?

Table 1 Cohort characteristics of the untreated patients

Characteristics	Stage 0, no. (%)	Stage I, no. (%)	Stage II, no. (%)	Stage III, no. (%)	Stage IV, no. (%)	Total, no. (%)
Patients, n	2,234	1,390	1,702	1,096	6,265	12,687
Age, y						
<40	76 (3.4)	35 (2.5)	73 (4.3)	24 (2.2)	121 (1.9)	329 (2.6)
40–54	926 (41.5)	217 (15.6)	259 (15.2)	115 (10.5)	805 (12.8)	2,322 (18.3)
55–69	684 (30.6)	381 (27.4)	362 (21.3)	234 (21.4)	2,004 (32.0)	3,665 (28.9)
≥70	548 (24.5)	757 (54.5)	1,008 (59.2)	723 (66.0)	3,335 (53.2)	6,371 (50.2)
Ethnicity						
White	1,673 (74.9)	1,074 (77.3)	1,197 (70.3)	769 (70.2)	4,935 (78.8)	9,648 (76.0)
Asian	144 (6.4)	88 (6.3)	127 (7.5)	68 (6.2)	317 (5.1)	744 (5.9)
Black	287 (12.8)	197 (14.2)	330 (19.4)	233 (21.3)	923 (14.7)	1,970 (15.5)
Other	130 (5.8)	31 (2.2)	48 (2.8)	26 (2.4)	90 (1.4)	325 (2.6)
ER status						
Positive	603 (27.0)	1,076 (77.4)	1,197 (70.3)	631 (57.6)	3,154 (50.3)	6,661 (52.5)
Negative	82 (3.7)	115 (8.3)	198 (11.6)	181 (16.5)	634 (10.1)	1,210 (9.5)
Unknown	1,549 (69.3)	199 (14.3)	307 (18.0)	284 (25.9)	2477 (39.5)	4,816 (38.0)
PR status						
Positive	485 (21.7)	932 (67.1)	1,010 (59.3)	492 (44.9)	2,447 (39.1)	5,366 (42.3)
Negative	140 (6.3)	242 (17.4)	376 (22.1)	313 (28.6)	1,249 (19.9)	2,320 (18.3)
Unknown	1,609 (72.0)	216 (15.5)	316 (18.6)	291 (26.6)	2,569 (41.0)	5,001 (39.4)
Tumor grade*						
I	222 (9.9)	400 (28.8)	208 (12.2)	61 (5.6)	313 (5.0)	1,204 (9.5)
II	426 (19.1)	523 (37.6)	624 (36.7)	302 (27.6)	1,333 (21.3)	3,208 (25.3)
III	297 (13.3)	209 (15.0)	416 (24.4)	341 (31.1)	1,321 (21.1)	2,584 (20.4)
IV	68 (3.0)	2 (0.1)	13 (0.8)	15 (1.4)	75 (1.2)	173 (1.4)
Unknown	1,221 (54.7)	256 (18.4)	441 (25.9)	377 (34.4)	3,223 (51.4)	5,518 (43.5)

*, by the "World Health Organization International Classification of Diseases for Oncology, Second Edition". Grade I codes for well differentiated, grade II codes for moderately differentiated, grade III codes for poorly differentiated, grade IV codes for undifferentiated. ER, estrogen receptor; PR, progesterone receptor.

appeared to have the same progression time as ER negative patients, but the survival of stage IV patients was much longer in ER positive patients than ER negative patients.

Notably, we found that the anatomical T and N stages of the primary tumors progressed similarly, but a slightly faster trend was observed in ER positive patients than ER negative patients, which was somewhat unexpected, as most basic research evidence and clinical evidence suggest that ER positive patients have better prognosis

and slower progress. Our study found that the survival difference between ER positive and ER negative patients does not mainly depend on the progression time of the primary tumor, but mainly depends on the progress of the metastasis. The breast is an estrogen-enriched organ, which may provide a micro-environmental advantage for the growth of ER positive tumor cells. After metastasis, an estrogen-deficient environment makes ER positive patients have a lower growth rate than ER negative patients. The
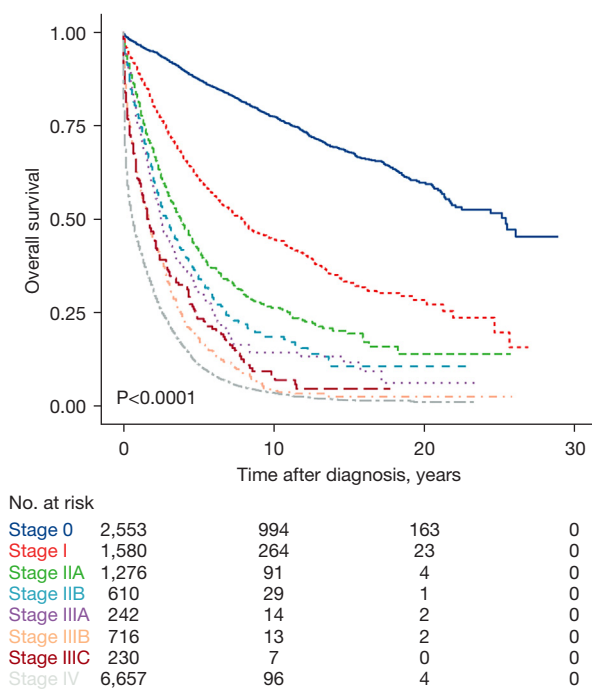
Figure 2 Survival analysis and comparison based on the clinical stages of the untreated patients.

Table 2 MST gap derived from OS between two cancer stages of untreated patients

Cancer stages	ΔMST, years
Stages 0–I	5
Stages I–IIA	4.2
Stages IIA–IIB	1.2
Stages IIB–IIIA	0.3
Stages IIIA–IIIB	1.3
Stages IIIB–IIIC	0.3
Stages IIIC–IV	1.0

Δ, the gap of time between two cancer stages. MST, medium survival time; OS, overall survival.

situation of growth rate for ER negative patients may be just the opposite. Currently, the theory that surgery to combat cytokine storms has a greater effect on ER negative cells is widely accepted (10,11). The current explanation for the double-peak recurrence of 2–3 and 8–9 years after surgery states that the two recurrence peaks represent the growth of cytokine storm sensitive and insensitive cancer cells (12-14). We suspect that the two recurrence peaks may be the contribution of both ER negative and ER positive cells.

The estimated median survival time of ER negative stage V patients was shorter than that of ER positive patients, indicating that the growth rate of metastatic lesions in ER negative patients was faster than that of ER positive patients, and ER negative patients should receive early aggressive treatment. Conversely, in relation to ER positive patients, as the growth rate of metastatic lesions is slower, they should be followed and managed in a long-term manner. Research should be conducted to identify predictive biomarkers that could distinguish between patients who have early or late metastasis. There is evidence that circulating tumor cells (CTCs) could serve as an indicator for that purpose (15). Further, the modern technology classification of molecular subtypes of CTC, which are independent of the primary site, may have potential application value.

Our results also showed that the progression time was not very even. The progression time from stage IIB to IIIA, and IIIB to IIIC was only 0.3 years, respectively. This implied that the anatomic tumor, node, metastasis (TNM) staging system needs further optimization. The molecular background and treatments in modern oncology will also affect the prognosis. The concept of an effective tumor burden, including anatomy and molecular portraits and treatment factors, needs to be urgently defined. There are untreated and treated patients in the SEER database, which will have certain value in future stage optimization.

Our study had several potential limitations. First, the use of SEER data instead of cohort study data and the lack of an exact treatment regimen may limit the strength of the evidence, and the group of untreated patients analyzed in our study may be biased. Second, as the data came from a single country, the analysis of data from other countries or databases may be necessary to further confirm our findings. Finally, 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 attempted to study the natural course of breast cancer in untreated breast cancer patients to identify patterns of progression of primary and metastatic foci. We found that ER negative patients had the same progression
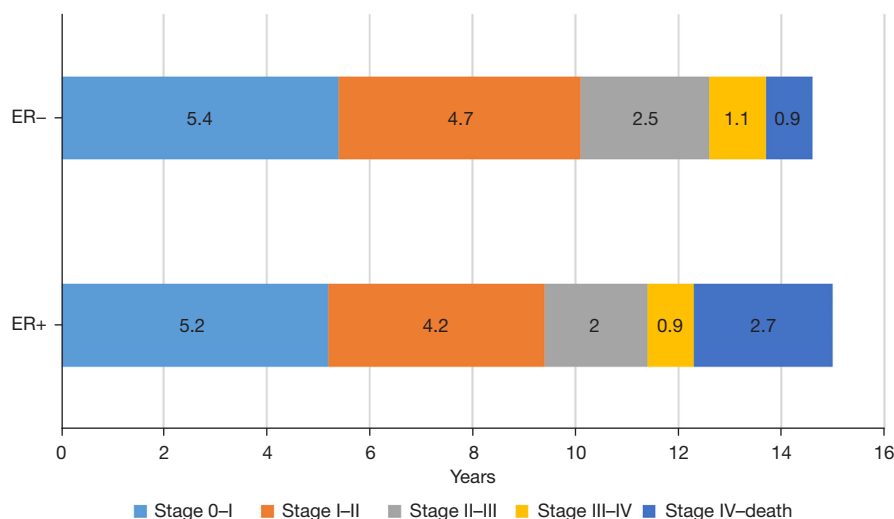
Figure 3 Progression times between subsequent stages in untreated ER positive and ER negative patients. The light blue bar represents the progression time from stages 0 to I; the orange, grey, and yellow bars represent the progression times from stages I to II, stages II to III, and stages III to IV, respectively; the dark-blue bar represents the survival time of stage IV patients. ER, estrogen receptor; ER–, estrogen receptor negative; ER+, estrogen receptor positive.
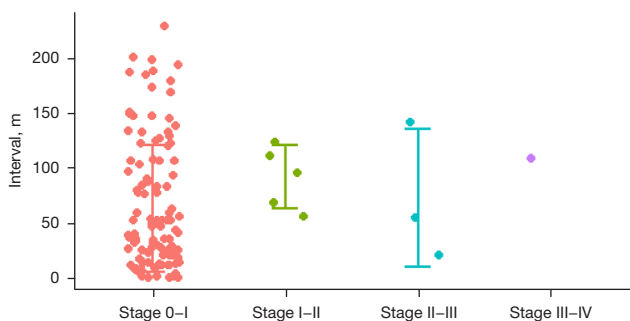
Figure 4 Scatter plot of progression time in untreated breast cancer patients enrolled multiple times in the cohort.

time as ER positive patients in the primary site but had a much worse progression time in the metastasis site.

Acknowledgments

Footnote

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. Harbeck N, Penault-Llorca F, Cortes J, et al. Breast cancer. Nat Rev Dis Primers 2019;5:66.
2. Weigelt B, Peterse JL, van 't Veer LJ. Breast cancer metastasis: markers and models. Nat Rev Cancer 2005;5:591-602.
3. von Fournier D, Weber E, Hoeffken W, et al. Growth rate of 147 mammary carcinomas. Cancer 1980;45:2198-207.
4. Heuser L, Spratt JS, Polk HC Jr. Growth rates of primary breast cancers. Cancer 1979;43:1888-94.
5. Lee SH, Kim YS, Han W, et al. Tumor growth rate of invasive breast cancers during wait times for surgery assessed by ultrasonography. Medicine (Baltimore) 2016;95:e4874.
6. Heuser L, Spratt JS Jr, Polk HC Jr, et al. Relation between mammary cancer growth kinetics and the intervals between screenings. Cancer 1979;43:857-62.
7. 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.
8. Hu T, Wu J, Long M, et al. Modeling effective tumor burden of primary lesion and metastatic lymph node in breast cancer patients from the SEER database. Gland Surg 2022;11:236-44.
9. Therneau TM, Grambsch PM. Modeling Survival Data: Extending the Cox Model. Springer, New York; 2000.
10. Esquivel-Velázquez M, Ostoa-Saloma P, Palacios-Arreola MI, et al. The role of cytokines in breast cancer development and progression. J Interferon Cytokine Res 2015;35:1-16.
11. Purohit A, Newman SP, Reed MJ. The role of cytokines in regulating estrogen synthesis: implications for the etiology of breast cancer. Breast Cancer Res 2002;4:65-9.
12. Duncan LJ, Coldham NG, Reed MJ. The interaction of cytokines in regulating oestradiol 17 beta-hydroxysteroid dehydrogenase activity in MCF-7 cells. J Steroid Biochem Mol Biol 1994;49:63-8.
13. Reed MJ, Purohit A. Breast cancer and the role of cytokines in regulating estrogen synthesis: an emerging hypothesis. Endocr Rev 1997;18:701-15.
14. Simard J, Gingras S. Crucial role of cytokines in sex steroid formation in normal and tumoral tissues. Mol Cell Endocrinol 2001;171:25-40.
15. Sparano J, O'Neill A, Alpaugh K, et al. Association of Circulating Tumor Cells With Late Recurrence of Estrogen Receptor-Positive Breast Cancer: A Secondary Analysis of a Randomized Clinical Trial. JAMA Oncol 2018;4:1700-6.