Published in final edited form as:

Cancer Causes Control. 2012 September ; 23(9): 1587–1591. doi:10.1007/s10552-012-0026-7.

Validation of AJCC TNM staging for breast tumors diagnosed before 2004 in cancer registries

Marilyn L. Kwan¹, Reina Haque², Valerie S. Lee¹, W-L Joanie Chung², Chantal C. Avila², Heather A. Clancy¹, Virginia P. Quinn², and Lawrence H. Kushi¹

¹Division of Research, Kaiser Permanente Northern California

²Department of Research & Evaluation, Kaiser Permanente Southern California

Abstract

Purpose—American Joint Committee on Cancer (AJCC) Tumor (T), Nodal (N) and Metastatic (M) staging is commonly used in clinical practice for treatment decisions, yet before 2004, Surveillance Epidemiology and End Results (SEER)-affiliated cancer registries did not routinely include TNM staging defined by AJCC criteria, reporting instead SEER Summary Staging.

Methods—We developed and validated an algorithm to determine AJCC TNM staging from Extent of Disease information for 17,133 female breast cancer cases diagnosed from 1988–2003 in the cancer registries of Kaiser Permanente Northern and Southern California. Test characteristics (percent agreement, Cohen’s kappa, sensitivity, specificity) were calculated to compare derived TNM with gold-standard TNM available in the registry.

Results—Agreement for TNM variables was excellent (range 0.91–1.00 for percent agreement and Cohen’s kappa). The sensitivity and specificity, respectively, of the algorithm for AJCC TNM Version 6 staging was: Stage 0 (0.99, 1.00), Stage I (0.97, 0.98), Stage II (0.91, 0.96), Stage III (0.69, 0.99), Stage IV (0.92, 1.00). Stage III had lower sensitivity due to reclassification of supraclavicular lymph node positivity from M1 (Stage IV) to N3 (Stage IIIC) in AJCC Version 6.

Conclusions—Derived AJCC staging for breast tumors diagnosed before 2004 is feasible and accurate using cancer registry data.

Keywords

breast cancer; American Joint Committee on Cancer; tumor; nodal; metastatic cancer staging; cancer registry; validation study

Introduction

American Joint Committee on Cancer (AJCC) Tumor (T), Nodal (N) and Metastatic (M) staging is commonly used in clinical practice for treatment decisions [1]. Before 2004, however, many cancer registries that conformed to the standards of the North American Association of Central Cancer Registries (NAACR) [2], including the NCI-supported Surveillance Epidemiology and End Results (SEER) registries [3], did not routinely include TNM staging defined by AJCC criteria, reporting instead SEER Summary Staging [4]. As reported in the SEER annual reports, SEER Summary Stage is used to examine secular trends in cancer incidence, yet it is less useful in investigations of cancer prognosis and treatment. Thus, numerous studies that rely on cancer registries for diagnostic information

cannot include clinically-relevant TNM staging information for cases diagnosed before 2004.

Kaiser Permanente Northern California (KPNC) and Kaiser Permanente Southern California (KPSC) health plans operate their own individual cancer registries and report case counts to the California Cancer Registry, regional registries, and the SEER Program [5, 6]. Together, KPNC and KPSC provide comprehensive medical care to approximately 6.7 million members (about 2% of the U.S. population), including over 6,000 women diagnosed with breast cancer each year.

Direct TNM staging was not collected prior to 2004 in the cancer registry of KPNC. However, the registry did include information on extent of disease (EOD), which includes extent and size of primary tumor, presence of metastases, and lymph node involvement as required by NAACR and SEER [7]. In contrast, the cancer registry of KPSC collected both T,N, and M staging and EOD variables since 1988, in accordance with cancer center accreditation standards of the American College of Surgeons [8].

We developed an algorithm to map EOD data elements to AJCC TNM stages for female breast cancer cases diagnosed from 1988–2003 identified from the KPNC Cancer Registry. We then validated this algorithm using data from the KPSC Cancer Registry in 17,133 cases diagnosed from 1996–2003 with available EOD and TNM information as the gold standard.

Materials and Methods

Using the SEER Comparative Staging Guide for Cancer (1993) [7] and the AJCC Cancer Staging Manual 6th edition (2002) [1], TNM stages were derived from EOD variables in the KPNC Cancer Registry for 48,279 female breast cancer cases diagnosed from 1988–2003 (Table 1). The T, N, and M components were combined to derive a composite stage (0-IV) using AJCC Version 6 methods. Furthermore, stage was calculated using an adaptation of the AJCC system as described by the 1993 SEER publication to enable inclusion of all tumor histologic types but updated to incorporate Version 6 coding (SEER-modified AJCC staging).

The algorithm was applied to the same EOD variables in the KPSC Cancer Registry for 17,819 female breast cancer cases diagnosed from 1996–2003. Derived TNM variables and stage from this algorithm were compared to gold-standard pathologic TNM variables and stage available in the KPSC Cancer Registry. Any missing pathologic TNM variables, which is TNM staging done by the pathologist only, in the KPSCCR (n=493, n=491, and n=496, respectively) were replaced with available clinical TNM variables, as comparisons between pathologic and clinical TNM variables were found to be minimal. Tumors that could not be staged (n=678, 3.9%) or with incomplete TNM information (n=8) were excluded from the analysis, leaving a final sample size of 17,133 cases. Test characteristics of the staging algorithm compared to the pre-defined gold standard (percent agreement, Cohen's kappa, sensitivity, and specificity) were calculated for overall stage and individual TNM variables. Percent agreement was defined as the ratio of the number of times the derived stage by the EOD staging algorithm and the gold-standard TNM stage agree divided by the total number of stages determined. Kappa was defined as the proportion of agreement after removing the proportion of agreement which would occur by chance (Cohen's kappa). Within each stage category only, sensitivity was defined as the proportion of tumors with the known stage from the KPSC Cancer Registry that was correctly staged by the EOD staging algorithm (e.g., Stage I was correctly identified as Stage I). In contrast, within each stage category only, specificity was defined as the proportion of tumors without the known stage from the KPSC Cancer Registry that was correctly not staged as the known stage by the

EOD staging algorithm (e.g., Stage I was not identified as Stage 0, II, III, or IV). Thus, overall sensitivity and specificity were not calculable for all stages combined. Finally, KPNC and KPSC distributions of breast cancer diagnoses by stage from 1996–2003 were compared.

This study was approved by the Institutional Review Boards of KPNC and KPSC. All analyses were two-sided and performed using SAS version 9.2; $P < 0.05$ was considered statistically significant.

Results and Discussion

Agreement between the derived TNM variables and the gold-standard TNM variables was excellent (percent agreement and Cohen's kappa): T (93% and 0.91); N (96% and 0.94); M (100% and 0.93). In both the AJCC TNM Version 6 and SEER modified AJCC TNM staging systems, consistent agreement and high sensitivity and specificity were observed overall and by stage, except for somewhat lower kappa and sensitivity for Stage III (Table 2). The sensitivity and specificity, respectively, for AJCC Version 6 were: Stage 0 (0.99, 1.00), Stage I (0.97, 0.98), Stage II (0.91, 0.96), Stage III (0.69, 0.99), Stage IV (0.92, 1.00), Overall (0.92, 0.89). Slightly higher test results were found for SEER-modified AJCC staging. The overall distributions for derived TNM staging among KPSC and KPNC breast cancer cases diagnosed from 1996–2003 were similar by stage and by years of diagnosis (Table 3).

In summary, agreement between the derived T, N, and M variables and the gold-standard T, N, and M variables was excellent (range 0.91–1.00) for all breast cancer cases diagnosed from 1996–2003 at KPSC. Consistent high sensitivity and specificity were observed overall and within each stage except for somewhat lower agreement and sensitivity for Stage III. The overall distributions for derived TNM staging for KPSC and KPNC were comparable by stage and by diagnosis years.

The lower agreement and sensitivity of the algorithm for Stage III was due to the reclassification of metastasis to the supraclavicular lymph node(s) from M1 (Stage IV) to N3 (Stage IIIC) in AJCC Version 6 [9]. Thus, in our study, some misclassification between Stages IIIC and IV occurred, as the years of diagnosis under consideration were 1996–2003 (AJCC Versions 4 and 5). However, we used an algorithm based on AJCC Version 6 (2004–2009) to be more consistent with current staging guidelines.

Of note, beginning with 2010 cancer diagnoses, AJCC Version 7 is being used [10]. The only major change in TNM staging for breast tumors from Version 6 to Version 7 is the new sub-division of Stage I to Stage IA and IB [11]. Stage IB now includes small tumors with micrometastases in lymph nodes (N1mi), which were previously included as part of Stage IIA. Given this change, our algorithm based on AJCC Version 6 can be successfully applied to accommodate AJCC Version 7, with the small modification of T0 and T1 tumors with nodal micrometastases being classified as Stage IB instead of previously Stage IIA. This modification will result in a minor stage shift of some tumors from IIA to IB. However, we are unable to assess the proportion of cases that would shift stages as micrometastasis is normally categorized under pathologic N (pN1mi), and our Kaiser Permanente cancer registries only capture information on clinical N.

Determination of AJCC stage using EOD information for breast cancer cases diagnosed before 2004 has been done previously by the California Cancer Registry for internal purposes using the SEER Comparative Staging Guide for Cancer [7, 12]. However, to our knowledge, validation of an EOD/AJCC staging algorithm has not been performed in an independent population with previously collected TNM variables.

In conclusion, SEER EOD variables for breast tumors available in cancer registries also reporting to the SEER program can be used to derive AJCC TNM staging. Epidemiologic studies of breast cancer prognosis using cancer registry data can accurately apply current cancer staging criteria to historical breast cancer cases.

Acknowledgments

We would like to thank Michael D. Oehrli, MPA, CTR of the Kaiser Permanente Northern California Cancer Registry and Gerri L. Salazar, CTR of the Kaiser Permanente Southern California Cancer Registry for their consultation on this study.

Grant Support

This work was supported, in part, by grants from the National Cancer Institute (R01 CA136743 to R. H., RC2 CA148185, and the Cancer Research Network, U19 CA079689).

References

1. American Joint Committee on Cancer. American Joint Committee on Cancer (AJCC) Cancer Staging Manual 6th Edition. 2002.
2. [Accessed January 10, 2012] North American Association of Central Cancer Registries. Available from: <http://www.naaccr.org/>.
3. [Accessed December 15, 2011] Surveillance Epidemiology and End Results (SEER). Available from: <http://seer.cancer.gov/>.
4. YJ, et al., editors. SEER Summary Staging Manual - 2000: Codes and Coding Instructions. Bethesda, MD: National Institutes of Health; 2001.
5. Oehrli, MD.; Quesenberry, CP.; Leyden, WA. 2009 Annual Report on Trends, Incidence, and Outcomes. Kaiser Permanente, Northern California Cancer Registry; 2009.
6. Kaiser Permanente Southern California Cancer Registry. Kaiser Permanente Southern California 2009 Cancer Report. Pasadena, CA: 2009.
7. SEER Program: Comparative Staging Guide for Cancer. National Institutes of Health; 1993.
8. American College of Surgeons.
9. American Joint Committee on Cancer. The AJCC Comparison Guide: Fifth Versus Sixth Edition. 2002. p. 20-23.
10. American Joint Committee on Cancer. American Joint Committee on Cancer (AJCC) Cancer Staging Manual 7th Edition. 2009.
11. American Joint Committee on Cancer. <http://www.cancerstaging.org/staging/changes2010.pdf>
12. [Accessed December 15, 2011] California Cancer Registry. Available from: <http://www.ccrca.org/>.

Table 1

Derivation of tumor (T), nodal (N), and metastatic (M) variables from extent of disease (EOD) variables (extension, primary tumor size, lymph node) for female breast cancer cases diagnosed at Kaiser Permanente Northern California

Extension (Metastases)*	Primary Tumor Size	T variable	M variable [†]
00 In-situ: Noninfiltration; intraductal without infiltration; lobular neoplasia		Tis	M0
05 and ICD-O-3 code (C50.0) and ICD-O-3 codes(C50.1–50.9)		Tis T0	M0 M0
10, 11, 12, 13, 14, 15, 16, 17, 18 20, 21, 22, 23, 24, 25, 26, 27, 28 30, 31, 33, 34, 35, 38	001,002 003–<005 005–<010 010–<020 020–<050 050–990 998: Diffuse; widespread: > ¾ breast; inflammatory carcinoma 999: not stated 997: Paget's disease of nipple no demonstrable tumor	T1 T1a T1b T1c T2 T3 T3 Txa Error [‡]	M0 Error [‡]
40		T4a	M0
50		T4b	M0
60 (both 40 and 50)		T4c	M0
70		T4d	M0
80		T4d ^l	M1
85		T4d ^l	M1
99	001,002 003–<005 005–<010 010–<020 020–<050 050–990 998: Diffuse; widespread: > ¾ breast; inflammatory carcinoma 999: not stated 997: Paget's disease of nipple no demonstrable tumor	T1 T1a T1b T1c T2 T3 T3 Txa Error [‡]	MX [‡] Error [‡]
Lymph Node		N variable	M variable
0 No lymph node involvement		N0	MX [§]
1		N1a	MX [§]
2, 3, 4		N1b	MX [§]
5		N2	MX [§]
6		N1x	MX [§]
7		N3	MX [§]
8		Nxx ^{**}	M1
9		Nx	MX [§]

* When extension (metastases) is unknown (code=99), tumor size could be available, and thus, assignment of T variable. However, M variable is coded to MX per the SEER Comparative Staging Guide for Cancer [7].

[†]M variable is determined as follows: If not coded to 80 or 85 for further extension or metastasis, respectively, other than actual coding of unknown extension (code=99), M0 (no distant metastasis) is assumed.

[‡]Note on page S.12 of the SEER Comparative Staging Guide for Cancer [7] indicating that when size code 997 is used with any extension code other than 05, it is considered an error and data should be reviewed if possible.

[§]Placeholder for M stage. Since Lymph Node codes for lymph node involvement (not information on tumor itself), only code=8 describes distant lymph nodes involvement that should be a stage IV. This placeholder is for programming purposes only and is not to be interpreted separately.

^{**}Placeholder for T variable or N variable since M variable (M1) will take precedence over staging to Stage IV.

Test characteristics of the Kaiser Permanente Northern California (KPNC) extent of disease (EOD) staging derivation for AJCC TNM Version 6 and SEER-modified AJCC TNM Staging Systems of female breast cancer cases diagnosed from 1996–2003 at Kaiser Permanente Southern California (KPSC)

Table 2

AJCC TNM Version 6 Staging*	Stage 0	Stage I	Stage II	Stage III	Stage IV	Overall [‡]
Percent Agreement	1.00	0.97	0.94	0.97	1.00	0.92
(95% CI) [‡]	(1.00, 1.00)	(0.97, 0.98)	(0.94, 0.95)	(0.97, 0.97)	(1.00, 1.00)	(0.92, 0.93)
Kappa Statistic	0.99	0.94	0.88	0.75	0.94	0.89
(95% CI) [‡]	(0.99, 1.00)	(0.94, 0.95)	(0.87, 0.88)	(0.73, 0.77)	(0.92, 0.95)	(0.89, 0.90)
Sensitivity	0.99	0.97	0.91	0.69	0.92	Not
(95% CI) [‡]	(0.99, 0.99)	(0.96, 0.97)	(0.90, 0.92)	(0.66, 0.71)	(0.90, 0.94)	calculable
Specificity	1.00	0.98	0.96	0.99	1.00	Not
(95% CI) [‡]	(1.00, 1.00)	(0.97, 0.98)	(0.96, 0.97)	(0.99, 0.99)	(1.00, 1.00)	calculable
SEER-modified AJCC TNM Staging*	Stage 0	Stage I	Stage II	Stage III	Stage IV	Overall [‡]
Percent Agreement	1.00	0.98	0.96	0.97	1.00	0.95
(95% CI) [‡]	(1.00, 1.00)	(0.98, 0.98)	(0.95, 0.96)	(0.97, 0.97)	(0.99, 1.00)	(0.95, 0.96)
Kappa Statistic	1.00	0.95	0.90	0.77	0.94	0.93
(95% CI) [‡]	(0.99, 1.00)	(0.95, 0.96)	(0.90, 0.91)	(0.75, 0.79)	(0.92, 0.95)	(0.93, 0.94)
Sensitivity	1.00	0.98	0.95	0.72	0.92	Not
(95% CI) [‡]	(0.99, 1.00)	(0.98, 0.99)	(0.94, 0.96)	(0.70, 0.75)	(0.90, 0.94)	calculable
Specificity	1.00	0.98	0.96	0.99	1.00	Not
(95% CI) [‡]	(1.00, 1.00)	(0.97, 0.98)	(0.95, 0.96)	(0.99, 0.99)	(1.00, 1.00)	calculable

* From the AJCC Cancer Staging Manual 6th edition [1] and the SEER Comparative Staging Guide for Cancer [7], respectively.

[‡] Test characteristics were estimated by comparing derived TNM staging data from extent of disease data in the KPSC Cancer Registry to previously collected, gold-standard TNM staging information in the KPSC Cancer Registry. Overall sensitivity and specificity were not calculable for all stages combined since these test characteristics were specific to each stage category.

[‡] Overall statistics for AJCC TNM Staging Version 6 exclude Not Applicable for mixed stage per [1]

§ Overall statistics for SEER-modified AJCC TNM Staging exclude Not Applicable for mixed stage and Error per [7].

Table 3
Distributions of TNM staging among KPSC and KPNC female breast cancer cases diagnosed from 1996–2003

AJCC TNM Version 6 Staging*							
Years of Diagnosis	Site	Stage 0	Stage I	Stage II	Stage III	Stage IV	Total
1996–1997	KPSC	456 (17.16)	1426 (22.35)	1248 (20.46)	181 (18.76)	112 (20.93)	3423 (100)
	KPNC	666 (20.01)	1766 (22.78)	1322 (21.42)	123 (17.60)	122 (19.33)	3999 (100)
1998–1999	KPSC	572 (21.53)	1434 (22.48)	1340 (21.97)	263 (27.25)	122 (22.80)	3731 (100)
	KPNC	867 (26.05)	1757 (22.67)	1454 (23.56)	176 (25.18)	148 (23.45)	4402 (100)
2000–2001	KPSC	753 (28.34)	1715 (26.89)	1685 (27.63)	274 (28.39)	138 (25.79)	4565 (100)
	KPNC	831 (24.97)	1903 (24.55)	1540 (24.96)	214 (30.62)	163 (25.83)	4651 (100)
2002–2003	KPSC	876 (32.97)	1804 (28.28)	1826 (29.94)	247 (25.60)	163 (30.47)	4916 (100)
	KPNC	964 (28.97)	2325 (30.00)	1855 (30.06)	186 (26.61)	198 (31.38)	5528 (100)

SEER-modified AJCC TNM Staging*							
Years of Diagnosis	Site	Stage 0	Stage I	Stage II	Stage III	Stage IV	Total
1996–1997	KPSC	466 (17.47)	1492 (22.96)	1297 (20.25)	196 (19.27)	112 (20.93)	3563 (100)
	KPNC	666 (20.01)	2008 (23.46)	1480 (21.68)	145 (18.57)	122 (19.33)	4421 (100)
1998–1999	KPSC	573 (21.48)	1467 (22.58)	1507 (23.53)	279 (27.43)	122 (22.80)	3948 (100)
	KPNC	832 (24.99)	2008 (23.46)	1615 (23.65)	207 (26.50)	148 (23.45)	4810 (100)
2000–2001	KPSC	753 (28.22)	1726 (26.57)	1745 (27.25)	281 (27.63)	138 (25.79)	4643 (100)
	KPNC	867 (26.04)	2103 (24.57)	1793 (26.26)	227 (29.07)	163 (25.83)	5153 (100)
2002–2003	KPSC	876 (32.83)	1812 (27.89)	1855 (28.97)	261 (25.66)	163 (30.47)	4967 (100)
	KPNC	964 (28.96)	2441 (28.52)	1940 (28.41)	202 (25.86)	198 (31.38)	5745 (100)

* From the AJCC Cancer Staging Manual 6th edition [1] and the SEER Comparative Staging Guide for Cancer [7], respectively.