SHORT REPORT

Completeness of breast cancer staging in the Danish Cancer Registry, 2004–2009

Anne Gulbech Ording[1]
Malene Schou Nielsson[1]
Trine Frøslev[1]
Søren Friis[2]
Jens Peter Garne[3]
Mette Søgaard[1,4]

[1]Department of Clinical Epidemiology, Aarhus University Hospital, Aarhus, [2]Danish Cancer Society Research Center, Danish Cancer Society, Copenhagen, [3]Department of Breast Surgery, [4]Department of Clinical Microbiology, Aalborg Hospital, Aarhus University Hospital, Aalborg, Denmark

Background: The purpose of this study was to investigate the completeness of TNM (Tumor, Node, Metastasis) staging for breast cancer in the Danish Cancer Registry.

Methods: We identified 26,488 patients with a first diagnosis of breast cancer between 2004 and 2009 from the Danish Cancer Registry. We obtained information on comorbidity through the Danish National Patient Registry. We estimated the completeness of TNM registration in the Danish Cancer Registry and stratified the analysis by gender, age, year of cancer diagnosis, and comorbidity. We designed an algorithm categorizing breast cancer into localized, regional, distant, or unknown stage based on TNM codes.

Results: The overall completeness of TNM registration was 85.4%. The completeness varied little by gender and study year, but decreased from 91.3% in patients aged 0–39 years to 57.0% in patients aged 80 years or more, and from 87.9% among patients with a low level of comorbidity to 69.7% among patients with a high level of comorbidity.

Conclusion: The completeness of the TNM registration varied substantially by age and level of comorbidity. Thus, depending on the outcome under study, stage-specific analyses may yield biased results. The completeness of TNM should be considered in study designs using TNM information.

Keywords: neoplasm staging, breast neoplasms, registries

Introduction

Breast cancer is the most common cancer and leading cause of cancer death among women worldwide.[1] Among Danish women, breast cancer accounted for 29% of all incident cancers, and was the cause of 16% of all cancer deaths among women during the period 1999–2006.[2] To improve survival of Danish patients with breast cancer, nationwide biannual mammographic screening was introduced in 2007 and offered to women between 50 and 69 years of age.[3] Cancer stage is an important predictor of prognosis, with 5-year survival of 98% for localized breast cancer decreasing to 27% for metastatic disease.[4] Therefore, analyses of temporal changes in stage-specific incidence and mortality from breast cancer are important for evaluating the effectiveness of early detection and treatment programs.

Cancer registries are valuable resources for monitoring cancer incidence and survival. The Danish Cancer Registry (DCR) has recorded incident cases of cancer on a nationwide basis since 1943 and has been shown to have accurate and almost complete ascertainment of cancer cases. The International Classification of Disease revision 7 (ICD-7) was used until 2003 to categorize cancer sites, and has been converted to the International Classification of Disease revision 10 (ICD-10).[5,6] The DCR

Correspondence: Anne G Ording
Department of Clinical Epidemiology,
Aarhus University Hospital,
Olof Palmes Allé 43-45,
8200 Aarhus N, Denmark
Tel +45 8716 8063
Fax +45 8716 2215
Email ao@dce.au.dk

11

includes information on diagnosis, cancer type, topography, morphology, and stage according to the Tumor, Node, Metastasis (TNM) classification.[7] Information on TNM classification may be prone to underreporting and misclassification, eg, of tumor size (T) and lymph node (N) status.[8] Missing information may potentially lead to biased results, if the missing information is not random. Thus, to draw valid inferences from stage-specific analysis, completeness needs to be quantified. Currently, there is a lack of studies on completeness of TNM classification in the DCR. Therefore, we conducted this study to evaluate completeness of TNM classification for breast cancer. In addition, we aimed to determine whether the completeness varied by gender, patient age, calendar period, cancer stage, or level of comorbidity.

Materials and methods

We performed this study in Denmark, which has about 5.4 million inhabitants. All residents are provided with free, tax-supported medical care. Since 1968, the Danish Civil Registration System has assigned a unique 10-digit personal identification number to all Danish residents,[9] encoding date of birth and gender. This number is used in all Danish registers, allowing unambiguous individual-level data linkage.

Ascertainment of patients with breast cancer

From the DCR, we identified all patients with a primary diagnosis of invasive breast cancer (ICD-10 code C50) between 2004 and 2009.[5,6] Tumor stage was recorded as local, regional, or distant in the DCR until 2003.[5] Since 2004, stage has been recorded using the TNM classification.[7] From the DCR, we also obtained information on date of diagnosis, age, and gender.

Comorbidity

Data on the presence of comorbidity were obtained from the Danish National Patient Registry (DNPR).[10] This registry contains data on all admissions to nonpsychiatric hospitals in Denmark since 1977 and outpatient contacts since 1995, including the personal identification number, date of admission/contact and discharge, and diagnosis codes (ICD-10 from 1994 and onwards). We defined pre-existing comorbidity using the Charlson Comorbidity Index (CCI), based on hospital diagnoses within 10 years preceding the date of breast cancer diagnosis. The CCI is based on disease categories which are each weighted according to the adjusted risk of one-year mortality.[11] Excluding breast cancer, we defined

the level of comorbidity as low (CCI score 0), medium (CCI score 1–2), and high (CCC score $\geq$3).

Statistical analysis

We calculated the completeness and corresponding 95% confidence intervals, of the TNM registration overall, by each component individually (ie, T, N, and M), and according to the stage categories (see Appendix). To reduce the prevalence of patients with breast cancer of unknown stage, we converted the TNM classification into summary staging using an algorithm that allowed categorization of tumors with certain missing TNM stage components into localized, regional, distant, and unknown stage (see Appendix). Missing data were allowed if the information available on other T, N, or M components provided sufficient and meaningful information to stage the tumors. The "unknown" category represented tumors of high T class, which are known to have considerable risk of lymph node or distant metastasis, but lacked information of either or both.

Completeness was defined as the number of individuals with TNM recordings, and defined stage categories, divided by the total number of patients. We stratified completeness by gender, age (0–39 years, 40–59 years, 60–79 years, and $\geq$80 years), year of cancer diagnosis, and CCI score (low, medium, high). Analyses were performed using SAS (version 9.2; SAS Institute Inc, Cary, NC).

Results

We identified 26,488 patients with incident breast cancer between 2004 and 2009. The median patient age was 63 years (interquartile range 54–72 years). A total of 26,350 (99.5%) breast cancer patients were women, and 138 (0.5%) were men.

The overall completeness of TNM registration was 85.4%. There was a slightly better registration for women than for men (85.4% versus 81.2%). There were no major changes in TNM completeness during the six-year study period, though completeness was slightly lower in 2009 compared with previous years. Completeness declined markedly with advancing age, from 91.3% among patients aged 0–39 years to 57.0% among patients aged 80 years and older. Completeness also declined with increasing CCI score, from 87.9% among patients with a low level of comorbidity to 69.7% among patients with high levels of comorbidity. Similar patterns were found for the individual components of the TNM classification, although the N and M recordings had somewhat more missing values than T. These data are presented in Table 1.

Table 1 Proportions and 95% confidence intervals for completeness of TNM registration among breast cancer patients in the Danish Cancer Registry by year of breast cancer diagnosis, gender, age group, and Charlson Comorbidity Index score

	TNM completeness		T completeness		N completeness		M completeness		Total
	%	(95% CI)	%	(95% CI)	%	(95% CI)	%	(95% CI)	n
Overall	85.4	(85.0–85.8)	94.7	(94.4–94.9)	89.6	(89.2–90.0)	89.7	(89.3–90.1)	26,488
Year of diagnosis									
2004	86.0	(84.9–87.1)	94.7	(94.0–95.4)	89.3	(88.3–90.3)	91.0	(90.1–91.9)	3904
2005	86.6	(85.5–87.7)	95.6	(94.9–96.2)	89.8	(88.8–90.7)	92.1	(91.2–92.9)	3940
2006	84.8	(83.7–85.9)	95.8	(95.2–96.4)	90.0	(89.1–90.9)	89.9	(89.0–90.8)	4103
2007	85.9	(84.8–86.9)	96.0	(95.4–96.6)	90.5	(89.6–91.3)	90.9	(90.0–91.7)	4122
2008	87.0	(86.0–87.9)	95.4	(94.8–96.0)	91.3	(90.5–92.1)	90.4	(89.5–91.2)	4737
2009	82.9	(81.9–83.9)	91.6	(90.8–92.3)	87.4	(86.5–88.2)	85.5	(84.6–86.4)	5682
Gender									
Female	85.4	(85.0–85.9)	94.7	(94.4–95.0)	89.6	(89.3–90.0)	89.7	(89.3–90.1)	26,350
Male	81.2	(74.0–87.0)	93.5	(88.4–96.7)	85.5	(78.9–90.6)	88.4	(82.3–92.9)	138
Age group (years)									
0–39	91.3	(89.5–93.0)	96.6	(95.3–97.6)	94.9	(93.4–96.1)	94.2	(92.6–95.5)	994
40–59	91.5	(90.9–92.0)	96.7	(96.3–97.0)	95.4	(95.0–95.8)	93.5	(93.0–94.0)	9740
60–79	87.3	(86.7–87.8)	94.9	(94.5–95.3)	91.2	(90.7–91.7)	91.3	(90.8–91.8)	12,632
≥80	57.0	(55.3–58.8)	86.9	(85.7–88.1)	63.6	(61.9–65.3)	70.0	(68.4–71.6)	3122
CCI score									
Low	87.9	(87.5–88.4)	95.4	(95.1–95.7)	91.9	(91.5–92.3)	91.4	(91.0–91.8)	20,274
Medium	78.8	(77.7–79.9)	93.0	(92.2–93.6)	83.8	(82.8–84.8)	84.9	(83.9–85.9)	5100
High	69.7	(67.0–72.4)	89.2	(87.3–91.0)	74.5	(71.9–77.0)	80.7	(78.3–82.9)	1114

Abbreviations: CI, confidence interval; CCI, Charlson Comorbidity Index; TNM, Tumor, Node, Metastasis classification.

Using the algorithm allowing inclusion of some Tx, Nx, and Mx codes in the classification of tumor stage (Appendix) decreased the number of unknown tumors from 3867 when not allowing any x's in the definite stage categories to 2551 using the algorithm allowing some x's in these categories (data not shown). Thus, with our algorithm, the proportion of tumors with unknown stage was reduced from 14.6% to 9.6%. Similar to the primary analyses of TNM completeness, the proportion of tumors of unknown stage was highest among older patients and those with high CCI scores. The proportion of patients with localized breast cancer at the time of diagnosis increased from 43.2% in 2004 to 49.8% in 2009. These data are presented in Table 2.

Discussion

In this nationwide study, we found reasonable completeness of TNM registration in the DCR. The completeness was relatively constant during the study period, except for a slight decline from 2008 through 2009, which was due to a delay in the recording of TNM. However, we found that completeness varied substantially with patient age and level of comorbidity. When we applied a stage algorithm allowing some missing TNM components in the various categories, the proportion of cases with definite staging increased.

The main strengths of our study included the population-based design and use of nationwide data from the DCR,

which has virtually complete ascertainment of breast cancer.[12] However, we only examined completeness of TNM registration, and our study did not allow for estimation of the accuracy of TNM registration. Furthermore, the impact of classifying some unknown stage breast tumors into categories with known tumors may not be appropriate for all purposes.

Moreover, we may have misclassified comorbidity due to inaccuracy of hospital and outpatient diagnoses as well as lack of information on comorbidities diagnosed in primary care. Even so, a recent study reported consistently high positive predictive values for all of the CCI diseases based on diagnoses recorded in the DNPR.[13]

Compared with other cancer sites, the proportion of unstaged breast cancer patients is typically relatively low,[8] though other studies, in agreement with our findings, have shown declining completeness of TNM registration with increasing age and comorbidity score.[8,14] In particular, Yancik et al[15] found that women older than 70 years underwent fewer lymph node dissections than younger women. Other studies have also shown that elderly breast cancer patients and patients with comorbidity are less likely to receive breast cancer treatment according to treatment guidelines and have a poorer prognosis than younger patients with no severe comorbid conditions.[16,17] The reasons are likely multifactorial, and may include less complete

Table 2 Proportions and 95% confidence intervals for staging of breast cancer patients in the Danish Cancer Registry by year of breast cancer diagnosis, gender, age group, and Charlson Comorbidity Index score

	Cancer stage[a]								Total
	Localized		Regional		Distant		Unknown		
	%	(95% CI)	%	(95% CI)	%	(95% CI)	%	(95% CI)	n
Overall	46.0	(45.4–46.6)	39.2	(38.6–39.8)	5.2	(5.0–5.5)	9.6	(9.3–10.0)	26,488
Year of diagnosis									
2004	43.2	(41.7–44.8)	41.4	(39.9–43.0)	6.4	(5.6–7.2)	9.0	(8.1–9.9)	3904
2005	43.7	(42.1–45.2)	42.3	(40.8–43.8)	6.0	(5.3–6.8)	8.0	(7.2–8.9)	3940
2006	45.2	(43.7–46.7)	40.7	(39.2–42.2)	5.2	(4.5–5.9)	8.9	(8.1–9.8)	4103
2007	44.4	(42.9–45.9)	41.4	(39.9–42.9)	5.8	(5.1–6.6)	8.3	(7.5–9.2)	4122
2008	47.5	(46.1–48.9)	38.7	(37.3–40.1)	4.7	(4.1–5.3)	9.1	(8.3–10.0)	4737
2009	49.8	(48.5–51.1)	33.2	(32.0–34.4)	3.9	(3.4–4.4)	13.1	(12.3–14.0)	5682
Gender									
Female	46.0	(45.4–46.6)	39.2	(38.6–39.8)	5.2	(5.0–5.5)	9.6	(9.3–10.0)	26,350
Male	37.7	(29.9–46.0)	43.5	(35.4–51.8)	7.2	(3.8–12.5)	11.6	(7.1–17.7)	138
Age group (years)									
0–39	40.1	(37.1–43.2)	49.3	(46.2–52.4)	4.0	(2.9–5.4)	6.5	(5.1–8.2)	994
40–59	45.8	(44.8–46.8)	44.9	(43.9–45.9)	3.6	(3.3–4.0)	5.7	(5.3–6.2)	9740
60–79	48.5	(47.6–49.3)	37.8	(36.9–38.6)	5.8	(5.4–6.2)	8.0	(7.5–8.5)	12,632
≥80	38.2	(36.6–40.0)	24.0	(22.6–25.6)	8.4	(7.4–9.4)	29.4	(27.8–31.0)	3122
CCI score									
Low	46.2	(45.5–46.9)	40.8	(40.1–41.5)	4.9	(4.6–5.2)	8.1	(7.8–8.5)	20,274
Medium	46.2	(44.9–47.6)	34.2	(32.9–35.5)	5.8	(5.2–6.4)	13.8	(12.9–14.8)	5100
High	40.4	(37.5–43.3)	32.5	(29.8–35.3)	9.2	(7.7–11.1)	17.9	(15.7–20.2)	1114

Note: [a]Based on the staging algorithm (Appendix 1).
Abbreviations: CI, confidence interval; CCI, Charlson Comorbidity Index; TNM, Tumor, Node, Metastasis classification.

diagnostic assessment related to age, comorbidity and disabilities, patient preferences, apparent signs of distant metastases obviating the need for staging in treatment decisions, or competing medical conditions requiring treatment prior to cancer therapy. However, cautious interpretation of our results is necessary, as age and comorbidity are usually correlated.

TNM completeness does not necessarily reflect the clinical reality. For example, a breast cancer case who for various reasons was not treated surgically might be coded Nx, although axillary lymph node biopsy revealed positive lymph node metastases. As a second example, for breast cancer treated with curative intent surgery, no other diagnostic work-up other than a lung x-ray is performed. These cases are likely registered as M0 by most surgeons, although some might prefer to denote them as Mx.

Our findings may have important implications for studies using information on breast cancer stage. Because completeness was lower among the elderly and those with a high level of comorbidity, the data on TNM stage was not missing at random. Consequently, study results may potentially be biased when staging constitutes the exposure or a confounding factor, and this may lead to incorrect conclusions. Thus, patients with missing data on TNM stage should be carefully

handled in the statistical analyses, eg, by applying missing imputation methods,[18] or bias analysis.[19] For instance, population-based cancer registry data is important for monitoring the effectiveness of mammography screening programs. Stage-specific analyses are especially crucial, and stage-recording practices may affect the evaluation of screening program and trends in stage-specific incidence and prognosis. However, most patients with screening mammography-detected breast cancer undergo surgery, and consequently have high TNM completeness.

In conclusion, our study showed that completeness of TNM registration for breast cancer varied differentially with age and level of comorbidity. Therefore, careful consideration should be given to the methodological implications in studies of cancer-related outcomes using data from the DCR.

Acknowledgment

Disclosure

The authors report no conflicts of interest in this work.

References

1. Jemal A, Bray F, Center MM, Ferlay J, Ward E, Forman D. Global cancer statistics. *CA Cancer J Clin*. 2011;61(2):69–90.
2. Tryggvadottir L, Gislum M, Bray F, et al. Trends in the survival of patients diagnosed with breast cancer in the Nordic countries 1964–2003 followed up to the end of 2006. *Acta Oncol*. 2010;49(5):624–631.
3. Vejborg I, Mikkelsen E, Garne JP, et al. Mammography screening in Denmark. *Dan Med Bull*. 2011;58(6):C4287.
4. Jemal A, Siegel R, Ward E, et al. Cancer statistics, 2008. *CA Cancer J Clin*. 2008;58(2):71–96.
5. Storm HH, Michelsen EV, Clemmensen IH, Pihl J. The Danish Cancer Registry – history, content, quality and use. *Dan Med Bull*. 1997;44(5): 535–539.
6. Gjerstorff ML. The Danish Cancer Registry. *Scand J Public Health*. 2011;39(Suppl 7):42–45.
7. Page DL, Fritz AG, Balch CM, et al. *AJCC Cancer Staging Manual*. 6th ed. New York, NY: Springer; 2002.
8. Merrill RM, Sloan A, Anderson AE, Ryker K. Unstaged cancer in the United States: a population-based study. *BMC Cancer*. 2011;11:402.
9. Pedersen CB. The Danish Civil Registration System. *Scand J Public Health*. 2011;39(Suppl 7):22–25.
10. Lynge E, Sandegaard JL, Rebolj M. The Danish National Patient Register. *Scand J Public Health*. 2011;39(Suppl 7):30–33.
11. Charlson ME, Pompei P, Ales KL, MacKenzie CR. A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. *J Chronic Dis*. 1987;40(5):373–383.
12. Jensen AR, Overgaard J, Storm HH. Validity of breast cancer in the Danish Cancer Registry. A study based on clinical records from one county in Denmark. *Eur J Cancer Prev*. 2002;11(4):359–364.
13. Thygesen SK, Christiansen CF, Christensen S, Lash TL, Sorensen HT. The predictive value of ICD-10 diagnostic coding used to assess Charlson comorbidity index conditions in the population-based Danish National Registry of Patients. *BMC Med Res Methodol*. 2011;11:83.
14. Koroukian SM, Xu F, Beaird H, Diaz M, Murray P, Rose JH. Complexity of care needs and unstaged cancer in elders: a population-based study. *Cancer Detect Prev*. 2007;31(3):199–206.
15. Yancik R, Wesley MN, Ries LA, Havlik RJ, Edwards BK, Yates JW. Effect of age and comorbidity in postmenopausal breast cancer patients aged 55 years and older. *JAMA*. 2001;285(7):885–892.
16. Patnaik JL, Byers T, Diguiseppi C, Denberg TD, Dabelea D. The influence of comorbidities on overall survival among older women diagnosed with breast cancer. *J Natl Cancer Inst*. 2011;103(14):1101–1111.
17. Ring A, Sestak I, Baum M, et al. Influence of comorbidities and age on risk of death without recurrence: a retrospective analysis of the arimidex, tamoxifen alone or in combination trial. *J Clin Oncol*. 2011;29(32): 4266–4272.
18. Eisemann N, Waldmann A, Katalinic A. Imputation of missing values of tumour stage in population-based cancer registration. *BMC Med Res Methodol*. 2011;11:129.
19. Lash TL, Silliman RA. A sensitivity analysis to seperate bias due to confounding from bias due to predicting misclassification by a variable that does both. *Epidemiology*. 2000;11(5):544–549.

I5

Appendix table

Appendix Algorithm for breast cancer staging according to the TNM classification

Tumor stage			
Localized	T1–4	N0	M0
	T1–2	N0	Mx
	T1	Nx	M0x
Regional	T1–4,x	N1–3	M0
Distant	Any T	Any N	M1
Unknown	T2–4,x	Nx	M0–x
	T3–4,x	N0	Mx
	T1–4,x	N1–3	Mx
	T0	N1–3	M0–1x
	T0	N0,x	M1

Note: There were six patients with T0 tumors (no evidence of primary tumor). These were all categorized in the unknown category.
Abbreviation: TNM, Tumor, Node, Metastasis.