Cancer Epidemiology 59 (2019) 129–134

Contents lists available at ScienceDirect

Cancer Epidemiology

journal homepage: www.elsevier.com/locate/canep

A validated algorithm for register-based identification of patients with recurrence of breast cancer—Based on Danish Breast Cancer Group (DBCG) data

Linda Aagaard Rasmussen[a,b,*], Henry Jensen[a], Line Flytkjær Virgilsen[a], Lise Bech Jellesmark Thorsen[c,d], Birgitte Vrou Offersen[c,d], Peter Vedsted[a]

[a] Research Centre for Cancer Diagnosis in Primary Care (CaP), Research Unit for General Practice, Bartholins Allé 2, 8000 Aarhus C, Denmark
[b] Department of Public Health, Aarhus University, Bartholins Allé 2, 8000 Aarhus C, Denmark
[c] Department of Experimental Clinical Oncology, Aarhus University Hospital, Noerrebrogade 44, 8000 Aarhus C, Denmark
[d] Department of Oncology, Aarhus University Hospital, Noerrebrogade 44, 8000 Aarhus C, Denmark

ARTICLE INFO

Keywords:
Breast neoplasms
Recurrence
Algorithms
Validation studies
Registries
Denmark

ABSTRACT

Background: Cancer recurrence is not routinely and completely registered in Danish national health registers, which challenges register-based research. The aim of this study was to develop and validate a register-based algorithm to identify patients with recurrence of breast cancer (BC).

Methods: We conducted a cohort study based on data from Danish national health registers and used the Danish National Patient Register and the Danish National Pathology Register as sources to identify BC recurrence. We used data from the Danish Breast Cancer Group (DBCG) validated against medical records on recurrence status and recurrence date for 471 women with early stage unilateral BC as the gold standard of BC recurrence to assess the accuracy of the algorithm to identify BC recurrence.

Results: The algorithm displayed a sensitivity of 97.3% (95% confidence interval (CI): 93.2–99.3), a specificity of 97.2% (95% CI: 94.8–98.7) and a positive predictive value of 94.4% (95% CI: 89.2–97.3). The concordance correlation coefficient for the agreement between recurrence dates generated by the algorithm and the gold standard was 0.97 (95% CI: 0.96-0.98), and the date was estimated within +/−30 days of the gold standard in 66% of the patients and within +/−60 days in 76% of the patients.

Conclusion: The developed algorithm almost perfectly identified BC recurrence and with reasonable timing compared to the gold standard.

1. Introduction

The ageing population and improvements in diagnostics and treatment have increased the number of breast cancer (BC) survivors over the last decades [1–3]. The incidence of BC was 4682 in Denmark in 2015, which corresponds to an incidence rate of 123.2 per 100,000 (age-standardised European rate) [4]. International studies report recurrence rates of 12–49% in patients diagnosed with BC, depending on tumour stage, adjuvant therapy and follow-up time [5–9].

Insight into when, where and how cancer recurrence presents is essential to provide optimal care for cancer survivors [3,10,11]. Recurrence is not routinely registered in most patient registers, and identification of patients with cancer recurrence remains a challenge. BC recurrence is registered in the clinical cancer database by the Danish Breast Cancer Group (DBCG). Yet, these registrations are incomplete, and follow-up ends after 10 years [12].

Three studies have reported on the identification of BC recurrence through administrative data. A recent Danish study yielded a sensitivity (SEN) of 88% and a positive predictive value (PPV) of 73% [13]. A UK study identified 92% of all recurrences [14], and a US study yielded a SEN of 89% and a specificity (SPE) of 99% [6]. Other studies have developed algorithms to identify recurrence of two or more cancers,

Abbreviations: BC, breast cancer; CCC, concordance correlation coefficient; CI, confidence interval; DBCG, Danish Breast Cancer Group; DCR, Danish Cancer Register; GDPR, general data protection regulations; GP, general practitioner; ICD-10, International Classification of Diseases, 10th revision; IQR, inter quartile range; NPaR, National Pathology Register; NPR, National Patient Register; NPV, negative predictive value; PPV, positive predictive value; SEN, sensitivity; SNOMED, systematized nomenclature of medicine; SPE, specificity; TNM, tumor node metastasis; VPN, virtual private network

* Corresponding author at: Bartholins Allé 2, 8000 Aarhus C, Denmark.
E-mail address: linda.rasmussen@ph.au.dk (L. Aagaard Rasmussen).

https://doi.org/10.1016/j.canep.2019.01.016
Received 24 October 2018; Received in revised form 20 December 2018; Accepted 27 January 2019
Available online 08 February 2019
1877-7821/

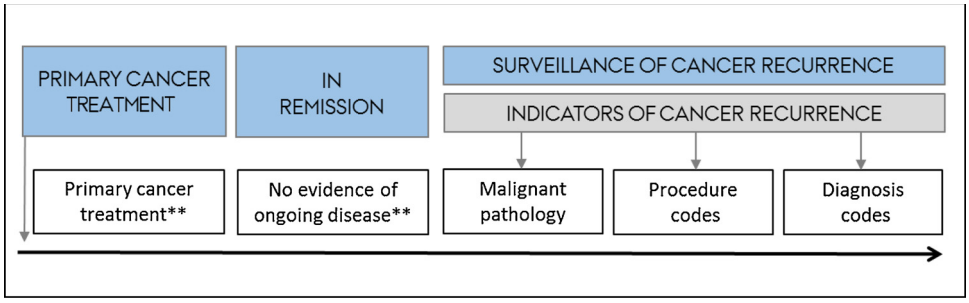

Fig. 1. Schematic overview of the algorithm.

including BC [15–18]. These studies have been criticised for moderate performance, infrequent recurrence, small sample size and sampling of non-representative populations from academic centres and single institutions [11].

Two recent Danish studies have reported promising results on register-based algorithms to identify patients with recurrence of colorectal cancer [19] and bladder cancer [20]. This indicates a possibility to improve the performance of an algorithm to identify women with BC recurrence.

The objective of this study was to develop and validate an alternative register-based algorithm to identify patients with recurrence of BC in a population of women curatively treated for BC in Denmark. Furthermore, the study aimed to assess the accuracy of the cancer recurrence diagnosis date derived from the algorithm.

2. Material and methods

We conducted a cohort study based on data from Danish national health registers. The unique personal registration number assigned to all Danish residents was used to link data at the individual level [21].

2.1. Data sources

Data were retrieved from four Danish national registers. We obtained vital status and migration from the Danish Civil Registration System [21], BC diagnosis and tumour stage from the Danish Cancer Register (DCR) [22], and information on tumour stage, procedure and diagnosis codes for all cancer-related contacts in the Danish National Patient Register (NPR) [23]. The Danish National Pathology Register (NPaR) [24] provided data on Systematized Nomenclature of Medicine (SNOMED) classification registrations [25], which allowed identification of malignant morphology (codes M8 and M9). The fifth digit of the morphology code indicates behaviour (e.g. 4: “direct spread to surrounding tissue”, 6: “malignant metastasis” and 7: “malignant recurrence”).

2.2. Gold standard

The gold standard of BC recurrence was based on data from a study by Thorsen et al. [26] and DBCG [27] data on 529 Danish women operated for unilateral early-stage node-positive BC in 2003-2007. To determine if the women had BC recurrence, Thorsen and colleagues reviewed DBCG registrations of recurrence and additionally, the NPaR was reviewed for registrations of recurrence according to the SNOMED classification. If no recurrence was registered, the general practitioner (GP) was asked about recurrence. If lack of recurrence was confirmed by the GP, the patient was categorised with no recurrence. If recurrence was reported by the GP or the GP did not reply, medical records were reviewed to confirm or exclude recurrence. The gold standard did not distinguish between recurrence and second primary BC as these two events are often pooled together in research on disease-free survival [6].

2.3. Study population

Patients were eligible for inclusion if registered in the NPR with a lumpectomy or mastectomy between 60 days before and 200 days after the date of the BC diagnosis, which allowed for a period of preoperative oncological treatment. Patients were excluded if registered in the DCR with a previous cancer (except for non-melanoma skin cancer) or registered in the DCR or the NPR with distant metastasis within 90 days of surgery (C78, C79 or CxxxM of the International Classification of Diseases, 10th revision (ICD-10) or distant tumour stage based on the Tumor, Node, Metastasis (TNM) classification system) [28]. Patients were excluded if aged younger than 18 years or had emigrated or died within 90 days of first cancer treatment.

2.4. The algorithm

The end date of primary treatment was defined as the date of lumpectomy or mastectomy, or the date of the last radiotherapy or chemotherapy in case of adjuvant oncological treatment (Fig. 1). Adjuvant chemotherapy and radiotherapy was defined by procedure codes in the NPR with an indication diagnosis of BC (C50*) or relevant lymph node diagnosis (C773, C779). Adjuvant radiotherapy was defined as > 15 fractions within intervals of no more than 30 days. Adjuvant chemotherapy was defined as 5–7 series within intervals of no more than 60 days. Furthermore, adjuvant radiotherapy could follow adjuvant chemotherapy if initiated within 90 days after ended chemotherapy, or vice versa.

A subsequent period without register-based evidence of ongoing disease was required to ensure that the patient was in remission. The final day of this period was 90 days after surgery or 30 days after ended adjuvant treatment, whichever came last. Indicators of ongoing disease was: 1) a SNOMED code with M8-M9 morphology, 2) a procedure code of radiotherapy or chemotherapy with a malignant indication diagnosis (ICD-10: C00-C96 and D37-D48), 3) a surgical procedure code of mastectomy, lumpectomy or lymph node resection (apart from axillary lymph nodes within 30 days after primary surgery) with a malignant indication diagnosis, 4) a distant tumour stage based on TNM registration or 5) a malignant diagnosis code. Diagnosis codes of BC and non-melanoma skin cancer were exempted.

If the patient was assessed to be in remission, the cancer recurrence surveillance period began. A patient was classified by the algorithm to have BC recurrence if one of the following five indicators was present: 1) registration in the NPaR of SNOMED morphology codes M8-M9 with 4, 6 or 7 in the fifth digit and a morphology similar to a morphology code registered in the NPaR within 90 days of the primary BC diagnosis date or the date of primary BC surgery, 2) registration in the NPR of radiotherapy procedure codes with BC (C50*) or metastases (C76-C79 and CxxxM) as the indication diagnosis, 3) registration in the NPR of surgical procedure codes for BC surgery (resection of breast (KHAB), mastectomy (KHAC), surgery for local recurrence of BC (KHAF)) with C50* as the indication diagnosis, 4) registration in the NPR of an ICD-10 diagnosis code of BC recurrence (C509X) or 5) registration in the

NPR of an ICD-10 diagnosis code of distant metastasis (C76-C79 and CxxxM, except for C775 and C779, which may refer to regional lymph nodes at the time of primary diagnosis).

If a second primary cancer diagnosis other than C50* was present in the DCR or in the NPR before or within 30 days of a metastasis diagnosis code, the metastasis diagnosis code was disregarded as it could be related to the second primary cancer. The recurrence date estimated by the algorithm was defined as the date with a registration of an indicator of recurrence. If the same patient had more than one indicator of cancer recurrence, the first registered date was regarded as the date of cancer recurrence.

2.5. Analyses

SEN, SPE, PPV and negative predictive value (NPV) with 95% confidence interval (CI) were generated using the concordant and discordant frequencies between recurrences identified by the algorithm and the gold standard. Furthermore, Cohen's kappa coefficient test for agreement was calculated. Stratified analyses of SEN, SPE, PPV and NPV were conducted for recurrences identified by indicators of 1) a pathology test result, 2) a cancer-related procedure and 3) a diagnosis of metastasis or recurrence.

The following sensitivity analyses were conducted: 1) minimum period from BC surgery to recurrence surveillance was increased to 180 days, 2) stratification on year of the gold standard recurrence date (2003–2006, 2007–2009 and 2010–2016) and 3) stratification on age at primary BC surgery (18–49 years, 50–59 years and 60–69 years).

Lin's concordance correlation coefficient (CCC) [29] assessed the strength of agreement between the date of recurrence identified by the algorithm and the gold standard. The agreement is considered "substantial" for CCC > 0.95 and "almost perfect" for CCC > 0.99 [30]. Data were analysed using Stata 15.1 statistical software (StataCorp LP, College Station, TX, USA).

Table 1
Patient characteristics stratified on cancer recurrence status in the gold standard population.

	Cancer recurrence n (%)	No cancer recurrence n (%)	All patients n (%)
N	149 (100)	322 (100)	471 (100)
Age groups, years			
-18–49	50 (34)	97 (30)	147 (31)
-50–59	56 (38)	124 (39)	180 (38)
-60–69	43 (29)	101 (31)	144 (31)
Tumour stage of primary cancer			
-Regional lymph node positive	94 (63)	228 (71)	322 (68)
-Missing tumour stage	55 (37)	94 (29)	149 (32)
Comorbidity at primary cancer[a]			
-CCI: 0	46 (31)	129 (40)	175 (37)
-CCI: 1–2	6 (4)	21 (7)	27 (6)
-CCI: 3+	97 (65)	172 (53)	269 (58)

[a] Charlson Comorbidity Index at the time of the primary breast cancer diagnosis.

3. Results

A total of 487 patients fulfilled the inclusion criteria (Fig. 2), and 471 patients constituted the final study population; 149 (32%) of these had cancer recurrence according to the gold standard.

The characteristics of the population stratified by cancer recurrence status according to the gold standard are presented in Table 1. The median follow-up time for the gold standard population was 7.5 years (inter quartile range (IQR): 5–9) since the primary BC surgery.

3.1. Performance of the algorithm to detect BC recurrence

The algorithm yielded a SEN of 97% (95% CI: 93–99) and a SPE of

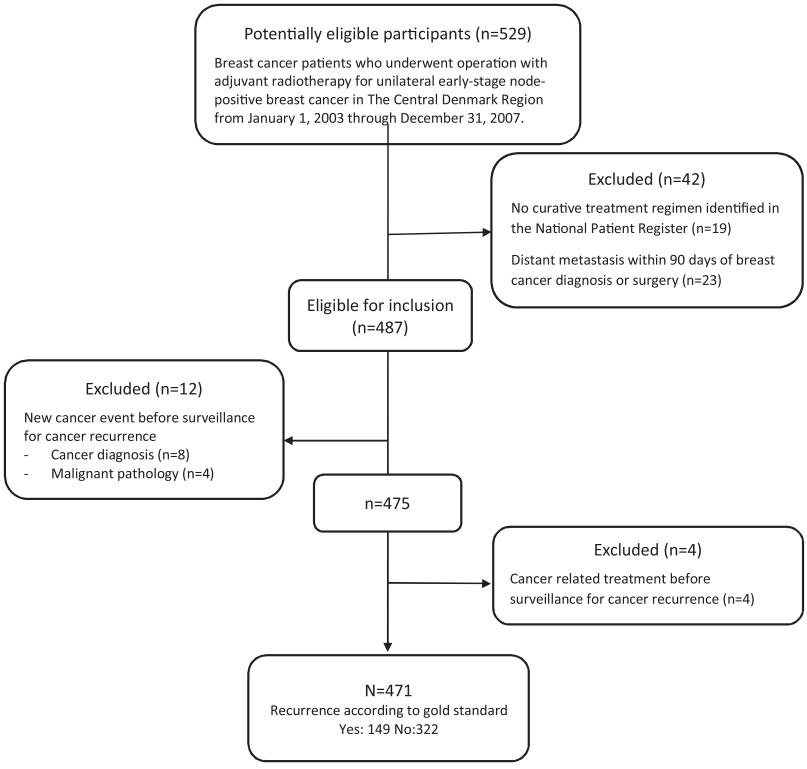

Fig. 2. Flowchart of study population.

Table 2
Performance of the algorithm for breast cancer recurrence.

Algorithm performance	% (95% CI) [a]
Sensitivity	97.3 (93.2–99.3)
Specificity	97.2 (94.8–98.7)
Positive predictive value	94.2 (89.2–97.3)
Negative predictive value	98.7 (96.8–99.7)
Kappa agreement	97.2 (95.3–98.5)
Kappa, ƙ (95% CI)	0.94 (0.90–0.97)

CI: confidence interval.
[a] % (95% confidence interval) if nothing else stated.

Table 3
Agreement between recurrence identified by the gold standard and the algorithm.

	Gold standard recurrences		
Algorithm recurrences	Yes	No	Total
Yes	145	9	154
No	4	313	317
Total	149	322	471

97% (95% CI: 95–99) (Table 2).

Of the 154 women classified with recurrence by the algorithm, 145 were true positives and 9 were false positives; false positive rate: 1.9% (Table 3).

The performance varied across the individual indicators of recurrence; the highest SEN was seen at 87% for diagnosis codes of metastasis and recurrence (Table 4). The majority of false positives were due to registration of radiotherapy. When omitting radiotherapy as an indicator of recurrence, the SEN decreased to 94%, and the PPV increased to more than 97% (precise estimates are not reported due data protection regulations for less than 3 observations).

The sensitivity analyses showed similar results as the overall analysis. The only difference was that the SEN tended to be higher in women aged 18–49 years compared to older age groups, with a dose-response trend across age groups. The reverse was found for SPE (Table A.1 in Appendix).

3.2. Concordance of recurrence date

The recurrence date derived by the algorithm showed substantial concordance with the recurrence date of the gold standard (CCC: 0.97, 95% CI: 0.96-0.98) (Figure A.1, Table A.2 in Appendix). The median difference between the recurrence date estimated by the algorithm and the recurrence date of the gold standard was 17 days (IQR: 6–56 days). Indicators from pathology results estimated the recurrence date most accurately; 65% of the cases were set at +/- 30 days of the gold standard recurrence date. The corresponding proportion of indicators for diagnosis codes and procedure codes was 53% and 40%, respectively.

4. Discussion

4.1. Main findings

We developed and validated a register-based algorithm to identify BC patients with recurrence in Denmark. The algorithm identified 97% of the patients with recurrence according to the gold standard and reached a positive predictive value of 94%. The recurrence dates generated by the algorithm showed substantial agreement with the recurrence dates of the gold standard.

4.2. Strengths and limitations

The most important strength of this study is that the algorithm is based on the continuously updated and high quality Danish national registries, which ensures complete data of high validity [21–24]. A further strength is the inclusion of pathology test results as an indicator of recurrence in the algorithm, which increases the SEN and the accuracy of the estimated recurrence date. The Danish tax-funded public healthcare system with free and equal access for all citizens minimised the risk of selection bias. Furthermore, the DBCG [27] provides national guidelines on BC diagnosing, treatment and follow-up, which ensures uniform national procedures and strengthens the generalisability of the algorithm.

Another important strength is that the gold standard originated from a population-based cohort study with limited exclusion criteria [26]. This ensures higher generalisability to the entire BC population compared to using data from a randomised controlled trial as a gold standard as these often include a more restricted, younger and healthier population with uniform pathways [31]. The information on recurrence in the gold standard is based on registrations of recurrence in the DBCG registry and the NPaR (pathology results), in addition to information from the GP and/or medical record review. As the pathology examination is considered the gold standard against which all clinical conclusions are compared [24], and it is mandatory to send a discharge summery to the GP after hospital admissions, the risk of misclassification of recurrence status in the gold standard was minimal. Thus, the gold standard is considered to be highly valid.

The sensitivity analyses indicated that the algorithm was robust to changes over time. However, corrections could be necessary if major changes are introduced in the treatment regimens for primary BC and BC recurrence. Finally, the evaluation of the accuracy of the recurrence date strengthens the study.

The limitations of the study primarily concern the risk of misclassification of recurrence and recurrence dates from missing or

Table 4
Performance of indicators of breast cancer recurrence.

	Pathology codes[a] % (95% CI[d])	Procedural codes[b] % (95% CI)	Diagnosis codes[c] % (95% CI)
Sensitivity	> 34.0 (26.5–42.2)[e]	64.4 (56.2–72.1)	86.6 (80.0–91.6)
Specificity	> 99.1 (97.3–99.8) [e]	97.8 (95.6–99.1)	100 (98.9–100)
Positive predictive value	> 94.4 (84.6–98.8) [e]	93.2 (86.5–97.2)	100 (97.2–100)
Negative predictive value	> 76.3 (71.9–80.3) [e]	85.6 (81.6–89.0)	94.2 (91.1–96.4)

[a] Histology similar to primary breast cancer histology and coded as recurrent, metastatic or direct spread to surrounding tissue.
[b] Radiotherapy and breast cancer surgery.
[c] C76-C79 (except C775 and C779), CxxxM and C509X.
[d] CI: confidence interval.
[e] Very few false positives were identified by pathology codes. To adhere to data protection regulations, the number of false positives was increased to three, and the rest of the numbers were changed accordingly to estimate sensitivity, specificity, positive predictive value and negative predictive value, i.e. the true estimates were higher, as indicated by the > sign.

incorrect registrations. Missing data are of less concern as patients with BC recurrence are likely to be in contact with a hospital. However, comorbid and frail women may be at higher risk of being missed by the algorithm as their delicate state may contraindicate biopsy and cancer treatment. This is supported by the tendency to higher SEN in the younger age groups.

A period of 90 days with no evidence of active disease after the last day of treatment was required to identify patients in remission after treatment of the primary cancer (Fig. 1). This arbitrary cut-off could have led to misclassification of disease status. All false positives identified by radiotherapy procedure codes appeared shortly after the end of the disease-free period and are expected to concern women with delayed initiation of adjuvant radiotherapy after definitive surgery. Omitting radiotherapy as indicator of recurrence increased the PPV and decreased the SEN. Another approach to avoid these false positives could be to prolong the required period with no evidence of ongoing disease, although our sensitivity analysis showed similar results.

The algorithm missed four women with distant recurrence according to the gold standard. Four of these women received a malignant pathology test result within 45 days of the recurrence date defined by the gold standard. However, only two were registered with the same morphology as for the primary BC although denoted" primary cancer" by the fifth digit in the SNOMED code. Hence, all four women were missed by the algorithm. This illustrates the importance of correct registrations to achieve high performance of the algorithm. Furthermore, few of these women died within one week of the recurrence date, which indicates no treatment options and minimal diagnostic workup.

The gold standard population comprised of early-stage node-positive BC patients. These women have higher risk of recurrence than node-negative BC patients, which is reflected in the recurrence rate of 32% in the present study. This may have provided a higher PPV in the current study population compared to a corresponding node-negative gold standard population. Furthermore, the gold standard did not distinguish between recurrence and second primary BC. One should be aware of the bias that this may introduce, depending on the focus of the research.

4.3. Comparison with other studies

A recently published algorithm by Cronin-Fenton [20] reported lower SEN and PPV than the present study, although the study by Cronin-Fenton and colleagues was based on the same registers. Our superior SEN could be caused by the recurrence indicators as, in our study, these were also based on radiotherapy, BC surgery procedure codes, and the diagnosis code for breast metastasis (C509 M). In an effort to avoid false positives we required a pathology specimen to be of the same morphology as the primary BC, and the fifth position in the SNOMED code to be denoted 4, 6 or 7; this may explain our superior PPV compared to Cronin-Fenton et al.

Studies from the UK and the US have attempted to identify BC recurrence using administrative data [6,14–16,18,32]. When Chubak et al. [6] evaluated radiotherapy as an indicator of recurrence, only 40–50% of these patients had recurrence according to the gold standard. We required a co-diagnosis of BC, metastasis or BC recurrence for procedural codes as indicators and reached a PPV of 93% (Table 4).

It has been suggested that metastasis codes are not well suited for identification of recurrent cancer [17]. Nevertheless, the diagnosis codes for metastasis and recurrence reached a SEN of 87% in our study (Table 4). Two previous studies have reported PPVs of 87% and 94% when using codes for metastasis and second primary cancer [6,16]. In our study, a metastasis diagnosis code was disregarded as indicator of recurrence if a second primary cancer was diagnosed before the metastasis as we could not determine if the metastasis originated from the BC or the second primary cancer. This may have caused the high PPV of 100% (Table 4) when evaluating diagnosis codes alone.

Hasset et al. [16] included diagnosis codes of secondary malignant neoplasms and chemotherapy codes as indicators of colorectal, breast and prostate cancer recurrence. For BC, the algorithm reached a SEN of 80%, but only a PPV of 30%. For chemotherapy alone, the PPV was as low as 11%. Chemotherapy appears to be a source of many false positives. Therefore, we disregarded it as an indicator of recurrence in the present study. In Warren et al. [15], the inclusion criteria and the gold standard of recurrence was death of BC. Hence, women successfully treated for recurrence were not included in the study. The study is also disadvantaged by only including patients with recurrence according to the gold standard as it excludes assessment of the SPE and PPV of the algorithm. These results underline the importance of having a representative gold standard population and considering both SPE and SEN when evaluating the performance of an algorithm. Wide criteria for indicators of recurrence would yield a high SEN, but too wide criteria would generate false positives and introduce bias.

Only two studies have assessed the accuracy of determining a date of recurrence through an algorithm [16,32]. In Hasset et al. [16], recurrence dates produced by their algorithm fell within 30 days of the gold standard date in 20–36% of the cases. We managed to identify the recurrence date more accurately; 66% of the dates fell within 30 days of the gold standard recurrence date (Table A.2 in Appendix). Our determination of a recurrence date differed by a median of 17 days compared to the gold standard, which was much less than the median difference of 40 days reported by Lamont et al. [32]. The inclusion of pathology codes as a recurrence indicator in our study may explain why our algorithm provided more accurate estimates of recurrence date.

5. Conclusion

We have developed an alternative algorithm to identify breast cancer recurrence using Danish routine healthcare data. The algorithm showed very high sensitivity and specificity. Furthermore, the algorithm defined the recurrence date fairly accurately. Thus, the algorithm could be an important instrument for future research in the field of breast cancer recurrence.

Ethics

The patient pathway for cancer recurrence research project (Study 1, id 119) has been approved and is registered in the Record of Processing Activities at the Research Unit of General Practice in Aarhus in accordance with the provisions of the General Data Protection Regulation (GDPR).

This type of study did not require approval from the Committee on Health Research Ethics in the Central Denmark Region as the study did not involve human biological material.

Author contributions

All authors contributed to the conception and design of the study or to the data acquisition. LR performed the statistical analysis and drafted the manuscript. All authors interpreted the data. HJ, LFV, LBJT, BVO and PV provided critical revision of the manuscript for important intellectual content. All authors read and approved the final version of the manuscript.

Data availability

The data supporting the findings of this study are stored and maintained electronically at Statistics Denmark and can only be accessed by approved collaborative partners through a secured virtual private network (VPN). In line with confidentiality provisions, the data are not publicly available as they contain information that could compromise the privacy of the research participants.

Funding

This work was supported by the Danish Cancer Society and Aarhus University. The funders had no role in the study.

Declarations of interest

None.

Acknowledgement

Not applicable.

Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at doi:https://doi.org/10.1016/j.canep.2019.01.016.

References

[1] H.B. Hovaldt, N.P. Suppli, M.H. Olsen, M. Steding-Jessen, D.G. Hansen, H. Moller, et al., Who are the cancer survivors? A nationwide study in Denmark, 1943-2010, Br. J. Cancer 112 (April (9)) (2015) 1549–1553, https://doi.org/10.1038/bjc.2015.68.

[2] Danish Health Authority, Nye Kræfttilfælde I Danmark [Cancer Incidence in Denmark 2016], (2016) Available at:https://sundhedsdatastyrelsen.dk/da/tal-og-analyser/analyser-og-rapporter/sygdomme/cancerregisteret Accessed 15 October, 2018.

[3] G. Rubin, A. Berendsen, S.M. Crawford, R. Dommett, C. Earle, J. Emery, et al., The expanding role of primary care in cancer control, Lancet Oncol. 16 (September (12)) (2015) 1231–1272.

[4] NORDCAN, The NORDCAN Project. Cancer statistics for the Nordic Countries, (2019) Available at:http://www-dep.iarc.fr/NORDCAN/English/frame.asp Accessed 15 October, 2018.

[5] A. Bodilsen, B.V. Offersen, P. Christiansen, J. Overgaard, Pattern of relapse after breast conserving therapy, a study of 1519 early breast cancer patients treated in the Central Region of Denmark 2000-2009, Acta Oncol. 55 (August (8)) (2016) 964–969.

[6] J. Chubak, O. Yu, G. Pocobelli, L. Lamerato, J. Webster, M.N. Prout, et al., Administrative data algorithms to identify second breast cancer events following early-stage invasive breast cancer, J. Natl. Cancer Inst. 104 (June (12)) (2012) 931–940.

[7] Early Breast Cancer Trialists' Collaborative Group (EBCTCG), S. Darby, P. McGale, C. Correa, C. Taylor, R. Arriagada, et al., Effect of radiotherapy after breast-conserving surgery on 10-year recurrence and 15-year breast cancer death: meta-analysis of individual patient data for 10,801 women in 17 randomised trials, Lancet 378 (Novemeber (9804)) (2011) 1707–1716.

[8] M.B. Jensen, T.O. Nielsen, A.S. Knoop, A.V. Laenkholm, E. Balslev, B. Ejlertsen, Mortality and recurrence rates among systemically untreated high risk breast cancer patients included in the DBCG 77 trials, Acta Oncol. 57 (January (1)) (2018) 135–140.

[9] T. Laurberg, C.D. Lyngholm, P. Christiansen, J. Alsner, J. Overgaard, Long-term age-dependent failure pattern after breast-conserving therapy or mastectomy among Danish lymph-node-negative breast cancer patients, Radiother. Oncol. 120 (July (1)) (2016) 98–106.

[10] I. Soerjomataram, J.W. Coebergh, Epidemiology of multiple primary cancers, Methods Mol. Biol. 471 (2009) 85–105.

[11] J.L. Warren, K.R. Yabroff, Challenges and opportunities in measuring cancer recurrence in the United States, J. Natl. Cancer Inst. 107 (May (8)) (2015) 10–17.

[12] P. Christiansen, N. Al-Suliman, K. Bjerre, S. Moller, Danish Breast Cancer Cooperative Group, Recurrence pattern and prognosis in low-risk breast cancer patients—data from the DBCG 89-A programme, Acta Oncol. 47 (July (4)) (2009) 691–703.

[13] D. Cronin-Fenton, A. Kjaersgaard, M. Norgaard, J. Amelio, A. Liede, R.K. Hernandez, et al., Breast cancer recurrence, bone metastases, and visceral metastases in women with stage II and III breast cancer in Denmark, Breast Cancer Res. Treat. 167 (January (2)) (2018) 517–528.

[14] G.S. Mannu, J. Broggio, J. Charman, S. Darby, Identifying recurrence in breast cancer patients from routinely collected data in England, Eur. J. Surg. Oncol. 42 (May (5)) (2016) S33–S34.

[15] J.L. Warren, A. Mariotto, D. Melbert, D. Schrag, P. Doria-Rose, D. Penson, et al., Sensitivity of medicare claims to identify cancer recurrence in elderly colorectal and breast cancer patients, Med. Care 54 (August (8)) (2016) e47–e54.

[16] M.J. Hassett, D.P. Ritzwoller, N. Taback, N. Carroll, A.M. Cronin, G.V. Ting, et al., Validating billing/encounter codes as indicators of lung, colorectal, breast, and prostate cancer recurrence using 2 large contemporary cohorts, Med. Care 52 (October (10)) (2014) e65–e73.

[17] N. Chawla, K.R. Yabroff, A. Mariotto, T.S. McNeel, D. Schrag, J.L. Warren, Limited validity of diagnosis codes in medicare claims for identifying cancer metastases and inferring stage, Ann. Epidemiol. 24 (September (9)) (2014) 666–672.

[18] D. McClish, L. Penberthy, A. Pugh, Using medicare claims to identify second primary cancers and recurrences in order to supplement a cancer registry, J. Clin. Epidemiol. 56 (August (8)) (2003) 760–767.

[19] T.L. Lash, A.H. Riis, E.B. Ostenfeld, R. Erichsen, M. Vyberg, O. Thorlacius-Ussing, A validated algorithm to ascertain colorectal cancer recurrence using registry resources in Denmark, Int. J. Cancer 136 (May (9)) (2015) 2210–2215.

[20] L.A. Rasmussen, H. Jensen, L.F. Virgilsen, J.B. Jensen, P. Vedsted, A validated algorithm to identify recurrence of bladder cancer: a register-based study in Denmark, Clin. Epidemiol. 10 (November) (2018) 1755–1763.

[21] M. Schmidt, L. Pedersen, H.T. Sorensen, The Danish Civil Registration System as a tool in epidemiology, Eur. J. Epidemiol. 29 (August (8)) (2014) 541–549.

[22] M.L. Gjerstorff, The Danish Cancer Registry, Scand. J. Public Health 39 (July (7 Suppl)) (2011) 42–45.

[23] M. Schmidt, S.A. Schmidt, J.L. Sandegaard, V. Ehrenstein, L. Pedersen, H.T. Sorensen, The Danish national patient registry: a review of content, data quality, and research potential, Clin. Epidemiol. 7 (November) (2015) 449–490.

[24] R. Erichsen, T.L. Lash, S.J. Hamilton-Dutoit, B. Bjerregaard, M. Vyberg, L. Pedersen, Existing data sources for clinical epidemiology: the Danish National Pathology Registry and Data Bank, Clin. Epidemiol. 2 (August) (2010) 51–56.

[25] SNOMED international, SNOMED International, Available at:https://www.snomed.org/ Accessed 15 October, 2018 (2019).

[26] L.B. Thorsen, B.V. Offersen, H. Dano, M. Berg, I. Jensen, A.N. Pedersen, et al., DBCG-IMN: a population-based cohort study on the effect of internal mammary node irradiation in early node-positive breast cancer, J. Clin. Oncol. 34 (February (4)) (2016) 314–320.

[27] The Danish Breast Cancer Cooperative Group, The Danish Breast Cancer Cooperative Group, Available at:www.dbcg.dk Accessed 15 October, 2018 (2019).

[28] Union for International Cancer Control, TNM Classification of Malignant Tumours, (2018) Available at:https://www.uicc.org/resources/tnm Accessed 15 October, 2018.

[29] L.I. Lin, A concordance correlation coefficient to evaluate reproducibility, Biometrics 45 (March (1)) (1989) 255–268.

[30] G.B. McBride, A Proposal for Strength of Agreement Criteria for Lin's Concordance Correlation Coefficient, National Institute of Water & Atmospheric Research, Hamilton, New Zealand, 2005.

[31] M.D. Galsky, How I treat bladder cancer in elderly patients, J. Geriatr. Oncol. 6 (January (1)) (2015) 1–7.

[32] E.B. Lamont, J.E. Herndon 2nd, J.C. Weeks, I.C. Henderson, C.C. Earle, R.L. Schilsky, et al., Measuring disease-free survival and cancer relapse using medicare claims from CALGB breast cancer trial participants (companion to 9344), J. Natl. Cancer Inst. 98 (September (18)) (2006) 1335–1338.