



Estimating the incidence of breast cancer recurrence using administrative data

Steven Habbous^{1,2} · Andriana Barisic¹ · Esha Homenauth¹ · Sharmilaa Kandasamy¹ · Katharina Forster¹ · Andrea Eisen^{1,3} · Claire Holloway^{1,4}

Received: 20 September 2022 / Accepted: 9 November 2022 / Published online: 23 November 2022
© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2022

Abstract

Background Breast cancer is the most common cancer among women, but most cancer registries do not capture recurrences. We estimated the incidence of local, regional, and distant recurrences using administrative data.

Methods Patients diagnosed with stage I-III primary breast cancer in Ontario, Canada from 2013 to 2017 were included. Patients were followed until 31/Dec/2021, death, or a new primary cancer diagnosis. We used hospital administrative data (diagnostic and intervention codes) to identify local recurrence, regional recurrence, and distant metastasis after primary diagnosis. We used logistic regression to explore factors associated with developing a distant metastasis.

Results With a median follow-up 67 months, 5,431/45,857 (11.8%) of patients developed a distant metastasis a median 23 (9, 42) months after diagnosis of the primary tumor. 1086 (2.4%) and 1069 (2.3%) patients developed an isolated regional or a local recurrence, respectively. Patients with distant metastatic disease had a median overall survival of 15.4 months (95% CI 14.4–16.4 months) from the time recurrence/metastasis was identified. In contrast, the median survival for all other patients was not reached. Patients were more likely to develop a distant metastasis if they had more advanced stage, greater comorbidity, and presented with symptoms ($p < 0.0001$). Trastuzumab halved the risk of recurrence [OR 0.53 (0.45–0.63), $p < 0.0001$].

Conclusion Distant metastasis is not a rare outcome for patients diagnosed with breast cancer, translating to an annual incidence of 2132 new cases (17.8% of all breast cancer diagnoses). Overall survival remains high for patients with locoregional recurrences, but was poor following a diagnosis of a distant metastasis.

Keywords Breast cancer · Recurrence · Metastasis · Local · Regional · Distant · Locoregional · Administrative data

Introduction

Breast cancer is the most common cancer among women, yet most cancer registries do not capture breast cancer recurrence [1]. This has led to an information gap that hinders health system monitoring and quality improvement efforts for a variety of reasons. First, given the large number of new breast cancer cases diagnosed each year, the burden to diagnose and treat recurrences is expected to be significant [2, 3]. Second, survival from primary breast cancer has improved with early detection and advances in treatment, but this has also resulted in a rise in the number of women living with metastatic disease [4]. Third, understanding the risk factors for recurrence could be used to identify high-risk populations and can also be used to guide patient-informed treatment strategies for the primary carcinoma [5–7].

Andriana Barisic and Esha Homenauth contributed equally.

✉ Steven Habbous
Steven.habbous@ontariohealth.ca

¹ Ontario Health (Cancer Care Ontario), 525 University Ave, Toronto, ON M5G2L3, Canada

² Department of Epidemiology & Biostatistics, Western University, London, ON N6A 5C1, Canada

³ Department of Medical Oncology, Sunnybrook Health Sciences Centre, Toronto, ON M4Y 1H1, Canada

⁴ Department of Surgery, University of Toronto, Toronto, ON M5T1P5, Canada

Most population-level statistics on incidence, risk factors, and mortality are inferred from hospital-based cohort studies or randomized controlled trials, which may not generalize to the population because of differences in surveillance and referral practices [8–11]. Conversely, an algorithmic approach applied to population-level data has the benefits of automation and more comprehensive data capture with less loss to follow-up. To this end, researchers have sought to use administrative data to capture potential recurrences using diagnostic codes, evidence of subsequent treatment, survival estimates, and other indicators of healthcare utilization [12–16]. Validated against a gold standard, typically medical chart review, these methods have demonstrated utility, despite study-to-study variation in the definition of recurrence and the methodology used to identify it.

Leveraging this groundwork, we sought to estimate the population-level incidence of breast cancer recurrences in Ontario, Canada. We further estimate the percent of distant metastasis to the brain, bone, lung, pleura, liver, and other sites.

Methods

Cohort ascertainment

Patients were identified from the Ontario Cancer Registry (OCR) with a breast cancer diagnosis between 01/Jan/2013 and 31/Dec/2017 [17]. Patients were excluded if they were < 18 years of age at diagnosis, were diagnosed after death or by death certificate, had an invalid Ontario health card number, a non-Ontario postal code, or did not access the healthcare system through the Ontario Health Insurance Program (OHIP) within 1 year of diagnosis. We further excluded patients with: 1) lymphomas, sarcomas, or other non-breast histologies ($n = 298$); 2) stage 0, IV, unknown, or missing stage at presentation ($n = 5517$); and 3) unknown laterality for the primary cancer diagnosis ($n = 242$).

Data sources used to identify evidence of breast cancer recurrence

The Discharge Abstract Database (DAD) and the National Ambulatory Care Reporting System (NACRS) collect information for all inpatient and outpatient hospital encounters, respectively. The Activity Level Reporting (ALR) database includes data on radiation as reported by the regional cancer centers in the province. The ALR database also includes information on anti-neoplastic systemic therapy administered in regional cancer centers and community hospitals by drug name. The provincial New Drug Funding Program (NDFP) collects information on newer more costly

intravenous systemic therapies, including the indication as specific for recurrent or metastatic breast cancer.

We looked for evidence of recurrence starting from the date of the primary breast cancer diagnosis until the date of censor, defined as the earliest of (1) the next cancer diagnosis (any new cancer in the OCR, excluding non-melanoma skin); (2) last contact date with the healthcare system (DAD, NACRS, OHIP); (3) death (OCR or RPDB); or (4) December 31, 2021. This was referred to as the recurrence follow-up period (the time the patient is “at-risk” of developing a recurrence).

Primary laterality

The OCR collects information on laterality (left, right, or bilateral) for the primary diagnosis. When assigning a new breast cancer to an existing case, the case resolution process considers histology, laterality, and temporality of the subsequent breast cancer. To allow for some misclassification, we considered patients who had a new contralateral breast cancer diagnosed within 6 months of the primary diagnosis to have had bilateral primary breast cancer (using the primary cancer diagnosis date as the diagnosis date) [18].

Covariates

Patient sociodemographic characteristics included age at primary diagnosis, rurality, and neighborhood-level marginalization derived from the 2016 Canada Census using the Postal Code Conversion File, PCCF + v7B. Clinical characteristics included Charlson comorbidity score (DAD/NACRS), symptomatic presentation, screening through the Ontario Breast Screening Program, and overall stage, hormone receptor (HR) status (positive for either estrogen or progesterone receptors), and HER2 status from the Collaborative Staging database [17]. HER2- patients who received trastuzumab within 1 year of diagnosis were reclassified as HER2 +.

Indicators of recurrence

Flags for local, regional, and distant recurrence/metastasis were created (Fig. 1). Patients were classified into 4 mutually exclusive categories based on the presence of: (1) distant metastasis (DM) flag; (2) regional recurrence (RR) flag; (3) local recurrence (LR) flag; and (4) no recurrence/metastasis.

Ipsilateral ductal carcinoma in situ (DCIS)

We captured the first ipsilateral DCIS from the OCR (ICD-O-3 behavior code 2) after the date of primary breast cancer diagnosis.

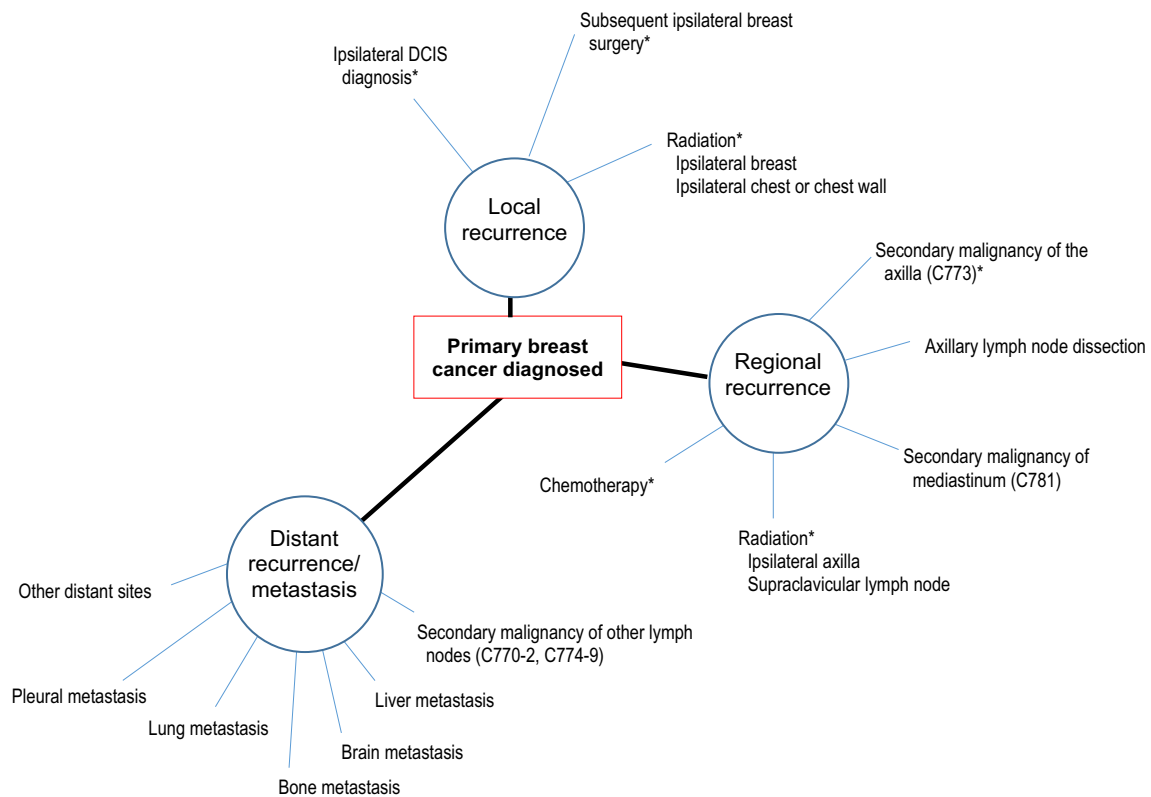


Fig. 1 Indicators for local, regional, and distant metastases from primary breast cancer. Indicators were searched starting from the date of primary surgical resection until end of follow-up. Indicators with an

asterisk (*) were searched starting 1-year after the primary resection. **not considered in this study, but reported by others

Second breast surgery

We used DAD/NACRS to identify a second breast cancer surgery occurring > 12 months after diagnosis to reduce the likelihood of misclassifying revisions or reconstructions after the primary resection (eTable S1). To determine the laterality of the second surgery, we used the intervention location attribute. The accuracy of this attribute code was verified (overall agreement 96.4%) using the laterality of the primary cancer diagnosis from the OCR and the laterality of the primary surgical resection (eTable S2).

Radiation

Radiation (ALR) delivered to the ipsilateral or contralateral breast, ipsilateral or contralateral axilla, supraclavicular region, or chest/chest wall were only considered if administered > 12 months after the primary diagnosis to avoid erroneously capturing primary treatment (eTable S3). No time restriction was placed on radiation administered to other sites. The accuracy of the laterality for primary radiation treatment was 98.9% (eTable S2).

Systemic therapy

We used the ALR and NDFP databases to capture specific chemotherapeutic agents suggestive or indicated for recurrent/metastatic breast cancer (eTable S4). For drugs or policies specific for breast cancer recurrence/metastasis (e.g., pertuzumab), no time restriction was applied (e.g., the earliest treatment was captured any time after diagnosis). For systemic agents that could also be used to treat primary breast cancer or recurrence/metastasis (e.g., doxorubicin), evidence of recurrence was considered if the agent was administered > 12 months after the primary cancer diagnosis.

Intervention codes for possible metastatic disease

We captured interventions (i.e., resection, fixation, destruction, repair) conducted on the four most common sites of metastatic breast cancer (brain, bone, liver, lung) (Supplementary Table S1 for DAD/NACRS procedure codes).

Diagnostic codes for secondary malignancy

We used ICD-10 diagnostic codes to flag secondary malignancies (eTable S5).

Statistical methods

To assess the internal validity of these codes to capture recurrence/metastasis, we used overall survival (OS) as an outcome indicator since OS was derived independent of any of the recurrence/metastasis flags, assuming that patients with a recurrence would have worse OS than those who did not [15]. OS was presented using Kaplan–Meier plots. Factors associated with recurrence/metastasis were assessed using multivariable logistic regression, reporting odds ratios (OR) with 95% confidence intervals (CI).

Software and privacy

All analyses were conducted at Ontario Health using Statistical Analysis Software (SAS Institute Inc., Cary NC). Research ethics was not required as per the Ontario Health privacy office.

Results

After exclusions, 45,857 patients were included (Fig. 2). The recurrence follow-up period was a median 67.1 (IQR 51.8, 84.8) months [mean 65.8 (SD 24.6) months]. Patients were a mean 63.1 (SD 13.6) years of age at diagnosis, 21,187 (46%) of patients were diagnosed with stage I breast cancer, 35,773 (78%) had ductal carcinoma, 31,590 (69%) had HR +/HER2- disease, 39,726 (87%) had no comorbidity, and 23,948 (52%) were symptomatic at the time of diagnosis (Table 1).

Local recurrence

A total 2159 (4.7%) of patients had a LR flag a median 23 (IQR 14–39) months after the primary diagnosis (Table 2). This was driven predominantly by ipsilateral breast surgery (1,448 patients) or radiation to the ipsilateral breast, chest, or chest wall (1,008 patients). Only 27 patients had an ipsilateral DCIS diagnosis during the follow-up period. Using OS as an outcome indicator, patients who had a LR flag had worse OS compared to patients who had no LR (log-rank $p < 0.0001$) (Fig. 3A).

Regional recurrence

A total 1125 (2.5%) of patients had a RR flag a median 22 (13, 40) months after diagnosis, driven by radiation to the

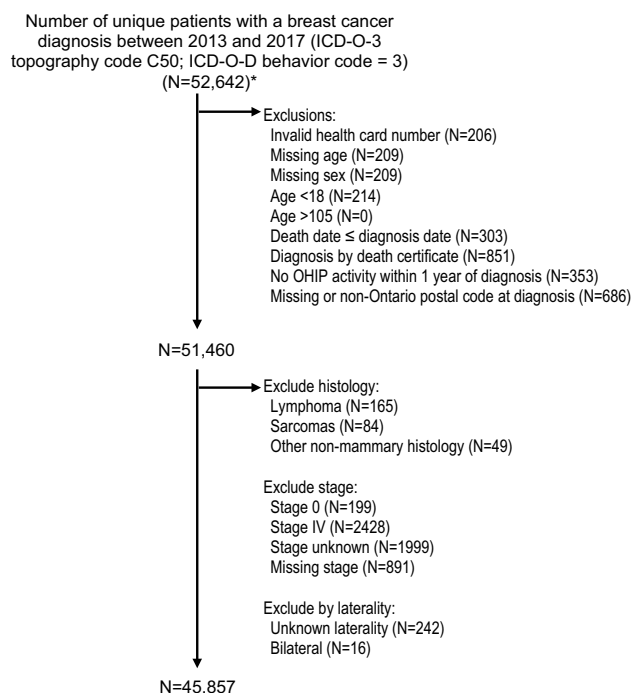


Fig. 2 Kaplan–Meier plots for overall survival among patients who were classified as having a local recurrence (A), regional recurrence (B), or distant metastasis using the strict definition of only secondary diagnostic codes or radiation intent (C). The overall summary of the recurrence/metastatic event is presented in (D)

ipsilateral axilla, supraclavicular node or internal mammary chain ($n = 605$) or a diagnosis of a secondary malignancy in the axilla as the most responsible diagnosis ($n = 395$). Using OS as an outcome indicator, patients who had a RR flag had worse OS compared to patients who did not (log-rank $p < 0.0001$) (Fig. 3B).

Distant metastasis

A total 5431 (11.8%) of patients had a DM a median 23 (9, 42) months after diagnosis. Patients most frequently had an indicator for metastasis to the bone ($n = 2614$; 5.7%), liver ($n = 1809$; 3.9%), lung ($n = 1698$; 3.7%), brain ($n = 1129$; 2.5%), some other lymph nodes outside the ipsilateral axillary, internal mammary or supraclavicular regions ($n = 822$; 1.8%), and pleura ($n = 726$; 1.6%). A total 2128 (4.6%) of patients had a DM to some other site. Using OS as an outcome indicator, patients who developed a DM had worse OS compared to patients who did not (log-rank $p < 0.0001$) (Fig. 3C; eFigure S1 by site of metastasis).

Systemic therapy

A total 2653 (5.8%) of patients received chemotherapy not indicated for primary breast cancer and 1626 (3.5%)

Table 1 Patient Characteristics

	Full cohort (<i>n</i> = 45,857)	<i>N</i> (%) with distant metastasis (<i>n</i> = 5431)	Relative odds of developing a distant metastasis			
			Model 1 ^a		Model 2 ^a	
			OR (95% CI)	<i>p</i> value	OR (95% CI)	<i>p</i> value
Demographic characteristics						
Age, years ^b						
18–49	8102 (18%)	1102 (20%)	1.0 (ref)	0.02	1.0 (ref)	0.02
50–74	28,859 (63%)	2919 (54%)	1.09 (0.96–1.23)		1.10 (0.97–1.25)	
75–105	8896 (19%)	1410 (26%)	1.32 (1.06–1.64)		1.31 (1.06–1.63)	
Continuous	63.1 (SD 13.6)	63.7 (SD 15.2)	1.01 (0.96–1.06)	0.63	1.00 (0.95–1.05)	0.96
Sex						
Female	45,516 (99.3%)	5354 (98.58%)	1.0 (ref)	0.002	1.0 (ref)	0.002
Male	341 (0.7%)	77 (1.42%)	1.59 (1.19–2.12)		1.60 (1.20–2.13)	
Clinical characteristics						
Route to diagnosis ^c						
OBSP screened	13,110 (29%)	738 (14%)	1.0 (ref)	< .0001	1.0 (ref)	< .0001
GP screened	8799 (19%)	883 (16%)	1.46 (1.31–1.63)		1.46 (1.31–1.64)	
Symptomatic	23,948 (52%)	3810 (70%)	1.87 (1.70–2.06)		1.87 (1.70–2.06)	
Overall stage						
I	21,187 (46%)	927 (17%)	1.0 (ref)	< .0001	1.0 (ref)	< .0001
II	18,531 (40%)	2483 (46%)	2.85 (2.62–3.11)		2.89 (2.65–3.15)	
III	6139 (13%)	2021 (37%)	8.96 (8.16–9.84)		9.17 (8.34–10.1)	
Charlson comorbidity						
Missing	17,439 (38%)	1826 (34%)	0.85 (0.79–0.91)		0.85 (0.79–0.91)	
0	22,287 (49%)	2671 (49%)	1.0 (ref)	< .0001	1.0 (ref)	< .0001
1	4030 (9%)	572 (11%)	1.12 (1.00–1.25)		1.12 (1.00–1.25)	
2	1200 (3%)	178 (3%)	1.01 (0.84–1.22)		1.00 (0.83–1.21)	
3+	901 (2%)	184 (3%)	1.56 (1.29–1.89)		1.51 (1.25–1.83)	
Biomarker status						
HR +/HER2+	4667 (10%)	644 (12%)	1.0 (ref)	< .0001	1.0 (ref)	< .0001
HR +/HER2-	31,590 (69%)	3106 (57%)	0.90 (0.81–0.99)		0.56 (0.48–0.65)	
HR-/HER2+	2055 (4%)	363 (7%)	1.25 (1.07–1.45)		1.28 (1.10–1.49)	
HR-/HER2-	4332 (9%)	916 (17%)	1.78 (1.58–2.01)		1.10 (0.92–1.31)	
Missing	3213 (7%)	402 (7%)	1.14 (0.99–1.32) ^e		0.79 (0.67–0.95) ^e	
Primary trastuzumab						
No	40,229 (88%)	4651 (86%)	-	-	1.0 (ref)	< .0001
Yes	5628 (12%)	780 (14%)	-		0.53 (0.45–0.63)	
Histology						
Ductal	35,773 (78%)	4131 (76%)	1.0 (ref)	< .0001	1.0 (ref)	< .0001
Lobular	4059 (9%)	251 (5%)	1.00 (0.89–1.11)		0.99 (0.89–1.11)	
Ductal and lobular	2003 (4%)	527 (10%)	0.99 (0.85–1.15)		0.98 (0.84–1.14)	
Mucinous	889 (2%)	37 (1%)	0.41 (0.29–0.58)		0.41 (0.29–0.59)	
Adenocarcinoma	887 (2%)	101 (2%)	0.86 (0.68–1.10)		0.86 (0.68–1.10)	
Other histology	2246 (5%)	384 (7%)	1.14 (1.00–1.30)		1.14 (1.00–1.30)	
Laterality						
Left	23,485 (51%)	2760 (51%)	1.0 (ref)	0.43	1.0 (ref)	0.44
Right	22,372 (49%)	2671 (49%)	1.03 (0.96–1.09)		1.03 (0.96–1.09)	
Sociodemographic characteristics						
Rurality ^d						
Urban	40,373 (89%)	4769 (88%)	1.0 (ref)	0.43	1.0 (ref)	0.47
Rural	5040 (11%)	621 (12%)	0.96 (0.85–1.07)		0.96 (0.86–1.08)	

Table 1 (continued)

	Full cohort (<i>n</i> = 45,857)	<i>N</i> (%) with distant metastasis (<i>n</i> = 5431)	Relative odds of developing a distant metastasis			
			Model 1 ^a		Model 2 ^a	
			OR (95% CI)	<i>p</i> value	OR (95% CI)	<i>p</i> value
Residential instability						
1 (least unstable)	8704 (19%)	943 (18%)	1.0 (ref)	0.19	1.0 (ref)	0.18
2	8523 (19%)	988 (18%)	1.04 (0.94–1.16)		1.04 (0.94–1.16)	
3	8479 (19%)	943 (18%)	0.94 (0.84–1.05)		0.94 (0.84–1.04)	
4	8734 (19%)	1141 (21%)	1.05 (0.94–1.17)		1.05 (0.94–1.17)	
5 (most unstable)	10,777 (24%)	1350 (25%)	0.99 (0.88–1.10)		0.98 (0.88–1.09)	
Material deprivation						
1 (least deprived)	10,365 (23%)	1076 (20%)	1.0 (ref)	0.008	1.0 (ref)	0.01
2	9508 (21%)	1090 (20%)	1.09 (0.99–1.20)		1.09 (0.99–1.20)	
3	9002 (20%)	1042 (19%)	1.08 (0.97–1.19)		1.07 (0.97–1.19)	
4	8558 (19%)	1113 (21%)	1.20 (1.08–1.33)		1.19 (1.08–1.33)	
5 (most deprived)	7784 (17%)	1044 (19%)	1.19 (1.06–1.33)		1.18 (1.06–1.33)	
Dependency						
1 (least dependent)	8623 (19%)	1014 (19%)	1.0 (ref)	0.27	1.0 (ref)	0.24
2	8281 (18%)	981 (18%)	0.98 (0.88–1.09)		0.98 (0.88–1.09)	
3	8188 (18%)	961 (18%)	0.97 (0.87–1.08)		0.97 (0.87–1.08)	
4	8520 (19%)	968 (18%)	0.89 (0.80–0.99)		0.89 (0.79–0.99)	
5 (most dependent)	11,605 (26%)	1441 (27%)	0.94 (0.84–1.05)		0.94 (0.84–1.05)	
Ethnic concentration						
1 (least concentrated)	8623 (19%)	1118 (21%)	1.0 (ref)	0.003	1.0 (ref)	0.002
2	8566 (19%)	1042 (19%)	0.93 (0.84–1.04)		0.94 (0.84–1.04)	
3	8188 (18%)	1030 (19%)	0.94 (0.84–1.05)		0.94 (0.84–1.05)	
4	9203 (20%)	1009 (19%)	0.82 (0.73–0.92)		0.82 (0.73–0.92)	
5 (most concentrated)	9988 (22%)	1166 (22%)	0.82 (0.73–0.93)		0.82 (0.73–0.93)	

OR odds ratio, CI confidence interval, HR hormone receptor (either estrogen or progesterone), HER2 human epidermal growth factor receptor 2

^aadjusted for all variables shown in the table. Receipt of trastuzumab within 1 year of diagnosis was considered a mediator of the effect of biomarker status on the likelihood of recurrence. Thus, we provided two models, one omitting the mediator in the logistic regression model (Model 1) and another including it (Model 2)

^bage at diagnosis of the primary breast cancer. Odds ratio for continuous age reflects a 10-year increase in age

^cthe patients' primary breast cancer was detected by the Ontario Breast Screening Program (OBSP), was screened by their general practitioner (GP), or was otherwise symptomatic

^dsource (or adapted from): Statistics Canada Postal Code Conversion File and Postal Code Conversion File Plus (version 7B, received May 2019) which is based on data licensed from Canada Post Corporation. The patients' postal code at diagnosis was used

^eanalysis was repeated with the missing category assigned as an active value. The ORs for the other levels were only subtly changed

received chemotherapy for a recurrent/metastatic indication any time during the recurrence follow-up period. A total 1473 (3.2%) of patients received chemotherapy indicated for breast cancer at least 1 year after diagnosis. Using either of these classifications, a total 3253 (7.1%) of patients received chemotherapy for recurrence/metastasis a median 26 (12, 45) months after diagnosis.

Summary of recurrences

Because systemic therapy could be provided for either RR or DM, patients receiving chemotherapy without evidence

of DM were considered to have had a RR. This decision was made a priori, but was also supported when OS was used as an outcome indicator (eFigure S2). A total 5431 (11.8%) patients developed a DM a median 23 (9, 42) months after diagnosis; 1086 (2.4%) of patients had a RR a median 13 (10, 34) months after diagnosis; and 1069 (2.3%) of patients had an isolated LR a median 26 (16, 42) months after diagnosis (eFigure S3 for time-until-event). The remaining 38,271 (83.5%) patients had no evidence of progressive disease during the recurrence follow-up period. Patients with DM had a median survival of 53.3 (95% CI 51.9–54.7) months from the time of primary diagnosis and a median

Table 2 Indicators of breast cancer recurrence

Administrative data	<i>n</i> with indicator (<i>N</i> =45,857)	Median (IQR), months since diagnosis ^a
	<i>N</i>	Median (IQR)
Local sites		
Ductal carcinoma in situ (ipsilateral only) ^b	27 (<1%)	73 (64, 80)
Second breast surgery (ipsilateral only) ^b	1448 (3.2%)	26 (17, 41)
Breast/chest radiation (ipsilateral only) ^b	1008 (2.2%)	17 (12, 33)
Local recurrence (ipsilateral DCIS, ipsilateral breast surgery, ipsilateral breast radiation, or ipsilateral chest wall radiation)	2159 (4.7%)	23 (14, 39)
Regional sites		
Lymph node		
Axilla (C773) ^{b,c,d}	394 (0.9%)	28 (19, 45)
Axillary lymph node dissection (ipsilateral only) ^b	169 (0.4%)	26 (16, 42)
Mediastinum (C781) ^b	84 (0.2%)	46 (29, 60)
Radiation, axilla or supraclavicular lymph node (ipsilateral only) ^b	605 (1.3%)	15 (12, 32)
Regional recurrence (C773, ipsilateral ALND, C781, radiation to the ipsilateral axilla, ipsilateral supraclavicular lymph node, or internal mammary chain)	1125 (2.5%)	22 (13, 40)
Distant sites		
Liver metastasis		
Secondary diagnostic code (C787)	1809 (3.9%)	33 (18, 51)
Neoplasm of uncertain behavior of liver, gallbladder and bile ducts (D376) ^b or Malignant neoplasm of liver and intrahepatic bile ducts (C22) ^b	104 (0.2%)	30 (13, 53)
Liver resection	20 (<1%)	12 (5, 48)
Liver destruction	8 (<1%)	13 (6, 20)
Liver radiation	57 (0.1%)	50 (26, 66)
Any liver (C787, liver resection, destruction, radiation)	1833 (4.0%)	33 (18, 51)
Any liver (C787, liver resection, destruction, radiation, D376, C22)	1870 (4.1%)	33 (17, 51)
Brain metastasis		
Brain (C793)	1129 (2.5%)	30 (16, 47)
Brain radiation	861 (1.9%)	30 (17, 48)
Any of the above (C793, brain radiation)	1257 (2.7%)	29 (16, 48)
Bone metastasis		
Bone (C795)	2614 (5.7%)	32 (16, 50)
Neoplasm of uncertain behavior of bone and articular cartilage (D480) ^b or Malignant neoplasms of bone and articular cartilage (C40, C41) ^b	81 (0.2%)	31 (20, 53)
Bone radiation	1445 (3.2%)	33 (19, 50)
Bone resection	109 (0.2%)	37 (18, 54)
Bone fixation	808 (1.8%)	39 (20, 58)
Bone repair	698 (1.5%)	39 (24, 57)
Bone metastasis (C793, surgery, radiation, fixation, repair)	4170 (9.1%)	33 (16, 51)
Bone metastasis (C793, surgery, radiation, fixation, repair, D480, C40-C41)	4187 (9.1%)	33 (16, 51)
Lung metastases		
Lung (C780)	1698 (3.7%)	30 (16, 49)
Pleura (C782, C783)	726 (1.6%)	33 (16, 53)
Malignant neoplasm of bronchus and lung (C34)	714 (1.6%)	13 (3, 40)
Lung/bronchus/pleural resection	106 (0.2%)	20 (5, 35)
Lung/bronchus destruction	6 (<1%)	24 (18, 65)
Lung radiation	147 (0.3%)	22 (6, 42)
Any lung of the above (C780, C782, C783)	2024 (4.4%)	30 (14, 48)
Any lung of the above (C780, C782/3, lung/pleural resection, destruction, radiation)	2142 (4.7%)	30 (14, 48)
Other sites		

Table 2 (continued)

Administrative data	<i>n</i> with indicator (<i>N</i> =45,857)	Median (IQR), months since diagnosis ^a
	<i>N</i>	Median (IQR)
Other lymph node (C770-2, C774-9) ^b	822 (1.8%)	35 (20, 52)
Other distant sites (digestive organs C784-C786, C788; urogenital organs C790, C791; skin C792; other nervous system (C794); ovary (C796), breast, genital, and other sites C798; and other unspecified sites C799)	2128 (4.6%)	25 (8, 46)
Radiation by intent		
Radiation for metastasis (any site) ^c	366 (0.8%)	66 (51, 83)
Radiation with palliative intent (any site)	2319 (5.1%)	27 (14, 43)
Any distant metastasis		
Definition 1 (strict)—only secondary diagnostic codes	5431 (11.8%)	23 (9, 42)
Definition 2 (moderate)—also includes treatment for liver, brain, bone, and pulmonary	6756 (14.7%)	26 (10, 46)
Definition 3 (loose)—also includes primary diagnostic codes	6863 (15.0%)	25 (10, 45)
Systemic chemotherapy		
Category I: agent delivered any time during the recurrence follow-up period	2653 (5.8%)	29 (12, 48)
Category II: agents delivered during the recurrence follow-up period but administered at least 1 year after diagnosis	1473 (3.2%)	34 (21, 49)
Category III: Protocols delivered any time during the recurrence follow-up period	1626 (3.5%)	27 (13, 44)
Any chemotherapy above (category 1, 2, or 3)	3253 (7.1%)	26 (12, 45)
Final definitions		
Distant metastasis (strict definition)	5431 (11.8%)	23 (9, 42)
Regional recurrence only (regional recurrence or any chemotherapy)	1086 (2.4%)	13 (10, 34)
Local recurrence only	1069 (2.3%)	26 (16, 42)
No recurrence or metastasis	38,271 (83.5%)	—

^aMedian (IQR—25th, 75th percentile) from the time of the primary breast cancer diagnosis

^brestricted to 12 months after the diagnosis date

^ccannot determine laterality for a diagnostic code

^drestricted to DAD where the diagnosis type is not 1 (not a pre-admission diagnosis)

^emetastasis intent was a new intent field in ALR

OS of 15.4 months (95% CI 14.4–16.4 months) from the time the recurrence/metastasis was captured (Fig. 3D–E). In contrast, the median survival for all other patients was not reached.

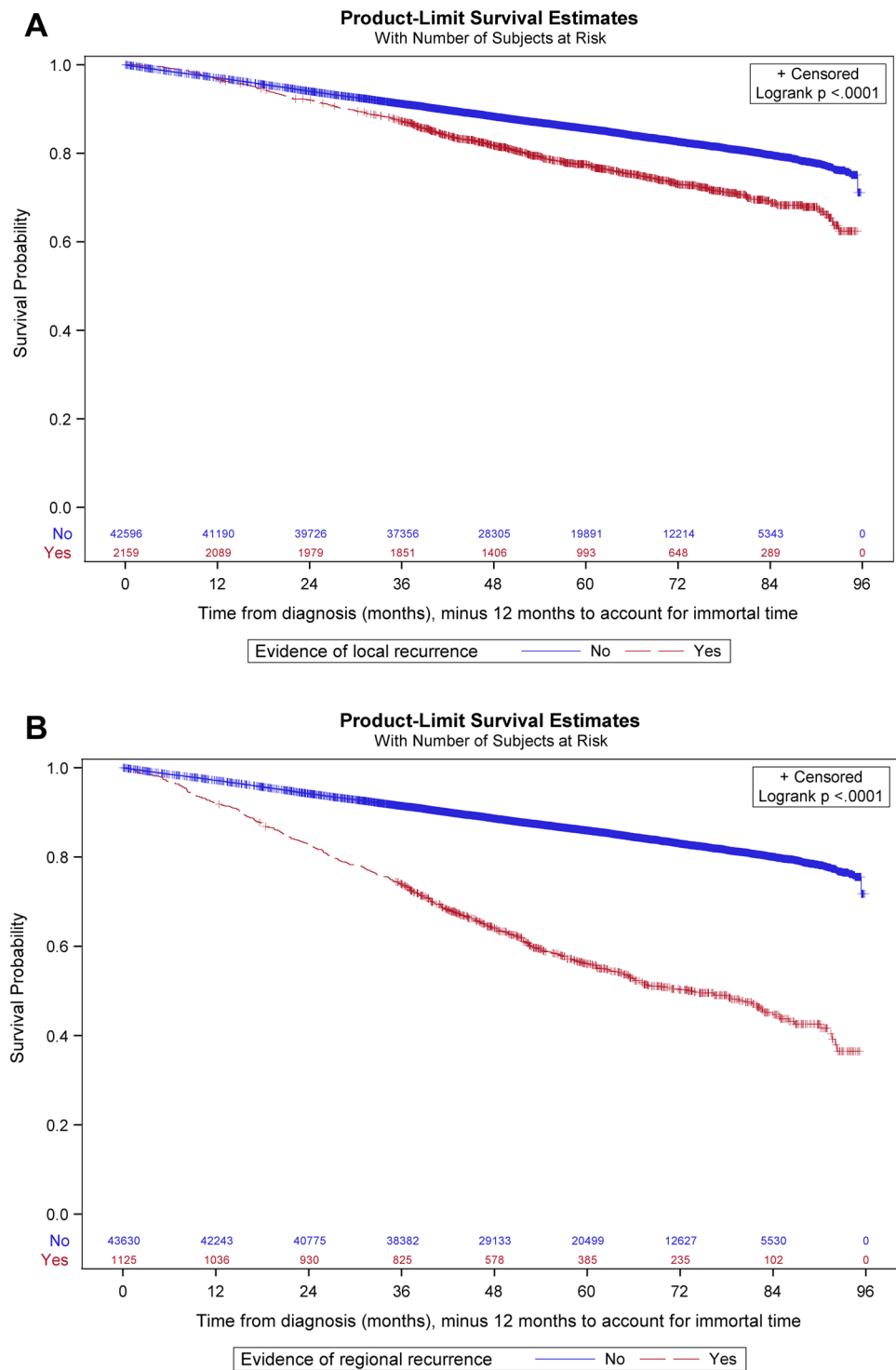
In the subset of patients diagnosed with primary breast cancer in 2013 (for follow-up time for up to nine years), a total 1184/8713 (13.6%) of patients were found to have developed a DM during the recurrence follow-up period. There was a non-linear trend in the occurrence of the events with a decreasing risk of capturing a DM as the time since diagnosis increased: 29.6% of DM were identified within the first year of diagnosis, 18.0% within the second year, and 171 (14.4%) within the third year (Fig. 4).

Factors associated with recurrence

We explored factors associated with disease progression (LR, RR, or DM) (eTable S6 for descriptive statistics). After adjustment, patients were more likely to develop a

DM (versus regional, local, or no recurrence) if they were older [OR 1.32 (1.06–1.64) for age 75 + versus < 50 years], male [OR 1.59 (1.19–2.12)], had stage II [OR 2.85 (2.62–3.11)] or stage III [OR 8.96 (8.16–9.84)] versus stage I, were symptomatic at primary breast cancer presentation [OR 1.87 (1.70–2.06) versus screen-detected], or had 3 + comorbidities [OR 1.56 (1.29–1.89) versus none] (Table 1). Patients were least likely to develop DM if their primary tumor was of mucinous histology [OR 0.41 (0.29–0.58)]. There were no obvious trends related to neighborhood-level socioeconomic characteristics. In the statistical model where we do not adjust for receipt of trastuzumab, HER2 + patients were more likely to develop a recurrence among HR + patients [OR 1.12 (1.01–1.23)], but less likely among HR- patients [OR 0.70 (0.61–0.81)]. After additionally adjusting for receipt of trastuzumab for the primary breast cancer (a mediator for the effect of HER2 status on the risk of recurrence), the effect of

Fig. 3 Product-limit survival estimates

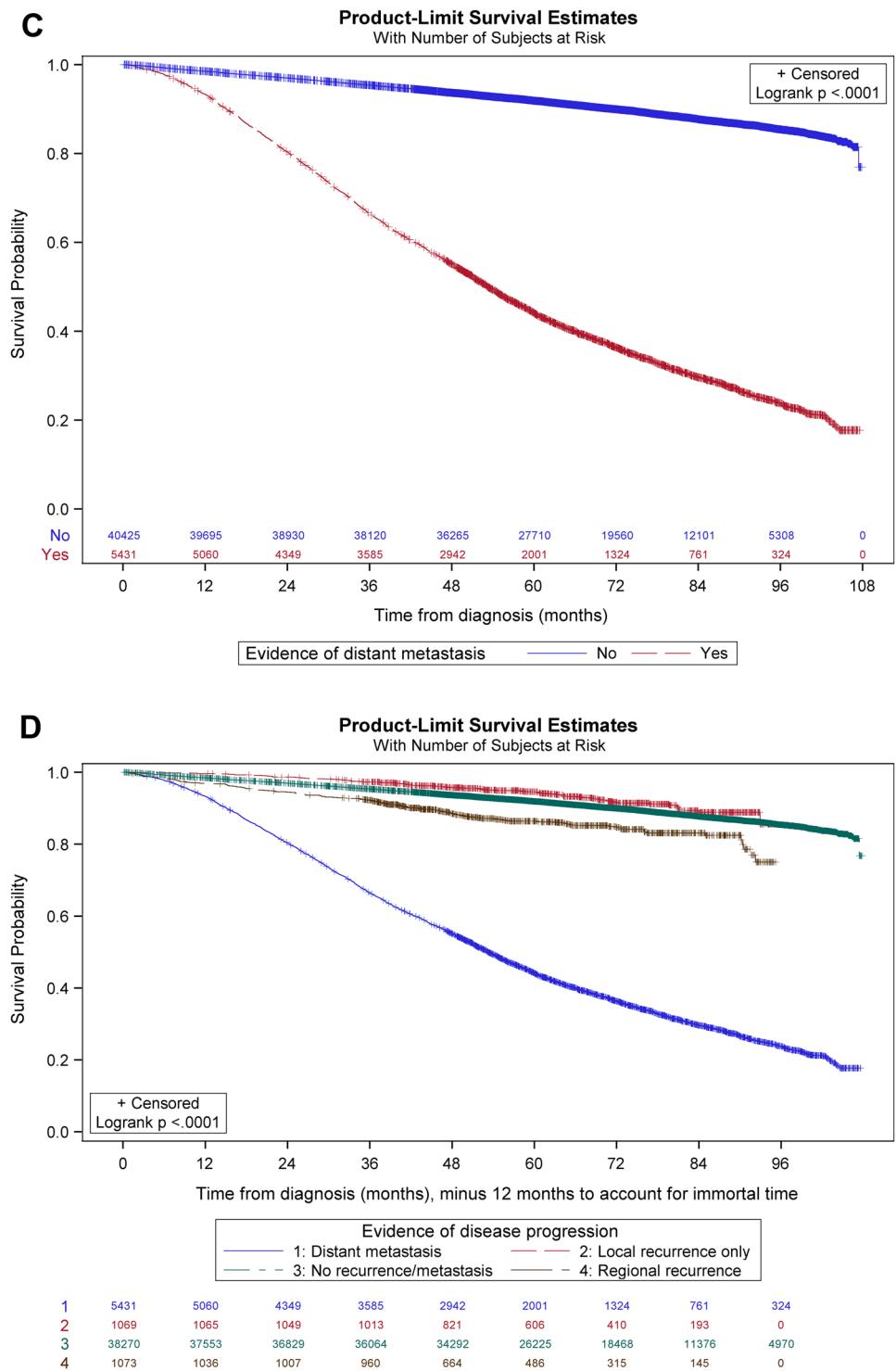


HER2+ on the risk of recurrence increased such that HER2+ patients were more likely to develop a recurrence if they were HR+ [OR 1.79 (1.53–2.10), $p < 0.0001$], but not HR- [OR 1.16 (0.95–1.42), $p = 0.13$]. Receipt of primary trastuzumab halved the risk of recurrence [OR 0.53 (0.45–0.63), $p < 0.0001$].

Discussion

In the present study, we estimated that 2.3%, 2.4%, and 11.8% of stage I-III breast cancer patients had an isolated LR, a (loco)-regional recurrence only, and DM over a median follow-up of 67 months, respectively. OS after isolated LR was similar to the recurrence-free population and moderately

Fig. 3 (continued)

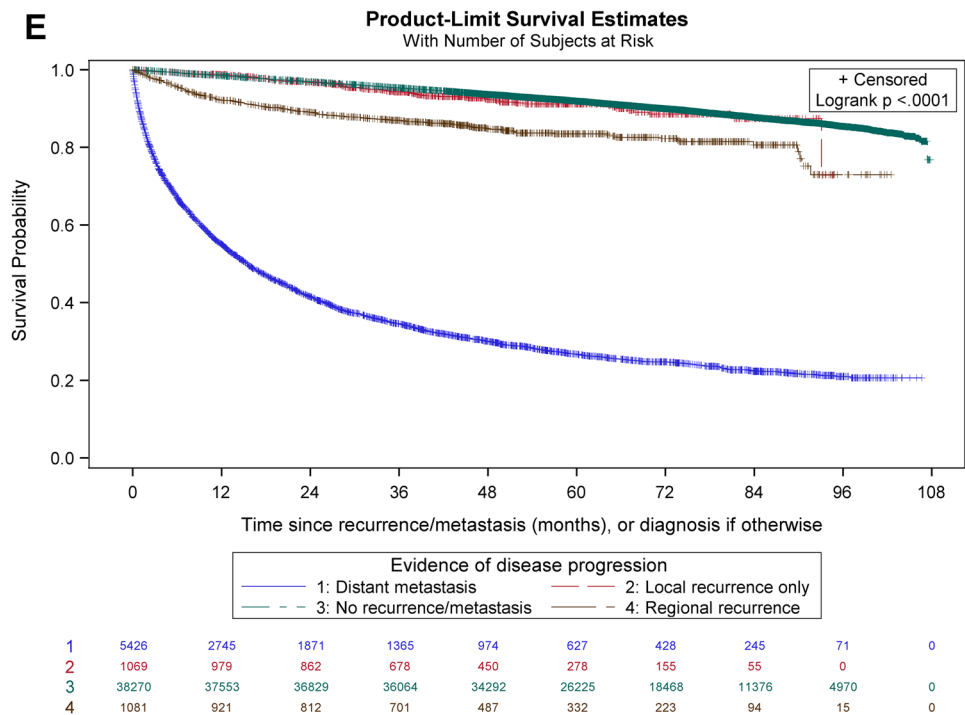


worse for patients with (loco)-regional recurrence. However, OS was substantively reduced after a diagnosis of metastatic disease, with a median survival of 15.4 months. Patients presenting with symptomatic breast cancer and more advanced stage were more likely to develop a DM.

Recurrence rates reported in the literature vary due to differences in data availability, follow-up time, the definition

of recurrence, and whether DM or RR were included [14]. However, our estimates are aligned with several international investigations. One European study demonstrated a 5-year cumulative incidence of 1.1%, 1.2%, and 7.6% for local, regional, and DM, respectively, after a median follow-up of 72 months [19]. Similar results were reported by one study from China (3.9% for locoregional and 8.8% for DM

Fig. 3 (continued)



after a median follow-up of 75 months) [20]. One study in Brazil demonstrated a LR rate of 6.8% a median 28 months after primary breast-conserving surgery and was consistent with estimates from a recent systematic review [21, 22]. One registry-based study from the Netherlands reported a probability of recurrence of 20% after a median follow-up of 10 years, demonstrating a peak risk during the second and eighth years after diagnosis [23]. Other studies using more statistical approaches yielded estimates ranging from 7.6% over a period of 52 months to 19% over 20 years [14, 15].

Several studies demonstrated that the use of administrative data to identify a recurrence is valid and appropriate. Compared with manual chart review, indicators of a recurrence limited only to subsequent treatment (surgery, chemotherapy, or radiation) yielded sensitivity and specificity ranging from 78 to 81% [24, 25] However, this approach will not capture untreated DM [26] One Danish group supplemented treatment with ICD-10 diagnostic codes for secondary malignancies, demonstrating high agreement with medical chart review in a small sample of patients ($n = 471$, sensitivity = 97%, specificity = 97%) [27]. One Ontario study achieved a sensitivity of 85.3% and specificity of 93.8% using treatment indicators and cause of death from patients treated at three academic tertiary cancer centers [28]. Investigators in Alberta, Canada used several classification and regression tree (CART) models to optimize different measures of discrimination. Their overall predictive accuracy was good using a model that incorporated subsequent treatment, referral to oncologists, visits to cancer centers, visits to oncologists or general surgeons, primary cancer diagnostic codes, and cause of death due to cancer [13]. A similar study from the United States yielded lower accuracy using CART, but employed fewer indicators of recurrence, limited to registry data, treatment, secondary non-breast malignancy diagnostic codes, and a diagnosis of breast carcinoma in situ [12, 29]. To define DM, we opted to exclude any codes for treatments and instead relied on diagnostic codes for secondary malignancies. This was done because the reason for treatments (e.g., surgery) could be multifactorial and not specific to breast cancer metastasis. Prior work to identify

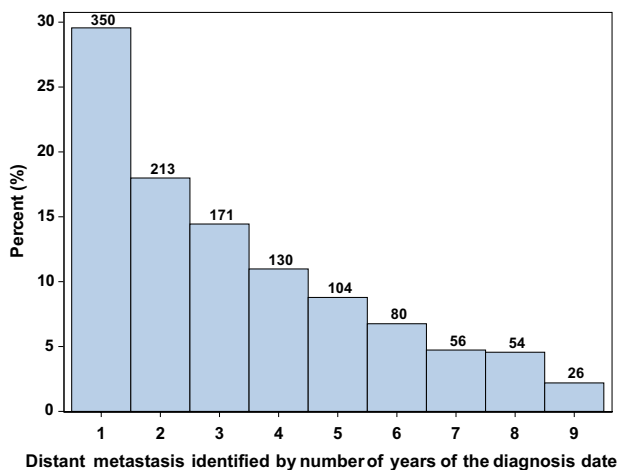


Fig. 4 Distant metastasis identified by number of years of the diagnosis date

metastatic brain cancers indeed found that surgical resections of brain tissue were frequently for new primary brain cancers instead of metastases [30].

Each year in Ontario, we anticipate approximately 12,000 new breast cancer diagnoses [31]. We therefore anticipate that 1632 (13.6% of 12,000) patients each year will present with a DM from a breast cancer that was diagnosed within the previous nine years. This is in addition to the approximately 500 patients per year diagnosed with stage IV breast cancer at initial presentation. This yields at least 2,132/12,000 (17.8%) patients per year diagnosed with metastatic breast cancer (de novo or after a previous diagnosis). This annual incidence surpasses that of many other primary cancer types, and with clinical trials geared toward this population demonstrating positive survival benefits of treatment, more accurate prospective data collection on these cases is needed [32]. Due to the lack of population-level data on this population, evidence is lacking on the effectiveness of treatments and the geographic variability in treatment options provided. Our findings indicate that this is not an inconsequential number of people, and efforts should be taken to obtain more detailed information on the type of recurrence/metastasis (e.g., location of lesions, number of lesions), treatment options, and the date of diagnosis.

Limitations

Although this is the first population-based study estimating breast cancer recurrence in Ontario, there are some limitations. First, although recurrences can occur at any time, for select indicators of loco-regional recurrence, we searched starting 12 months after the diagnosis date. This was important to reduce the number of false positives (e.g., misclassifying ongoing treatment or re-treatment for residual disease for the primary cancer as a recurrent event), but occurred at the expense of false negatives (e.g., missing an early recurrent event) [12, 13]. We believe we are justified in this trade-off because LR within the first year of diagnosis are less likely and discriminating recurrences from residual disease is not always clear. Second, although our results have face validity and reasonable convergent validity (e.g., using OS as an outcome indicator; reporting similar incidence estimates as the literature; and reporting similar factors associated with recurrence as the literature and in the expected directions), we were unable to systematically validate our findings using a gold standard (e.g., through an integrated electronic medical record system). Despite this, our definition uses similar indicators used by previous studies that demonstrated good agreement with chart review [24, 25, 27]. We believe those studies are generalizable to our patients since many of these studies were derived from jurisdictions with similar healthcare systems. Third, we may have overestimated the effect of recurrence on mortality since patients

treated with oral hormonal therapy alone (e.g., those with more indolent cancers who respond well to hormonal therapy) may have been excluded. Unlike intravenous therapies, oral hormonal therapies like tamoxifen and aromatase inhibitors are incompletely captured in our databases. Patients can remain on hormonal therapy and switch between agents for many years after their primary cancer diagnosis. However, our databases do not give us the ability to distinguish prolonged therapy from the primary cancer versus new therapy for a recurrence. Thus, we opted to omit oral treatments like hormones from the study.

Strengths

Despite the above limitations, one strength of this study is that it was population-based and therefore less likely to miss recurrences because of patient movement within the province. This is one caveat of medical chart review since access to medical records may be restricted to specific institutions, yet patients may be followed elsewhere. At a population level, we provide estimates that are important for system level monitoring and planning for several reasons. First, survival for patients with breast cancer is high, with 5-year OS estimated at 85% for patients with stage I-III disease [33]. Thus, using OS as an outcome in comparative studies often demands large sample sizes with lengthy follow-up for statistical inferences. Clinical trialists therefore often use surrogate endpoints such as recurrence-free survival [34]. This may circumvent some of the problems of using OS as an outcome when mortality is uncommon, but recurrence is not always a surrogate for OS and must be interpreted cautiously [35, 36]. Isolated LR may be a poor predictor of OS since it can be clinically managed, but may instead be considered an indicator affecting patient quality-of-life. DM may instead be considered a better outcome indicator because it was highly prognostic and 82% of all DM were captured within 5 years of diagnosis. Second, since breast cancer is the most common cancer among women, uncommon or even rare events can have a large burden on population-level health system resources. Third, the incidence of a recurrence/metastasis (15.5% of breast cancer survivors) may warrant breast cancer survivors' consideration for organized screening programs [5, 37, 38].

In future work, comprehensive electronic medical record systems would enable machine learning techniques to use information from imaging reports, pathology reports, and clinic progress notes to classify LR/RR/DM. One recent study using natural language processing of text from progress notes and pathology reports yielded an AUC 0.93 and 0.87 in the training and validation datasets, respectively [39]. Pathology reports alone would only be available to a non-representative sample of patients; those who had a biopsy or surgical resection to diagnose recurrence. Tissue

specimens may be more common for some metastatic sites (breast, liver) but uncommon in others (brain, bone). Thus, imaging reports and clinic notes would provide a more representative corpus of documents for text-mining.

Conclusion

In conclusion, in our study using administrative data, 15.5% of breast cancer patients developed a recurrence/metastasis within 9 years after a diagnosis of breast cancer. Symptomatic patients with stage III primary breast cancer are at elevated risk of developing a DM.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s10549-022-06812-z>.

Acknowledgements Parts of this material are based on data and information compiled and provided by the Canadian Institute of Health Information (CIHI). However, the analyses, conclusions, opinions and statements expressed herein are those of the author, and not necessarily those of CIHI. We acknowledge support of the Ministry of Health and Long-Term Care in this report. All views expressed are those of the authors of this report and do not necessarily reflect those of Ontario or the Ministry.

Author contributions All authors contributed substantially to this work, including conception and study design (SH, KF, AE, CH, SK), analysis (SH, EH, AB), and manuscript preparation (all).

Funding No specific funding sources were used.

Data availability Ontario Health is prohibited from making the data used in this research publicly accessible if it includes potentially identifiable personal health information and/or personal information as defined in Ontario law, specifically the Personal Health Information Protection Act (PHIPA) and the Freedom of Information and Protection of Privacy Act (FIPPA). Upon request, data de-identified to a level suitable for public release may be provided.

Declarations

Conflict of interests All co-authors have no conflicts of interest to disclose.

Ethics approval Research ethics was not required as per the Ontario Health privacy office.

Consent to participate Patient consent was not required for this study.

References

- Warren J, Yabroff K (2015) Challenges and opportunities in measuring cancer recurrence in the United States: PubMed: NCBI. *J Natl Cancer Inst* 107:8
- Vuksanovic D, Sanmugarajah J, Lunn D, Sawhney R, Eu K, Liang R (2021) Unmet needs in breast cancer survivors are common, and multidisciplinary care is underutilised: the Survivorship Needs Assessment Project. *Breast Cancer* 28(2):289–297
- Koch L, Bertram H, Eberle A et al (2014) Fear of recurrence in long-term breast cancer survivors: still an issue. Results on prevalence, determinants, and the association with quality of life and depression from the cancer survivorship: a multi-regional population-based study. *Psychooncology* 23(5):547–554
- Mariotto AB, Etzioni R, Hurlbert M, Penberthy L, Mayer M (2017) Estimation of the number of women living with metastatic breast cancer in the United States. *Cancer Epidemiol Biomarkers Prev* 26(6):809–815
- Muradali D, Kennedy EB, Eisen A, Holloway CMB, Smith CR, Chiarelli AM (2017) Breast screening for survivors of breast cancer: a systematic review. *Prev Med (Baltim)* 103:70–75
- Darby S, McGale P, Correa C et al (2011) 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(9804):1707–1716
- Dong Y, Zhang WW, Wang J, Sun JY, He ZY, Wu SG (2019) The 21-gene recurrence score and effects of adjuvant radiotherapy after breast conserving surgery in early-stage breast cancer. *Futur Oncol* 15(14):1629–1639
- Bradley R, Burrett J, Clarke M et al (2015) Aromatase inhibitors versus tamoxifen in early breast cancer: patient-level meta-analysis of the randomised trials. *Lancet* 386(10001):1341–1352
- Dent R, Trudeau M, Pritchard KI et al (2007) Triple-negative breast cancer: clinical features and patterns of recurrence. *Clin Cancer Res* 13(15):4429–4434
- Siotos C, Naska A, Bello RJ et al (2019) Survival and disease recurrence rates among breast cancer patients following mastectomy with or without breast reconstruction. *Plast Reconstr Surg* 144(2):169e–177e
- Voogd AC, Nielsen M, Peterse JL et al (2001) Differences in risk factors for local and distant recurrence after breast-conserving therapy or mastectomy for stage I and II breast cancer: pooled results of two large European randomized trials. *J Clin Oncol* 19(6):1688–1697
- Chubak J, Yu O, Pocobelli G et al (2012) Administrative data algorithms to identify second breast cancer events following early-stage invasive breast cancer. *J Natl Cancer Inst* 104(12):931–940
- Xu Y, Kong S, Cheung WY et al (2019) Development and validation of case-finding algorithms for recurrence of breast cancer using routinely collected administrative data. *BMC Cancer* 19:1
- Ritzwoller DP, Hassett MJ, Uno H et al (2018) Development, validation, and dissemination of a breast cancer recurrence detection and timing informatics algorithm. *J Natl Cancer Inst* 110(3):273–281
- Mariotto AB, Zou Z, Zhang F, Howlader N, Kurian AW, Etzioni R (2018) Can we use survival data from cancer registries to learn about disease recurrence? The case of breast cancer. *Cancer Epidemiol Biomarkers Prev* 27(11):1332–1341
- Haque R, Shi J, Schottinger JE et al (2015) A hybrid approach to identify subsequent breast cancer using pathology and automated health information data. *Med Care* 53(4):380–385
- Habbous S, Homenauth E, Barisic A et al (2022) Differences in breast cancer diagnosis by patient presentation in Ontario: a retrospective cohort study. *C Open* 10(2):E313–E330
- Padmanabhan N, Subramanyan A, Radhakrishna S (2015) Synchronous bilateral breast cancers. *J Clin Diagnostic Res* 9(9):10–13
- Botteri E, Bagnardi V, Rotmensz N et al (2009) Analysis of local regional recurrences in breast cancer after conservative surgery. *Ann Oncol* 21(4):723–728
- Luo C, Zhong X, Deng L, Xie Y, Hu K, Zheng H (2019) Nomogram predicting locoregional recurrence to assist decision-making of postmastectomy radiation therapy in patients with T1–2N1 breast cancer. *Int J Radiat Oncol Biol Phys* 103(4):905–912

21. Tovar JR, Zandonade E, Amorim MHC (2014) Factors associated with the incidence of local recurrences of breast cancer in women who underwent conservative surgery. *Int J Breast Cancer* 2014:639534
22. Spronk I, Schellevis FG, Burgers JS, de Bock GH, Korevaar JC (2018) Incidence of isolated local breast cancer recurrence and contralateral breast cancer: a systematic review. *Breast* 39:70–79
23. Geurts YM, Witteveen A, Bretveld R et al (2017) Patterns and predictors of first and subsequent recurrence in women with early breast cancer. *Breast Cancer Res Treat* 165(3):709–720
24. Hassett MJ, Ritzwoller DP, Taback N et al (2014) Validating billing/encounter codes as indicators of lung, colorectal, breast, and prostate cancer recurrence using two large contemporary cohorts. *Med Care* 52(10):e65–e73
25. Warren JL, Mariotto A, Melbert D et al (2016) Sensitivity of Medicare claims to identify cancer recurrence in elderly colorectal and breast cancer patients. *Med Care* 54(8):e47–e54
26. Wapnir IL, Khan A (2019) Current strategies for the management of locoregional breast cancer recurrence. *Oncol (United States)* 33(1):19–25
27. Rasmussen LA, Jensen H, Virgilsen LF, Thorsen LBJ, Offersen BV, Vedsted P (2019) A validated algorithm for register-based identification of patients with recurrence of breast cancer—based on danish breast cancer group (DBCG) data. *Cancer Epidemiol* 59:129–134
28. Holloway CMB, Shabestari O, Eberg M et al (2022) Identifying breast cancer recurrence in administrative data: algorithm development and validation. *Curr Oncol* 29(8):5338–5367
29. Kroenke CH, Chubak J, Johnson L, Castillo A, Weltzien E, Caan BJ (2016) Enhancing breast cancer recurrence algorithms through selective use of medical record data. *JNCI J Natl Cancer Inst* 108(3):336
30. Habbous S, Forster K, Darling G et al (2020) Incidence and real-world burden of brain metastases from solid tumors and hematologic malignancies in Ontario: a population-based study. *Neuro-oncology Adv* 3:1
31. Ontario Cancer Statistics 2020. 2020. <https://www.cancercare.ontario.ca/en/statistical-reports/ontario-cancer-statistics-2020/ch-1-estimated-current-cancer-incidence>. Accessed November 11, 2020
32. Pagani O, Senkus E, Wood W et al (2010) International guidelines for management of metastatic breast cancer: can metastatic breast cancer be cured? *JNCI J Natl Cancer Inst* 102(7):456
33. Coleman M, Forman D, Bryant H et al (2011) Cancer survival in Australia, Canada, Denmark, Norway, Sweden, and the UK, 1995–2007 (the International Cancer Benchmarking Partnership): an analysis of population-based cancer registry data. *Lancet* 377(9760):127–138
34. Miksad RA, Zietemann V, Gothe R et al (2008) Progression-free survival as a surrogate endpoint in advanced breast cancer. *Int J Technol Assess Health Care* 24(4):371–383
35. Kemp R, Prasad V (2017) Surrogate endpoints in oncology: When are they acceptable for regulatory and clinical decisions, and are they currently overused? *BMC Med* 15:1
36. Hernandez-Villafuerte K, Fischer A, Latimer N (2018) Challenges and methodologies in using progression free survival as a surrogate for overall survival in oncology. *Int J Technol Assess Health Care* 34(3):300–316
37. Seely JM, Alhassan T (2018) Screening for breast cancer in 2018—what should we be doing today? *Curr Oncol* 25(Suppl 1):S115–S124
38. Jones T, Duquette D, Underhill M et al (2018) Surveillance for cancer recurrence in long-term young breast cancer survivors randomly selected from a statewide cancer registry. *Breast Cancer Res Treat* 169(1):141–152
39. Zeng Z, Espino S, Roy A et al (2018) Using natural language processing and machine learning to identify breast cancer local recurrence. *BMC Bioinformatics* 19(Suppl 17):498

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.