2019

Metastatic breast cancer incidence, site and survival in Australia, 2001-2016: A population-based health record linkage study protocol

Sarah J. Lord
The University of Notre Dame Australia, sally.lord@nd.edu.au

Belinda E. Kiely

Sally-Anne Pearson

Benjamin Daniels

Dianne L. O'Connell

See next page for additional authors

Follow this and additional works at: https://researchonline.nd.edu.au/med_article

Part of the Medicine and Health Sciences Commons

This article was originally published as:

Original article available here:
10.1136/bmjopen-2018-026414
This is an Open Access article distributed in accordance with the Creative Commons Attribution-Non Commercial 4.0 International (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, provided the original work is properly cited and the use is non-commercial. See: https://creativecommons.org/licenses/by-nc/4.0/

This article originally published in BMJ Open available at: https://bmjopen.bmj.com/content/9/2/e026414
No changes have been made to this article.

BMJ Open

Metastatic breast cancer incidence, site and survival in Australia, 2001–2016: a population-based health record linkage study protocol

Sarah J Lord, Belinda E Kiely, Sallie-Anne Pearson, Benjamin Daniels, Dianne L O’Connell, Jane Beith, Max K Bulsara, Nehmat Houssami

ABSTRACT

Introduction Advances in systemic therapy for early and metastatic breast cancer (BC) over the last two decades have improved patients’ survival, but their impact on metastatic disease outcomes at a population level is not well described. The aim of this study is to investigate changes in the incidence, site and survival of metastatic disease for women with a first diagnosis of BC in 2001–2002 vs 2006–2007.

Methods and analysis Population-based retrospective cohort study of women with first primary invasive BC registered in the New South Wales (NSW) Cancer Registry in 2001–2002 and 2006–2007. We will use linked records from NSW hospitals, dispensed medicines, outpatient services and death registrations to determine: women’s demographic and tumour characteristics; treatments received; time to first distant metastasis; site of first metastasis and survival. We will use the Kaplan-Meier method to estimate cumulative incidence of distant metastasis, distant recurrence-free interval and postmetastasis survival by extent of disease at initial diagnosis, site of metastasis and treatment-defined tumour receptor type (hormone receptor-positive, human epidermal growth factor receptor-2-positive, triple negative). We will use Cox proportional hazards regression to estimate the relative effects of prognostic factors, and we will compare systemic therapy patterns by area-of-residence and area-level socioeconomic status to examine equity of access to healthcare.

Ethics and dissemination Research ethics committee approval was granted by the Australian Institute of Health and Welfare (#E2017/2/255), NSW Population and Health Services (#HREC/17/CIPHS/19) and University of Notre Dame Australia (#017.144S). We will disseminate research findings to oncology, BC consumer and epidemiology audiences through national and international conference presentations, lay summaries to BC consumer groups and publications in international peer-reviewed oncology and cancer epidemiology journals.

INTRODUCTION

Breast cancer (BC) is the most common cancer diagnosed in Australian women, affecting one in eight women by the age of 85 years; and the second most common cause of cancer death for women after lung cancer.1 Most women present with early (BC) and have surgery, with or without radiotherapy, as well as adjuvant systemic therapy to reduce their risk of subsequent relapse. Approximately 7% of women present with distant metastasis at diagnosis.2

Over the last two decades, there have been major advances in systemic therapy for both early and metastatic BC, including new cytotoxic therapies, endocrine therapies targeted to hormone receptor-positive tumours and therapies targeted to human epidermal growth factor receptor-2 (HER2)-positive tumours. In Australia, the introduction of these new therapies has been associated with reduced BC mortality rates.3

While cancer registries report BC mortality rates annually, metastatic BC outcomes are not reported routinely. Thus, clinicians do not have access to up-to-date information to counsel women about their risk of distant metastasis and survival postmetastasis with
current treatment protocols. This is a significant clinical challenge because following initial treatment of the primary tumour, a major source of anxiety for women is fear of recurrence. Studies of patients and BC consumer groups in Australia and internationally report that prognostic information represents an important unmet need. To address this gap, we require contemporary population-based estimates of the incidence, sites and survival of metastatic BC.

A large German cancer registry study demonstrated the 5-year risk of metastasis for women with an initial diagnosis of early BC fell from 27% to 15% from 1978 to 1984 to 1995–2003. A subsequent study using the same registry with women diagnosed in 2007–2013 indicated a further reduction in 5-year risk of metastasis to approximately 8%. However, follow-up time for this later cohort was limited; and investigators did not have access to treatment data.

The anatomical site of first metastasis appears to have also changed over time. For women with a diagnosis of early BC up to 2003, large cohort studies have reported a reduction in bone being the first site of metastasis over time, with an increased proportion with first metastasis to visceral sites such as the liver and brain. These findings raise the hypothesis that adjuvant therapies may delay or prevent bone metastases, or control BC subtypes that spread to bone, but have less impact on more aggressive subtypes that spread to visceral sites. There is also some evidence that tumour subtype is associated with site of spread. For example, women with HER2-positive tumours may be at higher risk of metastasis to the brain than other tumour subtypes. It is unknown whether newer treatments have altered these patterns of metastatic sites.

We previously estimated the 5-year risk of metastatic BC at 10% from a population-based sample of 6644 Australian women with an initial diagnosis of non-metastatic BC in 2001–2002. This study demonstrated a marked difference in 5-year risk of metastasis between women with node-negative disease (5%) versus those with node positive or locally advanced disease at presentation (18%). It is not yet known whether recent advances in adjuvant therapy have helped to reduce this disparity. Our previous research also found an independent association between risk of metastasis and area-defined socioeconomic status, after adjusting for other prognostic factors, but we did not have access to treatment data to explore the impact of different adjuvant therapies on metastasis among socioeconomic groups.

There is also limited evidence on the impact of changes in systemic treatment on metastatic BC survival at a population level. Trials of systemic therapies for metastatic BC have demonstrated improvements in overall survival, and trial meta-analyses have indicated improvements in survival over time. In contrast, large population-based studies have not consistently reported an overall improvement in metastatic BC survival over time, with no improvement for women with metastasis within 5 years of an initial diagnosis of early BC and conflicting findings for women with metastasis at initial diagnosis of BC. These studies do not examine the impact of the increased availability of targeted treatment on metastatic BC survival for specific tumour subtypes. Our recent Australia-wide population-based study of women treated with trastuzumab for metastatic BC has reported an improvement in median survival from 27 months to 38 months for women treated in 2001–2008 vs 2009–2015 respectively, with a longer treatment duration observed for the latter group.

Observational evidence indicates that patients exposed to systemic therapy in the adjuvant setting may progress more rapidly once metastases occur, although some have argued against this interpretation. Single centre observational studies including treatment data suggest that while metastatic BC survival has improved for women with hormone receptor-positive disease and HER2-positive disease, survival has not improved for women with triple negative disease where there are no targeted treatments.

To address the above-defined knowledge gaps, we will conduct a programme of research with the aim to assess the impact of systemic adjuvant therapy advances on the incidence, sites and survival of metastatic BC at a population level. This research will extend existing population-based evidence by describing metastatic disease outcomes for treatment-defined subgroups; and includes long-term data to report outcomes for women with a distant recurrence-free interval of greater than 5 years and long-term survivors following metastases (table 1).

OBJECTIVES

Primary research questions

1. Has the 5-year cumulative incidence of metastatic BC changed for women with an initial diagnosis of early BC in the period 2006–2007 vs 2001–2002, overall and within subgroups categorised by extent of disease at initial diagnosis (localised, regional, unknown) and treatment-defined tumour receptor subtypes?


Secondary questions

4. What is the annual hazard of developing distant metastases for women with an initial diagnosis of early BC in 2001–2002 and 2006–2007, overall and by extent of disease at initial diagnosis (localised, regional, unknown) and treatment-defined tumour receptor subtypes?

5. To what extent does metastatic BC survival vary by the presence of distant metastasis at initial diagnosis and time to distant metastasis for those initially diagnosed with non-metastatic BC; initial metastasis site and treatment-defined tumour receptor type?
Table 1  Study characteristics and comparison with recent population-based cohort studies assessing changes in metastatic breast cancer (BC) incidence, site and survival over time

| Lead author publication year country region | Initial diagnosis period | N     | Observation years/period | Extent at diagnosis | Systemic therapy/receptor subtype | Systemic therapy/receptor subtype | Adjuvant treatment patterns | Systemic treatment patterns | Systemic treatment patterns | MBC site | MBC site | MBC site | MBC site | MBC site | MBC site | MBC site | MBC site | MBC site | MBC site | MBC site |
|---------------------------------------------|-------------------------|-------|--------------------------|--------------------|----------------------------------|----------------------------------|--------------------------------|-----------------------------|-----------------------------|-----------------------------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|
| Present study                               |                         |       |                          |                    |                                  |                                  |                              |                             |                             |                             |          |          |          |          |          |          |          |          |          |          |          |
| Australia                                   | 2001–2002               | 8133  | 15+                      | All                | Subgroup            |                                  |                                  |                              |                             |                             |          |          |          |          |          |          |          |          |          |          |          |
| New South Wales                             | 2006–2007               | 8715  | 10+                      | All                | Subgroup            |                                  |                                  |                              |                             |                             |          |          |          |          |          |          |          |          |          |          |          |
| state-wide                                 | /to June 2017           |       |                          |                    |                                  |                                  |                              |                             |                             |                             |          |          |          |          |          |          |          |          |          |          |          |
| Hölzel*                                    | 2017*                   | 16    | 2017†                    | All                | Metastatic BC         |                                  |                                  |                              |                             |                             |          |          |          |          |          |          |          |          |          |          |          |
| Germany                                    | 1978–1984               | 9952  | 15+                      | All                | Subgroup            |                                  |                                  |                              |                             |                             |          |          |          |          |          |          |          |          |          |          |          |
| Munich                                     | 1998–2006               | 30 684| 7–15+                    | All                | Subgroup            |                                  |                                  |                              |                             |                             |          |          |          |          |          |          |          |          |          |          |          |
|                                            | 2007–2013               | 29 826| /to 2013                 | All                | Subgroup            |                                  |                                  |                              |                             |                             |          |          |          |          |          |          |          |          |          |          |          |
| van den Hurk*                              | 2011†                   | 8     | /to Oct 2008             | All                | Metastatic BC         |                                  |                                  |                              |                             |                             |          |          |          |          |          |          |          |          |          |          |          |
| Germany                                    | 1978–1984               | 4978  | 10                       | All                | Subgroup            |                                  |                                  |                              |                             |                             |          |          |          |          |          |          |          |          |          |          |          |
| Munich                                     | 1985–1994               | 10 201| 10                       | All                | Subgroup            |                                  |                                  |                              |                             |                             |          |          |          |          |          |          |          |          |          |          |          |
|                                            | 1995–2003               | 18 592| /to Oct 2008             | All                | Subgroup            |                                  |                                  |                              |                             |                             |          |          |          |          |          |          |          |          |          |          |          |
| Ording 2017†                               |                        |       |                          |                    |                                  |                                  |                              |                             |                             |                             |          |          |          |          |          |          |          |          |          |          |          |
| Denmark                                    | 1997–2001               | 1905  | 5                        | All                | Metastatic BC         |                                  |                                  |                              |                             |                             |          |          |          |          |          |          |          |          |          |          |          |
| nation-wide                                | 2002–2006               | 1664  | 5                        | All                | Metastatic BC         |                                  |                                  |                              |                             |                             |          |          |          |          |          |          |          |          |          |          |          |
|                                            | 2007–2011               | 819   | /to Jan 2013             | All                | Metastatic BC         |                                  |                                  |                              |                             |                             |          |          |          |          |          |          |          |          |          |          |          |

■ outcome assessed in specified time period; □ outcome assessed in part or subset.

*Participants for periods 1978–1987 and 1988–1997 are not population-based samples; metastatic BC survival time is reported by site stratified by HR subtype (positive, negative).
†Metastatic BC incidence and survival is reported for women with metastasis within 5 years of initial BC diagnosis; metastatic BC survival by HR subtype is reported from multivariable Cox regression.

BC, breast cancer; HR, hormone receptor; M0, distant metastasis occurring after initial diagnosis of non-metastatic BC; M1, distant metastasis at initial diagnosis of BC; NSW, New South Wales.
6. Does the use of systemic therapy for early and metastatic BC vary by a woman’s age, area of residence or area-level socioeconomic status?

METHODS
Study design
A population-based retrospective cohort study using health record linkage to access routinely collected health data.

Setting
New South Wales (NSW) is the most populous state in Australia with a population of 7.48 million in 2016.24 Australian citizens and permanent residents have access to subsidised healthcare from Australia’s publicly funded universal healthcare system, including free treatment in public hospitals, outpatient services and subsidised access to prescription medicines.

Participants
All women aged 18 years or older with a first primary invasive BC (ICD-10 C50.0–C50.9) registered in the NSW Cancer Registry in January 1, 2001–December 31, 2002 (cohort 1); or January 1, 2006–December 31, 2007 (cohort 2). We will exclude women with no NSW hospital admission records, because they are likely to have moved outside of NSW for their initial treatment, and therefore subsequent metastasis and death may not be captured. We will also exclude women with a non-breast primary cancer recorded in the Cancer Registry prior to, or at the same date as, their initial BC diagnosis to avoid misclassification of first metastasis as breast/non-breast.

The NSW Cancer Registry records approximately one third of all new BCs diagnosed in Australia and reports similar outcomes to those observed nationally, therefore the study sample will be broadly representative of BC cases across Australia during the study period.

Data sources
We will obtain study data from seven health datasets as described below. The key study variables are listed in table 2.

The NSW Cancer Registry receives notifications of all cancer cases diagnosed in residents of NSW, including new and recurrent cases of invasive BC and cancer deaths. Notification is a statutory requirement for public and private hospitals, pathology laboratories, departments of radiation oncology, outpatient departments, day procedure centres and nursing homes. Data collection follows International Association of Cancer Registries conventions. Demographic information includes patients’ date of birth, area of residence and country of birth. Data on tumour characteristics include extent of disease at diagnosis (localised, regional, distant, unknown) and International Classification of Diseases for Oncology version-3 (ICD-O-3) morphology codes.

The NSW Registry of Births, Deaths and Marriages (RBDM) records the date of all deaths registered in NSW. The date and cause are certified by a registered medical practitioner or a coroner if a coronial inquiry is required. The majority of non-coronial deaths are registered in the RBDM within 4 weeks of death.

The NSW Admitted Patient Data Collection (APDC) codes episodes of care recorded for all inpatient separations (discharges, transfers, deaths) from all public, private and repatriation hospitals, private day procedures centres and public nursing homes in NSW. A trained medical coder at each hospital reviews the patient’s medical record to classify the principal diagnosis, additional diagnoses and procedures (including surgery, radiotherapy and chemotherapy) using the International Classification of Disease version-10 (Australian-modified) (ICD–10AM) coding system.25

NSW BreastScreen is the Australian national BC screening programme. Women aged 40 years and over are eligible for free mammograms every 2 years. Invitations and reminders are mailed to women in the target screening population (women aged 50–69 years, extended to 74 years after July 2013). Screening participation rates were approximately 57% during the study period. BreastScreen records screening episodes, cancer diagnosis, histopathology, primary tumour size and grade, reason for histopathology (screening or interval cancer) and primary treatment.

Commonwealth Government Pharmaceutical Benefits Scheme (PBS) claims provide information on all subsidised medicines dispensed in Australia outside of public hospitals, including dates of prescription and supply.26 PBS claims include an Anatomical Therapeutic Classification (ATC) code and a PBS item code. The ATC code provides information about the class of drug using a five-level hierarchy. At the first level, ‘L’ classification refers to antineoplastic and immunomodulating agents; and lower levels refer to different classes of agents. The PBS item code provides information about the generic drug name, form, strength, administration route and approved indication.

For patients treated in NSW public hospitals, chemotherapy is almost entirely administered in outpatient clinics, thus we will have good ascertainment of cancer medications for all patients with cancer, irrespective of whether they receive care in public or private hospitals. From April 2012, PBS claims capture all PBS-listed medicines dispensed. Prior to April 2012, PBS claims were not recorded for medicines that cost less than the required patient copayment (up to $A35.40 or $A5.80 for concession holders in 2012). Nearly, all cancer medicines are priced higher than this co-payment cost, thus we will have comprehensive capture of cancer medicines over the study period.

The Commonwealth Government Herceptin Program subsidised trastuzumab for women with HER2-positive metastatic BC from December 2001 to June 2015, after which it was subsidised by the PBS.27 The PBS has subsidised trastuzumab for treatment of early BC since October 2006.


4

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Data extracted to code variable</th>
<th>Data source</th>
</tr>
</thead>
<tbody>
<tr>
<td>Demographic</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age</td>
<td>Date of birth (month, year) at primary diagnosis</td>
<td>Cancer Registry</td>
</tr>
<tr>
<td>Rurality</td>
<td>Accessibility/Remoteness Index of Australia at diagnosis</td>
<td>Cancer Registry</td>
</tr>
<tr>
<td>Socioeconomic</td>
<td>SocioEconomic Indexes for Areas at primary diagnosis</td>
<td>Cancer Registry</td>
</tr>
<tr>
<td>Country of birth</td>
<td>Country of birth (Australia/New Zealand, other)</td>
<td>Cancer Registry</td>
</tr>
<tr>
<td>Tumour</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Screen detected</td>
<td>Reason for histopathology recorded as screen detected; or screening ≤6 months of primary diagnosis if reason not recorded.</td>
<td>BreastScreen</td>
</tr>
<tr>
<td>Extent of disease</td>
<td>Local, regional, distant, unknown at diagnosis</td>
<td>Cancer Registry</td>
</tr>
<tr>
<td>Morphology</td>
<td>ICD-O-3 morphology code (ductal, lobular and mixed ductal/lobular, other) at primary diagnosis</td>
<td>Cancer Registry</td>
</tr>
<tr>
<td>Receptor status</td>
<td>PBS item codes for endocrine therapy, anti-HER2 therapy</td>
<td>PBS*, Herceptin Program</td>
</tr>
<tr>
<td>Other primary cancer/s</td>
<td>Number of other primary cancers, date of diagnosis for each (month, year) and ICD-O-3 morphology and topography codes</td>
<td>Cancer Registry</td>
</tr>
<tr>
<td>Treatment</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Primary surgery</td>
<td>APDC procedure codes for breast surgery (breast conserving, mastectomy, other) and axillary surgery (sentinel node biopsy, axillary dissection, other)</td>
<td>APDC</td>
</tr>
<tr>
<td>Primary radiotherapy</td>
<td>APDC procedure codes and MBS item codes for radiotherapy (MBS Group T2)</td>
<td>APDC, MBS</td>
</tr>
<tr>
<td>Systemic therapy</td>
<td>APDC procedure codes and MBS item codes for chemotherapy (MBS Group T1, subgroup 11) PBS item codes for first-line adjuvant therapy PBS item codes for first-line therapy for metastatic BC</td>
<td>APDC, MBS, PBS, Herceptin Program</td>
</tr>
<tr>
<td>Outcomes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Time to first distant metastasis</td>
<td>Date of primary BC diagnosis</td>
<td>Cancer Registry</td>
</tr>
<tr>
<td></td>
<td>Date of first notification where extent of disease is coded as distant</td>
<td>Cancer Registry</td>
</tr>
<tr>
<td></td>
<td>Date of first APDC principal or additional diagnosis codes for secondary malignant neoplasms (C77.0–C77.9, C78.0–C78.8, C79.0–C79.88), if no other prior primary cancer</td>
<td>APDC</td>
</tr>
<tr>
<td></td>
<td>Date of first PBS claim where code specifies MBC or not specified and≥12 months following primary diagnosis</td>
<td>PBS</td>
</tr>
<tr>
<td></td>
<td>Date of first MBS claim for new chemotherapy or radiotherapy ≥12 months following primary diagnosis if no other prior primary cancer</td>
<td>MBS</td>
</tr>
<tr>
<td>Site of first distant metastasis</td>
<td>First APDC principal or additional diagnosis code for secondary malignant neoplasms: C77.0–C77.9; C78.0–C78.7; C79.0–C79.8</td>
<td>APDC</td>
</tr>
<tr>
<td></td>
<td>First radiotherapy field code</td>
<td>MBS, APDC</td>
</tr>
<tr>
<td>Post-metastasis survival time</td>
<td>Date and cause of death (breast, non-breast/unknown)</td>
<td>Cancer Registry†, RBDM‡</td>
</tr>
</tbody>
</table>

We have requested all PBS claims with an Anatomical Therapeutic Classification code for antineoplastics and immunomodulating agents (LO1–LO4).

†The Cancer Registry collects fact and cause of death from the RBDM and the Australian Bureau of Statistics; and periodically updates these records with data from the National Death Index to capture deaths of patients who may have moved outside of NSW.

‡We will request death data directly from the RBDM to supplement the death data recorded in the Cancer Registry so the study can include the most recent death data available.

APDC, Admitted Patient Data Collection; BC, breast cancer; HER-2, human epidermal growth factor receptor-2; ICD-O-3, International Classification of Diseases for Oncology, third edition; MBS, Medical Benefits Schedule; PBS, Pharmaceutical Benefits Scheme; RBDM, Registry of Births, Deaths and Marriages.

Commonwealth Government Medicare Benefits Schedule (MBS) claims provide information on all subsidised outpatient consultations, investigations and non-pharmaceutical treatments including radiotherapy services (simulation, dosimetry and treatment) and date of service. The MBS item code provides information about
the type of service, such as chemotherapy services (group T1, subgroup 11) and radiotherapy services (group T2). For radiotherapy treatment services, this information includes field site (Radiation Oncology Category 3, T2).

**Health record linkage**

Health record linkage will be used to identify relevant records from each of the data collections for each study participant (figure 1). To assemble the two study cohorts, the Cancer Registry data custodian will provide the Centre for Health Record Linkage (CHeReL) with a list of registry record identifiers for eligible women in each study period. CHeReL will create a Project Person Number (PPN) for each person represented and perform the record linkage for the NSW-held datasets (Cancer Registry, APDC, RBDM and BreastScreen). The Australian Institute of Health and Welfare (AIHW) will perform the record linkage for the commonwealth-held data sets (PBS, MBS, Herceptin Program). Each data custodian will extract the approved study variables plus the PPN and forward this dataset to the investigators. We will use the PPN to link the datasets received from each data custodian.

**Outcomes**

**Primary outcomes**

The primary study outcomes are time to first distant metastasis and time to BC death. For women with an initial diagnosis of non-metastatic disease, we will define first distant metastasis as the date of the first Cancer Registry record of metastatic BC, first PBS or MBS claim for treatment of metastatic BC (chemotherapy, radiotherapy), or for women without other known non-BC primaries, the first APDC record of distant metastasis, whichever occurs earliest.

For PBS and MBS claims, we will define metastatic BC if the item code indication specifies metastatic BC; or if the treatment commenced 12 months or more after an early BC diagnosis. A 12-month period was chosen for this definition as treatment of the primary tumour is generally completed within 12 months. We anticipate women may change hormone therapy for their early disease, thus if a new treatment is commenced with a treatment break of less than 90 days after the initial treatment period then two clinicians will review the information for a decision on distant metastasis status.

For APDC records for women with no known prior non-BC primaries, we will define metastatic BC from the first episode of care where an ICD-10AM diagnosis code for secondary malignant neoplasms (excluding to axillary/upper limb lymph nodes).

We will determine BC-specific death from the cause of death recorded in the Cancer Registry or NSW RBDM (for recent deaths).

**Secondary outcomes**

We will determine site of first metastasis by ICD-10AM diagnosis code recorded in the APDC or MBS item code for radiotherapy. For women with a non-BC primary (diagnosed after their primary BC), we will only extract this information if a matching Cancer Registry record is available to verify metastatic BC status.
We will use death from any cause in a secondary analysis of overall survival following the first record of distant metastasis.

**Covariables**

We will extract women’s demographic and tumour characteristics at baseline from the Cancer Registry, APDC and BreastScreen (table 2). We will classify comorbidities using the Charlson co-morbidity index\(^28\) using ICD-10AM codes recorded in the APDC.\(^29\)

We will extract treatment types from the APDC (type of primary surgery), PBS (class of first-line systemic therapy for early BC and metastatic BC), Herceptin Program (first-line systemic therapy for metastatic BC) and MBS (radiotherapy field) (table 2). We will classify treatment as adjuvant therapy if commenced within 12 months of the initial early BC diagnosis and before the first recorded metastatic event. We will classify treatment as metastatic therapy if commenced at/after the first recorded metastatic event.

**Study size**

To identify a 1% difference in the cumulative incidence of metastatic BC for women with an initial diagnosis of non-metastatic BC in 2006–2007 vs 2001–2002, we require a total sample size of 1644. This provides a power of 99% to detect a 5-year cumulative incidence rate of 9% in 2006–2007 compared with 10% cumulative incidence in 2001–2002\(^2\) assuming a Poisson distribution and accepting a 5% significance level. Thus, with an estimated 6044 women in cohort 1 and 6850 women in cohort 2 (localised 62%, node positive or locally advanced 38%), we have sufficient power to address our primary research question.

We will also have over 99% power to detect a 1% difference in 5-year cumulative incidence in the subgroup of women with localised BC at diagnosis, using our initial study finding of a 5.3% risk in cohort 1, which requires a minimum of 628 women per group.

**Statistical methods**

Cumulative incidence of distant metastasis and distant recurrence-free interval

For women with an initial diagnosis of non-metastatic BC, we will estimate time to the first recorded distant metastasis from the Cancer Registry date of primary BC diagnosis to the inferred date of first distant metastasis. We will use the Kaplan-Meier method to calculate the cumulative incidence function and present a cumulative incidence curve for distant metastasis. We will also estimate the distant recurrence-free interval. We define distant recurrence as the first record of a distant metastasis or BC death.\(^30\) For this analysis, we will censor women with no distant metastasis or BC death recorded at 30 September 2016 (the ‘reference’ date), representing the last date Cancer Registry notifications of metastatic disease are available for the study. We will use the log-rank test to test for differences in the distant recurrence-free interval by: extent of disease at initial diagnosis (localised, regional, unknown); period of diagnosis (2001–2002, 2006–2007); area of residence (categorised as major city, regional, remote, using the Accessibility and Remoteness Index of Australia) and area-level socioeconomic status (SocioEconomic Indexes for Areas (SEIFA) quintiles). We will estimate annual hazard rates for distant metastasis by extent of disease at initial diagnosis and by period (2001–2002, 2006–2007).

We will use Cox proportional hazards regression to investigate associations between baseline characteristics and the distant recurrence-free interval including: age (<50, 50–69, ≥70 years); stage (localised, regional, unknown); tumour morphology (ductal, lobular and mixed ductal/lobular, other); area of residence (major city, regional or remote); SEIFA quintile; country of birth (Australia/New Zealand, other) and period of diagnosis (2001–2002, 2006–2007). We will test proportional hazards using the Harrell-Lee approach (1986) and if violated, fit a time stratified proportional hazards model. Bivariable models will be used to assess each variable, and a multivariable model will be developed using backwards elimination to eliminate variables that do not contribute to the model. Results will be reported as a HR, 95% CI and p value using the Wald test.

We will perform subgroup analyses for women classified by tumour receptor type according to type of adjuvant treatment dispensed as: hormone receptor positive (endocrine treatment); HER2-positive (anti-HER2 therapies) and triple negative cancer (no endocrine or anti-HER2 therapy).

**Metastasis site**

We will compare the distribution of the first metastatic site that is, bone, lung, liver and brain at 8 years postdiagnosis between cohort 1 vs cohort 2 and by adjuvant therapy-defined tumour receptor type.

**Postmetastasis survival**

We will estimate postmetastasis survival from the date of the first record of distant metastasis to the date of BC death. We will use the Kaplan Meier method to calculate the cumulative incidence function and present a cumulative incidence curve for BC death. We will undertake a secondary analysis of overall survival by defining the endpoint as death from any cause. We will also estimate relative metastatic BC survival using data for mortality rates for Australian women from the AIHW National Mortality Database.

We will compare postmetastasis survival by extent of disease at initial diagnosis (distant metastasis ‘M1’, no distant metastasis ‘M0’) using the log rank test. We will perform separate analyses for M1 and M0 groups to examine postmetastasis survival by site of metastatic spread (bone, lung, liver, brain, other); age at first record of distant metastasis (<50, 50–69, ≥70 years) and treatment-defined tumour receptor type. We will use Cox proportional hazards regression to investigate the impact on postmetastasis BC survival of baseline variables, time to distant metastasis and first-line treatment type.
for metastatic BC, using the approach described above for investigation of the distant recurrence-free interval. We will also use regression methods described by Fine and Gray to take into account non-BC deaths as a competing risk.\textsuperscript{31}

Analyses will be conducted using SPSS V.24 and SAS V.9.4 statistical software.

**Patient and public involvement**

A BC consumer representative from the Breast Cancer Action Group NSW contributed to the development of the study research questions, outcome measures and application for funding. A consumer representative from the Breast Cancer Network Australia will contribute to the interpretation and reporting of the results and dissemination to BC consumer groups. Study participants will not be identified or contacted for dissemination of results.

**Data access, transfer and disposal**

Investigators will access study data through the Sax Institute’s Secure Unified Research Environment (SURE) facility by using a secure project workspace setup on a Virtual Desktop Infrastructure environment. Electronic files can only enter and leave the project workspace in the SURE facility through a ‘Curated Gateway’ that is protected by three layers of perimeter firewalls.

The CHeReL will destroy the linkage keys created for the project 12 months after the record linkage has been completed. Once the study has been completed, and results reported, all of the study datasets, outputs and working files will be archived in SURE for a 7-year period then securely destroyed.

**Dissemination plan**

We will disseminate study findings to oncology, BC consumer and epidemiology audiences through national and international conference presentations such as the Clinical Oncology Society of Australia’s Annual Scientific Meeting and the San Antonio Breast Cancer Symposium and by submission for publication in international peer-reviewed oncology and cancer epidemiology journals. We will follow study reporting guidelines such as the ‘REporting of studies Conducted using Observational Routinely-collected health Data’ statement when presenting study findings.\textsuperscript{32} We will also prepare lay summaries of study findings for dissemination to BC consumer groups and for media release to the public.

Programming code used in the analyses will be available from the corresponding author. We are not permitted to share study data with other individuals for any purpose without specific approval from the relevant data custodians and human research ethics committees.

**DISCUSSION**

This project will provide valuable evidence about the extent to which advances in systemic adjuvant therapy since 2005 have had an impact on metastatic BC incidence, sites and survival for Australian women and variations by tumour receptor type (hormone receptor, HER2, triple negative) as defined by treatment received.

The major strength of this study is that the large population-based cohort will allow us to better describe outcomes from metastatic BC and for subgroups of women defined by adjuvant treatment regimens and baseline prognostic factors such as extent of disease at diagnosis and time to distant metastasis. With a minimum observation period of 13.75 years for cohort 1 and 8.75 years for cohort 2, this study will also allow us to report outcomes relevant to women with longer time to distant metastasis.

A limitation is that tumour receptor type will be inferred from treatment records. However, an Australian validation study of 167 BC cases has reported endocrine treatment is a very good proxy for oestrogen receptor status with a sensitivity of 93% (95% CI: 86% to 96%) and a specificity of 95% (95% CI: 83% to 99%) when using a 4.5-year lookback period for pharmaceutical dispensing claims.\textsuperscript{33} The authors’ review of false negative findings (oestrogen receptor-positive cancers not identified in the lookback period) indicated that cases were missed if initial diagnosis was close to the last date of the lookback period (resulting in a short follow-up period) or before the lookback period commenced. In the study proposed here, we will have all dispensing claims from time of diagnosis, thus we anticipate identification of receptor status to be even higher.

Internationally, our study findings will make an important contribution to BC care by providing information that can be used to counsel women with a new BC diagnosis about prognosis, including risk of first distant metastasis up to, and beyond, 5 years; sites of spread and the association between time and site of first distant metastasis and postmetastasis survival. This information will be valuable to inform long-term life planning and discussions about routine surveillance. Clinical trials of treatments are not designed or powered to address these questions.

Our findings will also help to improve the design of clinical trials of new treatment strategies by providing up-to-date estimates of distant metastasis event rates for more realistic sample size calculations. By identifying subgroups of women with poor prognosis, the study findings may also inform priorities for research to further develop effective targeted treatment.

Within Australia, a finding of variations in adjuvant therapy use and metastatic BC outcomes by geographical and socioeconomically classified areas may help to identify and redress inequities in delivery or access to cancer services. This information will be valuable to inform planning of specialised cancer treatment services and to improve access to specialised services.

**Author affiliations**

1The School of Medicine, University of Notre Dame Australia - Darlinghurst, Darlinghurst, New South Wales, Australia

2National Health and Medical Research Council (NHMRC) Clinical Trials Centre, The University of Sydney, Camperdown, New South Wales, Australia

3The Centre for Big Data Research in Health, The University of NSW, Kensington, NSW, Australia
The Institute of Health Research and the School of Medicine, University of Notre Dame, Fremantle, Western Australia, Australia

Contributors
SL, NH: conceived the research questions. SL, NH, BK, SAP, BD, DO'C, JB, MB: contributed to the development of the research questions and methods. All authors contributed to drafting the study protocol.

Funding
This research is supported by a National Health and Medical Research Council (NHMRC) Project Grant (ID: 1125433); National Breast Cancer Foundation (NBCF) Breast Cancer Research Leadership Fellowship (2016–2020), Nehmat Houseam; the NHMRC Centre of Research Excellence in Medicines and ID (ID: 1060407) and a Cooperative Research Centre Project (CRC-P) Grant from the Australian Government Department of Industry, Innovation and Science (ID: CRC-P-439).

Competing interests
None declared.

Ethics approval
Approved by the Australian Institute of Health and Welfare (#EO2017/2/255), NSW Population and Health Services (#HREC/17/CIPHS/19) and University of Notre Dame Australia (#0171445).

Provenance and peer review
Not commissioned; peer reviewed for ethical and funding approval prior to submission.

Open access
This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commerically, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: http://creativecommons.org/licenses/by-nc/4.0/.

REFERENCES