

Application Form for Chapter 6, Part 4 of the Public Health Act 2005

Note: where boxes are provided, at least one requires a tick

There are 10 parts to this form and all must be completed and submitted

1. Application Type

☒ Initial

☒ HREC approval letter attached

☐ Amendment

No _____

Reference number of initial PHA

grant _____

☐ HREC amendment approval letter attached where appropriate, for example, if changes to the protocol or changes to information requested have occurred (check with the HREC for guidance)

2. HREC information

Name of Reviewing HREC	Queensland University of Technology
HREC Approval Number (also insert HREC approval number in page header)	4884
HREC approval expiry date (or if no expiry date then indicate that approval is ongoing and provide the due date of annual report to HREC)	13 January 2027 (subject to receipt of satisfactory progress or conditional reports)
Project name (also insert Project name in page footer)	Full title: Understanding geographical variations in breast cancer vulnerability in Queensland using linked administrative health datasets: a proof-of-concept study Abbreviated title: Urban Health Indicators - Breast Cancer

3. Applicants

Coordinating Principal Investigator (CPI) / Principal Investigator (PI)

(Please note: CPI/PI must be the person to whom the HREC approval is addressed)

Title	Professor	Name	Peter Baade
Institution and Position Title	Cancer Council Queensland (Senior Manager, Descriptive Epidemiology) Queensland University of Technology (Adjunct Professor)		
Postal Address	PO Box 201, Spring Hill, Queensland 4004		
Telephone	(07) 3634 5317	Email	peterbaade@cancerqld.org.au

Contact person for notices (if different from CPI/PI)

Title	eg Dr, Mr, Ms	Name	
Institution and Position Title			
Postal Address			
Telephone		Email	

All researchers or persons being given the Information (insert more rows if required)

Title and Name of Person	Position Title	Name of Institution
Professor Kerrie Mengersen	Distinguished Professor	Queensland University of Technology
Dr Paramita Dasgupta	Research Fellow	Cancer Council Queensland
Dr Jessica Cameron	Research Fellow	Cancer Council Queensland
Dr James Retell	Data Analyst	Cancer Council Queensland
Dr Belinda Goodwin	Manager, Health systems and psycho-oncology	Cancer Council Queensland
Dr Larry Myers	Post-doctoral research fellow	Cancer Council Queensland
Dr Derrick Lopez	Research Fellow	The University of Western Australia

4. Locations where project will be conducted

e.g. RBWH, PAH

Cancer Council Queensland

5. Description of the HREC approved research

Please answer the following questions in the space below.

Please note: information provided here must not deviate from the HREC approved protocol.

5.1 Describe the research project including the research objectives and expected outcomes

Project outline

The project is one of the several projects within the Australian Urban Research Infrastructure Network (AURIN) Urban Health Indicators Project (AusUrb-HI). The AusUrb-HI project aims to develop new indicators of health and its determinants by integrating health, socio-economic and other urban data sets to provide a more holistic and spatially explicit understanding of the mental and physical health of the urban population. The AusUrb-HI project has provided funding to carry out this specific study, as well as input into the aims, objectives and data access processes.

This ethics applications relates specifically to the cancer arm of the AusUrb-HI project. Separate ethics applications will be submitted for other arms of the AusUrb-HI project. The cancer arm of the AusUrb-HI project will collaborate with the Australian Cancer Atlas Research Team (<https://atlas.cancer.org.au/>) to develop indicators quantifying the association between:

- (i) demographic, socioeconomic, health service utilisation and environmental factors and
- (ii) incidence and survival rates for cancer in urban and regional areas across Queensland. This proof-of-concept study will initially focus on breast cancer.

Background information

Cancer is a leading cause of death in Australia. Approximately 50,000 Australians died from cancer last year. According to Cancer Council Australia, 150,000 cases per year are projected by 2020 and 1 in 2 Australians will be diagnosed with cancer by the age of 85.

Released in 2018 as the first of its kind, the Australian Cancer Atlas is an online, interactive platform that provides modelled estimates of incidence and relative survival for 20 cancer types for over 2000 small areas across Australia. The term "small areas" is defined by the Statistical Area 2 (SA2) classification used by the Australian Bureau of Statistics. The Australian Cancer Atlas was developed by statisticians, cancer researchers, visualisation experts and IT specialists from Cancer Council Queensland and Queensland University of Technology, supported by the Australian Institute of Health and Welfare and FrontierSI (formerly the Cooperative Research Centre for Spatial Information), with additional input from government, community members, consumers, and the media.

To date, the Australian Cancer Atlas has attracted almost 550 media mentions, 600 downloads and 58,000 unique users. It is also currently being replicated in New Zealand and the Netherlands, with the methods also being utilised in Ontario, Canada. In its short life, it has demonstrated Australia's leadership in spatial technology while using health data in fresh and creative ways to guide research, advocacy, and health service delivery, and this was recognised by the national peak body catering for the spatial information industry, the Surveying and Spatial Sciences Institute, from whom the Atlas received the 2019 Award for Spatial Enablement and also the JK Barrie industry award for overall excellence. Thus, the Atlas has set standards for how health statistics can be made accessible to a diverse range of stakeholders, including the broader community. It has informed action for public health, allowed improved decision making and enabled service planners and policy makers to identify geographic patterns of inequity.

Through the development of spatial statistical models and innovative visualisation methods, the Australian Cancer Atlas provides a unique foundation to unlock insights into the extent and burden of cancer caused by geographical variation across the country. For example, in Queensland alone, it has been estimated that removing these geographical disparities could save almost 5,400 lives over nine years. These types of analyses have contributed to changes in government policy to reduce these inequalities, including a

doubling of travel subsidies for patients living outside major cities to access cancer treatment.

However, to date there has been little, if any investigation of the geographical variation between small areas for different cancer types within urban areas, and between different urban areas. Given that 86% of Australians live in urban areas, these investigations are critical for health managers, government, and the public.

In this proof of concept of study, investigators will focus on breast cancer in women diagnosed in Queensland. Breast cancer is the most common cancer diagnosed among Australian women. In Queensland, there are more than 3,500 new cases of breast cancer diagnosed each year.

Study justification

The current version of the Australian Cancer Atlas only includes cancer incidence and survival estimates and does not include any possibly explanatory variables such as sociodemographic or health services utilisation data. While further development of the Atlas, started in March 2021, will provide additional measures relating to cancer burden, our current ability to uncover insights about geographic variation in cancer incidence and survival remains frustratingly restricted: we can learn what variation exists, but not why. Linked health data may provide further insights as to why this variation exists. In addition, to date there has been little, if any investigation of geographical variation between small areas for different cancer types within urban areas, and between different urban areas. Given that 86% of Australians live in urban areas, these investigations are critical for health managers, government and the public.

In this proof-of-concept study, the investigators will use linked health datasets to explore these geographical variations, with a particular focus on urban and regional areas of Queensland. This project will focus solely on breast cancers diagnosed in Queensland, with future aims of expanding this to other types of cancer types diagnosed within Queensland, and eventually, on a national scale. These broader projects will be the subject of separate ethics applications in the future.

Hypotheses

Geographical variation in health services utilisation, including breast cancer screening and diagnostic services, are important, independent determinants of disparities in breast cancer survival outcomes.

Research questions/ aims

The overall aim of this case study is to develop new indicators of breast cancer vulnerability in urban areas. The key questions that will be addressed in the study are as follows:

1. What is the extent and nature of the geographical variation in breast cancer screening, incidence, surgical treatment, and survival among women living in the urban and regional areas of Queensland?
2. What are the key area-level demographic, socioeconomic, health service and environmental variables affecting screening participation, incidence, surgical treatment and survival for breast cancer?
3. How many breast cancers could be prevented, and how many lives saved, if the geographical variation within urban and regional areas was removed?
4. What are effective visual methods by which to communicate these associations to a wide audience?

In order to answer these questions, the project will develop a suite of new indicators:

- (a) A Spatial Variability Indicator (SVI), which will quantify the spatial variation in breast cancer screening, incidence, hospitalisation and survival in an urban region. This indicator will enable comparison between urban regions and also facilitate identification of small areas of most concern within a region. This will address Research Question 1 above.
- (b) A Cancer Vulnerability Indicator (CVI), based on identified key drivers for breast cancer screening, breast cancer incidence, survival, and hospitalisations. This indicator will quantify the relative importance of each of the key drivers and enable identification of small areas in an urban region that are vulnerable to higher cancer incidence, lower survival, and/or increased hospitalisations. This will address Research Question 2 above.
- (c) A Variation Adjusted Cancer Indicator (VACI), which will quantify the number of breast cancers prevented and lives saved if the geographical variation was removed. This will address Research Question 3 above.

Outcome measures:

- (a) Number of women screened for breast cancer, number diagnosed with breast cancer, and number receiving related healthcare (e.g. breast cancer surgery, radiotherapy, chemotherapy) within 45 days of diagnosis by small geographical area (SA2);
- (b) Number of women screened for breast cancer, number diagnosed with breast cancer, and number receiving related healthcare within 45 days of diagnosis for each SA2, by key area-level demographic, socioeconomic, availability of health services and environmental variables; and
- (c) Number of cancers that would be diagnosed at an earlier stage, and the number of breast cancer deaths that would be prevented, if these outcomes among women living in the small geographical areas in the lowest three quartiles were instead equivalent to the outcomes for women living in the geographical areas in the highest 25th percentile.

5.2 Outline the methodology (include the number of patients required and justification for this number as well as for the date range for the data items and if any patients aged <18 years)

Study design

This is a retrospective cohort study of women diagnosed with breast cancer using linked health datasets from the Queensland Cancer Register, BreastScreen Queensland and Queensland Hospital Admitted Patients Data Collection.

The Queensland Cancer Register maintains a register of all cases of cancer diagnosed in Queensland since 1982 (except keratinocyte cancers). Notification of cancer is a statutory requirement for all public and private hospitals, nursing homes and pathology services throughout Queensland. In 2018 there were over 3,700 women diagnosed with invasive breast cancer in Queensland, and over 700 diagnosed with breast cancer in situ. The Queensland Cancer Register is routinely linked with the National Death Index to obtain mortality information for people diagnosed with cancer.

The Queensland Hospital Admitted Patients Data Collection collects demographic data and clinical information on all admitted patients separated from both public and licensed private hospitals and private day surgeries in Queensland. There are nearly 1 million hospital separations in Queensland each year, with about two-thirds (66%) of these in public hospitals.

BreastScreen Queensland is part of the BreastScreen Australia Program which is jointly funded by the commonwealth, state and territory governments. The program commenced in 1991 and provides the only free population-based breast cancer screening services in Queensland. A total of 380,122 women aged 50-74 were screened by BreastScreen Queensland during 2018-2019.

Study participants

There are two non-exclusive study cohorts who are identified from the following two datasets:

Dataset A from BreastScreen Queensland consists of all eligible women aged 50-74 years who had a mammography screen between 2015 and 2021 irrespective of results. The age group of 50-74 years is consistent with the current target age group for breast cancer screening. A total of 380,122 women aged 50-74 were screened by BreastScreen Queensland during 2018-2019. It is estimated there will be about 680,000 women included in the BreastScreen Queensland cohort (Dataset A) for the period 2015-2021.

Dataset B is based on a cohort extracted from the Queensland Cancer Register consisting of all women (estimated 75,000 women) aged 20 years and over who were diagnosed with in situ or invasive breast cancer in Queensland between 1 January 2000 and 31 December 2019. We will be linking this Queensland Cancer Register dataset to the Queensland Hospital Admitted Patients Data Collection (Dataset C) and the BreastScreen Queensland (Dataset D) datasets via the Statistical Services Branch of Queensland Health. The final linked dataset (linking Dataset B with Datasets C and D) will consist of records from the Queensland Hospital Admitted Patients Data Collection and BreastScreen Queensland that link to women within the Queensland Cancer Register cohort (irrespective of age at diagnosis).

Supplementary data

In addition to person-level data in Datasets A and B, we will also be accessing aggregated non-identifiable data from (i) Australian Bureau of Statistics Census, (ii) Social Health Atlas, and (iii) Location of Medical Services

(i) Population-level data from Australian Bureau of Statistics

We will utilise the following population-level information defined at the SA2 level using Australian Bureau of Statistics (ABS) data based on the ABS Census including the 2001, 2006, 2011 and 2016 Census and other data sources. Specific years of data available will vary depending on the indicator, and include data interpolated for between-Census years. The data are publicly available and no ethics approval is required. The following ABS Census data have already been extracted for each SA2, supplemented with additional data corresponding to earlier years if available:

- Population density (people per square kilometre, 2000-2019)
- Income (per cent of population for each income category, 2014-2019, yearly)
- Education (per cent of population in each Education category, 2011, 2016)
- Unemployment rate (per cent of population unemployed, 2011, 2016)
- Elderly (per cent of population aged 65 years and over, 2000-2019, yearly)
- Infants (per cent of population aged under 1 year of age, 2000-2019, yearly)
- Single parents of dependants (per cent of population who are single with dependants, 2011, 2016)
- Unpaid childcare (per cent of families with children in unpaid childcare, 2011, 2016)
- Persons requiring significant assistance (per cent of population requiring specific assistance, 2011, 2016)
- Persons with a disability (per cent of population with a disability, 2015, 2018)
- Living alone (per cent of population living alone, 2011, 2016)
- English second language (per cent of population having English as a second language, 2011, 2016)
- Indigenous (per cent of population who are identified as Aboriginal or Torres Strait Islander, 2011, 2016)
- Mobile homes (per cent of population living in a mobile home, 2014-2019)
- Crowded dwellings (per cent of population living in crowded dwellings, 2016)
- Renters (per cent of population who are renting their usual residence, 2011, 2016)
- Mortgage payers (per cent of population who are paying a mortgage, 2011, 2016)

(ii) Population-level data from the Social Health Atlas

We will utilise the following population-level information defined at the SA2 level from the Social Health Atlas. The Social Health Atlas (<https://phidu.torrens.edu.au/social-health-atlases>) is developed and managed by the Public Health Information Development Unit at Torrens University, South Australia, and provides information about a range of health-related measures by various geographical classifications, including Population Health Areas, which are derived from single or composite SA2 areas. The data are publicly and freely available via a Creative Commons Attribution-NonCommercial-ShareAlike 3.0 Australia licence (CC BY-NC-SA 3.0 AU), enabling these data to be used for non-commercial purposes. Since the estimates are provided as area-specific percentages, no ethics approval is required. The following data will be used, primarily relating to the 2015 calendar year and later:

- Condition: Cardiovascular disease (CVD) / asthma
- Condition: High blood pressure / chronic obstructive pulmonary disease (COPD)
- Condition: Overweight/Obese
- Condition: Obese
- Condition: Smoker
- Condition: High alcohol consumption
- Condition: High distress levels
- Condition: Low exercise levels
- Condition: Adequate fruit intake

(iii) Location of Medical Services

We will utilise the following population-level information defined at the SA2 level for Location of Medical Services. The data are publicly available through Queensland Health

websites and specific government departments and health-related organisations, and no ethics approvals are required. The following data will be used:

- Location (based on SA2) of public and private hospitals (number and type per SA2)
- Location (based on SA2) of general practitioners (number per SA2)
- Location (based on SA2) of other medical services (number and type per SA2)

Statistical plan

Sample size determination and power

Given that the study cohort reflects the whole population of women diagnosed with invasive or in-situ breast cancer in Queensland during the study time period, there are no specific sample size calculations. The confidence or credible intervals of model coefficients for each of the statistical models will be examined to ensure the estimates are able to be interpreted. Bayesian spatial and spatio-temporal models, as used in the Australian Cancer Atlas and/or subsequent development, have the advantage of borrowing information from neighbouring geographical areas, thus providing more stable estimates even in the presence of small numbers. Model convergence will be assessed through standard methods (Duncan et al., 2020).

Data analysis and statistical methods

Previous Australian studies, including those relating to breast cancer, have examined the association between socioeconomic disadvantage, geographic remoteness, and health indicators based on hierarchical or multilevel models (Baade et al, 2013; Dasgupta et al, 2014; Bentley et al, 2008; Turrell et al, 2007a; Turrell et al, 2007b; Kavanagh et al, 2006; Turrell et al, 2008), sometimes incorporating spatial information derived from a geographic information systems analysis (Baade et al, 2013; Dasgupta et al., 2013; Turrell and Giskes, 2008). Other studies, including those relating to breast screening participation and outcomes (Hsieh et al, 2016; Hsieh et al, 2015) have utilised Bayesian hierarchical models as the general statistical framework for data analysis and modelling (Lawson et al, 2003). These Bayesian models follow the general spatial model of Besag, York and Mollie (1991), including extensions to spatial clusters (Lawson and Denison, 2002) and other spatially correlated outcomes (Lawson et al, 2003).

Logistic and Poisson models with extra-binomial and extra-Poisson variation and allowance for zero-inflation will be considered as potential methods to combine individual and area-level data, allowing for sparse disease outcomes where appropriate (Jackson et al, 2008). Assessment of temporal changes will consider the spatio-temporal models proposed by Bannerjee et al (2003, 2004), Berardinelli et al (1995), Knorr-Held (2000), Utazi et al (2018) and Gomez-Rubio et al (2019) among others.

Separate models will be investigated for the separate indicators, including screening participation, breast cancer incidence (diagnosis), treatment type and intervals and survival. Individual-level data obtained from the data linkage for this study, along with the area-level data from the Census and other data sources will be combined using the multilevel format.

Outcome measures

A Spatial Variability Indicator (SVI), which will quantify the spatial variation in each of the study indicators, including breast cancer screening, incidence, treatment, and survival in urban and regional areas. This indicator will enable comparison between urban regions and also facilitate identification of small areas of most concern within a region. This work will build on the Tango's Maximised Excess Events Test (Tango, 2000) used in the Atlas of cancer in Queensland (Cramb et al, 2011) and the Australian Cancer Atlas

(atlas.cancer.org.au) but expand it to consider the specific characteristics of the variation. This work will feed directly into the future expansion of the Australian Cancer Atlas.

A Cancer Vulnerability Indicator (CVI), based on identified key drivers for breast cancer screening, incidence, survival and hospitalisations. This indicator will quantify the relative importance of each of the key individual- and area-level characteristics associated with the outcomes and enable identification of small areas in an urban region that are vulnerable to higher breast cancer incidence, lower survival, and/or more appropriate and timely access to recommended treatment.

A Variation Adjusted Cancer Indicator (VACI), which will quantify the number of breast cancer-related deaths avoided within five years of diagnosis if the geographical variation were removed. We have previously quantified the numbers of deaths for the Atlas of cancer in Queensland (Cramb et al, 2011b), and estimated that if the survival among cancer patients in more rural or disadvantaged areas were raised to the Queensland average, over 1,200 cancer deaths (9% of total) within five years of diagnosis could have been prevented. Different strategies to quantify the concept of preventable deaths will be investigated.

Participant consent

We are seeking waiver of consent for both cohorts based on the following justifications:

- a. there is no direct involvement of patients in the research, but we will use data that were derived from their previous encounters with health service providers dating back to 1997 (depending on the data source);
- b. since the names or addresses of people are not included in the data extract provided to the research team, any requirement to seek consent will increase the threat to privacy due to the need to identify these people for the purpose of seeking consent;
- c. it is reasonable to assume that most patients would have consented to our research proposal if they had been asked;
- d. there will be an estimated 1 million+ and 75,000 women in Datasets A and B respectively, making it very difficult to obtain informed consent from this large number of women, some of whom would have died while others may be in poor health as a result of their illness. Excluding women who have died would seriously affect the scientific merit of the study (death is one of the outcomes of interest in the study);
- e. attempts to collect consent from patients (or their carers) who have experienced adverse events could induce unnecessary anxiety or distress;
- f. deidentified data provided to the research team to be used in the analyses provides no potential threat to the well-being of the patients concerned because the screening, diagnosis and treatments have already occurred. The benefit is that we will determine geographical variations in screening, incidence and outcomes of women with breast cancer, with the ultimate aim of preventing this cancer and providing best care for women diagnosed with breast cancer in the future, irrespective of their geographical location;
- g. measures to ensure privacy and the confidentiality of information during the data linkage will follow well-established procedures developed by the Statistical Services Branch of Queensland Health. These methods are now well established as being protective of the privacy of individuals.
- h. Storage of data will provide security for the de-identified data provided to the Research team, and the statistical models proposed for these analyses are specifically designed to preserve the confidentiality associated with small area data.

5.3 Justify the use of Information and how the benefits of the project (to the public) outweigh the risks for the individual/s whose Information will be used

Potential risks associated with the project

Given that no individuals will be contacted or identified during this study, the risk of harm or discomfort to any participants whose deidentified data will be extracted is negligible. It is possible that through the results of this study, specific geographical areas will be identified as having a substantially greater or lower vulnerability of breast cancer. When this result is solely due to random variation due to small numbers of cases, it may lead to unnecessary worry and distress among people living in those areas, or those who know people living in those areas. Importantly, it is likely that areas with a "real" higher or lower vulnerability and related outcomes exist, and so it is important that these areas are identified to then motivate and inform efforts to intervene and result these disparities. This project will provide a unique opportunity to identify these areas.

How will risks be mitigated and managed

The statistical methodology to be used in this study is specifically designed to reduce the possibility of spuriously inflated (or reduced) estimates due to small numbers. In addition, a key component of our approach is to incorporate meaningful estimates of uncertainty around any estimates. Combined with novel visualisation methods, we intend to reduce the likelihood of incorrect interpretation and thus reduce unnecessary worry and distress. When there is evidence that a specific geographical area does have poorer cancer screening participation, we will investigate, at an ecological level, possible drivers of those differences, and use those ecological findings to communicate the poorer outcomes more effectively. Cancer Council's 131120 Help Line will be made available for people to call should they have concerns about the results in their particular area. Any people who call about their concerns and distress about the results will be provided with additional information, and offered support and guidance as to how they might reduce their own risk and those around them.

Benefits associated with research

The current version of the Australian Cancer Atlas only includes cancer incidence and survival estimates and does not include any possibly explanatory variables such as sociodemographic or health services utilisation data. While further development of the Atlas, started in March 2021, will provide additional measures relating to cancer burden, our current ability to uncover insights about geographic variation in cancer incidence and survival remains frustratingly restricted: we can learn what variation exists, but not why. Linked health data may provide further insights as to why this variation exists. In addition, to date there has been little, if any investigation of geographical variation between small areas for different cancer types within urban areas, and between different urban areas. Given that 86% of Australians live in urban areas, these investigations are critical for health managers, government and the public. In this proof-of-concept study, the investigators will use linked health datasets to explore these geographical variations, with a particular focus on urban and regional areas of Queensland. This project will focus solely on breast cancers diagnosed in Queensland, with future aims of expanding this to other types of cancer types diagnosed within Queensland, and eventually, on a national scale.

6. Description of the data items required (Information) and data custodian authorisation

Please copy and paste this entire section 6 for each data source and corresponding data custodian authorization (i.e. copy and paste 6.1 and 6.2 together for each data source).

HREC approval number: 4884

6.1 Data items

Data source	BreastScreen Queensland (Dataset A)
Cohort Scope (e.g. inclusion and exclusion criteria such as age group, diagnoses, hospital admission date range, geographic area, any control group/s)	Dataset A will include details of all Queensland women who have been screened for breast cancer for the period 1 January 2015 to 30 June 2021, irrespective of whether they have a breast cancer diagnosis or not.
Data Scope (e.g. descriptors such as demographics, diagnoses, clinical factors, pathology, radiology, therapy, outcomes)	Dataset A will include details of all Queensland women who have been screened for breast cancer for the period 1 January 2015 to 30 June 2021, irrespective of whether they have a breast cancer diagnosis or not.
Data item/s to be given to researchers (e.g. name, date of birth, sex, antibiotics, histology, readmissions)	<p>We will not be requiring any identifying patient information, with patients only identified by a unique project specific identification number. We have requested the following variables:</p> <ul style="list-style-type: none"> • Unique person ID number (specific to this project) • Age at Breast Screen episode (whole years) • Month and year of Breast Screen episode • Age at death (if relevant, in whole years) • Indigenous status • Non-English-speaking background (Yes/No) • Episode number • Screening outcome • Assessment outcome (where applicable) • Surgical Histopathology data (Invasive cancer/DCIS/Other/Nodes sampled-involved/Sentinel Node Biopsy) for the "Dominant Lesion" • Primary treatment data (type of surgery, other treatments, location of surgical unit, evidence of metastases) • SA2 (2016 ASGS) of residence at time of screening episode <p>BreastScreen Queensland will also provide SA2-specific participation rate estimates that have been adjusted for known issues relating to the geographical data.</p>
Date range (a specific start date and end date is required e.g. 01/01/2004 to 31/07/2014)	1 January 2015 to 30 June 2021
Frequency of request (e.g. once / six monthly)	Once
Data items to be used for linkage (if applicable)	BreastScreen Queensland (Dataset A) will not be linked to any other datasets

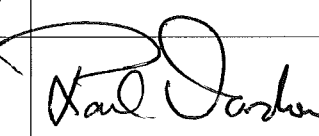
6.2 Data custodian authorisation for the aforementioned data items (collectively, Information)

Part 4, Public Health Act 2005 Application form, version: September 2021
Urban Health Indicators - Breast Cancer

Data custodian declaration:

I have considered this application and have seen all the relevant documents (including the HREC approval letter) that are required for this research and confirm that our unit is able to provide the data services indicated in this application within present resources.

In supplying the Information for an approved research project, no warranties are made as to: the fitness of the Information; the appropriateness or validity of the proposed research methods; or the appropriateness of the purpose for which the Information is being given.

Repository name	BreastScreen Queensland (Dataset A)		
Data custodian to list any special conditions or limitations that may be applied as a condition by the decision-maker (e.g. cost per record or whether authorization is dependent on availability of staff at the time of data request):			
Data Custodian Name	Paul Varchen.		
Position	Acting Executive Director, Preventive Health Branch.		
Unit	Cancer Screening Unit		
Hospital / Facility	Queensland Health		
Signature of data custodian		Date	14/6/2022

Signature of data custodian must be no more than twelve (12) months old at time of submission to the Director-General's delegate.

Data source	Queensland Cancer Register (Dataset B)
Cohort Scope (e.g. inclusion and exclusion criteria such as age group, diagnoses, hospital admission date range, geographic area, any control group/s)	All women aged ≥ 20 years who were diagnosed with breast cancer (ICD-O-3 = C50, in-situ and invasive) in Queensland between 1 January 2000 and 31 December 2019. This dataset will form the study cohort from which the subsequent data linkage will take place
Data Scope (e.g. descriptors such as demographics, diagnoses, clinical factors, pathology, radiology, therapy, outcomes)	All women aged ≥ 20 years who were diagnosed with breast cancer (ICD-O-3 = C50, in-situ and invasive) in Queensland between 1 January 2000 and 31 December 2019. This dataset will form the study cohort from which the subsequent data linkage will take place.
Data item/s to be given to researchers (e.g. name, date of birth, sex, antibiotics, histology, readmissions)	<ul style="list-style-type: none"> • Study ID (person based, deidentified, created by SSB) • Tumour ID (deidentified, based on QCR Tumourid) • Sex (females only) • Indigenous status (Aboriginal or Torres Strait Islander, other Australian, not stated) • Morphology • Differentiation • Behaviour code (in-situ, invasive) • Basis of diagnosis • Laterality • Tumour size (mm) • Number of lymph nodes examined • Number of lymph nodes positive • Age at diagnosis in whole years • Age at death (if applicable) in whole years • Cause of death (if applicable) • Survival time in days between diagnosis and death (if applicable) or diagnosis and last follow-up time. • Month and year at diagnosis • Month and year at death (if applicable) • Marital status (married, never married, widowed, divorced, separated, unknown) • Occupation (professional, white collar, blue collar, not in labour force, not stated) • Country of birth (recoded to Australia, NZ/UK/North America, elsewhere, not stated) • SA2 (2016 ASGC) of usual residence at diagnosis Variables described above
Date range (a specific start date and end date is required e.g. 01/01/2004 to 31/07/2014)	Women aged 20 years and over who were diagnosed with breast cancer in Queensland between 1 January 2000 and 31 December 2019
Frequency of request (e.g. once / six monthly)	Once

Data items to be used for linkage (if applicable)	<p>For the linkage process, Queensland Cancer Register will provide the Statistical Services Branch with identifying data:</p> <p>Queensland Cancer Register patient identification number</p> <p>Queensland Cancer Register tumour id</p> <p>QCR notification hospital UR number</p> <p>QCR notification hospital facility number</p> <p>full name</p> <p>date of birth</p> <p>sex</p> <p>address, suburb, postcode</p> <p>via secure electronic file transfer (kiteworks). The Statistical Services Branch will use deterministic and probabilistic linkage to link the datasets.</p>
--	---

6.2 Data custodian authorisation for the aforementioned data items (collectively, Information)

Data custodian declaration:

I have considered this application and have seen all the relevant documents (including the HREC approval letter) that are required for this research and confirm that our unit is able to provide the data services indicated in this application within present resources.

In supplying the Information for an approved research project, no warranties are made as to: the fitness of the Information; the appropriateness or validity of the proposed research methods; or the appropriateness of the purpose for which the Information is being given.

Repository name	Queensland Cancer Register (Dataset B)		
Data custodian to list any special conditions or limitations that may be applied as a condition by the decision-maker (e.g. cost per record or whether authorization is dependent on availability of staff at the time of data request):			
Data Custodian Name	Danica Cossio		
Position	Manager		
Unit	Queensland Cancer Control and Analysis Team		
Hospital / Facility			
Signature of data custodian		Date	

Signature of data custodian must be no more than twelve (12) months old at time of submission to the Director-General's delegate.

Data source	Queensland Hospital Admitted Patient Data Collection (Dataset C)
Cohort Scope (e.g. inclusion and exclusion criteria such as age group, diagnoses, hospital admission date range, geographic area, any control group/s)	All women aged ≥20 years who were diagnosed with breast cancer (identified from the Queensland Cancer Register) in Queensland between 1 January 2000 and 31 December 2021.
Data Scope (e.g. descriptors such as demographics, diagnoses, clinical factors, pathology, radiology, therapy, outcomes)	Where the cohort links to the QHAPDC, Statistical Services Branch (SSB) will provide the following data items for all episodes of care where there was a discharge from a public or private hospitals in Queensland from 1 January 2000 to 30 June 2021. Data for private hospitals only available from 1 July 2007 onward.
Data item/s to be given to researchers (e.g. name, date of birth, sex, antibiotics, histology, readmissions)	<ul style="list-style-type: none"> • Patient Study ID (created by SSB) • Facility ID (Public facilities only, deidentified ID numbers for private hospitals) • Principal diagnosis (ICD-10-AM – can include those relating to breast cancer or other conditions) • Other diagnosis (ICD-10-AM – All, list of specific comorbidities to be provided) • Procedure (ICD-10-AM – surgical procedure codes to be provided), All • Date of separation (month, year) • SA2 (2016 ASGS) of usual residence at time of separation • Age at admission (5 year age groups up to 85 years and over)
Date range (a specific start date and end date is required e.g. 01/01/2004 to 31/07/2014)	From 1 January 2000 (Public) and 1 July 2007 (private) to 30 June 2021
Frequency of request (e.g. once / six monthly)	Once
Data items to be used for linkage (if applicable)	<p>For the linkage process, the Statistical Services Branch will be provided with identifying data:</p> <p>Hospital UR number Hospital facility number full name date of birth sex address</p> <p>via secure electronic file transfer. The Statistical Services Branch will use deterministic and probabilistic linkage to link the datasets (B & C).</p>

6.2 Data custodian authorisation for the aforementioned data items (collectively, Information)

Data custodian declaration:

HREC approval number: 4884

I have considered this application and have seen all the relevant documents (including the HREC approval letter) that are required for this research and confirm that our unit is able to provide the data services indicated in this application within present resources.

In supplying the Information for an approved research project, no warranties are made as to: the fitness of the Information; the appropriateness or validity of the proposed research methods; or the appropriateness of the purpose for which the Information is being given.

Repository name	Queensland Hospital Admitted Patient Data Collection (Dataset C)		
Data custodian to list any special conditions or limitations that may be applied as a condition by the decision-maker (e.g. cost per record or whether authorization is dependent on availability of staff at the time of data request):			
Data Custodian Name	Ms Ann Harrington		
Position			
Unit	Statistical Services Branch		
Hospital / Facility	Queensland Health		
Signature of data custodian		Date	

Signature of data custodian must be no more than twelve (12) months old at time of submission to the Director-General's delegate.

Data source	BreastScreen Queensland (Dataset D)
Cohort Scope (e.g. inclusion and exclusion criteria such as age group, diagnoses, hospital admission date range, geographic area, any control group/s)	All women aged ≥ 20 years who were diagnosed with breast cancer (identified from the Queensland Cancer Register) in Queensland between 1 January 2000 and 30 June 2019. For the women identified in this breast cancer cohort, we will require data on all instances of their breast screening since 1997 regardless of age at screening
Data Scope (e.g. descriptors such as demographics, diagnoses, clinical factors, pathology, radiology, therapy, outcomes)	All BreastScreen Queensland records with a QCR number completed from 1 January 1997 to 30 June 2021
Data item/s to be given to researchers (e.g. name, date of birth, sex, antibiotics, histology, readmissions)	<ul style="list-style-type: none"> • Unique person ID number (specific to this project) • Age at Breast Screen episode (whole years) • Month and year of Breast Screen episode • Age at death (if relevant, in whole years) • Indigenous status • Non-English-speaking background (Yes/No) • Episode number • Screening outcome • Assessment outcome (where applicable) • Surgical Histopathology data (Invasive cancer/DCIS/Other/Nodes sampled-involved/Sentinel Node Biopsy) for the "Dominant Lesion" • Primary treatment data (type of surgery, other treatments, location of surgical unit, evidence of metastases) • SA3 of residence at time of screening episode
Date range (a specific start date and end date is required e.g. 01/01/2004 to 31/07/2014)	1 January 1997 to 30 June 2021.
Frequency of request (e.g. once / six monthly)	Once
Data items to be used for linkage (if applicable)	For the linkage process, BreastScreen Queensland already has a Queensland Cancer Register patient identification number on their database that provides a direct link with the Queensland Cancer Register. This will be provided to the Statistical Services Branch via secure electronic file transfer. The Statistical Services Branch will use deterministic linkage to link the two datasets (B and D).

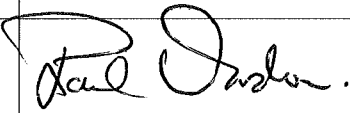
6.2 Data custodian authorisation for the aforementioned data items (collectively, Information)

Data custodian declaration:

Part 4, Public Health Act 2005 Application form, version: September 2021
Urban Health Indicators - Breast Cancer

I have considered this application and have seen all the relevant documents (including the HREC approval letter) that are required for this research and confirm that our unit is able to provide the data services indicated in this application within present resources.

In supplying the Information for an approved research project, no warranties are made as to: the fitness of the Information; the appropriateness or validity of the proposed research methods; or the appropriateness of the purpose for which the Information is being given.

Repository name	BreastScreen Queensland (Dataset D)		
Data custodian to list any special conditions or limitations that may be applied as a condition by the decision-maker (e.g. cost per record or whether authorization is dependent on availability of staff at the time of data request):			
Data Custodian Name	Paul Vardon		
Position	Acting Executive Director, Preventive Health Branch		
Unit	Cancer Screening Unit		
Hospital / Facility	Queensland Health		
Signature of data custodian		Date	14/6/2022

Signature of data custodian must be no more than twelve (12) months old at time of submission to the Director-General's delegate.

7. Data Linkage

7.1 Is data linkage required for this project?

☒ Yes ☐ No

7.2 Have you consulted with the data linkage service?

☒ Yes ☐ No

7.3 Who is linking the data?

(If Queensland Health please name the department/branch undertaking the linkage)

Statistical Analysis and Linkage, Statistical Services Branch, Queensland Health

8. Privacy and confidentiality

8.1 In what form will the Information be disclosed to you?

☒ Electronic ☐ Paper ☐ Both paper and electronic

8.2 Who is giving the information to you (including the name of the relevant person/s giving the information and to whom the information is given).

Dataset A: Deidentified data will be provided by secure electronic transmission by Mr Nick Ormiston-Smith (or authorised delegate) to the Coordinating Principal Investigator. The data will only be made available to researchers named on the ethics application.

Linked Dataset (based on Datasets B, C & D): Deidentified data will be provided by secure electronic transmission by the Statistical Services Branch) to the Coordinating Principal Investigator. The data will only be made available to researchers named on the ethics application.

8.3 How will the privacy of any individual that may be capable of being identified from the Information or its use in combination with other information be protected?

There will be about 680,000 women included in the BreastScreen Queensland cohort (Dataset A) for the period 2015-2021, and about 75,000 women from the Queensland Cancer Register diagnosed with breast cancer between 2000 and 2019 (Dataset B). It is not known how many records from the QHAPDC (Dataset C) and BreastScreen Queensland (Dataset D) will be linked to the approximately 75,000 women in the breast cancer cohort. In Queensland there are 528 SA2 areas covering Queensland. Although the distribution of women per SA2 will vary widely, this means, on average, approximately 1700 women (Dataset A) and 142 women (Dataset B) would be in each SA2. Numbers would be smaller when considering year-specific or age group-specific estimates. The release and reporting of statistical data by data custodians and researchers needs to consider two key issues: data privacy and statistical stability. Data privacy relates to the responsibility of data custodians to protect the identity of individuals in their data, and ensure that this is not compromised by the release of that data for reporting purposes. Statistical stability relates to the inherent random fluctuation of statistics based on small numbers of cases; the smaller the numbers, the more they fluctuate, potentially leading to incorrect interpretation. These issues are particularly relevant when considering geographical data. To address both these issues for geographical data, we will use a specific statistical method known as "spatial smoothing". While standard methods typically only adjust for age and sex in each area, spatial smoothing recognises the geographical structure of the data and includes data from the neighbouring geographical areas when calculating the spatial estimates. This additional data provides greater stability to the estimates. In addition, because the spatial estimates are modelled, rather than observed, spatial smoothing reduces any risk of identifiability for specific individuals. These protective effects of spatial smoothing are most pronounced in areas where it is needed the most, that is, those with the smallest numbers of cases. Smoothed estimates are designed to reflect the real differences in the underlying rate or risk between areas.

In protecting the privacy of individuals, the following points are also relevant to our research proposal:

- (a) All patients will be identified by a project specific identification number;
- (b) There will be no contact with members of the study cohort;
- (c) The research team is not involved in clinical work, therefore the risk of patient identification is very low;
- (d) Names and addresses are not included in our datasets. We are not requesting any potentially identifiable fields such as full admission and discharge dates (we are only requesting MMY);
- (e) We have a highly experienced team of researchers covering various disciplines who are aware of issues of confidentiality and privacy through our numerous similar studies that have addressed these issues.
- (f) Data will be stored on the secure network at Cancer Council Queensland. Access to the study datasets is restricted to named research team members who are staff at Cancer Council Queensland. All Cancer Council Queensland employees are subject to confidentiality requirements. All research conducted at Cancer Council Queensland conforms to national standards of conduct (Australian Code for the Responsible Conduct of Research) to ensure ethical, honest, well-documented and meticulous research.
- (g) All publications and presentations arising from this project will not contain any identifying information, and no individual, medical practice or hospital will be identified or identifiable in such material.
- (h) Upon completion of the analyses and the study, the datasets will be archived in the same network location and with the same security policy and access as was used during the study.
- (i) The data will be used in a manner consistent with its use by Government agencies when they report on patient health outcomes. However, our team has the technical

expertise for using advanced statistical methods and providing insights beyond what is available in Government reports. The results of these enhanced analyses will be published in national or international peer reviewed journals, reported at scientific conferences, and be translated into lay language, with assistance from our consumer representatives, to be disseminated through the websites of the collaborating organisations.

- (j) When disseminating results of the study, researchers will suppress small cell counts (e.g. $n < 5$) to protect participant confidentiality.

Note that the project specific identification number (see (a) above) will be a random code with a unique value for each person. A link between the project specific identification number and the Study IDs created by SSB will be retained by SSB in case of future data validation questions sent to SSB by the research team. The research team will have no access to this link.

8.4 How will Information security be maintained, throughout its entire research pathway?

(e.g. where will the information be transferred to (including name of institutions and physical locations) and by whom, the method/software that will be used to transfer the information, how it will be shared and with whom)

Dataset A and the linked dataset (ie. Dataset B linked with Datasets C and D) will be securely transferred from Queensland Health to the research team using Kiteworks software.

All computer files containing original data used in the study will be kept in password-protected computer networks at the Cancer Council Queensland accessible only by the named investigators. These staff members have individual passwords, and computers screens will be locked upon an individual leaving their workstation or after five minutes of inactivity. The Cancer Council Queensland computing network is protected by global virus protection and a firewall, together with regular Microsoft Windows updates.

All computers used for data storage and analysis are inside offices within the Cancer Council Queensland which are secured by swipe card access.

All unit record data obtained for this project will be securely destroyed 7 years after publication of the final output. This is in accordance with the NHMRC's guidelines on data retention in their Australian code for the responsible conduct of research, which recommends a minimum of five years from the date of publication. Once the 7 years has elapsed, all original unit record data files will be deleted from the secure network and any backup drives containing those original files. Information generated as part of this project (aggregated results of statistical analyses) will be stored indefinitely.

All approved applicants will view/analyse the data while physically located in Australia.

9. Checklist

This checklist is provided to help in submitting your application (all boxes require a tick before submission).

<input checked="" type="checkbox"/>	Does the PHA application title match the HREC approval letter title? (Section 2)
<input checked="" type="checkbox"/>	Have you included the HREC reference number, expiry date and attached a copy of the HREC approval letter and all HREC amendment approvals? (Section 3)
<input checked="" type="checkbox"/>	Have you listed all people who will be given the requested Information and their institution/s? (Section 3)
<input checked="" type="checkbox"/>	Have you answered all sections clearly and provided documents to support your application?
<input checked="" type="checkbox"/>	Have you listed all data repositories, data items, and data dates required? (Section 6)
<input checked="" type="checkbox"/>	Have you specified how privacy, confidentiality and Information security will be maintained? (Section 8)
<input checked="" type="checkbox"/>	Do you have relevant data custodian authorization for all Information required? (Section 6)
	If all the above boxes are ticked, please proceed to Section 10 and sign the undertaking of confidentiality

10. Coordinating Principal Investigator / Principal Investigator undertaking of confidentiality

In the course of being, as relevant, given, holding, using or further disclosing Information for research purposes in relation to this application, if granted, I acknowledge that:

- I will be given Information which, if inappropriately dealt with or managed or insufficiently secured may cause loss or damage to individuals, public or private facilities or communities.
- I will not disclose Information in any publication whatsoever (including, for example, any report or communication disclosed to persons or entities other than those listed in a grant of this application).
- I will not use Information for any purpose other than for performing the specific activities detailed in my application if granted.
- I will ensure that the Information is not disclosed to any person or entity other than those listed in a grant of this application and is kept confidential.
- I will ensure that the Information is securely stored such that it cannot be accessed by any person or entity other than those authorised to be given the Information as listed in a grant of this application, including in relation to disposing of it securely.
- I agree to submit a further application to amend this application, if granted, or a new application (as applicable) if:
 - I seek additional information (this includes but is not limited to, additional data from the same data sets, new data sets and new sites)
 - I want to extend the time period for the grant of this application, if granted
 - another researcher not listed in a grant of this application, if granted, seeks to be given Information for the purposes of the research listed in the grant of application, if granted.

In signing this undertaking, I will take all steps necessary to ensure that all researchers, including myself, involved in the research given Information described in this application, if granted, will adhere to the obligations specified above and the conditions set out in a grant

HREC approval number: 4884

of this application, if granted. I believe that the public interest benefits of this research outweigh the public interest detriments, including maintaining confidentiality and privacy of patients who may be identified by the Information, and will provide for increased knowledge and improved health outcomes (section 284(2)(a)(i)).

I also acknowledge that if this application is granted by the Director-General of Queensland Health as chief executive under the Public Health Act 2005, or delegate, the research project must satisfy a Research Governance review process prior to commencement of the research.

Signature



Coordinating Principal Investigator /
Principal Investigator Name (please print)

Professor Peter Baade

Date

27th April 2022

Please note. This section cannot be signed by any person other than the person nominated on the section 3.

References

- Baade P, Dasgupta P, Aitken J, Turrell G. Geographic remoteness, area-level socioeconomic disadvantage, and colorectal cancer survival: a cross-sectional multilevel study. *BMC Cancer*. 2013 Oct 24;13(1):493.
- Banerjee, S., Carlin, B.P., Gelfand, A.E. (2004) Hierarchical Modeling and Analysis for Spatial Data. Chapman and Hall.
- Banerjee, S., Wall, M.M., Carlin, B.P. (2003) Frailty modelling for spatially-correlated survival data, with application to infant mortality in Minnesota. *Biostatistics* 4, 123-142.
- Bentley R, Kavanagh A, Subramanian SV, Turrell G. Area disadvantage, individual socioeconomic position, and premature cancer mortality in Australia 1998-2000: a multilevel analysis. *Cancer Causes and Control* 2008; 19:183-193.
- Bernardinelli, L., Clayton, D.G., Pascutto, C., Montomoli, C., Ghislandi M., Songini, M. (1995) Bayesian analysis of space-time variation in disease risk. *Statistics in Medicine* 14, 2433-2443.
- Besag, J., York, J., Mollie, A. (1991) Bayesian image restoration with two applications in spatial statistics. *Annals of the Institute of Statistical Mathematics* 43, 1-59.
- Cramb SM, Mengersen KL, Baade PD (2011b). Atlas of Cancer in Queensland: geographical variation in incidence and survival, 1998 to 2007. Viertel Centre for Research in Cancer Control, Cancer Council Queensland. Brisbane, Queensland
- Cramb, S. M., K. L. Mengersen and P. D. Baade (2011a). Developing the atlas of cancer in Queensland: methodological issues. *International Journal of Health Geographics* 10: 9.
- Dasgupta P, Cramb SM, Aitken JF, Turrell G, Baade PD. Comparing multilevel and Bayesian spatial random effects survival models to assess geographical inequalities in colorectal cancer survival: A case study. *International Journal of Health Geographics*. 2014 Oct 4;13:36
- Duncan E, Cramb S, Baade P, Mengersen K, Saunders T, Aitken J. (2020) Developing a Cancer Atlas using Bayesian Methods: A Practical Guide for Application and Interpretation. <https://atlas.cancer.org.au/developing-a-cancer-atlas/>
- Gomez-Rubio V, Moraga P, Molitor J, Rowlingson B. DCluster: Model-based detection of disease clusters. R package version 1.0-1. <https://CRAN.R-project.org/package=DCluster>. 2018.
- Hsieh JC, Cramb SM, McGee JM, Dunn NAM, Baade PD, Mengersen KL. Spatially varying coefficient inequalities: Evaluating how the impact of patient characteristics on breast cancer survival varies by location. *PLoS One*. 2016 May 5;11(5)
- Hsieh JC, Cramb SM, McGee JM, Dunn NAM, Baade PD, Mengersen KL. Geographic variation in the intended choice of adjuvant treatments for women diagnosed with screen-detected breast cancer in Queensland. *BMC Public Health* 2015, 15:1204
- Jackson, C., Best, N., Richardson, S. (2008) Hierarchical related regression for combining aggregate and individual data in studies of socio-economic disease risk factors. *Journal of the Royal Statistical Society A*. 171, 159-178.
- Kavanagh AM, Turrell G, Subramanian SV. Does area-based social capital matter for the health of Australians? A multilevel analysis of self-rated health in Tasmania. *International Journal of Epidemiology* 2006; 35:607-613.
- Knorr-Held, L. (2000) Bayesian modelling of inseparable space-time variation in disease risk. *Statistics in Medicine* 19, 2555-2567.
- Lawson, A.B. and Densson, , D.G.T. (2002) Spatial cluster modelling. Chapman and Hall.

- Lawson, A.B., Browne, W.J., Vidal Rodeiro, C.L. (2003) Disease Mapping with WinBUGS and MLwiN. Wiley.
- Tango, T. (2000). A test for spatial disease clustering adjusted for multiple testing. *Statistics in Medicine* 19 (2): 191-204.
- Turrell G, Giskes K. Socioeconomic disadvantage and the purchase of takeaway food: a multilevel analysis. *Appetite* 2008; 51:69-81.
- Turrell G, Kavanagh A, Draper G, Subramanian SV. Do places affect the probability of death in Australia? A multilevel study of area-level disadvantage, individual-level socioeconomic position, and all-cause mortality, 1998-2000. *Journal of Epidemiology and Community Health* 2007; 61:13-19.
- Turrell G, Sanders AE, Slade GD, Marceles W, Spencer AJ. The independent contribution of neighbourhood disadvantage and individual-level socioeconomic position to self-reported oral health: a multilevel analysis. *Community Dentistry and Oral Epidemiology* 2007; 61:13-19.
- Utazi CE, Afuecheta EO, Nnanatu CC. A Bayesian latent process spatiotemporal regression model for areal count data. *Spatial and Spatio-Temporal Epidemiology*. 2018;25:25-37.