STUDY PROTOCOL

Queensland Cardiovascular Data Linkage (QCard): A population-based cohort study [version 1; peer review: awaiting peer review]

Joshua Byrnes, Son Nghiem, Clifford Afoakwah, Paul A. Scuffham
Centre for Applied Health Economics, Griffith University, Brisbane, Queensland, 4111, Australia

Abstract
Queensland is Australia's hotspot for cardiovascular disease (CVD). Critically, beyond modifiable lifestyle risk factors; socio-demographic differences and environmental factors account for significant variations in healthcare use and outcomes among cardiac patients across the country. To better understand the impacts of these factors on the health of cardiac patients, there is a need for a comprehensive and robust longitudinal cohort study that can unpack the underlying dynamics. This paper describes the protocol for the Queensland Cardiovascular Linkage (QCard) Study. The QCard is a longitudinal linkage cohort study of cardiac patients who were first hospitalised with any cardiac condition in 2010, with follow up hospitalisations until December 2015. The primary aim of the QCard is to identify and characterise the nature and impact of socio-demographic inequality among those presenting for the first time with the most common form of CVD in Australia (heart disease) in Queensland from 2010 with minimum 5-years follow-up of subsequent healthcare utilisation and outcomes. A secondary aim is to undertake an exploration of the impact of environmental and specific health service factors on healthcare use and survival time in the same QCard cohort. Administrative public and private hospital inpatient, outpatient and emergency department data for all of Queensland will be linked with individual primary care data and pharmaceutical data. These data will also be linked to regional socio-demographic data and environmental data, as well as data that describes the features of each hospital in the region. The findings from the study will provide critical information for cardiac patients, clinicians and health policymakers. Such information ranges from identifying most vulnerable cardiac patients who may require targeted needs to providing estimates for cost-effective ways of evaluating healthcare interventions that seek to improve the health of cardiac patients.

Keywords
Healthcare use, health outcomes, inequalities, cost, health disparities, heart disease
Corresponding author: Son Nghiem (s.nghiem@griffith.edu.au)

Author roles: Byrnes J: Conceptualization, Investigation, Methodology, Supervision, Writing – Original Draft Preparation, Writing – Review & Editing; Nghiem S: Conceptualization, Data Curation, Formal Analysis, Investigation, Methodology, Writing – Original Draft Preparation, Writing – Review & Editing; Afoakwah C: Conceptualization, Formal Analysis, Methodology, Writing – Original Draft Preparation, Writing – Review & Editing; Scuffham PA: Conceptualization, Data Curation, Funding Acquisition, Investigation, Methodology, Supervision, Validation, Writing – Original Draft Preparation, Writing – Review & Editing

Competing interests: No competing interests were disclosed.

Grant information: PS was assigned the following National Health and Medical Research Council grants [NHMRC1044897, NHMRC1055214, NHMRC1136923].

The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Copyright: © 2020 Byrnes J et al. This is an open access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

How to cite this article: Byrnes J, Nghiem S, Afoakwah C and Scuffham PA. Queensland Cardiovascular Data Linkage (QCard): A population-based cohort study [version 1; peer review: awaiting peer review] F1000Research 2020, 9:282
https://doi.org/10.12688/f1000research.23261.1

Introduction
Cardiovascular diseases (CVDs) remain a global health concern due to the number of deaths attributable to the disease. In Australia, particularly Queensland, the prevalence and incidence of CVD continue to impose a tremendous burden on the health system. Queensland, with 8 of the nation’s top 20 regions for death due to heart disease, was identified by the National Heart Foundation of Australia (NHFA) as Australia’s hotspot for cardiovascular disease (CVD) – a national health priority area. Queensland has the second-highest incidence rate of CVD after the Northern Territory. The incidence of acute coronary events (heart attacks) was estimated to be 20% higher than the national average in 2014, and one-tenth of all hospital expenditure was related to CVD (Queensland Health, 2016). Also, 29% of all total deaths are CVD related, which was 5% higher than the national rate in 2014. Coronary Heart Disease (CHD) and stroke were 9% and 8% higher than the national death rates in 2014, respectively.

Despite the common modifiable risk factors (such as lifestyle) associated with CVD, socio-demographic differences, environmental factors and hospital characteristics account for significant variations in healthcare use and outcomes among cardiac patients. Socio-demographic inequality that drives patient ranges from age, gender to ethnicity differences contribute to substantial variations in cardiac health within a population.

The role of the environment on cardiac health can also be tremendous. Environmental factors such as ambient temperatures, air pollution, and macroeconomic shocks drive healthcare use and outcomes among cardiac patients. Socio-demographic inequality that drives patient ranges from age, gender to ethnicity differences. Also, micro-economic factors such as education, wealth and remoteness of usual residence contribute to substantial variations in cardiac health within a population.

Finally, hospital characteristics, the availability of cardiac services, access to evidence-based cardiac care, community healthcare and variation in clinical practice also play an important role in determining the health of cardiac patients. Cardiac services such as cardiac surgery, telehealth, exercise stress test, and Holter monitoring varies substantially across hospitals, which ultimately predict variations in health outcomes. The variations in health services across hospitals in Australia and its leading hospital outcomes are often reflected in how hospitals are classified.

To better understand the impact of systematic socio-demographic inequality, environmental factors, and hospital characteristics on healthcare use and outcomes among cardiac patients, there is the need for robust longitudinal cohort patient-level data that can characterise patients’ outcomes over time. This paper describes the QCard data, which is a longitudinal patient-level cohort linkage data of cardiac patients from Queensland, Australia, and can be merged with numerous other data sets pertaining, for example to, environmental factors. Figure 1 describes a model of how the QCard data will be used to explore the impact of individual unmodifiable risk factors, modifiable factors of healthcare delivery and the impact of wider environmental effects on healthcare use and outcome among cardiac patient post-first CVD hospitalisation.

The QCard study has two broad aims. The primary aim of the QCard is to identify and characterise the nature and impact of socio-demographic inequality among those presenting for the first time with the most common form of CVD in Australia (heart disease) in all regions, from metropolitan to remote, in Queensland and their subsequent healthcare utilisation and outcomes (e.g., survival, readmission, complications). A secondary aim is to undertake an exploration of the impact of environmental and specific health service factors (e.g., air pollution, temperature, policies) on the healthcare use and survival time in the same QCard cohort.

Protocol
Study design
QCard is a longitudinal patient-level cohort linkage study of all CVD hospitalisations in 2010 with subsequent hospitalisations over a six-year period in Queensland, Australia. That is, the QCard will utilise secondary data that encapsulates all primary, secondary and tertiary healthcare used by all cardiac patients who were first hospitalised in 2010 and follow up utilisation until December 2015.

Study setting and population
Queensland is the second-largest in the geographical area and the third most populated state in Australia and has an area of 665,615 square miles ranging from -10 to -30 latitude. The QCard data will include all patients who were admitted to hospitals with a diagnosis of CVD in 2010 with subsequent admissions until December 2015 in Queensland (QLD), Australia. The cohort will represent all CVD admissions in QLD during the study period. We will use ICD-10-AM (International Statistical Classification of Diseases and Related Health Problems, 10th Revision, Australian Modification) codes from the admission data to define cardiovascular disease as follows: All cardiovascular diseases: I00-I99, Acute Coronary Syndrome (ACS): I200, I210-I214, Unstable Angina: I200, ST-Elevation myocardial infarction: I210-I213, Non-ST Elevation myocardial infarction: I214, Chronic Heart Syndrome (CHS): I20-I25, Heart Failure: I50, Cerebrovascular Disease (CeVD): I60-I69, and Chronic rheumatic heart disease: I05-I09. The study is expected to have a sample of 200,000 first-ever cardiac hospitalisations, with a survival rate of approximately 71%.

Inclusion criteria: Any CVD-related inpatient admission to a public health facility in 2010 for people aged 18 and above.
Exclusion criteria: patients aged under 18 or admitted to private health facilities as private patients. Non-admitted patients were also excluded – that is those attending outpatients or the emergency department who were not admitted. In addition, participants were excluded if they had a CVD-related inpatient hospitalisation before 2010.

Ethics
Ethical approval for the QCard study was obtained from the Human Research Ethics Committee (Ref No: 2017/001) by Griffith University and Queensland Health. Data will be stored and accessed using the Secure Unified Research Environment. Individual patient consent was waived for this study as it involved no direct contact with, or identification of, patients. Access to data was authorised under the Public Health Act of Queensland. Only un-identifiable, and non-reidentifiable, data were obtained.

Data sources
The data required for the study were submitted to data custodians in each of the 16 Hospital and Health Services in Queensland for approval. After receiving approval from the data custodians, the data were extracted and provided by the Statistical Unit of Queensland Health. The cohort data will be extracted from the Queensland Hospital Admitted Patient Data Collection (QHAPDC), Emergency Department Data Collection (EDDC), Registrar General Deaths Database (RGGD), comorbidity, and costs database. The cohort will also be linked with national data on the Medicare Benefits Schedule (MBS) and the Pharmaceutical Benefits Scheme (PBS) from the Australian Institute of Health and Welfare (AIHW). These data sets will contain information about types of medication and health services used, total expenditures and out-of-pocket contribution by patients for services, imaging and diagnostic procedures and prescription medicines received in the community. Table 1 summarises the data sources the expected variables available in them.

Environmental data
External data on environmental factors will be linked using residential postcodes of patients. All study participants have their postcodes recorded in the data; hence, we will be able to identify inequality at all geographical levels including the 78 local government areas (LGAs) in Queensland, Statistical Area Level 3 (SA3), Statistical Area Level 4 (SA4), State Suburb (SSC) and Postal Area (POA) levels. Specifically, environmental factors on air pollution will be sourced from Queensland Government Department of Environment and Science (QGDES), ambient temperature from Australian Bureau of Meteorology (BoM), and macroeconomic factors and socio-economic status (SES) from Australian Bureau of Statistics (ABS). These data will be used to investigate the impact of environmental shocks and conditions on healthcare use and survival time of cardiac patients. Effects of macroeconomic fluctuations and policy changes on health services usages and outcomes are also investigated.

Hospital characteristics
Data on hospital characteristics such as type of hospital (public/private), number of beds, locations and types of funding (block vs activity-based), peer group codes (acute, sub-acute and non-acute) and remoteness index will be collected. Such data will be sourced from Hospital and Health Services (HHSs), Independent Hospital Pricing Authority (IHPA) and Australian Institute of Health and Welfare (AIHW) and will cover 2010 and then annually thereafter.

Data linkage
Data sets will be linked using the individual identifier code, generated by the AIHW. The episode identifier code, generated by the Queensland Health Statistical Analysis Unit,
and date of services will also be used to match data sets. In particular, individual and episode identifier codes will be used to link admissions data with the cost and sub-acute and non-acute data sets (see Figure 2). Data from the death registry were linked using only the individual identifier. Admission data will be linked with the morbidity data using the individual and episode identifiers.

The selected data sets will also be linked with the MBS and PBS data. The MBS has records of federally subsidised health services accessible to all eligible Australians. The PBS has records of federally subsidised prescription medications accessible to all eligible Australians. The MBS and PBS data contain detailed information of MBS schedule fee, MBS fee charged, MBS benefits paid by the Commonwealth government, PBS patient contribution, PBS benefit by the Commonwealth government as well as the quantities of drugs dispensed.

We will also link external data, including socio-economic index, macroeconomic indicators and meteorological information to the main data set using residential postcodes.

Finally, postcode-level data on socio-economic and environmental variables will be linked with the main data using the individual postcodes and time (month and year) of admission or separation. Hospital level variables will also be merged unto the admission data using hospital names. These external data sets will not be at the individual levels but based on place of usual residence, hospitals of visits, and the time of hospital event.

To protect the privacy of patients, an individual probabilistic identification code, generated by the AIHW, will be used. The data will be stored and analysed using the Secure Unified Research Environment (SURE). Only authorised researchers can access the data remotely using password and security tokens with

<table>
<thead>
<tr>
<th>Table 1. Main data sources and variables.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Admission</strong></td>
</tr>
<tr>
<td>Facility names</td>
</tr>
<tr>
<td>Age of patients</td>
</tr>
<tr>
<td>Sex of patients</td>
</tr>
<tr>
<td>Regions of origin</td>
</tr>
<tr>
<td>Marital status</td>
</tr>
<tr>
<td>Indigenous status</td>
</tr>
<tr>
<td>Residential status</td>
</tr>
<tr>
<td>State of residence</td>
</tr>
<tr>
<td>Statistical local area</td>
</tr>
<tr>
<td>Statistical area level 2</td>
</tr>
<tr>
<td>Occupation</td>
</tr>
<tr>
<td>Medicare eligibility</td>
</tr>
<tr>
<td>Hospital insurance cover</td>
</tr>
<tr>
<td>Admission/separation date</td>
</tr>
<tr>
<td>Length of stay at ICU</td>
</tr>
<tr>
<td>Original referral code</td>
</tr>
<tr>
<td>Separation mode</td>
</tr>
<tr>
<td>Usual accommodation</td>
</tr>
<tr>
<td>Employment status</td>
</tr>
<tr>
<td>Pensions status</td>
</tr>
<tr>
<td>Referral for further care</td>
</tr>
<tr>
<td>Diagnosis-related group</td>
</tr>
<tr>
<td><strong>Sub-acute non-acute (SNAP)</strong></td>
</tr>
<tr>
<td>SNAP types/summary</td>
</tr>
<tr>
<td>SNAP admission/separation date</td>
</tr>
<tr>
<td>Multidisciplinary care plan</td>
</tr>
</tbody>
</table>

Note: ICU=Intensive Care Unit; SNAP=Subacute and Non-acute Assessment and Pathways; ICD = International Classification of Disease; MBS = Medicare Benefits Schedule; PBS = Pharmaceutical Benefits Scheme.
time-synchronised one-time passwords. No data will be exported outside the SURE. Also, results of data analyses will be screened by the SURE personnel to ensure no identifiable information is included before exporting to an external gateway.

**Main outcome measures**

The outcomes that will be investigated in this study are put into primary and secondary outcomes: The primary outcomes are mortality, time-to-death, readmission and days alive and out of the hospital. Mortality in the study will be measured as deaths after the first CVD hospitalisation. This outcome will be traced from the RGDD database. Using the QHAPDC and RGDD data, we will be able to measure survival time by looking at when a patient was first hospitalised and the date of death. Alive and out of hospital will adjust the patient’s survival time for the number of days spent in the hospital.

The secondary outcomes will be healthcare utilisation across all care sectors. More specifically, indicators such as general practitioners (GP) visits, use of cardiac-specialised services (which will be traced from the Medical Benefits Scheme (MBS) data), emergency department and inpatient hospital admissions, in-hospital adverse events, and pharmaceutical use from the Pharmaceutical Benefits Scheme (PBS) data.

**Data analysis and statistical plan**

The study will adopt two analytical approaches. The first analytical strategy will be to explore the extent to which socio-economic and spatial inequality permeates among cardiac patients, as shown in Figure 1. Techniques as Bayesian Spatial Analysis (BSA) will be used to investigate the extent to which spatial variations across the 78 local government areas in Queensland associate with variations in health service utilisation and mortality among cardiac patients. The number of deaths, admissions and GP visits in each LGA-period-sex-age group will be specified using a Poisson model. The Poisson model can estimate LGA level deaths and use of health services relative to that expected if all LGAs have the same deaths and use of health service as the national rates. In addition to the BSA, the Maximum Likelihood Estimation (MLE) techniques such as the logit and probit model will be employed to investigate the impact of socio-demographic factors (such as gender, age, and ethnicity) and cardiac-specific conditions (such as CHD, HF, AF etc.) on healthcare use and outcomes among cardiac. These techniques will enable us to identify most the most vulnerable patients who may require targeted public policies.

The second analytical approach will seek to explore the impact of environmental factors on health outcomes. Particularly, environmental effects on the survival time of cardiac patients and will utilise parametric and semi-parametric survival models such as the Cox proportional hazard model, Weibull, Gompertz, and Exponential. By using both parametric and semi-parametric survival models, the study will be able to provide reliable estimates on the impact of environmental factors on survival time under different assumptions placed on the baseline hazards of cardiac patients. In addition, generalised linear models with relevant link function and distributional assumptions will be used to estimate the impact of hospital characteristics and environmental factors on other outcomes (e.g., readmissions) and healthcare use such as hospital admissions, GP visits, healthcare costs and drug use. Unobservable characteristics across patients will be mitigated.
using a random-effects approach, while a fixed-effect approach will be used to address unobserved characteristics at community or health service levels. These approaches will reduce potential biases by unobserved or missing variables (confounders).

Data will be analysed using STATA 15.0 or subsequent upgrades and R 3.6.1.

Dissemination of information
Findings of this study will be disseminated via publication in peer-review journals, conference presentations, policy briefings and public lectures.

Study status
At the time of writing, all health service data have been obtained and linked. A baseline paper that describe characteristics of the cohort is in preparation. Also environmental data on air pollution and ambient temperature have been obtained from the Queensland Government Department of Environment and Science (QGDES) and Australian Bureau of Meteorology (ABM) respectively and are being analysed. Collaborations with three other institutions in Australia have been developed to investigate a wide range of research topics from this cohort.

Discussion
QCard is a population-based linkage cohort study of cardiac patients in Queensland, Australia who was first hospitalised with any CVD condition in 2010, with follow up admissions until December 2015. The study will investigate cardiac outcome measures such as survival time and healthcare use, which reflect the burden of cardiovascular diseases to the patient, health system and society at large. A nuanced analysis of these outcomes will unpack any existing inequalities within the population as well as better our understanding on how public health policies can be strategised to offer cost-effective models of care for cardiac patients.

Although the QCard study is similar to other cohort-studies on CVD, it differs from these studies in some key elements. First, the QCard study includes different cardiac conditions ranging from mild conditions like varicose veins to serious an urgent condition like acute myocardial infarction. The advantage of having such heterogeneous cardiac data is that it allows for a comparative analysis of any prevailing inequality in healthcare use and outcomes across different conditions, as well as identifying which public health policies work best for which cardiac conditions. Second, this study will interrogate the impact of environmental factors on the health outcomes of cardiac patients. This approach will bring to bear the contemporaneous impact of environmental shocks on the health of cardiac patients, hence enabling cost-effective ways of modelling healthcare interventions for patients with CVD. Finally, and perhaps most importantly, the QCard study is unique regarding the context of patients being investigated. As mentioned earlier, CVD is a major health concern in Queensland regarding its incidence, prevalence and death tolls in the state. Therefore, a comprehensive and robust study on CVD in Queensland will provide a context-specific understanding of the disease and how policy can minimise its burden.

In conclusion, the QCard study is a comprehensive population-based study of cardiac patients in Queensland which will explore the socio-demographic inequalities in healthcare use and outcomes as well as the interacting impacts of environmental factors on health outcomes. The findings from the study will provide useful information for cardiac patients, clinicians and health policymakers. Such information ranges from identifying most vulnerable cardiac patients to providing estimates for cost-effective ways of planning healthcare interventions that can improve the health of cardiac patients.

Data availability
Underlying data
No data are associated with this article.

References

outdoor air quality on childhood asthma. J Health Econ. 2004; 23(6): 1209–36.


The benefits of publishing with F1000Research:

- Your article is published within days, with no editorial bias
- You can publish traditional articles, null/negative results, case reports, data notes and more
- The peer review process is transparent and collaborative
- Your article is indexed in PubMed after passing peer review
- Dedicated customer support at every stage

For pre-submission enquiries, contact research@f1000.com