

Protocol

# The Australian Genetic Heart Disease Registry: Protocol for a Data Linkage Study

Alexandra Butters<sup>1,2,3</sup>, MPH; Bianca Blanch<sup>4</sup>, PhD; Anna Kemp-Casey<sup>5</sup>, PhD; Judy Do<sup>1,2</sup>, BPsych; Laura Yeates<sup>1,2,3,4,6</sup>, Grad Dip (GenCouns), BSc; Felicity Leslie<sup>1,2</sup>, BAppSci; Christopher Semsarian<sup>3,4,6</sup>, PhD; Lee Nedkoff<sup>7,8</sup>, PhD; Tom Briffa<sup>7</sup>, PhD; Jodie Ingles<sup>1,2,6</sup>, PhD; Joanna Sweeting<sup>1,2</sup>, PhD

<sup>1</sup>Clinical Genomics Laboratory, Centre for Population Genomics, Garvan Institute of Medical Research, Darlinghurst, Australia

<sup>2</sup>Clinical Genomics Laboratory, Centre for Population Genomics, Murdoch Children's Research Institute, Melbourne, Australia

<sup>3</sup>Faculty of Medicine and Health, The University of Sydney, Sydney, Australia

<sup>4</sup>Agnes Ginges Centre for Molecular Cardiology, Centenary Institute, The University of Sydney, Sydney, Australia

<sup>5</sup>Clinical and Health Sciences, University of South Australia, Adelaide, Australia

<sup>6</sup>Department of Cardiology, Royal Prince Alfred Hospital, Sydney, Australia

<sup>7</sup>School of Population and Global Health, The University of Western Australia, Perth, Australia

<sup>8</sup>Victor Change Cardiac Research Institute, Sydney, Australia

**Corresponding Author:**

Joanna Sweeting, PhD

Clinical Genomics Laboratory

Centre for Population Genomics

Garvan Institute of Medical Research

384 Victoria Street

Darlinghurst, 2010

Australia

Phone: 61 2 9359 8049

Email: [joanna.sweeting@populationgenomics.org.au](mailto:joanna.sweeting@populationgenomics.org.au)

## Abstract

**Background:** Genetic heart diseases such as hypertrophic cardiomyopathy can cause significant morbidity and mortality, ranging from syncope, chest pain, and palpitations to heart failure and sudden cardiac death. These diseases are inherited in an autosomal dominant fashion, meaning family members of affected individuals have a 1 in 2 chance of also inheriting the disease ("at-risk relatives"). The health care use patterns of individuals with a genetic heart disease, including emergency department presentations and hospital admissions, are poorly understood. By linking genetic heart disease registry data to routinely collected health data, we aim to provide a more comprehensive clinical data set to examine the burden of disease on individuals, families, and health care systems.

**Objective:** The objective of this study is to link the Australian Genetic Heart Disease (AGHD) Registry with routinely collected whole-population health data sets to investigate the health care use of individuals with a genetic heart disease and their at-risk relatives. This linked data set will allow for the investigation of differences in outcomes and health care use due to disease, sex, socioeconomic status, and other factors.

**Methods:** The AGHD Registry is a nationwide data set that began in 2007 and aims to recruit individuals with a genetic heart disease and their family members. In this study, demographic, clinical, and genetic data (available from 2007 to 2019) for AGHD Registry participants and at-risk relatives residing in New South Wales (NSW), Australia, were linked to routinely collected health data. These data included NSW-based data sets covering hospitalizations (2001-2019), emergency department presentations (2005-2019), and both state-wide and national mortality registries (2007-2019). The linkage was performed by the Centre for Health Record Linkage. Investigations stratifying by diagnosis, age, sex, socioeconomic status, and gene status will be undertaken and reported using descriptive statistics.

**Results:** NSW AGHD Registry participants were linked to routinely collected health data sets using probabilistic matching (November 2019). Of 1720 AGHD Registry participants, 1384 had linkages with 11,610 hospital records, 7032 emergency department records, and 60 death records. Data assessment and harmonization were performed, and descriptive data analysis is underway.

**Conclusions:** We intend to provide insights into the health care use patterns of individuals with a genetic heart disease and their at-risk relatives, including frequency of hospital admissions and differences due to factors such as disease, sex, and socioeconomic status. Identifying disparities and potential barriers to care may highlight specific health care needs (eg, between sexes) and factors impacting health care access and use.

**International Registered Report Identifier (IRRID):** DERR1-10.2196/48636

(*JMIR Res Protoc* 2023;12:e48636) doi: [10.2196/48636](https://doi.org/10.2196/48636)

## KEYWORDS

data linkage; genetic heart diseases; health care use; cardiomyopathies; arrhythmia; cardiology; heart; genetics; registry; registries; risk; mortality; national; big data; harmonization; probabilistic matching

## Introduction

Genetic heart diseases are a significant cause of cardiac morbidity and mortality in the youth and are characterized by distinct variability in the onset and severity of symptoms [1]. The most prevalent is hypertrophic cardiomyopathy (HCM), which affects up to 1 in 500 people [2]. They are generally inherited as autosomal dominant traits, conferring a 50% risk to first-degree relatives. A diagnosis relies on clinical investigations with a cardiologist and, in some cases, genetic testing, which is ideally conducted in a multidisciplinary clinic [3].

Disease symptoms can be variable ranging from asymptomatic to episodes of unexplained syncope, palpitations, and chest pain to the most serious outcomes of heart failure and sudden cardiac death (SCD). Accordingly, some individuals with a genetic heart disease require life-saving medical intervention via emergency department presentation and hospitalization. Treatment options may include prescription medication, cardioversion, and surgeries such as implantable cardioverter-defibrillator insertion, pacemaker insertion, septal reduction therapies (eg, alcohol septal ablation and septal myectomy), or heart transplant. The implantable cardioverter defibrillator is the only proven therapy to prevent SCD in individuals with a genetic heart disease [4]. The frequency with which these individuals interact with the health care system is unknown, including for noncardiac reasons [5].

Routinely collected health data are data collected for purposes other than research or without specific a priori research questions developed before collection [6]. These data provide both population- and person-based information, meaning we can examine common patterns of health care for the population [7,8]. In jurisdictions with universal health care such as Australia, we can observe an individual's interaction with hospitals and emergency departments over time and link them to mortality data. Studies that have used routinely collected data to investigate genetic heart diseases, or their consequences such as SCD, focus predominantly on mortality data and may restrict to specific subpopulations such as infants or children [9-11]. There are limitations in correctly identifying individuals with a genetic heart disease in these data collections due to

insufficient International Classifications of Diseases (ICD) codes [5], inadequate recognition of these diseases outside of specialized settings, and infrequent interaction with health facilities for many patients who remain largely well and are generally managed via outpatient clinics.

Here we report a protocol for the linkage of Australian Genetic Heart Disease (AGHD) Registry participants [12,13] with state-wide hospitalization, emergency department and mortality registries, and the national mortality registry to establish a rich and comprehensive data resource to examine the health care use of individuals with a genetic heart disease and their families for the first time.

## Methods

### Study Setting, Design, and Population

This is a retrospective cohort study of people in New South Wales (NSW), Australia, enrolled in the AGHD Registry from 2007 to 2019. The AGHD Registry operates nationally and has actively recruited participants since 2007 [12,13]. The goal of the AGHD Registry is to enroll every individual in Australia with a genetic heart disease and their at-risk relatives to understand and improve diagnosis, prevention, and treatment options. Eligibility criteria include a confirmed diagnosis of any inherited cardiomyopathy (HCM, dilated cardiomyopathy, arrhythmogenic cardiomyopathy, and restrictive cardiomyopathy), left ventricular noncompaction, or primary arrhythmia syndromes (long QT syndrome, Brugada syndrome, or catecholaminergic polymorphic ventricular tachycardia; [Table 1](#)) or at-risk relatives of those with a confirmed diagnosis.

Participant demographic, clinical, and genetic information (provided by the AGHD Registry data custodian) was linked with whole-population routinely collected health data, specifically hospital admissions (NSW Admitted Patient Data Collection), emergency department presentations (NSW Emergency Department Data Collection), and in the event of death, mortality registry data (NSW Registry of Births, Deaths, and Marriages; Australian Coordinating Registry Cause of Death Unit Record File; [Table 2](#)). Although the AGHD Registry is an ongoing registry, linkage to the routinely collected data is not continuous and only occurs on application to the data custodians with ethics approval.

**Table 1.** Classification and description of genetic heart diseases included in the AGHD<sup>a</sup> Registry.

Genetic heart disease type and specific condition	Definition	Diagnostic tests
<b>Inherited cardiomyopathies</b>		
Hypertrophic cardiomyopathy	Abnormal thickening of the heart muscle on either the left or right ventricle	Electrocardiogram, echocardiogram, cardiac magnetic resonance imaging (MRI), and Holter monitor
Dilated cardiomyopathy	Enlargement of the left ventricle of the heart affects the ability of the heart to pump blood effectively	Electrocardiogram, echocardiogram, cardiac MRI, and Holter monitor
Arrhythmogenic cardiomyopathy	A disease of the heart muscle where the normal heart muscle cells on either the left or right ventricle are replaced by fat and scar tissue. This change can lead to heart rhythm abnormalities, or the affected ventricle may become enlarged and not pump blood effectively	Electrocardiogram, echocardiogram, cardiac MRI, and Holter monitor
Left ventricular noncompaction	Characterized by deep trabeculations (finger-like projections) in the heart muscle wall of predominantly the left ventricle, but may also appear in the right ventricle	Electrocardiogram, echocardiogram, and cardiac MRI
<b>Primary arrhythmia syndromes</b>		
Long QT syndrome	Characterized by abnormal electrical activity in the heart that often presents in children and teenagers	Electrocardiogram, echocardiogram, and exercise or stress test
Catecholaminergic polymorphic ventricular tachycardia (CPVT)	CPVT is a rare condition often present in children and teenagers. A cardiac event can occur during periods of high emotion or exercise as these situations cause the body to release adrenalin or noradrenalin. These persons have an abnormal response to adrenaline, which causes the heart to beat fast and irregularly	Electrocardiogram, echocardiogram, exercise or stress test, and Holter monitor
Brugada syndrome	A rhythm disorder that causes the ventricles to beat abnormally fast, affecting the ability of the heart to pump blood effectively	Electrocardiogram, echocardiogram, and drug infusion study

<sup>a</sup>AGHD: Australian Genetic Heart Disease.

**Table 2.** Data sets included in the Australian Genetic Heart Disease (AGHD) Registry data linkage study.

Data set	Purpose	Description	Variables
AGHD Registry	Cohort definition and linked data	National registry collecting clinical and genetic data of patients with genetic heart disease and their families	<p>Demographics</p> <ul style="list-style-type: none"> <li>Sex</li> <li>Birth date (mm/yyyy)</li> <li>Ethnicity (based on the Australian Standard Classification of Cultural and Ethnic Groups [14])</li> <li>Vital status (alive or deceased)</li> <li>Postcode</li> <li>Disease status (definite, possible, or at-risk)</li> <li>Smoker status</li> </ul> <p>Clinical history</p> <ul style="list-style-type: none"> <li>Clinical diagnosis</li> <li>Age at diagnosis</li> <li>Reason for diagnosis</li> <li>Genetic status (likely pathogenic, pathogenic variant identified, or not identified)</li> <li>Sudden cardiac events ever</li> <li>Heart failure ever</li> <li>Syncope ever</li> <li>Nonsustained ventricular tachycardia ever</li> <li>Atrial fibrillation ever</li> <li>Left ventricular outflow tract obstruction ever</li> <li>Interventions (implantable cardioverter defibrillator, pacemaker, ablations, myectomy, and heart transplant)</li> <li>Family history</li> </ul> <p>Investigations</p> <ul style="list-style-type: none"> <li>Echocardiography parameters</li> <li>Electrocardiography parameters</li> </ul>
NSW <sup>a</sup> Admitted Patient Data Collection	Linked data	Records all admitted patient services provided by NSW public hospitals, private hospitals, and private day procedure centers	<p>Demographics</p> <ul style="list-style-type: none"> <li>Sex</li> <li>Birth date (mm/yyyy)</li> <li>Country of birth</li> <li>Residential statistical local area (SLA) at admission</li> <li>Health insurance on admission</li> </ul> <p>Episode of care</p> <ul style="list-style-type: none"> <li>Episode start date</li> <li>Episode end date</li> <li>Condition onset flag</li> <li>Diagnosis codes ICD-AM<sup>b</sup> (up to 50)</li> <li>Australian Refined Diagnosis-Related Group</li> <li>Hours in intensive care unit</li> <li>Hours on mechanical ventilation</li> <li>Mode of separation</li> <li>Procedure codes ICD-AM (up to 50)</li> <li>Procedure date</li> <li>Service-related group</li> </ul> <p>Facility information</p> <ul style="list-style-type: none"> <li>Hospital type</li> <li>Local health district (LHD)</li> <li>Facility identifier</li> <li>Facility transferred to</li> <li>Facility transferred from</li> </ul>

Data set	Purpose	Description	Variables
NSW Emergency Department Data Collection	Linked data	Provides information about patient presentations to the emergency departments of public hospitals in NSW	Demographics <ul style="list-style-type: none"> <li>• Sex</li> <li>• Birth date (mm/yyyy)</li> <li>• Country of birth</li> <li>• Residential SLA at admission</li> <li>• Health insurance on admission</li> </ul> Episode of care <ul style="list-style-type: none"> <li>• Episode start date and time</li> <li>• Episode end date and time</li> <li>• Diagnosis code ICD-AM (up to 50)</li> <li>• Type of visit</li> <li>• Mode of separation</li> <li>• Referred to on departure</li> </ul> Facility information <ul style="list-style-type: none"> <li>• LHD</li> <li>• Facility identifier</li> </ul>
NSW Registry of Births, Deaths, and Marriages	Linked data	Includes death registrations	<ul style="list-style-type: none"> <li>• Birth date (mm/yyyy)</li> <li>• Death date (mm/yyyy)</li> <li>• LHD at death</li> </ul>
Australian Coordinating Registry Cause of Death Unit Record File	Linked data	Includes death registrations	<ul style="list-style-type: none"> <li>• Death date (mm/yyyy)</li> <li>• Death age</li> <li>• Place at birth</li> <li>• State/territory of usual residence</li> <li>• Sex</li> <li>• Place of occurrence for external cause of death</li> <li>• Underlying cause of death ICD code</li> <li>• Contributing cause of death ICD codes (up to 20 comorbidity codes)</li> </ul>

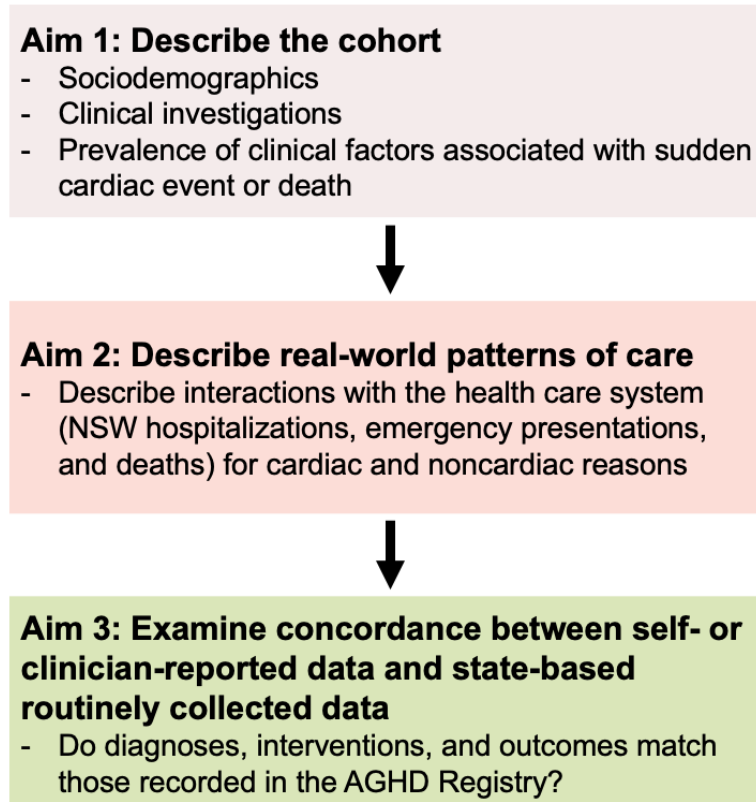
<sup>a</sup>NSW: New South Wales.

<sup>b</sup>ICD-AM: International Statistical Classification of Diseases and Related Health Problems, Australian Modification.

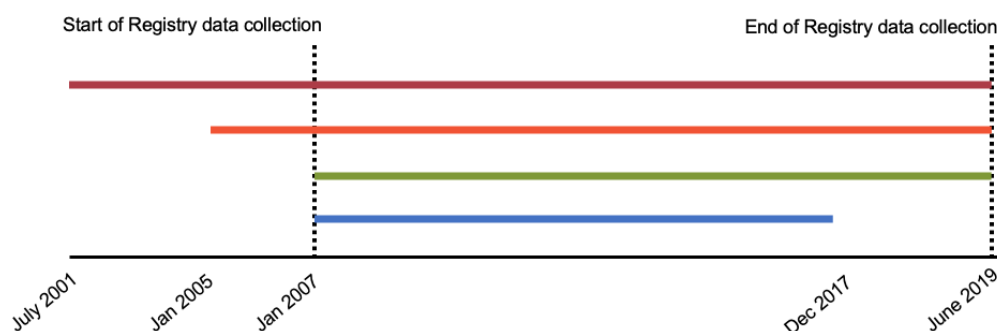
## Aims

Our objective is to characterize AGHD Registry participants using routinely collected whole-population health data sets to investigate the health care use of individuals with a genetic heart disease and their at-risk relatives (Figure 1). We will also examine the concordance between data sources to see whether

diagnoses, interventions, and outcomes match those recorded in the AGHD Registry. In our analyses, we will determine if patterns of care vary according to patient demographics such as age, sex, level of residential, and relative socioeconomic disadvantage of patient residence (Figure 2). We will evaluate episodes of care for any medical condition or reason.

**Figure 1.** Aims of the Australian Genetic Heart Disease (AGHD) Registry data linkage study. NSW: New South Wales.**Figure 2.** Period of data requested per routinely collected health data set.

**Routinely collected data—requested time data periods**  
 New South Wales Admitted Patient Data Collection  
 New South Wales Emergency Department Data Collection  
 New South Wales Registry of Births, Deaths, and Marriages  
 Australian Coordinating Registry Cause of Death Unit Record File



## Recruitment and Consent

Participants have been recruited to the AGHD Registry since 2007. Two recruitment methods were used for the data linkage aspects of the study:

1. For participants enrolled in the AGHD Registry between 2007 and April 2017, an opt-out approach was coupled with a waiver of consent. This involved sending a letter to all AGHD Registry participants outlining the study and asking them to contact us if they wish to withdraw their consent for this study. If the participant did not contact us, then the waiver of consent was applied.

2. An amendment to include permission for access to linked data sets was made to the participant information and consent form. Every person enrolled in the AGHD Registry from May 2017 provided written, informed consent for their data to be linked to the administrative data collections mentioned herein.

As part of the AGHD Registry recruitment process, extensive demographic, clinical, and genetic information was collected from the individuals and their cardiologist.

Although participants can withdraw their consent at any time for the AGHD Registry, data already included in the linkage



before their withdrawal of consent cannot be removed as all extracted data held by the researchers are deidentified.

### Data Sources and Variables

**Table 2** lists and describes the data sets used in this study. The AGHD Registry was used for cohort definition and linking variables with the routinely collected health data. We linked to the following data sets.

#### *NSW Admitted Patient Data Collection*

The Admitted Patient Data Collection collates information on all admitted patient services provided by NSW public (including psychiatric hospitals and multipurpose services) and private hospitals (including day procedure centers) [15]. Data include patient demographics (ie, sex, birth date, country of birth, and statistical local area of residence at admission), episodes of care (ie, date and time of admission and separation [the completion of treatment for an admitted patient because of death, discharge, or transfer to another facility or care type], referral source, diagnoses, procedures, and service referred upon discharge), and facility information (ie, hospital type, facility, and local health district [LHD]). Up to 50 relevant diagnoses and procedures were included. These are coded according to standardized codes from the International Statistical Classification of Diseases and Related Health Problems, Australian Modification and the Australian Refined Diagnosis-Related Group.

This study included data on admissions between July 1, 2001, and June 30, 2019 (Figure 2). When linked with the AGHD Registry data, this period allowed us to examine patterns of care before and after a genetic heart disease diagnosis. The requested variables will allow us to examine the frequency and type of interactions and care received during hospitalizations.

#### *NSW Emergency Department Data Collection*

The Emergency Department Data Collection provides information about patient presentations to the emergency departments of public hospitals and contracted private hospitals in NSW [15]. Each record represents a presentation to an emergency department for emergency care and treatment. Data include demographics (same as Admitted Patient Data Collection), episodes of care (eg, date of arrival and separation, diagnoses, and status at separation), and facility information (eg, facility identifier and LHD). Up to 50 relevant diagnoses are included and are coded according to standardized codes from the International Statistical Classification of Diseases and Related Health Problems, Australian Modification.

Data for the Emergency Department Data Collection were available from January 1, 2005, to June 30, 2019 (Figure 2). By combining the hospital data with the AGHD Registry data, we can analyze the emergency department episodes before and after their diagnosis during this period. These data will allow us to explore the frequency and reason for an emergency department presentation for any medical reason.

#### *Mortality Data*

Two data sets contain mortality information for deaths occurring in NSW: the NSW Registry of Births, Deaths, and Marriages death registrations and the Australian Coordinating Registry

Cause of Death Unit Record File [15]. All deaths are certified by a medical practitioner or coroner (if a coronial inquiry is required) and registered by the NSW Registry of Births, Deaths, and Marriages. Details of all registered deaths are forwarded to the Australian Coordinating Registry. Details include the date of death, underlying cause, and up to 20 contributing causes of death. The causes of death are coded using the ICD, Tenth Revision, international coding system rather than the International Statistical Classification of Diseases and Related Health Problems, Tenth Revision, Australian Modification used in Australian hospitals. The Australian Coordinating Registry Cause of Death Unit Record File data are finalized considerably later than the NSW Registry of Births, Deaths, and Marriages death registration data because of the time needed to confirm details on the death certificate.

Data from the NSW Registry of Births, Deaths, and Marriages were available from January 1, 2007, to June 30, 2019, and data from the Australian Coordinating Registry Cause of Death Unit Record File were available from January 1, 2007, to December 31, 2017 (Figure 2). These variables will allow us to determine the reasons and causes of death.

### Data Linkage Methods

Linkage of each AGHD Registry participant's demographic, clinical, and genetic information with whole-population routinely collected health data was conducted by the Centre for Health Record Linkage (CHeReL). The AGHD Registry data custodian (the individual responsible for the registry data and any related activities) provided CHeReL with 2 separate files. To protect privacy, identifier information, including given names, last names, and current and past residential addresses (street, suburb, postcode, and state), was separated from the content information (registry clinical and genetic information). The 2 files from each data set were stored and handled separately. CHeReL's data linkage unit applied a combination of matching methods to the identifier data to identify and distinguish between individuals. The identifier data belonging to each individual were then assigned an arbitrary person number, which replaced the name, address, and other identifying details. Using an encrypted version of the arbitrary person number, CHeReL's data integration unit created a research project-specific person number (PPN). This PPN was used to join the AGHD Registry content data to the NSW state-wide Admitted Patient Data Collection, Emergency Department Data Collection, and the mortality registries (Registry of Births, Deaths, and Marriages and Australian Coordinating Registry Cause of Death Unit Record File). Once linkages were complete, the data sets were released to the research team in a fully deidentified form, ensuring the privacy and confidentiality of all individuals. The AGHD Registry custodian does not have access to the linked data to ensure that the linked data cannot be reidentified. These deidentification measures preclude the withdrawal of any individuals included in the data linkage study after the linkage takes place. Access to the deidentified linked data is only provided to individuals named in the approved ethics protocol. The PPN combines records for the same person across different data sets.

## Statistical Analysis

All data harmonization and analyses will be conducted using the R software (version 4.2.3; R Foundation for Statistical Computing). Harmonization of the data sets will involve systematic identification and removal of any duplicate records, checking for a consistent number of person-level records across the data sets, consistency across variables (eg, admission dates before discharge dates), checking for outliers, and identification and flagging of missing or invalid data. We will analyze and report each data set using descriptive statistics, including mean, median, SD, IQR, and sample size (number). We will also conduct descriptive analysis stratified by factors including age, sex, relative socioeconomic disadvantage, and advantage of patient residence and LHD of facility using the Australian Bureau of Statistics Socio-Economic Indexes for Areas [16] and genetic status (likely pathogenic or pathogenic gene identified vs no gene identified). Additionally, Cox regression models will be developed to identify independent factors associated with adverse outcomes, such as heart failure, atrial fibrillation, stroke, and death.

## Patient and Public Involvement

Patients and the public have not and will not be involved in the study design.

## Ethics Approval

Technical feasibility for the study was granted by the CHeReL, NSW Ministry of Health (ID#2016.38). Ethical approval has been granted by the NSW Population and Health Services Research Ethics Committee (2019/ETH0153) and Sydney Local Health District (Royal Prince Alfred Hospital zone) Human Research Ethics Committee (2020/ETHO00220). The current approval permits data storage at the Garvan Institute of Medical Research. We will disseminate project findings at scientific conferences and in peer-reviewed journals.

## Results

Currently, the research team has received all deidentified data from CHeReL. The record linkage occurred in November 2019. A total of 1384 NSW-based AGHD Registry participants have been linked with 11,610 hospital records, 7032 emergency department records, and 60 death records. Data assessment and cleaning processes have been undertaken, with descriptive data analysis underway. Initial results are expected to be published by early 2024.

## Discussion

This study represents the first in Australia to provide a comprehensive clinical data set on health care use patterns of individuals with a genetic heart disease and their at-risk relatives. It overcomes the limitations of previous studies, which rely on ICD codes to identify patients, by linking data from the AGHD Registry with routinely collected health data in NSW. The study will shed light on health care use among individuals with a genetic heart disease in Australia, examining hospitalizations, emergency department presentations, and mortality rates over approximately 20 years, stratified by clinical

diagnosis, age, sex, socioeconomic status, and genetic status. We will investigate the impact of these factors on patterns of care, including interactions with health care providers and on disease severity and outcomes. Overall, we aim to improve the diagnosis, prevention, and treatment options for individuals with a genetic heart disease by providing a better understanding of their health care use and identifying areas for improvement in health care provision. Ultimately, the study's results will have a significant impact on the care of individuals with a genetic heart disease and their at-risk relatives.

Routinely collected data are commonly used in patterns of care studies. These studies are prolific when investigating cancer populations [17], as they provide the opportunity to describe a patient population and common treatment, as well as examine potential differences such as sex [18] and residential location [19]. In relation to cardiovascular diseases, patterns of care studies have predominantly focused on coronary heart disease or heart failure [18-22]. However, the population of individuals diagnosed with genetic heart diseases differs from those with these conditions, often being younger at diagnosis, with different risk factors and lifestyle constraints such as exclusion from high-level and competitive sports [23]. Therefore, studies such as that described here are needed to examine the unique genetic heart disease population.

By linking the AGHD Registry to routinely collected health data sets, we have overcome a significant limitation of previous studies that relied exclusively on the ICD diagnosis coding system to identify patients with genetic heart diseases [5]. A previous systematic review by our team found that for countries such as Australia, only 3 of the 12 genetic heart diseases included in the search strategy had an explicit ICD diagnosis code [5]. Codes that best match the other 9 genetic heart diseases being investigated were broad descriptive diagnoses (such as other cardiomyopathies used to describe arrhythmogenic cardiomyopathy), which also encompasses other genetic heart diseases. Additionally, the ICD diagnosis coding system fails to explicitly identify rare diseases, with only 500 of the 6000 rare diseases having an ICD diagnosis code [24].

Procedure codes may provide a little more information with specific codes for implantable cardioverter-defibrillator implantation and other interventions and surgeries associated with genetic heart diseases. However, as not all individuals undergo surgical procedures because of their condition, these codes also cannot identify all patients with a genetic heart disease. Because of these coding issues, studies to date have largely focused on those genetic heart diseases where discrimination from linked data sets is possible, that is, HCM and Marfan syndrome.

The clinical course for many patients with a genetic heart disease can be mild, meaning interactions with hospital and emergency departments could be infrequent or not required. However, when hospitalization is necessary, these patients often require urgent and potentially life-saving medical intervention. To date, there is limited research available regarding the health care system interactions of individuals with causative variants, and only one paper has used linked primary care hospital and mortality records in patients with HCM. Our study aims to address these



gaps in knowledge and provide answers to these important questions. We will be able to observe the interaction of individuals with less severe clinical course who may never require treatment at a hospital or emergency department for genetic heart disease but present for other reasons.

Because of the absence of studies exploring care patterns associated with genetic heart diseases, it remains unclear whether care varies based on specific factors. Previous studies examining nongenetic cardiac conditions have found that patient factors, such as age, sex, and residential location, impact care patterns [15,18,20,25]. It is vital to examine the impact of socioeconomic factors, as the AGHD Registry contains a significant proportion of patients seen at specialist genetic heart diseases clinics in NSW, which have been shown to overrepresent socioeconomically advantaged individuals [26]. We will be able to investigate if patterns of care or outcomes vary based on the location of the facility and the relative socioeconomic disadvantage of that location. For instance, we will examine whether areas with a relatively high disadvantage have fewer implantable cardioverter defibrillator surgeries than other facilities, as observed in other countries [27].

Among studies using more than one data source, validating the agreement between data types such as self-reporting and routinely collected health data is important. Multiple studies have examined the concordance between these data types for many medical conditions including cardiac diseases [18-21,28-31]. One study found that the agreement between self-report data and health records was substantially lower for persons with heart failure than for other chronic illnesses such

as diabetes, hypertension, myocardial infarction, and stroke [28]. Patient factors associated with discordance include sex, age, education level, and socioeconomic status [30,32]. However, no studies have investigated the concordance in individuals with a genetic heart disease between self-reported and routinely collected health data.

There are limitations in coding accuracy, especially for the Emergency Department Data Collection, which could affect the validity of the data collected. The study relies on AGHD Registry participants who are predominantly from a single tertiary center, which has been shown to lack socioeconomic diversity [29]. We will be unable to ascertain the prevalence and incidence of genetic heart diseases accurately as the AGHD Registry does not include all patients in NSW with a genetic heart disease. However, even if our sample is not representative, the relationships between variables are still valid [33].

This study aims to provide valuable insights into the health care use patterns of individuals with a genetic heart disease and their at-risk relatives. By linking data from the AGHD Registry with routinely collected health data in NSW, this study has established a comprehensive clinical data set that can identify reasons for hospitalization and emergency presentation, with the potential to reveal disparities based on disease, sex and sociodemographic status, and opportunities for preventive care. Overall, this study's findings may have a significant impact on the care of individuals with a genetic heart disease and their at-risk relatives, with the potential for better outcomes and quality of life.

## Acknowledgments

JJ acknowledges a National Heart Foundation of Australia Future Leader Fellowship (#106732). This work was supported in part by New South Wales Health Early-Mid Career Research Cardiovascular Grant. LY is a recipient of a cofunded National Heart Foundation of Australia/National Health and Medical Research Council (NHMRC) PhD scholarship (#102568/#191351). CS is supported by an NHMRC Investigator Grant (#2016822) and an NSW Health Cardiovascular Disease Clinician Scientist Grant. LN is supported by a National Heart Foundation Future Leader Fellowship (#105038).

## Data Availability

The NSW Health data sets generated during and analyzed during this study are not publicly available as they contain sensitive health information and an application must be made to the data custodians to use this data (through the Centre for Health Record Linkage). The Australian Genetic Heart Disease Registry data are available from the Registry custodian (A/Prof Jodie Ingles) on reasonable request.

## Conflicts of Interest

JJ receives research grant support from Bristol Myers Squibb. All remaining authors have nothing to disclose.

## References

1. Girolami F, Frisso G, Benelli M, Crotti L, Iacone M, Mango R, et al. Contemporary genetic testing in inherited cardiac disease: tools, ethical issues, and clinical applications. *J Cardiovasc Med (Hagerstown)* 2018;19(1):1-11 [FREE Full text] [doi: [10.2459/JCM.0000000000000589](https://doi.org/10.2459/JCM.0000000000000589)] [Medline: [29176389](https://pubmed.ncbi.nlm.nih.gov/29176389/)]
2. Semsarian C, Ingles J, Maron MS, Maron BJ. New perspectives on the prevalence of hypertrophic cardiomyopathy. *J Am Coll Cardiol* 2015;65(12):1249-1254 [FREE Full text] [doi: [10.1016/j.jacc.2015.01.019](https://doi.org/10.1016/j.jacc.2015.01.019)] [Medline: [25814232](https://pubmed.ncbi.nlm.nih.gov/25814232/)]
3. Austin R, Quinn MCJ, Afoakwah C, Metke-Jimenez A, Leroux H, Atherton J, et al. Investigation of current models of care for genetic heart disease in Australia: a national clinical audit. *Int J Cardiol* 2021;330:128-134 [FREE Full text] [doi: [10.1016/j.ijcard.2021.02.010](https://doi.org/10.1016/j.ijcard.2021.02.010)] [Medline: [33581180](https://pubmed.ncbi.nlm.nih.gov/33581180/)]

4. Olde Nordkamp LRA, Wilde AAM, Tijssen JGP, Knops RE, van Dessel PFHM, de Groot JR. The ICD for primary prevention in patients with inherited cardiac diseases: indications, use, and outcome: a comparison with secondary prevention. *Circ Arrhythm Electrophysiol* 2013;6(1):91-100 [FREE Full text] [doi: [10.1161/CIRCEP.112.975268](https://doi.org/10.1161/CIRCEP.112.975268)] [Medline: [23275262](https://pubmed.ncbi.nlm.nih.gov/23275262/)]
5. Blanch B, Sweeting J, Semsarian C, Ingles J. Routinely collected health data to study inherited heart disease: a systematic review (2000-2016). *Open Heart* 2017;4(2):e000686 [FREE Full text] [doi: [10.1136/openhrt-2017-000686](https://doi.org/10.1136/openhrt-2017-000686)] [Medline: [29209507](https://pubmed.ncbi.nlm.nih.gov/29209507/)]
6. Benchimol EI, Smeeth L, Guttman A, Harron K, Moher D, Petersen I, RECORD Working Committee. The REporting of studies Conducted using Observational Routinely-collected health Data (RECORD) Statement. *PLoS Med* 2015;12(10):e1001885 [FREE Full text] [doi: [10.1371/journal.pmed.1001885](https://doi.org/10.1371/journal.pmed.1001885)] [Medline: [26440803](https://pubmed.ncbi.nlm.nih.gov/26440803/)]
7. Emery J, Boyle D. Data linkage. *Aust Fam Physician* 2017;46(8):615-619 [Medline: [28787562](https://pubmed.ncbi.nlm.nih.gov/28787562/)]
8. Smith M, Flack F. Data linkage in Australia: the first 50 years. *Int J Environ Res Public Health* 2021;18(21):11339 [FREE Full text] [doi: [10.3390/ijerph182111339](https://doi.org/10.3390/ijerph182111339)] [Medline: [34769852](https://pubmed.ncbi.nlm.nih.gov/34769852/)]
9. Winkel BG, Holst AG, Theilade J, Kristensen IB, Thomsen JL, Ottesen GL, et al. Nationwide study of sudden cardiac death in persons aged 1-35 years. *Eur Heart J* 2011;32(8):983-990 [FREE Full text] [doi: [10.1093/eurheartj/ehq428](https://doi.org/10.1093/eurheartj/ehq428)] [Medline: [21131293](https://pubmed.ncbi.nlm.nih.gov/21131293/)]
10. Morris VB, Keelan T, Leen E, Keating J, Magee H, O'Neill JO, et al. Sudden cardiac death in the young: a 1-year post-mortem analysis in the Republic of Ireland. *Ir J Med Sci* 2009;178(3):257-261 [doi: [10.1007/s11845-009-0294-8](https://doi.org/10.1007/s11845-009-0294-8)] [Medline: [19221830](https://pubmed.ncbi.nlm.nih.gov/19221830/)]
11. Fragkouli K, Vougiouklakis T. Sudden cardiac death: an 11-year postmortem analysis in the region of Epirus, Greece. *Pathol Res Pract* 2010;206(10):690-694 [FREE Full text] [doi: [10.1016/j.prp.2010.05.005](https://doi.org/10.1016/j.prp.2010.05.005)] [Medline: [20932458](https://pubmed.ncbi.nlm.nih.gov/20932458/)]
12. Ingles J, Semsarian C. The Australian genetic heart disease registry. *Int J Cardiol* 2013;168(4):e127-e128 [doi: [10.1016/j.ijcard.2013.08.036](https://doi.org/10.1016/j.ijcard.2013.08.036)] [Medline: [23998550](https://pubmed.ncbi.nlm.nih.gov/23998550/)]
13. Ingles J, McGaughan J, Vohra J, Weintraub RG, Davis A, Atherton J, et al. Establishment of an Australian National Genetic Heart Disease Registry. *Heart Lung Circ* 2008;17(6):463-467 [FREE Full text] [doi: [10.1016/j.hlc.2008.05.603](https://doi.org/10.1016/j.hlc.2008.05.603)] [Medline: [18722159](https://pubmed.ncbi.nlm.nih.gov/18722159/)]
14. Australian Standard Classification of Cultural and Ethnic Groups (ASCCEG). Australian Bureau of Statistics. 2016. URL: <http://www.abs.gov.au/AUSSTATS/abs@.nsf/DetailsPage/1249.02016?OpenDocument> [accessed 2023-08-01]
15. Mehta RH, Rathore SS, Radford MJ, Wang Y, Wang Y, Krumholz HM. Acute myocardial infarction in the elderly: differences by age. *J Am Coll Cardiol* 2001;38(3):736-741 [FREE Full text] [doi: [10.1016/s0735-1097\(01\)01432-2](https://doi.org/10.1016/s0735-1097(01)01432-2)] [Medline: [11527626](https://pubmed.ncbi.nlm.nih.gov/11527626/)]
16. Technical paper: socio-economic indexes for areas (SEIFA). Australian Bureau of Statistics. 2016. URL: <https://tinyurl.com/yc26sf23> [accessed 2020-04-27]
17. Vinod SK, Delaney GP, Bauman AE, Barton MB. Lung cancer patterns of care in south western Sydney, Australia. *Thorax* 2003;58(8):690-694 [FREE Full text] [doi: [10.1136/thorax.58.8.690](https://doi.org/10.1136/thorax.58.8.690)] [Medline: [12885986](https://pubmed.ncbi.nlm.nih.gov/12885986/)]
18. Ayanian JZ, Epstein AM. Differences in the use of procedures between women and men hospitalized for coronary heart disease. *N Engl J Med* 1991;325(4):221-225 [FREE Full text] [doi: [10.1056/NEJM199107253250401](https://doi.org/10.1056/NEJM199107253250401)] [Medline: [2057022](https://pubmed.ncbi.nlm.nih.gov/2057022/)]
19. Gamble JM, Eurich DT, Ezekowitz JA, Kaul P, Quan H, McAlister FA. Patterns of care and outcomes differ for urban versus rural patients with newly diagnosed heart failure, even in a universal healthcare system. *Circ Heart Fail* 2011;4(3):317-323 [FREE Full text] [doi: [10.1161/CIRCHEARTFAILURE.110.959262](https://doi.org/10.1161/CIRCHEARTFAILURE.110.959262)] [Medline: [21430285](https://pubmed.ncbi.nlm.nih.gov/21430285/)]
20. Adams KF, Fonarow GC, Emerman CL, LeJemtel TH, Costanzo MR, Abraham WT, ADHERE Scientific Advisory Committee and Investigators. Characteristics and outcomes of patients hospitalized for heart failure in the United States: rationale, design, and preliminary observations from the first 100,000 cases in the acute decompensated heart failure national registry (ADHERE). *Am Heart J* 2005;149(2):209-216 [doi: [10.1016/j.ahj.2004.08.005](https://doi.org/10.1016/j.ahj.2004.08.005)] [Medline: [15846257](https://pubmed.ncbi.nlm.nih.gov/15846257/)]
21. McClellan WM, Flanders WD, Langston RD, Jurkovitz C, Presley R. Anemia and renal insufficiency are independent risk factors for death among patients with congestive heart failure admitted to community hospitals: a population-based study. *J Am Soc Nephrol* 2002;13(7):1928-1936 [FREE Full text] [doi: [10.1097/01.asn.0000018409.45834.fa](https://doi.org/10.1097/01.asn.0000018409.45834.fa)] [Medline: [12089390](https://pubmed.ncbi.nlm.nih.gov/12089390/)]
22. Quinn JR, Pesis-Katz I, Mukamel DB. Variations in patterns of care activities for nursing home residents with heart failure. *Res Gerontol Nurs* 2013;6(1):29-35 [doi: [10.3928/19404921-20121204-02](https://doi.org/10.3928/19404921-20121204-02)] [Medline: [23244568](https://pubmed.ncbi.nlm.nih.gov/23244568/)]
23. Sweeting J, Ingles J, Ball K, Semsarian C. Challenges of exercise recommendations and sports participation in genetic heart disease patients. *Circ Cardiovasc Genet* 2015;8(1):178-186 [FREE Full text] [doi: [10.1161/CIRCGENETICS.114.000784](https://doi.org/10.1161/CIRCGENETICS.114.000784)] [Medline: [25691687](https://pubmed.ncbi.nlm.nih.gov/25691687/)]
24. Rath A, Olry A, Dhombres F, Brandt MM, Urbero B, Ayme S. Representation of rare diseases in health information systems: the Orphanet approach to serve a wide range of end users. *Hum Mutat* 2012;33(5):803-808 [FREE Full text] [doi: [10.1002/humu.22078](https://doi.org/10.1002/humu.22078)] [Medline: [22422702](https://pubmed.ncbi.nlm.nih.gov/22422702/)]
25. Jneid H, Fonarow GC, Cannon CP, Hernandez AF, Palacios IF, Maree AO, Get With the Guidelines Steering Committee and Investigators. Sex differences in medical care and early death after acute myocardial infarction. *Circulation* 2008;118(25):2803-2810 [FREE Full text] [doi: [10.1161/CIRCULATIONAHA.108.789800](https://doi.org/10.1161/CIRCULATIONAHA.108.789800)] [Medline: [19064680](https://pubmed.ncbi.nlm.nih.gov/19064680/)]
26. Ingles J, Johnson R, Sarina T, Yeates L, Burns C, Gray B, et al. Social determinants of health in the setting of hypertrophic cardiomyopathy. *Int J Cardiol* 2015;184:743-749 [doi: [10.1016/j.ijcard.2015.03.070](https://doi.org/10.1016/j.ijcard.2015.03.070)] [Medline: [25827935](https://pubmed.ncbi.nlm.nih.gov/25827935/)]

27. Winther-Jensen M, Hassager C, Lassen JF, Køber L, Torp-Pedersen C, Hansen SM, et al. Association between socioeconomic factors and ICD implantation in a publicly financed health care system: a Danish nationwide study. *Europace* 2018;20(7):1129-1137 [FREE Full text] [doi: [10.1093/europace/eux223](https://doi.org/10.1093/europace/eux223)] [Medline: [29016817](https://pubmed.ncbi.nlm.nih.gov/29016817/)]
28. Okura Y, Urban LH, Mahoney DW, Jacobsen SJ, Rodeheffer RJ. Agreement between self-report questionnaires and medical record data was substantial for diabetes, hypertension, myocardial infarction and stroke but not for heart failure. *J Clin Epidemiol* 2004;57(10):1096-1103 [doi: [10.1016/j.jclinepi.2004.04.005](https://doi.org/10.1016/j.jclinepi.2004.04.005)] [Medline: [15528061](https://pubmed.ncbi.nlm.nih.gov/15528061/)]
29. Violán C, Foguet-Boreu Q, Hermosilla-Pérez E, Valderas JM, Bolívar B, Fàbregas-Escurriola M, et al. Comparison of the information provided by electronic health records data and a population health survey to estimate prevalence of selected health conditions and multimorbidity. *BMC Public Health* 2013;13:251 [FREE Full text] [doi: [10.1186/1471-2458-13-251](https://doi.org/10.1186/1471-2458-13-251)] [Medline: [23517342](https://pubmed.ncbi.nlm.nih.gov/23517342/)]
30. Yasaitis LC, Berkman LF, Chandra A. Comparison of self-reported and Medicare claims-identified acute myocardial infarction. *Circulation* 2015;131(17):1477-1485 [FREE Full text] [doi: [10.1161/CIRCULATIONAHA.114.013829](https://doi.org/10.1161/CIRCULATIONAHA.114.013829)] [Medline: [25747935](https://pubmed.ncbi.nlm.nih.gov/25747935/)]
31. Misirli G, Bamia C, Dilis V, Benetou V, Zilis D, Trichopoulou A. Validation of self-reported incident cardiovascular disease events in the Greek EPIC cohort study. *Ital J Public Health* 2012;9(3):e7538 [FREE Full text] [doi: [10.2427/7538](https://doi.org/10.2427/7538)]
32. Teh R, Doughty R, Connolly M, Broad J, Pillai A, Wilkinson T, et al. Agreement between self-reports and medical records of cardiovascular disease in octogenarians. *J Clin Epidemiol* 2013;66(10):1135-1143 [doi: [10.1016/j.jclinepi.2013.05.001](https://doi.org/10.1016/j.jclinepi.2013.05.001)] [Medline: [23860185](https://pubmed.ncbi.nlm.nih.gov/23860185/)]
33. Willett WC, Blot WJ, Colditz GA, Folsom AR, Henderson BE, Stampfer MJ. Merging and emerging cohorts: not worth the wait. *Nature* 2007;445(7125):257-258 [doi: [10.1038/445257a](https://doi.org/10.1038/445257a)] [Medline: [17230171](https://pubmed.ncbi.nlm.nih.gov/17230171/)]

## Abbreviations

**AGHD:** Australian Genetic Heart Disease  
**CHeReL:** Centre for Health Record Linkage  
**HCM:** hypertrophic cardiomyopathy  
**ICD:** International Classification of Diseases  
**LHD:** local health district  
**NSW:** New South Wales  
**PPN:** project-specific person number  
**SCD:** sudden cardiac death

*Edited by A Mavragani; submitted 01.05.23; peer-reviewed by I Kedan, O Balogun; comments to author 21.07.23; revised version received 25.07.23; accepted 26.07.23; published 20.09.23*

*Please cite as:*

Butters A, Blanch B, Kemp-Casey A, Do J, Yeates L, Leslie F, Semsarian C, Nedkoff L, Briffa T, Ingles J, Sweeting J  
*The Australian Genetic Heart Disease Registry: Protocol for a Data Linkage Study*  
*JMIR Res Protoc* 2023;12:e48636  
URL: <https://www.researchprotocols.org/2023/1/e48636>  
doi: [10.2196/48636](https://doi.org/10.2196/48636)  
PMID:

©Alexandra Butters, Bianca Blanch, Anna Kemp-Casey, Judy Do, Laura Yeates, Felicity Leslie, Christopher Semsarian, Lee Nedkoff, Tom Briffa, Jodie Ingles, Joanna Sweeting. Originally published in *JMIR Research Protocols* (<https://www.researchprotocols.org>), 20.09.2023. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in *JMIR Research Protocols*, is properly cited. The complete bibliographic information, a link to the original publication on <https://www.researchprotocols.org>, as well as this copyright and license information must be included.