

International Journal of Population Data Science

Journal Website: www.ijpds.org



Swansea University
Prifysgol Abertawe

Using novel data linkage of congenital heart disease biobank data with administrative health data to identify cardiovascular outcomes to inform genomic analysis

Samantha J. Lain^{1,*}, Gillian M. Blue^{2,3}, Bridget R. O'Malley^{2,3}, David S. Winlaw⁴, Gary Sholler^{1,2}, Sally L. Dunwoodie^{5,6}, Natasha Nassar¹, and the Congenital Heart Disease Synergy Study group*

Submission History

Submitted:	29/03/2023
Accepted:	22/08/2023
Published:	14/11/2023

¹Child Population and Translational Health Research, Children's Hospital at Westmead Clinical School, The University of Sydney, NSW, Australia, 2006

²Heart Centre for Children, Sydney Children's Hospital Network, Children's Hospital at Westmead, NSW, Australia

³Faculty of Medicine and Health, The University of Sydney, Camperdown, NSW, Australia

⁴Cincinnati Children's Hospital Medical Centre, Heart Institute, and University of Cincinnati College of Medicine, Cincinnati, OH, USA

⁵Victor Chang Cardiac Research Institute, Sydney, New South Wales, Australia

⁶School of Clinical Medicine, Faculty of Medicine and Health, University of New South Wales, Australia

*The Congenital Heart Disease Synergy Study group: Sally L. Dunwoodie, David Winlaw, Eleni Giannoulatou, Natasha Nassar, Edwin Kirk, Gavin Chapman, Gillian Blue, Gary Sholler, Samantha Lain.

Abstract

Introduction

Contemporary care of congenital heart disease (CHD) is largely standardised, however there is heterogeneity in post-surgical outcomes that may be explained by genetic variation. Data linkage between a CHD biobank and routinely collected administrative datasets is a novel method to identify outcomes to explore the impact of genetic variation.

Objective

Use data linkage to identify and validate patient outcomes following surgical treatment for CHD.

Methods

Data linkage between clinical and biobank data of children born from 2001–2014 that had a procedure for CHD in New South Wales, Australia, with hospital discharge data, education and death data. The children were grouped according to CHD lesion type and age at first cardiac surgery. Children in each 'lesion/age at surgery group' were classified into 'favourable' and 'unfavourable' cardiovascular outcome groups based on variables identified in linked administrative data including: total time in intensive care, total length of stay in hospital, and mechanical ventilation time up to 5 years following the date of the first cardiac surgery. A blind medical record audit of 200 randomly chosen children from 'favourable' and 'unfavourable' outcome groups was performed to validate the outcome groups.

Results

Of the 1872 children in the dataset that linked to hospital or death data, 483 were identified with a 'favourable' cardiovascular outcome and 484 were identified as having a 'unfavourable' cardiovascular outcome. The medical record audit found concordant outcome groups for 182/192 records (95%) compared to the outcome groups categorized using the linked data.

Conclusions

The linkage of a curated biobank dataset with routinely collected administrative data is a reliable method to identify outcomes to facilitate a large-scale study to examine genetic variance. These genetic hallmarks could be used to identify patients who are at risk of unfavourable cardiovascular outcomes, to inform strategies for prevention and changes in clinical care.

Keywords

congenital heart disease; data linkage; genetic variation; cardiovascular outcome

*Corresponding Author:

Email Address: Samantha.lain@sydney.edu.au (Samantha J. Lain)

Introduction

Congenital heart disease (CHD) pertains to structural heart abnormalities present at birth and is the most common congenital anomaly. In Australia, approximately half of babies born with CHD will require surgical or catheter-based interventions at some time, and a third will require surgical intervention in the first year of life [1]. There is substantial variability in outcomes after surgery for CHD that cannot easily be explained [2, 3]. It is hypothesised that genetic changes may explain some of this variability, although the specific genes and pathways affected are not well described.

In contemporary care of CHD, perioperative care for most conditions is largely standardized. For example, in the management of transposition of the great arteries, there is a uniform approach to surgery, well-established routines for use of the heart-lung machine, and protocols for optimal post-operative care [4]. Despite internal consistency in surgical technique and care pathways, there is heterogeneity in outcomes that while not causing death, may cause delays in recovery as well as short- and long-term morbidity. Genetic variation in pathways mediating responses to stressors and cardiac function may be relevant, in addition to prenatal, socio-demographic and maternal factors.

Genetic testing is indicated for patients with a wide range of CHD morphologies [5]. In specific sub-groups of patients, a genetic diagnosis can be established to explain the abnormal cardiac development that led to the CHD subtype. Beyond these insights into causation of structural heart disease, specific types of genetic variation have been shown to be associated with worse clinical outcomes after surgery for CHD, including elevated risk of mortality and longer time on the ventilator [2]. Tests of association are limited by the relatively small number of patients affected and limited sample sizes. The identification of composite 'favourable' and 'unfavourable' outcomes is an alternative approach that avoids excessive segmentation of the cohort according to CHD morphology and seeks to eliminate assumptions as to the cause of various outcomes.

To obtain insights into the clinical and genetic outcomes of infants with CHD requires bringing together various pieces of information that are collected as part of research and clinical care. The use of data linkage to develop a categorization of patient outcomes, is a novel and cost-effective method which may facilitate a large-scale study of genetic variants and their association with patient outcome following surgical treatment for CHD.

The aim of this study was to (i) use data linkage of routinely collected administrative and clinical data to identify cardiovascular outcomes following surgery for CHD, and (ii) validate these outcomes against the medical records of a random sample of children.

Methods

Study setting and participants

The study population included children who were born from 2001–2014, had primary surgery for a CHD, and had a DNA sample collected and stored at The Children's Hospital at

Westmead, New South Wales (NSW), Australia. As children with syndromes, such as Down syndrome (trisomy 21), have poorer postsurgical outcomes [6], they were excluded. Children with isolated patent ductus arteriosus (PDA) were also excluded.

Data sources

The Kids Heart BioBank (KHB) is a clinical dataset with associated biological sample collected for children who have had surgery for CHD in The Children's Hospital at Westmead (CHW) in NSW. The Heart Centre at CHW is the largest centre for treating CHD in NSW. Recruitment to the KHB began in 2003 and informed consent was given by the child's parent/guardian to collect diagnosis and treatment information, and a blood sample for DNA analysis. The KHB was linked to three administrative datasets to identify infant and child outcomes, including the Admitted Patient Data Collection (APDC), Register of Births, Deaths and Marriages (RBDM) and the National Assessment Program – Literacy and Numeracy (NAPLAN). The APDC is an administrative database of all public and private hospital admissions in NSW and is based on information from hospital medical records, with diagnoses coded according to the 10th revision of the International Classification of Diseases Australian modification (ICD10-AM) and procedures coded according to the Australian Classification of Health Interventions (ACHI). The RBDM includes a registration of all deaths in NSW. The NAPLAN is an annual nationwide assessment of numeracy and literacy of all Australian children in grades 3, 5, 7 and 9 (aged 9–14 years). Children who are 'exempt' from sitting the NAPLAN testing, those with significant intellectual disability or co-existing conditions, were also included.

Data linkage methods

Data linkage was performed by the NSW Centre for Health Record Linkage (CHeReL). The CHeReL uses identifying information such as name, address, date of birth and gender for probabilistic record linkage process, through *ChoiceMaker* software [7]. *ChoiceMaker* uses 'blocking' and 'scoring' to identify definite and possible matches. A combination of a probabilistic decision, which is computed using a machine learning technique, and absolute rules, which include upper and lower probability cut-offs, to determine whether each potential match denotes or possibly denotes the same person. At the completion of the process, each record is assigned a record identification number to allow linked records for the same individual with de-identified linked data forwarded to researchers for analysis. A random sample of 1,000 Person IDs from the linkage was selected and reviewed and found a false positive linkage rate of 5/1,000 Person IDs (0.5%).

Ethics approval for the study was attained from the NSW Population and Health Services Research Ethics Committee (2019/ETH11615).

Congenital heart disease groups

Given the heterogeneity in the type, severity and clinical management of CHD, children in the KHB were divided into mutually exclusive groups defined by CHD lesion type and

age at first surgery. CHD lesion is recorded upon recruitment into KHB and was categorised into 3 groups based on broad similarity of anatomical abnormality (Table 1). These 3 lesion groups were then divided into 3 sub-groups defined by age at surgery; (i) neonatal surgery (0–30 days), (ii) infant surgery (31 days–1 year) and (iii) childhood surgery (1–5 years), to form 9 mutually exclusive 'lesion/age at surgery' groups (Table 1). The date of first surgery was identified in the APDC as when a patient first had a cardiac procedure identified from a list of ACHI codes (Supplementary Appendix 1) previously used to identify CHD cases [8].

Study outcomes

Clinical cardiovascular outcomes were determined in consultation with a paediatric cardiac surgeon (DW) and paediatric cardiologist (GS) based on data from the APDC and were chosen to identify 'unfavourable' and 'favourable' outcomes. Hospital admissions up to five years following the date of the initial cardiac surgery were examined. Severe cardiac-related morbidity or mortality based on recorded diagnosis of cardiac arrest, stroke and/or seizure/convulsion were identified (Supplementary Appendix 2). Total time in an intensive care unit (ICU) and total completed hours on mechanical ventilation during cardiac surgery admission and any subsequent admission for both cardiac and non-cardiac issues, were aggregated per child. Time receiving mechanical ventilation includes both invasive and non-invasive mechanical ventilation. Total length of stay (LOS) in hospital up to 5 years following primary cardiac procedure, including readmissions to any hospital in NSW, were aggregated per child. All hospital admissions up to 5 years following the first surgery were included as many complications related to CHD, such as infections and seizures, are non-cardiac. Length of stay and time in ICU were chosen to identify outcome groups as these variables are good overall markers for a complicated post-surgical clinical course and associated with poor cardiovascular outcomes, [9] and this unbiased approach seeks to eliminate assumptions as to the cause of various outcomes.

Children in each of the 9 'lesion/age at surgery' groups were classified into 3 'unfavourable' outcome categories using a hierarchical approach; (1) registration of death or a recorded diagnosis of cardiac arrest, stroke and/or seizure/convulsion, (2) $\geq 90^{\text{th}}$ centile of total time in ICU in the 5 years following primary cardiac procedure, (3) $\geq 90^{\text{th}}$ centile of total days in hospital and 75–90th centile of total time in ICU up to 5 years following primary cardiac procedure (Table 2).

As 'favourable' clinical diagnoses are not recorded in hospital discharge data, those with the shortest total length of time in ICU, shortest time on mechanical ventilation and total days in hospital up to 5 years following primary procedure were classified into 3 'favourable' outcome categories (Table 2). To ensure that children identified as having a 'favourable' cardiovascular outcome did not leave NSW following their initial surgery and have a 'unfavourable' cardiovascular outcome recorded in another jurisdiction, children over the age of 10 years old were excluded if they did not link to NSW school attendance data, identified by lack of a NAPLAN school test record. Children who were classified as 'exempt' from NAPLAN were also excluded from having a 'favourable' outcome as poor neurocognitive outcomes in CHD patients are

associated with poor cardiovascular function and insufficient cerebral oxygenation [10].

Clinical audit and validation

For each 'lesion/age at surgery' group, we selected approximately 25% of patients with 'favourable' and 25% with 'unfavourable' cardiovascular outcomes, to establish a cohort of ~ 475 'unfavourable' cardiovascular outcome cases and ~ 475 'favourable' controls. Once a list of children with 'unfavourable' and 'favourable' outcomes were chosen, a list of linkage numbers was sent to the CHeReL who re-identified cases and sent these to the KHB data custodian (GB) for validation. The data custodian was blinded to whether the children had favourable or unfavourable outcomes and were only told children were classified to either group 1 or 2.

The KHB data custodian randomly selected 100 children from each group (total $n = 200$) using the list of children sent by CHeReL. A random sample of children with 'favourable' or 'unfavourable' outcomes were audited as we want to ensure those we have identified, truly have the outcome to which they are categorised. This is to ensure we do not have misclassification of outcomes that would impact the results of the genetic analysis. We did not audit the records of children from the cohort that were not selected as having a 'favourable' or 'unfavourable' outcome as it is not the aim of the study to ensure we identified all cases.

Medical records from The Children's Hospital at Westmead were accessed to validate 'unfavourable' or 'favourable' cardiovascular outcome defined by record linkage. A clinical researcher, blinded to outcome category, abstracted details from medical records and classified each record as 'unfavourable' or 'favourable' using clinical judgement to assess whether patients followed an expected clinical course for their lesion type (see Supplementary Appendix 3). Various documents, such as discharge summaries, operation reports, and nursing notes were reviewed to assess clinical variables. All cases were assessed in the context of the specific CHD lesion group, typical surgery/intervention and typical length of stay for that lesion. As only the medical records for the Children's Hospital at Westmead in NSW were available subsequent admissions to other hospitals in NSW could not be examined in the validation process although medical record-based coding information were available for these admissions.

Results

There were 2,330 children in the KHB born from 2001–2014 (Figure 1). Following exclusion of those with a syndrome, with an isolated PDA, and that did not link to an APDC record, there were 1,872 children included in the cohort. The median total hours in ICU up to 5 years following the primary CHD surgery for each group is shown in Table 3. As expected, the total time in ICU decreases as age of first surgery increases (Table 3). Those groups who had surgery in the first 30 days of life (neonatal) had median time in ICU from 273–358 hours, while the groups who had surgery after 1 year of age (childhood) had median ICU times from 24–71 hours. A similar pattern was seen for total time received mechanical ventilation and total days/length of stay in hospital (Table 3).

Table 1: CHD lesion and age at surgery groups

CHD Lesions	Lesion groups	Age at surgery group/Procedure example
• Functional single ventricle • Heterotaxy/and or complex heart disease • Dextro-transposition of the great arteries • Truncus arteriosus • Double outlet right ventricle + significant malposition • Total anomalous pulmonary venous return	Group A	Neonatal surgery/Norwood procedure Infant surgery/Repair of atrioventricular septal defect Childhood surgery/Fontan operation
• Tetralogy of Fallot • Common Atrioventricular canal • Coarctation of the aorta • Double outlet right ventricle (Fallot type) • Ebstein anomaly • Partial anomalous pulmonary venous return • Left Ventricular Outflow Tract Obstruction (isolated) • Right Ventricular Outflow Tract Obstruction (isolated)	Group B	Neonatal surgery/Arterial switch operation Infant surgery/Repair of tetralogy of Fallot Childhood surgery/Repair of aortic valve stenosis
• Atrial septal defect (ASD) • Ventricular septal defect (VSD) • Double outlet right ventricle (isolated) • Vascular ring	Group C	Neonatal surgery/Closure of VSD Infant surgery/Closure of VSD Childhood surgery/Closure of secundum ASD

Table 2: Definitions of 'unfavourable' and 'favourable' cardiovascular outcome categories

Outcome category	Definition
Unfavourable – category 1	Recorded death or diagnosis of cardiac arrest, stroke, intracerebral haemorrhage, cerebral infarction, seizure/convulsion
Unfavourable – category 2	Total time in intensive care unit (neonatal and/or paediatric ICU) during cardiac surgery and readmissions that is $\geq 90^{\text{th}}$ centile for 'lesion/age at surgery group'
Unfavourable – category 3	Total length of stay (LOS) in hospital (both during cardiac procedure and readmissions) that is $\geq 90^{\text{th}}$ centile for 'lesion/age at surgery group' and total time in ICU $\geq 75^{\text{th}}$ centile
Favourable – category 1	Total time in ICU that is $\leq 25^{\text{th}}$ centile, total hours on mechanical ventilation $\leq 25^{\text{th}}$ centile & total LOS $\leq 50^{\text{th}}$ centile
Favourable – category 2	Total time in ICU that is $\leq 25^{\text{th}}$ centile & total hours on mechanical ventilation $\leq 50^{\text{th}}$ centile & total LOS $\leq 50^{\text{th}}$ centile for 'lesion/age at surgery group'
Favourable – category 3	Total time in ICU that is $< 50^{\text{th}}$ centile & total hours on mechanical ventilation $< 50^{\text{th}}$ centile & total LOS $\leq 50^{\text{th}}$ centile for 'lesion/age at surgery group'

ICU – intensive care unit; LOS – length of stay.

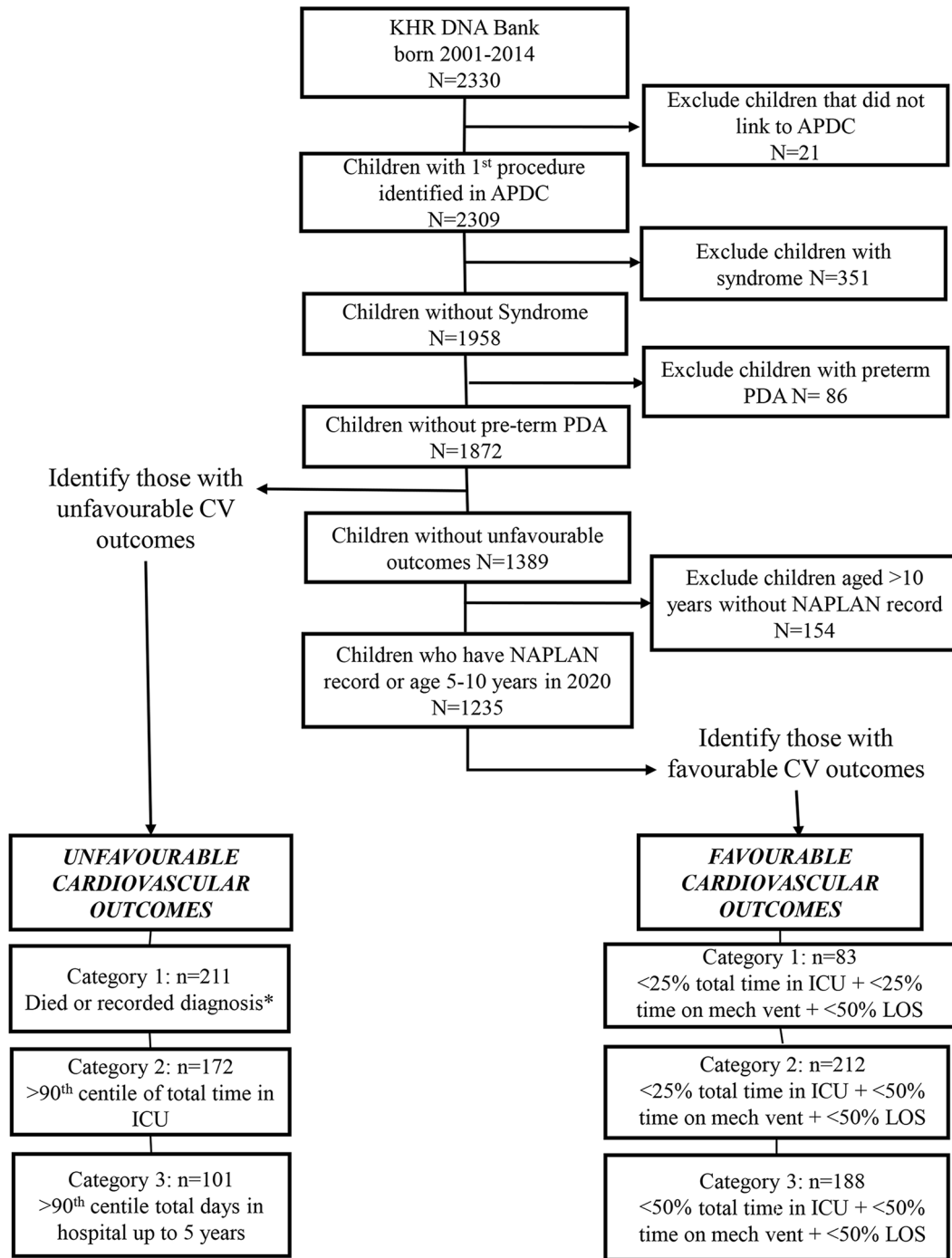
The number of children in each 'lesion/age at surgery' group that were classified into 'unfavourable' and 'favourable' outcome categories are shown in Supplementary Appendix 4. As 'unfavourable-category 1' is defined by specific diagnoses, the proportion of children in each 'lesion/age at surgery' group differs, with 21% of those in the 'lesion group A – neonatal surgery' group being classified in 'unfavourable-category 1' compared to only 5% of those in group 'lesion group B – childhood surgery'.

After excluding those with a 'unfavourable' outcome ($n = 484$) and those that did not link with a NAPLAN record or were exempt from sitting NAPLAN ($n = 154$), 1,235 children remained (Figure 1). To identify the required children with

'favourable' outcomes, those with total time in ICU less than the 50^{th} centile and time received mechanical ventilation less than 50^{th} centile for each 'lesion/age at surgery' group were identified ($n = 483$). These children were classified into 3 'favourable' outcome categories based on total time in ICU and total LOS in hospital (Table 3). Children with 'favourable' and 'unfavourable' outcomes were identified equally across the study period.

An audit of 200 randomly selected medical records resulted in the clinical researcher evenly classifying children across the 6 outcome categories. Records for the initial cardiac surgery for six of the 200 could not be found, and there were incomplete records for an additional two children. There were 10 records

Figure 1: Flow chart of study population



*see Table 1 for diagnoses and definitions of outcome groups

that were classified in discordant outcome categories between data linkage and the medical record audit. Three of these discrepancies were due to additional information recorded during readmissions to other hospitals identified by data linkage that could not be seen in the medical records of the one children's hospital. Six patients with septal defects were identified in linked data as 'unfavourable' but 'favourable' in medical records. This is due to the low ICU length of stay cut-offs required to identify 'unfavourable' outcomes for these groups. Ninety per cent of patients with septal defects

and surgery during childhood had an ICU stay less than 2 days.

There was one discordant record identified in linked data as 'favourable' but 'unfavourable' in medical records. This patient had total ICU stay and total hospital LOS just below the 50th centile for their lesion/age at surgery group. Excluding records that could not be found, 95% (182/192) of linked data outcome groups matched the outcome groups classified using clinical judgment and review based on information in the medical records.

Table 3: Total time in intensive care unit and total length of stay in hospital up to five years following primary cardiac surgery by lesion/age at surgery groups

CHD lesion Group	Age at surgery*	n	Total time in ICU (hours) median [IQR]	90 th centile time in ICU cut-off (days)	Total LOS in hospital (days) median [IQR]	90 th centile total LOS in hospital cut-off (days)	Total time on mechanical ventilation (hours) median [IQR]
Lesion Group A	Neonatal	381	358 [223–592]	1020	33 [21–60]	90	120 [51–229]
	Infant	77	174 [113–314]	502	29 [18–45]	64	48 [0–140]
	Child	34	71 [47–124]	277	18 [9–29]	38	0 [0–52]
Lesion Group B	Neonatal	188	273 [173–461]	794	28 [19–40]	62	96 [28–178]
	Infant	350	91 [48–152]	291	11 [7–18]	30	34 [0–98]
	Child	186	26 [22–46]	72	6 [5–7]	11	0 [0–0]
Lesion Group C	Neonatal	38	326 [145–695]	1928	26 [21–43]	84	91 [30–146]
	Infant	323	51 [31–108]	168	8 [6–12]	21	23 [0–50]
	Child	296	24 [21–27]	47	5 [4–6]	9	0 [0–0]

*Neonatal surgery = ≤ 30 days old, infant surgery = 31 days–1 year, child surgery = 1–5 years.

ICU = intensive care unit, LOS = length of stay, IQR = inter-quartile range.

Discussion

This study used record linkage between a clinical dataset of children who had surgery for CHD and routinely collected administrative data to identify cardiovascular outcomes. DNA samples for children identified with 'favourable' and 'unfavourable' cardiovascular outcomes will then be examined to understand if genetic differences can explain variability in outcomes following surgery. An audit of medical records of a random sample of children in the outcome groups found that record linkage is a reliable method to identify and classify cardiovascular outcomes, with 95% of outcome groups matching with groups identified using clinical judgement and information from medical records.

The linkage of routinely collected administrative data to identify outcomes for clinical trials has been used for the last two decades [11], however linking administrative data to inform studies examining genetic variability in outcomes is novel. There are many advantages to using linked administrative data to identify outcomes compared to other traditional outcome collection methods such as patient interview or medical record review. As administrative data is collected for other purposes, this readily available data source is a cost and time-efficient method of outcome collection [11]. Data linkage minimises loss to follow-up as patients can be identified if they have moved address. A recent study attempting to contact and survey young adults by mail who were diagnosed with childhood Type 1 diabetes had a 4% response rate [12]. Other biases are also minimised, such as recall bias from patient surveys or observer bias associated with outcome ascertainment as outcome collection is independent from treatment allocation.

Data linkage of administrative data can also identify wider health service use compared to accessing patient records available in one hospital. Although children in this cohort generally access one main children's hospital for cardiac care, they can attend other hospitals for additional health issues, such as the child admitted to another hospital for seizures, which would not have been identified without data linkage. As administrative hospital discharge data is population-based,

any hospital admission in the state of NSW is captured. Longitudinal data linkage can also identify short and long-term outcomes.

The range of routinely collected administrative data available for linkage also broadens the scope of outcomes that can be examined beyond health outcomes, such as education and development [13]. The linkage of numerous databases has many advantages when examining a child's post-surgery follow-up. In defining 'favourable' outcomes we needed to ensure that a child had not left the jurisdiction and had a 'unfavourable' cardiovascular outcome in another state or country. A whole-population dataset, such as NAPLAN where every child attending school in NSW has a NAPLAN record, can ensure that the child is present in the state in the years following their surgery. Furthermore, the addition of multiple datasets may also improve the identification of outcomes. Hospital admission data can under-report mortality from 10% [14] to 20% [15]. The linkage of mortality data from a death registry ensured that we identified all children in the cohort who have died.

The use of administrative data to identify cardiovascular outcomes also has disadvantages. Hospital discharge data does not collect detailed clinical data that may be required. Postsurgical details of cardiovascular outcomes, such as adequacy of cardiac output, were not available in the linked hospital data. However low cardiac output syndrome and other postsurgical cardiovascular outcomes are strongly associated with increased ICU stay which is reliably collected in the administrative dataset [9]. Age at surgery and complexity of surgical procedure are also predictors of length of time in ICU [9], however we have examined outcomes in lesion/age at surgery groups to adjust for this.

The quality of the data recorded in administrative data can also be varied. Specific diagnoses need to be explicitly stated in the medical records to be coded as coders are not clinicians and rely on consistent documentation [16]. Validation studies of serious cardiovascular events recorded in administrative data versus medical records has shown good positive predictive value: stroke, 91–97% [17], cardiac arrest, 92.5% [18] and seizures 64%–97% [19]. High positive

predictive value indicates a high proportion of routinely collected data that were true positives. Aside from using diagnosis codes to identify serious cardiovascular events, it was decided not to use specific diagnosis codes to identify other cardiovascular complications. Instead, we calculated total length of hospital stay, time on mechanical ventilation and time in ICU as proxies for unfavourable cardiovascular outcomes. In NSW, the main reasons for collecting hospital discharge data is to plan health services, track indicators of health status, and monitor the utilisation of hospital services [20], and as a result, variables that identify service utilisation are well collected. Total length of stay and time in ICU are well collected in administrative data with 97–99% of admission and discharge dates accurate within 1 day [21, 22]. The audit performed on a random sample of records in this study support this high level of accuracy, with 95% of records matching to medical records.

We found there was little variability in outcomes within some of the lesion/age groups. For example, 90% of patients with septal defects who had surgery during childhood had less than 2 days in ICU. For these children, establishing a cut-off of time in ICU to classify children as having ‘unfavourable’ outcomes was difficult, and as seen in the validation study, did not always correspond with direct assessment of the medical record. Later outcomes may be more valuable for certain lesion/age groups where duration of ICU stay is non-discriminatory.

The use of linked data to identify outcomes is time and cost-efficient as study participants do not need to be contacted to collect outcomes. However privacy of the participants must be protected, so data-separation principle was up-held, whereby no person had access to identifying information and linked administrative data [23]. All linked data were de-identified and to re-identify participants in the outcome categories for the medical record review, data was sent via the data linkage unit (CHeReL) for re-identification and the researcher performing the record review was blinded to outcome categories. When the DNA samples are retrieved, they will be sent for sequencing and analysed without any identifying information, only with outcome categories to ensure the data-separation principle is up-held.

Previous studies that have focused on the relationship between genetic variation and clinical outcomes amongst people with CHD have not used data linkage with administrative datasets. These include studies addressing clinical associations with copy number variation to 3 years post operatively [2], associations between damaging de novo variants and ventricular function [24], association between polygenic risk scores associated with higher diastolic pressure and post-procedural outcomes, and the relationship between de novo variants and post procedural clinical outcome with median follow-up times of 2.65 years [25]. In our study, we examined outcomes within 5 years of primary cardiac surgery to identify conditions that developed following post-surgical discharge, however additional linkage can identify outcomes into adulthood.

Conclusion

This study has shown that the linkage of a clinical and biobank dataset with routinely collected administrative data is a reliable method to identify outcomes to facilitate a large-scale study of the relationship between genetic architecture and clinical outcome. Cardiovascular outcomes following cardiac surgery are impacted by lesion type and severity, but even within the same lesion and repair there is individual variability in outcomes following surgery. In the future, identification of genetic hallmarks could be used to identify patients who are at risk of unfavourable cardiovascular outcomes, to inform strategies for prevention and modification of clinical care.

Acknowledgements

This study and investigators have been supported as part of a National Health and Medical Research Committee (NMHRC) Synergy Grant (APP1181325) and New South Wales Ministry of Health-funded Luminesce Alliance. SLD is supported by NMHRC Principal Research Fellowship ID1135886 and Leadership Fellowship ID2007896. NN is supported by NMHRC Investigator grant (APP1197940) and Financial Markets Foundation for Children.

We would like to thank the NSW Ministry of Health for access to the population health data used in this study, and the NSW Centre for Health Record Linkage for linking the datasets.

Statement of conflicts of interest

All authors have no conflicts of interest.

Ethics statement

Ethics approval for the study was attained from the NSW Population and Health Services Research Ethics Committee (2019/ETH11615).

References

1. Australian Institute of Health and Welfare. Congenital heart disease in Australia. In: AIHW, editor. Canberra <https://www.aihw.gov.au/reports/heart-stroke-vascular-diseases/congenital-heart-disease-in-australia/summary> Accessed: 20/9/2023; Cat. no. CDK 14.; 2019.
2. Kim DS, Kim JH, Burt AA, Crosslin DR, Burnham N, Kim CE, et al. Burden of potentially pathologic copy number variants is higher in children with isolated congenital heart disease and significantly impairs covariate-adjusted transplant-free survival. *Jnl Thor Card Surg.* 2016;151(4):1147–51. e4. <https://doi.org/10.1016/j.jtcvs.2015.09.136>
3. Kulik TJ, Sleeper LA, VanderPluym C, Sanders SP. Systemic ventricular dysfunction between stage one and stage two palliation. *Pediatric Cardiology.* 2018;39:1514–22. <https://doi.org/10.1007/s00246-018-1923-7>

4. Salve GG, Betts KS, Ayer JG, Chard RB, Nicholson IA, Orr Y, et al., editors. A simplified approach to predicting reintervention in the arterial switch operation. *Sem in Thor Cardio Surg*; 2022: Elsevier. <https://doi.org/10.1053/j.semtcvs.2021.04.058>
5. Wilde AA, Semsarian C, Márquez MF, Shamloo AS, Ackerman MJ, Ashley EA, et al. European heart rhythm association (EHRA)/Heart Rhythm Society (HRS)/Asia Pacific Heart Rhythm Society (APHRS)/Latin American Heart Rhythm Society (LAHRS) expert consensus statement on the state of genetic testing for cardiac diseases. *Heart rhythm*. 2022. <https://doi.org/10.1016/j.hrthm.2022.03.1225>
6. Fudge Jr JC, Li S, Jaggars J, O'Brien SM, Peterson ED, Jacobs JP, et al. Congenital heart surgery outcomes in Down syndrome: analysis of a national clinical database. *Peds*. 2010;126(2):315–22. <https://doi.org/10.1542/peds.2009-3245>
7. Borthwick A, Buechi M, Goldberg A. Key Concepts of the ChoiceMaker 2 Record Matching System. *Proceedings of the KDD-03 Workshop on Data Cleaning, Record Linkage, and Object Consolidation*; Washington DC2003.
8. He WQ, Nassar N, Schneuer FJ, Lain SJ, group CHDSS, Dunwoodie SL, et al. Examination of validity of identifying congenital heart disease from hospital discharge data without a gold standard: Using a data linkage approach. *Paediatric and Perinatal Epidemiology*. 2023. <https://doi.org/10.1111/ppe.12976>
9. Pagowska-Klimek I, Pychynska-Pokorska M, Krajewski W, Moll JJ. Predictors of long intensive care unit stay following cardiac surgery in children. *Eur Jnl of Card-thor Surg*. 2011;40(1):179–84. <https://doi.org/10.1016/j.ejcts.2010.11.038>
10. Verrall CE, Blue GM, Loughran-Fowlds A, Kasparian N, Gecz J, Walker K, et al. 'Big issues' in neurodevelopment for children and adults with congenital heart disease. *Open Heart*. 2019;6(2):e000998. <https://doi.org/10.1136/openhrt-2018-000998>
11. Mc Cord KA, Ewald H, Agarwal A, Glinz D, Aghlmandi S, Ioannidis JP, et al. Treatment effects in randomised trials using routinely collected data for outcome assessment versus traditional trials: meta-research study. *BMJ*. 2021;372. <https://doi.org/10.1136/bmj.n450>
12. James S, Pryke A, Cusumano J, Jenkins A, Benitez-Aguirre P, Craig M, et al. 20-Year Outcomes of Childhood-Onset Type 1 Diabetes: the CANDID Incident Cohort Survey. *Diab Med*. 2020:e14473-e. <https://doi.org/10.1111/dme.14473>
13. Lawley CM, Winlaw DS, Sholler GF, Martin A, Badawi N, Walker K, et al. School-age developmental and educational outcomes following cardiac procedures in the first year of life: A Population-Based Record Linkage Study. *Ped Card*. 2019;40(3):570–9. <https://doi.org/10.1007/s00246-018-2029-y>
14. Maruszewski B, Lacour-Gayet F, Monro JL, Keogh BE, Tobota Z, Kansy A. An attempt at data verification in the EACTS Congenital Database. *Eur Jnl Card-thor Surg*. 2005;28(3):400–4. <https://doi.org/10.1016/j.ejcts.2005.03.051>
15. Gibbs JL, Monro JL, Cunningham D, Rickards A. Survival after surgery or therapeutic catheterisation for congenital heart disease in children in the United Kingdom: analysis of the central cardiac audit database for 2000-1. *BMJ*. 2004;328(7440):611. <https://doi.org/10.1136/bmj.38027.613403.F6>
16. Welke KF, Karamlou T, Diggs BS. Databases for assessing the outcomes of the treatment of patients with congenital and paediatric cardiac disease—a comparison of administrative and clinical data. *Card in the Yng*. 2008;18(S2):137–44. <https://doi.org/10.1017/S1047951108002837>
17. Porter J, Mondor L, Kapral MK, Fang J, Hall RE. How reliable are administrative data for capturing stroke patients and their care. *Cerebr Dis Extra*. 2016;6(3):96–106. <https://doi.org/10.1159/000449288>
18. Shelton SK, Chukwulebe SB, Gaieski DF, Abella BS, Carr BG, Perman SM. Validation of an ICD code for accurately identifying emergency department patients who suffer an out-of-hospital cardiac arrest. *Resus*. 2018;125:8–11. <https://doi.org/10.1016/j.resuscitation.2018.01.021>
19. Shui IM, Shi P, Dutta-Linn MM, Weintraub ES, Hambidge SJ, Nordin JD, et al. Predictive value of seizure ICD-9 codes for vaccine safety research. *Vacc*. 2009;27(39):5307–12. <https://doi.org/10.1016/j.vaccine.2009.06.092>
20. Australian Institute of Health and Welfare. National Hospitals Data Collection. Canberra; 2021. Accessed 5 May 2022.
21. Garland A, Yogendran M, Olafson K, Scales DC, McGowan K-L, Fransoo R. The accuracy of administrative data for identifying the presence and timing of admission to intensive care units in a Canadian province. *Med care*. 2012;50(3):e1–e6. <https://doi.org/10.1097/MLR.0b013e318245a754>
22. Scales DC, Guan J, Martin CM, Redelmeier DA. Administrative data accurately identified intensive care unit admissions in Ontario. *Jnl Clin Epi*. 2006;59(8):802–7. <https://doi.org/10.1016/j.jclinepi.2005.11.015>
23. Kelman CW, Bass AJ, Holman CD. Research use of linked health data—a best practice protocol. *ANZ Jnl Pub Hlth*. 2002;26(3):251–5. <https://doi.org/10.1111/j.1467-842x.2002.tb00682.x>
24. Lewis MJ, Hsieh A, Qiao L, Tan R, Kazzi B, Channing A, et al. Association of Predicted Damaging De Novo Variants on Ventricular Function in Individuals With Congenital Heart Disease. *Circ: Gen Prec Med*. 2023:e003900. <https://doi.org/10.1161/CIRCGEN.122.003900>

25. Boskovski MT, Homsy J, Nathan M, Sleeper LA, Morton S, Manheimer KB, et al. De novo damaging variants, clinical phenotypes, and post-operative outcomes in congenital heart disease. *Circ: Gen Prec Med*. 2020;13(4):e002836. <https://doi.org/10.1161/CIRCGEN.122.003900>



Supplementary Appendix 1: ACHI codes for cardiac condition related procedure

Procedure descriptions	Procedure codes
Insertion, removal and replacement of a cardiac pacemaker or electrode	38253, 38256, 38259, 38278, 38281, 38350, 38353
Procedure on the atrium, cardiac vessels and aorta	38456, 38457, 38458, 38460, 38464, 38466, 38470, 38473, 38475, 38477, 38480, 38481, 38483, 38485, 38487-38490, 38493, 38497, 38500, 38503-38505, 38507-38509, 38512, 38515, 38518, 38521, 38524, 38530, 38533, 38550, 28553, 38556, 38559, 38562, 38565, 38568, 38571
Procedures on the ventricle	38615, 38618, 38621, 38624, 38627-00
Procedures on the myocardium	22075, 38418, 38275, 38577, 38588, 38560, 38650, 90206
Closure of cardiac collateral vessels	38700.00-38700.03
Percutaneous interventions including angioplasty, atrial septostomy, balloon valvuloplasty, baffle and conduit procedures and procedures on the pulmonary and coronary arteries	38270, 38300, 38303, 38306, 38309, 38312, 38315, 38318, 38637, 38706, 38715, 38721, 38727, 38733, 38739, 38742, 38745, 38748, 38751, 38754, 38757, 38760, 38763, 38766
Cardiopulmonary bypass	38588.00, 38600.00, 38603.00, 90225.00, 38627.02, 38653

Supplementary Appendix 2: Severe cardiac-related outcomes and associated diagnosis codes recorded in admitted patient data collection

Description	ICD10-AM Diagnosis code
Cardiac arrest	I46
Intracerebral haemorrhage	I61
Other no traumatic intracranial haemorrhage	I62
Cerebral infarction	I63
Stroke	I64
Epileptic seizures	G40
Convulsions	R56



Supplementary Appendix 3: Medical record audit

Powerchart was used to review 200 randomly selected patient medical records to determine whether they had ‘unfavourable’ or ‘favourable’ outcomes. All cases were assessed in the context of the specific lesion, typical intervention and typical length of stay for that lesion. For example, a single ventricle pathway usually requires up to three staged procedures and consequently a longer ICU and hospital stay than a patient requiring an ASD closure.

The clinical variables reviewed included:

- Whether the patient was deceased or alive
- Any reported complications either during the operation or post-operatively
- Length of ICU stay
- Total hospital length of stay
- Any readmission to ICU within the same hospital admission
- Multiple admissions to hospital for cardiac interventions (atypical for the lesion type)
- Type of procedure completed (surgery vs cardiac catheterization)
- Presence of comorbidities

Various documents were reviewed to assess the above clinical variables. Discharge summaries provided details relating to the total length of stay, reason for visit, how many separate visits an individual has had and details of any major operative/post-operative complications. This information provided an overview and guided the more detailed review of operation reports, nursing notes, ICU discharge summaries and any other available documentation to identify additional information.

Criteria for a ‘favourable’ outcome classification included:

- A shorter than average ICU stay relative to the lesion

- Discharge from hospital in good condition and in a shorter time frame than average (for specific lesion)
- No or minor operative or post-operative complications
- No cardiac surgery re-interventions

Criteria for a ‘unfavourable’ outcome classification included:

- Patient deceased
- Major operative or post-operative complications (*as per list below)
- Longer than average ICU admission relative to lesion
- Longer than average total hospital stay relative to lesion

*Major operative and early post-operative complications:

- Excessive bleeding
- Re-intervention the same day as the operation
- CPR or unexpected initiation of cardiovascular support measures during ICU stay
- Extracorporeal membrane oxygenation (ECMO)
- Re-admission to ICU within the same hospital admissionArrhythmia disrupting cardiovascular stability
- Major organ damage (e.g brain, renal, hepatic) requirement targeted assessment and management
- Sepsis/infection

NOTE: Only patient medical records at The Children’s Hospital at Westmead were reviewed, health data recorded elsewhere was not accessible. All criteria were assessed within the context of patient-specific lesions. Re-interventions included any additional treatments/interventions required as a result of a complication. When assessing length of stay, classification was determined based on the ‘typical’ length of stay of the other participants with the same lesion type.



Supplementary Appendix 4: Numbers in each 'unfavourable' and 'favourable' cardiovascular outcome category

Lesion Group	Age at surgery*	n	Unfav – cat 1 n (row %)	Unfav – cat 2 n (row %)	Unfav – cat 3 n (row %)	Total Un-favourable	Fav – cat 1 n (row %)	Fav – cat 2 n (row %)	Fav – cat 3 n (row %)	Total favourable
Group A	Neonatal	381	80 (21%)	27 (7%)	–	107 (28%)	11 (3%)	40 (10%)	49 (13%)	100 (26%)
	Infant	77	9 (12%)	7 (9%)	*	21 (27%)	*	*	13 (17%)	19 (25%)
	Child	34	*	*	*	8 (24%)	*	*	*	8 (24%)
Group B	Neonatal	188	33 (18%)	16 (8%)	12 (6%)	61 (32%)	6 (3%)	12 (6%)	33 (18%)	51 (27%)
	Infant	350	43 (12%)	30 (9%)	20 (6%)	93 (27%)	23 (7%)	30 (9%)	27 (8%)	80 (23%)
	Child	186	9 (5%)	17 (9%)	14 (7%)	40 (21%)	8 (4%)	32 (17%)	14 (7%)	54 (29%)
Group C	Neonatal	38	*	*	*	11 (29%)	*	7 (18%)	*	11 (29%)
	Infant	323	16 (12%)	37 (11%)	28 (9%)	81 (25%)	18 (6%)	34 (10%)	35 (11%)	87 (27%)
	Child	296	12 (11%)	32 (10%)	18 (6%)	62 (21%)	11 (4%)	50 (17%)	12 (4%)	73 (25%)
TOTAL		1872	211	172	101	484	83	212	188	483

*cell size = 5 or less.

