CJC Pediatric and Congenital Heart Disease 3 (2024) 57–66

Original Article

The Quebec Congenital Heart Disease Registry: A Model of Prospective Databank to Facilitate Research in Congenital Cardiology

Laurence Watelle, MD,[a] Louis-Olivier Roy, BSc,[a] Jonathan Lauzon-Schnitka,[a] Garrett Newell, MD,[a] Anne Dumas, MD,[a] Alexandre Nadeau, BSc,[a] Wei Ting Xiong, MD,[a] Kevin Rego,[a] Camille Beaulieu,[a] Emilie Groulx-Boivin,[b] Marie-Ève Roy-Lacroix, MD,[c] Laurence Vaujois, MD,[d] Christian Drolet, MD,[d] Adrian Dancea, MD,[b] Nagib Dahdah, MD,[e] Jean-Luc Bigras, MD,[e] and Frédéric Dallaire, MD, PhD;[a] on behalf of the Canadian Congenital and Pediatric Cardiology Research Network (CCPCRN) investigators

[a] *Department of Pediatrics, Faculty of Medicine and Health Sciences, Université de Sherbrooke, and Centre de recherche du Centre Hospitalier Universitaire de Sherbrooke, Sherbrooke, Québec, Canada*

[b] *Division of Cardiology, Montreal Children's Hospital, McGill University, Montreal, Québec, Canada*

[c] *Department of Obstetrics and Gynecology, Faculty of Medicine and Health Sciences, Université de Sherbrooke, and Centre de recherche du Centre Hospitalier Universitaire de Sherbrooke, Sherbrooke, Québec, Canada*

[d] *Department of Pediatrics, CHU de Québec, Université Laval, Québec, Québec, Canada*

[e] *Division of Pediatric Cardiology, CHU Ste-Justine, Université de Montréal, Montréal, Québec, Canada*

ABSTRACT

Background: A national registry of congenital heart disease (CHD) would facilitate project initiation, decrease costs, increase statistical power, and avoid duplication. Establishing such registries poses numerous challenges, but the current Canadian research ecosystem in CHD is well positioned to meet them. We assessed the feasibility of building a province-wide CHD registry by automatically identifying people with CHD and extracting their native cardiac anatomy from multiple clinical data sources, without the need for manual data entry.

Methods: We designed a CHD registry of all fetuses and children with at least 1 echocardiographic report confirming CHD since 2000. We interfaced the registry with several clinical and echocardiography data sources from all paediatric cardiology programmes in Québec.

RÉSUMÉ

Contexte : Un registre national des cardiopathies congénitales (CC) pourrait faciliter le lancement de projets de recherche, en diminuer les coûts, en améliorer la puissance statistique tout en évitant les redondances. La mise en place de tels registres pose de nombreux défis, mais l'écosystème de recherche canadien dans le domaine de la CC est bien placé pour y répondre. Nous avons évalué la faisabilité de la mise en place d'un registre des CC à l'échelle provinciale par l'identification automatique des personnes atteintes de CC et l'extraction de leur anatomie cardiaque native à partir de plusieurs sources de données cliniques, sans nécessiter de saisie manuelle de données.

Méthodologie : Nous avons conçu un registre des CC incluant tous les fœtus et les enfants pour qui au moins un rapport d'évaluation électrocardiographique confirmait la présence d'une CC depuis 2000.

Investigators studying congenital heart diseases (CHD) are faced with significant challenges. There are many forms of CHD, and each is heterogeneous in its presentation and complications. This heterogeneity decreases the number of cases that can be included at any given centre, therefore reducing statistical power. Cardiac lesions in repaired and unrepaired CHD can have insidious chronic effects, the consequences of which can manifest after many years, even decades.[1-3] Finally, the so-called hard outcomes often used in cardiology research (mortality, transplantation, malignant arrhythmia, and other major adverse cardiac events) often have low incidence,[3] which reduces result precision and lowers the probability of observing unequivocal signals.

Received for publication November 19, 2023. Accepted December 12, 2023.

Corresponding author: Dr Frédéric Dallaire, University of Sherbrooke, 3001, 12e Ave Nord, Sherbrooke, Québec J1H 5N4, Canada, Tel.: +1-819-346-1110.

E-mail: frederic.a.dallaire@usherbrooke.ca

https://doi.org/10.1016/j.cjcpc.2023.12.001

Results: We extracted 885,287 echocardiogram reports and 70,121 clinical records. We identified CHD in 43,452 children and 4682 fetuses. There were 1128 (2.3%) cases with files in multiple institutions, and patients with more complex CHD were 3 times more likely to be seen in more than 1 institution. So far, the registry has been used to build and link CHD cohorts for 7 distinct projects.

Conclusions: We demonstrated the feasibility of a baseline CHD registry in Québec without the need for manual data entry, in which other CHD research projects could be nested. This could serve as a blueprint to expand the registry and to develop an integrated approach where data gathered in caring for patients with CHD serve as data layers that incrementally contribute to a national cohort, for which data remain easily accessible and usable.

Le registre a été mis en relation avec plusieurs sources de données cliniques et échocardiographiques provenant de tous les programmes en cardiologie pédiatrique au Québec.

Résultats : Nous avons extrait 885 287 rapports d'échocardiographie et 70 121 dossiers cliniques. La présence d'une CC a été établie chez 43 452 enfants et 4 682 fœtus. Dans 1 128 cas (2,3 %), un dossier existait dans plus d'un établissement. Les patients présentant des CC plus complexes étaient 3 fois plus susceptibles d'être suivis dans plus d'un établissement. Jusqu'à présent, le registre a été utilisé pour établir et mettre en relation des cohortes de patients atteints de CC pour sept projets de recherche distincts.

Conclusions : Nous avons démontré la faisabilité de la mise en place d'un registre de référence des CC au Québec sans recours à la saisie manuelle de données, dans lequel peuvent se nicher d'autres projets de recherche sur les CC. Notre démarche pourrait servir de prototype pour une expansion du registre et pour une approche d'intégration des données recueillies dans la prestation de soins aux patients atteints de CC, afin de former des couches de données qui s'ajoutent au fur et à mesure à une cohorte nationale, avec des données faciles à obtenir et à utiliser.

Meeting these challenges requires innovation in study design, data collection, and analysis to improve sample size, population homogeneity, duration of follow-up, data quality, and, ultimately, internal and external validity. We envisioned that the ideal cohort of patients with CHD would be populational (not only a selected sample), longitudinal (significant events are continuously captured throughout the lifespan), well phenotyped (there is an accurate description of anatomy, risk factors, and interventions), inclusive (atypical malformations are well documented rather than excluded), and accessible (barriers to accessing, using, and exploiting the data from the cohort should be minimized).

If all these characteristics are desirable, they seldom are reunited, as investigators often compromise on one to get the other. Registries of surgeries and interventions provide us with high-quality data on anatomy and interventions[4-6] but must often compromise on follow-up duration. Very well-phenotyped cohorts with longer follow-up have often had to focus on specific types of CHD.[7,8] National registries based on health system data are informative at the population level,[9,10] but often have to compromise on data granularity and accuracy.

There is a need for a framework to build a durable and valuable cohort of patients with CHD in Canada. Attempting to manually collect granular data on all patients with CHD across the lifespan is desirable but would require unrealistic amounts of resources and commitment. Rather, we propose that a cohort with CHD should be built incrementally by gradually incorporating new layers of data accumulated over the course of research projects on specific topics and CHD lesions. As a proof of concept, we report here the feasibility of a province-wide clinical CHD registry that feeds automatically on several data sources from all paediatric cardiology divisions in Québec. This provides an exhaustive list of patients that serves as a foundation on top of which additional research data can be layered and linked. We then extrapolate on how this approach could be expanded nationally.

Methods

Objectives

The objectives of this study were: (1) to build the foundation of a province-wide CHD registry by automatically identifying people with CHD and extracting their native cardiac anatomy from multiple clinical data sources; (2) to test automatization of data harmonization, specifically on the classification of CHD diagnosis; (3) to assess the feasibility of performing annual data updates; (4) to find ways of handling data from patients receiving care in multiple institutions through accurate indexing of patients; and (5) to promote and facilitate data sharing, availability, and interoperability.

Study design and population

This is a retrospective and prospective registry of fetuses and children with confirmed CHD in Québec, with retrospective inclusion from 2000 to 2017, and prospective data collection thereafter with periodic updates starting in 2019. The participating institutions were the Centre mère-enfant soleil du Centre Hospitalier Universitaire de l'Université Laval, the Centre Hospitalier Universitaire de Sherbrooke, the Montreal Children's Hospital, and the Centre Hospitalier Universitaire Ste-Justine. The targeted population included all patients and fetuses with any form of CHD seen at one of these sites during the study period. The main data sources were the raw clinical and echocardiography data from all 4 paediatric cardiology divisions in Québec.

Development of the databank

The registry was designed as a databank that could be used as a foundation for CHD research. The databank is housed in the data centre of the Canadian Congenital and Pediatric Cardiology Research Network (CCPCRN).[11,12] The CCPCRN provides a management framework that specifies processes for scientific review, data access, data storage, and

Table 1. Examples of cross-match tables

Final registry code (ICD-11)	Montreal Children Hospital	Centre mère-enfant soleil du CHUL	CHU Sainte-Justine (2000-2015 echo system)	CHU Sainte-Justine (current echo system [from 2015])
01.01.03 (congenitally corrected transposition of great arteries)	01.01.03 (L-TGA-corrected transposition)	LTGV (transposition L corrigée G.V.)	1405 (L-transposition des gros vaisseaux) 835CR (S-L-L) CH119 (transposition corrigée anatomiquement)	01.01.03 (transposition des gros vaisseaux anatomiquement corrigée/L-TGV)
04.07.01 (partial anomalous pulmonary venous connection(s))	04.07.01 (PAPVC [partial anomalous]) 04.07.10 (partial anomalous pulmonary venous connection)	RVPP (retour veineux pulm. anorm. part.)	13 (anom. part. du retour veineux pulmonaire dans l'OD) 17 (anom. part. du ret. vein. pulm. ds le sinus coronaire) CH30 (anomalie partielle du retour veineux pulmonaire)	04.07.10 (partial anomalous pulmonary venous connection) 04.07.01 PAPVC (partial anomalous) 04.07.04 and 04.07.05 (retour veineux pulmonaire/anomalie partielle)

In the body of the table, we voluntarily reproduced the many acronyms and the bilingual nature of the data as examples of what was found in source data.
CHU, Centre Hospitalier Universitaire; CHUL, Centre Mère-enfant Soleil du Centre Hospitalier Universitaire de l'Université Laval; ICD-11, International Classification of Disease, 11th edition.

data ownership. The CCPCRN data centre was responsible for data management, harmonization, access, and storage.

Face-to-face meetings with investigators from participating centres took place in 2017 to determine the key baseline variables that should be part of the registry. It was decided that the baseline data to be extracted would populate data fields on demographics (names, date of birth, health insurance number, and mother-child link in cases of fetuses), native cardiac anatomy with up to 20 diagnostic codes, percutaneous and surgical cardiovascular interventions, tracking information on the data sources used to populate the fields, and links with current, past, or future research databases on specific CHD lesions.

Algorithm for classification of the native cardiac anatomy and procedures

We developed interfaces between the registry and the various clinical and echocardiography data sources of each centre. We designed scripts that automatically converted local diagnostic coding systems used by the participating institutions into a uniform and interoperable nomenclature.

The native cardiac anatomy was classified using the codes proposed by the International Society for Nomenclature of Paediatric and Congenital Heart Disease. This is a set of 318 codes developed by CHD nomenclature experts set to serve as the foundation for the 11th version of the International Classification of Disease.[13] This nomenclature has been proposed to serve as a basis for CHD research in Canada.[14] Each patient could be assigned up to 20 specific codes, which could then be distilled down to a primary CHD category if needed (see Supplemental Table S1 for the complete list of diagnostic codes with the associated CHD category).

Cross-map tables and scripts were developed for each source dataset to format the extracted data and make them compatible with the registry. Examples of cross-match tables are presented in Table 1. The participating institutions either had an in-house coding system or used some version of the International Paediatric and Congenital Cardiac Code coding systems.[13] For one institution, individual echocardiography reports were not searchable, but a structured clinical database with house CHD codes and dates of follow-up at the patient level was available. For these, comma separated variable structured datasets were exported from the sources systems. For 2 institutions, there was a period during which they did not use a structured coding system. For these, free-text echocardiography reports were exported, and a natural language algorithm was developed to extract keywords and match them to precise diagnostic codes. Once exported locally, each dataset containing the relevant data elements was encrypted and securely sent to the CCPCRN data centre for treatment and analysis.

The cross-map tables and scripts were applied to a selection of random data (200 subjects per centre), and manual comparisons were made. Discrepancies were then used to refine the algorithms and cross-map tables. Two to three iterations of this process were made before the final scripts were applied to all patients.

Retrospective data extraction up to 2017 was done in 2018. Updates were then obtained regularly, and the algorithms were applied to newly obtained data. Data quality control was done by manually comparing automatically classified data with the original data for a random sample of 5% of patients in each centre.

The Society of Thoracic Surgeons procedural short list and long list were used as the basis to categorize surgeries and procedures. The development scripts to harvest surgical and percutaneous procedures, and the testing of these scripts and cross-map tables is ongoing (data not shown).

Incremental data collection

At each annual update, newly extracted data are compared with data held in the registry. For patients already included in the registry, new diagnostic codes or interventions are added when needed. We collected information on the data sources interrogated and number of occurrences of each diagnostic code for each source of data. The number of occurrences for a specific code for a given patient can be used as a metric of accuracy of a diagnosis.

Patient indexing, linkage, and universal unique identifiers

We developed an algorithm to compare patients' identifiers to avoid duplication of a patient treated at multiple sites or of patients with name changes, spelling errors, and data entry errors across data sources. Deterministic matching was used for unique identifiers (provincial health care number and medical chart number), and probabilistic matching was used for names and date of birth. In 2019, the protocol was amended to enable linkage of patient identifiers to test the feasibility of assigning a universal unique identifier (UUID) to patients in the registry.[15] This UUID allows patient tracking across multiple projects without sharing identifiers.

Ethical considerations

The registry was approved by the research ethics board of the Centre de Recherche du Centre Hospitalier Universitaire de Sherbrooke, with delegated approval by all participating institutions. Based on the favourable balance between the risk of a breach of confidentiality and the potential benefits for patients care through increased knowledge, a waiver of consent was granted, conditionally to the possibility of opting out. We have put in place a simple online opt-out mechanism. Because most patients have regular paediatric cardiology follow-up, we display posters in waiting rooms to inform patients of the existence of the registry and on what to do to opt out.

Special permission by the research ethics boards was given to obtain and store patient identifiers to assess the feasibility and accuracy of patient indexing. Patients' identifiers are used for indexing of patients during the maintenance and updates of the registry. A unique code links patients' identifiers with diagnostic and clinical data. Although kept on the same server, identifier data are kept in a separate form with a dedicated access level. Registry updates are executed within the secure informational structure of the CCPCRN data centre. Identifiers' data may be exported for linkages with administrative databases. In these cases, the identifiers' data are stripped from diagnostic data, encrypted, and sent via secure servers.

Analysis

We used the SAS software (Cary, NC) for scripts, data manipulation, data management, and statistical analysis. Some scripts were done using Python. Data access and interrogation of the registry data were managed with the REDCap electronic data capture tools[16] hosted at the CCPCRN data centre. Descriptive statistics were used to report the registry numbers. As there was no need for hypothesis testing, no statistical tests were performed.

Results

Content of the registry and number of patients

As of June 2023, we interrogated and analysed 885,287 echocardiography reports from the 3 institutions where fetal and paediatric echocardiography reports were available. In addition, we searched 70,121 clinical records in the institution with a clinical database to identify eligible patients. Our algorithms identified 49,868 reports or clinical data entries with at least 1 CHD diagnosis. Of them, we excluded 1734 patients with a single occurrence of isolated tricuspid or pulmonary valve regurgitation, leaving 48,134 patients for analysis (43,452 children and 4682 fetuses).

Our approach was purposefully very broad in including patients in the registry, and we expected to also harvest data on children with cardiac lesions of little consequence. In our dataset, 9980 of 43,452 (23.0%) children had only 1 occurrence of a single diagnostic code on echocardiography, without evidence of follow-up echocardiography. Otherwise, 21,703 of 43,452 (49.9%) children and 925 of 4682 (19.8%) fetuses had more than 1 diagnostic code, and 4280 of 43,452 (9.8%) children and 210 of 4682 (4.5%) fetuses had 5 or more diagnostic codes. A total of 25,650 of 43,452 (58.9%) children had more than 1 echocardiography (from echocardiography data sources) or more than 1 follow-up appointment (in clinical database), and 10,120 of 43,452 (23.3%) children had 5 or more years of follow-up. The first occurrence of a CHD code was recorded at a median age of 0.16 years (interquartile range: 0.02-9.3 years). For the most recent years when data were complete, between 1217 and 1847 new children and 349 to 373 new fetuses were included in the registry each year.

Type of CHD included

Table 2 lists the number of patients included according to the CHD category. More than half (24,302 of 43,452 [52.3%]) of children were included because of an isolated septal anomaly or an isolated pulmonary or aortic valve anomaly. As expected, more severe CHD were over-represented in fetuses compared with children. Despite being relatively rare, lesions such as functionally univentricular heart (1.0% of all CHD) were well represented and enabled identification of an interestingly high number of patients (eg, 444 cases of single ventricles).

Indexing of patients

There were 1128 cases with files in multiple institutions (1100 of 43,452 [2.5%] children and 28 of 4682 [0.6%] fetuses). As expected, there was an over-representation of more complex cardiopathies in the subgroup followed in more than 1 centre. CHD categories such as functionally univentricular heart, pulmonary atresia, double outlet right ventricle, common arterial trunk, and tetralogy of Fallot were 3 times more

Table 2. List of the number of patients according to CHD categories

CHD category	Children Number (%)	Fetuses Number (%)
Congenital anomaly of the ventricular septum	9587 (22.1)	1105 (23.6)
Congenital anomaly of the aortic valve (including sub- and supravalval stenosis)	5716 (13.2)	295 (6.3)
Congenital anomaly of atria and/or atrial septum	4847 (11.2)	25 (0.5)
Congenital anomaly of pulmonary valve (including sub- and supravalvlar stenosis)	4152 (9.6)	202 (4.3)
Congenital arterial duct (ductus arteriosus) anomaly	3880 (8.9)	46 (1.0)
Congenital anomaly of mitral valve	2253 (5.2)	45 (1.0)
Congenital anomaly of aorta and/or its branches (excluding arch obstruction)	1894 (4.4)	240 (5.1)
Other congenital left ventricular anomalies	1452 (3.3)	29 (0.6)
Congenital anomaly of aortic arch	1368 (3.1)	502 (10.7)
Congenital anomaly of tricuspid valve	1359 (3.1)	79 (1.7)
Tetralogy of Fallot	1174 (2.7)	282 (6.0)
Congenital anomaly of pulmonary arterial tree	1084 (2.5)	53 (1.1)
Common atrioventricular junction (common atrioventricular canal)	870 (2.0)	255 (5.4)
Transposition of the great arteries (discordant ventriculoarterial connections)	663 (1.5)	191 (4.1)
Congenital anomaly of position and spatial relationships of thoracoabdominal organs	446 (1.0)	99 (2.1)
Functionally univentricular heart	444 (1.0)	404 (8.6)
Other congenital right ventricular anomalies	442 (1.0)	30 (0.6)
Congenital anomaly of pulmonary vein(s)	360 (0.8)	38 (0.8)
Congenital anomaly of mediastinal vein	320 (0.7)	284 (6.1)
Congenital anomaly of coronary arteries	273 (0.6)	16 (0.3)
Double outlet right ventricle	262 (0.6)	195 (4.2)
Congenital pulmonary atresia	220 (0.5)	99 (2.1)
Common arterial trunk (truncus arteriosus)	80 (0.2)	56 (1.2)
Others	306 (0.7)	109 (2.3)

CHD, congenital heart disease.

likely to have been seen in more than 1 institution during childhood (data not shown). The proportion of children with a univentricular heart seen in multiple institutions was 34 of 444 (7.6%). For children with tetralogy of Fallot, it was 143 of 1174 (12.2%).

We identified 841 of 48,134 (1.7%) cases of patients with name changes or spelling errors. Of these, 72 cases were due to a temporary baby name born in one hospital and later transferred to another hospital. We were able to retrieve the unique health insurance number of 35,276 of 43,452 (81.2%) children and 4308 of 4682 (92.0%) pregnant women with fetuses with CHD, which will greatly ease the tracking of patients across institutions and facilitate linkage with administrative data.

Patient identification and linkage with CHD studies

Since its inception, we have successfully linked the patients in the registry with cohorts of patients with CHD that were constituted in parallel of the registry development. These

Table 3. Examples of research projects nested in or linked to the registry

Study and population	Linkage and progress
TRIVIA study. Population-based cohort of tetralogy of Fallot in Québec	Cohort retrospectively linked with the registry. Successful linkage of the TRIVIA cohort with administrative data for >93% of patients, which could be reused in the future if needed.
FREQUENCY study. Population-based cohort of mother-infant pairs for prenatal CHD screening	Fetuses with CHD included in the FREQUENCY study were retrospectively included in the registry. Second phase of the FREQUENCY study ongoing. The linkage of the registry data with administrative data will enable assessment of prenatal detection rates in Québec over >15 years.
Aortic dilatation in the bicuspid aortic valve study. Retrospective cohort of patients with bicuspid aortic valve to study the rate of aortic root and ascending aorta dilatation in children	Study completed. Patient identification done in part with the registry. Complete linkage with registry data.
Effect of medical treatment to prevent aortic dilatation in bicuspid aortic valve Retrospective observational study	Ongoing study (manuscript in preparation). Patient identification done with the registry for the Québec sites. Data for previous bicuspid aortic valve project linked and reused.
Paediatric outcomes of congenitally corrected transposition of the great arteries study Pan-Canadian retrospective observational cohort study	Ongoing study (analysis in progress). Patient identification done with the registry for the Québec sites.
CANFON Connection study. Pan-Canadian prospective observational cohort study of patients with singe ventricle and Fontan circulation	Ongoing study (enrolment). Patient identification done with the registry for the Québec sites. Registry data will help for linkage with administrative data.
QUALITY study. Online study collecting quality of life data testing recruitment strategy using social media	Ongoing study (enrolment). Participating patients are linked with the registry to determine their baseline cardiac anatomy. The online quality of life questionnaires developed and implemented in this study were reused in the CANFON study above.

CHD, congenital heart disease.

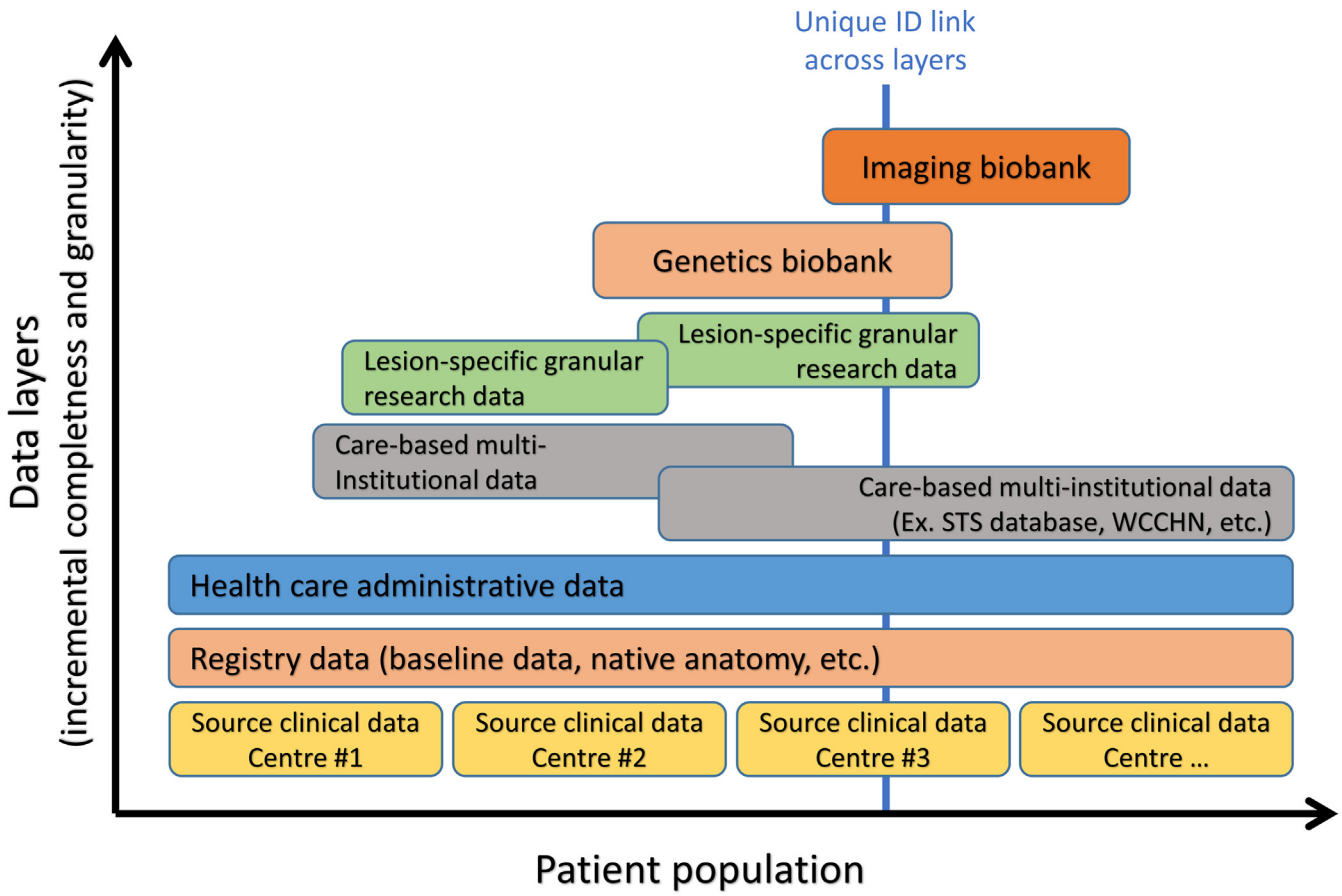

Figure 1. Data layer model. Each data source serves as a layer that increases data completeness. Registry data linked with clinical source data from each centre serve as the foundation of this model. A unique identifier, such as a universal unique identifier, links a patient's data across all layers. ID, identifier; STS, Society of Thoracic Surgeons; WCCHN, Western Canadian Children's Heart Network.

patients were included in the TRIVIA study (cohort of tetralogy of Fallot),[17-19] the FREQUENCY study (cohort of mother-infant pairs for prenatal CHD screening),[20,21] and a study on aortic dilatation in the context of a bicuspid aortic valve.[22] The registry has also served as the basis for patient identification and linkage for other ongoing studies (Table 3). In addition to rapid patient identification across the province, datasets for each of these studies can now be easily linked, which facilitate secondary uses of data. For example, the linkage of the previous study data on ascending aorta dilatation in patients with bicuspid valve[22] was used to automatically populate data fields for an ongoing retrospective cohort on the effect of medical therapy for slowing ascending aorta dilatation in patients with bicuspid aortic valve. Each of these linkages increased the completeness and accuracy of data on these patients while reducing the burden of data collection.

Discussion

We have demonstrated the feasibility of establishing the foundations of a CHD registry by integrating raw clinical data from different hospitals and multiple clinical sources stored in various formats. We showed that indexing identifiers across multiple sites is necessary and feasible to avoid artificial duplication of patients, and that it is possible to securely maintain the link between research data and patient identifiers, without breaching confidentiality. We also showed that automation is feasible to classify native cardiac anatomy diagnoses from raw clinical data. These represent the first data layers of an incremental CHD cohort that may serve as a reliable source of data for research projects, with the possibility of systematic linking of research datasets to improve efficiency.

A CHD registry as the foundation of a multilayer data model

On the one hand, a sustainable national longitudinal CHD research registry with manual data entry would be impractical and costly. On the other hand, having several cohorts of patients on targeted lesions scattered across the country quickly becomes complex to integrate. We propose here a hybrid approach in which newly collected data from research projects are viewed as layers of data that increment the completeness of a single CHD cohort databank. In this model, a registry of patients with CHD, such as the one presented in this study, serves as the foundation over which other datasets are layered (Fig. 1). Because all these layers are linked, new research projects only need to collect the missing data elements, which then themselves become new layers. This model has several

benefits but does have its challenges, none of which are insurmountable.

Benefits and challenges of a registry-based multilayer data model

The multilayer approach adds significant efficiency in data collection. It avoids duplication of data collection and allows deeper phenotyping for specific patient populations when needed. This facilitates the future use of data collection tools developed from previous projects, and it decreases the need for redundant administrative steps to access research data. Interesting possibilities are the reuse of previously developed quality of life and physical activity questionnaires, the reuse of echocardiography data already collected from a previous project on the same CHD lesion, and the possibility of benefiting from previous data linkages between clinical and health care administrative data (see also Table 3).

The multilayer model requires a common and uniform baseline dataset. There are benefits and increased efficiency in having such a baseline comprehensive cohort of patients with CHD under a single umbrella protocol. These include a more straightforward process to access and link data, less duplication of data collection, and enhanced possibilities for data interoperability and harmonization. Because the overall number of eligible patients is high, this is feasible only with some level of automatization of baseline demographic and cardiac data extraction from local clinical data sources, such as what was demonstrated in the current study.

In our study, 7%-12% of patients with moderate-to-severe CHD had encounters in multiple institutions. Similarly, a preliminary interrogation of the datasets from the Western Canadian Children's Heart Network indicated that >13% of patients received care in more than 1 province (personal communication). Excluding these patients from studies, or not accounting for the potential duplication of their data in multi-institutional datasets, could significantly bias study results. Our approach enabled efficient patient indexing, which is a first step to tackling these issues.

Patient indexing also facilitates linkages between clinical and administrative data and between paediatric and adult centres to track patients transitioning from paediatric to adult care. We expect that exporting this indexing approach will be an uphill battle in the current context of evolving and sometimes heterogeneous privacy regulations. There exist methods to securely cross-match patients' identifiers without revealing their identity, such as the use of 1-way encryption and UUIDs.[15,23] Figure 2 shows a schematic example of how such a patient indexer based on scrambled patient identifiers may enable the linkage of multiple datasets. Capacity building, advocacy, awareness, and sharing of best practices and harmonized methods on the managing and storing of sensible patients' information locally will be key to meet these challenges. Our study served as a test case to explore these options. Other projects are ongoing to test ways of meeting privacy requirements while ensuring validity of results and security of confidential information.

It is expected that not all layers of data can be hosted in a single centralized dataset. For example, datasets from health care administrative data must often remain on the secure severs of each province and cannot be centralized. Furthermore, valuable granular data from biological and imaging biobanks will be hosted at various institutions in the country. If these datasets are properly planned and built to include a patient unique identifier linking each dataset, they can effectively function as a single databank, although its parts are stored in different places and settings. A baseline CHD registry can serve as the anchor to which other datasets are linked, therefore facilitating data integration. However, it requires a higher skill level in database management and analysis (ie, distributed analysis expertise), which may not be readily available in all centres. Capacity building and sharing of expertise in these techniques will be paramount in maximizing the chances of success.

International examples of CHD registers and data integration

Elsewhere, there have been excellent models of national registries of patients with CHD. Typical examples are the Swedish Registry of Congenital Heart Disease and the Danish Register of Congenital Heart Disease, which both prospectively collect data on CHD, with linkage with several other governmental and administrative data sources.[9,24] In the United States, efforts to link several large datasets that hold data on patients with CHD have been successfully undertaken by the Cardiac Networks United.[25] Although data from these linkages are not population based (selection bias due to insurance coverage and referral bias from the inclusion of larger reference centres), impactful research has already come out of this initiative. However, privacy laws in the United States only allow for linkage using indirect identifiers, such as dates of admission dates, which decreases the success rates of data linkages (a recent study reported approximately 75% linkage rate).[26] In Canada, an interesting model is that of the Canadian Neonatal Network.[27] This network focuses on research relating to neonatal care and adopted a hybrid model of research and quality improvement. They successfully engaged most Canadian neonatal units into entering data for preterm babies under a single umbrella protocol, on which research projects are built and targeted extra data collection can be done.

Perspectives

The current Canadian context is favourable to build a durable cohort of patients with CHD that would be a foundation for CHD research. First, the government-funded health system and the relatively small number of institutions that care for patients with CHD, even when including nonsurgical programmes, enable the constitution of representative cohorts of patients with favourable external validity. Secondly, there exists an increasingly strong culture of collaboration between Canadian investigators and institutions in paediatric and congenital cardiology.[11] Thirdly, the Canadian research ecosystem has specific expertise in leveraging the valuable population-based data contained in health care administrative databanks,[28,29] in performing linkage of clinical and administrative data for various sources,[17,19,21] and in providing data infrastructure to facilitate data access without compromising security and confidentiality.[11] Our study provides evidence that development of well-integrated collaborative approaches can improve CHD research in Canada.

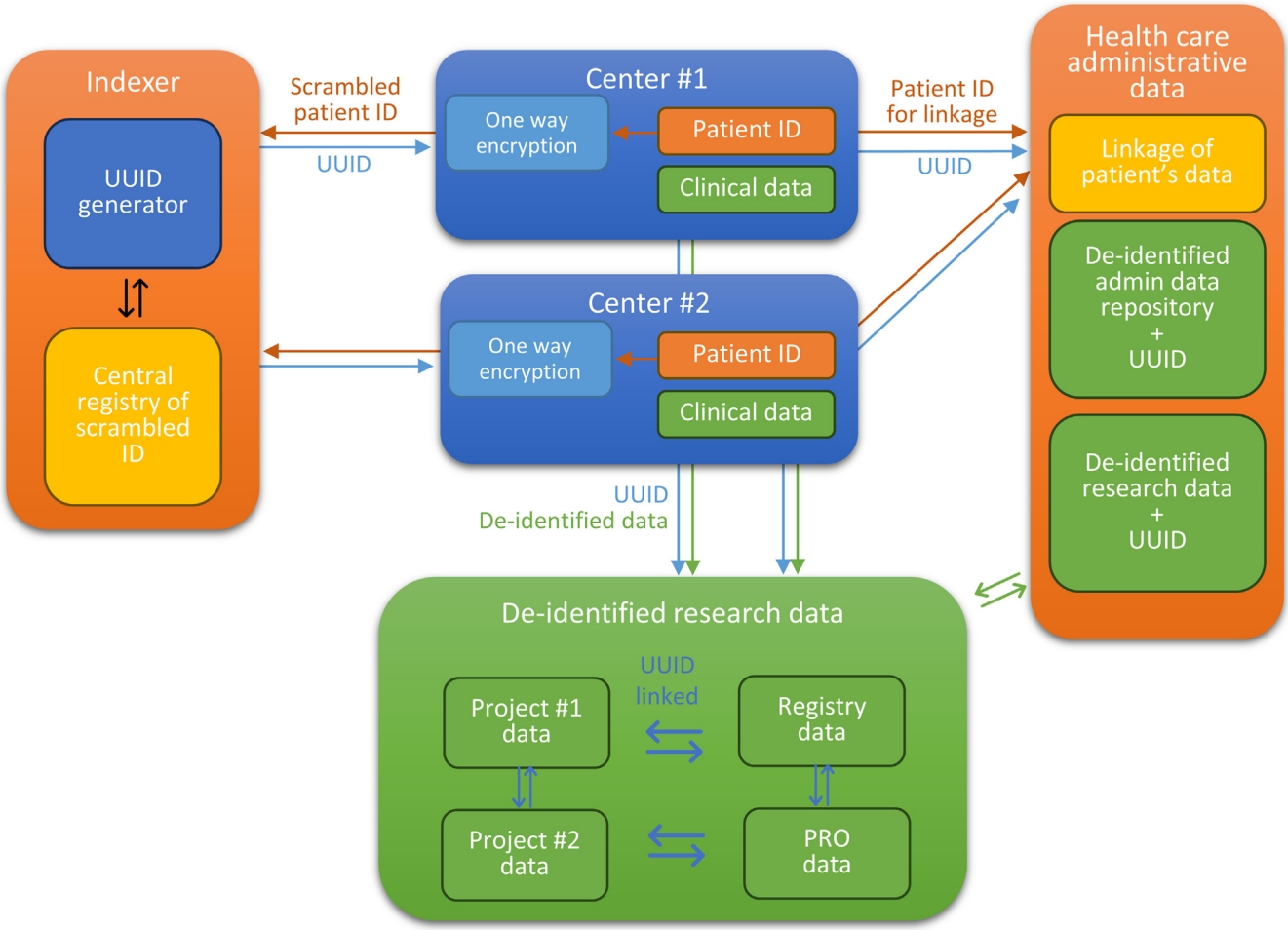

Figure 2. Schematic of a patient indexer based on scrambled patient identifiers. Patients' identifiers are scrambled and sent to a patient indexer. The indexer searches for matches for the same identifiers from other sources. A universal unique identifier (UUID) is generated and sent back to the centre. Research data are stripped from the patient identifier, and the UUID is then used to link patients across sources. PRO, patient reported outcomes.

Exporting our methods in other provinces is feasible, although not without challenges. The next steps are to engage centres where echocardiographic and demographic data can be retrieved and work to tailor the conversion scripts to each system. Upstream discussions with research ethics boards and privacy experts are needed to determine acceptable yet practical ways to index patients while preserving confidentiality. Exploration of possibilities of harmonization and linkage with current research cohorts and datasets will also be paramount to leverage existing data. These are challenges we should not shy away from, as finding solutions to better integrate research efforts will ultimately yield more efficient knowledge generation and thus improved care.

Conclusions

We showed that it is possible to create a baseline registry of patients with CHD by using data from various sources with different formats from several centres. The amount of upstream work needed to design scripts and cross-match tables was important, but periodic updates were then simpler and straightforward. This enabled the inclusion of all CHD lesions, including mild lesions, as the automatic extraction of data could process a larger number of patients without the need from manual data entry. Our indexing strategy effectively dealt with patients with encounters in multiple institutions. This number is expected to be much more important when adult and paediatric institutions participate in the same registry. This registry has already served as the basis for lesion-specific projects and has decreased data collection burden by enabling reuse of collected data and linkages between projects and datasets. We expect that this approach could serve as a blueprint to expand the registry to other provinces and to effectively approach CHD research in a more integrated way.

Ethics Statement

This study complies with the Tri-Council Policy Statement on Ethical Conducts for Research Involving Humans.

Patient Consent

The authors confirm that patient consent is not applicable to this article. A waiver of consent was obtained based on the

favourable balance between the risk of a breach of confidentiality and the potential benefits for patients care through increased knowledge. Patients had the option to opt out.

Funding Sources

This project is funded by the Fondation de la recherche pédiatrique. Dr Frédéric Dallaire received salary support from the Fonds de Recherche du Québec—Santé. The funders had no role in the design and conduct of the study.

Disclosures

The authors have no conflicts of interest to disclose.

Editorial Disclaimer

Given his role as Associate Editor, Frédéric Dallaire had no involvement in the peer review of this article and has no access to information regarding its peer review.

References

1. Ministeri M, Alonso-Gonzalez R, Swan L, Dimopoulos K. Common long-term complications of adult congenital heart disease: avoid falling in a H.E.A.P. *Expert Rev Cardiovasc Ther*. 2016;14:445–462.
2. Wald RM, Valente AM, Marelli A. Heart failure in adult congenital heart disease: emerging concepts with a focus on tetralogy of Fallot. *Trends Cardiovasc Med*. 2015;25:422–432.
3. Valente AM, Gauvreau K, Assenza GE, et al. Contemporary predictors of death and sustained ventricular tachycardia in patients with repaired tetralogy of Fallot enrolled in the INDICATOR cohort. *Heart*. 2014;100:247–253.
4. Jacobs JP, Mayer Jr JE, Mavroudis C, et al. The Society of Thoracic Surgeons Congenital Heart Surgery Database: 2017 update on outcomes and quality. *Ann Thorac Surg*. 2017;103:699–709.
5. Martin GR, Beekman RH, Ing FF, et al. The IMPACT registry: IMproving Pediatric and Adult Congenital Treatments. *Semin Thorac Cardiovasc Surg Pediatr Card Surg Annu*. 2010;13:20–25.
6. Pasquali SK, Kaltman JR, Gaynor JW, et al. Enhancing efficiency and scientific impact of a clinical trials network: the Pediatric Heart Network Integrated CARdiac Data and Outcomes (iCARD) Collaborative. *Cardiol Young*. 2019;29:1121–1126.
7. Messroghli DR, Pickardt T, Fischer M, et al. Toward evidence-based diagnosis of myocarditis in children and adolescents: rationale, design, and first baseline data of MYKKE, a multicenter registry and study platform. *Am Heart J*. 2017;187:133–144.
8. Valente AM, Gauvreau K, Assenza GE, et al. Rationale and design of an international multicenter registry of patients with repaired tetralogy of Fallot to define risk factors for late adverse outcomes: the INDICATOR cohort. *Pediatr Cardiol*. 2013;34:95–104.
9. Olsen M, Videbaek J, Johnsen SP. The Danish Register of Congenital Heart Disease. *Scand J Public Health*. 2011;39:50–53.
10. Svensson B, Idvall E, Nilsson F, Liuba P. Health-related quality of life in children with surgery for CHD: a study from the Swedish National Registry for Congenital Heart Disease. *Cardiol Young*. 2017;27:333–343.
11. Dallaire F, Battista MC, Greenway SC, et al. The Canadian Pediatric Cardiology Research Network: a model national data-sharing organization to facilitate the study of pediatric heart diseases. *CJC Open*. 2021;3:510–515.
12. Canadian Congenital and Pediatric Cardiology Research Network. A comprehensive and unified structure. Available at: http://ccpcrn.ca. Accessed March 6, 2022.
13. International Society for Nomenclature of Paediatric and Congenital Heart Disease. The cardiac code. Available at: http://ipccc.net/download-the-ipccc/. Accessed January 16, 2021.
14. Beland MJ, Harris KC, Marelli AJ, et al. Improving quality of congenital heart disease research in Canada: standardizing nomenclature across Canada. *Can J Cardiol*. 2018;34:1674–1676.
15. Roy LO, Rangsitratkul C, Dallaire F. Establishing a Canadian-based universal unique identifier for congenital and pediatric cardiology research. *Can J Cardiol*. 2022;38:135.
16. Harris PA, Taylor R, Minor BL, et al. The REDCap consortium: building an international community of software platform partners. *J Biomed Inform*. 2019;95:103208.
17. Blais S, Marelli A, Vanasse A, et al. The 30-year outcomes of tetralogy of Fallot according to native anatomy and genetic conditions. *Can J Cardiol*. 2021;37:877–886.
18. Blais S, Marelli A, Vanasse A, et al. The TRIVIA cohort for surgical management of tetralogy of Fallot: merging population and clinical data for real-world scientific evidence. *CJC Open*. 2020;2:663–670.
19. Blais S, Marelli A, Vanasse A, et al. Comparison of long-term outcomes of valve-sparing and transannular patch procedures for correction of tetralogy of Fallot. *JAMA Netw Open*. 2021;4:e2118141.
20. Noel C, Gagnon MH, Cardinal MP, et al. Rationale and design of the FREQUENCY Study: The Fetal Cardiac Registry of Quebec to Improve Resource Utilization in Fetal Cardiology. *J Obstet Gynaecol Can*. 2019;41:459–465.e12.
21. Cardinal MP, Gagnon MH, Tetu C, et al. Incremental detection of severe congenital heart disease by fetal echocardiography following a normal second trimester ultrasound scan in Quebec, Canada. *Circ Cardiovasc Imaging*. 2022;15:e013796.
22. Blais S, Meloche-Dumas L, Fournier A, Dallaire F, Dahdah N. Long-term risk factors for dilatation of the proximal aorta in a large cohort of children with bicuspid aortic valve. *Circ Cardiovasc Imaging*. 2020;13,e009675.
23. Maaroufi M, Landais P, Messiaen C, Jaulent MC, Choquet R. Federating patients identities: the case of rare diseases. *Orphanet J Rare Dis*. 2018;13:199.
24. Bodell A, Björkhem G, Thilén U, Naumburg E. National quality register of congenital heart diseases—can we trust the data? *J Congenit Cardiol*. 2017;1:1–8.
25. Gaies M, Anderson J, Kipps A, et al. Cardiac Networks United: an integrated paediatric and congenital cardiovascular research and improvement network. *Cardiol Young*. 2019;29:111–118.

26. Bates KE, Donohue J, Zhang W, et al. Assessing enrollment of eligible infants in the national pediatric cardiology quality improvement collaborative (NPC-QIC) through linkage to the pediatric cardiac critical care consortium (PC4) registry [e-pub ahead of print]. *Cardiol Young*. https://doi.org/10.1017/S1047951123001671. Accessed July 12, 2023.

27. The Canadian Neonatal Network. Available at: https://www.canadianneonatalnetwork.org/. Accessed September 17, 2023.

28. Islam S, Kaul P, Tran DT, Mackie AS. Health care resource utilization among children with congenital heart disease: a population-based study. *Can J Cardiol*. 2018;24:1289–1297.

29. Marelli AJ, Ionescu-Ittu R, Mackie AS, et al. Lifetime prevalence of congenital heart disease in the general population from 2000 to 2010. *Circulation*. 2014;130:749–756.

Supplementary Material

To access the supplementary material accompanying this article, visit *CJC Pediatric and Congenital Heart Disease* at https://www.cjcpc.ca// and at https://doi.org/10.1016/j.cjcpc.2023.12.001.