



# Novel linkage of the Society of Thoracic Surgeons database and hospital electronic health records

## Original Article

**Cite this article:** Atkins CA, Nellis JR, Vekstein AM, D'Ottavio A, DeLaRosa JM, Chiswell K, Turek JW, Beckerman Z, Raskind Hood C, Book WM, Hoffman TM, Hartman RJ, Walsh M, Li JS, and Welke KF (2024). Novel linkage of the Society of Thoracic Surgeons database and hospital electronic health records. *Cardiology in the Young*, page 1 of 7. doi: [10.1017/S1047951124026520](https://doi.org/10.1017/S1047951124026520)

Received: 6 August 2024  
Revised: 27 August 2024  
Accepted: 31 August 2024



### Keywords:

Society of Thoracic Surgeons; CHD; electronic health records; database; linkage

### Corresponding author:

Christopher A. Atkins;  
Email: [christopher.atkins@duke.edu](mailto:christopher.atkins@duke.edu)

Joseph R. Nellis is the co-first author, contributed equally along with Christopher A. Atkins to this work.

Christopher A. Atkins<sup>1</sup> , Joseph R. Nellis<sup>2</sup>, Andrew M. Vekstein<sup>2</sup>, Alfred D'Ottavio<sup>3</sup>, Jesse M. DeLaRosa<sup>3</sup>, Karen Chiswell<sup>3</sup>, Joseph W. Turek<sup>2</sup> , Ziv Beckerman<sup>2</sup>, Cheryl Raskind Hood<sup>4</sup>, Wendy M. Book<sup>4,5</sup>, Timothy M. Hoffman<sup>6</sup>, Robert J. Hartman<sup>6</sup>, Michael Walsh<sup>7</sup>, Jennifer S. Li<sup>1</sup> and Karl F. Welke<sup>8</sup>

<sup>1</sup>Division of Cardiology, Department of Pediatrics, Duke University Medical Center, Durham, NC, USA; <sup>2</sup>Division of Cardiovascular and Thoracic Surgery, Department of Surgery, Duke University Medical Center, Durham, NC, USA; <sup>3</sup>Duke Clinical Research Institute, Durham, NC, USA; <sup>4</sup>Emory Rollins School of Public Health, Department of Epidemiology, Atlanta, Georgia; <sup>5</sup>Emory School of Medicine, Division of Cardiology, Atlanta, Georgia; <sup>6</sup>Division of Cardiology, Department of Pediatrics, UNC School of Medicine, Chapel Hill, NC, USA; <sup>7</sup>Division of Cardiology, Department of Pediatrics, Wake Forest Baptist Health, Winston-Salem, NC, USA and <sup>8</sup>Congenital Heart Center, Atrium Health Levine Children's Hospital, Charlotte, NC, USA

### Abstract

**Background:** Society of Thoracic Surgeons Congenital Heart Surgery Database is the largest congenital heart surgery database worldwide but does not provide information beyond primary episode of care. Linkage to hospital electronic health records would capture complications and comorbidities along with long-term outcomes for patients with CHD surgeries. The current study explores linkage success between Society of Thoracic Surgeons Congenital Heart Surgery Database and electronic health record data in North Carolina and Georgia. **Methods:** The Society of Thoracic Surgeons Congenital Heart Surgery Database was linked to hospital electronic health records from four North Carolina congenital heart surgery using indirect identifiers like date of birth, sex, admission, and discharge dates, from 2008 to 2013. Indirect linkage was performed at the admissions level and compared to two other linkages using a "direct identifier," medical record number: (1) linkage between Society of Thoracic Surgeons Congenital Heart Surgery Database and electronic health records from a subset of patients from one North Carolina institution and (2) linkage between Society of Thoracic Surgeons data from two Georgia facilities and Georgia's CHD repository, which also uses direct identifiers for linkage. **Results:** Indirect identifiers successfully linked 79% (3692/4685) of Society of Thoracic Surgeons Congenital Heart Surgery Database admissions across four North Carolina hospitals. Direct linkage techniques successfully matched Society of Thoracic Surgeons Congenital Heart Surgery Database to 90.2% of electronic health records from the North Carolina subsample. Linkage between Society of Thoracic Surgeons and Georgia's CHD repository was 99.5% (7,544/7,585). **Conclusions:** Linkage methodology was successfully demonstrated between surgical data and hospital-based electronic health records in North Carolina and Georgia, uniting granular procedural details with clinical, developmental, and economic data. Indirect identifiers linked most patients, consistent with similar linkages in adult populations. Future directions include applying these linkage techniques with other data sources and exploring long-term outcomes in linked populations.

### Introduction

The Society of Thoracic Surgeons Congenital Heart Surgery Database is the largest congenital heart surgery database worldwide with over 500,000 entries since 1994.<sup>1,2</sup> While the Society of Thoracic Surgeons Congenital Heart Surgery Database is one of the most granular active databases collecting intraoperative and post-operative data and outcomes for patients undergoing surgery for CHD, it does not include information beyond 30 days after discharge for the primary episode of care. Society of Thoracic Surgeons Congenital Heart Surgery Database has been previously linked to other databases to capture additional information.<sup>3-5</sup> Linkage to hospital electronic health records, however, has not been previously reported. Such a linkage would combine the long-term outcomes for all patients with CHD who undergo surgical repair, including additional medical and surgical encounters, resource utilisation, other medical diagnoses, and medication use found in hospital electronic health record with the depth and detail of the patients' pre-operative characteristics, surgical course, and perioperative

© The Author(s), 2024. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.



complications provided in the Society of Thoracic Surgeons Congenital Heart Surgery Database.<sup>6–8</sup>

This study sought to (1) explore linkage between Society of Thoracic Surgeons Congenital Heart Surgery Database and electronic health record data at surgical institutions in North Carolina and Georgia using indirect versus direct identifiers; (2) describe linkage success across institution and state, by year and by diagnosis; and (3) compare linked and unlinked patients to assess representativeness of a linked population to a general congenital cardiac surgical population.

## Material and methods

### Data sources

The Society of Thoracic Surgeons Congenital Heart Surgery Database contains patient characteristics, procedural details, and in-hospital and 30-day outcomes for patients undergoing surgery for CHD at 119 participating centres (as of 2021) in the United States, representing over 90% of congenital heart surgery programmes in the United States, Canada, and Saudi Arabia (one institution).<sup>2</sup> Society of Thoracic Surgeons data harvests were obtained from four participating institutions in North Carolina. One additional North Carolina institution was excluded from the analysis due to it not reporting outcomes to the Society of Thoracic Surgeons Congenital Heart Surgery Database. The Centers for Disease Control and Prevention's Surveillance of Congenital Heart Defects Across the Lifespan Project, hereinto referred to as "Centers for Disease Control and Prevention's CHD LS Project," uses large clinical and administrative databases from five national sites to understand population-based CHD epidemiology and outcomes.<sup>7</sup> Duke University and Emory University were both participants in Centers for Disease Control and Prevention's CHD LS Project. In the current study, linkage of Society of Thoracic Surgeons Congenital Heart Surgery Database data and electronic health records from four congenital heart surgery programmes in North Carolina that consistently participated in both the Society of Thoracic Surgeons Congenital Heart Surgery Database and Centers for Disease Control and Prevention's CHD LS Project was conducted using an indirect approach. In addition, an alternative linkage technique was conducted using direct identifiers between Society of Thoracic Surgeons data from two Georgia facilities and data from Georgia's CHD repository that was developed as part of the Centers for Disease Control and Prevention's CHD LS Project; this linkage effort served as a comparison and was selected given their high direct linkage success. Analysis of the Society of Thoracic Surgeons Congenital Heart Surgery Database and Centers for Disease Control and Prevention's CHD LS Project was approved by the Institutional Review Board of Duke University and other participating centres with need for individual patient consent waived (IRB# Pro00014556 and Pro00068049). The CHD LS Project at Emory University received approval from its Institutional Review Board with a complete waiver of Health Insurance Portability and Accountability Act authorisation and informed consent (#IRB00064051). The Institutional Review Board and institutional data use agreements did not allow for sharing of direct patient identifiers across institutions, but separate comparison of direct identifiers of Georgia CHD repository was completed. All database management at North Carolina was performed in an Oracle system Database version 19c, while at Georgia, database management was conducted using various applications on a Linux-based system. For both sites, all

statistical analysis was performed in SAS version 9.4 (SAS Institute, Cary, North Carolina).

### Patient population

Society of Thoracic Surgeons Congenital Heart Surgery Database data harvests were queried for unique surgical admissions, confirmed through Society of Thoracic Surgeons Patient ID, admission, and discharge dates, from 2008 to 2013. Patient records with a missing admission or discharge date, date of birth (or date that could not be imputed from age at surgery and surgery date), sex assigned at birth, primary diagnosis, or primary procedure were excluded. Data from the electronic health record from each participating institution were obtained by searching records for International Classification of Diseases (ICD)-9 CM CHD-related diagnosis codes and inpatient admissions with ICD-9-CM/CPT codes for cardiovascular surgical procedures from 2008–2013, except for one of the four North Carolina participating hospitals (Hospital 4) that only contributed data from the years 2010 to 2012. Participants from Georgia included individuals whose cardiac-related surgery was recorded in one of two Society of Thoracic Surgeons databases from participating facilities and cases with CHD from Georgia's CHD repository that was further developed as part of Centers for Disease Control and Prevention's CHD LS Project.

### Linkage technique

First, patient admissions from the Society of Thoracic Surgeons Congenital Heart Surgery Database were linked to cardiovascular surgical admissions from the electronic health record by matching with indirect identifiers including date of birth, sex, admission, and discharge dates. Linkage was performed at the level of admission rather than procedures; many patients were expected to undergo multiple procedures during the same hospitalisation, and electronic health record contains data on admissions which can confound the linkage.

In addition, two comparative linkages were performed using "direct identifiers." First, for a subset of patients undergoing surgery at one of the North Carolina institutions (referred to as Hospital 2 on graphs below), linkage was conducted between their Society of Thoracic Surgeons Congenital Heart Surgery Database and electronic health record data using a medical record number as a direct identifier, which was documented in the electronic health record and from the Society of Thoracic Surgeons demographic form from version 3.0 onwards (effective from September 2009). The medical record number was available in 1,026/1,696 (60.5%) of the Society of Thoracic Surgeons Congenital Heart Surgery Database patient records from Hospital 2. Second, another direct linkage effort was performed for a Georgia cohort between Society of Thoracic Surgeons data and the Georgia's CHD repository data in order to have a comparison linkage rate. The linkage between these Georgia sources also utilised medical record number as a direct identifier.

### Assessment of linkage success

Linkage results were summarised as the number and percent of Society of Thoracic Surgeons Congenital Heart Surgery Database admissions linked to admissions from the electronic health record through indirect and then direct identifiers. Summaries of the linkage success were then stratified by site and year of surgery. In order to identify potential biases in the linkage process by indirect identifiers, the patient and

procedural characteristics of all eligible Society of Thoracic Surgeons Congenital Heart Surgery Database admissions for patients successfully linked by indirect identifiers versus those who could not be linked were compared to identify whether or not specific CHD diagnoses or age groups were less frequently linked by indirect identifiers. Statistical comparisons were made using the Wilcoxon rank sum test for continuous variables and the chi-squared or Fisher's exact test, as appropriate, for binary and categorical variables. A  $p$  value  $< 0.05$  was considered statistically significant.

## Results

### Indirect linkage success: overall, by site, by year, and by index operation

A total of 4,685 patients were available to be linked by indirect identifiers. Using indirect identifiers, 79% (3,692/4,685) of Society of Thoracic Surgeons Congenital Heart Surgery Database admissions at the four queried North Carolina hospitals were linked to hospital admissions. The remaining 993 Society of Thoracic Surgeons Congenital Heart Surgery Database admissions were for patients who were not linked using indirect identifiers to the electronic health record. Linkage success was consistent across three of the four sites, including 81% (1,238/1,521) at Hospital 1, 75% (1,418/1,889) at Hospital 2, and 85% (799/940) at Hospital 3. However, linkage was poor at Hospital 4 with only 31% (104/335) success. The proportion of admissions that successfully linked was consistent year-by-year across the study period (Figure 1) with linkage across all sites in 2008 at 77% and in 2013 at 76%.

Linkage success across all primary procedures for the index operations in the Society of Thoracic Surgeons Congenital Heart Surgery Database ranged from 38% to 94% depending upon procedure (Table 1). Atrial septal defect closure (93.9%), partial anomalous pulmonary venous return repair (92.8%), and atrio-ventricular canal defect repair (92.6%) had the highest linkage success, while extra-corporeal membrane oxygenation procedures (37.7%), pacemaker/automatic implantable cardioverter defibrillator implants (54.5%), and heart transplantation (58.3%) had the lowest linkage success.

### Representativeness of linked patients

Admission characteristics (index operation) of linked versus unlinked patients are presented in Table 2, with no single trend towards higher or lower risk. Successfully linked patients were of similar age and size to unlinked patients. In terms of pre-operative risk factors, linked cases were more commonly reoperations (33% vs. 26%,  $p < 0.001$ ), less frequently required pre-operative mechanical ventilation (7% vs. 12%,  $p < 0.001$ ), and less commonly presented with renal dysfunction (1.6% vs. 3.8%,  $p < 0.001$ ).

Finally, Table 3 demonstrates the frequency of operative complications in the linked versus unlinked patients. While linked patients were representative of the overall population for many complications, including unplanned reoperation or catheterisation, chylothorax, cardiac arrest, severe cardiac dysfunction, and sepsis, there were some key differences. Linked patients had lower incidence of neurologic deficit (1.1% vs. 2.4%,  $p < 0.001$ ) and prolonged mechanical ventilation (9.5% vs. 16.8%,  $p < 0.001$ ), but higher incidence of post-operative arrhythmia (14.4% vs. 9.2%,  $p < 0.001$ ).

### Indirect versus direct linkage

The medical record number was available for 1026 Society of Thoracic Surgeons Congenital Heart Surgery Database patients with 1071 admissions at Hospital 2, and 925 (90.2%) of these patients could be directly linked by medical record number to the electronic health record. In comparison, 75% of this patient group were successfully linked by indirect methods to the same patient. With parallel direct linkage techniques of index cases in lifespan between the Georgia Society of Thoracic Surgeons data and Georgia CHD repository, there was 99.5% (7,544/7,585) linkage success.

### Comment

The Society of Thoracic Surgeons Congenital Heart Surgery Database is the world's largest and most robust registry of congenital heart surgery patient characteristics, operative data, and outcomes. However, the database is inherently limited by a focus on in-hospital and 30-day outcomes. Hospital electronic health records house long-term data, including additional medical and surgical encounters, resource utilisation, other medical diagnoses, and medication use. The potential to link these two entities for future investigation unites the strengths of each database, including granular diagnosis and procedural details from the Society of Thoracic Surgeons Congenital Heart Surgery Database with the long-term outcomes in hospital electronic health records, which has particular significance in the paediatric population as demonstrated in Cardiac Networks United.<sup>9,10</sup> This linkage would allow better understanding of survival, healthcare use, reproductive health, and long-term outcomes of people living with CHDs. In doing so, we can provide improved counselling on the natural history of CHD and further investigate outcomes to optimise care of these patients.

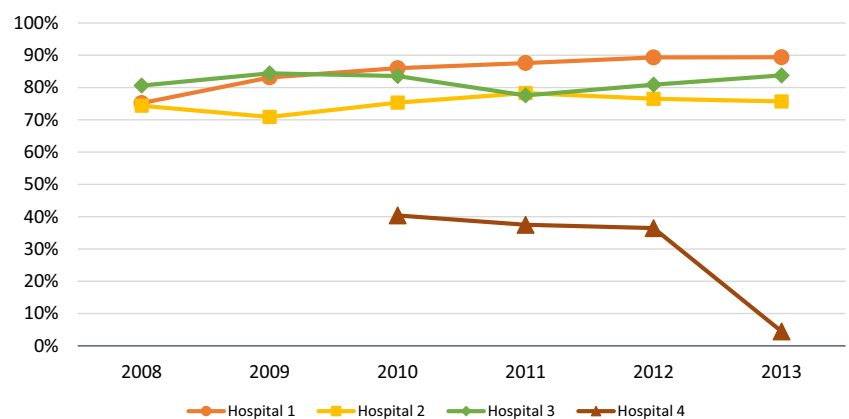
In an exploratory analysis linking Society of Thoracic Surgeons Congenital Heart Surgery Database data from four institutions in North Carolina with hospital electronic health record, the majority of individuals were successfully linked using indirect identifiers including date of birth, sex, and dates of admission and discharge. In general, linkage is more likely to be successful when a direct identifier is used, including at Hospital 2 or with Georgia's databases. The perfect linkage in the final comparative group was likely related to a single Georgia wide-based paediatric institution's participation who provided data as part of their Centers for Disease Control and Prevention CHD Lifespan Project and the Society of Thoracic Surgeons database; both data sources were exceptionally managed and had consistent reporting of medical record numbers. Parallel direct linkage techniques successfully build upon the existing CHD surveillance infrastructure and now utilised integrated electronic health record through direct linkage utilising medical record numbers submitted to Society of Thoracic Surgeons Congenital Heart Surgery Database for all patients. In this instance, the source for CHD LS Project clinical records and Society of Thoracic Surgeons submission data both utilise a common medical record number identifier that ensures correct linkage as any submitted Society of Thoracic Surgeons record would require a medical record number that can be linked back to the CHD LS Project data.

Examining linkage success by index operation for potential patterns, more common elective procedures, such as atrial septal defect, ventricular septal defect, atrio-ventricular canal defect repair, or pulmonary artery conduit, are more likely to be linked, whereas emergent procedures like extra-corporeal membrane

**Table 1.** Specific operation stratified by success of linkage by indirect identifiers at four North Carolina hospitals (in descending order of linkage success)

	Linked by indirect identifiers (N = 3692)	Not linked by indirect identifiers (N = 993)	All eligible STS-CHSD admissions (N = 4685)	Linkage success (%)
Atrial septal defect repair	275	18	293	93.9
PAPVR repair	64	5	69	92.8
Atrio-ventricular canal defect repair	214	17	231	92.6
Valve replacement	262	33	295	88.2
Aortic valve repair for aortic stenosis	112	14	126	88.9
Pulmonary conduit placement/reoperation	101	13	114	88.6
Fontan	148	19	167	88.6
Ventricular septal defect repair	335	44	379	88.4
Modified Blalock–Thomas–Taussig shunt	95	14	109	87.1
Tetralogy of Fallot repair	206	31	237	86.9
Bidirectional Glenn	111	17	128	86.7
Aortic coarctation repair	193	33	226	85.4
Valvuloplasty	91	22	113	80.5
Pulmonary artery reconstruction/plasty	45	11	56	80.4
Pulmonary artery banding	42	11	53	79.2
TAPVC repair	56	16	72	77.8
Norwood procedure	117	39	156	75.0
Aortic arch repair	41	19	60	68.3
Vascular ring repair	33	17	50	66.0
Patent ductus arteriosus closure	191	105	296	64.5
Orthotopic heart transplant	28	20	48	58.3
Pacemaker/AICD implantation/procedures	131	110	241	54.4
ECMO cannulation/procedure	49	81	130	37.7
Other procedures	650	249	899	N/A
Missing procedural information	8	13	21	N/A

ECMO = Extra-corporeal membrane oxygenation; PAPVR = partial anomalous pulmonary venous return; TAPVC = total anomalous pulmonary venous connection; AICD = automatic implantable cardioverter defibrillator; STS-CHSD = Society of Thoracic Surgeons Congenital Heart Surgery Database.

**Figure 1.** Indirect linkage success across four North Carolina institutions in Society of Thoracic Surgeons Congenital Heart Surgery Database by year

**Figure 1.** Indirect linkage success across four North Carolina institutions in Society of Thoracic Surgeons Congenital Heart Surgery Database by year. All eligible admissions in the Society of Thoracic Surgeons Congenital Heart Surgery Database were linked via indirect identifiers to hospital electronic health record. The percentage of successfully linked admissions was steady across the study period, with 77% of admissions at all sites linked in 2008 compared to 76% in 2013. Linkage success increased at Hospital 3 (75% in 2008 to 89% in 2013) and Hospital 1 (81% in 2008 to 84% in 2013).

**Table 2.** Patient characteristics stratified by success of linkage by indirect identifiers at four North Carolina hospitals

	Linked by indirect identifiers (N = 3692)	Not linked by indirect identifiers (N = 993)	All eligible STS-CHSD admissions (N = 4685)	p value
<b>Demographics</b>				
Age (years)	7.16 [0.13–8.55]	7.92 [0.02–14.13]	7.32 [0.08–10.19]	0.246
Male gender	1996 (54.1%)	581 (58.5%)	2577 (55.0%)	0.012
Height (cm)	90.8 [54.0–126.5]	95.4 [50.0–154.0]	91.7 [53.0–135.0]	0.543
Weight (kg)	22.3 [4.0–27.1]	26.4 [3.3–49.0]	23.2 [3.8–32.7]	0.973
Body surface area (m <sup>2</sup> )	0.70 [0.25–0.98]	0.78 [0.21–1.45]	0.72 [0.24–1.11]	0.505
<b>Pre-operative risk factors*</b>				
Prior cardiac surgery	1229 (33.3%)	253 (25.5%)	1482 (31.6%)	<0.001
Mechanical ventilation	249 (6.7%)	119 (12.0%)	368 (7.9%)	<0.001
shock	224 (6.1%)	51 (5.1%)	275 (5.9%)	0.268
Gastrostomy tube present	93 (2.5%)	31 (3.1%)	124 (2.6%)	0.293
Renal dysfunction	60 (1.6%)	38 (3.8%)	98 (2.1%)	<0.001
Stroke/intracranial haemorrhage	47 (1.3%)	22 (2.2%)	69 (1.5%)	0.029
Seizures	54 (1.5%)	20 (2.0%)	74 (1.6%)	0.216

\*Most frequent pre-operative risk factors reported for the index operation of the admission. Categorical variables are presented as n (%). Continuous variables are presented as median [Q1–Q3]. STS-CHSD = Society of Thoracic Surgeons Congenital Heart Surgery Database.

**Table 3.** Society of Thoracic Surgeons operative outcomes stratified by success of linkage by indirect identifiers at four North Carolina hospitals

Complication*	Linked with indirect Identifiers (N = 3692)	Not linked with indirect identifiers (N = 993)	All eligible STS-CHSD admissions (N = 4685)	p value
Arrhythmia	532 (14.4%)	91 (9.2%)	623 (13.3%)	<0.001
Chylothorax	66 (1.8%)	20 (2.0%)	86 (1.8%)	0.637
Severe cardiac dysfunction	49 (1.3%)	17 (1.7%)	66 (1.4%)	0.361
Cardiac arrest	90 (2.4%)	21 (2.1%)	111 (2.4%)	0.553
Neurologic deficit	39 (1.1%)	24 (2.4%)	63 (1.3%)	<0.001
Respiratory insufficiency requiring mechanical ventilation >7 days	351 (9.5%)	167 (16.8%)	518 (11.1%)	<0.001
Sepsis	55 (1.5%)	22 (2.2%)	77 (1.6%)	0.110
Unplanned reoperation	92 (2.5%)	26 (2.6%)	118 (2.5%)	0.821
Unplanned interventional catheterisation	43 (1.1%)	17 (1.7%)	59 (1.3%)	0.150

\*Most frequent complications reported for index operation of the admission. STS-CHSD = Society of Thoracic Surgeons Congenital Heart Surgery Database.

oxygenation cannulation were less likely to be linked. While there were no systematic (higher or lower risk, increased or decreased operative outcomes) differences between linked and unlinked patients who would potentially introduce bias into future analyses, there were a few key differences that must be acknowledged as potential limitations of this technique. Linkage techniques appear to capture a high proportion of patients with risk factors like prior cardiac surgery but miss some patients with pre-operative mechanical ventilation or renal impairment. Most importantly, the frequency of some operative outcomes was significantly lower among linked patients, notably neurologic deficit and prolonged ventilation. While it is unclear why, we speculate that those with these types of complications were likely to stay in the hospital longer and thus have more discrepancy with respect to their

discharge date and potential changes to indirect identifier information. Thus, a linked cohort may not represent the overall population for some important outcomes.

Linkage between the Society of Thoracic Surgeons databases and external data sources has been utilised in the past to assess longer-term outcomes. Most notably, the Centers for Medicare and Medicaid Services administrative claims data for patients 65 years of age and older has been linked to the Society of Thoracic Surgeons Adult Cardiac Surgical Database via a set of indirect identifiers, including surgical site, date of birth, admission, and discharge dates. Initial linkage studies which focused on isolated coronary artery bypass grafting found this technique to achieve up to 97% completeness (number of Centers for Medicare and Medicaid Services coronary artery bypass grafting hospitalisations

linked to an Society of Thoracic Surgeons registry divided by number of Centers for Medicare and Medicaid Services coronary artery bypass grafting hospitalisations at Society of Thoracic Surgeons sites).<sup>11</sup> There were, however, some key differences between linked and unlinked patients in terms of sex, race, elective versus urgent/emergent status, and, most importantly, frequency of in-hospital and 30-day mortality.<sup>12</sup> Subsequent studies employing these linkage techniques have achieved linkage success ranging from 63% to 78% depending on the population of interest.<sup>13–15</sup> Thus, successful linkage and potential limitations in terms of representativeness to the overall population described in the current study appear consistent with established approaches utilised in the adult cardiac surgery population.

Within the congenital cardiac surgery population, registry linkage has previously been attempted between the Society of Thoracic Surgeons Congenital Heart Surgery Database and Congenital Heart Surgeons Society Database, which follows long-term outcomes in twelve common populations with a CHD diagnoses at 74 centres in North America.<sup>3,4</sup> Of note, participation in Congenital Heart Surgeons Society and Society of Thoracic Surgeons Congenital Heart Surgery Database is voluntary and requires patient consent. Thus, the goal of linking Society of Thoracic Surgeons Congenital Heart Surgery Database and Congenital Heart Surgeons Society data was not only to conduct future analyses utilising the detailed long-term follow-up in Congenital Heart Surgeons Society, but also to optimise enrolment in Congenital Heart Surgeons Society by identifying potential patients submitted to Society of Thoracic Surgeons Congenital Heart Surgery Database. Similar linkage has been performed between the Society of Thoracic Surgeons Congenital Heart Surgery Database and paediatric administrative data contained in the Pediatric Health Information System database, achieving 87.4% matching to exact criteria.<sup>5</sup> This linkage provides a potential to add details about long-term pharmaceutical use, imaging studies, and billing to the Society of Thoracic Surgeons Congenital Heart Surgery Database.

A Society of Thoracic Surgeons Congenital Heart Surgery Database and hospital electronic health record linkage creates a unique dataset within congenital heart surgery that would be more broadly applicable to a larger population of patients. Historically, long-term outcomes following congenital heart surgery have come from single institution studies. While helpful, studies are limited by era effect, lack of follow-up, and case volume, particularly for high-complexity, low-volume cases. Future studies employing this linkage could connect surgical procedures at a young age with lifelong outcomes of significance in the paediatric population.

### Limitations

Beyond the potential challenges with indirect linkage techniques discussed previously, several limitations to this analysis should be noted. First, this study is a proof-of-concept study demonstrating the potential to link the Society of Thoracic Surgeons Congenital Heart Surgery Database to the hospital electronic health record in two states. Linkage success may decline when applied more broadly, especially given potential differences in data collection. Even among the four institutions in North Carolina in the present study, linkage success varied greatly, likely revealing the limitations in data modelling including administrative logistics. In particular for this dataset, there were issues with defining the exact admission and discharge dates that were utilised for linkage purposes versus the billing encounters that represented transactions outside of the

health system. In the context of surgery for CHD, families may seek surgical care away from their home state, and all their follow-up care may be in another state as well, limiting the success of linkage and generalizability. In addition, this study focused on utilising ICD-9-CM diagnosis codes, the given electronic health record standard during this study's time frame. Since then, there has been development of ICD-10-CM diagnosis codes and updates to the Society of Thoracic Surgeons database that should yield similar results found in our study, but have not been directly measured. In addition, all CHD is not captured within the Society of Thoracic Surgeons database and future studies could evaluate linkage for catheterisation-based data registries such as Improving Pediatric and Adult Congenital Treatments. Lastly, confidentiality and privacy breaches have become more widespread in the technological era and present potential concerns for direct linkage of patient data. While this is a significant concern, if future studies were to utilise this linkage, contributing sites could link their own data with direct identifiers prior to it leaving their health system and thus eliminate the need for these identifiers to leave their firewall. In addition, given this concern, there has been an advent of health information technology companies specifically created for the exchange of healthcare data that utilise hashed identifiers instead of direct identifiers, so coordinating centres can then link with other sources using de-identified data points.

### Conclusions

Large multi-centre registries are a powerful tool in modern observational cardiac surgery research; however, each database presents limitations in depth of data or length of patient follow-up. This study demonstrates a technique for linkage between the Society of Thoracic Surgeons Congenital Heart Surgery Database and hospital electronic health record database in North Carolina and Georgia, uniting granular procedural details with hospital clinical data. Through indirect identifiers alone, patients were linked with linkage success rates similar to linkage studies employing the adult cardiac surgery populations. While unlinked patients had notable differences in characteristics and outcomes, there were no systematic trends in these differences to suggest that linked patients did not represent the general population. Future directions include applying these linkage techniques to other datasets in cardiac transplantation or interventional cardiology and exploring long-term outcomes in the linked population to further optimise our care of these patients in regards to survival, healthcare use, and reproductive health.

**Acknowledgements.** None.

**Financial support.** This work was supported by the Centers for Disease Control 1 NU50DD004933-01-00: Surveillance of Congenital Heart Disease Across the Lifespan and the National Institutes of Health (5T32HL069749-17; Dr. Vekstein).

**Competing interests.** None.

### References

1. Jacobs JP, Shahian DM, Prager RL *et al.* Introduction to the STS national database series: outcomes analysis, quality improvement, and patient safety. *Ann Thorac Surg* 2015; 100: 1992–2000.
2. Kumar SR, Mayer JE, Overman DM, Shashidharan S, Wellnitz C, Jacobs JP. The Society of thoracic surgeons congenital heart surgery database: 2021 update on outcomes and research. *Ann Thorac Surg* 2021; 112: 1753–1762.

3. Jacobs JP, Pasquali SK, Austin E et al. Linking the Congenital Heart Surgery Databases of the Society of Thoracic Surgeons and the Congenital Heart Surgeons' Society: part 1—Rationale and methodology. *World J Pediatr Congenit Heart Surg* 2014; 5: 256–271.
4. Jacobs JP, Pasquali SK, Austin E et al. Linking the Congenital Heart Surgery Databases of the Society of Thoracic Surgeons and the Congenital Heart Surgeons' Society: part 2—lessons learned and implications. *World J Pediatr Congenit Heart Surg* 2014; 5: 272–282.
5. Pasquali SK, Jacobs JP, Shook GJ et al. Linking clinical registry data with administrative data using indirect identifiers: implementation and validation in the congenital heart surgery population. *Am Heart J* 2010; 160: 1099–1104.
6. Sarno LA, Cortright L, Stanley T, Tumin D, Li JS, Sang CJ. Clinical and socioeconomic predictors of work participation in adult congenital heart disease patients. *Cardiol Young* 2020; 2: 1–5.
7. Glidewell MJ, Farr SL, Book WM et al. Individuals aged 1–64 years with documented congenital heart defects at healthcare encounters, five U.S. surveillance sites, 2011–2013. *Am Heart J* 2021; 238: 100–108.
8. Williams JL, Torok RD, D'Ottavio A et al. Causes of death in infants and children with congenital heart disease. *Pediatr Cardiol* 2021; 42: 1308–1315.
9. Anderson BR, Dragan K, Crook S et al. Improving longitudinal outcomes, efficiency, and equity in the care of patients with congenital heart disease. *J Am Coll Cardiol* 2021; 78: 1703–1713.
10. Gaies M, Anderson J, Kipps A et al. Cardiac Networks United Executive Committee and Advisory Board. Cardiac Networks United: an integrated paediatric and congenital cardiovascular research and improvement network. *Cardiol Young* 2019; 29: 111–118.
11. Jacobs JP, Edwards FH, Shahian DM et al. Successful linking of the Society of Thoracic Surgeons Adult Cardiac Surgery Database to Centers for medicare and medicaid services medicare data. *Ann Thorac Surg* 2010; 90: 1150–1157.
12. Jacobs JP, Shahian DM, He X et al. Completeness, and representativeness of the Society of Thoracic Surgeons Adult Cardiac Surgery Database. *Ann Thorac Surg* 2016; 101: 33–41.
13. Brennan JM, Edwards FH, Zhou Y, O'Brien S, Booth ME. Long-term safety and effectiveness of mechanical versus biologic aortic valve prostheses in older patients. *Circulation* 2013; 127: 1647–1655.
14. Riad FS, Grau-Sepulveda M, Jawitz OK et al. Anticoagulation in new-onset postoperative atrial fibrillation: an analysis from the Society of Thoracic Surgeons Adult Cardiac Surgery Database. *Heart Rhythm O2* 2022; 3: 325–332.
15. Lowenstern AM, Vekstein AM, Grau Sepulveda M, Badhwar V, Thourani VH et al. Impact of Transcatheter Mitral Valve Repair Availability on Volume and Outcomes of Surgical Repair. *J Am Coll Cardiol* 2023; 81: 521–532.