

Original Article

Cite this article: Prospero CJ, Trachtenberg FL, Pemberton VL, Pasquali SK, Anderson BR, Ash KE, Bainton J, Dunbar-Masterson C, Graham EM, Hamstra MS, Hollenbeck-Pringle D, Jacobs JP, Jacobs ML, John R, Lambert LM, Oster ME, Swan E, Waldron A, and Nathan M (2019) Lessons learned in the use of clinical registry data in a multi-centre prospective study: the Pediatric Heart Network Residual Lesion Score Study. *Cardiology in the Young* 29: 930–938.

doi: [10.1017/S1047951119001148](https://doi.org/10.1017/S1047951119001148)

Received: 19 December 2018

Revised: 13 March 2019

Accepted: 23 April 2019

First published online: 17 June 2019

Key words:

Research efficiency; prospective studies; registry data

Author for correspondence:

Meena Nathan MD, MPH, Department of Cardiac Surgery, Bader 273, Boston Children's Hospital, 300 Longwood Ave, Boston, MA 02115, USA. Tel: +1 617 355 7932; Fax: +1 617 730 0214; E-mail: meena.nathan@cardio.chboston.org

Lessons learned in the use of clinical registry data in a multi-centre prospective study: the Pediatric Heart Network Residual Lesion Score Study

Carol J. Prospero¹, Felicia L. Trachtenberg², Victoria L. Pemberton³, Sara K. Pasquali⁴, Brett R. Anderson⁵, Kathleen E. Ash⁶, Jessica Bainton⁷, Carolyn Dunbar-Masterson⁸, Eric M. Graham⁹, Michelle S. Hamstra⁶, Danielle Hollenbeck-Pringle², Jeffrey P. Jacobs¹⁰, Marshall L. Jacobs¹¹, Rija John¹², Linda M. Lambert¹³, Matthew E. Oster¹⁴, Elizabeth Swan¹⁵, Abigail Waldron¹⁶, Meena Nathan¹⁷, for the Pediatric Heart Network Investigators

¹Nemours Cardiac Center, Alfred I. duPont Hospital for Children, Wilmington, DE, USA; ²New England Research Institutes, Watertown, MA, USA; ³National Heart, Lung, and Blood Institute, NIH, Bethesda, MD, USA; ⁴Division of Pediatric Cardiology, Department of Pediatrics, University of Michigan C.S. Mott Children's Hospital, Ann Arbor, MI, USA; ⁵Division of Pediatric Cardiology, New York Presbyterian/Morgan Stanley Children's Hospital, New York, NY, USA; ⁶Division of Pediatric Cardiology, Cincinnati Children's Hospital Medical Center, Cincinnati, OH, USA; ⁷Division of Pediatric Cardiology, The Hospital for Sick Children, Toronto, ON, Canada; ⁸Department of Cardiology, Boston Children's Hospital, Boston, MA, USA; ⁹Division of Pediatric Cardiology, Medical University of South Carolina, Charleston, SC, USA; ¹⁰Department of Surgery, Johns Hopkins All Children's Hospital, Saint Petersburg, FL, USA; ¹¹Division of Cardiac Surgery, Johns Hopkins Heart and Vascular Institute, Baltimore, MD, USA; ¹²Division of Congenital Heart Surgery, Texas Children's Hospital, Houston, TX, USA; ¹³Division of Pediatric Cardiothoracic Surgery, University of Utah, Salt Lake City, UT, USA; ¹⁴Division of Pediatric Cardiology, Emory University School of Medicine, Children's Healthcare of Atlanta, Atlanta, GA, USA; ¹⁵Division of Pediatric Cardiology, Riley Hospital for Children, Indianapolis, IN, USA; ¹⁶Division of Cardiology, Children's Hospital Philadelphia, Philadelphia, PA, USA and ¹⁷Department of Cardiac Surgery, Boston Children's Hospital, Boston, MA, USA

Abstract

Background: Using existing data from clinical registries to support clinical trials and other prospective studies has the potential to improve research efficiency. However, little has been reported about staff experiences and lessons learned from implementation of this method in pediatric cardiology. **Objectives:** We describe the process of using existing registry data in the Pediatric Heart Network Residual Lesion Score Study, report stakeholders' perspectives, and provide recommendations to guide future studies using this methodology. **Methods:** The Residual Lesion Score Study, a 17-site prospective, observational study, piloted the use of existing local surgical registry data (collected for submission to the Society of Thoracic Surgeons-Congenital Heart Surgery Database) to supplement manual data collection. A survey regarding processes and perceptions was administered to study site and data coordinating center staff. **Results:** Survey response rate was 98% (54/55). Overall, 57% perceived that using registry data saved research staff time in the current study, and 74% perceived that it would save time in future studies; 55% noted significant upfront time in developing a methodology for extracting registry data. Survey recommendations included simplifying data extraction processes and tailoring to the needs of the study, understanding registry characteristics to maximise data quality and security, and involving all stakeholders in design and implementation processes. **Conclusions:** Use of existing registry data was perceived to save time and promote efficiency. Consideration must be given to the upfront investment of time and resources needed. Ongoing efforts focussed on automating and centralising data management may aid in further optimising this methodology for future studies.

With the recent decline in federal research funding and the increase in costs and complexity of conducting multi-centre studies and clinical trials, investigators and research leaders have sought methods to improve efficiency.¹ One method has involved leveraging data from existing clinical registries.^{1–6} Registries collect pre-specified clinical data for a variety of purposes, including outcomes tracking, national benchmarking, quality improvement, and public reporting, and are also used to facilitate research activities.

Using registry data for clinical studies and trials has been termed the “next disruptive technology” in research.⁷ This has been hypothesised to have the potential to improve efficiency and

reduce redundancies in research data collection and management since many registries are already capturing some or all of the data of interest within a large, engaged group of sites. The field of cardiology is well suited to take advantage of this methodology given the availability of multiple existing clinical registries and databases, standardised nomenclature and definitions, and a collaborative environment among centres.^{1,3,8,9} Clinical registry data have been utilised to support prospective research in a few select studies in the field to date.^{10–12} However, little has been reported about experience with this method in pediatric cardiology.

We conducted a survey across multiple stakeholders to understand the use of clinical registry data to support a prospective multi-centre observational study conducted within the Pediatric Heart Network. Our aims were to: (1) describe the process of using local registry data in conjunction with standard data collection in a large, multi-institutional study, (2) understand the perceptions of stakeholders involved in data collection and management, and (3) provide recommendations that may aid in guiding future studies using this methodology.

Materials and methods

Pediatric Heart Network

The Pediatric Heart Network was established in 2001 with funding from the National Heart, Lung, and Blood Institute of the National Institutes of Health. Consisting of 10 core clinical sites, a data coordinating center, and multiple auxiliary sites, the Pediatric Heart Network conducts observational studies and randomised clinical trials in pediatric acquired heart disease and congenital heart disease.¹³ Data collection for these studies is routinely performed by trained research co-ordinators at the clinical sites and requires substantial financial support for the time necessary to collect and enter data.

Residual Lesion Score Study

The Residual Lesion Score Study is a prospective, multi-centre, observational cohort study conducted by the Pediatric Heart Network to assess the association between residual lesions following specified cardiovascular surgical operations and early and mid-term outcomes, with 1149 infants consented and enrolled at 17 centres between July 2015 and August 2017. The Residual Lesion Score Study combined two methods for data collection: (1) the traditional method of data collection utilised by the Pediatric Heart Network, which is done by trained research staff and (2) the extraction of existing local registry data already being collected at the sites for submission to the Society of Thoracic Surgeons–Congenital Heart Surgery Database. This was the first prospective study within the Pediatric Heart Network to pilot the use of registry data for a proportion of the study variables.

To verify the reliability of the local registry data for use in the Residual Lesion Score Study, the completeness and accuracy of the study variables of interest were examined through a retrospective audit of 500 patients at Pediatric Heart Network sites.¹⁴ The previously published results of this audit indicated that 94.7% of the local registry data elements of interest were both complete and accurate.¹⁴ This work was facilitated by the Integrated CARDiac Data and Outcomes Collaborative, which functions across the Pediatric Heart Network to integrate data sources to plan, implement, and conduct studies more efficiently.

Registry data

The Society of Thoracic Surgeons–Congenital Heart Surgery Database is the largest worldwide clinical data registry for congenital and pediatric heart surgery and includes perioperative data for all surgical cases performed at 129 participating centres from North America. Local registry data are collected by clinicians and/or trained data managers using standardised definitions and entered into compliant software for submission to the Society of Thoracic Surgeons–Congenital Heart Surgery Database. Data are submitted to the Society of Thoracic Surgeons–Congenital Heart Surgery Database data warehouse as part of regular data harvests and undergo a central validation process as well as site audits to ensure completeness and accuracy.^{15–17}

Process for use of the registry data in the Residual Lesion Score Study

Based on the previously published audit results,¹⁴ approximately 240 individual variables, which included demographics, pre-operative risk factors, procedure specific risk factors, operative characteristics, and major adverse events (approximately 10% of the total Residual Lesion Score Study variables), were selected for extraction from each site's local registry in the format designed for submission to the Society of Thoracic Surgeons–Congenital Heart Surgery Database. Among the study variables that were available in the local clinical registry, about 6% did not meet the reliability and completeness criteria and were therefore also collected manually by the site co-ordinators. The remaining study variables, such as echocardiographic variables, longitudinal outcomes, and other data that are not collected in the local registry, were obtained by chart review or from Residual Lesion Score Study-specific data collection forms completed at the time of surgery, site and core lab review of echocardiograms, or longitudinal follow-up.

Prior to study initiation, several different methods for extracting registry data were considered. The methodology promoting the greatest efficiency was thought to involve a direct feed to the Pediatric Heart Network Data Coordinating Center (which performed the data management and analysis for the Residual Lesion Score Study) from the Society of Thoracic Surgeons–Congenital Heart Surgery Database data warehouse, which receives and quality checks local registry data from each site. However, challenges related to potential cost, timing, and approval of such a design precluded the use of this method. Alternatively, the study team elected to work with each individual site to develop methods to extract local registry data from its Society of Thoracic Surgeons-compliant software.

In order to establish the appropriate data collection processes at the sites, study staff underwent centralised training on the protocol and data collection methods. Programming queries to extract specified data from each site's clinical registry in an identical format across 15 study sites using six different software packages was achieved after bi-monthly conference calls over a 6-month period. (Two of the 17 study sites entered all data directly into the Electronic Data Capture System and did not utilise registry data.) The queries, which were developed by programmers at the site or by the software vendors, were then tested at each site to ensure that data were accurately retrieved in the appropriate format. This process required several rounds of testing and revisions. For the Residual Lesion Score Study, research co-ordinators managed registry data collection for 1015/1149 enrolled patients. Table 1 shows enrolment by site. Cumulative registry data were extracted

Table 1. Study enrolment by site (n=1149).

Site	Patients enrolled
A*	121
B	153
C	68
D	89
E	68
F	92
G	14
H	82
I	56
J	29
K	136
L	20
M	93
N	19
O*	13
P	40
Q	56

*Sites A and O did not participate in the registry process.

monthly from sites for approximately 24 months. The data were reviewed at each site and then submitted to the data coordinating center where all data were merged and checked for missing and inconsistent data. As the clinical registry software was updated (once during the study period), the query required revision and retesting. Table 2 outlines the steps involved in the use of registry data for the Residual Lesion Score Study.

Survey methods

In order to understand staff perceptions about the process of utilising registry data in the Residual Lesion Score Study, a brief survey was developed and administered to each staff member involved in the data collection at 15 of the 17 clinical sites and the Pediatric Heart Network Data Coordinating Center. Two sites (one of which did not participate in the Society of Thoracic Surgeons-Congenital Heart Surgery Database) entered all data directly into the Electronic Data Capture system for the Residual Lesion Score Study and were therefore excluded from the survey. The survey was sent to principal investigators, co-investigators, research coordinators, registry data managers, and Pediatric Heart Network Data Coordinating Center staff via the Research Electronic Data Capture system in December 2017, with a 6-week response period, and two reminder e-mails being sent to non-responders during this window.¹⁸ The Pediatric Heart Network “Lead Co-ordinator” at each site was asked to complete an additional section about the processes; otherwise, all surveys were identical. Partially completed surveys were accepted. The survey sections are outlined below. (See Supplementary figure 1 for the full survey.)

Demographics included the respondents’ site and role in the Residual Lesion Score Study.

Process was completed by the lead co-ordinator at each site and gathered information about the steps required to use local registry

data, the staff involved in this process, problems encountered, and other practical issues.

Perceptions included Likert scale questions to assess staff perceptions about the time and training burden of using the local registry data and its reliability compared to data collected by study co-ordinators. The responses were rated on a five-point scale that included strongly agree, agree, neither agree nor disagree, disagree, and strongly disagree.

Recommendations were open-ended questions to address pros, cons, and recommendations for future studies using these methods.

The Nemours Cardiac Center site in Wilmington, Delaware, administered the staff survey; the Nemours Institutional Review Board reviewed the survey and determined that this did not constitute human patient research.

Analysis

Responses to the survey were summarised using frequencies by study role and compared using Kruskal–Wallis tests. Responses from open-ended (write-in) questions were described and summarised. All analyses were conducted using SAS v9.4 (SAS Institute Inc., Cary, NC, United States of America), and statistical significance was tested at level 0.05.

Results

The survey response rate was 98% (54/55) and included responses from one or more survey recipients at each of the 15 eligible centres as well as the data coordinating center. The distribution of respondents was as follows: 15 lead study co-ordinators (28%), 14 principal investigators (26%), 10 registry data managers (19%), 5 other study co-ordinators (9%), 5 co-investigators (9%), and 5 Pediatric Heart Network Data Coordinating Center staff (9%).

Process

The lead research co-ordinators reported that the monthly process to extract registry data, review results, remove protected health information, and upload data to the Pediatric Heart Network Data Coordinating Center involved one to four staff members at each site (Fig 1). A little over half (n = 8; 53%) stated that the time required to complete the registry process at the site each month was 30–90 minutes, with another two (13%) reporting times greater than 90 minutes (Fig 2). In addition, the research co-ordinators regularly reviewed and responded to queries concerning possible data discrepancies and missingness sent from the Pediatric Heart Network Data Coordinating Center.

Perceptions

Overall, 57% (n=31) of respondents agreed/strongly agreed that using local registry data *in addition to* standard chart abstraction saved the research staff time and 74% (n=40) agreed/strongly agreed that this process would save time in future Pediatric Heart Network studies. There were no significant differences across staff roles in response to these questions (Table 3). The majority (n=37; 71%) of respondents agreed/strongly agreed that using local registry data *instead of* routine data collection would save time in future studies (e.g. use of registry data for all study variables rather than a portion of the study). There was uniform agreement across study roles that using local registry data *instead*

Table 2. Steps for using clinical registry data for the RLS Study.

Pre-study processes	Processes for registry data extraction at sites	Processes at Data Coordinating Center
<ul style="list-style-type: none"> • Audit at study sites to assess completeness and accuracy of local registry data • Study protocol training and certification of all site staff including both the local registry team and study co-ordinators • Development of programming to extract local registry data in identical format across 15 study sites using six different software packages – supported through collaboration of the study team, PHN, software vendors, and local registry teams at the sites • Testing and revision of programming as needed before being released • For sites not participating in the audit, QC of the local clinical registry data for the first 10 patients was done to assess accuracy and completeness 	<ul style="list-style-type: none"> • Site-specific query used to extract registry data for consented patients monthly for duration of study (approximately 24 months) • Research co-ordinator at each site reviewed local registry data query results, removed PHI, uploaded results to the FTP site to share with the PHN DCC each month • Research co-ordinator at each site collected some pre-determined registry variables, as well as other non-registry variables through traditional chart review and entered into the EDC system • Sites were provided with monthly data discrepancy/missingness lists and asked to enter corrections into the EDC system • Sites resolved problems with registry data (e.g. variable names, formatting, and PHI) and resent data to DCC • As needed, programming revised as older database versions were converted to upgraded versions and during regularly scheduled STS-CHSD updates 	<ul style="list-style-type: none"> • DCC statistician manually reviewed the data files monthly from each of the 15 sites • DCC performed standard data checks for out-of-range data, potential spurious values, and missingness and issues were validated with the sites • DCC also notified sites of issues with the registry data that resulted in incorrect variable names or formats, and inclusion of PHI during transmission. • Registry data were converted to SAS and merged with other study data • Final registry data were compared with data from the EDC system and discrepancies were resolved with sites

DCC = Data Coordinating Center; EDC = Electronic Data Capture; FTP = File Transfer Protocol; PHI = Protected Health Information; QC = Quality Control; RLS = Residual Lesion Score; STS-CHSD = Society of Thoracic Surgeons Congenital Heart Surgery Database.

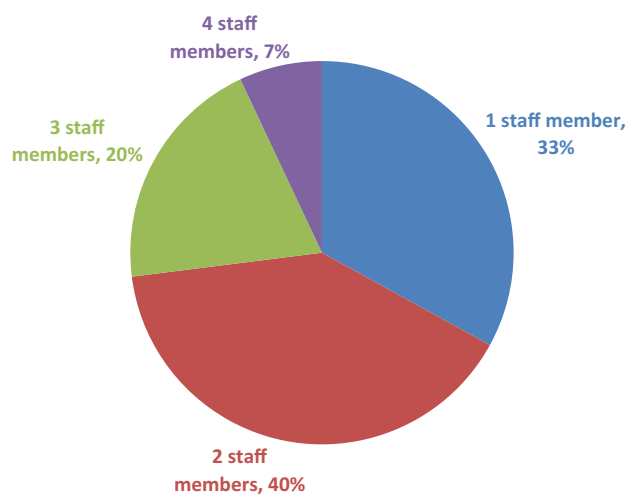


Figure 1. Staff needed each month at the study site to complete local registry data process. Percentage of lead study co-ordinators indicating the number of staff needed each month to complete the local registry process (extracting, reviewing, and uploading local registry data) at their site (n = 15).

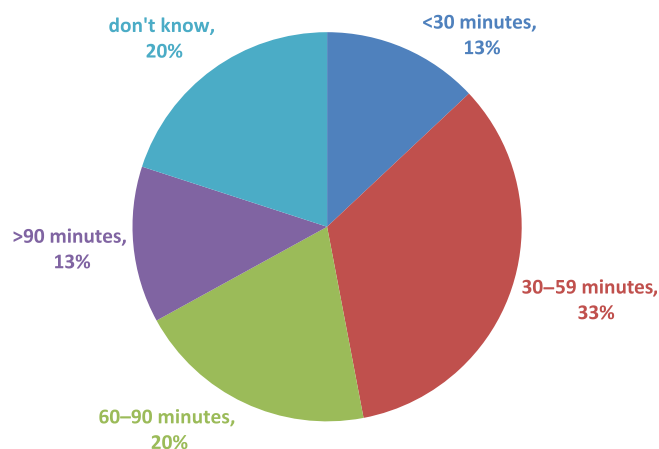


Figure 2. Time needed each month at the study site to complete local registry process. Percentage of lead study co-ordinators indicating the time needed each month to complete the local registry process (extracting, reviewing, and uploading local registry data) at their site (n=15).

of routine data collection would save time in future studies (Table 3).

Only 27% (n=14) of respondents agreed/strongly agreed that using the local registry data required a significant amount of additional training; however, more than half of the respondents (n = 29; 55%) agreed/strongly agreed that staff spent a significant amount of time developing and testing the registry programming to extract the data. There were no significant differences among staff roles for this question (Table 3). When asked about their perceptions of the reliability of clinical registry data, 27% (n = 14) of respondents agreed/strongly agreed that it was more reliable than data collection and entry by research co-ordinators. This included 70% (7/10) of the Society of Thoracic Surgeons database managers compared to 13–25% (7/42) of other study staff (p = 0.03).

Pros, cons, and recommendations identified by survey respondents

Pros, cons, and recommendations for using registry data were elicited from respondents in a series of open-ended questions. The most frequent responses are summarised as follows:

Pros identified by survey respondents:

- Using local registry data saved time and effort, particularly for the research co-ordinator, and eliminated the need for data collection and entry of those fields available in the local registry.
- The local registry data variables were well defined and consistent across sites providing reliable and accurate data.

Cons identified by survey respondents:

- Some sites did not routinely collect all of the registry data fields applicable to the study, which led to missing data

Table 3. Staff perceptions of using local clinical registry data from the STS-CHSD by study role, frequency, and percent.

		Co-investigator	Data Coordinating Center staff	Lead research co-ordinator	Principal investigator	Registry data manager	Other research co-ordinator	p Value
Response rate		5/5 100	5/5 100	15/15 100	14/15 93	10/10 100	5/5 100	
Using the STS registry data <i>in combination with</i> medical record extraction in the PHN RLS Study has saved the research staff time	Strongly disagree	0 0.00	2 40.00	0 0.00	0 0.00	0 0.00	0 0.00	0.22
	Disagree	0 0.00	1 20.00	3 20.00	2 14.29	1 10.00	1 20.00	
	Neither disagree nor agree	4 80.00	0 0.00	3 20.00	4 28.57	0 0.00	2 40.00	
	Agree	0 0.00	2 40.00	7 46.67	3 21.43	7 70.00	2 40.00	
	Strongly agree	1 20.00	0 0.00	2 13.33	5 35.71	2 20.00	0 0.00	
Using registry data <i>in combination with</i> medical record extraction in <i>future</i> PHN studies will save the research staff time	Strongly disagree	0 0.00	1 20.00	0 0.00	0 0.00	0 0.00	0 0.00	0.11
	Disagree	0 0.00	2 40.00	1 6.67	1 7.14	0 0.00	0 0.00	
	Neither disagree nor agree	2 40.00	0 0.00	4 26.67	1 7.14	1 10.00	1 20.00	
	Agree	2 40.00	2 40.00	9 60.00	8 57.14	7 70.00	4 80.00	
	Strongly agree	1 20.00	0 0.00	1 6.67	4 28.57	2 20.00	0 0.00	
Using registry data <i>instead of</i> medical record extraction in <i>future</i> PHN studies will save the research staff time	Strongly disagree	0 0.00	0 0.00	0 0.00	1 7.69	0 0.00	0 0.00	0.94
	Disagree	0 0.00	0 0.00	1 6.67	0 0.00	0 0.00	0 0.00	
	Neither disagree nor agree	1 25.00	1 20.00	4 26.67	2 15.38	3 30.00	2 40.00	
	Agree	3 75.00	4 80.00	8 53.33	6 46.15	5 50.00	2 40.00	
	Strongly agree	0 0.00	0 0.00	2 13.33	4 30.77	2 20.00	1 20.00	
Research staff completed a significant amount of additional training in order to be able to use the STS registry data for the RLS Study	Strongly disagree	0 0.00	0 0.00	2 13.33	2 14.29	0 0.00	0 0.00	0.43
	Disagree	2 50.00	0 0.00	4 26.67	3 21.43	3 30.00	3 60.00	
	Neither disagree nor agree	1 25.00	3 60.00	8 53.33	2 14.29	5 50.00	1 20.00	
	Agree	1 25.00	1 20.00	1 6.67	6 42.86	2 20.00	0 0.00	
	Strongly agree	0 0.00	1 20.00	0 0.00	1 7.14	0 0.00	1 20.00	
Research staff spent a significant amount of additional time preparing and finalising the query in order to use the STS registry data for the RLS Study	Strongly disagree	0 0.00	0 0.00	0 0.00	2 14.29	0 0.00	0 0.00	0.14
	Disagree	0 0.00	0 0.00	2 13.33	2 14.29	1 10.00	0 0.00	
	Neither disagree nor agree	3 75.00	1 20.00	4 26.67	2 14.29	6 60.00	1 20.00	
	Agree	1 25.00	1 20.00	5 33.33	7 50.00	1 10.00	1 20.00	
	Strongly agree	0 0.00	3 60.00	4 26.67	1 7.14	2 20.00	3 60.00	

(Continued)

Table 3. (Continued)

		Co-investigator	Data Coordinating Center staff	Lead research co-ordinator	Principal investigator	Registry data manager	Other research co-ordinator	p Value
The STS registry data is more reliable than the medical record extraction data at most sites	Strongly disagree	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00	0.034
	Disagree	0 0.00	3 60.00	2 13.33	2 14.29	0 0.00	0 0.00	
	Neither disagree nor agree	3 75.00	1 20.00	1 066.67	1 178.57	4 40.00	3 60.00	
	Agree	1 25.00	1 20.00	3 20.00	0 0.00	4 40.00	2 40.00	
	Strongly agree	0 0.00	0 0.00	0 0.00	1 7.14	2 20.00	0 0.00	
The STS registry data is more reliable than the medical record extraction data at your site	Strongly disagree	0 0.00	0 0.00	0 0.00	1 7.14	0 0.00	0 0.00	0.026
	Disagree	1 25.00	2 50.00	6 40.00	3 21.43	0 0.00	1 20.00	
	Neither disagree nor agree	2 50.00	1 25.00	7 46.67	8 57.14	3 30.00	3 60.00	
	Agree	1 25.00	1 25.00	2 13.33	1 7.14	4 40.00	1 20.00	
	Strongly agree	0 0.00	0 0.00	0 0.00	1 7.14	3 30.00	0 0.00	

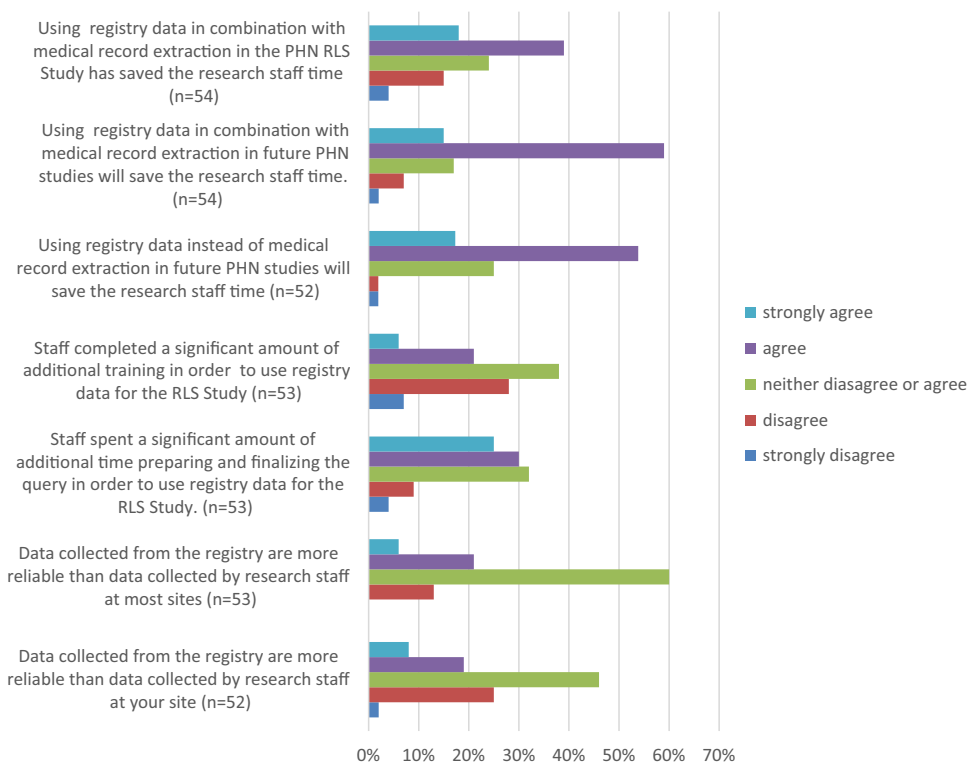


Figure 3. Staff perceptions of using local clinical registry data from the Society of Thoracic Surgeons Congenital Heart Surgery Database. PHN = Pediatric Heart Network; RLS = Residual Lesion Score.

that subsequently had to be manually collected by the co-ordinator.

- The programing of local data abstraction was complicated, time-consuming, and involved multiple staff at each site to test

and finalise the process. Multiple software platforms were involved, and extraction programs had to be updated whenever new versions of the software were released. Early in the study, several sites experienced technical difficulties uploading registry

data to the website at the Pediatric Heart Network Data Coordinating Center, which required time to resolve.

- Using local registry data resulted in extra steps for the individual sites as well as for the data coordinating center staff. The data coordinating center had to manage two completely different processes for data collection and cleaning.
- Initially, sites submitted local registry data twice per year to The Society of Thoracic Surgeons-Congenital Heart Surgery Database. This was based on bi-annual deadlines and harvest schedules for the local registry data and did not correspond to monthly submissions of local registry data to the Pediatric Heart Network Data Coordinating Center. Therefore, some local teams had to alter their data collection and cleaning processes for study patients.
- In the processes utilised for the Residual Lesion Score Study, coordinators were responsible for manually stripping protected health information from local registry data prior to sending to the Pediatric Heart Network Data Coordinating Center; this resulted in cases of inadvertent disclosure of protected health information by sites.

Recommendations identified by survey respondents:

- Stakeholders should be involved early and throughout the design and implementation of this methodology.
- Methods to simplify the programming and processes to extract registry data should be considered.
- As appropriate, less frequent registry data extractions could save time for both the sites and the Data Coordinating Center; however, this decrease in frequency of data extraction may not be feasible when data are needed in near real time.
- Consideration should be given to the unique aspects of a clinical registry, including data collection processes and timelines.
- Strategies should be developed to manage protected health information appropriately; processes should be automated as appropriate to avoid human error.
- Registry data are most valuable for studies in which it will be the main source of data.

Discussion

The Residual Lesion Score Study served as a pilot for the Pediatric Heart Network to assess the feasibility of using local registry data for a proportion of study variables. Overall, staff perceived that the local registry could be used as a reliable source for obtaining research data and that it saved time for research coordinators by eliminating the need for data collection and entry for approximately 10% of the study variables. The survey respondents also identified several challenges associated with using local registry data in a prospective, multi-centre study.

Study design

Our survey results highlight the significant investment of time and resources necessary upfront to plan and execute this type of design. As reported by others, collaboration across multiple stakeholders was key.^{11,19} In the Residual Lesion Score Study, this involved engagement of individuals across the network conducting the study, registry experts, teams at the local site, and industry representatives from various database software companies. It is important to recognise that while gains from this type of research design may be seen at the site level, they come at a potential cost related to

the collaboration and effort needed upfront for study design and data management efforts. In our case, many of the individuals involved generously volunteered their time. These factors should be considered when setting up study timelines and budgets, and there should be enough variables collected from the clinical registry so that the process adds value.

Process for extracting and integrating registry data

Our study demonstrates some of the challenges related to extracting local registry data at the site level. This challenge was due in part to the existence of multiple software platforms for data collection within and across sites, as well as differences across sites in personnel and resources related to registry data management and expertise.

Several methodological options can aid in addressing these challenges. First, in cases where data extraction from local sites is still required, a standard program has recently been developed that can be uniformly applied across different sites and different software platforms to automatically extract local surgical registry data, strip protected health information, and produce a standardised data extract (M. Boskovski, personal communication 30 May, 2018 via conference call). This method was successfully utilised in a recent study conducted by the Pediatric Cardiac Genomics Consortium, which merged data from local surgical registries at study sites with genetic data to evaluate the impact of copy number variants on outcomes in children undergoing heart surgery. This approach could cut down significantly on the time and effort necessary by data coordinating centers for data cleaning and could also eliminate issues of inadvertent sharing of protected health information.

The ideal design to maximise efficiency would likely involve direct extraction of registry data from the central registry data warehouse. This strategy would minimise burden on individual sites and on the study analytic and data management team, as registry data extraction could occur through a single centralised process by registry experts after data cleaning was performed. This strategy would also accrue the full benefit of all data quality measures employed by the central registry warehouse. Previously, these methods have been used successfully in the pediatric cardiovascular population to support the conduct of the Vasoactive-Inotropic Score Study, which utilised data from the Pediatric Cardiac Critical Care Consortium Registry and in an ongoing clinical trial: Steroids to Reduce Systemic Inflammation after Neonatal Heart Surgery Trial.^{11,20,21} This strategy has also been used in adult cardiovascular disease trials.¹⁹ It is important to note that while more efficient, this methodology may involve costs that would need to be integrated into the overall study budget. There may also be potential challenges with data sharing.

The potential efficiencies realised with utilising clinical registry data are also likely most apparent when they are used for all or nearly all of the data collection for the study. Both our quantitative and qualitative survey data consistently identified this theme. In this pilot phase, only approximately 10% of study variables could be included from the registry data, but the other types of studies have been performed using a much higher percentage of study variables. For example, the Thrombus Aspiration during ST-Elevation Myocardial Infarction in Scandinavia study was a multi-centre trial, which reported the use of registry data for all study variables, with substantial cost savings.^{22,23} The Study of Access Site for Enhancement of Percutaneous Coronary Intervention for Women collected a large proportion of study

variables from a clinical registry and reported a decrease in coordinator workload by approximately 65%.¹⁹ Linking multiple databases and registries may also maximise the number of variables available and further increase efficiency.⁵

Our findings highlight the reality that managing multiple data sources is challenging and requires additional steps for the clinical sites and the study data coordinating center. For the Residual Lesion Score Study, sites extracted registry data regularly over approximately 24 months and the process involved about 30–60 minutes per month at many sites. While this may seem like a small investment of time, it is important to emphasise that this only accounted for approximately 10% of the study variables and does not take into consideration time spent completing other study requirements. Additionally, the Pediatric Heart Network Data Coordinating Center staff survey responses were less favorable overall than those of the clinical site staff. Although the perception was that this process saved time for the research staff, respondents from the data coordinating center perceived that a greater amount of time was needed to manage two separate methods of data collection. The potential for increased burden on the data coordinating center was unexpected and was not accounted for in the study budget or staffing. Impact on the data coordinating center was highest early in the study, as problems with the registry data were identified and had to be resolved. Additional data checks were required at the end of the study to compare some elements of the clinical registry data with data also collected in the Electronic Data Capture system for the same or related data elements such as non-matching data or data for events that were expected to occur. For example, if data elements were originally missing in the registry, the site was instructed to enter them into the Electronic Data Capture system; if these data later became available in the registry, they were cross-checked. While this study did not collect the actual time spent by all study personnel, it would be important for studies considering this approach to understand that the amount of time spent may increase for some roles, while decreasing for others. Some of these challenges may be mitigated by optimising the design, data flow, and data management strategies as described above.

Nuances of registry data collection

It is essential to understand the nuances of the specific registries that will be utilised, including timing of registry data collection and submission, data definitions, missingness, and accuracy of requisite data fields. For example, in the Residual Lesion Score Study, monthly data submission was desirable for study purposes, but the local clinical registry data used in the study were only submitted twice a year to the Society of Thoracic Surgeons-Congenital Heart Surgery data warehouse. The need for monthly submission of data for the Residual Lesion Score Study required some local teams to alter their data collection and cleaning processes for study patients. As conveyed in our survey results, less frequent study data submissions would decrease this additional effort both at the site and at the data coordinating center and may be most efficient with a single data extract from the registry data warehouse. However, in some contexts, such as during certain types of clinical trials, less frequent submission of study data may not be feasible and more “real time” data may be necessary to assess patient eligibility or adverse events. Several registries now allow for real-time submission and analysis of data; in fact, the Society of Thoracic Surgeons transitioned to a “continuous harvest” in 2017 with capabilities for near real-time submission of data.

Most registries also have their own set of unique standards for data variables and definitions, data quality checks, type of staff entering data (clinical versus administrative), auditing procedures, and other processes, which can all affect the quality of the data.⁹ All data collection processes can be prone to error, and data quality can vary across registries, sites, and staff. According to our survey, the majority of registry data managers perceived that data from the registry are more reliable than data collected by the research staff, whereas a fair number of research staff disagreed. It is likely that each group was biased towards its own process and may have lacked an understanding of the other’s procedures and training for ensuring data reliability.

To increase data accuracy, study variables not meeting adequate completeness based on the audit study¹⁴ were collected by both registry extract and site co-ordinators. The data coordinating center then compared the data from these two sources and issued queries for mismatched data. Additionally, the sites were queried for data missing in the registry. These additional data checks added to site and data coordinating center burden but increased data quality. Audits may be used after a study is initiated to confirm data quality, especially for key variables, but care should be taken to balance this additional burden with the desire for data quality.

Limitations

The site survey had a high response rate but was limited to a single study conducted by the Pediatric Heart Network, and the information gathered may not be fully applicable across other settings. The survey was administered between December 2017 and January 2018. Residual Lesion Score Study enrolment was completed in August 2017, with final clinical registry data extraction completed in January 2018. Respondents may not have recalled the details of processes used during initial query development and data extraction and may have answered questions differently had the survey been administered earlier in the study rather than towards the end. Conversely, respondents may also have answered differently had the survey been administered later in the study, as the Pediatric Heart Network Data Coordinating Center issued many additional data queries during final data cleaning. While there was limited staff turnover during the Residual Lesion Score Study, the survey may not have adequately captured the full experience or perceptions at sites that did experience turnover.

Implications

Despite the challenges identified and the amount of time invested prior to launch of the Residual Lesion Score Study, most staff perceived that this “hybrid” approach to data collection leverages local registry data and saves time. Most staff also believed that studies embedded completely within a registry would save even more time. Future studies utilising registry data should (1) engage study team members and other stakeholders when designing the study, (2) consider the best approach and timing for extracting registry data while adhering to study timelines and protecting health information, and (3) understand the nuances of the clinical registry and how they impact the research study. Efforts geared towards automating and centralising data management processes for studies using registry data may aid in further optimising this methodology for future studies.

Acknowledgments. We would like to thank the registry data managers and the software vendors who graciously volunteered their time to support the Residual Lesion Score Study. We would also like to acknowledge the research staff at the

clinical sites and the data coordinating center for their time and effort completing the survey for this project and their valuable work on the Residual Lesion Score Study.

Financial Support. The study was supported by grants (U24HL135691, U10HL068270, HL109818, HL109778, HL109816, HL109743, HL109741, HL109673, HL068270, HL109781, HL135665 and HL135680) from the National Heart, Lung, and Blood Institute, National Institutes of Health. Meena Nathan was supported by a K23 grant (NHLBI/NIH HL119600). Brett Anderson was supported by a K23 grant (NHLBI/NIH HL133454). The contents of this work are solely the responsibility of the authors and do not necessarily represent the official views of the National Heart, Lung, and Blood Institute.

Conflicts of Interest. Jeffrey P. Jacobs, MD is Chair of The Society of Thoracic Surgeons Workforce on National Databases. Eric Graham serves as a research consultant for Bayer.

Supplementary material. To view supplementary material for this article, please visit <https://doi.org/10.1017/S1047951119001148>.

References

- Pasquali SK, Jacobs JP, Farber GK, et al. Report of the National Heart, Lung, and Blood Institute working group: an integrated network for congenital heart disease research. *Circulation* 2016; 133: 1410–1418.
- James S, Rao SV, Granger CB. Registry-based randomized clinical trials – a new clinical trial paradigm. *Nat Rev Cardiol* 2015; 12: 312–316.
- Jones WS, Roe MT, Antman EM, et al. The changing landscape of randomized clinical trials in cardiovascular disease. *J Am Coll Cardiol* 2016; 68: 1898–1907.
- Roe MT, Mahaffey KW, Ezekowitz JA, et al. The future of cardiovascular clinical research in North America and beyond-addressing challenges and leveraging opportunities through unique academic and grassroots collaborations. *Am Heart J* 2015; 169: 743–750.
- Vener DF, Gaies M, Jacobs JP, Pasquali SK. Clinical databases and registries in congenital and pediatric cardiac surgery, cardiology, critical care, and anesthesiology worldwide. *World J Pediatr Congenit Heart Surg* 2017; 8: 77–87.
- Zannad F, Pfeffer MA, Bhatt DL, et al. Streamlining cardiovascular clinical trials to improve efficiency and generalisability. *Heart* 2017; 103: 1156–1162.
- Lauer MS, D'Agostino RB, Sr. The randomized registry trial – the next disruptive technology in clinical research? *N Engl J Med* 2013; 369: 1579–1581.
- Jacobs ML, Jacobs JP, Hill KD, et al. The Society of Thoracic Surgeons Congenital Heart Surgery Database: 2017 update on research. *Ann Thorac Surg* 2017; 104: 731–741.
- Riehle-Colarusso TJ, Bergersen L, Broberg CS, et al. Databases for congenital heart defect public health studies across the lifespan. *J Am Heart Assoc* 2016; 5.
- Frobert O, Lagerqvist B, Olivecrona GK, et al. Thrombus aspiration during ST-segment elevation myocardial infarction. *N Engl J Med* 2013; 369: 1587–1597.
- Gaies MG, Jeffries HE, Niebler RA, et al. Vasoactive-inotropic score is associated with outcome after infant cardiac surgery: an analysis from the pediatric cardiac critical care consortium and virtual PICU system registries. *Pediatr Crit Care Med* 2014; 15: 529–537.
- Rao SV, Hess CN, Barham B, et al. A registry-based randomized trial comparing radial and femoral approaches in women undergoing percutaneous coronary intervention: the SAFE-PCI for women (study of access site for enhancement of PCI for women) trial. *JACC Cardiovasc Interv* 2014; 7: 857–867.
- Mahony L, Sleeper LA, Anderson PA, et al. The pediatric heart network: a primer for the conduct of multicenter studies in children with congenital and acquired heart disease. *Pediatr Cardiol* 2006; 27: 191–198.
- Nathan M, Jacobs ML, Gaynor JW, et al. Completeness and accuracy of local clinical registry data for children undergoing heart surgery. *Ann Thorac Surg* 2017; 103: 629–636.
- Clarke DR, Breen LS, Jacobs ML, et al. Verification of data in congenital cardiac surgery. *Cardiol Young* 2008; 18: 177–187.
- Jacobs JP, Jacobs ML, Mavroudis C, et al. Nomenclature and databases for the surgical treatment of congenital cardiac disease – an updated primer and an analysis of opportunities for improvement. *Cardiol Young* 2008; 18: 38–62.
- The Society for Thoracic Surgeons. STS Congenital Heart Surgery Database Data Specifications Version 3.22. 2013; https://www.sts.org/sites/default/files/documents/CongenitalDataSpecsV3_22.pdf Accessed 18 June, 2018.
- Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap) – a metadata-driven methodology and workflow process for providing translational research informatics support. *J Biomed Inform* 2009; 42: 377–381.
- Hess CN, Rao SV, Kong DF, et al. Embedding a randomized clinical trial into an ongoing registry infrastructure: unique opportunities for efficiency in design of the Study of access site for enhancement of percutaneous coronary intervention for women (SAFE-PCI for women). *Am Heart J* 2013; 166: 421–428.
- STeroids to REduce Systemic Inflammation after Neonatal Heart Surgery. <https://ClinicalTrials.gov/show/NCT03229538>. Accessed 8 January, 2018.
- Hill KD, Kannankeril PJ. Perioperative corticosteroids in children undergoing congenital heart surgery: five decades of clinical equipoise. *World J Pediatr Congenit Heart Surg* 2018; 9: 294–296.
- Frobert O, Lagerqvist B, Gudnason T, et al. Thrombus aspiration in ST-elevation myocardial infarction in Scandinavia (TASTE trial). A multicenter, prospective, randomized, controlled clinical registry trial based on the Swedish angiography and angioplasty registry (SCAAR) platform. Study design and rationale. *Am Heart J* 2010; 160: 1042–1048.
- Wachtell K, Lagerqvist B, Olivecrona GK, James SK, Frobert O. Novel trial designs: lessons learned from thrombus aspiration during ST-segment elevation myocardial infarction in Scandinavia (TASTE) trial. *Curr Cardiol Rep* 2016; 18: 11.