Cardiology in the Young

cambridge.org/cty

Original Article

Cite this article: Anderson BR, Gotlieb EG, Hill K, McHugh KE, Scheurer Mark A., Mery CM, Pelletier GJ, Kaltman JR, White OJ, Trachtenberg FL, Hollenbeck-Pringle D, McCrindle BW, Sylvester DM, Eckhauser AW, Pasquali SK, Anderson JB, Schamberger MS, Shashidharan S, Jacobs JP, Jacobs ML, Boskovski M, Newburger JW, Nathan M, and on behalf of the Pediatric Heart Network (2020) Registry-based trials: a potential model for cost savings? *Cardiology in the Young* **30**: 807–817. doi: 10.1017/S1047951120001018

Received: 7 April 2020
Revised: 9 April 2020
Accepted: 9 April 2020
First published online: 8 May 2020

Keywords:
Registry-based trial; cost savings; pragmatic trial; trial design

Author for correspondence:
B. Anderson, MD, MBA, MS, NewYork-Presbyterian/Morgan Stanley Children's Hospital, 3959 Broadway, CH-2 North, New York, NY 10032-3784, USA. Tel: +1 (212) 305 8509; Fax: +1 (212) 305 4429. E-mail: bra2113@cumc.columbia.edu

*A complete list of non-author contributors appears in Appendix 1.

Registry-based trials: a potential model for cost savings?

Brett R. Anderson[1], Evelyn G. Gotlieb[2], Kevin Hill[3], Kimberly E. McHugh[4], Mark A. Scheurer[4], Carlos M. Mery[5], Glenn J. Pelletier[6], Jonathan R. Kaltman[7], Owen J. White[8], Felicia L. Trachtenberg[9], Danielle Hollenbeck-Pringle[9], Brian W. McCrindle[10], Donna M. Sylvester[11], Aaron W. Eckhauser[12], Sara K. Pasquali[13], Jeffery B. Anderson[14], Marcus S. Schamberger[15], Subhadra Shashidharan[16], Jeffrey P. Jacobs[17], Marshall L. Jacobs[18], Marko Boskovski[19], Jane W. Newburger[20,21], Meena Nathan[21,22] and on behalf of the Pediatric Heart Network*

[1]Division of Pediatric Cardiology, NewYork-Presbyterian/Morgan Stanley Children's Hospital, Columbia University Irving Medical Center, New York, NY, USA; [2]The Wharton School, University of Pennsylvania, Philadelphia, PA, USA; [3]Department of Pediatrics, Duke Clinical Research Institute, Duke University Medical Center, Durham, NC, USA; [4]Department of Pediatrics, Medical University of South Carolina, Charleston, SC, USA; [5]Texas Center for Pediatric and Congenital Heart Disease, Dell Children's Medical Center, University of Texas Dell Medical School, Austin, TX, USA; [6]Division of Cardiac Surgery, Nemours Cardiac Center, Alfred I duPont Hospital for Children, Wilmington, DE, USA; [7]National Heart Lung and Blood Institute, National Institutes of Health, Bethesda, MD, USA; [8]CardioAccess, Fort Lauderdale, FL, USA; [9]New England Research Institutes, Watertown, MA, USA; [10]Department of Cardiology at SickKids, Labatt Family Heart Centre, University of Toronto, The Hospital for Sick Children, Toronto, Canada; [11]Department of Cardiology, The Children's Hospital of Philadelphia, Philadelphia, PA, USA; [12]Section of Pediatric Cardiothoracic Surgery, Division of Cardiothoracic Surgery, Department of Surgery, University of Utah, Salt Lake City, UT, USA; [13]Division of Pediatric Cardiology, C.S. Mott Children's Hospital, Ann Arbor, MI, USA; [14]Division of Pediatric Cardiology, Cincinnati Children's Hospital and Medical Center, Cincinnati, OH, USA; [15]Division of Pediatric Cardiology, Riley Children's Hospital, Indianapolis, IN, USA; [16]Division of Cardiovascular Surgery, Children's Healthcare of Atlanta, Atlanta, GA, USA; [17]Division of Cardiovascular Surgery, Department of Surgery, Johns Hopkins All Children's Heart Institute, St. Petersburg, FL, USA; [18]Division of Cardiac Surgery, Department of Surgery, Johns Hopkins School of Medicine, Baltimore, MD, USA; [19]Division of Cardiac Surgery, Harvard Medical School, The Brigham and Women's Hospital, Boston, MA, USA; [20]Department of Cardiology, Boston Children's Hospital, Boston, MA, USA; [21]Department of Pediatrics, Harvard Medical School, Boston, MA, USA and [22]Department of Cardiac Surgery, Boston Children's Hospital, Boston, MA, USA

Abstract

Background/Aims: Registry-based trials have emerged as a potentially cost-saving study methodology. Early estimates of cost savings, however, conflated the benefits associated with registry utilisation and those associated with other aspects of pragmatic trial designs, which might not all be as broadly applicable. In this study, we sought to build a practical tool that investigators could use across disciplines to estimate the ranges of potential cost differences associated with implementing registry-based trials versus standard clinical trials. *Methods:* We built simulation Markov models to compare unique costs associated with data acquisition, cleaning, and linkage under a registry-based trial design versus a standard clinical trial. We conducted one-way, two-way, and probabilistic sensitivity analyses, varying study characteristics over broad ranges, to determine thresholds at which investigators might optimally select each trial design. *Results:* Registry-based trials were more cost effective than standard clinical trials 98.6% of the time. Data-related cost savings ranged from $4300 to $600,000 with variation in study characteristics. Cost differences were most reactive to the number of patients in a study, the number of data elements per patient available in a registry, and the speed with which research coordinators could manually abstract data. Registry incorporation resulted in cost savings when as few as 3768 independent data elements were available and when manual data abstraction took as little as 3.4 seconds per data field. *Conclusions:* Registries offer important resources for investigators. When available, their broad incorporation may help the scientific community reduce the costs of clinical investigation. We offer here a practical tool for investigators to assess potential costs savings.

© The Author(s), 2020. Published by Cambridge University Press.

Many clinical decisions in medicine are made with either low-quality evidence or no evidence at all.[1] For adult cardiovascular disease, for example, only 11% of clinical guideline recommendations are Level of Evidence A.[2] Both the National Institutes of Health and the Patient-Centered Outcomes Research Institute have recognised the critical need for more randomised clinical trials, but highlight concerns about expenses associated with such investigations (https://rethinkingclinicaltrials.org/).

In 2013, when Fröbert and colleagues published the Thrombus Aspiration during ST Segment Elevation Myocardial Infarction (TASTE) trial, they raised the scientific community's awareness of what is now known as "registry-based trials" or "trials within a registry", a type of pragmatic study design that leverages existing registry infrastructure and data to streamline patient accrual and data collection and promote inclusion of patients from broad and heterogeneous populations.[3] That same year, the Study of Access Site for Enhancement of Percutaneous Coronary Intervention (SAFE-PCI)[4] and the Randomized Evaluation of Decolonization versus Universal Clearance to Eliminate MRSA (REDUCE MRSA) trial[5] two additional high-profile, registry-based trials, were published. Editorials estimated study expenses as low as $20–50 per subject[3-6] – massive departures from the $2000 to $40,000 per subject typically estimated for standard trials.[7-9] Registry-based trials were quickly hailed as "the next disruptive technology in clinical research",[10] lauded for decreasing study costs and increasing generalisability.[11-13] Yet, these landmark investigations employed other pragmatic, cost-saving methodologies beyond the incorporation of registry data.[14] They tested focused, resource-limited interventions that occurred within the scope and flow of clinical practice and whose mediators and effects were both already measured accurately and entirely within one single national registry with robust and validated data. While these studies offer tremendous lessons in creativity for trialists, not all questions lend themselves to pragmatic designs, and reported costs associated with these trials were based on short-hand calculations, leaving room for interpretation. It is not known the extent to which incorporation of registry data alone might reduce study costs and the range of studies for which this strategy might be less costly.

In this study, we sought to build a practical tool that investigators could use to estimate the ranges of potential cost savings associated with implementing registry-based trials versus standard clinical trials and to assess the trial characteristics driving these cost differences. We focused on costs associated with data acquisition, cleaning, and linkage, across a broad range of hypothetical trial characteristics, to provide concrete data for investigators across disciplines. Distribution-based inputs and sensitivity analyses were used to estimate potential costs. Potential cost savings associated with trial design or recruitment were considered supplementary and beyond the scope of this analysis.

Methods

Study design

We built a Markov economic model to analyse ranges of potential cost differences associated with data acquisition, cleaning, and linkage. We compared costs unique to registry-based trials versus standard clinical trial implementation. Study costs common to both designs (such as those associated with intervention, randomisation, contract execution, safety monitoring, follow-up imaging, or interviews) were excluded. Costs that would be incurred even in the absence of study participation were also excluded. These included registries' set-up costs, dues, data abstraction, etc. Further, we assumed centres have staff versed in chart abstraction. Sensitivity analyses accounted for uncertainty in model inputs, testing assumptions over broad ranges with probabilistic distributions. These incorporate, for example, differences in local electronic medical records, staff experience, wages, etc. Supplemental Figure 1 delineates the cost difference calculations. Supplemental Figure 2 provides a pictorial representation of tested models.

Cost savings associated with the use of registries in trial design and recruitment were considered to be highly study specific. Thus, while potentially important, they were considered to be separate and supplemental to this analysis. For example, trials on rare disease that might have relied on expensive advertisements with uncertain reach might derive great benefit from relying instead on a disease-specific registry. Alternately, a trial predicated on the recruitment of an already captive population – such as hospitalised patients – might find little to no utility in registry utilisation. We encourage investigators to consider these aspects of registry utilisation as they conceptualise their trials, as well.

Data sources

Because no published standards exist regarding the amount of time it takes to collect and manage research data, estimates of time were obtained through a series of semi-structured interviews with investigators, research coordinators, data managers from 18 academic institutions, programmers from four registry-based investigations,[15-18] and statisticians and project managers at HealthCore (formerly The New England Research Institutes), a contract research organisation with expertise in all aspects of trial implementation, including trial design, project management, data coordination and integration, research analytics, and quality assurance (www.healthcore.com). Amongst others, HealthCore coordinates trial activities for the National Heart Lung and Blood Institute-sponsored Pediatric Heart Network, which has led over 30 paediatric cardiovascular multi-centre investigations over the past 17 years (see Appendix 1 and Supplemental Figure 3).

Salary estimates were obtained via Glassdoor, using 2018 national averages and distributions and calculations of percent effort based on the time spent on each task, assuming a 40-hour work week. Estimates were validated by comparing against HealthCore's contracting records. It was assumed that physician investigator time was similar across study types and thus excluded from the models. Fringe and indirect cost rates were estimated based on discussions with the Department of Health and Human Services and manual review of publicly disclosed rate agreements from 40 institutions across the United States. All model inputs and data sources are detailed in Table 1.

Methods of data acquisition

We tested four methods of data acquisition in our models, one representing a standard clinical trial design and three representing potential approaches to obtaining registry data. Figure 1 characterises the flow of data to registries – from patient interviews and electronic medical records to locally held registry-specific databases and finally to centralised (multi-centre) data repositories – and the possible mechanisms for data acquisition.

Standard clinical trial

For standard clinical trials, data are obtained and entered manually by research assistants, nurses, and/or research coordinators at participating centres (heretofore referred to collectively as "coordinators"), either through patient interviews or review of original source records. We assumed that some errors would be made during manual data entry that might be identified and corrected during data cleaning (such as missing data and data out of range). We assumed that data coordinating centre analysts (heretofore referred to as "statisticians") would spend time identifying errors and that study coordinators at each centre would review source documents and adjudicate these errors.

https://doi.org/10.1017/S1047951120001018 Published online by Cambridge University Press

Table 1. Input values and data sources used in base case models and sensitivity analyses.

Description	Base case	Low	High	Distribution	Data source
Costs and other resources					
Registry-related costs/resources					
Programmer salary	$74,831	$37,746	$150,986	Gamma Mean: 74,831 SD: 8000	Glassdoor (USA)
Programmer time to write initial query	4.2 hours*	2.75 hours	11 hours	Normal Mean: 5.5 SD: 0.5	Informed by interviews with programmers
Programmer time to implement query per centre	2 hours*	0.67 hours	5 hours	Normal Mean: 2.5 SD: 1.25	Informed by interviews with programmers
Programmer support time per centre	2 hours	0.5 hours	4 hours	Normal Mean: 2.0 SD: 0.5	Informed by interviews with programmers
Data manager salary	$69,638	$36,292	$145,168	Gamma Mean: 69,638 SD: 9000	Glassdoor (USA)
Statistician salary	$80,203	$40,101	$160,406	Gamma Mean: 80,203 SD: 10,500	Glassdoor (USA)
Research coordinator time to upload the query	0.33 hours	0.25 hours	0.42 hours	Normal Mean: 0.33 hour SD: 0.08 hour	Informed by interviews with research coordinators
Number of software platforms per registry accessed	1	1	100	Uniform	Informed by interviews with investigators
Data coordinating centre analyst (statistician) time to write code to clean and link data registry case	30 hours*	10 hours	80 hours	Normal Mean: 40 SD: 5	Informed by interviews with trial statisticians
Individual centre data manager time to run query	1.0 hour	0.2 hour	3.0 hour	Normal Mean: 1.0 SD: 0.25	Informed by interviews with data managers
Number of centres involved in the study	10	1	200	Uniform	Informed by interviews with investigators
Data manager time to run the query	0.76 hours*	0.2 hours	3.0 hours	Normal Mean: 1.0 SD: 0.25	Informed by interviews with data managers
Cost of accessing data registry	$0	$0	$500,000	Uniform	Informed by interviews with investigators
Standard abstraction-related costs/resources					
Clinical research coordinator salary	$57,596	$29,407	$117,628	Gamma Mean: 57,596 SD: 9000	Glassdoor (USA)
Statistician salary	$80,203	$40,101	$160,406	Gamma Mean: 80,203 SD: 10,500	Glassdoor (USA)
Data manager salary	$69,638	$36,292	$145,168	Gamma Mean: 69,638 SD: 9000	Glassdoor (USA)
Statistician time to clean data in standard case	30 hours	15 hours	60 hours	Normal Mean: 30 SD: 5	Informed by interviews with trial statisticians
Data manager time to clean data in standard case	10 hours	5 hours	20 hours	Normal Mean: 10 SD: 1.5	Informed by interviews with trial statisticians
Research coordinator time to manually enter data (per patient per variable)	1.5 minutes	10 seconds	15 minutes	Beta Mean: 0.025 hours SD: 0.0125 hours	Informed by interviews with research coordinators

(*Continued*)

https://doi.org/10.1017/S1047951120001018 Published online by Cambridge University Press

Table 1. (*Continued*)

Description	Base case	Low	High	Distribution	Data source
Number of patients involved in the study	500	100	10,000	Beta Mean: 500 SD: 250	Informed by interviews with investigators
Number of discrete variables able to be abstracted from the registry	200	50	1000	Beta Mean: 200 SD: 100	Informed by interviews with investigators
Fringe rate	0.5	0.2	0.7	Normal Mean: 0.5 SD: 0.1	Department of Health and Human Services
Indirect cost rate	0.7	0.5	0.9	Normal Mean: 0.7 SD: 0.1	Department of Health and Human Services
Probabilities					
Registry-related probabilities					
Percentage of variables able to be abstracted from the registry	0.90	0.01	1.00	Beta Mean: 0.90 SD: 0.05	Informed by interviews with investigators
Probability that the centre will need additional query support/query rewrite	0.75	0.25	1.00	Beta Mean: 0.75 SD: 0.05	Informed by interviews with programmers
Probability that the query data out of range (i.e. does need to be entered again manually)	0.02	0.001	0.05	Beta Mean: 0.02 SD: 0.005	Informed by interviews with trial statisticians
Standard trial-related probabilities					
Percentage of erroneously manually entered variables	0.02	0.001	0.05	Beta Mean: 0.02 SD: 0.005	Informed by interviews with trial statisticians

*Indicates that base case values include an "experience" or "efficiency" factor. All registry time-related variables were scaled using an experience function to account for the fact that, in the current era, staff may not be familiar with registry-based trial techniques and that there is a learning curve associated with implementation of these methods. The experience factor was defined by 1/(1.5^experience level), with an assumed a hyper-exponential distribution, with lambda = 1.5 and p = 0.4. Experience level grossly represents the number of times study personnel have previously implemented a registry-based trial. It was assumed that personnel involved in standard clinical trials are well versed in the nature of the work, as this has been the long-term trial strategy. SD = standard deviation.

Registry-based trial

Registry data are typically gathered and stored locally in registry-specific databases and then shared with central data repositories that link and clean the data they receive from all contributing sites. Each contributing institution typically retains ownership of its locally held data, but the registry sponsor usually owns and governs use of data in the central repository. Access to data, once it enters a central repository, varies greatly, depending on the financial and legal structure of each registry. Some registries allow unlimited access to unblinded data at little or no additional cost.[19] Others allow access to de-identified data, but restrict or charge for access to patient identifiers.[20-22] Finally, the most restrictive registries release no data from the central repositories and require all analytics to be performed at the registry itself. This may result in high fees and restrictions in the types of questions the investigator is allowed to research.[23,24]

Depending on registry and study characteristics, optimal data access strategies may vary. Centralised data access – accessing data from the central repository – is typically the most straightforward.[16,17] It also offers the benefit of including the greatest number of sites and assures uniform data linkage and cleaning. When the costs of accessing centralised data are high, or when registries place significant barriers to access, some investigators may prefer to use locally held data by soliciting cooperation from individual sites. In these cases, investigators work with collaborators at peer institutions, helping them to gather and share locally held data. This can be accomplished either via focused data queries at each site[15] or a more generic data harvest, with *post hoc* data manipulation.[25,26]

We modelled each of these data acquisition strategies. For all registry-based trials, we assumed that a percentage of valid data could be abstracted from a registry (varied from 1 to 100%). The percentage of data that can be abstracted indicates the data richness of the registry. The number of data points needed to be abstracted is a product of the number of variables included in the study and the number of patients in the study. If 40% of data were available in the registry, this would indicate that, for a study with 1000 patients and 100 variables, 40,000 data points could be abstracted from the registry and 60%, or 60,000 data points, would need to be abstracted manually. For model simplicity, we assumed that all variables abstracted took the average amount of time, such that each data element abstracted was weighted equally. Sensitivity analyses consider alternate assumptions.

In our primary, base case model, we also assumed that only one registry was accessed. As described above, we assumed that the costs associated with creating and maintaining registries are sunk costs, paid whether or not a trial is conducted, and therefore were beyond our models' scope.

Data access strategies

(1a) Accessing Locally Held Data via Study-Specific Query: In this design, data managers at each centre run focused queries on locally held data that result in abstraction and transmission of

https://doi.org/10.1017/S1047951120001018 Published online by Cambridge University Press

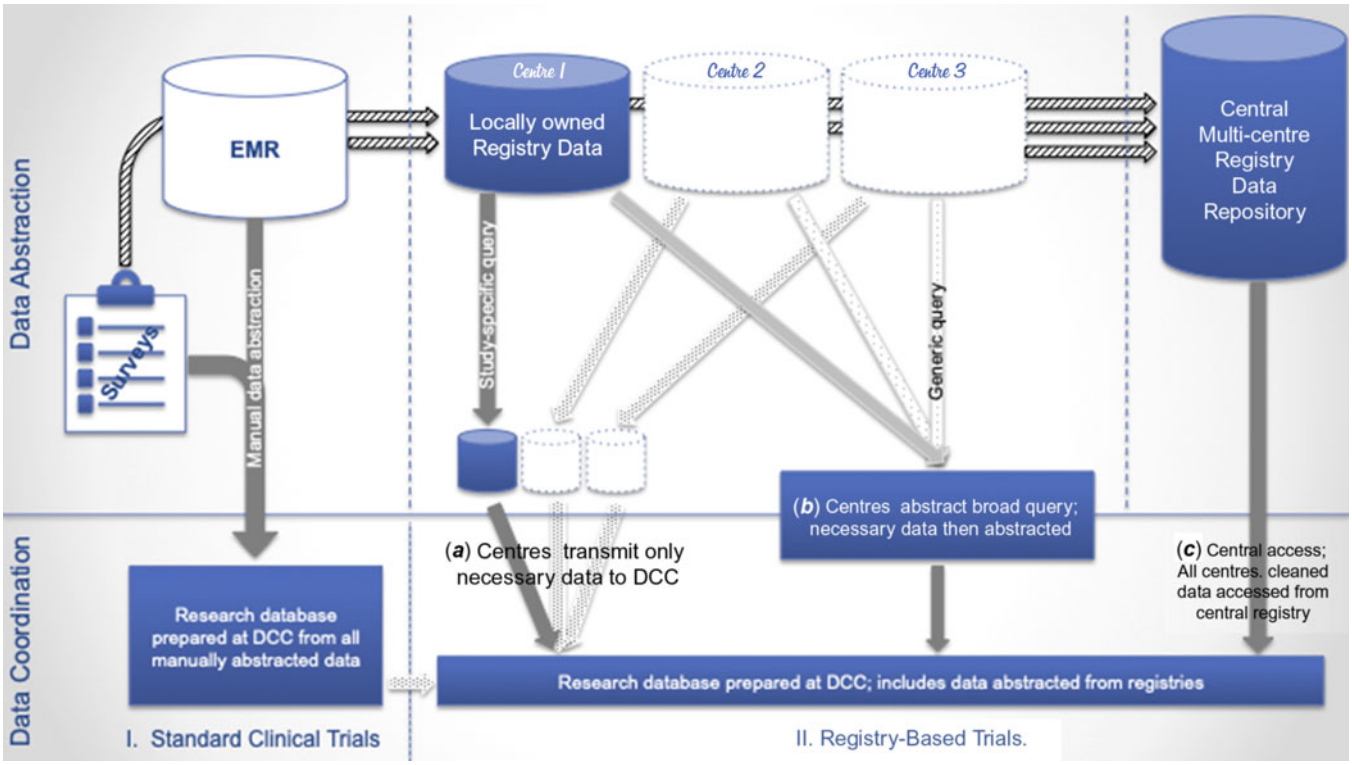


Figure 1. Data capture – the flow of data from electronic medical records to registries to research. Here we delineate the normal flow of data from electronic medical records and surveys to registries and the different ways in which investigators can access these data. (I) Under standard clinical trial design, data are all manually abstracted and transmitted to a data coordinating centre for linkage and cleaning. (II) Data from electronic medical records and surveys are also entered into locally held databases at each centre and then transmitted from each centre to a central, multi-centre data repository. These data can be accessed using the following strategies: (***a***) at each site via study-specific queries that select only the necessary data for transmission to a data coordinating centre, (***b***) at each site via a more generic query that abstracts less specific data that is cleaned either at each site or at the data coordinating centre *post hoc*, or (***c***) from the central data repository that has already cleaned and linked the data. While centres own locally held data, there may be fees associated with accessing central repository data.

study-specific patients and preselected variables. This design requires the most work for local data managers, but provides the most stringent patient privacy protections. For this design, we assumed a programmer (or series of programmers) would write initial queries to pull required data from local data repositories and that these queries would need to be tailored to local data platforms. Data managers at each site would run these queries. The resultant data would be uploaded by coordinators and linked with other centres' data by statisticians at a data coordinating centre. A series of query errors might be identified and corrected at this stage, including inaccuracies in query scripts or data capture, local issues with data entry, etc. Query adjudication would involve variable time inputs from programmers, data managers, coordinators, and/or statisticians. In the base-case model, we assumed that all data could be collected during one final data abstraction, after query errors were rectified. If multiple separate data sources/registries were accessed, input times would be multiplied by the number of sources. Data not available in the registries would be abstracted and entered manually (see Table 1).

(1b) Accessing Locally Held Data via Generic Query with Post Hoc Data Manipulation: In this design, data managers run generic queries at each centre to abstract all data from the local database. *Post hoc* data manipulation is then used to extract study-specific patients and preselected variables. These generic queries pull data on all patients in the registry within specified eligibility criteria (all patients, for example, admitted during a specified time period with a specified diagnosis or undergoing a specified procedure). Depending on the nature of the data, institutional review boards may require that only data on consented patients be transmitted to a data coordinating centre. In these cases, data managers at each centre would have to manipulate the data (either manually or using statistical software) to abstract only appropriate cases and/or variables. Research coordinators would then upload datasets to a data repository. Additional time might be spent by programmers, research coordinators, and/or data managers at each centre to implement and revise queries. Finally, statisticians at a data coordinating centre would clean and link data from all contributing sites (see Table 1).

(2) Accessing Data from Central Data Platform: In this design, data are accessed directly from registries, central repositories, or data warehouses, with flat fees paid that result in clean, linked data from all sites. These flat fees cover expenses associated with writing, revising, and implementing data queries and are typically defined to cover all data abstraction associated with the life cycle of the trial. As above, data not available in the registry would be abstracted and entered manually (Table 1). In this arm, we model the maximum tolerable fees that would allow this strategy to be the most cost effective, under various study conditions.

Statistical analyses

We constructed Markov decision-tree models using TreeAge Pro 2018 software (TreeAge, Williamstown, MA). To test the impact of uncertainty in our primary, base case estimates, we conducted one-way, two-way, and probabilistic sensitivity analyses. These analyses allowed us to determine which study characteristics had the greatest effects on cost differences and to determine input thresholds at which

https://doi.org/10.1017/S1047951120001018 Published online by Cambridge University Press

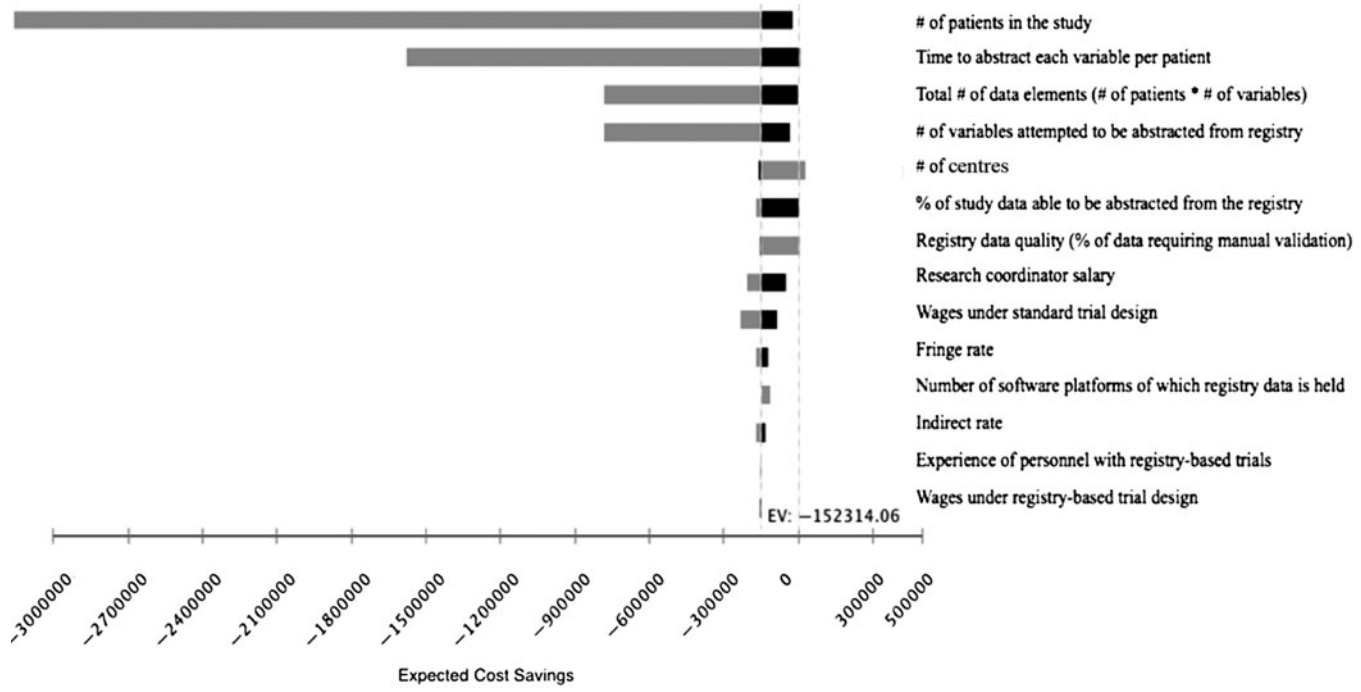


Figure 2. Tornado diagram. The tornado diagram graphically displays the effects on cost savings of changes in the 14 most impactful study parameters when comparing the costs associated with data acquisition, cleaning, and linkage under a standard trial design versus a registry-based trial when accessing locally held data via a study-specific data query. The vertical line represents the expected cost savings ($152,314) associated with a registry-based trial when all study parameters are at base case estimates. Grey bars represent high range values for each parameter; black lines represent low range values for the parameter. Dashed line represents the point at which standard trial and registry-based trial designs would be equally cost effective. This diagram assumes only one study characteristic changes at a time.
*Wages under standard and registry-based trials were analysed based on a wage index variable, introduced to account for variation in personnel salaries.

investigators might optimally select each trial design. We performed probabilistic sensitivity analyses using Monte Carlo simulation, whereby we calculated cost difference estimates from the assigned distributions for each variable. We repeated these estimates 100,000 times and used these estimates to calculate 95% confidence intervals. We independently tested the costs associated with standard data acquisition against each of the three registry-based methods and compared cost difference. Input ranges and distributions for probabilistic sensitivities are characterised in Table 1.

Results

Cost savings

Data collection, cleaning, and linkage were projected to cost $183,319 using a standard clinical trial design for a study involving 100,000 independent data elements (the number of patients multiplied by the number of variables per patient), assuming all study parameters at base case estimates. Data collection, cleaning, and linkage for the same study conducted using a registry with a focused data query were estimated to cost $31,005. In Monte Carlo simulation, allowing all trial characteristics to vary simultaneously, the costs associated with data collection, cleaning, and linkage for standard clinical trials varied from $16,744 to $678,137 (95% confidence interval, CI). In contrast, the associated costs for registry-based trials varied from $6825 to $99,858 (95% CI), resulting in an estimated savings of $4363 to $591,657. The base case estimated savings were $152,314. A registry-based trial was more cost effective 98.6% of the time when registry data were available. Savings ranges were similar for all three registry-based trial designs.

Effects of study parameters on cost estimates

Figure 2 displays the effects of changes in key study parameters on cost savings. The savings associated with registry-based trial designs were driven predominantly by the total number of individual data elements able to be abstracted from a registry and by the amount of time (percent effort) it took research coordinators to manually abstract and enter each data element. In one-way sensitivity analyses, assuming local access to registry data via a study-specific query, leaving all other study parameters at base case values, a standard clinical trial design became less costly if <3768 independent data elements (patients multiplied by the number of variables per patient) were able to be abstracted from a registry, if research coordinators spent <3.4 seconds per variable per patient to obtain and enter each data element, or if >252 centres participated. Regional increases in wages and prior experience with registry-based trials increased cost savings associated with registry-based designs, though effects were modest.

Table 2 and Figure 3 detail the effects on cost savings of changing two study parameters at a time. Standard data abstraction was more cost effective only in studies involving very low coordinator abstraction times, large number of centres, large number of software platforms, low percentages of data able to be abstracted from registries, and large percentages of data requiring manual review.

Selecting a data access strategy

When conducting registry-based trials, it was consistently less cost effective to abstract locally held data via generic query with post hoc data manipulation than to abstract locally held data via study-specific query. It was most cost effective to access data directly from a central registry when registry fees were less than

https://doi.org/10.1017/S1047951120001018 Published online by Cambridge University Press

Table 2. Estimated cost savings associated with registry utilisation versus standard trial design, as a function of study size and other key trial characteristics, assuming registry data accessed locally.

	Cost savings (95% CI)		
	Study size*		
	10,000 data elements	100,000 data elements	1,000,000 data elements
Research coordinator time per data element			
½ minute	$−613 (−$20,512 to $15,904)	$4547 (−$17,676 to $34,849)	$522,050 ($35,922–1,946,536)
1½ minutes	$9,864 (−$15,421 to $54,562)	$152,314 ($4363–591,657)	$1,576,810 ($113,08–5,847,937)
5 minutes	$46,828 (−$6,674 to $193,799)	$522,050 ($31,799–1,988,146)	$5,248,845 ($384,909–19,425,286)
# of centres			
1	$15,532 ($1,903–57,100)	$157,982 ($11,327–575,888)	$1,584,478 ($114,378–5,762,109)
10	$9864 (−$2,120 to $50,753)	$152,314 ($4363–591,657)	$1,576,810 ($110,353–$5,756,799)
200	−$109,790 (−$131,148 to −$72,972)	$32,660 (−$69,542 to $425,076)	$1,457,156 ($82,468–5,573,837)
% data abstracted from registry			
1	−$4635 (−$4,846 to −$2,216)	−$3052 (−$4370 to $2,227)	$12,776 ($−4387 to $63,527)
90	$9864 ($1014–35,040)	$152,314 ($4363–591,657)	$1,576,810 ($107,742–5,850,295)
100	$11,493 ($1941–40,432)	$169,771 ($12,445–632,672	$1,752,745 ($123,501–6,481,290)
Research coordinator salary			
$29,407	$2310 ($734–8018)	$75,041 ($5147–280,119)	$802,351 ($57,089–2,893,852)
$57,596	$9864 ($1182–37,485)	$152,314 ($4363–591,657)	$1,576,810 ($114,226–5,674,802)
$117,528	$25,925 ($2010–99,466)	$316,602 ($21,551–1,152,697)	$3,223,370 ($236,132–11,586,442)
% of data requiring manual validation			
0.1	$10,237 ($1280–39,175)	$155,449 ($10,218–579,240)	$1,607,563 ($113,372–5,853,231)
2	$9864 ($1068–38,643)	$152,314 ($4363–591,657)	$1,576,810 ($109,967–5,754,198)
5	$9275 ($870–37,738)	$147,365 ($8738–555,247)	$1,528,254 ($105,710–5,595,522)
Wage index, standard trial design**			
0.5	$2170 ($539–8089)	$73,395 ($4597–279,521)	$785,644 ($53,629–2,877,068)
1	$9864 ($1113–38,582)	$152,314 ($4363–591,657)	$1,576,810 ($110,353–5,756,798)
1.5	$17,558 ($1768–69,244)	$231,233 ($14,889–858,067)	$2,367,976 ($167,109–8,637,726)
Fringe rate			
0.2	$7892 ($909–31,124)	$121,851 ($7827–459,078)	$1,261,449 ($88,296–4,635,207)
0.5	$9864 ($1134–39,020	$152,314 ($4363–591,657)	$1,576,810 ($110,371 to $5,794,010)

(*Continued*)

https://doi.org/10.1017/S1047951120001018 Published online by Cambridge University Press

Table 2. (*Continued*)

	Cost savings (95% CI)		
	Study size*		
	10,000 data elements	100,000 data elements	1,000,000 data elements
0.7	$11,180 ($1384–44,222)	$172,622 ($11,088–650,361)	$1,787,051 ($125,087–6,566,543)
# software platforms			
1	$9864 ($1094–38,582)	$152,314 ($4363–591,657)	$1,576,810 ($110,352–5,756,799)
20	$2573 (−$3555 to $30,199)	$145,023 ($3660–561,827)	$1,569,519 ($103,076–5,748,996)
100	−$28,129 (−$18,702 to −$9,852)	$114,321 (−$14,443 to $526,209)	$1,538,817 ($75,754–5,716,833)
Indirect rate			
0.5	$8704 ($989–37,035)	$134,395 ($8558–508,693)	$1,391,303 ($97,015–5,085,144)
0.7	$9864 ($1121–38,593)	$152,314 ($4363–591,657)	$1,576,810 ($109,951–5,763,163)
0.9	$11,024 ($1250–43,133)	$170,234 ($10,839–636,745)	$1,762,318 ($122,886–6,441,183)
Registry personnel experience level***			
0.7	$9864 ($1113–38,583)	$152,314 ($4363–591,657)	$1,576,810 ($110,352–5,756,799)
5	$14,374 ($3358–49,401)	$156,924 ($12,531–575,122)	$1,581,320 ($114,201–5,762,129)
10	$15,191 ($3855–50,951)	$157,641 ($13,092–575,700)	$1,582,137 ($114,888–5,762,609)
Wage index, registry-based trial design**			
0.5	$12,626 ($1217–49,615)	$155,076 ($10,114–572,914)	$1,579,572 ($111,907–5,760,444)
1	$9864 ($1113–$38,582)	$152,314 ($4363–591,657)	$1,576,810 ($110,352–5,756,799)
1.5	$7102 ($1079–23,315)	$149,553 ($9306–564,946)	$1,574,049 ($108,700–5,754,717)

Cost savings are calculated varying two study characteristics at a time and holding all other characteristics at base case values defined in Table 1.
*Study size = number of total data elements (# of patients * # of discrete variables)
**Wage index = a multiplier of base case salaries for a specific trial design's personnel that accounts for regional or other variations in salary.
***Experience = a multiplier of time to conduct each component of a registry-based trial; time decreases with prior experience with registry-based trials.

$97,051 (including indirect costs), assuming that there were greater than 89,948 discrete data fields and assuming that manual data abstraction and entry took more than 1.3 minutes per data field. The costs of local data access increased with the number of separate queries that needed to be run.

Discussion

We used simulation modelling to build a practical tool for investigators across disciplines that can be used to estimate the potential cost savings associated with incorporating registries into clinical trial design across a broad range of study characteristics. Our data suggest that in the vast majority of cases, incorporating registries resulted in cost savings, decreasing costs from a few thousand dollars to over $500,000 per study. The size of the study had the greatest impact on cost savings, though even relatively small trials typically benefitted from registry inclusion.

The magnitude of cost savings we estimate is lower than some of those previously described. Editorials following some of the earliest registry-based trials estimated their total study costs between $20 and $50 per subject.[3-6] Previous editorials have compared these estimates to the costs of standard clinical trials, which typically range between $2000 and $40,000 per patient.[7-9] With even just 500 patients, as in our base model, this would result in cost savings between $1M and $20M per trial (as compared to $4363 to $591,657, we estimate here). Yet, these are not balanced comparisons. The registry-based trials described in these editorials reduced costs in multiple ways. For the TASTE trial, the first prominent registry-based trial, for example, the primary intervention was coronary artery thrombus aspiration, a quick and resource-limited procedure, either performed or not performed

https://doi.org/10.1017/S1047951120001018 Published online by Cambridge University Press

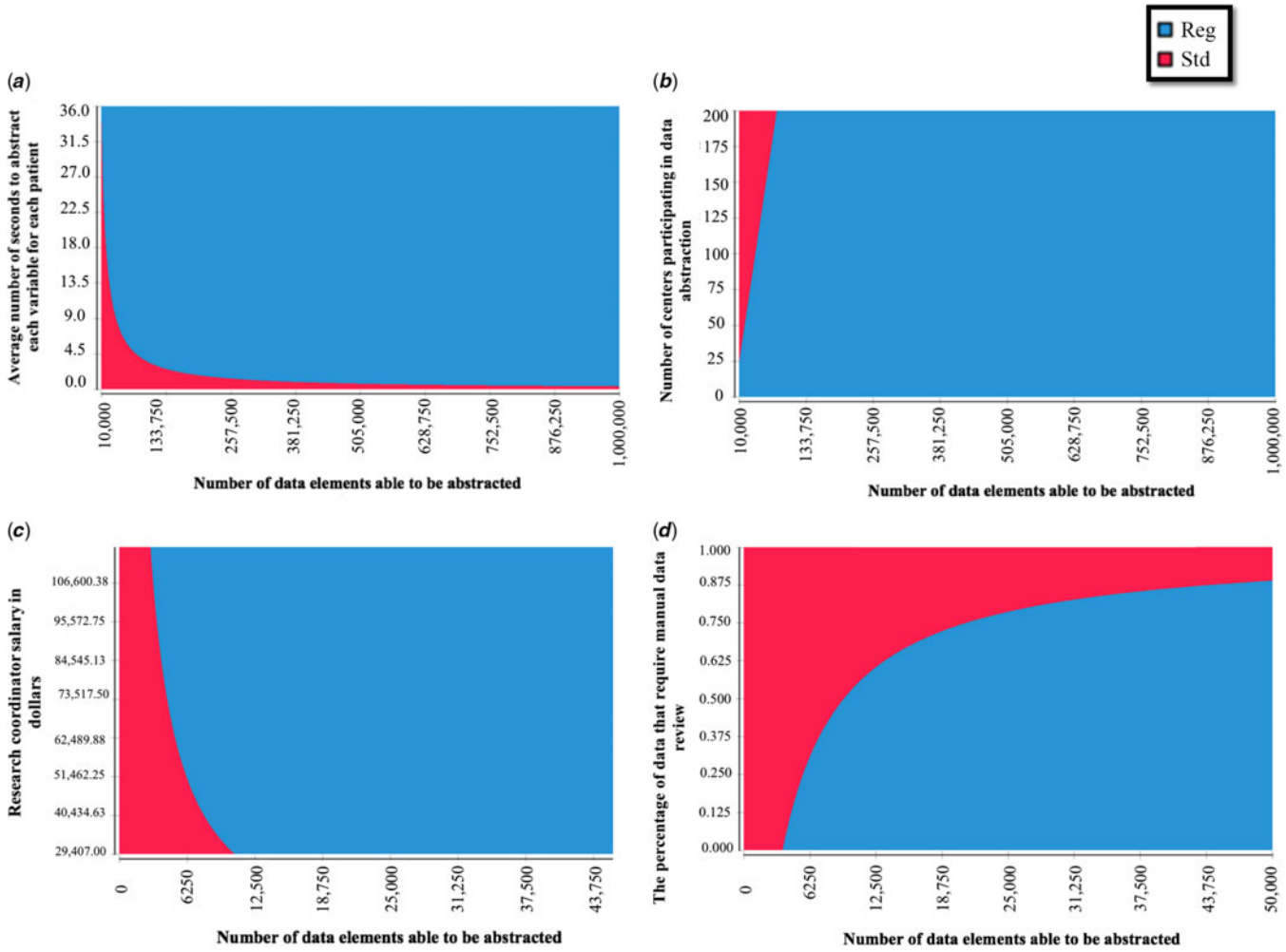


Figure 3. Two-way sensitivity analyses. Graphs display the interactive effects on cost savings of two simultaneous changes in study parameters when comparing a standard trial design versus a registry-based trial when accessing locally held data via a study-specific data query. (***a***) Displays the effects of changes in the total number of data elements able to be abstracted versus the average number of seconds it takes a research coordinator to abstract each variable for each patient; (***b***) displays the effects of changes in the total number of data elements able to be abstracted versus the number of centres participating in data abstraction; (***c***) displays the effects of changes in the total number of data elements able to be abstracted versus the salaries of research coordinators; and (***d***) displays the effects of changes in the total number of data elements able to be abstracted versus the quality of the registry data (the percentage of data that require manual data review). Red denotes scenarios in which a standard trial design would be less costly. Blue denotes scenarios in which registry-based trials would be less costly.

alongside a clinically indicated percutaneous coronary intervention, in patients with ST elevation myocardial infarctions. Further, only registry data were gathered, no study-specific follow-up was conducted, no data adjudications were conducted, and data access fees were minimal.[3] In the SAFE-PCI trial, similarly, the intervention was a quick and resource-limited procedure, associated with no to minimal additional costs, and follow-up was limited.[4]

It is instructive to see the creativity that went into investigations such as the TASTE and SAFE-PCI trials and the incredible data that were generated on relatively limited budgets. There is much that can and should be learned from their varied, pragmatic designs. That having been said, there are research questions for which more costly interventions or supplemental data may be required. We demonstrated that these investigations might still be able to reduce costs by incorporating the use of existing data for a portion of data capture.

In our study, we sought to model the potential cost savings of utilising a registry as part of data collection across a broad range of trials. We conducted sensitivity analyses to empower investigators to make informed decisions about the feasibility of incorporating a registry-based study design into their workflow. In 2009, 25 international trialists worked together to design the PRECISE (Pragmatic Explanatory Continuum Indicator Summary) tool to help investigators develop trials best aligned with their study intents, helping them to define where their research question falls on the spectrum from "*explanatory*" to "*real world*" implementation (precis-2.org).[14,27,28] Ultimately, we recommend that researchers consider incorporating registries into their investigations, even when their studies are highly explanatory or when only seemingly small portions of data can be abstracted.

It should be noted that trial design and patient recruitment offer additional and highly important areas for potential cost savings. It has been reported that up to 45% of funded trials fail to enroll 80% of targeted patients and that roughly half request recruitment-related extensions.[29,30] Many have argued that utilising registries in trial design and recruitment may help improve study feasibility and reduce the need for costly recruitment/marketing strategies.

https://doi.org/10.1017/S1047951120001018 Published online by Cambridge University Press

Disease-specific and research registries now exist to facilitate patient identification.[31,32] The potential benefits associated with using registries in trial design and recruitment are not captured in our models. We encourage investigators to consider these potential benefits as well, in the context of their individual research questions.

There are some additional limitations to our models that should be considered. First, we conducted a simulation using reasonable estimates. Some trials, however, will vary in ways that are outside the scope of our models. Our reasonable estimates are based on North American costs, and thus our findings may not be generalisable to other countries. Second, there might be additional costs if data from different sources have to be stored in separate data platforms. Third, our estimates of time are based on interviews. We incorporate wide ranges to account for uncertainty, yet there are no established standards. Further, not all registry data are subject to the highest level of rigour and auditing. This ought to be weighed seriously when designing a trial, as it might present issues with the acceptance of trial results or add additional costs if external audits are performed. Additionally, the costs of developing and maintaining registries are not captured in this study. There is growing attention to the rapidly rising financial burden for hospitals associated with registry participation. These costs are steep and real, yet they fall outside the scope of this study. There may be ways to reduce costs in conjunction with the usage of registry data, including the implementation of automated mechanisms for reliable data abstraction and validation and the incorporation of machine learning to reduce the overhead associated with these registries. Such advances might further reduce the costs of registry-based trials, either by improving data quality or reducing data access fees, but for purposes of this study, these are treated as sunk costs.

Conclusion

Our data suggest that registry-based trials, broadly incorporated, can confer cost-saving opportunities for the majority of trials for which even a small amount of relevant and accurate data are available. We offer here a practical tool for investigators interested in assessing these cost savings for their own trials.

Acknowledgments. We thank the Collaborators listed in Appendix 1 for contributing to our data collection in semi-structured interviews. This work was supported by the US National Institutes of Health.

Supplementary material. To view supplementary material for this article, please visit https://doi.org/10.1017/S1047951120001018

Financial support. This study was supported by grants (U24HL135691, U10HL068270, HL109818, HL109778, HL109816, HL109743, HL109741, HL109673, HL068270, HL109781, HL135665, HL135680) from the National Heart, Lung, and Blood Institute, National Institutes of Health (NHLBI, NIH). Dr. Brett Anderson receives support from K23 HL133454 from NHLBI, NIH. Dr. Meena Nathan receives support from K23 HL119600 from NHLBI, NIH. The authors have no other financial relationships relevant to this article to disclose.

Conflict of interest. None.

References

1. Califf RM, Sanderson I, Miranda ML. The future of cardiovascular clinical research: informatics, clinical investigators, and community engagement. JAMA 2012; 308: 1747–1748. doi: 10.1001/jama.2012.28745.
2. Tricoci P, Allen JM, Kramer JM, et al. Scientific evidence underlying the ACC/AHA clinical practice guidelines. JAMA 2009; 301: 831–841. doi: 10.1001/jama.2009.205.
3. Frobert O, Lagerqvist B, Olivecrona GK, et al. Thrombus aspiration during ST-segment elevation myocardial infarction. NEJM 2013; 369: 1587–1597. doi: 10.1056/NEJMoa1308789.
4. 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. doi: 10.1016/j.ahj.2013.06.013.
5. Huang SS, Septimus E, Kleinman K, et al. Targeted versus universal decolonization to prevent ICU infection. NEJM 2013; 368: 2255–2265. doi: 10.1056/NEJMoa1207290.
6. Thabane L, Kaczorowski J, Dolovich L, et al. Reducing the confusion and controversies around pragmatic trials: using the Cardiovascular Health Awareness Program (CHAP) trial as an illustrative example. Trials 2015; 16: 387. doi: 10.1186/s13063-015-0919-3.
7. Detsky AS. Using cost-effectiveness analysis to improve the efficiency of allocating funds to clinical trials. Stat Med 1990; 9: 173–184.
8. Eisenstein EL, Lemons PW, 2nd, Tardiff BE, et al. Reducing the costs of phase III cardiovascular clinical trials. Am heart J 2005; 149: 482–488. doi: 10.1016/j.ahj.2004.04.049.
9. Sertkaya A, Wong HH, Jessup A, et al. Key cost drivers of pharmaceutical clinical trials in the United States. Clin Trials 2016; 13: 117–126. doi: 10.1177/1740774515625964.
10. Lauer MS, D'Agostino RB, Sr. The randomized registry trial–the next disruptive technology in clinical research? NEJM 2013; 369: 1579–1581. doi: 10.1056/NEJMp1310102.
11. Ramsberg J, Neovius M. Register or electronic health records enriched randomized pragmatic trials: the future of clinical effectiveness and cost-effectiveness trials? Nordic J Health Econ 2015; 5: 62. clinical trials, pragmatic trials, disease registries, electronic health records, costs. doi: 10.5617/njhe.1386.
12. Ford I, Norrie J. Pragmatic trials. NEJM 2016; 375: 454–463. doi: 10.1056/NEJMra1510059.
13. Li G, Sajobi TT, Menon BK, et al. Registry-based randomized controlled trials- what are the advantages, challenges, and areas for future research? J Clin Epi 2016; 80: 16–24. doi: 10.1016/j.jclinepi.2016.08.003.
14. Loudon K, Treweek S, Sullivan F, et al. The PRECIS-2 tool: designing trials that are fit for purpose. BMJ 2015; 350: h2147. doi: 10.1136/bmj.h2147.
15. Nathan M, Trachtenberg FL, Van Rompay MI, et al. Design and objectives of the pediatric heart network's residual lesion score study. J Thorac Cardiovasc Surg 2019. doi: 10.1016/j.jtcvs.2019.10.146.
16. Anderson BR, Kumar SR, Gottlieb-Sen D, et al. The congenital heart technical skill study: rationale and design. World J Pediatr Congenit Heart Surg 2019; 10: 137–144. doi: 10.1177/2150135118822689.
17. Hill K, Baldwin H, Bichel D, et al. Rationale and design of the steroids to reduce systemic inflammation after infant heart surgery (stress) trial. Am Heart J 2019.
18. Boskovski M. The Brigham and Women's Hospital, Harvard Medical School, 2019.
19. ResearchMatch. https://www.researchmatch.org/.
20. Pediatric Health Information System Database. https://www.childrenshospitals.org/programs-and-services/data-analytics-and-research/pediatric-analytic-solutions/pediatric-health-information-system.
21. Healthcare Cost and Utilization Project. https://www.hcup-us.ahrq.gov/databases.jsp.
22. Surveillance, Epidemiology, and End Results. https://seer.cancer.gov/registries/.
23. Society of Thoracic Surgeons. https://www.sts.org/registries-research-center/sts-national-database.
24. American College of Cardiology. https://cvquality.acc.org/NCDR-Home/registries.
25. Anderson BR, Dragan K, Glied S, et al. Building capacity to assess longitudinal outcomes and health expenditures for children with congenital heart disease. J Am Coll Card 2019; 73: 649. doi: 10.1016/s0735-1097(19)31257-4.

https://doi.org/10.1017/S1047951120001018 Published online by Cambridge University Press

26. Boscoe FP, Schrag D, Chen K, et al. Building capacity to assess cancer care in the Medicaid population in New York State. Health Serv Res 2011; 46: 805–820. doi: 10.1111/j.1475-6773.2010.01221.x.
27. Thorpe KE, Zwarenstein M, Oxman AD, et al. A pragmatic-explanatory continuum indicator summary (PRECIS): a tool to help trial designers. J Clin Epidemiol 2009; 62: 464–475. doi: 10.1016/j.jclinepi.2008.12.011.
28. Schwartz D, Lellouch J. Explanatory and pragmatic attitudes in therapeutical trials. J Chronic Dis 1967; 20: 637–648. doi: 10.1016/0021-9681(67)90041-0.
29. McDonald AM, Knight RC, Campbell MK, et al. What influences recruitment to randomised controlled trials? A review of trials funded by two UK funding agencies. Trials 2006; 7: 9. doi: 10.1186/1745-6215-7-9.
30. Carlisle B, Kimmelman J, Ramsay T, et al. Unsuccessful trial accrual and human subjects protections: an empirical analysis of recently closed trials. Clin Trials 2015; 12: 77–83. doi: 10.1177/1740774514558307.
31. Tan MH, Thomas M, MacEachern MP. Using registries to recruit subjects for clinical trials. Contemp Clin Trials 2015; 41: 31–38. doi: 10.1016/j.cct.2014.12.012.
32. Strasser JE, Cola PA, Rosenblum D. Evaluating various areas of process improvement in an effort to improve clinical research: discussions from the 2012 Clinical Translational Science Award (CTSA) Clinical Research Management workshop. Clin Transl Sci 2013; 6: 317–320. doi: 10.1111/cts.12051.

Appendix 1

Collaborators

The following members of the Pediatric Heart Network are author and non-author contributors, surveyed as part of this investigation:

Clinical Site Investigators, Coordinators, and Data Managers: *The Hospital for Sick Children,* Steven Schwartz (PI), Brian McCrindle, Andreea Dragulescu, Jessica Bainton, Martha Rolland, Patricia Walter (research coordinator), Elizabeth Radojewski (research coordinator)*; *Children's Hospital Boston*, Jane W. Newburger, Meena Nathan (PI), Ron Lacro, Carolyn Dunbar-Masterson (research coordinator), Lisa-Jean Buckley, Steven Colan, Alexander Tang, Heather Doyle (data manager); *Children's Hospital of Philadelphia*, J. William Gaynor (PI), Matt O' Connor, Tonia Morrison (research coordinator), Devida Long, Abigail Waldron, Somaly Srey (research coordinator)*, Geraldine Altema, Donna M. Sylvester (research coordinator), Andrea Kennedy (data manager); *Medical University of South Carolina,* Eric Graham (PI), Kimberly E. McHugh, Mark A. Scheurer, Andrew Atz, Scott Bradley, Jason Buckley, Shahryar Chowdhury, Kalyan Chundru (research coordinator), Geoffrey Forbus, Anthony Hlavacek, Minoo Kavarana, Arni Nutting, Carolyn Taylor, Sinai Zyblewski, Sally Enyon (data manager); *Primary Children's Hospital and the University of Utah,* L. LuAnn Minich (co-PI), Philip T. Burch* (co-PI), Aaron Eckhauser, Linda M. Lambert (research coordinator), Eric R. Griffiths, Richard V. Williams, Randy Smout (data mangager); *University of Michigan*, Richard G. Ohye (PI), Sara Pasquali, Jimmy Lu, Tammy Doman (research coordinator), Theron Paugh* (data manager), Leslie Wacker (data manager); *Cincinnati Children's Medical Center*, Nicolas Madsen, Jeffery B. Anderson, Teresa Barnard – deceased (research coordinator), Michelle Hamstra (research coordinator), Kathleen (Ash) Rathge (research coordinator), Michael Taylor, Mary Suhre, Kelly Brown* (data manager); *Riley Children's Hospital-Indiana University,* Marcus Schamberger (PI), Leanne Hernandez (research coordinator), Liz Swan* (research coordinator), Cindy Dwight (research coordinator), Mark Turrentine, Mark Rodefeld, Mary Stumpf, Connie Dagon (data manager); *Children's Hospital of Atlanta,* Kirk Kanter (PI), Subhadra Shashidh, Janet Fernandez, Raejanna Ashley, Leslie Smithley*, Cheryl Stone*,Kelly Oliver*, Matthew Ferguson, Dawn Boughton (data manager); *Texas Children's Hospital,* Carlos Mery (PI), Iki Adachi, Giles W. Vick, Liz McCullum (research coordinator)*, Rija John, Joyce Chan* (research coordinator), Carmen Watrin (data manager), and Karla Dyer (research coordinator); *Columbia University Irving Medical Center*, Brett Anderson (Co-PI), Emile Bacha (Co-PI), Katrina Golub (research coordinator), Glennie Mihalovic (research coordinator)*, Rosalind Korsin (research coordinator)*, Linfang Li (data manager); *Duke University Medical Center*, Kevin Hill; *Children's Hospital of Wisconsin,* Ronald Woods (PI), Michelle Otto (research coordinator), Jessica Stelter, Sydney Allen, James Tweddell*, Garrick Hill, Benjamin Goot, Angie Klemm (data manager); *Johns Hopkins All Children's Heart Institute* Jeffrey Jacobs, Thieu Nguyen, Jade Hanson (research coordinator), Dawn Morelli, Jean Wilheim (data manager); *Nemours Cardiac Center*, Christian Pizarro (PI), Glenn Pelletier, Gina Baffa, Carol Prospero (research coordinator), Kimberly Hoffman (data manager); *University of Minnesota Masonic Children's Hospital,* Shanthi Sivanandam (PI), Susan Anderson (research coordinator), Dan Nerheim, Brian Harvey (research coordinator), Amy Ash; *Children's Hospital and Clinics of Minnesota,* David Overman (PI), Marko Vezmar, Melissa Buescher; *Children's Mercy Hospital,* Geetha Raghuveer (PI), James O'Brien, Anitha Parthiban, Jennifer Panuco (research coordinator), Jennifer Marshall, Rita France, Elizabeth Turk (data manager); *John's Hopkins University*, Marshall Jacobs.

The following members of the Pediatric Heart Network are additional non-author contributors:

National Heart, Lung, and Blood Institute: Gail Pearson, Victoria Pemberton, Mario Stylianou, Kristin Burns.

Network Chair: *University of Texas Southwestern Medical Center*, Lynn Mahony.

Data Coordinating Center: *HealthCore* (formerly *New England Research Institutes*), Margaret Bell*, Tammi Mansolf*, Susan Pham*, Tiffany Eldridge*, Michael Leong*, Maria Van Rompay, Amanda Aing*, Allison Crosby-Thompson.

Echocardiography Core Laboratory: Boston Children's Hospital, Steven Colan, Jami Levine, Edward Marcus.

Protocol Review Committee: Michael Artman*, Timothy Feltes*, Jeffrey Krischer*, G. Paul Matherne (Chair), Frank Evans, Joga Gobburu, Andrew Mackie, Patrick McQuillen, Shelley Miyamoto, Joanna Shih, Janet Simsic, Sally Hunsberger*.

Data and Safety Monitoring Board: J. Philip Saul (Chair), Dianne L. Atkins, Preetha Balakrishnan, Craig Broberg, David J. Driscoll, Frank Evans, Sally A. Hunsberger, Liza-Marie Johnson, Thomas J. Knight, Paul Lipkin, Holly Taylor, Diane Atkins*, Mark Galantowicz*, David Gordon*, Sally Hunsberger*, John Kugler*, Helen Parise*, Holly Taylor*.

*No longer at the institution listed.

https://doi.org/10.1017/S1047951120001018 Published online by Cambridge University Press