




Feasibility of using real-world data in the evaluation of cardiac ablation catheters: a test-case of the National Evaluation System for Health Technology Coordinating Center

Sanket S Dhruva ¹, Guoqian Jiang,² Amit A Doshi,³ Daniel J Friedman,⁴ Eric Brandt,⁵ Jiaping Chen,⁵ Joseph G Akar,⁴ Joseph S Ross,^{6,7} Keondae R Ervin,⁸ Kimberly Collison Farr,⁵ Nilay D Shah,² Paul Coplan,⁹ Peter A. Noseworthy ¹⁰, Shumin Zhang,⁹ Thomas Forsyth,⁵ Wade L Schulz,^{7,11} Yue Yu,² Joseph P Drozda, Jr. ⁵

To cite: Dhruva SS, Jiang G, Doshi AA, *et al.* Feasibility of using real-world data in the evaluation of cardiac ablation catheters: a test-case of the National Evaluation System for Health Technology Coordinating Center. *BMJ Surg Interv Health Technologies* 2021;**3**:e000089. doi:10.1136/bmjst-2021-000089

► Additional supplemental material is published online only. To view, please visit the journal online (<http://dx.doi.org/10.1136/bmjst-2021-000089>).

Received 08 April 2021

Accepted 24 September 2021



© Author(s) (or their employer(s)) 2021. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ.

For numbered affiliations see end of article.

Correspondence to

Dr Sanket S Dhruva;
Sanket.Dhruva@ucsf.edu

ABSTRACT

Objectives To determine the feasibility of using real-world data to assess the safety and effectiveness of two cardiac ablation catheters for the treatment of persistent atrial fibrillation and ischaemic ventricular tachycardia.

Design Retrospective cohort.

Setting Three health systems in the USA.

Participants Patients receiving ablation with the two ablation catheters of interest at any of the three health systems.

Main outcome measures Feasibility of identifying the medical devices and participant populations of interest as well as the duration of follow-up and positive predictive values (PPVs) for serious safety (ischaemic stroke, acute heart failure and cardiac tamponade) and effectiveness (arrhythmia-related hospitalisation) clinical outcomes of interest compared with manual chart validation by clinicians.

Results Overall, the catheter of interest for treatment of persistent atrial fibrillation was used for 4280 ablations and the catheter of interest for ischaemic ventricular tachycardia was used 1516 times across the data available within the three health systems. The duration of patient follow-up in the three health systems ranged from 91% to 97% at ≥7 days, 89% to 96% at ≥30 days, 77% to 90% at ≥6 months and 66% to 84% at ≥1 year. PPVs were 63.4% for ischaemic stroke, 96.4% for acute heart failure, 100% at one health system for cardiac tamponade and 55.7% for arrhythmia-related hospitalisation.

Conclusions It is feasible to use real-world health system data to evaluate the safety and effectiveness of cardiac ablation catheters, though evaluations must consider the implications of variation in follow-up and endpoint ascertainment among health systems.

INTRODUCTION

Recent policy changes have increased the salience of real-world data (RWD), which are those data collected during routine clinical care, to generate real-world evidence (RWE) that can support regulatory decision-making.¹

Key messages

What is already known about this subject?

- Recent legislation and policy have increased the importance of using real-world evidence to support regulatory decision-making for medical devices.
- Little is known about the ability to use health system data to conduct research about medical device safety and effectiveness.

What are the new findings?

- We used a decentralised model of research to conduct a retrospective cohort analysis at three health systems, identifying cardiac ablation catheters of interest and patients in whom they were used for specific indications.
- We found a distribution of follow-up duration between the health systems (ranging from 66% to 84% at ≥1 year), indicating some variation in long-term follow-up of ablation patients; patients were identified in whom follow-up was generally adequate to support both periprocedural and longer term clinical outcome ascertainment.
- We found a distribution of positive predictive values for identification of safety and effectiveness outcomes (ranging from 55.7% to 100%), which were generally adequate compared with clinician chart review after code-based algorithms were used to reduce false-positive cases.

How might these results affect future research or surgical practice?

- These results demonstrate the feasibility of examining the safety and effectiveness of cardiac ablation catheters using real-world health system data.
- Lessons from this study that can help future researchers use real-world health system data to evaluate medical device safety and effectiveness.
- The results from this study also identify opportunities to enhance the completeness and quality of real-world data.

The 21st Century Cures Act of 2016 placed increasing emphasis on the use of RWE, including for secondary indication approvals for drugs.² In 2017, the US Food and Drug Administration (FDA) Center for Devices and Radiological Health released a Guidance Document about this topic, entitled ‘Use of RWE to Support Regulatory Decision-Making for Medical Devices’.³ FDA’s Guidance discusses potential RWE use to support expanded indications of use for medical devices. Examples of employing RWE for this purpose include four premarket indication expansion decisions based on data from the Transcatheter Valve Therapeutics Registry.⁴ However, registries are specialised RWD sources because they require significant effort and resources, including general reliance on trained abstractors to manually extract and input many data elements.⁵ Additionally, registries are usually specialised to a disease state or condition with a limited number of variables because of the cost and time needed for that abstraction and, therefore, are unavailable for all medical devices.

To further the goal of developing timely and robust RWE for informing regulators and clinicians regarding medical device effectiveness and safety, the National Evaluation System for health Technology (NEST) was created.⁶ The most ubiquitous sources of RWD and with the greatest clinical detail are health system databases, including electronic health records. However, health system data are not specifically designed for research purposes¹ and their ability to contribute reliable evidence for medical device safety and effectiveness evaluations, including for label expansions, remains uncertain. Since the fall of 2018, the NEST Coordinating Center (NESTcc) has supported multiple test-case studies to investigate the use of RWD, including health system data, in RWE generation that can be used to inform regulatory decision-making. Several test-cases are feasibility assessments focused on the availability of pertinent variables—including medical device use, covariates and safety and effectiveness outcomes.⁷ Here, we report on one of the first such test-cases to be completed. In this test-case, we sought to assess the feasibility of using RWD related to two cardiac ablation catheters that were generated during routine clinical practice and extracted from electronic information systems at three health systems to conduct a study that could inform regulatory decision-making for clinical indication expansion. We describe the process of collecting the necessary data, evaluating its reliability and lessons learnt that can inform future work.

METHODS

Project origination

The study was proposed to NESTcc by Johnson & Johnson, with the ultimate objective of evaluating the safety and effectiveness of two cardiac ablation catheters when used in routine clinical practice. The specific catheters of interest are the ThermoCool Smarttouch (ST) catheters, initially approved by FDA in February 2014, and the

ThermoCool Smarttouch Surround Flow (STSF) catheters, initially approved by FDA in August 2016.

After independent review, NESTcc funded the project. NESTcc currently includes 16 network collaborators (healthcare providers, academic research institutions, payers and professional registries) that collect, curate and analyse RWD that may be used for regulatory decision-making. Among its network collaborators, NESTcc identified three health systems that were interested in the proposal and that had significant experience with these devices: Mercy Health, Mayo Clinic and Yale-New Haven Hospital. Johnson and Johnson and the three NESTcc network collaborators, with Mercy Health serving as the lead, developed a full research plan that was approved by the NESTcc. Institutional Review Board (IRB) approval was obtained at Mercy Health (IRB submission number 1349229–1, acknowledged as research but not human subjects research), Mayo Clinic (IRB application number 19–001493, exempt from the requirement for IRB approval) and Yale University (IRB submission number 2000024523, approved for medical record review only).

Overview of participating health systems

Mercy Health is a health system operating in four states in the Midwest with 39 hospitals, 12 outpatient surgery centres and 35 urgent care sites and caring for a population of approximately 4.2 million active patients. Mayo Clinic partially owns and operates the Mayo Clinic health system of 70 hospitals and clinics, serving a population of approximately 1.3 million patients annually. Yale-New Haven Health System is a health system of five hospitals in the Northeast, caring for approximately 2 million patients annually. All three health systems use the same system of electronic health records (EHRs) maintained by the company Epic (Epic Systems, Verona, Wisconsin). For inventory management, the health systems use different systems: OptiFlex at Mercy, QSight at YNHH and Supply + at Mayo Clinic and have different methods for tracking medical devices.

Overall goals and overview of original data collection

The study tested the feasibility of the three independent health systems to obtain RWD from their electronic information systems, including EHRs, to examine the safety and effectiveness of improved irrigation technology, called Surround Flow (SF) to the tip of the ThermoCool cardiac ablation catheter. The informatics methods have been described separately.⁸

The first portion of the study sought to examine the feasibility of demonstrating equivalent safety and effectiveness of the ThermoCool Smarttouch catheter that does not have the SF technology as compared with the ThermoCool catheter that does have this technology, ThermoCool Smarttouch SF. While both catheters have labelled indications for treating paroxysmal atrial fibrillation (AF), only ThermoCool Smarttouch- Surround Flow- (STSF) is labelled for persistent AF; this indication was obtained by an investigational device exemption

clinical trial. The second portion of the study sought to compare the safety and effectiveness of the ThermoCool Smarttouch SF catheter with the ThermoCool Smarttouch and the Navistar catheters for ischaemic ventricular tachycardia (VT); the latter two catheters have a labelled indication for recurrent drug/device refractory sustained monomorphic VT due to prior myocardial infarction, while the Smarttouch SF catheter does not. The Navistar catheter is an earlier form of the ThermoCool catheter initially approved by the FDA in 2006 before contact force monitoring (Smarttouch) and improved cooling irrigation (SF) were added.

This research study first sought to identify the medical devices (ablation catheters) of interest in the health system electronic information systems and then to link to the pertinent patient populations who received treatment with these catheters. Afterwards, the performance of codes/algorithms to identify key safety and effectiveness outcomes of interest (ischaemic stroke, cardiac tamponade, acute heart failure and arrhythmia-related hospitalisation) was compared with clinician chart review in a small sample of patients at each health system (up to 25 patients per health system). All analyses were conducted individually at each health system using a decentralised model⁹; summary results were shared across researchers from the three institutions, but no patient-level data were shared.

Device, procedure and patient population identification

We identified patients who had received treatment with the medical devices (ablation catheters) of interest. The catheters were identified first (as opposed to the patient cohort), because the use of the catheters was necessary to ensure the feasibility of the study. Different strategies were used at each health system to identify the specific catheters used for ablation, including the device identifier (DI) component of the US FDA required unique DIs (UDIs) that were supplied by Johnson & Johnson and were available in the FDA's Global Unique Device Identification Database. These DIs were identified in supply chain and point-of-care inventory management data. Mayo Clinic also used catalogue numbers. For data prior to 2016, Mercy identified device information used in the procedures from a combination of Healthcare Common Procedure Coding System codes and device billing information from specific charge codes in the EHR (online supplemental table 1). Then, devices were linked to the specific patients who received treatment with them.

Once ablation catheters were identified, we determined types of the ablation procedure and arrhythmia for which the patients underwent ablation. We identified patients who had either AF or VT ablations performed with these catheters using Current Procedural Terminology (CPT) codes associated with the procedures (online supplemental table 2). We further identified the subset of these patients who had persistent AF and VT using International Classification of Diseases (ICD)-9-Clinical Modification (CM) (for VT only) and ICD-10-CM diagnosis codes for their arrhythmia prior to or during the clinical

encounter that included the ablation procedure (online supplemental table 3). ICD-9 and 10-CM codes were used after filtering based on CPT codes to add additional specificity, since the CPT codes do not identify the precise patient population of interest. Among patients undergoing AF ablation, defining the persistent AF phenotype was possible with ICD-10-CM, but not with ICD-9-CM, because the latter system does not include codes for AF subtypes. Patients with diagnosis codes only for unspecified AF were excluded, because of inability to subtype their AF. Patients with codes for long-standing persistent AF were included as persistent AF.

Among patients undergoing ablation for VT, classifying patients by ischaemic and non-ischaemic VT required identifying the subset who had prevalent ischaemic heart disease (online supplemental table 4). This decision was based on the assumption that ischaemic heart disease codes, including those for both acute myocardial infarction as well as chronic ischaemic heart disease, would distinguish patients with ischaemic from non-ischaemic VT.

Finally, in order to understand the availability of longitudinal data for outcome ascertainment, we determined the duration of follow-up through in-person encounters of patients receiving catheter ablation within each of the three health systems at 7 days, 30 days, 6 months and 1 year. Only encounters at the end of, or subsequent to, the time period of interest were included. Follow-up was defined as an encounter identified using an algorithm of in-person contact that included both face-to-face visits and remote contact, such as telephone visits, with any representative at the given health system. These encounters were not limited to those with cardiology clinicians.

Clinical outcomes

Pertinent clinical outcomes for patients undergoing ablation for AF or VT were then identified using ICD-9-CM and ICD-10-CM diagnosis and procedure codes and CPT codes (CPT codes were used for cardiac tamponade) based on previous publications,^{10 11} domain knowledge and the National Institute of Health's Value Set Authority Center (online supplemental table 5).¹² A panel of four cardiologists, including three cardiac electrophysiologists, then reviewed the list of codes and consensus was reached on the codes for inclusion using a modified Delphi process. The diagnosis codes listed as the primary discharge diagnoses were then applied to health system data to identify patients with the clinical outcome of interest. Physicians on the research team then performed manual chart review to determine the positive predictive value (PPV) of these code algorithms relative to clinician verification within the patient population of interest. When there were 25 or fewer events, all charts were reviewed. When there were more than 25, then 25 randomly selected patient charts for each of three serious safety outcomes (ischaemic stroke, cardiac tamponade and acute heart failure) and a single effectiveness outcome (arrhythmia-related hospitalisation) were reviewed. The goal of this analysis

Table 1 Number of times that the ablation catheters of interest were used across all procedures, by health system

Health system	ThermoCool ST	ThermoCool ST SF
Mercy Health (1 January 2014–7 August 2019)	1329	740
Mayo Clinic (1 January 2014–31 December 2018)	2545	401
Yale New Haven Hospital (1 October 2017–30 June 2019)	406	375
Total	4280	1516

was to determine whether a specific outcome of interest, focusing on outcomes commonly assessed in randomised controlled trials of ablation catheters,^{13 14} was accurately identified in a broad population of patients with the algorithm. The number of charts selected for each clinical event reflected both the number of patients identified with a given event in a broad patient cohort within each health system's data and resources for performing chart review. We calculated the 95% CIs based on the efficient-score method.¹⁵

RESULTS

Device identification

Device data were obtained from 1 January 2014 through 7 August 2019 at Mercy Health, from 1 January 2014 through 31 December 2018 at Mayo Clinic and from 1 October 2017 through 30 June 2019 at Yale New Haven Hospital (YNHH). In total, 4280 ablations were performed for patients with the catheter being investigated for persistent AF (ThermoCool ST), ranging from 406 at YNHH to 2545 at Mayo Clinic (although there were more than 3 years fewer data at YNHH due to inability to ascertain device). Overall, 1516 ablations were performed with the catheter being investigated for ischaemic VT

(Thermocool ST SF), ranging from 375 at YNHH to 740 at Mercy Health (table 1).

Patient population

EHR data were obtained from 1 January 2014 through 20 February 2020 at Mercy Health; 1 January 2014 through 31 December 2019 at Mayo Clinic; and 1 February 2013 through 13 August 2019 at YNHH (additional data through 31 December 2019 were obtained for longitudinal follow-up). Overall, a total of 357181 patients with AF were identified, including 27864 patients with persistent AF and 266001 patients with ICD codes for 'unspecified' and 'other' AF (patients were allowed to have multiple diagnoses for types of AF). In total, 59425 patients with VT were identified, including 39092 with ischaemic VT.

A large proportion of patients had ICD-10-CM codes for various combinations of paroxysmal, persistent and chronic AF as well as the non-specific ICD-9-CM codes. In these instances, codes for paroxysmal AF did not necessarily appear in the record first with codes for persistent AF entered at a later date, even though this is the well-established disease progression.¹⁶ The ultimate decision was to use the ICD-10-CM diagnosis codes for the ablation procedure. If there were multiple codes, then the most advanced in terms of expected disease progression was selected.

With regards to ablation, we identified 8676 ablation procedures for AF, ranging from 1299 at YNHH to 4906 at Mayo Clinic (table 2). We identified 1865 ablation procedures for VT, ranging from 198 at Mercy to 1140 at Mayo Clinic. An additional 8676 ablations had another primary diagnosis or a missing primary diagnosis.

Evaluation of follow-up

The duration of patient follow-up as ascertained using information from each of the three health system's EHRs ranged from 91% to 97% at ≥ 7 days, 89% to 96% at ≥ 30 days, 77% to 90% at ≥ 6 months, and 66% to 84% at ≥ 1 year (figure 1). Investigation of follow-up at YNHH identified that YNHH's electrophysiology laboratory has allowed non-YNHH physicians to perform procedures

Table 2 Cardiac catheter ablation procedure counts within the populations of interest (atrial fibrillation and ventricular tachycardia)

Health system	Ablation procedures (AF population)	Ablation procedures (VT population)	Ablation procedures ('other primary' and 'missing primary')	Total
Mercy Health (1 January 2014–20 February 2020)	2471	198	2530	5199
Mayo Clinic (1 January 2014–31 December 2019)	4906	1140	3142	8932
Yale New Haven Hospital (1 February 2013–13 August 2019)	1299	527	3004	4830
Total	8676	1865	8676	18961

AF, atrial fibrillation; VT, ventricular tachycardia.

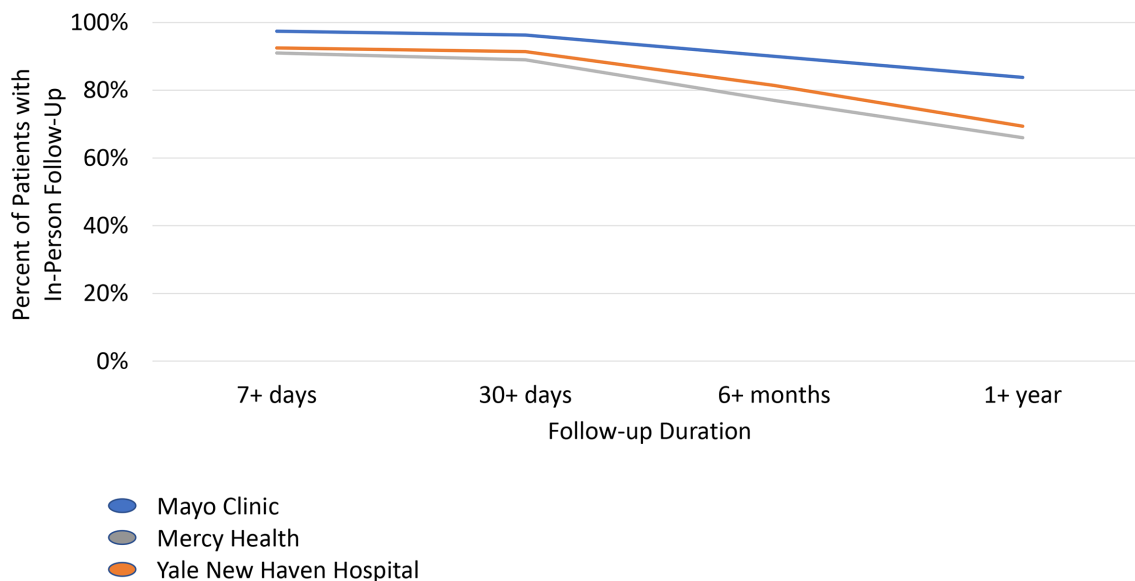


Figure 1 In-person follow-up encounters for patients after catheter ablation. Mercy data are for all ablation procedures from 2009 through 2018; Yale New Haven Hospital data are for all ablation procedures between February 2013 and August 2019; Mayo Clinic data are for all ablation procedures between January 2009 and December 2020.

there, and some YNHH physicians follow patients at non-YNHH clinics; therefore, these patients' follow-up would not be within YNHH's EHR.

Positive predictive values of codes/algorithms for clinical outcomes

Overall, PPVs were 96.4% for acute heart failure, ranging from 80% to 100% with between 10 and 25 charts reviewed at each health system (table 3). For ischaemic stroke, the overall PPV was 63.4%; this included 29.2%, 91.7% and 100% at the three health systems with between 4 and 24 charts reviewed. The PPV for cardiac tamponade was 100%, with two charts reviewed at one health system. The overall PPV for the effectiveness endpoint, arrhythmia-related hospitalisation, was 55.7% and ranged from 26.7% to 84.0% with between 15 and 25 charts reviewed.

Data resource use

A key factor in assessing the feasibility of using the NESTcc Network Collaborator EHR databases for evaluation of safety and effectiveness is whether the sample size is sufficient. Based on the feasibility study's catheter ablation

procedure counts and the anticipated increase in procedure numbers, it was expected that the sample size will be adequate to conduct a study using RWD from two of these participating health systems with high follow-up rates to examine expansion of indication of cardiac catheters to treat ischaemic VT and persistent AF. This phase 2 study is further evaluating and refining codes/algorithms for identification of patient populations, covariates and outcomes and will be focused on both short-term and long-term safety and effectiveness outcomes and could be the first label expansion study that solely uses electronic data from health systems.

DISCUSSION

In this study, we determined that the participating health systems had used adequate numbers of ThermoCool catheters and could obtain data of sufficient quality from their electronic information systems to evaluate the safety and effectiveness of cardiac ablation catheters. There were several important lessons learnt from this feasibility

Table 3 Positive predictive values of algorithms for primary diagnosis codes of safety and effectiveness outcomes based on physician-led chart reviews

Clinical outcome	Mercy Health	Mayo Clinic	Yale New Haven Hospital
Acute heart failure	21/21 (100%; 95% CI, 80.8% to 100%)	25/25 (100%; 95% CI, 83.4% to 100%)	8/10 (80.0%; 95% CI, 44.2% to 96.5%)
Ischaemic stroke	22/24 (91.7%; 95% CI, 71.5% to 98.5%)	7/24 (29.2%; 95% CI, 13.4% to 51.2%)	4/4 (100%; 95% CI, 39.6% to 100%)
Arrhythmia hospitalisation	9/21 (42.9%; 95% CI, 22.6% to 65.6%)	21/25 (84.0%; 95% CI, 63.1% to 94.7%)	4/15 (26.7%; 95% CI 8.9% to 55.2%)
Cardiac tamponade	None reported	Not explored	2/2 (100%; 95% CI, 19.8% to 100%)



study, which will not only inform our investigations of the ablation catheters in question but also the use of RWD generally for medical device-specific studies.

A key strength was the successful use of a decentralised model for research.⁸ All three health systems retained their data behind their individual firewalls, but data were collected using common definitions that will enable research using distributed analytics. This will provide a much larger sample size than from a single health system. The infrastructure was additionally built for two similar studies (focused on persistent AF and ischaemic VT); these commonalities suggest that a reusable infrastructure can be created for answering multiple research questions about real-world medical device safety and effectiveness.

In addition, several challenges were encountered in the feasibility stage of this research as well as insights that have led us to conclude that the challenges are all addressable. An overall challenge was that diagnostic codes lacked sufficient resolution for creating patient cohorts, which may necessitate expanding data types used for computed phenotypes. ICD-9-CM codes lack specificity for subtypes of AF. Because our ablation indication of interest was persistent AF, this prevents analysis of specific phenotypes of atrial fibrillation and would require alternative extraction approaches. Fortunately, the ICD-10-CM transition (1 October 2015) provides additional detail to AF subtypes. However, many patients had non-specific ICD-10-CM codes recorded; if the specific AF subtype cannot be documented, these patients may be dropped as they cannot be determined to have persistent AF. One option to address this challenge is to create and validate an algorithm (eg, such as including the use of antiarrhythmic medications) that may increase probability a patient has persistent AF. We also can use the Mayo Clinic cardiac ablation registry, a quality improvement database that captures most ablations with nurse-abstracted data, to validate our coding algorithms for identifying persistent AF. Additionally, Mercy has previously validated natural language processing algorithms to identify patients with persistent AF in Mercy's notes text. To ensure a consistent definition of AF, we started with use of the ICD-10-CM diagnosis code at the time of the ablation procedure. If there were multiple codes, we considered refining based on disease progression in future decision-making.

Similarly, while there are codes for identification of VT ablation, there are no specific ICD-9-CM or ICD-10-CM codes for identification of ablation of ischaemic VT. Patients were determined to have ischaemic VT using ICD-9-CM and ICD-10-CM diagnosis codes for ischaemic heart disease,¹⁷ assuming these codes in a patient's history would indicate an ischaemic aetiology of VT. The accuracy of these diagnostic codes, and what look-back period to use, for ischaemic heart disease is unclear.

Second, using health system data to create a cohort of patients who underwent the procedure of interest using a specific device has challenges. Ascertainment of pertinent covariates using prior diagnosis codes may be limited unless patients were cared for routinely within a health

system prior to the procedure. Missing data for this reason has been termed EHR-discontinuity and can risk biasing studies through misclassification of confounding variables and outcomes.¹⁸ A solution is to use an algorithm to identify patients with high data completeness in the EHR.¹⁸ For patients receiving ablation within a referral centre, chart histories of covariates (eg, prior ablation) may not be in discrete fields and, thus, may be incompletely ascertained if only automated extraction is used and, thus, limit accurate risk adjustment. However, these data may be available to a manual abstractor. Another potential solution is linkage with insurer claims data, such as Medicare fee-for-service claims, or registry data. (There are currently multiple national AF ablation registries and sometimes local registries, such as the Mayo Clinic cardiac ablation registry.) Another challenge is that other covariates that describe how the catheter was used may not be in the EHR or captured by current coding algorithms; this is particularly important as different operators may use different ablation techniques, and ablation techniques, such as the use of high-power short-duration ablation, may evolve over time. A solution is to use procedural data captured by the ablation technology during the procedure that records numerous parameters employed by the operator, including power, contact force, lesion duration, continuous versus point-by-point ablation and lesion sets, which are not captured with ICD codes. Additionally, important covariates such as centre ablation volume and operator experience can be obtained from health system data.

Third, ensuring that patients have sufficient follow-up within the given health system is important for long-term outcome ascertainment, particularly at regional referral centres. In-hospital complications will nearly always be captured. However, postdischarge, patients may receive care at outside health systems. If they experience a complication, they may present to the nearest emergency department, geographically distant from the centre where the ablation procedure was performed. Follow-up decreased progressively from 7 days to 30 days to 6 months and then to 1 year. We found that at least two-thirds of patients had follow-up at 1 year for all three participating health systems. Follow-up beyond 1 year is likely to be challenging. However, because we did not capture care at outside health systems, it is possible that this may undercount pertinent clinical conditions or events, including hospitalizations; although possible, we do not think that such missed events are likely to occur often when patients continue to see clinicians within the same health system. Low follow-up data capture rates in individual health system EHRs must, nevertheless, be identified. Strategies to improve data capture, such as through patient-centred digital health data sharing platforms that track outcomes in multiple EHR systems,¹⁹ could increase the amount of available follow-up data. Another possible solution is identification of a subpopulation that likely has follow-up within the health system where the ablation is performed (eg, within a close geographic radius or receiving primary and/or cardiology care within the health system). Finally,

linkage to claims or registry data could provide additional follow-up.

Fourth, although a small number of patients were identified with the four outcomes of interest, there was variability in the PPVs of using diagnosis codes to identify the safety outcomes of interest across outcomes and across healthcare systems, despite filtering on primary discharge diagnosis codes. For example, the PPVs for identification of stroke at one health system were relatively low, which may have resulted from the inability of the algorithm to distinguish patients with acute strokes from those with history of prior stroke. Using primary discharge diagnosis codes for hospitalizations for serious events like stroke may be less subject to misclassification than codes from outpatient visits. However, some patients with peri-procedural safety events may be missed by reliance solely on primary diagnosis codes, since safety events that occur during the procedure may be entered as secondary diagnoses while the primary diagnosis is used for the morbidity that led to the procedure. Both primary and secondary diagnoses codes should likely be used. Mid-term and long-term evaluation for events like heart failure that have a high background rate in the target population may make it difficult to differentiate between a prior event and a treatment-related occurrence using diagnostic codes alone. Finally, additional clinically important outcomes, such as atrio-oesophageal fistula and pulmonary vein stenosis, which are expected to be identified only at longer duration of follow-up, also need to be identified.

There are multiple possible solutions for improving accuracy of outcome ascertainment, which are similar to the strategies to address cohort identification. For example, prior studies have validated a variety of diagnosis codes against chart review in electronic health record studies.¹¹ Additionally, tools for phenotype development and evaluation using machine learning approaches are increasingly being made available to assist these efforts.²⁰ Ultimately, comparing the diagnostic codes or algorithms (composed of several codes for presence, absence, timing, setting and coincident diagnoses and treatments) with clinician review of electronic health records to determine extent of concordance between codes and clinical judgement may be necessary. While our study only had the resources to examine PPVs and sometimes identified smaller numbers of adverse events than would be expected based on the peer-reviewed literature, our study is limited because we did not assess negative predictive values, sensitivity, specificity, and accuracy of outcome ascertainment among patients with adequate follow-up within health systems to ensure that outcomes are not missed. If coding algorithms fail to perform adequately in this regard, an alternative approach to safety event identification will be considered. A group of patients at high risk of the outcome of interest who have negative results on the algorithm to detect the outcome of interest will require manual review, which can be time-consuming. Such analyses should examine specific sections of a patient's EHR data, such as clinical notes where, if a given clinical event

had occurred, it would be recorded. For example, cardiology or cardiac electrophysiology notes should capture ablation-related complications. This approach has been used successfully in validating algorithms to detect opioid overdose²¹ and addiction²² in postmarketing studies overseen by the FDA. In our study, it is possible that clinicians who performed the chart review were, in some cases, the same as those performing the procedures; it is possible that this could bias towards undercounting of adverse events. While this potential source of bias is not critical in a feasibility study of this type, efforts will be needed to ensure that methods are unbiased and consistent across different sites in a study of device effectiveness and safety.

Finally, effectiveness outcomes need to be carefully chosen to ensure that they are meaningful patient-oriented metrics. While we did not evaluate efficacy end points in this study, for future research, we decided to use clinical outcomes that reflect sequelae of arrhythmia recurrence: for AF, this includes rehospitalisation for AF or interventions addressing a new atrial tachyarrhythmia, including cardioversion, repeat ablation or new antiarrhythmic drug prescription; for ischaemic VT, this includes hospitalisation for VT or repeat ablation as well as heart failure, since that can be a sequela of VT. These are meaningful measures of effectiveness since the goal of ablation is to prevent additional treatment for an arrhythmia or serious consequences thereof, primarily heart failure, recurrent VT and mortality. Arrhythmia ICD diagnosis and procedure codes alone (ie, AF or VT codes) for outcome identification may have low PPVs because clinicians may add these codes of past events to follow-up visits since they can help with medical history or reimbursement. Even relying on primary diagnosis codes may not be sufficient, and manual chart review may be necessary to improve PPV and, therefore, it may not be easily achievable. In the case of AF, for example, physicians may maintain the diagnosis for clinic visits if patients are continued on therapeutic anticoagulation for thromboembolism prophylaxis. We found that PPVs using this approach for arrhythmia-related hospitalisation had variation across health systems (27%, 43% and 84%); but, because the numbers assessed were small, the CIs around the point estimates of several outcomes were overlapping. Additionally, for patients with AF, a blanking period (ie, time period when arrhythmia recurrences are not included) needs to be considered, given the high recurrence rate during this 60-day or 90-day period postablation.²³ Other data sources that could be helpful for outcome ascertainment, such as results from electrocardiograms, outpatient rhythm monitors or cardiac implantable electronic device (ie, pacemaker, implantable cardioverter defibrillator or implantable loop recorder) interrogations, were evaluated and found to be not feasible to easily obtain from the EHR databases used in our study; however, novel approaches are creating methods to import these data in standardised formats.

CONCLUSIONS

RWD collected during routine care present tremendous promise for use in medical device evaluations. The current feasibility study demonstrates the potential for evaluating the safety and effectiveness of new technology added to cardiac ablation catheters along with the challenges inherent in performing studies using health system data. The feasibility study also describes strategies to overcome these challenges and to help make RWD fit-for-purpose to generate RWE that can be used to support decisions by regulators, payers, clinicians and patients.

Author affiliations

¹Department of Medicine, University of California San Francisco School of Medicine, San Francisco, California, USA

²Department of Health Sciences Research, Mayo Clinic, Rochester, Minnesota, USA

³Mercy Clinic, St. Louis, Missouri, USA

⁴Department of Internal Medicine, Cardiovascular Medicine, Yale School of Medicine, New Haven, Connecticut, USA

⁵Mercy Research, Chesterfield, Missouri, USA

⁶Department of Internal Medicine, Yale School of Medicine, New Haven, Connecticut, USA

⁷Center for Outcomes Research and Evaluation, Yale-New Haven Hospital, New Haven, Connecticut, USA

⁸National Evaluation System for health Technology Coordinating Center (NESTcc), Medical Device Innovation Consortium, Arlington, Virginia, USA

⁹Medical Device Epidemiology and Real-World Data Science, Johnson & Johnson, New Brunswick, New Jersey, USA

¹⁰Department of Cardiovascular Medicine, Mayo Clinic, Rochester, Minnesota, USA

¹¹Department of Laboratory Medicine, Yale School of Medicine, New Haven, Connecticut, USA

Twitter Joseph P Drozda, Jr. @jpdrozda

Acknowledgements This project was supported by a research grant from the Medical Device Innovation Consortium (MDIC) as part of the National Evaluation System for Health Technology (NEST), an initiative funded by the U.S. Food and Drug Administration (FDA). Its contents are solely the responsibility of the authors and do not necessarily represent the official views nor the endorsements of the Department of Health and Human Services or the FDA. While MDIC provided feedback on project conception and design, the organisation played no role in collection, management, analysis and interpretation of the data, nor preparation, review and approval of the manuscript. The research team, not the funder, made the decision to submit the manuscript for publication.

Contributors All named authors meet the International Committee of Medical Journal Editors (ICMJE) criteria for authorship for this article. Manuscript writing: SD and JD. Manuscript revision for important intellectual content: GJ, AAD, DJF, EB, JC, JGA, JR, KRE, KCF, NDS, PC, PN, SZ, TF, WLS, YY. Individual health system site leadership: JR, NDS, JD. Clinical expertise: SD, AAD, DJF, JGA, PN. Informatics and analytic expertise: GJ, EB, JC, TF, WLS, YY. Medical device expertise: KRE, PC, SZ. Program management: KCF.

Funding Funding for this publication was made possible, in part, by the Food and Drug Administration through grant 1U01FD006292-01. Views expressed in written materials or publications and by speakers and moderators do not necessarily reflect the official policies of the Department of Health and Human Services; nor does any mention of trade names, commercial practices, or organization imply endorsement by the US Government. We also acknowledge the analytic assistance of H. Patrick Young, Yale School of Medicine.

Competing interests SD receives research funding from the National Heart, Lung, and Blood Institute (NHLBI, K12HL138046) of the National Institutes of Health (NIH), from the Medical Device Innovation Consortium as part of the National Evaluation System for health Technology Coordinating Center (NESTcc), Food and Drug Administration (FDA), Greenwall Foundation, and Arnold Ventures. DJF has received educational grants from Boston Scientific, Medtronic, and Abbott; research grants from the American Heart Association, National Cardiovascular Data Registry, Boston Scientific, Abbott, Medtronic, Merit Medical, and Biosense Webster; consulting fees from Abbott and AtriCure. JGA has received consulting fees from

Biosense Webster. JSR received research support through Yale University from Johnson and Johnson to develop methods of clinical trial data sharing, from the Food and Drug Administration to establish Yale-Mayo Clinic Center for Excellence in Regulatory Science and Innovation (CERSI) program (U01FD005938), from the Medical Device Innovation Consortium as part of NESTcc, from the Agency for Healthcare Research and Quality (R01HS022882), from the National Heart, Lung and Blood Institute of the National Institutes of Health (NIH) (R01HS025164, R01HL144644), and from the Laura and John Arnold Foundation to establish the Good Pharma Scorecard at Bioethics International and to establish the Collaboration for Research Integrity and Transparency (CRIT) at Yale. NDS has received research support through Mayo Clinic from the Food and Drug Administration to establish the Yale–Mayo Clinic Center for Excellence in Regulatory Science and Innovation program (U01FD005938), from the Centers of Medicare and Medicaid Innovation under the Transforming Clinical Practice Initiative, from the Agency for Healthcare Research and Quality (U19HS024075, R01HS025164, R01HS025402, R03HS025517), from the National Heart, Lung, and Blood Institute of the National Institutes of Health (R56HL130496, R01HL131535), from the National Science Foundation, and from the Patient Centered Outcomes Research Institute to develop a Clinical Data Research Network. PC is an employee of Johnson & Johnson; the manufacturer of the ThermoCool catheters, and Biosense Webster, is a Johnson & Johnson company. However, this study was a feasibility study to evaluate the suitability of the databases to conduct a study and no evaluation of the safety and effectiveness of any medical device was conducted in the study. PAN receives research funding from National Institutes of Health (NIH, including the NHLBI and the National Institute on Aging (NIA)), Agency for Healthcare Research and Quality (AHRQ), FDA, and the American Heart Association (AHA). He is a study investigator in an ablation trial sponsored by Medtronic. PAN and Mayo Clinic are involved in potential equity/royalty relationship with AliveCor. SZ is an employee of Johnson & Johnson; the manufacturer of the ThermoCool catheters and Biosense Webster, is a Johnson & Johnson company. WLS was an investigator for a research agreement, through Yale University, from the Shenzhen Center for Health Information for work to advance intelligent disease prevention and health promotion; collaborates with the National Center for Cardiovascular Diseases in Beijing; is a technical consultant to Hugo Health, a personal health information platform, and cofounder of Refactor Health, an AI-augmented data management platform for healthcare; is a consultant for Interpace Diagnostics Group, a molecular diagnostics company. In the past 36 months, JPD has received research support from Medtronic and Johnson & Johnson. His non-dependent son is an employee of Boston Scientific.

Patient consent for publication Not applicable.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement No data are available. Data are unavailable for sharing because they contain a significant amount of personal health information.

Open access This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>.

ORCID iDs

Sanket S Dhruva <http://orcid.org/0000-0003-0674-2032>

Peter A. Noseworthy <http://orcid.org/0000-0002-4308-0456>

Joseph P Drozda, Jr. <http://orcid.org/0000-0001-9181-8534>

REFERENCES

- 1 Sherman RE, Anderson SA, Dal Pan GJ, *et al*. Real-world evidence - what is it and what can it tell us? *N Engl J Med* 2016;375:2293–7.
- 2 *21st Century Cures Act. PI 114-255* 2016 United States
- 3 Food & Drug Administration. Use of real-world evidence to support regulatory decision-making for medical devices, 2017. Available: www.fda.gov/downloads/medicaldevices/deviceregulationandguidance/guidancedocuments/ucm513027.pdf [Accessed July 1, 2021].
- 4 Pappas G, Berlin J, Avila-Tang E, *et al*. Determining value of coordinated registry networks (CRNs): a case of transcatheter valve therapies. *BMJ Surg Interv Health Technol* 2019;1:e000003.
- 5 Dhruva SS, Ross JS, Desai NR. Real-world evidence: promise and peril for medical product evaluation. *PT* 2018;43:464–72.
- 6 Shuren J, Califf RM. Need for a national evaluation system for health technology. *JAMA* 2016;316:1153–4.
- 7 Fleurence RL, Shuren J. Advances in the use of real-world evidence for medical devices: an update from the national evaluation system for health technology. *Clin Pharmacol Ther* 2019;106:30–3.

- 8 Jiang G, Dhruva SS, Chen J, *et al.* Feasibility of capturing real-world data from health information technology systems at multiple centers to assess cardiac ablation device outcomes: a fit-for-purpose informatics analysis report. *J Am Med Inform Assoc* 2021;28:2241–50.
- 9 Fleurence RL, Blake K, Shuren J. The future of registries in the era of real-world evidence for medical devices. *JAMA Cardiol* 2019;4:197–8. doi:10.1001/jamacardio.2018.4933
- 10 Kokotailo RA, Hill MD. Coding of stroke and stroke risk factors using international classification of diseases, revisions 9 and 10. *Stroke* 2005;36:1776–81.
- 11 Saczynski JS, Andrade SE, Harrold LR, *et al.* A systematic review of validated methods for identifying heart failure using administrative data. *Pharmacoepidemiol Drug Saf* 2012;21 Suppl 1:129–40.
- 12 Bodenreider O, Nguyen D, Chiang P, *et al.* The NLM value set authority center. *Stud Health Technol Inform* 2013;192:1224.
- 13 Marrouche NF, Brachmann J, Andresen D, *et al.* Catheter ablation for atrial fibrillation with heart failure. *N Engl J Med* 2018;378:417–27.
- 14 Packer DL, Mark DB, Robb RA, *et al.* Effect of catheter ablation vs antiarrhythmic drug therapy on mortality, stroke, bleeding, and cardiac arrest among patients with atrial fibrillation: the CABANA randomized clinical trial. *JAMA* 2019;321:1261–74.
- 15 Newcombe RG. Two-sided confidence intervals for the single proportion: comparison of seven methods. *Stat Med* 1998;17:857–72.
- 16 Proietti R, Hadjis A, Alturki A, *et al.* A systematic review on the progression of paroxysmal to persistent atrial fibrillation: shedding new light on the effects of catheter ablation. *JACC Clin Electrophysiol* 2015;1:105–15.
- 17 Initiative S. Coding trend analyses: ischemic heart disease, 2019. Available: https://www.sentinelinitiative.org/sites/default/files/surveillance-tools/validations-literature/Ischemic_Heart_Disease_Trend_Analysis_Report.pdf
- 18 Lin KJ, Singer DE, Glynn RJ, *et al.* Identifying patients with high data completeness to improve validity of comparative effectiveness research in electronic health records data. *Clin Pharmacol Ther* 2018;103:899–905.
- 19 Dhruva SS, Ross JS, Akar JG, *et al.* Aggregating multiple real-world data sources using a patient-centered health-data-sharing platform. *NPJ Digit Med* 2020;3:60.
- 20 Swerdel JN, Hripcsak G, Ryan PB. PheValuator: development and evaluation of a phenotype algorithm evaluator. *J Biomed Inform* 2019;97:103258.
- 21 Green CA, Perrin NA, Hazlehurst B, *et al.* Identifying and classifying opioid-related overdoses: a validation study. *Pharmacoepidemiol Drug Saf* 2019;28:1127–37.
- 22 Carrell DS, Albertson-Junkans L, Ramaprasan A, *et al.* Measuring problem prescription opioid use among patients receiving long-term opioid analgesic treatment: development and evaluation of an algorithm for use in EHR and claims data. *J Drug Assess* 2020;9:97–105.
- 23 Calkins H, Kuck KH, Cappato R, *et al.* 2012 HRS/EHRA/ECAS expert consensus statement on catheter and surgical ablation of atrial fibrillation: recommendations for patient selection, procedural techniques, patient management and follow-up, definitions, endpoints, and research trial design: a report of the heart rhythm society (HRS) task force on catheter and surgical ablation of atrial fibrillation. developed in partnership with the european heart rhythm association (EHRA), a registered branch of the european society of cardiology (ESC) and the european cardiac arrhythmia society (ECAS); and in collaboration with the american college of cardiology (ACC), american heart association (AHA), the asia pacific heart rhythm society (APHRS), and the society of thoracic surgeons (STS). endorsed by the governing bodies of the american college of cardiology foundation, the american heart association, the european cardiac arrhythmia society, the european heart rhythm association, the society of thoracic surgeons, the asia pacific heart rhythm society, and the heart rhythm society. *Heart Rhythm* 2012;9:e21:632–96.