


# Lessons on the use of real-world data in medical device research: findings from the National Evaluation System for Health Technology Test-Cases

Journal of **Comparative Effectiveness Research**

Justin W Timbie<sup>\*,1</sup> , Alice Y Kim<sup>2</sup>, Lawrence Baker<sup>3</sup> , Rosemary Li<sup>4</sup> & Thomas

W Concannon<sup>3</sup> 

<sup>1</sup>RAND, Arlington, VA 22202, USA

<sup>2</sup>RAND, Santa Monica, CA 90401, USA

<sup>3</sup>RAND, Boston, MA 02116, USA

<sup>4</sup>RAND, Pittsburgh, PA 15213, USA

\*Author for correspondence: [jtimbie@rand.org](mailto:jtimbie@rand.org)

**Aim:** Although the US FDA encourages manufacturers of medical devices to submit real-world evidence (RWE) to support regulatory decisions, the ability of real-world data (RWD) to generate evidence suitable for decision making remains unclear. The 2017 Medical Device User Fee Amendments (MDUFA IV), authorized the National Evaluation System for health Technology Coordinating Center (NESTcc) to conduct pilot projects, or 'Test-Cases', to assess whether current RWD captures the information needed to answer research questions proposed by industry stakeholders. We synthesized key lessons about the challenges conducting research with RWD and the strategies used by research teams to enhance their ability to generate evidence from RWD based on 18 Test-Cases conducted between 2020 and 2022. **Materials & methods:** We reviewed study protocols and reports from each Test-Case team and conducted 49 semi-structured interviews with representatives of participating organizations. Interview transcripts were coded and thematically analyzed. **Results:** Challenges that stakeholders encountered in working with RWD included the lack of unique device identifiers, capturing key data elements and their appropriate meaning in structured data, limited reliability of diagnosis and procedure codes in structured data, extracting information from unstructured electronic health record (EHR) data, limited capture of long-term study end points, missing data and data sharing. Successful strategies included using manufacturer and supply chain data, leveraging clinical registries and registry reporting processes to collect and aggregate data, querying standardized EHR data, implementing natural language processing algorithms and using multidisciplinary research teams. **Conclusion:** The Test-Cases identified numerous challenges working with RWD but also opportunities to address these challenges and improve researchers' ability to use RWD to generate evidence on medical devices.

**Plain language summary: What is this article about?:** The purpose of the study was to evaluate the potential of real-world data (RWD) to support regulatory decision-making for medical devices by synthesizing information from pilot projects conducted by the National Evaluation System for health Technology Coordinating Center (NESTcc). We reviewed documents from 18 completed pilot projects and conducted 49 interviews with representatives from participating organizations to identify challenges and successful strategies in using RWD. We sought to derive lessons that could enhance the use of RWD for regulatory purposes and inform future research and investments in this area.

**What were the results?:** The results showed that the pilot projects faced multiple challenges, including difficulties in identifying specific medical devices due to the lack of unique device identifiers (UDIs), incomplete data capture and challenges in measuring long-term outcomes. Successful strategies included leveraging supply chain data, clinical registries and standardized data models, as well as involving multidisciplinary teams. Despite these efforts, the projects highlighted the need for improved data capture and sharing practices to make real-world data (RWD) more reliable for regulatory decision-making.

**What do the results of the study mean?:** The results indicate that while real-world data (RWD) has the potential to support regulatory decision-making for medical devices, significant challenges in data capture, identification and long-term outcome measurement must be addressed. For medical device industry stakeholders and regulators, this means prioritizing improvements in data collection practices, standardization and collaboration to ensure RWD can be reliably used for regulatory purposes.

**Shareable abstract:** New study synthesizes findings from pilot projects to evaluate the use of real-world data for regulatory decisions in medical devices, highlighting key challenges and successful strategies. #MedTech #RWD

First draft submitted: 5 May 2024; Accepted for publication: 23 July 2024; Published online: 16 August 2024

**Keywords:** claims • clinical trial • electronic health record • medical device • patient registry • real-world data • real-world evidence • regulatory

Vast amounts of data are produced during the routine delivery of medical care and its reimbursement. These data, also known as real-world data (RWD), include healthcare claims, population and condition-based registries, patient-generated health data (PGHD) and information contained in electronic health records (EHRs) [1]. Both regulators and manufacturers of medical products have expressed enthusiasm for using RWD to generate generalizable evidence on the benefits and risks of medical products at a larger scale and, potentially, lower cost, than traditional studies that are typically conducted under highly controlled conditions [2–4]. Although the FDA Center for Devices and Radiological Health (CDRH) has historically accepted clinical evidence on the potential benefits and risks of medical products derived from RWD (also known as real-world evidence, or RWE) to supplement evidence from clinical trials, CDRH recently encouraged manufacturers to submit RWE as the primary source of evidence for informing regulatory decisions [5]. However, many stakeholders remain uncertain about the ability of RWD to capture important patient-level details, device and procedure information and long-term outcomes [6–9], and whether regulators will consider the resulting RWE to be ‘fit-for-purpose’ for specific regulatory decisions [10].

In August 2016, the FDA awarded a cooperative agreement to establish the National Evaluation System for health Technology Coordinating Center (NESTcc) [11] to the Medical Device Innovation Consortium, a public–private partnership that seeks to accelerate patient access to new medical device technologies. NESTcc was founded to establish clear and efficient pathways for manufacturers and other stakeholders to generate timely, reliable and cost-effective evidence on medical devices using RWD specifically. As part of its commitments outlined in the 2017 Medical Device User Fee Amendments (MDUFA IV) [12], NESTcc was responsible for conducting pilot projects that would draw on diverse sources of RWD to generate evidence on medical device performance in a range of clinical areas. It was envisioned that these pilots might eventually lead to future studies that could support pre-market approvals and clearances, expanded indications, post-market safety and surveillance studies and coverage decisions [13].

Beginning in 2017, NESTcc began establishing a research network comprising academic institutions, healthcare providers and registry stewards (Table 1). These ‘network collaborators’ contributed RWD sources, typically on very large populations, as well as expertise in analyzing RWD [14]. NESTcc then issued calls to medical device manufacturers and other industry stakeholders to suggest concepts for Test-Cases designed to test the suitability of RWD for specific regulatory uses [11]. Organizations that submitted these concepts were referred to as ‘submitting organizations’. Test-Case teams comprising network collaborators and submitting organizations developed study protocols with specific study objectives, analyzed the data, provided monthly progress reports and submitted a final report of the study results. Ultimately, 21 Test-Cases covering different regulatory use cases were funded.

We identified key lessons on the use of RWD that emerged from the 18 Test-Cases that were completed by December 2023. Drawing on interviews with Test-Case team members and reviews of documents generated by each team, we identified the key challenges in using RWD to address the research questions proposed by submitting organizations and the strategies used by Test-Case teams to address these challenges. Based on these findings, we identified opportunities to enhance the use of RWD to support regulatory decision making.

**Table 1. Participants in the Test-Cases.**

Participant type	Organization type	Name of organization
Network collaborator	Health system	Duke University Health System
		Lahey Hospital & Medical Center
		Mayo Clinic
		Mercy Health
		Vanderbilt University Medical Center
		Yale New Haven Health System
	Clinical data research network	Insight CRN
		OneFlorida
		PEDSnet
		Stakeholders, Technology and Research CRN
	Public-private partnership	MDEpiNet
		HealthCore
	Claims aggregator	HealthCore
HealthCore		
Healthcare research organization	Regenstrief Institute	
	Regenstrief Institute	
Medical school with physician practices, affiliated hospitals	Weill Cornell Medicine	
	Weill Cornell Medicine	
Submitting organization	Medical device manufacturer	Abbott
		Adhesys Medical
		AventaMed
		Becton, Dickinson and Company
		Cook Medical
		Intrinsic Therapeutics
		Johnson & Johnson
		Medtronic
		W. L. Gore & Associates
		American Academy of Orthopaedic Surgeons
	Professional society	American Academy of Orthopaedic Surgeons
		American Sleep Apnea Association
	Patient advocacy organization	American Sleep Apnea Association
		American Sleep Apnea Association
	Regulatory agency	FDA
		FDA
	Public-private partnership	MDEpiNet
MDEpiNet		
Research organization	OrthoCRN	
	OrthoCRN	
Prescription digital therapeutics company	Pear Therapeutics	
	Pear Therapeutics	
Health system	Yale New Haven Health System	
	Yale New Haven Health System	

## Methods

We reviewed study protocols and reports that were generated by Test-Case teams, conducted interviews with representatives of participating organizations and conducted a thematic analysis of interview transcripts. The study protocol was approved by our institution's Human Subjects Protection Committee.

### Review of Test-Case documents

We reviewed the study protocol, final results report and post-study self-assessment for each Test-Case. Each document was submitted by the lead network collaborator to NESTcc with input from one or more supporting network collaborators and/or the submitting organization. All study protocols, including statistical analysis plans, were designed prospectively and submitted and approved by NESTcc before the research began. From the study protocols, we abstracted objectives, data sources, end points and data extraction and validation plans. From the final results reports and self-assessments, we identified successes, challenges, methodological changes, conclusions and participants' overall experience conducting the research.

### Interviews with Test-Case teams

We conducted a total of 49 interviews with Test-Case team members by videoconference between October 2020 and September 2022. Interviewees included 35 representatives of network collaborators, including 20 physician researchers and 15 non-physician researchers. We also interviewed 14 representatives of submitting organizations, including 8 representatives of medical device manufacturers (two directors of research, two directors of data science,

two directors of regulatory affairs, one director of health economics and outcomes research and one chief operating officer) and six other individuals representing research organizations, medical specialty societies and the FDA.

We sought to conduct interviews with all network collaborators and the submitting organization for each Test-Case to gather perspectives from participants with diverse roles. We developed semistructured interview protocols with information gathered from the document review that were tailored to each stakeholder and designed to gain deeper insight into the contents of the Test-Case documents and probe for information that might not have been reported in the documents. Interview topics included challenges conducting the research, views on the strength of evidence, research impact, future directions for the research and overall lessons learned. A team of three researchers conducted the interviews: one researcher led the interview, another probed for clarity or additional detail and a third took notes. Interviews typically lasted one hour, were recorded after receiving consent from participants, and were professionally transcribed.

### Data analysis

We developed an iterative, thematic coding structure based on the interview protocol and study aims. One researcher reviewed a subset of interview transcripts, coded transcripts by theme, created additional themes as needed and summarized all themes and supportive quotations in a findings document. A second researcher then reviewed both the transcripts and the findings document and refined the themes as needed. The first researcher then applied the same thematic coding structure to the remaining interview transcripts. The second researcher then reviewed all remaining transcripts to ensure that all relevant themes were included in the findings document.

### Limitations

Although the Test-Cases were diverse in their focus, research on medical devices that were not examined in the Test-Cases or that used different types of RWD might identify different lessons or different opportunities to optimize use of RWD. In addition, we did not conduct interviews with representatives of three of the 21 Test-Cases that had not been completed as of December 2023. Two of these Test-Cases were prospective studies that were awarded in the second round of Test-Cases and for which enrollment was delayed; the third Test-Case was canceled because of contracting delays. In addition, 10 of 59 stakeholders that we contacted declined to participate in an interview. These individuals were typically network collaborators who joined Test-Cases well after their launch, contributed a smaller sample size to the analysis, or played a more limited role in the project (e.g., statistical analysis alone). We had limited insight into the views of payers and FDA representatives because payers did not participate in the Test-Cases and FDA staff participated in only a few of them. Finally, research on specific devices can be highly sensitive, even in pilot studies, and some stakeholders might have been reluctant to share information about their experience or the quality of the evidence generated from their Test-Cases.

## Results

### Overview of the 21 Test-Cases

The 21 Test-Cases reflect research concepts submitted by 16 organizations. These organizations primarily included medical device manufacturers but also a research organization, professional society, patient advocacy organization, public-private partnership and the FDA (Table 1). The Test-Cases studied devices and technologies spanning ten medical specialties, with 12 of 21 Test-Cases addressing cardiac or orthopedic conditions (Table 2). The devices examined included implantables (e.g., heart valves, stents and orthopedic implants), devices used in surgical procedures (e.g., ablation catheters), diagnostic tests, durable medical equipment and mobile health applications such as one that was used to collect patient-reported outcomes.

The Test-Cases differed in their objectives, data sources and other design features (Table 3). Fourteen Test-Cases explored questions and data that could eventually support pre-market decisions including label expansions and pre-market approvals or clearances; ten covered post-market or surveillance use cases. Submitting organizations and network collaborators specified objectives for their Test-Cases, such as identifying members of the target population who received the device of interest in RWD, identifying specific device models and measuring patient-level outcomes. For example, one Test-Case used both EHR data and registry data to assess one-year outcomes for patients undergoing treatment with a manufacturer's cardiac ablation catheter. Nineteen of 21 Test-Cases analyzed data from EHRs, sixteen Test-Cases used a retrospective design and nearly one-half required at least some manual data extraction from EHRs.

**Table 2. Medical specialties and devices covered by the Test-Cases.**

Specialty	Device or procedure
Cardiology	Cardiac ablation catheter; cardiac implantable device leads; electrode renal denervation system; mechanical aortic heart valve; cardiovascular stent device; Apple Watch ECG diagnostic + mHealth <sup>†</sup>
Cardiovascular	Iliac branch endoprosthesis/stent
Dental	Craniomaxillofacial distractors
Dermatology	Suture, staples and adhesive strips
Otolaryngology	Ear tubes
Oncology	Lung cancer <i>in vitro</i> diagnostic test; surgical ablation device
Orthopedics	Annular closure device; intervertebral body fusion device; knee and hip implants; total knee arthroplasty
Respiratory	Positive airway pressure therapy
Sleep Medicine	mHealth for insomnia, prescription digital therapeutic (mobile app) <sup>†</sup>
Urology	mHealth for surgical mesh (mobile app); surgical mesh

<sup>†</sup> Findings covering these devices or procedures are not included in the analysis as these Test-Cases were not completed as of December 2023. ECG: Electrocardiogram.

**Table 3. Objectives and study design of the Test-Cases.**

Attribute	Details	Test-Cases (n)
Objective	Identify individuals in the target population	18
	Identify specific devices	13
	Measure patient and/or procedural characteristics	18
	Measure patient-level outcomes	21
	Identify comparison group	7
	Assess concordance among different real-world data sources	10
	Collect patient-generated data	4
Study type	Retrospective	16
	Prospective	5
Data source	Registry	6
	Claims	7
	Electronic health records	19
	Literature	1
	Patient-generated health data	4
	Biological samples	1
Method of data collection	Automated and manual extraction	9
	Automated extraction	11
	Literature review	1
	Qualitative data collection	1
	Survey	3
	Wearable device	2
	Mobile app	3
Data collection features	Common data model	10
	Unique device identifiers	5
	Natural language processing	1
	Regular expression matching	2
	Hashtag method for supplementing registry records	1
Timing of outcome assessment	Post-operative	1
	1 month	3
	3 months	1
	6 months	1
	1 year	5
	1+ year	1
	1–2 years	1
	2 years	5
	3 years	1
	N/A	2

### Challenges working with real-world data

Test-Case teams encountered several challenges working with RWD. Interviewees emphasized that although RWD hold many advantages, they are inherently more complex than experimental trial data, and best practices for addressing data gaps and collecting and analyzing data efficiently are still emerging.

*Identifying devices in the absence of unique device identifiers*

Few teams were able to leverage unique device identifiers (UDIs) for their Test-Cases because their organizations did not routinely capture them in their EHRs. Among those that did, the timing with which UDI capture began often varied across organizations, meaning different collaborators might contribute different years to the analysis. When UDIs were not available in structured tables, some Test-Cases attempted to search free-text notes in their EHR, but researchers often found that these notes did not consistently document device manufacturers or models. Supply chain databases provided a valuable source of UDIs for some network collaborators (as we discuss below), but some teams also had difficulty accessing these databases within their organizations.

*Capturing key elements & their appropriate meaning in structured data*

Most Test-Cases used EHR data that had been previously standardized using a common data model (CDM), such as the PCORnet CDM [15]. However, CDMs did not always cover all the fields needed to address the objectives of each Test-Case. For example, in one Test-Case involving ablation catheters, information such as health status scores (e.g., Eastern Cooperative Oncology Group scale), indicators of liver function (e.g., cirrhosis), tumor characteristics (e.g., size and location) and device details (e.g., ablation probe type and number of ablation cycles) were not available in the CDM. Test-Cases often had to combine standardized EHR data with data extracted from other sources or even unstructured data. Even if standardized EHR data were available, ensuring that the appropriate clinical meaning could be ascribed to each data element, particularly diagnosis and procedure data, could be a challenge. For instance, difficulty distinguishing a prior condition from a current diagnosis made it hard for some teams to attribute study end points to use of a device, as opposed to a preexisting condition:

*"It became a challenge to identify acute ischemic stroke. Participating organization D found that acute ischemic stroke associated with a hospitalization had a positive predictive value of 40 percent. And the problem we found as we delved into it is that once a patient had a stroke, whenever they're admitted to the hospital, stroke continues to be on the diagnosis list making it very difficult if you use ischemic stroke as your safety event. We've identified that we need to have a much more refined algorithm to separate true ischemic stroke from history of stroke."*

– Participating organization E

*Reliability of diagnosis & procedure codes in structured data*

Diagnosis and procedure information extracted from structured data could be unreliable for several reasons. Test-Cases had difficulty determining the laterality of joint replacement procedures, distinguishing mechanical heart valves from bioprosthetic valves, or identifying subtypes of knee arthroplasty procedures because of inadequate granularity in coding systems. Even when these systems contained sufficient detail, variation within and between network collaborators in coding practices created measurement challenges. For example, one Test-Case found that their collaborators varied in the use of codes for cardiac lead replacement procedures, which made it difficult for the team to distinguish mechanical device failures from non-mechanical failures. Similarly, one Test-Case team was unable to differentiate chronic and acute forms of a condition:

*"There's a thousand different ways that people will code acute otitis media or otitis media with effusion, and I think unless it's something that you do prospectively and you're actually following the patients, it's very hard to do retrospectively. There are hundreds of codes. So you would think something as straightforward as otitis media, which is bread and butter for pediatric health systems, would be a little clearer. We realized that we can't distinguish the two with data in the electronic health record. You just can't. So that's why we ended up combining them."*

– Participating organization F

When available diagnosis codes did not allow reliable measurement of conditions, some Test-Cases sought to use supplemental data. For example, one Test-Case used data from registries or information found in clinical notes to classify patients into one of three atrial fibrillation subtypes.

*Extracting information from unstructured RWD*

Test-Cases that required highly detailed information on the condition or treatment of interest were often required to search unstructured EHR data, such as medical notes or reports. For example, one Test-Case team that examined wound closure devices were required to search clinical notes to determine the type of suture material used, cleanliness

of the wound and layer of wound closure. Medical device details and pathology data, such as tumor characteristics, were particularly difficult to locate in unstructured text fields:

*"The other place we see this [challenge] is in pathology, so, what histology level of the tumor is it exactly? where does one find that in an EHR? Those things are typically not easily searchable. So, unless someone has already abstracted it because the site has their own tumor registry, it's difficult to get that information."*

– Participating organization A

Abstracting information from notes also created challenges because clinicians might describe a single procedure using very different terminology or might not provide the level of detail required. Two Test-Cases built text-matching algorithms to extract data from notes but had varying degrees of success locating the required information.

#### *Limited capture of long-term study end points*

Regulators often require data on medical device performance over several years, but EHR data might not capture long-term study end points for patients who change providers or relocate. This challenge can be significant at academic health systems that have large referral populations and may not provide ongoing care for patients after an index procedure. For one Test-Case, only 16 percent of patients had any end point data beyond 30 days after a major surgery. To improve the measurement of mortality end points, which often occur outside of hospitals, some collaborators linked their EHR data to national or state death registries to improve the validity of mortality outcomes. Difficulties with long-term follow-up of end points caused some Test-Cases to focus only on short-term outcomes and put less emphasis on long-term end points:

*"I would be cautious to really interpret the results very strongly because of the under-ascertainment of the outcomes. I do think that this [Test-Case] taught [us] that long-term outcome ascertainment in the absence of a prospectively collected outcome or registry link or claims link probably is not reasonable."*

– Participating organization G

#### *Data completeness*

Key data elements might be missing from the RWD used in Test-Cases because the databases were not designed to collect the information or because the data elements were inconsistently captured. Test-Cases that used EHR data might have access to prescription orders but not information about whether prescriptions are filled. Registries might contain rich specialty-specific data, such as histology, diagnoses and malignancy, but only for a selected population, such as patients with malignant tumors only. In other cases, missing data resulted from incomplete documentation of information that was typically captured in either structured data or medical notes. In one Test-Case focusing on children with acute otitis media, information about hearing or speech concerns, balance issues, or missed days of school might be captured in clinical notes or through routine questionnaires, but missing data could be a problem because clinical notes might not be standardized, physicians might use questionnaires to varying degrees and patients might not always respond to these questionnaires.

#### *Limits to data sharing as a result of privacy & proprietary concerns*

Negotiations around data sharing caused delays for some Test-Cases. Some interviewees felt that these delays were driven by a lack of trust on the part of administrators and lawyers who worried that sharing individual-level data increased the risk of reidentifying individual patients. Some Test-Cases reported delays even if no protected health information was shared. In some Test-Cases, privacy concerns ultimately limited the extent to which collaborators were willing to share data. For example, in one Test-Case, these concerns caused collaborators to analyze their own data separately rather than as a pooled sample:

*"The intention was never to aggregate individual-level data across the health systems. The assumption was always that, with the three health systems, we were going to each do the work in parallel, and then we were going to summarize our results in aggregate because we [didn't] have the authority to pool our data."*

– Participating organization C

In another Test-Case, the leaders of a health system were reluctant to share with their researchers supply chain data that contained information on cardiac devices implanted within their system.

### Keys to success

Interviewees described different data sources, methods and other assets that were key to the successful completion of their Test-Cases and could improve the efficiency and scale of follow-on research using RWD.

#### *Using manufacturer data to support research site selection & data collection*

Medical device manufacturers in three Test-Cases shared either patient-level or aggregated data with network collaborators to support the research. Aggregated sales data helped collaborators estimate likely sample sizes, and, in future studies, could inform the selection of research sites by identifying the highest-volume users of a device. This approach may be particularly helpful for research on rare diseases, where procedure codes found in EHRs may be unreliable, or when UDIs are not available in a site's EHR data. In one Test-Case, the submitting organization provided access to patient-level data that included patient names and procedure dates from a tracking database established by the manufacturer for FDA-required surveillance.

#### *Using supply chain data to identify use of medical devices*

Health systems maintain detailed records of their purchases and inventories of medical devices. Although supply chain data may not be used routinely by researchers, some stakeholders recognized the importance of these data in identifying use of specific devices when a health system did not regularly capture UDIs within its EHR. These data were considered useful because they contain information on use of specific devices on specific dates, and they are likely to be complete. In at least two Test-Cases, supply chain data were linked with EHR data to identify the dates on which a device was used in medical procedures, which, in turn, helped to link the use of devices to specific patients. Organizations could also use supply chain data to provide initial projections of device use to support the design of future studies.

#### *Using clinical data contained in registries*

A few Test-Cases used cardiac, orthopedic, cancer and other procedure- and condition-based registries to provide a rich source of curated, clinical data on patients receiving medical devices. Two Test-Cases that used an atrial fibrillation ablation registry and a vascular surgery registry cited the large number of research sites across the country that actively contributed to them as an indicator of their data quality. Test-Cases used registries to avoid the need to abstract data manually from EHRs or to serve as a gold standard for validating information extracted from EHRs. In other Test-Cases, state- and health system-based patient registries were used to support identification of the target population and measure selected outcomes. For example, in two oncology Test-Cases, network collaborators used state tumor registries, liver cancer-specific patient databases and local registries designed for lung cancer screening or biomarker development and validation to extract detailed patient information, such as the number, diameter and shape of tumors.

#### *Using common data models to extract EHR data*

Several Test-Cases accessed versions of their EHR data that had been standardized using the PCORnet CDM to automate the extraction of selected covariates and outcomes. These Test-Cases highlighted the value of CDMs in eliminating the need to map data elements required for the Test-Case to specific variables and tables in an organization's raw EHR data. In addition, for Test-Cases involving multiple network collaborators, CDMs allowed the creation of a common algorithm that could be used to extract data from the collaborators' EHRs. One benefit of using the PCORnet CDM in particular was that it eliminated the need for further validation of some study end points because data quality checks are already built into the data standardization process at organizations that participate in PCORnet [15].

#### *Using natural language processing & other advanced forms of text matching to extract data*

Only a single Test-Case used natural language processing (NLP) to extract clinical information from unstructured EHR data likely due to the time required to develop and validate these algorithms. In this Test-Case, the team leveraged a validated NLP algorithm to identify three atrial fibrillation subtypes, described earlier, which avoided the need to conduct chart reviews. Another Test-Case used regular expression matching to identify both the use of the device of interest (surgical mesh) and the four surgical procedures typically used to implant the device. The Test-Case used a 'waterfall' method for implementing the algorithm, such that the lead network collaborator developed a prototype that other collaborators adapted locally.

### *Using multidisciplinary research teams*

Many Test-Cases highlighted the importance of engaging physician specialists and data scientists in their Test-Cases to leverage their complementary expertise. Early involvement in the design of the Test-Cases by clinicians with experience using the device of interest helped teams select appropriate safety and/or effectiveness outcomes, specify the definitions to operationalize each outcome and determine which specific device features and variables were needed to address each study objective. Interviewees also felt that extensive involvement of data scientists was essential given the complex structure of EHR data, the challenges of linking data sets and the need for iterative refinement of algorithms to extract the data. Other stakeholders felt that a strong team of outcomes researchers and epidemiologists that are skilled in observational studies was important to serve as a bridge between clinical experts and data scientists.

## Discussion

The NESTcc Test-Cases represent one of the largest systematic efforts to date to explore the potential for using RWD in diverse regulatory use cases. They build on efforts launched by the FDA nearly a decade ago to encourage medical product developers to submit RWE to support regulatory decisions around product effectiveness—a major shift from the FDA’s historical reliance on RWE to support decisions on product safety. The agency’s efforts have included establishing the RWE Program and a guiding framework identifying high-level considerations that agency staff would use when reviewing RWE [16], publishing examples of RWD used in recent regulatory decisions [17] and releasing more specific guidance around use of RWE to support regulatory decisions for pharmaceuticals and biologics [18] and medical devices [19]. Previous studies have suggested that expanding the appropriate use of RWE in regulatory decisions is a ‘learning process’ for both industry and FDA and medical device stakeholders needed more specific guidance around appropriate uses to guide their research and regulatory strategies [9]. The Test-Cases, which spanned a wide range of devices and regulatory use cases, were designed to identify lessons that would help manufacturers improve their regulatory submissions while helping regulators develop more specific guidance that was informed by empirical studies.

The Test-Cases were valuable in identifying both challenges and successful strategies when working with RWD. In general, the extent of the challenges depended on the specificity of the data needed to address each Test-Case’s research questions and the data assets available within each organization. Test-Cases that sought to measure the performance of specific devices in narrowly defined patient populations and measure outcomes over one or more years faced varying degrees of difficulty depending on whether their organizations were able to capture and link data required to measure important patient and device characteristics and clinical outcomes. Network collaborators who could identify devices using UDIs in structured data tables and ensure longitudinal data capture within their health system or through regional partnerships were generally more likely to meet their study objectives. Follow-on research studies that build on the Test-Cases remain ongoing in some cases, and, in at least one case, has led to a successful regulatory decision [20].

The limitations of claims data for use in outcomes research have long been known. Similarly, prior studies have documented challenges with the completeness, structure and standardization of EHR data and the need for linkage to other data sources to support their use for clinical or population health research [21–23]. Some of these limitations have led researchers to pursue the development of coordinated registry networks, which are built around an existing or new registry and use linkages to other RWD to develop a curated source of clinically rich information on large populations [24–26]. In fact, the majority of examples of RWE that have supported regulatory decisions described in a recent FDA report have featured analyses derived from CRNs [17]. However, NESTcc focused largely on using EHR data for their Test-Cases in recognition that EHRs had a limited history of use in medical device research.

Our study adds to the growing literature documenting the potential value as well as the challenges using RWD. Recent studies have estimated a substantial return-on-investment and time savings from using RWD compared with traditional studies [27,28]. Mixed-methods studies synthesizing industry stakeholder views on the use of RWD have highlighted other benefits including developing more robust measures of device safety and performance and identifying opportunities to improve devices, but they also identified challenges conducting research with medical devices, particularly the identification of and tracing the use of devices throughout the healthcare delivery system—a challenge also highlighted by the Test-Cases [29]. Recent research that has examined factors associated with UDI implementation suggest that organizational attributes, such as the extent of external collaborations (including those with FDA), use of systems approaches to innovation and technology and both leadership and staff educational

**Table 4. Opportunities to improve use of real-world data identified through the Test-Cases.**

Opportunity	Key stakeholders	Benefits
Expand capture of UDIs and tracking the extent of UDI capture, by specialty, within health systems	Health systems, regulators and accrediting bodies	<ul style="list-style-type: none"> <li>Increases opportunities for health systems to participate in RWE studies</li> <li>Facilitates selection of research sites</li> <li>Enables larger-scale studies</li> <li>Strengthens incentives for UDI adoption</li> </ul>
Expand use of medical device supply chain data	Researchers and health systems	<ul style="list-style-type: none"> <li>Facilitates identification of target populations when EHRs lack UDIs</li> </ul>
Expand prospective collection of high-value, structured EHR data	Hospitals and health systems	<ul style="list-style-type: none"> <li>Supports research studies and health system's own quality improvement and value-based care initiatives</li> </ul>
Identify opportunities for effectively deploying NLP by assessing structure of clinical notes across sites	Organizations involved in RWE generation, network collaborators	<ul style="list-style-type: none"> <li>Helps manufacturers optimize selection of research sites where NLP could be used</li> </ul>
Explore methods for generating evidence on cost	Payers and medical device manufacturers	<ul style="list-style-type: none"> <li>Expands evidence generation on the value of medical devices</li> </ul>
Develop and disseminate libraries of computable phenotypes and end points	Medical device manufacturer or research organizations	<ul style="list-style-type: none"> <li>Facilitates reuse of algorithms</li> <li>Speeds up early phases of research studies</li> <li>Increases consistency in measurement</li> </ul>
Expand use and breadth of CDMs	Hospitals, health systems, research organizations, CDM developers	<ul style="list-style-type: none"> <li>Helps scale RWE generation</li> <li>Facilitates data sharing and harmonizes research processes</li> <li>Allows better capture of indication information and end points</li> </ul>
Create additional linkages between EHR and other data	Organizations involved in RWE generation, health systems	<ul style="list-style-type: none"> <li>Improves the richness of EHR data</li> <li>Improves measurement of long-term study end points (especially mortality)</li> </ul>
Foster sustainable, collaborative research networks	Research organizations or networks	<ul style="list-style-type: none"> <li>Facilitates data collection, sharing and analysis by aligning research infrastructure and processes</li> <li>Regional collaborations may allow more complete measurement of long-term study end points</li> </ul>

CDM: Common data model; EHR: Electronic health record; NLP: Natural language processing; UDI: Unique device identifier; RWE: Real-world evidence.

efforts around UDIs are associated with UDI implementation in health systems [30]. Other recent efforts to improve use of RWD have focused on standardizing and improving the rigor of documentation to support health technology assessments and regulatory decisions, including detailed descriptions of the provenance, completeness, accuracy and reliability of RWD [31,32].

Our findings suggest several steps that stakeholders across the medical device ecosystem could take to improve the scale, reliability and efficiency of research with RWD—particularly when using data derived from EHRs (Table 4). First, hospitals and health systems could take steps to capture UDIs routinely in their EHR data to be able to meaningfully participate in studies that generate RWE to support regulatory decisions. Use of automated device data capture systems can ensure complete capture of patients who can receive a device while also improving supply chain management, reducing clinician burden and eliminating documentation errors [33]. Regulators, accrediting bodies, or other groups could facilitate research site selection for these studies and provide stronger incentives for greater adoption of UDIs by tracking the extent of UDI capture, by specialty, within health systems. To address lack of UDIs in EHRs, research organizations could expand their use of medical device supply chain data. Researchers engaged in medical device research could work with their health system leaders to develop both the business case and the governance policies needed for appropriate use of their health system's medical device supply chain data to identify members of target populations and expand opportunities to participate in generating RWE.

Improving data capture beyond UDIs will also be important for expanding generation of RWE. Hospitals and health systems could explore the business case for expanding the collection of specific, high-value data elements in structured data fields in ways that would not only support research studies but also a health system's own quality improvement and value-based care initiatives. They could also identify opportunities for effectively deploying NLP. For example, organizations participating in RWE generation could assess the potential for use of NLP in different use cases to help manufacturers optimize the selection of research sites. Network collaborators could conduct additional exploratory analyses of site-level variation in the structure of clinical notes and reports to identify future opportunities for refinement. Payers and medical device manufactures could also coordinate efforts to expand the systematic capture and appropriate use of cost data and to facilitate engagement with CDM developers, which could help expand generation of evidence on the value of medical devices. Payers, in particular, prefer detailed

information about the costs and benefits of treatment options that are tailored to the populations for whom they manage health benefits, such as seniors or working-age adults, evidence that RWD can provide efficiently.

Opportunities also exist around expanding and curating existing research tools. For example, organizations representing medical device manufacturers or supporting medical device research could establish and curate libraries of computable phenotypes and end points (i.e., clinical conditions and outcomes that can be assessed through a computerized query involving one or more data elements) [34] along with information about their prior use in research, validation results and other metadata. Doing so could facilitate reuse of algorithms, improve research efficiency and increase consistency in measurement. In addition, hospitals and health systems across the United States could adopt CDMs to help scale the generation of RWE, and organizations that routinely engage in collaborative research could adopt the same CDMs to harmonize research processes. Developers of CDMs could expand their scope to include additional data elements needed for research on medical devices. Hospitals and health systems could expand the use of EHR modules that store detailed information on medical devices in structured fields to facilitate CDM expansion and capture indications for device use and end points.

In light of data gaps and challenges with long-term follow-up of end points which will be particularly important for regulators, creating more linkages between EHR and other data could support a wide range of research. Health systems could take steps to link EHR data with state or national death indexes to better capture mortality end points and expand use of privacy-preserving record linkage among EHR, registry and claims data to improve measurement of long-term study end points. Expanding linkages to death data may also require changes to state laws to facilitate access to healthcare organizations [35]. Organizations involved with RWE generation could consider linking to state all-payer claims databases and leveraging patient-enabled sharing of historical claims data through Blue Button technology. Research organizations could consider developing regional networks to ensure reliable measurement of study end points in EHR data for the majority of patients within a market [36]. Such networks could also develop common elements of research infrastructure, such as CDMs, and align administrative processes to facilitate data collection, sharing and analysis.

## Conclusion

The NESTcc Test-Cases represent one of the largest initiatives undertaken to systematically assess the potential for RWD to support regulatory decision making. Interviewees overwhelmingly viewed the Test-Cases as valuable in helping to identify lessons about the use of RWD that were generalizable across medical devices and conditions. RWD used in the Test-Cases contained rich clinical information on large and diverse patient populations that could generate information on patients efficiently over the span of 1–2 year research projects. However, the Test-Cases revealed several challenges facing manufacturers who seek to generate and use RWE for regulatory purposes—most importantly, reliably capturing UDIs and ensuring complete measurement of end points over time. Additional engagement by key stakeholders around these lessons could help shape future research and investments to address these gaps. Capturing lessons from follow-on studies along with continued engagement with FDA officials around the strengths and limitations of the evidence generated in the Test-Cases, will be critical to understanding the extent to which RWD can generate evidence that regulators view as fit-for-purpose to support regulatory decision-making.

## Author contributions

JW Timbie and TW Concannon were responsible for study conception and design. JW Timbie, AY Kim, L Baker and TW Concannon were responsible for acquisition of data. JW Timbie, AY Kim, L Baker, R Li and TW Concannon were responsible for data analysis. JW Timbie, AY Kim and TW Concannon were responsible for drafting and revision of the manuscript.

## Acknowledgments

The authors thank S Siami, S Mazzatenta, J Gasvoda, S McNabb, R Fleurence, R Rath, R Zusterzeel, K Ervin, D Kim, P Salcedo, P Kremer and R Smith from the National Evaluation System for health Technology Coordinating Center (NESTcc) for their guidance and support during the project. They also thank S Case, R Dickerson and G Gahlon from RAND for their assistance with data collection and analysis.

## Disclaimer

The contents of this paper 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 organization played no role in collection, management, analysis and interpretation of the data. Views expressed

in this publication 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 United States Government.

#### Financial disclosure

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 US FDA through grant no. 1U01FD006292-02. The authors have no other relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript apart from those disclosed.

#### Competing interests disclosure

The authors have no competing interests or relevant affiliations with any organization or entity with the subject matter or materials discussed in the manuscript. This includes employment, consultancies, honoraria, stock ownership or options, expert testimony, grants or patents received or pending, or royalties.

#### Writing disclosure

No writing assistance was utilized in the production of this manuscript.

#### Ethical conduct of research

The authors obtained institutional review board approval from RAND's Human Subjects Protection Committee for the research described.

#### Open access

This work is licensed under the Attribution-NonCommercial-NoDerivatives 4.0 Unported License. To view a copy of this license, visit <https://creativecommons.org/licenses/by-nc-nd/4.0/>

### Summary points

- The NESTcc Test-Cases, which drew on research topics spanning ten medical specialties, represent one of the largest initiatives ever undertaken to systematically assess the potential for real-world data (RWD) to support regulatory decision making for medical devices.
- Participants reported difficulty extracting information on the use of specific devices, which was an important early step in many Test-Cases. Supply chain data within some healthcare organizations helped to identify use of a specific device when a health system did not regularly capture unique device identifiers (UDIs) in its electronic health record (EHR).
- Most Test-Cases sought to investigate the use of a device in association with a specific condition or procedure. Although Test-Cases were often able to extract this information easily from structured data, diagnosis and procedure information could be unreliable because of limitations of coding systems and variation in coding practices across network collaborators.
- Test-Cases that required highly detailed information on the condition or treatment being studied were often required to search unstructured EHR data, such as medical notes or reports. A few Test-Cases used cardiac, orthopedic, cancer and other procedure- and condition-based registries to provide a curated source of rich, clinical data on patients receiving medical devices.
- Patient attrition can limit the ability of EHR data to capture information on patients over long periods. This was particularly pronounced at academic health systems that had large referral populations but might not provide ongoing care for patients referred to their system.
- Many Test-Cases highlighted the importance of engaging physician specialists and data scientists in their Test-Cases to leverage their complementary expertise.
- Opportunities to strengthen collection of RWD include expanding the capture of unique device identifiers within health systems, developing libraries of computable phenotypes and end points and creating more linkages between EHR and claims data.
- Coordinated efforts by medical device stakeholders to address these opportunities could enhance the ability of RWD to generate evidence at larger scales, enhance longitudinal measurement of end points and improve the efficiency of research.
- Although the Test-Cases were diverse in their focus, studies focusing on other types of medical devices, medical conditions and RWD sources might identify different lessons and opportunities to optimize use of RWD.

## References

Papers of special note have been highlighted as: • of interest

1. U.S. Food & Drug Administration. Real-World Evidence (2023). <https://www.fda.gov/science-research/science-and-research-special-topics/real-world-evidence>
2. European Medicines Agency. Real-world evidence provided by EMA: support for regulatory decision-making (2024). [https://www.ema.europa.eu/en/documents/other/guidance-real-world-evidence-provided-ema-support-regulatory-decision-making\\_en.pdf](https://www.ema.europa.eu/en/documents/other/guidance-real-world-evidence-provided-ema-support-regulatory-decision-making_en.pdf)
3. RM Califf. Realizing the Promise of Real-World Evidence (2023). <https://www.fda.gov/news-events/fda-voices/realizing-promise-real-world-evidence>
4. Johnson & Johnson. What is real-world evidence and how does it help patients? A Q&A with Josh Gagne, Vice President and Global Head of Epidemiology, Johnson & Johnson (2022). <https://www.jnj.com/office-of-chief-medical-officer/what-is-real-world-evidence-and-how-does-it-help-patients>
5. Center for Devices and Radiological Health, Administration CfBEaRUSFD. Use of Real-World Evidence to Support Regulatory Decision-Making for Medical Devices: Guidance for Industry and Food and Drug Administration Staff (2017). <https://www.fda.gov/media/99447/download>
6. Bartlett VL, Dhruva SS, Shah ND, Ryan P, Ross JS. Feasibility of using real-world data to replicate clinical trial evidence. *JAMA Netw. Open.* 2(10), e1912869 (2019).
7. Hampson G, Towse A, Dreitlein WB, Henshall C, Pearson SD. Real-world evidence for coverage decisions: opportunities and challenges. *J. Comp. Eff. Res.* 7(12), 1133–1143 (2018).
8. O'Neill T, Miksad R, Miller D *et al.* ISPOR, the FDA, and the evolving regulatory science of medical device products. *Value Health* 22(7), 754–761 (2019).
9. Timbie JW, Kim AY, Concannon TW. Use of real-world evidence for regulatory approval and coverage of medical devices: a landscape assessment. *Value Health* 24(12), 1792–1798 (2021).
- **This qualitative analysis summarizes stakeholder views on the use of RWD including industry and FDA perspectives.**
10. Duke-Margolis Center for Health Policy. A Framework for Regulatory Use of Real-World Evidence (2017). [https://healthpolicy.duke.edu/sites/default/files/2020-08/rwe\\_white\\_paper\\_2017.09.06.pdf](https://healthpolicy.duke.edu/sites/default/files/2020-08/rwe_white_paper_2017.09.06.pdf)
11. Medical Device Innovation Consortium. Medical Device Innovation Consortium (MDIC) Awarded \$3 million for National Evaluation System for health Technology Coordinating Center (2016). <https://mdic.org/wp-content/uploads/2016/09/Final-Press-Release-MDIC-awarded-NEST-Coordinating-Center-20160913.pdf>
12. U.S. Food & Drug Administration. Medical Device User Fee Amendments (MDUFA) (2021). <https://www.fda.gov/industry/fda-user-fee-programs/medical-device-user-fee-amendments-mdufa>
13. National Evaluation System for health Technology Coordinating Center. An Invitation to Submit Concepts for Real-World Evidence Test-Cases to the NEST Coordinating Center (NESTcc) (2017). <https://nestcc.org/wp-content/uploads/2017/12/12.4.17-NESTcc-Call-for-Concepts.pdf>
14. National Evaluation System for health Technology Coordinating Center. Research Network (Undated). <https://nestcc.org/research-network/>
15. National Patient-Centered Clinical Research Network. Data (Undated). <https://pcornet.org/data/>
16. U.S. Food & Drug Administration. Framework for FDA'S Real-World Evidence Program (2018). <https://www.fda.gov/media/120060/download>
17. Center for Devices and Radiological Health; U.S. Food & Drug Administration. Examples of Real-World Evidence (RWE) Used in Medical Device Regulatory Decisions: selected examples with file summaries, details on real-world data source, populations, and descriptions of use (Undated). <https://www.fda.gov/media/146258/download>
- **The examples of RWD contained in this review cover a wide range of data sources and regulatory use cases.**
18. Food and Drug Administration Center for Drug Evaluation and Research, Center for Biologics Evaluation and Research, Oncology Center of Excellence. Real-World Data: Assessing Electronic Health Records and Medical Claims Data To Support Regulatory DecisionMaking for Drug and Biological Products: Guidance for Industry (2021). <https://www.fda.gov/media/152503/download>
19. U.S. Food & Drug Administration Center for Devices and Radiological Health, Center for Biologics Evaluation and Research. Use of Real-World Evidence to Support Regulatory Decision-Making for Medical Devices: Draft Guidance for Industry and Food and Drug Administration Staff (2023). <https://www.fda.gov/regulatory-information/search-fda-guidance-documents/draft-use-real-world-evidence-support-regulatory-decision-making-medical-devices>
20. Al-Faruque F. FDA's Califf: expect to see more RWE-based regulatory decisions (2023). <https://www.raps.org/News-and-Articles/News-Articles/2023/9/FDA%E2%80%99s-Califf-Expect-to-see-more-RWE-based-regulatory>
21. Chan KS, Fowles JB, Weiner JP. Review: electronic health records and the reliability and validity of quality measures: a review of the literature. *Med. Care Res. Rev.* 67(5), 503–527 (2010).
22. Richesson RL, Marsolo KS, Douthit BJ *et al.* Enhancing the use of EHR systems for pragmatic embedded research: lessons from the NIH Health Care Systems Research Collaboratory. *J. Am. Med. Inform. Assn* 28(12), 2626–2640 (2021).

23. Casey JA, Schwartz BS, Stewart WF, Adler NE. Using electronic health records for population health research: a review of methods and applications. *Annu. Rev. Publ. Health* 37, 61–81 (2016).
24. The Medical Device Epidemiology Network. Coordinated Registry Networks (Undated). <https://www.mdepinet.net/coordinated-registry-networks>
25. Tsougranis G, Eldrup-Jorgensen J, Bertges D *et al.* The Vascular Implant Surveillance and Interventional Outcomes (VISION) coordinated registry network: an effort to advance evidence evaluation for vascular devices. *J. Vasc. Surg.* 72(6), 2153–2160 (2020).
26. Marinac-Dabic D, Aryal S, Azarbaijani Y *et al.* The Women’s Health Technologies Coordinated Registry Network (WHT-CRN) (2020). <https://aspe.hhs.gov/sites/default/files/private/pdf/259016/wht-crn-final-reoprt-to-aspe.pdf>
27. 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.* 1(1), e000003 (2019).
28. Poulou BK, Avila-Tang E, Schwartzman H *et al.* Determining the value of the abdominal core health quality collaborative to support regulatory decisions. *Hernia* doi:10.1007/s10029-024-02990-5 (2024) (Epub ahead of print).
29. McDermott O, Kearney B. The value of using real-world evidence as a source of clinical evidence in the European medical device regulations: a mixed methods study. *Expert Rev. Med. Devices* 21(1–2), 149–163 (2024).
30. Dhruva SS, Ridgeway JL, Ross JS, Drozda JP Jr, Wilson NA. Exploring unique device identifier implementation and use for real-world evidence: a mixed-methods study with NESTcc health system network collaborators. *BMJ Surg. Interv. Health Technol.* 5(1), e000167 (2023).
31. Miksad RA, Abernethy AP. Harnessing the power of real-world evidence (RWE): a checklist to ensure regulatory-grade data quality. *Clin. Pharmacol. Ther.* 103(2), 202–205 (2018).
32. Fleurence RL, Kent S, Adamson B *et al.* Assessing real-world data from electronic health records for health technology assessment: the SUITABILITY Checklist: a Good Practices Report of an ISPOR Task Force. *Value Health* 27(6), 692–701 (2024).
33. Tcheng JE, Nguyen MV, Brann HW *et al.* The medical device unique device identifier as the single source of truth in healthcare enterprises – roadmap for implementation of the clinically integrated supply chain. *Med. Devices Evid. Res.* 14, 459–14467 (2021).
34. Richesson R, Wiley LK, Gold S, Rasmussen L. Electronic Health Records-Based Phenotyping (2020). <https://rethinkingclinicaltrials.org/chapters/conduct/electronic-health-records-based-phenotyping/electronic-health-records-based-phenotyping-introduction/>
35. Wenger NS, Sanz Vidorreta FJ, Dudley MT, Walling AM, Hogarth M. Consequences of a health system not knowing which patients are deceased. *JAMA Intern. Med.* 184(2), 213–214 (2024).
36. Shenkman E, Hurt M, Hogan W *et al.* OneFlorida Clinical Research Consortium: linking a clinical and translational science institute with a community-based distributive medical education model. *Acad. Med.* 93(3), 451–455 (2018).