

Evaluating AI-Enabled Clinical Decision and Diagnostic Support Tools Using Real-World Data



Duke | **MARGOLIS CENTER**
for Health Policy

March 11, 2022

Authors

Robbert Zusterzeel, Senior Director, IQVIA*

Benjamin A. Goldstein, Associate Professor of Biostatistics and Bioinformatics, Duke University

Barbara J. Evans, Professor of Law and Professor of Engineering, University of Florida

Thomas Rodes, Policy Analyst, Duke-Margolis Center for Health Policy

Kerra Mercon, Senior Research Assistant, Duke-Margolis Center for Health Policy (former)

Christina Silcox, Digital Health Policy Fellow, Duke-Margolis Center for Health Policy

* Work performed as an independent contractor for Duke-Margolis

Acknowledgments

The authors would like to thank many people for contributing their time and expertise to inform and improve this white paper. The paper would not have been possible without the input of expert perspectives from a July 2020 Duke-Margolis private virtual workshop. Additionally, the Duke team held many informational meetings and calls with various stakeholders across the artificial intelligence and health care ecosystem over the course of this project. We are extremely grateful for their time and thoughtful feedback on working drafts.

Funding

This project was funded by the Gordon and Betty Moore Foundation.

Disclosures

Any opinions expressed in this paper are solely of those of the authors and do not represent the views or policies of other organizations external to Duke.

Executive Summary

Artificial Intelligence (AI) holds great potential for improving health and health care in the United States and globally through knowledge discovery, detection and monitoring of diseases, development of novel digital therapeutics, and augmentation or automation of clinical decision-making for diagnosis and treatment of patients. Some of these AI-enabled software tools will be medical devices (“Software as a Medical Device” or SaMD) that are regulated by the U.S. Food and Drug Administration (FDA), while others will fall outside of FDA’s authority. For all AI-enabled SaMDs, it will be critical to continuously evaluate the tool after deployment and over time to ensure that it is performing within the range expected. This may be a particular challenge for clinical decision support (CDS) tools developed with machine-learning (AI/ML) due to non-standardized electronic health records systems and ever-changing workflows.

Using high-quality real-world data (RWD) to generate real-world evidence (RWE) of the clinical performance of SaMDs may allow evaluations to be done more efficiently and can create a broader sense of how well the software tool works in multiple subgroups of interest (individuals

of different ages, races, geographical areas, associated co-morbidities, etc.). This report explores what type of data would be needed to perform these types of evaluations, if those elements exist in common RWD sources, and current challenges in collecting and using such data. While many of the key takeaways apply more broadly, this report focuses on challenges in the context of post-market performance evaluations of AI/ML-enabled CDS tools. These insights and recommendations are drawn from a two-day virtual private workshop held in July 2020, a literature review, and informational calls.

The report provides recommendations for consideration in the continued development of a future regulatory model for software-based medical devices. We separate these recommendations into three areas: performance measures, data access and privacy, and data sharing and security. Performance measures correlate with the specific benefit-risk ratios of CDS tools related to changes in their accuracy as the data environments change, new workflows or standards of care are introduced, or patient populations shift, and can further vary with potential inappropriate usage. Specific data elements that should be monitored include algorithm inputs and outputs, algorithm use characteristics, and patient outcomes. These elements are ideally standardized within and across data sources with consideration of performance bias in diversity of patient populations and generalizability across health systems. The accessibility of data to conduct performance evaluations depends on privacy protection regulations, including HIPAA, the Common Rule, and state and local laws, which can apply differently depending on the entity using the data. Some of these regulations may even require updates to better enable future real-time evaluation of CDS tools. Finally, when the correct data are available and can be accessed, the physical exchange should ideally occur on secure platforms governed by data use agreements.

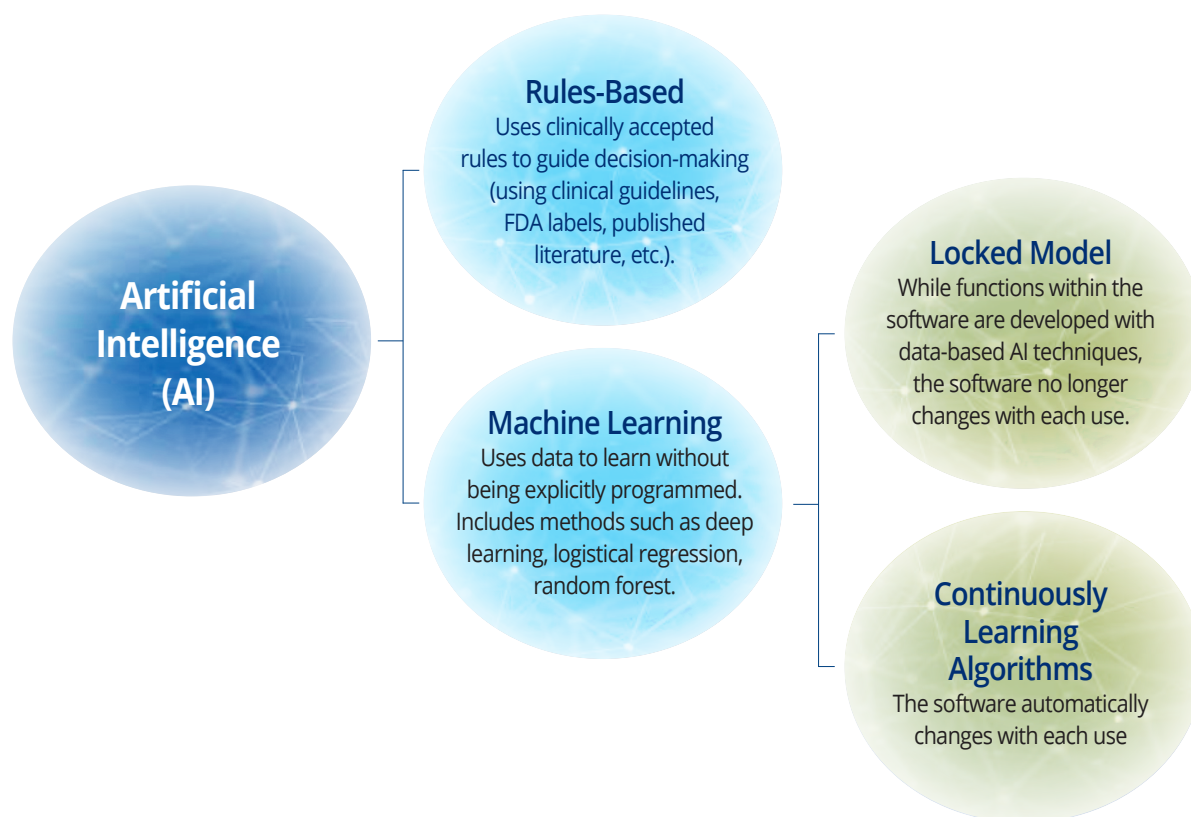
This report explores what type of data would be needed to perform these types of evaluations, if those elements exist in common RWD sources, and current challenges in collecting and using such data.

Introduction

The U.S. Food and Drug Administration (FDA) is interested in the use of RWD for the development and continued evaluation of artificial intelligence (AI)-enabled software products used in healthcare settings. Software tools built with AI are increasingly being recognized for their potential to improve healthcare. The definitions of AI and its subparts vary widely with a lack of consensus but for this paper are mostly based on previously published definitions in prior Duke-Margolis white papers. AI can be divided into two categories depending on how it is programmed.

- **Rules-based AI** uses previously validated information (e.g., clinical guidelines, risk calculators, published studies) to set up a series of clinically accepted weights or decision steps that lead to a prediction, diagnosis, or recommendation.
- **Data-based AI**, often referred to as **machine learning (AI/ML)**, is trained using sets of labeled input data (called “training data”) and uses programmed processes to derive relationships between the inputs and the so-called “labels”.^a The derived relationships are then used to predict how new input data would be labeled, which becomes a prediction, diagnosis, or recommendation for the clinician in an AI/ML-enabled CDS tool (operational data). Once a tool that was developed through ML is in use, the model can continue to learn or be “fine-tuned” by incorporating new data (learning data). These changes can be made automatically each time a new labeled example is received (**continuously learning algorithms**) or the labeled examples can be stored for periodic updating of the tool (**locked models**) (**Figures 1 and 2**).

FIGURE 1 | BREAKDOWN OF RULES-BASED AI AND MACHINE LEARNING



^a This describes a specific type of machine-learning called “supervised machine learning” which is how most current AI/ML-enabled CDS tools are trained.

The Role of AI/ML-Enabled CDS in Healthcare

AI/ML-enabled CDS is used to aid decision-making around triage, diagnosis, and treatment of individual patients by incorporating AI/ML-enabled CDS tools into clinical workflows. A rules-based CDS tool could help the physician determine whether an individual patient needs an intervention based on medical professional society guidelines programmed into the tool. For example, to inform patient-specific dosage decisions, the tool could be fed a patient's International Normalized Ratio (INR – a measure that outlines how quickly the blood coagulates) and track whether it drops or rises above a certain level. Using that input, the algorithm would make a recommendation for the appropriate dosage of anticoagulants to treat the patient.

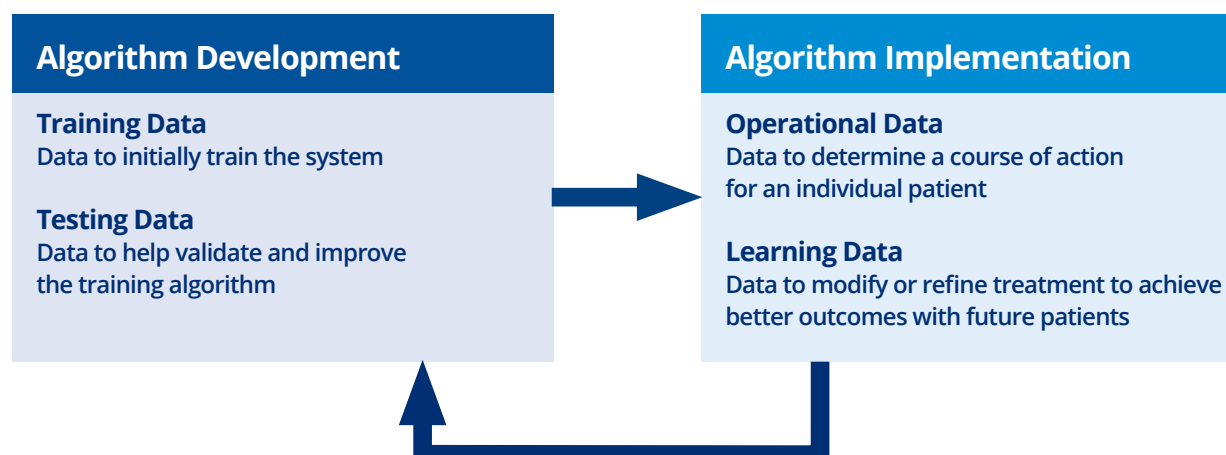
AI/ML-enabled CDS is used to aid decision-making around triage, diagnosis, and treatment of individual patients by incorporating AI/ML-enabled CDS tools into clinical workflows.

ML-enabled CDS tools, however, can map many clinical features to a particular outcome by generating a probabilistic assessment of the likelihood of the event or diagnosis. For example, a tool may analyze the pixel values of a digital x-ray and be able to predict if and where the foot is fractured. The example is relatively low risk given

that it is overseen or monitored by a trained professional and a timely treatment action or intervention is important but not critical to prevent or mitigate long-term irreversible consequences. However, other examples of AI/ML products, such as adaptive implantable defibrillators that detect specific patterns in heart rhythms to determine whether to automatically administer a life-saving electrical shock, are higher-risk. FDA's Center for Devices and Radiological Health (CDRH) recently published a list of legally marketed AI-enabled medical devices that shows the range of clinical risk levels.¹

Software tools used in healthcare, like all medical products, should be safe and effective for patients and reduce burden on health care workers whenever possible. Some but not all software tools used in healthcare are considered "software as a medical device" (SaMD). CDRH regulates products classified as SaMD using their standard risk classification system (Class I [low risk], II [moderate], and III [high risk]). The 21st Century Cures Act, passed in late 2016, clarified FDA's authority over certain clinical software tools and CDRH subsequently released draft Guidance on CDS tools in 2019 that outlines when a CDS tool is considered a medical device, with a focus on whether the user can "independently review the basis" of the recommendation.² This distinction is important for AI/ML-enabled tools, which in many (but not all) instances are unable to provide a human-comprehensible "reason" for the prediction.

FIGURE 2 | DATA USED IN BUILDING, TESTING, AND USING MACHINE LEARNING TOOLS



The Use of RWD in Regulatory Surveillance and Decision-Making

FDA defines RWD as “data relating to patient health status and/or the delivery of health care routinely collected from a variety of sources.” RWD can exist in many forms and includes data derived from electronic health records (EHRs), insurance claims and billing data, data from product and disease registries, and patient-generated health data from mobile devices or other settings.³

FDA has considered RWD to support regulatory decisions and issued a Guidance document in 2017 on the use of these types of data to develop Real-World Evidence (RWE) to support regulatory activities.³ Under the appropriate conditions, analysis of high-quality RWD can produce RWE regarding benefits and risks of medical devices for pre- and post-market regulatory purposes (i.e., before and after a device has received marketing authorization and is used in clinical settings, respectively). RWD is more often used in the post-market period to evaluate long-term safety and effectiveness in broader patient populations. In March 2021, the FDA published a report with 90 examples illustrating the use of RWE in medical device regulatory decisions.⁴ In addition, FDA’s Center for Drug Evaluation and Research (CDER) and Center for Biologics Evaluation and Research (CBER) have recently issued common draft Guidance on the use of RWD in drug and biologics regulatory applications⁵ for different data sources,^{6,7} including the standardization of RWD elements.⁸

FDA has expressed interest in RWD and RWE for the development and evaluation of SaMD and AI/ML-enabled SaMD partly because of the rapid product lifecycle and frequent updates associated with the development of such tools.

FDA has expressed interest in RWD and RWE for the development and evaluation of SaMD and AI/ML-enabled SaMD partly because of the rapid product lifecycle and frequent updates associated with the development of such tools. In 2017, FDA introduced a software pre-certification pilot program.⁹ The overall goal of the pre-certification pilot was to provide more streamlined and efficient regulatory oversight of software-based

medical devices developed by manufacturers who have demonstrated a robust culture of quality and organizational excellence. The pilot program’s scope was SaMD generally and included AI/ML SaMD. The most recent draft of the working model was released in January 2019.¹⁰ This model continued to envision strong post-market surveillance, with an emphasis on Real-World Performance Analytics (RWPA) that included continued evaluations on accuracy, safety, and effectiveness, among other measures.

More recently, a 2021 FDA Action Plan specifically on AI/ML-based software stated that “gathering performance data on the real-world use of the SaMD may allow manufacturers to understand how their products are being used, identify opportunities for improvements, and respond proactively to safety or usability concerns.”¹¹ Therefore, RWD could potentially be used not only for safety surveillance, but also for regularly updating software tools. For example, FDA has been exploring the use of “change control plans” submitted by developers when applying for FDA authorization that would allow updates to software tools as described in those pre-approved plans without additional regulatory approval. While official guidance is being developed,¹² the Medical Device Innovation Consortium (MDIC) has been working on a template for change control plans.¹³ These updates could either happen iteratively through a series of locked models or by using data in continuously learning models^b in the future.¹¹ An example of a pre-specified change control plan for locked algorithm updates was included in a recently cleared FDA application for a tool to assist medical professionals in the acquisition of cardiac ultrasound images.¹⁴

Finally, in October 2021, the FDA, along with other global regulatory agencies, released a guiding principles document on good machine learning practices for medical device development that is intended to be expanded upon in the future.¹⁵ Representativeness of the intended patient population, clinically useful outputs, and continued monitoring for performance are some of the key aspects of these guiding principles as well.

^b This paper does not go into detail about continuous learning models as, to our knowledge, none have been authorized for use by FDA. However, it is very possible that these tools would use RWD as their continuing training data and may use RWD for automated performance checks.

Post-Market Evaluation of AI/ML-Enabled CDS Tools

The evaluation of a CDS tool consists of four distinct phases, each with its own goals and processes: 1) retrospective model evaluation;^c 2) silent period evaluation;^d 3) prospective evaluation;^e and 4) ongoing monitoring. It is generally accepted that a cleared or approved AI/ML-enabled SaMD product performs well enough to use given its pre-market performance characteristics. However, as with any other medical device, it is important to monitor its use in the post-market to substantiate performance in broader real-world settings. In this report the focus is solely on the ongoing monitoring of a CDS tool during the post-market timeframe (phase 4).

Key Considerations for Evaluation and Monitoring

Generally, CDS tools should be monitored for two potential categories of issues in the post-market: *changes in performance of the CDS tool* and *inappropriate or variable usage of the CDS tool*.

Changes in Performance of the CDS Tool

The performance of the CDS tool itself may change over time. This can occur for many reasons, including changes in the clinical workflow, how data are entered into databases, and patient populations or standards of care. While the goal of silent evaluation is to verify the validity of the data flow, back-end EHR systems are dynamic—their underlying architecture undergoes frequent changes and updates known as “refreshes.” These refreshes may “break” software tools that rely on data streams from outdated data structures. It is important to know which data elements feed a CDS tool and what may happen to the tool if certain data are no longer accessible.

Clinical workflows and clinical standards of care also evolve over time. Most AI/ML-based CDS tools are not learning *causal* associations. Instead, the tools are learning risk factors that affect the probability of an outcome. For example, a tool may “learn” that a patient getting

a laboratory test for Lyme disease means they are more likely to have Lyme disease, since clinicians only order the test for patients who may have been exposed or display symptoms consistent with the disease. However, if that laboratory test becomes standard of care – i.e., the test is ordered for all patients – then the test is no longer informative for this reason, and the previously learned association is invalid. When clinical workflows change, it is important to understand how that relates to the performance of the CDS tool itself.

Finally, changes in the underlying patient population may affect CDS tool performance, and the characteristics of specific sites where the tool is applied are therefore important. The characteristics of patient populations can vary widely between hospitals. The patient population is likely to differ in some respects in new settings where the CDS tool is implemented, as opposed to the setting(s) where it was developed and tested. If these differences are substantial, the performance of the tool may change.

The topic of site-specific performance of AI/ML software tools was raised multiple times at a February 2020 FDA public meeting on AI in radiological imaging due to concerns that performance may differ across care locations or that AI/ML-enabled software tools may need to be customized to individual care systems’ IT infrastructure.¹⁶ It was notable that this was emphasized in a meeting addressing radiology software tools, as imaging data are generally more standardized and interoperable than many other types of health data.

The importance is further highlighted by two recent examples: a sepsis prediction model integrated into the EHR software system from a large vendor¹⁷ and a pneumonia screening algorithm for x-rays.¹⁸ The sepsis model was developed at three sites and displayed good performance. However, an independent analysis at a large academic hospital that was not part of the model development found that its performance was

^c During retrospective model evaluation the proposed model is tested on retrospective data. The retrospective data are distinct from data that were used to develop the algorithm. The goal is to ensure that the tool has the expected operating characteristics with respect to discrimination, calibration and decision rule performance.

^d During silent period evaluation the tool is implemented into the real-time environment but is not used by clinicians. The goal is to ensure that the tool performs prospectively in the way it performed retrospectively. Since data flows through an EHR system differently than how it is stored, it is important to test that predictions or scores are calculated in the expected way.

^e During prospective evaluation the tool is placed into production and used. The goal is to ensure it is properly used by the target user (e.g., nurses, doctors, care managers) and that relevant clinical decisions are made in response to the tool.

largely inadequate, mainly due to a low sepsis detection rate. For the pneumonia screening algorithm, training was performed on more than 150,000 x-rays from three hospital systems. The study demonstrated that real-world performance on chest x-rays from other hospitals was significantly lower than on the x-rays used to validate the model from the original hospital.

Inappropriate or Variable Usage of the CDS Tool

CDS tools require human interactions and decision-making. Every CDS tool is different in the way it presents information to, or interacts with, the user. Some tools will display a clear decision rule (risk vs. no-risk) with an associated action the user should take, while others display more qualitative decision rules (high/medium/low risk) without a specific associated action. Others will simply present a predicted risk and leave it up to the user to decide how to use that information. The less directive a CDS tool is, the more influential clinical judgement becomes. As such, it is important to monitor what clinical actions are taken in response to the tool's outputs to determine overall clinical effectiveness. In addition, usability of the CDS tool should be tested pre-market (e.g., through human factors review) and included in performance evaluation in the post-market period.

Relatedly, usage of CDS tools often varies. It is important to note that FDA does not regulate how medicine is practiced, and healthcare professionals are allowed to use a lawfully marketed medical device “off-label” in clinical practice. However, FDA does have interest in making sure devices are clearly labeled to avoid user confusion about the manufacturer's intended use and ensuring that manufacturers adequately disclose risks of off-label uses. In addition, FDA monitors the safety of common off-label uses and can respond if new significant risks are detected.

Additional Considerations

Within the monitoring framework discussed above there are several additional considerations that can lead to variability in performance, create potential undesired effects, or drive the (inappropriate) use of CDS tools.

Health Equity and Performance Disparities

Since many CDS tool algorithms are trained on retrospectively collected RWD, it is possible for them to “learn” systemic and structural biases that exist within

Since many CDS tool algorithms are trained on retrospectively collected RWD, it is possible for them to “learn” systemic and structural biases that exist within the standard provision of care based on factors such as age, sex, race, socio-economic, and insurance status.

the standard provision of care based on factors such as age, sex, race, socio-economic, and insurance status. In these scenarios, it is possible for the use of CDS tools – particularly CDS tools that direct additional care or services – to perpetuate those biases. As such, it is important to assess actual model performance within different population subgroups based on these factors or on characteristics that are relevant to the intended use of the specific tool. For example, a CDS tool that is trained within one demographic subgroup may not perform as well in other demographic subgroups. When such problems occur, it may be necessary to recalibrate or retrain the model with more representative source data. The recent guiding principles by regulators¹⁵ and the CDRH 2022-2025 Strategic Priorities¹⁹ have emphasized these points as well.

Confounding by Medical Intervention

One of the challenges of ongoing monitoring is that patients are receiving interventions in response to the tool's output, and the chosen interventions may vary from one setting to the next. This can largely depend on variations in adoption of clinical guidelines, system protocols and procedures, and institutional policy requirements. If a tool is effective, then high-risk patients will receive interventions that (hopefully) prevent adverse health outcomes. When this occurs, it could appear, from the naïve evaluator's perspective, that the patient was erroneously classified as being at high risk. This has been referred to as *confounding by medical intervention*.²⁰ While clear solutions do not currently exist, the issue highlights the importance of tracking not only what a patient's assessed risk was, but also what actions were taken in response to that risk, in order to properly evaluate the CDS tool.

Recommendations

Given the analyses discussed above, the following data elements will generally be required to evaluate software tool performance:

- The model score, prediction, or recommendation (model output).
- The observed outcome or an alternative comparator (comparison to output).
- Model inputs (operational data). It may also be important to assess whether the inputs for the algorithm were “correct” (as opposed to whether the inner workings of the algorithm are sound) in the sense that they were concordant with the patient’s actual status. For example, if having diabetes is an important input variable, data collection should include checking whether the patient’s records and the data input to the algorithm contain the same measured glucose values. If the data sources differ, the tool will produce a different output than a human provider, even if it is working as intended.
- Demographic subgroup analysis variables.

If the effect on patient outcomes is important for regulatory use, then additional data points may need to be captured, regardless of whether providers follow the algorithm’s recommendations. These would include:







- Did the healthcare professional act on the algorithm’s output/recommendations? How so?
- What other actions (external to algorithm use) were performed that might have influenced outcomes, potentially confounding the relationship between algorithm performance and patient health status? (e.g., early action or additional treatments)?

The required data elements may not always be available depending on the use-case of the CDS tool. In particular, collecting information on relevant external factors other than the CDS tool is quite complicated, as some interventions may not be captured. In addition, even if relevant external factors were captured, it is also important to have temporal information regarding such factors to accurately determine how the CDS tool, or other actions pertaining to a patient, influenced the care pathway or a patient’s health outcomes. In other words, can we capture the real-world context around whether those external factors or interventions occur after the CDS tool’s recommendation? Is it possible they were influenced by that recommendation? Answering these questions requires highly detailed data collection and can be difficult even with complete data. Regardless, public reporting of post-market performance should provide sufficient information to ensure transparency, build public trust, and support ongoing regulatory oversight.

Table 1 summarizes the considerations for the information and data that should ideally be captured in the post-market setting related to CDS tool performance assessment, where those data may be located, and how they may be stored. Collection of additional data should be considered based on the specifics of the CDS tool and its intended use and risk determination.

Finally, there are other features beyond accuracy and improvements in patient outcomes that contribute to the overall “performance” of an AI-enabled software tool (such as ease of use or compatibility with on-premise IT systems), but those are generally beyond the scope for this paper. An initiative that dives into some of these issues is the Accelerated Digital Clinical Ecosystem (ADVICE).²¹

TABLE 1 | Real-World Data Generally Needed for Post-Market Evaluation of AI/ML-Enabled CDS Tools

Category of Information	Data Elements*	Potential Location	Examples	Comments
Algorithm Inputs/Subgroup Analysis	Demographics		Age, sex, race	<ul style="list-style-type: none"> Race is an important data element to allow subgroup performance analysis but should be considered carefully before used as an input.
	Socio-economic factors		Insurance status	<ul style="list-style-type: none"> Insurance status is a commonly used proxy, more direct measures are not generally available in common sources of RWD
	Medical history		Diagnoses, procedures, medications, vitals, labs, symptoms, imaging, physiological monitors, data internal to a medical device	<ul style="list-style-type: none"> Symptoms (e.g., subjective physical observations [sleep, appetite etc.]) are commonly unstructured (free text/notes) Data internal to a medical device (e.g., waveforms from an implantable defibrillator) are not commonly found in RWD sources (but could potentially be extracted from the device) Data definitions may not be standardized within or across health systems
Algorithm Outputs	Based on intended use		Risk score, diagnosis, suggested action	<ul style="list-style-type: none"> Not always stored in common RWD sources; developers could build in this functionality
Comparator (observed outcome or another comparator)	Based on intended use		Vitals/physiological monitoring Outcomes such as diagnosis, (re-)hospitalization, death	<ul style="list-style-type: none"> Real-world diagnosis will be available, but may not be a reliable comparator if subjective or if influenced by the algorithmic prediction RWD comparator outputs may not always be available; consensus expert opinion utilizing health records can be used in cases where observed outcomes are not available, or if the outcomes were potentially adverted by the prediction
Patient Interventions	Based on intended use		Medication administration, procedures, other patient actions	<ul style="list-style-type: none"> Billable procedures and treatments will be captured within RWD sources, but may be difficult to determine causality Some interventions may not be captured in any RWD sources (e.g., change in patient position to prevent pressure ulcers)

KEY

Electronic Health Record (EHR) or ancillary EHR data systems



Claims



Registry



Internal device data

*Specific data elements depend on the intended use of the CDS tool and can vary. EHR electronic health record; RWD real-world data

Existence and Quality of RWD for Post-Market Evaluation of CDS Tools

Current RWD Collection Methods

The scope of data elements discussed above may or may not be available in hospitals or care settings and may also depend on that health system's EHR systems and workflows. Some of these RWD elements would be in databases that co-exist with hospital EHR systems rather than integrated into the main EHR record itself. If the particular healthcare system where the CDS tool is deployed is not integrated with patients' primary and specialist care, medical history data may be incomplete. This would create potential issues if the CDS tool, for example, relies on inputs related to medications that are prescribed or filled in one setting but such data are not available in the other, or where patient data are only available during a certain timeframe of a patient's life and important longitudinal information is missing. Other data elements may not traditionally live in EHRs yet exist in claims and billing data, data from product and disease registries, patient-generated health data, and other potential ancillary data sources.

CDS tools can operate in a wide variety of care settings, but the input and output data needed for post-market evaluation can likely be found within the same data systems where CDS tools are already operating. Collecting these data would be relatively straightforward. However, information such as whether and how clinicians use the algorithm's output, or patients' long-term health outcomes, may be found in different data systems, or may not be recorded and digitized at all. For example, administering a medication will most likely be captured, but changing the position of a patient at risk of pressure ulcers may not be documented.

Challenges in RWD Collection

Availability of Appropriate and Representative Source Data

Manufacturers attempting to use RWD to evaluate post-market performance, particularly if one of the endpoints involves longer-term patient outcomes, may need to link across multiple different databases covering a fuller spectrum of an individual patient's health journey. This linkage process generally requires unique patient identifiers, meaning that de-identified data, which is often

easier to access, may be less useful for this purpose. And while linking across data sources can add additional information about patients, it is generally time-consuming and expensive. It can also introduce bias if there are differences in the data quality between these sources.

Manufacturers may find that establishing appropriate data collection mechanisms at the source – that is, at sites where the algorithm is used – is more efficient and accurate until ongoing interoperability efforts bear more fruit. Care will need to be taken to ensure that sites selected for post-market evaluation are diverse geographically, demographically, and socio-economically to effectively evaluate generalizability of products.

However, there are promising approaches in the medical imaging space where such interoperability efforts are starting to pay off, including the “assess-AI” platform of the American College of Radiology's (ACR) Data Science Institute. This platform is a clinical data registry capturing algorithm effectiveness at the point of care as well as metadata related to specific imaging exams (including device used and relevant patient information).²² Data for the registry are collected through an existing software tool that is widely available across hospital systems. The creation of such a centralized registry operated by a respected non-profit entity addresses many of the challenges discussed in this report. It allows manufacturers to meet their post-market surveillance requirements and clinical sites to obtain regular reporting on the algorithm's performance in patients. However, as discussed earlier, imaging data are generally more standardized and interoperable relative to other types of health data. A centralized registry may therefore not necessarily be the right or most efficient choice in all situations, but ACR's effort is one approach that might be adapted to further facilitate real-world performance assessments.

As AI/ML algorithms are deployed in the real world, it is crucial to assess whether they meet their labeled standards across different patient demographics, morbidities, and health care systems. FDA's AI/ML Action Plan highlights the need for improved methods to evaluate and address algorithmic bias and to promote algorithm robustness.¹¹ As discussed earlier, it is possible for CDS tools to “learn” systemic and structural biases that exist within the

Socio-economic status and social determinants of health, which are major drivers of health care delivery and outcomes, are generally not well captured in RWD sources.

standard provision of care based on demographic factors such as age, sex, race, and socio-economic status. In addition, socio-economic status and social determinants of health, which are major drivers of health care delivery and outcomes, are generally not well captured in RWD sources. This is also true for patient-reported outcomes (PROs) and genomic data. Health system factors such as workflows, available medical equipment, and common insurance requirements can also introduce bias into the software tools. These biases can affect the resulting algorithm, which may cause the algorithm to perpetuate and potentially magnify health disparities or make mistakes as it is scaled across different health systems. It will be tempting for companies to collect RWD in large academic healthcare centers or integrated care systems where staff are more experienced in data analysis and data are generally of higher quality and more standardized, but some of the most convenient, most usable sources of data will have systemic biases that must be accounted for. It is critical for FDA and manufacturers to carefully consider performance in diverse sets of both patients and health systems where these tools are being used.

Data Quality

Even if all the required data for adequate post-market performance evaluation are available and properly retrievable across all their sources, there may still be issues regarding the quality of that data in the absence of generally accepted quality standards. CDRH's 2017 Guidance on RWE for regulatory decision-making highlights the importance of relevance and reliability of RWD in medical devices.³ Relevant RWD captures sufficient detail on device use, exposures, and outcomes in the appropriate populations. Relevance also encompasses RWD's usability – data elements must be analyzable using proper statistical methods in order to guide sound scientific and clinical judgment. Reliability refers to the assessment of the data collection process itself, including whether the people

and processes in place during that collection and analysis provide adequate assurance that errors are minimized, and that data quality and integrity are sufficient. This includes pre-specification of data elements and their definitions and an understanding of the specific sources of the data as well as completeness and consistency across sites and over time.³ Common data models, such as OMOP, i2b2, Sentinel and PCORnet can help with such pre-specification, but any of these will require enhancements to accommodate relevant medical device-related fields.

There are many factors that influence data quality within specific use-cases of CDS tool performance evaluation that need to be considered. Multiple groups within the medical device space have built on CDRH's RWE Guidance to address these further, including the Duke-Margolis Center for Health Policy^{23,24} and the National Evaluation System for health Technology Coordinating Center (NESTcc).²⁵ In the drug and biologics space, FDA's common CDER and CBER Guidance specifically addresses the standardization of RWD elements for such purposes.⁸

Recent findings from the pilot testing of the FDA's software pre-certification program, published in September 2020, further exemplify the difficulties of collecting appropriate data for post-market evaluations of AI/ML-enabled CDS tools.²⁶ As discussed earlier, the program calls for a pre-specified RWPA plan that would enable appropriate data collection and reporting mechanisms. However, the report found that further refinements are needed to identify additional measures to support RWPA. Other work also needs to be conducted in outlining the mechanics for collecting RWPA data from multiple sources. This includes reducing reliance on solely manual collection of information and focusing on ways to use modern technology and leverage data from external sources. These action items were repeated as an important priority for CDRH in the January 2021 AI Action Plan.¹¹

The COVID-19 pandemic has shown that the creation of centralized medical data is possible in the U.S.,²⁷ even with the reality of a fragmented system of health records compared to other countries and the presence of overlapping or sometimes contradictory privacy laws at the federal, state, and local levels (discussed later). NIH's N3C database currently includes more than six million de-identified COVID-19 patient records and has become one of the largest in the world, exemplifying that significant issues related to data sharing can be overcome for the right reasons and if the right incentives are put in place.

Recommendations

To use RWD to efficiently evaluate the safety and performance of AI/ML-enabled CDS tools, policymakers and other stakeholders in this space will need to address the following persisting issues:

- Inconsistent and inefficient methods for capturing relevant data on algorithm accuracy and performance;
- A lack of data on patient health status and long-term outcomes;
- A lack of data on how physicians use (or do not use) algorithms' outputs;
- Inconsistent or lacking data across the continuum of a patient's care;
- Systemic biases within data sources that make it difficult to find representative data; and
- Inconsistent data elements or definitions for data elements across sites of data collection.

Many of these issues also affect the use of RWD for evaluation of other medical products. Living frameworks developed by the Duke-Margolis Center for Health Policy^{23,24} and NESTcc^{25,28} around RWD quality and methods, recent CDER/CBER draft Guidance on RWD use,⁵⁻⁸ and further maturation of data linkage and exchange can provide a foundation for the establishment of improved data collection mechanisms.

In addition, rich, interoperable datasets continuously collected from real-world clinical practice will be needed to train AI effectively. Programs such as the National Institutes of Health's (NIH) Bridge2AI are focused on bringing together technological and biomedical experts with social scientists and humanists to help create datasets that are specifically suitable for ML purposes.²⁹

Privacy and Access to RWD for Regulatory Surveillance

For RWD to be accessed legally and ethically, patient privacy must be a priority. Privacy and data access compliance issues vary, depending on whether RWE will be used by private-sector device manufacturers and software developers (for brevity, "developers") or by the FDA. Multiple laws apply to these issues, including the Health Insurance Portability and Accountability Act of 1996 (HIPAA), the Common Rule, State law, and FDA's human

subject protections. As discussed earlier, the 2017 FDA *Guidance on Use of Real-World Evidence to Support Regulatory Decision-Making for Medical Devices*³ and a more recent FDA publication⁴ gave examples of how RWE might be used for medical devices and CDS tools subject to regulation as medical devices.³⁰ Some of these examples describe uses of RWE by device manufacturers, for example, to expand indicated uses for a device already in clinical use, to conduct post-market surveillance or post-approval studies required as a condition of FDA approval, or to serve as control groups in pre-market studies for new devices. Other examples describe uses of RWE by FDA, for example, to investigate emerging safety issues not detected during pre-market review. This section considers both regulators and developers and focuses mainly on the surveillance aspects of using RWD for CDS tools.

For RWD to be accessed legally and ethically, patient privacy must be a priority.

Selection of Applicable Laws and Regulations on Data Privacy

Health Insurance Portability and Accountability Act of 1996 (HIPAA): The HIPAA regulations set forth rules to protect covered information while advancing the ability to have the safe, secure sharing of health data, including limits on who can access and share patients' health data and standards for data security when those data are collected or shared. These rules apply to "covered entities," which generally include healthcare providers, health insurers, and their business associates who help process or analyze data from providers and insurers.

Common Rule: The Common Rule, originally promulgated in 1991 and revised in 2018, is a set of regulations governing the ethics of research involving human subjects. These regulations are designed to protect human research subjects from harm and apply to all research conducted or supported by the federal government. Many U.S. academic institutions, through institutional review boards (IRBs), hold their researchers to these regulations regardless of funding.

FDA Human Subject Protections: These protections similarly aim to prevent harms to human subjects participating in research or clinical investigations to generate data for submission to FDA. In addition, FDA has regulations aimed at ensuring the quality of data from FDA-regulated studies. Depending on various factors (e.g., funding source and aims of research), a given study may be regulated by the Common Rule, the FDA human subject protections, both, or neither.

State, Tribal, and Local law: State, Tribal, and local law also apply to health care information stored about patients. HIPAA does not override State, Tribal, and local law provisions that are more stringent (more protective) than HIPAA.

Access to and Use of RWD by Private Software Developers

Developers can use any data (real-world data or clinical trial data) if the patients involved formally consent to that use. However, obtaining consent can be logistically difficult or even impracticable and, importantly, the patients that do consent may not be representative of the patient population on whom the software tool is being used. Because of that, developers may want to access more complete datasets, and there are methods to do that under HIPAA, each with their own benefits and challenges. When a software developer is *not subject* to the HIPAA Privacy Rule, there are often heightened privacy concerns about releasing RWD into its possession, even if de-identified, so healthcare providers may be more reluctant to share RWD.

Whether a developer is covered by the HIPAA Privacy Rule^{31,32} depends on specific facts of its business model. Independent software developers (i.e., developers that are not affiliated with health care systems) often are not HIPAA-covered entities, although they might still fall under the Privacy Rule as business associates of

hospitals and clinics with which they enter contracts.^{33,34} In contrast, academic medical centers and teaching hospitals that develop CDS tools in-house typically would be HIPAA-covered entities because they are engaged in the provision of health care.

The Privacy Rule does have various provisions allowing HIPAA-covered healthcare providers to share RWD with developers without patient authorization. For example, providers can share RWD in de-identified format.³⁵ Yet many providers use HIPAA's safe-harbor de-identification method,³⁶ which can reduce the utility of RWD, for example, by deleting patients' zip codes that might help uncover biases in CDS tools. The Privacy Rule also allows statistical de-identification, which could support modern computational privacy protections (privacy-by-design), but IRBs and data-holding institutions continue to favor the safe-harbor method,^{35,37,38} and further workforce development may be needed to foster wider use of, and trust in, statistical de-identification. More broadly, de-identification can make it difficult to spot duplicative data entries or to link data from multiple sources to assemble useful longitudinal health records, as discussed earlier. Partially

de-identified data can be shared as a “limited data set,” subject to a data use agreement restricting use, redisclosure, and re-identification.^{39,40} Identifiable data can also be shared, for example, if an IRB or privacy board approves a waiver of individual authorization, but many IRBs are reluctant to allow health systems to share data in identifiable format.

Alternatively, the Privacy Rule allows HIPAA-covered providers to share data with FDA-regulated developers in the private sector to enable them to perform various tasks (such as adverse-event reporting or post-market surveillance) that FDA requires with respect to products already on the market.⁴¹ Such disclosures serve a public health purpose. Unfortunately, because FDA’s pre-certification program is, at present, a voluntary program, it does not appear to set compliance “requirements” that would enable access to RWD using this provision of the Privacy Rule.

Public health disclosures of RWD must comply with HIPAA’s “minimum necessary” standard,⁴² but that standard can allow identifiers to be shared if they are necessary to fulfill the requestor’s intended use of the data. Thus, HIPAA allows a covered entity to share such data if the recipient provides justification (for example, if the recipient explains that it needs the data to study biases in FDA-regulated software tools). However, if those data, once received, were to be repurposed for further model training or to expand or refine the labeling, that new use would be considered “research” (commercial research) and is outside the scope of HIPAA’s public health exception. A fresh HIPAA authorization for the research use (or an IRB/Privacy Board waiver of authorization) should be obtained.

When the software developer is a HIPAA-covered entity, access to RWD is simplified.

When the software developer is a HIPAA-covered entity, access to RWD is simplified. A covered entity still must comply with the “minimum necessary” standard and only use data in research to the extent necessary to fulfill the purpose of the research. When CDS tools are developed in-house at a hospital or academic medical center, several additional HIPAA provisions can help provide access to needed RWD from that institution and potentially from

other institutions. For example, the HIPAA Privacy Rule allows data to be used for business operational purposes (e.g., quality improvement studies, such as a study of how software tools perform in clinical practice) without individual authorization.^{43,44} This and various other HIPAA provisions may allow access to RWD to support AI/ML development, validation, and post-marketing studies; however, a fact-specific analysis is always required to assess whether a given proposed use fits within one of the available HIPAA pathways for access.

The Privacy Rule also allows disclosure of data to public health authorities such as the FDA for various activities including public health surveillance and investigations.⁴⁵ This would allow the aggregated results of studies, as well as the underlying data, to be reported to FDA as necessary to support the Food, Drug, and Cosmetic Act.

It is unclear to what degree private-sector software developers will be subject to the Common Rule, which is triggered when an entity is conducting research that is federally funded, or to State or Tribal requirements that can be triggered depending on where the research takes place and what it involves. Regulatory uses of RWE by a private-sector company do not clearly fit the Common Rule’s definition of regulated “research,” and that is especially true under the new Common Rule, which expressly excludes public health surveillance from its research definition.^{46,47} Even if an activity does constitute research, it generally would not fall under the Common Rule if it is privately, rather than federally, funded. FDA’s human-subject protections for medical devices^{48,49} only apply to persons participating in a device clinical trial or persons on whose specimens FDA-regulated research is performed.⁵⁰⁻⁵² The observational uses of RWE contemplated here seemingly do not fall under the authority of FDA’s human-subject protections.

There is a lack of consistency and clarity about what is required to make post-market performance assessment of CDS tools possible as intended within the FDA’s pre-certification program. The above provisions of the HIPAA Privacy Rule are helpful, but they do not fully resolve problems developers face in gaining access to RWD for regulatory uses. The Privacy Rule’s provisions are all permissive in that they *allow* but do not *require* providers to share RWD.

Use of RWD by the FDA

FDA's own uses of RWD are governed not by the HIPAA Privacy Rule but by the Privacy Act of 1974.⁵³ The agency has a long history of handling sensitive data in the course of its regulatory decision-making and maintains a high level of public trust as a result. FDA, as part of the U.S. Department of Health and Human Services, is subject to the Common Rule in its own research uses of data, even though research by FDA-regulated private-sector entities often falls outside the Common Rule for reasons already discussed. To the extent FDA uses RWD in its own regulatory decision-making, these activities seemingly would be characterized – in the vast majority of cases – as public health surveillance or other

public health practice activities that are not “research” that falls under the Common Rule. The recent revisions to the Common Rule make clear that “public health surveillance activities, including the collection and testing of information or biospecimens, conducted, supported, requested, ordered, required, or authorized by a public health authority” are excluded from the definition of “research” that the Common Rule regulates.⁴⁷

Recommendations

Ensuring flows of RWD to support regulatory decision-making ultimately may require separate state or federal legislation spelling out when healthcare providers can be required to share RWD with software developers to support appropriate regulatory oversight. Such legislation could, for example, take the form of future state healthcare facility licensure statutes requiring facilities that implement CDS tools to supply RWD to developers to help them detect any problems. To protect patients' privacy, those same statutes could set limits on how developers can

use the RWD, for example, by imposing privacy standards or limiting downstream re-disclosures of RWD once it is in the developer's hands.

State comprehensive privacy law requirements that are more stringent than the Privacy Rule are not federally preempted and still apply in addition to the Privacy Rule.^{54,55} These laws could address perceived weaknesses in HIPAA's protections for RWD and bolster public trust in regulatory uses of RWD. Colorado⁵⁶ and Virginia,⁵⁷ following California's example, recently passed or signed legislation related to relevant privacy protections. However, they currently do not seem to impose additional restrictions on data or entities already subject to HIPAA and other federal privacy protections.

Both FDA and state governments will need to take further action in order to make RWD a viable and trusted tool for regulatory decision-making.

Ensuring flows of RWD to support regulatory decision-making ultimately may require separate state or federal legislation spelling out when healthcare providers can be required to share RWD with software developers to support appropriate regulatory oversight.

Mechanisms for Safe Data Sharing and Use

Data Sharing Methods

Beyond the aforementioned challenges with locating and obtaining complete and relevant data, as well as complying with applicable laws and ethics, the relevant parties involved in this work must also determine how to physically exchange data safely and securely.

This step comes with its own difficulties. FDA finds cybersecurity to be essential⁹ in medical device software, so the data architecture underlying the algorithm must be secure. FDA also aims to “explore less reliance on solely manual collection of information and more focus on ways to use technology, such as automated remote access to digital data, to collect SaMD product information once it is on the market.”¹¹ This is indeed a critical step toward streamlining data collection, but again requires very rigorous cybersecurity measures to ensure patients’ privacy is protected. Finally, in some instances of such data exchange, the inclusion of PHI or PII is required, or acceptable, increasing the need for more secure data platforms.

Beyond the aforementioned challenges with locating and obtaining complete and relevant data, as well as complying with applicable laws and ethics, the relevant parties involved in this work must also determine how to physically exchange data safely and securely.

The sharing and use of data can be governed by the execution of data sharing or data use agreements (DSA/DUAs). These agreements are legally enforceable contracts that clearly document what data are being shared and how the data can be used. HIPAA allows much data to be exchanged without DSA/DUAs, except for limited data sets. As discussed earlier, “limited data set” refers to a limited set of identifiable patient information as defined in the HIPAA Privacy Regulations. Because a “limited data set” is still considered PHI, the Privacy Regulations contemplate that the privacy of individuals will be protected by requiring covered entities to enter into DUAs with recipients of these data sets.

While DSA/DUAs are not always required, their use should be encouraged as they can further protect the privacy of people whose data are used. For example, a DSA/DUA could outline that data are allowed to be shared with software developers for the sole purpose of post-market surveillance of their CDS tool (i.e., to improve their algorithm), but subject to restrictions on reuse of the data without patient authorization (whether by de-identifying the data or by waiving authorization for further uses).

At a high level, the agreements should incorporate language that protects the entity providing the data, ensures that the data will not be misused, prevents miscommunication, and is enforceable. Specific concerns that the provider and receiving entity should address include the intended use and constraints on use of the data, data confidentiality and security, methods of data sharing, the period of the agreement, potential financial costs of data sharing, data destruction policy after use, and other relevant items related to the specifics of a project or its intended purpose.

The Office of the National Coordinator for Health Information Technology (ONC) within the department of Health and Human Services has developed a trusted exchange framework (TEF) and common agreement (CA), together named TEFCAs,⁵⁸ to establish a universal floor of interoperability across the U.S. The CA establishes an infrastructure model and governing approach to securely share clinical information, while the TEF describes a common set of foundational principles for trust policies and practices to help facilitate information exchange. While this framework is generally intended for the exchange of health information between doctors, nurses, pharmacists, or other healthcare providers and patients as part of routine care, its concepts could be applied to facilitate the development of standardized agreements with “plug-and-play” options depending on the context of use and data requirements.

To further simplify data sharing, ONC has also released its draft version 3.0 of the United States Core Data for Interoperability (USCDI) standard.⁵⁹ The USCDI sets a foundation for broader sharing of electronic health information to support patient care by providing

a standardized set of data elements for nationwide, interoperable health information exchange. The draft version 3.0 includes data elements for patient demographics, diagnoses, procedures, medications, laboratory tests, vital signs, diagnostic imaging, clinical notes, unique device identifiers and others. In addition to other common data standardization models, such as PCORnet, Sentinel, i2b2, and OMOP, the USCDI can be helpful in further standardizing and exchanging relevant RWD more efficiently. The NESTcc recently released an active surveillance roadmap outlining how the design, build, and current testing of a cloud environment for medical device post-market surveillance will incorporate many of the characteristics around data quality, data sharing, analytics, privacy, security, and interoperability across multiple participating health systems.⁶⁰

Reusable DSA/DUAs, in combination with standardized data models,⁶¹ can be used for efficient post-market data collection for AI/ML-enabled CDS tool performance evaluations as well as for other purposes that require RWD. Public-private partnerships within the medical device space, including the medical device industry, research organizations, hospitals, clinicians, payers, patient groups and the CDRH at FDA, can play a key role in facilitating the development of such standardized agreements and physical mechanisms for secure data transfer.

Data Sharing Platforms and Security

The availability of cloud computing and storage has revolutionized how health data can be stored and analyzed. However, given the sensitivity of health data, the security of data exchange platforms, such as clouds, should be a top priority. This is particularly essential across systems for data linkage between multiple institutions or entities. Security is complicated and requires strict adherence to comprehensive frameworks, including implementation of frequent security updates to protect against new potential threats. There are multiple security compliance frameworks for certifying platforms for health data storage and exchange, such as ISO 27001, FISMA, the NIST cybersecurity framework, FedRAMP and others, most of them incorporating HIPAA-related controls. Depending on the specific use cases the platform needs to support, one or multiple of these compliance frameworks may need to be applied.

Recommendations

Regulators and other stakeholders involved in the development of a future regulatory model for AI/ML-enabled CDS tools should continue to consider how to implement cloud platforms for sharing RWD that 1) contain appropriate security controls, 2) are audited by third-party experts, and 3) comply with existing security compliance frameworks. These platforms should include standardized tools and modules that are HIPAA-compliant

and are governed by re-usable DSAs/DUAs. Public-private partnerships that include stakeholders from across the medical device and healthcare ecosystem, such as NESTcc, are best positioned to facilitate this work and should pay attention to ongoing efforts within the field.

Leveraging Experiences and Data from International Settings

Other regulatory jurisdictions are considering AI/ML software tools similar to those under development in the U.S. and are considering policies to mobilize RWD for oversight of such tools.

Leveraging the experience obtained outside of the U.S. can be helpful in pre-market decisions and in the timely detection of potential safety issues during real-world use. FDA has used medical device clinical data from outside the U.S. (OUS) if the data meet certain standards.⁶² In a recent FDA publication including examples of RWE in support of regulatory applications, there were 22 submissions including OUS data.⁴

However, additional challenges must be addressed in order to incorporate OUS RWD in the timely, iterative, and consistent post-market performance evaluation of CDS tools. It is important to realize that CDS tools “work” on data collected about patients, not on patients directly, unlike many other medical devices. For example, an MRI machine operates similarly regardless of its geographical location, but software tools can perform very differently if there are significant underlying differences in the ways their data inputs are interpreted and recorded between countries. These differences could be caused by potential differences in patient populations, healthcare delivery and quality, coverage and reimbursement of medical

products, clinical data architectures and storage, and overall data standardization.

Furthermore, the accessibility of relevant data from international settings may be problematic as well, due to applicable comprehensive privacy regulations (e.g., the General Data Protection Regulation [GDPR] in the European Union). Making efficient and secure international data sharing possible on the large scale that would be ideal for AI/ML-enabled CDS tools will require a deep understanding of the data and their characteristics across different geographic contexts and the ability to identify mechanisms for data exchange under internationally applicable privacy laws.

As with other multi-stakeholder collaborations, this can best be achieved by international partnerships that include stakeholders from across the healthcare ecosystem (such as NESTcc, the international medical device regulators forum [IMDRF] and others). Notable examples are the Get-Real Institute in Europe, which facilitates the adoption and implementation of RWE in health-care decision-making,⁶³ and the addition of an international partner from the United Kingdom within NESTcc’s research network.⁶⁴ It will be worthwhile to closely monitor the learnings from these collaborations on how to address issues with international data sharing.

Summary and Conclusions

FDA has shown interest in manufacturers using RWD/RWE for the post-market performance evaluation of SaMD, including for AI/ML-enabled CDS tools, because of their rapid product lifecycles and frequent updates. However, such post-market evaluation can be complicated as software tools should ideally be monitored for potential changes in performance over time (e.g., changes in clinical workflows, patient populations, standards of care, data entry and variability across sites) as well as for inappropriate or variable usage (e.g., monitoring clinical actions taken in response to the tool and the appropriateness of those actions). To achieve this, specific data elements that should be captured within RWD sources include algorithm inputs and outputs, “gold standard” comparators, patient outcomes, and details on whether the healthcare professional acted on the algorithm’s recommendations. However, these data may not always be available or be of high enough quality or may require complicated linkage between siloed data sources. In addition, patient privacy protections, such as those enacted by HIPAA, the Common Rule, state law, and FDA’s human subject protections, may complicate data accessibility depending on whether RWD/RWE will be used by private-sector device manufacturers and software developers or by the FDA itself. Finally, parties involved in data collection must also determine how to exchange data securely and how the sharing and use of data should be governed from a legal perspective. Ensuring the flow of RWD to support consistent, timely, and efficient post-market performance evaluation of CDS tools as envisioned under FDA’s pre-certification program and the 2021 AI/ML action plan ultimately will require the capture of appropriate and high-quality data related to algorithm accuracy and variables that influence patient outcomes, separate legislation by federal and state legislators spelling out when healthcare providers can be required to share RWD with software developers, and outlining best practices for data sharing using secure software platforms governed by DSA/DUAs.

References

- ¹ U.S. Food and Drug Administration. Artificial Intelligence and Machine Learning (AI/ML)-Enabled Medical Devices. Accessed January 18th, 2022, <https://www.fda.gov/medical-devices/software-medical-device-samd/artificial-intelligence-and-machine-learning-ai-ml-enabled-medical-devices>
- ² U.S. Food and Drug Administration. Clinical Decision Support Software Draft Guidance for Industry and Food and Drug Administration Staff. Accessed January 18th, 2022, <https://www.fda.gov/media/109618/download>
- ³ U.S. Food and Drug Administration. Use of Real-World Evidence to Support Regulatory Decision-Making for Medical Devices Guidance for Industry and Food and Drug Administration Staff. Accessed January 18th, 2022, <https://www.fda.gov/regulatory-information/search-fda-guidance-documents/use-real-world-evidence-support-regulatory-decision-making-medical-devices>
- ⁴ U.S. Food and Drug Administration. Examples of Real-World Evidence (RWE) Used in Medical Device Regulatory Decisions Accessed January 18th, 2022, <https://www.fda.gov/media/146258/download>
- ⁵ U.S. Food and Drug Administration. Considerations for the Use of Real-World Data and Real-World Evidence To Support Regulatory Decision-Making for Drug and Biological Products Draft Guidance for Industry. Accessed January 18th, 2022, <https://www.fda.gov/media/154714/download>
- ⁶ U.S. Food and Drug Administration. Real-World Data: Assessing Electronic Health Records and Medical Claims Data To Support Regulatory Decision-Making for Drug and Biological Products Draft Guidance for Industry. Accessed January 18th, 2022, <https://www.fda.gov/media/152503/download>
- ⁷ U.S. Food and Drug Administration. Real-World Data: Assessing Registries to Support Regulatory Decision-Making for Drug and Biological Products Guidance for Industry Draft Guidance for Industry. Accessed January 18th, 2022, <https://www.fda.gov/media/154449/download>
- ⁸ U.S. Food and Drug Administration. Data Standards for Drug and Biological Product Submissions Containing Real-World Data Draft Guidance for Industry. Accessed January 18th, 2022, <https://www.fda.gov/media/153341/download>
- ⁹ U.S. Food and Drug Administration. Digital Health Software Precertification (Pre-Cert) Program. Accessed January 18th, 2022, <https://www.fda.gov/medical-devices/digital-health-center-excellence/digital-health-software-precertification-pre-cert-program>
- ¹⁰ U.S. Food and Drug Administration. Developing a Software Precertification Program: A Working Model v1.0. Accessed February 20th, 2022, <https://www.fda.gov/media/119722/download>
- ¹¹ U.S. Food and Drug Administration. Artificial Intelligence/Machine Learning (AI/ML)-Based Software as a Medical Device (SaMD) Action Plan. Accessed January 18th, 2022, <https://www.fda.gov/media/145022/download>
- ¹² U.S. Food and Drug Administration. CDRH Proposed Guidances for Fiscal Year 2022 (FY2022). Accessed January 18th, 2022, <https://www.fda.gov/medical-devices/guidance-documents-medical-devices-and-radiation-emitting-products/cdrh-proposed-guidances-fiscal-year-2022-fy2022>
- ¹³ Medical Device Innovation Consortium. MDICx: AI/ML Framework Public Comment Q&A. Accessed February 14th, 2022, <https://mdic.org/event/ivd-ai-ml-public-comment-announcement/>
- ¹⁴ U.S. Food and Drug Administration. De Novo Classification Request for Caption Guidance. Accessed January 18th, 2022, https://www.accessdata.fda.gov/cdrh_docs/reviews/DEN190040.pdf
- ¹⁵ U.S. Food and Drug Administration, Health Canada, Medicine and Healthcare Products Regulatory Agency. Good Machine Learning Practice for Medical Device Development: Guiding Principles. Accessed January 18th, 2022, <https://www.fda.gov/media/153486/download>
- ¹⁶ U.S. Food and Drug Administration. Public Workshop - Evolving Role of Artificial Intelligence in Radiological Imaging. Accessed January 18th, 2022, <https://www.fda.gov/medical-devices/workshops-conferences-medical-devices/public-workshop-evolving-role-artificial-intelligence-radiological-imaging-02252020-02262020>
- ¹⁷ Wong A, Otles E, Donnelly JP, et al. External Validation of a Widely Implemented Proprietary Sepsis Prediction Model in Hospitalized Patients. *JAMA Intern Med.* Jun 21 2021;doi:10.1001/jamainternmed.2021.2626
- ¹⁸ Zech JR, Badgeley MA, Liu M, Costa AB, Titano JJ, Oermann EK. Variable generalization performance of a deep learning model to detect pneumonia in chest radiographs: A cross-sectional study. *PLoS Med.* Nov 2018;15(11):e1002683. doi:10.1371/journal.pmed.1002683
- ¹⁹ U.S. Food and Drug Administration. Center for Devices and Radiological Health: 2022-2025 Strategic Priorities. Accessed February 14th, 2022, <https://www.fda.gov/media/155888/download>
- ²⁰ Paxton C, Niculescu-Mizil A, Saria S. Developing predictive models using electronic medical records: challenges and pitfalls. *AMIA Annu Symp Proc.* 2013;2013:1109-15.
- ²¹ ADvICE. ADvICE: Accelerated Digital Clinical Ecosystem. Accessed April 8th, 2021, <https://advice.ucsf.edu/>
- ²² American College of Radiology. Assess-AI. Accessed January 18th, 2022, <https://www.acrdsi.org/DSI-Services/Assess-AI>
- ²³ Duke-Margolis Center for Health Policy. Characterizing RWD Quality and Relevancy for Regulatory Purposes. Accessed January 18th, 2022, <https://healthpolicy.duke.edu/sites/default/files/2020-08/Characterizing%20RWD%20for%20Regulatory%20Use.pdf>
- ²⁴ Duke-Margolis Center for Health Policy. Determining Real-World Data's Fitness for Use and the Role of Reliability. Accessed January 18th, 2022, https://healthpolicy.duke.edu/sites/default/files/2019-11/rwd_reliability.pdf
- ²⁵ National Evaluation System for health Technology Coordinating Center. National Evaluation System for health Technology Coordinating Center Data Quality Framework. Accessed January 18th, 2022, <https://nestcc.org/nestcc-data-quality-framework/>
- ²⁶ U.S. Food and Drug Administration. Developing the Software Precertification Program: Summary of Learnings and Ongoing Activities. Accessed January 18th, 2022, <https://www.fda.gov/media/142107/download>
- ²⁷ MIT Technology Review. It took a pandemic, but the US finally has (some) centralized medical data. Accessed January 18th, 2022, <https://www.technologyreview.com/2021/06/21/1026590/us-covid-database-n3c-nih-privacy/>

- ²⁸ National Evaluation System for health Technology Coordinating Center. National Evaluation System for health Technology Coordinating Center Research Methods Framework. Accessed January 18th, 2022, <https://nestcc.org/nestcc-methods-framework/>
- ²⁹ National Institutes of Health. Bridge to Artificial Intelligence (Bridge2AI). Accessed January 18th, 2022, <https://commonfund.nih.gov/bridge2ai>
- ³⁰ U.S. House of Representatives. 21 USC CHAPTER 9, SUBCHAPTER V: DRUGS AND DEVICES. §360j(o)(1)(E) Accessed January 18th, 2022, <https://uscode.house.gov/view.xhtml?path=/prelim@title21/chapter9/subchapter5&edition=prelim>
- ³¹ Code of Federal Regulations. Title 45 Subtitle A Subchapter C Part 160. Accessed January 18th, 2022, <https://www.ecfr.gov/cgi-bin/text-idx?SID=7cc9760c0f6ba628c3ef7cf380c5c168&mc=true&node=pt45.2.160&rgn=div5>
- ³² Code of Federal Regulations. Title 45 Subtitle A Subchapter C Part 164. Accessed January 18th, 2022, <https://www.ecfr.gov/cgi-bin/text-idx?SID=7cc9760c0f6ba628c3ef7cf380c5c168&mc=true&node=pt45.2.164&rgn=div5>
- ³³ Code of Federal Regulations. Title 45 Subtitle A Subchapter C Part 160 §102(b). Accessed January 18th, 2022, <https://www.ecfr.gov/cgi-bin/text-idx?SID=7cc9760c0f6ba628c3ef7cf380c5c168&mc=true&node=pt45.2.160&rgn=div5>
- ³⁴ Code of Federal Regulations. Title 45 Subtitle A Subchapter C Part 164 §500 Accessed January 18th, 2022, <https://www.ecfr.gov/cgi-bin/text-idx?SID=7cc9760c0f6ba628c3ef7cf380c5c168&mc=true&node=pt45.2.164&rgn=div5>
- ³⁵ Code of Federal Regulations. Title 45 Subtitle A Subchapter C Part 164 §502(d). Accessed January 18th, 2022, <https://www.ecfr.gov/cgi-bin/text-idx?SID=7cc9760c0f6ba628c3ef7cf380c5c168&mc=true&node=pt45.2.164&rgn=div5>
- ³⁶ Code of Federal Regulations. Title 45 Subtitle A Subchapter C Part 164 §514(b)(2). Accessed January 18th, 2022, <https://www.ecfr.gov/cgi-bin/text-idx?SID=7cc9760c0f6ba628c3ef7cf380c5c168&mc=true&node=pt45.2.164&rgn=div5>
- ³⁷ Code of Federal Regulations. Title 45 Subtitle A Subchapter C Part 164 §514(a). Accessed February 20th, 2022, <https://www.ecfr.gov/cgi-bin/text-idx?SID=7cc9760c0f6ba628c3ef7cf380c5c168&mc=true&node=pt45.2.164&rgn=div5>
- ³⁸ Code of Federal Regulations. Title 45 Subtitle A Subchapter C Part 164 §514(b)(1). Accessed February 20th, 2022, <https://www.ecfr.gov/cgi-bin/text-idx?SID=7cc9760c0f6ba628c3ef7cf380c5c168&mc=true&node=pt45.2.164&rgn=div5>
- ³⁹ Code of Federal Regulations. Title 45 Subtitle A Subchapter C Part 164 §514(e)(3)(i). Accessed January 18th, 2022, <https://www.ecfr.gov/cgi-bin/text-idx?SID=7cc9760c0f6ba628c3ef7cf380c5c168&mc=true&node=pt45.2.164&rgn=div5>
- ⁴⁰ Code of Federal Regulations. Title 45 Subtitle A Subchapter C Part 164 §514(e)(4). Accessed January 18th, 2022, <https://www.ecfr.gov/cgi-bin/text-idx?SID=7cc9760c0f6ba628c3ef7cf380c5c168&mc=true&node=pt45.2.164&rgn=div5>
- ⁴¹ Code of Federal Regulations. Title 45 Subtitle A Subchapter C Part 164 §512(b)(1)(ii). Accessed January 18th, 2022, <https://www.ecfr.gov/cgi-bin/text-idx?SID=7cc9760c0f6ba628c3ef7cf380c5c168&mc=true&node=pt45.2.164&rgn=div5>
- ⁴² Code of Federal Regulations. Title 45 Subtitle A Subchapter C Part 164 §502(b). Accessed January 18th, 2022, <https://www.ecfr.gov/cgi-bin/text-idx?SID=7cc9760c0f6ba628c3ef7cf380c5c168&mc=true&node=pt45.2.164&rgn=div5>
- ⁴³ Code of Federal Regulations. Title 45 Subtitle A Subchapter C Part 164 §502(a)(1)(ii). Accessed January 18th, 2022, <https://www.ecfr.gov/cgi-bin/text-idx?SID=7cc9760c0f6ba628c3ef7cf380c5c168&mc=true&node=pt45.2.164&rgn=div5>
- ⁴⁴ Code of Federal Regulations. Title 45 Subtitle A Subchapter C Part 164 §506. Accessed January 18th, 2022, <https://www.ecfr.gov/cgi-bin/text-idx?SID=7cc9760c0f6ba628c3ef7cf380c5c168&mc=true&node=pt45.2.164&rgn=div5>
- ⁴⁵ Code of Federal Regulations. Title 45 Subtitle A Subchapter C Part 164 §512(b)(1)(i). Accessed January 18th, 2022, <https://www.ecfr.gov/cgi-bin/text-idx?SID=7cc9760c0f6ba628c3ef7cf380c5c168&mc=true&node=pt45.2.164&rgn=div5>
- ⁴⁶ Code of Federal Regulations. Title 45 Subtitle A Subchapter A Part 46 §102(l). Accessed January 18th, 2022, <https://www.ecfr.gov/cgi-bin/text-idx?SID=7cc9760c0f6ba628c3ef7cf380c5c168&mc=true&node=pt45.1.46&rgn=div5>
- ⁴⁷ Code of Federal Regulations. Title 45 Subtitle A Subchapter A Part 46 §102(l)(2). Accessed January 18th, 2022, <https://www.ecfr.gov/cgi-bin/text-idx?SID=7cc9760c0f6ba628c3ef7cf380c5c168&mc=true&node=pt45.1.46&rgn=div5>
- ⁴⁸ Code of Federal Regulations. Title 21 Chapter 1 Subchapter A Part 50. Accessed January 18th, 2022, <https://www.ecfr.gov/cgi-bin/text-idx?SID=7cc9760c0f6ba628c3ef7cf380c5c168&mc=true&node=pt21.1.50&rgn=div5>
- ⁴⁹ Code of Federal Regulations. Title 21 Chapter 1 Subchapter A Part 56. Accessed January 18th, 2022, <https://www.ecfr.gov/cgi-bin/text-idx?SID=7cc9760c0f6ba628c3ef7cf380c5c168&mc=true&node=pt21.1.56&rgn=div5>
- ⁵⁰ Code of Federal Regulations. Title 21 Chapter 1 Subchapter A Part 50 §3(g). Accessed January 18th, 2022, <https://www.ecfr.gov/cgi-bin/text-idx?SID=7cc9760c0f6ba628c3ef7cf380c5c168&mc=true&node=pt21.1.50&rgn=div5>
- ⁵¹ Code of Federal Regulations. Title 21 Chapter 1 Subchapter A Part 56 §102(e). Accessed January 18th, 2022, <https://www.ecfr.gov/cgi-bin/text-idx?SID=7cc9760c0f6ba628c3ef7cf380c5c168&mc=true&node=pt21.1.56&rgn=div5>
- ⁵² Code of Federal Regulations. Title 21 Chapter 1 Subchapter H Part 812 §3(p). Accessed January 18th, 2022, <https://www.ecfr.gov/cgi-bin/text-idx?SID=7cc9760c0f6ba628c3ef7cf380c5c168&mc=true&node=pt21.8.812&rgn=div5>
- ⁵³ U.S. House of Representatives. 5 USC CHAPTER 5, SUBCHAPTER II: ADMINISTRATIVE PROCEDURE. §552(a). Accessed January 18th, 2022, <https://uscode.house.gov/view.xhtml?req=granuleid:USC-prelim-title5-section552&num=0&edition=prelim>
- ⁵⁴ Code of Federal Regulations. Title 45 Subtitle A Subchapter C Part 160 §202. Accessed January 18th, 2022, <https://www.ecfr.gov/cgi-bin/text-idx?SID=7cc9760c0f6ba628c3ef7cf380c5c168&mc=true&node=pt45.2.160&rgn=div5>
- ⁵⁵ Code of Federal Regulations. Title 45 Subtitle A Subchapter C Part 160 §203. Accessed January 18th, 2022, <https://www.ecfr.gov/cgi-bin/text-idx?SID=7cc9760c0f6ba628c3ef7cf380c5c168&mc=true&node=pt45.2.160&rgn=div5>
- ⁵⁶ Colorado General Assembly. SB21-190: Protect Personal Data Privacy. Accessed January 18th, 2022, https://leg.colorado.gov/sites/default/files/documents/2021A/bills/2021a_190_enr.pdf

- ⁵⁷ General Assembly of Virginia. SB 1392: Consumer Data Protection Act. Accessed January 18th, 2022, <https://lis.virginia.gov/cgi-bin/legp604.exe?211+ful+SB1392ES1+pdf>
- ⁵⁸ Office of the National Coordinator for Health Information Technology. Trusted Exchange Framework and Common Agreement (TEFCA). Accessed January 18th, 2022, <https://www.healthit.gov/topic/interoperability/trusted-exchange-framework-and-common-agreement-tefca>
- ⁵⁹ Office of the National Coordinator for Health Information Technology. United States Core Data for Interoperability (USCDI) Draft 3.0. Accessed February 13th, 2022, <https://www.healthit.gov/isa/sites/isa/files/2022-01/Draft-USCDI-Version-3-January-2022-Final.pdf>
- ⁶⁰ National Evaluation System for health Technology Coordinating Center. National Evaluation System for health Technology Coordinating Center Active Surveillance Roadmap. Accessed January 18th, 2022, <https://nestcc.org/wp-content/uploads/2021/08/NESTcc-Active-Surveillance-Roadmap-for-Public-Comment.pdf>
- ⁶¹ Jose J, Hendricks-Sturup R. A HIPAA “Expert Pathway” Method to Achieve More Efficient Access to Protected Healthcare Information. Accessed January 18th, 2022, <https://law.mit.edu/pub/ahipaaexpertpathway/release/4>
- ⁶² U.S. Food and Drug Administration. Acceptance of Clinical Data to Support Medical Device Applications and Submissions Frequently Asked Questions. Accessed January 18th, 2022, <https://www.fda.gov/media/111346/download>
- ⁶³ GetReal Institute. GetReal Institute. Accessed February 14th, 2022, <https://www.getreal-institute.org/>
- ⁶⁴ National Evaluation System for health Technology Coordinating Center. NESTcc Announces First International Research Network Collaborator. Accessed January 18th, 2022, <https://nestcc.org/nestcc-announces-first-international-research-network-collaborator/>