Generating Real-World Evidence by Strengthening Real-World Data Sources

Using "real-world evidence" to bring new treatments to patients as part of the 21st Century Cures Act (the "Cures Act") is a key priority for the Department of Health and Human Services (HHS). Specifically, the Cures Act places focus on the use of real-world data to support regulatory decision-making, including the approval of new indications for existing drugs in order to make drug development faster and more efficient. As such, the Cures Act has tasked the Food and Drug Administration (FDA) to develop a framework and guidance for evaluating real-world evidence in the context of drug regulation.[1]

> "Every day, health care professionals are updating patients' electronic health records with data on clinical outcomes resulting from medical interventions used in routine clinical practice. As our experience with new medical products expands, our knowledge about how to best maximize their benefits and minimize potential risks sharpens with each data point we gather. Every clinical use of a product produces data that can help better inform us about its safety and efficacy."
>
> **JACQUELINE CORRIGAN-CURAY, MD, JD**
> **DIRECTOR OF THE OFFICE OF MEDICAL POLICY IN FDA'S CENTER FOR DRUG EVALUATION AND RESEARCH**

Under the FDA's framework, **real-world evidence (RWE)** is generated by different study designs or analyses, including but not limited to, randomized trials like large simple trials, pragmatic trials, and observational studies. RWE is the clinical evidence about the use, potential benefits, and potential risks of a medical product based on an analysis of **real-world data (RWD)**. RWD are data about patient health status and/or health care delivery that are routinely collected from a variety of sources, including electronic health records (EHRs), claims data, registries, and patient-generated health data.[2]

Exhibit A. FDA RWE Framework. RWD and RWE are playing an increasing role in health care decisions, from developing guidelines and decision support tools for clinical practice to supporting clinical trial designs and observational studies that generate innovative treatment approaches.[3]

Sources for Real-World Data
EHRs, insurance claims, patient registries, digital health applications

Real-World Evidence
Clinical evidence regarding the use, potential benefits, and potential risks of a medical product derived from analysis of real-world data

Real-World Data
Data relating to patient health status and/or the delivery of health care routinely collected from a variety of sources

High quality, real-time health data are often difficult to access because of lack of data standardization, cost, patient privacy, or other legal and intellectual property restrictions.[4] Further, due to a variety of interoperability issues, it is often difficult to bring data together from different RWD sources. Under the Office of the Secretary Patient-Centered Outcomes Research Trust Fund (OS-PCORTF) portfolio, the Assistant Secretary for Planning and Evaluation (ASPE) supports and coordinates a range of cross-agency projects that are working to strengthen the availability of RWE by standardizing RWD sources so they are fit for use and by building linkages across RWD sources so data are easier to share and analyze.

Standardizing RWD Sources So They Are Fit For Use.

The need for standardized data is a key challenge for researchers trying to analyze the unstructured narrative or "free text" fields in pathology, post-market, biomarker, and EHR reports.[5] The process of extracting this narrative data manually is labor-intensive and expensive. To address this issue, the FDA and several federal partners have undertaken projects under the OS-PCORTF portfolio.

Development of a Natural Language Processing (NLP) Web Service for Public Health Use

The FDA and the Centers for Disease Control and Prevention (CDC) collaborated on a project, the *Development of a Natural Language Processing (NLP) Web Service for Public Health Use*, which was completed in 2019. The goal of this project was to develop a structured and standardized way to code free text within the EHR using NLP tools. NLP allows computers and other technologies to interpret human language. This removes the manual process for researchers, while increasing the completeness, timeliness, and accuracy of narrative text data, which can then be used for public health research. The project team developed the Clinical Language Engineering Workbench (CLEW), a cloud-based, open source, web service that hosts NLP and machine learning tools, clinical NLP services, and gives users the opportunity for tool development. The team tested the CLEW NLP tools through two pilot projects: a cancer pathology registry (CDC) and the Safety Surveillance program (FDA). Using the CLEW, the pilot projects successfully converted free text clinical data into standardized coded data—meaning that quality RWD that can be used to inform drug safety surveillance and other types of research.

KEY PRODUCTS

- The CLEW Platform is available to CDC researchers on the agency's website
- The project systems, code, and documentation (e.g., user guidance, lessons learned, and a technical report) are all publicly available on the CDC's Github page and the FDA's Github page
- The project published a systematic review of existing clinical NLP systems: Natural language processing systems for capturing and standardizing unstructured clinical information: a systematic review, Journal of Biomedical Informatics 2017
- The project published a paper on how to create an annotated dataset for training NLP models: Generation of an annotated reference standard for vaccine adverse event reports, Vaccine 2018

PROJECT AGENCIES: FDA AND CDC

Harmonization of Various Common Data Models and Open Standards for Evidence Generation

The FDA, Office of the National Coordinator of Health Information Technology (ONC), and the National Institutes of Health (NIH)--specifically, the National Center for Advancing Translational Sciences (NCATS), National Cancer Institute (NCI), and the National Library of Medicine (NLM)--are addressing the need for standardized RWD through the *Harmonization of Various Common Data Models and Open Standards for Evidence Generation* project, which began in 2017. The project is enabling the use of data among four major research networks–FDA Sentinel, Accrual to Clinical Trials (ACT) Network and Informatics for Integrating Biology & the Bedside (i2b2), Patient-Centered Outcomes Research Network (PCORnet), and Observational Medical Outcomes Partnership (OMOP)–to support research across a range of health issues. The four research networks

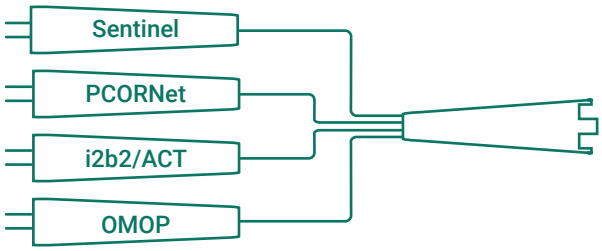
KEY PRODUCTS

- The project published an implementation guide for harmonizing common data models in HL7 Fast Healthcare Interoperability Resources (FHIR) along with mapping tools: CDM-to-BRIDG, BRIDG-to-SDTM, and a BRIDG-to-FHIR for each research network

PROJECT AGENCIES: FDA, NIH (NCATS, NCI, NLM), AND ONC

(see Exhibit B) each contain information on treatments and patient outcomes from data sources like EHRs and administrative claims data. Harmonizing the data from these research networks will allow researchers to leverage large amounts of RWD from across the networks and will allow them to investigate trends related to patient demographics (e.g., elderly, pediatric, and non-US populations). In turn, this will help answer research questions about the safety of different cancer treatment options, and it will generate RWE that supports patient and provider decision-making. The team is currently leveraging the project infrastructure to advance FDA and NIH's understanding of the coronavirus and potentially identify more effective treatments. The centralized database, designed as part of the National COVID-19 Cohort Collaborative (N3C) project, could provide sharper insight into coronavirus patients' needs. NIH anticipates that it could serve as a foundation for addressing future public health emergencies.[6]

Exhibit B. Harmonizing Research Networks. *The Harmonization of Various Common Data Models and Open Standards for Evidence Generation project* seeks to align data from across four research networks to generate RWE that supports patient and provider decision-making.



THE SOLUTION:

- Use a converter between various adapters
- Allow researchers to ask a question once and receive results from many different sources using a common agreed-upon standard structure or common data model (CDM)

Standardization and Querying of Data Quality Metrics and Characteristics for Electronic Health Data Project

KEY PRODUCTS

- The Data Quality Metrics Authoring and Querying Platform is a cloud-based, open-source tool for using and authoring quality metrics
- The Technical Documentation Report provides technical documentation for software developers and other technical users
- User Documentation supports researchers in using the platform and tools

PROJECT AGENCY: FDA

The FDA's *Standardization and Querying of Data Quality Metrics and Characteristics for Electronic Health Data Project* was designed to fill the gap in standards that describe the quality and completeness of electronic health data. As researchers increasingly use networks to diversify their data sources, it can be difficult to understand the sources and characteristics of the available data, and to assess whether the data are fit-for-purpose. Adding to the uncertainty is the fact that research networks have their own data quality and validation processes. To address these areas, the project has created an online platform with an open-source toolkit that contains visualization and analytic tools, and a data quality model to help researchers assess data sources across diverse networks. Metadata standards help create a fingerprint for each data source that gives researchers vital information about the data, such as the setting (e.g., labs, hospitals) and the measures (e.g., how many responses fall outside of an expected range) that allow them to determine whether the data set is suitable for their own studies.

Source Data Capture from Electronic Health Records: Using Standardized Clinical Research Data

Researchers rely on robust clinical data to generate meaningful findings, and yet the information systems into which clinical care data and clinical research data are captured, organized, and stored can be very different, creating barriers to research. Given that clinical data is captured in EHRs, the FDA conducted a project in 2018, *Source Data Capture from Electronic Health Records: Using Standardized Clinical Research Data (One Source)* to automate the flow of structured EHR data into data systems that support clinical trials. Removing manual processes helped make data transfer faster and more efficient, it reduced costs and burden for health care providers and research staff, and it improved the data quality. The open-source forms, source code, and standards enhancement considerations developed in the course of this project are available to the public and organizations interested in using EHR data for conducting clinical research.

KEY PRODUCTS

The final report details activities and products to support widespread use of the OneSource platform and tools:

- Source code for Integrating the Healthcare Enterprise (IHE) Retrieve Form for Data Capture standard and EHRs (Epic Integration code base)
- Detailed Gap analysis between Case Report Forms and University of California, San Francisco (UCSF) EHR system
- Guidelines for future source data capture implementations in supporting both health care and clinical research
- Lessons learned and considerations for future work in standards and implementations

PROJECT AGENCIES: FDA

Linking Patient-Generated and EHR Data.

Leveraging patient-generated health data to generate insight into patient-experienced outcomes in the real world has gained increased attention from scientists, industry, and regulators in recent years.[7] As researchers look to collect more patient-generated data, there is a need for user-friendly solutions that enable patients to engage with their health data while also lessening burden on them. Health-supportive apps and app-based donation of data to research programs can address both.

Collection of Patient-Provided Information through a Mobile Device Application for Use in Comparative Effectiveness and Drug Safety Research

Under the *Collection of Patient-Provided Information through a Mobile Device Application for Use in Comparative Effectiveness and Drug Safety Research* project, FDA developed and piloted a mobile app, the FDA MyStudies App, to capture patient-generated data from pregnant women. This project, which ended in 2018, focused on medical product exposures during pregnancy, outcomes, risk factors, and confounding factors. The app quickly engaged the study population of women in their first trimester of pregnancy using data from the EHR, and captured sensitive information including continued alcohol, smoking, and illicit drug use during pregnancy. The MyStudies App also linked the patient-generated data to electronic health data in FDA's Sentinel CDM, which is comprised largely of administrative and claims data from health insurance plans, to further enrich it for research analysis.[8] To assist other researchers in capturing and using patient-generated data, the FDA released open-source code for two versions of the MyStudies App: one built on Apple's ResearchKit (iOS) framework, and the other on the open-source ResearchStack framework, which runs on Google's Android. Through this app, FDA is enhancing usability and allowing for integration of other RWD sources and the generation of RWE. Since the project ended, the Crohn's & Colitis Foundation has built upon the FDA MyStudies App to capture patient experience data beyond the clinical care system to establish a comprehensive picture of patients' disease journeys. In addition, FDA is making the app available to investigators as a free platform to obtain informed consent securely from patients for eligible clinical trials when face-to-face contact is not possible or practical due to COVID-19 control measures.[9]

KEY PRODUCTS

- The MyStudies app is available on the Google Cloud Platform as well as the ResearchStack (Android) framework and Apple's ResearchKit (iOS) framework.
- The source code for the app is available on GitHub so that developers can improve on its capabilities

PROJECT AGENCIES: FDA

Use of the ADAPTABLE Trial to Strengthen Methods to Collect and Integrate Patient-Reported Information with Other Data Sets and Assess Its Validity

KEY PRODUCTS

- The Patient-Reported Data Assessment Tool and technical and user documentation are publicly available on Github
- Key data elements have been added to the NIH CDE Repository

PROJECT AGENCIES: NIH

When patient-generated data can be linked to existing clinical data, they are even more useful for researchers. The NIH's project, *Use of the ADAPTABLE Trial to Strengthen Methods to Collect and Integrate Patient-Reported Information with Other Data Sets and Assess Its Validity*, which was completed in 2019, developed, tested, and evaluated patient-generated data standards to make sure any patient-reported data that might be used in research would be complete, consistent, and valid. Based on this evaluation, the NIH team produced a Patient-Reported Data Assessment Tool that allows investigators to evaluate the quality of their own patient-generated data compared to EHR data using a menu-driven query tool. The patient-generated data provides information on relevant clinical events or encounters patients experience outside of the clinical trial setting, while also supplementing EHRs with important patient-reported and demographic information that is of use to researchers.

Developing a Strategically Coordinated Registry Network (CRN) to Support Research on Women's Health Technologies

KEY PRODUCTS

- Women's Health Technology Medical Device Epidemiology Network Website is intended to raise awareness of the WHT-CRN project
- The WHT-CRN Common Core Data Set can be used by researchers and providers in and outside the CRN to standardize data for analysis
- HL7 WHT-CRN FHIR IG provides guidelines for how to use FHIR to capture and exchange data among the registries. This IG will be revised an expanded by the FDA's COP follow-on project

PROJECT AGENCIES: FDA, ONC, AND NIH/NLM

Lack of standardized data from diverse sources, including patient-reported outcomes (PROs) creates gaps in research and evidence-based guidance for women's health issues. In a recently completed project, *Developing a Strategically Coordinated Registry Network to Support Research on Women's Health Technologies (WHT-CRN),* the FDA, NIH/National Library of Medicine (NLM), and ONC merged data from registries, claims, EHRs, and PRO sources on women's health devices to help researchers study their safety and effectiveness. As part of this work, the project team identified data elements that are captured most often and harmonized them across registries so they can be recorded in a standardized way and shared among the CRN. When the WHT-CRN project concluded in 2019, a related project launched, *Bridging the PCOR Infrastructure and Technology Innovation through Coordinated Registry Networks (CRN) Community of Practice (COP).* In this follow-on project, the FDA plans to strengthen the PCOR infrastructure by applying the lessons learned from the WHT-CRN to the CRN Community of Practice (COP) in 12 clinical areas. The goal is to enhance data collection, sharing, and research across participating CRNs, creating a minimum data set that includes device identification numbers, adding patient-generated data, and linking the data to other data sources (i.e., clinical, claims, registries). This project will continue building infrastructure to generate evidence around the safety and effectiveness of medical devices used in women's health and beyond.

Linking EHR Data to Other RWD Sources.

Linking clinical data to other RWD sources can help paint a more complete picture of a public health issue. For example, research on childhood obesity interventions is often limited because researchers cannot easily link pediatric health-related data stored across different health information systems to assess effectiveness.

Childhood Obesity Data Initiative (CODI): Integrated Data for Patient-Centered Outcomes Research Project

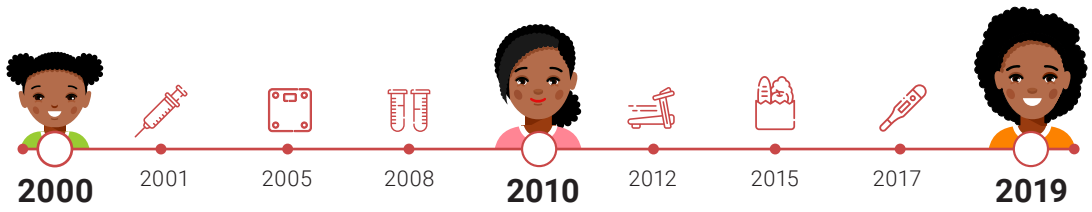
The CDC's *Childhood Obesity Data Initiative (CODI): Integrated Data for Patient-Centered Outcomes Research Project*, which began in 2018, is working to address this problem by leveraging existing information technology tools in innovative ways to facilitate access to childhood obesity data across health systems and sectors. To do so, the project is developing and testing an expanded common data model including pediatric obesity-related information, patient linkage and deduplication tools, and data query services to bring together data stored across different organizations to create an individual-level, linked longitudinal record that includes individual and community-level risk factors, weight management interventions delivered in clinical and community settings, and clinical outcomes across health information systems, (as shown in the graphic). The data linkage and query tools will be piloted in the Colorado Health Observation Regional Data Service (CHORDS) and will support evidence generation about effective treatment strategies and patient outcomes. This RWD and RWE will help health professionals in developing and tailoring weight management interventions to the specific needs of children.

ANTICIPATED PRODUCTS

- The project team conducted a technical environmental scan, which provides documentation of the current business and technical processes and tasks for capturing the required childhood obesity data
- The team has published the CODI Data Models Implementation Guide for use of the record linkage and deduplication services
- Linkage and deduplication tools and services will be made available on the CDC's cloud-based Surveillance Data Platform

PROJECT AGENCY: CDC

Exhibit C. The CODI Solution. CODI will bring together data stored across different organizations to create an individual-level, linked longitudinal record.[10]



Looking to the Future. As the availability and use of RWD rapidly advances, the OS-PCORTF has funded many new projects to further interoperable data sharing and linkages among key data sources. The widespread adoption and use of health care information systems and new health technologies to capture, store, manage, or transmit health data will change the types of data traditionally used to support real-world decision making. The introduction of these new data sources may contribute to regulatory decisions and support optimizing treatment decisions by health care practitioners.[11]

ADDITIONAL OS-PCORTF FUNDED PROJECTS

- The Agency for Healthcare Research and Quality (AHRQ)'s and the Office of the Assistant Secretary for Preparedness and Response (ASPR)'s *Assessing and Predicting Medical Needs in a Disaster* (FY 2018) project is working to build a data platform to expand the Healthcare Cost and Utilization Project (HCUP) data set to include quarterly emergency department and in-patient data.
- ASPE's and the Administration for Children and Families' (ACF) Linking State *Medicaid and Child Welfare Data for Outcomes Research on Treatment for Opioid Use Disorder (OUD)* (FY 2019) project is working to link Medicaid and child welfare records. These data sets will contain linked patient-level data, including Medicaid enrollment, patient diagnoses, services, and claims, along with child welfare outcomes (e.g., length of time in foster care, repeat maltreatment) to generate RWE to inform treatment of OUD.
- CDC's *Making Electronic Health Record (EHR) Data More Available for Research and Public Health* (FY 2019) project is working to develop an application to extract data from multiple clinical organizations using multiple EHR platforms. This will enable greater interoperability across EHR systems in order to generate RWE about specific health problems.

REFERENCES

1. Caffrey M. Getting Ready for the Use of Real-World Evidence. In Focus Blog. February 2019. https://www.ajmc.com/focus-of-the-week/getting-ready-for-the-use-of-realworld-evidence
2. FDA. Framework for FDA's Real-World Evidence Program. December 2018. www.fda.gov/media/120060/download
3. Gottlieb S. Getting Ready for the Use of Real-World Evidence. In Focus Blog. February 2019. https://www.healthitanswers.net/fda-budget-matters-a-cross-cutting-data-enterprise-for-real-world-evidence-2/
4. Onwudiwe NC, Tenenbaum K, Boise BH et al. Real World Evidence: Implications and Challenges for Medical Product Communications in an Evolving Regulatory Landscape. Food and Drug Law Institute: Update Magazine. August/September 2018. https://www.fdli.org/2018/08/update-real-world-evidence-implications-and-challenges-for-medical-product-communications-in-an-evolving-regulatory-landscape/
5. CDC. Natural Language Processing Workbench Web Services. July 2017. https://www.cdc.gov/cancer/npcr/informatics/nlp-workbench/index.htm
6. National Center for Advancing Translational Sciences (NCATS). National COVID Cohort Collaborative (N3C). https://www.ohdsi.org/wp-content/uploads/2020/05/N3C-OHDSI-Community-Call-5.26.20.pdf
7. McDonald L, Malcolm B, Ramagopalan S, Syrad H. Real-world data and the patient perspective: the PROmise of social media? BMC Med. 2019;17(1):11. Published 2019 Jan 16. doi:10.1186/s12916-018-1247-8
8. Sentinel Initiative. Sentinel System's Story for the Public. Retrieved from https://www.sentinelinitiative.org/
9. FDA. COVID MyStudies Application (App). https://www.fda.gov/drugs/science-and-research-drugs/covid-mystudies-application-app
10. CDC. Childhood Obesity Data Initiative. October 2019. https://www.cdc.gov/obesity/initiatives/codi/childhood-obesity-data-initiative.html
11. Onwudiwe NC, Tenenbaum K, Boise BH et al. Real World Evidence: Implications and Challenges for Medical Product Communications in an Evolving Regulatory Landscape. Food and Drug Law Institute: Update Magazine. August/September 2018. https://www.fdli.org/2018/08/update-real-world-evidence-implications-and-challenges-for-medical-product-communications-in-an-evolving-regulatory-landscape/