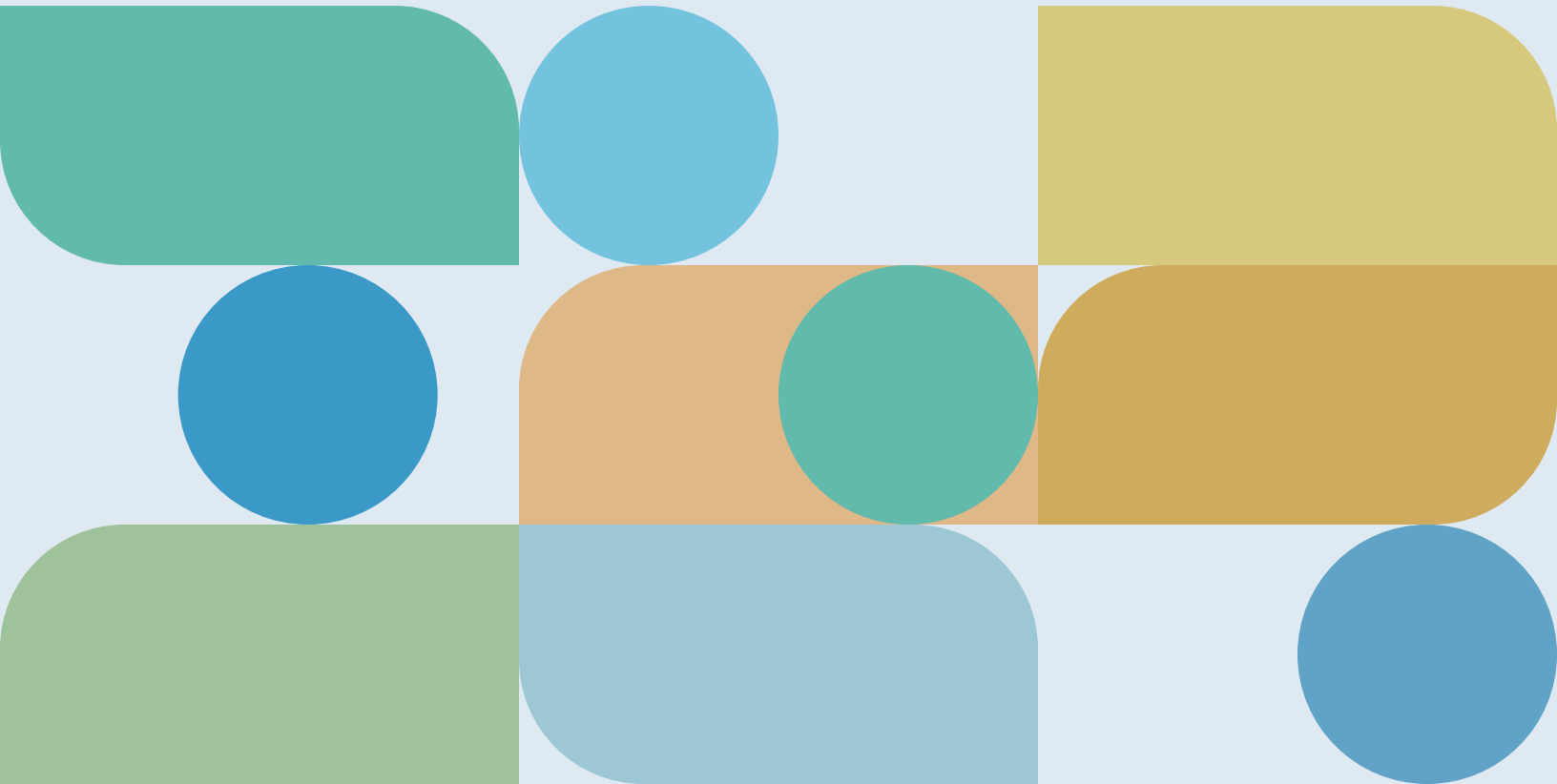
MILKEN INSTITUTE

MARCH 2026

The Power of Real-World Insights in a Modern Clinical Research Enterprise

KRISTIN SCHNEEMAN AND SARAH SHEEHAN



ABOUT THE MILKEN INSTITUTE

The Milken Institute is a nonprofit, nonpartisan think tank focused on accelerating measurable progress on the path to a meaningful life. With a focus on financial, physical, mental, and environmental health, we bring together the best ideas and innovative resourcing to develop blueprints for tackling some of our most critical global issues through the lens of what's pressing now and what's coming next.

ABOUT MILKEN INSTITUTE HEALTH

Milken Institute Health develops research and programs to advance solutions in biomedical innovation, public health, healthy aging, and food systems.

ABOUT FASTERCURES

The Milken Institute's FasterCures is working to build a system that is effective, efficient, and driven by a clear vision: patient needs above all else. We believe that transformative and life-saving science should be fully realized and deliver better treatments to the people who need them.

TABLE OF CONTENTS

EXECUTIVE SUMMARY

- **Real-world insights (RWI) are transforming clinical research decision-making**. By integrating health-care, consumer, behavioral, and digital data, RWI provide a more complete, continuous view of patients' lives, enabling better trial design, faster execution, and more representative enrollment.
- **RWI complement—but are distinct from—real-world evidence (RWE)**. While RWE supports regulatory and clinical conclusions, RWI inform earlier, operational, and strategic decisions such as feasibility, site selection, recruitment strategy, and risk mitigation that determine trial success.
- **Linked health and consumer data can materially improve recruitment, efficiency, and diversity**. Combining electronic health records (EHRs), pharmacy data, consumer demographics, and social determinants of health reduces screen failures, shortens timelines, and supports enrollment of populations historically underrepresented in research.
- **AI and advanced analytics amplify the value of RWI, but adoption remains limited**. AI-enabled analysis has demonstrated substantial gains in speed, cost, and predictive capability, yet most organizations have not fully implemented these tools due to infrastructure, data quality, trust, and investment barriers.
- **Data quality, interoperability, and trust are the primary constraints to scale**. Inconsistent standards, incomplete data provenance, and complex privacy and consent requirements limit broader adoption, underscoring the need for shared standards, robust governance, and patient-centered consent models.
- **Realizing the full promise of RWI requires coordinated, cross-sector action**. Strategic partnerships among life sciences companies, data providers, health systems, retailers, regulators, and patient organizations are essential to build trusted, interoperable data ecosystems that support modern, patient-centered clinical research.

INTRODUCTION

The clinical research enterprise is undergoing a fundamental transformation as traditional data sources used to inform trial design and conduct are being augmented with many additional types of research-related health and consumer data. RWI represent a paradigm shift from the sole use of data from episodic clinical encounters to inform trial design and conduct toward continuous, comprehensive visibility into the medical, behavioral, social, and economic factors that impact patient participation in research and the scientific and operational success of trials.

The convergence of EHRs, consumer transaction data, wearable device metrics, social determinants of health information, and other types of research-related data creates unprecedented opportunities to improve patient recruitment, optimize trial design, reduce the cost of trial conduct, and proactively enhance trial efficiency. This comprehensive approach to data integration holds promise in helping to bridge critical information gaps that have limited clinical trial efficiency, representativeness, and generalizability to real-world patient populations.

DEFINING REAL-WORLD INSIGHTS

What Are RWI?

RWI are derived from the analysis of research-relevant data collected outside traditional clinical research settings. These insights complement and enrich clinical trial and health-care delivery data by providing more complete picture of patient characteristics, health status, and behaviors. RWI can be sourced from traditional real-world health data as well nontraditional real-world data (RWD).

Traditional RWD

- EHR data
- Claims data
- Pharmacy data
- Pharmacy claims data
- Prescription fill data
- Clinical services data

Nontraditional RWD

- Registry data
- Consumer data
- Consumer geographic and demographic data
- Consumer transaction and financial data
- Consumer-generated data from digital platforms and wearables

Traditional RWD collected during routine clinical care include:

- **EHR data**—digitally stored patient health data (including demographics, diagnoses, medications, laboratory results, imaging reports, genetic data, and clinical notes) collected during routine clinical care. These data can enhance clinical trial conduct by enabling more efficient patient identification and recruitment through automated screening of large patient populations based on specific inclusion and exclusion criteria.[1] Additionally, EHR data can support trial operations by providing baseline patient characteristics, reducing data collection burden by supporting automated extraction of relevant clinical measurements, and enabling long-term safety monitoring and outcomes research.[2]
- **Claims data**—derived from administrative records generated when providers submit billing information to insurance companies or government payers. These data typically include patient demographics, procedure and diagnostic codes (e.g., ICD-10, codes), service dates, provider types, and costs, but often lack detailed clinical information such as lab results or disease severity information. Claims data can support clinical trial conduct by enabling trial feasibility assessments and site selection through analysis of disease prevalence across diverse populations and geographic regions.[3]
- **Pharmacy data**
    - **Pharmacy claims data**—generated when a prescription is filled and submitted for payment to an insurer or other payer. Claims data include drug name, dosage form, drug strength, fill date, days of supply, financial information, and de-identified patient and prescriber codes, facilitating longitudinal tracking of medication safety and concordance patterns.[4]
    - **Prescription fill data**—generated when a prescription is filled, independent of when a claim is filed with an insurer or another payer. Prescription fill data include drug name and strength, dosage form, fill date, days of supply, financial information, and de-identified patient and prescriber codes.

Most importantly, the data capture medication refill patterns, including for patients who are uninsured or pay out of pocket for their prescriptions.[5]

- **Clinical services data**—generated when patients receive basic clinical services at retail clinics. This might include data from basic labs and diagnostic tests or participation in chronic disease management programs (e.g., weight management programs) and notes from clinical visits. These data can support patient identification, outreach, and engagement for clinical research.[6]

Nontraditional RWD often generated outside the context of regular clinical encounters include:

- **Registry data**—organized systems that use observational study methods to collect uniform data, forming longitudinal datasets containing information about individuals with a specific disease or treatment. These datasets often include patient demographics, clinical characteristics, treatment patterns, and outcomes that offer detailed, real-world insights beyond what is captured in claims or EHRs.[7]
- **Consumer data**
  - **Consumer geographic and demographic data**—consumer-generated information on age, sex, geographic location, income, and socioeconomic status that can be used to support trial site selection, flag patients who may be eligible for clinical trials, and understand trial access barriers.[8] Consumer geographic and demographic data help capture information on social determinants of health and environmental factors that shape health, behaviors, access to trials, and trial participation.[9]
  - **Consumer transaction and financial data**—data generated from consumers' health and wellness purchases that may be used to identify health status or behaviors. These data may include purchases of over-the-counter (OTC) medication, supplements, diet-related products, self-care items (e.g., scales), and diagnostic tests (e.g., blood sugar test strips).[10]
  - **Consumer-generated data from digital platforms and wearables**—consumer-input data collected from digital platforms and websites such as apps used by consumers to track their care journeys or insurance and health statuses.[11] Each dataset could help support clinical trial recruitment, track medication concordance, and understand health status and behaviors that may have implications for trial participation. Data generated by wearable smartwatches, fitness trackers, or other connected medical devices can also generate activity and lifestyle information, changes in biological processes that can be used as endpoints in clinical research, and treatment concordance insights.[12]

Who Uses RWI?

RWI can be used in the clinical research enterprise by sponsors, contract research organizations (CROs), patient organizations, and their partners to understand disease burden, treatment concordance patterns, and patient behaviors and experiences. RWI can also be used to support hypothesis-generating observations for clinical research, inform clinical trial recruitment and enrollment, and identify real-world factors that could influence clinical trial access and retention.

RWI and RWE Are Different

Although they may rely on analyzing some of the same RWD sources, RWI and RWE are different. RWE is generated through studies with predefined hypotheses, appropriate statistical methodologies, and controls for confounding factors to support trials and health-care decision-making.[13] RWE is typically derived from RWD collected during routine health-care encounters and **intended to answer specific research questions about treatment effectiveness, safety, or clinical outcomes that can inform regulatory decisions, clinical guidelines, or health-care policies**.[14]

RWI are derived from analysis of RWD collected initially for both health care and non-health care related purposes that can be **applied to enrich decision-making in the health-care context, such as business or scientific decision-making related to clinical trial conduct**.

Although they may be sourced from the same types of data, **RWI and RWE differ in their intended purpose, level of pre-planned statistical analysis, and regulatory acceptance**. RWI are not typically used to support clinical or regulatory decision-making, but might reveal interesting patterns or trends about a patient population that warrant further investigation through structured RWE studies or controlled clinical trials. Table 1 illustrates the specific inputs, outputs of analyses, and end users of RWI and RWE, though there is some overlap, and these lists are not exhaustive.

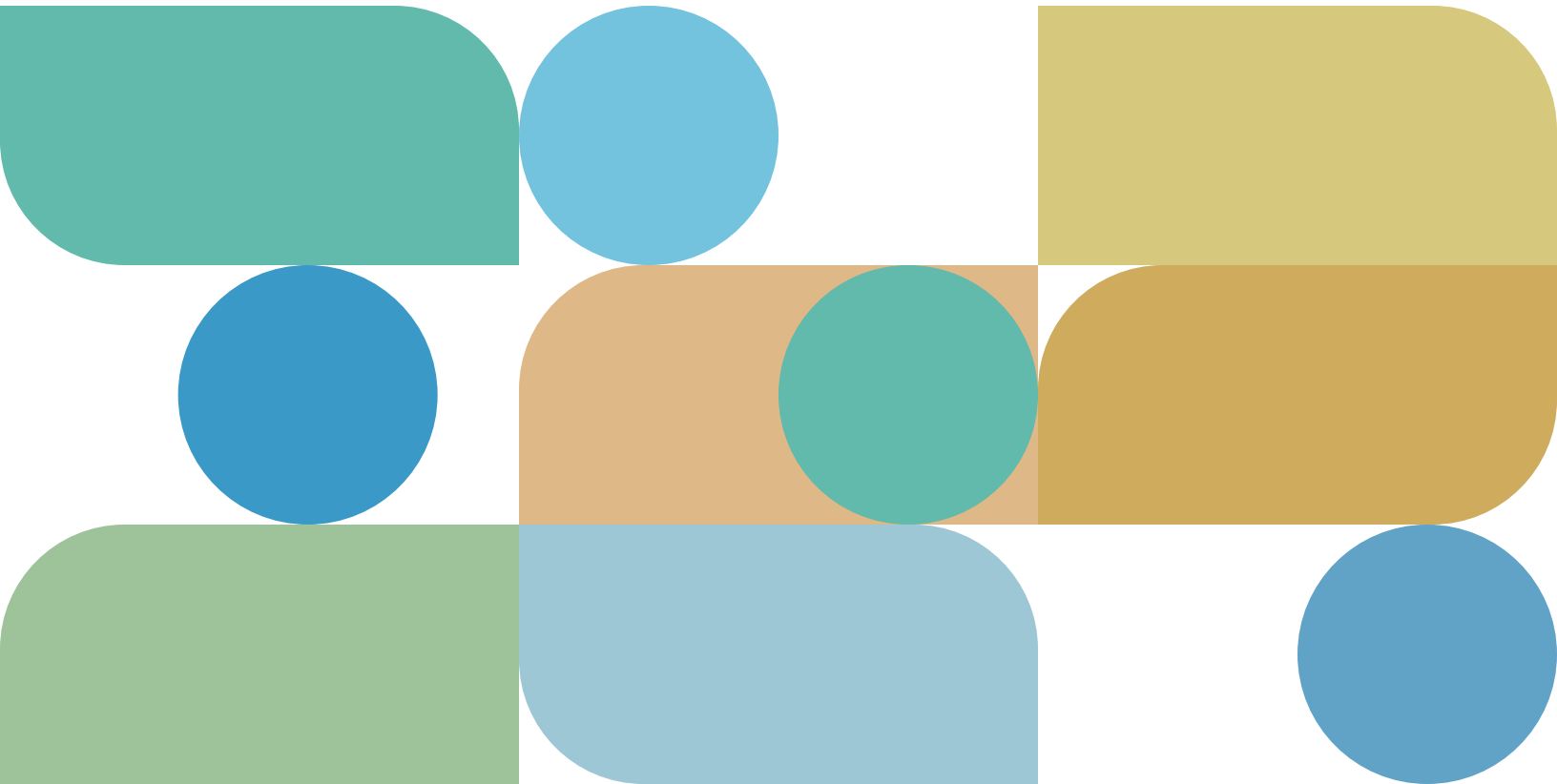


Table 1: Distinguishing Between RWI and RWE as Products of RWD Analysis

	Common Inputs	Common Outputs	End Users
RWI	• Pharmacy claims, prescription fill data • Wearable device and mobile health application data • Social determinants of health data • Consumer transaction and financial data	• Prescribing pattern and medication concordance studies • Trial eligibility screening • Trial feasibility studies • Patient health-care behavior and decision-making studies • Disease burden and risk-prediction studies	• Life sciences companies • Health-care providers • Payers • Patients and patient advocacy organizations • Researchers and academics
RWE	• EHRs • Medical and pharmacy claims data • Patient and disease registry data • Patient-reported outcome measures • Laboratory and diagnostic test results • Genetic data	• Comparative effectiveness research findings • Safety surveillance reports and risk assessments • Treatment pathway analyses • Health economic evaluations (e.g., cost-effectiveness studies) • Clinical practice guidelines and recommendations • Regulatory submissions and post-market surveillance reports • Long-term health outcomes	• Regulatory agencies • Policymakers • Life sciences companies • Health-care providers • Payers • Patients and patient advocacy organizations • Researchers and academics

Source: Milken Institute (2026)

THE VALUE OF RWI FOR CLINICAL RESEARCH

The value proposition presented by using RWI to support clinical evidence generation and trial-related decision-making is immense (see Table 2). Leveraging RWI in clinical research can bridge the information gap generated by relying solely on health-care data collection to support clinical trial recruitment, enrollment, and conduct.

> **RWI—and the data that support them—account for the medical, social, economic, and behavioral realities of patients' daily lives, which impact whether and how they participate in research, as well as how they access and use approved medical products.**

Unlike traditional health-care data that capture only episodic clinical encounters, RWI powered by novel sources of data can provide continuous, comprehensive visibility into factors that impact trial conduct, patient access to trials, and, ultimately, health outcomes. The value proposition of RWI extends beyond operational efficiency in trials to encompass opportunities for data use to enhance patient engagement, accelerate therapeutic access, and advance personalized approaches to treatment. RWI derived from pharmacy, consumer, and other types of data can drive value for the following end-user groups.

Table 2: The Value Proposition of RWI by End-User Group

End Users	Example RWI Application
Life sciences and CROs	De-identified consumer and pharmacy data can enhance trial feasibility assessments, hypothesis generation, patient recruitment strategies, and treatment concordance monitoring.
Data and tokenization vendors	Merging consumer demographic, purchase, and pharmacy data can support the creation of richer, tokenized datasets used to study treatment patterns, adherence, outcomes, and population health trends.
Health systems and payers	Purchase and prescription data can be linked to health-care claims and outcomes data, enabling predictive analytics for treatment concordance in the post-market setting, among other capabilities.
Academic institutions	Health and consumer data can be shared with academic institutions to support important dimensions of research, such as the identification of novel research questions and the validation of findings from other studies.
Patient advocacy and trade organizations	Health and consumer data can provide patient and trade organizations with concrete evidence to support policy positions and funding requests, transforming advocacy with data-driven arguments.

Source: Milken Institute (2026)

THE IMPACT OF RESEARCH-RELEVANT RWI

RWI generation is already transforming traditional trial design and execution using comprehensive patient insights. The current state of play for RWI generation and use can be captured in four broad domains:

1. Enhancing patient identification and recruitment
2. Improving trial design and protocol optimization
3. Driving cost optimization, operational efficiency, and competitive advantage
4. Improving risk mitigation and predictive analytics

Enhancing Patient Identification and Recruitment

Data partners and sponsors can collaborate to identify eligible patients more precisely by combining EHRs with consumer behavior patterns, lifestyle factors, and social determinants of health.[15] This creates a more complete patient profile that goes beyond traditional patient characteristics, allowing for targeted recruitment strategies that can promote representative enrollment and reduce trial accrual timelines.[16]

Technology-powered analysis of RWD is already transforming trial recruitment efficiency due to the increasing regulatory acceptability of AI, as demonstrated in particular by a National Institutes of Health (NIH)-developed AI algorithm that matches potential volunteers with clinical trials.[17] Current platforms enable comprehensive RWD analysis that facilitates quicker, more inclusive recruitment processes by analyzing sociodemographic data alongside clinical information to reflect real-world patient population diversity. For example, one of these platforms used AI-enabled analysis of RWD to enroll eligible patients and reduce trial site activation timelines from the standard eight months to 16 days.[18] The field is moving toward broader adoption, with an expected 14 percent market growth rate for AI-enabled clinical trials over the next eight years.[19]

Improving Trial Design and Protocol Optimization

RWD can provide valuable insights into patient demographics, health status, and disease burden that help sponsors and their data partners design more representative and feasible clinical trial protocols.[20] These insights can be particularly useful for refining inclusion and exclusion criteria by identifying appropriate patient characteristics and comorbidities that ultimately improve patient recruitment and the generalizability of trial results to broader patient populations.[21]

The use of RWI in clinical trial design and feasibility assessments is becoming increasingly sophisticated. RWD, specifically EHR data, are being used to speed the development of compliant trial protocols and support sponsors in identifying the most advantageous clinical trial sites.[22] While specific examples of prospective clinical trials that used RWD to design their eligibility criteria from the outset are not readily documented in the literature, RWD-driven approaches are increasingly being used to evaluate existing eligibility criteria and inform future trial design.[23]

Cost Optimization, Operational Efficiency, and Competitive Advantage

By improving patient selection, reducing screen failures, minimizing dropout rates, and enabling remote monitoring capabilities, the use of RWI can increase trial efficiency and reduce overall trial costs by accelerating timelines.[24] Sponsors can also use RWI generated from combined RWD to support a drug's value proposition and inform their go-to-market strategy during drug development, and facilitate market advantage.[25]

Over the last 10 years, there have been several instances of sponsors leveraging RWD to improve operational efficiencies and shorten enrollment timelines, including using EHR and claims to pre-identify eligible patients and simulate trial feasibility.[26]

> **Recently, automated prescreening processes informed by real-world data have decreased the screening workload by 92 percent and increased efficiency in screening by 450 percent.[27]**

RWD based on epidemiological datasets and registries have also been used to simulate eligibility criteria, reducing screen failures and protocol amendments—one of the most expensive drivers of trial inefficiency.[28]

Risk Mitigation and Predictive Analytics

Leveraging RWI can enable sophisticated predictive modeling to identify patients at risk for adverse events, protocol deviations, or dropout before they occur.[29] This proactive approach allows for intervention strategies that protect patient safety while preserving trial integrity and reducing costly protocol amendments.

RWD is increasingly utilized in clinical trials to predict and mitigate risks, enhancing trial efficiency and patient safety. Existing applications include the use of machine learning models applied to RWD to assist in identifying high-risk patients who may require closer monitoring during the trial.[30] Newer research also indicated that RWD can be used to inform adaptive amendments to a trial protocol based on the accrual of interim data, allowing sponsors to reallocate resources and prioritize less risky interventions without compromising trial rigor.[31]

RWI USE CASES ACROSS THE CLINICAL RESEARCH ENTERPRISE

The use cases below represent applications of real-world insights possible now and in the near term, with efforts to build on pilot programming and to invest in data infrastructure and analysis capabilities that mature our ability to leverage RWI routinely and at scale.

Using Geographic and Demographic Data to Inform Clinical Trial Protocol Development

Analyzing demographic data such as age, sex, race and ethnicity, and socioeconomic indicators that consumers have consented to share can ensure that recruitment goals are feasible across patient subgroups and in alignment with US Food and Drug Administration (FDA) and NIH guidance for diversity in clinical trials. These data have been used in clinical trial feasibility studies to estimate the size of an intended patient population and to prevent unnecessarily narrow trial eligibility criteria, which can exclude portions of an intended population.[32]

Geographic data have also been used to complete an assessment of how burdensome, in terms of travel time and distance, participation in a clinical trial may be for patients.[33] Building capacity for the use of RWI to inform trial feasibility studies may create a pathway for trials designed with decentralized aspects to reduce burden and increase volume and diversity of trial participation.

Early Disease Detection and Trend Forecasting with Purchase History Data to Enable Proactive Enrollment

Analyzing purchase patterns for health-related products (e.g., glucose meters, OTC pain relievers, and nutritional supplements) can help retailers and their clinical research partners flag individuals at higher risk for certain conditions and enable more effective clinical trial screening. Aggregating sales of symptom-related products (e.g., cough medicine and thermometers) can also support efforts to identify geographic hotspots of emerging illnesses or outbreaks, enabling site selection and proactive recruitment for trials in affected areas.

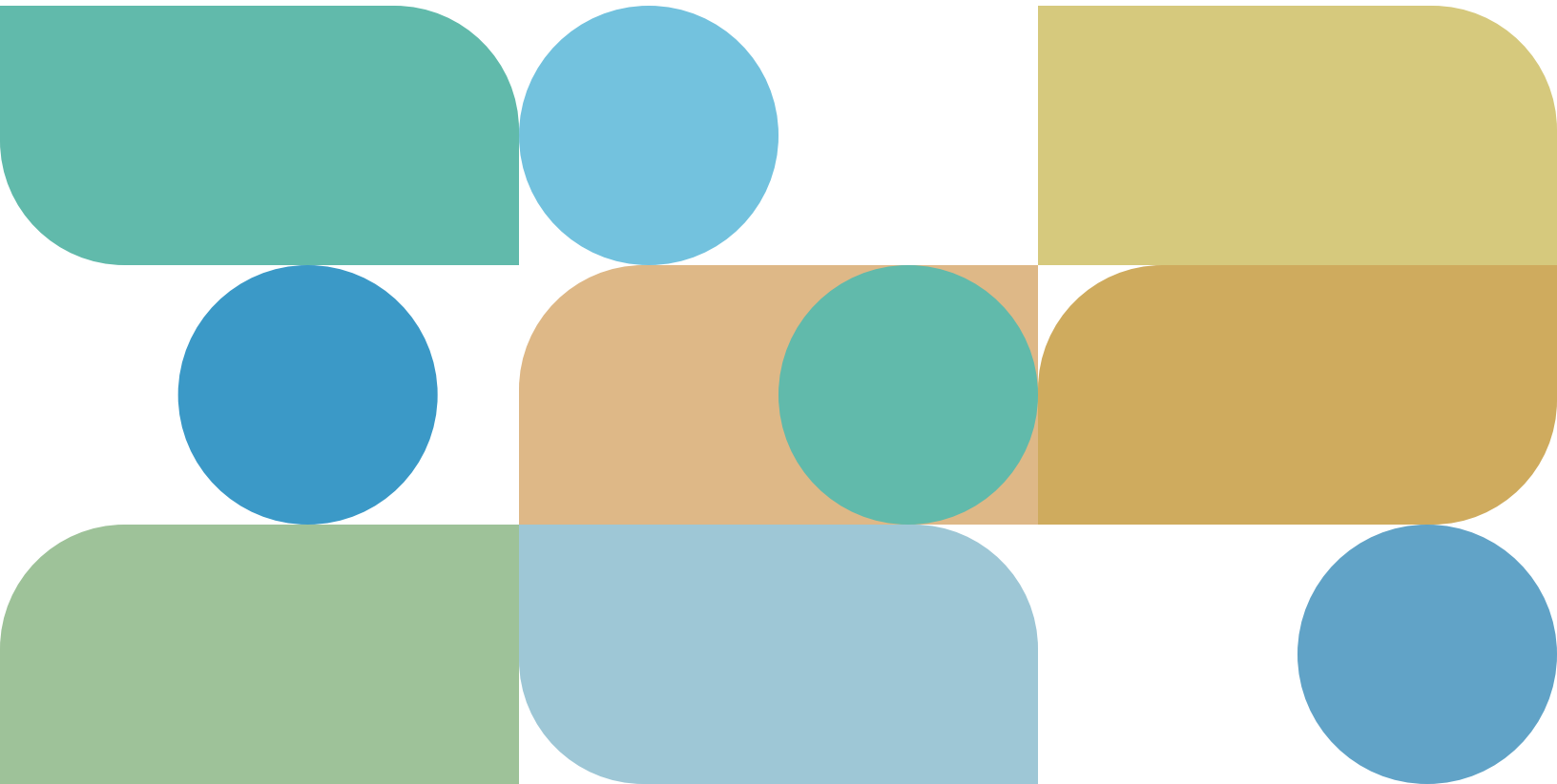
Leveraging Loyalty Card Data to Examine Population Health Outcomes

Loyalty card data, particularly when linked with traditional health data through deterministic or probabilistic matching (e.g., claims, EHR, and pharmacy data), can be used to investigate the impact of diet and consumption trends on health outcomes. These data have also proven valuable for evaluating public health policies (e.g., taxation and legislative reforms), allowing for data analysis from large populations across diverse regions and socioeconomic groups.[34] Use of loyalty card data to inform health research enables real-time and retrospective tracking of factors that may impact population health outcomes and can be useful in assessing household-level patterns and investigating how sociodemographic differences may factor into health care access, consumption trends, or health outcomes.

Predicting and Monitoring Real-World Treatment Concordance in the Post-market Setting

Using retail pharmacy purchase data to monitor whether patients are visiting retail pharmacies, filling prescriptions, or buying ancillary care products can allow researchers to assess the status of and barriers to real-world treatment concordance. Pharmacy data can be used to detect irregularities in prescription fill patterns, offering real-world insights to facilitate patient-level interventions to improve treatment concordance.

Prescription fill data can be linked with clinical outcomes data to inform post-market safety and effectiveness assessments using real-world evidence. These data can also be used to improve traditional approaches to monitoring or predicting concordance, such as the use of claims data or patient self-report, which miss cash payments for prescriptions and have performed poorly in concordance prediction accuracy.[35]



LEVERAGING AI AND MACHINE LEARNING TO IMPROVE RWI GENERATION

The use of artificial intelligence and machine learning (AI/ML) solutions to analyze large volumes of RWD from EHRs, wearables, and non-health-related data streams can help sponsors and their data partners generate and leverage RWI to optimize trial design and conduct.[36] AI and ML algorithms are already helping to accelerate patient recruitment by matching eligible participants to appropriate studies, supporting real-time monitoring of trial safety, and generating synthetic control arms that reduce the need for placebo groups through the creation of virtual patient profiles.[37]

> **Sponsors that effectively leverage RWD and AI/ML solutions to analyze RWD in their trials can save up to 60 percent on drug development costs and bring products to market 30 percent faster.[38]**

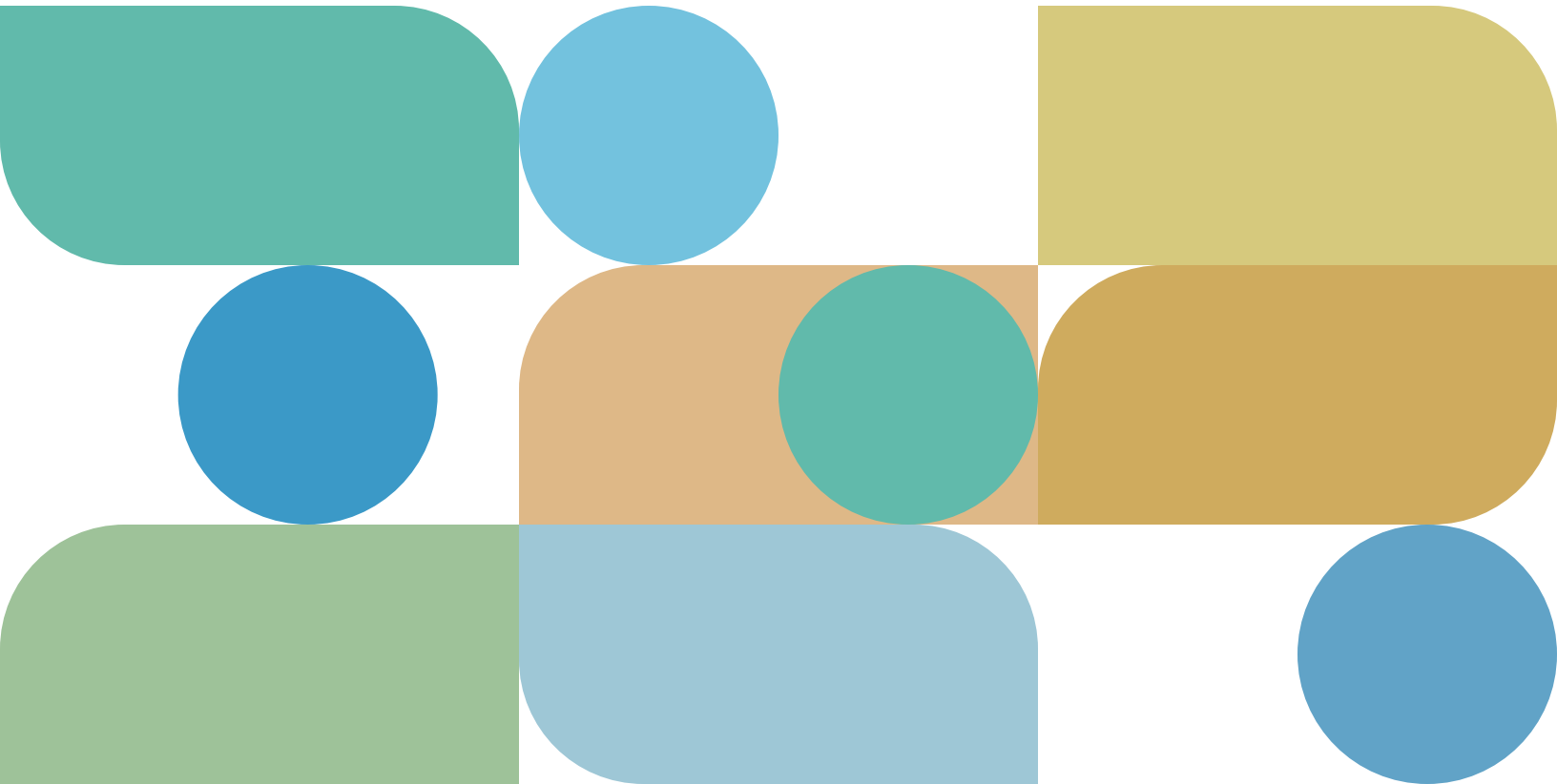
However, as of late 2024, only a small percentage of companies—11 percent of nearly 80 pharmaceutical companies and CROs surveyed in a Tufts study—reported fully implementing AI/ML to enable clinical trial activities. An additional 22 percent of companies reported partially implementing AI/ML solutions. These companies described 36 case examples of AI/ML use in clinical trials: 33 percent of those use cases featured AI-enabled activities in clinical trial design and planning; 56 percent of the cases covered AI-enabled activities in trial conduct; and 11 percent featured AI-enabled activities to support regulatory submissions. While AI implementation has grown to support the analysis of RWD in clinical research, its use seems to have grown almost exclusively within larger companies with the resources to build AI infrastructure and a larger number of clinical trials to leverage that infrastructure. Significant investment in AI/ML capabilities and technical infrastructure is needed to advance the routine use of the solutions in RWI generation that supports business and scientific decision-making related to clinical research. On average, companies in the Tufts study reported an investment of $3.3 million–$4.4 million to implement AI/ML-enabled solutions to support clinical trial conduct.[39]

While meaningful use cases demonstrate the utility of AI and ML solutions in RWD analysis to inform site selection and reduce risk in trial conduct, insufficient trust in training data and persistent data harmonization challenges limit scalability (see Table 3).[40]

Table 3: Challenges to Scaled Implementation of AI/ML Solutions in Clinical Research

Challenge	Description
Data quality and standardization	Varying data collection standards and coding systems across data partners and types, missing data, and incomplete documentation of data provenance impede data transformation processes.[41]
Trust and generalizability	Lack of trust in the quality of AI model training data and data governance procedures, as well as underreporting of demographic characteristics, exacerbate algorithmic bias and reduce generalizability.[42]
Technical infrastructure	Challenges exist with integrating workflows and IT systems supporting trial conduct with modern AI/ML platforms and insufficient computational resources for complex analyses.[43]
Resource investment	Significant capital and workforce investment is required for the development and ongoing maintenance of AI/ML infrastructure to support its use in trial design and conduct.[44]

Source: Milken Institute (2026)



REGULATION OF RWI IN CLINICAL RESEARCH

Ensuring Privacy, Security, and Trust for Patient Data Used to Support RWI

Unique privacy, security, and trust challenges arise related to the use of RWI supported by data collected initially for not only health-related purposes, but also data from federated datasets comprising consumer health and other consumer information. Organizations interested in partnering to share data that power RWI should prioritize implementing rigorous informed consent, data governance, and infrastructure security processes.

Importance of Addressing Challenges to Informed Consent

RWI generation presents unique challenges for informed consent processes that differ from traditional clinical trial consent procedures.[45] The informed consent process to support RWI generation must address these challenges and keep pace with the evolving landscape of data use, in which health information may be combined with consumer data, social determinants of health information, and digital health metrics to generate comprehensive RWI.

Retrospective analysis of previously collected data, regardless of the purpose or time frame during which the data were collected, is the backbone of RWI generation. This generation requires the development of broad consent frameworks, such as those being implemented by sites in NIH's Clinical and Translational Science Awards (CTSA) Program Network, that can accommodate future research activities while still providing patients and consumers with meaningful information about how their data might be used for research purposes.[46]

Broad-Scale Informed Consent

The Common Rule, which governs informed consent for research, was revised in 2017 to enable emerging, data-driven types of research. The rule states that "broad consent for the storage, maintenance and secondary research use of identifiable private information … is permitted as an alternative to the informed consent requirements."[47] Broad-scale informed consent allows patients to give consent in advance to be re-contacted, and for their health data and biospecimens to be used for future research. The majority of federally funded CTSA institutions are pursuing it for at least one of these purposes.

Success depends on leadership support, strong IT systems, and clear patient and staff education. The biggest hurdles are integrating consent into routine workflows, aligning IT systems, and balancing research with other priorities. And while most patients express support, willingness is lower among racial and ethnic minorities and individuals with less formal education, underscoring the need for trust-building and inclusive communication.

Integrating consumer health and financial data at scale with other datasets to generate RWI presents complex logistics and data reuse challenges, particularly in securing informed consent.

When health data are combined with data produced outside of conventional clinical or academic settings or collected through devices or digital platforms, informed consent processes that mimic those seen in health research are often not feasible or available.[48] Individually administered informed consent processes might be operationally infeasible, and organizations must take other meaningful steps to ensure patients are informed of how their individual data contribute to future research outcomes or how privacy protection techniques such as de-identification and statistical disclosure control affect their privacy risks.

Organizations may use one of several approaches to broad-scale consent, including an authorization model that allows participants to choose in advance how their data are used and how they prefer to be contacted by study administrators related to data reuse.[49] Approaches to broad-scale consent for data reuse also include dynamic consent, in which participants manage their consent preferences on an ongoing basis through a digital platform that allows for two-way communication between participants and researchers.[50]

Characterizing the Data Protection Landscape for RWI Generation

Regulatory and data security requirements for stakeholders supplying data to support clinical research are comprehensive and binding. There are several federal and state regulations that data stewards should be aware of, especially to support the use of merged health and consumer data to support business and scientific decision-making as part of clinical trials (see Table 4). About a dozen states have either standalone health data privacy laws or general consumer data privacy laws that specifically regulate health and biometric data outside the Health Insurance Portability and Accountability Act (HIPAA). There have also been attempts to standardize consumer health data laws at the federal level, including the American Data Privacy and Protection Act of 2022 and American Privacy Rights Act of 2024. However, neither bill has been passed into law.

Table 4: Federal and State Regulations Governing Health and Consumer Data in Clinical Trials

Regulation	Clinical Research Impact	Key RWI-Related Provisions
Federal Health Data Regulations		
HIPAA	Controls the use of protected health information (PHI) from covered entities such as health-care providers, health plans, and clearinghouses	• Authorization for use and disclosure of PHI • De-identification standards • Requirement to make reasonable efforts to limit the use of PHI to the bare minimum needed to accomplish the intended purpose
FDA Regulations (21 CFR Parts 50, 56, 312, 812)	Governs how data used in clinical trials must be collected, managed, and reported	• Data integrity requirements such as source data verification, retention, and data provenance information
The Common Rule (45 CFR 46)	Outlines data sharing requirements, among other requirements, with binding applications for federally funded clinical research	• Data reuse exemptions for secondary research • Privacy and confidentiality protections • Informed consent disclosure requirements
Federal Consumer Data Regulations		
Federal Trade Commission Act (15 USC §45)	Governs consumer data collection and use practices, including health-related consumer data used in clinical research	• Privacy policy compliance measures • Reasonable data security measures • Consumer consent for material changes to privacy practices, data use policies, or terms of service
Electronic Communications Privacy Act	Controls access to digital communications and stored electronic data, including data used in clinical research	• Restrictions on access to stored digital communications • Limits on disclosure of consumer information to third parties and requirements for consumer notice of disclosure

Selected State Health and Consumer Data Regulations		
California Consumer Privacy Act	Governs the personal information of California residents, including personal information used in clinical research	• Broad definition of personal information in act affects research data use • Right to know, delete, correct, and opt out • Data minimization and purpose limitation • Sensitive personal information protections • Third-party disclosure requirements
California Confidentiality of Medical Information Act	Additional layer of protection beyond HIPAA for California health data, including data used in clinical research	• Patient authorization for disclosure • Stricter than HIPAA in some areas • Specific breach notification requirements
Texas Medical Privacy Act	Additional consent requirements for Texas health data, including health data used in clinical research	• Patient authorization for disclosure • Stricter consent requirements than HIPAA • Prohibition on PHI sales • Narrower de-identification standards than HIPAA • Higher bar for institutional review board (IRB) waivers • Additional consent requirements and fewer research activities permitted without patient consent, as compared to HIPAA
Washington State My Health My Data Act (similar bills passed in NV, CT, and MD)	Protection of personal health data that falls outside the ambit of HIPAA; impacts all organizations that conduct business in Washington and that collect, process, share, or sell consumer health data implicated by the act	• Covers data used to make inferences about a consumer's health status • Requires organizational disclosures and consumer consent for the collection, sharing, and use of health information • Prohibits the sale of consumer health data without obtaining signed consumer authorization

Source: Milken Institute (2026)

Addressing Real and Perceived Data Quality, Standardization, and Interoperability Issues

Common Data Quality Issues

If unaddressed, several data quality challenges can compromise sponsors' and their data partners' ability to use RWI based on health and consumer health data in support of clinical trial conduct. Common RWD quality issues include:

- **Inconsistent data definitions**—When different data sources use distinct definitions for the same health concept, combining sources for analysis can impact reliability.[51]
- **Missing data**—Missing values for variables of interest can introduce bias and impact the reliability of conclusions drawn from RWD analysis to support decision-making.[52]
- **Temporal misalignment**—RWD is often collected for non-research purposes from diverse patient populations over long or variable periods of time, which can introduce bias when time frames of interest are misaligned; this underscores the importance of close temporal alignment across datasets.[53]
- **Gaps in data validation**—Inadequate validation of data at source, uncertainties around data provenance, and lack of audit trails make it hard to assess and promote RWD reliability.[54]

Approaches to Data Standardization and Supporting Technical, Semantic, and Cross-Organization Interoperability

Health information exists across disparate systems with inconsistent formats, terminologies, and quality standards, while consumer data from retailers, digital platforms, and wearables may follow entirely different structural and semantic conventions. Without comprehensive data standardization frameworks, attempts to integrate these diverse data sources can result in analytical challenges, reduced data quality, and limited research utility that undermines the potential value of real-world insights.[55]

The Health Level 7 Fast Healthcare Interoperability Resources (FHIR) standardization framework covers the exchange of clinical health data and consumer-generated health data. It provides the most comprehensive framework currently available for standardizing the combination of traditional health data with health-related consumer data across multiple domains and use cases.[56]

Data standardization is the chief enabler for interoperability, ensuring that health and health-related information collected across different contexts can be meaningfully combined, analyzed, and applied to inform clinical trial decision-making. The successful integration of RWI into clinical trials requires comprehensive interoperability across three critical dimensions:

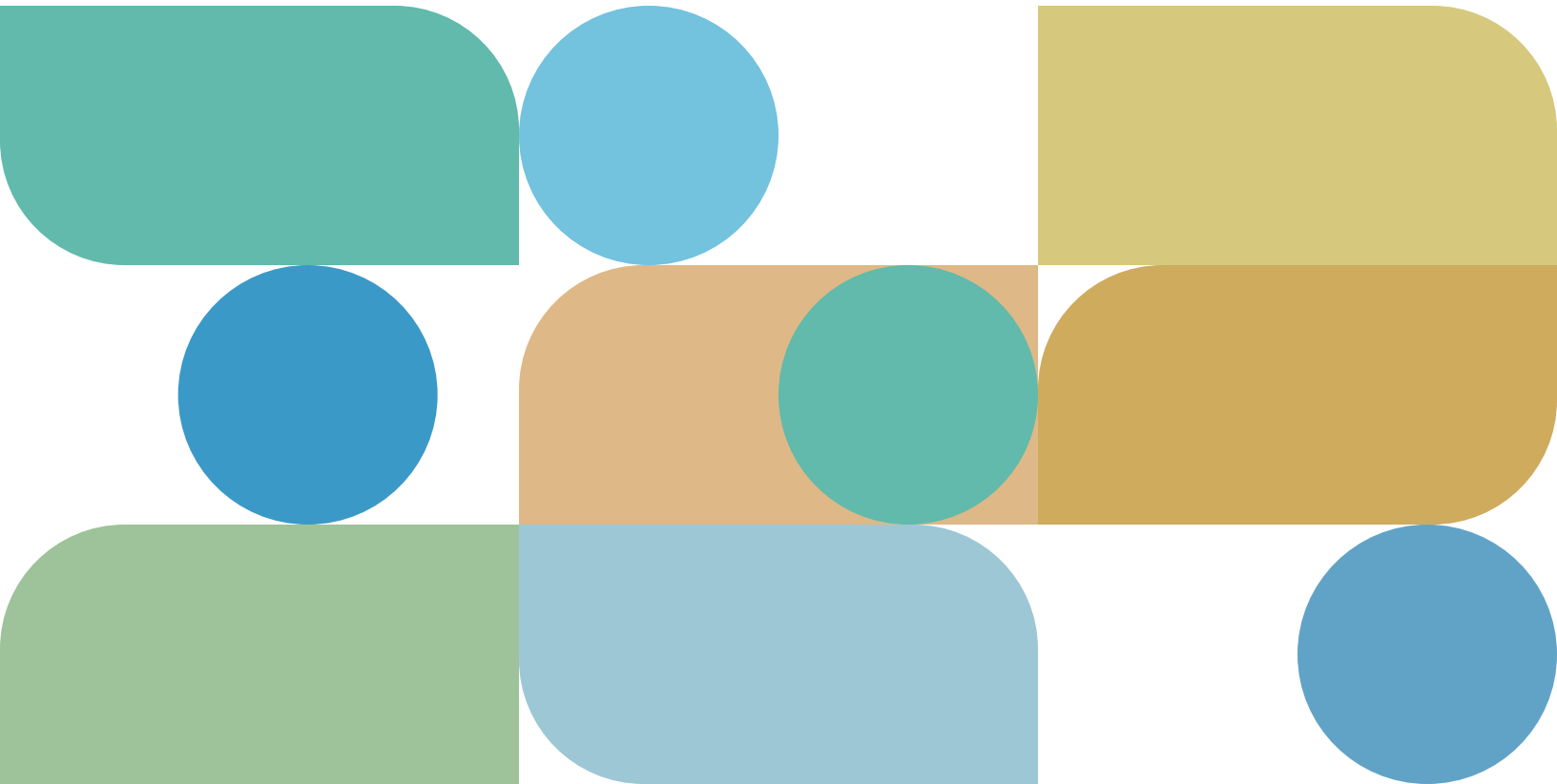
- Technical interoperability through standardized application programming interfaces (APIs) and data exchange protocols that enable cross-platform communication
- Semantic interoperability via common vocabularies and data models that ensure consistent interpretation of clinical and consumer data elements across diverse sources

- Organizational interoperability through aligned governance frameworks and collaborative processes that enable data sharing while maintaining compliance with privacy regulations and quality standards

Technical interoperability is achieved by implementing standardized APIs and data exchange protocols such as FHIR or Observational Medical Outcomes Partnership common data standards to enable seamless, secure communication between disparate health-care systems, consumer platforms, and research databases.[57]

Developing semantic interoperability centers on establishing common vocabularies, terminologies, and ontologies essential to enable meaningful data interpretation across different domains.[58] This includes mapping between clinical terminologies (e.g., ICD-10 and SNOMED-CT) and consumer data categories, as well as developing crosswalks between different coding systems used by various data partners, no matter the type of data supplied, to support trial start-up or conduct.[59]

This kind of interoperability can be codified within and across organizations through the creation of unified health data governance structures (e.g., Trusted Exchange Framework and Common Agreement) that enable effective collaboration among health-care partners, technology vendors, retailers, and research institutions.[60] Other steps toward cross-organizational interoperability include standardizing data use agreements, establishing common consent frameworks, and developing shared quality assurance processes.



CONCLUSION

The scaled use of RWI generated outside of traditional health-care settings presents a transformative opportunity to enhance the design, conduct, and impact of clinical trials. By reliably linking traditional health-care data with rich consumer information, RWI can improve patient identification and recruitment, optimize protocol design, reduce operating costs, and accelerate trial timelines.

> **These insights enable a more continuous and comprehensive understanding of the patient journey and offer pathways to representative enrollment, generalizable evidence, and quicker therapeutic access.**

Realizing the full promise of RWI within the clinical research enterprise requires that we address critical data-related challenges and appropriately accommodate regulatory and ethical protections related to the use of merged health and consumer data. The increased use of AI and ML presents solutions to the operational challenges, but data quality, standardization, and interoperability still pose significant barriers to the use of combined data in research. Similarly, the issues of trust and the need for transparent and effective approaches to data use consent must also be addressed to advance the use of RWI.

Moving forward, solutions will be tied to coordinated action on data quality and meaningful steps to ensure public trust and ethical data use. Data standardization and exchange frameworks should be more broadly adopted to enable interoperability across diverse data streams. Investments in a robust federated infrastructure, data governance practices, and transparent, patient-centered consent models—including dynamic consent approaches—are necessary to ensure both ethical integrity and public trust around data use and reuse. Strategic partnerships among life sciences companies, data vendors, health systems, and patient advocacy groups will also be critical in building sustainable pipelines for comprehensive, representative data that are useful and usable in clinical research.

RWI have the potential to reshape clinical research by making it more efficient, representative, and responsive to patient needs. By proactively addressing the technical, operational, and ethical challenges outlined, partners in the clinical research ecosystem can unlock the potential of RWI to drive a more modern, patient-centered, and impactful research enterprise.

ENDNOTES

1. Martin R. Cowie et al., "Electronic Health Records to Facilitate Clinical Research," *Clinical Research in Cardiology* 106, no. 1 (2017): 1–9, https://doi.org/10.1007/s00392-016-1025-6; Jeffrey Kirshner et al., "Automated Electronic Health Record–Based Tool for Identification of Patients with Metastatic Disease to Facilitate Clinical Trial Patient Ascertainment," *JCO Clinical Cancer Informatics* 5 (July 1, 2021): 719–27, https://doi.org/10.1200/CCI.20.00180.

2. Kimberly A. McCord and Lars G. Hemkens, "Using Electronic Health Records for Clinical Trials: Where Do We Stand and Where Can We Go?" *Canadian Medical Association Journal* 191, no. 5 (February 4, 2019): E128–E133, https://doi.org/10.1503/cmaj.180841; Alba Gutiérrez-Sacristán et al., "Development and Validation of an Open-Source Pipeline for Automatic Population of Case Report Forms from Electronic Health Records: A Pediatric Multi-Center Prospective Study," *eBioMedicine* 108 (October 2024), https://pubmed.ncbi.nlm.nih.gov/39288532/; Alison Callahan et al., "Medical Device Surveillance with Electronic Health Records," *npj Digital Medicine 2*, no. 1 (September 25, 2019): 94, https://doi.org/10.1038/s41746-019-0168-z.

3. Hua Chen et al., "Demographic Clinical Trial Diversity Assessment Methods: Use of Real-World Data," *Contemporary Clinical Trials Communications* 44 (April 2025): 101432, https://doi.org/10.1016/j.conctc.2025.101432.

4. Constance P. Fontanet et al., "Comparison of Measures of Medication Adherence from Pharmacy Dispensing and Insurer Claims Data," *Health Services Research 57*, no. 3 (August 13, 2021): 524–36, https://doi.org/10.1111/1475-6773.13714.

5. Alexander Galozy and Slawomir Nowaczyk, "Prediction and Pattern Analysis of Medication Refill Adherence Through Electronic Health Records and Dispensation Data," *Journal of Biomedical Informatics* 112 (2020): 100075, https://doi.org/10.1016/j.yjbinx.2020.100075.

6. Ed Miseta, "Walmart's Healthcare Research Institute Is Ready to Aid Underserved Communities," *Clinical Leader*, October 26, 2022, https://www.clinicalleader.com/doc/walmart-s-healthcare-research-institute-is-ready-to-aid-underserved-communities-0001.

7. Lars Hulstaert et al., "Enhancing Site Selection Strategies in Clinical Trial Recruitment Using Real-World Data Modeling," *PLOS One 19*, no. 3 (March 11, 2024): e0300109, https://doi.org/10.1371/journal.pone.0300109.

8. *Feasibility Decision Tree: A Tool to Prioritize the Recruitment of a Representative Population During Site Selection*, MRCT Center Diversity Toolkit, version 1.1 (MRCT Center, December 2022), https://mrctcenter.org/wp-content/uploads/2022/12/12-Feasibility-Decision-Tree-Updated.pdf.

9. Sara King et al., "Leveraging Social Determinants of Health to Enhance Recruitment of Underrepresented Populations in Clinical Trials," *Methodist DeBakey Cardiovascular Journal* 20, no. 5 (November 5, 2024): 81–88, https://doi.org/10.14797/mdcvj.1447.

10. Alisha Suhag et al., "Shopping Data for Population Health Surveillance: Opportunities, Challenges, and Future Directions," *Journal of Medical Internet Research* 27, no. 1 (April 9, 2025): e75720, https://doi.org/10.2196/75720; Hannah R. Brewer et al., "Association Between Purchase of Over-the-Counter Medications and Ovarian Cancer Diagnosis in the Cancer Loyalty Card Study (CLOCS): Observational Case-Control Study," *JMIR Public Health and Surveillance* 9, no. 1 (January 26, 2023): e41762, https://doi.org/10.2196/41762.

11. Gregory Daniel et al., *Mobilizing mHealth Innovation for Real-World Evidence Generation: An Action Plan* (Duke-Margolis Center for Health Policy, March 2020), https://healthpolicy.duke.edu/sites/default/files/2020-03/duke-margolis_mhealth_action_plan.pdf.

12. Caroline Marra et al., "Quantifying the Use of Connected Digital Products in Clinical Research," *npj Digital Medicine* 3 (April 3, 2020): 50, https://doi.org/10.1038/s41746-020-0259-x.

13. "Advancing Real-World Evidence Program," Food and Drug Administration, accessed September 9, 2025, https://www.fda.gov/drugs/development-resources/advancing-real-world-evidence-program; EUPATI, "Real-World Data (RWD) and Real-World Evidence (RWE)," *EUPATI Toolbox*, accessed September 9, 2025, https://toolbox.eupati.eu/resources/patient-toolbox/real-world-data-rwd-real-world-evidence-rwe/.

14. "Real-World Evidence," Food and Drug Administration, accessed September 9, 2025, https://www.fda.gov/science-research/science-and-research-special-topics/real-world-evidence; Valton Costa et al., "The Relevance of the Real-World Evidence in Research, Clinical, and Regulatory Decision Making," *Frontiers in Public Health* 13 (February 17, 2025), https://doi.org/10.3389/fpubh.2025.1512429.

15. Chenyu Li et al., "Realizing the Potential of Social Determinants Data in EHR Systems: A Scoping Review of Approaches for Screening, Linkage, Extraction, Analysis, and Interventions," *Journal of Clinical and Translational Science* 8, no. 1 (October 10, 2024): e147, https://doi.org/10.1017/cts.2024.571; Kevin Lybarger et al., "Leveraging Natural Language Processing to Augment Structured Social Determinants of Health Data in the Electronic Health Record," arXiv preprint, April 14, 2023, https://doi.org/10.48550/arXiv.2212.07538.

16. Yingcheng Sun et al., "A Framework for Systematic Assessment of Clinical Trial Population Representativeness Using Electronic Health Records Data," *Applied Clinical Informatics* 12, no. 4 (September 8, 2021): 816–825, https://doi.org/10.1055/s-0041-1733846.

17. "NIH-Developed AI Algorithm Matches Potential Volunteers to Clinical Trials," *National Institutes of Health*, November 18, 2024, https://www.nih.gov/news-events/news-releases/nih-developed-ai-algorithm-matches-potential-volunteers-clinical-trials.

18. Samantha Mallahan, "A Technology-Enabled Clinical Trial Program's Impact on Patient Screening and Trial Enrollment in 2024," *ASCO*, accessed September 12, 2025, https://www.asco.org/abstracts-presentations/252988.

19. "AI in Clinical Trials Market Size and Share, 2024 - 2032", Global Market Insights Inc., accessed September 12, 2025, https://www.gminsights.com/industry-analysis/ai-in-clinical-trials-market.

20. Jennifer H. Benbow et al., "Increasing Inclusiveness of Patient-Centric Clinical Evidence Generation in Oncology: Real-World Data and Clinical Trials," *American Society of Clinical Oncology Educational Book* 42 (May 13, 2022): 116-126, https://doi.org/10.1200/EDBK_350574.

21. "Real-World Data Meets Clinical Trial Design: Optimizing Protocols and Site Selection," *Fierce Biotech*, October 21, 2024, https://www.fiercebiotech.com/sponsored/real-world-data-meets-clinical-trial-design-optimizing-protocols-and-site-selection; Anatol-Fiete Näher et al., "Generalizability in Real-World Trials," *Clinical and Translational Science* 17, no. 7 (July 24, 2024): e13886, https://doi.org/10.1111/cts.13886.

22. Simon Dagenais et al., "Use of Real-World Evidence to Drive Drug Development Strategy and Inform Clinical Trial Design," *American Society for Clinical Pharmacology & Therapeutics*, (November 28, 2021), https://ascpt.onlinelibrary.wiley.com/doi/10.1002/cpt.2480; "New Oracle Cloud Services Help Pharmas Accelerate Clinical Trial Site Feasibility Assessment and Patient Recruitment," Oracle Newsroom, October 29, 2024, https://www.oracle.com/news/announcement/new-oracle-cloud-services-help-pharmas-accelerate-clinical-trial-site-feasibility-assessment-and-patient-recruitment-2024-10-29.

23. Ruishan Liu et al., "Evaluating Eligibility Criteria of Oncology Trials Using Real-World Data and AI," *Nature 592*, no. 7855 (April 7, 2021): 629–633, https://doi.org/10.1038/s41586-021-03430-5.

24. Scott R. Evans et al., "Real-World Data for Planning Eligibility Criteria and Enhancing Recruitment: Recommendations from the Clinical Trials Transformation Initiative," *Therapeutic Innovation and Regulatory Science* 55, no. 3 (January 3, 2021): 545–552, https://doi.org/10.1007/s43441-020-00248-7; Harsha K Rajasimha, "The Advent of Patient-Centric Technologies to Combat the High Dropout Rates: Revolutionizing Clinical Trials," *Telehealth and Medicine Today* 9, no. 2 (April 30, 2024), https://doi.org/10.30953/thmt.v9.474; Zhiying Fu et al., "Reducing Clinical Trial Monitoring Resources and Costs with Remote Monitoring," *Journal of Medical Internet Research* 25, no. 1 (June 27, 2023): e42175, https://doi.org/10.2196/42175; Sarfaraz K. Niazi, "Advice to the FDA to Improve Its Proposed Guidelines," *Pharmaceuticals* 17, no. 11 (October 24, 2024): 1424, https://doi.org/10.3390/ph17111424.

25. Ritupriya Yamujala, "Is Lab Data the Secret Weapon for More Efficient Clinical Trials?" *MMIT Network*, December 8, 2022, https://www.mmitnetwork.com/thought-leadership/is-lab-data-the-secret-weapon-for-more-efficient-clinical-trials.

26. Robert J. Stanton et al., "Using Epidemiological Data to Inform Clinical Trial Feasibility Assessments: A Case Study," *Stroke* 54, no. 4 (February 28, 2023): 1009–1014, https://doi.org/10.1161/STROKEAHA.122.041650; Jeffrey R. Curtis et al., "Use of Health Plan Combined with Registry Data to Predict Clinical Trial Recruitment," *Clinical Trials* 11, no. 1 (December 17, 2013): 96–101, https://doi.org/10.1177/1740774513512185.

27. Yizhao Ni et al., "Automated Clinical Trial Eligibility Prescreening: Increasing the Efficiency of Patient Identification for Clinical Trials in the Emergency Department," *Journal of the American Medical Informatics Association* 22, no. 1 (July 16, 2014): 166–178, https://doi.org/10.1136/amiajnl-2014-002887.

28. Stanton, "Using Epidemiological Data to Inform Clinical Trial Feasibility Assessments: A Case Study."

29. "Revolutionizing Clinical Trials: How AI Is Improving Clinical Trial Retention," *OmniScience Insights*, accessed September 10, 2025, https://www.omniscience.bio/news-insights/revolutionizing-clinical-trials-how-ai-is-improving-clinical-trial-retention; Emmette Hutchison et al., "Modeling Clinical Trial Attrition Using Machine Intelligence," medRxiv, November 14, 2021, https://www.medrxiv.org/content/10.1101/2021.11.12.21266277v1.full; Anu Patel et al., "Predictive Modeling of Drug-Related Adverse Events with Real-World Data," *Clinical Pharmacology & Therapeutics* 115, no. 4 (February 12, 2024): 847–859, https://doi.org/10.1002/cpt.3201.

30. Lorinda Coombs et al., "A Machine Learning Framework Supporting Prospective Clinical Decisions Applied to Risk Prediction in Oncology" *npj Digital Medicine* 5, no. 1 (August 16, 2022): 117, https://doi.org/10.1038/s41746-022-00660-3.

31. Keer Chen et al., "A Hybrid Prior Bayesian Method for Combining Domestic Real-World Data and Overseas Data in Global Drug Development," ResearchGate, accessed February 12, 2026, https://doi.org/10.48550/arXiv.2505.12308.

32. Stanton et al., "Using Epidemiological Data to Inform Clinical Trial Feasibility Assessments."

33. Julia E. Clougherty et al., "Geography, Generalisability, and Susceptibility in Clinical Trials," *The Lancet Respiratory Medicine* 9, no. 4 (April 2021): 330–332, https://doi.org/10.1016/S2213-2600(21)00046-1.

34. Romana Burgess et al., "Exploring the Use of Supermarket Loyalty Card Data in Health Research," *Public Health* 247 (October 2025): 105848, https://doi.org/10.1016/j.puhe.2025.105848.

35. Alexis A. Krumme et al., "Can Purchasing Information Be Used to Predict Adherence to Cardiovascular Medications?" *BMJ Open* 6, no. 11 (November 2016), https://doi.org/10.1136/bmjopen-2015-011015.

36. Burgess, "Exploring the Use of Supermarket Loyalty Card Data in Health Research."

37. "NIH-Developed AI Algorithm Matches Potential Volunteers to Clinical Trials," National Institutes of Health, November 18, 2024, https://www.nih.gov/news-events/news-releases/nih-developed-ai-algorithm-matches-potential-volunteers-clinical-trials; Rabie Adel El Arab et al., "Bridging the Gap: From AI Success in Clinical Trials to Real-World Healthcare Implementation," *Healthcare* 13, no. 7 (March 22, 2025): 701, https://doi.org/10.3390/healthcare13070701; "The Pivotal Role of AI in Clinical Trials," *BiopharmaTrend*, October 7, 2025, https://www.biopharmatrend.com/artificial-intelligence/the-pivotal-role-of-ai-in-clinical-trials-from-digital-twins-to-synthetic-control-arms-176/.

38. Asif Dhar, "Overcoming Generative AI Implementation Blind Spots in Health Care," Deloitte Center for Health Solutions, January 30, 2024, https://www.deloitte.com/us/en/insights/industry/health-care/how-to-prepare-for-generative-ai-in-health-care.html.

39. Ken Getz, "New Insights on the Impact of AI-Enabled Solutions," *Applied Clinical Trials Online*, 34, no. 3 (June 11, 2025), https://www.appliedclinicaltrialsonline.com/view/new-insights-on-the-impact-of-ai-enabled-solutions.

40. Moe Alsumidaie, "How Novartis Is Using AI for Clinical Trial Feasibility and Site Selection," *Clinical Trial Vanguard*, December 30, 2024, https://www.clinicaltrialvanguard.com/conference-coverage/how-novartis-is-using-ai-for-clinical-trial-feasibility-and-site-selection/; Henry P. Foote et al., "Embracing Generative Artificial Intelligence in Clinical Research and Beyond: Opportunities, Challenges, and Solutions," *JACC Journals* 4, no. 3 (February 10, 2025), https://doi.org/10.1016/j.jacadv.2025.101593.

41. James L. Cross et al., "Bias in Medical AI: Implications for Clinical Decision-Making," *PLOS Digital Health* (November 7, 2024), https://doi.org/10.1371/journal.pdig.0000651.

42. Getz, "New Insights on the Impact of AI-Enabled Solutions."

43. Peng Lean Chong et al., "Integrating Artificial Intelligence in Healthcare: Applications, Challenges, and Future Directions," *Future Science OA* 11, no. 1 (July 4, 2025): 2527505, https://doi.org/10.1080/20565623.2025.2527505.

44. Getz, "New Insights on the Impact of AI-Enabled Solutions."

45. Jennifer Zhe Zhang et al., "Informed Consent in Pragmatic Trials," *Journal of Medical Ethics* (December 19, 2022), https://doi.org/10.1136/medethics-2021-107765; Arun Bhatt, "Ethical Considerations for Real-World Evidence Studies," *Perspectives in Clinical Research* 15, no. 3 (Jul–Sep 2024): 152–54, https://doi.org/10.4103/picr.picr_256_23.

46. Redonna Chandler et al., "Broad-Scale Informed Consent," *Journal of Clinical and Translational Science* 3, no. 5 (September 23, 2019): 253–260, https://doi.org/10.1017/cts.2019.397.

47. "45 CFR 46," US Department of Health and Human Services, accessed September 17, 2025, https://www.hhs.gov/ohrp/regulations-and-policy/regulations/45-cfr-46/index.html.

48. Effy Vayena and Alessandro Blasimme, "Health Research with Big Data: Time for Systemic Oversight," *Cambridge University Press* 46, no. 1 (January 1, 2021): 119–129, https://doi.org/10.1177/1073110518766026.

49. Timothy Caulfield et al., "DNA Databanks and Consent: A Suggested Policy Option Involving an Authorization Mode," *Springer Nature* 4, no. 1 (January 3, 2003), https://doi.org/10.1186/1472-6939-4-1.

50. Jane Kaye et al., "Dynamic Consent: A Patient Interface for Twenty-First Century Research Networks," *European Journal of Human Genetics* 23, no. 2 (May 7, 2014): 141–146, https://doi.org/10.1038/ejhg.2014.71.

51. Andrew J. Goodwin et al., "Timing Errors and Temporal Uncertainty in Clinical Databases—A Narrative Review," *Frontiers in Digital Health* 4 (August 17, 2022), https://doi.org/10.3389/fdgth.2022.932599.

52. Hyun Kang, "The Prevention and Handling of the Missing Data," *Korean Journal of Anesthesiology* 64, no. 5 (May 24, 2013): 402–406, https://doi.org/10.4097/kjae.2013.64.5.402.

53. Russanthy Ruthiran Velummailum et al., "Data Challenges for Externally Controlled Trials: Viewpoint," *JMIR Publications* 25, no. 1 (April 5, 2023), https://doi.org/10.2196/43484.

54. Arun Bhatt, "Data Quality—The Foundation of Real-World Studies," *Perspectives in Clinical Research* 14, no. 2 (April–June 2023): 92–94, https://journals.lww.com/picp/fulltext/2023/14020/data_quality___the_foundation_of_real_world.9.aspx.

55. Sandra Barbosu, "Harnessing AI to Accelerate Innovation in the Biopharmaceutical Industry," *Information Technology & Innovation Foundation*, (November 15, 2024), https://itif.org/publications/2024/11/15/harnessing-ai-to-accelerate-innovation-in-the-biopharmaceutical-industry/.

56. "US Core Implementation Guide (v8.0.1: STU 8)," HL7 International, accessed September 17, 2025, https://www.hl7.org/fhir/us/core/.

57. Carina Nina Vorisek et al., "Fast Healthcare Interoperability Resources (FHIR) for Interoperability in Health Research: Systematic Review," *JMIR Publications* 10, no. 7 (July 19, 2022), https://doi.org/10.2196/35724; Katarzyna Grzeslak, "A Guide to the OMOP Common Data Model," IQVIA, June 19, 2025, https://www.iqvia.com/blogs/2025/06/a-guide-to-the-omop-common-data-model.

58. Markus Kreuzthaler et al., "Linguistic and Ontological Challenges of Multiple Domains Contributing to Transformed Health Ecosystems," *Frontiers in Medicine* 10 (March 14, 2023): 1073313, https://doi.org/10.3389/fmed.2023.1073313.

59. Blanda Helena de Mello et al., "Semantic Interoperability in Health Records Standards: A Systematic Literature Review," *Springer Nature* 12 (January 26, 2022), https://doi.org/10.1007/s12553-022-00639-w.

60. "TEFCA," Assistant Secretary for Technology Policy, accessed September 17, 2025, https://www.healthit.gov/topic/interoperability/policy/trusted-exchange-framework-and-common-agreement-tefca.

ACKNOWLEDGMENTS

The Milken Institute is grateful to Walmart for its support of the Institute's independent work on inclusive infrastructure for community-based research initiatives.

ABOUT THE AUTHORS

Kristin Schneeman is the senior director on the FasterCures team at the Milken Institute, where she oversees most policy and programmatic activities, focused on improving the research and development environment, engaging patients in research and health, and accelerating innovation. Among other initiatives, she leads The Research Acceleration and Innovation Network program, which provides a platform for knowledge sharing and relationship building to support the growth of venture philanthropy in medical research. Schneeman has over 30 years of experience in public policy, politics, academia, and the media. She served as a senior adviser and policy director to a gubernatorial candidate in Massachusetts, as a policy aide to a US congressman, and as the front-line manager and chief-of-staff for a senior adviser to Vice President Al Gore. At Harvard University, she directed research projects on future challenges facing governments and on complex negotiations in business, politics, and international relations. Schneeman began her career as a producer of documentary films, for which she was the recipient of an Emmy Award in 1990.

Sarah Sheehan is a research consultant specializing in data and digital health strategy, with a focus on leveraging digital health technologies and real-world data to support high-quality clinical evidence generation and technology-enabled care delivery. Previously, Sheehan served as program lead for CancerX at the Digital Medicine Society, leading research portfolio design, partner engagement, and program delivery for a national digital health accelerator embedded within the White House Cancer Moonshot. Prior to their time with CancerX, Sheehan was an assistant research director at Duke University, directing biomedical innovation and regulatory science initiatives focused on integrating clinical research and care, modernizing clinical trials, and advancing the use of real-world evidence in regulatory decision-making. Before joining Duke, Sheehan was a health-care research analyst at the US Government Accountability Office and a research assistant at the Centers for Disease Control and Prevention, where their work examined the efficacy of drug-development policies designed to address unmet medical needs.

LOS ANGELES | WASHINGTON | NEW YORK | MIAMI | LONDON | ABU DHABI | SINGAPORE