Navigating eTMF Complexity in Real-World Data Research

Chris Huang, Director Product Management, Oracle Health & Life Sciences
Tom Stonehouse, Senior Consulting Practice Manager, Oracle Life Sciences

14 Oct 2025

Meet the Speakers

Chris Huang

Title: Director, Product Management

Organization: Oracle Health & Life Sciences

Tom Stonehouse

Title: Senior Consulting Practice Manager

Organization: Oracle Life Sciences

Disclaimer and Disclosures

- *The views and opinions expressed in this presentation are those of the author(s) and do not necessarily reflect the official policy or position of CDISC.*

- *The author(s) have no real or apparent conflicts of interest to report.*

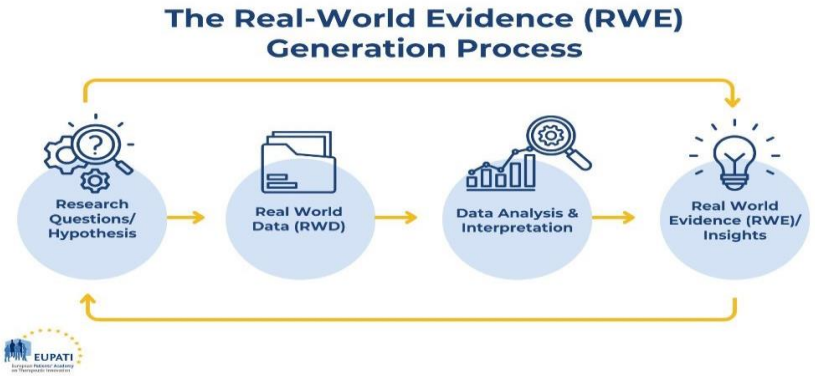
Agenda

1. How RWE studies differ from traditional clinical trials
2. Why TMF matters
3. Complexities in RWE TMF Management
4. Best Practices and Solutions

How RWE studies differ from traditional clinical trials (and implications for documentation)

How Do RWE Studies Differ from Traditional Clinical Trials?

Aspect	Traditional Clinical Trials	RWE Studies
Design	Prospective, controlled, *randomised*	Often observational (retrospective or prospective), *non-randomised*
Data Source	Purpose-collected for the trial	Existing data from routine care/practice
Population	Selected, with strict entry criteria	Broad, real-world populations
Setting	Research sites	Clinics, hospitals, registries, community
Documentation Needs	Standardized, well-defined (TMF)	More varied; includes source data agreements, transformation records, privacy documentation, etc.



Why Is RWE Research Important?

- **Complements randomized controlled trials (RCTs)**
  - RCTs are the gold standard but have limitations like strict inclusion criteria and artificial study settings.
  - RWE provides insights into how interventions perform in broader, everyday clinical practice.
- **Addresses evidence gaps**
  - Helps evaluate effectiveness and safety in underrepresented populations or long-term real-world settings.
- **Supports regulatory, clinical, and reimbursement decisions**
  - Agencies like FDA and EMA increasingly use RWE for label expansion, post-marketing commitments, and accelerated approvals.

Framingham Heart Study

Three Generations of Dedication

Since 1948, its extensive data infrastructure and follow-up have enabled deep phenotyping, risk prediction algorithms, and multi-generational discovery.

Nurses' Health Study

Exemplifies scale and longitudinal rigor, with over 40 years of data on lifestyle, medical history, genomics, and outcomes—requiring meticulous documentation, consent versioning, and linkage logs over time.

Why TMF matters in Real-World Evidence (RWE) research

Question to audience

How many of you are working in RWE studies?

Growth in RWE Use and Acceptance

FDA Real-World Evidence Submission Trends

Between 2019 and 2023, FDA recorded 80+ submissions citing RWD/RWE; by 2022, ~75% of NDAs/BLAs included RWE components.

2019
Tracking of RWE Submissions Begins
FDA began tracking submissions citing real-world evidence, signaling a shift toward data-driven regulatory evaluations.

2020
Broader Use in Applications
More submissions incorporated RWE, supporting both primary and supplementary roles in decision-making.

2021
Momentum Continues
Use of RWE in submissions continued to rise, reflecting growing acceptance and reliance.

2022
~75% of NDAs/BLAs Include RWE
By 2022, roughly 75% of new drug and biologic license applications included RWE components influencing decisions.

EMA: Integration of RWE in Marketing Authorizations

RWE use has risen sharply; 40%+ of initial authorizations cite RWE and it is becoming a standard part of application data.

2022 >40% of Initial MAs Cite RWE
EMA annual report shows over 40% of initial marketing authorizations used real-world evidence.

2023–2024 RWE Common in Applications
Marketing authorization applications now commonly include RWD/RWE, highlighting EMA's increased focus on RWE for evaluations.

2024+ Standard Component Going Forward
RWE is expected to be a standard component in future EMA applications, solidifying its role in drug evaluation.

Takeaway: RWE is now mainstream across FDA and EMA pathways—moving from tracked usage to a standard component of applications.

Growth in RWE Use and Acceptance

2010

RWE Publications Begin Rising

From just a few hundred articles in 2010, global publications on real-world evidence steadily climbed, reflecting growing interest over the decade.

2015

Steady Growth in RWE Articles

By 2015, the accumulation of RWE publications demonstrated consistent growth, signaling increasing value placed on real-world data.

2020

RWE Publications Surge Past 2,000

In 2020, RWE publications surpassed 2,000 annually worldwide, highlighting a major expansion in research output and engagement with RWE.

2022

Pharma Increases RWE Investment

A Tufts survey reported that over 60% of pharma/biotech companies had increased their RWE investments over the prior three years, indicating growing commitment.

2023

87% Plan More RWE Investment

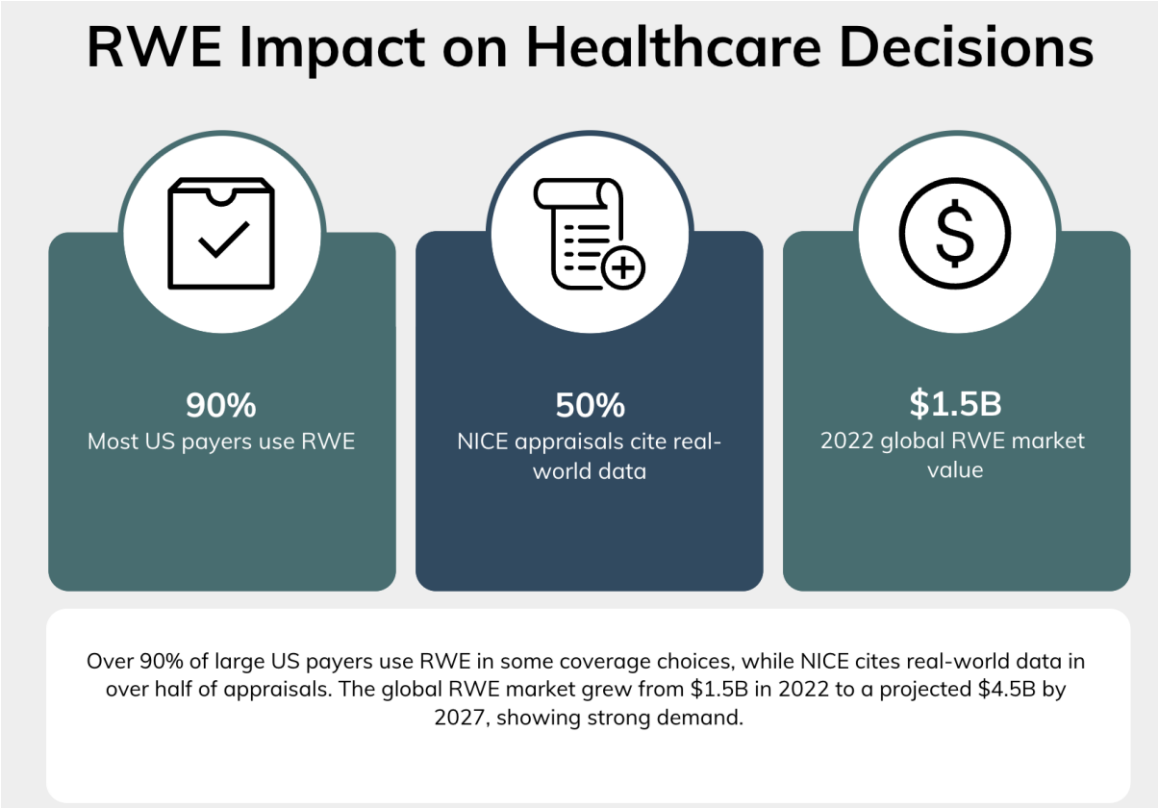
Industry leaders planned to further increase RWE investments, underscoring the expanding role of RWE in drug development and biopharma strategy.

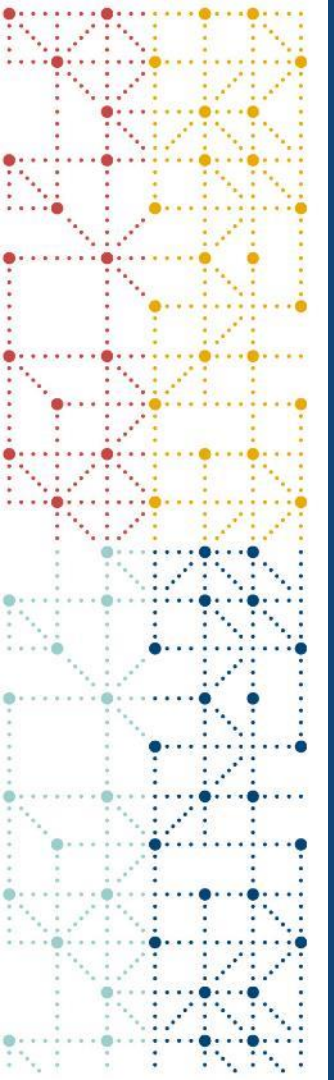
2027

Projection: Continued Scale & Formalization

Organizations are expected to further standardize RWE governance and automation, with broader payer/HTA reliance and continued market growth (with prior forecasts pointing toward ~$4.5B by 2027).

Growth in RWE Use and Acceptance





Complexities in RWE TMF Management

System & Alignment Challenges

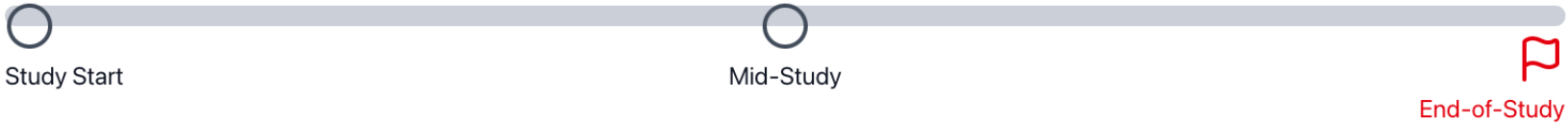
Problem: Early misalignment on systems and expectations multiplies downstream effort and cost.

Sponsor

Which system? eTMF vs. eISF vs. repository → unclear system-of-record

CRO

What to collect? RWE document set vs. trial-oriented SOPs



Visibility only emerges at **End-of-Study**

Impact: Rework, duplicate filing, inconsistent metadata, audit-time surprises.

Operational Complexity

Problem: Forcing trial TMF patterns onto RWE creates unnecessary workload, cost and risk.

RWE Studies	Trial TMF Pattern Applied
~10–15 Core Documents	100s of Document Types

SOP Misalignment
CROs plan lean RWE indices, sponsors enforce GCP-driven TMF.

Risk Aversion
Sponsors over-collect "just in case"; inspectors never penalize for too much.

Fragmented Visibility
Sponsors don't see CRO filing until end-of-study handover.

Inflexible TMF Models
eTMFs built for trials don't flex for RWE (direct-to-patient, EHR, claims).

Regulatory Ambiguity
Unclear oversight levels → sponsors default to higher collection.

Impact: Escalated costs and extended timelines driven by additional review cycles, corrections, and re-indexing of unneeded artifacts.

Multiple, Non-Standard Data Sources

RWE studies may involve:

- **EHRs** — Clinical records
- **Claims** — Payer billing
- **Registries** — Disease cohorts
- **Wearables** — Sensors & apps
- **Patient-Generated** — PROs, diaries

Each source uses different structures, standards, and extraction processes.

eTMF Complexity

Heterogeneous Sources

- EHR exports
- Claims feeds
- Registry pulls
- Wearable data
- Patient-reported

Per-Source Documentation

- **Use & Transfer Agreements** — DUAs/DTAs aligned to provider terms.
- **Provenance Logs** — Origin, extraction, transformation steps.
- **Linkage & Transform Protocols** — Merging logic across claims/EHR/registries.
- **De-identification & Privacy** — HIPAA/GDPR compliance and audit trail.

Ensuring consistency and traceability across sources is a major organizational and technological challenge.

Unique Documentation Needs

Unlike traditional trials, RWE studies often require documentation that reflects complex data sourcing and secondary data use.

Key examples include:

Data Use and Sharing Agreements

Contracts with hospitals or data providers outlining terms for data access and permitted use.

Data Provenance Logs

Detailed records showing where data came from, how it was extracted, transformed, and validated.

Data Linkage and Transformation Protocols

Documentation of methods used to combine different datasets (e.g., merging claims and EHR data), including algorithms or logic applied.

Data Privacy and De-identification Reports

Evidence that patient identities were protected in compliance with regulations (GDPR, HIPAA, etc.).

Greater Focus on Privacy, Ethics, and Compliance

Privacy Risks

- Handling real-world data often means working with large, sensitive datasets from routine healthcare — raising risks for data breaches or non-compliance.
- Emphasize least-privilege access, data minimization, and de-identification controls.

Ethical Oversight

- RWE studies increasingly seek IRB/EC approvals, even for retrospective or secondary data analyses — adding documentation to the eTMF.
- Maintain consent language mapping and data-use justifications for reuse.

Regulatory Expectations

- EMA, FDA, and other regulators may inspect RWE documentation, especially when evidence supports product labeling or safety.
- Ensure audit-ready traceability: provenance logs, DUAs/DTAs, de-ID reports, and IRB/EC approvals.

Practical takeaway: Build a privacy-first eTMF with explicit ethics documentation and regulator-ready evidence chains.

Inspection Readiness Challenges

Increased Scrutiny

Regulatory audits for RWE focus on **traceability, provenance, and transparency** — elements often not standard in traditional TMFs.

Potential Gaps

Missing or inconsistent documentation about data origins or transformations can delay submissions, threaten inspection readiness, or result in findings.

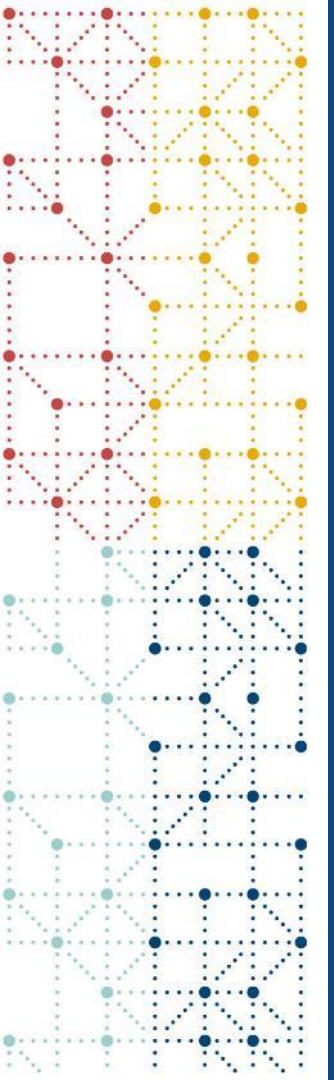
Lack of TMF Specs for RWD

TMF reference models are trial-oriented; **RWD-specific indices and filing rules** are inconsistent, creating ambiguity in what to collect and where to file it.

Inspector Focus — Evidence Chain

- **Source Agreements**
  DUAs/DTAs, access terms, permitted use
- **Provenance & ETL**
  Origin, extraction, transforms, QC steps
- **Privacy & De-ID**
  HIPAA/GDPR safeguards, re-ID risk eval
- **Traceability Links**
  Linkage logic, audit trails, versioning

Action: Define an RWD-aware TMF index, add provenance/traceability artifacts, and standardize filing rules to be inspection-ready.



Best Practices and Solutions

Best Practices and Solutions

Adaptable & Flexible Tools (eTMF)

- Configurable folders, metadata & indices for RWE data types (EHR, claims, registries, devices).
- Templates/checklists that adjust to retrospective vs. prospective and decentralized models.
- Built-in provenance & lineage capture for audit-ready traceability.

Leveraging Industry Standards

- Anchor to RWS-DI and TMF Reference Model v4+ for "what to collect."
- Map artifacts to FDA/EMA RWE expectations; document IRB/EC/Privacy determinations.
- Use controlled vocabularies/metadata for cross-source consistency.
- Enable interoperability with FHIR and related health data standards.

Agentic Automation

- AI classification & autofiling of RWE artifacts; metadata enrichment.
- Gap analysis with real-time alerts; auto-generated templates & filing rules.
- Decision trails & audit logs for explainability and inspection readiness.

Outcome: A configurable, standards-aligned, and automated eTMF that reduces effort, strengthens compliance, and scales with RWE complexity.

"Choose Your Own Adventure" – Intelligent TMF Configuration for RWD Studies

Study Type Selection
- ☐ Interventional (Drug)
- ☐ Device / Combination Product
- ☐ Observational / Registry
- ☑ Real-World Evidence (RWE)
- ☐ Expanded Access / Compassionate use

Data Source(s)
- ☑ EHR
- ☐ Registry
- ☑ Claims & Billing Data
- ☑ Patient-generated health data
- ☐ Public Health databases
- ☐ Healthcare provider networks
- ☐ EDC / Labs

Data Collection Model
- ☐ Traditional (Site-based)
- ☑ Hybrid / DCT
- ☐ Fully Decentralized / Virtual

SUBMIT

Smart TMF Template Generated
- Auto-applies the expected artifact structure (based on RWS-DI or TMF RM)
- Includes region-specific placeholders

Automated Upfront Configuration
- TMF Specialist receives guided checklist
- Completeness tracking tailored to selected inputs

Outcome: Faster setup, fewer errors, audit-ready from day 1

Example Standards-based Tool

Leveraged Today

RWS-DI: TMF Reference Model–Aligned Tool for RWS

- **Standardization** — Defines the minimal RWS document set, mapped to CDISC TMF taxonomy.
- **Regulatory Alignment** — Anchored to FDA/EMA expectations; supports audit readiness.
- **Efficiency** — Omits ~40% of trial-centric artifacts, reducing burden and speeding close-out.
- **Traceability** — Maintains provenance and linkage for inspection transparency.
- **Interoperability** — Designed to align with TMF RM v4 and FHIR standards.
- **Future-Proofing** — Provides a scalable base for automation and evolving CDISC guidance.

Key Differences between the TMF Reference Model and the RWS-DI

Comparators	TMF Reference Model	RWS-DI
Applicability	Clinical Trials	Clinical Studies that are not Clinical Trials*
Zones	11	10
Sections	40	23
Artifacts	249	130
...of which "core"	197	11
...of which recommended	52	119
Terminology	Trial	Study
Terminology	Subjects	Patients
Terminology	Safety Reporting	Pharmacovigilance

CDISC TMF RWE Working Group

Working toward **TMF Reference Model V4** support for Real-World Evidence (RWE)

Collaborative Initiative

Sponsors, CROs, and tech partners co-develop RWE-ready TMF guidance under CDISC governance.

Bridging Standards

Extending TMF Reference Model V4 to cover non-interventional studies, registries, decentralized data, and real-world sources.

Core RWE Artifacts

Aligning minimal required documents with RWS-DI to right-size filing while preserving compliance & inspection readiness.

Future-Ready Operations

Enable scalability, automation, and interoperability (FHIR/HL7 mappings) for mixed RCT + RWE portfolios.

Road to V4 (RWE Support)

→ Outcome: V4 aims to deliver a **right-sized, RWE-inclusive TMF** with artifact definitions, mappings (incl. FHIR), and guidance that reduce burden while maintaining inspection readiness.

"From trial-centric to RWE-inclusive: V4 is the bridge."

Best Practices and Solutions

Leveraging AI

Advanced cloud technologies enable collection, storage, and analysis of petabytes of real-world data (RWD). Much of this information—especially clinician notes and medical imagery—is unstructured. AI techniques (ML/NLP) can curate these data and surface previously hidden patterns, provided they are applied with strong clinical oversight and validation.

Cloud-Scale Data Platform

Elastic storage & compute to ingest petabytes; metadata catalogs for datasets and lineage; secure access controls.

Unstructured & Heterogeneous Sources

Clinician notes, images, EHR exports, device data—arriving in diverse formats that require normalization and ETL.

ML/NLP Pipelines

De-identification, entity extraction, text/image embedding, and feature engineering to create analysis-ready data.

Pattern Discovery

Efficient search and curation reveal relationships and signals previously hidden across modalities and sites.

Robust Validation
Separate training/validation sets; bias checks; performance monitoring.

Clinical Oversight
Clinician-led review of outputs; relevance and safety vetting.

Governance
Document pipelines, decisions, and audit trails in the TMF.

AI-Driven Best Practices and Solutions for RWE

Data-driven Document Mapping

- AI-powered classification of eTMF documents.
- Metadata enrichment to reduce manual errors.

Template & Checklist Generation

- Automated configuration templates based on RWD-specific needs.
- Custom rule engines for hybrid/decentralized trials.

Gap Analysis & Compliance

- Predictive models for gap identification.
- Real-time alerts for missing or expired documents.

Continuous Learning

- User feedback integration to refine recommendations.
- Periodic AI model updates for compliance alignment.

Best Practice Reminders

- Configurable workflows for RWD-specific filing.
- Regulatory alignment to ensure audit readiness.

Handling Huge Data Volumes

- Manage petabytes of structured and unstructured RWD.
- Enable scalable AI for intelligent mapping, template generation, and rule engines.
- Ensure efficient data ingestion with traceability and compliance.

Final Insights: TMF in Real-World Studies

Alignment is critical

Sponsors and CROs must set expectations up front to avoid over-collection, rework, and cost inflation.

RWS ≠ Clinical Trials

Right-size TMF for real-world studies—don't blindly apply full GCP trial models where they don't fit.

Standards bring clarity

Leverage CDISC TMF reference models and the RWS-DI to define the minimal, defensible document set.

Future is agentic + automated

Use AI to map, classify, and file at scale; apply rule engines for validation and auditability.

Education & collaboration

Continuous sponsor–CRO–regulator dialogue is essential to sustain quality and reduce burden.

"Right-size the TMF for RWS: Collect what matters, align early, and let automation do the rest."

Thank You!