PHUSE REAL WORLD DATA AUTUMN EVENT 2026

Beyond Standard Checks:

A Reusable Data Quality Validation Framework for OMOP CDM in Real-World Evidence Studies

Sowmya Lakshmi Gudimella

Associate Biostatistical Programmer

Amgen



Data Quality in Real-World Data: Established Approaches

Common themes recur across regulatory guidance, research literature, OMOP/OHDSI, and industry practice.



REGULATORY [1–3]

- Fit-for-purpose
- Reliability & relevance
- Standardization & traceability
- Systematic quality assessment

Emphasizes data suitability, reliability, standards, and traceability.



RESEARCH LITERATURE [4,5]

- Completeness
- Conformance
- Plausibility
- Correctness / consistency

Identifies recurring DQ dimensions and structured assessment approaches.



OMOP / OHDSI [6,7]

- Conformance
- Completeness
- Plausibility
- Automated DQ assessment

Operationalizes systematic DQ assessment for OMOP CDM data.



INDUSTRY PRACTICE [8,9]

- Profiling
- Rule-based validation
- Monitoring
- QC / reporting

Emphasizes reusable rules, monitoring, and operational quality management.

Across these approaches, data quality is addressed through defined concepts, systematic assessment, and repeatable validation methods.

From Established DQ Concepts to OMOP Validation

Established DQ concepts provide a foundation for assessing whether standardized OMOP data are fit for downstream use.

ESTABLISHED DQ CONCEPTS [4,5]

- Completeness
- Conformance
- Plausibility
- Consistency / Integrity
- Reliability / Relevance

Recurring concepts identified across regulatory, research, OMOP, and industry approaches.

OMOP CDM [6,7]



Standardized structure

Standardized vocabulary

Provides a common representation of heterogeneous healthcare data.

PRACTICAL QUALITY QUESTIONS

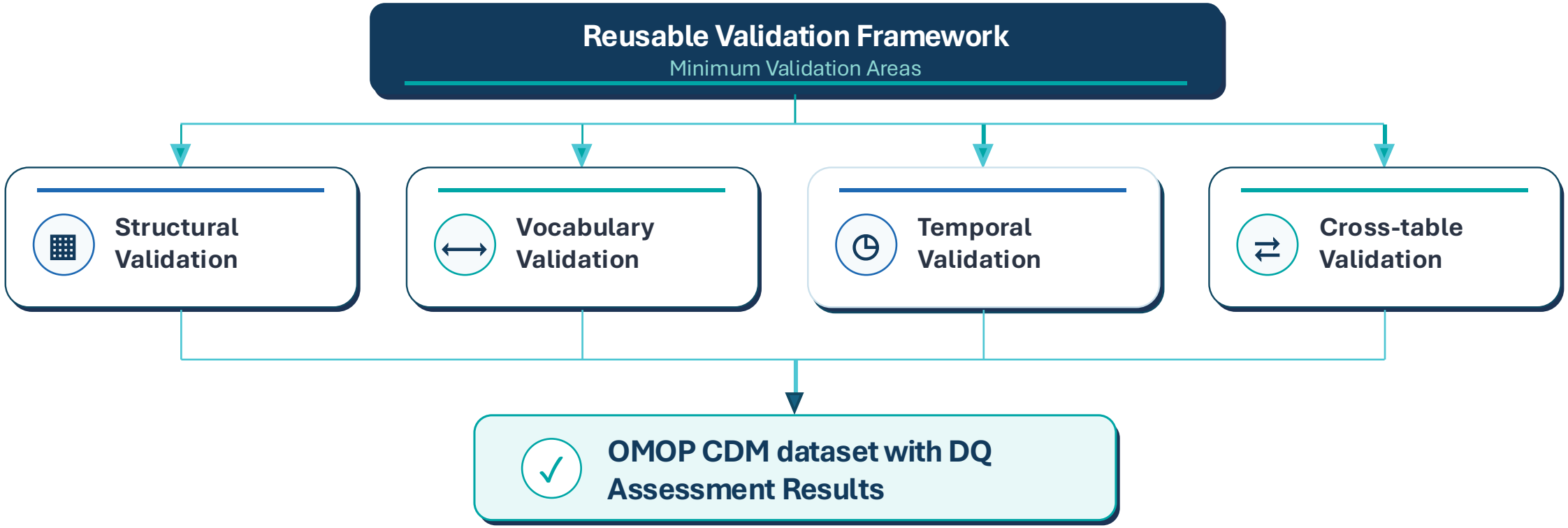
- Is the data complete and conformant?
- Are clinical concepts represented and events timed appropriately?
- Are data relationships internally coherent?

These questions help determine whether standardized data are suitable for downstream analysis.

Systematic assessment of these quality needs provides the basis for a reusable validation approach.

Standardization organizes data. Validation establishes confidence.

Modular Validation Framework for OMOP CDM



Minimum validation areas.
Additional checks may be considered based on study objectives, data source, and use case.

- Trusted Evidence
- Reusable Across Studies
- Reliable Real-World Evidence

Standardization organizes data. Validation establishes confidence.

Structural validation

OMOP PERSON table: structural issues

Example structural issues

person ID (PK)*	gender ID*	birth year*	race ID	ethnicity ID
1001	8507	1968	8527	38003563
NULL	8532	1975	8516	38003563
1003	NULL	1984	8527	38003563
1004	8507	NULL	NULL	NULL

■ Missing PK ■ Mandatory NULL ■ Incomplete row

Validation Results

Structural validation confirms that required structural checks pass before downstream quality assessment.

- ✓ **Mandatory columns verified**
- ✓ **Primary key integrity verified**
- ✓ **Correct data types**
- ✓ **Structural completeness verified**

Ready for downstream quality checks

Complete and structurally sound data is the foundation of reliable analysis

Vocabulary validation

OMOP CONDITION_OCCURRENCE

Example vocabulary mapping workflow

- **ICD-10: E11 — Diabetes Mellitus**
- **ICD-10: I10 — Hypertension**
- **ICD-10: I50 — Heart failure**
- **Local code: ABC-123 — Unknown Diagnosis**

OMOP Standard Vocabulary



Maps source concepts to standard concepts using OMOP vocabulary relationships

Mapping Results

- Diabetes mellitus — **Standard Concept**
- Hypertension — **Standard Concept**
- Heart failure — **Mapping Review Required**
- Local code ABC-123 — **No Standard Concept Found**

Validation Results

- Standard concepts verified
- Valid mapping relationships
- Vocabulary integrity confirmed
- Unmapped codes flagged for correction

Ready for downstream quality checks

Accurate concept mapping ensures consistent clinical interpretation

Temporal validation

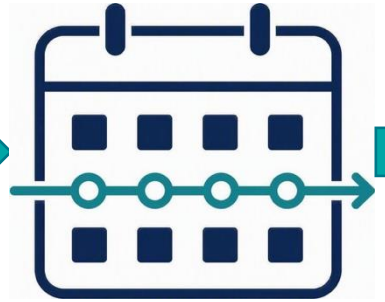
OMOP CONDITION_OCCURRENCE table

Temporal validation

Example temporal inconsistencies

- × **Condition occurrence after death**
- × **Visit end date before visit start date**
- × **Drug exposure before birth date**
- ✓ **Valid chronological event**

TEMPORAL VALIDATION RULES



Ensures logical chronological order of clinical events

Validation Results

- ✓ **Valid event chronology**
- ✓ **Dates within observation period**
- ✓ **Logical event sequencing**
- ✓ **No impossible clinical timelines**

Ready for downstream quality checks

Chronological consistency prevents clinically implausible events.

Cross-table validation

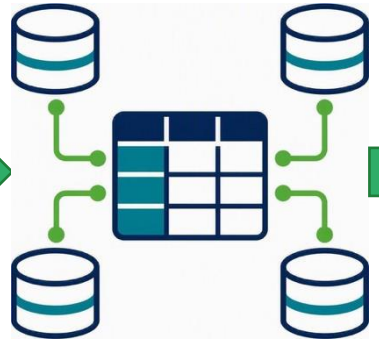
OMOP table relationships

Example relationship issues

PERSON	VISIT OCCURRENCE	CONDITION OCCURRENCE

- ✗ CONDITION_OCCURRENCE -> missing VISIT_OCCURRENCE
- ✗ VISIT_OCCURRENCE -> missing PERSON
- ✗ DRUG_EXPOSURE -> invalid VISIT_OCCURRENCE
- ✓ All foreign-key relationships valid

Cross-table validation rules



Verifies primary and foreign key relationships across OMOP tables

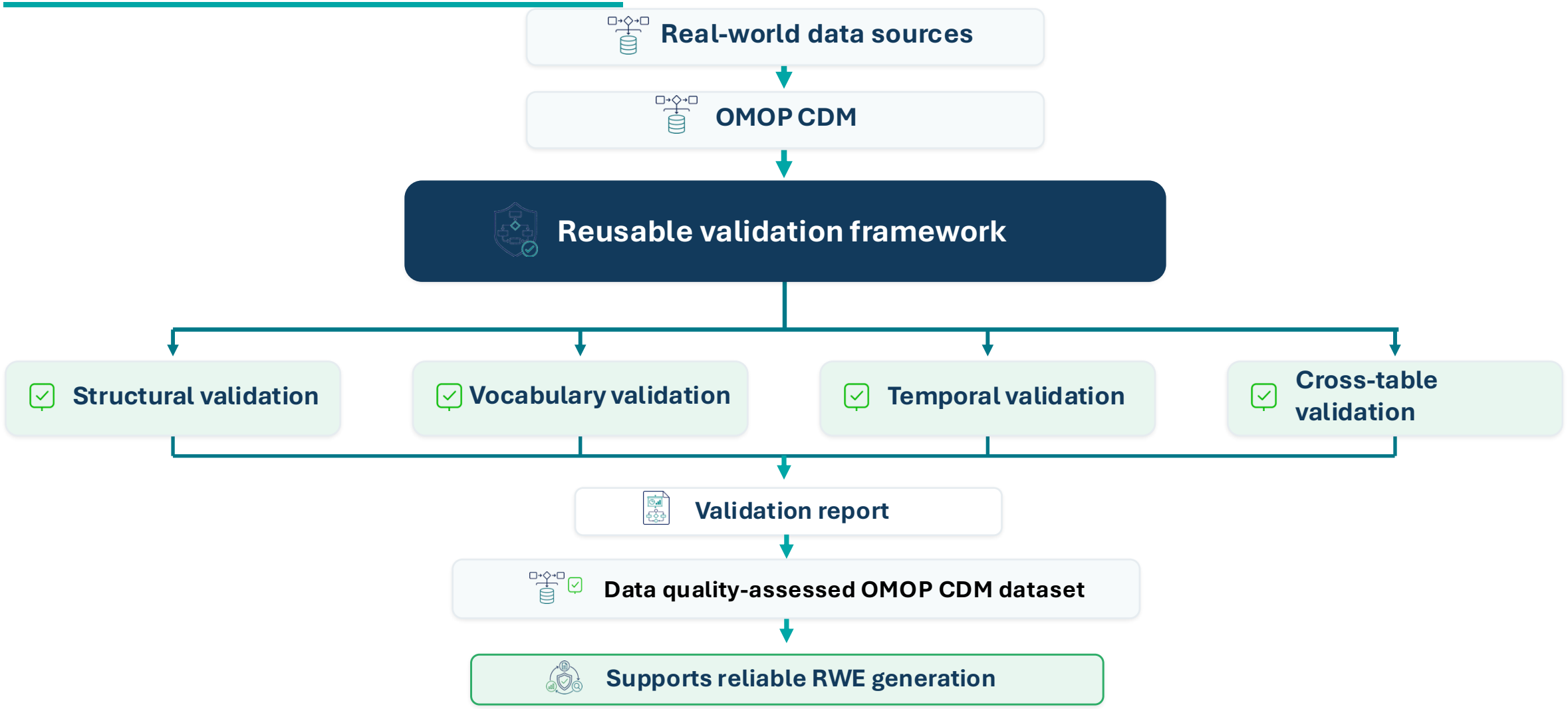
Validation Results

- ✓ Valid person–visit relationships
- ✓ Valid visit–condition links
- ✓ Foreign key integrity verified
- ✓ Relational integrity verified

Ready for downstream quality checks

Relational integrity preserves consistency across OMOP domains

Framework implementation workflow



One workflow transforms standardized data into validated, analysis-ready datasets.

Benefits of the Reusable Validation Framework



Faster Validation

- Reduces manual QC effort
- Speeds study start-up



Reusable Across Studies

- Shared validation assets
- Reusable across protocols



Improved Data Quality

- Helps identify issues earlier
- Supports cleaner CDM outputs



Standardized Validation

- Consistent validation rules
- Standardized QC outputs



Reliable Real-World Evidence

- Can increase confidence in downstream analyses
- Supports more reliable real-world evidence



Scalable Framework

- Adds modules as needs grow
- Supports multiple OMOP studies

One Reusable Framework, Consistent Validation, Analysis-Ready OMOP Datasets, Trusted Real-World Evidence

Key Takeaways



Validation is essential

OMOP standardization alone does not guarantee data quality or fitness for downstream use. Structural, vocabulary, temporal, and cross-table validation provide complementary quality assessment.



Modular and reusable framework

Reusable validation modules can reduce repetitive QC effort and support consistent validation across OMOP studies.



Trusted real-world evidence

Systematically assessed OMOP datasets can increase confidence in downstream analyses and support reliable real-world evidence generation.

Better validation. Better OMOP. Better evidence

References

Selected sources from the literature evidence matrix supporting the presentation's scientific foundation.

REGULATORY & RESEARCH

1. U.S. Food and Drug Administration (FDA). RWD/RWE Guidance. 2023.
   fda.gov/.../considerations-use-real-world-data-and-evidence

2. U.S. Food and Drug Administration (FDA). Data Standards for RWD. 2023.
   fda.gov/.../data-standards-submissions-containing-rwd

3. European Medicines Agency (EMA). Data Quality Framework. 2023.
   ema.europa.eu/.../data-quality-framework-medicines-regulation

4. Kahn et al. Harmonized framework for assessing EHR data quality. 2016.
   pubmed.ncbi.nlm.nih.gov/27274012

5. Weiskopf & Weng. Framework for assessing EHR data quality for reuse. 2013.
   pubmed.ncbi.nlm.nih.gov/22733976

OMOP & PRACTICE

6. OHDSI. Book of OHDSI: Data Quality. 2019.
   ohdsi.github.io/TheBookOfOhdsi/DataQuality

7. OHDSI. Data Quality Dashboard. 2020.
   ohdsi.github.io/DataQualityDashboard

8. Ataccama. Data Quality Overview.
   docs.ataccama.com/.../data-quality-overview

9. DARWIN EU. Improving interoperability. 2024.
   darwin-eu.org/.../improving-interoperability

Literature grounds the framework; practical validation remains the focus.