**TAXIS Network Study Protocol**

**Version:** 0.5

**Date:** 15-Sep-2025

**Authors:** J. Marc Overhage, MD, PhD, Steve Bandeian, MD, JD

**Acknowledgement:** The analysis is based in part on work from the Observational Health Data Sciences and Informatics collaborative. OHDSI ([http://ohdsi.org)](http://ohdsi.org/) is a multi-stakeholder, interdisciplinary collaborative to create open-source solutions that bring out the value of observational health data through large-scale analytics.

**List of abbreviations:**

CDM – Common Data Model

OMOP – Observational Medical Outcomes Partnership

OHDSI – Observational Health Data Sciences and Informatics

**1. Abstract**

The TAXIS Network Study is a federated, multi‑site program to discover, validate, and publish reusable clinical relationships among **all OMOP clinical concept domains** (conditions, drugs, procedures, measurements/labs, devices, observations, and visits). Sites execute a standard TAXIS study package locally to generate **de‑identified aggregate statistics** for concept pairs (and, where applicable, directed sequences). The Coordinating Center (CC) performs quality control, harmonization, and **LLM‑assisted, safeguarded classification** with human‑in‑the‑loop adjudication to label relationship types (e.g., *treats*, *adverse effect of*, *diagnostic for*, *part‑of procedure pathway*, *co‑morbidity cluster*). Network‑level, site‑agnostic artifacts are released openly; no patient‑level data leave the sites. **Phase 1** focuses on condition‑condition; **Phase 2+** extends to all domain pairs.

**2. Amendments & Milestones**

|  |  |  |
| --- | --- | --- |
| Protocol Version | Planned Date | Notes |
| Version 1.0 | 1-Oct-2025 |  |

|  |  |  |
| --- | --- | --- |
| Milestone | Planned Date | Notes |
|  |  |  |
| Release TAXIS package v1.1 | 2025‑10 | At 2025 symposium |
| Local approvals (IRB/Privacy) | Site‑specific | Template text provided |
| Phase 1 data window (Dx‑Dx) | 2025‑10 → 2026‑01 | Rolling intake |
| Phase 2 pilots (Drg‑Dx; Proc‑Dx; Meas‑Dx) | 2025‑12 → 2026‑03 | Narrow pilots first |
| Aggregation/validation (all phases) | 2025‑11 → 2026‑04 | QC + classification |
| Public release(s) | 2026‑06 | Iterative network artifacts |

**3. Rationale & Background**

Knowledge graphs for clinical decision support and research require high‑quality, **typed relationships** across concept domains (e.g., *metformin treats type 2 diabetes*; *troponin diagnostic for myocardial injury*; *PCI follows diagnostic cath*). TAXIS extends established OHDSI/OHDSI‑like federated methods by (a) **eventizing** heterogeneous domain data into comparable episodes, (b) computing **pairwise and sequential statistics** with domain‑aware windows, and (c) applying a reproducible, **guard‑railed LLM** and human adjudication layer to assign **semantically rich edge types**.

For OHDSI users, these tools simplify cohort building, improve phenotype accuracy, and streamline confounding control. For instance, users studying sepsis can easily identify related upstream and downstream conditions using the TAXIS relationships without manual mapping, thereby improving both efficiency and analytical rigor. Subsequent steps, including linking diagnoses to services and mapping service-to-service relationships, will unlock insights into care pathways, service variation, and potential improvements in care delivery. These steps will enable more dynamic modeling of real-world clinical practice.

**4. Objectives & Research Questions**

**Primary Objective:** Create a validated, reusable **multi‑domain relationship dataset** with statistical descriptors and typed edges suitable for populating a clinical knowledge graph.

**Specific Aims:**

1. Define domain‑agnostic and domain‑specific **event construction** methods and co‑occurrence/sequential windows across OMOP domains.
2. Generate site‑level, de‑identified **aggregate statistics** for concept pairs and sequences; pool across the network to identify robust associations.
3. Classify relationships with a safeguarded **LLM + human adjudication** workflow and publish open, network‑level artifacts.

**5. Study Design Overview**

Multicenter, observational, federated study across OMOP v5+ sites. The TAXIS package runs locally to produce masked **site‑level aggregates** for pre‑specified domain‑pair analyses; CC performs QC, harmonization, classification, and dissemination. No patient‑level data or direct identifiers are transferred.

**6. Methods**

**6.1 Domains & Pair Types in Scope**

OMOP standard concept domains initially targeted: **Condition, Drug, Procedure, Measurement (Lab), Device, Observation** (with Visit for context). We prioritize the following **pair types** (Phase wave in parentheses):

* **Condition ↔ Condition (Dx‑Dx)** – co‑morbidity clusters; complications; sequelae. *(Phase 1)*
* **Drug → Condition (Drg‑Dx)** – *treats*, *induces/adverse effect of* (ADR), *prophylaxis for*. *(Phase 2)*
* **Condition → Drug (Dx‑Drg)** – *treated with*, *contraindicated with* (flagged via heuristics). *(Phase 2)*
* **Procedure → Condition (Proc‑Dx)** – *performed for/diagnostic of/treats*; *complication of procedure*. *(Phase 2)*
* **Condition → Procedure (Dx‑Proc)** – *indication for*. *(Phase 2)*
* **Procedure ↔ Procedure (Proc‑Proc)** – *pathway step/sequence*, *bundled/part‑of*. *(Phase 3)*
* **Measurement ↔ Condition (Meas‑Dx)** – *diagnostic for*, *monitoring for*, *risk marker for*. *(Phase 2)*
* **Drug ↔ Drug (Drg‑Drg)** – *co‑administration*, *potential interaction* (signal only). *(Phase 3)*
* **Device ↔ Condition/Procedure (Dev‑Dx/Proc)** – *implanted for*, *used during*. *(Phase 3)*

Additional pairs (e.g., Observation↔Condition) may be developed as well.

**6.2 Domain‑Agnostic Event Construction ("Eventization")**

All domains are mapped into **events** with: event\_id, concept\_id, domain, start\_datetime, end\_datetime (nullable), person\_id (local only), visit\_id (optional), provenance (source table), severity/value (for measurements). Event construction rules:

* **Condition**: diagnosis-group based condition\_episodes by person, merging same condition within group specific intervals
* **Drug**: drug\_eras (OHDSI convention) with allowed gaps; capture ingredient and clinical drug levels (RxNorm).
* **Procedure**: single‑day events; create **procedure episodes** if repeated within 7 days.
* **Measurement**: each result an event; derive **abnormal flags** by reference range; optional binning for quantitative values.
* **Device**: exposure window from implant to removal if available; else single‑day.
* **Observation**: single‑day; roll‑up to curated groupers where applicable.

Each participating site will have the option to exclude any concepts necessary to adhere to their individual data governance guidelines.

**6.3 Co‑occurrence & Sequential Windows (by Pair Type)**

Default **person‑level** windows (sites may tighten):

| **Pair Type** | **Co‑occurrence Window** | **Directionality/Sequence Rule** |
| --- | --- | --- |
| Dx‑Dx | ±30 days | A before B / same‑day / B before A |
| Drg‑Dx (treats/ADR) | Drug start ≤7 days **before** Dx suggests *treats*; Dx within ≤14 days **after** drug start suggests *ADR* | Time‑anchored around drug start |
| Dx‑Drg | Dx ≤30 days **before** drug start suggests *treated with* | Dx index → drug start |
| Proc‑Dx | Proc within ≤7 days **after** Dx (*treatment*); Dx within ≤7 days **after** Proc (*complication*) | Relative to procedure time |
| Dx‑Proc | Dx ≤30 days **before** Proc (*indication*) | Dx index → procedure |
| Proc‑Proc | Second proc ≤30 days **after** first (*pathway step*) | Ordered pair |
| Meas‑Dx | Meas same‑day/≤3 days **before** Dx (*diagnostic for*); Meas after Dx (*monitoring*) | Relative to Dx time |

**6.4 Collected Summary Statistics**

| **Name** | **Description** |
| --- | --- |
| **cooc\_obs** | Number of distinct persons who had both Concept A and Concept B within the time window (co-occurrence observations). |
| **cooc\_event\_count** | Total number of co-occurring event pairs across all persons (counts multiple events per person). |
| **a\_before\_b** | Count of times Concept A occurred before Concept B within the time window. |
| **same\_day** | Count of times Concept A and Concept B occurred on the same date for the same person. |
| **b\_before\_a** | Count of times Concept B occurred before Concept A within the time window. |
| **nA** | Number of distinct persons with Concept A. |
| **nB** | Number of distinct persons with Concept B. |
| **total\_persons** | Total number of distinct persons in the dataset (denominator for expected rates). |
| **expected\_obs** | Expected number of co-occurrences of Concept A and Concept B if independent, based on marginals (nA, nB, total\_persons). |

**6.4 Site‑Level Aggregate Computation**

For each pair type, compute **distinct person counts** and event counts within the window; record marginals (nA, nB, population). Derive domain‑agnostic statistics (expected counts, **lift** with CI, **Z‑scores**, **odds ratios** with Haldane‑Anscombe correction, directionality proportions).

| **Name** | **Description** |
| --- | --- |
| **lift** | Ratio of observed to expected co-occurrences ( >1 = more frequent than expected). |
| **lift\_lower\_95 / lift\_upper\_95** | 95% confidence interval bounds for the lift measure. |
| **z\_score** | Standardized difference between observed and expected co-occurrences (signal strength). |
| **ab\_h** | Haldane-smoothed count of both A and B (for odds ratio stability). |
| **a\_only\_h** | Haldane-smoothed count of persons with only Concept A. |
| **b\_only\_h** | Haldane-smoothed count of persons with only Concept B. |
| **neither\_h** | Haldane-smoothed count of persons with neither A nor B. |
| **odds\_ratio** | Odds of A and B co-occurring compared to not co-occurring. |
| **or\_lower\_95 / or\_upper\_95** | 95% confidence interval bounds for odds ratio. |
| **directionality\_ratio** | Proportion of directional occurrences (A before B vs B before A). |
| **dir\_prop\_a\_before\_b** | Proportion of ordered pairs where A precedes B. |
| **dir\_lower\_95 / dir\_upper\_95** | 95% confidence interval bounds for the directionality ratio. |
| **confidence\_a\_to\_b** | Probability that a person with Concept A also has Concept B. |
| **confidence\_b\_to\_a** | Probability that a person with Concept B also has Concept A. |

Apply a **small‑cell suppression** threshold of ≥50 distinct persons in any reported cell.

**6.5 Candidate Selection**

Pairs advance to validation if meeting thresholds, e.g., **lift ≥1.5**.

**6.6 LLM‑Assisted Classification (Centralized)**

**Purpose:** Map statistical associations to **typed clinical relationships** using constrained prompts, and deterministic settings. Only the concept names, the co-occurrence count and the actual to expected ratios are available to the LLM. Discordant findings will trigger **human adjudication.**

**Allowed edge labels (condition to condition):**

* 1: 'A causes B',
* 2: 'B causes A',
* 3: 'A indirectly causes B',
* 4: 'B indirectly causes A',
* 5: 'A and B share common cause',
* 6: 'Treatment of A causes B',
* 7: 'Treatment of B causes A',
* 8: 'A and B have similar initial presentations',
* 9: 'A is subset of B',
* 10: 'B is subset of A',
* 11: 'No clear relationship'
* )

**Safeguards:** Predefined label set; no novel mechanisms; **version‑pinned models**; prompt+input hashing; audit trail; **human‑in‑the‑loop** for low‑confidence or conflicting directionality.

**6.7 Aggregation & Public Artifacts**

CC harmonizes site descriptors and aggregates masked counts/statistics to produce:

* A **public, typed relationship table** spanning multiple domains (site‑agnostic).
* Distributional summaries and data dictionary.
* A **private, site‑specific** store for QC and sensitivity analyses only.

**6.8 Exclusions & Local Overrides**

Sites may exclude sensitive concepts or raise suppression thresholds (e.g., ≥20). Excluded pairs are omitted from outputs.

**6.9 Quality Control (QC)**

**Statistical**: Recomputation of derived metrics from site-specific counts

**Benchmark against literature:** Human-curated relationship datasets will be used to cross-check results. For example, medication indications published in compendia can be used to cross-check medication-condition pairs.

**Cross-site reproducibility:** Compare results from multiple OMOP instances (e.g., IHIE vs Johns Hopkins) and check overlap/discordance.

**7. Roles & Responsibilities**

* **Coordinating Center (TCC):** Maintains package and plug‑ins; receives aggregates; QC; classification; aggregation; governance; dissemination.
* **Participating Sites:** Local execution; privacy controls; secure transfer; respond to QC.
* **Steering Committee:** Scientific oversight; SAP updates.
* **Data Governance Board:** Reviews exclusions; privacy/security monitoring.

**8. Data Management & Security**

**Flow:** Local execution → masked aggregates + site descriptor → secure transfer (TLS) → CC storage → RBAC with audit.  
**Incident response:** Purge to the extent technically possible; notify site; post‑mortem.  
**No PHI/patient‑level data** transferred; strict no‑reidentification clause.

**9. Protection of Human Subjects**

Aggregate, de‑identified data only; **small‑cell suppression (≥10)**. Anticipated Not Human Subjects Research at CC; sites obtain local determinations per policy.

**10. Software, Reproducibility & Documentation**

* **Package architecture:** OHDSI study package with R wrapper and OHDSI SQL, versioned on github
* **TAXIS web site:** supports uploads of site data and downloads with relationship tags and aggregation of counts, statistical analysis, LLM validation and human review.
* **LLM logs:** concept pairs, model version; temperature; relationship, rational maintained

**11. Dissemination Plan**

Phased public releases by pair‑type family (Dx‑Dx → Drg‑Dx → Proc‑Dx/Meas‑Dx → others). Network‑level tables, dashboards, white paper, and manuscript; site‑level data remain private.

**12. Risks & Mitigations**

* **Residual disclosure risk:** Suppression; aggregation.
* **Spurious associations:** Multi‑site replication; domain‑specific thresholds; human adjudication.
* **Operational errors:** Containerization; preflight checks; schema validation.

**13. Limitations**

Aggregate design limits patient‑level adjustment for confounding; event/window choices affect sensitivity/specificity; ADR/contraindication signals are exploratory without formal causality assessment.

**14. Timeline**

See §2; Gantt chart in repository.

**15. References & Study Artifacts**

* TAXIS package (core + plug‑ins); data dictionary; QC checklist; governance templates.

**Appendix A. Pair‑Type Registry (Initial)**

| **Pair Type** | **Domains** | **Default Window** | **Primary Metrics** | **Edge Labels (examples)** |
| --- | --- | --- | --- | --- |
| Dx‑Dx | Condition↔Condition | ±30d | Lift, Z, OR | causes; complication of; co‑morbidity; associative |
| Drg‑Dx | Drug→Condition | −7→+14d around drug start | PRR, ROR, Lift | treats; adverse effect of; prophylaxis for |
| Dx‑Drg | Condition→Drug | Dx ≤30d before drug | Lift | treated with; contraindicated with\* |
| Proc‑Dx | Procedure↔Condition | −7/+7d | Lift, OR | treats; diagnostic for; complication of |
| Dx‑Proc | Condition→Procedure | Dx ≤30d before proc | Lift | indication for |
| Proc‑Proc | Procedure→Procedure | ≤30d | Support, Confidence | sequence; pathway step; part‑of |
| Meas‑Dx | Measurement↔Condition | −3/+0 (diagnostic), +0→ | LR+/LR‑, Lift | diagnostic for; monitoring for; risk marker |
| \*Heuristic flag; requires human review. |  |  |  |  |

**Appendix B. Site‑Level Output Schema (Generic)**

**Concept fields:** domain\_a, concept\_a\_id, concept\_a\_name, concept\_a\_class, vocabulary\_a; domain\_b, concept\_b\_id, concept\_b\_name, concept\_b\_class, vocabulary\_b.  
**Event counts:** nA\_persons, nB\_persons, population\_persons; cooc\_persons; cooc\_events; a\_before\_b\_events; same\_day\_events; b\_before\_a\_events.  
**Derived:** expected\_cooc; lift; lift\_ci\_low/high; z\_score; ab\_h, a\_only\_h, b\_only\_h, neither\_h; odds\_ratio; or\_ci\_low/high; directionality\_ratio; conf\_a\_to\_b; conf\_b\_to\_a.  
**Domain‑specific (conditional):** PRR, ROR, EBGM/IC; LR+/LR‑; support, confidence.  
**Provenance:** site\_id (pseudonymous); data\_refresh\_date; vocab\_version; package\_version; pair\_type; window\_id; run\_hash.

**Appendix C. LLM Classification Guardrails (Multi‑Domain)**

* Fixed label set per pair‑type; no mechanistic speculation.
* Deterministic settings; prompt+input hashing; cache; audit trail.
* **Human adjudication triggers:** CI includes null (e.g., lift CI spans 1), conflicting directionality, high impact (e.g., potential contraindication), or low model confidence.

**Appendix D. Governance, Privacy & Security Templates**

* DUA with no‑redistribution/no‑reidentification; incident response SOP; key rotation/access control SOP.

**Appendix E. Sensitivity Analyses Menu**

Alternative windows per pair‑type; higher suppression thresholds; inpatient vs outpatient stratification; leave‑one‑site‑out stability.