Poster Code: RWD139

Advancing Computable Operational Definitions with Ontology Mapping and Synthetic Data Simulation for More Consistent Real-World Evidence

Scan for poster + more

Aaron Kamauu, MD MS MPH[1] Sadie Nordstrand, BS[2] Joseph Flinders, BS[2] Adam Hansen, PhD[2]

[1]Navidence, Salt Lake City, UT, USA; [2]Geneial, Missouri City, TX, USA

So What?

Ontology mapping with GenOMA enables stringent, symptom-anchored TRD computable operational definitions (CODefs) to be applied to free-text clinical notes and patient-reported surveys — unlocking cohorts in unstructured data sources that could not otherwise support rigorous Real-World Evidence (RWE).

BACKGROUND

The Problem

~30% of MDD patients develop Treatment-Resistant Depression

6× variability in TRD prevalence estimates (Fife 2018)

2 rich free-text data sources previously inaccessible to rigorous CODefs

The Gap / Problem

Computable Operational Definitions (CODefs) promise reproducibility — but CODef selection and performance depends on whether the data source contains structured, standardized fields. Stringent, symptom-anchored TRD definitions require standardized drug, dose, duration, and response data. Two rich but underutilized data sources — clinician-written EHR narrative notes and patient-reported registry surveys — carry exactly this information in prose form. Ontology mapping can add rigor to the use of CODef logic, and synthetic data and demonstrate this value.

HEOR / RWE Relevance

Treatment-resistant depression (TRD) definitions, as a subset of major depressive disorder (MDD), determine clinical trial eligibility, payer formulary and coverage decisions, and cross-study evidence pooling. When "TRD" means different things across datasets, the evidence base is fragmented for all stakeholders — clinicians, payers, regulators, and researchers alike.

OBJECTIVES

Primary Objective

To evaluate how ontology mapping with GenOMA affects the applicability and performance of two published TRD Computable Operational Definitions (CODefs) when applied to synthetic clinician-written EHR narrative notes and synthetic patient-reported survey data.

Secondary Objectives

- Compare CODef applicability (% evaluable criteria) across raw vs. GenOMA-mapped data formats
- Assess CODef performance across 4 dataset conditions x 2 CODefs
- Demonstrate how definitional choice interacts with data source to affect cohort identification

DISCLOSURES & ACKNOWLEDGMENTS

Aaron Kamauu: Co-Founder & CEO, Navidence Inc. Sadie Nordstrand, Joseph Flinders, Adam Hansen: Employees, Geneial. GenOMA developed by Geneial. TRD CODefs from Navidence published CODef library. No additional conflicts of interest to disclose.

REFERENCES

1. Cepeda 2018
2. Fife D, et al. Curr Med Res Opin. 2018;35(8):1297-1305.
3. McIntyre RS, et al. Lancet. 2023;401(10380):957-976.
4. Rush 2006
5. Hayasaka 2015
6. Sforzini L, et al. J Affect Disord. 2022;296:245-254.
7. Chandler GM, et al. J Clin Psychiatry. 2010.
8. APA 2010 practice guideline
9. GenOMA methods paper

METHODS

Study Design & Data Source

Simulation study. A synthetic MDD cohort (N=500) with known ground-truth TRD status is projected into two RWD formats. Two published TRD CODefs are applied to each of 4 dataset conditions and outputs are compared to ground-truth labels. No formal hypothesis testing.

Study Population

N=500 synthetic patients across 6 clinically meaningful archetypes with known ground-truth TRD status. Parameters from STAR*D. [1-5]

Synthetic Cohort Archetypes (N=500)

Ground Truth Cohort (N=500)
6 archetypes | known TRD labels | per-patient medication history

Project to 2 RWD Formats

EHR Narrative Notes
Clinician-written prose
Medication, dose, duration, response

- Raw Format: Unstructured text, No standardization
- GenOMA-Mapped: standardized drug identifiers and MGH ATRQ categories

Patient-Reported Survey
Free-text drug names, vague dosing
Subjective response language

- Raw Format: Brand/generic confusion, Vague dates & doses
- GenOMA-Mapped: standardized drug identifiers and MGH ATRQ categories

Apply 2 Published TRD CODefs to each condition:

Rx+ATRQ (Stringent)
MDD Dx + ≥2 adequate trials (different classes, therapeutic dose, ≥6 wks) + MGH ATRQ <25%/trial

Rx-Only (Proxy)
MDD Dx + ≥2 adequate trials + subsequent different-class treatment as non-response proxy

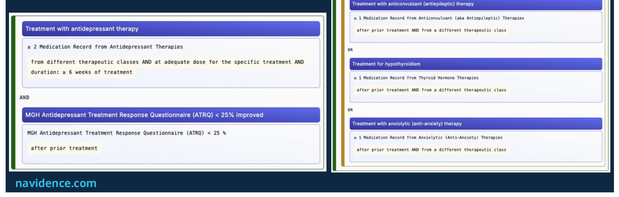
Compare CODef Output to Ground-Truth TRD Labels
Sensitivity / PPV / F1

Outcome Measures

Per-patient TRD predictions from each CODef compared to ground-truth TRD labels. We report sensitivity (recall), PPV (precision), and F1 across all five datasets and both CODef variants.

CODefs Evaluated

- Rx+ATRQ (Stringent): MDD Dx + ≥2 adequate antidepressant trials (different classes, therapeutic dose, >=6 wks) + MGH ATRQ <25% improvement/trial. Anchors on quantified non-response.
- Rx-Only (Proxy): MDD Dx + >=2 adequate trials + subsequent different-class treatment as proxy for non-response. Used when symptom-score data is unavailable.



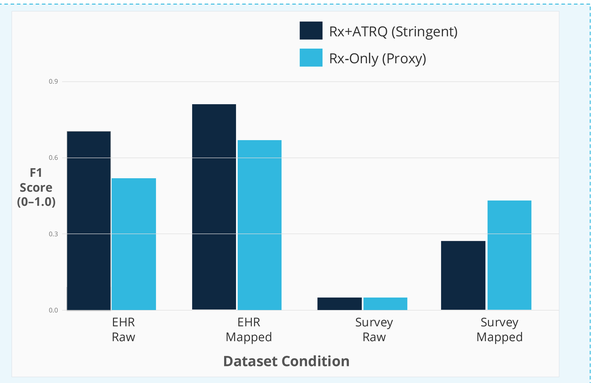
navidence.com

RESULTS

N=500 Synthetic Cohort Patients (Ground Truth)

30% Ground-Truth TRD Prevalence

8 CODef × Dataset Analysis Conditions



F1 score by dataset condition and CODef variant. Ontology mapping increases F1 for both CODefs across both data sources, with the largest gains on survey data (Rx+ATRQ: 0.05 → 0.27; Rx-Only: 0.05 → 0.43).
Rx-Only outperforms Rx+ATRQ on Survey Mapped data — patients self-report medication more readily than full ATRQ trial detail, suggesting CODef choice should be data-format-aware.

Condition	CODef	Sensitivity	Specificity	PPV	NPV	F1	Est. TRD Prevalence
EHR Raw	Rx+ATRQ	58.0%	97.7%	91.6%	84.4%	0.71	19.0%
EHR Raw	Rx-Only	36.7%	98.0%	88.7%	78.3%	0.52	12.4%
EHR Mapped	Rx+ATRQ	73.3%	96.6%	90.2%	89.4%	0.81	24.4%
EHR Mapped	Rx-Only	54.0%	96.6%	87.1%	83.0%	0.67	18.6%
Survey Raw	Rx+ATRQ	2.7%	100.0%	100.0 %	70.6%	0.05	0.8%
Survey Raw	Rx-Only	2.7%	100.0%	100.0 %	70.6%	0.05	0.8%
Survey Mapped	Rx+ATRQ	16.0%	98.6%	82.8%	73.2%	**0.27**	5.8%
Survey Mapped	Rx-Only	28.7%	98.0%	86.0%	76.2%	**0.43**	10.0%

CODef predictions versus ground-truth TRD labels (N=500; true prevalence 30%). Specificity and PPV remain consistently high across all conditions; sensitivity is the dominant axis of variation. Ontology mapping improves F1 for both CODefs across both data sources, w/ the largest raw → mapped gains on Survey data (bold).
Estimated TRD prevalence underestimates ground truth in every condition, reflecting low sensitivity rather than low precision.

Single-Patient Walkthrough | One synthetic True TRD patient through the pipeline

① Ground Truth

62 y/o M, MDD recurrent – 3 failed medication trials in distinct classes

- Escitalopram (SSRI)
  - 10.9mg × ~6 mo — "no significant improvement"
- Trazodone (Other antidepressant)
  - 364.8mg × 3 mo — "minimal change"
- Esketamine
  - 64.2mg × 3 mo — "no change"

✓ TRD POSITIVE (Ground Truth)

② EHR Notes | Raw Format → GenOMA Mapped

"Medication trials this episode:
· escitalopram 10.9mg, several months, no significant improvement
· trazodone 364.8mg × 3 mo, minimal change
· esketamine 64.2mg × 3 mo, no change"

✓ generic names mapped to 3 distinct AD classes
✓ doses, durations, and "failed" phrases parsed

✓ Rx+ATRQ CODef – 3 failed trials detected | ✕ Rx-Only – no evidence of non-AD trial

③ Patient Survey (raw)

"I tried Lexapro for maybe 2 months. It didn't change anything ... then Desyrel 364.8 milligrams for about 3 months. It barely helped ... then Spravato, I think the standard amount for a few months. It had no effect."

⚠ *Lexapro/Desyrel/Spravato are not in the generic-only dictionary*

✕ Rx + ATRQ – false negative, no AD trials detected | ✕ Rx-Only – false negative, no AD trials detected

④ Patient Survey (GenOMA-mapped)

Drug identifiers normalized
- Lexapro → escitalopram (SSRI)
- Desyrel → trazodone (Other AD)
- Spravato → esketamine (Esketamine)

Doses: assumed adequate based on available details
Durations: all ≥ 42 days
Phrases: "didn't change"/"barely helped"/"no effect" → all classified as failed

✓ Rx+ATRQ CODef – now able to detect 3 failed trials | ✕ Rx-Only – still no non-AD trial to detect

SYNTHETIC DATA GENERATION

Parameter	EHR notes	Survey
Pure-generic mentions	65%	40%
Pure-brand mentions	20%	60%
"Generic (brand)" parenthetical	15%	—
Brand-name misspellings	—	6%
Vague dose ("standard amount")	—	55%
Dose omitted	—	35%
Duration omitted	8%	20%
Copy-forward of inactive meds	10%	—
Titration sequences	20%	—
Ambiguous response phrasing	10%	15%

These parameters govern how each patient's structured trial history is rendered into EHR notes (clinician convention) and survey responses (patient self-report). Dashes denote parameters that don't apply to a given source — titration sequences and copy-forward are EHR-specific phenomena; vague dosing and misspellings are survey-specific. Values are literature-informed but not empirically calibrated against real corpora; see Limitations.

CONCLUSIONS

Key Findings

1. Raw free-text yields lower F1 of the stringent Rx+ATRQ CODef – ontology mapping significantly improves performance in both EHR and patient-survey formats.
2. GenOMA ontology mapping transforms both free-text formats into computable form, broadly enabling the stringent CODef in sources previously limited to proxy definitions.
3. Proxy CODefs (Rx-Only) identify fewer true cases than the stringent CODef when symptom-score data is available, but uniquely catch patients on survey data, where they outperform the stringent CODef.
4. CODef outputs represent a joint function of definition AND data source. Mapping improves CODef performance, but cannot recover information the text did not contain.

Implications

- Ontology mapping enables rigorous, symptom-anchored cohort definitions in EHR and patient-reported data – expanding evidence generation for TRD and any indication where key clinical signal is in unstructured text.
- Researchers should explicitly separate data-capture limitations from definitional limitations in CODef-based RWE studies.
- Publishers and regulators should require CODef applicability reporting alongside performance metrics.

LIMITATIONS

- Synthetic notes follow templated patterns; real EHR notes have greater idiosyncratic variation — production GenOMA performance may be lower than demonstrated.
- Circularity: projection phrasing libraries and extraction logic share conceptual alignment; this demonstrates computational feasibility, not real-world validation.
- Mapping accuracy is treated as near-perfect outside deliberately ambiguous cases; real-world errors would attenuate observed benefits.
- Direction of effect is robust; precise magnitudes depend on phrasing-library calibration.
- Single indication (TRD), two CODef variants — generalization requires parallel analysis for other conditions.