Performance of an AI-Powered Digital Phenotyping Platform for Patient Identification in Three Conditions With Systematic Undercounting in Real-World Data

Jigar Bandaria[1], Ombretta Palucci[2], Anna Swenson[1], Catherine Richards[1], Costas Boussios[1] | [1]OM1, Inc., Boston, MA, USA [2]OM1 B.V., Amsterdam, The Netherlands

Background

Patient identification in real-world data (RWD) sources often relies on standardized diagnostic codes, which fail to capture patients who are undiagnosed, misdiagnosed, or living with conditions lacking specific diagnostic codes. These patients remain unrecognized in care, so diagnosis and appropriate treatment are delayed or never happen.

AI applied to longitudinal patient histories can recognize the phenotype of a condition before, or without, a definitive code.

Objectives

To evaluate PhenOM®, an AI-powered digital phenotyping platform, across three conditions with distinct identification challenges:

- MASH (metabolic dysfunction-associated steatohepatitis): asymptomatic for years, with low-specificity billing codes
- Fabry disease (FD): rare, with heterogeneous multi-organ presentation
- Treatment-resistant depression (TRD): no ICD-10 code or DSM classification

Methods

- **Data:** longitudinal linked EMR and claims data from the OM1 Real-World Data Cloud (>370M de-identified U.S. individuals; average 10-year history), including unstructured clinical notes for a subset of patients, from which TRD attestations were derived.
- **Models:** individual-level phenotypic profiles built from structured history (diagnoses, procedures, medications, labs as normal/abnormal), used to score each patient's likelihood of the condition.
- **MASH:** ICD-10 K75.811; 21,593 cases among 4.8M adults sampled from the general population (0.45% prevalence).
- **FD:** ICD-10 E75.21 or FD-specific therapy2. The model was developed and evaluated on an FD-enriched dataset (4,978 cases, 1,000,000 controls); lift was then extrapolated to a population prevalence of 1 in 50,000, a conservative estimate based on published prevalence data.3
- **TRD:** psychiatrist-attested TRD in clinical notes within an MDD cohort4. By design, we fixed the study population at 20% TRD: 3,771 cases and 15,084 non-attested MDD controls (Figure 2). Only structured data were used as predictive features.
- **Evaluation:** for every model, all statistics were derived on a held-out dataset not used for training: AUROC; lift (prevalence in the top 10%, 1%, 0.1% of risk scores vs baseline); subgroup AUROC. Experts reviewed predictive features for clinical validity.

Results

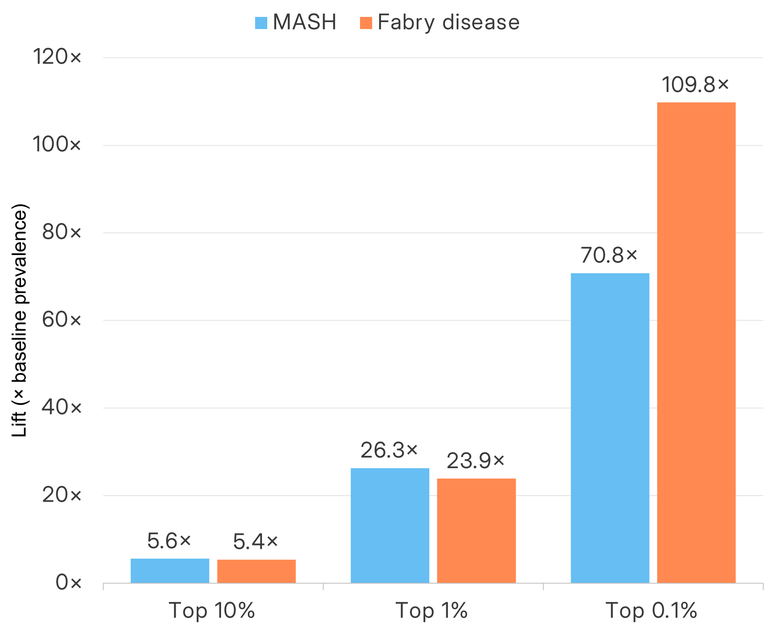
- **AUROC** was 0.82 (MASH), 0.82 (FD), and 0.87 (TRD), and was maintained by sex in FD (0.83 / 0.82) and TRD (0.88 / 0.87) and across TRD age and race subgroups (0.83–0.88).
- **Lift** rose steeply toward the top of the risk ranking: 5.6×, 26.3×, and 70.8× for MASH and 5.4×, 23.9×, and 109.8× for FD in the top 10%, 1%, and 0.1% (Figure 1).
- **TRD**: attested TRD rose from 1.6% in the lowest score decile to 86.8% in the highest (Figure 2).

MASH
0.82 AUROC

Fabry disease
0.82 AUROC
Male 0.83 · Female 0.82

TRD
0.87 AUROC
Male 0.88 · Female 0.87

Figure 1. Lift among the highest-risk patients: prevalence relative to cohort baseline



MASH

Baseline prevalence 0.45% (code-based prevalence in the population)

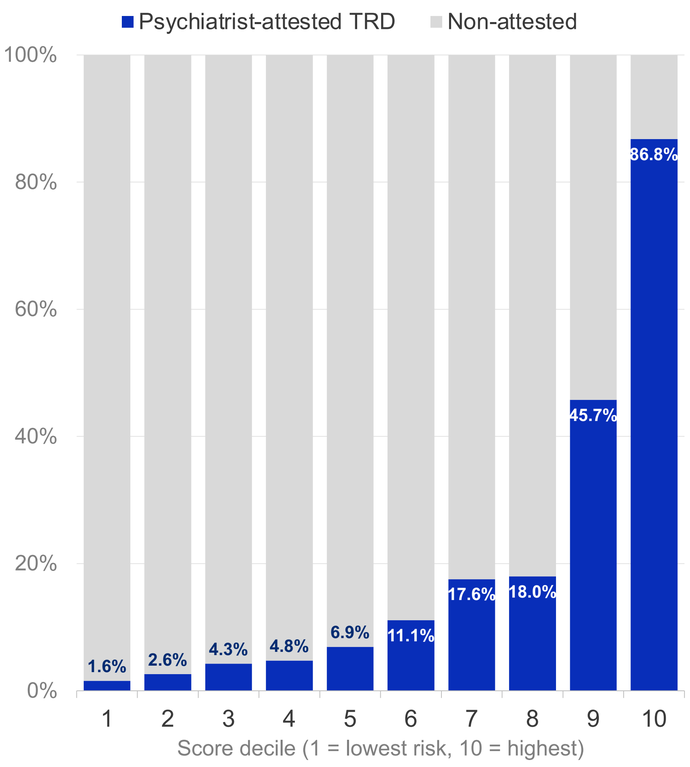
Risk group	Lift	Prevalence
Top 10%	5.6×	2.5%
Top 1%	26.3×	11.8%
Top 0.1%	70.8×	31.7%

Fabry disease

Lift from the enriched dataset, extrapolated to a conservative literature-based prevalence of 1 in 50,000[3]

Risk group	Lift	Projected prevalence
Top 10%	5.4×	1 in 9,261
Top 1%	23.9×	1 in 2,090
Top 0.1%	109.8×	1 in 455

Figure 2. Psychiatrist-attested TRD by model score decile in a 20% TRD / 80% non-attested MDD test population



Why a fixed 20% TRD population?

- Attested TRD captures only the TRD that clinicians document, and only in patients whose notes are available in the OM1 RWD Cloud.
- Note-based prevalence therefore understates true TRD. We fixed the test population at 20% attested TRD and 80% non-attested MDD patients (within the 6–55% real world reported range[5]).
- Against the 20% baseline, the top decile is 86.8% TRD; the top two deciles hold 66% of cases.
- Non-attested patients may include undocumented TRD, so top-decile precision is likely conservative.

Conclusions

- PhenOM identified patients with strong discrimination (AUROC 0.82–0.87) in three conditions where code-based definitions undercount, with performance maintained across subgroups.
- Likely cases concentrated in small, high-yield groups: ~25× lift in the top 1% and 71× (MASH) to 110× (FD) in the top 0.1%.
- AI-powered digital phenotyping can complement code-based cohort definitions in RWD studies and focus diagnostic screening where standard coding produces incomplete cohorts.

Conference

Presented at ISPOR Europe 2026, 8–11 November 2026, Vienna, Austria.
Disclosures: The original FD and TRD analyses were funded by Amicus Therapeutics and Compass Pathways Ltd, respectively.

References

1. Rinella ME, Lazarus JV, Ratziu V, et al. A multisociety Delphi consensus statement on new fatty liver disease nomenclature. J Hepatol. 2023;79(6):1542–1556.
2. Jefferies JL, Spencer AK, Lau HA, et al. A new approach to identifying patients with elevated risk for Fabry disease using a machine learning algorithm. Orphanet J Rare Dis. 2021;16:518.
3. Mehta A, Beck M, Eyskens F, et al. Fabry disease: a review of current management strategies. QJM. 2010;103(9):641–659.
4. Semeniuta D, Marci C, Bandaria J, et al. Identification of treatment resistant major depressive disorder using a machine learning algorithm. Presented at ICPE 2023; Halifax, Canada.
5. McIntyre RS, Alsuwaidan M, Baune BT, et al. Treatment-resistant depression: definition, prevalence, detection, management, and investigational interventions. World Psychiatry. 2023;22(3):394–412.

AUROC, area under the receiver operating characteristic curve; EMR, electronic medical record; FD, Fabry disease; MASH, metabolic dysfunction-associated steatohepatitis; MDD, major depressive disorder; RWD, real-world data; TRD, treatment-resistant depression.