AI–Derived Subgroups from Phase 2 Data Applied Post Hoc to Phase 3 MDD: Could Prior Design Adjustments Have Prevented Failure?

Joseph Geraci[1,2,3], Jan Sedway[1], Bessi Qorri[1], Mike Tsay[1], Adam Gogacz[1], Christian Cumbaa[1], Paul Leonchyk[1], Larry Alphs[1] , Luca Pani[4, 5]

[1] NetraMark Corp., Toronto ON, Canada; [2] Department of Pathology and Molecular Medicine, Queen's University, Kingston ON, Canada; [3] Arthur C. Clarke Centre for Human Imagination, School of Physical Sciences, University of California San Diego, CA USA; [4] Miller School of Medicine, University of Miami, FL USA; [5] University of Modena and Reggio Emilia, Modena, Italy

KEY FINDINGS

- Explainable ML-derived **multivariable** subgroups identified in Phase 2 recovered clinically meaningful drug–placebo separation when applied post hoc to a failed Phase 3 MDD trial
- Failure to control heterogeneity across phases can erase true treatment effects; **synergistic multivariable enrichment** improves Phase 2→3 generalization by reducing variance without overfitting.
- Using Phase 2–derived, clinically interpretable model derived subgroups can guide critical stratification decisions, offering a **regulator-aligned, operationally feasible** path to prospective MDD trial success.

BACKGROUND AND METHODOLOGICAL PROBLEM

- Late-phase clinical trials in MDD frequently fail as patient heterogeneity increases and placebo response amplifies from Phase 2 to Phase 3. Conventional subgrouping approaches—such as single-variable filters or exploratory post hoc analyses—often lack generalizability and may inflate false-positive findings.
- There is a critical need for stratification strategies that are **clinically interpretable, reproducible, and aligned with regulatory expectations**, while preserving true treatment signal and controlling variance in confirmatory trials.

Objective: To assess whether an explainable machine learning–based stratification framework could identify clinically meaningful, model-derived subgroups (MDS) in Phase 2 that, if prespecified, might have improved treatment–placebo separation in a subsequent failed Phase 3 trial of the same compound.

METHODS

Study Overview

Design: Retrospective translational analysis linking Phase 2 discovery to Phase 3 validation

Indication: Major Depressive Disorder (MDD) Treatment

Treatment Arms: REL-1017 / esmethadone, 25mg; Placebo

Primary Endpoint: Change from baseline in MADRS total score to primary endpoint (Day 7 for Phase 2; Day 28 for Phase 3)

Methods

A sub-insight learning (explainable machine learning) model was trained on baseline clinical variables from a positive Phase 2 study and applied without re-optimization to an independent Phase 3 dataset (post hoc translational analysis).

Phase 2: Discovery (N = 40; analysis set)

Model-Derived Subgroups (MDS) were restricted to **2–4 baseline variables** to preserve interpretability and manage the bias–variance trade-off.

The most robust MDS represented **38% of the Phase 2 sample at Day 7** and demonstrated **>17-point MADRS separation at Day 7**, compared with **~9 points in the overall population.(Cohen's D=1.37, p=.020)**. The loosened version of this MDS represented 55% of the Phase 2 sample at Day 7 with a >12. point separation (Cohen's D=1.03, p=.0299)

Low Appetite / Normal Blood Pressure MDS Characteristics

Reduced appetite: HAMD Somatic Symptoms—Gastrointestinal (SS GI) score = 2 (range 0–2; "difficulty eating without urging") at screening

Normal blood pressure: Systolic blood pressure (SBP) <132 mmHg (observed range: 108–143) at screening

Phase 3: Translation (N = 208; analysis set)

The top Phase 2-derived MDS was applied **post hoc** to an independent Phase 3 trial that did not meet its primary endpoint.

Two prespecified MDS operationalizations and corresponding single-variable comparators were evaluated:

MDS Applications

- **Strict MDS:** HAMD SS GI = 2 + SBP <132 mmHg
- **Loosened MDS:** HAMD SS GI >0 + relaxed SBP threshold

Single-Variable Comparators

- **HAMD SS GI:** SS GI = 2; SS GI >0
- **Systolic Blood Pressure:** SBP <132 mmHg; SBP <137 mmHg

RESULTS

Strict MDS Application

The **a priori Phase 2–derived MDS** demonstrated strong replication in Phase 3, with a **5.93-point drug–placebo separation** versus **0.94 points in the overall population** on Day 28 MADRS total score (**Cohen's d = 0.53; p = 0.025**)*. This subgroup comprised **38% of the Phase 3 sample**. Relative to the full dataset, this population showed **greater separation driven by both reduced placebo response (+1.4 points)** and **enhanced drug response (+3.6 points)**, yielding an additional **~5-point net separation**.

Loosened MDS Application

Relaxation of scale thresholds (HAMD-SS GI score 1–2; systolic blood pressure <137mmHg) increased subgroup size to **61% of the sample**, with attenuation of effect size (**Cohen's d = 0.35; p = 0.053**). Compared with the overall population, this group demonstrated **modest reduction in placebo response (+0.5 points)** with **enhanced drug response (+2.3 points)**, resulting in **~2.8 points of incremental separation**.

Single-Variable Comparator Analyses

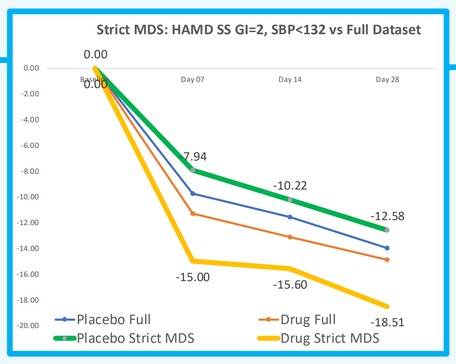
HAMD-SS GI

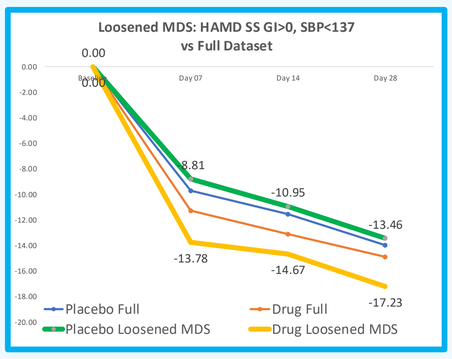
HAMD-SS GI score = 2 alone approached statistical significance (**Cohen's d = 0.36; p = 0.062**), encompassing **67% of participants (139/208)** with a **4.1-point separation which reduced placebo response (+1.7 points) and enhanced drug response (+1.5 points)** relative to the full dataset resulting in **~3.2 points of incremental separation**.

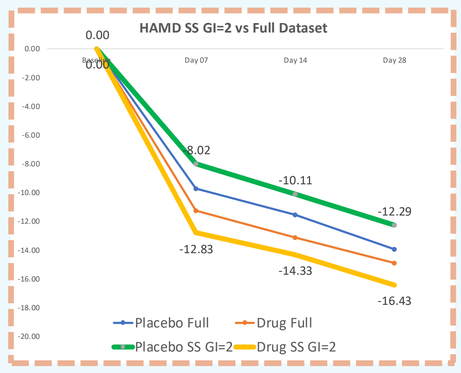
Further relaxation (HAMD-SS GI > 0) preserved the enhanced drug response but eliminated the reduced placebo response present in HAMD SS GI=2.

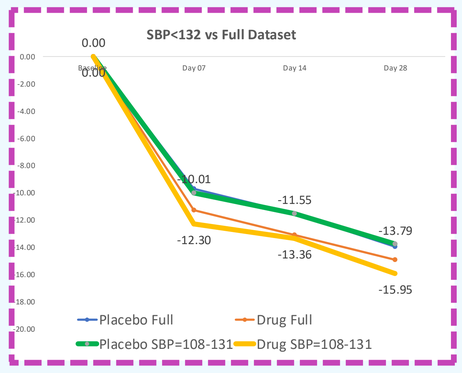
Systolic Blood Pressure

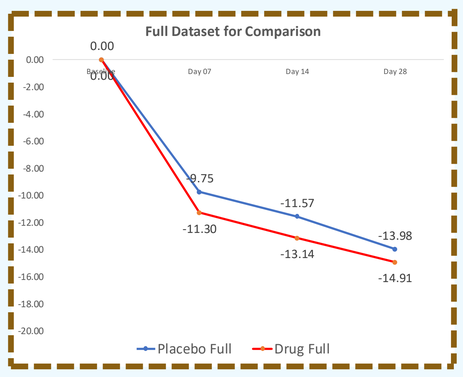
SBP thresholds (<132 or <137 mmHg) did not achieve statistical significance individually, though both showed **~2 points of additional separation with** enhanced drug response compared the full dataset. Only the <137 mmHg subgroup had reduced placebo response compared to the full dataset.



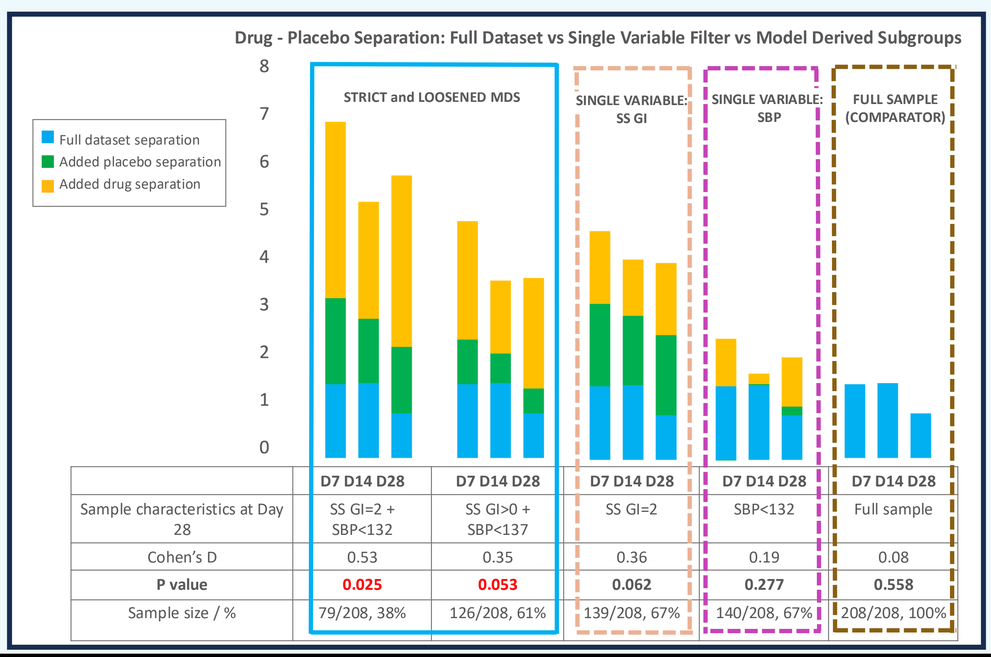








*Effect sizes shown as Cohen's d; p-values two-sided (unadjusted), computed on the Phase 3 analysis set.



	D7 D14 D28	D7 D14 D28	D7 D14 D28	D7 D14 D28	D7 D14 D28
Sample characteristics at Day 28	SS GI=2 + SBP<132	SS GI>0 + SBP<137	SS GI=2	SBP<132	Full sample
Cohen's D	0.53	0.35	0.36	0.19	0.08
P value	**0.025**	**0.053**	**0.062**	**0.277**	**0.558**
Sample size / %	79/208, 38%	126/208, 61%	139/208, 67%	140/208, 67%	208/208, 100%

CONCLUSIONS & IMPLICATIONS

Conclusions

Broadening subgroup definitions for greater inclusivity preserved treatment effect direction and improved feasibility, but single-variable reduction weakened the multivariable signal.

Prespecified, prospective use of Phase 2–derived MDS for Phase 3 stratification would likely have enabled detection of a treatment effect.

Implications for Future Trials

Phase 2–derived MDS support scientifically justified Phase 3 stratification for adequate and balanced representation of likely responders.

Multivariable MDS have better stability and robustness than single-variables and provide a stronger foundation for stratification.

Applying MDS enables a data-driven, interpretable, and regulator-aligned approach to Phase 3 trial design.

NETRAMARK

DISCLOSURES

Dr. Joseph Geraci is the founder of NetraMark and is a significant shareholder of NetraMark Holdings, which is a publicly traded company. Luca Pani, Jan Sedway, Christian Cumbaa, Bessi Qorri, Mike Tsay, Paul Leonchyk and Larry Alphs are employed by NetraMark. Dr. Luca Pani's Disclosures (past 3 years): Luca Pani is/has been a consultant or advisory board member for AbbVie, USA; BCG, Switzerland; Boehringer Ingelheim International GmbH, Germany; GH-Research, Ireland; Immunogen, USA; Johnson & Johnson USA; LB Pharmaceuticals, USA; Magdalena BioSciences, USA; Sanofi-Aventis-Genzyme, France and USA; Lundbeck, Denmark and Italy; Napo-Pharma, USA and EU; NetraMark, Canada; Pfizer Global, USA; Relmada Therapeutics, USA; Takeda, USA and owns shares/options from AiDvance Germany; Adapt UK, Enzeta USA, NetraMark Canada, and Relmada, USA.