From Data to Digital Phenotypes

Modelling High-Dimensional PROs and Actigraphy Data in Decentralized Psychiatric Trials

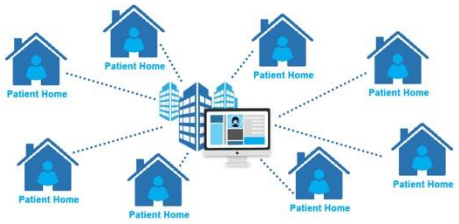
Sakshi Kaushik, PhD

Agenda

- Introduction & Challenge
- Hybrid Statistical Framework
- Results & Performance
- Discussion & Conclusion
- Limitations & Future Scope
- References

The Landscape: Decentralized Trials

Moving beyond the clinic: The rapid adoption of remote technologies is creating a new era of “Digital Psychiatry”.



Decentralized Trials

Remote monitoring allowing for real world evidence collection outside traditional setting

Actigraphy

Wearable-derived actigraphy data : Sleep efficiency, circadian rhythm, physical activity, etc..

ePROs

Subjective symptom data: mood, fatigue, concentration, anxiety, insight, etc..

The Analytical Challenge

Decentralized trials using wearable devices and ePROs generate vast amounts of high-dimensional longitudinal data.

Traditional methods struggle with:

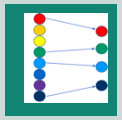
Multicollinearity	High correlation between predictors
Time dependency	Repeated measures
Sparse features	Small N, large P problem
Heterogeneity	Individual differences in treatment response

The Solution: Hybrid Framework

We propose a two-stage approach combining variable selection with modelling binary outcome over time

INPUT: High-dimensional Longitudinal Data

PROs (Mood, Anxiety, etc.)
Actigraphy data (Sleep, activity, etc.)



STAGE 1: Adaptive LASSO

Reduces dimensionality
Penalized regression

STAGE 2: Mixed-Effects Logistic Regression Model

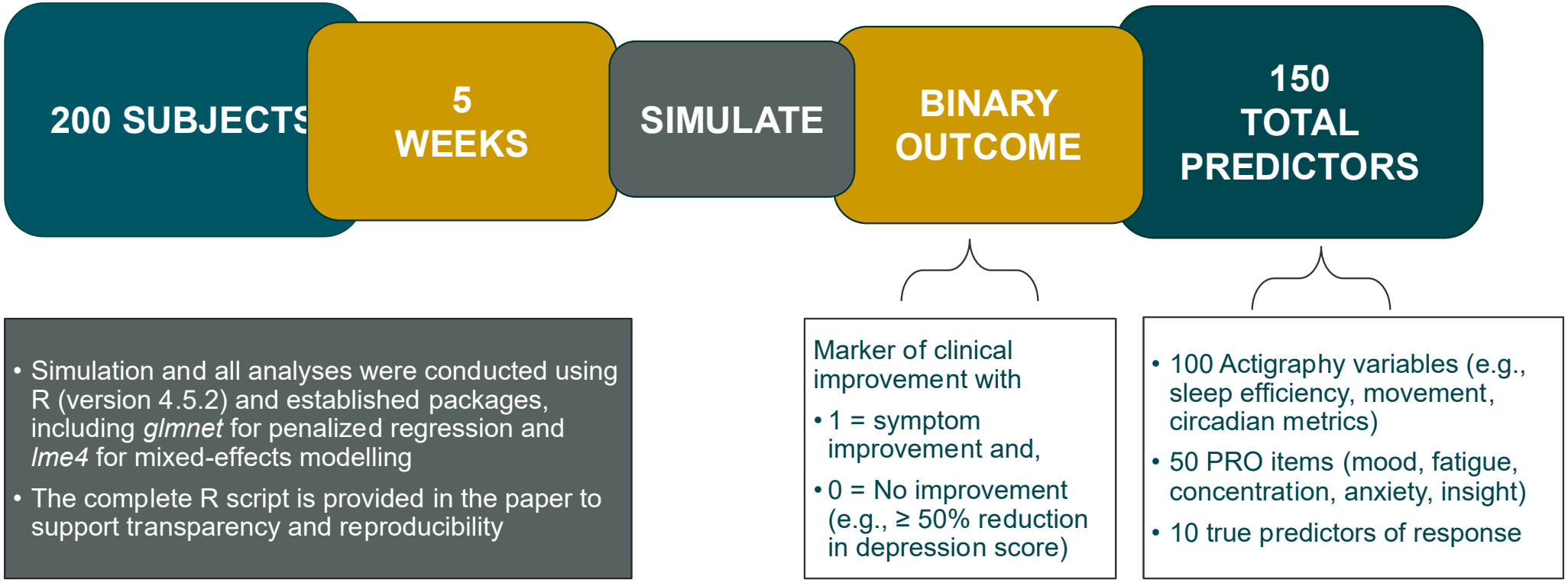
Inference (Random intercepts)
Repeated measurements

OUTPUT

Interpretable digital phenotypes

Simulation Study Design

To validate the framework, we simulated a realistic decentralized psychiatric trial environment



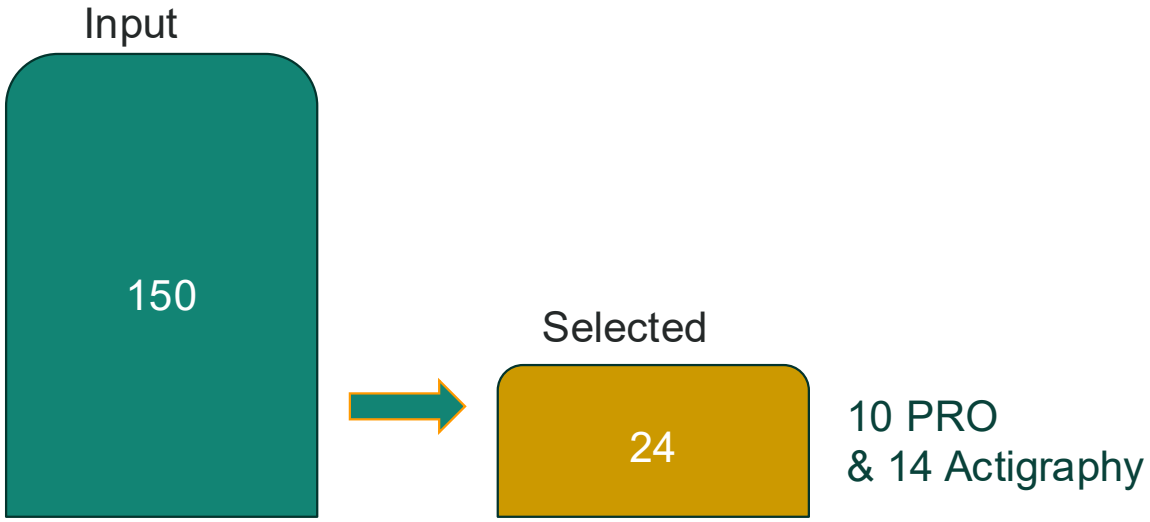
Results: Feature Selection Via Adaptive LASSO

The Adaptive LASSO model :

$$logit\,(\Pr(y_{it} = 1)) = x_{it}^{\intercal}\beta$$

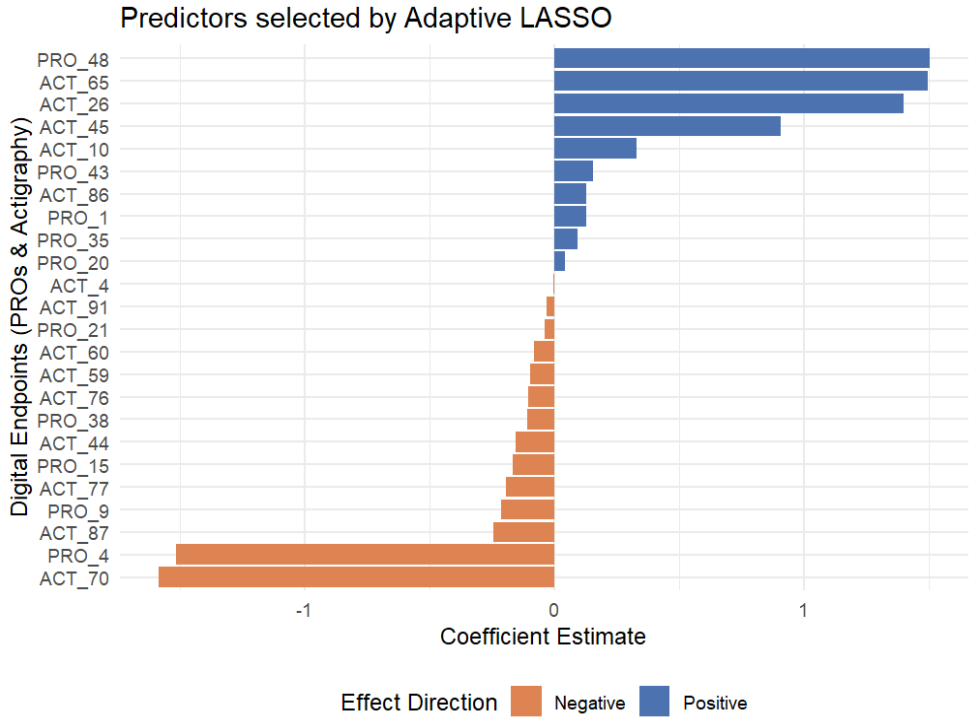
$$\hat{\beta} = arg\min_{\beta}\left\{-l(\beta) + \lambda\sum_{j=1}^{p+q} w_j|\beta_j|\right\}$$

where, $w_j = \frac{1}{|\hat{\beta}_j|^{\gamma}}$ (Zou, 2006; Li et. al., 2017)



Adaptive LASSO reduced the feature space by 84% with optimal $\lambda = 0.5268$ selected based on 10-fold cross validation

Coefficient plot of adaptive LASSO-selected predictors, with colours indicating direction of association



Let: $y_{it}\,\epsilon\,\{0,1\}$: denote a binary clinical improvement indicator variable for subject i at time t $(i = 1, \ldots, N; t = 1, \ldots, T)$

x_{it}: denote high-dimensional covariate matrix of size $(p+q)$, concatenating PRO and actigraphy features.

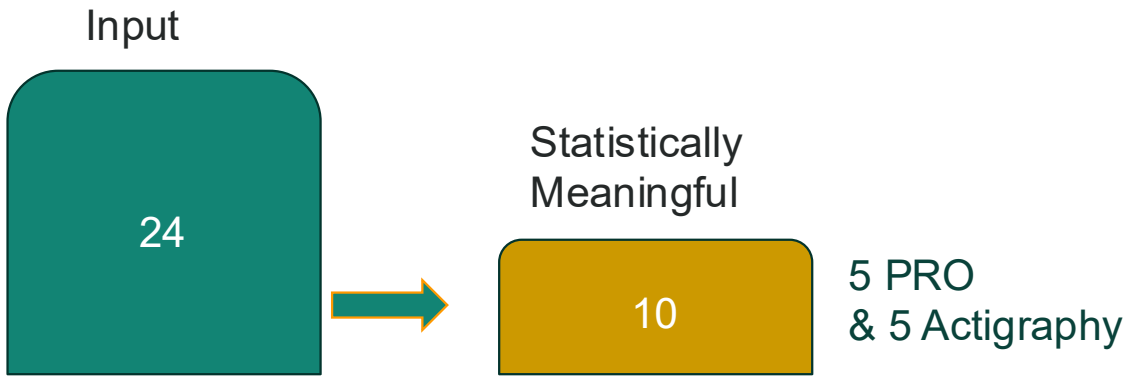
Results: Mixed-effects Logistic Regression

Selected predictors $z_{it} \subset x_{it}$ are analysed using below model:

$$logit\,(\Pr(y_{it} = 1)) = \alpha_i + z_{it}^{\intercal}\beta$$

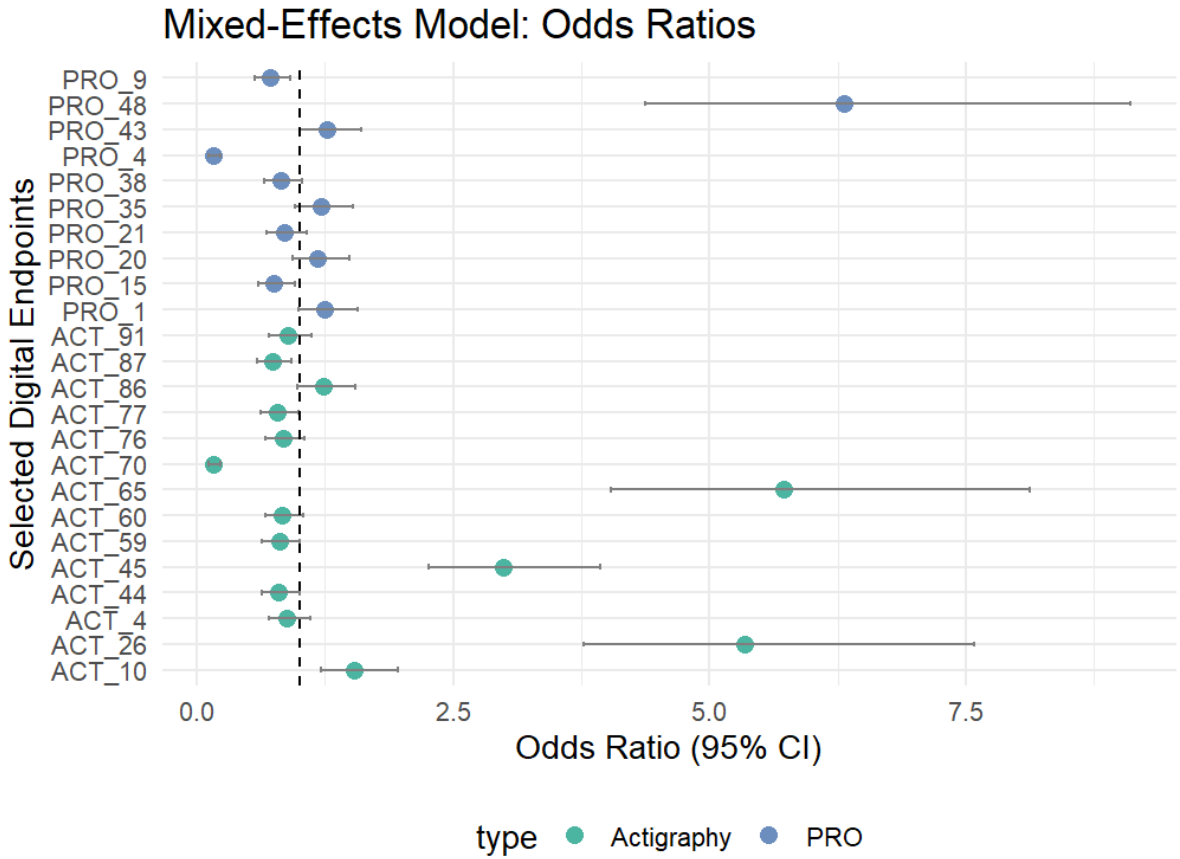
where,

$\alpha_i \sim N(0, \sigma^2)$ are subject-level random effect (Verbeke & Molenberghs, 2000; Fitzmaurice et. al., 2011)



Random intercept captures subject heterogeneity, and selected 10 statistically meaningful predictors (5 PRO & 5 actigraphy) having OR > 1

Forest plot of odds ratios and 95% CIs for fixed effects from mixed-effects logistic regression model



Model Performance: Adaptive LASSO

PERFORMANCE EVALUATIONS

Area under the curve = 0.9416

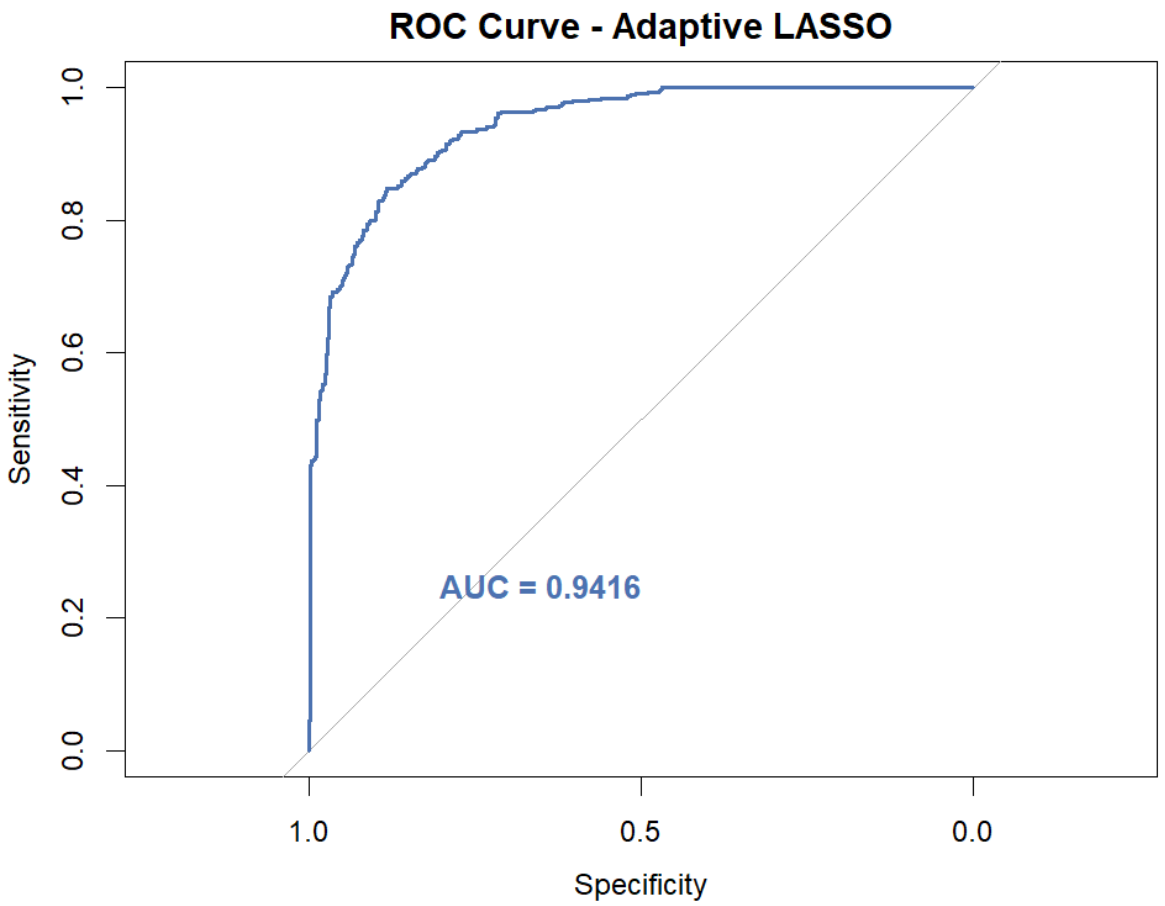
Classification accuracy* = 0.859

MSE = 0.098, RMSE = 0.313

Collectively, these results indicate that the selected predictors collectively retained the predictive signal

* Evaluated at a probability threshold of 0.5

ROC curve illustrating discriminative performance of adaptive LASSO

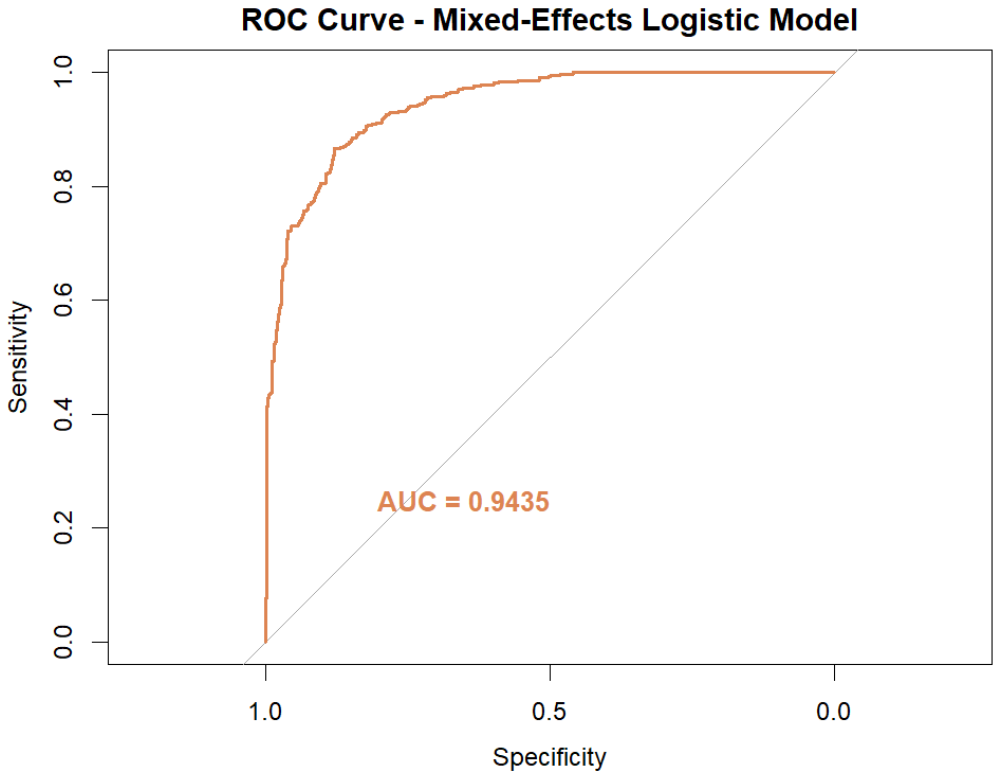


Model Performance: Mixed-effects Logistic Regression

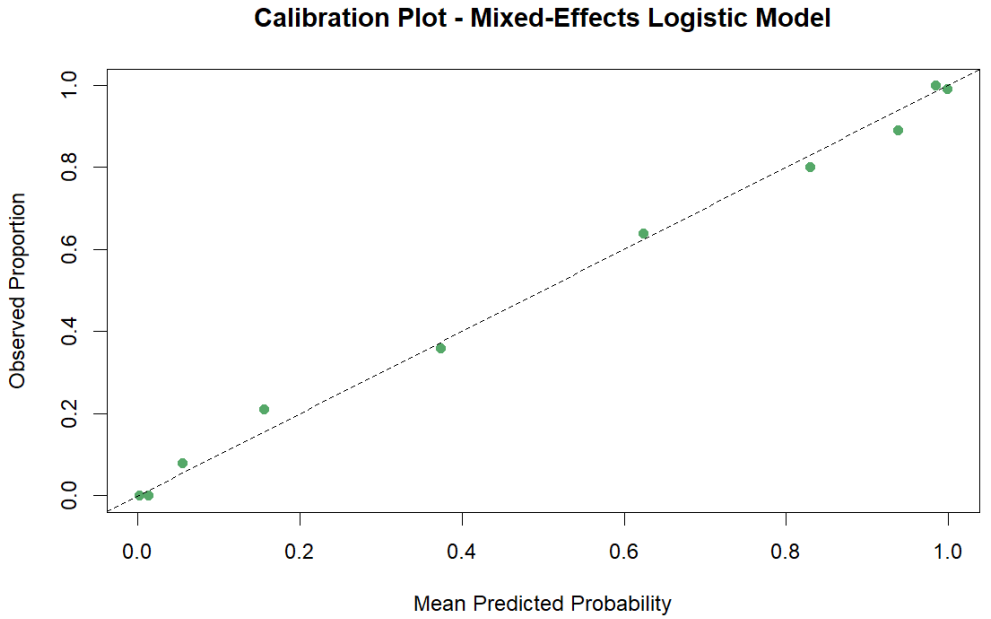
PERFORMANCE EVALUATIONS

Area under the curve = 0.9435	Classification accuracy* = 0.869	Brier score = 0.0963

ROC curve illustrating discriminative performance of Mixed-effects logistic model



Calibration plot comparing predicted probabilities and observed response rates for mixed-effects logistic regression model



Collectively, these results demonstrate that the mixed-effects logistic regression model is suitable to perform longitudinal analysis of high dimensional digital and PRO data.

* Evaluated at a probability threshold of 0.5

Discussion & Conclusion

- ✓ Structured, practical and transparent framework for high-dimensional digital data
- ✓ Supports hypothesis generation and endpoint refinement
- ✓ Clear separation of screening and longitudinal inference
- ✓ Balances interpretability and scalability
- ✓ Aligned with clinical exploratory clinical trial analyses

Practical Value & Clinical Relevance

Existing Tools

No new methods- thoughtful integration of proven statistical approaches

Protocol Template

Can inform future protocol and SAP development for digital endpoint trials

Implementation

Easily implementable using standard R packages

Suitability

Suitable for decentralized and hybrid trials and supports digital biomarker evaluation

Limitations & Future Scope

Limitations

- ✓ Simulated Data: Real-world validation needed
- ✓ Correlation: LASSO may select more correlated features
- ✓ Validation: Lack of standard cross-validation for mixed models
- ✓ Binary outcome: Currently limited to binary response

Future Scope

- ✓ Real-world Validation: Apply framework to actual psychiatric trial data
- ✓ Advanced Methods: Explore group LASSO, Bayesian hierarchical shrinkage approaches
- ✓ Temporal Dynamics: Incorporate time-varying effects
- ✓ New Outcomes: Extend to continuous or time-to-event endpoints

References

- Izmailova Es, Wagner JA, Perakslis ED (2018). Wearable devices in clinical trials: hype and hypothesis. Nat Rev Drug Discov.
- Ticona, E et al. (2023). Decentralised, patient-centric, site-less, virtual and digital clinical trials? Drug Discovery today, 28, 103520.
- Matcham F et al. (2019). Remote assessment of disease and relapse in major depressive disorder (RADAR-MDD). JMIR
- Zou, H. (2006). The Adaptive Lasso and Its Oracle Properties. Journal of the American Statistical Association.
- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. Journal of Statistical Software.
- Lei, Y., et al. (2024). A Machine Learning Framework for Digital Biomarkers. JMIR.
- Li, C., & Houts, C. (2021). Longitudinal Analysis of PROs in Clinical Trials. Journal of Biopharmaceutical Statistics.
- Steyerberg et al (2019). Clinical prediction models. Springer.

Thanks & Questions?

- Thank you
- Refer to the paper AS22 for details
- Feel free to contact at Sakshi.Kaushik@iconplc.com for any questions and comments

ICONplc.com