DeMMO: Longitudinal and Cross-Disease Modelling of Digital Mobility Outcomes via Multi-Task Learning

Menghui Zhou **Zhipeng Yuan** **Vitaveska Lanfranchi** **Po Yang**
School of Computer Science, University of Sheffield, Sheffield, United Kingdom
{menghui.zhou,zhipeng.yuan,v.lanfranchi,po.yang}@sheffield.ac.uk

Abstract

Digital mobility outcomes (DMOs) derived from wearable sensors characterise mobility in daily life and offer a promising means of monitoring disease progression. Yet most DMO studies examine one disease at one visit; they do not model how multivariate DMO relationships with multiple clinical outcomes evolve jointly across diseases. Technically, existing temporal multi-task frameworks can model progression within an individual disease, but they do not jointly model multiple prediction outcomes across diseases, particularly when disease cohorts do not share participants. To address these gaps, we propose DeMMO, an interpretable framework for longitudinal, multi-disease, and multi-outcome learning. DeMMO represents each disease–outcome objective by a longitudinal DMO coefficient matrix and combines temporal regularisation with stable and visit-specific feature selection. Its central technical contribution is an automatic cross-disease and cross-outcome relation-learning mechanism that learns signed relations directly from these longitudinal mappings, enabling selective information sharing without paired participants. We evaluate DeMMO on the recently released, large-scale, multicentre Mobilise-D dataset, which provides a new opportunity to study 24 harmonised real-world DMOs over five visits across multiple mobility-limiting conditions. Against nine strong linear, longitudinal, and deep-regression baselines, DeMMO achieves the best overall and outcome-specific prediction performance, with significant improvements over the strongest baselines. Stability selection further identifies reliable longitudinal DMO patterns for subsequent clinical validation and disease monitoring. The implementation code and experimental results are available at https://github.com/menghui-zhou/DeMMO.

1 Introduction

Disease monitoring typically relies on periodic clinical assessments that may require specialised resources, depend on evaluator judgement, or be affected by recall and individual perceptions (Delgado-Ortiz et al., 2023; Deane et al., 2014; Port et al., 2021; Prince et al., 2008). Assessments performed weeks or months apart also provide only intermittent observations and may miss gradual, fluctuating, or transient changes (Babrak et al., 2019; Coran et al., 2020). These limitations are particularly relevant to PD, MS, COPD, and PFF, which impose substantial personal and socioeconomic burdens (Yang et al., 2020; Bouleau et al., 2022; Boers et al., 2025; Williamson et al., 2017). Mobility reflects disease severity, progression, treatment response, and recovery, yet brief supervised tests measure what an individual is *capable of doing*, rather than what they *actually do* (Rochester et al., 2020). A six-minute walk test, for example, samples only approximately 0.06% of a week and cannot represent the diversity of daily mobility.

Advances in wearable sensors, low-power computing, and digital health offer a means of addressing this gap. Wearable inertial sensors can record movement over extended periods, while analytical algorithms transform acceleration and angular-velocity signals into digital mobility outcomes, or DMOs (Goldsack et al., 2020; Cerreta et al., 2020). DMOs characterise daily-life walking in terms of amount, pattern, pace, rhythm, and variability, capturing both *how much* and *how* individuals walk (Rochester et al., 2020; Kluge et al., 2021). By capturing walking bouts across different

days and contexts, DMOs provide denser and more representative longitudinal observations than episodic clinical assessments (Alcock et al., 2025). Most DMO studies validate individual measures cross-sectionally through correlations, linear models, or group comparisons (Polhemus et al., 2021; Mikolaizak et al., 2022; Megaritis et al., 2026; Eckert et al., 2026). Longitudinal evidence remains disease-specific and limited: studies in PD, hip-fracture recovery, and MS analysed small longitudinal samples or only a few prespecified measures (Mirelman et al., 2024; Engdal et al., 2024; Poleur et al., 2026). Consequently, the longitudinal relationships between comprehensive DMO profiles and multiple clinical outcomes remain unclear (Rábano-Suárez et al., 2025; Kirk et al., 2025). A major barrier has been the lack of large, harmonised longitudinal datasets combining repeated measurements of a common DMO set with multiple clinical outcomes across diseases (Rochester et al., 2020). Earlier resources were commonly cross-sectional, disease-specific, heterogeneous in sensing and aggregation, or limited to few assessment periods, preventing systematic study of stable, visit-specific, and cross-disease DMO patterns.

The recently released Mobilise-D dataset was developed to address this gap (Mikolaizak et al., 2022; Rochester et al., 2020; Kirk et al., 2024). It is the principal data resource produced by a major European public–private programme involving 35 partners from 13 countries. Conducted from 2019 to 2024, the programme represented a total investment of approximately EUR 49.9 million, including EUR 25.4 million contributed by the European Union. Mobilise-D first established the technical validity of its sensor and analytical pipeline through a separate multicentre study of 108 participants spanning healthy ageing and five mobility-limiting conditions. The subsequent CVS extended this foundation to clinical validation at an unprecedented scale: 2,400 participants across PD, MS, COPD, and PFF underwent a comprehensive clinical assessment and seven days of real-world mobility monitoring on five occasions, separated by six-month intervals over two years. The released resource combines harmonised walking-bout, daily, and weekly DMOs from a common lower-back sensor with clinical tests, clinician-reported assessments, and patient-reported outcomes. This combination of scale, longitudinal depth, cross-disease breadth, and standardised measurement makes Mobilise-D uniquely suited to studying how multivariate DMO–outcome relationships evolve during disease progression and functional recovery.

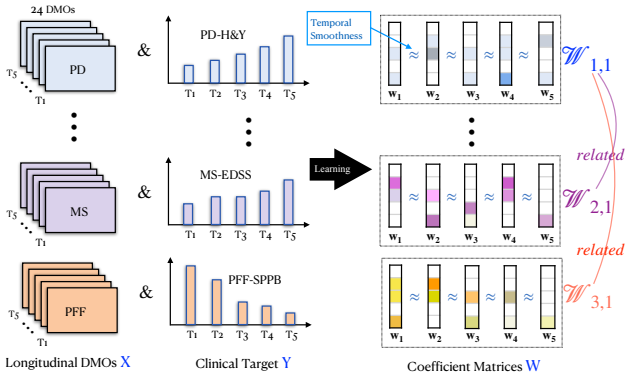

Figure 1: Overview of DeMMO. For each outcome, visit-specific coefficients map 24 longitudinal DMOs to clinical targets. DeMMO jointly learns temporally smooth mappings and selective within- and cross-disease relations.

Analysing this resource requires identifying stable and visit-specific effects among correlated DMOs, allowing DMO–outcome relationships to evolve over time, and learning shared structure across outcomes whose disease cohorts contain different participants.

We summarise the main contributions of this work as follows:

- **Technical contribution.** We expose a fundamental limitation of existing temporal multi-task learning: although it can model progression within a single disease, it cannot jointly learn multiple outcomes across diseases, let alone accommodate non-overlapping cohorts. To overcome this limitation, we introduce DeMMO and, at its core, a novel automatic cross-disease and cross-outcome relation-learning mechanism. It learns a signed relation graph directly from longitudinal DMO mappings, enabling selective knowledge transfer across unpaired cohorts while retaining outcome-specific temporal evolution and feature selection.
- **Practical contribution.** We fill an important gap in DMO research, where most studies examine one disease at one visit and therefore do not model multivariate DMO–outcome progression within and across diseases. Using the recently released Mobilise-D dataset, DeMMO jointly studies multiple outcomes over five visits and several mobility-limiting conditions. Against a broad and diverse set of nine strong baselines, spanning conventional

linear, structured longitudinal, and recent deep-regression methods, DeMMO consistently achieves superior prediction performance, with statistically significant improvements over the strongest baselines.

- **Potential clinical contribution.** We apply longitudinal stability selection (Zhou et al., 2024; Fan et al., 2026) to identify robust, outcome-specific DMO patterns. This advances a central goal of DMO research by identifying reliable candidate DMOs for clinical validation and disease monitoring.

Further discussion of DMO validation, longitudinal modelling, multi-task learning, and disease-progression methods is provided in Appendix A.

2 METHOD

2.1 MOBILISE-D LONGITUDINAL DATASET

To ground the proposed methodology in a concrete setting, we first introduce the recently released Mobilise-D Clinical Validation Study, a multicentre longitudinal DMO study of people with mobility-limiting conditions (Mazzà et al., 2021; Mikolaizak et al., 2022). The dataset contains mutually exclusive PD, MS, COPD, and PFF cohorts. Participants were scheduled for five visits, $T_1, \ldots, T_5$, followed by real-world monitoring with a body-worn inertial sensor. We analyse PD, MS, and PFF; COPD is excluded because the selected clinical outcomes are unavailable at T_2 and T_4. We use filtered weekly aggregates containing at least three valid monitoring days. Each record contains 24 DMOs describing walking amount, bout pattern, pace, rhythm, and variability. The quality-control variable `n_days_w` is not used as a predictor. We consider four regression objectives: H&Y stage (PD H&Y) (Hoehn & Yahr, 1967) and MDS–UPDRS Part III (PD MDS–UPDRS) (Goetz et al., 2008) for PD, EDSS for MS (MS EDSS) (Kurtzke, 1983), and SPPB for PFF (PFF SPPB) (Bellettiere et al., 2020). All outcomes retain their original scales. Because higher SPPB indicates better function, we define SPPB impairment as $12 - \mathrm{SPPB}$, so that higher values consistently indicate greater severity. Invalid values, including MDS–UPDRS scores of -1, are treated as missing. Complete-case processing is applied at the participant–visit level. A record is retained only when its clinical outcome and all 24 DMOs are available; missing values are not imputed. Participants may contribute records at any available visits and are not required to complete all five. Table 1 summarises the resulting dataset. Sample sizes decrease across visits because of loss to follow-up, unavailable recordings, insufficient monitoring days, and incomplete clinical assessments.

Table 1: Mobilise-D analytical dataset after requiring a valid target and complete observations for all 24 weekly DMOs. N denotes participants with at least one retained visit.

Disease	Target	N	T_1	T_2	T_3	T_4	T_5	Total	Range
PD	H&Y stage	574	528	439	400	363	346	2,076	0–4
PD	MDS–UPDRS III	574	528	439	400	363	346	2,076	1–84
MS	EDSS	578	535	448	402	346	352	2,083	0–7.5
PFF	SPPB impairment	469	418	358	282	146	108	1,312	0–11

2.2 LONGITUDINAL CLINICAL-PROGRESSION FORMULATION

We first consider one clinical outcome from a single disease cohort. This simplified setting introduces the longitudinal modelling, temporal regularisation, and DMO-selection components before their extension to multiple diseases and outcomes in Section 2.3.

At visit t, let $\mathbf{X}_t \in \mathbb{R}^{n_t \times p}$ denote the DMO feature matrix for n_t retained records, where $p = 24$, and let $\mathbf{y}_t \in \mathbb{R}^{n_t}$ contain the corresponding clinical outcomes. Unlike conventional progression models that use baseline measurements to predict future outcomes, Mobilise-D provides both DMOs and clinical outcomes at each of $T = 5$ visits. We therefore treat clinical-status estimation at each visit as a separate but related regression task: $\mathbf{y}_t = \mathbf{X}_t \mathbf{w}_t + \boldsymbol{\epsilon}_t$, where $\mathbf{w}_t \in \mathbb{R}^p$ is the visit-specific coefficient vector and $\boldsymbol{\epsilon}_t$ denotes observational noise. The coefficients are collected as $\mathbf{W} = [\mathbf{w}_1, \ldots, \mathbf{w}_T] \in \mathbb{R}^{p \times T}$. Each column represents the DMO–outcome relationship at one visit, while each row describes the longitudinal coefficient trajectory of one DMO. We use squared-error regression because the

clinical outcomes are numerical scores. To prevent visits with larger samples from dominating estimation, the loss is normalised within each visit: $\mathcal{L}(\mathbf{W}) = \sum_{t=1}^{T} \frac{1}{2n_t} \|\mathbf{y}_t - \mathbf{X}_t \mathbf{w}_t\|_2^2$.

Fused Temporal Smoothness. We model temporal smoothness using the fused Lasso, a well-established regulariser for longitudinal biomarker analysis (Zhou et al., 2012; 2023b). Let $\mathbf{R} \in \mathbb{R}^{(T-1)\times T}$ be the first-order difference matrix, with $R_{i,i} = -1$, $R_{i,i+1} = 1$, and all other entries zero. We use $\Omega_{\mathrm{FL}}(\mathbf{W}) = \|\mathbf{W}\mathbf{R}^{\mathsf{T}}\|_1$, which encourages adjacent visits to share coefficients while retaining temporally localised changes. This convex structure and its efficient optimisation have been established in previous work (Zhou et al., 2012; 2023b).

DMO Discovery via Sparse Group Lasso. Following Zhou et al. (2011), we use sparse group Lasso for interpretable DMO selection. Its element-wise term selects visit-specific effects, whereas its row-wise term selects DMOs across the longitudinal period: $\Omega_{\mathrm{SGL}}(\mathbf{W}) = \lambda_{\mathrm{L}}\|\mathbf{W}\|_1 + \lambda_{\mathrm{G}}\|\mathbf{W}\|_{2,1}$, where $\|\mathbf{W}\|_{2,1} = \sum_{j=1}^{p} \|\mathbf{w}_{j:}\|_2$. Thus, λ_{L} controls visit-specific sparsity and λ_{G} controls DMO-level sparsity. Sparse group Lasso is a mature convex feature-selection structure with efficient proximal algorithms (Tibshirani, 1996; Zhou et al., 2022a); selected DMOs are interpreted as candidates requiring subsequent clinical validation.

Combining the regression loss, fused temporal regularisation, and sparse group Lasso yields a representative formulation from an influential line of temporal multi-task learning research (Zhou et al., 2011; 2012; Cao et al., 2017; Zhou et al., 2022a; Zhou & Yang, 2023; Zhou et al., 2023b; Fan et al., 2026):

$$\min_{\mathbf{W}} \quad \sum_{t=1}^{T} \frac{1}{2n_t} \|\mathbf{y}_t - \mathbf{X}_t \mathbf{w}_t\|_2^2 + \lambda_{\mathrm{FL}} \|\mathbf{W}\mathbf{R}^{\mathsf{T}}\|_1 + \lambda_{\mathrm{L}} \|\mathbf{W}\|_1 + \lambda_{\mathrm{G}} \|\mathbf{W}\|_{2,1}. \tag{1}$$

However, Equation 1 models only one clinical outcome within a single disease cohort. It cannot jointly model multiple outcomes across diseases, let alone accommodate disease cohorts with non-overlapping participants. We next extend this formulation to address both settings.

2.3 Multi-Disease and Multi-Outcome Extension

Our analysis includes multiple prediction objectives across three cohorts: PD contributes two clinical outcomes, whereas MS and PFF each contribute one. The number of prediction objectives therefore differs from the number of diseases. Let $d \in \{1, \ldots, D\}$ index the disease cohorts, where $D = 3$, and let Q_d denote the number of clinical outcomes for disease d. The total number of disease–outcome prediction objectives is $M = \sum_{d=1}^{D} Q_d = 4$.

For disease d, outcome q, and visit t, let $\mathbf{X}_{t,q}^{(d)} \in \mathbb{R}^{n_{dtq} \times p}$ contain the complete DMO observations and let $\mathbf{y}_{t,q}^{(d)} \in \mathbb{R}^{n_{dtq}}$ contain the corresponding clinical outcomes. Here, n_{dtq} is the number of retained participant–visit records. Although all objectives share the same 24 DMO definitions, allowing $\mathbf{X}_{t,q}^{(d)}$ to depend on q accommodates outcome-specific differences in data availability. The visit-specific regression model is

$$\mathbf{y}_{t,q}^{(d)} = \mathbf{X}_{t,q}^{(d)} \mathbf{w}_{t,q}^{(d)} + \boldsymbol{\epsilon}_{t,q}^{(d)}, \tag{2}$$

where $\mathbf{w}_{t,q}^{(d)} \in \mathbb{R}^p$. The longitudinal coefficient matrix for disease d and outcome q is $\mathbf{W}_{d,q} = [\mathbf{w}_{1,q}^{(d)}, \ldots, \mathbf{w}_{T,q}^{(d)}] \in \mathbb{R}^{p \times T}$. Thus, each $\mathbf{W}_{d,q}$ represents one longitudinal clinical prediction objective rather than an entire disease.

2.4 Automatic Relation Learning Across Prediction Objectives

PD, MS, and PFF have distinct pathological mechanisms and clinical outcomes, but their DMO–outcome mappings may share functional patterns. Similarly, the two PD outcomes may reflect related but non-identical aspects of disease severity and motor impairment. These relationships cannot be specified reliably in advance. Outcomes from different diseases are not jointly observed because the cohorts contain different participants, while correlations between outcomes within the same disease do not necessarily reflect similarities in their longitudinal DMO mappings. We therefore learn relationships among prediction objectives directly from their model parameters.

For disease d and outcome q, let $\mathbf{u}_{d,q} = \mathrm{vec}(\mathbf{W}_{d,q}) \in \mathbb{R}^{pT}$. The representations of all M prediction objectives are collected as $\mathbf{U} \in \mathbb{R}^{pT \times M}$, with each column corresponding to one disease–outcome pair. We assume that each longitudinal DMO mapping is related to the remaining mappings. After assigning each disease–outcome pair a unique objective index, this relation is expressed as

$$\mathbf{U} \approx \mathbf{U}\mathbf{A}, \qquad \mathbf{A} = \mathbf{A}^{\mathsf{T}}, \qquad \mathrm{diag}(\mathbf{A}) = \mathbf{0}, \tag{3}$$

where $\mathbf{A} \in \mathbb{R}^{M \times M}$ is the learned relation matrix. Symmetry produces an undirected relation graph, while the zero diagonal excludes trivial self-relations.

The matrix $\mathbf{A}$ can be partitioned into blocks $\mathbf{A}^{(d,k)} \in \mathbb{R}^{Q_d \times Q_k}$ according to disease membership. A diagonal block $\mathbf{A}^{(d,d)}$ captures relationships among outcomes within disease d, whereas an off-diagonal block $\mathbf{A}^{(d,k)}$ captures cross-disease relationships. Symmetry implies $\mathbf{A}^{(d,k)} = (\mathbf{A}^{(k,d)})^{\mathsf{T}}$. We estimate the relation structure using

$$\Omega_{\mathrm{R}}(\mathbf{U}, \mathbf{A}) = \frac{\lambda_{\mathrm{R}}}{2} \|\mathbf{U} - \mathbf{U}\mathbf{A}\|_F^2 + \frac{\lambda_{\mathrm{A}}}{2} \|\mathbf{A}\|_F^2 , \tag{4}$$

where $\lambda_{\mathrm{R}} \geq 0$ controls cross-objective information sharing and $\lambda_{\mathrm{A}} > 0$ stabilises relation estimation. We use an ℓ_2 penalty rather than sparsity because the model contains only four prediction objectives and weak relationships may still provide useful information.

2.5 Unified Multi-Disease and Multi-Outcome Objective

Combining longitudinal regression, fused temporal regularisation, sparse group DMO selection, and automatic relation learning yields the proposed *DMO-enabled Multi-Disease and Multi-Outcome learning* framework (DeMMO):

$$\begin{aligned} \min_{\{\mathbf{W}_{d,q}\},\, \mathbf{A}} \quad & \sum_{d=1}^{D} \sum_{q=1}^{Q_d} \sum_{t=1}^{T} \frac{\|\mathbf{y}_{t,q}^{(d)} - \mathbf{X}_{t,q}^{(d)} \mathbf{w}_{t,q}^{(d)}\|_2^2}{2n_{dtq}} \\ & + \sum_{d=1}^{D} \sum_{q=1}^{Q_d} \big(\lambda_{\mathrm{FL}} \|\mathbf{W}_{d,q}\mathbf{R}^{\mathsf{T}}\|_1 + \lambda_{\mathrm{L}} \|\mathbf{W}_{d,q}\|_1 + \lambda_{\mathrm{G}} \|\mathbf{W}_{d,q}\|_{2,1}\big) \\ & + \frac{\lambda_{\mathrm{R}}}{2} \|\mathbf{U} - \mathbf{U}\mathbf{A}\|_F^2 + \frac{\lambda_{\mathrm{A}}}{2} \|\mathbf{A}\|_F^2 \\ \text{s.t.} \quad & \mathbf{A} = \mathbf{A}^{\mathsf{T}}, \qquad \mathrm{diag}(\mathbf{A}) = \mathbf{0}. \end{aligned} \tag{5}$$

As illustrated in Figure 1, the objective comprises four complementary components:

- longitudinal regression estimates clinical status at each visit;
- temporal smoothness regularisation encourages continuity across adjacent visits while permitting local changes;
- sparse group Lasso identifies longitudinally shared and visit-specific candidate DMOs; and
- automatic relation learning transfers information among related clinical outcomes within and across diseases.

Notably, DeMMO does not require shared participants across disease cohorts, and each prediction objective retains its own longitudinal coefficient matrix. The objective is biconvex in $\{\mathbf{W}_{d,q}\}$ and $\mathbf{A}$. For fixed $\mathbf{A}$, the coefficient subproblem is convex but non-smooth because of the fused-Lasso, Lasso, and group-Lasso penalties. For fixed coefficient matrices, relation learning is a strongly convex quadratic problem over symmetric matrices with zero diagonal. We alternate between these two subproblems.

2.6 Data Standardisation

All clinical outcomes are oriented so that higher values indicate greater disease severity or functional impairment. H&Y, MDS–UPDRS Part III, and EDSS retain their original directions, while SPPB

is transformed as $12 - \text{SPPB}$. Each outcome is then standardised using the mean and standard deviation estimated from its training records pooled across visits.

The 24 DMOs are similarly standardised using pooled training records from the corresponding disease cohort. Pooling across visits preserves longitudinal distributional changes that visit-specific standardisation could remove. All transformations and standardisation parameters are estimated exclusively from the training data and applied unchanged to the validation and test sets to prevent information leakage.

Additional results in the appendix. Due to space constraints, the complete alternating optimisation procedure, convergence analysis, and computational complexity are deferred to Appendix B. The resulting algorithm is computationally efficient: coefficient updates use established accelerated proximal-gradient operations, while relation learning adds little overhead because the four outcomes yield only six free relation parameters. Because convex sparsity-inducing penalties are known to shrink nonzero coefficients and may introduce estimation bias (Fan & Li, 2001), Appendix C develops two non-convex variants designed to reduce this bias. Their empirical results are reported in Appendix D. Despite their greater computational cost, both non-convex variants perform worse overall than the original DeMMO, suggesting that reducing sparsity-induced estimation bias is not beneficial in this setting. We conjecture that, because free-living DMOs contain substantial behavioural and measurement noise, the stronger shrinkage of the convex formulation provides useful regularisation, whereas the non-convex variants may retain more noise and generalise less effectively.

3 Experiments

We evaluate DeMMO on the Mobilise-D dataset described in Section 2.1. The analysis includes 24 weekly DMOs measured over five visits and four clinical prediction objectives: PD H&Y, PD MDS–UPDRS III, MS EDSS, and PFF SPPB impairment. COPD is excluded because the required clinical outcomes are unavailable at T2 and T4. The resulting objective-specific sample sizes are reported in Table 1. For each experimental run, participants are divided into training, validation, and held-out test sets in a 70%/10%/20% ratio. All available visits from the same participant remain in one split to prevent information leakage. Data processing and standardisation follow Section 2.6, with all parameters estimated from the training data only. We repeat the complete procedure over five matched participant-level splits using random seeds 42–46. Following common practice (Zhou et al., 2024; Fan et al., 2026), we use normalised mean squared error (nMSE) and weighted correlation (wR) to assess overall performance, and root mean squared error (RMSE) for visit-specific performance.

3.1 Compared Methods and Training Settings

We compare DeMMO with nine baselines spanning conventional linear regression, interpretable longitudinal models, and recent deep regression methods. All methods follow the common data-splitting and preprocessing protocol described above, with hyperparameters selected on the validation set and the held-out test set reserved for final evaluation. Methods that do not explicitly model time are fitted independently to each of the 20 outcome–visit tasks. All deep baselines use two 32-unit ReLU hidden layers without data augmentation.

DeMMO. DeMMO is trained using the alternating optimisation procedure described in Section B. Because a joint grid search over its five regularisation parameters is computationally expensive, we use two-stage selection. First, relation learning is disabled while λ_{FL}, λ_{L}, and λ_{G} are selected from $\{10^{-3}, 10^{-2}, 10^{-1}, 1\}$. These parameters are then fixed, and $(\lambda_{\mathrm{R}}, \lambda_{\mathrm{A}})$ is selected from $(0, 0)$ and $\{10^{-3}, 10^{-2}, 10^{-1}\}^2$. The selected model is refitted on the combined training and validation sets. Optimisation is limited to 50 outer and 500 inner iterations, with convergence tolerances of 10^{-6} and 10^{-7}, respectively. This two-stage strategy reduces model-selection cost but does not exhaust the full joint hyperparameter space. Section 3.2 provides the detailed hyperparameter-sensitivity analysis.

Baselines. **(1) Ridge regression.** We fit an independent ℓ_2-regularised linear model for each outcome–visit pair. Its penalty is selected from $\{10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}\}$ using validation nMSE. **(2) Lasso regression.** This sparse linear baseline independently fits each outcome–visit pair using an ℓ_1 penalty selected from $\{10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}\}$ on the validation set. **(3) FRoTS.** FRoTS is a

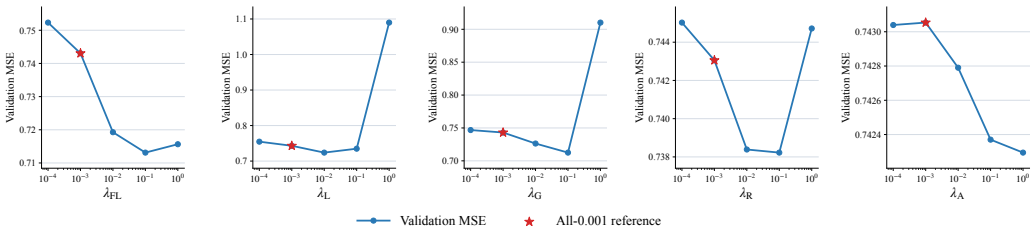

Figure 2: One-at-a-time sensitivity of validation nMSE; all other parameters are fixed at 0.001, with the shared reference marked by a red star.

Table 2: Test Performance on Mobilise-D (Mean $\pm$ Standard Deviation). The best mean in each column is shown in bold. The final row reports one-sided paired t-test p-values comparing DeMMO with the best-performing baseline over the five matched splits. $^*p < 0.05$.

	Overall		PD H&Y		PD MDS–UPDRS		MS EDSS		PFF SPPB impairment	
Method	nMSE ↓	wR ↑	nMSE ↓	wR ↑	nMSE ↓	wR ↑	nMSE ↓	wR ↑	nMSE ↓	wR ↑
Ridge	0.735 $\pm$ 0.031	0.497 $\pm$ 0.028	0.951 $\pm$ 0.033	0.283 $\pm$ 0.037	0.882 $\pm$ 0.053	0.366 $\pm$ 0.048	0.492 $\pm$ 0.054	0.721 $\pm$ 0.039	0.557 $\pm$ 0.065	0.677 $\pm$ 0.045
Lasso	0.741 $\pm$ 0.037	0.492 $\pm$ 0.034	0.958 $\pm$ 0.032	0.276 $\pm$ 0.045	0.889 $\pm$ 0.046	0.350 $\pm$ 0.065	0.495 $\pm$ 0.050	0.725 $\pm$ 0.039	0.562 $\pm$ 0.053	0.679 $\pm$ 0.038
FRoTS	0.740 $\pm$ 0.035	0.500 $\pm$ 0.028	0.960 $\pm$ 0.040	0.285 $\pm$ 0.034	0.889 $\pm$ 0.054	0.373 $\pm$ 0.041	0.491 $\pm$ 0.057	0.723 $\pm$ 0.038	0.561 $\pm$ 0.080	0.678 $\pm$ 0.053
MAGPP	0.757 $\pm$ 0.033	0.488 $\pm$ 0.027	0.984 $\pm$ 0.044	0.264 $\pm$ 0.035	0.907 $\pm$ 0.054	0.361 $\pm$ 0.042	0.504 $\pm$ 0.059	0.716 $\pm$ 0.040	0.575 $\pm$ 0.085	0.672 $\pm$ 0.054
MLP-MSE	0.762 $\pm$ 0.030	0.473 $\pm$ 0.022	0.976 $\pm$ 0.023	0.245 $\pm$ 0.038	0.929 $\pm$ 0.097	0.329 $\pm$ 0.076	0.493 $\pm$ 0.062	0.721 $\pm$ 0.039	0.597 $\pm$ 0.033	0.656 $\pm$ 0.020
MLP-L1	0.779 $\pm$ 0.028	0.451 $\pm$ 0.022	1.017 $\pm$ 0.013	0.186 $\pm$ 0.037	0.936 $\pm$ 0.058	0.306 $\pm$ 0.056	0.516 $\pm$ 0.083	0.716 $\pm$ 0.045	0.585 $\pm$ 0.054	0.667 $\pm$ 0.031
Rank-N	0.783 $\pm$ 0.028	0.403 $\pm$ 0.041	1.025 $\pm$ 0.014	0.070 $\pm$ 0.081	0.919 $\pm$ 0.035	0.262 $\pm$ 0.055	0.528 $\pm$ 0.066	0.710 $\pm$ 0.046	0.599 $\pm$ 0.044	0.653 $\pm$ 0.031
RankSim	0.762 $\pm$ 0.031	0.466 $\pm$ 0.029	0.978 $\pm$ 0.031	0.232 $\pm$ 0.054	0.916 $\pm$ 0.069	0.320 $\pm$ 0.064	0.503 $\pm$ 0.089	0.716 $\pm$ 0.050	0.595 $\pm$ 0.029	0.656 $\pm$ 0.023
ACCon	0.775 $\pm$ 0.026	0.460 $\pm$ 0.015	0.995 $\pm$ 0.052	0.232 $\pm$ 0.027	0.942 $\pm$ 0.089	0.307 $\pm$ 0.058	0.502 $\pm$ 0.056	0.713 $\pm$ 0.038	0.606 $\pm$ 0.069	0.648 $\pm$ 0.046
DeMMO	**0.722 $\pm$ 0.032**	**0.515 $\pm$ 0.030**	**0.929 $\pm$ 0.035**	**0.319 $\pm$ 0.048**	**0.874 $\pm$ 0.038**	**0.384 $\pm$ 0.038**	**0.479 $\pm$ 0.051**	**0.732 $\pm$ 0.039**	**0.549 $\pm$ 0.062**	**0.681 $\pm$ 0.044**
p-value	0.0015*	0.0148*	0.0215*	0.0140*	0.1807	0.0787	0.0896	0.0005*	0.0843	0.2500

robust temporal multi-task regression method (Zhou et al., 2023b) that jointly models five visits by decomposing coefficients into temporally smooth and visit-specific components. Its two penalties are selected from $\{10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}\}$ using validation nMSE. **(4) MAGPP.** MAGPP learns an automatic temporal relation graph for disease progression prediction (Zhou et al., 2024). It jointly estimates visit-specific coefficients and a 5×5 visit-relation graph under group and element-wise sparsity. Its four parameters are selected from $\{10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}\}$. **(5) MLP-MSE.** For each outcome–visit pair, we train a two-hidden-layer MLP with 32 ReLU units per layer using MSE, Adam, and validation-based early stopping. **(6) MLP-L1.** MLP-L1 uses the same architecture and training procedure as MLP-MSE but minimises MAE, isolating the effect of the loss. **(7) Rank-N-Contrast.** Rank-N-Contrast (Rank-N) exploits the ordering of continuous targets (Zha et al., 2024). Unit-width training-target bins pretrain a contrastive encoder, which is frozen before fitting a linear head to the continuous targets; out-of-range values use the nearest bin. **(8) RankSim.** RankSim encourages neighbourhood rankings in the representation to agree with continuous-label rankings (Gong et al., 2022); its ranking regulariser is combined with MSE without target binning. **(9) ACCon.** ACCon combines MSE with an angle-compensated contrastive regulariser that relates representation similarity to continuous-label distance (Zhao et al., 2025). Its weight is selected from $\{10^{-2}, 10^{-1}, 1, 10, 100\}$ using validation MSE.

3.2 Hyperparameter Sensitivity

We vary each parameter over $\{10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}, 1\}$ while fixing the others at 10^{-3}, and report mean validation MSE across the 20 outcome–visit tasks for seed 42 (Fig. 2); the test set is not used. All five penalties affect validation performance. The sparsity penalties are most influential and become excessive at 1, λ_{FL} favours moderate temporal smoothing, and the U-shaped λ_{R} curve supports selective rather than unrestricted cross-objective sharing. Performance is comparatively insensitive to λ_{A} over the tested range, likely because standardising the predictors and outcomes places the coefficient matrices across disease–outcome objectives on comparable scales.

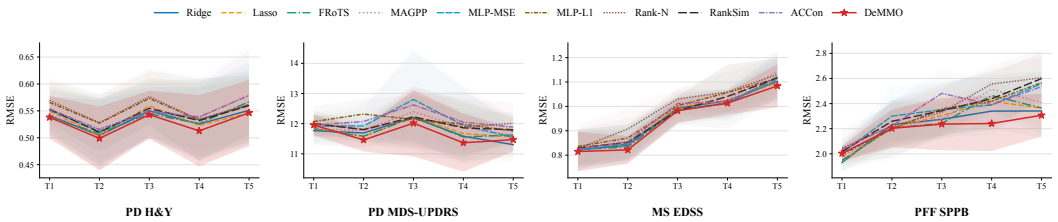

Figure 3: Visit-specific RMSE (mean ± standard deviation over five matched participant splits) for the four prediction objectives. Each panel reports T1–T5 performance for all baselines and DeMMO. Lower values indicate better predictive accuracy; no visit-level significance test is displayed.

3.3 Prediction Performance and Learned Cross-Objective Relations

DeMMO achieves the best mean nMSE and wR overall and for every outcome (Table 2). Relative to the strongest baseline, it reduces overall nMSE from 0.735 to 0.722 and increases wR from 0.500 to 0.515; both gains are significant ($p = 0.0015$ and $p = 0.0148$). Significant gains are also observed for both PD H&Y metrics and for MS EDSS wR. For PD H&Y, DeMMO achieves the lowest RMSE at T2–T4 and ranks first or second at every visit. For MS EDSS, it is best at T1, T2, and T5, ties for best at T3, and ranks second at T4. It also obtains the lowest PD MDS–UPDRS error at T2–T4, showing that the joint model remains competitive across outcomes with different scales and variability. Figure 3 summarises the visit-level comparison; detailed numerical results are provided in Appendix E, Tables 4–7. DeMMO ranks first or second in 18 of 20 tasks and is uniquely best in 12. It is particularly effective at PFF T3–T5, where follow-up samples are sparse and information sharing provides useful regularisation. Linear and temporal baselines generally outperform the deep methods, indicating that appropriate structure is more useful in this setting than greater model capacity. Errors increase towards later MS and PFF visits, whereas both PD outcomes are non-monotonic. This heterogeneity supports DeMMO's selective sharing of common structure while retaining outcome-specific DMO effects.

The optimiser learns the zero-diagonal cross-objective matrix $\mathbf{A}$. For interpretation, we report the complete relation matrix $\widetilde{\mathbf{A}} = \mathbf{A} + \mathbf{I}$, whose unit diagonal represents each objective's relation with itself. Figure 4 shows the element-wise mean of $\widetilde{\mathbf{A}}$ over all five matched splits. The strongest positive relation is between the two PD outcomes (0.49), followed by MS EDSS–PFF SPPB impairment (0.47), PD MDS–UPDRS–PFF (0.36), and PD H&Y–MS EDSS (0.30). The strong within-PD relation is clinically plausible because H&Y and MDS–UPDRS measure related aspects of PD severity. The cross-disease edges further show that unpaired cohorts can share functional mobility patterns despite different pathologies and clinical scales. Negative relations occur for PD MDS–UPDRS–MS EDSS (-0.32) and PD H&Y–PFF (-0.27). Because the matrix is learned from longitudinal coefficients, these values do not imply negative associations between diseases or raw scores. Instead, they indicate oppositely oriented DMO effects, allowing DeMMO to use both aligned and complementary structures.

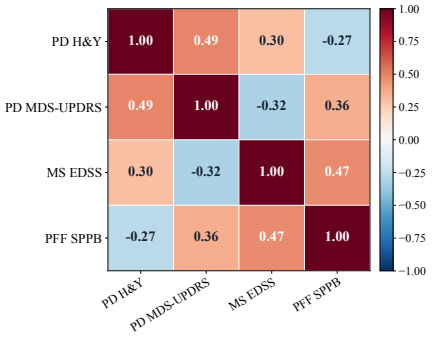

Figure 4: Mean complete relation matrix $\widetilde{\mathbf{A}} = \mathbf{A} + \mathbf{I}$ over all five matched splits.

3.4 Longitudinal Stability Selection

We perform longitudinal stability selection (Zhou et al., 2024; 2022b; 2023a) over ten randomly selected hyperparameter settings, each repeated on ten half-samples of the training participants. Figure 5 reports how often each DMO has a non-zero coefficient in the resulting 100 fits.

The profiles are outcome-specific but generally stable across visits. For MS EDSS, cadence P90 and walking-speed P90 in bouts longer than 30 seconds have the highest mean probabilities (0.868 and

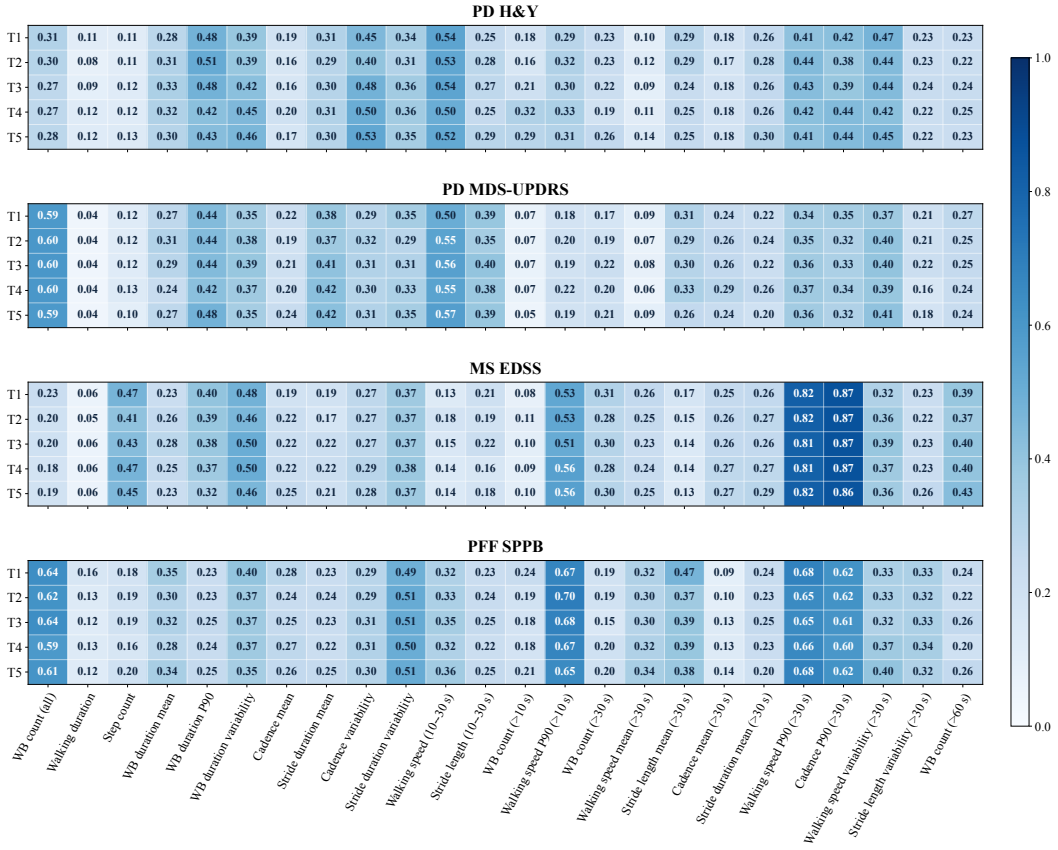

Figure 5: DMO selection probabilities across 100 half-sample fits (ten hyperparameter settings with ten repetitions) for each outcome and visit.

0.816). PFF selects a broader mix of pace, rhythm, and walking amount, including walking-speed P90, cadence P90, and total walking-bout count. PD H&Y most consistently selects mean walking speed in 10–30-second bouts, while PD MDS–UPDRS III also stably selects total walking-bout count.

These patterns broadly agree with the learned relations. The positively related PD outcomes share three of their five most stable DMOs, including mean walking speed in 10–30-second bouts. The positive MS–PFF pair likewise shares three high-speed or cadence measures. In contrast, PD MDS–UPDRS–MS shares no top-five DMO, while PD H&Y–PFF shares only long-bout walking-speed P90. Although relation signs do not determine feature-support overlap, the smaller overlap of the negatively related pairs indicates more distinct mobility signatures.

No DMO has a mean selection probability above 0.4 across all outcomes, and variation is greater across outcomes than visits. Thus, relevant DMOs tend to remain stable longitudinally, but their sets differ across outcomes. This argues against a universal DMO panel: DeMMO instead identifies robust, outcome-specific candidates while selectively sharing cross-outcome structure. Because correlated DMOs may substitute for one another, these candidates require subsequent clinical validation.

4 Conclusion

We proposed DeMMO, an interpretable framework for modelling longitudinal DMO relationships across multiple outcomes and diseases. Its automatic relation mechanism learns signed cross-outcome structure from DMO mappings and shares information across cohorts without requiring common participants. Across four outcomes and five Mobilise-D visits, DeMMO achieved an overall nMSE of 0.722 ± 0.032 and weighted correlation of 0.515 ± 0.030, significantly outperforming nine strong

baselines. Stability selection showed that DMO relevance is generally consistent across visits but differs across outcomes, supporting outcome-specific modelling rather than a universal DMO panel. The resulting DMO patterns are candidates for clinical validation; future work will assess their generalisability in independent cohorts, additional diseases, and longer follow-up studies.

AI USE STATEMENT

Generative AI tools were used to assist with language editing, LaTeX formatting, and code review. The authors reviewed all AI-assisted material, verified the reported claims and experimental results, and take responsibility for the final content of the paper and its associated artifacts.

REPRODUCIBILITY STATEMENT

The paper specifies the model, optimisation procedure, data splitting, preprocessing, hyperparameter selection, and evaluation protocol. Source code and privacy-safe intermediate and aggregate results are publicly available at https://github.com/menghui-zhou/DeMMO. Additional analyses are reported in the appendix.

REFERENCES

Lisa Alcock, Encarna Micó-Amigo, Tecla Bonci, and Silvia Del Din. Algorithms for gait: technical and clinical validity. In *Gait, Balance, and Mobility Analysis*, pp. 277–321. Elsevier, 2025.

Lmar M Babrak, Joseph Menetski, Michael Rebhan, Giovanni Nisato, Marc Zinggeler, Noé Brasier, Katja Baerenfaller, Thomas Brenzikofer, Laurenz Baltzer, Christian Vogler, et al. Traditional and digital biomarkers: two worlds apart? *Digital biomarkers*, 3(2):92–102, 2019.

Tian Bai, Shanshan Zhang, Brian L Egleston, and Slobodan Vucetic. Interpretable representation learning for healthcare via capturing disease progression through time. In *Proceedings of the 24th ACM SIGKDD international conference on knowledge discovery & data mining*, pp. 43–51, 2018.

John Bellettiere, Michael J Lamonte, Jonathan Unkart, Sandy Liles, Deepika Laddu-Patel, JoAnn E Manson, Hailey Banack, Rebecca Seguin-Fowler, Paul Chavez, Lesley F Tinker, et al. Short physical performance battery and incident cardiovascular events among older women. *Journal of the American Heart Association*, 9(14):e016845, 2020.

Elroy Boers, Angier Allen, Meredith Barrett, Adam V Benjafield, Mary B Rice, Jadwiga A Wedzicha, Leanne Kaye, Heather J Zar, Sanjeev Sinha, Obianuju Ozoh, et al. Forecasting the global economic and health burden of chronic obstructive pulmonary disease from 2025 through 2050. *Chest*, 2025.

A Bouleau, C Dulong, CA Schwerer, R Delgrange, K Bouaou, T Brochu, S Zinai, K Švecová, MJ Sá, A Petropoulos, et al. The socioeconomic impact of multiple sclerosis in france: Results from the petals study. *Multiple Sclerosis Journal–Experimental, Translational and Clinical*, 8(2): 20552173221093219, 2022.

Gavin Brittain, Michael Long, Ellen Buckley, Clemens Becker, Joren Buekers, Martí de Las Heras, Claudia Armengol, Sarah Koch, Julia Braun, Brian Caufield, et al. Testing the construct validity of real-world digital mobility outcomes in multiple sclerosis. *Multiple Sclerosis Journal*, pp. 13524585261464537, 2026.

Peng Cao, Xuanfeng Shan, Dazhe Zhao, Min Huang, and Osmar Zaiane. Sparse shared structure based multi-task learning for mri based cognitive performance prediction of alzheimer's disease. *Pattern Recognition*, 72:219–235, 2017.

Francesca Cerreta, Armin Ritzhaupt, Thomas Metcalfe, Scott Askin, João Duarte, Michael Berntgen, and Spiros Vamvakas. Digital technologies for medicines: shaping a framework for success. *Nature Reviews Drug Discovery*, 19(9):573–574, 2020.

Zhengping Che, Sanjay Purushotham, Kyunghyun Cho, David Sontag, and Yan Liu. Recurrent neural networks for multivariate time series with missing values. *Scientific reports*, 8(1):6085, 2018.

Edward Choi, Mohammad Taha Bahadori, Andy Schuetz, Walter F Stewart, and Jimeng Sun. Doctor ai: Predicting clinical events via recurrent neural networks. In *Machine learning for healthcare conference*, pp. 301–318. PMLR, 2016.

Philip Coran, Jennifer C Goldsack, Cheryl A Grandinetti, Jessie P Bakker, Marisa Bolognese, E Dorsey, Kaveeta Vasisht, Adam Amdur, Christopher Dell, Jonathan Helfgott, et al. Advancing the use of mobile technologies in clinical trials: recommendations from the clinical trials transformation initiative. *Digital Biomarkers*, 3(3):145–154, 2020.

Marta Francisca Corrà, Arash Atrsaei, Ana Sardoreira, Clint Hansen, Kamiar Aminian, Manuel Correia, Nuno Vila-Chã, Walter Maetzler, and Luís Maia. Comparison of laboratory and daily-life gait speed assessment during on and off states in parkinson's disease. *Sensors*, 21(12):3974, 2021.

Katherine HO Deane, Helen Flaherty, David J Daley, Roland Pascoe, Bridget Penhale, Carl E Clarke, Catherine Sackley, and Stacey Storey. Priority setting partnership to identify the top 10 research priorities for the management of parkinson's disease. *BMJ open*, 4(12):e006434, 2014.

Laura Delgado-Ortiz, Ashley Polhemus, Alison Keogh, Norman Sutton, Werner Remmele, Clint Hansen, Felix Kluge, Basil Sharrack, Clemens Becker, Thierry Troosters, et al. Listening to the patients' voice: a conceptual framework of the walking experience. *Age and Ageing*, 52(1):afac233, 2023.

Tobias Eckert, Martin Aursand Berge, Michael Long, Martí de Las Heras, Paula Alvarez, Hubert Blain, Julia Braun, Joren Buekers, Brian Caulfield, Monika Engdal, et al. Construct validity of real-world digital mobility outcomes in patients after proximal femoral fracture: a cross-sectional observational study. *Scientific Reports*, 16(1):9535, 2026.

Monika Engdal, Kristin Taraldsen, Carl-Philipp Jansen, Raphael Simon Peter, Beatrix Vereijken, Clemens Becker, Jorunn Laegdheim Helbostad, and Jochen Klenk. Real-world mobility recovery after hip fracture: secondary analyses of digital mobility outcomes from four randomized controlled trials. *Age and ageing*, 53(10):afae234, 2024.

Jianqing Fan and Runze Li. Variable selection via nonconcave penalized likelihood and its oracle properties. *Journal of the American statistical Association*, 96(456):1348–1360, 2001.

Xuanhan Fan, Menghui Zhou, Yu Zhang, Jun Qi, Yun Yang, and Po Yang. Beyond single scores: A multi-cognitive objective learning for ad progression prediction. *Pattern Recognition*, pp. 114348, 2026.

Mostafa Mehdipour Ghazi, Mads Nielsen, Akshay Pai, M Jorge Cardoso, Marc Modat, Sébastien Ourselin, Lauge Sørensen, Alzheimer's Disease Neuroimaging Initiative, et al. Training recurrent neural networks robust to incomplete data: application to alzheimer's disease progression modeling. *Medical image analysis*, 53:39–46, 2019.

Christopher G Goetz, Barbara C Tilley, Stephanie R Shaftman, Glenn T Stebbins, Stanley Fahn, Pablo Martinez-Martin, Werner Poewe, Cristina Sampaio, Matthew B Stern, Richard Dodel, et al. Movement disorder society-sponsored revision of the unified parkinson's disease rating scale (mds-updrs): scale presentation and clinimetric testing results. *Movement disorders: official journal of the Movement Disorder Society*, 23(15):2129–2170, 2008.

Jennifer C Goldsack, Andrea Coravos, Jessie P Bakker, Brinnae Bent, Ariel V Dowling, Cheryl Fitzer-Attas, Alan Godfrey, Job G Godino, Ninad Gujar, Elena Izmailova, et al. Verification, analytical validation, and clinical validation (v3): the foundation of determining fit-for-purpose for biometric monitoring technologies (biomets). *npj digital Medicine*, 3(1):55, 2020.

Yu Gong, Greg Mori, and Frederick Tung. Ranksim: Ranking similarity regularization for deep imbalanced regression. In *Proceedings of the 39th International Conference on Machine Learning*, pp. 7634–7649. PMLR, 2022.

Margaret M Hoehn and Melvin D Yahr. Parkinsonism: onset, progression, and mortality. *Neurology*, 17(5):427–427, 1967.

Alison Keogh, Lisa Alcock, Philip Brown, Ellen Buckley, Marina Brozgol, Eran Gazit, Clint Hansen, Kirsty Scott, Lars Schwickert, Clemens Becker, et al. Acceptability of wearable devices for measuring mobility remotely: observations from the mobilise-d technical validation study. *Digital Health*, 9:20552076221150745, 2023.

Cameron Kirk, Arne Küderle, M Encarna Micó-Amigo, Tecla Bonci, Anisoara Paraschiv-Ionescu, Martin Ullrich, Abolfazl Soltani, Eran Gazit, Francesca Salis, Lisa Alcock, et al. Mobilise-d insights to estimate real-world walking speed in multiple conditions with a wearable device. *Scientific reports*, 14(1):1754, 2024.

Cameron Kirk, Emma Packer, Ashley Polhemus, Mhairi K MacLean, Harry Bailey, Felix Kluge, Heiko Gaßner, Lynn Rochester, Silvia Del Din, and Alison J Yarnall. A systematic review of real-world gait-related digital mobility outcomes in parkinson's disease. *npj Digital Medicine*, 8 (1):585, 2025.

Felix Kluge, Silvia Del Din, Andrea Cereatti, Heiko Gaßner, Clint Hansen, Jorunn L Helbostad, Jochen Klucken, Arne Küderle, Arne Müller, Lynn Rochester, et al. Consensus based framework for digital mobility monitoring. *PloS one*, 16(8):e0256541, 2021.

John F Kurtzke. Rating neurologic impairment in multiple sclerosis: an expanded disability status scale (edss). *Neurology*, 33(11):1444–1452, 1983.

Hui Li, Yanlin Wang, Ziyu Lyu, and Jieming Shi. Multi-task learning for recommendation over heterogeneous information network. *IEEE Transactions on Knowledge and Data Engineering*, 34 (2):789–802, 2020a.

Yikuan Li, Shishir Rao, José Roberto Ayala Solares, Abdelaali Hassaine, Rema Ramakrishnan, Dexter Canoy, Yajie Zhu, Kazem Rahimi, and Gholamreza Salimi-Khorshidi. Behrt: transformer for electronic health records. *Scientific reports*, 10(1):7155, 2020b.

Tengfei Ma, Xuan Lin, Bosheng Song, Philip S Yu, and Xiangxiang Zeng. Kg-mtl: knowledge graph enhanced multi-task learning for molecular interaction. *IEEE Transactions on Knowledge and Data Engineering*, 35(7):7068–7081, 2022.

Claudia Mazzà, Lisa Alcock, Kamiar Aminian, Clemens Becker, Stefano Bertuletti, Tecla Bonci, Philip Brown, Marina Brozgol, Ellen Buckley, Anne-Elie Carsin, et al. Technical validation of real-world monitoring of gait: a multicentric observational study. *BMJ open*, 11(12):e050785, 2021.

Dimitrios Megaritis, Michael Long, Martí de Las Heras, Victoria Alcaraz-Serrano, Paula Alvarez, Clemens Becker, Julia Braun, Joren Buekers, Sara Buttery, Brian Caulfield, et al. The construct validity of real-world digital mobility outcomes in people with copd. *ERJ Open Research*, 12(2): 00993–2025, 2026.

M Encarna Micó-Amigo, Tecla Bonci, Anisoara Paraschiv-Ionescu, Martin Ullrich, Cameron Kirk, Abolfazl Soltani, Arne Küderle, Eran Gazit, Francesca Salis, Lisa Alcock, et al. Assessing real-world gait with digital technology? validation, insights and recommendations from the mobilise-d consortium. *Journal of neuroengineering and rehabilitation*, 20(1):78, 2023.

A Stefanie Mikolaizak, Lynn Rochester, Walter Maetzler, Basil Sharrack, Heleen Demeyer, Claudia Mazzà, Brian Caulfield, Judith Garcia-Aymerich, Beatrix Vereijken, Valdo Arnera, et al. Connecting real-world digital mobility assessment to clinical outcomes for regulatory and clinical endorsement–the mobilise-d study protocol. *PloS one*, 17(10):e0269615, 2022.

Riccardo Miotto, Li Li, Brian A Kidd, and Joel T Dudley. Deep patient: an unsupervised representation to predict the future of patients from the electronic health records. *Scientific reports*, 6(1): 26094, 2016.

Anat Mirelman, Jana Volkov, Amit Salomon, Eran Gazit, Alice Nieuwboer, Lynn Rochester, Silvia Del Din, Laura Avanzino, Elisa Pelosin, Bastiaan R Bloem, et al. Digital mobility measures: a window into real-world severity and progression of parkinson's disease. *Movement Disorders*, 39 (2):328–338, 2024.

Y Nesterov. A method for solving a convex programming problem with convergence rate o (1/k2). In *Soviet Mathematics. Doklady*, volume 27, pp. 367–372, 1983.

Yurii Nesterov. *Introductory lectures on convex optimization: A basic course*, volume 87. Springer Science & Business Media, 2003.

Minh Nguyen, Tong He, Lijun An, Daniel C Alexander, Jiashi Feng, BT Thomas Yeo, Alzheimer's Disease Neuroimaging Initiative, et al. Predicting alzheimer's disease progression using deep recurrent neural networks. *NeuroImage*, 222:117203, 2020.

Margaux Poleur, Barbara Willekens, Bertrand Degos, Damien Ricard, Vincent Van Pesch, Annick Mélin, Oihana Piquet, Alexis Tricot, Laurie Médard, Mona Michaud, et al. Stride-level measurement of gait as an early sensitive marker of disability progression in ambulatory patients with multiple sclerosis. *EClinicalMedicine*, 93, 2026.

Ashley Polhemus, Laura Delgado-Ortiz, Gavin Brittain, Nikolaos Chynkiamis, Francesca Salis, Heiko Gaßner, Michaela Gross, Cameron Kirk, Rachele Rossanigo, Kristin Taraldsen, et al. Walking on common ground: a cross-disciplinary scoping review on the clinical utility of digital mobility outcomes. *NPJ digital medicine*, 4(1):149, 2021.

Rebecca J Port, Martin Rumsby, Graham Brown, Ian F Harrison, Anneesa Amjad, and Claire J Bale. People with parkinson's disease: what symptoms do they most want to improve and how does this change with disease duration? *Journal of Parkinson's disease*, 11(2):715–724, 2021.

Stéphanie A Prince, Kristi B Adamo, Meghan E Hamel, Jill Hardt, Sarah Connor Gorber, and Mark Tremblay. A comparison of direct versus self-report measures for assessing physical activity in adults: a systematic review. *International journal of behavioral nutrition and physical activity*, 5 (1):56, 2008.

Pablo Rábano-Suárez, Natalia Del Campo, Isabelle Benatru, Caroline Moreau, Clément Desjardins, Álvaro Sánchez-Ferro, and Margherita Fabbri. Digital outcomes as biomarkers of disease progression in early parkinson's disease: a systematic review. *Movement Disorders*, 40(2):184–203, 2025.

Lynn Rochester, Claudia Mazzà, Arne Mueller, Brian Caulfield, Marie McCarthy, Clemens Becker, Ram Miller, Paolo Piraino, Marco Viceconti, Wilhelmus P Dartee, et al. A roadmap to inform development, validation and approval of digital mobility outcomes: the mobilise-d approach. *Digital biomarkers*, 4(Suppl. 1):13–27, 2020.

Vrutangkumar V Shah, James McNames, Martina Mancini, Patricia Carlson-Kuhta, Rebecca I Spain, John G Nutt, Mahmoud El-Gohary, Carolin Curtze, and Fay B Horak. Laboratory versus daily life gait characteristics in patients with multiple sclerosis, parkinson's disease, and matched controls. *Journal of neuroengineering and rehabilitation*, 17(1):159, 2020.

Satya Narayan Shukla and Benjamin M Marlin. Multi-time attention networks for irregularly sampled time series. In *... International Conference on Learning Representations*, volume 2021, pp. 14897, 2021.

Noah Simon, Jerome Friedman, Trevor Hastie, and Robert Tibshirani. A sparse-group lasso. *Journal of computational and graphical statistics*, 22(2):231–245, 2013.

Robert Tibshirani. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, 58(1):267–288, 1996.

Robert Tibshirani, Michael Saunders, Saharon Rosset, Ji Zhu, and Keith Knight. Sparsity and smoothness via the fused lasso. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 67(1):91–108, 2005.

Sam Williamson, Filipa Landeiro, Thomas McConnell, Lucy Fulford-Smith, M Kassim Javaid, Andrew Judge, and José Leal. Costs of fragility hip fractures globally: a systematic review and meta-regression analysis. *Osteoporosis international*, 28(10):2791–2800, 2017.

Wenya Yang, Jamie L Hamilton, Catherine Kopil, James C Beck, Caroline M Tanner, Roger L Albin, E Ray Dorsey, Nabila Dahodwala, Inna Cintina, Paul Hogan, et al. Current and projected future economic burden of parkinson's disease in the us. *npj Parkinson's Disease*, 6(1):15, 2020.

Kaiwen Zha, Peng Cao, Jeany Son, Yuzhe Yang, and Dina Katabi. Rank-n-contrast: learning continuous representations for regression. *Advances in Neural Information Processing Systems*, 36, 2024.

Yu Zhang and Qiang Yang. A survey on multi-task learning. *IEEE Transactions on Knowledge and Data Engineering*, 2021.

Botao Zhao, Xiaoyang Qu, Zuheng Kang, Junqing Peng, Jing Xiao, and Jianzong Wang. Accon: Angle-compensated contrastive regularizer for deep regression. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, pp. 22750–22758, 2025.

Jiayu Zhou, Lei Yuan, Jun Liu, and Jieping Ye. A multi-task learning formulation for predicting disease progression. In *Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 814–822, 2011.

Jiayu Zhou, Jun Liu, Vaibhav A Narayan, and Jieping Ye. Modeling disease progression via fused sparse group lasso. In *Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 1095–1103, 2012.

Menghui Zhou and Po Yang. Automatic temporal relation in multi-task learning. In *Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pp. 3570–3580, 2023.

Menghui Zhou, Yu Zhang, Tong Liu, Yun Yang, and Po Yang. Multi-task learning with adaptive global temporal structure for predicting alzheimer's disease progression. In *Proceedings of the 31st ACM International Conference on Information & Knowledge Management*, pp. 2743–2752, 2022a.

Menghui Zhou, Yu Zhang, Tong Liu, Yun Yang, and Po Yang. Efficient multi-task learning with adaptive temporal structure for progression prediction. *Neural Computing and Applications*, pp. 3570–3580, 2023a.

Menghui Zhou, Yu Zhang, Yun Yang, Tong Liu, and Po Yang. Robust temporal smoothness in multi-task learning. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 37, pp. 11426–11434, 2023b.

Menghui Zhou, Xulong Wang, Tong Liu, Yun Yang, and Po Yang. Integrating visualised automatic temporal relation graph into multi-task learning for alzheimer's disease progression prediction. *IEEE Transactions on Knowledge and Data Engineering*, 36(10):5206–5220, 2024.

Xiao Zhou, Yong Lin, Renjie Pi, Weizhong Zhang, Renzhe Xu, Peng Cui, and Tong Zhang. Model agnostic sample reweighting for out-of-distribution learning. In *International conference on machine learning*, pp. 27203–27221. PMLR, 2022b.

Appendix

A Related Work

A.1 Real-World DMO Estimation and Clinical Validation

Research on DMOs has examined whether mobility can be estimated accurately from wearable-sensor signals and whether the resulting measures are clinically meaningful (Goldsack et al., 2020; Cerreta et al., 2020; Polhemus et al., 2021). Consensus definitions have improved comparability across sensors, algorithms, and walking-bout definitions (Kluge et al., 2021). The Mobilise-D Technical Validation Study further demonstrated the feasibility of deriving real-world DMOs from a single wearable device across several mobility-limiting conditions (Mazzà et al., 2021; Keogh et al., 2023). Detection and cadence estimation were generally robust, whereas stride length and walking speed were more affected by context and impairment severity (Micó-Amigo et al., 2023; Kirk et al., 2024). Clinical evidence remains dominated by known-groups comparisons and concurrent associations, with less support for predictive validity, responsiveness, and ecological validity (Polhemus et al., 2021). Differences between laboratory and daily-life gait further motivate real-world validation (Shah et al., 2020; Corrà et al., 2021). Recent Mobilise-D studies have related individual DMOs to clinical measures and severity or recovery groups (Megaritis et al., 2026; Eckert et al., 2026), but do not show how multiple DMOs should be combined or how their relationships with clinical status evolve over time.

This literature reflects the distinction between technical and clinical validation. Technical studies establish that a sensor and algorithm can measure a walking construct with acceptable accuracy, whereas clinical validation asks whether that construct is associated with an outcome of interest and is fit for a specified context of use (Goldsack et al., 2020; Rochester et al., 2020). Mobilise-D was designed around this staged process: its technical validation study evaluated the measurement pipeline across conditions, while the Clinical Validation Study linked harmonised real-world DMOs to disease-specific clinical outcomes (Mazzà et al., 2021; Mikolaizak et al., 2022). Construct-validity analyses in COPD, PFF, and MS provide important outcome-specific evidence (Megaritis et al., 2026; Eckert et al., 2026; Brittain et al., 2026), but are primarily cross-sectional. DeMMO complements these studies by analysing the joint longitudinal contribution of the complete DMO panel rather than testing each DMO separately.

A.2 Longitudinal DMO Mining

Longitudinal DMO research remains limited. Reviews in PD found few studies with extended follow-up and a literature dominated by active or hospital-based tests, patient–control comparisons, and cross-sectional associations, leaving longitudinal change and predictive utility underexplored (Rábano-Suárez et al., 2025; Kirk et al., 2025). Emerging studies show that selected measures can capture change in PD, MS, and hip-fracture recovery (Mirelman et al., 2024; Poleur et al., 2026; Engdal et al., 2024). However, they are disease-specific and assess only a few prespecified outcomes; they do not jointly model how multivariate DMO–outcome relationships evolve across visits. To our knowledge, stable and visit-specific DMOs have not been systematically identified across multiple outcomes and mobility-limiting conditions. The Mobilise-D Clinical Validation Study provides the harmonised longitudinal data needed to address this gap (Mikolaizak et al., 2022).

Longitudinal evidence can address several distinct questions: whether a DMO changes over time, whether its change tracks clinical progression, whether it predicts a future endpoint, and whether its association with an endpoint is stable across visits. These questions are not interchangeable. A measure may show a population-level trend but contribute little to individual prediction, or remain predictive even when its marginal mean changes little. Existing studies have mainly examined trajectories of individual measures or their association with a single clinical endpoint (Mirelman et al., 2024; Engdal et al., 2024; Poleur et al., 2026). In contrast, our setting estimates a visit-specific multivariate mapping from 24 correlated DMOs to each clinical outcome. Stability selection then distinguishes effects that persist across visits and participant subsamples from those that are visit-specific, without interpreting statistical selection alone as clinical validation.

Attrition and incomplete follow-up add a further difficulty. Later visits often contain fewer observations, making independent visit-wise models unstable, whereas complete-case trajectory analysis would discard participants with partially observed follow-up. DeMMO retains all valid participant–visit records and borrows strength through coefficient-level temporal regularisation. It therefore targets the evolution of DMO–outcome associations rather than requiring every participant to have a complete five-visit trajectory.

A.3 Multi-Task Learning for Longitudinal Modelling

Multi-task learning (MTL) (Zhang & Yang, 2021; Ma et al., 2022; Li et al., 2020a) has been applied to disease progression by treating predictions at different future time points as related tasks. Studies in Alzheimer's disease use temporal smoothness, structured sparsity, and adaptive or robust regularisers to identify shared and time-specific biomarkers (Zhou et al., 2011; 2012; 2022a; 2023b; 2024). These methods generally predict multiple future outcomes from baseline features (Zhou et al., 2011; 2012; 2022b; 2024; Fan et al., 2026); in Mobilise-D, both DMOs and outcomes are repeatedly observed, so the goal is instead to model how their relationships evolve across visits. Mobilise-D also spans multiple outcomes and diseases. Recent MTL studies often define outcome relationships through patient-level correlations within one disease (Fan et al., 2026); this is unsuitable for non-overlapping disease cohorts. Cross-disease relationships must instead be learned from shared DMO-based prediction structures while preserving disease-specific patterns. Fused Lasso and sparse group Lasso provide mechanisms for temporal smoothness and structured selection (Tibshirani et al., 2005; Zhou et al., 2012; Simon et al., 2013). Here, they identify stable and visit-specific DMOs among 24 measures for parsimony and clinical interpretability. Existing MTL methods do not jointly accommodate repeated DMO–outcome relationships, multiple outcomes, and non-overlapping disease cohorts, motivating a framework that learns temporal, outcome-specific, and cross-disease structures simultaneously.

The task definition is central to this distinction. In conventional longitudinal MTL, tasks commonly correspond to future time points, and a shared baseline representation is used to predict several follow-up outcomes (Zhou et al., 2011; 2012; 2022b). Robust and adaptive extensions allow task-specific deviations or learn relations among these time points (Zhou et al., 2023b; 2024). In DeMMO, each disease–outcome pair is instead one prediction objective, and its coefficient matrix contains the complete sequence of visit-specific DMO mappings. Temporal dependence is modelled within each objective, while relation learning operates between objectives. This separation prevents a strong temporal relation from being confused with evidence that two clinical outcomes share the same DMO structure.

Cross-outcome learning also differs from ordinary multi-output regression. The two PD outcomes are observed in the same cohort, but MS and PFF outcomes come from different participants and have different clinical scales. Raw outcome covariance is therefore unavailable across diseases and would not, in any case, directly describe similarity between their DMO effects. DeMMO learns a symmetric signed relation graph from the longitudinal coefficient mappings. Positive and negative weights can represent aligned and oppositely oriented DMO patterns, while the zero diagonal excludes trivial self-relations. This parameter-level mechanism enables selective sharing across non-overlapping cohorts without pooling participant records or assuming identical effects.

A.4 Deep Learning for Disease Progression Modelling

Deep learning has been widely applied to large-scale electronic health records, medical images, and physiological signals. Representative models learn patient representations or temporal patterns to predict diseases, diagnoses, medications, and future conditions (Miotto et al., 2016; Choi et al., 2016; Li et al., 2020b; Bai et al., 2018). Recurrent models have also jointly predicted longitudinal imaging, diagnoses, cognitive scores, and ventricular volume in Alzheimer's disease while accommodating missing observations (Ghazi et al., 2019; Nguyen et al., 2020). These methods can capture nonlinear temporal dependencies but generally benefit from high-dimensional inputs, dense observations, or large patient populations. Mobilise-D instead contains 24 structured DMOs, at most five visits per participant, and smaller disease- and outcome-specific samples. Its main challenge is to identify interpretable, time-varying DMO relationships across clinical outcomes and diseases. A structured

multi-task model therefore provides more appropriate inductive biases and more interpretable effects than a high-capacity deep sequence model.

Deep models for irregular clinical time series can explicitly encode missing values, elapsed time, or attention across observations (Che et al., 2018; Shukla & Marlin, 2021). Such architectures are valuable when the principal signal lies in dense, heterogeneous event sequences. Here, however, the inputs are a small, clinically defined DMO panel observed on a fixed five-visit schedule, and the scientific objective includes identifying which DMO is relevant at which visit. Post-hoc feature attribution would not provide the same direct coefficient trajectories or structured sparsity.

We nevertheless include neural regression and continuous-label representation learning baselines in the empirical comparison. This tests whether nonlinear capacity or label-aware embeddings improve prediction in the available sample regime. The comparison is therefore not intended to claim that structured linear models dominate deep learning generally; rather, it evaluates whether the temporal, sparse, and cross-objective inductive biases of DeMMO are better matched to this particular longitudinal clinical problem.

B Optimization

The unified objective in Equation 5 is not jointly convex in the longitudinal coefficient matrices and relation matrix because of the multiplicative term $\mathbf{UA}$. However, it is convex in either block when the other is fixed. We therefore use alternating optimisation to update the longitudinal coefficient matrices $\{\mathbf{W}_{d,q}\}$ with $\mathbf{A}$ fixed, followed by the symmetric relation matrix $\mathbf{A}$ with $\{\mathbf{W}_{d,q}\}$ fixed.

For simplicity, we assign each disease–outcome pair a unique index $m \in \{1, \ldots, M\}$, where $M = 4$, and denote its coefficient matrix by $\mathbf{W}_m = [\mathbf{w}_{m1}, \ldots, \mathbf{w}_{mT}] \in \mathbb{R}^{p \times T}$. The corresponding data at visit t are denoted by $\mathbf{X}_{mt}$, $\mathbf{y}_{mt}$, and n_{mt}.

B.1 Updating the Longitudinal Coefficient Matrices

For fixed $\mathbf{A}$, the coefficient subproblem is decomposed into the smooth component

$$f(\{\mathbf{W}_m\}; \mathbf{A}) = \sum_{m=1}^{M} \sum_{t=1}^{T} \frac{1}{2n_{mt}} \|\mathbf{y}_{mt} - \mathbf{X}_{mt}\mathbf{w}_{mt}\|_2^2 + \frac{\lambda_{\mathrm{R}}}{2} \|\mathbf{U} - \mathbf{UA}\|_F^2 , \tag{6}$$

and the non-smooth component

$$g(\{\mathbf{W}_m\}) = \sum_{m=1}^{M} \left[\lambda_{\mathrm{FL}} \|\mathbf{W}_m \mathbf{R}^{\mathsf{T}}\|_1 + \lambda_{\mathrm{L}} \|\mathbf{W}_m\|_1 + \lambda_{\mathrm{G}} \|\mathbf{W}_m\|_{2,1}\right] . \tag{7}$$

The relation penalty $\lambda_{\mathrm{A}} \|\mathbf{A}\|_F^2/2$ is constant in this subproblem and is therefore omitted. We minimise $f + g$ using accelerated proximal gradient with backtracking line search.

The regression gradient for objective m at visit t is

$$\nabla_{\mathbf{w}_{mt}} f_{\mathrm{reg}} = \frac{1}{n_{mt}} \mathbf{X}_{mt}^{\mathsf{T}} \left(\mathbf{X}_{mt}\mathbf{w}_{mt} - \mathbf{y}_{mt}\right) . \tag{8}$$

For the relation term, define $\mathbf{B} = \mathbf{I} - \mathbf{A}$. Then $f_{\mathrm{rel}} = \lambda_{\mathrm{R}} \|\mathbf{UB}\|_F^2/2$, with gradient

$$\nabla_{\mathbf{U}} f_{\mathrm{rel}} = \lambda_{\mathrm{R}} \mathbf{UBB}^{\mathsf{T}}. \tag{9}$$

Because column m of $\mathbf{U}$ is the vectorised form of $\mathbf{W}_m$, its relation gradient is reshaped into a $p \times T$ matrix. The complete gradient is therefore

$$\nabla_{\mathbf{W}_m} f = \nabla_{\mathbf{W}_m} f_{\mathrm{reg}} + \mathrm{unvec}_{p \times T} \left([\nabla_{\mathbf{U}} f_{\mathrm{rel}}]_{:m}\right) . \tag{10}$$

Let $\mathbf{Z}_m^{(k)}$ be the extrapolated coefficient matrix at inner iteration k, and let L_k be the current estimate of the Lipschitz constant (Nesterov, 1983; 2003). The proximal-gradient step is

$$\mathbf{V}_m^{(k)} = \mathbf{Z}_m^{(k)} - \frac{1}{L_k} \nabla_{\mathbf{W}_m} f(\{\mathbf{Z}_m^{(k)}\}; \mathbf{A}), \tag{11}$$

followed by

$$\mathbf{W}_m^{(k+1)} = \mathrm{prox}_{g_m/L_k}\left(\mathbf{V}_m^{(k)}\right). \tag{12}$$

Composite proximal operator. The coefficient update involves a composite penalty consisting of fused Lasso, element-wise Lasso, and group Lasso terms. Although proximal operators cannot generally be composed in an arbitrary order, this fused sparse group-Lasso penalty admits an exact and efficient decomposition (Zhou et al., 2023b; 2012).

Specifically, the proximal problem is separable across prediction objectives and DMO rows. For objective m and DMO j, let $\mathbf{v}_{mj} \in \mathbb{R}^T$ denote the j-th row of $\mathbf{V}_m^{(k)}$. The corresponding update is

$$\mathbf{w}_{mj}^{(k+1)} = \arg\min_{\mathbf{w}\in\mathbb{R}^T} \frac{1}{2}\|\mathbf{w}-\mathbf{v}_{mj}\|_2^2 + \frac{\lambda_{\mathrm{L}}}{L_k}\|\mathbf{w}\|_1 + \frac{\lambda_{\mathrm{FL}}}{L_k}\|\mathbf{R}\mathbf{w}\|_1 + \frac{\lambda_{\mathrm{G}}}{L_k}\|\mathbf{w}\|_2. \tag{13}$$

This problem is solved exactly in two stages. First, the fused and element-wise sparse solution is obtained from

$$\widetilde{\mathbf{w}}_{mj} = \arg\min_{\mathbf{w}\in\mathbb{R}^T} \frac{1}{2}\|\mathbf{w}-\mathbf{v}_{mj}\|_2^2 + \frac{\lambda_{\mathrm{L}}}{L_k}\|\mathbf{w}\|_1 + \frac{\lambda_{\mathrm{FL}}}{L_k}\|\mathbf{R}\mathbf{w}\|_1. \tag{14}$$

Second, row-wise group shrinkage is applied:

$$\mathbf{w}_{mj}^{(k+1)} = \left(1 - \frac{\lambda_{\mathrm{G}}}{L_k\|\widetilde{\mathbf{w}}_{mj}\|_2}\right)_+ \widetilde{\mathbf{w}}_{mj}, \tag{15}$$

where $(a)_+ = \max(a, 0)$. The first stage identifies temporally fused and visit-specific coefficients, while the second removes the complete DMO trajectory when its magnitude is insufficient. This decomposition exactly evaluates the composite proximal operator and can be implemented using established fast fused-Lasso algorithms.

Backtracking line search is used to select L_k, and Nesterov acceleration is applied after each accepted update. The inner iterations terminate when the relative change in the coefficient-block objective falls below a prescribed tolerance.

B.2 Updating the Symmetric Relation Matrix

For fixed coefficient matrices, $\mathbf{U}$ is constant and the relation subproblem becomes

$$\begin{aligned} \min_{\mathbf{A}} \quad & \frac{\lambda_{\mathrm{R}}}{2}\|\mathbf{U}-\mathbf{U}\mathbf{A}\|_F^2 + \frac{\lambda_{\mathrm{A}}}{2}\|\mathbf{A}\|_F^2 \\ \text{s.t.} \quad & \mathbf{A} = \mathbf{A}^{\mathsf{T}}, \qquad \mathrm{diag}(\mathbf{A}) = \mathbf{0}. \end{aligned} \tag{16}$$

This is a strongly convex quadratic problem because $\lambda_{\mathrm{A}} > 0$. To satisfy symmetry and the zero-diagonal constraint exactly, we parameterise only the $K = M(M-1)/2$ upper-triangular entries.

For each pair $i < j$, define $\mathbf{E}_{ij} = \mathbf{e}_i\mathbf{e}_j^{\mathsf{T}} + \mathbf{e}_j\mathbf{e}_i^{\mathsf{T}}$, where $\mathbf{e}_i$ is the i-th standard basis vector in $\mathbb{R}^M$. Any feasible relation matrix can then be written as $\mathbf{A}(\mathbf{a}) = \sum_{i<j} a_{ij}\mathbf{E}_{ij}$, where $\mathbf{a} \in \mathbb{R}^K$. Let $\mathbf{b} = \mathrm{vec}(\mathbf{U})$ and define

$$\mathbf{C} = \left[\mathrm{vec}(\mathbf{U}\mathbf{E}_{12}), \mathrm{vec}(\mathbf{U}\mathbf{E}_{13}), \ldots, \mathrm{vec}(\mathbf{U}\mathbf{E}_{M-1,M})\right]. \tag{17}$$

Since $\mathrm{vec}(\mathbf{U}\mathbf{A}(\mathbf{a})) = \mathbf{C}\mathbf{a}$ and $\|\mathbf{A}(\mathbf{a})\|_F^2 = 2\|\mathbf{a}\|_2^2$, the constrained matrix problem reduces to

$$\min_{\mathbf{a}\in\mathbb{R}^K} \quad \frac{\lambda_{\mathrm{R}}}{2}\|\mathbf{b}-\mathbf{C}\mathbf{a}\|_2^2 + \lambda_{\mathrm{A}}\|\mathbf{a}\|_2^2. \tag{18}$$

Its unique minimiser is obtained from

$$\left(\lambda_{\mathrm{R}}\mathbf{C}^{\mathsf{T}}\mathbf{C} + 2\lambda_{\mathrm{A}}\mathbf{I}\right)\mathbf{a} = \lambda_{\mathrm{R}}\mathbf{C}^{\mathsf{T}}\mathbf{b}. \tag{19}$$

The relation matrix is then formed as $\mathbf{A} = \sum_{i<j} a_{ij}\mathbf{E}_{ij}$. This parameterisation satisfies both constraints by construction. In this study, $M = 4$ and hence $K = 6$, so the update can be computed efficiently using a standard linear-system solver.

Algorithm 1 Alternating Optimisation of the Longitudinal Coefficient Matrices and Cross-Objective Relation Matrix.

Require: Training data $\{\mathbf{X}_{mt}, \mathbf{y}_{mt}\}_{m=1,t=1}^{M,T}$; regularisation parameters $\lambda_{\mathrm{FL}}, \lambda_{\mathrm{L}}, \lambda_{\mathrm{G}}, \lambda_{\mathrm{R}}, \lambda_{\mathrm{A}}$; maximum number of outer iterations $R_{\max}$

Initialise $\{\mathbf{W}_m^{(0)}\}_{m=1}^{M}$ and $\mathbf{A}^{(0)} = \mathbf{0}$
for $r = 0, \ldots, R_{\max} - 1$ **do**
  Fix $\mathbf{A}^{(r)}$ and update $\{\mathbf{W}_m^{(r+1)}\}_{m=1}^{M}$ using accelerated proximal-gradient iterations with backtracking line search
  Evaluate the composite proximal operator row-wise using fused-Lasso signal approximation followed by group shrinkage
  Construct $\mathbf{U}^{(r+1)}$ by vectorising $\{\mathbf{W}_m^{(r+1)}\}_{m=1}^{M}$
  Construct $\mathbf{C}^{(r+1)}$ and $\mathbf{b}^{(r+1)}$ from $\mathbf{U}^{(r+1)}$
  Solve

$$\left(\lambda_{\mathrm{R}}\mathbf{C}^{\mathsf{T}}\mathbf{C} + 2\lambda_{\mathrm{A}}\mathbf{I}\right)\mathbf{a} = \lambda_{\mathrm{R}}\mathbf{C}^{\mathsf{T}}\mathbf{b}$$

  Form the symmetric, zero-diagonal relation matrix $\mathbf{A}^{(r+1)}$ from $\mathbf{a}$
  if the outer convergence criterion is satisfied **then**
    break
  end if
end for

Ensure: $\{\mathbf{W}_m\}_{m=1}^{M}$ and $\mathbf{A}$

B.3 Alternating Optimisation Algorithm

The complete optimisation procedure is summarised in Algorithm 1. We initialise the relation matrix as $\mathbf{A}^{(0)} = \mathbf{0}$. The coefficient matrices may be initialised either at zero or using independently fitted visit-specific regression models. During subsequent outer iterations, each coefficient subproblem is warm-started from the solution obtained in the preceding iteration.

The outer iterations terminate when the relative change in the complete objective satisfies

$$\frac{\left|\mathcal{J}^{(r+1)} - \mathcal{J}^{(r)}\right|}{\max\left(1, \left|\mathcal{J}^{(r)}\right|\right)} \leq \varepsilon, \tag{20}$$

where $\mathcal{J}^{(r)}$ denotes the unified objective value at outer iteration r. A maximum number of outer iterations is imposed as an additional safeguard.

B.3.1 Convergence and Computational Complexity

For fixed $\mathbf{A}$, the coefficient subproblem is convex but non-smooth. Because its composite proximal operator can be computed exactly, the accelerated proximal-gradient procedure with backtracking converges to the global minimiser of this subproblem. For fixed coefficient matrices, the relation-matrix subproblem is strongly convex when $\lambda_{\mathrm{A}} > 0$ and therefore has a unique minimiser.

If both block subproblems are solved exactly, or to sufficient numerical accuracy, each outer iteration does not increase the unified objective. Because the objective is bounded below, the sequence of objective values converges. Under standard regularity conditions, every accumulation point of the alternating sequence is a block-coordinate stationary point. However, the complete objective is biconvex rather than jointly convex because of the interaction between $\mathbf{U}$ and $\mathbf{A}$. Convergence to a global minimiser is therefore not guaranteed and may depend on the initialisation.

For the coefficient update, evaluating the regression gradients across all objectives and visits requires

$$\mathcal{O}\left(p \sum_{m=1}^{M} \sum_{t=1}^{T} n_{mt}\right)$$

operations. The relation-gradient term involves multiplication between the $pT \times M$ coefficient representation and the $M \times M$ relation matrices, requiring $\mathcal{O}\left(pTM^2\right)$ operations.

The composite proximal operator is separable across the Mp longitudinal coefficient trajectories. When a linear-time one-dimensional fused-Lasso solver is used, its overall cost is

$$\mathcal{O}\left(MpT\right).$$

Thus, if I_{in} accelerated proximal-gradient iterations are required, the coefficient-update cost per outer iteration is approximately

$$\mathcal{O}\left[I_{\text{in}}\left(p\sum_{m=1}^{M}\sum_{t=1}^{T}n_{mt}+pTM^2+MpT\right)\right]. \tag{21}$$

The relation update contains $K = M(M-1)/2$ free edge variables. Constructing its quadratic system depends on the pT-dimensional representations of the M objectives, while solving the resulting $K \times K$ linear system by a direct method requires $\mathcal{O}(K^3)$ operations. Because K depends only on the number of prediction objectives, this update remains inexpensive when M is small.

In the present study, $p = 24$, $T = 5$, $M = 4$, and $K = 6$. Consequently, the composite proximal operations and relation-matrix update are small, and the overall computational cost is dominated by repeated regression-gradient evaluations across participants, visits, and prediction objectives.

C Non-Convex Extensions of DeMMO

The fused-Lasso, Lasso, and group-Lasso regularisers in DeMMO provide a convex and computationally tractable formulation for temporal smoothing and DMO selection. However, convex sparsity penalties may over-shrink the coefficients of informative DMOs, particularly when their effects are weak, correlated, or confined to specific visits, thereby introducing estimation bias (Fan & Li, 2001). Such shrinkage may reduce both predictive performance and the ability to recover clinically meaningful longitudinal patterns.

Motivated by non-convex extensions of the fused sparse-group Lasso (Zhou et al., 2012), we introduce two variants, DeMMO-var1 and DeMMO-var2. Both use a concave square-root penalty to reduce shrinkage of sufficiently strong DMO effects while retaining sparsity. They differ in how DMO selection and temporal smoothness are organised. DeMMO-var1 treats them as separate components, allowing their strengths to be controlled independently. DeMMO-var2 couples them within a single DMO-level penalty, such that retention of a DMO depends jointly on its coefficient magnitude and temporal variation. Comparing these variants allows us to examine whether longitudinal DMO modelling benefits more from independent or coupled regularisation.

For concise notation, let $\mathbf{w}_{d,q,j:} \in \mathbb{R}^T$ denote the coefficients of DMO j across all visits for disease d and outcome q. Both variants retain the longitudinal regression loss $\mathcal{L}(\mathcal{W})$ and the relation-learning regulariser

$$\mathcal{R}_{\text{rel}}(\mathbf{U},\mathbf{A}) = \frac{\lambda_{\text{R}}}{2}\|\mathbf{U}-\mathbf{U}\mathbf{A}\|_F^2 + \frac{\lambda_{\text{A}}}{2}\|\mathbf{A}\|_F^2. \tag{22}$$

C.1 DeMMO-var1: Separate Selection and Temporal Smoothing

DeMMO-var1 replaces the Lasso and group-Lasso penalties with a composite $\ell_{(0.5,1)}$ penalty while retaining fused temporal regularisation as a separate component:

$$\begin{aligned}\min_{\mathcal{W},\mathbf{A}} \quad & \mathcal{L}(\mathcal{W}) + \lambda_{\text{NC}}\sum_{d,q,j}\sqrt{\|\mathbf{w}_{d,q,j:}\|_1} + \lambda_{\text{FL}}\sum_{d,q}\|\mathbf{W}_{d,q}\mathbf{R}^{\mathsf{T}}\|_1 + \mathcal{R}_{\text{rel}}(\mathbf{U},\mathbf{A}) \\ \text{s.t.} \quad & \mathbf{A} = \mathbf{A}^{\mathsf{T}}, \qquad \operatorname{diag}(\mathbf{A}) = \mathbf{0}.\end{aligned} \tag{23}$$

The outer square root promotes DMO-level sparsity, whereas the inner ℓ_1 norm permits visit-specific coefficients to be removed. Temporal smoothness is controlled independently by λ_{FL}.

C.2 DeMMO-var2: Coupled Selection and Temporal Smoothing

DeMMO-var2 instead incorporates coefficient sparsity and temporal variation within the same DMO-level penalty. Define

$$\phi_{d,q,j}(\mathcal{W}) = \|\mathbf{w}_{d,q,j:}\mathbf{R}^{\mathsf{T}}\|_1 + \beta\|\mathbf{w}_{d,q,j:}\|_1, \tag{24}$$

Algorithm 2 Non-Convex DeMMO Optimisation.

Require: Longitudinal data, variant parameters, ϵ, and tolerances
Initialise $(\mathcal{W}^{(0)}, \mathbf{A}^{(0)})$ from convex DeMMO
for $k = 0, 1, \ldots, K_{\mathrm{DC}} - 1$ **do**
  Compute $a_{d,q,j}^{(k)}$ for var1 or $b_{d,q,j}^{(k)}$ for var2 using Eq. (27)
  repeat
    Update $\mathcal{W}$ by accelerated proximal gradient on Eq. (28) or Eq. (29)
    Update the symmetric zero-diagonal $\mathbf{A}$ by solving the relation-learning linear system
  until the convex majoriser satisfies the block-convergence criterion
  Evaluate the original non-convex objective
  if its relative change is below $\varepsilon_{\mathrm{DC}}$ **then**
    break
  end if
end for
Ensure: $\mathcal{W}$ and $\mathbf{A}$

where β controls their relative contributions. The objective becomes

$$\begin{aligned} \min_{\mathcal{W},\mathbf{A}} \quad & \mathcal{L}(\mathcal{W}) + \lambda_{\mathrm{NC}} \sum_{d,q,j} \sqrt{\phi_{d,q,j}(\mathcal{W})} + \mathcal{R}_{\mathrm{rel}}(\mathbf{U}, \mathbf{A}) \\ \text{s.t.} \quad & \mathbf{A} = \mathbf{A}^{\mathsf{T}}, \qquad \mathrm{diag}(\mathbf{A}) = \mathbf{0}. \end{aligned} \tag{25}$$

Unlike DeMMO-var1, DeMMO-var2 determines whether to retain a DMO according to both the magnitude and temporal variation of its longitudinal coefficients.

C.3 Optimisation

Unlike convex DeMMO, the two variants require an additional difference-of- convex outer loop. We use the iterative reweighting procedure established for non-convex fused sparse-group Lasso models (Zhou et al., 2012; 2023b). For a non-negative quantity s, the concavity of the square root gives the majoriser

$$\sqrt{s + \epsilon} \leq \sqrt{s^{(k)} + \epsilon} + \frac{s - s^{(k)}}{2\sqrt{s^{(k)} + \epsilon}}, \tag{26}$$

where equality holds at $s = s^{(k)}$. Thus, at reweighting iteration k, the non-convex penalties are replaced by weighted convex ℓ_1 and fused-Lasso penalties. Define

$$a_{d,q,j}^{(k)} = \frac{\lambda_{\mathrm{NC}}}{2\sqrt{\|\mathbf{w}_{d,q,j:}^{(k)}\|_1 + \epsilon}}, \qquad b_{d,q,j}^{(k)} = \frac{\lambda_{\mathrm{NC}}}{2\sqrt{\phi_{d,q,j}(\mathcal{W}^{(k)}) + \epsilon}}, \tag{27}$$

where $\epsilon > 0$ prevents unbounded weights. Ignoring terms independent of $\mathcal{W}$, the coefficient majorisers for DeMMO-var1 and DeMMO-var2 are, respectively,

$$\mathcal{Q}_{\mathrm{var1}}^{(k)} = \mathcal{L}(\mathcal{W}) + \sum_{d,q,j} a_{d,q,j}^{(k)} \|\mathbf{w}_{d,q,j:}\|_1 + \lambda_{\mathrm{FL}} \sum_{d,q,j} \|\mathbf{w}_{d,q,j:}\mathbf{R}^{\mathsf{T}}\|_1 + \mathcal{R}_{\mathrm{rel}}, \tag{28}$$

$$\mathcal{Q}_{\mathrm{var2}}^{(k)} = \mathcal{L}(\mathcal{W}) + \sum_{d,q,j} b_{d,q,j}^{(k)} \left(\|\mathbf{w}_{d,q,j:}\mathbf{R}^{\mathsf{T}}\|_1 + \beta \|\mathbf{w}_{d,q,j:}\|_1 \right) + \mathcal{R}_{\mathrm{rel}}. \tag{29}$$

Both are weighted convex fused-Lasso problems. For fixed $\mathbf{A}$, we solve the coefficient block using accelerated proximal-gradient iterations with backtracking. The proximal step is separable across DMO trajectories and uses the same one-dimensional fused-Lasso solver as convex DeMMO, but with the row-specific weights in Eq. (27). For fixed $\mathcal{W}$, the symmetric relation matrix is updated using the same closed-form quadratic subproblem described in Section B.2.

Algorithm 2 summarises the complete procedure. We initialise both variants from the fitted convex DeMMO solution. This warm start provides a stable coefficient pattern and avoids starting the reweighting scheme near the singular point of the square-root penalty.

Table 3: Comparison of DeMMO and its two non-convex variants over five matched participant splits. Values are mean ± standard deviation. T1–T5 report visit-specific RMSE. The best mean within each objective and metric is shown in bold.

Objective	Method	nMSE ↓	wR ↑	T1	T2	T3	T4	T5
Overall	DeMMO-var1	0.728 ± 0.043	0.509 ± 0.040	–	–	–	–	–
	DeMMO-var2	0.731 ± 0.043	0.505 ± 0.038	–	–	–	–	–
	DeMMO	**0.722 ± 0.032**	**0.515 ± 0.030**	–	–	–	–	–
PD H&Y	DeMMO-var1	0.935 ± 0.043	0.306 ± 0.057	0.540 ± 0.041	0.500 ± 0.059	0.546 ± 0.047	**0.512 ± 0.068**	0.551 ± 0.061
	DeMMO-var2	0.941 ± 0.037	0.295 ± 0.049	0.543 ± 0.042	0.501 ± 0.062	0.548 ± 0.049	0.515 ± 0.067	0.549 ± 0.061
	DeMMO	**0.929 ± 0.035**	**0.319 ± 0.048**	**0.538 ± 0.040**	**0.499 ± 0.059**	**0.543 ± 0.044**	0.513 ± 0.066	**0.546 ± 0.061**
PD MDS–UPDRS	DeMMO-var1	0.876 ± 0.042	**0.385 ± 0.040**	11.966 ± 0.346	11.449 ± 0.429	12.038 ± 1.184	11.343 ± 1.031	11.497 ± 0.491
	DeMMO-var2	0.877 ± 0.046	0.382 ± 0.041	11.982 ± 0.313	**11.443 ± 0.391**	12.057 ± 1.282	**11.327 ± 1.018**	11.561 ± 0.503
	DeMMO	**0.874 ± 0.038**	0.384 ± 0.038	**11.955 ± 0.338**	11.460 ± 0.386	**12.017 ± 1.103**	11.365 ± 0.959	**11.470 ± 0.381**
MS EDSS	DeMMO-var1	0.486 ± 0.059	0.727 ± 0.044	0.819 ± 0.081	0.829 ± 0.063	0.989 ± 0.055	1.019 ± 0.052	1.094 ± 0.092
	DeMMO-var2	0.489 ± 0.058	0.726 ± 0.043	0.821 ± 0.082	0.833 ± 0.059	0.994 ± 0.046	1.022 ± 0.054	1.091 ± 0.086
	DeMMO	**0.479 ± 0.051**	**0.732 ± 0.039**	**0.815 ± 0.081**	**0.822 ± 0.057**	**0.982 ± 0.050**	**1.013 ± 0.047**	**1.084 ± 0.088**
PFF SPPB	DeMMO-var1	0.559 ± 0.073	0.673 ± 0.053	2.025 ± 0.096	2.210 ± 0.154	2.253 ± 0.210	2.280 ± 0.217	2.328 ± 0.161
	DeMMO-var2	0.561 ± 0.078	0.672 ± 0.057	2.044 ± 0.125	2.206 ± 0.149	2.240 ± 0.206	2.276 ± 0.238	2.354 ± 0.157
	DeMMO	**0.549 ± 0.062**	**0.681 ± 0.044**	**2.004 ± 0.062**	**2.203 ± 0.151**	**2.237 ± 0.209**	**2.241 ± 0.219**	**2.307 ± 0.170**

The majorisation step is tight and each inner block update decreases its convex surrogate. Consequently, the original objective is non-increasing and the procedure converges to a stationary point under the usual boundedness and accurate-subproblem assumptions. Global optimality is not guaranteed because the objectives remain non-convex.

Computational cost. Let K_{DC} be the number of reweighting iterations, K_{B} the number of alternating block updates per majoriser, and K_{PG} the number of proximal-gradient iterations. One gradient/proximal update has the same leading cost as convex DeMMO,

$$\mathcal{O}\left(p\sum_{m=1}^{M}\sum_{t=1}^{T}n_{mt}+pTM^2+MpT\right). \tag{30}$$

The complete coefficient cost is therefore

$$\mathcal{O}\left[K_{\mathrm{DC}}K_{\mathrm{B}}K_{\mathrm{PG}}\left(p\sum_{m,t}n_{mt}+pTM^2+MpT\right)\right], \tag{31}$$

in addition to the small relation-matrix solves. Convex DeMMO requires only one alternating optimisation sequence, whereas each non-convex variant solves a sequence of weighted convex DeMMO-like problems. In our implementation, K_{DC}, K_{B}, and K_{PG} are capped at 15, 12, and 300, respectively, with early stopping at every level. The convex warm start substantially reduces the realised number of iterations, but the two variants remain more computationally expensive. Their additional predictive flexibility therefore comes at the cost of longer model selection and training, which is the principal trade-off examined in our experiments.

D Experimental Evaluation of the Non-Convex Variants

Table 3 compares the original convex DeMMO with DeMMO-var1 and DeMMO-var2. We report overall and objective-specific nMSE and wR, together with RMSE at each of the five visits.

DeMMO performs best overall. It achieves an nMSE of 0.722 ± 0.032, compared with 0.728 ± 0.043 for DeMMO-var1 and 0.731 ± 0.043 for DeMMO-var2, while its wR of 0.515 ± 0.030 exceeds 0.509 ± 0.040 and 0.505 ± 0.038, respectively. DeMMO also obtains the lowest nMSE for every clinical outcome and the highest wR for three of the four outcomes; DeMMO-var1 is only marginally higher for PD MDS–UPDRS wR. At the visit level, DeMMO gives the lowest RMSE in 17 of the 20 tasks, whereas var1 leads at PD H&Y T4 and var2 leads at PD MDS–UPDRS T2 and T4.

The two variants retain DeMMO's outcome-specific sparsity, temporal continuity, and cross-objective relation learning, but use adaptive non-convex reweighting to reduce the estimation bias induced by the convex Lasso, group-Lasso, and fused-Lasso penalties. Their consistently weaker aggregate performance therefore suggests that shrinkage-induced bias is not a major limitation in this setting.

Table 4: Visit-specific RMSE for PD H&Y (mean ± standard deviation over five matched participant splits). The best mean in each column is shown in bold.

Method	T1	T2	T3	T4	T5
Ridge	0.541 ± 0.037	0.504 ± 0.065	0.550 ± 0.045	0.526 ± 0.069	0.551 ± 0.064
Lasso	0.549 ± 0.042	0.504 ± 0.067	0.559 ± 0.054	0.523 ± 0.071	**0.546 ± 0.065**
FRoTS	**0.538 ± 0.036**	0.509 ± 0.062	0.545 ± 0.042	0.526 ± 0.065	0.567 ± 0.076
MAGPP	0.539 ± 0.033	0.517 ± 0.066	0.547 ± 0.046	0.538 ± 0.061	0.580 ± 0.082
MLP-MSE	0.552 ± 0.040	0.512 ± 0.062	0.551 ± 0.051	0.534 ± 0.070	0.561 ± 0.073
MLP-L1	0.566 ± 0.040	0.527 ± 0.052	0.573 ± 0.050	0.533 ± 0.076	0.561 ± 0.055
Rank-N	0.570 ± 0.040	0.528 ± 0.054	0.577 ± 0.052	0.534 ± 0.076	0.563 ± 0.054
RankSim	0.553 ± 0.046	0.510 ± 0.063	0.555 ± 0.055	0.532 ± 0.074	0.560 ± 0.064
ACCon	0.554 ± 0.036	0.515 ± 0.057	0.549 ± 0.055	0.539 ± 0.071	0.579 ± 0.086
DeMMO	**0.538 ± 0.040**	**0.499 ± 0.059**	**0.543 ± 0.044**	**0.513 ± 0.066**	**0.546 ± 0.061**

Table 5: Visit-specific RMSE for PD MDS–UPDRS (mean ± standard deviation over five matched participant splits). The best mean in each column is shown in bold.

Method	T1	T2	T3	T4	T5
Ridge	**11.767 ± 0.389**	11.692 ± 0.285	12.200 ± 0.780	11.584 ± 0.884	**11.303 ± 0.348**
Lasso	11.874 ± 0.408	11.571 ± 0.299	12.232 ± 1.015	11.671 ± 1.006	11.559 ± 0.792
FRoTS	11.828 ± 0.506	11.584 ± 0.398	12.211 ± 0.772	11.558 ± 0.794	11.624 ± 0.755
MAGPP	11.807 ± 0.517	11.783 ± 0.568	12.393 ± 0.798	11.878 ± 0.748	11.606 ± 0.654
MLP-MSE	11.918 ± 0.289	11.925 ± 0.563	12.802 ± 1.673	11.945 ± 1.218	11.542 ± 0.365
MLP-L1	12.068 ± 0.513	12.317 ± 0.514	12.141 ± 0.804	11.984 ± 1.110	11.756 ± 0.392
Rank-N	12.001 ± 0.290	11.812 ± 0.402	12.155 ± 0.907	11.898 ± 0.927	12.015 ± 0.681
RankSim	12.004 ± 0.427	11.800 ± 0.248	12.227 ± 1.035	11.861 ± 0.902	11.798 ± 0.408
ACCon	11.951 ± 0.414	12.073 ± 0.573	12.623 ± 1.735	12.050 ± 1.244	11.887 ± 0.335
DeMMO	11.955 ± 0.338	**11.460 ± 0.386**	**12.017 ± 1.103**	**11.365 ± 0.959**	11.470 ± 0.381

We conjecture that this result reflects the substantial behavioural and measurement noise in free-living DMOs. By weakening shrinkage, the non-convex penalties may retain more noise together with the DMO signal, leading to poorer generalisation; the stronger shrinkage of convex DeMMO may instead provide beneficial regularisation. Convex DeMMO also shows lower overall variability and requires substantially less computation because it avoids the additional difference-of-convex iterations. These results favour the original DeMMO formulation for this noisy longitudinal DMO setting.

E Additional Experimental Results

Tables 4–7 provide the numerical results underlying Figure 3.

Outcome-specific observations. For PD H&Y, DeMMO is best or tied for best at all five visits, with its clearest advantages at T2–T4. For PD MDS–UPDRS, it achieves the lowest mean RMSE at T2–T4, while Ridge performs best at T1 and T5. DeMMO is also consistently competitive for MS EDSS: it is best at T1, T2, and T5, tied for best at T3, and only 0.001 above the best result at T4. The PFF task exhibits a different pattern. Ridge is strongest at T1 and MLP-L1 at T2, whereas DeMMO becomes best from T3 to T5. This later-visit advantage is particularly relevant because the PFF sample size decreases sharply across follow-up visits, suggesting that temporal and cross-objective information sharing provides useful regularisation when outcome-specific data are sparse.

Patterns across visits and methods. Across the four outcomes, DeMMO ranks first or second in 18 of the 20 outcome–visit comparisons and is uniquely best in 12. Its performance is not uniformly dominant, however: the two exceptions occur at PD MDS–UPDRS T1 and PFF T1, where simpler linear models perform better. Errors for MS EDSS and PFF generally increase at later visits, consistent

Table 6: Visit-specific RMSE for MS EDSS (mean ± standard deviation over five matched participant splits). The best mean in each column is shown in bold.

Method	T1	T2	T3	T4	T5
Ridge	0.821 ± 0.078	0.841 ± 0.060	0.993 ± 0.055	1.023 ± 0.050	1.102 ± 0.088
Lasso	0.830 ± 0.079	0.845 ± 0.065	0.997 ± 0.070	1.024 ± 0.045	1.098 ± 0.081
FRoTS	0.824 ± 0.072	0.833 ± 0.066	0.987 ± 0.067	**1.012 ± 0.060**	1.121 ± 0.094
MAGPP	0.823 ± 0.072	0.848 ± 0.068	1.004 ± 0.070	1.025 ± 0.069	1.142 ± 0.093
MLP-MSE	0.821 ± 0.083	0.846 ± 0.078	0.985 ± 0.051	1.015 ± 0.056	1.107 ± 0.113
MLP-L1	0.836 ± 0.070	0.870 ± 0.082	1.003 ± 0.037	1.057 ± 0.117	1.116 ± 0.080
Rank-N	0.827 ± 0.094	0.908 ± 0.085	1.030 ± 0.042	1.059 ± 0.084	1.130 ± 0.072
RankSim	0.830 ± 0.072	0.853 ± 0.075	**0.982 ± 0.059**	1.041 ± 0.131	1.116 ± 0.083
ACCon	0.826 ± 0.080	0.855 ± 0.073	1.012 ± 0.051	1.020 ± 0.046	1.113 ± 0.084
DeMMO	**0.815 ± 0.081**	**0.822 ± 0.057**	**0.982 ± 0.050**	1.013 ± 0.047	**1.084 ± 0.088**

Table 7: Visit-specific RMSE for PFF SPPB impairment (mean ± standard deviation over five matched participant splits). The best mean in each column is shown in bold.

Method	T1	T2	T3	T4	T5
Ridge	**1.931 ± 0.072**	2.233 ± 0.186	2.275 ± 0.157	2.340 ± 0.144	2.341 ± 0.211
Lasso	1.942 ± 0.084	2.209 ± 0.130	2.304 ± 0.180	2.418 ± 0.092	2.364 ± 0.158
FRoTS	1.948 ± 0.081	2.206 ± 0.219	2.242 ± 0.115	2.464 ± 0.119	2.365 ± 0.216
MAGPP	1.976 ± 0.078	2.203 ± 0.254	2.276 ± 0.122	2.514 ± 0.119	2.409 ± 0.222
MLP-MSE	2.010 ± 0.061	2.300 ± 0.233	2.351 ± 0.220	2.389 ± 0.240	2.558 ± 0.101
MLP-L1	1.995 ± 0.064	**2.198 ± 0.230**	2.358 ± 0.204	2.418 ± 0.229	2.567 ± 0.237
Rank-N	2.047 ± 0.142	2.220 ± 0.178	2.331 ± 0.185	2.556 ± 0.148	2.606 ± 0.229
RankSim	2.019 ± 0.097	2.259 ± 0.161	2.344 ± 0.201	2.439 ± 0.226	2.599 ± 0.179
ACCon	2.037 ± 0.053	2.224 ± 0.192	2.481 ± 0.192	2.391 ± 0.183	2.536 ± 0.124
DeMMO	2.004 ± 0.062	2.203 ± 0.151	**2.237 ± 0.209**	**2.241 ± 0.219**	**2.307 ± 0.170**

with their declining sample sizes and potentially greater follow-up heterogeneity, whereas both PD outcomes show non-monotonic trajectories. Deep regression and ranking methods rarely lead these visit-specific comparisons and often exhibit larger variability at later visits. Overall, the tables support selective rather than uniform information sharing: DeMMO preserves outcome-specific behaviour while gaining most where longitudinal observations become limited.