Mobile Health for Perioperative Care

Jacob Greenberg, MD, MSCI

Assistant Professor,

Director, AI Initiatives, Patient Safety & Quality Improvement
Department of Neurological Surgery
Washington University School of Medicine

AI For Health
2026

Disclosures

Research funding

NIH (NIAMS, NIMH), DoD, CSRS, SRS, AO Spine North America, Big Ideas Fund, Foundation for Barnes-Jewish Hospital, University of Missouri Spinal Cord Injury/Disease Research Program, Here & Next Program; Medtronic

Consulting

Medtronic, Kuros Biosciences

Outline

- Why Mobile Health for Spine Surgery
- Predicting Surgical Outcomes
- Monitoring Recovery
- Measuring Pain in Daily Life
- Cervical Myelopathy
- State of the Field

Why Mobile Health for Spine Surgery

Burden of Chronic Pain

1.7 billion

people with musculoskeletal disease worldwide

#6

cause of years lived with disability

361 per 1000

person-years: estimated rate of persistent high impact chronic pain in the United States

$135 billion

cost of neck and back pain in the United States (in 2016)

Spine Fusion for Degenerative Disease







Influences on Surgical Outcome

Pain Symptoms

Physical Function

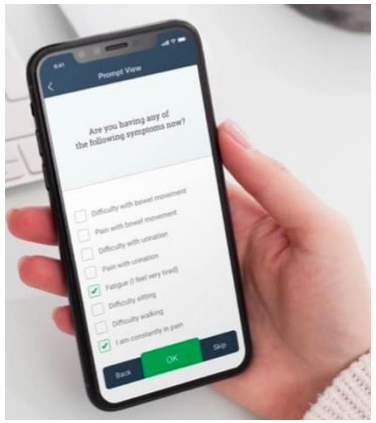
Mental Health

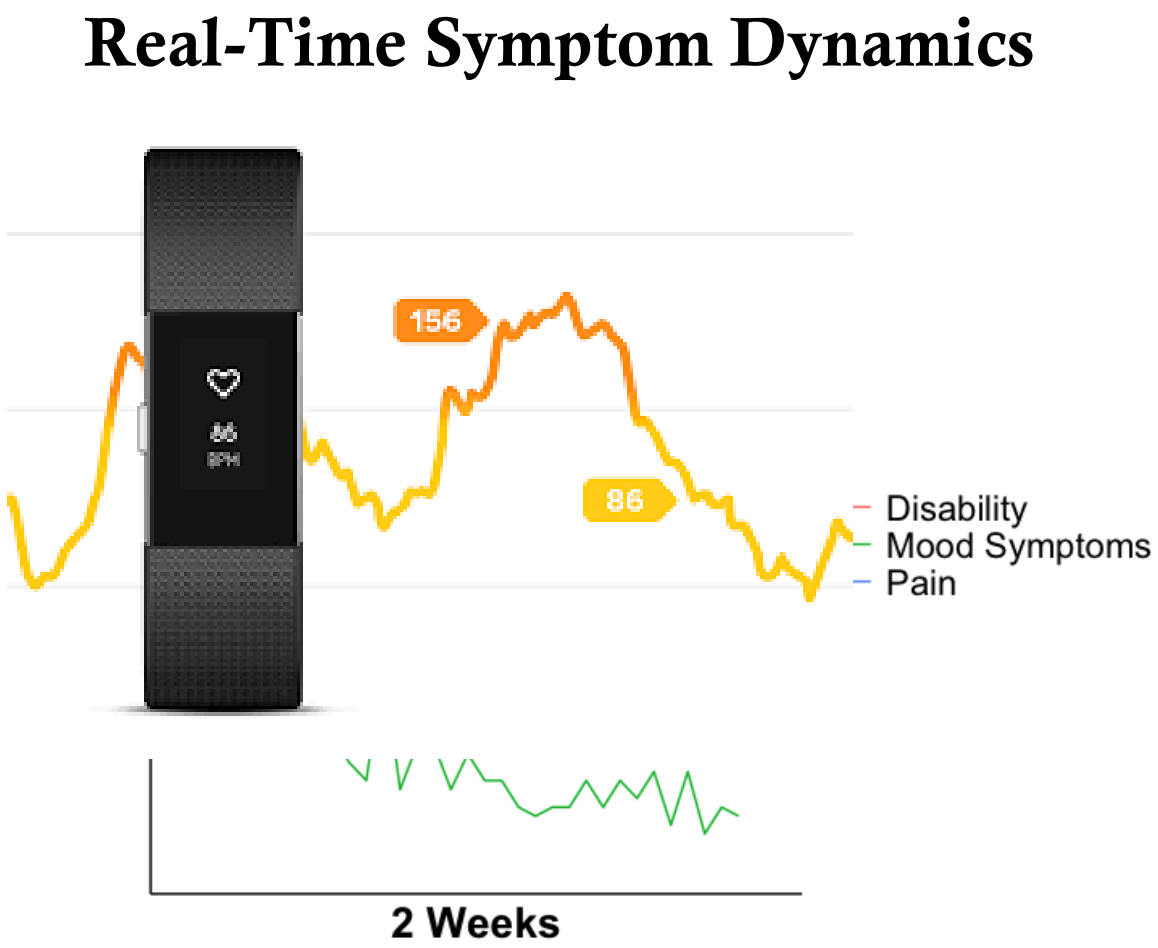
Traditional Methods to Monitor Surgical Recovery





Shortcomings of Current Patient-Reported Outcomes

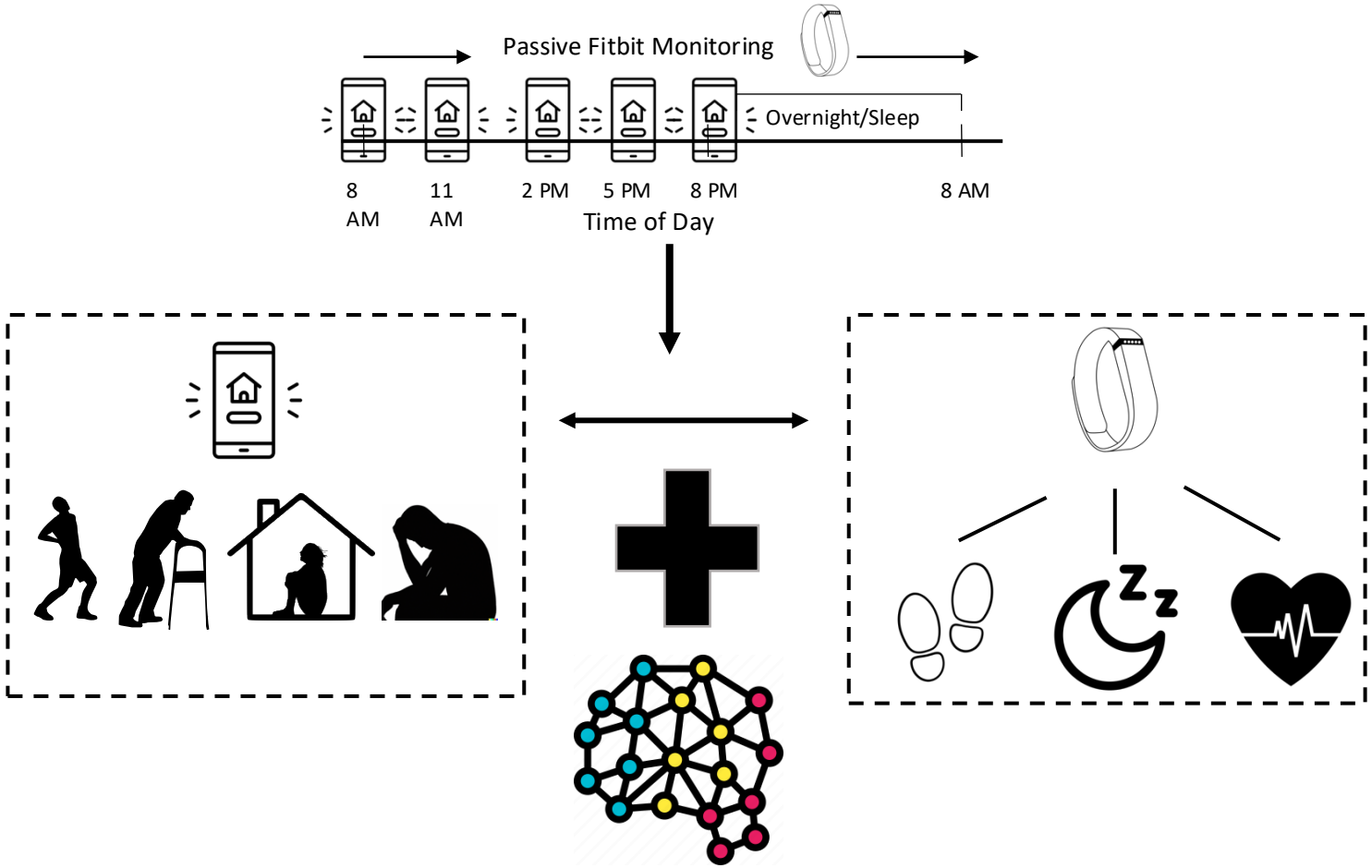




Study Objective

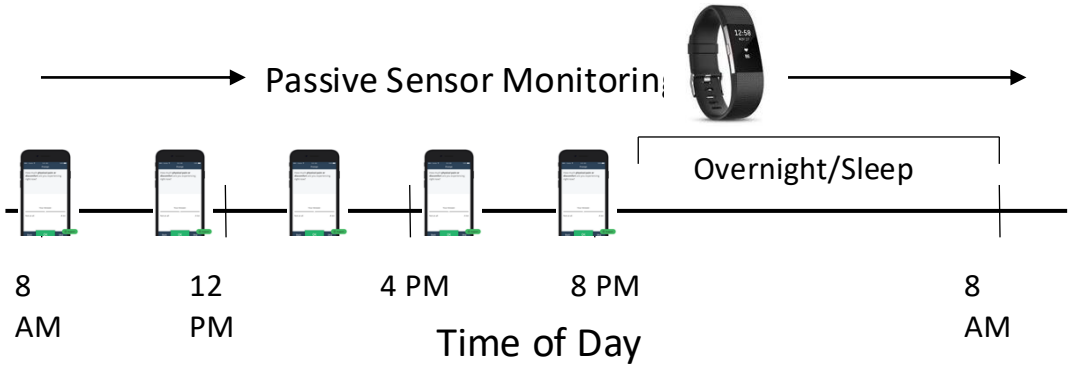
Use multimodal mobile health (mHealth) data to develop individualized disease profiles associated with spine fusion outcomes

Predicting Surgical Outcomes


Passive Fitbit Monitoring
Overnight/Sleep
8 AM
11 AM
2 PM
5 PM
8 PM
8 AM
Time of Day
Pain Interference
Quality of Recovery

Study Methods

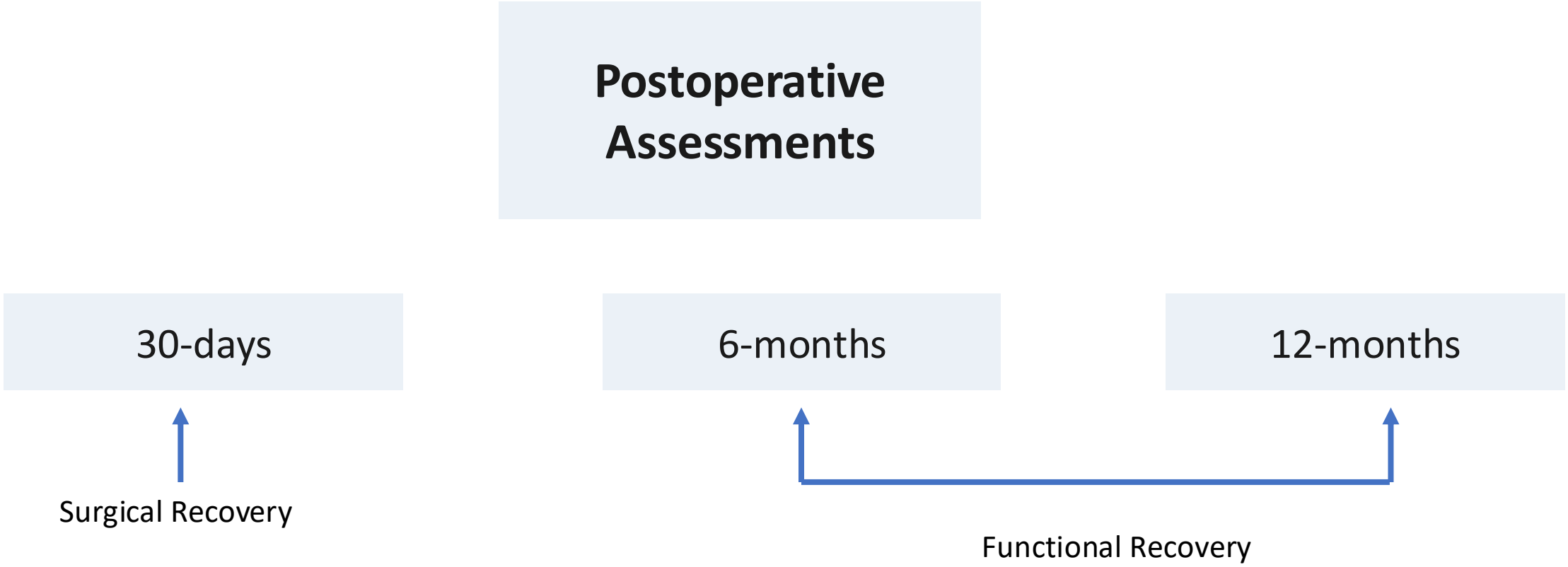
Preoperative mHealth Assessment



Traditional Cross-Sectional PROMs

- Pain (e.g., NRS, PROMIS)
- Disability (e.g., ODI, PROMIS)
- Mood (e.g., PHQ-9)
- Catastrophizing (e.g., PCS)
- Function status: Elixhauser, mFI-5

Study Methods



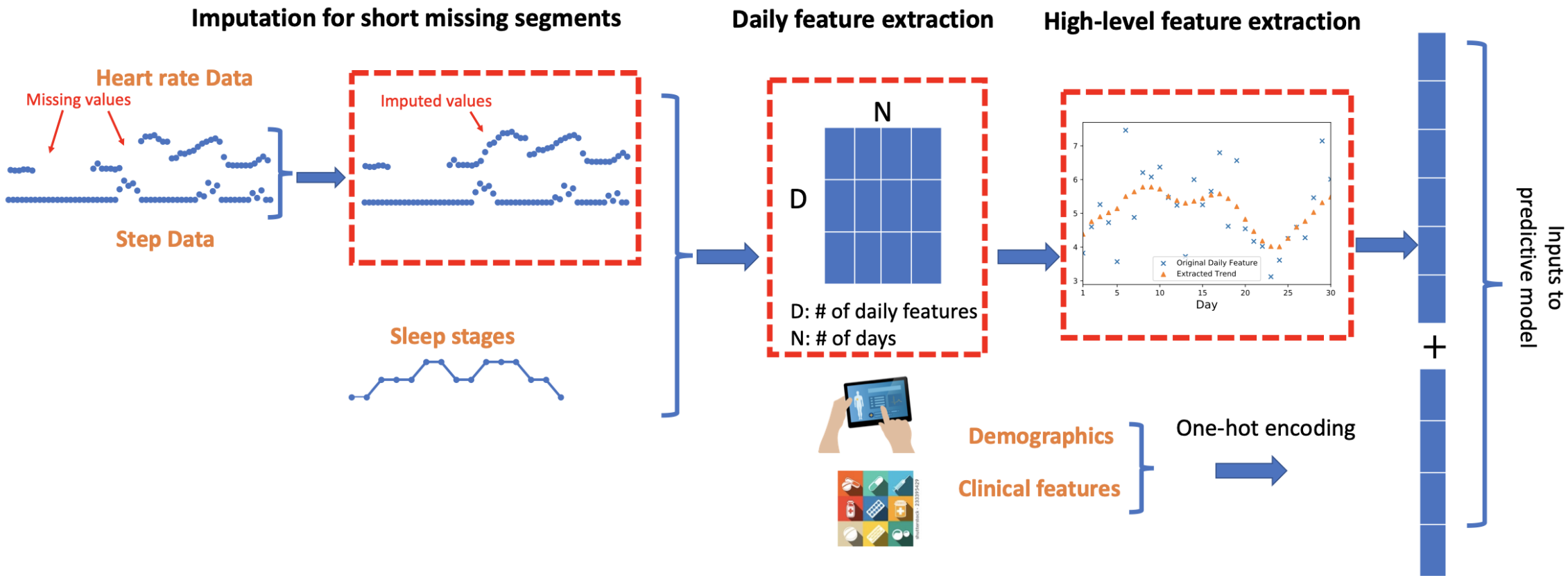
Handling Data Dimensionality

- Deep learning
  - Well-suited for:
    - High-dimensional data
    - Time-series data
  - Requires ”large” sample sizes
    - Most surgical mHealth studies currently include ~50-150 patients
- “Shallow” learning
  - Greater emphasis on feature selection and feature engineering

Data processing

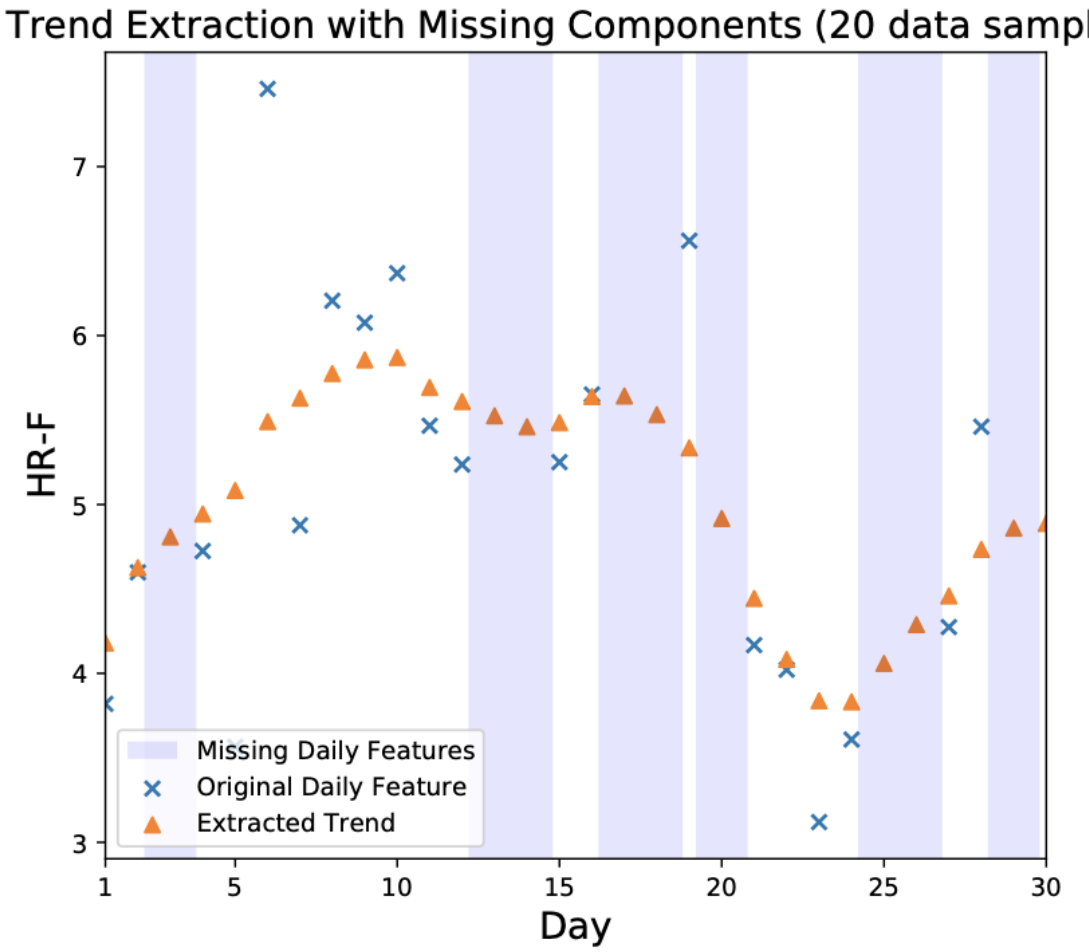
- EMA
  - Distribution, percentiles
  - Variability
  - Autocorrelation
  - Contemporaneous and longitudinal associations (e.g., pain and mood)
- Fitbit
  - Distribution, percentiles
  - Clinically meaningful groupings (e.g., time in mild vs. moderate-vigorous activity)

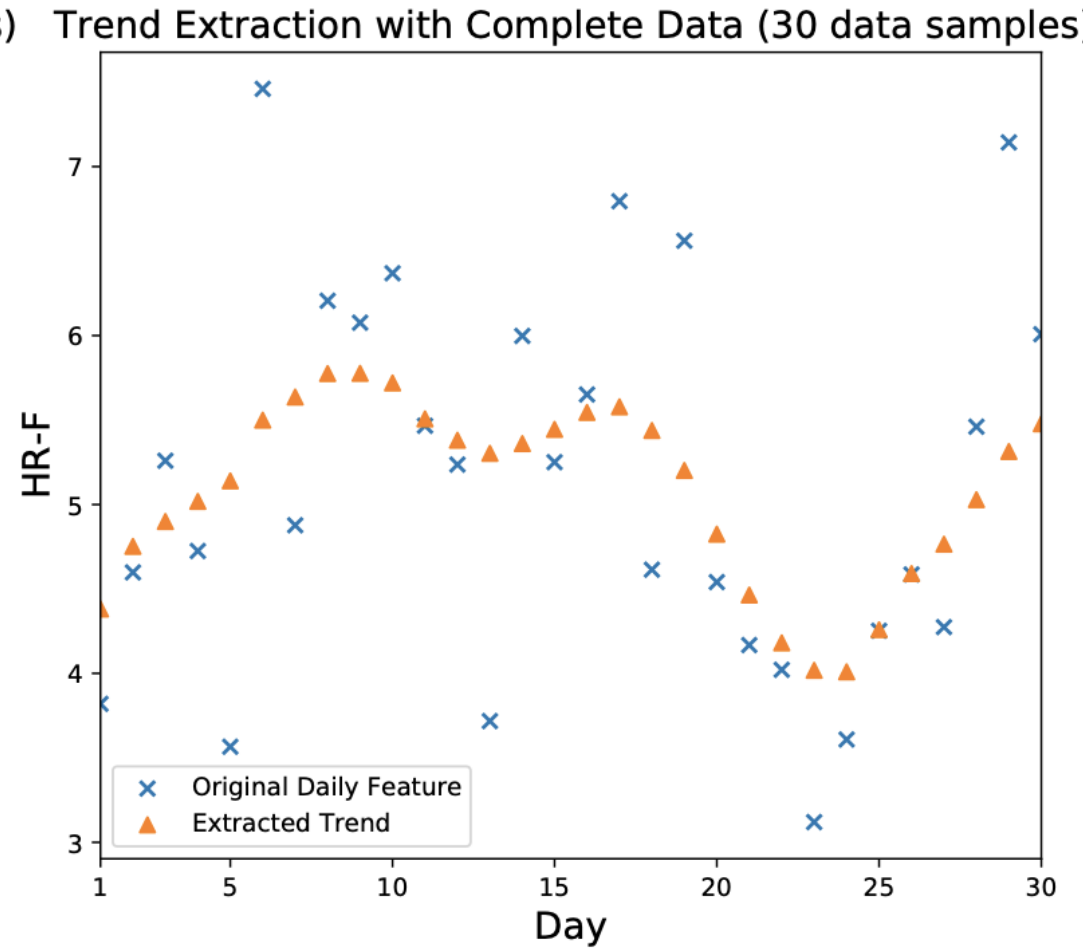






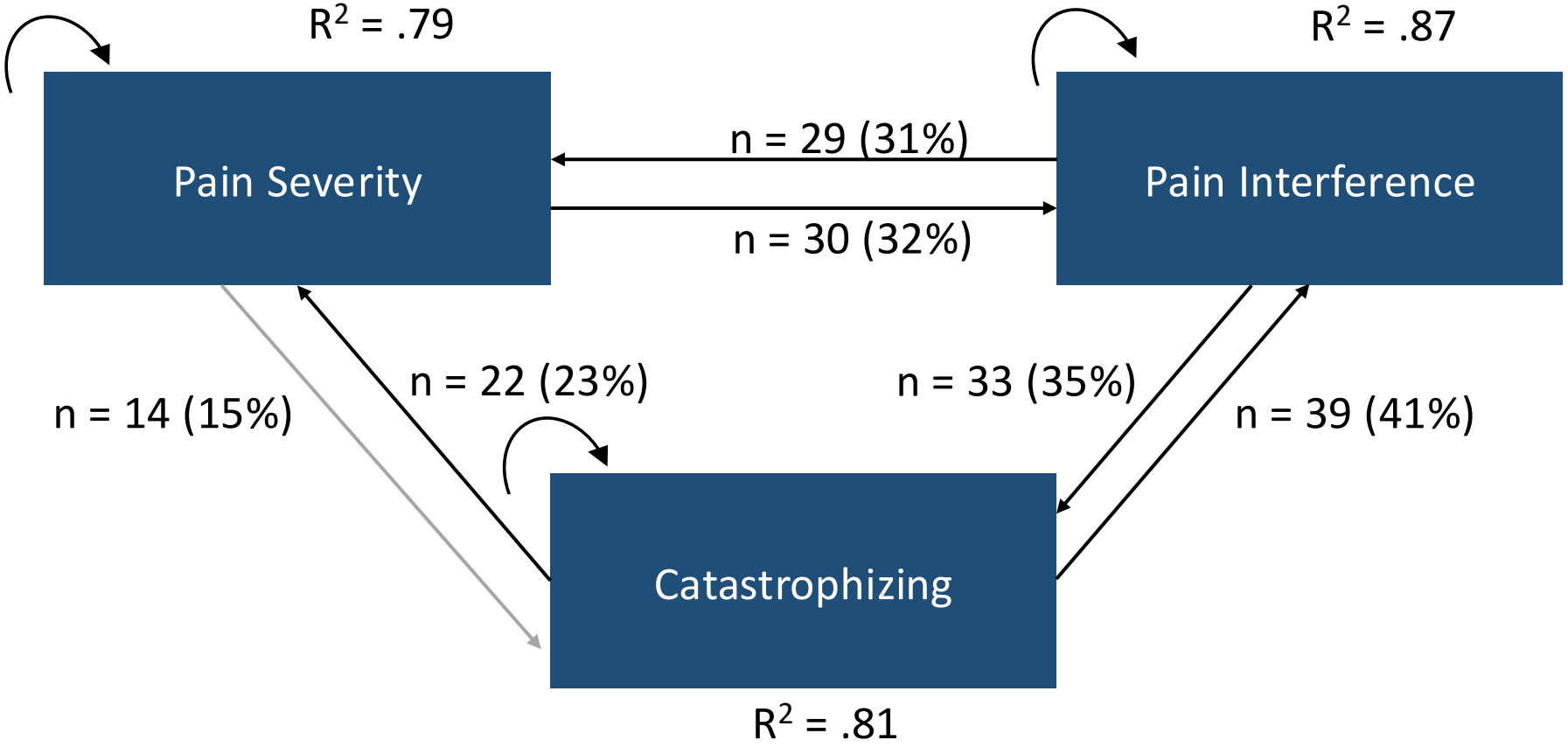
Zhang et al, 2022







Zhang et al, 2022

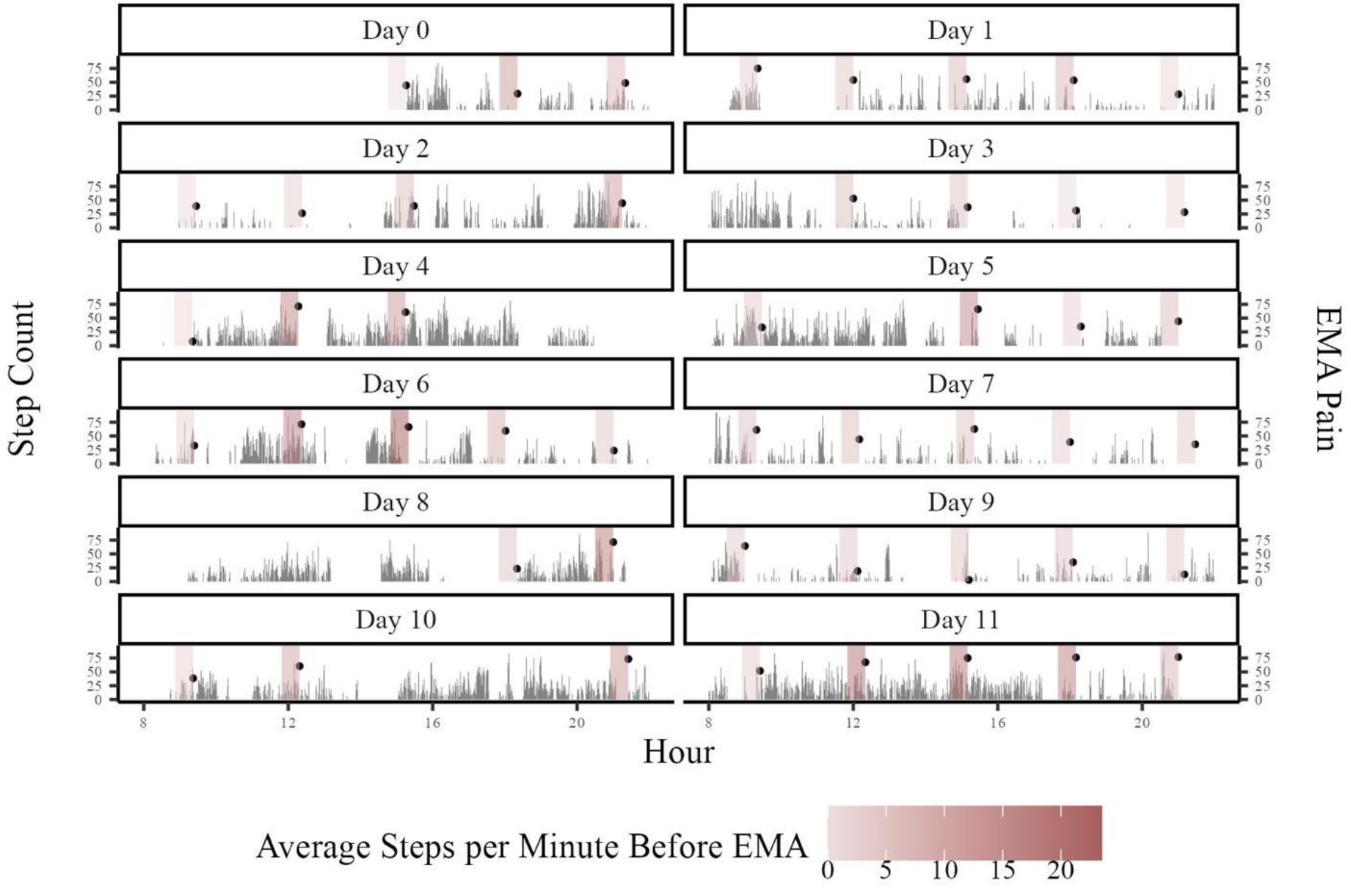

R^2 = .79
R^2 = .87
Pain Severity
Pain Interference
n = 29 (31%)
n = 30 (32%)
n = 14 (15%)
n = 22 (23%)
n = 33 (35%)
n = 39 (41%)
Catastrophizing
R^2 = .81



Data processing

- EMA-Fitbit
  - Longitudinal associations (e.g., pain and activity)





Day 0
Day 1
Day 2
Day 3
Day 4
Day 5
Day 6
Day 7
Day 8
Day 9
Day 10
Day 11
Step Count
EMA Pain
0
25
50
75
8
12
16
20
Hour
Average Steps per Minute Before EMA
0
5
10
15
20



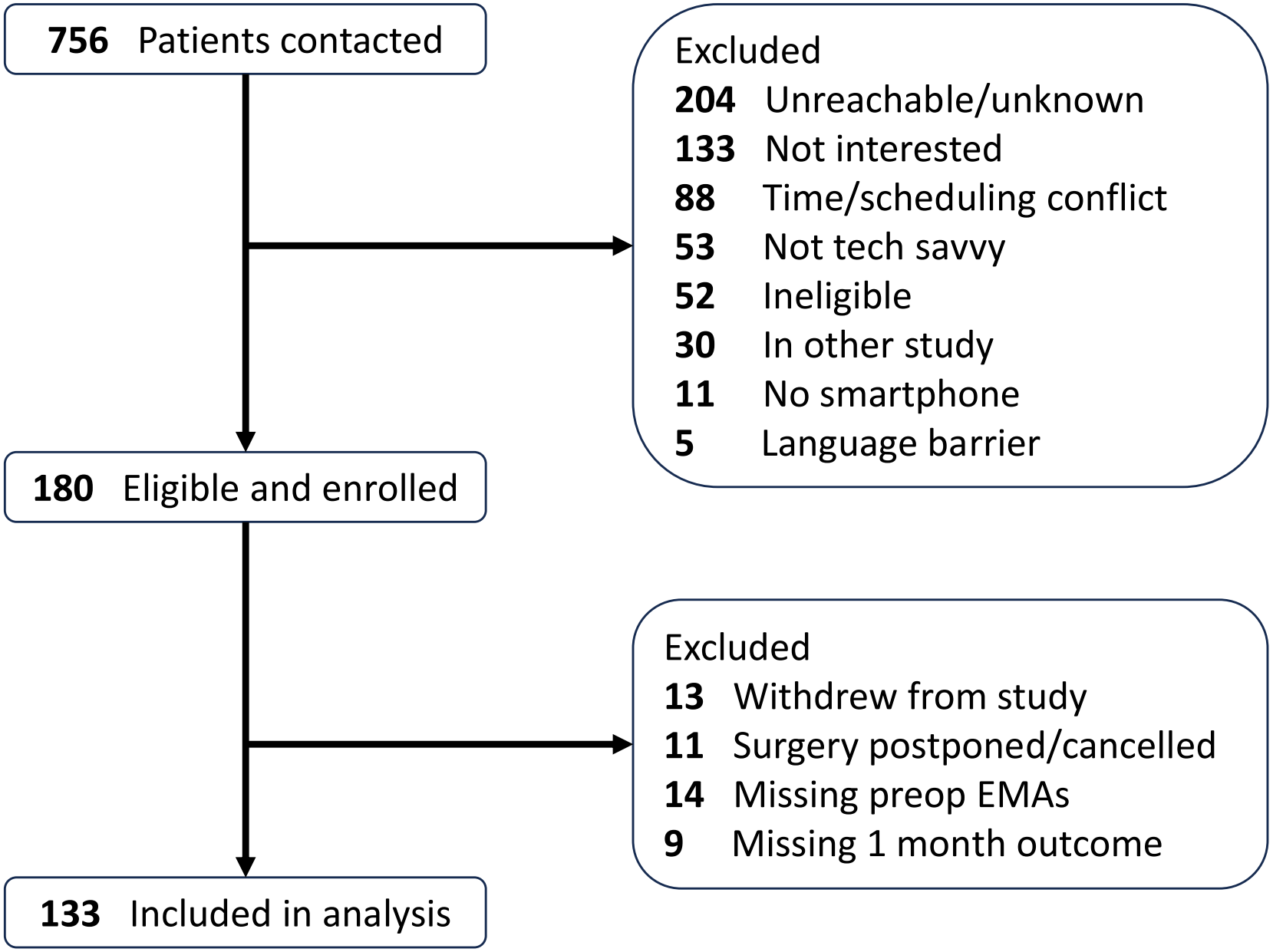
Outcomes

- PROMIS Pain Interference
  - Clinically meaningful improvement
- Quality of Recovery-15 scale
  - Good/excellent recovery

Prediction Modeling (Early + 1 Year Recovery)

- KNN multiple imputation
- Feature selection
  - Pearson correlations
  - P-values < 0.20
  - SelectKBest (grid search)
- Prediction modeling
  - Logistic and penalized regression
  - Other ML models: SVM, GBT, RD, KNN





756 Patients contacted
Excluded
204 Unreachable/unknown
133 Not interested
88 Time/scheduling conflict
53 Not tech savvy
52 Ineligible
30 In other study
11 No smartphone
5 Language barrier
180 Eligible and enrolled
Excluded
13 Withdrew from study
11 Surgery postponed/cancelled
14 Missing preop EMAs
9 Missing 1 month outcome
133 Included in analysis

Results

62 years

median age (IQR 53, 68)

56%

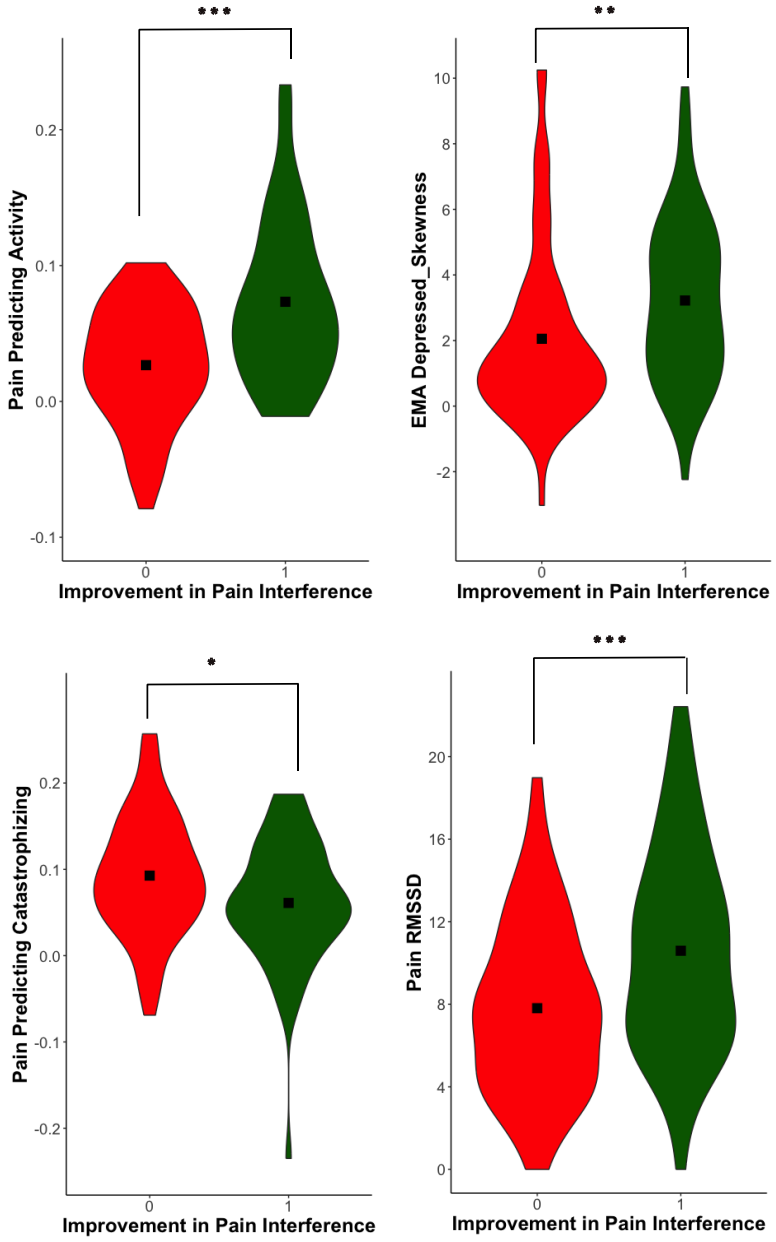
female

93%

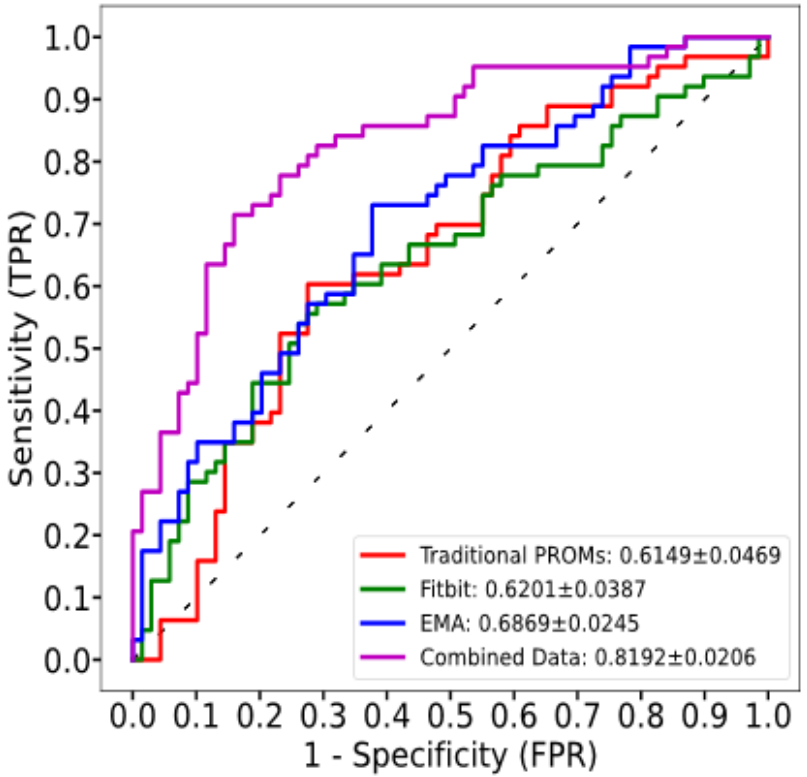
White

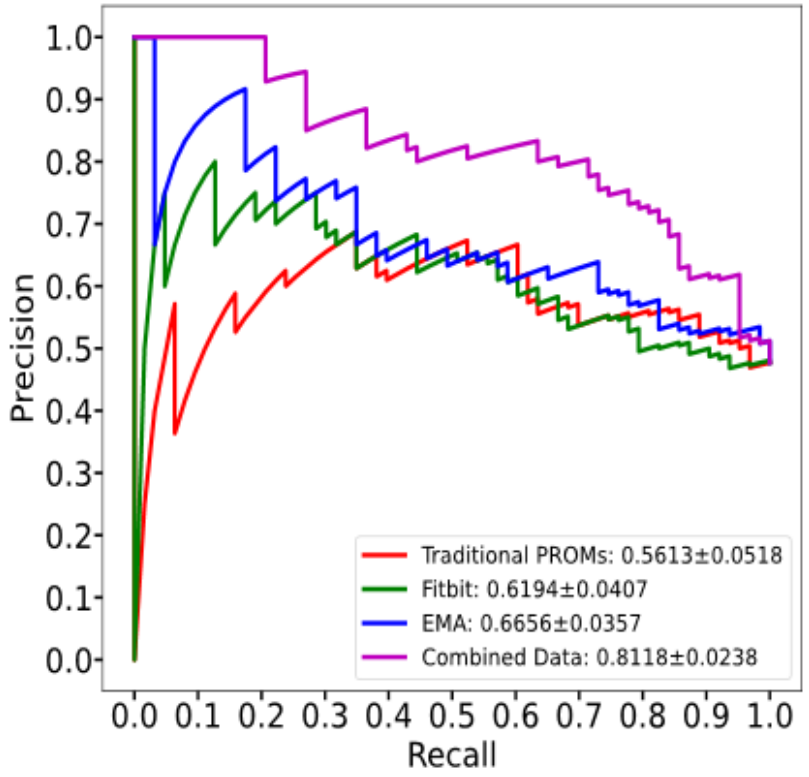
Change in Pain Interference (1 month)

- 63 (48%) achieved clinically meaningful improvement in PROMIS pain interference
- 27 variables with $p < 0.05$ in univariate analysis comparing improved vs. not improved



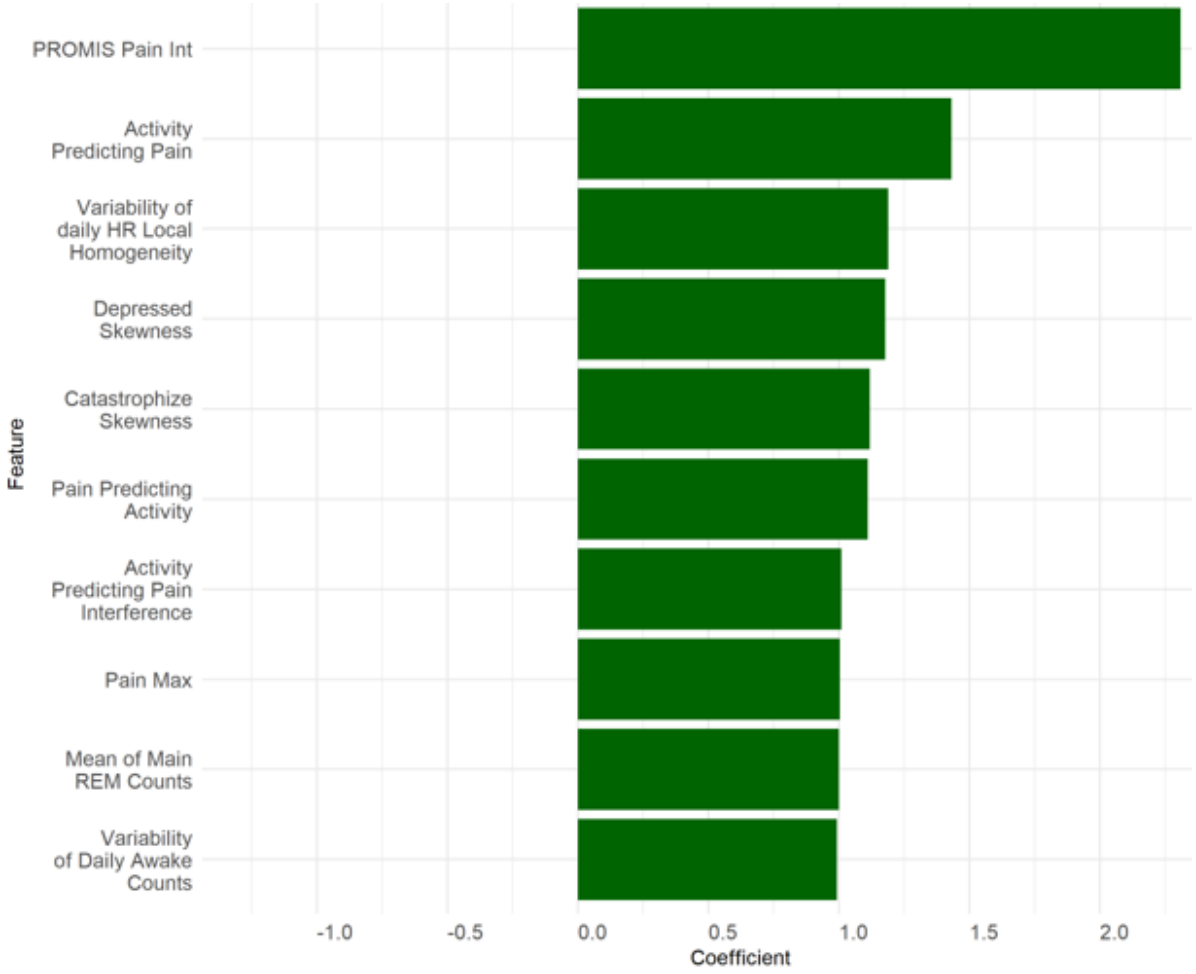
Change in Pain Interference (1 month)





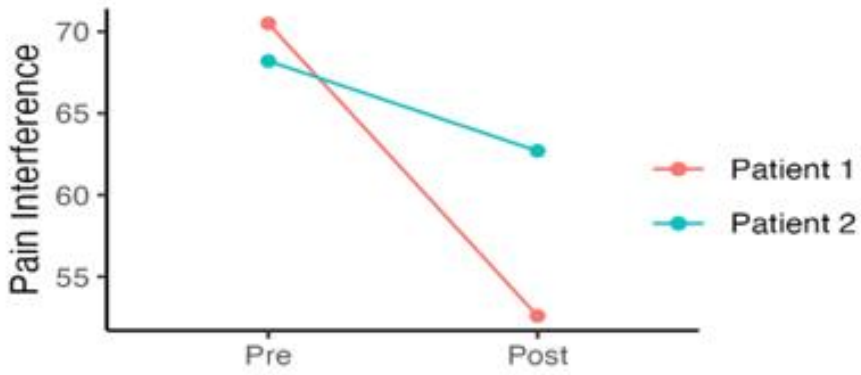


Change in Pain Interference (1 month)

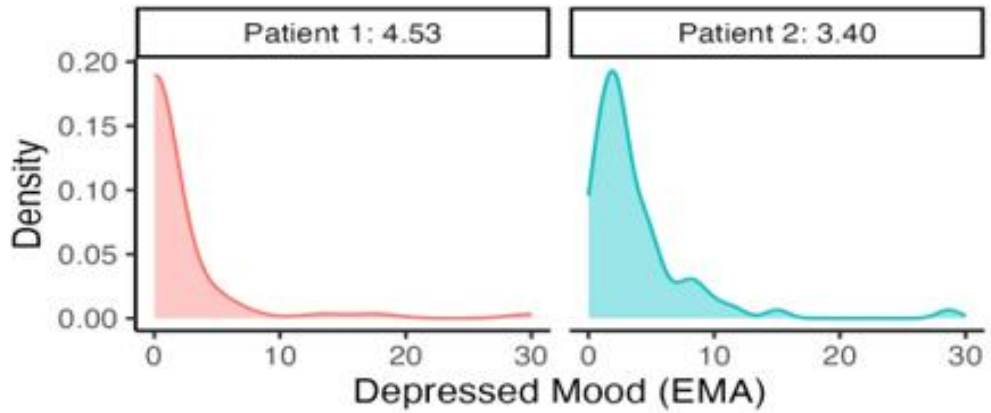




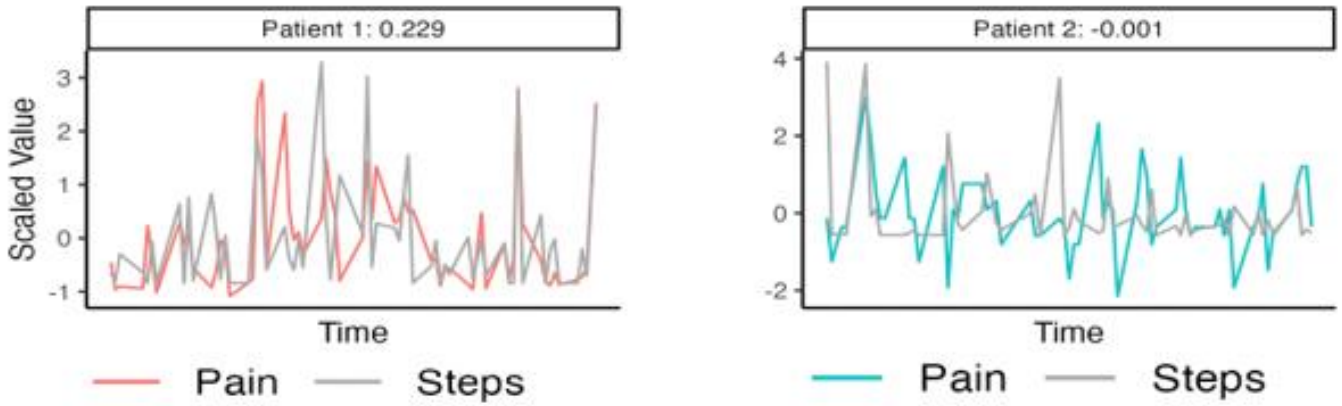
A) Pre-Post Change in PROMIS Pain Interference



B) Skewness of EMA depression items

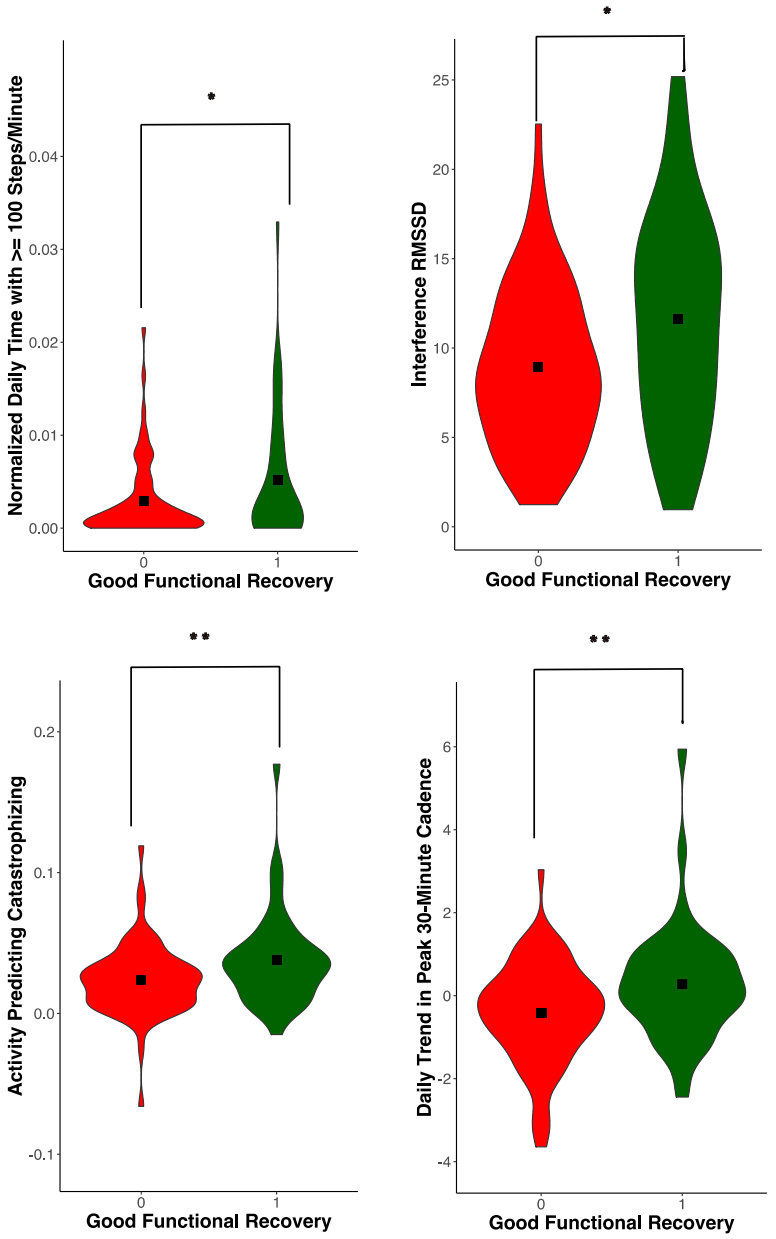


C) Dynamic association between step count and subsequent pain

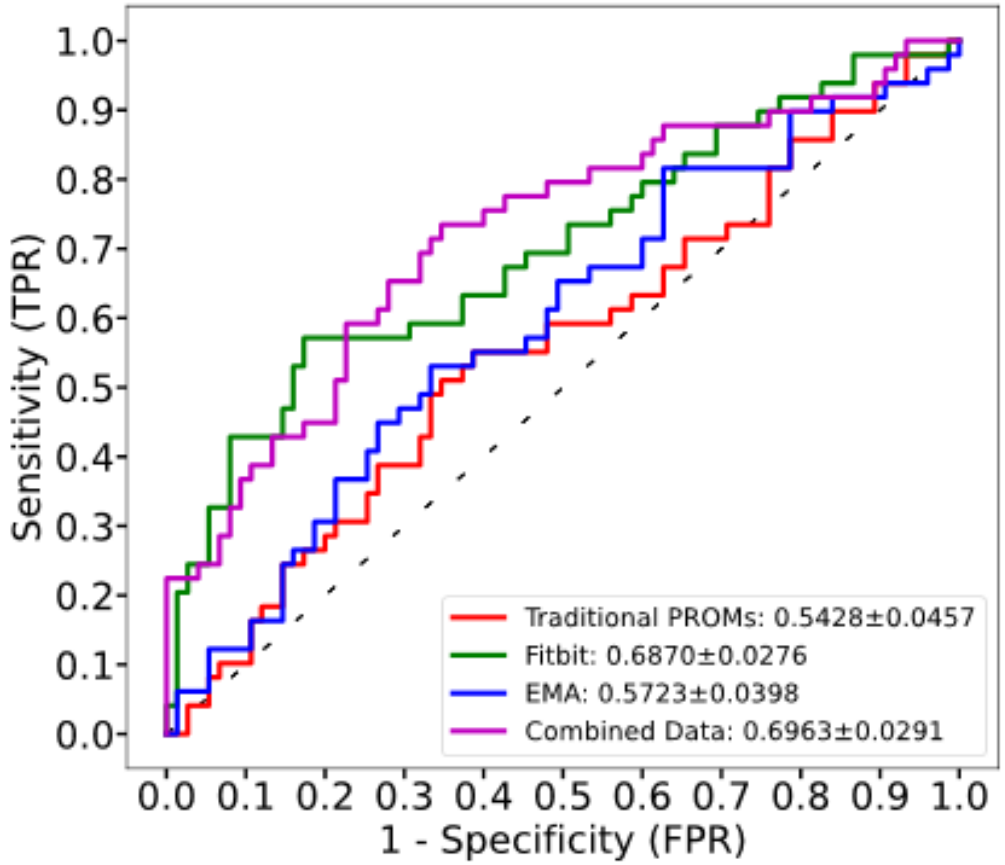


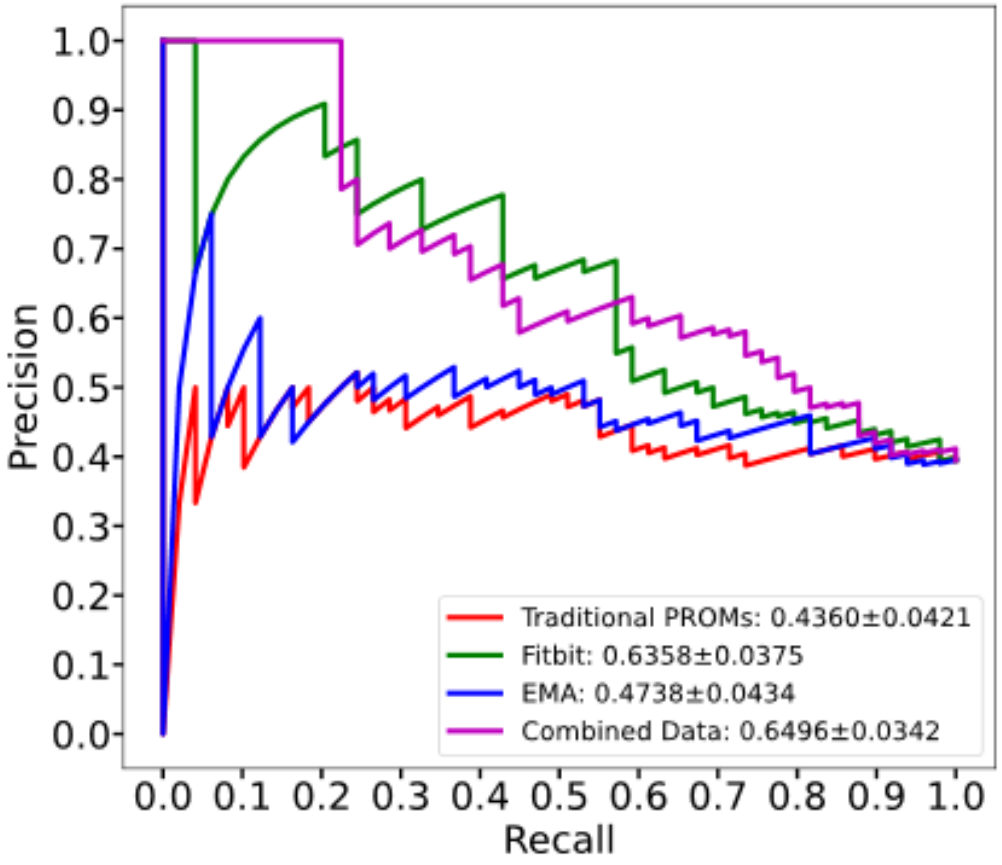
Good/Excellent Recovery (QOR-15)

- 49 patients (40%) had good or excellent recovery
- 17 variables with $p < 0.05$ in univariate analysis



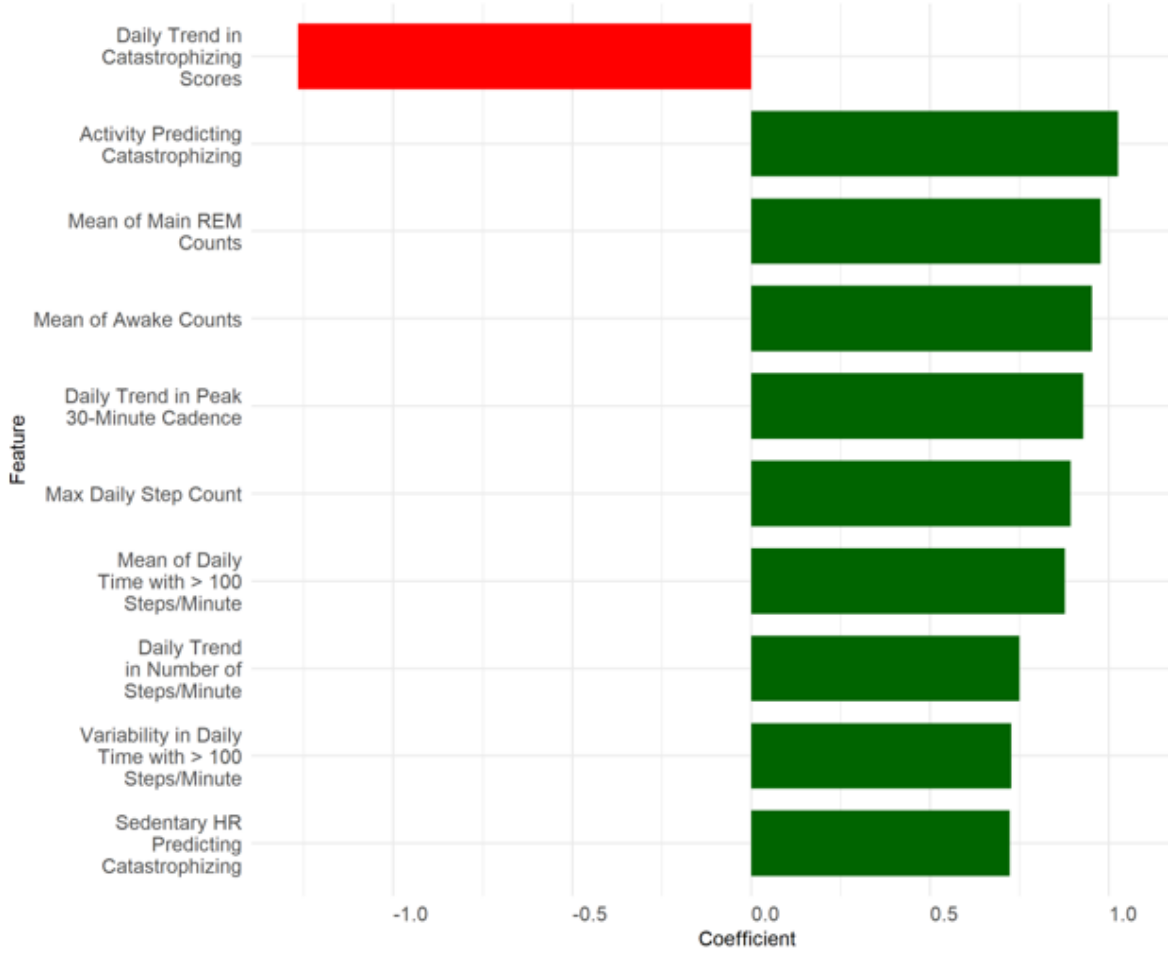
Good/Excellent Recovery (QOR-15)







Good/Excellent Recovery (QOR-15)



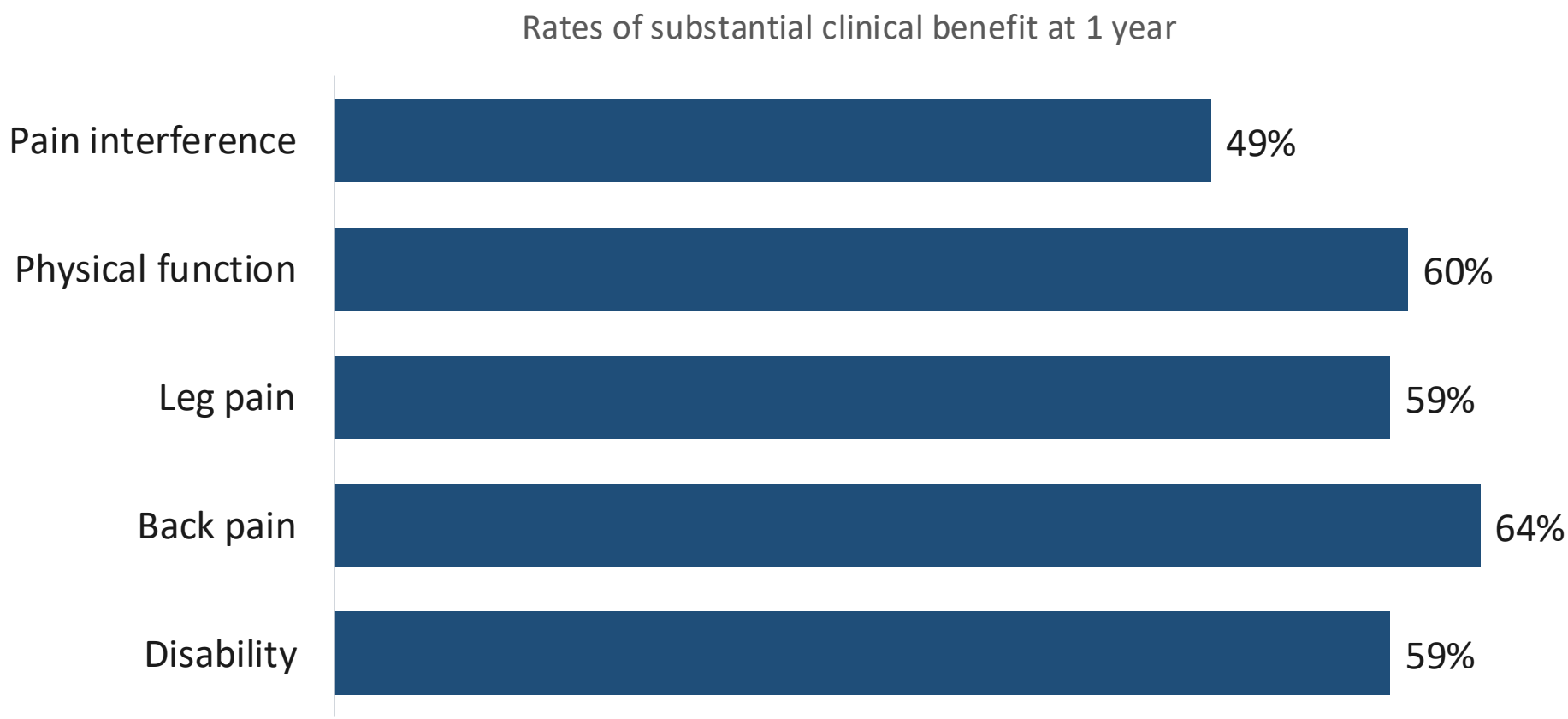


Multitask Learning

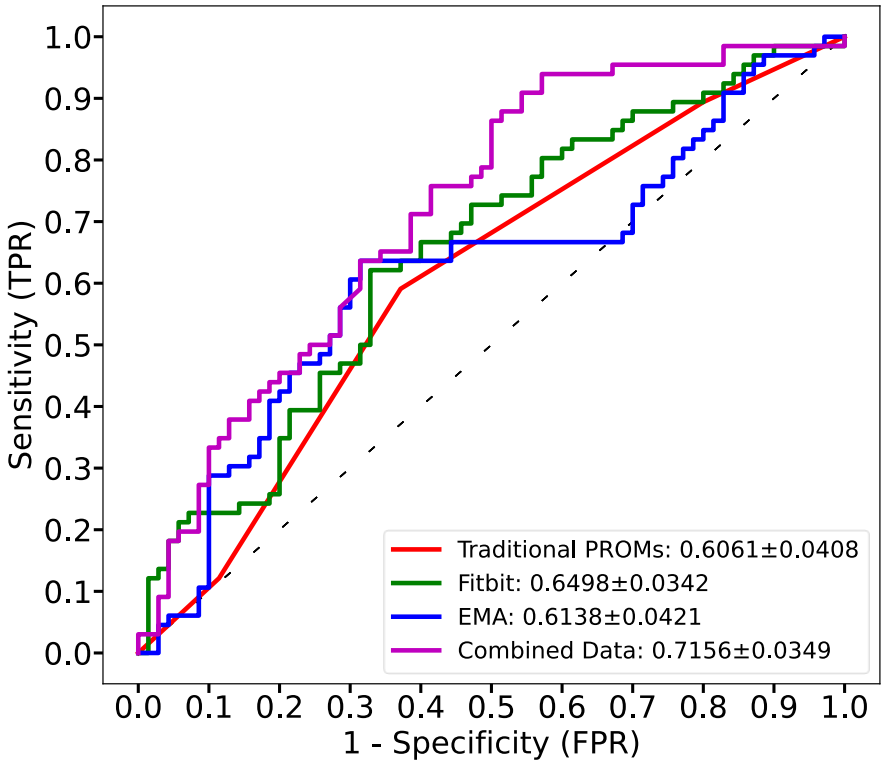
Model		Delta_Pinter (Range: 43.4)		Delta_PhysFun (Range: 38.1)		QOR (Range: 102)	
		MAE	RMSE	MAE	RMSE	MAE	RMSE
Single Task	Lasso	6.6028	8.5903	5.3776	6.9834	14.1834	19.0436
	Ridge	7.0256	9.4326	6.8531	8.8418	14.0231	17.4635
	GBT	5.7585	7.6121	5.3647	6.9851	14.5686	18.2378
	LR	9.7000	13.3693	9.6258	13.2938	23.4329	30.8910
	XGB	5.9730	7.7339	5.2697	6.7729	15.3370	18.5148
	SVM	6.3125	8.3677	5.3199	6.8038	14.2121	19.0337
	RF	6.2785	8.3329	5.2953	6.7379	14.2216	18.6168
	KNN	7.1342	8.9638	5.6555	6.9808	13.9044	18.9234
Multi Task	MTL-fixed	6.1442	8.0601	5.0702	6.4805	13.2057	17.4546
	Our method	**5.6151**	**7.4366**	**4.9630**	**6.1737**	**12.297**	**15.3717**

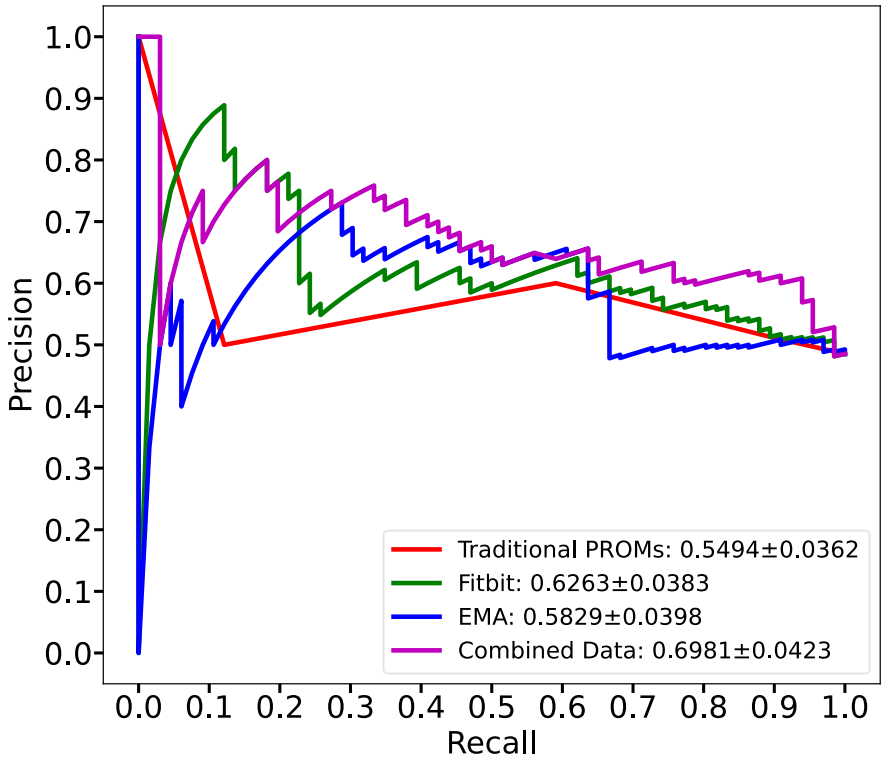


One Year Predictions

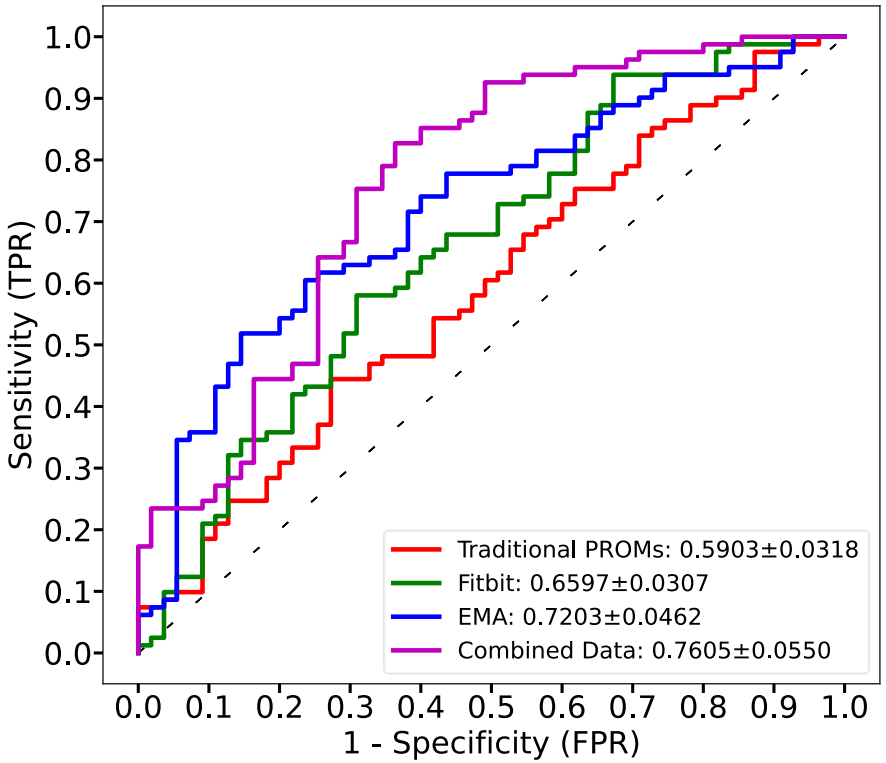


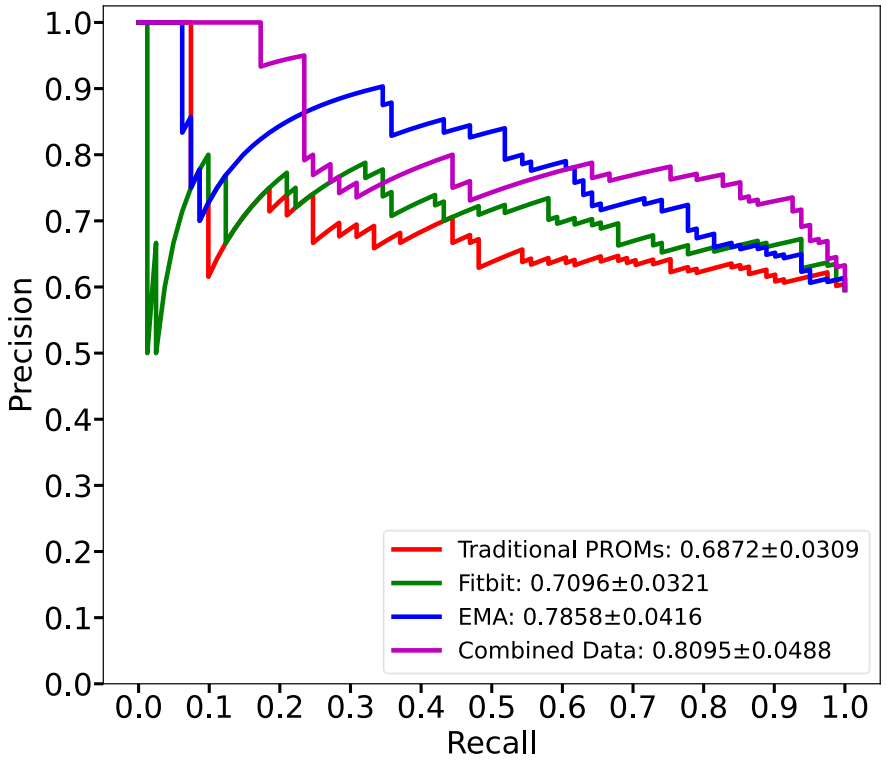
Pain Interference (1 Year)



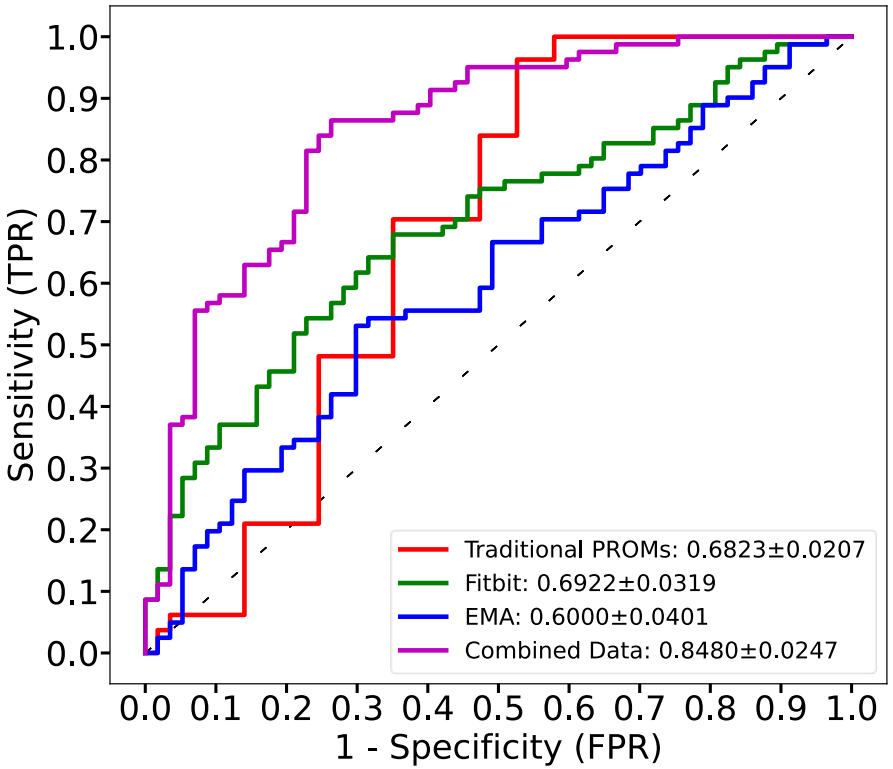


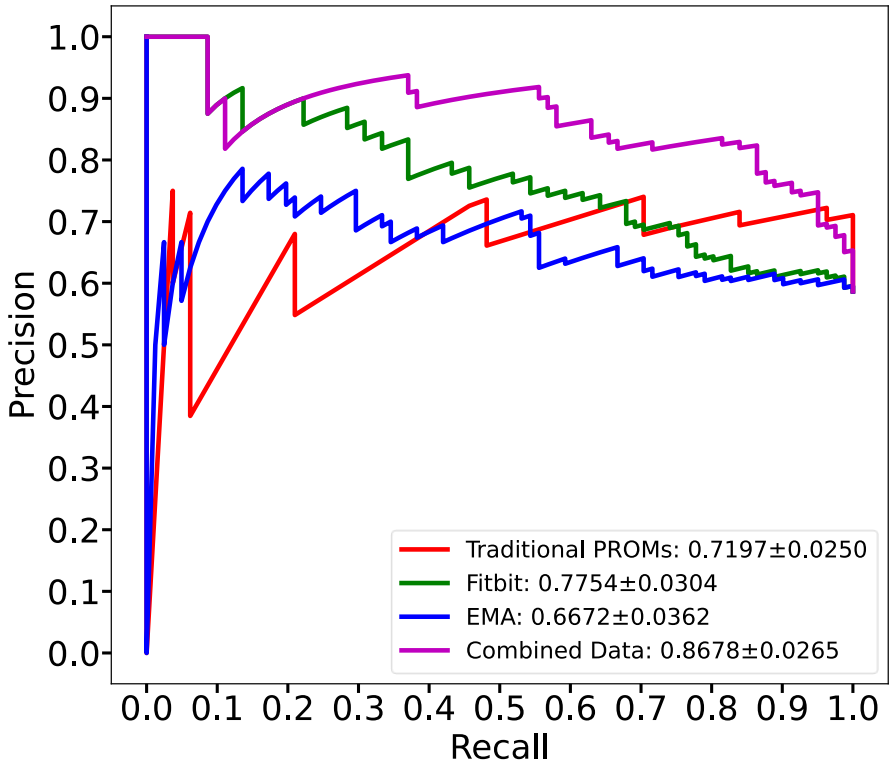
Physical Function (1 Year)



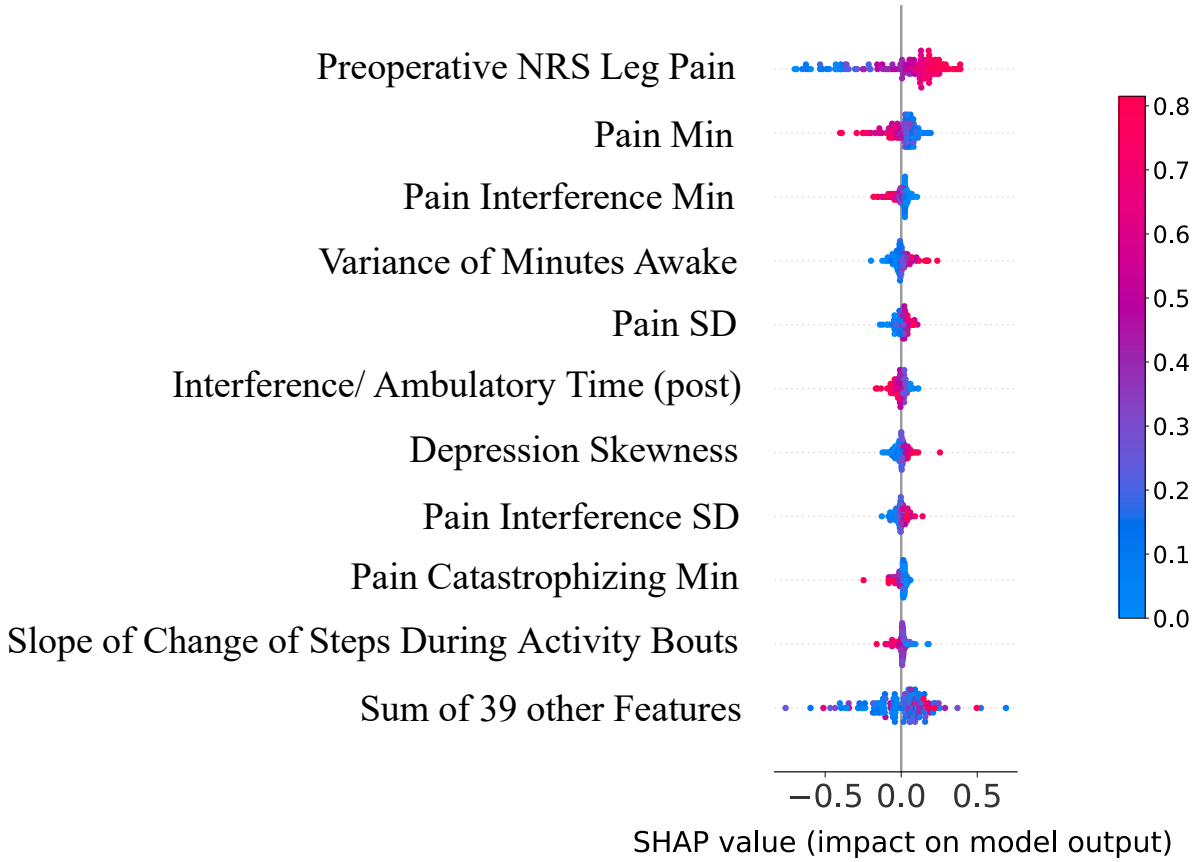


Leg Pain(1 Year)

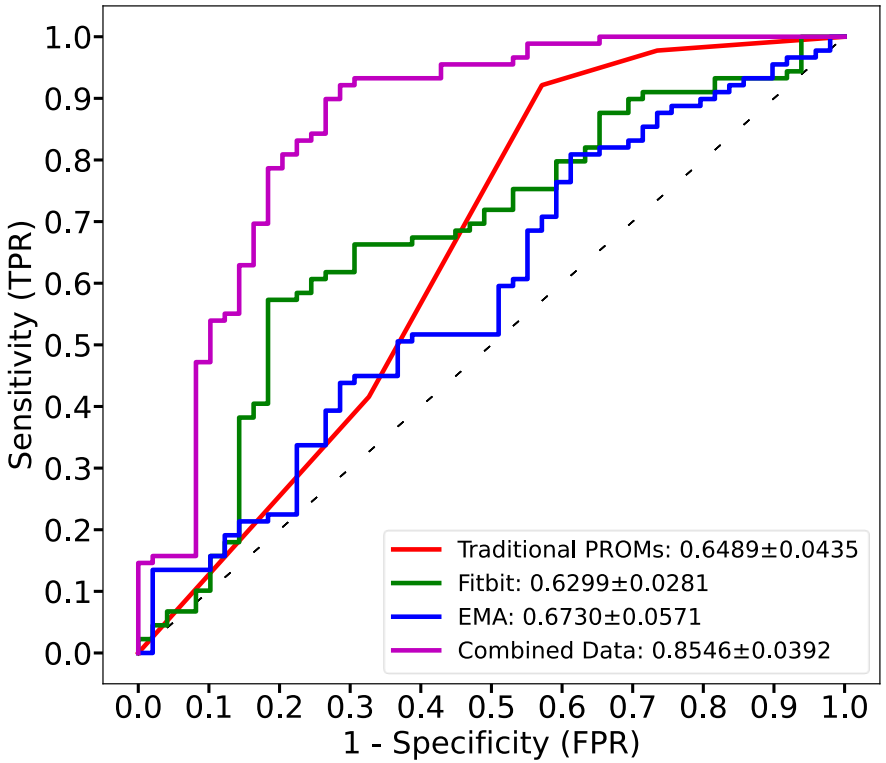


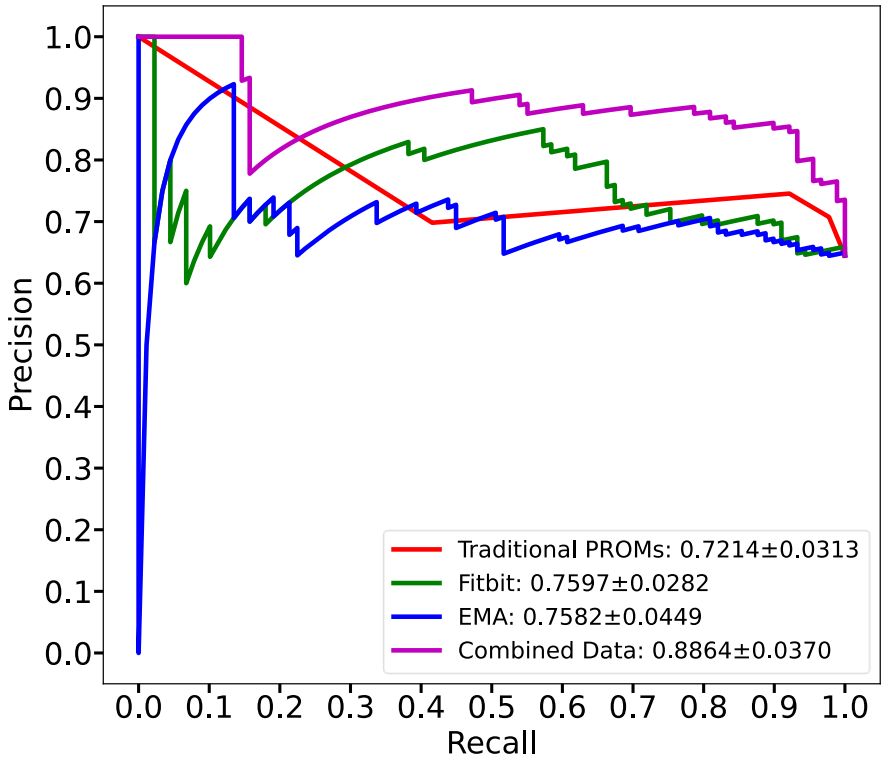


Leg Pain(1 Year)

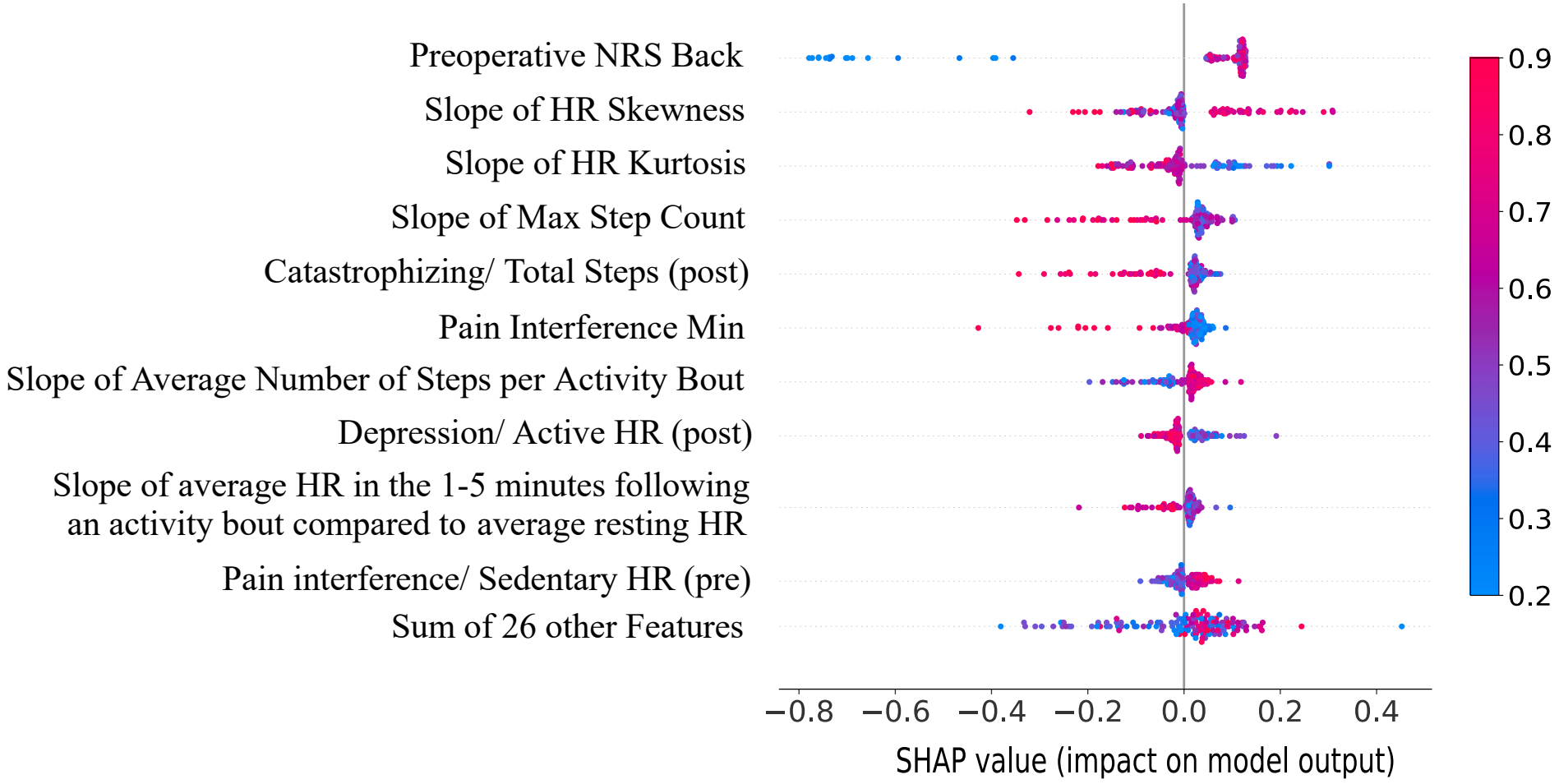


Back Pain(1 Year)

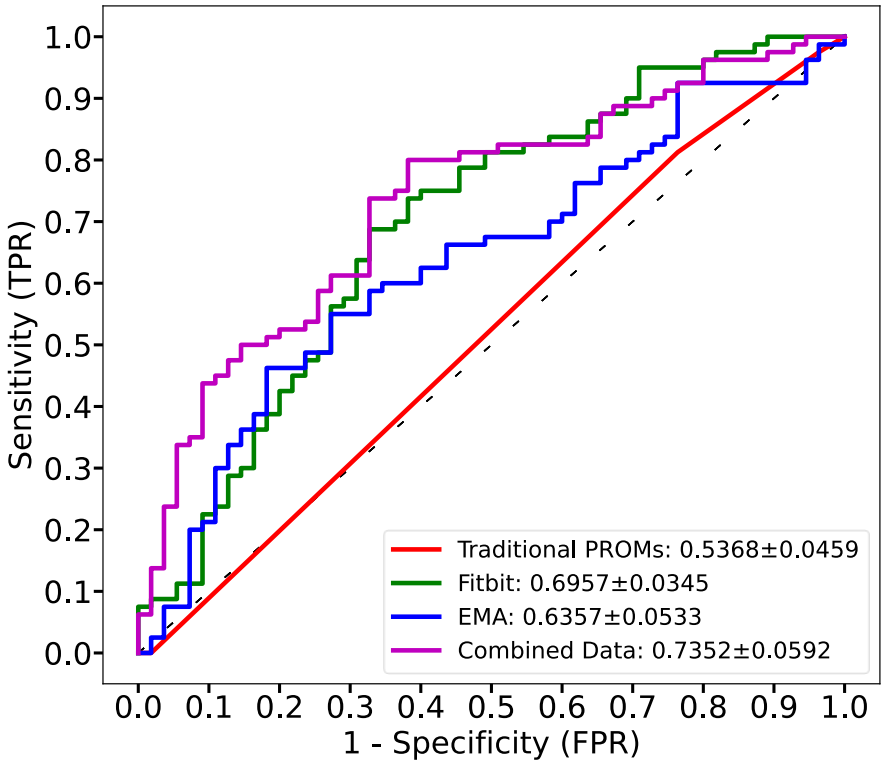


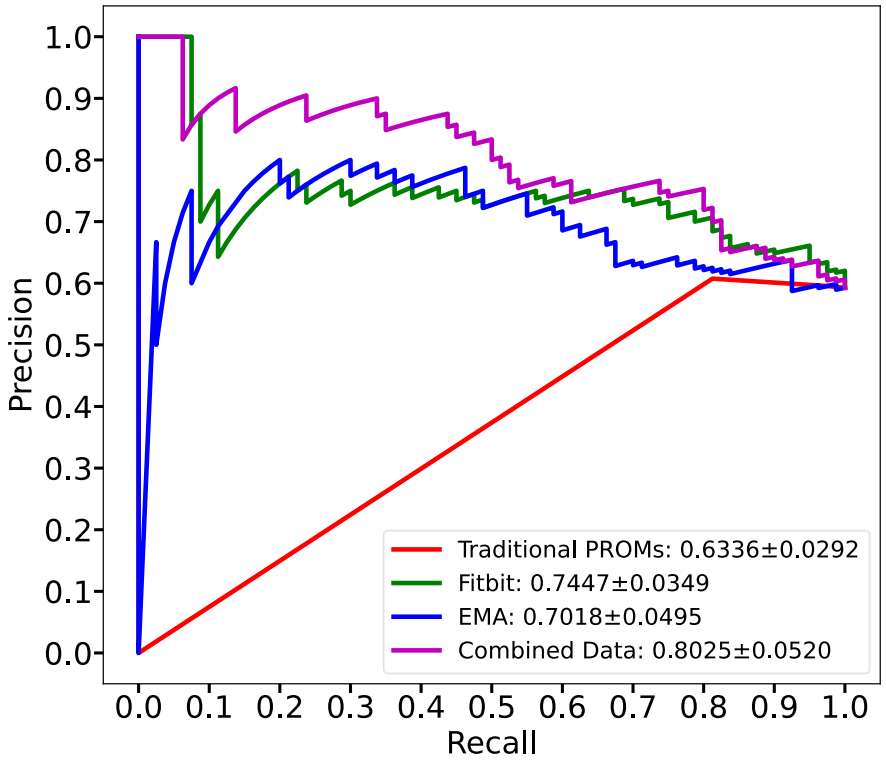


Back Pain(1 Year)

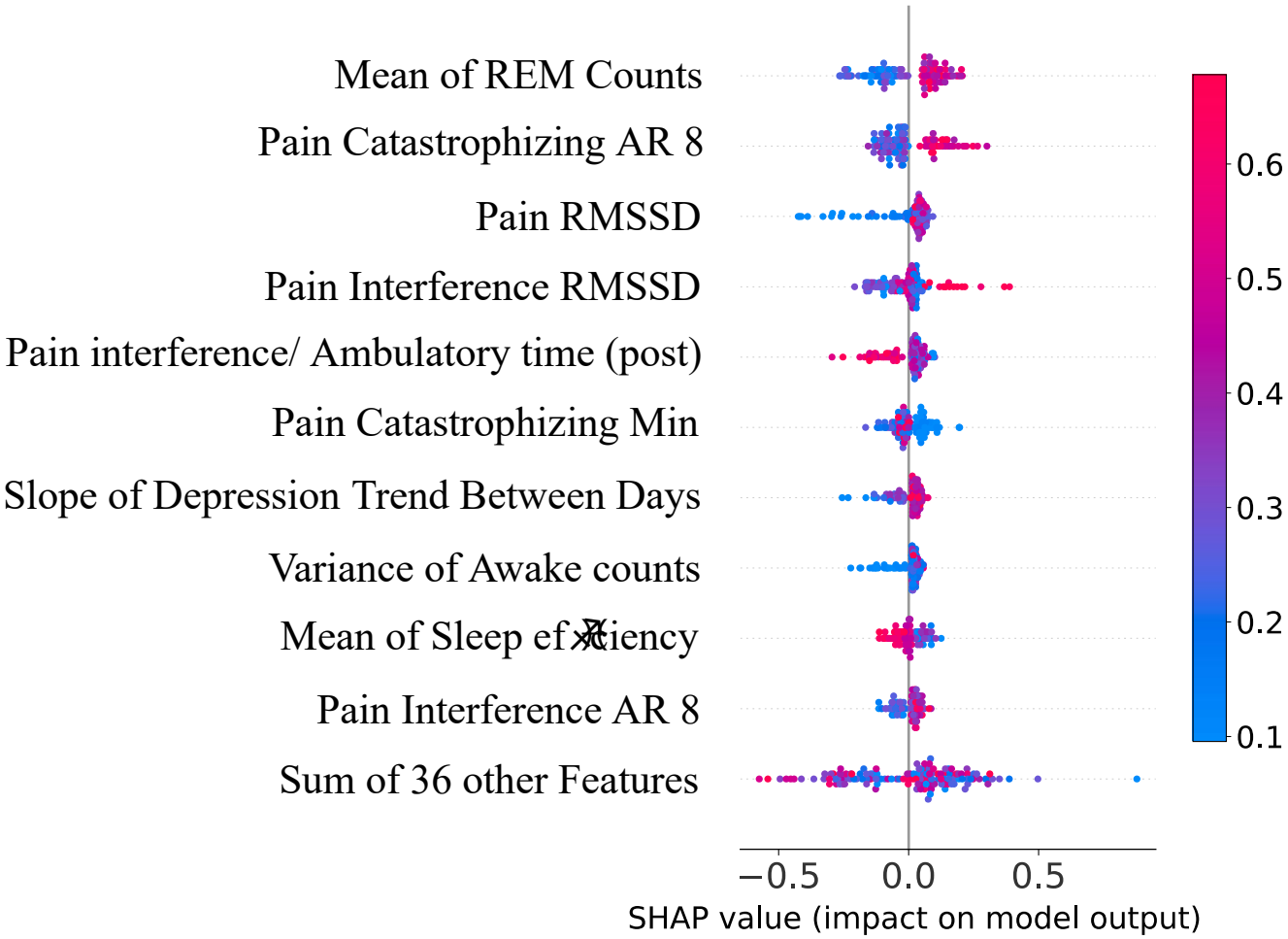


Disability (ODI) (1 Year)





Disability (ODI) (1 Year)



Study Conclusions

- Preoperative mHealth evaluations substantially improve predictions of lumbar surgery outcomes
- Revealed new variables with a potential impact on surgical outcomes
  - Pain variability
  - Pain-activity association
  - Activity intensity
- Identified the value of momentary symptom reporting and particularly passive wearable assessments

Greenberg et al., Neurosurgery 2024; Yakdan et al., Neurosurgery 2026

Monitoring Recovery

Beyond Prediction

- Clustering using mHealth data
- Developing interventions based on modifiable risk factors
- Passive symptom monitoring
- **Monitoring surgical recovery**



Studying Postoperative Recovery

- mHealth Data
  - EMA
  - Fitbit
- Domains
  - Pain
  - Mood
  - Function
  - Physiologic changes (i.e., heart-rate)



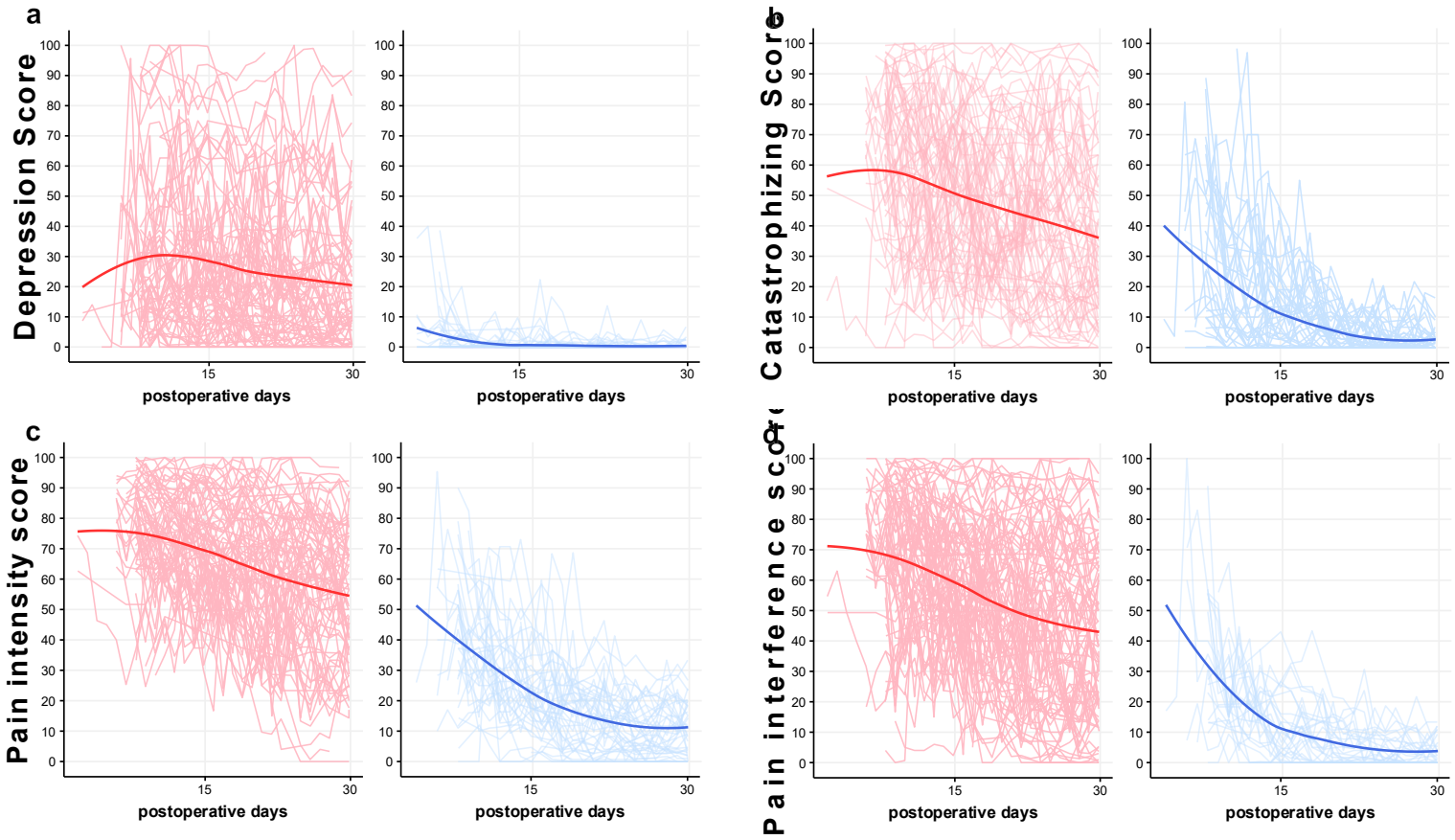
Unsupervised Clustering of *Postoperative* Recovery

- Methods
  - Functional PCA for clustering
- Goal: identify patients with similar patterns of recovery
  - Preoperative counseling
  - Comparing recovery across domains
  - Identifying risk factors for poor recovery

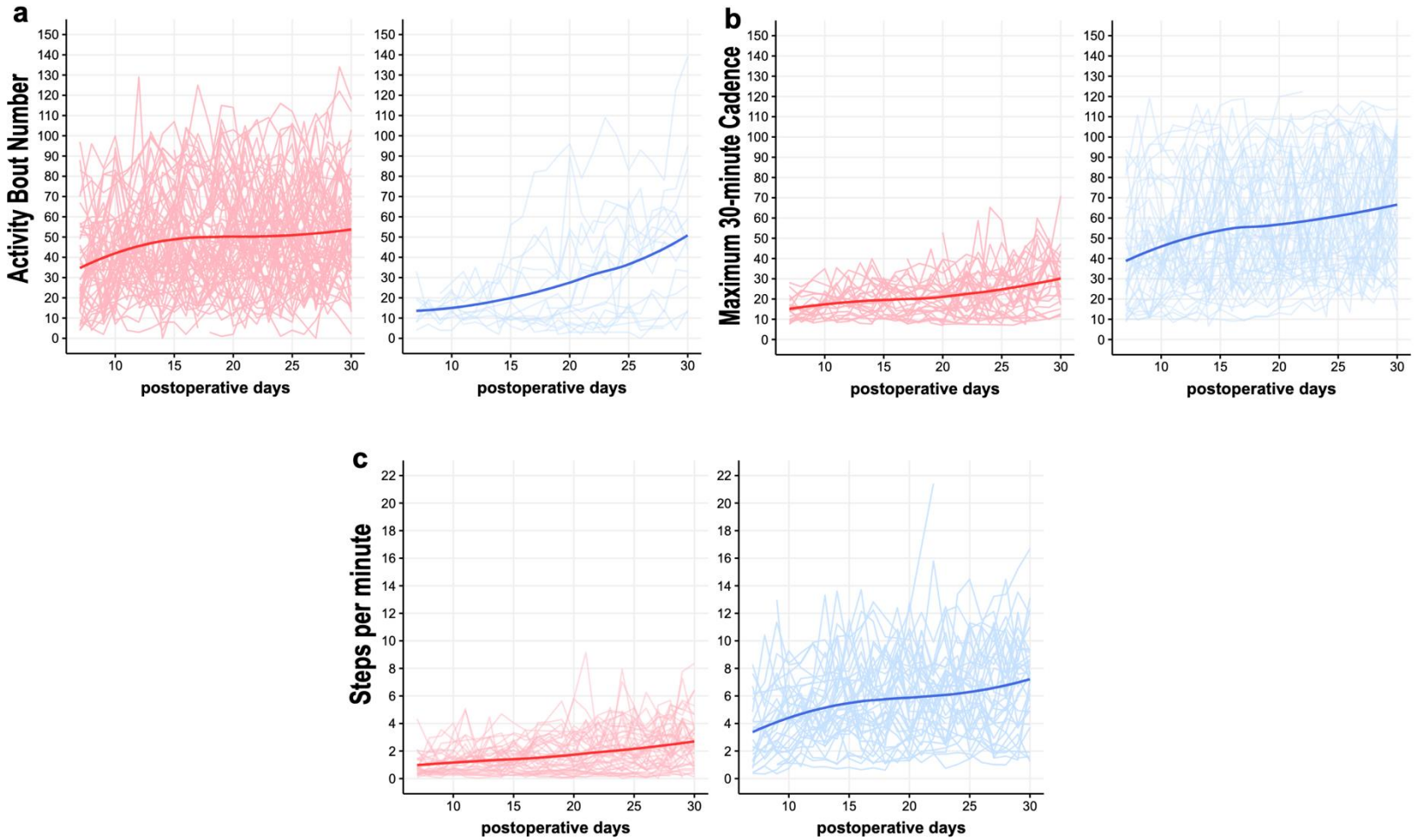


Yakdan et al., Spine J 2025

EMA Data


a
Depression Score
Catastrophizing Score
c
Pain intensity score
Pain interference score
postoperative days
postoperative days
postoperative days
postoperative days

Fitbit Data


a
Activity Bout Number
postoperative days
postoperative days
b
Maximum 30-minute Cadence
postoperative days
postoperative days
c
Steps per minute
postoperative days
postoperative days

Max 30-minute Cadence

Table 3: Comparison of 12-Month PROMs Between Favorable and Less Favorable Recovery Clusters

Recovery Domain	PROM	less-favorable Cluster Mean	favorable Cluster Mean	P-value
Depression	PROMIS Physical Function	43.05	44.27	0.41
	ODI Score	42.14	39.83	0.49
	PROMIS Pain Intensity	54.93	52.46	0.12
Pain Catastrophizing	PROMIS Physical Function	42.32	45.47	**0.03**
	ODI Score	44.67	35.65	**0.006**
	PROMIS Pain Intensity	55.74	51.10	**0.003**
Pain Interference	PROMIS Physical Function	42.65	46.16	**0.03**
	ODI Score	44.18	32.33	**0.001**
	PROMIS Pain Intensity	55.52	49.30	**<0.001**
Pain Intensity	PROMIS Physical Function	42.10	46.19	**0.006**
	ODI Score	46.03	32.24	**<0.001**
	PROMIS Pain Intensity	56.54	49.14	**<0.001**
Activity Bout Number	PROMIS Physical Function	42.71	44.81	0.35
	ODI Score	45.33	39.03	0.20
	PROMIS Pain Intensity	55.16	53.43	0.45
Maximum 30-minute Cadence	PROMIS Physical Function	39.43	47.01	**<0.001**
	ODI Score	48.87	35.63	**<0.001**
	PROMIS Pain Intensity	56.69	52.21	**0.01**
Steps per Minute	PROMIS Physical Function	41.30	48.14	**<0.001**
	ODI Score	44.88	34.49	**0.004**
	PROMIS Pain Intensity	56.21	50.80	**0.001**

Early Recovery Patterns Predict 1-Year Outcomes

129

patients after lumbar or thoracolumbar surgery; daily EMA + Fitbit for 30 days

2

dominant recovery patterns in each domain (functional PCA)

23% vs 7%

complication rate with less vs more favorable pain recovery

Favorable pain intensity and steps-per-minute recovery tracked with greater 12-month improvement in pain, function and disability

Counting favorable recovery domains predicted 12-month improvement better than traditional patient assessments

Yakdan et al., Spine J 2025

Wearable Benchmarks of Early Improvement

62 patients after elective lumbar surgery

50 candidate Fitbit metrics screened against clinically meaningful 1-month change in PROMIS pain interference and physical function

Selected metrics improved R^2 for 6-month PROMs; all but daily steps stayed significant after adjustment

Fitbit metric (change at 1 month)	Cutoff	AUC
Active time	7.6% decrease	0.71
Daily steps	1.55 steps/min decrease	0.70
Time moving ≥40 steps/min	0.7% decrease	0.67
Heart rate stability	15% increase	0.72

Zhang JK et al., J Neurosurg Spine 2025

Using EHR Data to Predict Cervical Myelopathy

CSM Is Diagnosed Late and Measured Poorly

Delayed diagnosis

Early signs (imbalance, clumsy hands) are vague and often missed

Clinic-only scores

mJOA and exam are episodic and rater-dependent

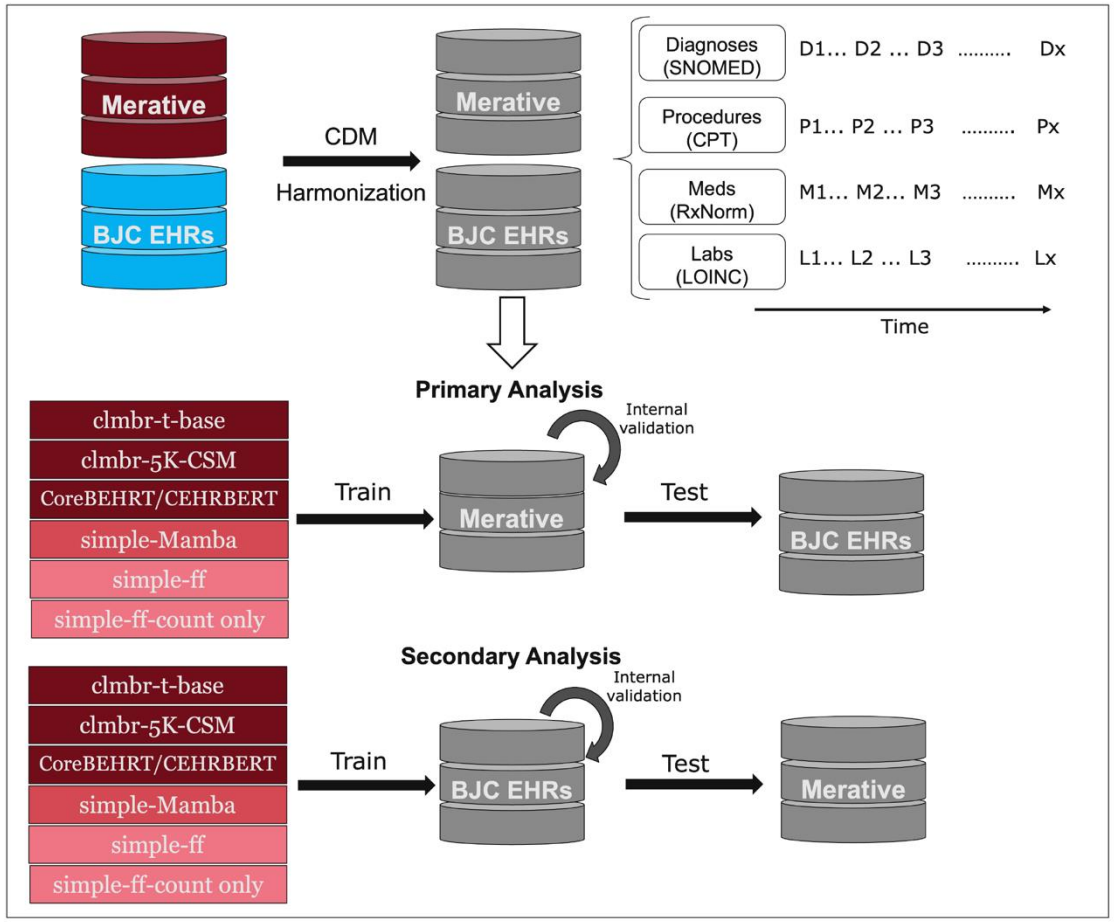
No data between visits

Decline at home goes unmeasured

Interviews with 15 patients and 14 clinicians: both groups wanted objective, longitudinal, at-home monitoring of gait, hand dexterity and daily function, built into the EHR and easy to use

Singh, Gonuguntla et al., medRxiv 2026

Predicting CSM From the EHR: Study Design



Models trained on national claims (Merative MarketScan) and tested in a separate health system (BJC), and vice versa

Yakdan, Warner et al., npj Digit Med 2026 (Fig. 5)



EHR Prediction: Foundation vs Clinically Guided Models

~1.9 million patients (MarketScan 30,772 CSM / 1.45M controls; BJC 13,200 CSM / 419,189 controls). Predicting CSM 6–30 months before diagnosis from diagnosis, procedure and medication codes.

Foundation models

CLMBR-T · 141–145M parameters

- Best on internal validation
- AUPRC 2.6–5.4× baseline prevalence
- Less stable in a new health system

Clinically guided models

Simple FF / Mamba · ≤3.8M parameters, 497 expert-chosen codes

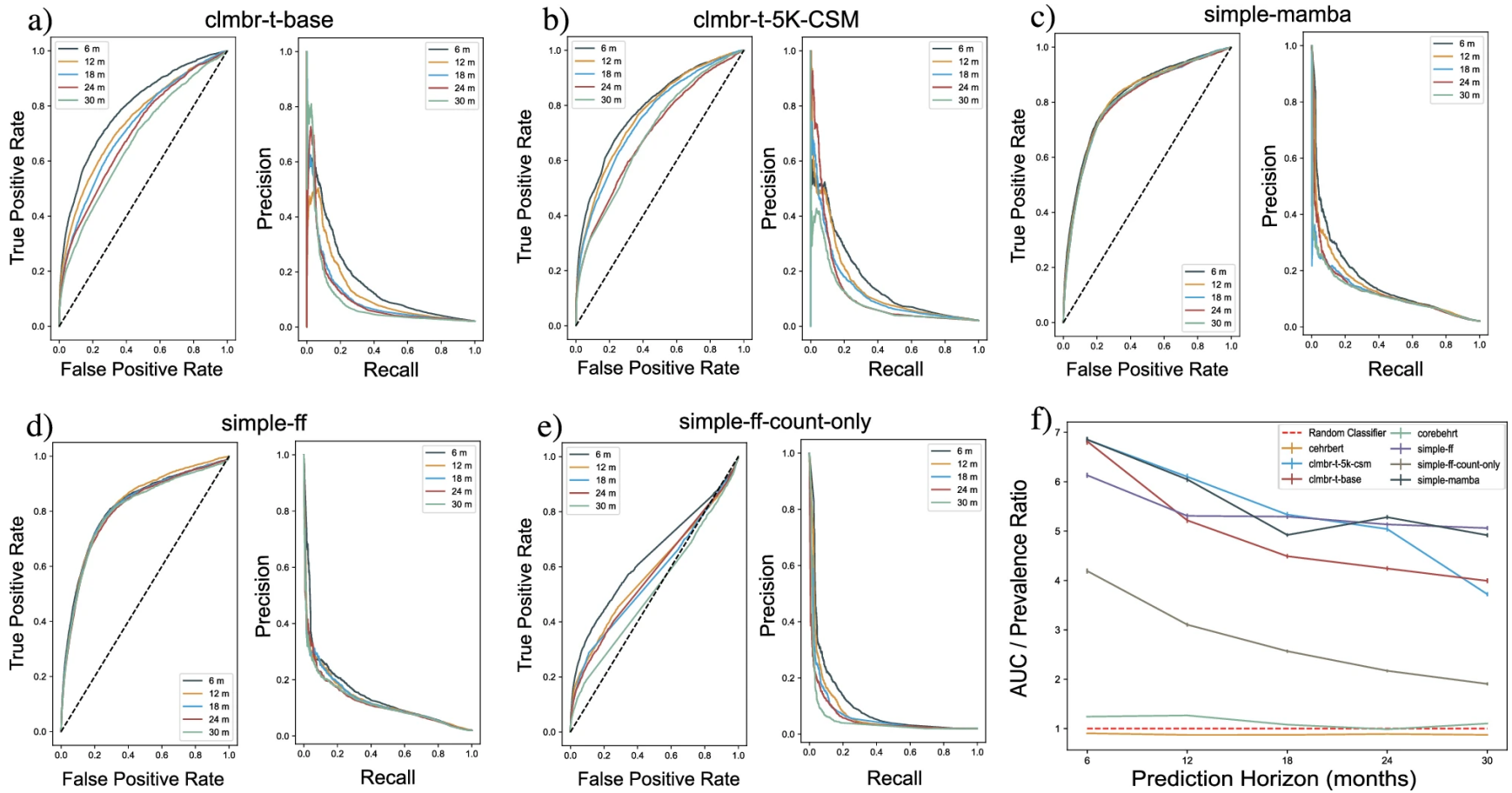
- Comparable or better on external validation
- 97–99% smaller
- More stable when moved between health systems

Takeaway: model complexity trades off against transferability

Yakdan, Warner et al., npj Digit Med 2026



Internal Validation Across Horizons



Merative internal validation: ROC and precision-recall curves at 6–30-month horizons (a–e); f: AUPRC lift ove

Yakdan, Warner et al., npj Digit Med 2026 (Fig. 1)



EHR Prediction: Results and What the Models Learned

2.6–5.4×

AUPRC lift over baseline prevalence on internal validation (best: foundation models)

up to 2.6×

AUPRC lift in a new health system (simple count-based model, 6 months before diagnosis)

6–30 mo

prediction horizon before a CSM diagnosis

What the models learned

- Clinical clues models picked up: unsteady gait, nerve conduction testing, physical therapy
- Performance rose with more healthcare encounters, especially neurosurgery or orthopedic visits
- Reverse validation (BJC → MarketScan): the simple count model outperformed the foundation models

Yakdan, Warner et al., npj Digit Med 2026

Using Smartphone Data to Monitor Cervical Myelopathy

SynapTrack: A Co-Designed Smartphone Platform

Tapping

Finger-tapping speed and rhythm

Pinching

Precision grasp targets and errors

Vibration

Vibratory detection threshold: first smartphone sensory test in CSM

Gait

12-step walk recorded by phone IMU at 100 Hz

Passive walking

Everyday walking speed from phone health data

Built with patients, surgeons and therapists; development approach published as a practical guide for neurosurgeon-led app development



Singh et al., In Preparation, 2026

SynapTrack Hand and Sensory Tasks: Early Validity

251

participants (151 CSM, 100 controls), single center

Measure	Result
Test-retest ICC (tapping, pinching, vibration)	**0.90–0.92**
Vibratory threshold vs mJOA (strongest)	**ρ = –0.54**
Tapping vs Box & Block / 9-Hole Peg	**ρ = 0.60**
AUROC, CSM vs controls	**0.76–0.78**

Reference tests: Box and Block, 9-Hole Peg, tuning fork, mJOA



Yakdan et al., medRxiv 2026

SynapTrack Gait: Phone vs Instrumented Mat

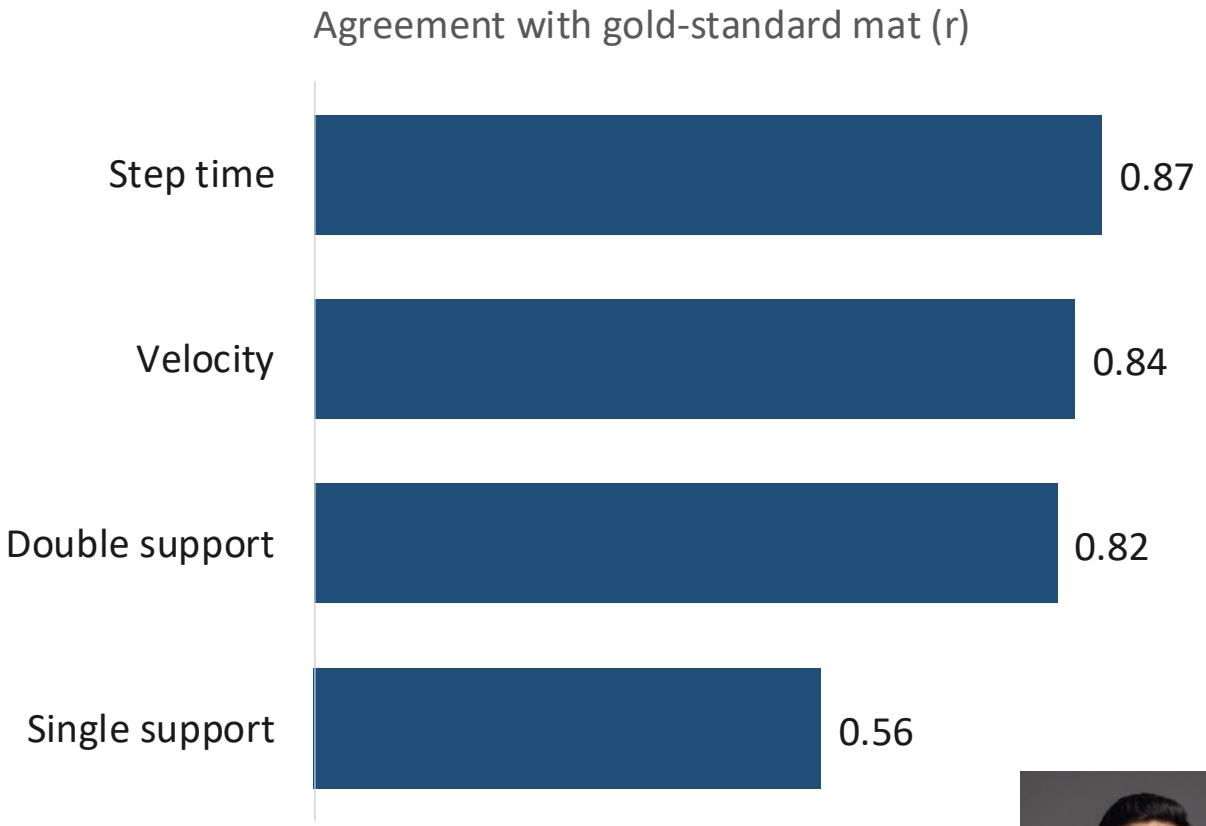
110

participants

55 CSM, 24 peripheral neuropathy, 31 controls; 12-step walk against a ProtoKinetics Zeno mat

Severity
spectral entropy vs mJOA r = −0.54; velocity r = 0.41

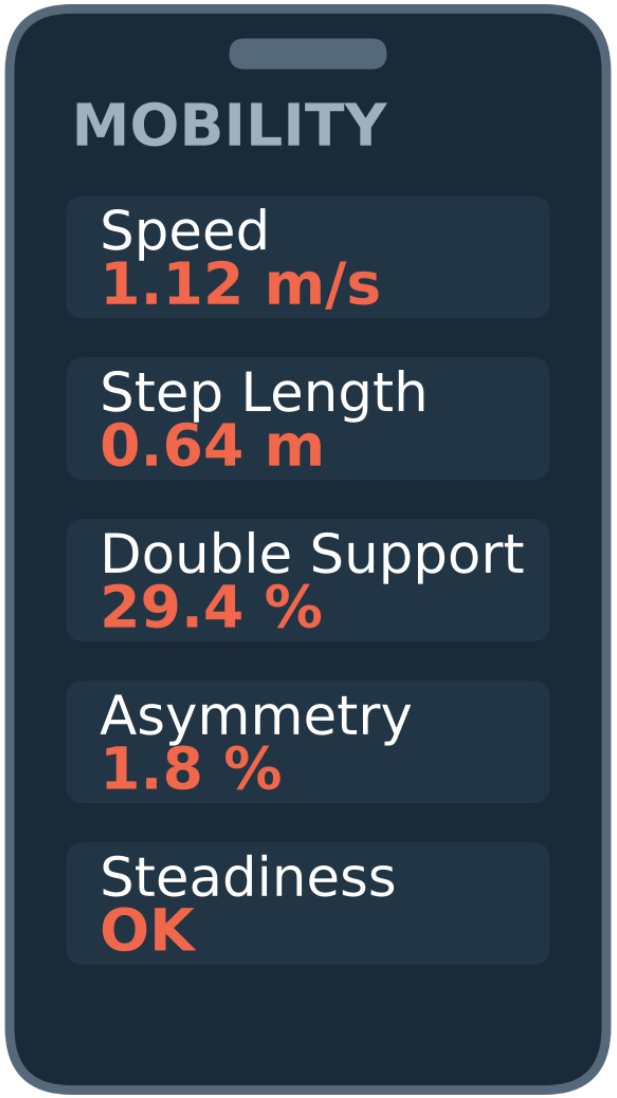
Discrimination
step RMS AUROC 0.80 vs controls





Lewis et al., medRxiv 2026

Using Apple Health Data to Track Myelopathy Severity



Walking speed how fast you walk

Step length how long your steps are

Double support % of each step with both feet

Walking asymmetry how uneven your steps are

Walking steadiness a composite of balance

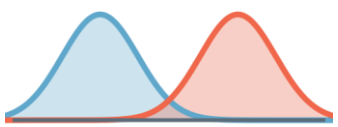
No app to open. No walking test. Nothing to wear.

Could these carry a signal of neurologic disease?



Singh et al., Passive Smartphone Gait Metrics in CSM, preprint 2026

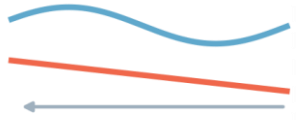
So we asked three questions.



Can gait distinguish myelopathy from no myelopathy?



Does it track how severe the disease is?



Was gait already changing before the diagnosis?

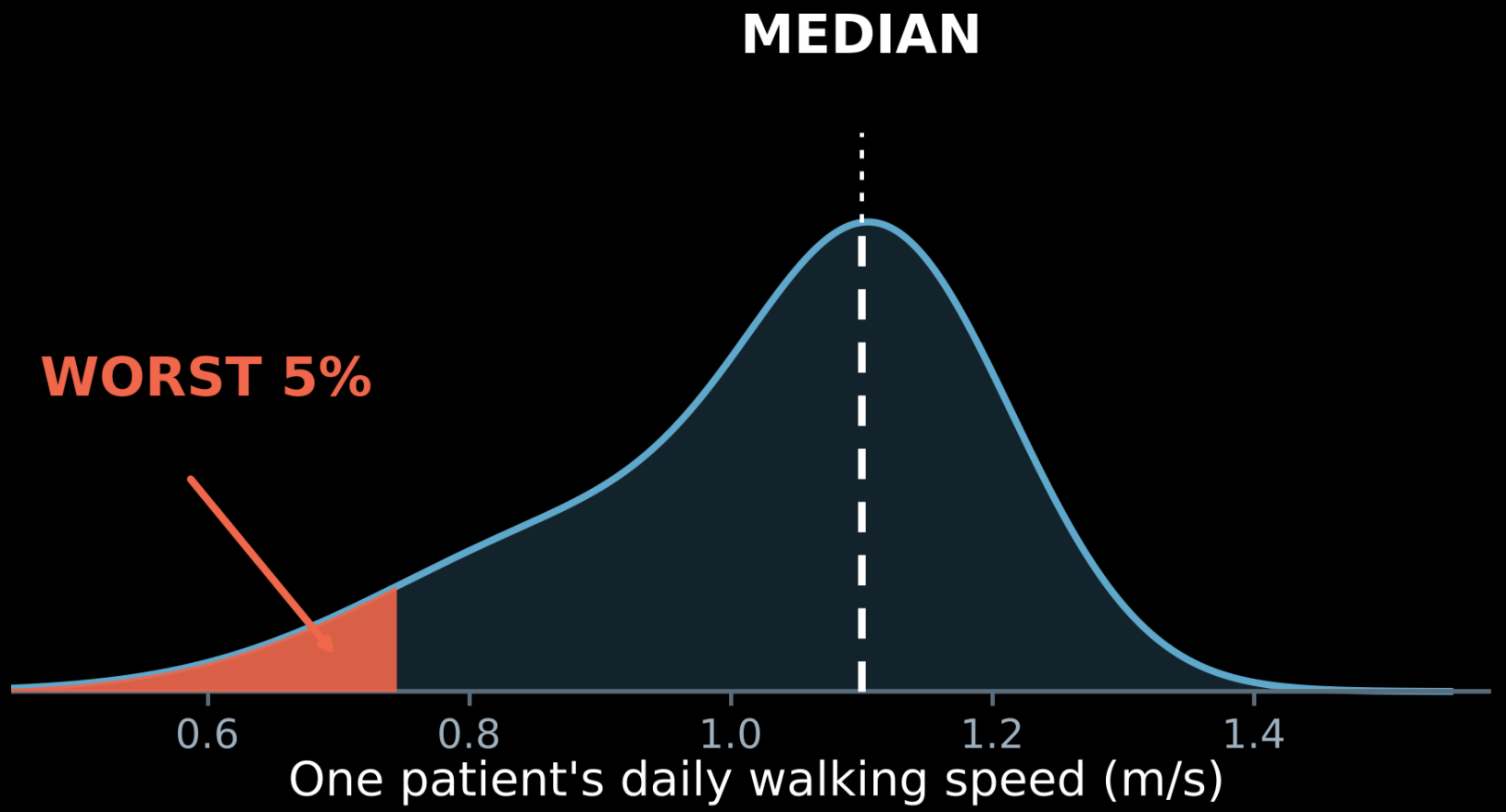


The record was already there.

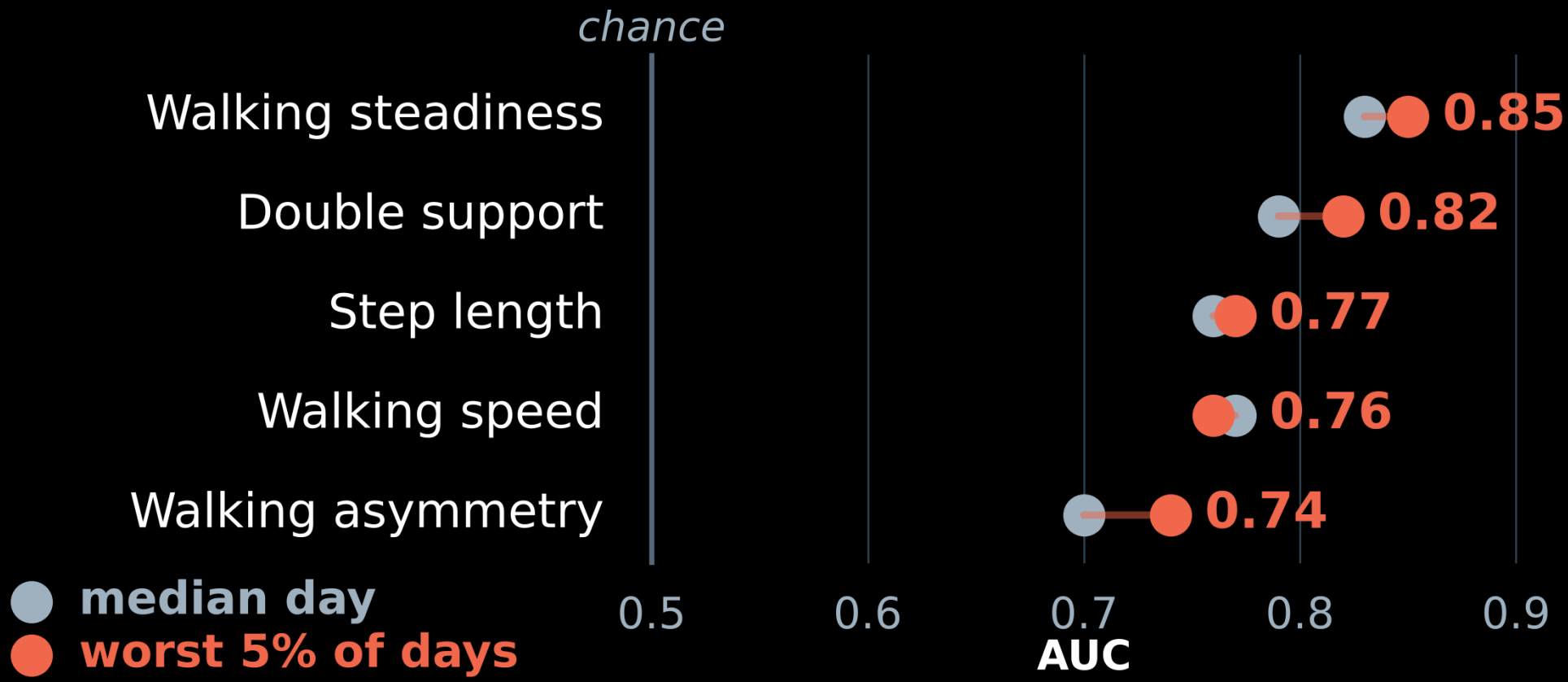
56 patients **42** controls



Maybe the average day isn't the important day.

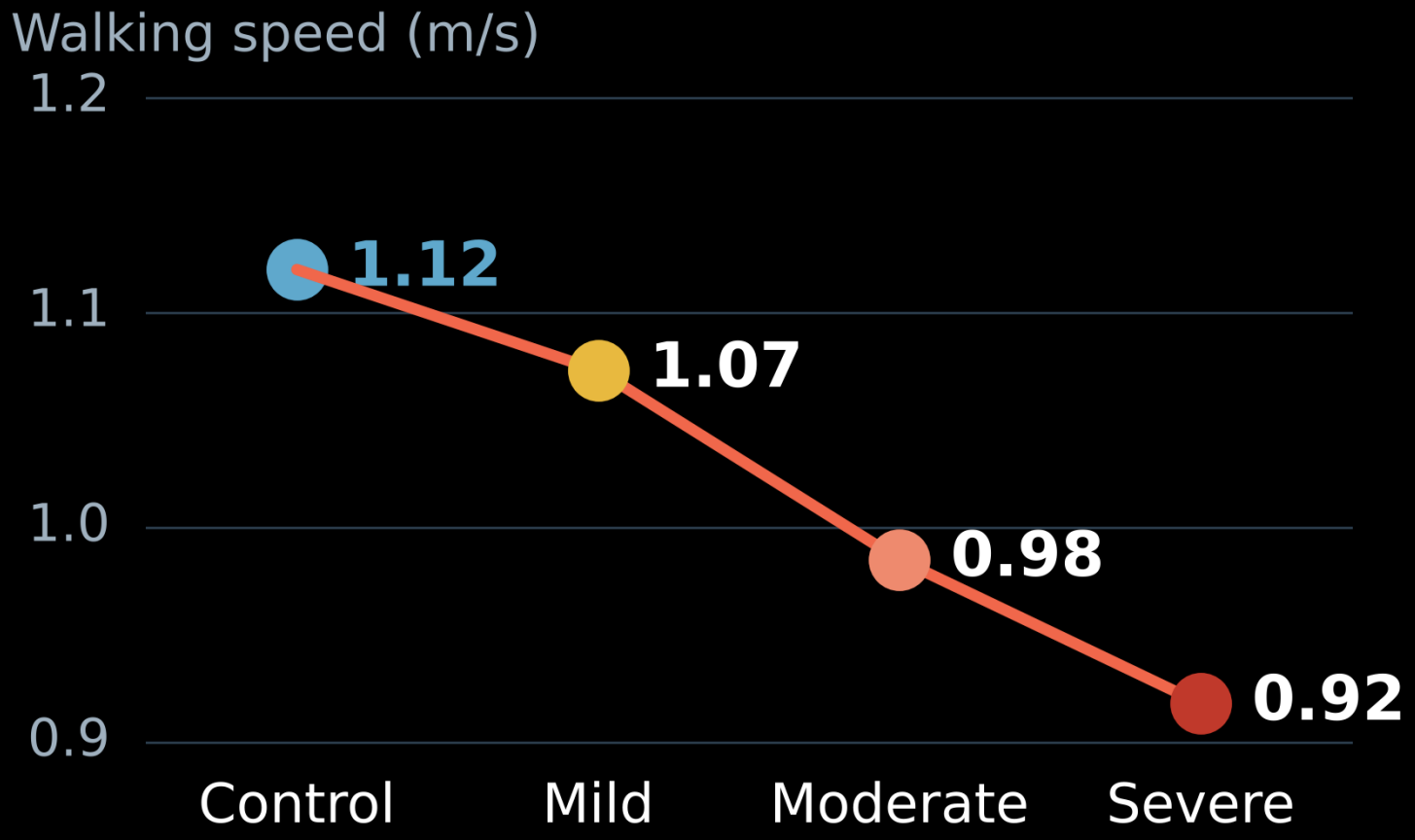


Every metric separated patients from controls.



In four of five metrics, the worst days separated the groups better.

Gait scaled with disease severity.



All five metrics changed in the same direction.

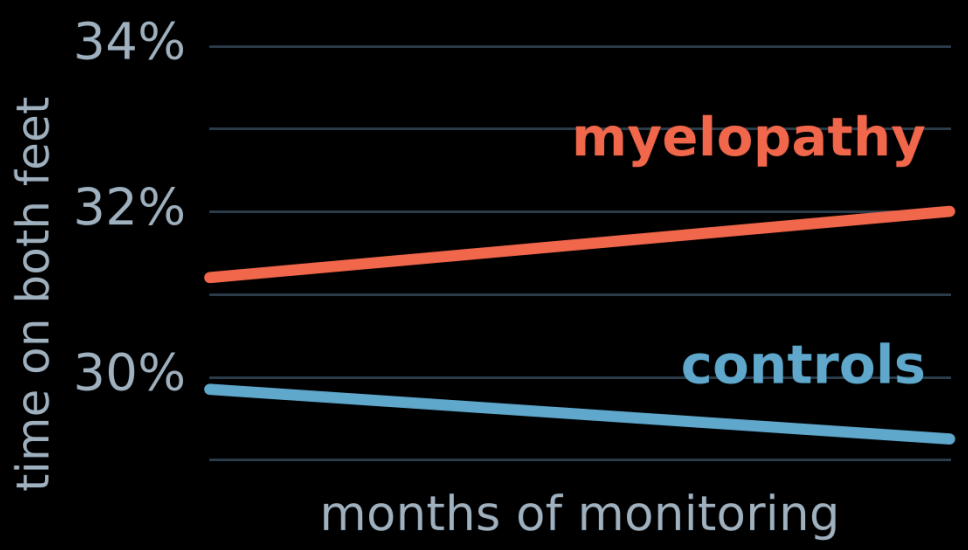
p < 0.001 for every one

For every metric, the worst days tracked severity more closely than the median.

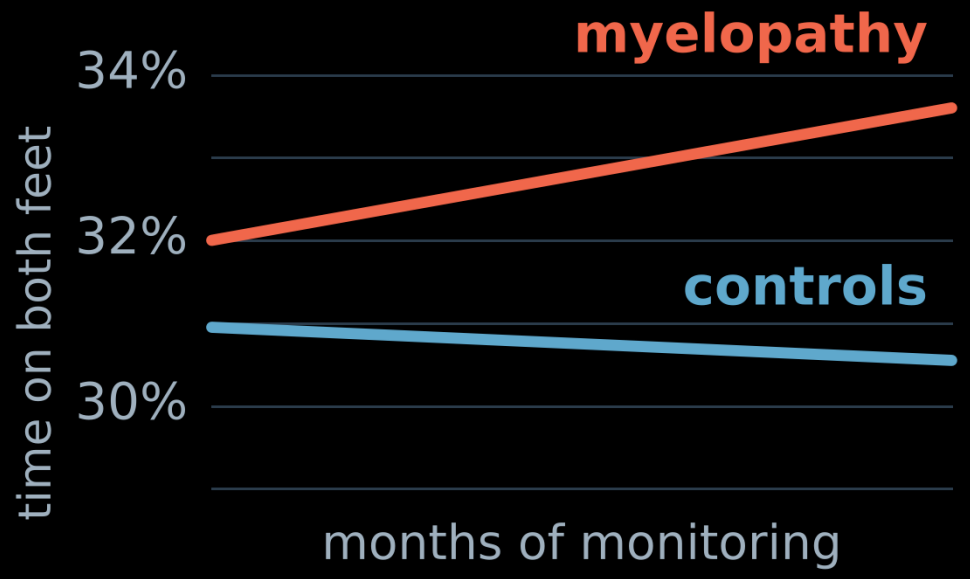
The signal was strongest in the bad days.

Double support: % of each step with both feet on the ground. Higher = less stable.

Median day



Worst 5% of days



Patients didn't just walk worse. On their worst days, they declined faster.

Each patient could signal their own decline.

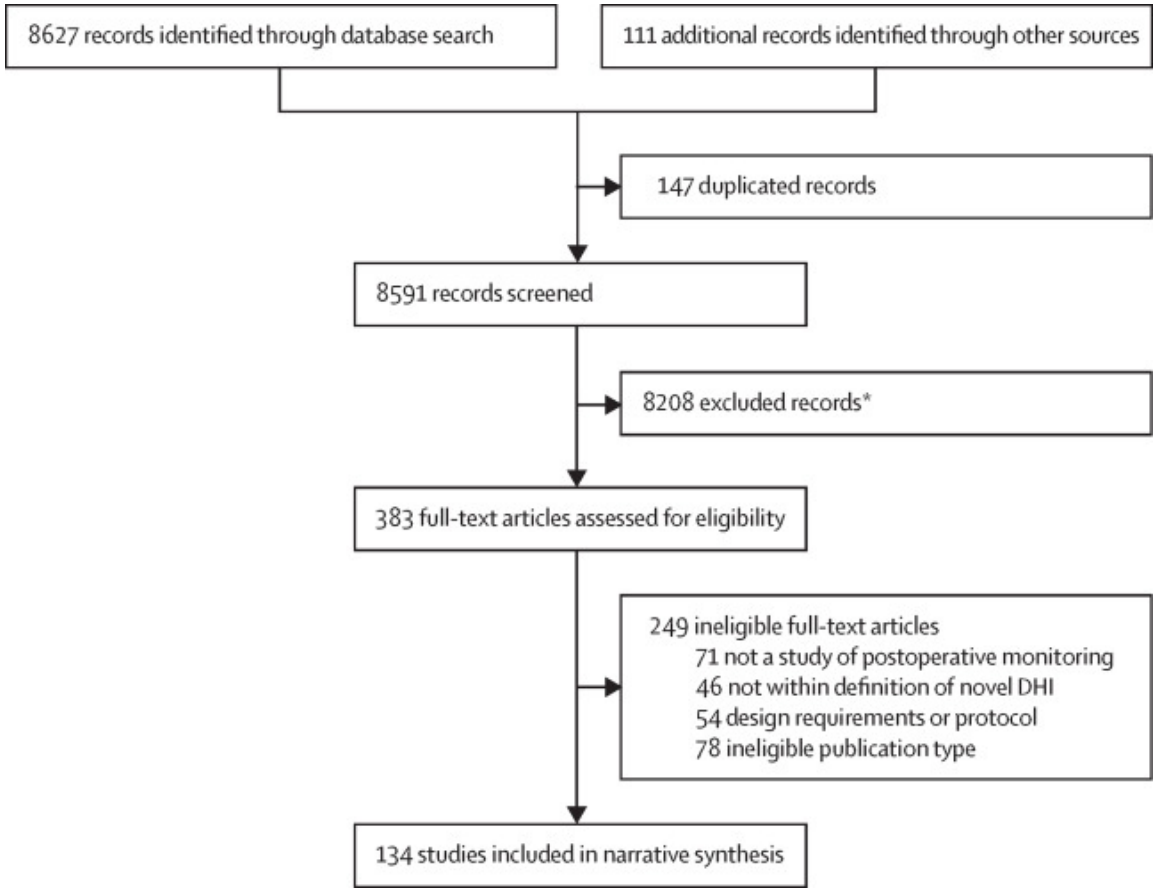
Flagged = worse by more than 1.5 SD than their own first 6 months, for ≥2 months in a row.



Two-thirds to three-quarters of patients were flagged.

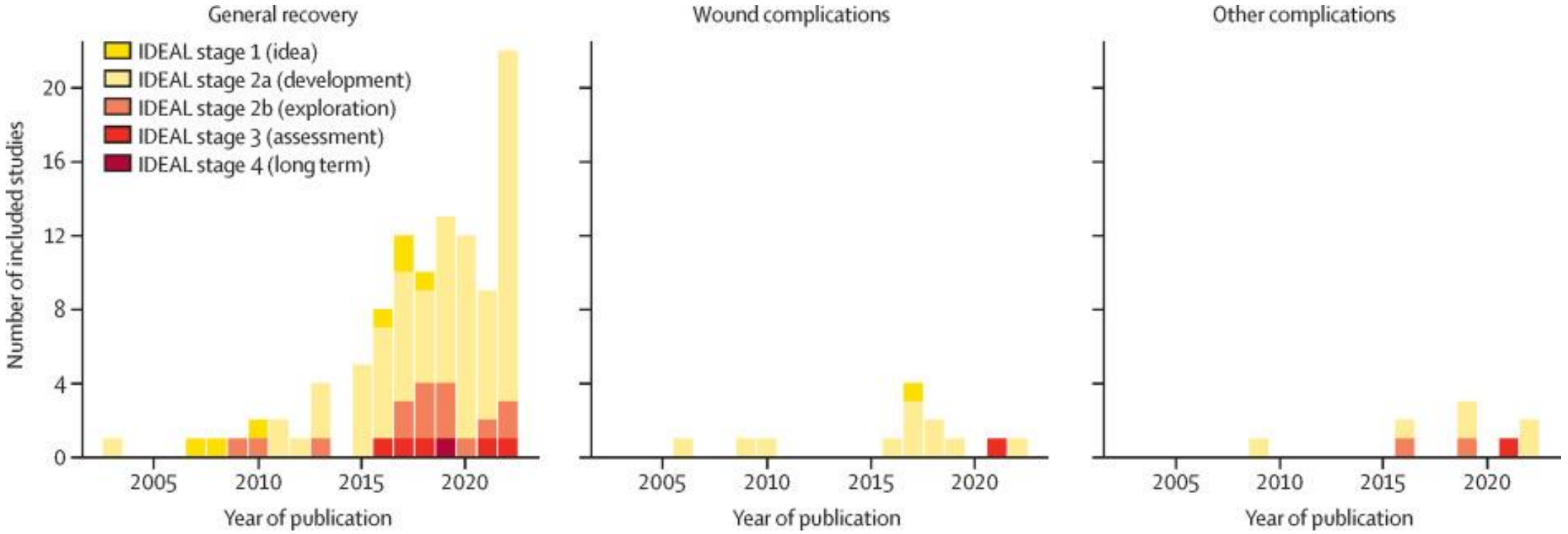
State of the Field

mHealth for Postop Monitoring: An Assessment of the Field



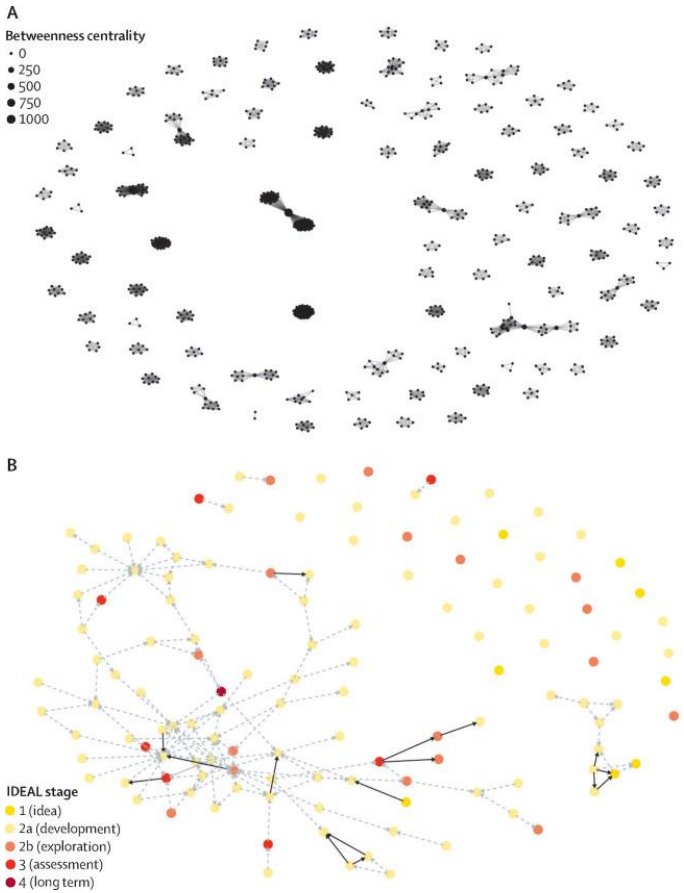
McLean et al, 2023; Lancet Digital Health

Maturity of mHealth Interventions to Date



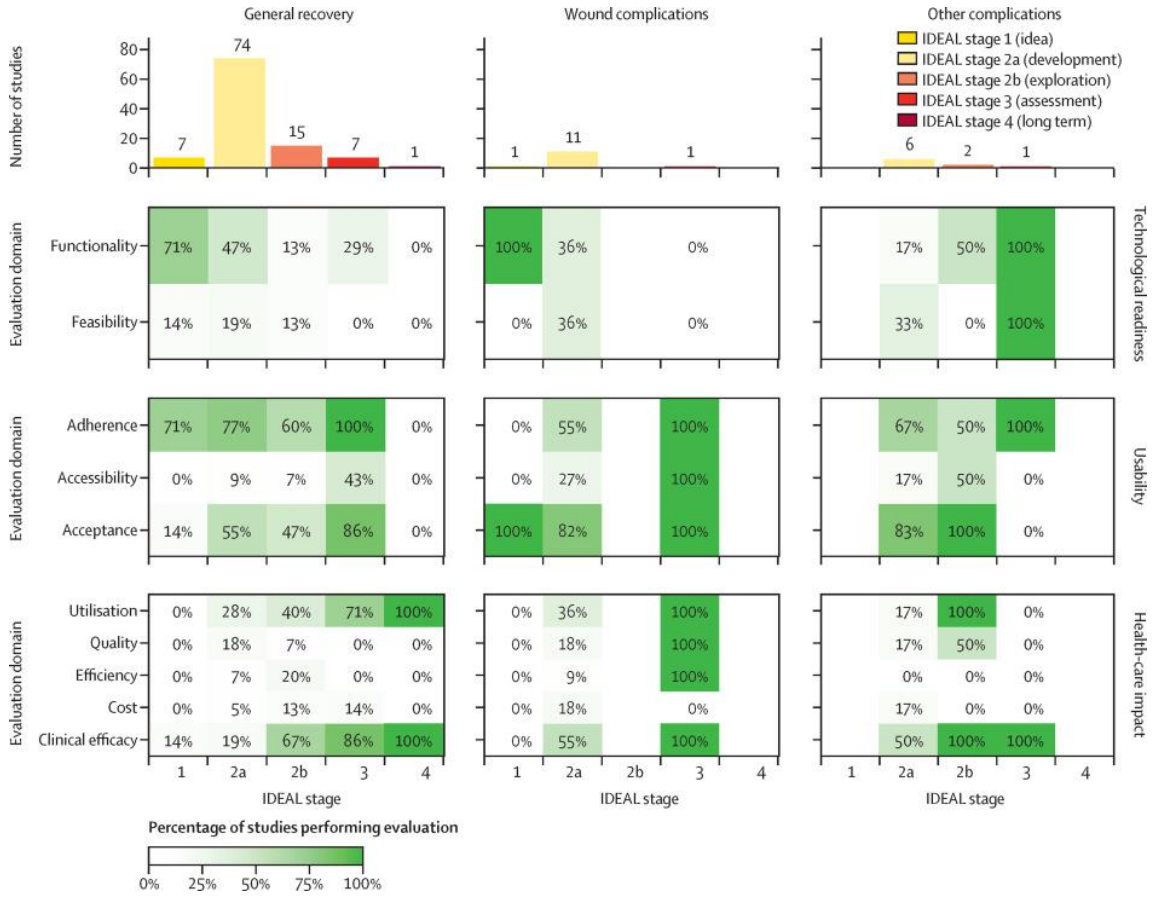
McLean et al, 2023; Lancet Digital Health

Evidence of Direct or Indirect Collaboration



McLean et al, 2023; Lancet Digital Health

Domains Evaluated by IDEAL Stage



McLean et al, 2023; Lancet Digital Health

Key Takeaways

- There has been tremendous growth in mHealth studies focused on *postoperative* monitoring alone, but most technologies are immature
- Limited evidence of collaboration or shared learning across efforts
- Substantial need to demonstrate clinical efficacy, quality and value
- Minimal focus on neurological recovery

Conclusions

- mHealth technology combined with AI methods have the potential to transform perioperative care
  - Preoperative evaluations
  - Targeted interventions
  - Postoperative monitor
  - Evaluating surgical outcomes

Conclusions

- mHealth technologies need further development, testing, and evidence of commercial viability
  - High-quality clinical evidence
  - Implementation studies
  - Commercialization and reimbursement strategies

Acknowledgments

Chenyang Lu, PhD (Computer Science)

Ziqi Xu (Computer Science)

Jingwen Zhang (Computer Science)

Ben Warner (Computer Science)

Madelyn Frumkin, MA (Psychology)

Salim Yakdan, MD (Neurosurgery)

Benjamin Steel (UC Berkely, Data Science)

Anthony Chen (WU, Computer Science)

Faraz Arkam, MD (Neurosurgery)

Pranay Singh (Neurosurgery)

Abigail Lewis (Neurosurgery)

Caitlin Kelleher, PhD (Computer Science)