Unsupervised Clustering of Longitudinal Clinical Measurements in Electronic Health Records

Arshiya Mariam[1,2], Hamed Javidi[1,2,3], Emily C. Zabor[1,4], Ran Zhao[1], Tomas Radivoyevitch[1], Daniel M. Rotroff[1,2,3,5*]

DOI: 10.1371/journal.pdig.0000628

[1] Department of Quantitative Health Sciences, Lerner Research Institute, Cleveland Clinic, Cleveland, Ohio, United States of America
[2] Center for Quantitative Metabolic Research, Cleveland Clinic, Cleveland, Ohio, United States of America
[3] Department of Electrical Engineering and Computer Science, Cleveland State University, Cleveland, Ohio, United States of America
[4] Taussig Cancer Institute, Cleveland Clinic, Cleveland, Ohio, United States of America
[5] Endocrinology and Metabolism Institute, Cleveland Clinic, Cleveland, Ohio, United States of America

PROBLEM STATEMENT & CONTEXT

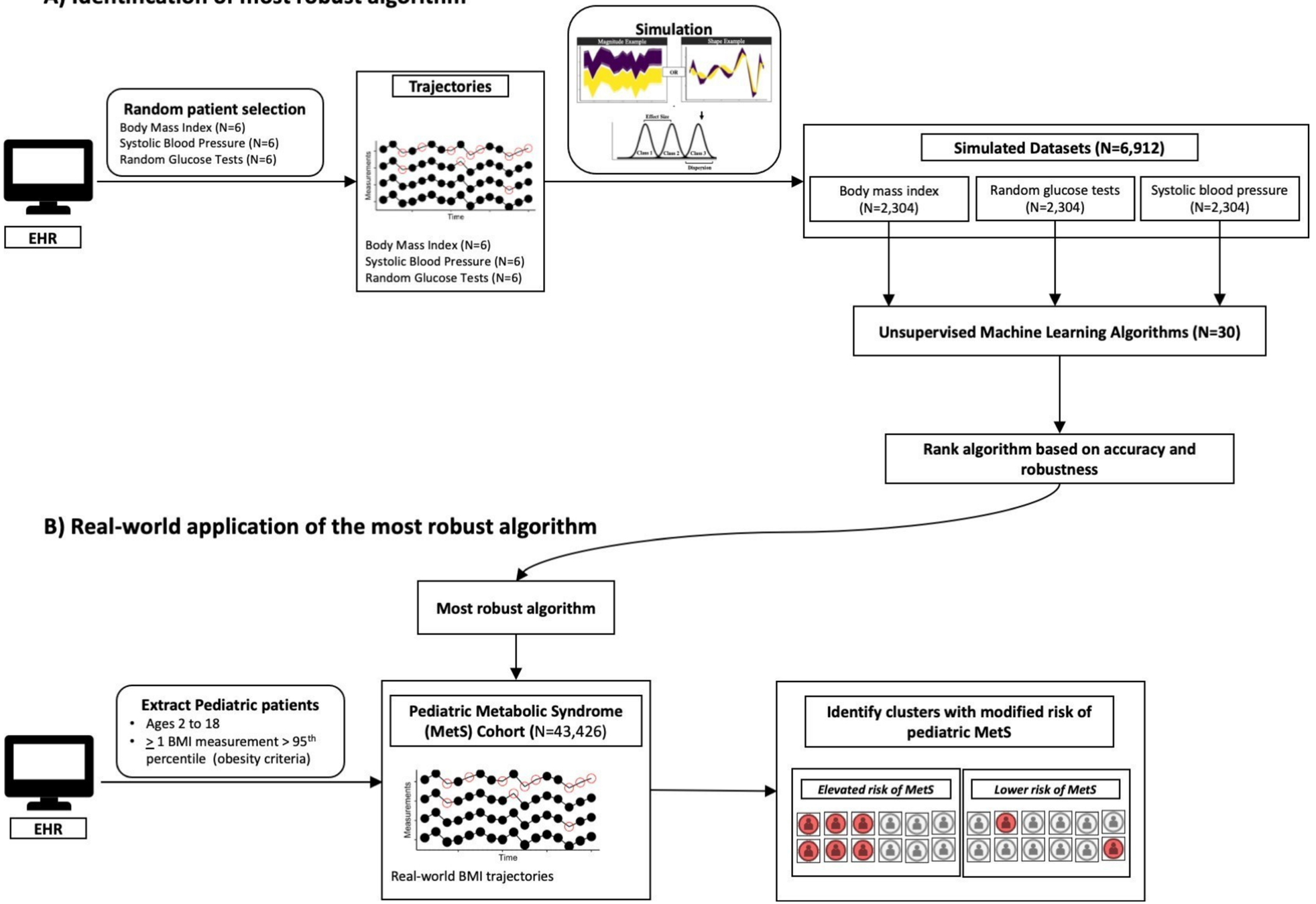
THE OPPORTUNITY:

1. Electronic Health Records contain longitudinal patient data with enormous research potential
2. Unsupervised algorithms from signal processing adapted for clinical use

THE CHALLENGE:

1. Clinical data has irregular sampling patterns (not routine)
2. High data missingness (healthcare utilization varies by patient)
3. Question remains: Which clustering algorithms perform best?

CRITICAL GAP: No systematic evaluation of unsupervised clustering on realistic clinical datasets with ground truth


A) Identification of most robust algorithm
Simulation
Magnitude Example
Shape Example
OR
Effect Size
Class 1
Class 2
Class 3
Dispersion
Random patient selection
Body Mass Index (N=6)
Systolic Blood Pressure (N=6)
Random Glucose Tests (N=6)
EHR
Trajectories
Measurements
Time
Body Mass Index (N=6)
Systolic Blood Pressure (N=6)
Random Glucose Tests (N=6)
Simulated Datasets (N=6,912)
Body mass index (N=2,304)
Random glucose tests (N=2,304)
Systolic blood pressure (N=2,304)
Unsupervised Machine Learning Algorithms (N=30)
Rank algorithm based on accuracy and robustness
B) Real-world application of the most robust algorithm
Most robust algorithm
Extract Pediatric patients
Ages 2 to 18
≥ 1 BMI measurement > 95th percentile (obesity criteria)
EHR
Pediatric Metabolic Syndrome (MetS) Cohort (N=43,426)
Measurements
Time
Real-world BMI trajectories
Identify clusters with modified risk of pediatric MetS
Elevated risk of MetS
Lower risk of MetS

RESEARCH OBJECTIVES

Objective 1: Algorithm Evaluation

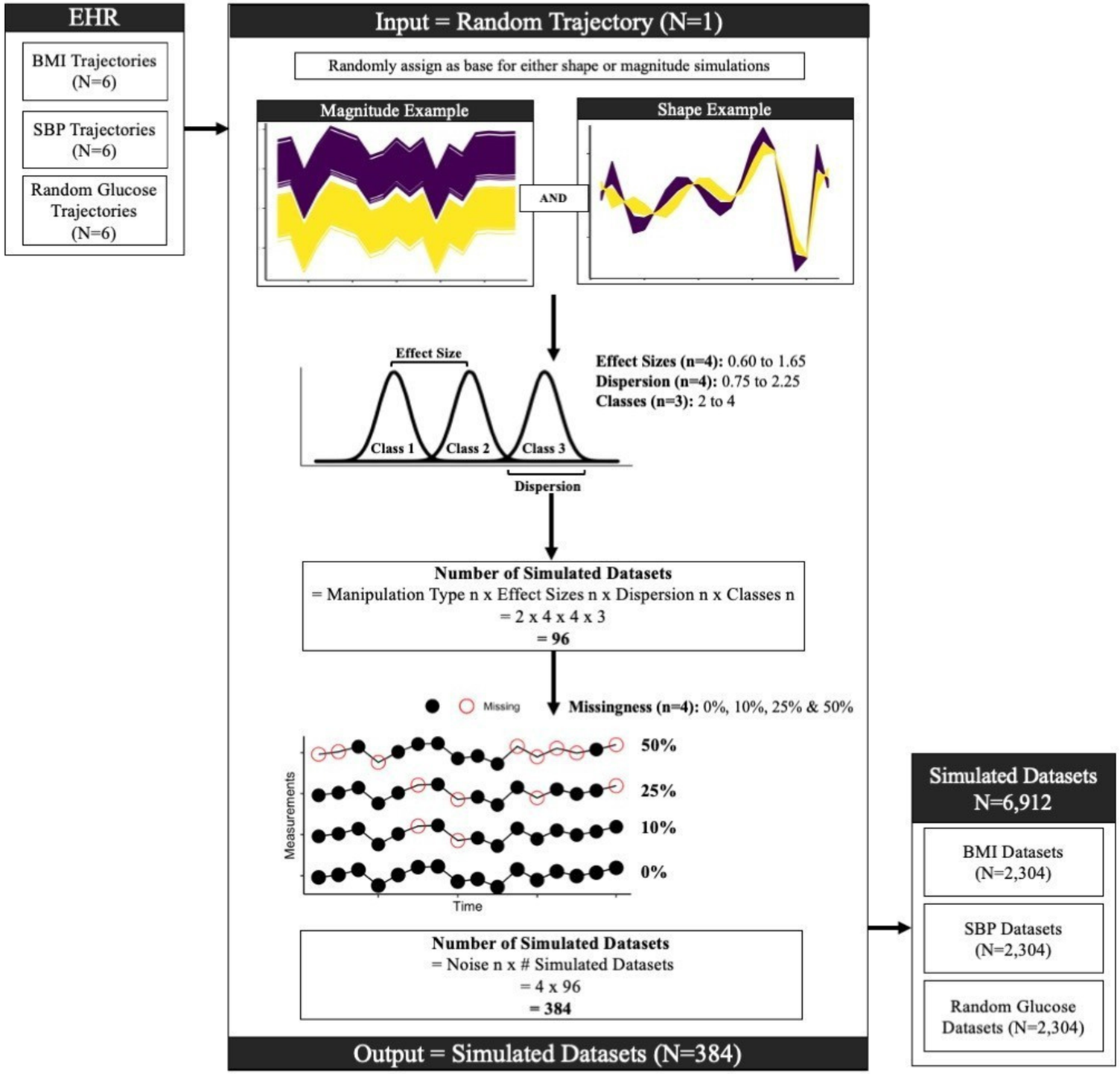
1. Evaluate 30 unsupervised clustering algorithms
2. Test on 6,912 simulated clinical datasets
3. Focus: Temporal matching + partitional/fuzzy clustering

Objective 2: Real-World Application

1. Apply best algorithm to real clinical cohort
2. 43,426 pediatric patients with obesity
3. Identify BMI trajectory patterns → MetS risk

Why Simulation?

1. Known ground truth = accurate accuracy measurement
2. Controlled parameters = understand performance under conditions


EHR
BMI Trajectories (N=6)
SBP Trajectories (N=6)
Random Glucose Trajectories (N=6)
Input = Random Trajectory (N=1)
Randomly assign as base for either shape or magnitude simulations
Magnitude Example
AND
Shape Example
Effect Size
Class 1
Class 2
Class 3
Dispersion
Effect Sizes (n=4): 0.60 to 1.65
Dispersion (n=4): 0.75 to 2.25
Classes (n=3): 2 to 4
Number of Simulated Datasets
= Manipulation Type n x Effect Sizes n x Dispersion n x Classes n
= 2 x 4 x 4 x 3
= 96
Missing
Missingness (n=4): 0%, 10%, 25% & 50%
50%
25%
10%
0%
Measurements
Time
Number of Simulated Datasets
= Noise n x # Simulated Datasets
= 4 x 96
= 384
Output = Simulated Datasets (N=384)
Simulated Datasets N=6,912
BMI Datasets (N=2,304)
SBP Datasets (N=2,304)
Random Glucose Datasets (N=2,304)

METHODOLOGY: DATASET PREPARATION

Clinical Measurements (3):

1. BMI (Body Mass Index)
2. SBP (Systolic Blood Pressure)
3. Random Glucose

Simulated Datasets:

1. real trajectories per measurement type
2. 16-year span with yearly measurements
3. Total: 6,912 datasets

Real Cohort (MetS):

1. N = 43,426 children (ages 2-18)
2. 55.7% male, 73.9% Caucasian
3. Mean follow-up: 8.47 years | MetS cases: 3.4%

METHODOLOGY: HOW 30 ALGORITHMS BUILT

30 Unique Algorithms = 3 Components × Options

Component 1: Assignment Method (2)

- Partitional: Hard assignment (1 cluster per point)
- Fuzzy: Probabilistic assignment (weights across clusters)

Component 2: Distance Measure (8)

- DTW, DTW-LB, LB-Improved, LB-Keogh (temporal matching)
- Euclidean, Manhattan, Soft-DTW, SBD (alternative metrics)

Component 3: Centroid Computation (6)

- PAM, DBA, Soft-DTW Centroids + other methods

2 × 8 × 6 = 30 Total Algorithms

EVALUATION METRICS

Adjusted Rand Index (ARI)
Measures: Agreement between predicted clusters & true clusters
Range: -1 to +1

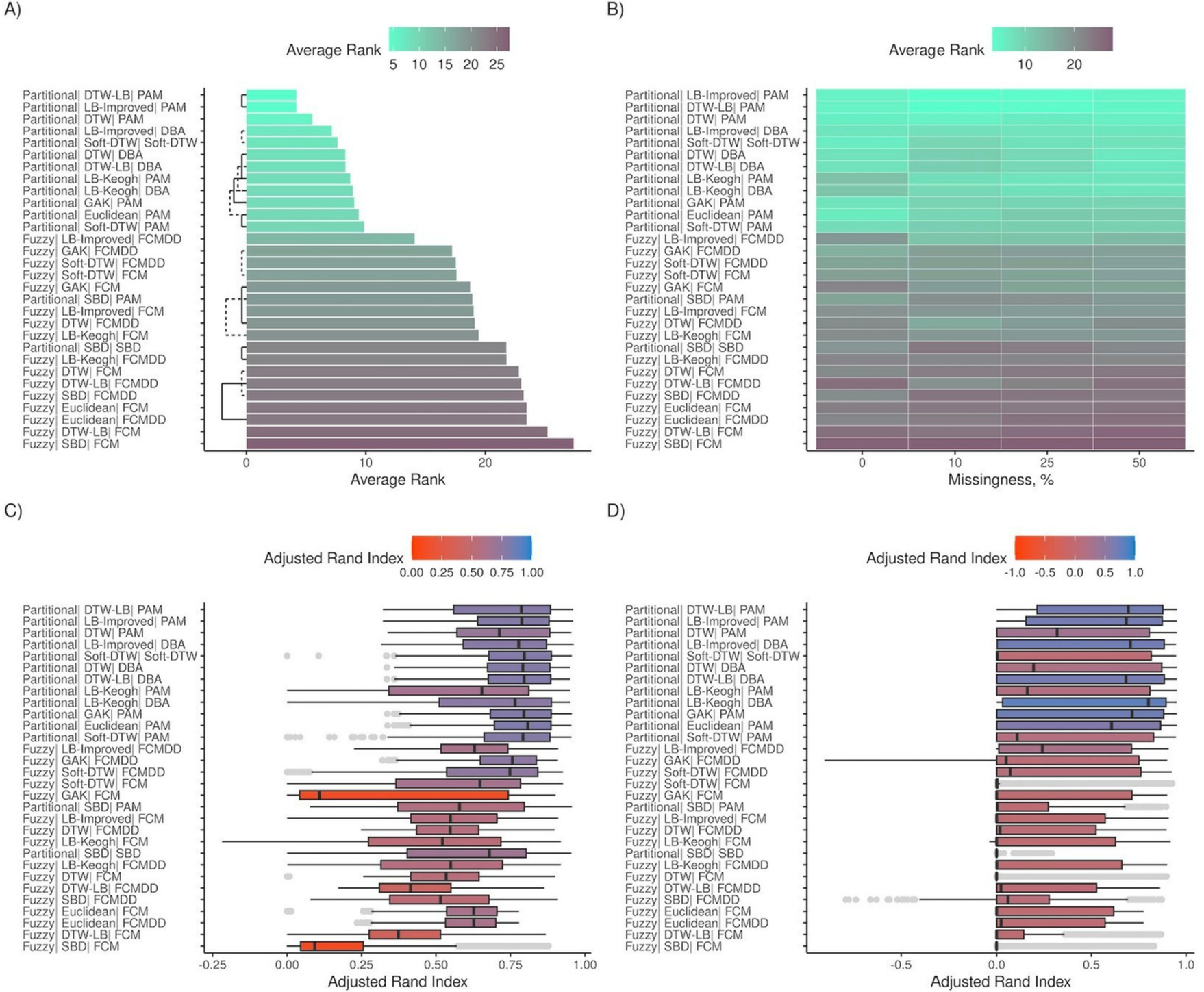
- +1 = Perfect clustering
- 0 = Random clustering
- -1 = Completely opposite

Interpretation:

ARI > 0.70	ARI 0.50-0.70	ARI < 0.30
Good	**Moderate**	**Poor**

Statistical Testing:

- Nemenyi Test (multiple comparisons) FDR correction
- Consensus >70% = stable clusters


A)
Average Rank
5 10 15 20 25
Partitional| DTW-LB| PAM
Partitional| LB-Improved| PAM
Partitional| DTW| PAM
Partitional| LB-Improved| DBA
Partitional| Soft-DTW| Soft-DTW
Partitional| DTW| DBA
Partitional| DTW-LB| DBA
Partitional| LB-Keogh| PAM
Partitional| LB-Keogh| DBA
Partitional| GAK| PAM
Partitional| Euclidean| PAM
Partitional| Soft-DTW| PAM
Fuzzy| LB-Improved| FCMDD
Fuzzy| GAK| FCMDD
Fuzzy| Soft-DTW| FCMDD
Fuzzy| Soft-DTW| FCM
Fuzzy| GAK| FCM
Partitional| SBD| PAM
Fuzzy| LB-Improved| FCM
Fuzzy| DTW| FCMDD
Fuzzy| LB-Keogh| FCM
Partitional| SBD| SBD
Fuzzy| LB-Keogh| FCMDD
Fuzzy| DTW| FCM
Fuzzy| DTW-LB| FCMDD
Fuzzy| SBD| FCMDD
Fuzzy| Euclidean| FCM
Fuzzy| Euclidean| FCMDD
Fuzzy| DTW-LB| FCM
Fuzzy| SBD| FCM
0 10 20
Average Rank
B)
Average Rank
10 20
Partitional| LB-Improved| PAM
Partitional| DTW-LB| PAM
Partitional| DTW| PAM
Partitional| LB-Improved| DBA
Partitional| Soft-DTW| Soft-DTW
Partitional| DTW| DBA
Partitional| DTW-LB| DBA
Partitional| LB-Keogh| PAM
Partitional| LB-Keogh| DBA
Partitional| GAK| PAM
Partitional| Euclidean| PAM
Partitional| Soft-DTW| PAM
Fuzzy| LB-Improved| FCMDD
Fuzzy| GAK| FCMDD
Fuzzy| Soft-DTW| FCMDD
Fuzzy| Soft-DTW| FCM
Fuzzy| GAK| FCM
Partitional| SBD| PAM
Fuzzy| LB-Improved| FCM
Fuzzy| DTW| FCMDD
Fuzzy| LB-Keogh| FCM
Partitional| SBD| SBD
Fuzzy| LB-Keogh| FCMDD
Fuzzy| DTW| FCM
Fuzzy| DTW-LB| FCMDD
Fuzzy| SBD| FCMDD
Fuzzy| Euclidean| FCM
Fuzzy| Euclidean| FCMDD
Fuzzy| DTW-LB| FCM
Fuzzy| SBD| FCM
0 10 25 50
Missingness, %
C)
Adjusted Rand Index
0.00 0.25 0.50 0.75 1.00
Partitional| DTW-LB| PAM
Partitional| LB-Improved| PAM
Partitional| DTW| PAM
Partitional| LB-Improved| DBA
Partitional| Soft-DTW| Soft-DTW
Partitional| DTW| DBA
Partitional| DTW-LB| DBA
Partitional| LB-Keogh| PAM
Partitional| LB-Keogh| DBA
Partitional| GAK| PAM
Partitional| Euclidean| PAM
Partitional| Soft-DTW| PAM
Fuzzy| LB-Improved| FCMDD
Fuzzy| GAK| FCMDD
Fuzzy| Soft-DTW| FCMDD
Fuzzy| Soft-DTW| FCM
Fuzzy| GAK| FCM
Partitional| SBD| PAM
Fuzzy| LB-Improved| FCM
Fuzzy| DTW| FCMDD
Fuzzy| LB-Keogh| FCM
Partitional| SBD| SBD
Fuzzy| LB-Keogh| FCMDD
Fuzzy| DTW| FCM
Fuzzy| DTW-LB| FCMDD
Fuzzy| SBD| FCMDD
Fuzzy| Euclidean| FCM
Fuzzy| Euclidean| FCMDD
Fuzzy| DTW-LB| FCM
Fuzzy| SBD| FCM
-0.25 0.00 0.25 0.50 0.75 1.00
Adjusted Rand Index
D)
Adjusted Rand Index
-1.0 -0.5 0.0 0.5 1.0
Partitional| DTW-LB| PAM
Partitional| LB-Improved| PAM
Partitional| DTW| PAM
Partitional| LB-Improved| DBA
Partitional| Soft-DTW| Soft-DTW
Partitional| DTW| DBA
Partitional| DTW-LB| DBA
Partitional| LB-Keogh| PAM
Partitional| LB-Keogh| DBA
Partitional| GAK| PAM
Partitional| Euclidean| PAM
Partitional| Soft-DTW| PAM
Fuzzy| LB-Improved| FCMDD
Fuzzy| GAK| FCMDD
Fuzzy| Soft-DTW| FCMDD
Fuzzy| Soft-DTW| FCM
Fuzzy| GAK| FCM
Partitional| SBD| PAM
Fuzzy| LB-Improved| FCM
Fuzzy| DTW| FCMDD
Fuzzy| LB-Keogh| FCM
Partitional| SBD| SBD
Fuzzy| LB-Keogh| FCMDD
Fuzzy| DTW| FCM
Fuzzy| DTW-LB| FCMDD
Fuzzy| SBD| FCMDD
Fuzzy| Euclidean| FCM
Fuzzy| Euclidean| FCMDD
Fuzzy| DTW-LB| FCM
Fuzzy| SBD| FCM
-0.5 0.0 0.5 1.0
Adjusted Rand Index

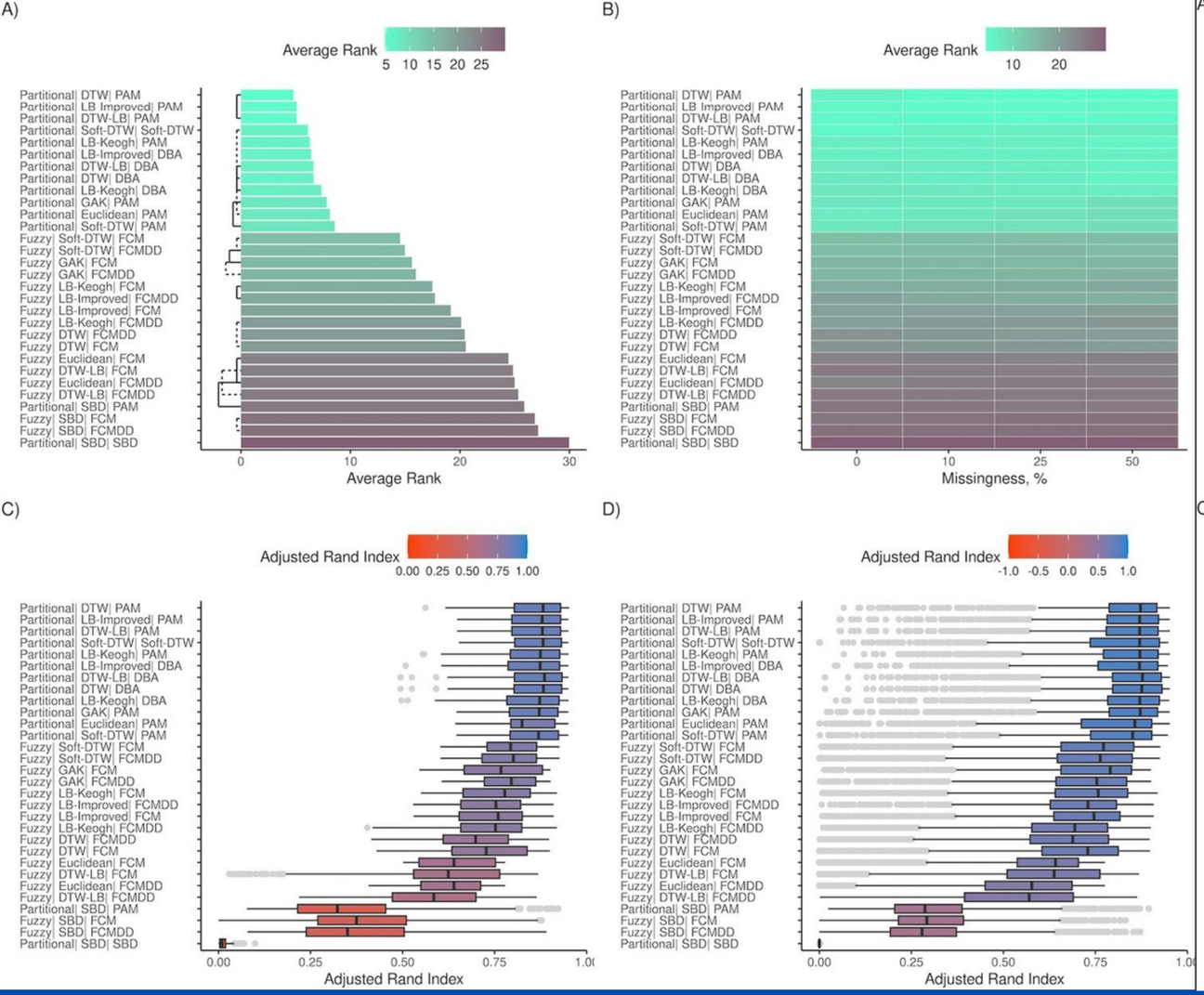
KEY RESULTS: ALGORITHM PERFORMANCE

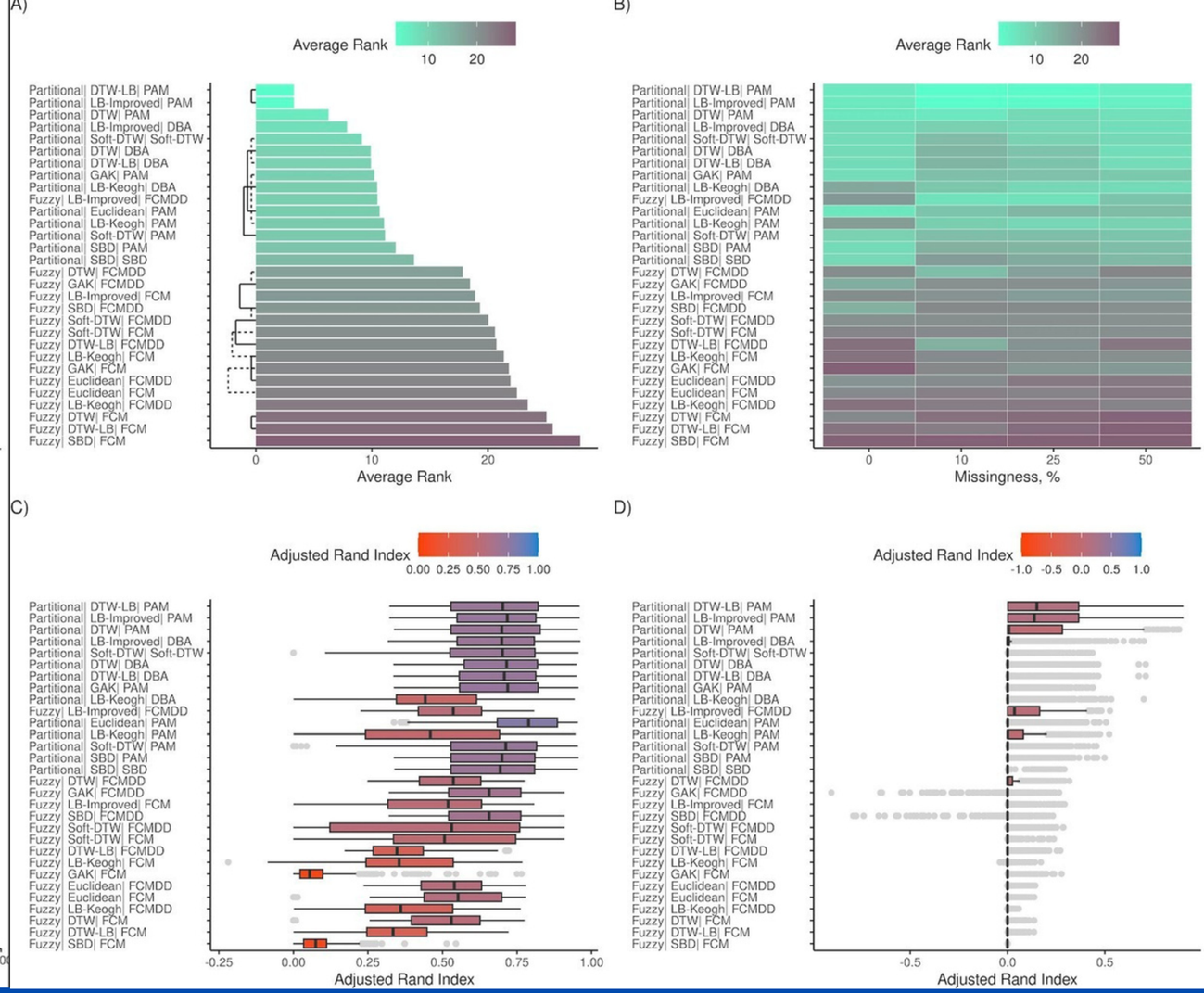
🏆 TOP 3 ALGORITHMS (Overall):

- **1st** DTW-LB + PAM — Mean Rank: 4.19 ($p < .05$)
- **2nd** (TIE) LB-Improved + PAM — Mean Rank: 4.19 ($p < .05$)
- **3rd** DTW + PAM — Mean Rank: 5.53 ($p < .05$)

KEY FINDINGS:

- **Partitional Clustering** ranked significantly higher than Fuzzy
- **SURPRISE** PAM outperformed DBA (which was designed for DTW!)
- **DTW variants** consistently superior to other metrics





MAGNITUDE VS SHAPE

CRITICAL FINDING: MAGNITUDE VS SHAPE

Magnitude-Based
(Classes differ by peak values)

- ✓ All algorithms performed well (ARI ~0.70)
- ✓ Missingness MINIMAL impact
- ✓ Easy to detect high vs low BMI

Shape-Based
(Classes differ by trajectory trend)

- ✗ Lower accuracies (ARI 0.55-0.60)
- ✗ 10% missing → 40% accuracy DROP!
- ✗ Hard to detect rising vs stable trends

⚠ **CLINICAL IMPLICATION** : If your EHR has missing data, use magnitude-based features (peak values) rather than shape-based features (trends)

REAL-WORLD APPLICATION: PEDIATRIC METS

Algorithm Selected:

Partitional Clustering + DTW + PAM

- One of top 3 most accurate
- Robust to variable trajectory lengths
- Handles real-world variation

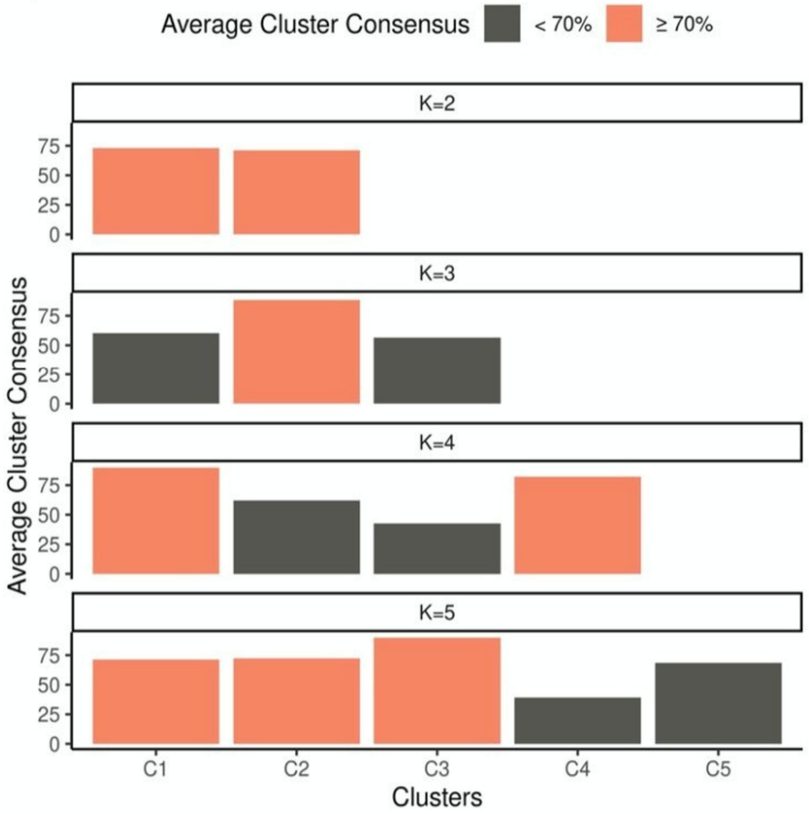
Optimal Clusters:

k = 5 identified

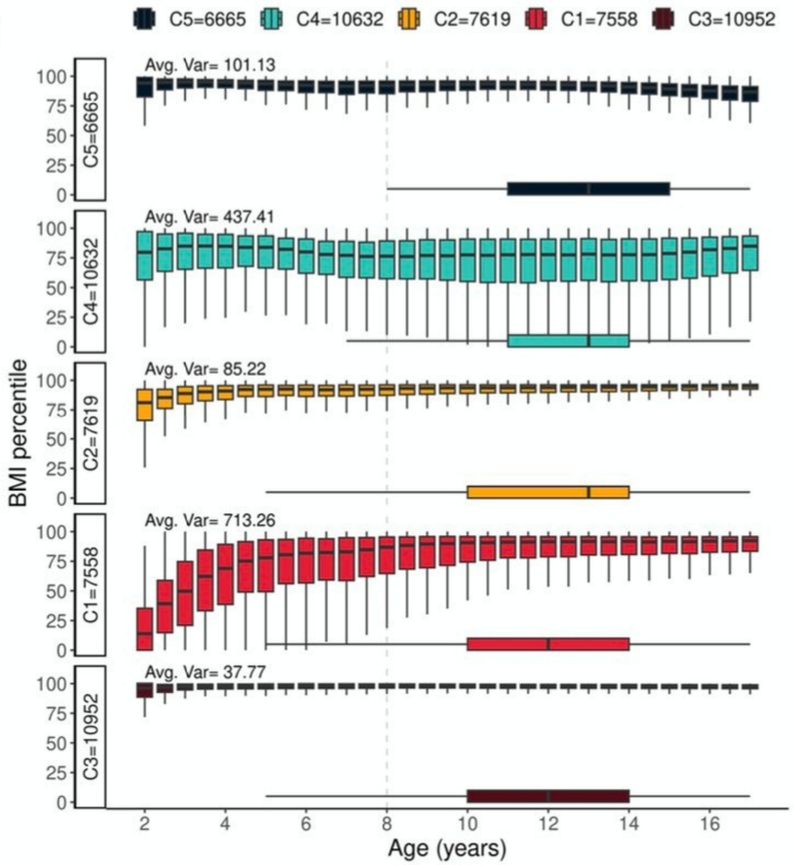
- 3 of 5 clusters >70% consensus ✓
- Stable, meaningful patterns
- Clinically interpretable

Cohort: N = 43,426 children | Ages 2-18 | 55.7% male | 73.9% Caucasian | Mean FU 8.47 years | 3.4% MetS cases

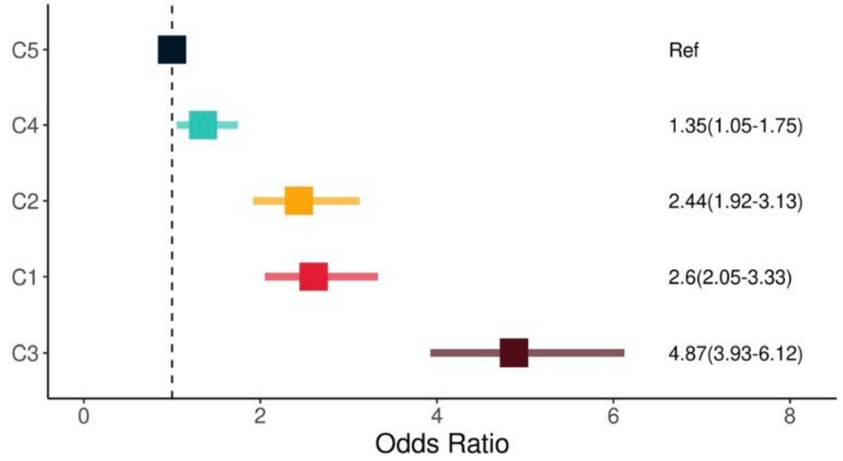
A)



B)



C)



D)

Clusters	*N*	*MetS N*	*MetS %*	*OR (95% CI)*	*Pr(>|z|)*
C1	*7558*	*260*	*3.44*	*2.6(2.05-3.33)*	*9.42e-15*
C2	*7619*	*246*	*3.23*	*2.44(1.92-3.13)*	*7.79e-13*
C3	*10952*	*685*	*6.25*	*4.87(3.93-6.12)*	*1.82e-44*
C4	*10632*	*193*	*1.82*	*1.35(1.05-1.75)*	*1.94e-02*
C5	*6665*	*90*	*1.35*	*Ref*	*Ref*

KEY FINDINGS: BMI TRAJECTORY CLUSTERS

Cluster C3: Consistently High BMI
OR 4.87 (95% CI: 3.93–6.12) ⚠️ **HIGHEST RISK** | 4.87× higher MetS odds

Cluster C1 & C2: Increasing BMI with Age
OR 2.44–2.60 | 2.4–2.6× higher MetS odds | Rising trajectory pattern

Cluster C4: Variable/Unstable BMI
OR 1.35 | Dysregulation pattern | High BMI variance

Cluster C5: Stable Low BMI (REFERENCE)
OR 1.0 | Protective pattern | Target for interventions

CLINICAL IMPLICATIONS & INSIGHTS

Key Takeaway: Longitudinal > Static Measurements
Not just WHAT the BMI is, but HOW IT CHANGES matters!

Clinical Workflow:

1. **Identify patient trajectory patterns in your EHR**
2. **Risk-stratify based on cluster membership**
3. **Allocate interventions based on risk cluster**
4. **Monitor for cluster transitions**
5. **Measure intervention outcomes per cluster**

✓ **Enables Precision Medicine**: Identify clinically meaningful subgroups WITHOUT predefined labels | Discover hidden phenotypes | Stratify interventions

LIMITATIONS & CONCLUSIONS

Key Limitations

- Limited trajectory diversity (6 per type)
- Simple imputation method (mean only)
- Selection bias in real cohort
- MetS definition ambiguous

Key Takeaways

- Limited trajectory diversity (6 per type)
- Simple imputation method (mean only)
- Selection bias in real cohort
- MetS definition ambiguous

CONCLUSION:

Systematic algorithm evaluation is CRITICAL for EHR clustering reliability

Longitudinal patterns > single measurements for patient stratification

CONCLUSION

KEY FINDINGS:
- DTW-based algorithms optimal for clinical time-series clustering
- Magnitude patterns robust to missingness; shape patterns vulnerable
- 5 distinct BMI trajectory clusters identified with MetS OR: 1.0-4.87
- Trajectory patterns superior to single measurements for risk prediction

CLINICAL SIGNIFICANCE:
→ Enables precision medicine: different trajectories → different interventions
→ C3 (consistently high BMI) requires urgent intervention (OR 4.87)
→ For EHRs with missing data (>10%), use magnitude features, not shape

CONTRIBUTION TO SCIENCE:
✓ First systematic evaluation of 30 algorithms on clinical EHR data
✓ Rigorous two-phase validation (simulation + real-world)
✓ Reproducible and evidence-based algorithm selection guidance

FUTURE DIRECTIONS:
- Advanced imputation methods • Broader measurement types
- Hierarchical clustering • Clinical integration studies

TAKEAWAY:
Longitudinal trajectory analysis > static measurements for clinical phenotyping and precision medicine implementation.