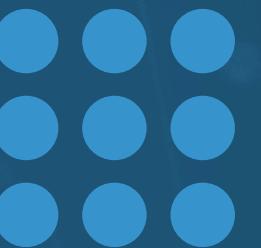
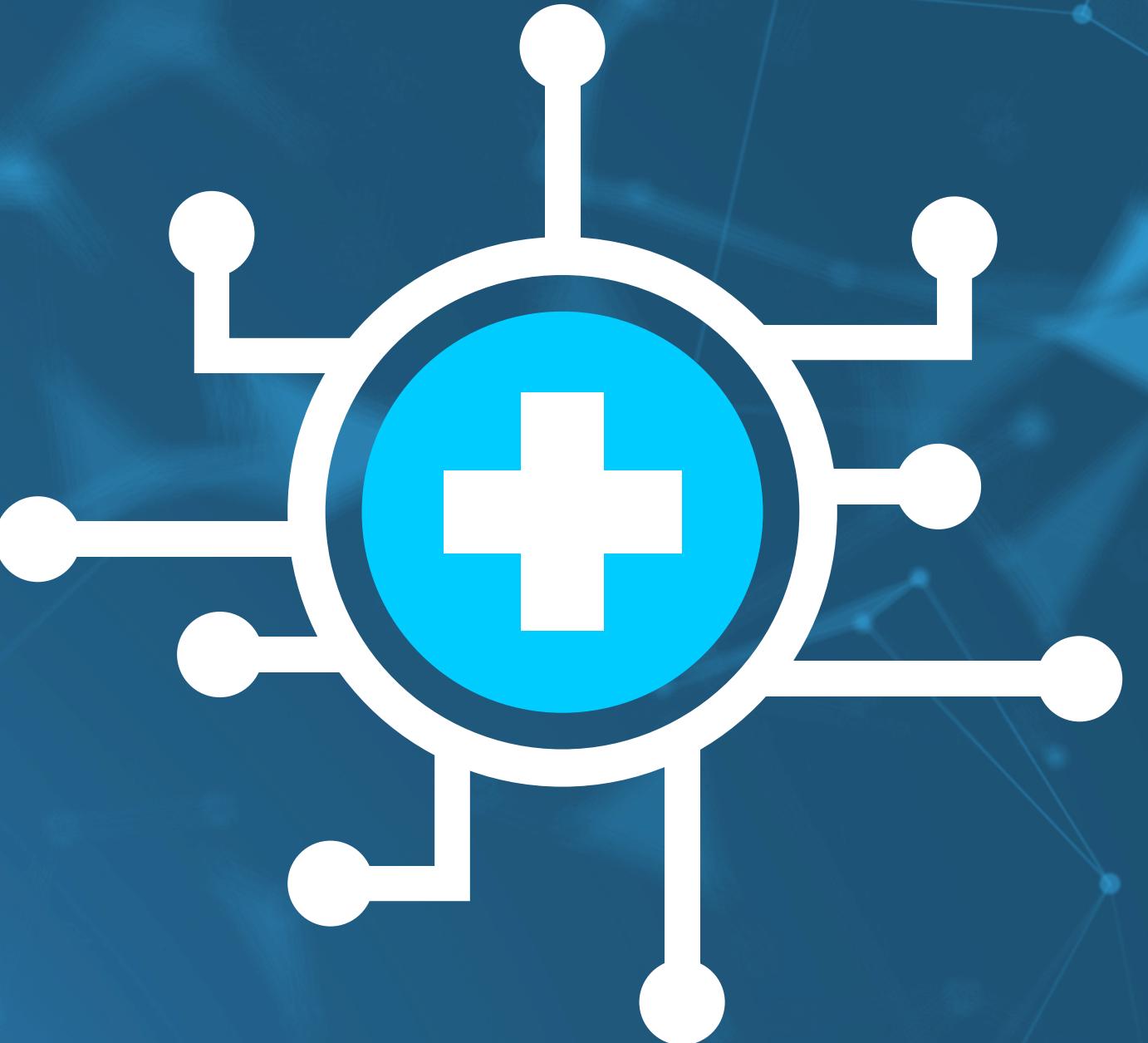


Lifespan Health May Meeting
AI Workshop
21st May 2025

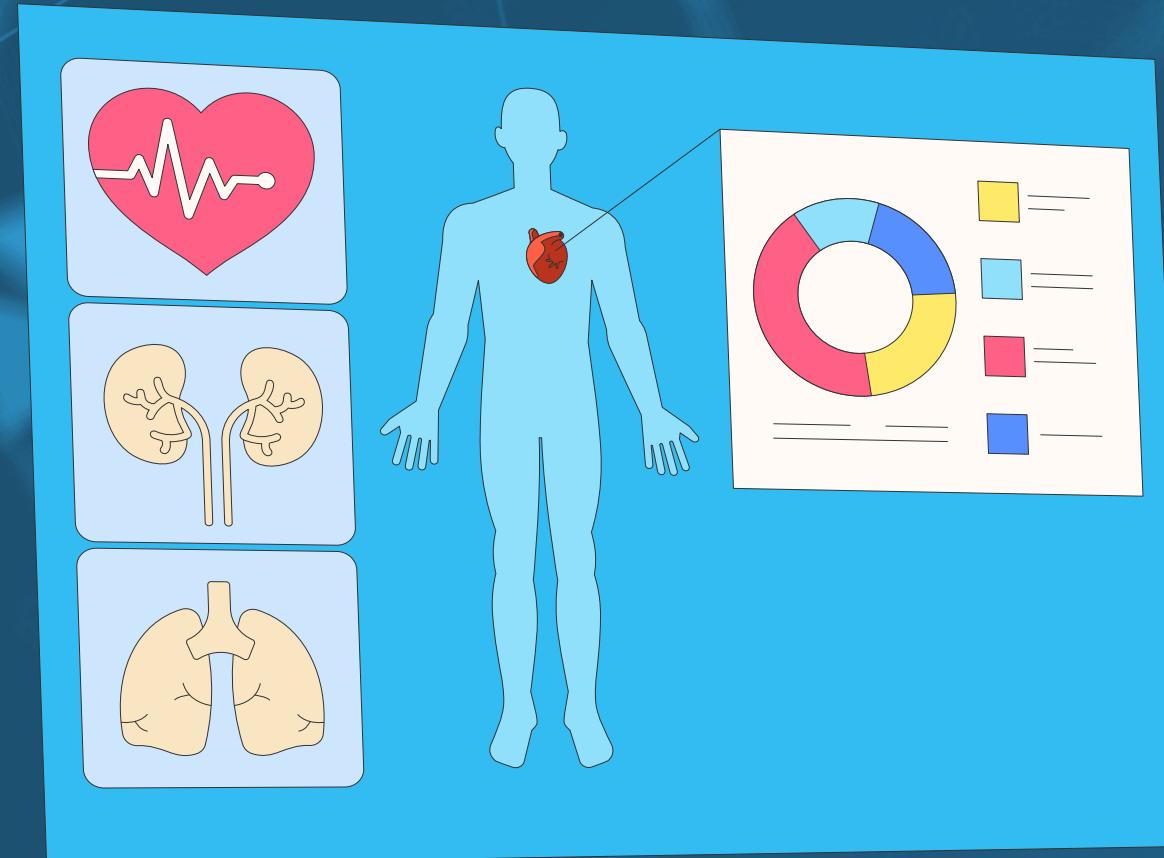
AI-Driven Identification of Symptom Clusters

A Case Study in Long Covid Pattern
Recognition

Jasmyn Gooding
PhD Researcher in Computer Science
MED Research Coordinator



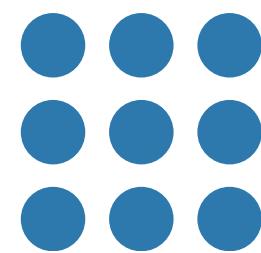
Pattern recognition



Helps uncover hidden structures in complex data
by grouping similar features or behaviors—
essential for understanding conditions that lack
clear-cut diagnostic markers.

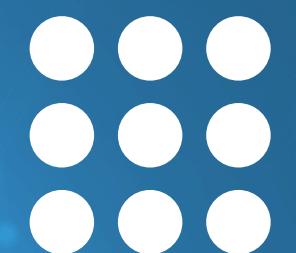
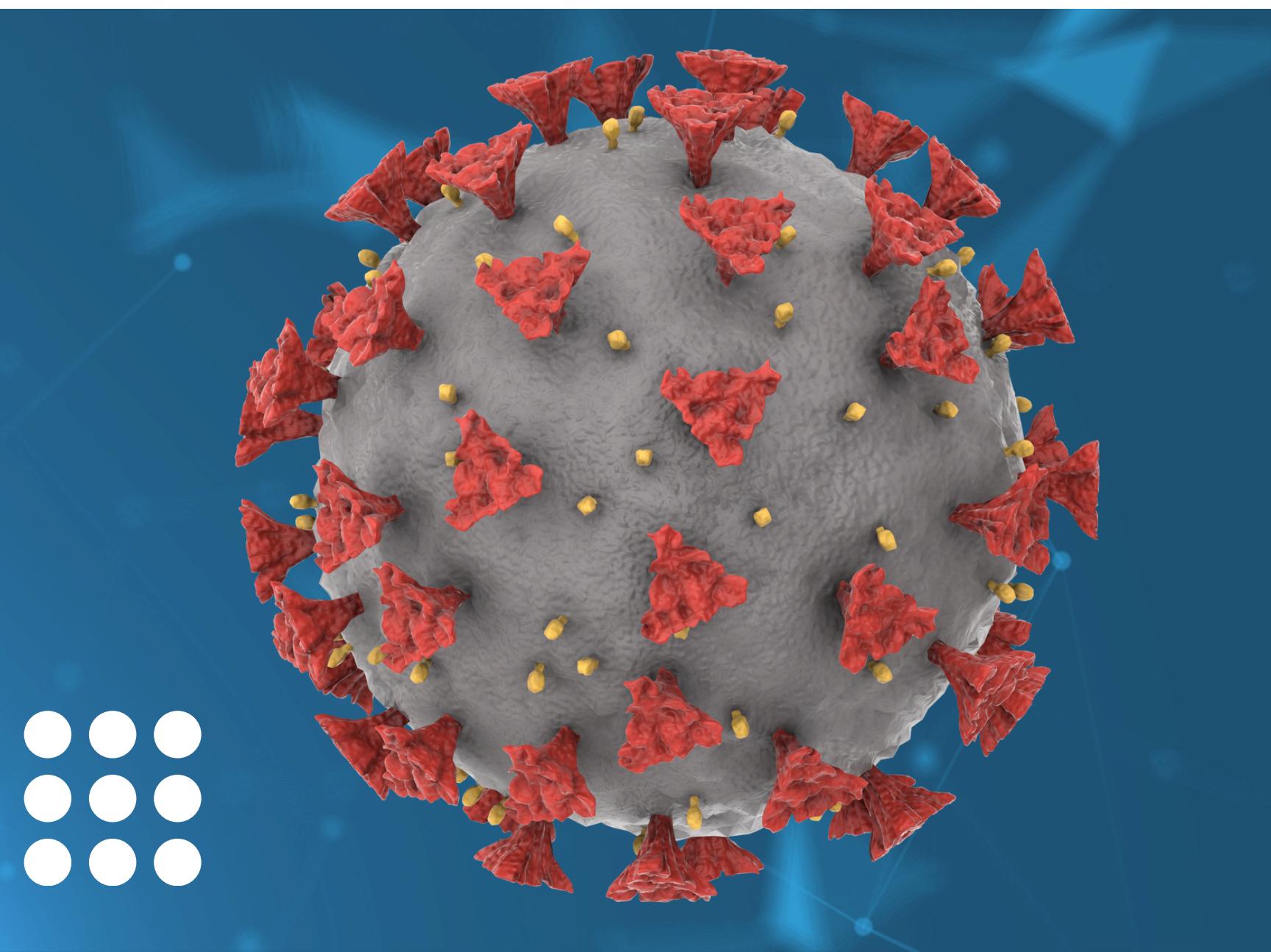
Symptoms like fatigue, brain fog, and palpitations
are non-specific and overlap with multiple
conditions, complicating traditional diagnostics.

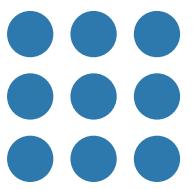
**Can AI help make sense of
complex symptom data?**



Clinical burden and diagnostic ambiguity

-
- * Clinical burden: Long Covid leads to frequent healthcare visits, extensive testing, and prolonged disability, straining both patients and healthcare systems.
 - * Diagnostic ambiguity: The absence of definitive biomarkers and overlapping symptoms with other conditions make accurate diagnosis and treatment planning difficult

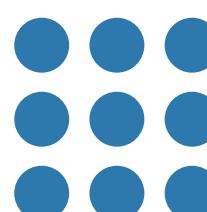




The Challenge for Clinicians

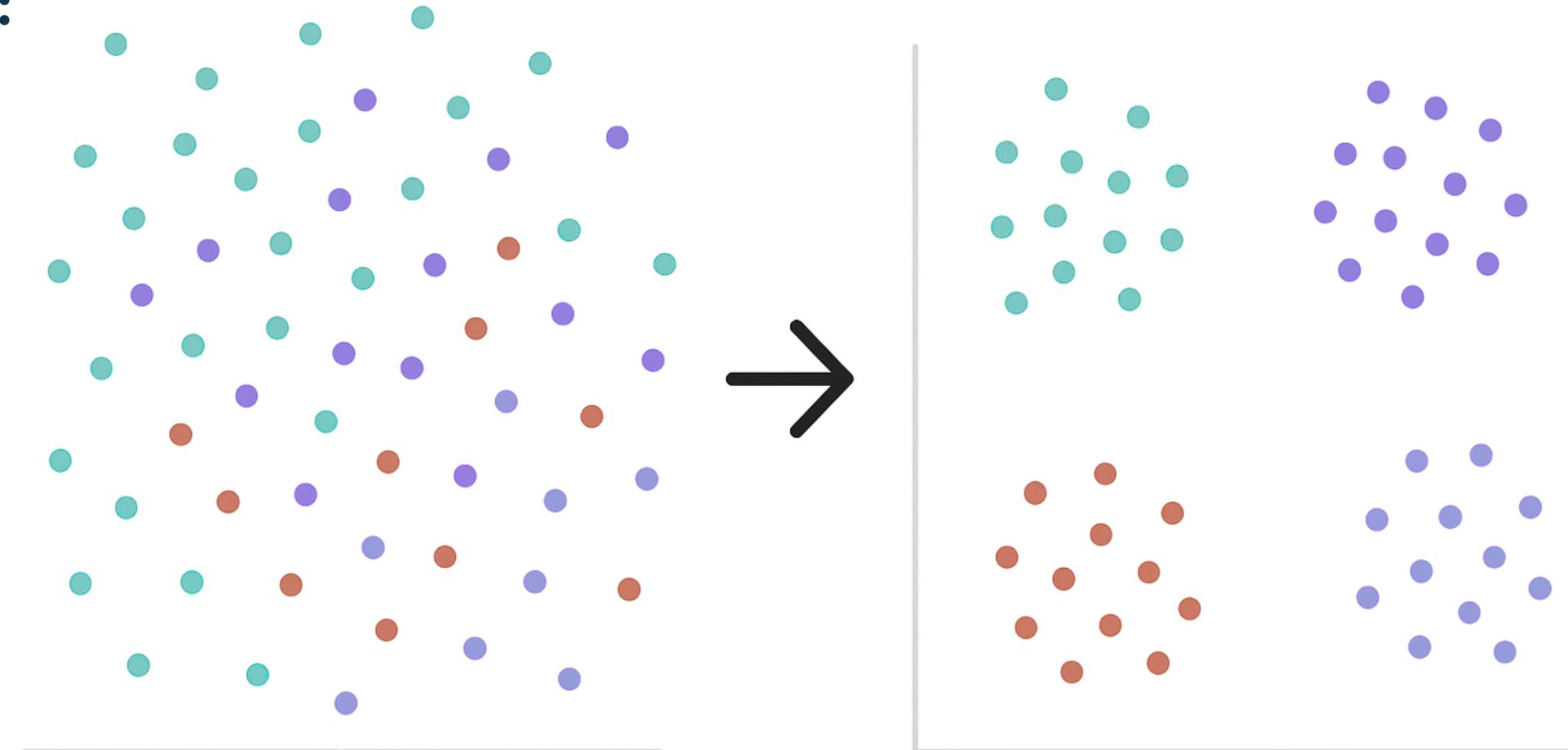
Fragmented care pathways: Patients often see * multiple specialists, leading to siloed assessments and delayed recognition of symptom patterns

Time and data constraints: Clinicians must make * sense of complex, evolving symptoms within limited consultation time and often without comprehensive data integration tools.



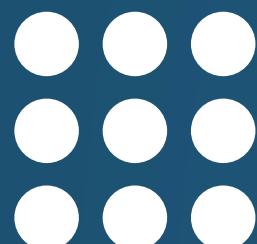
How to use AI?

Introducing Clustering:

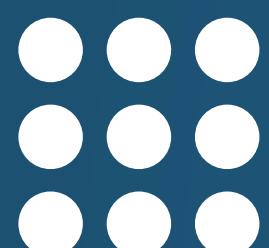
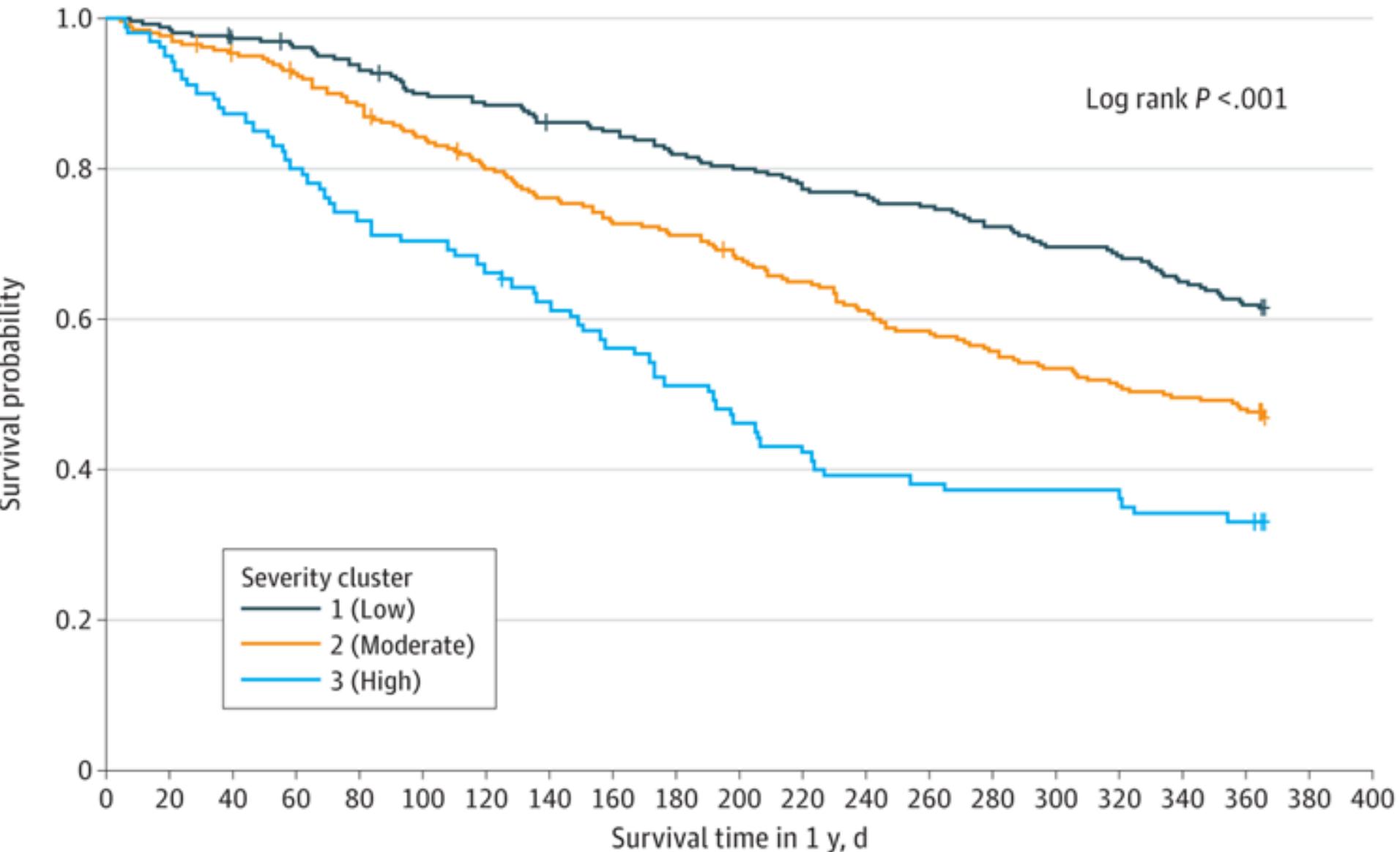
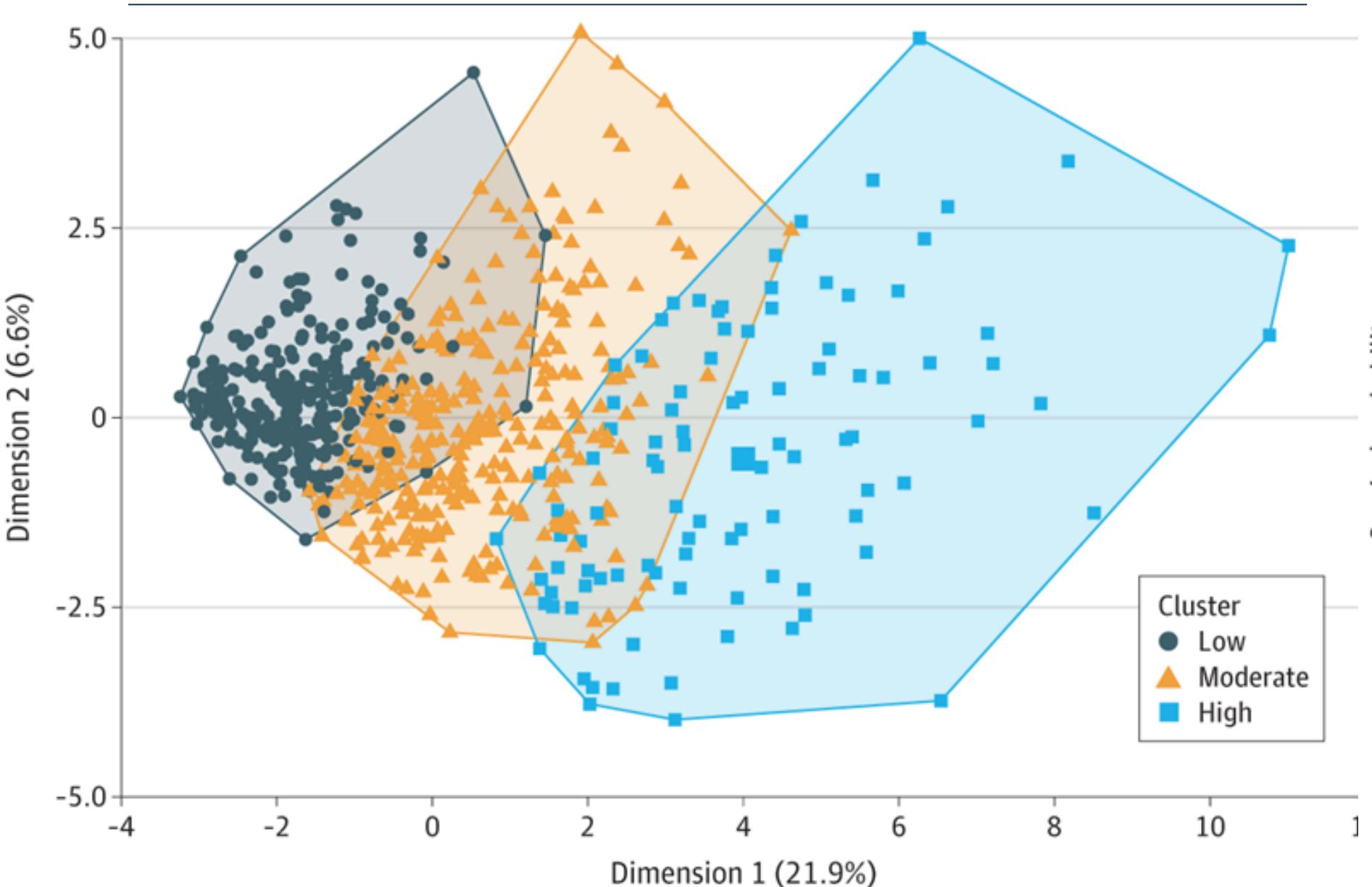


Unclustered Data

Clustered Data



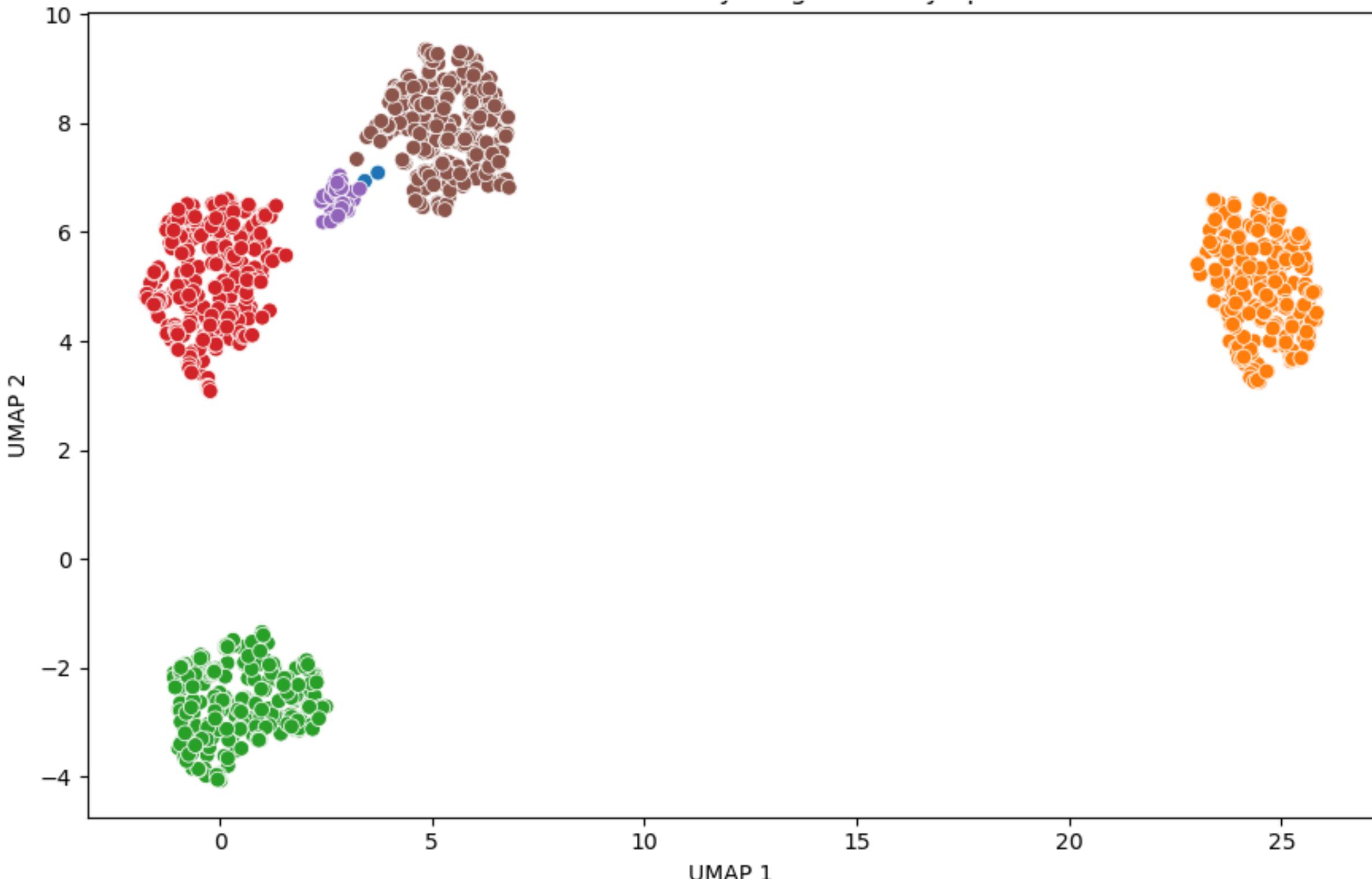
How to use AI?



Input

- Self Reported Symptoms

- Neurocognitive-dominant
- Cardiopulmonary-autonomic
- GI-dysautonomic
- Fatigue-predominant with mood overlap

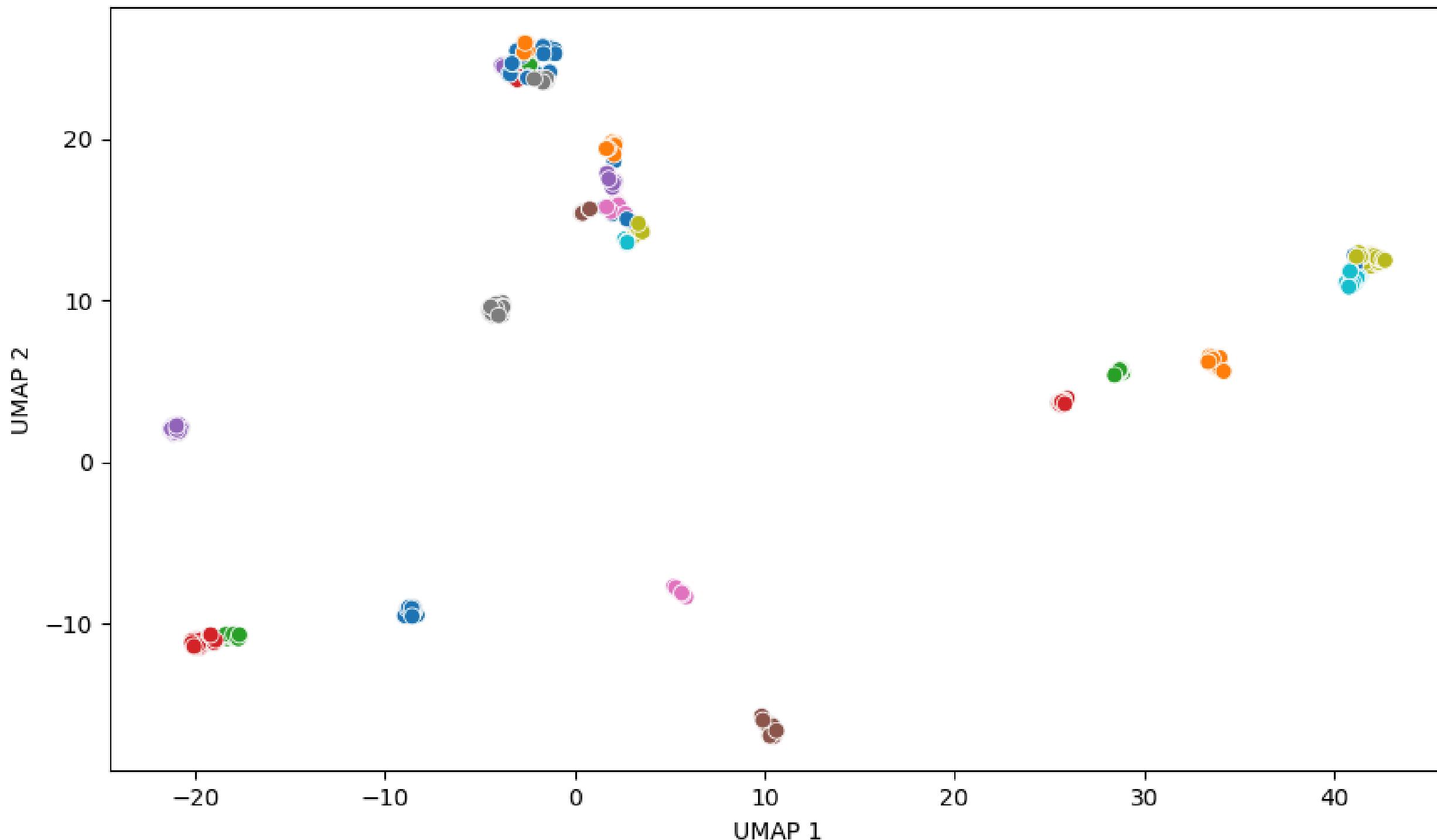


Input

- Self Reported Symptoms
(Co-occurrences,
longitudinally)

Next Steps:

- Determining what these clustered represent
- Adding Demographic Information
- Mapping to covid Strain



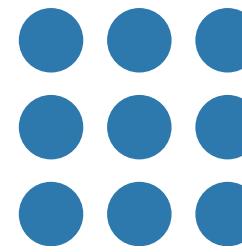
Impact:

Subtype identification: Helps define distinct phenotypes of Long Covid (e.g. neurocognitive, autonomic, gastrointestinal), moving beyond a one-size-fits-all label.

Toward diagnostic criteria: Cluster analysis can inform data-driven diagnostic frameworks, identifying symptom combinations that consistently co-occur.

Personalised interventions: Enables targeted care strategies tailored to each phenotype, improving clinical outcomes and resource allocation.

Task For Workshop:



Task (Group 3)

- Assigned one of five synthetic datasets, each simulating patient-reported Long Covid symptoms.
- Each dataset contains randomly assigned symptom patterns reflecting different underlying phenotypes (e.g. neurocognitive, cardiopulmonary, GI-dominant).
- Use UMAP to reduce dimensionality and visually explore potential cluster structure in the data.
- Apply HDBSCAN to perform unsupervised clustering and identify distinct symptom-based subgroups.

Discussion:

- Compare and interpret the clusters
- How many distinct groups emerge?
- What symptom profiles define them?
- How might they map to clinical care needs?
- How could use use clustering in your work?