Topological Data Analysis of the WISE Study (Women's Ischemia Syndrome Evaluation): Novel Patient Stratification and Mortality Prediction

Osama Dasa, MD, MPH, PhD[1], Himanshu Yadav, MS[2], Francheska A. Castro, MD[1], Mohammad Ruzieh, MD[1], Eileen Handberg, ARNP, PhD[1], and Carl J. Pepine, MD[1]

[1]Division of Cardiovascular Medicine, Department of Medicine, College of Medicine, University of Florida, Gainesville, Florida
[2]Department of Mathematics, College of Liberal Arts & Sciences, University of Florida, Gainesville, Florida

Disclosures: None

1 Background and Rationale

Traditional risk stratification methods in cardiovascular medicine often rely on linear models that may fail to capture the complex, multidimensional relationships inherent in clinical data. Women with ischemic heart disease present unique diagnostic and prognostic challenges, as they frequently exhibit atypical presentations and have been historically underrepresented in cardiovascular research.

2 Introduction

Topological Data Analysis (TDA) represents a paradigm shift in data analysis, offering a novel mathematical framework that can identify hidden patterns and relationships within high-dimensional clinical datasets. Unlike conventional statistical approaches, TDA preserves the geometric structure of data, allowing for the discovery of previously unrecognized patient phenotypes. By analyzing the "shape" of data rather than relying solely on statistical correlations, TDA can reveal clinically meaningful patient subgroups that may be missed by traditional clustering methods.

Our objective was to harness the power of TDA to identify distinct patient subsets within the WISE study cohort and evaluate their association with long-term mortality outcomes, potentially revolutionizing risk stratification approaches for women with suspected ischemic heart disease.

3 Methods

3.1 Study Population

We utilized the comprehensive 10-year follow-up dataset from the landmark WISE study, which included 936 symptomatic women who underwent coronary angiography for suspected ischemic heart disease. This well-characterized cohort provided a robust foundation for our topological analysis.

3.2 Data Processing

A total of 132 clinical variables were systematically analyzed, encompassing demographic characteristics, medical history, laboratory values, medications, procedural data, and cardiovascular risk factors. Prior to TDA implementation, data underwent rigorous preprocessing including normalization, missing value imputation, and outlier detection.

3.3 Topological Data Analysis

We employed the sophisticated Mapper algorithm through BlulightAI's advanced analytics platform. The Mapper algorithm constructs a topological network by creating overlapping neighborhoods in the data space, then clustering data points within each neighborhood. This approach generates a graph representation that captures both local and global structures within the dataset, revealing patient similarity patterns that might be obscured in traditional analysis.

3.4 Statistical Analysis

Following TDA-based patient stratification, we conducted comprehensive survival analysis using Kaplan-Meier curves to assess 10-year mortality differences between identified clusters. Log-rank tests were performed to evaluate statistical significance of survival differences.

4 Results

4.1 Cluster Identification

TDA successfully identified two distinct patient clusters with markedly different clinical profiles and outcomes. The topological network visualization clearly delineated these subgroups, demonstrating the power of geometric data analysis in clinical applications.

4.2 Cluster A Characteristics (n=386)

This subset was predominantly characterized by patients with a history of previous cardiovascular interventions and specific medication profiles. Key distinguishing features included nitrate usage, tamoxifen therapy, prior percutaneous transluminal coronary angioplasty (PTCA), myocardial infarction history, previous coronary artery bypass surgery, specific marital status patterns, and notably, elevated estradiol levels. This cluster appeared to represent a population with established coronary artery disease and more extensive intervention history.

4.3 Cluster B Characteristics (n=433)

The second cluster encompassed patients with a different cardiovascular risk profile, featuring more comorbid conditions and medical complexity. Prominent characteristics included congestive heart failure history, digitalis therapy, nitrate use, corticosteroid treatment, antiplatelet therapy, significant renal dysfunction (serum creatinine > 1.5 mg/dL), and peripheral vascular disease history. This phenotype suggested a more systemically compromised patient population.

4.4 Survival Analysis

Kaplan-Meier survival analysis revealed striking differences in long-term mortality between clusters. Cluster A demonstrated significantly elevated long-term mortality risk compared to Cluster B, highlighting the clinical relevance of TDA-derived patient stratification. The survival curves diverged early and maintained separation throughout the 10-year follow-up period.

5 Clinical Implications and Discussion

These findings represent a significant advancement in personalized cardiovascular medicine for women. The TDA-identified clusters provide clinically actionable insights that could inform treatment decisions and risk management strategies. The higher mortality risk in Cluster A, despite apparent cardiovascular intervention history, suggests that traditional markers of disease severity may not fully capture prognosis in this population.

6 Conclusion

This groundbreaking study demonstrates that Topological Data Analysis can effectively segregate women from the WISE study into two distinct subsets with unique clinical characteristics and significantly different long-term mortality risks. The novel risk stratification approach, informed by the inherent topological structure of clinical data, offers unprecedented insights into patient phenotyping that extend beyond conventional statistical methods.

This innovative methodology represents a paradigm shift toward precision medicine in cardiovascular care, potentially transforming how we identify high-risk patients and tailor therapeutic interventions. However, further validation in diverse patient cohorts and prospective clinical trials will be essential to establish the clinical utility and generalizability of TDA-based risk stratification in routine cardiovascular practice.

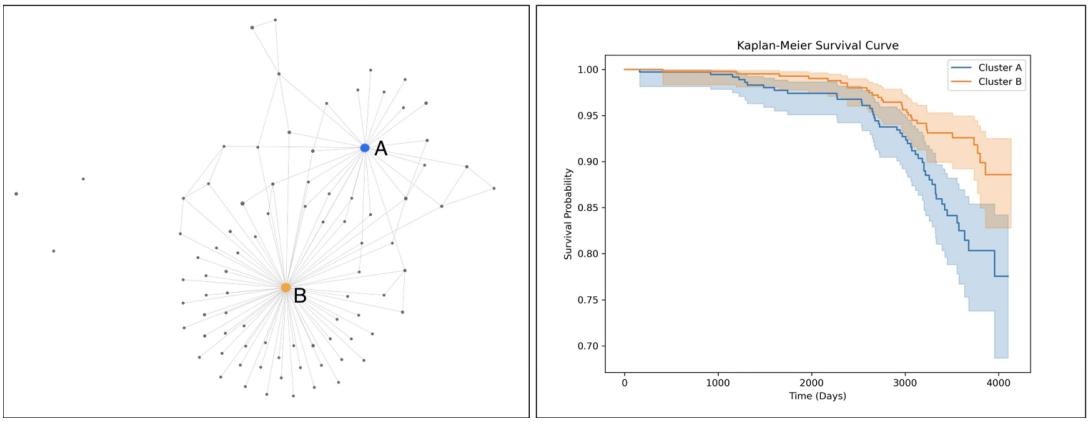

Figure 1: **Figure 1.** Left panel: Topological network graph displaying patient similarity mapper with two primary patient clusters (A and B). Right panel: Kaplan-Meier survival analysis demonstrating differential long-term mortality between the identified clusters.