Multivariate Pattern Analysis of Physical Activity and Cartilage Health: Data from the Osteoarthritis Initiative

Eni Halilaj1*, Trevor J. Hastie2, Garry E. Gold3, Scott L. Delp1
Departments of 1Bioengineering, 2Statistics, and 3Radiology, Stanford University

Introduction

• Data from activity monitoring sensors and cartilage MRI promise to generate new insights on the role of physical activity on OA progression.

• Reducing high-dimensional data to summary metrics and analyzing each outcome in isolation, however, limits the power of these rich data.

• Multivariate methods are more efficient in drawing insights from high-dimensional data where different features are correlated because they take advantage of the data covariance matrix.

• The aim of this study was to show that multivariate methods can identify relationships between physical activity and cartilage health that are missed by common univariate approaches.

Methods

20 subjects with similar characteristics (men, aged 45 – 60 years, BMI of 25 – 27 kg/m², and no radiographic evidence of OA) were selected from the incidence subcohort of the Osteoarthritis Initiative study.

• The relationship between physical activity and femoral cartilage health was tested:
  o using bivariate correlations among activity features \( x_1 \) – \( x_6 \) and cartilage MRI features \( y_1 \) – \( y_6 \)
  o using canonical correlation\(^2\), a multivariate analysis that seeks a linear combination of activity features \( x_1 \) – \( x_6 \) that is highly correlated with a linear combination of cartilage MRI features \( y_1 \) – \( y_6 \). This was done using canonical correlation between activity features \( x \) and cartilage MRI features \( y \). The canonical correlation is used to identify a linear combination of activity features \( x^* \) and cartilage MRI features \( y^* \) that are highly correlated.

• A permutations test was used to determine if the canonical correlation was statistically significant.

Results

• The mean (± SD) age, BMI, and amount of daily activity were 52 ± 4 years, 25.9 ± 1.2 kg/m², and 331,410 ± 102,810 counts, respectively.

• Bivariate correlation analyses yielded no significant associations among times spent in different activity intensity ranges and microstructural changes in different regions of cartilage \( r < .33, p > .17 \).

• Canonical correlation analysis identified a linear combination of times spent in different activity ranges that was significantly associated with a linear combination of \( T_2 \) changes in different cartilage regions.

• The ranges of activity with the strongest contribution were those encompassing sedentary time and vigorous activity \( x_1 \) and \( x_6 \).

• The medial weight-bearing region of femoral cartilage \( y_1 \) was affected the most by the identified combination of activities.

Conclusion

In this proof-of-concept study we showed that multivariate methods are more efficient than univariate ones in identifying associations between physical activity and cartilage health. Adoption of these methods, as well as more advanced machine learning algorithms, has the potential to generate new insight from large, high-dimensional data on OA.

References

1. 1. https://oai.epi-ucsf.org/

Acknowledgements

This work was supported by the Mobilize Center, a Big Data to Knowledge (BD2K) Center funded by NIH Grant 54EB020405.

The OAI is a public-private partnership comprised of five contracts funded by the National Institutes of Health. Private funding partners include Merck Research Laboratories; Novartis Pharmaceuticals Corporation, GlaxoSmithKline; and Pfizer, Inc.

*Correspondence: ehalilaj@stanford.edu