



STUDYING GENERALIZABLE MULTIVARIATE ASSOCIATIONS IN BIOLOGICAL SCIENCES: BRAIN STRUCTURE, BEHAVIOR AND PHYSICAL HEALTH

8/05/2025 ELIANA NICOLAISEN-SOBESKY, LEONARD SASSE,
DR. KAUSTUBH R. PATIL & PROF. DR. SARAH GENON

MEDICAL NEUROIMAGING



MEDICAL NEUROIMAGING: URUGUAY



Training:

- Licenciatura en Biología Humana
- Maestría en Ciencias Cognitivas
- Supervisor: PhD. Victoria Gradin
- Co-supervisor: PhD. Álvaro Cabana

Projects:

- Facultad de Psicología
 - Major depression, social anxiety, self-esteem
 - Electroencephalography
 - Neuroeconomics (social interactions, decision making)
- Facultad de Medicina (Dr. Daniel Olazábal)
 - Postpartum depression, oxytocine
- IIBCE Dr. Ángel Caputi
 - Electroencephalography, cognition



ORIGINAL ARTICLE

PSYCHOPHYSIOLOGY SPR WILEY

Event-related potentials during the ultimatum game in people with symptoms of depression and/or social anxiety

Eliana Nicolaisen-Sobesky^{1,2} | Valentina Paz^{1,3,4} | Francisco Cervantes-Constantino¹ | Eduardo Martínez-Montes⁵ | B. Gradin^{1,3}

ORIGINAL ARTICLE

PSYCHOPHYSIOLOGY SPR WILEY

Event-related potentials of social comparisons in depression and social anxiety

Valentina Paz^{1,2,3} | Eliana Nicolaisen-Sobesky^{1,4} | Gabriela Fernández-Theoduloz^{1,2} | Alfonso Pérez¹ | Francisco Cervantes Constantino¹ | Eduardo Martínez-Montes⁵ | Dominique Kessel⁶ | Álvaro Cabana^{1,3} | Victoria B. Gradin^{1,2}



Effect of self-esteem on social interactions during the Ultimatum Game

V. Paz, E. Nicolaisen-Sobesky, E. Collado, S. Horta, C. Rey, M. Rivero, P. Berriolo, M. Díaz, M. Otón, A. Pérez, G. Fernández-Theoduloz, Á. Cabana, V.B. Gradin^{*}
CIBPSi, Faculty of Psychology, Universidad de la República, Uruguay

Received: 2 August 2022 | Revised: 12 April 2023 | Accepted: 27 April 2023
DOI: 10.1002/dev.22400

RESEARCH ARTICLE

Developmental Psychobiology WILEY

Oxytocin system polymorphisms rs237887 and rs2740210 variants increase the risk of depression in pregnant women with early abuse

Olazábal Daniel Ernesto¹ | Bertoni Bernardo² | Grandi Graciela³ | Musetti Dora⁴ | Rey Grazzia⁵ | Sandberg Natalia¹ | Fernández Lucia¹ | Laporte Gabriela⁵ | Medici Florencia⁵ | Nicolaisen-Sobesky Eliana^{1,4}



Social Avoidance in Depression: A Study Using a Social Decision-Making Task

Gabriela Fernández-Theoduloz, Valentina Paz,
Eliana Nicolaisen-Sobesky, and Alfonso Pérez
Universidad de la República

Abraham P. Buunk
University of Groningen and Netherlands Interdisciplinary
Demographic Institute, Hague, the Netherlands

Álvaro Cabana and Victoria B. Gradin
Universidad de la República

Neural processing of iterated prisoner's dilemma outcomes indicates next-round choice and speed to reciprocate cooperation

Francisco Cervantes Constantino¹, Santiago Garat⁶, Eliana Nicolaisen-Sobesky^{1,2}, Valentina Paz³, Eduardo Martínez-Montes⁵, Dominique Kessel⁶, Álvaro Cabana^{1,3} and Victoria B. Gradin¹



MEDICAL NEUROIMAGING: GERMANY

communications biology

ARTICLE

<https://doi.org/10.1038/s42003-022-04244-5> **OPEN**

A cross-cohort replicable and heritable latent dimension linking behaviour to multi-featured brain structure

Eliana Nicolaisen-Sobesky^{1,8}, Agoston Mihalik^{2,3,4}, Shahrzad Kharabian-Masouleh^{1,5}, Fabio S. Ferreira^{2,3}, Felix Hoffstaedter^{1,5}, Holger Schwender⁶, Somayeh Maleki Balajoo^{1,5}, Sofie L. Valk^{1,5,7}, Simon B. Eickhoff^{1,5}, B. T. Thomas Yeo⁸, Janaina Mourao-Miranda^{2,3} & Sarah Genon^{1,5,8}

Cardiometabolic health, cortical thickness, and neurotrans systems: a large-scale multivariate study

Eliana Nicolaisen-Sobesky^{1,2*}, Somayeh Maleki Balajoo^{1,2}, Mostafa Mahdipour¹, Agoston Mihalik^{3,4,5}, Felix Hoffstaedter^{1,2}, Janaina Mourao-Miranda^{3,4}, Masoud Tahmasian^{1,2,6}, Simon B. Eickhoff^{1,2}, Sarah Genon^{1,2*}

On Leakage in Machine Learning Pipelines

Sasse, L.^{*1,2,3}, Nicolaisen-Sobesky, E.^{*1,2}, Dukart, J.^{1,2}, Eickhoff^{1,2}, S.B.; Götz, M.^{4,5}; Hamdan, S.^{1,2}; Komeyer, V.^{1,2,6}; Kulkarni, A.⁷; Lahnakoski, J.^{1,2}; Love, B.C.^{8,9,10}; Raimondo, F.^{1,2}; Patil, K.R.^{1,2,11}

hhu Heinrich Heine
Universität
Düsseldorf



JÜLICH
Forschungszentrum



Training:

- PhD in Medical Sciences (ongoing)
- Supervisor: Prof. Dr. Sarah Genon
- Posdoc (ongoing)
- Supervisor: Dr. Kaustubh R. Patil

Projects:

- Research Center Jülich & Heinrich-Heine Universität Düsseldorf
- Multivariate associations between:
 - brain structure and behavior
 - brain structure and risk factors
- Replicability of brain-behavior associations
- Leakage in machine learning

RESEARCH ARTICLE **OPEN ACCESS**

Discovery, Replicability, and Generalizability of a Left Anterior Hippocampus' Morphological Network Linked to Self-Regulation

Somayeh Maleki Balajoo^{1,2} | Anna Placht^{1,2} | Eliana Nicolaisen-Sobesky^{1,2} | Debo Dong^{2,3} | Felix Hoffstaedter^{1,2} | Sven G. Meuth⁴ | Nico Melzer⁵ | Simon B. Eickhoff^{1,2} | Sarah Genon^{1,2}

Lower motor performance is linked with poor sleep quality, depressive symptoms, and grey matter volume alterations

^{1,2}, Hanwen Bi^{3,4}, Eliana Nicolaisen-Sobesky^{2,3}, Felix Hoffstaedter^{2,3}, B.T. Thomas Orzegg^{1,6,11}, Simon B. Eickhoff^{2,3}, Masoud Tahmasian^{1,2,3*}

scientific reports

OPEN Empirical facts from search for replicable associations between cortical thickness and psychometric variables in healthy adults

Shahrzad Kharabian Masouleh^{1,11}, Simon B. Eickhoff^{1,11}, Somayeh Maleki Balajoo^{1,4}, Eliana Nicolaisen-Sobesky^{1,4}, Bertrand Thirion⁷ & Sarah Genon^{1,11}

GENERALIZABLE MULTIVARIATE ASSOCIATIONS

Multivariate associations:

- Canonical Correlation Analysis (CCA)



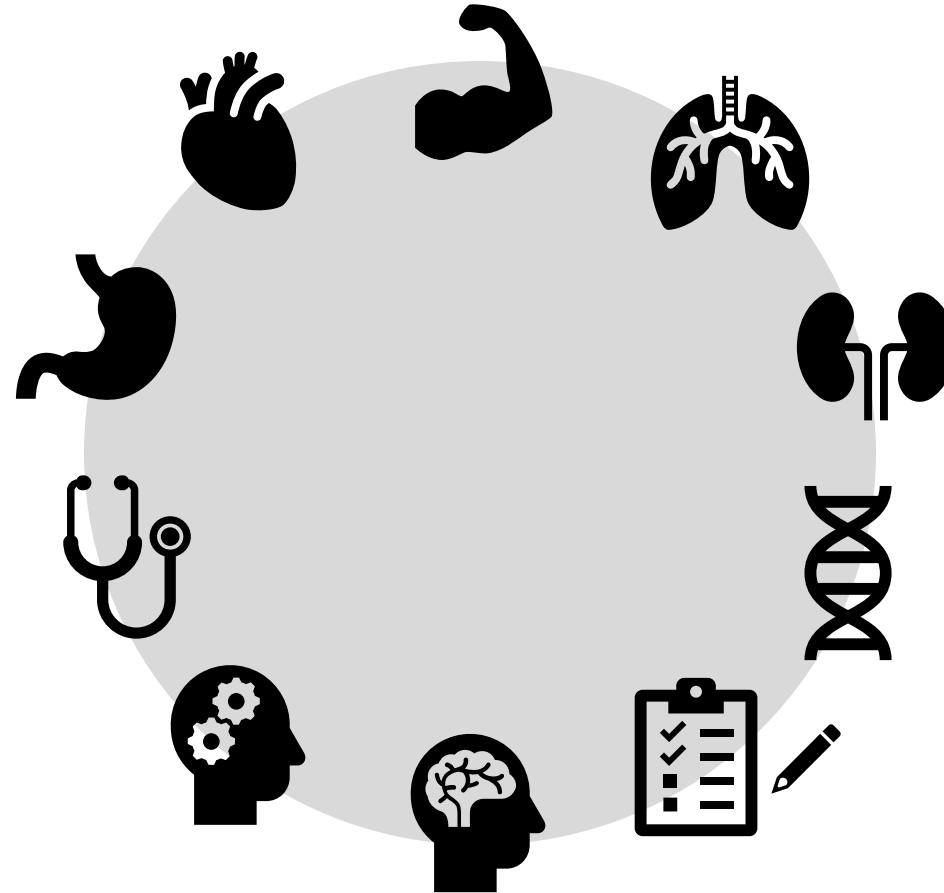
Generalizability:

- Machine learning
- Leakage



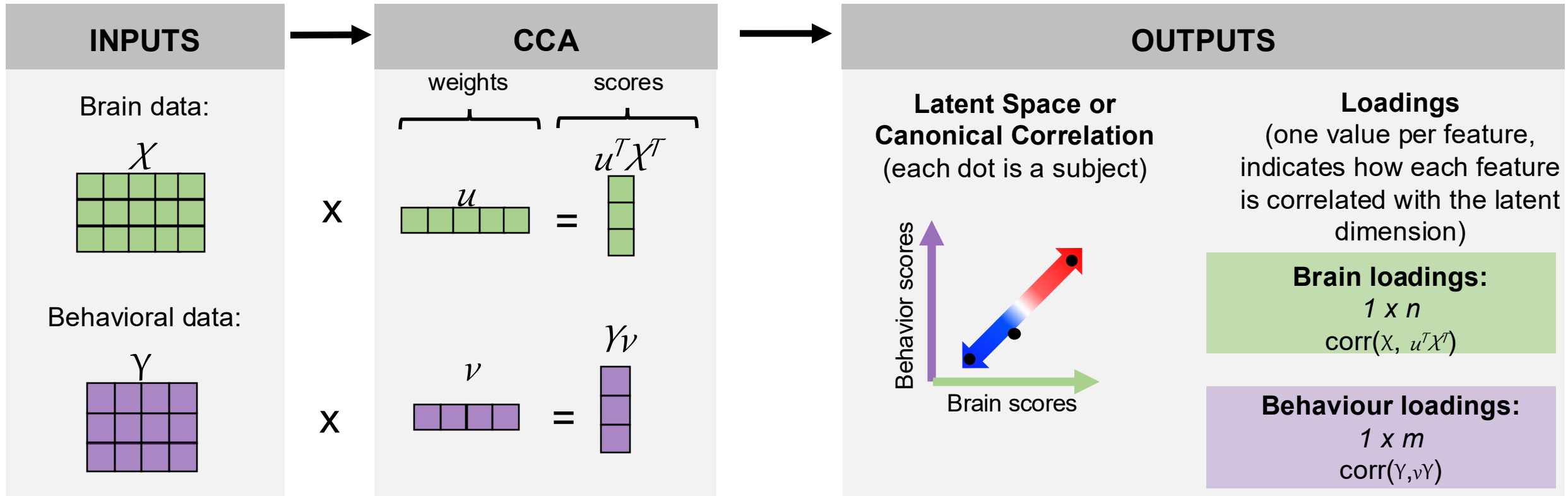
MULTIVARIATE ASSOCIATIONS IN BIOLOGY

They are everywhere and can obscure interpretation of results



CANONICAL CORRELATION ANALYSIS (CCA)

To study multivariate associations between two sets of data



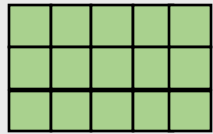
BRAIN STRUCTURE AND BEHAVIOR

Data and Methods

Sample: Human Connectome Project Young Adult (n=1047, 560 females, Age=28.78 ± 3.67)

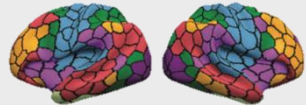


X

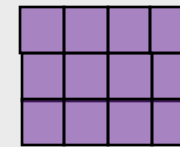


➤ **Brain data** (673 features):

- X matrix concatenating:
 - Grey matter volume (GMV, 273 regions)
 - Cortical thickness (CT, 200 regions)
 - Surface area (SA, 200 regions)
- Schaefer parcellation



Y



➤ **Behavioural data**
(60 features):

- Y matrix with features spanning:
 - Cognition
 - Emotion
 - Mental health

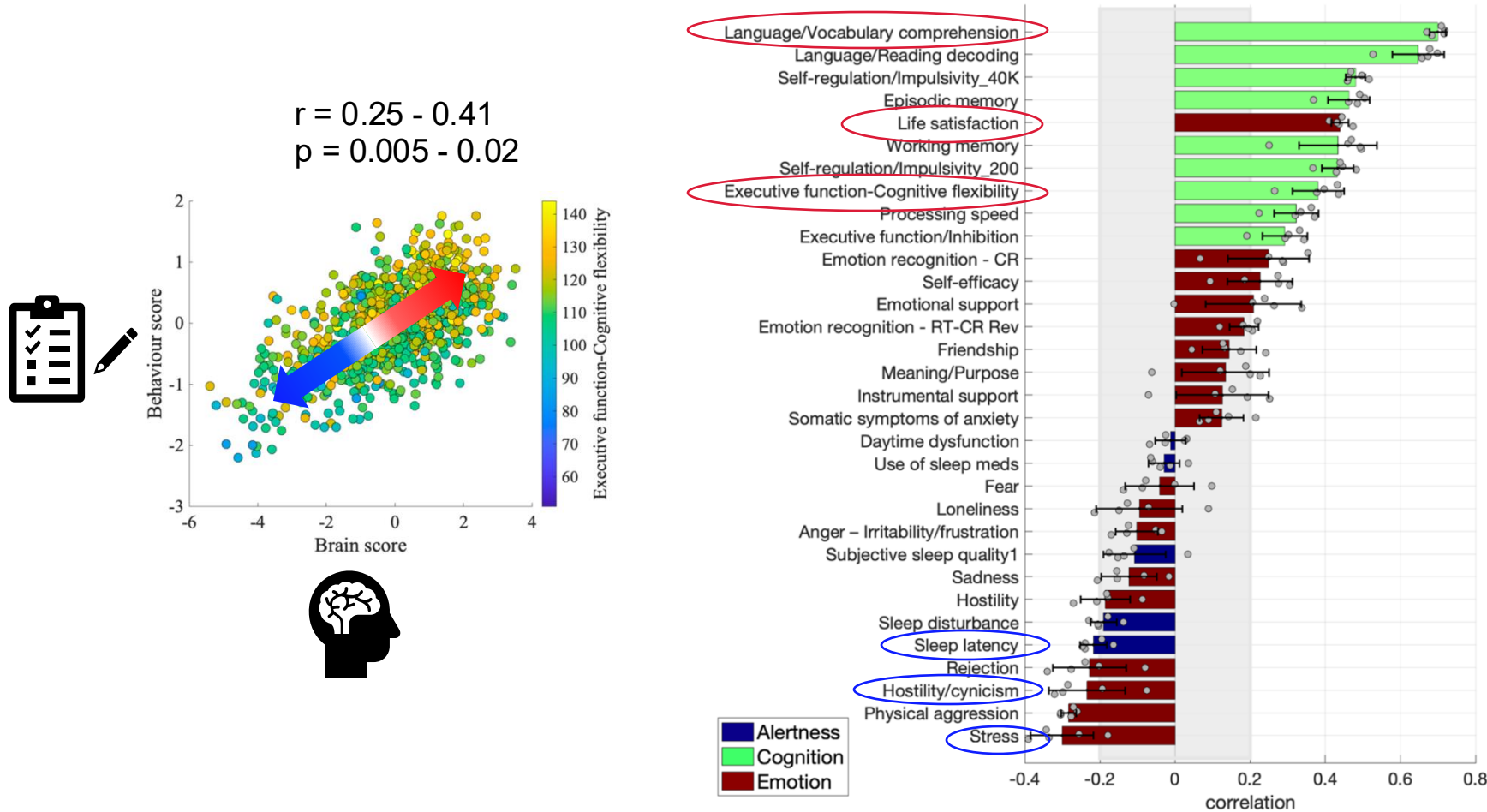


- CCA was combined with Machine Learning
- Age and gender were regressed out from X and Y.
- Brain data was normalised within subjects by brain size

Nicolaisen-Sobesky, ... Genon, 2022 Communications Biology.

BRAIN STRUCTURE AND BEHAVIOR

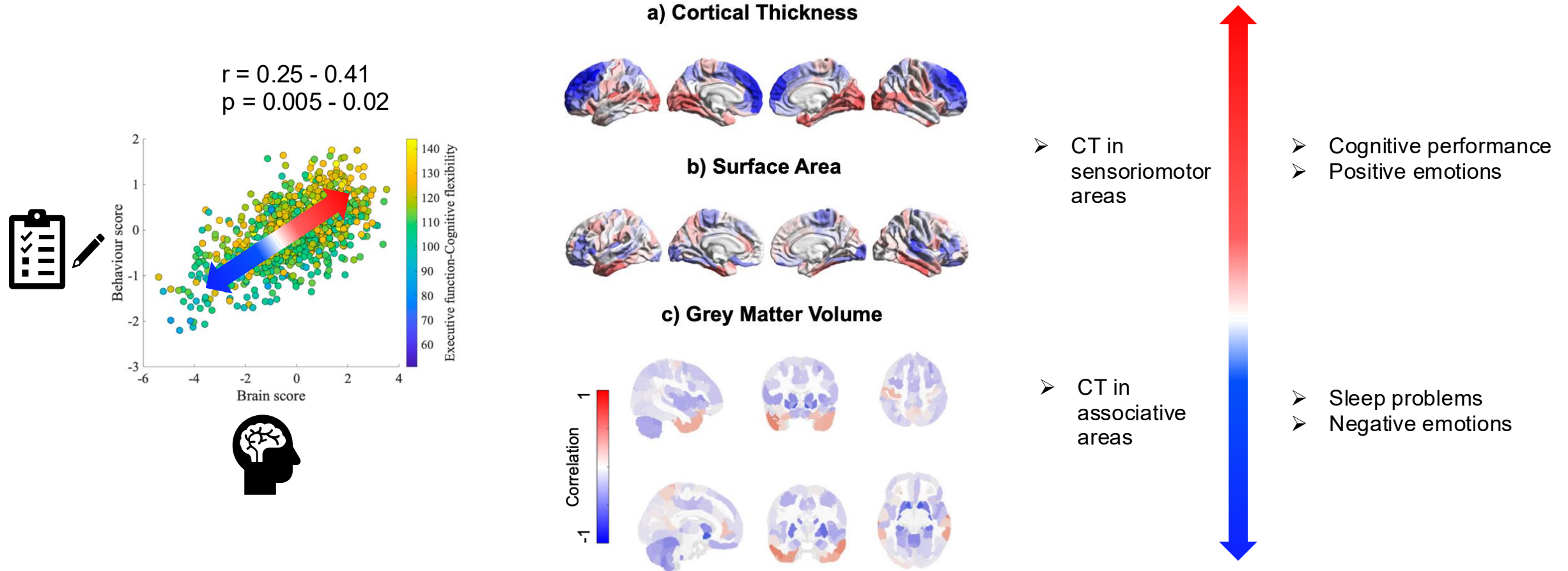
Results: Behavioral side of the latent dimension



Nicolaisen-Sobesky, ... Genon, 2022 Communications Biology.

BRAIN STRUCTURE AND BEHAVIOR

Results: Brain side of the latent dimension



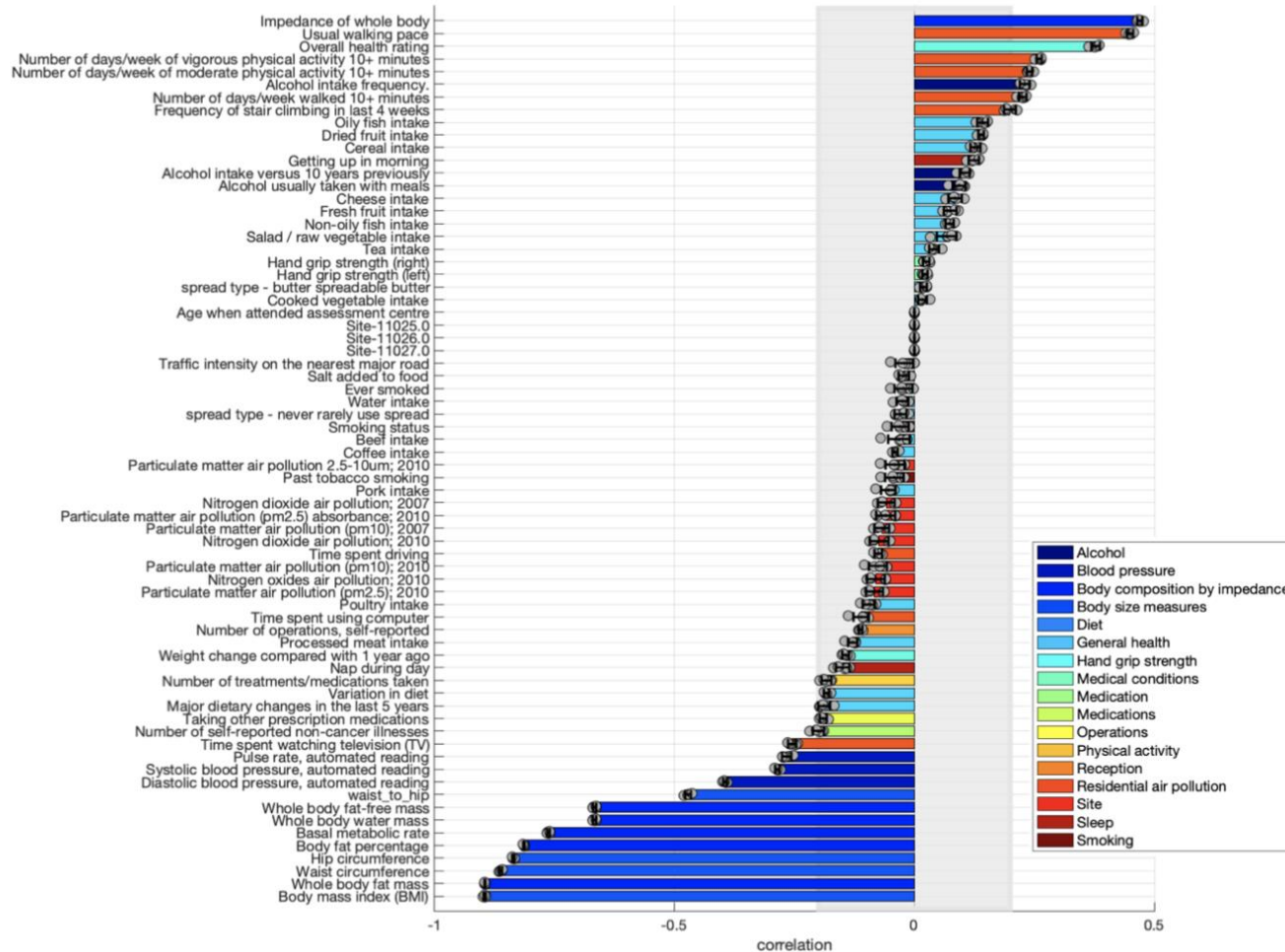
Nicolaisen-Sobesky, ... Genon, 2022 Communications Biology.

BRAIN STRUCTURE VS PHYSICAL HEALTH

UK Biobank (n=7370, 3685 females, Age=62.3 ± 7.6)

Cardiometabolic health, cortical thickness, and neurotransmitter systems: a large-scale multivariate study

Eliana Nicolaisen-Sobesky^{1,2*}, Somayeh Maleki Balajoo^{1,2}, Mostafa Mahdipour¹, Agoston Mihalik^{3,4,5}, Felix Hoffstaedter^{1,2}, Janaina Mourao-Miranda^{3,4}, Masoud Tahmasian^{1,2,6}, Simon B. Eickhoff^{1,2}, Sarah Genon^{1,2*}



- 

➤ Higher physical activity



➤ CT in:

 - Insula
 - Temporal lobe
 - Anterior cingulate cortex
- 
- Body Mass Index

➤ Basal metabolic rate

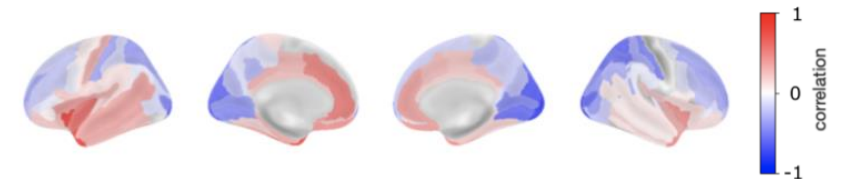
➤ Body fat percentage

➤ Hip circumference

➤ Waist circumference

➤ CT in:

 - Occipital pole
 - Superior parietal areas
 - Superior frontal areas



Nicolaisen-Sobesky, ... Genon, 2025 bioRxiv. Under review in PLOS Biology.

GENERALIZABLE MULTIVARIATE ASSOCIATIONS

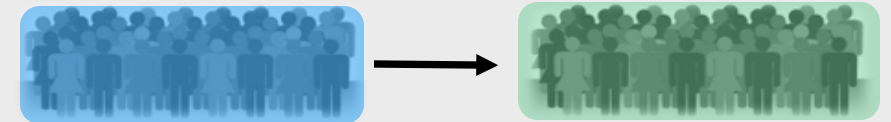
Multivariate associations:

- Canonical Correlation Analysis



Generalizability:

- Machine learning
- Leakage



WHAT IS GENERALIZABILITY?

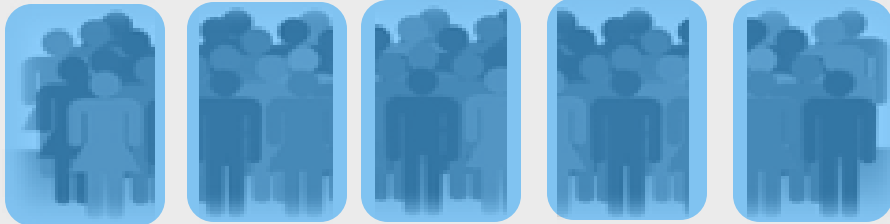
And cross-validation



Sample 1
(the sample we have now and we can analyse)



Sample 2
(for instance, from the future, another age range, another country, etc)



Fold 1

Fold 2

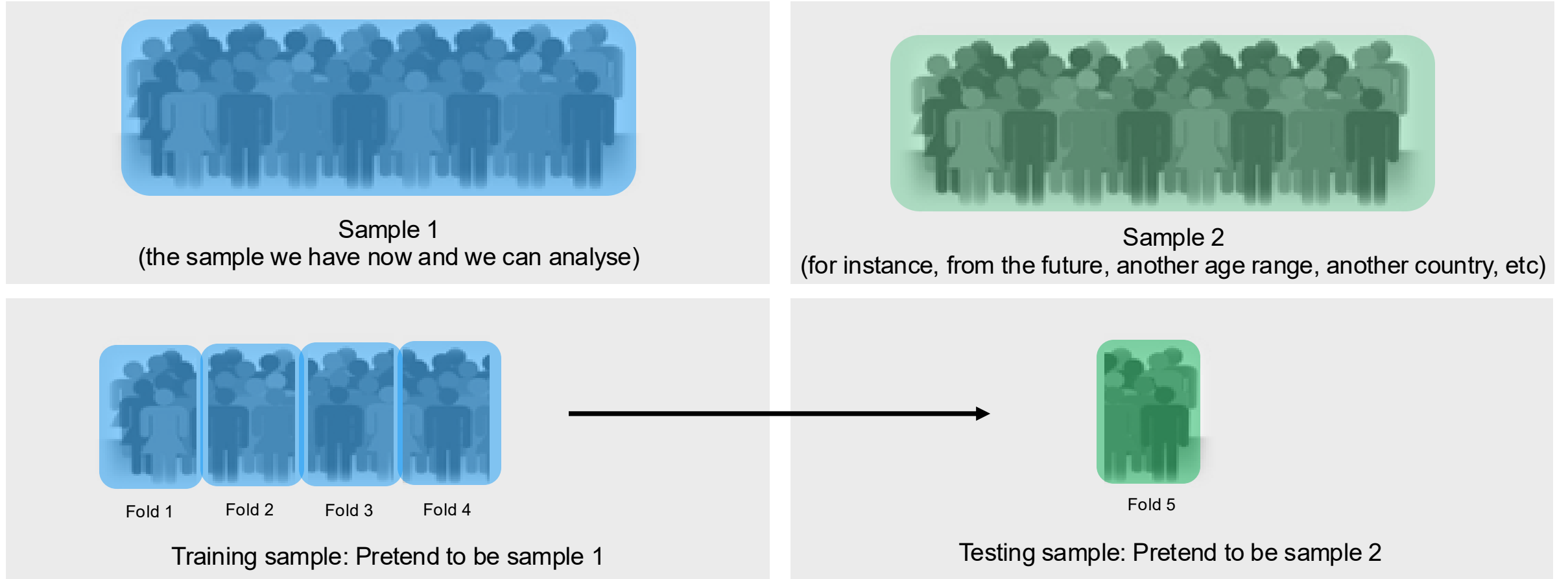
Fold 3

Fold 4

Fold 5

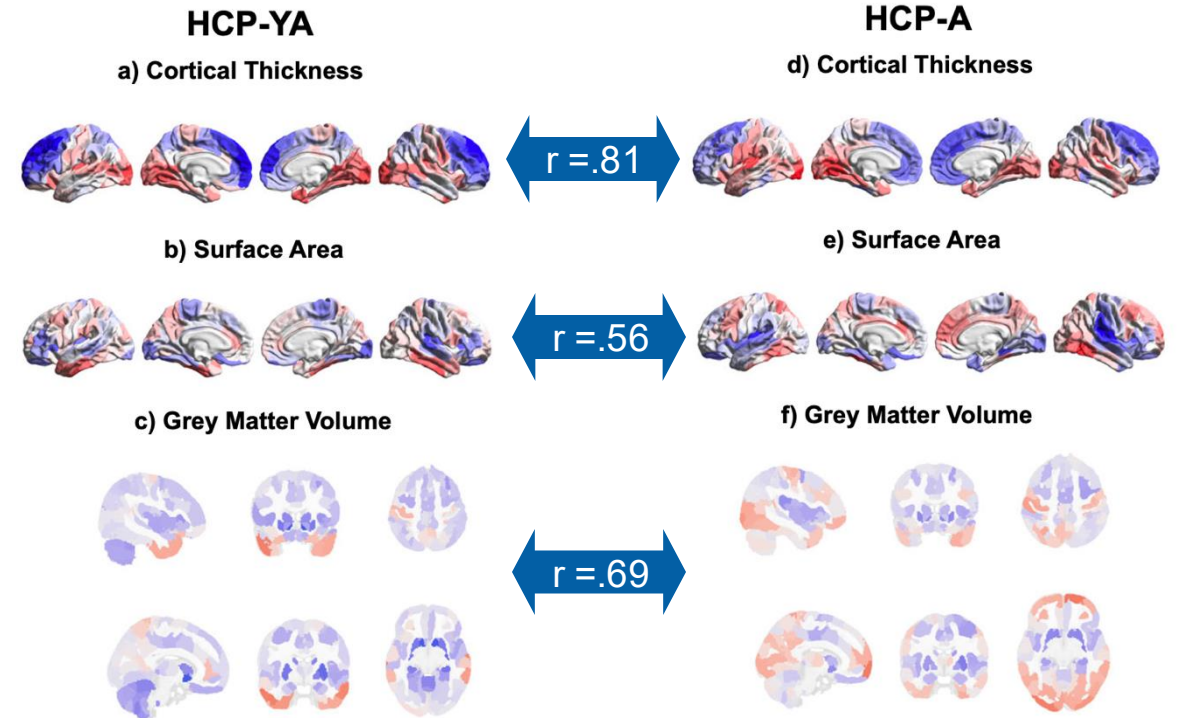
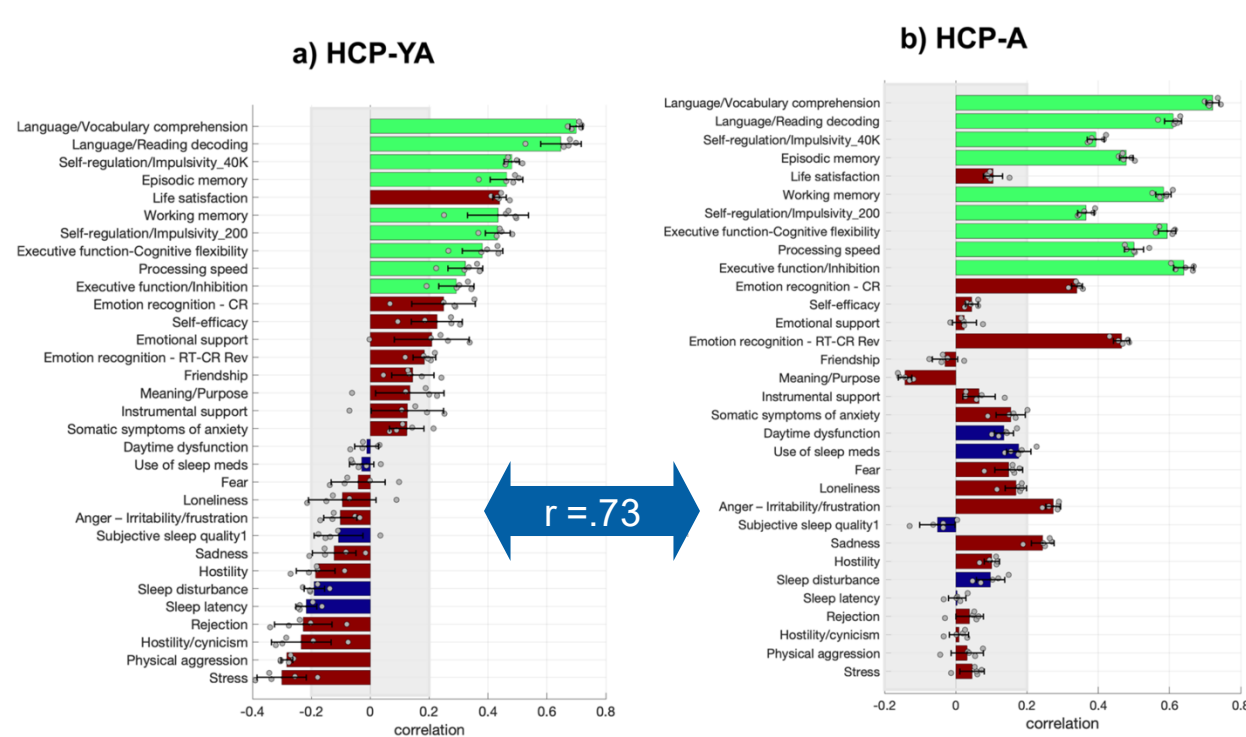
WHAT IS GENERALIZABILITY?

And cross-validation



THE LATENT DIMENSION REPLICATES ACROSS SAMPLES

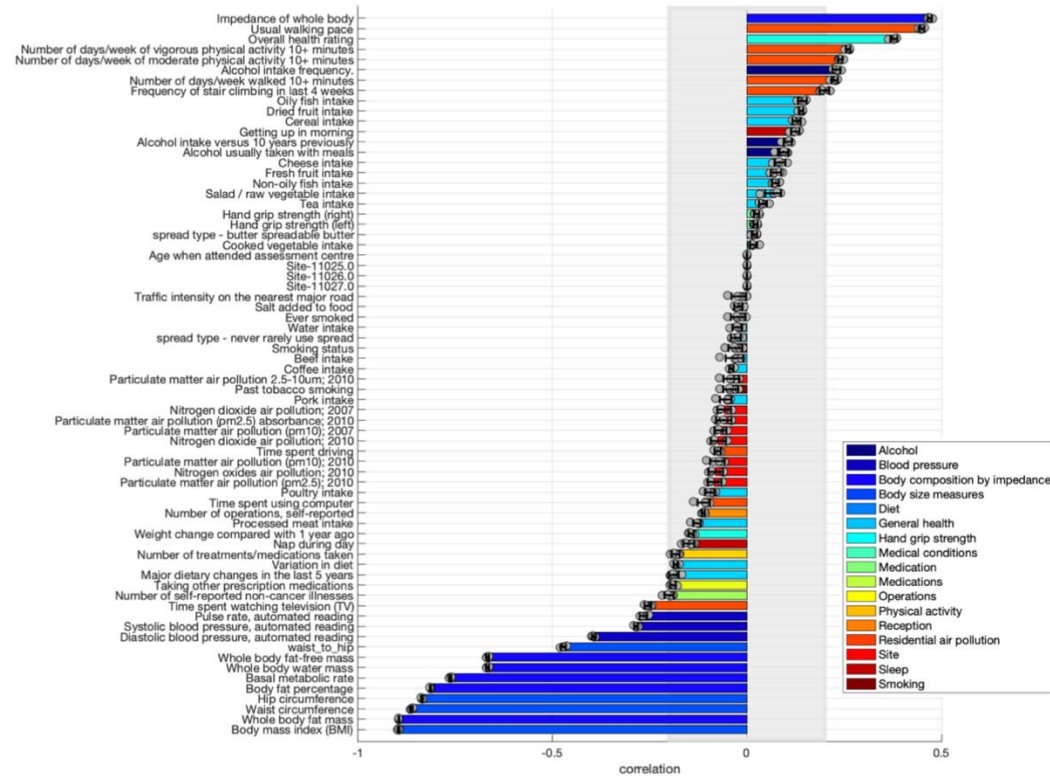
With different age ranges



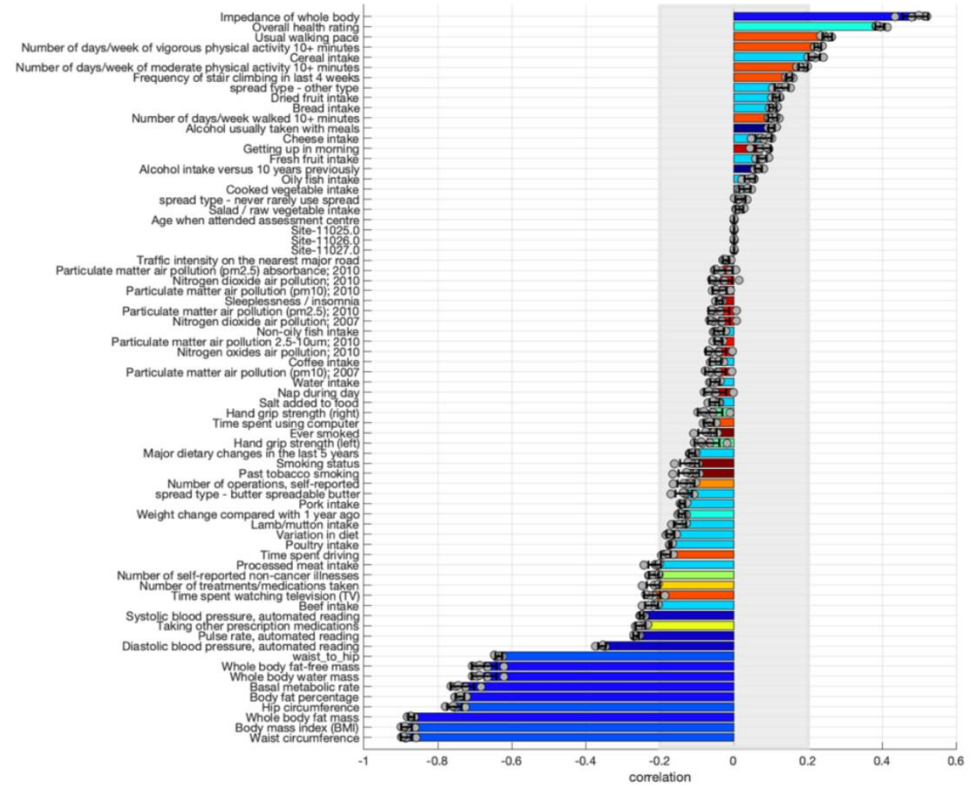
Nicolaisen-Sobesky, ... Genon, 2022 Communications Biology.

THE LATENT DIMENSION REPLICATES ACROSS SEXES

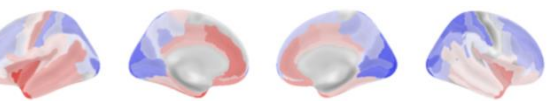
Women - raw Cortical Thickness - LD1



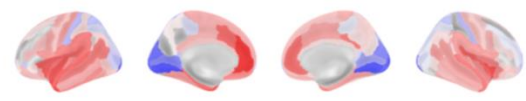
Men - raw Cortical Thickness - LD1



$r = .99$

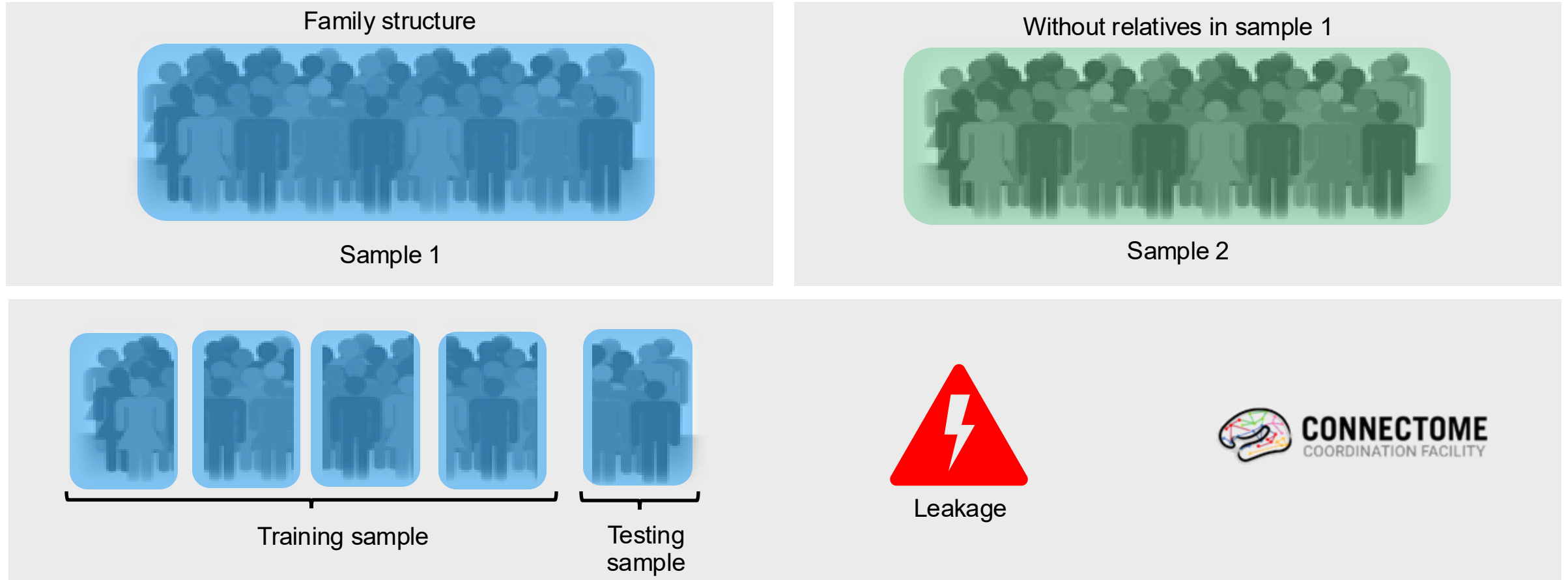


$r = .88$



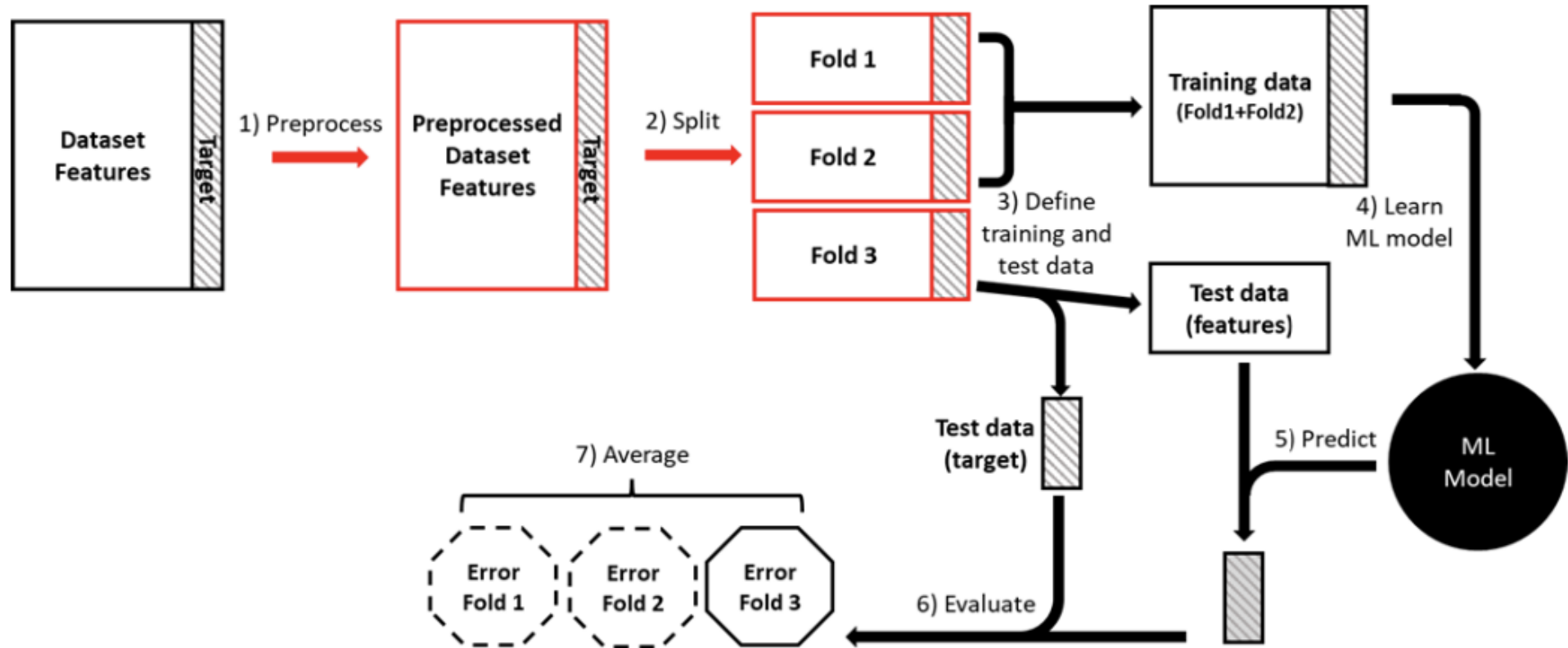
Nicolaisen-Sobesky, ... Genon, 2025 bioRxiv. Under review in PLOS Biology.

THREAT TO GENERALIZABILITY: LEAKAGE



Sasse, Nicolaisen-Sobesky, ... Patil, 2024 arXiv. Accepted in Journal of Big Data.

TEST-TO-TRAIN LEAKAGE



Sasse, Nicolaisen-Sobesky, ... Patil, 2024 arXiv. Accepted in Journal of Big Data.

HOW TO AVOID LEAKAGE

- Ensure **strict training-test set separation**.
- Ensure that performance metrics are calculated on **truly unseen** data that has not been used anywhere in the pipeline previously.
- More recommendations: Sasse, Nicolaisen-Sobesky et al., 2024 arXiv

On Leakage in Machine Learning Pipelines

Sasse, L.^{*1,2,3}; Nicolaisen-Sobesky, E.^{*1,2}; Dukart, J.^{1,2}; Eickhoff^{1,2}, S.B.; Götz, M.^{4,5};
Hamdan, S.^{1,2}; Komeyer, V.^{1,2,6}; Kulkarni, A.⁷; Lahnakoski, J.^{1,2}; Love, B.C.^{8,9,10};
Raimondo, F.^{1,2}; Patil, K.R.^{1,2,11}

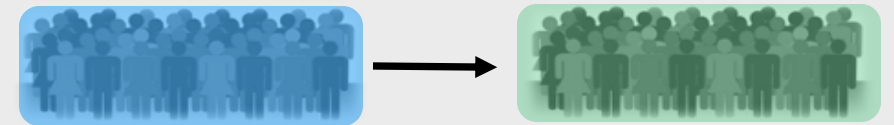
Sasse, Nicolaisen-Sobesky, ... Patil, 2024 arXiv. Accepted in Journal of Big Data.

CONCLUSIONS

- **Multivariate** effects in biology are common and they can be studied with **Canonical Correlation Analysis (CCA)**.
- Using CCA we identified latent dimensions linking:
 - Brain structure to behavior
 - Brain structure to risk factors for non-communicable diseases



- The appropriate use of **Machine Learning (ML)** pipelines improves the generalizability of the findings.
 - **Leakage** is a threat.



- I encourage researchers to use ML and CCA (combined or not) in biological sciences.

REFLEXIONES



- **Contactos:**

- Conocer investigadores en el Instituto donde trabajo y colaboradores de mis tutores
- Facilita oportunidades
- Canales de ofertas de trabajo (linkedin, etc)



- Oportunidad para **aprender:**

- Técnicas, métodos o temáticas
- Formas de hacer investigación



- Trabajar en el exterior es una **experiencia muy linda** pero que viene con **dificultades**.
- **No pienso que ir al exterior sea necesario** para tener una carrera exitosa.



- **La formación en ciencia en Uruguay es muy buena.**
- La **formación interdisciplinaria** que me brindaron **la Licenciatura en Biología Humana** y la **Maestría en Ciencias Cognitivas** me facilita trabajar en un ambiente interdisciplinario con profesionales de la ingeniería, medicina, filosofía, psicología, ciencias de la computación...

THANK YOU

Simposio IBRO-SNU Vuelta a Casa



Applied Machine Learning PI: Dr. Kaustubh R. Patil

Düsseldorf/Jülich (Germany)

Leonard Sasse

Prof. Dr. Jürgen Dukart

Prof. Dr. Simon Eickhoff

Sami Hamdan

Vera Komeyer

Dr. Federico Raimondo

Dr. Nicolás Nieto

Dr. Juha Lahnakoski

Ulm (Germany)

Prof. Dr. Michael Götz

Principal Global Services (India)

Dr. Abhijit Kulkarni

UCL (UK)

Prof. Dr. Bradley C Love



Cognitive and Clinical Neurosciences PI: Prof. Dr. Sarah Genon

Düsseldorf/Jülich (Germany)

Prof. Dr. Simon Eickhoff

Dr. Somayeh Maleki Balajoo

Prof. Dr. Holger Schwender

Dr. Shahrzad Kharabian-Masouleh

PD Dr. Masoud Tahmasian

Mostafa Mahdipour

Vincent Küppers

Dr. Felix Hoffstaedter

UCL (UK)

Prof. Dr. Janaina Mourao-Miranda

Dr. Agoston Mihalik

Dr. Fabio S Ferreira

Beheshti University (Iran)

Mahnaz Olfati

Singapore (Singapore)

Prof. Dr. B. T. Thomas Yeo

Leipzig (Germany)

Dr. Sofie Valk

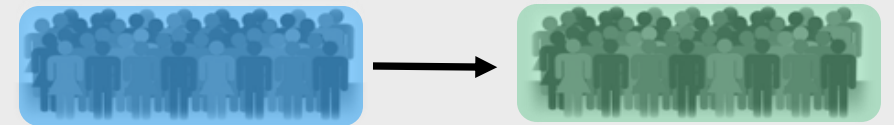


CONCLUSIONS

- **Multivariate** effects in biology are common and they can be studied with **Canonical Correlation Analysis (CCA)**.
- Using CCA we identified axes latent dimensions linking:
 - Brain structure to behavior
 - Brain structure to risk factors for non-communicable diseases



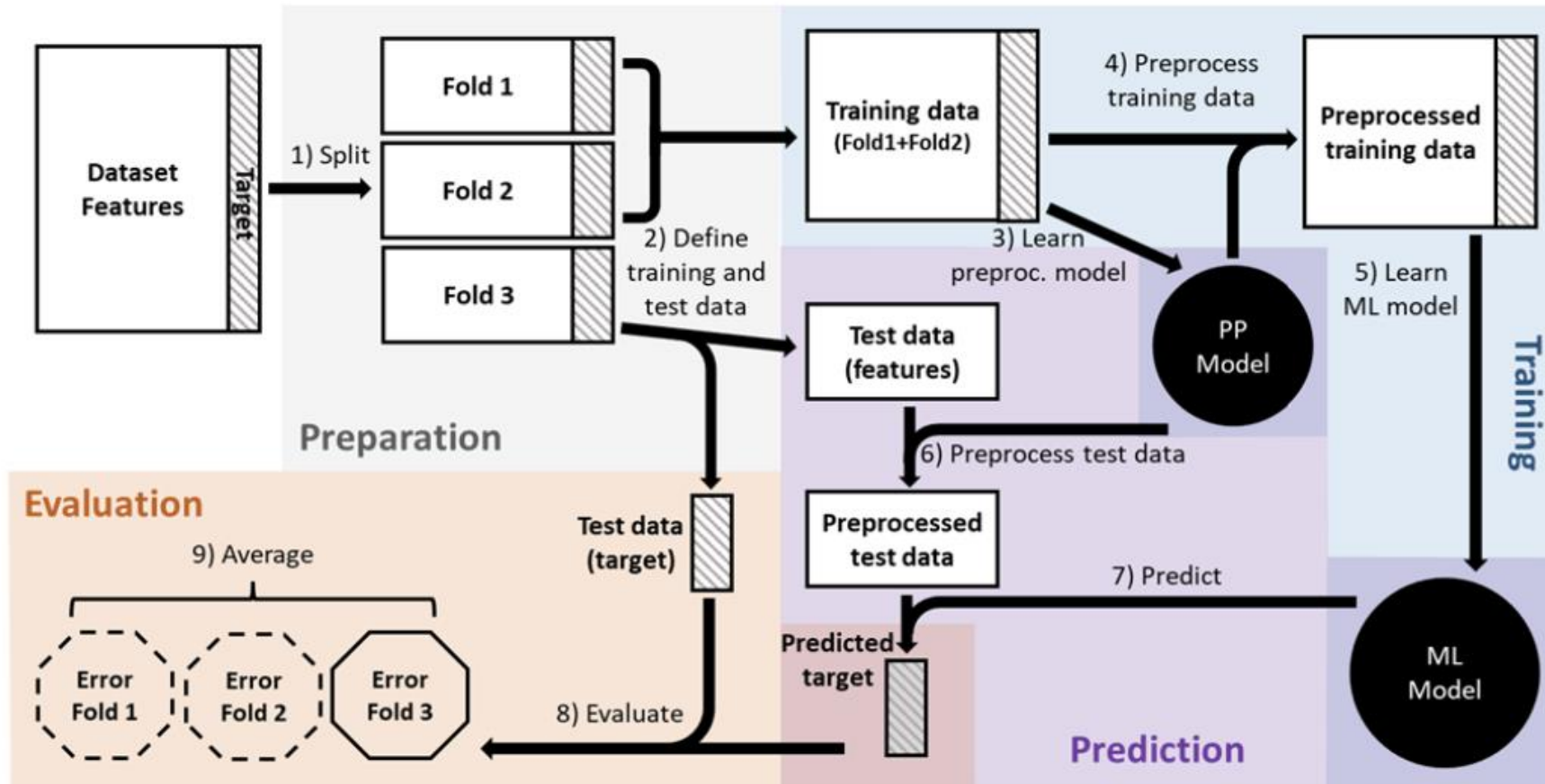
- The appropriate use of **Machine Learning (ML)** pipelines improves the generalizability of the findings.
 - **Leakage** is a threat.



- I encourage researchers to use ML and CCA (combined or not) in biological sciences.

LEAKAGE

Test-to-train leakage



Sasse, Nicolaisen-Sobesky, ... Patil, 2024 arXiv. Under review in Journal of Big Data.