Preregistration for 'More Than Just Associations: An Introduction to Causal Inference for Sport Science"

Simon Nolte

2024-04-12

This document contains the preregistration for the aforementioned project. As the project primarily involves conceptual work, there is no specific preregistration form available. Instead, a free-form approach will be adopted. It's important to note that preregistering a conceptual work presents certain challenges. The primary objective of the project is not to assess causal relationships in a practical research setting, but rather to showcase the general applicability of available methods and the necessity for the field to do so. Therefore, much of the project is exploratory in nature and cannot be preregistered in detail.

Meta Data

Title

'More Than Just Associations: An Introduction to Causal Inference for Sport Science'

Authors/Contributors

Simon Nolte[1], Robert Rein[2], Oliver Jan Quittmann[1]

[1]Institute of Movement and Neurosciences, German Sport University Cologne

[2]Institute of Exercise Training and Sport Informatics, German Sport University Cologne

Tasks and Roles (CRediT-Taxonomy)

SN: Conceptualization, Data curation, Formal analysis, Investigation, methodology, software, visualization, writing - original draft

RR: Methodology, Supervision, Writing - review & editing (paper only)

OJQ: Methodology, Supervision, Writing - review & editing (paper only)

Start Date

April 2024

Project Type

The project will be the primary author's master thesis and will result in a journal publication.

Funding

This project receives no external funding.

Background

To infer causality from data, a set of tools and methods exist, known as the statistical subfield of causal inference (Pearl et al., 2016) . The theory of causal inference has gained wide traction in many scientific fields in the recent years, to the extent that it became the core of modern data analysis textbooks in epidemiology (Lash et al., 2021) , social science (Llaudet & Imai, 2022) , and economics (Cunningham, 2021). Sport and exercise science has yet to discover causal inference.

Project Aim

This project aims to introduce the methods of causal inference to sport science. Using simulated and empirical examples, it will demonstrate how the use of graphical causal models can enhance study planning, data analysis, and result interpretation in sport and exercise research.

Methodology

Overview

The thesis and paper will feature somewhat different structures. The thesis will primarily concentrate on providing a theoretical introduction to the methods of causal inference. It will then use an existing dataset (as outlined in the following section) to illustrate these concepts. Furthermore, it will demonstrate the process of constructing a causal model through the development of a causal model for endurance performance.

On the other hand, the paper will maintain a theoretical conceptual focus and use fictional examples, potentially with simulated data.

Data Set

The data set for the thesis is from an existing investigation of physiological and performance parameters in trained runners (Quittmann et al., 2023). The project will use the main variables of the data set, namely: sex, anthropometrics (body mass, body length, body mass index), speed/times (5000-m time trial, 100-m sprint, endpoint of ramp test, onset of blood lactate), lactate data (sprint and ramp, post and delta each), gas exchange data (running economy at a fixed speed, maximum oxygen uptake, maximum respiratory exchange ratio, oxygen uptake at onset of blood lactate), maximum heart rate.

Incomplete observations will be dropped if variables with missing data are included in the model (unless imputation methods are demonstrated).

Data Analysis

All directed acyclic graphs analyzed on actual or simulated data will be analyzed using frequentist (multiple) linear models. While we generally sympathize with Bayesian data analysis methods, we feel that a frequentist approach makes the topic more accessible to most readers.

Data and Code Availability

All data and code of this project will be made freely available under open source licenses on GitHub and additionally archived using Zenodo (European Organization For Nuclear Research & OpenAIRE, 2013).

Deviations from the Preregistration

Deviations from the preregistration will be documented in a Transparent Changes document that is publicly available in the project repository. Major changes will be additionally stated in the final output documents.

References

Cunningham, S. (2021). *Causal Inference: The Mixtape*. Yale University Press.

European Organization For Nuclear Research, & OpenAIRE. (2013). *Zenodo*. https://doi.org/10.25495/7GXK-RD71

Foster, E. D., & Deardorff, A. (2017). Open Science Framework (OSF). *Journal of the Medical Library Association*, *105*(2), 203–206. https://doi.org/10.5195/jmla.2017.88

Lash, T. L., VanderWeele, T. J., Haneuse, S., & Rothman, K. (2021). *Modern Epidemiology* (4th ed.). Lippincott Williams & Wilkons.

Llaudet, E., & Imai, K. (2022). *Data Analysis for Social Science: A Friendly and Practical Introduction*. Princeton University Press.

Pearl, J., Glymour, M., & Jewell, N. P. (2016). *Causal Inference in Statistics: A Primer* (1st ed.). Wiley.

Quittmann, O. J., Foitschik, T., Vafa, R., Freitag, F. J., Sparmann, N., Nolte, S., & Abel, T. (2023). Is Maximal Lactate Accumulation Rate Promising for Improving 5000-m Prediction in Running? *International Journal of Sports Medicine*, *44*(4), 268–279. https://doi.org/10.1055/a-1958-3876