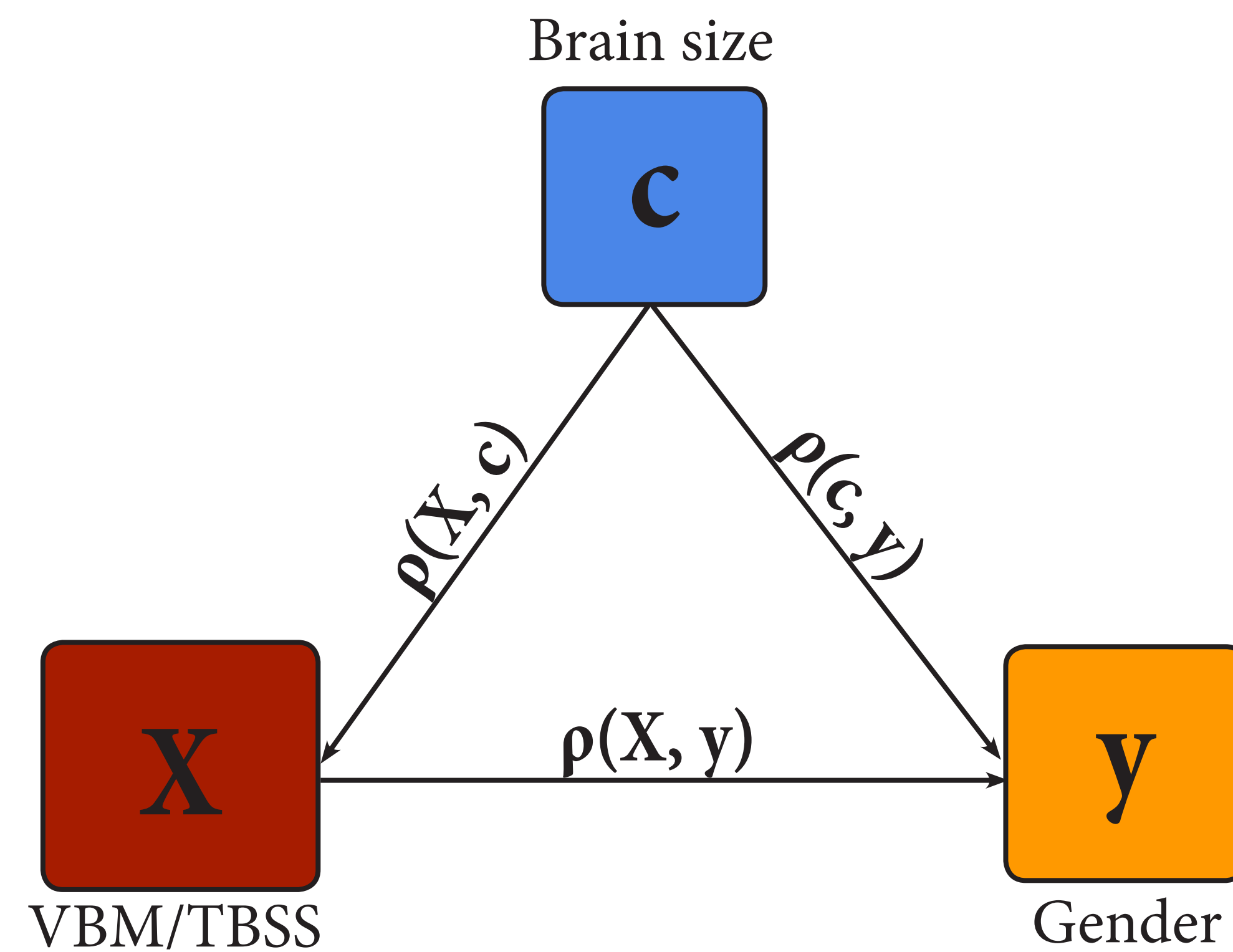


Controlling for confounds in multivoxel pattern analyses

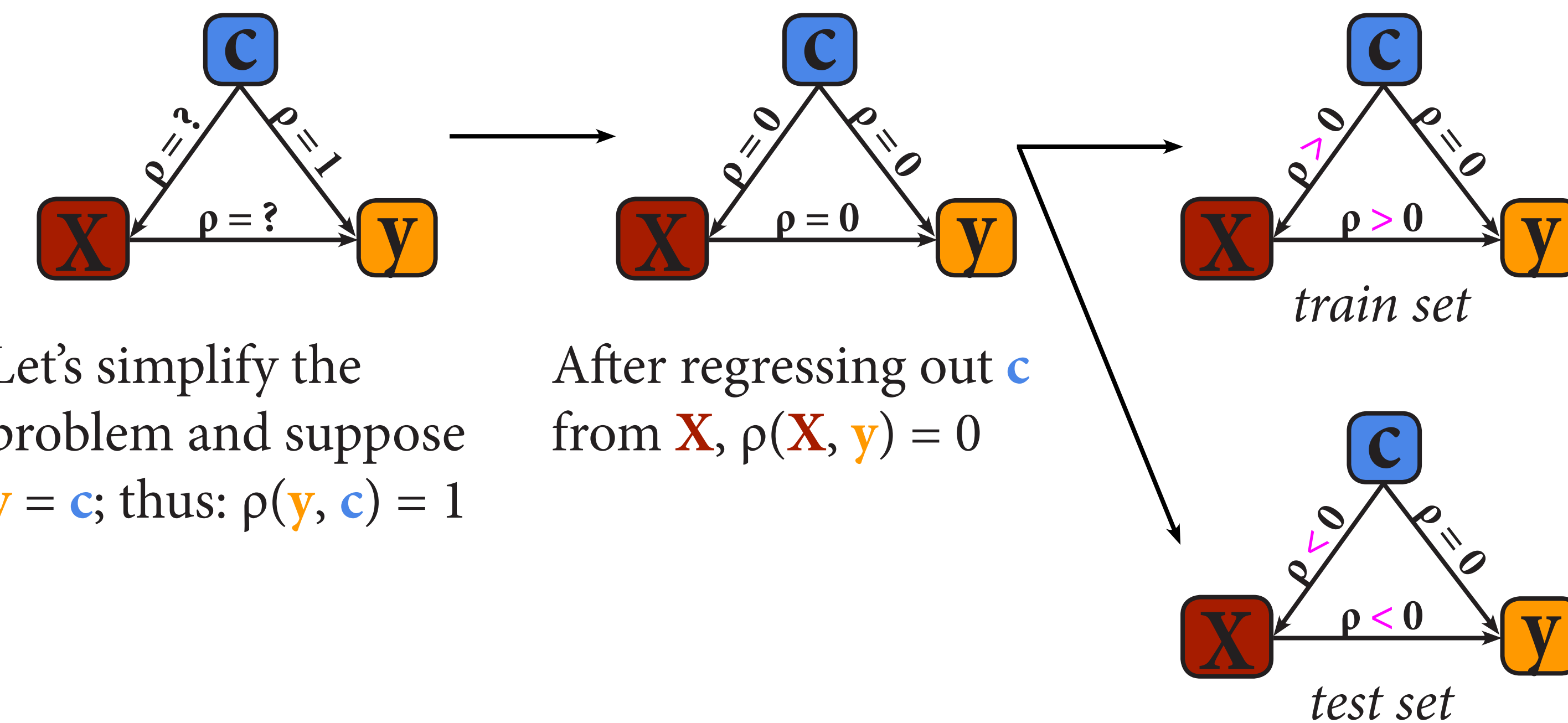
Steven Miletić, Lukas Snoek, & H. Steven Scholte
University of Amsterdam

Introduction

- MVPA has shown to be **more sensitive** than mass-univariate analysis, but arguably MVPA is also more sensitive to **confounds**¹
- Often, researchers deal with confounds (**c**) by regressing it out from the neural patterns (**X**)^{2,3}
- We show, however, that this introduces bias in cross-validated MVPA pipelines⁴, leading to **below-chance accuracy**³



What's going on?



Let's simplify the problem and suppose $y = c$; thus: $\rho(y, c) = 1$

After regressing out **c** from **X**, $\rho(X, y) = 0$

Due to random sampling, it may be that:

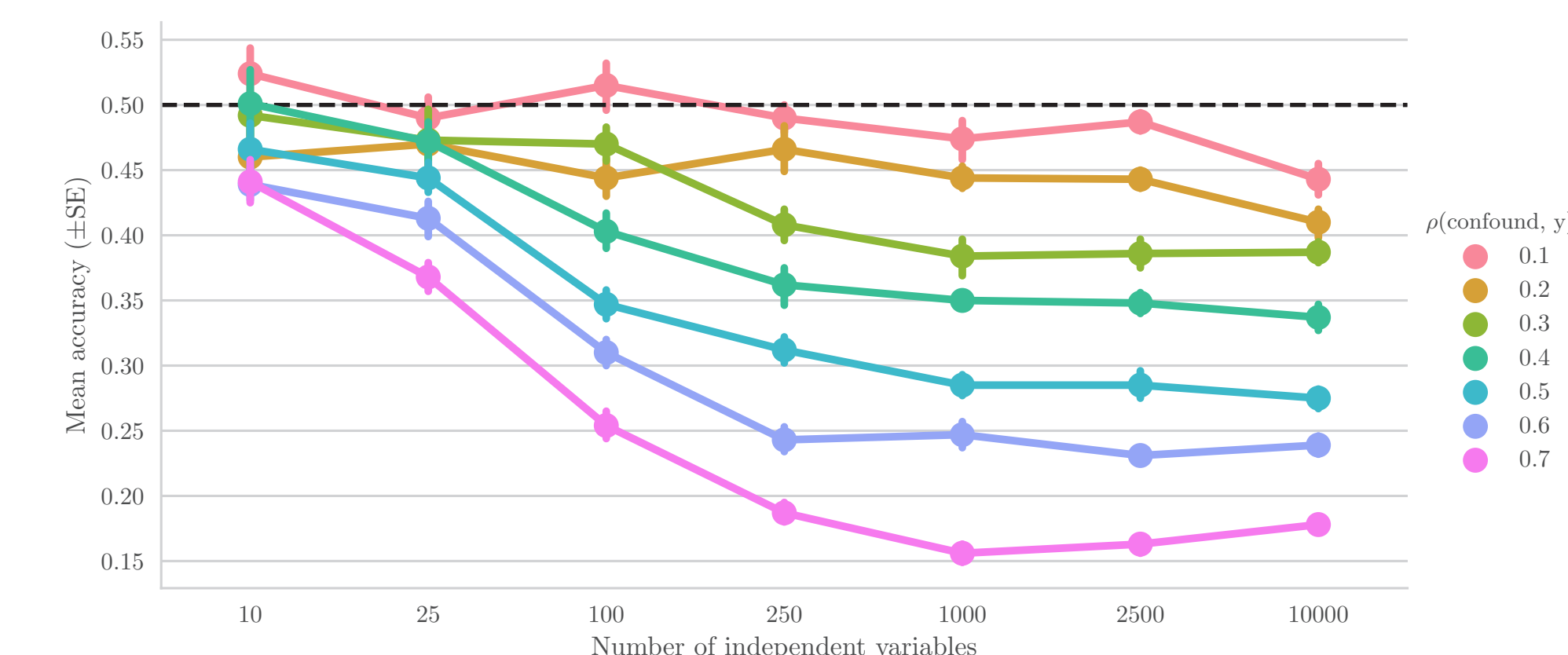
$$\rho(\mathbf{X}_{\text{train}}, \mathbf{y}_{\text{train}}) > 0$$

but, because

$$\rho(\mathbf{X}_{\text{full}}, \mathbf{y}_{\text{full}}) = 0$$

often:

$$\rho(\mathbf{X}_{\text{test}}, \mathbf{y}_{\text{test}}) < 0$$

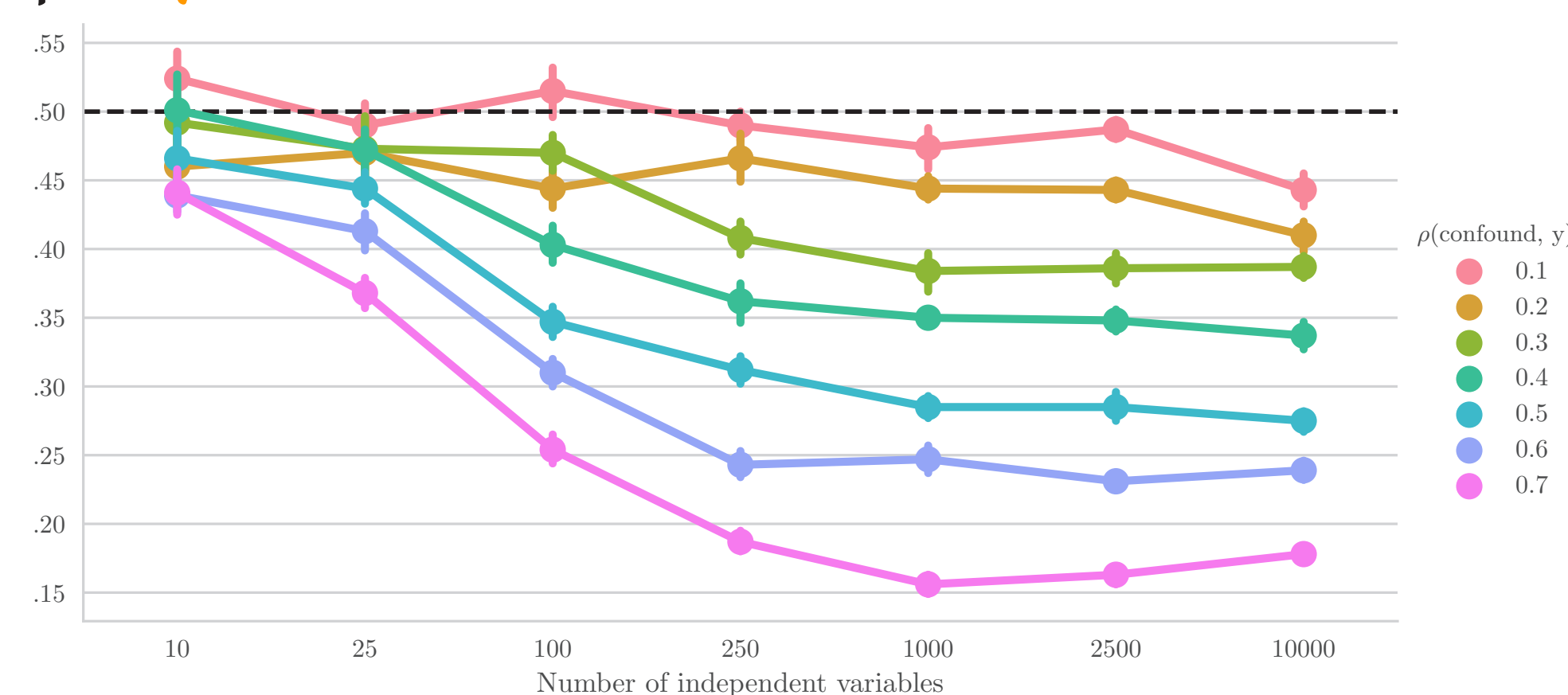


The strength of the negative bias depends on the number of independent variables, and the correlation coefficient between the confound and **y**

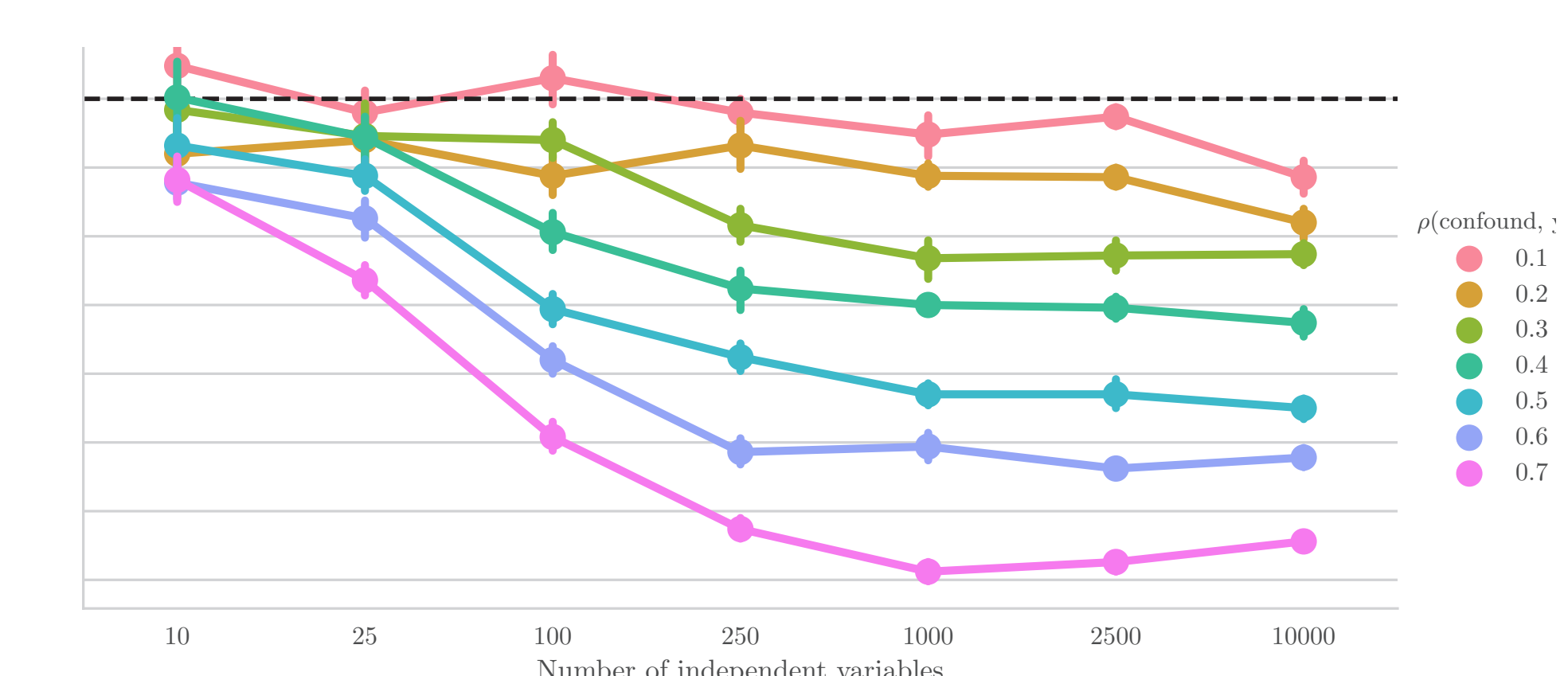
What's the problem?

Following the example to predict **gender (y)** from **VBM and TBSS-data (X)** in the face of the “confound” **brain size (c)**...

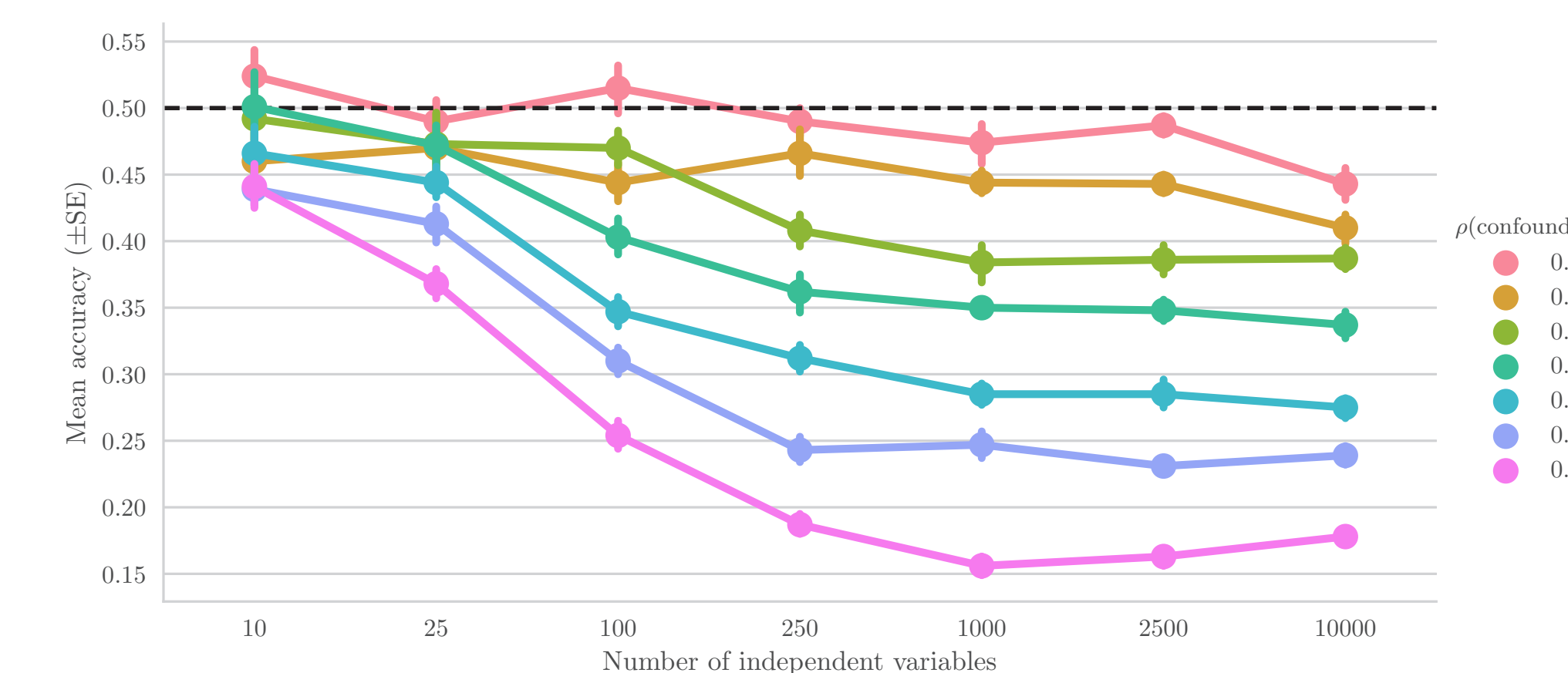
We know that brain size truly confounds $\rho(c, y) \dots$ ²



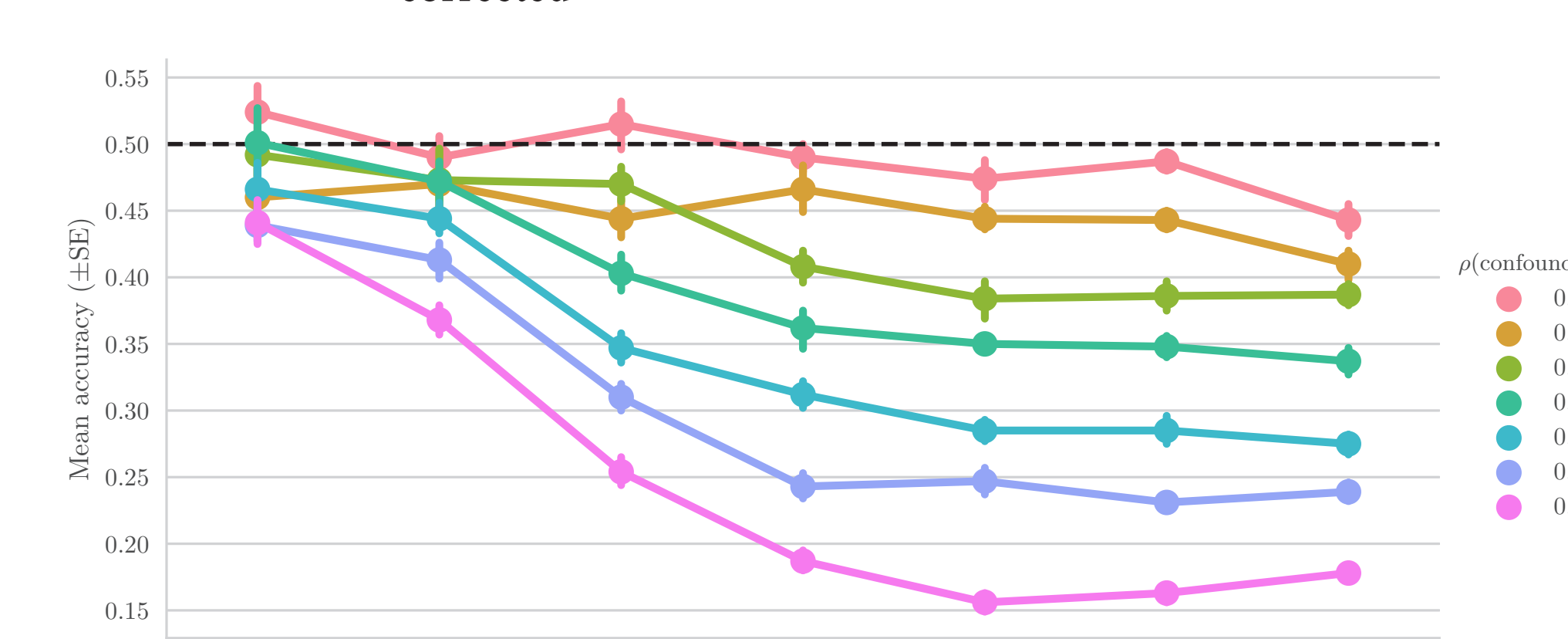
...and is related to our data...



So, $\rho(\mathbf{X}, \mathbf{y})_{\text{uncorrected}}$ is biased...

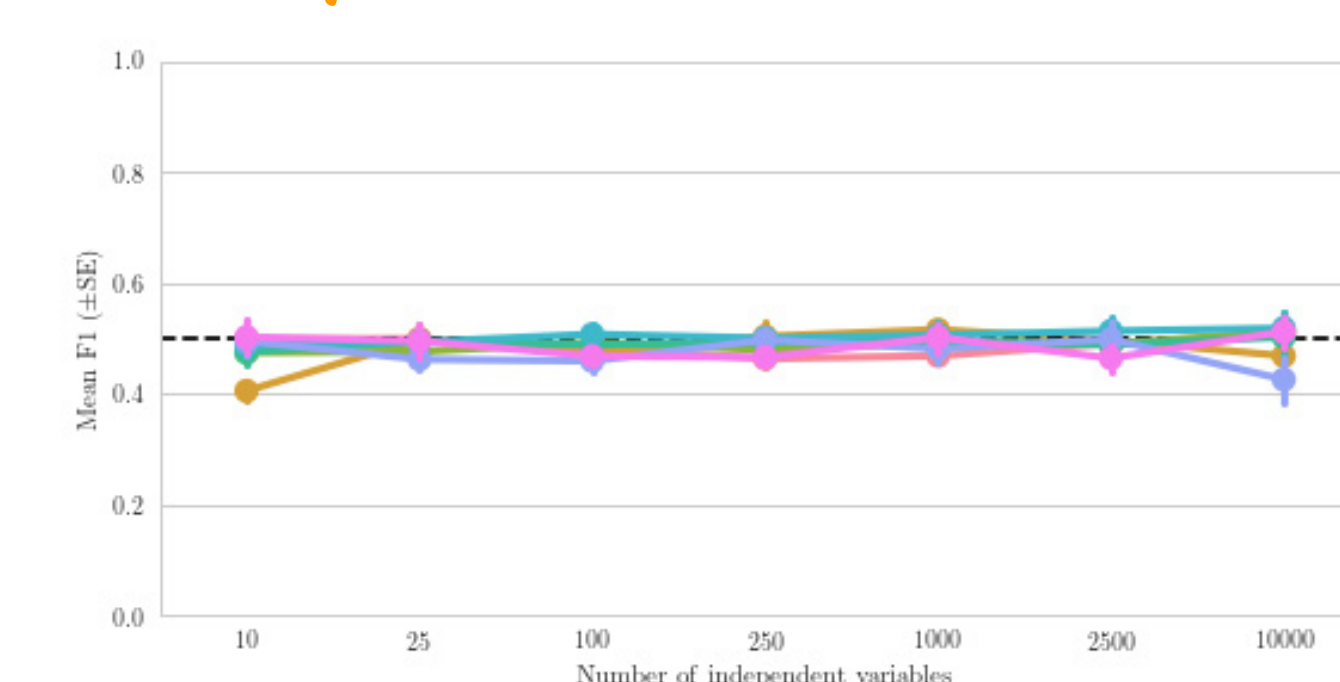


Yet, $\rho(\mathbf{X}, \mathbf{y})_{\text{corrected}} < \text{chance (50\%)}$

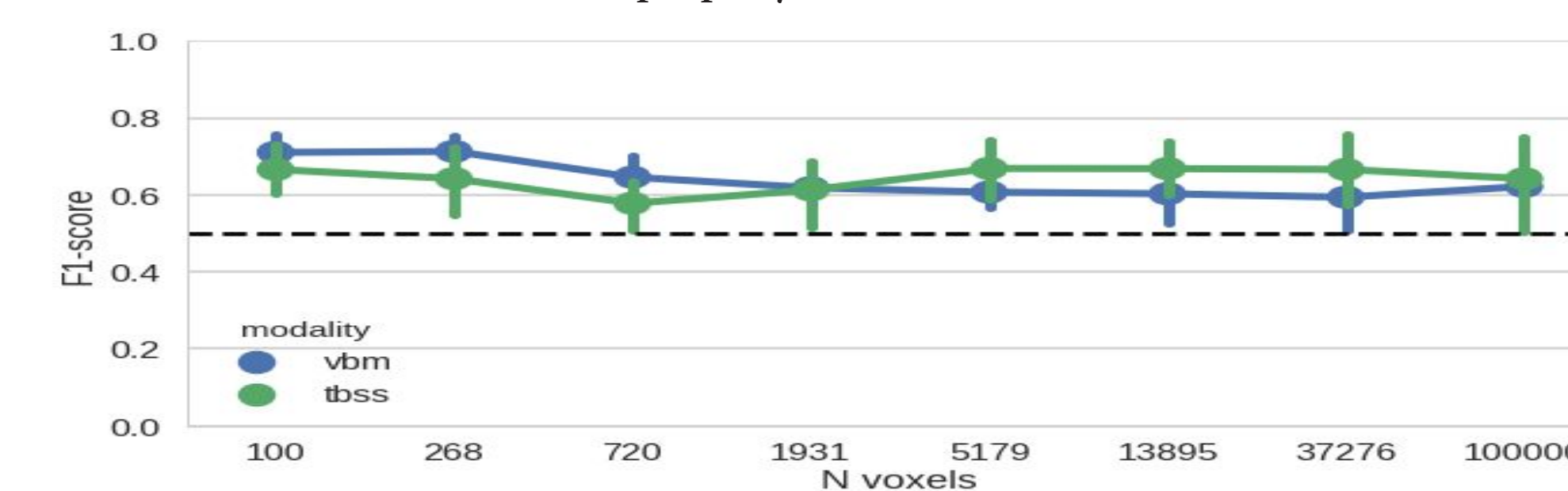


Solution

The problem can easily be solved by regressing out **c** from **X** within each fold! In simulations without a correlation (**X**, **y**)...



...and in our empirical example, where there is a relation, $\rho(\mathbf{X}, \mathbf{y})_{\text{corrected}}$ properly



Conclusion

- Confound regression introduces bias in cross-validated MVPA pipelines, especially when many voxels are used
- Like counterbalancing confounds⁴, confound regression should be done foldwise

References

- Naselaris & Kay, *TICS*, 2015
- Todd et al., *NeuroImage*, 2013
- Woolgar et al., *NeuroImage*, 2014
- Hebart & Baker, *Arxiv*, 2017
- github.com/lukassnoek/MVCA
- github.com/lukassnoek/skbold