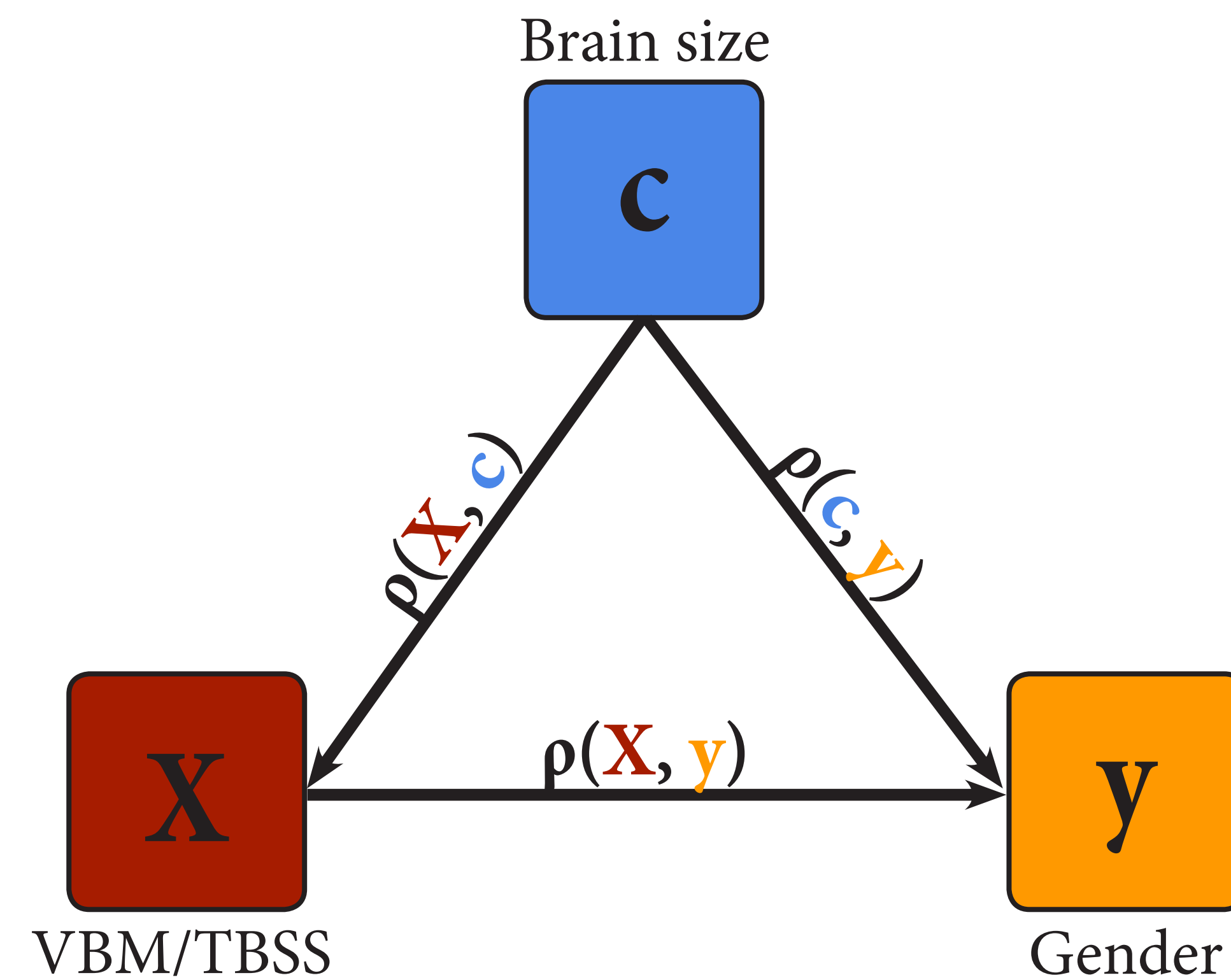




Introduction

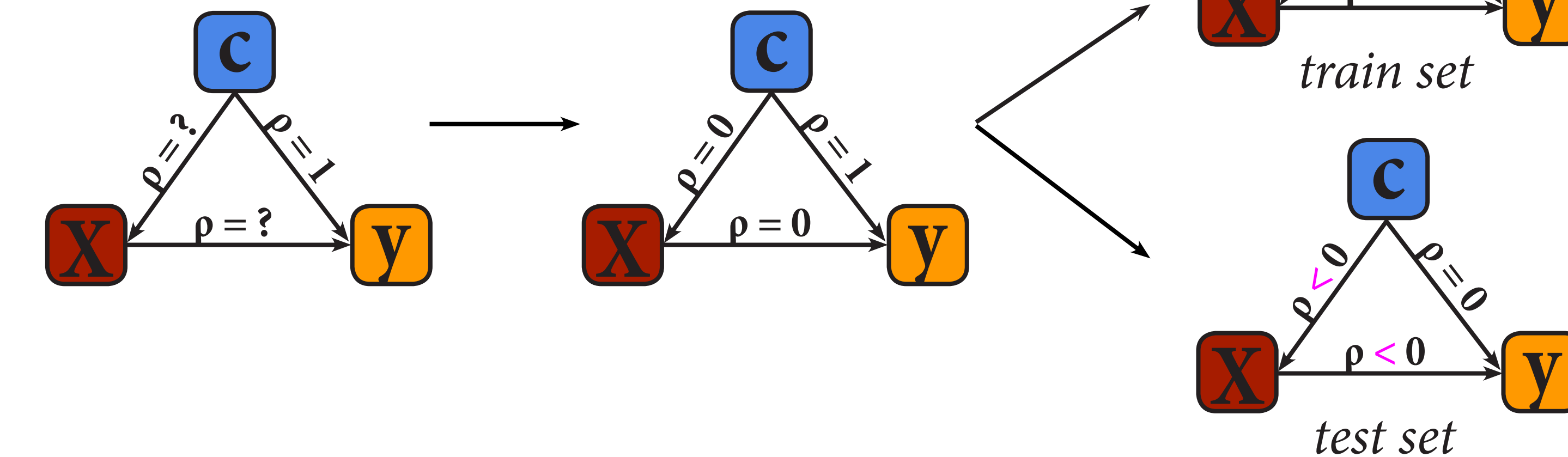
- Contrary to mass-univariate analyses, where confounds are often controlled for, it is **unclear how to handle confounds** in MVPA
- This poses a serious threat to the **generalizability of MVPA results** in both clinical and fundamental research - especially because MVPA is arguably **more sensitive** to confounds¹
- Here, we show how a previously proposed^{2,3} method of dealing with confounds (“confound regression”) leads to bias⁴, and causes **below-chance accuracy**³
- We introduce a **universal and unbiased method** of dealing with confounds in MVPA



What's going on?

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

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



Due to random sampling, it may be that:

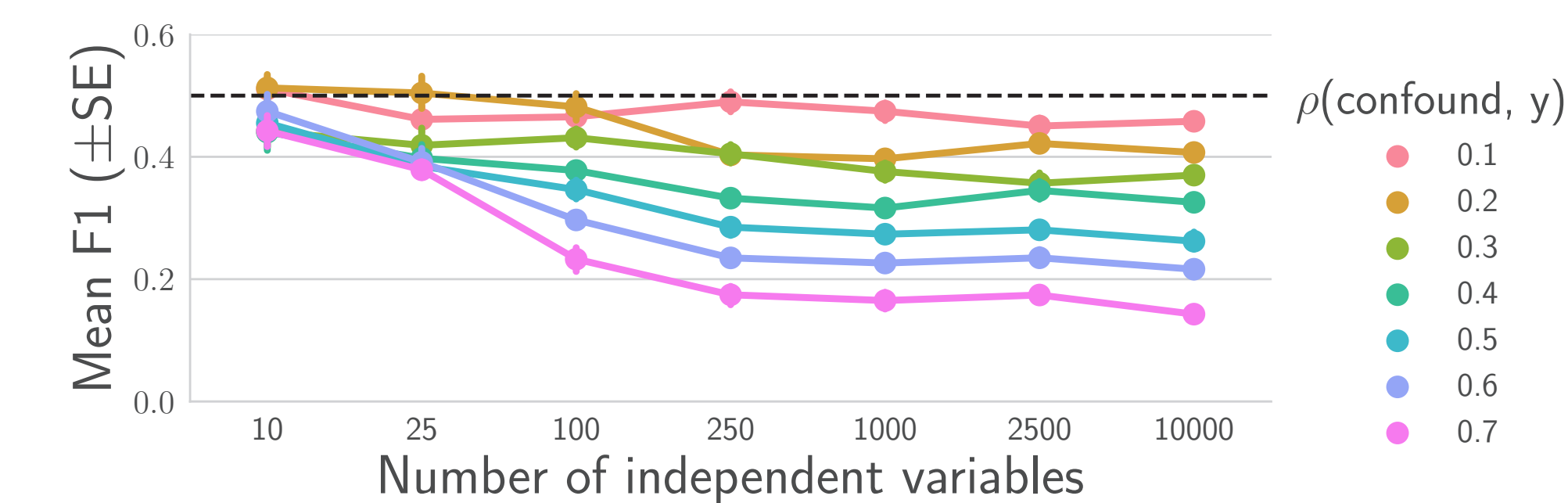
$$\rho(X_{\text{train}}, y_{\text{train}}) > 0$$

but, because

$$\rho(X_{\text{full}}, y_{\text{full}}) = 0$$

often:

$$\rho(X_{\text{test}}, y_{\text{test}}) < 0$$



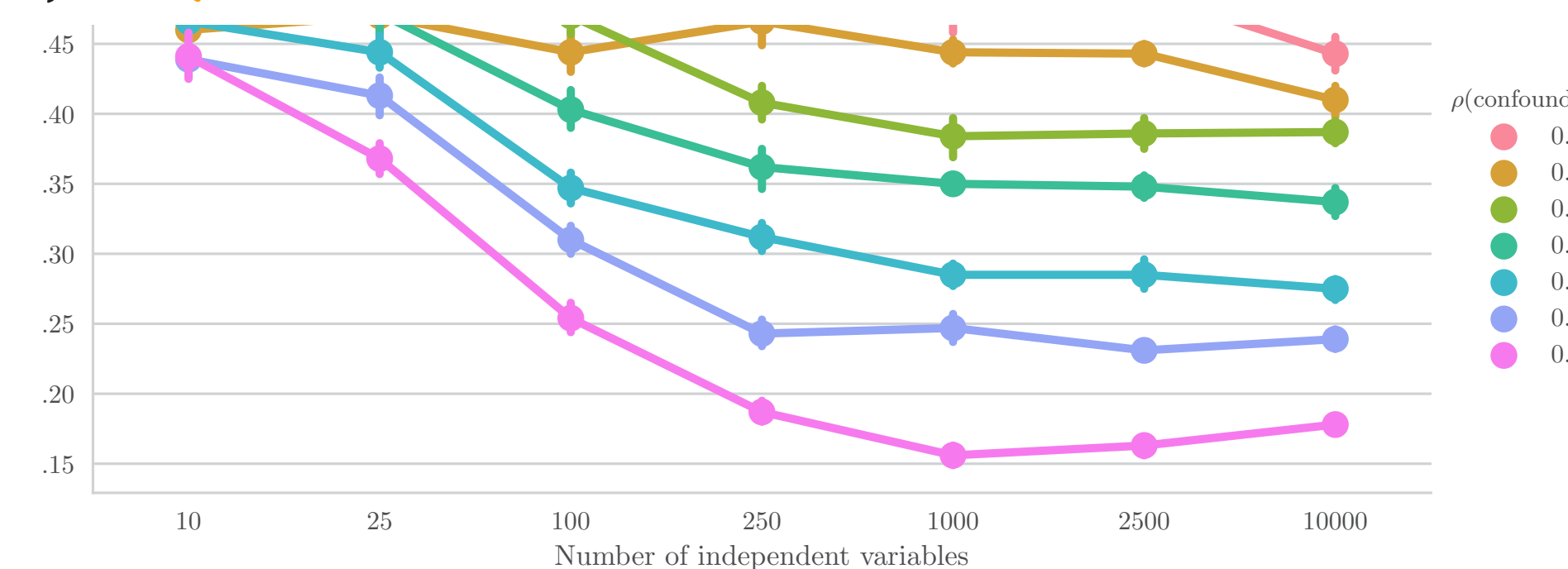
Simulations⁵ show how the strength of the bias depends on:

- the number of independent variables
- the correlation coefficient between the confound and y

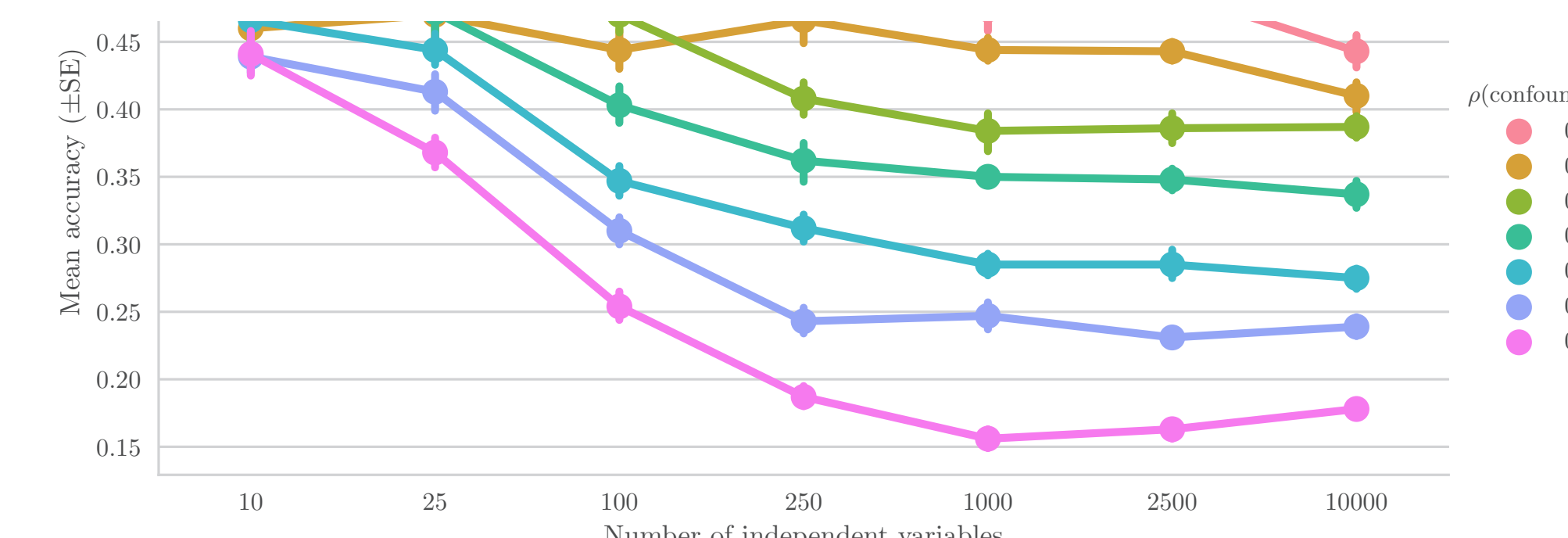
What's the problem?

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

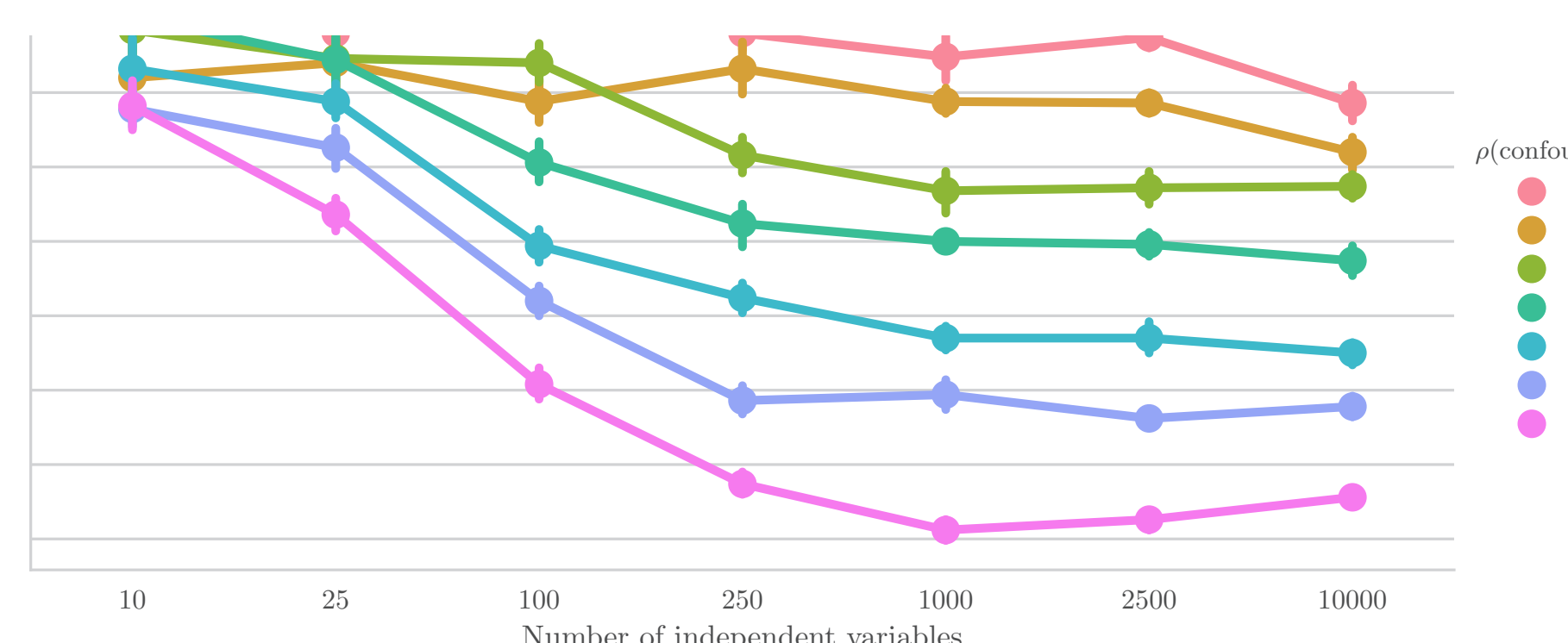
We know that brain size truly confounds $\rho(c, y) \dots^2$



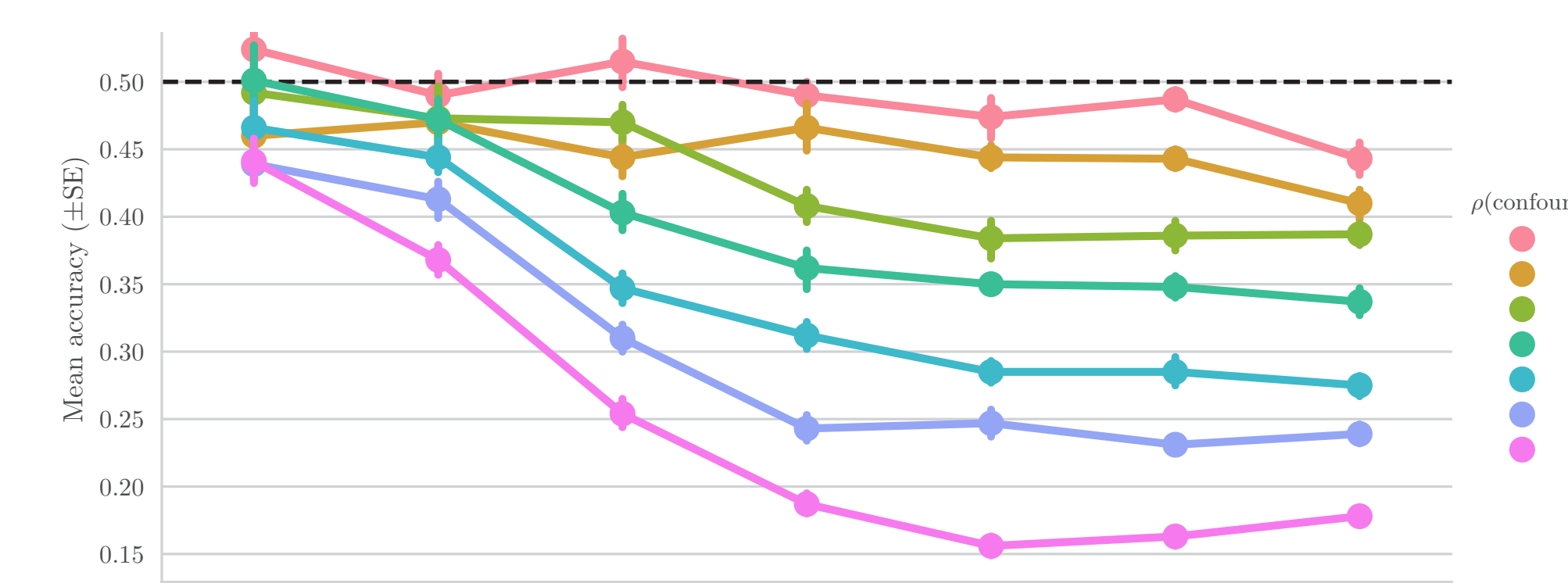
So, $\rho(X, y)_{\text{uncorrected}}$ is biased...



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

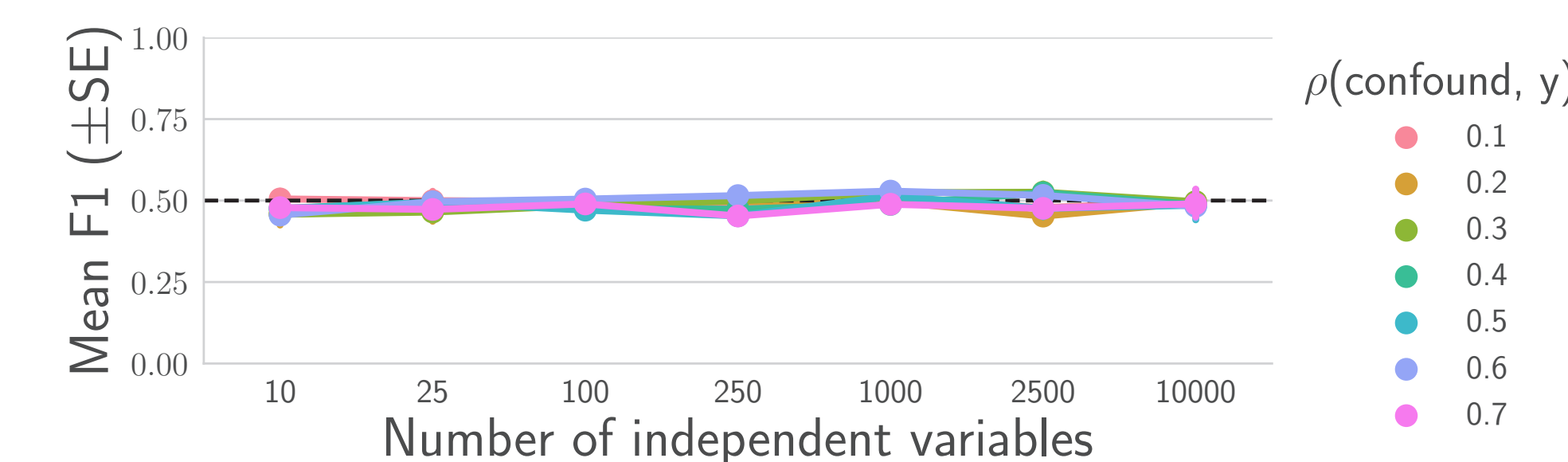


Yet, $\rho(X, y)_{\text{corrected}} < \text{chance (50\%)}$

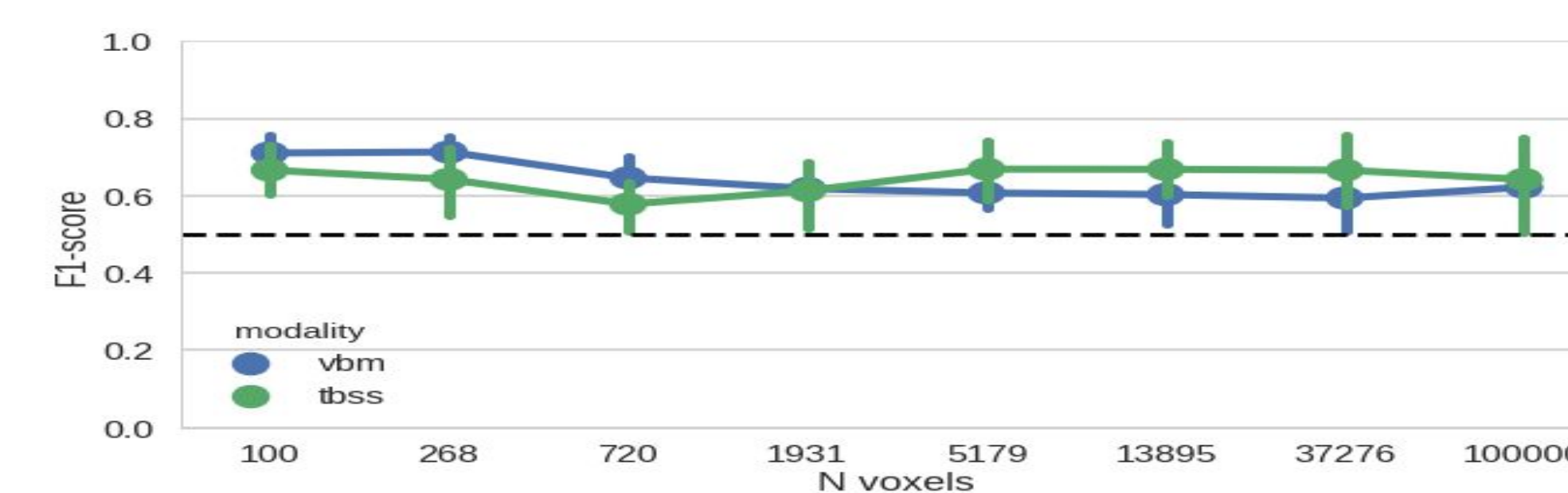


Solution

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



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



Conclusion

- Confound regression introduces bias in cross-validated MVPA pipelines, especially when many voxels are used
- Regressing out confounds foldwise is a universal and easy method, improving the generalizability of MVPA results

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