Caveats in Neuroimaging Practices: Covariate Selection and Result Reporting

Gang Chen[1], **Zhengchen Cai**[2], **Paul A. Taylor**[1]

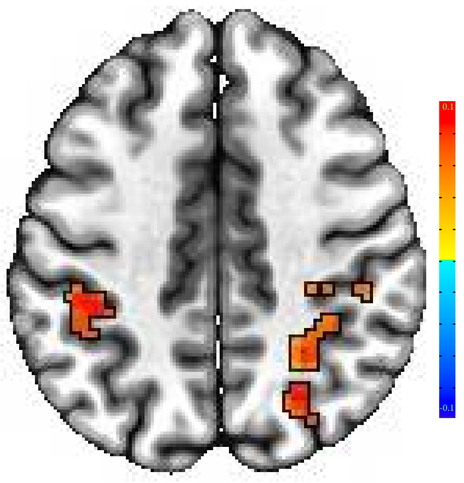
[1]Scientific and Statistical Computing Core, National Institute of Mental Health, NIH, USA

[2]The Neuro (Montreal Neurological Institute-Hospital), McGill University, Canada

Correspondence: gangchen@mail.nih.gov

Introduction

- Covariate Selection
  - Traditional justifications for selecting a covariate
    - Availability: Data for the variable were already collected
    - Prevalence: Previous studies used the variable as a covariate
    - Intention: The investigator intends to control for the variable
    - Statistics: Metrics such as p-value and R^2 are used to gauge the importance of a variable
  - ... but those all have problems (Chen et al, 2024), such as:
    - Circular Reasoning: Can a model be solely assessed by its own output?
    - Interpretation Ambiguities: predictivity $\neq$ mechanism understanding
    - Info Waste: Failing to incorporate domain knowledge about variable relationships
- Result Reporting
  - Common Practice: Mass univariate analysis
    - Controlling for Family-Wise Error: at the cluster level
  - Traditional justifications for stringent thresholding
    - Concern about multiple comparisons
    - Controlling for "false positives"



  - ... but those implementations also have problems (Chen et al, 2022):
    - Unrealistic Assumption: do effects in the brain take any values with equal likelihood?
    - Excessively Conservative: Overly penalizing due to adjustments for multiplicity
    - Artificial Dichotomy: Does statistical evidence render positive/negative dichotomy?
    - Disassociation with Neurology: Do cluster boundaries carry anatomical meaning?
    - Discrimination: Small regions are intrinsically penalized for their size
    - Ambiguity: A cluster partly straddling multiple regions
    - Info Waste: Reducing a cluster to a single peak

Motivating Example: A structural fMRI study

- Goal: Relationship between short-term memory (STM) and gray matter density (GMD)
  - Explanatory Variable: GMD
  - Response Variable: STM
- Potential Covariates
  - Sex, age, intracranial volume (ICV), APOE genotype, and body weight
- Questions to address
  - Explanatory vs Response Variable: is it justifiable to invert the roles of GMD and STM, making GMD the voxel-level response variable for easy implementation?
  - Covariates: Should all 5 covariates be included?
  - Interpretability: is it appropriate to report all parameter estimates from a single model?
  - Experimental Design: What variables could be omitted or might improve estimation?

Principles: Causality via Directed Acyclic Graphs (Pearl, 2009)

- Three Basic Types

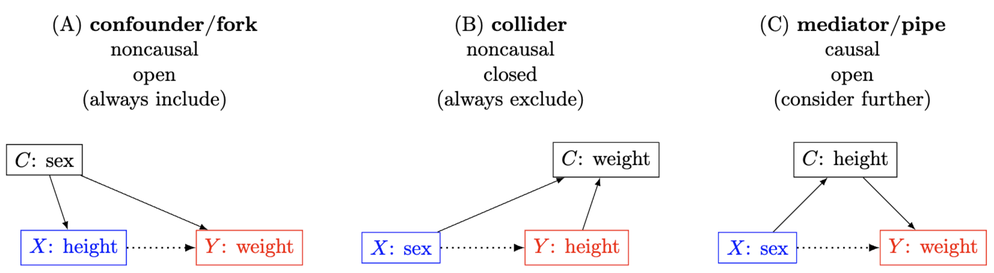


Figure 1: Three basic directed acyclic graph (DAG) types. The relationship under study is between an explanatory variable X (blue) and a response variable Y (red), as indicated by the dotted arrow. An arrow with a solid line indicates the *a priori* influence direction. A covariate C can be a confounder (A), a collider (B), or a mediator (C). The three variables of sex, height, and weight are used to illustrate the three DAG types.

  - Include confounders
  - Exclude colliders
  - Include/exclude mediators based on the focus on direct/total effects

- Four Auxiliary Types

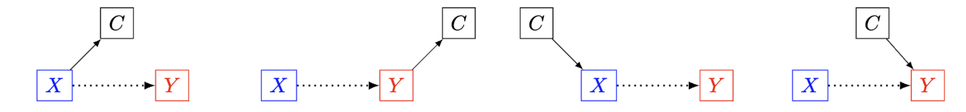


Figure 2: Covariate as a child or parent. The relationship under study is between an explanatory variable X (blue) and a response variable Y (red), as indicated by the dotted arrow. The explanatory variable X (A) or the response Y (B) may single-handedly influence the covariate C. Alternately, C may play the role of a parent, influencing X (C) or Y (D), but not both.

  - Only include the parent of the response variable for improved precision

Revisiting the Motivating Example

- Domain knowledge: Laying out relationships among variables

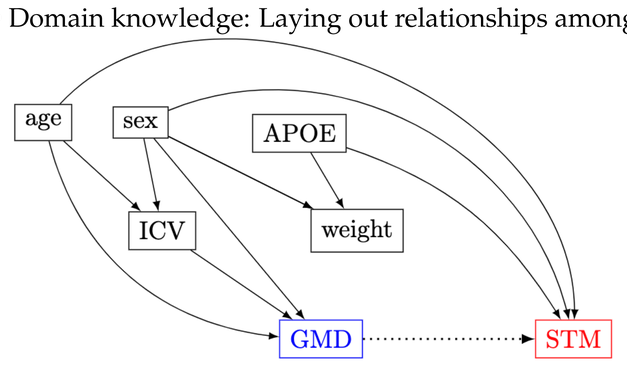


Figure 3: DAG for a structural MRI analysis. The research focus is on the association (dotted arrow) between gray matter density (GMD) (blue) and short-term memory (STM) (red). Five potential covariates are considered: sex, intrcranial volume (ICV), apolipoprotein E (APOE) genotype, weight and age. The relationships among the seven variables are based on subject knowledge.

- Addressing the Questions
  - Explanatory vs Response Variable: No, reversing the explanatory variable and the response variables leads to underestimation
- covariates
  - Sex: Confounder - should be included
  - Age: Confounder - should be included
  - APOE: Parent of response variable - inclusion improves precision
  - Weight: Collider - should be excluded (no harm when sex is included)
  - ICV: parent of explanatory variable - should be excluded
- Interpretability
  - Only report parameter estimation for the specific causal effect
  - Switching effects requires a separate model building process
- experimental design
  - Weight should be excluded from data collection
  - Additional variables like sleep hours may help

Result Reporting: Selection Bias in Neuroimaging

- Stringent Thresholding
  - Conditioning on descendants, leading to selection bias
  - Exacerbated by excessive penalty due to adjustment for multiplicity
  - Shares the same problem with the classic "double dipping" problem
  - Causes severe reproducibility problems (e.g., NARPS)

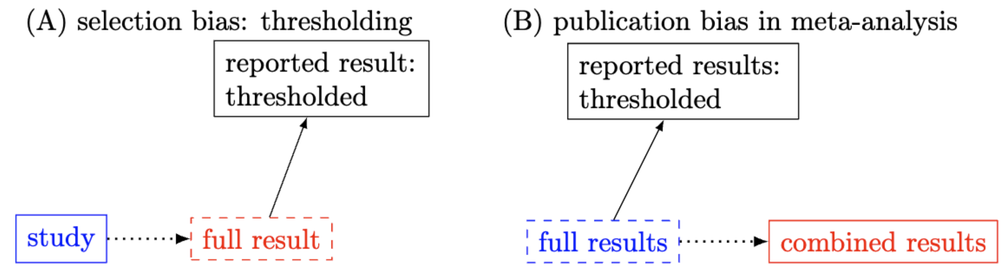
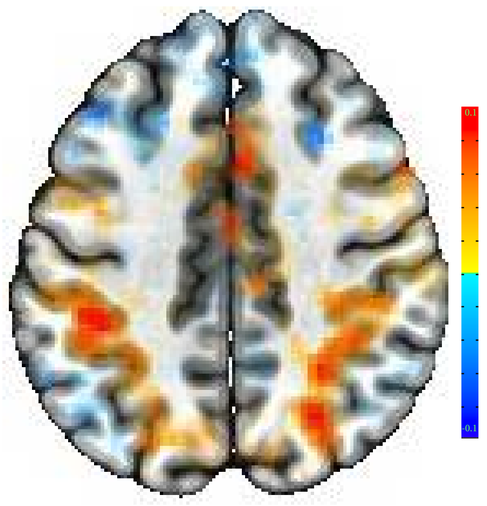


Figure 4: Two cases of selection bias in neuroimaging. (A) The reported result is a descendant of the full result, mediated by thresholding, leading to selection bias. (B) Distorted meta-analysis caused by conditioning on reported results that are descendants of the full results. Due to concerns about multiple testing, excessively stringent thresholding is typically practiced. This selection bias is further exacerbated when whole-brain results are reduced to single "peak" voxels.

- Recommended Result Reporting
  - Highlight results, but don't hide them (Taylor et al, 2023)
  - Perform analysis at region level (Chen et al, 2022)
  - Show estimated effect magnitude instead of statistic values (Chen et al., 2022)
  - Emphasize results with strong evidence
  - Maintain information continuity
  - Promote open science, accurate meta-analysis and reproducibility



Conclusions

- Covariate Selection: Should be based on variables relationships
  - Justifications in common practice: Neither sufficient nor necessary
  - Some effect estimations in the literature might be biased
- Result Reporting: Maintain information continuity
  - Mass univariate analysis causes inefficient modeling by disregarding data hierarchies
  - Stringent thresholding undermines open science, transparency and reproducibility

References

Chen, G., Cai, Z., Taylor, P.A., 2024. Through the lens of causal inference: Decisions and pitfalls of covariate selection. https://doi.org/10.1101/2024.01.11.575211

Chen, G., Taylor, P.A., Stoddard, J., Cox, R.W., Bandettini, P.A., Pessoa, L., 2022. Sources of Information Waste in Neuroimaging: Mishandling Structures, Thinking Dichotomously, and Over-Reducing Data. Aperture Neuro 2.

Pearl, J., 2009. Causal inference in statistics: An overview. Statistics Surveys 3, 96–146.

Taylor, P.A., Reynolds, R.C., Calhoun, V., Gonzalez-Castillo, J., Handwerker, D.A., Bandettini, P.A., Mejia, A.F., Chen, G., 2023. Highlight results, don't hide them: Enhance interpretation, reduce biases and improve reproducibility. NeuroImage 274, 120138.

Acknowledgments

The research was supported by the NIMH Intramural Research Programs of the NIH.