Model selection and causal effect estimation

A summary of the question

- You want to estimate causal effect of a variable X on a response variable Y, but there are "confounds" Z. E.g.:
  - X: characteristic(s) of child's linguistic input
  - Y: language development outcome (e.g., vocab test score)
  - Z: socioeconomic status of family
- You have a set of models $M_1, M_2, \ldots, M_k$ that are candidates for analyzing your dataset D
  - By assumption, D includes X, Y, Z, and maybe other info W
- Each model includes a characterization of the $Y \sim X$ relationship that is potentially scientifically interpretable
- Each model also has some predictive "score"
- How to use $\{M_i\}$ to estimate the causal effect of interest?

(e.g., Fu & Zheng 2024, https://doi.org/10.1007/s13158-024-00406-0)

Random-effects structure in this light

- Scenario: in addition to the foregoing, now our data are hierarchically organized with grouping factor F
- How should we think about random-effects specification in this light?