Defining the estimand of interest when integrating data across studies

* Underlying assumption
  + There is some true CATE that is universal across studies, and there is exactly one true Y(0) and one true Y(1) for each X
  + There is a true CATE but it depends on study membership, meaning Y(0) and Y(1) can be different for individuals with the same X but who come from different studies
    - Study is actually moderating the treatment effect; there is something at the study level that affects the way treatment impacts the outcome
  + CATE doesn’t depend on study, but CATE estimates do
    - Maybe study is a moderator, but we would want to know what about study makes it a moderator – “fake”/proxy
    - Y(0) and Y(1) don’t directly depend on study – we don’t observe some study-level things that they actually depend on
      * True Y(0) but it kind of varies based on study
      * Distributional potential outcomes?
      * We assume that study isn’t really a moderator but it’s a proxy for something else
  + Maybe think of it in terms of a random effects meta-analysis: we have a true mean and the study-specific taus vary randomly around that true mean tau
* Modeling approaches
  + If we assume that the true CATE depends on study membership, then it makes sense to have study-specific CATE functions
  + If we assume that there is a universal CATE, then we need to find a way to estimate a universal function or impute missing study variables
* Methods in simulations paper
  + Complete pooling: assumes universal CATE
  + Pooling with trial indicator: allows study to moderate the effect
  + Ensembling: allows for treatment effects to vary by study
  + Meta-analysis: assumes shared model for Y(0) but allows for study-level variance in treatment effect (??)
* Questions
  + Can we assume a universal CATE with the methods that estimate study-specific ones? And then I guess we just haven’t reached the final step yet?
  + Doesn’t it make sense for there to be some study-level variation in the treatment effect, but is that just a random error issue or is it something systematic?
  + Does study capture some unobserved variables and can it serve as sort of a proxy for that?
* Prediction intervals for non-parametric tau estimates
  + In random effects MA, we use 