Visualizing a treatment effect modifier or individual treatment responses

Suppose an expert collaborator has suggested a measured baseline covariate, M, which they suspect to be a treatment effect modifier. In specific, individuals with higher values of M are suspected to have a more positive response to treatment. How can M inform our study design?

At the end of these notes, I'll discuss a related, more theoretical concept: visualizing the *actual* individual treatment responses. In some ways, a very well articulated treatment effect modifier is just a proxy we might use to try to understand individual treatment responses. If we had the individual responses, we could visualize most of the relationships a modifier shows us, and more.

Matching on a suspected modifier

Is it important to match or sub-classify on M? The answer might depend on a few things.

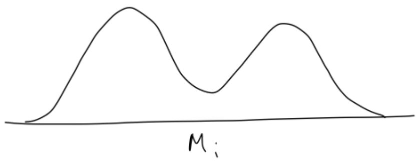
First, if the estimand of interest is the sample average treatment effect among the treated individuals, matching on M might not be as important as one might expect. A treated and control pair who are similar in terms of their prognostic score and M should have similar potential outcomes under each treatment assignment. However, if the estimand is the SATT, the unobserved Y(1) potential outcome of the control individual may not matter. Intuitively, when we estimate the SATT, we only care about the control individuals insofar as they approximate the unobserved potential outcomes of the treated individuals under the control assignment. As long as the control individuals are good propensity and prognostic score matches for the treatment group (i.e. they are close in terms of Y(0)), we may not actually care about their likely response to treatment (for more, see the commentary on matching on potential outcomes scores in the notes on visualizing the potential outcomes). Of course, this is likely to change if we're interested in a more symmetric treatment effect estimand like the ATE.

If we are interested in studying M more directly, things change dramatically. We may want to stratify by M prior to matching in order to get separate estimates of Conditional average treatment effect at different levels of M. If we were especially interested in confirming M as an effect modifier, we might even consider a factorial design, in which each treated individual is matched to 3 other observations: a control with a similar M value, a treated individual with an opposite M value, and a control individual with an opposite M value.

Regardless, M can give us some information on study design and interpretation.

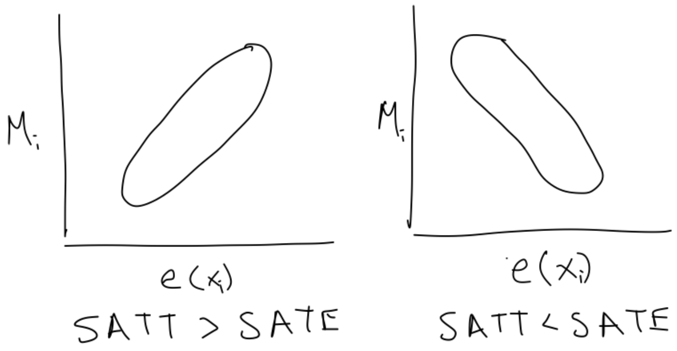
The marginal distribution of a potential modifier

A first consideration is the marginal distribution of M in the sample. Perhaps M is categorical, or it has a clear bimodal distribution (left). This might suggest a potential subgroup analysis comparing different levels of M, since individuals with high vs low values of M might have disparate responses to treatment. We might also consider matching by M or setting up a factorial design (see above). an important consideration for any of these designs is whether there are sufficient treated and control individuals at each level of M you are interested in studying.

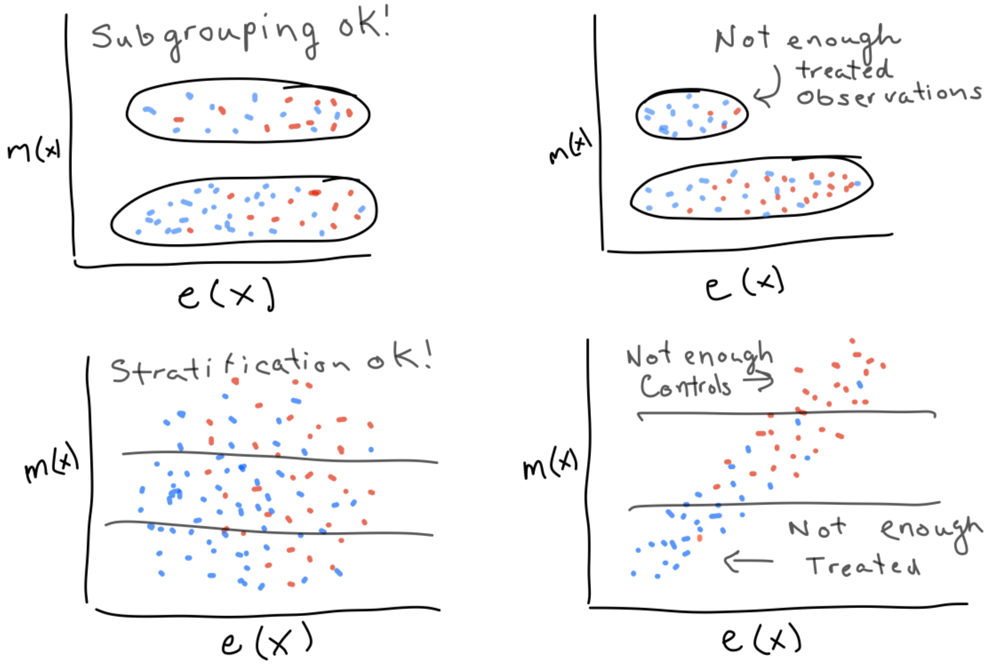


Modifier vs propensity score.

Even if we are not interested in effect modification directly, the relationship between the potential modifier and the propensity score can suggest how well the treatment effect may generalize.



On the left, above, treated individuals tend to have higher levels of the effect modifier. This means that the treated individuals are likely to have a better-than-average response to treatment, so the SATT is likely to be higher than the SATE. On the right, the opposite is true. This can suggest how well an estimate of SATT might generalize to other populations.



As alluded to above, the requirements for our data increase if we want to perform a subgroup analysis or estimate the CATE stratified by M, since each stratum must have sufficient treatment and control observations with sufficient overlap. The next drawing illustrates some scenarios in which stratifying or performing a subgroup analysis based on a modifier could lead to trouble. In the plots on the right, there are insufficient treated or control individuals in one or more strata, meaning estimating the treatment effect in these strata may be difficult. In fact, a researcher might question whether the treatment effect for these particular groups can be estimated at all. Regardless of the stratification scheme.

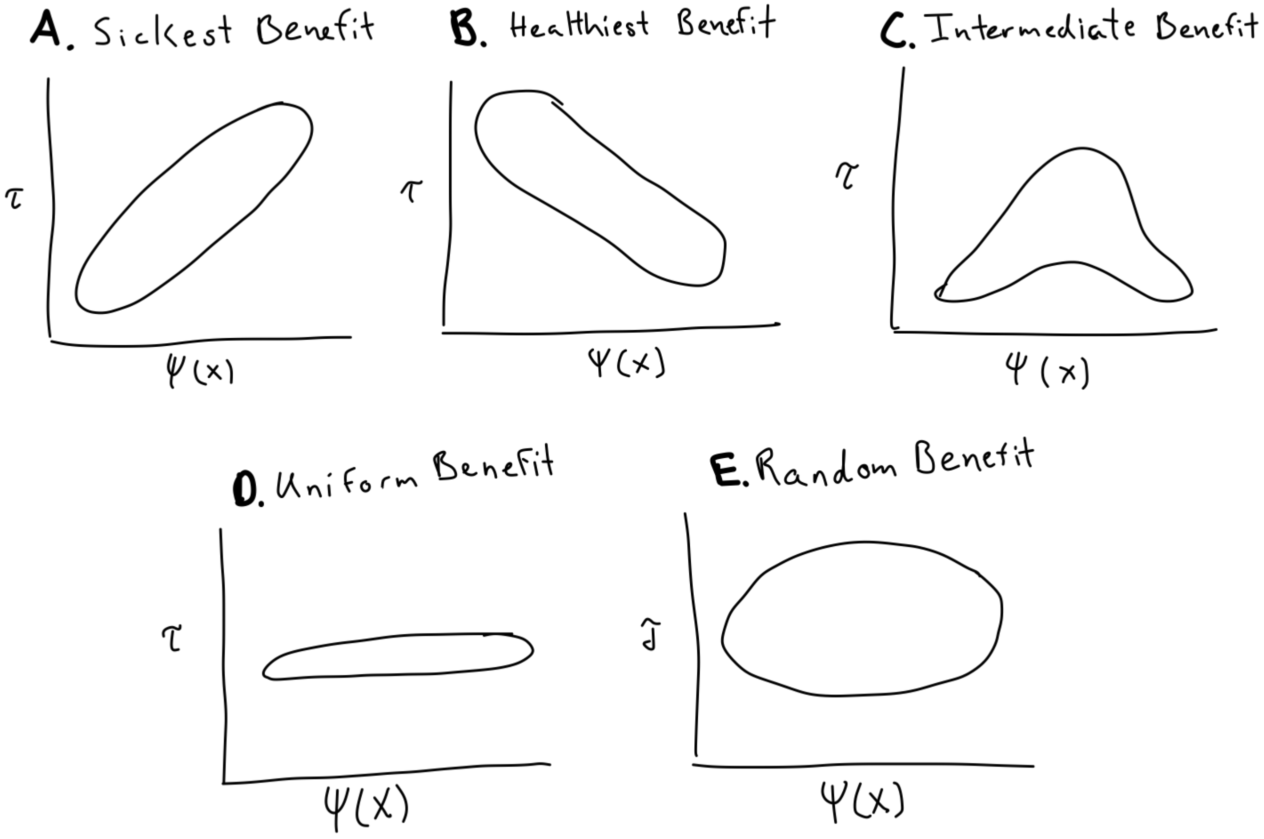
Modifier vs Prognostic Score.

Plotting the modifier vs the prognostic score can suggest something about the relationship between outcomes and response to treatment. If the prognostic score and the suspected modifier are highly correlated, that suggests whether “sicker” or “healthier” people are more likely to respond well to treatment.

This starts to get into some discussion about treatment allocation policies. This is a little easier to map out if we consider the actual individual treatment responses, rather than an effect modifier (see next page)

A more theoretical example: Individual treatment effects vs prognostic score.

A more theoretical discussion along these same lines might consider visualizing the true individual treatment effects versus the propensity and prognostic scores. Some interesting observations can be made when considering treatment effect versus prognosis. In particular, it is fairly common to assume in clinical risk modeling that the patients with the highest risk will benefit the most from treatment. The sketches below depict that assumed scenario alongside other possibilities. This is important because understanding the relationship between prognosis and treatment response explains where these high-risk-targeting programs are likely to succeed, and where they fail.



One immediate observation is that using prognostic score as a proxy for treatment response is only appropriate when treatment response and prognosis *are positively correlated*. In cases where the treatment response and prognosis are differently related (B and C) a policy which systematically treats the riskiest patients may may be ineffective at best, and at worst it may systematically deny treatment to those who stand to benefit most.

It is also important to note that treatment effect heterogeneity and prognosis need not be related at all (D and E). In cases where treatment response is mostly homogeneous, the question of allocating limited treatment resources may depend most on the research context and ethical considerations. If treatment response is highly variable but not correlated with prognosis, this suggests that some other characteristic (unrelated to the potential control outcome) may be driving treatment effect heterogeneity.