

Estimating program effects for subgroups is hard. Estimating effects for types of people who exist in theory, but whom we can't always identify in practice (i.e., latent subgroups) is harder. These challenges arise often, with noncompliance being a primary example. Another is estimating effects on groups defined by "counterfactual experience," i.e., by what opportunities would have been available absent treatment access. This paper tackles this difficult problem. We find that if one can predict, with some accuracy, latent subgroup membership, then bounding is a nice evaluation approach, relying on weak assumptions. This is in contrast to many alternatives that are tricky, often unstable, and/or rely on heroic assumptions.

What are latent subgroups again? Let's make things concrete. Early College High Schools expose students to college courses as they mature, thus easing transition to college itself. A lottery evaluation (Edmunds et al., 2017) found positive impacts on keeping students "on track" to graduate in the 9th grade. But how do these impacts vary? That is the question of this paper. We investigated whether ECHSs were differently effective for students who, absent ECHS availability, would go to a "low quality" (as estimated by the state) high school as compared to those who would go to a high quality one. This seems a simple problem: we have two groups of students, and we want to estimate impacts within each group. The problem is we do not necessarily *know* which group a student given an ECHS spot is in (thankfully, we do know which group students who are *not* given an ECHS spot is in). This is the "latent" part: we must infer group membership for those given an ECHS spot.

What kind of statistics is this? This paper is working in the realm of *principal stratification*, a "Tool for Understanding Variation in Program Effects Across [Latent] Subgroups" (Page et al., 2015). Principal stratification is a clear framework for understanding *what* one wants to estimate; however, it also makes it clear how hard it is to do so. Bounding is one avenue. With bounding, we examine the possible tradeoff of impacts between the two groups: if one group has a large impact, the other must have a small one, because we know the subgroup effects must average out to the overall average impact estimate.

Does bounding work? Usually bounds are overly wide to be useful. But if we cut up our sample into homogenous groups of students (i.e. *post stratify* it), we can substantially tighten them. Table 1, below, shows the benefits: bounding with covariates predicting where a child would go and how a child will do can reduce bound width by more than half. Bounding is a viable way of doing principal stratification without massive assumptions, at least in some contexts.

What about the ECHSs? Substantively, the paper shows that for those with access to a low-quality alternative, ECHSs were very helpful for keeping kids on track, with an estimated impact of 5 to 12 percentage points. For those with access to high-quality alternatives, the story is uncertain: estimated impacts could be anywhere from a harm of 13 percentage points to a benefit of 5 percentage points. Overall, this suggests that where we build programs like this matters.

Table 1. ECHS treatment effect bounds with and without covariate stratification

Stratification	ITT Bounds		Bounds Width	% Reduce	Adj. Bounds		Final Width
	LB	UB			LB	UB	
<i>High Quality Compliers</i>							
No Covariates	-34.1	1.4	35.5		-43.4	3.8	47.2
Covariates	- 8.4	2.5	10.9	69%	-13.2	5.1	18.4
<i>Low Quality Compliers</i>							
No Covariates	8.1	13.7	5.6		5.7	15.3	9.5
Covariates	7.4	9.7	2.3	60%	5.4	12.1	6.7

Note: Outcomes in percentage points. "% reduce" shows percent reduction due to using covariates. Final columns display bootstrap adjusted bounds and widths that takes into account estimation uncertainty.